CAPRI model documentation

Version: 18/01/2022

generated from https://www.capri-model.org/dokuwiki_help

Contents

Chapter 1

Introduction

1.1 Structure of the documentation

The documentation is structured as follows. Sections 1.2- 1.3 give an overview of capri system and its main software, gams. section 1.4 informs about the capri network. sections 1.5 and 1.6 describe historical development of the model and more recent examples of capri studies. chapter 2 provides with system requirements and the main installation instructions.

The rest of the document largely follows the workflow of the model: the different steps of building up the national, regional and global data base provide the foundations on which the system rests (3). subsequently the procedure needed to establish a baseline (chapter 4) is discussed. chapter 5 deals with the scenario impact analysis, giving descriptions for the regional supply models as well as for the global market model and their interactions in scenario runs. chapter 6 covers some elements of post model analysis, whereas chapter 7 covers options for spatial downscaling of the nuts2 results. at the very end (chapter 8), some developer tools for stability analysis are described.

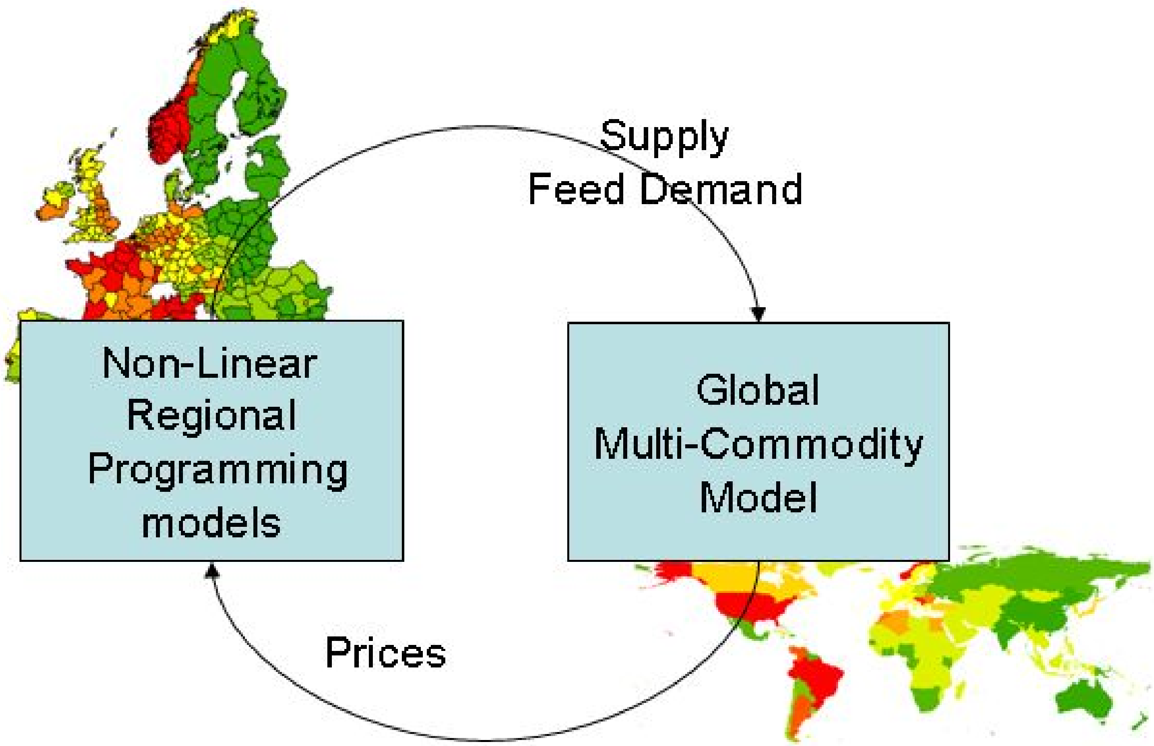
1.2 What is CAPRI

The Common Agricultural Policy Regional Impact (CAPRI) model is a global partial equilibrium model for the agricultural sector, with a focus on the European Union. It has been designed for ex-ante impact assessment of agricultural, environmental and trade policies. It has a supply module covering the EU and some auxiliary European countries[1] (regional programming models for about 280 European regions, detailed coverage of agricultural policies), embedded in a market module also covering regions in the rest of the world (global market model representing bilateral trade between 44 trade regions[2]). Thus it has global coverage but ignores potential interactions with non-agricultural sectors, except for land use.

The CAPRI modelling system itself consists of specific data bases, a methodology, its software implementation and the researchers involved in their development, maintenance and applications.

[1] EU-28 plus Turkey, Norway and the Western Balkan countries.
[2] Countries or country aggregates.

Figure 1: General structure of the CAPRI model



The data bases exploit wherever possible *well-documented, official and harmonised data sources,* especially data from EUROSTAT, FAOSTAT, OECD and extractions from the Farm Accounting Data Network (FADN)[3] . Specific modules ensure that the data used in CAPRI are mutually compatible and complete in time and space. They cover about 65 agricultural primary and processed products for the EU (see the Annex), from farm type to global scale including input and output coefficients.

The economic model builds on a *philosophy of model templates* which are structurally identical so that instances for products and regions are generated by populating the template with specific parameter sets. This approach ensures comparability of results across products, activities and regions, allows for low cost system maintenance and enables its integration within a larger modelling network such as SEAMLESS or the DG Clima modelling suite. At the same time, the approach opens up the chance for complementary approaches at different levels, which may shed light on different aspects not covered by CAPRI or help to learn about possible aggregation errors in CAPRI.

CAPRI is designed for scenario analysis. It is a comparative static model, which technically means that

[3]FADN data are used in the context of so called study contracts with DG AGRI, which define explicitly the scope for which the data can be used, who has access to the data and ensure the data are destroyed after the lifetime of the contract.

the market equilibrium simulated for a given point in time does not involve lags or leads of endogenous variables. If several points in time are simulated, these simulatons may be perfomed therefore in any order or in parallel[4]. Comparative static results are best interpreted as the long run outcome of some scenario, after all adjustments to the new equilibrium are completed. By contrast, dynamic or recursive dynamic models also trace the adjustment path over time, while considering lagged relationships that are ususally critical in adjustment processes. CAPRI simulations start from a so-called baseline, which is a special applicaiton of the model as discussed in a separate chapter of this documention. The CAPRI baseline integrates projections from external sources, typically the Agricultural Outlook published annually by the European Commission's Directorate-General for Agriculture and Rural Development (DG-AGRI) (European Commission 2017). The parameters describing the reactions in the sector are calibrated to the baseline scenario, making the model behave in accordance with the data and projections. The baseline mirrors the projected agricultural situation up to some point in time and usually assumes status quo policy assumptions (currently: CAP 2014-2020). In a simulated scenario, all conditions in the baseline are maintained – except for the changes to be analysed.

CAPRI contains two modules, market and supply, which interact (see Figure 1).

The *supply module* consists of independent aggregate non linear programming models representing activities of all farmers at regional or farm type level captured by the Economic Accounts for Agriculture (EAA). The models optimize regional agricultural income, given the prices for inputs and outputs, subsidy levels and other policy measures. These models are a kind of hybrid approach, as they combine a Leontief-technology for variable costs covering a low and high yield variant[5] with a non linear cost function which captures the effects of labour and capital on farmers' decisions. The non linear cost function allows for perfect calibration of the models and a smooth simulation response rooted in observed behaviour (see also Jansson and Heckelei 2011).

Around 55 agricultural inputs produced in about 60 activities are covered in the supply module. The activities include inputs to crop and livestock production from other sectors and intermediate inputs produced by the farms such as feed and young animals. The models capture in high detail the premiums paid under CAP, include NPK balances and a module with feeding activities covering nutrient requirements of animals.

Main constraints outside the feed block are arable and grassland − which are treated as imperfect substitutes -, and potential policy restrictions (set-aside obligations, milk and sugar quotas, environmental constraints). Prices are exogenous in the supply module and provided by the market module. Grass, silage and manure are assumed to be non tradable and receive internal prices based on their substitution value and opportunity costs. A land supply curve renders agriculture responsive to returns to land. Non-agricultural areas respond in line with a given total region area, giving rise to land transitions.

Market equilibria are calculated by iterations between the supply module and the market module.

The market module for marketable agricultural outputs is a *spatial, non-stochastic global multi-commodity* model for about 65 primary and processed agricultural products. About 80 world regions are modelled,

[4]This does not hold if land use transitions are simulated for environmental indicators but in a "basic" CAPRI run, these may be switched off.

[5]The two technological alternatives (for most activities), representing "high yield, high input" and "low yield, low input" technologies together define the average technology for the different production activities. Intensification or extensification in response to changing conditions is modelled in simulations partly by changing shares of these technologes and partly by other mechanisms like directly price dependent yields.

but aggregated to about 40 trade regions that trade bilaterally with each other, with the possibility of simultaneous import and export. It simulates supply, demand, and price changes in global markets considering international trade.

Agricultural supply is modelled in a simpler way than in the supply module, with behavioural functions for supply and feed demand. These are supplemented with other functions for processing, biofuel use, and human consumption. These functions apply flexible functional forms where calibration algorithms ensure full compliance with micro economic theory including curvature. The parameters are synthetic, i.e. to a large extent taken from the literature and other modelling systems.Consumers and traders are represented by economic agents that follow neo-classical micro-economic theory regarding behaviour, which makes it possible to compute welfare effects. Bi lateral trade flows and attached prices are modelled based on the Armington assumptions (Armington 1969). Policy instruments cover (bi lateral) tariffs, the Tariff Rate Quota (TRQ) mechanism and, for the EU, intervention stocks and subsidized exports. This market module delivers prices used in the supply module and allows for market analysis at global, EU and national scale.

As the supply models are solved independently at fixed prices, *the link between the supply and market modules* is based on an iterative procedure. After each iteration, during which the supply module works with fixed prices, the constant terms of the behavioural functions for supply and feed demand are calibrated to the results of the regional aggregate programming models aggregated to Member State level. Solving the market modules then delivers new prices. A weighted average of the prices from past iterations then defines the prices used in the next iteration of the supply module. Equally, in between iterations, CAP premiums are re calculated to ensure compliance with national ceilings and crop yields may respond to changing market prices.

Environmental indicators, primarily for nutrient surpluses and greenhouse gas (GHG) emissions, are calculated in CAPRI and may be directly addressed in some scenarios. Regarding nutrient surpluses, the supply module contains nutrient balance equations for nitrogen, phosphorous and potassium. It considers nutrient uptake by crops following a crop growth function, and supply of nutrients from mineral fertilizer, manure, crop residues, and, for nitrogen, atmospheric deposition and fixation. The balances also contain factors for over-fertilization, loss rates, and nutrient availability per source. From those balances nutrient surpluses can be calculated per region of the supply model. Technical information from the supply module is used to compute greenhouse gas emissions, based on IPCC methodology[6]. Globally, GHG emissions are computed based on estimated emission intensities per ton of product and production levels for globally traded commodities.

CAPRI allows for *modular applications* as e.g. regional supply models for a specific Member State may be run at fixed exogenous prices without any market module. In previous applications farm heterogeneity has been represented by a set of farm types for each NUTS2 region, each with its own supply model. The farm type model layer is currently being replaced with another solution such that it has been switched OFF in recent applications. Equally, the global market model can be run in stand-alone mode as well.

Post-model analysis includes the calculation of different income indicators as variable costs, revenues, gross margins, etc., both for individual production activities as for regions, according to the methodology of the EAA. A welfare analysis at Member State level, or globally, at country or country block level, covers agricultural profits, tariff revenues, outlays for domestic supports and the money metric measure to capture welfare effects on consumers. Outlays under the first pillar of the CAP are modelled in very

[6] Tier 1 or Tier 2 depending on the context.

high detail. Among the post model analysis options there are some designed to disentangle various contributions to scenario effects as explained in Chapter "Post model analysis". An important element of post model analysis is the option of *spatial down-scaling part* to clusters of 1 × 1 km grid cells, covering crop shares, crop yields, animal stocking densities, fertilizer application rates and derived environmental indicators. This is based on a statistical approach, handeled in file capdis.gms and covered in a separate Chapter of this documentation. Model results are presented as *interactive maps* and as thematic *interactive drill-down tables.* The CAPRI graphical user interface including the exploitation tools are documented in a separate user manual[7].

More information about the CAPRI model, including technical documentation, lists of peer-reviewed and other publications, and open access to the modelling system, is available at the model webpage: [8].

1.3 CAPRI uses the GAMS software

To solve the large-scale, non-linear optimization problems in the model, CAPRI uses a software called GAMS (General Algebraic Modelling System). GAMS is a programming language designed for solving optimization problems, widely used in economic modelling. Models in GAMS are defined by one or several text files (gms files) that contain definitions and solution methods for solving constrained optimization problems (such as the supply models of CAPRI) or systems of equations (such as the marked model of CAPRI), as well as commands for data handling and reporting.

Data used or produced by GAMS is generally stored in a file format called GDX (GAMS Data Exchange). CAPRI database and results are stored in gdx files, which can be loaded into the CAPRI Result Viewer in the Graphical User Interface where you can analyse and export the results. Without GAMS, you can view and analyse scenario results from previous scenario runs, but not run new simulations with CAPRI.

GAMS solves models using third-party solvers that are linked to GAMS. GAMS comes with a large library of such solvers, most of them specializing in particular types of problems or solution algorithms. CAPRI relies on a particular solver called CONOPT. While CAPRI itself is distributed free of charge for anyone to download and use, GAMS and the solvers such as CONOPT requires a license to work beyond demonstration mode.

1.4 The network

Methodological development, updating, maintenance and application of CAPRI are based on a *network approach*, in the first 15 years certainly dominated by the key developer Wolfgang Britz and a series of PhD projects supervised by Thomas Heckelei. In the meantime responsabilites have spread with main contributors in recent years being the Bonn team (U Bonn, EuroCARE), Thünen, SLU, JRC-Sevilla and JRC –Ispra. Over the years researchers from various universities and institutes (from Norway, Switzerland and Ireland) have contributed to CAPRI, which can be seen from the contributions to many publications.

The CAPRI modelling network may be defined as a 'club': there are currently no fees attached to its use but the entry in the network is controlled by the current club members. The members have agreed on a distribution of tasks to maintain and update the system. They as well contribute by acquiring

[7] `http://www.ilr.uni-bonn.de/em/staff/britz/ggig_e.htm`
[8] `http://www.capri-model.org`

new projects, by quality control of data, new methodological approaches, model results and technical solutions, and by organising events such as training sessions and preparing this documentation. It is currently considered if the club constitution needs an update as well.

1.5 CAPRI development and applications

CAPRI – 'Common Agricultural Policy Regionalised Impact analysis' is both the acronym for an EU-wide quantitative agricultural sector modelling system and of the first project centred around it[9]. The scope of the project has widened over time: the first phase (FAIR3-CT96-1849: CAPRI 1997-1999) provided the concept of the data base and the regional supply models, but linked these to a simple market model distinguishing the EU and rest-of-the-world. In parallel, a team at the FAL (now Thünen Institute, TI) in Braunschweig applied CAPRI to assess the consequences of an increased share of biological farming system (FAIR3-CT96-1794: Effects of the CAP-reform and possible further developments on organic farming in the EU). A further, relatively small project (ENV.B.2/ETU/2000/073: Development of models and tools for assessing the environmental impact of agricultural policies, 2001-2002, financed by DG-REGIO) added a dis-aggregation below administrative regions in form of farm type models, refined the existing environmental indicators and added new ones. A new EU research framework project with the original network (QLTR-2000-00394: CAP-STRAT 2001-2004) refined many of the approaches of the first phase, and linked a complex spatial global multi-commodity model into the system. The application of CAPRI for sugar market reform options in the context of another project improved the way the complex ABC sugar quota system is handled in the model.

Later, a larger project (EU research FP VI, Nr. 501981: CAPRI-Dynaspat) was conducted under the co-ordination of the team in Bonn to render the system recursive-dynamic, dis-aggregate results in space, include the new Member States and add a labour module and an indicator for energy use.

A PhD study (Pérez-Dominguez 2005) initiated (non-CO2) GHG accounting and modelling with CAPRI to analyse tradable permits for GHG emissions from agriculture. Subsequently several projects served to improve the representation of trade policies (FP VI, Nr. 502457: "EU MedAgPol", also FP VI: "EU-MercoPol") and extended the coverage the supply models to the New Member states including Bulgaria and Romania).

In 2006-2008 a first biofuel coverage in CAPRI has been achieved during an interim stay of Wolfgang Britz at JRC-Ispra which has been expanded in later years leading to follow up studies on bioenergy policies (Blanco et al. 2010, Britz and Delzeit 2013). In 2006-2007 CAPRI made contributions to study "Integrated measures in Agriculture to reduce Ammonia emission" together with MITERRA-Europe (Alterra, Wageningen) and GAINS (IASSA, Laxenburg) which led to an update of the N-cycle description in CAPRI.

From 2006-2012 CAPRI participated in the LIFE funded EC4MACS[10], the "European Consortium for Modelling of Air Pollution and Climate Strategies" which basically triggered a series of projects focussing on and improving long run projections in a modeling cluster with the PRIMES, GAINS and GLOBIOM models[11].

[9] http://www.ilr.uni-bonn.de/agpo/rsrch/capri/capri_e.htm

[10] See http://www.ec4macs.eu/home/index.html

[11] This group of projects combines, for example, the FP7 project CC-TAME (Climate Change - Terrestrial Adaptation and Mitigation in Europe) and several projects commissioned by DG CLIMA (just starting is "EUCLIMIT 5").

In line with the shift of the CAP focus towards sustainability, CAPRI contributed to CCAT – EU Cross compliance tool[12], an FP6 project coordinated by Wageningen University, for an integrated assessment of cross compliance impacts, and entered (also in 2007) CAPRI FARM[13] aiming at an analysis of farming sustainability.

GHG abatement options have also been investigated in two studies by the JRC (IES, Ispra[14], and IPTS, Seville[15]) that may be considered the initialisation of mitigation modelling with CAPRI, a research focus that has gained in importance up from 2009 to the present[16]. Recent applications cover the challenges of including agriculture in climate change mitigation strategies (Fellmann et al. 2018) and trade liberalisation impacts on GHG emissions abatement in the agricultural sector (Himics et al. 2018).

The current two level version of land supply derives from a study on agricultural and trade policy reform impacts on land-use across the EU, with a particular focus on land abandonment (Renwick et al. 2012).

Until summer 2013, again a EU framework project co-ordinated by the team in Bonn called "CAPRI-RD" ensured various updates, and added a layer of regional CGEs, while working on the integration of CAP pillar 2 measures into the system. While the latter have become an essential element of CAP representation in the system, the regional CGEs have not been applied since that time (Schroeder et al. 2015, but this might be also considered the starting point of Wolfgang Britz, the main developper of CAPRI up to 2013, to move more into CGE modelling[17]).

Sustainability in its various facets has been the topic driving model developments and extensions that are likely to be pursued in the next years.

- Beginning with a small explorative study in 2011 several studies led to the development and improvement of a "CAPRI water version" used in various projects[18] and studies on water-food linkages (Blanco et al. 2018).
- GHG accounting and modelling beyond non-CO2 required to address LULUCF effects in projects aiming at a complete coverage of the country area in the UNFCCC classification as well as transitions between those land categories and a closed carbon balance for agricultural areas[19]).
- Several efforts have been undetaken by JRC-Ispra, partly in house, partly in specific projects to achieve a more accurate representation of various environmental indicators. The detailed nutrient flow in CAPRI has been exploited to measure nitrogen footprint of food products in the EU (Leip

[12]See https://cordis.europa.eu/project/rcn/84125/factsheet/en

[13]See http://agrilife.jrc.ec.europa.eu/s_study3.html

[14]See https://ec.europa.eu/agriculture/sites/agriculture/files/external-studies/2010/livestock-gas/full_text_en.pdf

[15]See http://ftp.jrc.es/EURdoc/JRC69817.pdf

[16]The Ecampa studies (EcAMPA2 (EcAMPA3 report is still under preparation): https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/economic-assessment-ghg-mitigation-policy-options-eu-agriculture-ecampa-2) are prominent examples of this tradition with a focus on EU mitigation, while a more global long run orientation is pursued in the AGCLIM50 studies (e.g. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/challenges-global-agriculture-climate-change-context-2050-agclim50).

[17]See https://www.ilr.uni-bonn.de/em/rsrch/cgebox/cgebox_e.htm

[18]See, for example https://ec.europa.eu/jrc/en/publication/capri-water-20-upgraded-and-updated-capri-water-module and https://www.sim4nexus.eu/

[19]This started with an ERA NET project TRUSTEE in 2013 (https://www.trustee-project.eu/), was picked up in EcAMPA3 (beginning in 2017, technical report under preparation) and is pursued under SUPREMA as off 2018 (https://www.suprema-project.eu/

et al. 2014) and to assess the impacts of European livestock production (Leip et al. 2015). The representation of environmental constraints, involving restrictions for fertiliser applications, for ammonia emissions, lifestock density, is currently being improved and also led to a representation of manure trade between regions.

- Diet shifts of food consumers offer a great potential to achieve environmental relief (as well as health benefits), such that their representation in CAPRI has been improved in the context of various partly ongoing projects[20] and studies.

Apart from the wide area of sustainability aspects of trade modelling have also been repeatedly at the heart of targeted model improvements, mostly commissioned by JRC-IPTS[21] and thereby pursuing the CAPRI tradition of bilateral trade modelling.

Two areas of technical developments are also likely to be continued in the future. The first one is the improvement of linkages to the in house JRC model IFM CAP that permits to represent the diversity of CAP restrictions only amenable to modelling at the farm level. As IFM-CAP operates with exogenous prices, it requires prices as model inputs that may be provided by CAPRI. The ongoing SUPREMA project (mentioned in the context of LULUCF modelling already) pursues these linkages while trying to also watch for computational feasibility, given that IFM-CAP covers each FADM farm individually. The second strand of technical improvements is the initialisation of a "stable release cycle" for CAPRI, based on two JRC-IPTS projects that are currently pursued under SUPREMA.

The historical review has so far focussed on those studies and projects, that left clear marks in the current system as a heritage. In addition, the system was applied to a wide range of numerous different scenarios that often left smaller "traces" in the system but illustrate its capabilities and contributed to improvement in many details that are critical for serious impact assessments. The very first application in 1999 analysed the so called 'Agenda 2000' reform package of the CAP. Shortly afterwards, a team at SLI, Lund, Sweden applied CAPRI to analyse CAP reform option for milk and dairy. FAL, Braunschweig looked into the effects of an increase of organic production systems. WTO scenarios as well as scenarios on specific trade agreements were frequnetly untertaken. Moreover, CAPRI was applied to analyse sugar market reform options at regional level, linked to results of the WATSIM and CAPSIM models. In 2003, scenarios dealing with the CAP reform package titled 'Mid Term Review' were performed by the team in Bonn (Britz et al. 2003). In the wake of the sugar market reforms various reform options have been investigated (Adenaeuer et al. 2004).

In 2004 CAPRI was used to generate a baseline in close co-operation with DG Agri match DG Agri's outlook projections which has become a regular activity. Several studies have been launched in 2007 on particular aspects of the ongoing CAP reform (decoupling project for DEFRA, UK, modulation study by LEI for DG Agri and a milk quota expiry for JRC, IPTS, Seville). The Farm Type version of CAPRI has been used frequently to look at intrasectoral distribution of CAP reform impacts[22], direct payment harmonisation (Gocht et al. 2013), CAP greening (Gocht et al. 2017), and an EU-wide policy to extend grassland areas in order to increase carbon sink capacity (Gocht et al. 2016). A recent important

[20] See e.g. `https://www.susfans.eu/`. Diet shifts were already explored in CAPRI in earlier years (e.g. under the JRC –IPTS project AgCLIM50-2) but the effort devoted to demand modelling, data and analysis has increased.

[21] We may mention the Engage specific contract on "Detailed trade policy modelling with the CAPRI model" (2014, No 154208.X1), the running Engage2 specific contract on "Update of CAPRI tariff and trade database and split of Australia-New Zealand regional block" (2019, No 935680.X7), and pending H2020 proposals.

[22] See e.g. `https://www.sciencedirect.com/science/article/pii/S0161893816300011`

application, also involving the Farm Type layer, was the impact assessment of the proposals on the post-2020 CAP, involving CAPRI in a multi-model approach to determine effects on production, prices, trade, GHG emissions and the nitrogen balance (European Commission 2018).

Several analyses have investigated potential impact of climate change in EU agriculture by introducing changes in crop yields from biophysical models as exogenous shifts. This enables to analyse regional changes in production within the EU while considering market feedback, as well as the role of trade to counterbalance uneven effects of climate change across the world (Delincé et al 2015, Blanco et al. 2017, Pérez Dominguez and Fellmann, 2018).

As will be clear from this review the CAPRI system strongly benefitted from EU Commission support in various forms. Most of the initial developments were co financed by DG RSRCH through the series of past FP and H2020 projects and. Furthermore the DG-JRC (IPTS, Seville and IES Ispra) has actively contributed to improvements and extensions in various components of the system and also stimulated system development with a continuous flow of new research questions and matching projects. Since a number of years recurring demand for up-to-date and long run projections on the part of DG CLIMA is contributing to some regularity in the updating process for data base and projections. Nonetheless the CAPRI network faces the common problem of the commons such that the update process for documentation is in risk to lag behind the moving target of the current code. Readers identifying missing or obsolete sections are therefore invited to contact any of the authors.

Chapter 2

Getting started with capri

2.1 Installing CAPRI Stable Release 2 and earlier

CAPRI Stable Release (STAR) 2 and earlier releases are published with a full set of data, i.e. including all the intermediate data required to build the complete database and produce a calibrated baseline. The model with raw data and the consolidated model database are shipped in two compressed archives. For the latest release, STAR 2.7, they are called "STAR_2.7.zip" and "results_2.7.zip" respectively, and with similar naming pattern for other releases. Follow the steps below in order to install CAPRI on your system. Using CAPRI requires extensive knowledge of how the system works. Please look out for CAPRI training courses on the page of upcoming events. The following bare-bone instructions may nevertheless be sufficient to get the system up and technically running.

2.1.1 1. Ensure that you have Java and GAMS installed on your computer

See 2.3.

2.1.2 2. Download the two archives

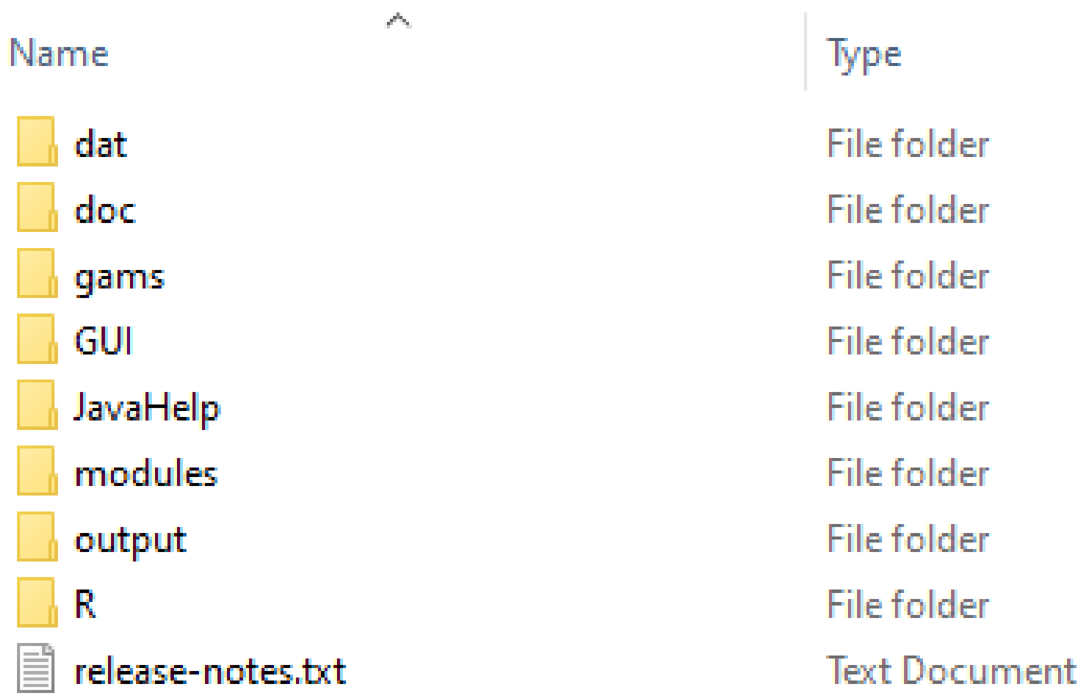
Click the links to "Code" and "Database" in the table of available CAPRI versions 2.2. We recommend using the latest available release.

2.1.3 3. Extract the files to a local hard drive

1. Create an installation folder to hold **model source files** on your local hard drive such that the path does not contain spaces. Thus, it will not work to install the model in "My Documents" (path containing space) or on a network drive (the access will be too slow). We call this the "CAPRI system folder". We assume here you created a folder `C:/CAPRI/STAR_2.7`.
2. Extract the files of the compressed archive with code, e.g. `STAR_2.7.zip` into the CAPRI system folder.

3. Extract the files of the compressed archive with data, e.g. `results_2.7.zip` into the subfolder `output/results` in the CAPRI system folder.

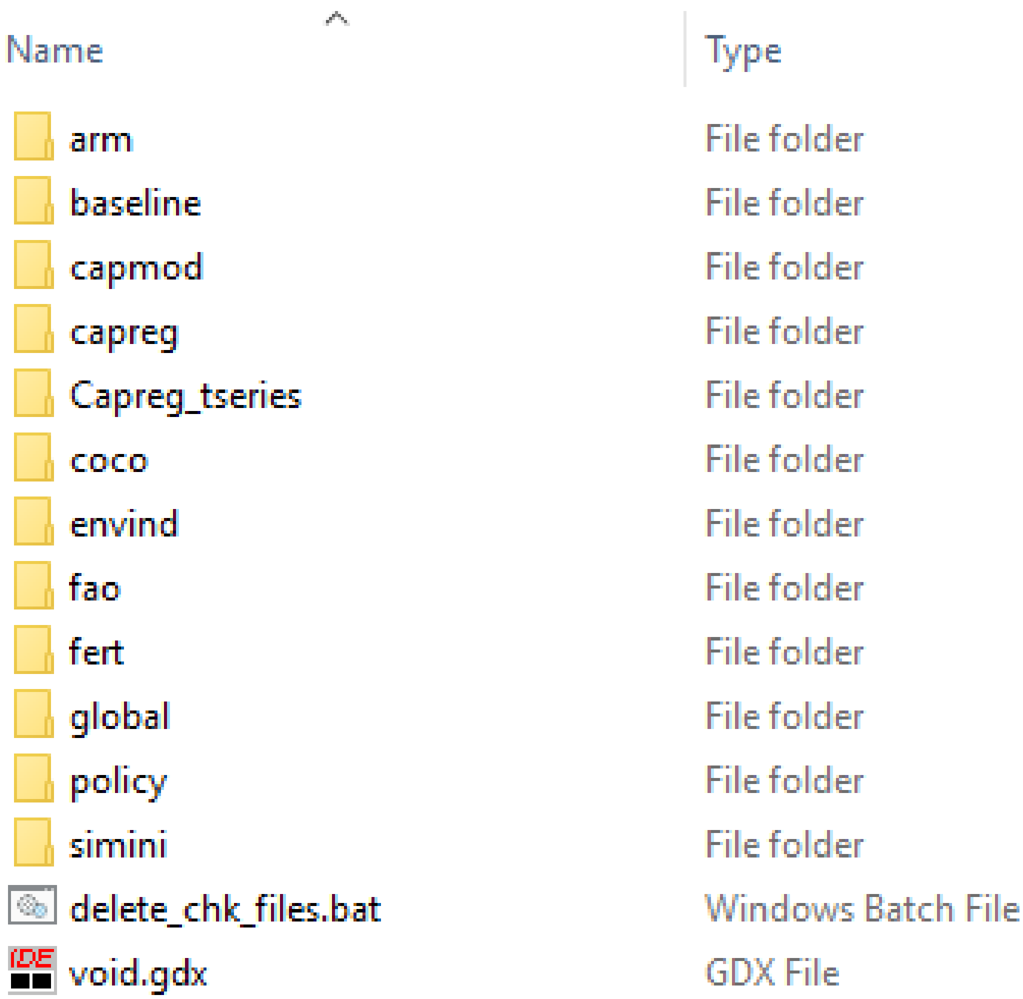
The resulting directory structure in your CAPRI system folder should look like this afterwards:



The directory structure inside the folder `output/results` should look like this:

2.1.4 4. Adjust the settings to your computer

1. Go to the subfolder "GUI", double-click "start_capri.bat". That should open the graphical user interface (GUI). If not, something is wrong with your Java-installation.
2. In the menu **settings**, choose **Edit settings.**
   - In the first tab: enter your name
   - In the second tab: verify that the paths to result and restart folders are set to **../output/results and ../output/restart.** The GAMS directory should read **../gams** and the data directory should read **../dat.**
   - In the third tab: enter the complete path to the GAMS executable. Probably it is something like **c:/gams/win64/30.3/gams.exe.** Also verify that the path to the Scratch directory is **../output/temp.** Click the button to "get number of processors".
   - In the fourth tab: do nothing.



List of result folders and files extracted and placed in ./output/results

- In the final tab: do nothing, or, if you have a programming text editor, enter the path to that editor in the proper field.

3. Click “Save in caprinew.ini” and accept the file name suggested for your settings.
4. Close the GUI, then start it again (using start_capri.bat), to make it process all your settings.

2.2 Download the current release

From 2016, the CAPRI model was made more accessible to the scientific community by the provision of supported stable releases. The stable releases are tagged and “frozen” versions of the model that can be referenced to and that will not change. The bug fixes and other code updates result in new releases with updated revision numbers. The release versions have been subjected to extensive testing, ensuring that all supported features technically work under different hardware/software settings.

The stable release consists of (a) the model code including raw data and (b) the compiled data base including calibrated baselines. The model is really self contained in (a), since all software and raw data needed to compile the data bases and construct the baselines is found there. Albeit (b), the databases, can be derived from (a), doing so is a somewhat complicated process. Therefore, we provide (b) as a shortcut. Furthermore, it turns out that generating (b) may generate slightly different results on different hardware/software combinations, and therefore it is convenient to use a common versioned database release.

The current release and selected previous releases can be downloaded as compressed (zipped) archives using the links in the following table. Note that the item "SVN-tag" only are available to developers with a an account in the SVN database. Information about each release is found in the release notes, 2.5 and in each release code base.

Release	Item	Link
STAR 2.7	Code	`https://1drv.ms/u/s!Aj7_RXyD8q-MhJlpsZVKK4eRDCjM2w?e=gLdPBw`
	Database	`https://1drv.ms/u/s!Aj7_RXyD8q-MhJlqh8nnJYSkfOm7hw?e=5dbyr1`
	SVN-tag	`https://svn1.agp.uni-bonn.de/svn/capri/tags/star/2.7`
STAR 2.6	Code	`https://1drv.ms/u/s!Aj7_RXyD8q-Mg-08OEMGjf1A8H-wuQ?e=9QnIDU`
	Database	`https://1drv.ms/u/s!Aj7_RXyD8q-Mg-076HkDqqg6mVOOHQ?e=eDOTmO`
	SVN-tag	`https://svn1.agp.uni-bonn.de/svn/capri/tags/star/2.6`
STAR 2.5	Code	`https://1drv.ms/u/s!Aj7_RXyD8q-Mg4BDjQ4AbWbXXzY1_g?e=NvyyLT`
	Database	`https://1drv.ms/u/s!Aj7_RXyD8q-Mg4BBF3d-lsULH9gH3A?e=fIIFOd`
	SVN-tag	`https://svn1.agp.uni-bonn.de/svn/capri/tags/star/2.5`
STAR 2.4	Code	`https://1drv.ms/u/s!Aj7_RXyD8q-MgtQmzmdEn_3FuNaqyA`
	Database	`https://1drv.ms/u/s!Aj7_RXyD8q-MgtQlHdUSz6F5Uv_Sgw`
	SVN-tag	`https://svn1.agp.uni-bonn.de/svn/capri/tags/star/2.4`
STAR 1.4	Code	`https://1drv.ms/u/s!Aj7_RXyD8q-MgpgRyPwElFGkhK5QJw`
	Database	`https://1drv.ms/u/s!Aj7_RXyD8q-MgpgSo2ebWOOY-c3oXA`
	SVN-tag	`https://svn1.agp.uni-bonn.de/svn/capri/tags/star/1.4`

2.3 System requirements

CAPRI requires that you have a windows computer with Java and GAMS (`http://www.gams.com`, distribution 32.2 or later is recommended) installed. You need a license for the CONOPT solver. Many tasks in CAPRI utilize parallel computing. It is therefore an advantage if you have a machine that can run many threads in parallel.

Regarding Java: There are some licensing implications when using Oracle's platform. Public updates released after January 2019 will not be available for business, commercial, or production use without a commercial license. Since then the GUI for java can also be executed using a open and free installation of `https://jdk.java.net/15/`.

2.4 Build databases and baselines

The stable releases are shipped with the data needed to make simulations. Advanced users may want to run the data consolidation steps themselves. It can be done using the following steps. Note that this is not needed if you downloaded and installed the pre-compiled database in the steps described above.

Compiling the database from scratch takes a long time. Depending on the hardware you use, it can take up to several days. In order to keep track of all the various settings required for completing all the steps, the release contains a "batch execution file" that instructs the GUI how to carry out a sequence of tasks without intervention of the user. Here is what you need to do:

1. In the GUI, choose the menu GUI –> Batch execution
2. In the dialogue, choose the file `<gamsdir>/GUI/batchfiles/build_database_and_baseline.txt`
3. Uncheck the option "Only compile the GAMS programs"
4. Click "Start batch execution".
5. Wait for the program to finish (hours, days).
6. Click "Open HTML report" and verify that all steps were completed without errors (RC=0).

The current batch execution file (STAR 2.7) does not take over the path to your GAMS installation as entered in the settings dialogue during installation, but you will have to open the text file and enter the appropriate path manually.

2.5 Release notes (cumulative)

STAR 2.7 Most importantly, the dairy market should now be linked with the supply models, and the new feed version is used by default. Other minor issues below:

- Priors missing in GHG emission trend estimation now exist (again?)
- The fallback structure for missing nutrient balance data in West Balkan and Turkey taken over from Trunk.
- Import surge of cheese in market baseline prevented by adjusted bounds widening. This problem does not seem to exist in trunk.
- FAOSTAT bug fixed (exportGUI task failed)

STAR 2.6 Update of the graphical user interface (GUI) to allow the use of the policy editor.

STAR 2.5 Several modifications of the premium payments, for the Basic Payment Schemes, where the results were found not to fit observations.

STAR 2.4 This maintenance release addresses ...

- Pillar 1 payments of the CAP were missing for "new" member states when running CAP 2014-2020 for a year before 2020.

- Several minor tweaks and fixes.

STAR 2.3 This maintenance release addresses stability of feeding, reporting for fertilizers, and also includes some cleaning up of code:

- The feed distribution was revised to become more stable under repeated starts (P. Witzke)
- Several options were removed from the GUI with respect to the estimation of GHG emission trends
- A large pack of updates to the GHG emission estimations were imported from the development branch ClipByFood
- The inner fertilizer allocation model can be turned OFF in simulations (on by default)
- The GUI now reports a decomposition of the NUTNED_ equation, accessible under the theme "Fertilization"
- Some items were excluded from checks in COCO to make the program run through with GAMS 25.0 and 25.1
- The legacy data set on ghg emissions from EDGAR was removed.

STAR 2.2 This maintenance release resolves several important calibration issues.

- The market model did not calibrate properly due to an inconsistently included "BREXIT" policy
- The supply models did not calibrate properly due to missing parameters for manure trade
- The stability tests did not give true results due to an inconsistent use of results_in and results_out, essentially mixing fertilizer parameters of two runs
- Several minor bug fixes, e.g. in linking supply and demand and scaling of NMIN
- A final overhaul of feeding is still due in Maintenance Release 2.3.

STAR 2.1 This maintenance release implements the revised treatment of fertilizers and feed. For fertilizers, a bi-level programming approach has been implemented, where the flows of fertilizers are modelled as a Bayesian estimator ensuring an interior solution that is "close to" the calibrated flows when simulating and close to a prior distribution when calibrating. For feed, the distribution of feeding stuffs to animals was revised to improve plausibility and stability, but without principal changes of the way the model works.

In addition to the feed and fertilizer modifications, the following bugs or minor issues were addressed:

- Changing the order of work steps and tasks in the GUI to reflect the order in which the steps can be carried out (e.g. FAOSTAT first)
- An issue with price experiments in Threads-mode was resolved
- Start at an infrastructure to report the results of the fertilizer allocation to DATAOUT (reports/fert_dist_results.gms)
- Relaxing the winter cover requirement in one Finnish region to avoid infeasibility in baselines.

- The income computations for EAA had inconsistent prices for FODDER, so that the regional farms had costs <> revenues for fodder, which should not be possible.
- The Dual Analysis of the supply models had not updated versions of some constraints, so that there were an “unexplained rest”. (sugar beet and greening restrictions)

Finally, the testing routines were augmented and slightly revised, to include individual testing of simulations with supply and market models standalone.

STAR 2.0 This new series of releases contains two key modifications:

- It calibrates to the CAP post 2014 instead of the old MTR scenario. This required adding the first-order conditions of the greening restrictions in the PMP algorithm, which was not trivial.
- It includes the possibility to set BREXIT ON when building the database. This allows the user to build a model where the UK is a separate market model region with bilateral trade instruments with the EU. This is now the standard setting, albeit it implements free trade between UK and EU27.

Some modifications that were scheduled for this release were not included, because they were not sufficiently stable in testing at the key date for the release. In particular, the following components are essentially unchanged from STAR 1.3 but scheduled for inclusion in a subsequent release.

- Revision of the fertilizer distribution. This was found to be numerically unstable in STAR 1.0, and is still so as testing revealed.
- Revision of the feed distribution to give more stable results and a more plausible allocation of feeding stuffs.

Finally, this release has some known issues in addition to the points mentioned above:

- New Norwegian data was made available but not in time to complete the testing phase.
- There are occasional problems to reproduce the baseline in 2030 when market and supply are allowed to interact. The precise circumstances causing this to happen are still unclear.

STAR 1.3 A maintenance release addressing the following issue in STAR 1.2:

- Including the most recent Graphical User Interface (GUI)
- A bug with the Basic Payment Scheme in Greece led to missing payments in several Greek regions.
- Changing the way the grassland maintenance requirement works in Greening (lower bound on grass land)
- Modified reporting for EU28
- Modified report tables for the GUI

It was observed that when the GUI batch execution file “build_database_and_baseline.txt” was executed with this model version, the baseline calibration of the market model sometimes failed.In that case, a manual re-start of that task directly from the GUI using the default settings worked. Furthermore, it was observed that the baseline reproduction run (i.e. calibrating to mtr_rd_cal and then simulating mtr_rd_ref) resulted in small changes in some of the “new” member states, in particular Serbia.

STAR 1.2 A maintenance release addressing two minor problems encountered in STAR 1.1.

- A bug fix in the rural development policy logic (gams/policy/rd_logic.gms). The bug may have caused problems when building a regionalized database if a particular folder (results_out/capmod) was missing.
- A bug in the batch execution file “build_database_and_baseline.txt” that prevented the farm type databases from being built.

STAR 1.1 This is a maintenance release addressing some issues that surfaced since STAR 1.0 was published.

- The scaling of Japanese prices was wrong, leading to biased results in the market model
- Several issues relating to the implementation of the second pillar payments, causing them to be missing or wrong in CAPMOD (simulation) and also CAPREG (regional database)
- Renaming all the standard scenario files in the folder pol_input/CAP_AFTER_2014, so that the (CAPMOD) result file names become shorter and more instructive
- A randomly appearing issue with farm type trends. GAMS had problems deleting grid computing handles under full system load (parallel computing)
- A bug in the user interface that caused the batch execution (e.g. build_database_and_baseline.txt) to launch Turkey (only!) in the wrong way
- Setting the default number of processors in the “build_database_and_baseline.txt” batch to “4”, so that it is fairly safe to start without any modification.

STAR 1.0 This release attempts to provide a CAPRI model where a wide selection of tasks from baseline construction to simulation can be carried out. With other versions of CAPRI, it has been a general feature that when some tasks were maintained, others ceased to work, so that there were multiple model versions where some problems had been resolved but where not everything worked properly.

Since it is utopic to aspire that all mechanisms ever built into CAPRI would work simultaneously, a selection of “supported features” was created. Features of the model that are not “supported” are simply not tested, and so they may or may not perform as intended. The list of supported features is documented in programmatic form in the GUI batch execution file “supported_features.txt”.

Some features that should be supported still fail to work properly. In particular, we note that the following technical problems persist:

- “Generate GAMS child processes on different threads”, causing many procedures in the model to run in parallel as gams child processes if set ON, is not entirely stable. Recommendation is to keep OFF for reproducibility.

- “Dampening of high activity level elasticities” must still be kept OFF. It is unclear whether this feature will survive or rather be replaced by some general adaptation of elasticities for long run experiments in combination with a revised calibration procedure
- Numerical instability of the calibration of fertilizer distribution among crops. Repeated runs do not give identical results.
- Numerical instability of the calibration of animal feed to various animals. Repeated runs do not give identical results.
- Occasional failure of task “Generate farm type trends” for random regions. Remedy: Re-run “Build regional time series” and “Build regional database” for that country, for nuts2 and farm types, and redo all tasks from “Build global database” onwards.

The release has not been systematically tested from a content point of view. Nevertheless, release candidates have been used in a few applications, where some issues have surfaced. In particular, the distribution of rural development funds needs to be revised. Such revision has partially been done already in various projects, but the modifications need to be consolidated and integrated into a maintenance release. Similar improvements have accumulated in the areas of market model tariff data and greenhouse gas emissions, also foreseen to be integrated in a future maintenance release, after thorough testing.

2.6 Trunk result files for download by revision

Result folder zip	trunk revision	Date	batch file
Download link	8300	24.10.2020	`https://svn1.agp.uni-bonn.de/!/#capri/view/head/trunk`
Download link	8872	04.10.2020	`https://svn1.agp.uni-bonn.de/!/#capri/view/head/trunk`

Chapter 3

The capri data base

Models and data are almost not separable. Methodological concepts can only be put to work if the necessary data are available. Equally, results obtained with a model mirror the quality of the underlying data. The CAPRI modelling team consequently invested considerable resources to build up a data base suitable for the purposes of the project. From the beginning, the idea was to create wherever possible sustainable links to well-established statistical data and to develop algorithms which can be applied across regions and time, so that an automated update of the different pieces of the CAPRI data base could be performed as far as possible.

The main guidelines for the different pieces of the data base are:

- Wherever possible link to harmonised, well documented, official and generally available data sources to ensure wide-spread acceptance of the data and their sustainability.
- Completeness over time and space. As far as official data sources comprise gaps, suitable algorithms were developed and applied to fill these.
- Consistency between the different data (closed market balances, perfect aggregation from lower to higher regional level etc.)
- Consistent link between 'economic' data as prices and revenues and 'physical data' as farm and market balances, crop rotations, herd sizes, yields and input demand.

According to the different regional layers interlinked in the modelling system, data at Member State level (in terms of modelling) currently EU28 plus Norway, Turkey and Western Balkan countries need to fit to data at regional level administrative units at the so-called NUTS 2 level, about 300 European regions and data at global level, currently 44 "non supply-model-regions. A further layer consists of georeferenced information at the level of clusters of 1×1 km grid cells which serves as input in the spatial down-scaling part of CAPRI. This data base is discussed along with the methodology and not in the current chapter. As it would be impossible to ensure consistency across all regional layers simultaneously, the process of building up the data base is split in several parts:

- Building up the data base at national or Member State level. It integrates the EAA (valued output and input use) with market and farm data, with areas and herd sizes and a herd flow model for young animals (Section 3.2).

- Building up the data base at regional or NUTS 2 level , which takes the national data basically as given (for purposes of data consistency), and includes the allocation of inputs across activities and regions as well as consistent acreages, herd sizes and yields at regional level (Section 3.3).
- The input allocation step is a key step in the establishment of the database. It allows the calculation of regional and activity specific economic indicators such as revenues, costs and gross margins per hectare or head and is covered in a separate Section 3.4.
- Building up the global data base, which includes supply utilisation accounts for the other regions in the market model, bilateral trade flows, as well as data on trade policies (Most Favourite Nation Tariffs, Preferential Agreements, Tariff Rate quotas, export subsidies) (Section 3.5).
- Given the extent of public intervention in the agricultural sector, policy data complete the database. They are partly supply oriented CAP instruments like premiums and quotas and partly data on trade policies (Most Favourite Nation Tariffs, Preferential Agreements, Tariff Rate quotas, export subsidies) plus data domestic market support instruments (market interventions, subsidies to consumption), see Section 3.6.

The basic principle of the CAPRI data base is that of the 'Activity Based Table of Accounts' which roots in the combination of a physical and valued input/output table including market balances, activity levels (acreages and herd sizes) and the EAA.

3.1 Production Activities as the core

`Authorship:Peter Witzke`

The economic activities in the agricultural sector are broken down conceptually into 'production activities' (e.g. cropping a hectare of wheat or fattening a pig). These activities are characterised by physical output and input coefficients. For most activities, total production quantities can be found in statistics and output coefficients derived by division of activity levels (e.g. 'soft wheat' would produce 'soft wheat' and 'straw', whereas 'pigs for fattening' would produce 'pig meat' and NPK comprised in manure). However, for some activities other sources of information are necessary (e.g. a carcass weight of sows is necessary to derive the output coefficient for the pig fattening process). For manure output engineering functions are used to define the output coefficients. The way the different output coefficients are calculated is described in more detail below.

The second part characterising the production activities are the *input coefficients.* Soft wheat, to pick up our example again, would be linked to a certain use of NPK fertiliser, to the use of plant protection inputs, repair and energy costs. All these inputs are used by many activities, and official data regarding the distribution of inputs to activities are not available. The process of attributing total input in a region to individual activities is called input allocation. It is methodologically more demanding than constructing output coefficients. Specific estimators are developed for young animals, fertilisers, feed and the remaining inputs, which are discussed below.

Multiplied with average farm gate prices for outputs and inputs respectively, output coefficients define farm gate revenues, and input coefficients variable production costs. The average farm prices used in the

CAPRI data base are derived from the EEA and hence link physical and valued statistics. However, in some cases as young animals and manure which are not valued in the EEA, own estimates are introduced.

In order to finalise the characterisation of the income situation in the different production activities, subsidies paid to production must be taken into account. The CAPRI data base features a rather complex description of the different CAP premiums allocated to the individual activities. However, subsidies outside of the CAP for the EU Member States have received less attention (in line with smaller amounts).

The following table gives an example for selected activity related information from the CAPRI data base.

Table 1: Example of selected data base elements for a production activity

	SWHE [Soft wheat production activity]	Description	Unit
Outputs			
SWHE	7853.84	Soft wheat yield	kg/ha
STRA	9817.30	Straw yield	kg/ha
Inputs			
NITF	175.52	Organic and anorganic N applied	kg/ha
PHOF	49.57	Organic and anorganic P applied	kg/ha
POTF	62.51	Organic and anorganic K applied	kg/ha
SEED	70.91	Seed input	const Euro 1995/ha
PLAP	59.85	Plant protection products	const Euro 1995/ha
REPA	53.27	Repair costs	const Euro 1995/ha
ENER	25.15	Energy costs	const Euro 1995/ha
INPO	79.25	Other inputs	const Euro 1995/ha
Income indicators			
TOOU	825.26	Value of total outputs	Euro/ha
TOIN	522.13	Value of total inputs	Euro/ha
GVAP	303.13	Gross value added at producer prices	Euro/ha
PRME	328.86	CAP premiums	Euro/ha
MGVA	631.99	Gross value added at producer prices plus premiums	Euro/ha
Activity level and data relating to CAP			
LEVL	609.91	Hectares cropped	1000 ha
HSTY	5.22	Historic yield used to define CAP premiums	t/ha
SETR	8.63	Set aside rate	%

3.1.1 Technology variants for production activities

For most activities there are two technologies available, typically a low and a high yield variety. Usually they are defined to cover each 50% of the activity level observed in ex post data, but with some particularities in the sugar sector (see *'/sugar/techf.gms'*).

3.1.2 Linking production activities and the market

The connection between the individual activities and the markets are the activity levels. Total soft wheat produced is the sum of cropped soft wheat hectares multiplied with the average soft wheat output coefficient. In cases like pig meat, as mentioned before, several activities are involved to derive production.

The produced quantities enter the farm and market balances. Production plus imports as the resources are equal to the different use positions as exports, stock changes, feed use, human consumption and processing. These balances are only available at Member State, not at regional level. Production establishes the link to the EAA as well, as average farm gate prices are unit values derived by dividing the values from the EAA by production quantities.

The three basic identities linking the different elements of the data base are expressed in mathematical terms as following. The first equation implies that total production or total input use (code in the data base: GROF or gross production/gross input use at farm level) can be derived from the input and output coefficients and the activity levels (LEVL):

$$GROF_j = \sum_j LEVL_j \cdot IO_j \tag{3.1}$$

The second type of identities refers to the farm and market balances:

$$\begin{aligned} GROF_{io} - SEDF_{io} - LOSF_{io} - INTF_{io} &= NETF_{io} \\ NETF + IMPT_{io} &= EXPT_{io} + STCM_{io} \\ &+ FEDM_{io} + LOSM_{io} \\ &+ SEDM_{io} + HCOM_{io} \\ &+ INDM_{io} + PRCM_{io} \\ &+ BIOF_{io} \end{aligned} \tag{3.2}$$

The farm balance positions are seed use (SEDF) and losses (LOSF) on farm (only reported for cereals) and internal use on farm (INTF, only reported for manure and young animals). NETF or net trade on farm is hence equal to valued production/input use and establishes the link between the market and the agricultural production activity. Adding imports (IMPT) to NETF defines total resources, which must be equal to exports (EXPT), stock changes (STCM), feed use on market (FEDM), losses on market (LOSM), seed use on market (SEDM), human consumption (HCOM), industrial use (INDM), processing (PRCM), and use for biofuel production (BIOF).

The third identity defines the value of the EAA in producer prices (EAAP) as sold production or purchased input use (NETF) in physical terms multiplied with the unit valued price (UVAP):

$$EAAP_{io} = UVAP_{io} NETF_{io} \tag{3.3}$$

The following table shows the elements of the CAPRI data base as they have been arranged in the tables of the data base.

Table 2: Main elements of the CAPRI data base

	Activities	Farm- and market balances	Prices	Positionsform from the EAA
Outputs	Output coefficients	Production, seed and feed use, other internal use, losses, stock changes, exports and imports, human consumption, processing	Unit value prices from the EAA with and without subsidies and taxes	Value of outputs with or without subsidies and taxes linked to production
Inputs	Input coefficients	Purchases, internal deliveries	Unit value prices from the EAA with and without subsidies and taxes	Value of inputs with or without subsidies and taxes link to input use
Income indicators	Revenues, costs, Gross Value Added, premiums			Total revenues, costs, gross value added, subsidies, taxes
Activity levels	Hectares, slaughtered heads or herd sizes			
Secondary products		Marketable production, losses, stock changes, exports and imports, human consumption, processing	Consumer prices	

3.2 The Complete and Consistent Data Base (COCO) for the national scale

The COCO database is built by the application of two modules:

COCO1 module:
Prepare national database for all EU27 Member States the Western Balkan Countries, Turkey and Norway.

It is basically divided into three main parts:

- A data import “part” that is not a single “module” but rather a collection activity to prepare a large set of very heterogeneous input files
- Including and combining these partly overlapping input data according to some hierarchical overlay criteria, and

- Calculating complete and consistent time series while remaining close to the raw data.

Data preparation (part 1) and overlay (part 2) form a bridge between raw data and their consolidation to impose completeness and consistency. The overlay part tries to tackle gaps in the data in a quite conventional way: If data in the first best source (say a particular Eurostat table from some domain) are unavailable, look for a second best source and fill the gaps using a conversion factor to take account of potential differences in definitions. To process the amount of data needed in a reasonable time this search to second, third or even fourth best solutions is handled as far as possible in a generic way in the GAMS code of COCO where it is checked whether certain data are given and reasonable. However there are a few special topics that are explained in separate sections.

COCO2:
The finishing step estimates consumer prices, consumption losses, and some supplementary data for the feed sector (by-products used as feedstuffs, animal requirements on the MS level, contents and yields of roughage). Both tasks run simultaneously for all countries and build on intermediate results from the main (COCO1) part of COCO like human consumption and processing quantities.

3.2.1 Overview and data requirements for the national scale

An overview on the key data collection, assingments and corrections in main program coco1.gms is given in the following figure.

The different steps will be explained in more detail in the following sections.

The CAPRI modelling system is, as far as possible, fed by statistical sources available at European level which are mostly centralised and regularly updated. Farm and market balances, economic indicators, acreages, herd sizes and national input output coefficients were initially almost entirely from EUROSTAT. In the course of time, more and more special data sets have been added to fill gaps or resolve problems detected in EUROSTAT data, such as specific data on Western Balkan Countries or on the biofuel sector.

The main sources used to build up the national data base are shown in the following.

Figure 2: Overview on key elements in the consolidation of European data at the Member state level (in coco1.gms)

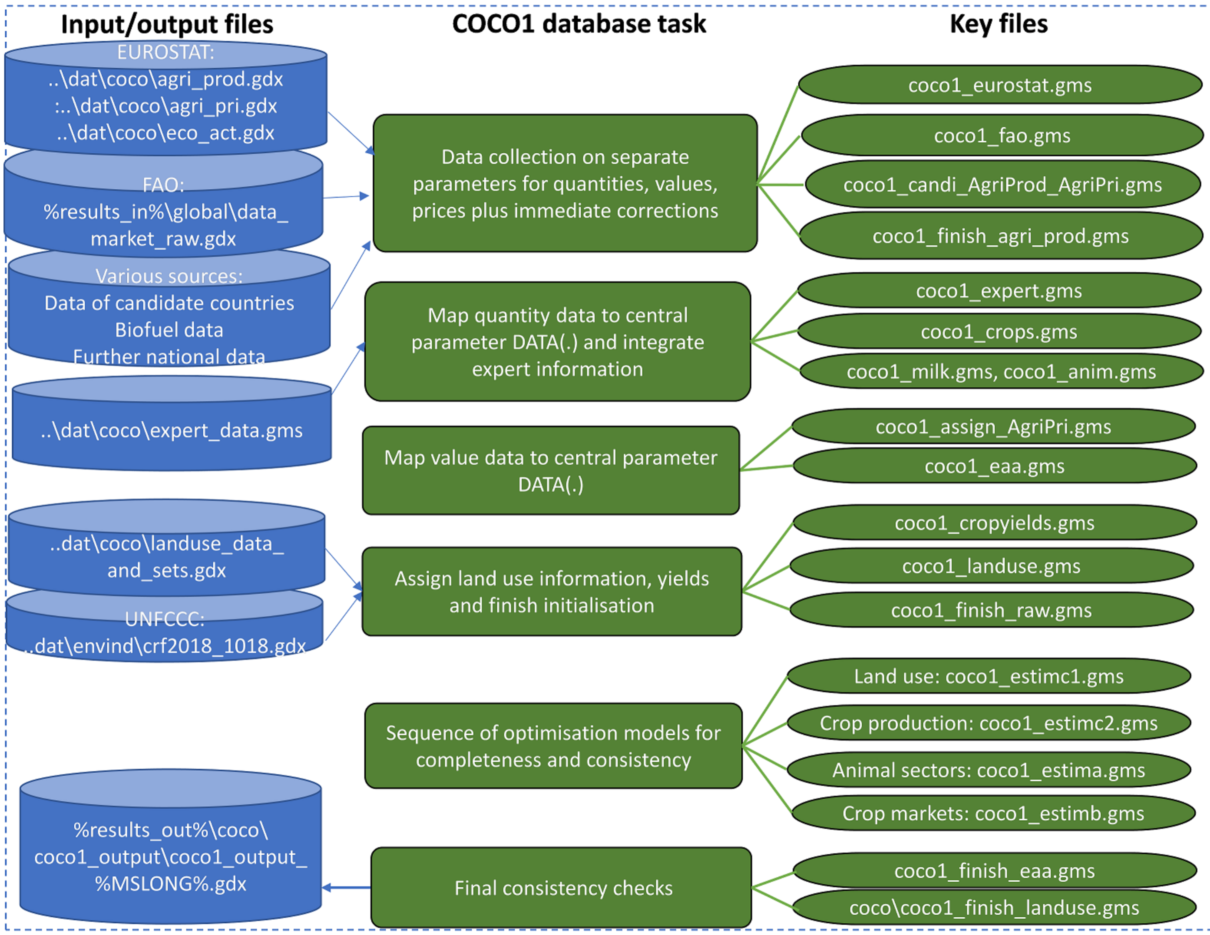


Table 3: Data items and their main sources

Data items	Source
Activity levels	Eurostat: Crop production statistics, Land use statistics, herd size statistics, slaughtering statistics, statistics on import and export of live animals For Western Balkan Countries and Turkey: Eurostat supplemented with national statistical yearbooks, data from national ministries, FAOstat production statistics and others
Production, farm and market balance positions	Eurostat: Farm and market balance statistics, crop production statistics, slaughtering statistics, statistics on import and export of live animals For Western Balkan Countries and Turkey: Eurostat supplemented with national statistical yearbooks, data from national ministries, FAOstat production statistics and others
Sectoral revenues, costs, and producer prices	Eurostat: Economic Accounts for Agriculture (EAA) and price indices for gap filling, otherwise unit value calculation For Western Balkan Countries and Turkey: Supplemented with national statistical yearbooks, data from national ministries, results from AgriPolicy, FAOstat price statistics
Consumer prices	Derived from macroeconomic expenditure data (Eurostat, supplemented with UNSTATS) and food price information from various sources
Output coefficients	Derived from production and activity levels, engineering knowledge

3.2.2 Data Import

A large set of very heterogeneous input files (in terms of organisation and format) is collected, currently covering the following years:

Table 4: Temporal coverage of national data by region

Member State	Range
EU15 Member States without Germany	1984 – 2018/2019
Germany and (12) New Member States	1989 – 2018/2019
Western Balkan (WB) Countries and Turkey	1995 – 2018/2019
Norway	1984 – 2017

Eurostat data **First step: Data download and format conversion** Data are originally downloaded in "TSV-format", as offered by Eurostat for bulk data users. The TSV-format is a flat file format for time series. Data can be selected for all EU MS and some Candidate Countries. Availability differs by country, of course (almost nothing for the Kosovo, Montenegro, Bosnia & Herzegonina). In the process of downloading the TSV files are also converted in GAMS readable form (csv or gdx). The following themes and table groups of Eurostat are accessed:

Agriculture, forestry and fisheries

- Agriculture ("agr")
  - Economic Accounts for Agriculture (Table Group "aact", saved on CAPRI parameter "p_ecoact"
  - Agricultural prices and price indices (Table Group "apri", saved on CAPRI parameter "p_agripri"
  - Agricultural product related physical information (production, activity levels from Table Group "apro", saved on CAPRI parameter "p_agriprod"
  - Older, discontinued Eurostat series that still provide useful information (requiring some ad hoc extrapolations), for example (a) market balance information for products other than cereals, oilseeds and wine, critical for "COCO1", (b) relative price level indices of food products (MS relative to EU average) for COCO2, (c) availability and production of feedingsstuffs (useful for COCO2 completions on feed from by-products)

Economy and Finance

- National annual accounts ("nama10")
  - Annual national accounts -> National Accounts detailed breakdowns (by industry, by product, by consumption purpose) -> Final consumption expenditure of households by consumption purpose (COICOP 3 digit),
  - General indicators to National Accounts - Population and employment
  - GDP and main components - Current prices, volumes, price indices
- Prices ("prc")

– Harmonized indices of consumer prices (prc_hicp) here: HICP (2005=100) -annual Data, and HICP - Item weights

Second step: data selection and code mapping

The second step is data selection and code mapping performed by the GAMS program *'coco_input.gms'*. Cross sets linking Eurostat codes to COCO codes define the subset of data series subsequently used.

The mapping rules are collected in two sub-programs called by *'coco_input.gms'*, for example:

- *'gams/coco/ eurostat_agriculture_mapping.gms'* for the tables from Eurostat's "Agriculture and Fisheries" Statistics
- *'eurostat_ econfinc_mapping.gms'* for the tables from Eurostat's "Economy and Finance" Statistics

Example from file *'Eurostat _agriculture_mapping.gms'*. The results of the program run are gdx-files loaded by files (e.g. coco/coco1_eurostat.gms) which are in turn loaded by coco1.gms or coco2.gms.

```
SET EcoActMAP(ASS_COLS,ASS_ROWS, eco_act_ori_eurostat) "mapping" /
EAAP.CERE. aact_eaa01_01000_PROD_PP_MIO_EUR
EAAP.SWHE. aact_eaa01_01110_PROD_PP_MIO_EUR
EAAP.DWHE. aact_eaa01_01120_PROD_PP_MIO_EUR /;



SET AgriProdMAP(ASS_COLS,ASS_ROWS, agri_prod_ori_eurostat) "mapping" /
CERE.LEVL.( apro_cpnh1_C1000_AR,apro_cpnh1_h_C1000_AR)
SWHE.LEVL.( apro_cpnh1_C1110_AR,apro_cpnh1_h_C1110_AR)
SWH1.LEVL.( apro_cpnh1_C1111_AR,apro_cpnh1_h_C1111_AR) /;
```

Western Balkan Countries and Turkey For those countries Eurostat data need completion in almost every area which is handled in country specific xls files. The structure of these supplementary Excel country sheets and the definitions of the data are tailored to COCO. The resulting sheets in these xls files are uniform across countries, in order to ease data extraction for the modelling part by applying macros. However, each national information system has its own peculiarities and hence, not all data are fully harmonised across countries. Various sources are assessed and combined in a case by case manner: Eurostat data, if already available and plausible, are handled as the preferred data source. Data collected from the national statistical yearbooks have second priority, followed by expert data collected in from earlier projects. Finally FAO data provides often the fall-back solution for any remaining missing time series.

The final sheet in each of these country specific xls files is the interface to the GAMS programing world of COCO. An Excel macro "SELECT_data_all" collects the time-series compiled in other sheets and puts them into this final sheet with the appropriate COCO code. Another macro finally exports the numbers into text files like "dat/coco/bosnia_coco.gms". Because the xls file are quite complex due to various linkages, we do not read directly from them. This avoids unplanned changes and permits convenient tracing of data changes via the CAPRI versioning system svn.

Supplementary data for Romania and Bulgaria Country level data from national experts were compiled in Excel files that help in particular to complete the meat and milk sectors.

FAO data selection Two FAO data sources are combined:

- For all regions FAO data (mapped in the context of module "global database" to CAPRI codes and hence consistent across modules) serve as a fall back option under certain conditions, defined in the code. This fall back function of FAO data has gained in importance since Eurostat discontinued the publication of most market balances since 2014. In some cases also activity level (area) information may be taken from FAO.
- Some particular data like disaggregate data on herds of chicken, ducks, turkeys and geese are compiled in a separate include file dat/coco/fao_add.gms because these data types are usually not loaded for global database.

Other additional input data COCO1: Biofuels

- Production, market balance and feedstock quantities for biodiesel and bioethanol are collected from a multitude of sources:
  - EU project `http://www.elobio.eu` (production, demand, biodiese and bioethanol, 1999-2007)
  - Eurostat, Energy balances and demand (tables nrg_xxxx) production, demand, trade for diesel, gasoline, biodiesel and bioethanol, 2001-15)
  - Eurostat, Production and trade (PRODCOM), ethanol and biodiesel, 2000-14
  - PRIMES model[1] database (production, biodiesel and bioethanol, 2000-07)
  - US Energy Information Administration (EIA), production of biodiesel and bioethanol, 2000-12, incl. some non-EU countries
  - DG Agri Ethanol balances (production partly with split by feedstocks and MS, demand and trade)
  - Aglink ex post database (most data for Turkey, also EU biofuel production from non-standard sources (NAGR).
  - USDA GAIN reports (market balances for Serbia, feedstocks for biodiesel in EU)
  - FAOstat (market balances for palm oil)
- Prices at the pump and retail prices for diesel and gasoline are from Eurostat's energy database (`http://epp.eurostat.ec.europa.eu/portal/page/portal/energy/data/database`), supplemented with IEA Statistics 2016 for Turkey.
- Taxes for diesel, gasoline, biodiesel and bioethanol are collected from DG Energy website and publications, and EURACTIV, EU news & policy debates, Brussels (`http://www.euractiv.com/en/enterprise-jobs/fuel-taxation/article-117495`)
- Some supplementary Aglink data give information on feedstock composition, tariffs and world market prices for crude oil, biodiesel and bioethanol.

[1]PRIMES MODEL, EC3MLAB of ICCS, National University of Athens.

- Trade data for undenatured ethyl alcohol, denatured ethyl alcohol, fatty acid mono-alkyl esters, crude palm oil, palm and fraction and palm kernel and fraction are collected from Eurostat's COMEXT data (2000-14).
- Market balances for palm oil are taken from FAOstat and supplemented with COMEXT.

COCO1: Sugar Quotas

- All sugar quotas 1999 until 2006 from the annual sugar yearbook.
- Buy-back 2006 in the restructuring program from CAP monitor 16 January 2008.
- Sugar quotas renounced by member states following sugar reform (2006-2010), information from Wirtschaftliche Vereinigung Zucker e.V. (WVZ) and Verein der Zuckerindustrie e.V. (VdZ), Bonn (`http://www.zuckerwirtschaft.de/1_3_2_1.htm`) and KWS SAAT AG, Einbeck (`http://www.kws.de/ca/fh/thd/`)

COCO1: Milk

- Market balances for casein and whey powder were only available on EU level from ZMP, Bonn, which was closed down in 2009.
- DG Agri partly completes gaps in Eurostat series and offers this consolidated database for download. This is used to close gaps in gams/coco/coco1_eurostat.

COCO1: Producer prices for cotton

Import unit values for cotton seeds, cotton lint, flax and hemp are additionally selected from COMEXT.

COCO1: Expert data

Data from experts, which will overwrite all Eurostat data, is included for special issues for some Member States (e.g. grass yields for the Netherlands).

This also applies at the moment for all Norwegian input data such that Eurostat data are currently ignored. However, as Eurostat completeness has also improved on Norway, this procedure might be reconsidered in the future.

COCO1: Land use data

The raw data on land use are currently prepared outside the CAPRI system. Source code and input files are available at EuroCARE, Bonn (R:/Coco_input/land_use). Relevant (raw) information is stored in dat/coco/landuse_data_and_sets.gdx. The data base comprises information on land use classes from various sources, which are again partly discontinued but useful for the early years:

- REGIO - Eurostat, land use, REGIO domain(NUTS2 level - yearly, 1984-2014)
- ENVIO - Eurostat, land use, env_la_luc1.xls (MS level - 1985, 1990,1995, 2000)
- LANDCOVER - Eurostat, land cover(MS level – 2009, 2012, 2015)
- Corine Land Cover (CLC), 44clc_nuts2.xls (NUTS2 level - 1990, 2000, 2006, 2012)
- FAO - area.xls(MS level - yearly, 1984-2016)

- MCPFE (Ministerial Conference on the Protection of Forests in Europe), jointly published by FAO and UNECE (MS level - 1990, 2000, 2005, 2010, 2015)
- FSS - Eurostat, FSS(NUTS2 level - 1990, 1993, ..., 2007, 2010, 2013), only added in coco1/landuse
- UNFCCC (1990-2016), also covers land transitions and settlement data. Official data for LULUCF accounting, merged with other data in coco1_landuse.

COCO2: Economic data

- Eurostat: Economy and Finance, Exchange rates, Bilateral exchange rates, Euro/ECU exchange rates. Data is already prepared in Excel for premature introduction of Euro in price data from the International Labour Organisation (ILO).
- Eurostat, population. To complete early years data from and old Eurostat domain (AGRIS, Population) are also loaded.
- GDP price index expressed in Euros

COCO2: Expenditures

Consumer expenditures on food items are included from:

- Eurostat: Old domain SEC2 for data up to 1997 (HIST)
- Instituto Nacional de Estadística m(INE): Anuario de Estadística Agroalimentaria (AEA), Consumer expenditure on food items in Spain close to HIST definitions up to 1996
- Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI): Consumer expenditure on food items for DEW 1985-92 in Mio DM
- Statistisches Bundesamt Deutschland (SBA): Weighted average of expenditure shares in German household types 2 and 3 (1985-91)
- Eurostat, Final consumption expenditure of households by consumption purpose (COICOP 3 digit)
- United Nations Statistics Division (UNSTATS): Household consumption expenditure in USD
- Eurostat, PRICE: Consumer expenditure weights are used as indicators for budget shares
- Eurostat: Economy and Finance, GDP and main components, Final consumption expenditure of households: Total private consumption of households in current prices (Table “a_gdp_c”)

COCO2: Consumer food prices and consumer food price indices

Food price indices from:

- Eurostat, PRICE, 2005=100.
- Several national sources for western Balkan regions
- Eurostat: Old domain FOOD of section AGRICULTURE: Aggregate food price index with old Eurostat methodology and base 1985
- INTERNATIONAL LABOUR ORGANIZATION Geneva (ILO): LABORSTA Labour Statistics Database, retail prices of selected food unit, prices indices of selected food unit, discontinued after 2008

- Eurostat: Detailed average prices – 2008 - 2015 [table prc_dap15] is used to extend the ILO consumer price series.

COCO2: By-products

- FAO: Food Balance Sheets, Commodity Balances, Livestock and Fish Primary Equivalents: Imports and exports quantities for fish meal, dried cassava, gluten deed and meal, as well as feed quantities for fish meal.
- Eurostat: Purchase prices for fish meal, dried sugar beet pulp, soya cake, and wheat bran
- Eurostat: data (at most up to 2010) from discontinued tables (“food_in afeed1” and “bilares”) on production of feedingstuffs and availability of feedingstuffs
- FAO: Food Balance Sheets, Commodity Balances, Crop Primary Equivalents: Milled rice and total sugar unit value
- Netherlands Economic Institute (NEI): Purchase prices for sugar, calculated by the average of Intervention Price and CAOBISCO price

COCO2: Milk Products

- Zentrale Markt- und Preisberichtstelle (ZMP): Producer prices of selected milk products (only available for some countries)
- Agrarmarkt Informations-Gesellschaft mbH (AMI): AMI-Marktbilanz Milch 2011 (only available for some countries)
- DG AGRI (Réponses au questionnaire (art. 8 du Règlement (CEE) n° 536/93), (art. 15 R 1392/2001) and (art. 26 R 595/2004)): Data on direct sales of raw milk and farm processing in DG AGRI definitions for quota administration

COCO2: Others

- Eurostat: External trade, External trade detailed data, COMEXT, EU27 Trade Since 1988 By CN8, Reporter EU15: Auxiliary trade data for wheat, soft wheat and durum wheat, export values and quantities for cotton and cotton seeds, data on imports and exports of most relevant by-products
- Statistisches Jahrbuch ueber Ern., Landw. U. Forsten, 1999, 2006 und 2010 (Aufkommen u Verbrauch von Futtermitteln): Net imports and feed from domestic production of by-products in Germany
- USDA: Prices for soya, rape and sunflower cake and oil, prices for corn gluten feed

3.2.3 COCO1: Overlay from various sources

The main program coco1.gms starts with a number of declarations of sets and parameters to handle the collection and overlay of “raw data”, often given in a classification different from the target one (sets COLS, ROWS).

A recurrent characteristic of COCO is to solve the problem: if the first best source has gaps in a particular country, or even is entirely empty, select the second or even third best source to fill the gaps.

```
*   General SET declarations
------------------------------------------------------------------ 12 line(s) not disp
$INCLUDE 'coco\Coco1_Sets.gms'
------------------------------------------------------------------ 1 line(s) not displ
*   user supplied settings, for example: sets T, TT1... but also globals like %RESDIR%
$INCLUDE 'RunSetsCoco1.gms'
```

Including standard and supplementary data from Eurostat ('coco1_eurostat.gms') The main program coco1.gms proceeds by importing data from Eurostat prepared beforehand (in coco_input.gms). The main data (on p_agriProd, p_ecoAct, and p_agriPri) are processed step by step and corrections made on selected data for all MS[2].

```
*               - Collecting stadard Eurostat data
*               - Adding non-standard data for several MS (UNFCCC, DG AGRI, COMEXT)
*               - adding special (national) sources for problem cases
*               - Fixing (most) evident scaling, classification, definition problems in particular or several MS
----------------------------------------------------------------- 4 line(s) not displayed ----------------------
$include 'coco\coco1_eurostat.gms'
```

Data from FAOstat ('coco1_fao.gms') The general fall-back option for missing data is FAOstat which requires a few corrections compared to the standard mappings in the context of module "global database", including:

- Rebooking of "other use" to processing (PRCM) or other balance positions
- Disaggregation of olives (table olives, olives for oil), grapes (table grapes, grapes for wine), wheat (common, durum)
- Checks for data changes after sugar reform 2006
- Country specific fixes like in coco1_eurostat.gms.

```
*      FAO market and commodity balances and levels
*               - Load FAO data
*               - Fixing (most) evident scaling, classification, definition problems in particular or several MS
*
*---------------------------------------------------------------
$include 'coco\coco1_fao.gms'
```

Data from additional sources for the Western Balkan Countries and Turkey ('coco1_croatia_data.gms' and 'coco1_candi_AgriProd_AgriPri.gms') Croatia is the first country singled out from the special data input for the Western Balkan Countries and Turkey. Croatia is by now mostly sourced from Eurostat, as the other EU members, but a few supplementary expert data have been retained. For the other Western Balkan regions and Turkey, *'coco1_candi_AgriProd_AgriPri.gms'* further adapts the WB data from the country specific xls files to match the COCO definitions that also apply to EU28 countries (on parameters p_agriProd and p_agriPri).

[2]Eurostat offers data for Belgium and Luxembourg separately, whereas the database combines both countries to the model region "BL000" (Belgium and Luxembourg). The key reason is that Eurostat offers data mainly for the aggregate Belgium and Luxembourg up to the year 1999, especially for all market balances. Furthermore, Luxembourg has a rather small agricultural sector (2004 total output was about EUR 250 million) with some similarities to Belgium.

```
*       Special data for Croatia
*                 - Expert data for early years (before 2005)
*                 - Data from statistical year books
*
*---------------------------------------------------------------
$ifi %ms% == HR $include 'coco\coco1_croatia_data.gms'
--------------------------------------------------------------- 8 line(s) not dis
*   Include additional Western balkan data and data adjustments (also applied to TR):
*
$include 'coco\coco1_candi_AgriProd_AgriPri.gms'
```

The include file handles the following:

1. Similar to EU-28 MS there are many case-by-case adjustments correcting different scaling and definitions (live weight <-> carcass weight, reaggregations for wine and fruits…).
2. In many cases, market balances are simply incomplete. As a fall back solution, domestic demand is calculated from production and net trade and disaggregated with shares taken from a sister country aggregate (Romania, Bulgaria, Greece, Slovenia, Hungary). Other corrections with "borrowed" information are:
   (a) Trade data are frequently missing in the WBs, such that FAO data are included where available.
   (b) Production of oilcakes and sugar is estimated from raw products, if missing, using the sister country aggregate processing coefficients;
   (c) The production of milk products is estimated from processing coefficients in Serbia which has a quite complete series;
3. Price information is also completed relying on the sister country aggregates.

Final completions and revisions for all Member States ('coco1_finish_agriprod.gms') Based on the availability of second and third best options various finalising steps are applied to the quantity data. It should be noted that the CAPRI database tries to estimate market balances (needed for separate behavioural function for feed, food, processing, biofuel demand) in spite of Eurostat discontinuing the publication of market balances for most products since 2014. For this purpose the old Eurostat market balances are still loaded and combined with more recent production data. This triggers the need for data completions and estimations in the most recent years (which are also most critical for projections). In 2019 market balance data have returned to the Eurostat server for cereals and oilseeds, but only for a single year (2017) => It is likely that adjustments like the following will also be needed in the future:

- Completion of production data from the (discontinued) Eurostat market balance statistics (model code "USAP") with quantity information given from the production statistics (code "GROF") or from agricultural account statistics (model code "EAAQ") using a correction factor calculated from overlapping years.
- Additional gap filling using FAO data for special cases and general cases of missing data (e.g. for balances). An additional difficulty is that FAO commoditiy balances are currently (2019) also ending in 2013 (especially valuable for recent years).

- Domestic use can be calculated (under some conditions) from imports, export and usable production. If only domestic use is given for some products, the sub-positions, such as industrial use, processing, human consumption, feed on market, total seed and total losses are allocated with the average shares in data for other years, from the same country. As a fall back solution, the average shares from other countries are used.
- For the milk products whey powder and casein, the disaggregation of demand is mainly based on EU data collected by the German “Zentrale Markt- und Preisberichtstelle für Erzeugnisse der Land-, Forst- und Ernährungswirtschaft GmbH” (ZMP) and some auxiliary assumptions.
- As data for oilseeds are critical for all countries, the implied processing coefficient is checked for plausibility. If the national coefficient is lower than 60% or above 150% the average coefficient for all EU-15 MS, the data for usable production of the country are corrected by multiplying the processing data with the average EU-15 coefficient. Domestic use and all sub-positions are subsequently re-calculated.
- Some additional calculations to prepare the use of animal herd data in coco1_anim:
  - Some calculations to combine FAO and FSS data on poultry herds
  - Completions acknowledging seasonality in cattle and sheep and goats herd countings
  - Aggregations and residual calculations to the COCO animal categories from animal types in Eurostat (say “Heifers for raising, 1-2 years”)

The file handling the previous actions is ‘coco1_finish_agriprod.gms’:

```
*      Final completions and revisions to finish agric production for all MS before assigning to DATA
*                - complete USAP with GROF or EAAQ or milk production statistics
*                - use FAOstat for problem cases (eg fruits, poultry) in Eurostat and for missing data
*                  (rule considers the number of obs and internal consistency in Eurostat vs FAOstat)
*                - Completion of DOMM and components (if USAP and trade is given)
*                - Complete secondary milk prods (for whey powder and casein based on German ZMP)
*                - Check+complete+revise oilseeds/oils balances if processing yields are messy (bad for biodiesel)
*                - Special completions for animal survey data, partly considering seasonality
*                - Read biofuel data and convert to standard AgriProd format
*
*---------------------------------------------------------------
$include 'coco\coco1_finish_AgriProd.gms'
debug("%MS%000",ass_cols,ass_rows,tt1,"%system.fn%_%system.incline%") =  p_AgriProd("%MS%000",ass_cols,ass_rows,tt1)
                                                                      +  p_EcoAct("%MS%000",ass_cols,ass_rows,tt1)
                                                                      +  p_AgriPri("%MS%000",ass_cols,ass_rows,tt1);
* execute_unload "%results_out%\coco\debug\debug_%ms%" debug;
*$batinclude "util\debug.gms" %system.fn% %system.incline%
*$EXIT
```

The previous code snippet also shows for the interested reader two frequently used debugging devices:

1. The key parameters at a certain point in the program flow (above: p_agriProd, p_agriPri, p_ecoAct) are copied to a debugging parameter “debug” (better name would be: “p_debug”). At the end of a coco1 run (or if desired also at this point) the parameter is unloaded into a file “results/coco/debug/debug_%MS%.gdx” such that the various assignments, corrections, deletions that have occurred up to a certain program line may be inspected in one file.
2. The next command “$batinclude “util/debug” %system.fn% %system.incline% unloads the whole memory, incuding all parameters but also sets and other symbols, at this point into a debugging file in the gams/temp folder. This may be useful to analyse “difficult” cases of debugging.

Finally the biofuel sector is prepared.

EU biofuel sector data ('coco1_finish_agriprod.gms' and 'prepare_biofuel_data.gms') The first issue to note is that market balances for sugar beet and sugar are compiled in such a way that all biofuel use of beets is converted into biofuel use of sugar, as if the beets were first processed to sugar and only then converted to ethanol. The advantage of this approach is that sugar is part of the market model and thus may enter the behavioural functions for biofuel feedstock use whereas beets only exist in the supply part of CAPRI. A second advantage is that biofuel feedstock use was indeed booked under sugar in some MS and under beets in others such that our approach ensures a standardisation of booking principles.

Biofuel production

There is no differentiation made between fuel- or non-fuel (undenatured or denatured) quantities in production, import and export positions of ethanol. But the consumption position of ethanol is differentiated in fuel-ethanol consumption and non-fuel-ethanol consumption. Hence data on fuel and non-fuel production and consumption of ethanol was required. In the case of biodiesel this differentiation is irrelevant. The ex-post data on biofuel production are coming from diverse sources which is unavoidable to complete the data for years as of 2002 up to the present, if necessary with the help of second and third best solutions or assumptions (compare *biofuel/prepare_biofuel_data.gms*).

The overlay considers data availability and consistency across sources:

- For ethanol we consider DG agri as the first best source as it does not only cover production and demand, but also a break down by feedstocks (cereals, beets, wine, fruits, potatoes, other).
- Some countries (Croatia, Turkey, Bulgaria, Romania, Serbia) are supplemented from other sources (AGLINK-COSIMO, USDA, Eurostat PRODCOM). AGLINK also supplements production other than from agricultural feedstocks.
- Eurostat PRODCOM, Energy balances and PRIMES serve to extrapolate or backcast the DG Agri information to years with missing data.
- Ethanol trade by MS is taken from COMEXT but scaled to be in line with DG AGri data for the whole EU.
- Production of biodiesel is usually from the energy balances while trade is from COMEXT. If data are complete and results reliable, demand is computed residually. In cases of missing data or implausible results, demand is taken from Energy balances, PRIMES, or the EloBio project and trade is calculated as a residual with some rules.

Feedstock demand

In addition to market balances for the fuels the CAPRI data base requires the shares of the raw products on the production of biodiesel and bioethanol at the level of CAPRI products. For bioethanol, this information is partly provided by the DG Agri balances, hence this has been selected to be the major source. The detailed recording follows from the existence of support measures for distillation of wine, fruits and potatoes which triggered a detailed monitoring of ethanol markets. However, for biodiesel the statistical sources are scarce. It turns out that the most consistent estimates for EU regions are apparently produced by USDA services, covering rape, sunflower, soya, palm oil but also used cooking oils, tallow and other oils. As these data do not cover single MS an estimation procedure has been devised (in *biofuel/calc_feedstock_shares.gms*). The initialisation of this estimated feedstock composition relied on the observed increase in INDM according to Eurostat (or more precisely the COCO initialisation when entering *'prepare_biofuel_data.gms'*) which is assumed to be the main source to "cut out" the required biofuel processing quantities (BIOF) by MS from market balances that so far did not include BIOF.

A special case was palm oil, as the CAPRI database (COCO) doesn't cover an industrial use position for this product so far. EUROSTAT-COMEXT delivers data on import and export quantities of crude palm oil (HS 151110) for EU Member states. Thereby an increase of palm oil imports was observed within the relevant ex post period (2002-2005). Thus the following assumptions were made to derive approximated values for palm oil processing to biodiesel: (a) Import quantities minus export quantities are equal to domestic consumption of palm oil as domestic production in European Member states can be neglected. (b) The average aggregated consumption quantity of palm oil before 2002 was assumed to be completely used for human consumption as no significant biodiesel consumption took place. By subtracting this constant share of human consumption from the observed consumption quantities after 2002 gave an estimate for the quantities used for industrial processing

Given that many data sources are combined and several aggregation conditions should be maintained, it turned out necessary to set up a small optimisation problem with the following properties (see towards the end of *'prepare_biofuel_data.gms'*):

- The estimation tries to stay close to the initial feedstock composition
- Extra terms penalise deviations from DG Agri (first best souce for ethanol) and implausibly high shares for palm oil
- Technical conversion coefficients (see below) link standard feedstock use and estimated production which has to aggregate with non-standard feedstocks (NAGR) to total production of biofuels. Non-standard feedstocks are those not endogenous in the CAPRI market model (potatoes, fruits and other for bioethanol, used cooking oils, tallow and other for biodiesel)
- Total domestic use (with data modifications heavily penalised in the objective) is consistently broken down into biofuel use, other industrial use and non-industrial (e.g. food) use to avoid disturbing the initialisation in previous include files based on Eurostat data.

Technology parameters

Conversion coefficients for 1st generation biofuels were collected from different sources. The AgLink-Cosimo model includes a set of conversion coefficients which are in line with the CAPRI product definitions and have become the main source for CAPRI. The table below displays the set of conversion coefficients used for 1st generation biofuels and corresponding by-products.

Table 5:Conversion coefficients for 1st generation biofuel production

Conversion coefficients (t/t)		Ethanol	Byproducts
Grains	Wheat	0.274	0.266 DDGS
	Barley	0.247	0.266 DDGS
	Oats	0.247	0.266 DDGS
	Rye	0.247	0.266 DDGS
	Corn (dry milling)	0.335	0.292 DDGS
Other	Table Wine	0.100	
Sugar Crops	Sugar	0.517	
	Sugar beets	0.079	0.004 Vinasses*
		Biodiesel	Byproducts
Veegetable oils	Rape oil	0.922	0.100 Glycerine
	Soy oil	0.922	0.100 Glycerine
	Sunflower oil	0.922	0.100 Glycerine
	Palm oil	0.922	0.100 Glycerine

Note: The beet coefficient has been increased in the meantime from 0.079 to 0.086.

Fuel prices and taxes

For a specification of processing-, biofuel supply- and demand-functions in the base year, ex post prices are required. Furthermore, given the structure of the CAPRI market module (described in Section 5.4), a differentiation of producer, consumer and import price is also needed. these differentiated prices are not covered in any statistical database for biofuels but they can be derived indirectly by given information on taxes, tariffs and subsidies from the world market price which is available. thus beside ex post prices information on consumer (excise) taxes, import tariffs and further subsidies are required. the aglink-cosimo database includes ex post world market prices for ethanol and biodiesel. this price was taken as the base value to calculate the differentiated prices in the respective countries. the import tariffs for ethanol and biodiesel were also taken from the aglink-cosimo database. as the consumer taxes for ethanol and biodiesel in most instances correspond to a reduced excise tax on fossil fuels the consumer taxes for gasoline and diesel were taken as a base value. this tax information was acquired from euractiv[3] where levels of diesel and petrol taxation in 2002 are published for european member states. for the required time period (2002-2005) taxation levels were calculated with respect to com(2002)410[4] which set minimum excise tax rates for non-commercial diesel and petrol since 2006. to identify the excise tax exemptions and producer subsidies, if existent, for the single member states the obligatory 'member states reports on the implementation of directive 2003/30/ec of 8 may 2003 on the promotion of the use of biofuels or other renewable fuels for transport' were consulted which are published by the commission[5]. Three different types of tax regulations for biofuels were identified which are applied among the different Member states: an absolute tax for biofuels, an absolute reduction of the excise tax on fossil fuels and a relative reduction of the excise tax on fossil fuels. All differentiated in taxation for blended biofuels or pure biofuels. Based on this information the different ex post prices for the period 2002-2005 were

[3]fixme `http://www.euractiv.com/en/taxation/fuel-taxation/article-117495`, 20.07.2009.

[4]proposal for a council directive amending directive 92/81/eec and directive 92/82/eec to introduce special tax arrangements for diesel fuel used for commercial purposes and to align the excise duties on petrol and diesel fuel (com(2002)410).

[5]`http://ec.europa.eu/energy/renewables/biofuels/ms-reports-dir-2003-30-en.htm`.

recalculated. As the envisaged biofuel demand function will be a function of (among other variables) the relation between fossil fuel consumer prices and biofuel consumer prices the acquisition of fossil fuel prices was required additionally. To hold consistency between the biofuel and fossil fuel prices the price information for fossil fuels were also taken from the AgLink-Cosimo database which provides EU market prices for diesel and petrol. For the recalculation of consumer prices in individual Member states the already collected taxation levels for fossil fuels were applied. Because there exists a significant difference between the physical energy content and the density of biodiesel, ethanol, petrol and diesel a direct comparison of prices (in €/t) is not possible. For this reason the prices as well as the taxation levels were converted into Euro per ton oil equivalent (toe).

Assigning data to database array So far data processing has focussed on the key Eurostat Table Group “apro” (collected on parameter p_agriProd). The next parts of COCO will collect data from other sources, including the other two Table Groups for prices and Economic accounts (“apri”, “aact”) to a single GAMS array “data”. This data collection activity happens in files coco1_expert.gms to coco1_eaa.gms with a summary of the details given below.

```
*    Assignment of Expert Data
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_expert.gms'
---------------------------------------------------------------- 7 line(s) not displayed ----------
*    Crop levels + market balance positions (for ALL products except dairy) from Agric Production
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_crops.gms';
---------------------------------------------------------------- 6 line(s) not displayed ----------
*    Balance sheets for milk production and dairy industry
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_milk.gms';
---------------------------------------------------------------- 7 line(s) not displayed ----------
*    Parameter initialisation for herd sizes for animals
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_anim.gms';
---------------------------------------------------------------- 5 line(s) not displayed ----------
*      Assign AgriPri data to array DATA after checking for evident nonsence
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_assign_AgriPri.gms'
---------------------------------------------------------------- 6 line(s) not displayed ----------
*      Estimate an EAA for Balkan countries and TR
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_candi_EcoAct.gms';
---------------------------------------------------------------- 7 line(s) not displayed ----------
*      Economic accounts for agriculture based on Eurostat
---------------------------------------------------------------- 2 line(s) not displayed ----------
$include 'coco\coco1_eaa.gms';
```

Include file ‘*coco1_expert.gms*’

This file collects expert data for specific countries that receive priority over all other data sources in the initialisaiton. The most relevant case is Norway where nearly all data are provided and checked by NIBIO (Norwegian Institute of Bioeconomy Research).

Include file ‘*coco1_crops.gms*’

This sub-module assigns the areas, crop production data and most market balance positions from Eurostat’s Table Group “apro” . However, it is necessary to first deal with a double counting in the land use statistics of Eurostat with cotton both counted among textile crops as well as oil crops. This is fixed by having the aggregate activity “textile crops” producing both other oilseeds (i.e. cotton seeds) as well as textiles (here cotton lint) and removing cotton from the other oils area.

After this special case the crop areas from Eurostat’s production statistics are copied to the LEVL position of the “data” array. Data from Eurostat’s land use statistics are the second best choice in case of missing areas.

Inappropriate aggregation (ignoring gaps in the component series) has been frequently observed in past experiences with Eurostat data such that aggregates are added up, if possible, from any given sub-components. This principle applies to “GRAS” (permanent grass land = meadows PMEA+ pastures PPAS), and some other aggegates.

In terms of gross production (GROF) it has to be mentioned that preference is given to the market balance information “USAP” over the production statistics “GROF”), as the former may be expected to be consistent with the trade and demand positions. Thus we set (considering the time lag between balance data and production statistics):

For products with market balance $DATA(GROF,t) = p\_agriProd(USAP,t+1)$
Remaining products $DATA(GROF,t) = p\_agriProd(GROF,t)$

Some special assingments handle SEDF and LOSF for cereals and the residual calculation of production of “OOIL” starting from oil crops (OILC).

More important is a procedure to ensure a complete initialisation of fodder production quantities, an area with widespread gaps in the raw data. This procedure estimates fodder yields (of “PMEA”, “PPAS”, “TGRA”, “FCLV”, “FLUC”, “FPGO”, “FAGO” and “MAIF”) from the relationship of known fodder yields to those in other EU countries. To ensure completeness, cereal yields are also considered such that fodder yields may be estimated, in the worst case, from the fodder yields in other EU countries, corrected by the ratio of cereal yields in the MS under consideration to EU cereal yields.

Contrary to the program name, *all* balance positions for crops *and animals*, except milk positions, are assigned to the “data” array in ‘*coco1_crops.gms*’. Specific treatments are necessary for fruits, table grapes and olives for oil and residual calculations anre undertaken for missing human consumption, total domestic use, and usable production.

In several cases upper or lower limits are assigned for qunatities and areas where it turned out that missing data are often completed in the optimisation part of COCO in an unsatisfactory way. The empirical basis for these limits is diverse. It may rest on production statistics (if production is given there but missing in the market balances), on sugar quotas for the sugar beet sector, or in some cases (fruits, vegetables) on a moving average over given observations.

Include file ‘*coco1_milk.gms*’

This file assigns the data for dairy products and raw milk from Eurostat’s “apro” Table Group, with some re-aggregations and additional lower and upper limits for the optimisation parts of COCO1.

Gross production of raw milk is usually given from the farm balance data (COMI = CMLK, cow milk + BMLK, buffalo milk. SGMI = EMLK, ewes milk + GMLK goats milk).

Gaps are more frequent for deliveries to dairies (“PRCM”) which are preferably derived from the aggregate processing volume of raw milk according to farm balance (to ensure consistency with gross production) or, as a second best solution added up from the components in the dairy collection data (e.g. collection of CMLK, BMLK, EMLK, GMLK). Often there are also data to disaggregate the non-delivered parts of raw milk into direct sales (e.g. HCOM.COMI), feed use (INTF.COMI), use for farm cheese, butter and other processing products (INDM.COMI) and finally losses and home consumption of liquid milk (LOSM.COMI) and to identify on farm production (e.g. FARM.CHES).

Whereas production data and deliveries to dairies may be distinguished into "COMI" and "SGMI", the dairy statistics on derived products obtained or associated market balances do not permit such distinction. As a consequence, the dairy sector is treated as if all raw milk from cows, sheep etc. was collected and merged into single raw milk at dairy ("MILK"). The marketable production for this aggregate milk, at the dairy level, is set to the sum of the processing volumes from cow and buffalo milk, sheep and goat milk (from the farm balance). Finally, the balance sheets for the secondary milk products are usually taken from the "apro" data selected from Eurostat.

The content of milk products is initialised using two types of information: statistical data on fat content of dairy products (and protein content for raw milk) and default technical coefficients for the content of milk products, in terms of milk fat and protein (this is the only initial information for protein, apart from raw milk, where statistical data on protein content are available). The initial information on the fat content of dairy products is rendered complete and reliable by discarding statistical information on contents that are implausibly far away from standard technical coefficients.

Include file '*coco1_anim.gms*'

Assigning herd size, process length, activity level, yield and production data often requires significant reaggregations from the slaughtering statistics and therefore explanations in this documentation:

The first best source for tons of slaughtered meat of the main animal categories (SLGT.IPIG, ILAM, ICAT and ICHI) is the usable production (USAP) from the balance sheets because this is likely to be consistent with market balances. As a second best source we use the slaughtering statistics, but with a correction factor. Export and imports of live animals expressed in carcass weight are partly taken from the slaughtering statistics or from the balance sheets, depending on availability. It is useful to remember that total production of meats in heads (e.g. "GROF.IPIG") is set equal to the sum of all slaughtered heads plus exported heads minus imported heads. Accordingly, the production of meat in tons equals the sum of slaughtered tons plus exported tons minus imported tons.

Herd size data are initialised based on the data prepared in '*coco1_finish_agriProd.gms*', taking an average of the available countings related to a calendar year. In the cattle sector we take the weighted average 0.25*December(t-1)+0.5*May-June(t)+0.25*December(t) to assign the average herd size in the calendar year. For dairy cows and suckler cows this average herd size this is also the activity level. The input coefficient for dairy cows ("DCOW.ICOW") and suckler cows ("SCOW.ICOW") reflects the number of slaughtered heads (of cows), in relation to the total herd size of cows with a fall back value in case of missing data of 0.2. The slaughter weight of cows is cows' meat production divided by slaughtered heads. A particularity is the culling of cows in the UK due to the mad cow disease, because culled cows do not show up in the slaughtering statistics and yet they have top be considered for reasonable replacement rates. This is solved by estimating the total killings of cows (near zero slaughterings + cullings GROF.ICOW) in the period 1996-2005 from typical replacement rates in the pre-crisis period and booking the estimated cullings on losses (LOSF.ICOW for heads, LOSF.BEEF for tons of culled cows).

For cattle other than cows the activity level definition is more complex. In the case of heifers and bulls for fattening, the activity level equals the number of slaughtered heads plus net exports of live animals. If slaughtered heads of heifers and bulls are unavailable, 45% of total cattle slaughterings (net of cow and calves if available) are used as a default value. Heifers for raising will be used to replace dairy and suckler cows, therefore the number of raised heifers (activity level) may be recalculated from cows slaughterings and the change in the cows' herd size over the next two years.

In the same manner the number of heifers needed as input (GROF.IHEI) for each year is equal to the activity levels of heifers for raising and heifers for fattening. The number of female calves raised (activity level) in the current year is equals the number of heifers used as inputs in the following year. Similarly the number of young bulls raised equals next year's production of adult male cattle in heads. In countries with complete statistical data there are only two activity levels that cannot be fully inferred from statistical data alone: As the statistics do not distinguish slaughterings and trade of male and female calves we are using a male share of 51% to estimate the split of male and female calves. This also permist to calculate the total number of calves of each sex needed as input for each year as calves for raising plus calves for fattening and correspondingly the output coefficient of cows. Conversely the output coefficients of calves in terms of beef may be calculated from statistical data on slaughtered calves in tons and heads.

Herd size data usually may be mapped exactly to particular cattle categories in the CAPRI data base, including the distinction of heifers for raising and for fattening. The only exception is the distinction of the herd size of male and female calves which is assigned according to the estimated split in the related activity levels. Having assigned both the herd size as well as activity levels permits to assign: average process length in days = activity level / herd size * 365. The average process length in turn is related to the daily growth of animals according to another accounting identity: final (live) weight = beginning (live) weight + daily growth * (process length – empty days). This accounting identity will be imposed in the COCO1 estimation procedure, but module coco1_anim assigns bounds (parameters UppLim and LowLim) for the process length such that the implied daily growth values remain in a reasonable range. For heifers there is also an upper bound for the process length for statistical reasons: female animals older than 36 months are classified as "cows", whether they have calved or not.

Activity levels and slaughter weights for animal types other than cattle are more straightforward to obtain. The herd size of fattenened pigs beyond 20 kg, of piglets up to 20 kg and sows (+ boars) is the average number according to the four possible annual counting (April, May/June, August and December). The number of fattened pigs (flow of animals) equals total slaughtered pigs minus slaughtered sows. The output coefficient (piglets) per sow equals the number of slaughtered pigs plus the increase in the sows herd size. The input coefficient is an estimate of sows slaughterings per sow (inferred from stock data on young sows and the stock change of all sows). The production of pork from pigs for fattening is calculated from total meat production less the pork from sows, assuming that a sow produces 120 kg of meat.

Two particularities in the pig sector are worth mentioning. The first is that as of 2011 the COCO database includes the herd size of piglets < 20kg (on code PIGL00.HERD) even though there is no explicit activity level "raising of piglets". Instead the piglets raised are one of the outputs of activity sows with total production of piglets given on code GROF.YPIG. Accordingly we cannot store the process length for raising of piglets in a column for "raising of piglets" but introduce a new code "PIGF.YDAYS" such that in the completed data base we find the relationship PIGF.YDAYS = GROF.YPIG / PIGL00.HERD * 365. Including the piglets turned out useful because it permits to make use of statistical data on the total pigs population which is sometimes available even though pig slaughterings in heads are missing.

The second pig sector particularity relates to the requirement functions for pigs, stored in the form of a table (*/dat/feed/porkreq.gms*) that relates daily growth to final slaughter weights. For consistency reasons the same table is used to define bounds for the permissible process length.

In the poultry sector we have herd size data for chicken broilers, turkeys, ducks, and geese (yearly average, mainly from FAO) and hens from Eurostat (average of this and last year's December counting). The first four give the total herd size of poultry for fattening whereas the herd size of hens also equals the activity level. The output coefficient for eggs relies on usable production from the balance sheets divided by the herd size of hens. A replacement rate of 80% is assumed for laying hens. The activity level of poultry fattening is the difference of total produced poultry heads minus slaughtered hens. The output coefficients and production in terms of meat are straightforward to calculate from here. With activity level and aggregate herd size of poultry for fattening being defined it is possible to calculate the implied process length. The information on the shares of chicken broilers, turkeys, ducks, and geese is used to specify technical bounds for the daily growth and process length. In addition the technical literature also permitted to specify typical empty days for cleaning of stables (or seasonality in the case of geese and ducks). The differentiation of poultry for fattening is only maintained temporarily in COCO1 because it helped to use statistical information for the specification of some technical coefficients that strongly depend on the shares of turkeys. Subsequent CAPRI modules (like CAPREG) will only use the COCO results for the aggregate poultry fattening activity (POUF).

The herd size data for sheep and goats are assigned in the same way as for cattle. The herd size of sheep and goats for milk is at the same time the activity level. The number of slaughtered lambs (sheep and goats) is the total slaughtering number (including net exports of young animals) minus the slaughtering of adults. This estimate for slaughtered lambs in heads also defines the activity level of sheep and goats for fattening. The total output in tons set equal to the meat production. A particularity in the sheep and goat sector is the strong seasonality in some countries. Empty days are specified based on the share of the December counting (sheep in continuous systems) to the May-June counting (sheep in seasonal + continuous systems). These enter the specification of bounds for the process length in sheep and goats fattening.

Include file '*coco1 assign_AgriPri.gms*'

Before assigning the prices from p_agriPri to the tareget parameter data 3 issues are addressed:

- Price differences in the original series between MS suggested that not all series have been already expressed "per nutrient".
- Prices for dairy products CHES and COCM need aggregation from more specific series
- Outliers are identified according to limits for plausible differences to the EU average

Include file '*coco1_ candi_EcoAct.gms*'

Except for Macedonia, which reports EAA data to Eurostat, all other candidate countries receive an EAA initalisation from previously assigned GROF times PRIC. Input positions are assigned based on shares borrowed from an average across selected EU MS.

Include file '*coco1_eaa.gms*'

In this file EAA data from Eurostat are assigned from parameter p_ecoAct to data(.), including unit values. For a number of aggregates special assignments are needed to obtain monetary values matching with the aggregates used elsewhere in COCO.

Unit values at producer price are preferably calculated as a quotient from the value at producer price and the quantity as selected from the EAA statistics. However some checks are used to discard grossly implausible (outlier) unit values.

To serve as a fall back option for the EAA unit values, the previously assigned prices from the p_agriPri parameter are corrected to acknowledge the typical differences between producer prices (UVAP) and selling prices (PRIC). Finally, if price indices are still missing for single items, those from product groups are used.

Prices for energy positions heating gas EGAS and fuel EFUL may be used to infer quantity variables in CAPREG from value information. A special section takes care for completeness.

Finally production of non-physical items from the EAA (some outputs like NURS, FLOW and inputs other than heating gas EGAS and fuel EFUL) may be calculated by the quotient of EAA value and a price index. As we will also express the output "quantity" for heterogenous items "other industrial crops" (OIND), "other crops" (OCRO) and "other animal products" (OANI) in values at constant prices (currently 2005), the complete list of non-physical items with quantity information given as values in constant prices is (using the codes from the end of this documentation):

Outputs: NURS,FLOW,SERO,RQUO,NASA,OIND,OCRO,OANI.

Inputs: IPHA,WATR,REPM,REPB,ELEC,ELUB,INPO,PLAP,SEED,SERI.

With coco1_eaa.gms passed, the presumably best raw data are collected on the central parameter data(.), but a few additional completions are possible to improve the internal consistency of the initialisation before proceeding to the main consolidation steps:

```
*      Further assignments for market balances: (1) SEDT/LOST+components, (2) PRCY, (3) STKM
-------------------------------------------------------------- 73 line(s) not displayed ----------------
*      Calculate positions for residuals
-------------------------------------------------------------- 2 line(s) not displayed -----------------
$include "coco\coco1_resid.gms"
-------------------------------------------------------------- 5 line(s) not displayed -----------------
*      Crop output coefficients
-------------------------------------------------------------- 2 line(s) not displayed -----------------
$include "coco\coco1_cropyields.gms"
-------------------------------------------------------------- 5 line(s) not displayed -----------------
*      Adjust gras yield with respect to expert knowledge from Oenema/Velthof
-------------------------------------------------------------- 2 line(s) not displayed -----------------
$include "coco\coco1_gras.gms"
-------------------------------------------------------------- 4 line(s) not displayed -----------------
*      read land use information from different sources
-------------------------------------------------------------- 2 line(s) not displayed -----------------
$include "coco\coco1_landuse.gms"
-------------------------------------------------------------- 8 line(s) not displayed -----------------
*      Finish raw data assignments
*               - assing missing UVAPs from EAAP or PRIC for non crops
*               - check and act if one of GROF or LEVL or YILD is missing (gives abort in final results)
*               - check seed quantities and limit to technically reasonable quantities per ha
*               - introduce LowLims and UPPLims for FEDM,HCOM,SEDF,SEDM to limit yearly change
*               - store interim data at end of this file = priors for following optimisation models
* -------------------------------------------------------------------
$include "coco\coco1_finish_raw.gms"
```

Include file *'coco1_resid.gms'*

This file calculates residuals from the given data for aggregates and sub-positions for crops. The residual activity level and market balance position is defined as a difference between the group level and the sum of individual crops. This calculation is not carried out if there are gaps in some components or if the total is smaller than the sum of given components.

Include file '*coco1_cropyields.gms*'

Yields are evidently calculated for each crop activity by dividing the gross production by the production level for this activity. However, this sub-module also applies a Hodrick-Prescott (HP) filter to smooth out problems with yields from activities with small production areas. This optimisation program has tight bounds around observed production and area data (± 100 t or ± 100 ha). The HP objective penalises peaks in the data as frequently encountered (partly due to rounding errors) with small areas or quantities. The tight leeway around observed values is irrelevant for moderately important crops in the sense that the result will be almost identical to the original data. For 'unimportant' crops, however, the HP filter term will lead to some smoothing of peaks in the data and thus, in general, to more plausible yields for these crops[6].

Include file '*coco1_gras.gms'*

In most countries grass is the most important 'crop' in terms of area use yet, often the data on grass areas and production are one of the weakest parts of crop statistics. When relying solely on statistical data, the COCO database frequently showed unbelievable grass yields in some MS. This sub-module assigns grass yields, based on expert knowledge, to be used as priori information together with statistical data in part 2 of the COCO routine. The key information is expert data[7] on typical grass yields in dry matter for 2002 in all EU-28 MS and WBs. To convert this expert information, for a single year, into expert time series for grass yields, the expert data for 2002 are linked to the yields of activity aggregate cereals, assuming that long run yield growth and yearly fluctuations run approximately in parallel. The yields for pasture, meadows and other fodder on arable land are adjusted accordingly.

Include file '*coco1_landuse.gms*'

This file allows to process information from various sources on the same item, in particular areas for various land use items ("LEVL"). In order to handle the different sources, new rows are defined, indicating from which source the information on land use area is coming which is typically only offered for a selected years or a limited period:

- **LEVAgriProd** - Eurostat national land use data (Eurostat Table: "apro_cpp_luse", discontinued). As these data are annually available since the 80s and give important land use categories (total area ARTO with inland waters INLW, arable land ARAC, permanent grassland GRAS, forest land FORE, etc) this would be our preferred source, if all series were complete and reliable.
- **LEVCLC** - Land use levels derived from Corine Land Cover (CLC) using a transformation matrix to LUCAS in two steps

[6]For example in France, in 2000, 100 ha only represented 0.002% of the soft wheat area, but 100 ha of tobacco represented 16 % of the total area, as tobacco is irrelevant in France. These irrelevant items will be those where unrealistic yields will be frequently found and where deviations from Eurostat data will be acceptable.

[7]These were estimates worked out in September 2006 by Oene Oenema and Gerard Velthof from Alterra, Wageningen, in the context of a service contract for DG-ENV (Integrated measures in agriculture to reduce ammonia emissions, No 070501/2005/422822/MAR/C1) with the participation of EuroCARE.

- – Original Corine Land Cover (44 classes, aggregated to the NUTS2 level[8] obtained from JRC, Ispra for 1990, 2000, 2006, 2012. To link the Corine information to the CAPRI land use classes we used as an interim step so-called contingency tables from CLC to LUCAS categories provided by JRC Ispra at NUTS2 level. This allows to map the Corine classes (like complex cultivation patterns – "complexCultiv") to the *most probable* land cover class from the LUCAS survey (in the example "complexCultiv" -> annual crops) which may be aggregated then to the CAPRI land use aggregates (annual crops LUCAS -> arable crops, CAPRI code ARAC). However, while this mapping to the *"most probable"* category in LUCAS preserves the original information as much as possible, it has disadvantages, for example, that certain LUCAS categories like "fallow land" are not mapped at all because they are not the most probable matching LUCAS category for any of the CLC classes.
  – To acknowledge that the Corine Classes may be mapped to several LUCAS categories we multiplied them with the "profiles", giving the distribution of each Corine category according to the LUCAS classes. In this case, only 26.7% of the "complexCultiv" area is mapped to annual crops, but 7.3% are mapped to "temporary pastures", 6.4% to "permanent grassland with sparse tree/shrub vegetation" and so forth. The transformed Corine data often give the most detailed area coverage and thus assume a role as a kind of fall back information in case that other information is missing.

- **LEVRegio** - Eurostat regional land use data (Eurostat Table: "agr_r_landuse", discontinued). Inspite of using the same codes as for the national data, the national totals, aggregated from the NUTS2 regions are not always in line with LEVAgriProd. Furthermore a few categories are missing (no inland waters, no other wooded land). However there are few alternative annual series available to regionalise the national data in CAPREG.
- **LEVFAO** - Land use data from the resource FAOSTAT domain FIXME [9] with annual time series on agricultural land use but also some non agricultural area categories (forest, inland waters, other land, total area).
- **LEVLucas** – directly using the LUCAS data is an option that has been considered but not implemented in CAPRI so this code is not used at the moment.
- **LEVLandCov** - Eurostat land cover data for 2009, 2012, 2015 at the MS level. Agricultural land is only distinguished into cropland CROP and grassland GRAS, but 5 nonagricultural areas are neatly aggregating up to the total country (Artificial ARTIF, shrubland (considered similar to "other wooded land" OWL), bare land & wetlands (mapped to "other sparcely vegetated or bare OSPA) and waters WATER.
- **LEVEnvio** - Eurostat land cover data from the environment section (Table "env_la_luc1" FIXME [10], discontinued). Total area is classified into about 40 categories, but data are only given for a number of years (1950, 1970, 1980, 1985, 1990, 1995, 2000) and with many gaps, in particular for the subcategories.

[8] Data for some countries and years affected by evident problems have been removed. For example the 2006 CLC data only covered parts of Greece, hence are no usable to calculate totals at the MS level.

[9] See `http://faostat3.fao.org/home/index.htmlDOWNLOAD`.

[10] Apparently these data are currently under revision because they are not accessible on the Eurostat website anymore since about June 2012. However they are still accessible (in July 2012) via `http://eu22.eu/land-use.2/land-use-by-main-category/`.

- **LEVMcpfe** – Data from the Ministerial Conference on the Protection of Forests in Europe C&I database for quantitative indicators. This gives validated data on the forest sector (forest land FORE, other wooded land OWL) and some non forestry data (inland waters INLW, total country area ARTO), but data were only given for 1990, 2000, 2005, 2010, 2015.
- **LEVFSS** - Eurostat farm structure survey data (Table "ef_lu_ovcropaa"). Gives a very detailed and reliable description of agricultural area use, but only for the survey years (1990, 1993, 1995, 1997, 2000, 2003, 2005, 2007, 2010, 2013). As CAPRI_regLU these data are also used in the subsequent regionalisation steps of the CAPRI data consolidation because NUTS2 data are offered. The main disadvantage for our purposes is the complete lack of nonagricultural data coverage.
- **LEVcrf** – The UNFCCC common reporting format (CRF) data (1990-2016), also cover land transitions and give settlement data. Official data for LULUCF accounting.

These sources each provide information on some "land use classes" (Table 7 of Annex) at least. These land use classes might be related to agricultural activities (like "olive groves" OLIVGR, covering the activities "tables olives" TABO and "olives for oil" OLIV) or they may refer to nonagricultural land uses ("artificial land" OART). Land use classes are in turn related to land use aggregates (Table 7 of Annex).

Include file '*coco1_finish_raw.gms*'

This file includes some final checks and adjustments before moving on to the optimisation part of coco.

- For seed quantities technical limits for reasonable seed use per ha are imposed.
- For all non crop products producer prices are assigned from the EAAP/UVAP positions or PRIC
- For all products with one of activity level, production or yield missing some correcting actions are taken.
- For FEDM, HCOM, SEDF and SEDM lower and upper limits are introduced to limit yearly changes in the subsequent estimation routines.

3.2.4 COCO1 Estimation procedure

COCO was primarily designed to fill gaps or to correct inconsistencies found in statistical data and, additionally, to easily integrate data from non EUROSTAT sources in the model. However, given the task of having to construct consistent time series on yields, market balances, EAA positions and prices for all EU Member States, and therefore thousands of series, a heavy weight was put on a transparent and uniform *econometric solution* so that manual corrections were avoided, to some extent at least. Regarding the construction of the data base, three principal problems had to be solved:

1. Gaps had to be filled in time series, either before the first available point, inside the range where observations are given, or beyond it.
2. Some time series were missing altogether and had to be estimated, e.g. when there are data on animal production but none on meat output per head.
3. Corrections of given statistical data should be minimised, if possible.

In order to take into account logical relation between the time series to fill, and eventually to make minimal corrections in the light of consistency definitions, simultaneous estimation techniques are used in this exercise. In order to use to the greatest extent the information contained in the existing data, the following principles are applied:

1. Accounting identities positions of the market balance summing up to zero, the difference between stocks as the stock change and similar restrictions constrain the estimation outcome.
2. Relations between aggregated time series (e.g. total cereal area) and single time series are used as additional restrictions in the estimation process.
3. Bounds for the estimated values based on engineering knowledge or derived from first and second moments of times series ensure plausible estimates and/or bind estimates to original data. Additionally, bounds are constructed from more disaggregated time series, if the aggregate is missing.
4. As many time series as technically possible are estimated simultaneously to use the full extent of the informational content of the data constraints (1) and (2).

The first three points neatly conform to the Bayesian Highest Posterior Density (HPD) approach proposed in Heckelei et al. 2005. The reader may notice that the problem is quite similar to system estimation in economics. Consider a system of supply curves. A standard approach to estimate such a system includes the specification of a functional form consistent with profit maximisation and the imposition of various constraints (homogeneity, symmetry, convexity) on the parameters to be estimated. Our approach is quite similar, as our goal asks for consistent estimates as well. Instead, we introduce explicit data constraints involving the fitted values for each point and take the fitted values later as the content of the data base.

The estimation is prepared in the following steps:

1. Estimate independent trend lines for the time series.
2. Estimate a Hodrick-Prescott filter using given data where available and otherwise the trend estimate as input.
3. Define 'target values' which are (a) given data, (b) the results from the Hodrick-Prescott filter times R^2 plus the last (1-R^2) times the average of nearest observations. The target values may be considered modes of a prior distribution.
4. Specify a 'standard deviation' for each data point which is different for given data and gaps.

The concept is put to work by a minimisation of normalised least squares under constraints:

$$
\begin{aligned}
min_{y_{i,t}} & \sum_{i,t \in obs} wgt^{dat}((y_{i,t} - y_{i,t}^{dat})/abs(y_{i,t}^{trd} - y_{i,t}^{dat}))^2 \\
& + \sum_{i,t \notin obs} wgt^{ini}((y_{i,t} - y_{i,t}^{ini})/s_{i,t})^2 \\
& + \sum_{i,t} wgt^{hp}((y_{i,t+1} - y_{i,t}) - (y_{i,t} - y_{i,t-1})/s_{i,t})^2 \\
& + \sum_{i,t} wgt^{up}((max(y_{i,t}^{up}, y_{i,t}) - y_{i,t}^{up}))/abs(y_{i,t}^{up}))^2 \\
& + \sum_{i,t} wgt^{lo}((min(y_{i,t}^{lo}, y_{i,t}) - y_{i,t}^{lo}))/abs(y_{i,t}^{lo}))^2
\end{aligned}
\tag{3.4}
$$

s.t.

$y_{i,t}^{LO} < y_{i,t} < y_{i,t}^{UP}$

Accounting identities defined on$y_{i,t}$

Identity of land use from different sources

where i represents the index of the elements to estimate (crop production activities or groups, herd sizes etc.), t stands for the year, wgtx are weights attached to the different parts of the objective ($wgt^{dat} = wgt^{hp} = 10, wgt^{ini} = 1, wgt^{up} = wgt^{lo} = 100$), and

$y_{i,t}$ = the fitted value for item i, year t

$y_{i,t}^{dat}$ = the observed data for item i, year t

$obs = \{(i,t)|y_{i,t}^{dat}0\}$, the set of data points with nonzero data

$y_{i,t}^{trd}$ = the trend value of an initial t rend line through the given data

$y_{i,t}^{ini}$ = initial supports for gaps: preliminary Hodrick-Prescott filter result (from step 2) times R^2 plus the last (1-R^2) times the average of nearest observations

$s_{i,t}, (i,t) \notin obs = 0.1 \cdot y_{i,t}^{ini} + s_{i,t}^{trd}$, weighted sum of the initial support for gaps and the standard error of the initialising trend

$s_{i,t}, (i,t) \in obs = 0.1 \cdot y_{i,t}^{dat} + s_{i,t}^{trd}$, weighted sum of given data and the standard error of the initialising trend

$y_{i,t}^{lo}, y_{i,t}^{up}$ = 'soft' bounds, triggering a high additional penalty if violated

$y_{i,t}^{LO}, y_{i,t}^{UP}$ = 'hard' bounds, defining the feasible space

The general weighing of the different terms evidently reflects the acceptability of certain types of deviations which is lowest (= 1) for deviations of the fitted value from the HP filter initialisation as these are considered quite poor, preliminary estimates (derived from independent trends). The weights are 10 times higher for deviations from given data and for the smoothing HP filter term. Finally there are extra penalty terms for fitted values moving beyond plausible 'soft' bounds $y_{i,t}^{lo}, y_{i,t}^{up}$. The 'hard' bounds

$y_{i,t}^{LO}, y_{i,t}^{UP}$ are constraining the feasible space for a number of solution attempts. However, if it turns out that certain constraints would persistently preclude feasibility of the data consolidation problem, they are relaxed in a stepwise fashion, but this widening of bounds is monitored on a parameter to check.

The denominators used to normalise the different terms are ‘standard deviations’ of the prior distribution in the framework of a HPD estimation but they are specified in view of practical considerations. Essentially they provide another weighting for particular (i,t) deviations depending on their acceptability, but these weights are specific to the particular data point. All denominators are derived from the variable in question such that they acknowledge the fact that the means of the time series entering the estimation deviate considerably. The normalisation hence leads to minimisation of relative deviations instead of absolute ones which could not be summed in a reasonable way.

It should be mentioned that the above representation of the COCO objective function is a quite simplified one: It is evident that the above lacks safeguards against division by zero or very small values which are included in the GAMS code. Furthermore there are different types of gaps which are not reflected above to avoid clutter (Are there gaps in a series with some data or is the series empty? Is the mean based on data or estimated from $y_{i,t}^{lo}, y_{i,t}^{up}$?)

Equation 4 indicates that accountancy restrictions are added. These restrictions can be balances (land, milk contents, young animals), aggregation conditions, definitions for processing coefficients and yields etc. They are quite similar to those applied for the ex ante trend projections as discussed in detail in Section 3.3 but the COCO1 accounting identities tend to acknowledge more details or have to establish the data base that is subsequently given for the ex ante trend projections, for example related to the split of high and low yield animal activites (DCOL, DCOH, BULL, BULH, HEIL, HEIH):

```
*  ----- Relationship of of technologies (high,low) to total DACTs (only HEIF, BULF)
*        NOTE: Omitting p_splitFac(DCOW,EstR) also omits this equation from EGRP=MILK_ACT
*
   e_splitDACT("%MS%000",EstC,EstR,T) $  SUM(DACT_TO_COLS(DACT,EstC),
                                           p_NobsP("%MS%000",DACT,"LEVL")*p_splitFac(EstC,EstR))..

        v_EstimY("%MS%000",EstC,EstR,T)
           =E= SUM(DACT_TO_COLS(DACT,EstC) $ p_splitFac(EstC,EstR),
                                           p_splitFac(EstC,EstR)*v_EstimY("%MS%000",DACT,EstR,T));
```

The fixed yield variation imposed in this way is ± 20% and each of the variants corresponds a fixed 50% of the total activity level whereas other accounting equations ensure that the process length DAYS and the daily growth DAILY vary accordingly.

In the **dairy sector** the strategy of an update in 2015 has been to obtain a fairly detailed data consolidation with a distinction of milk processed and dairy products obtained in dairies and on farm, using most of the available data sources. For the subsequent modules this disaggregate description of the dairy sector is consolidated to some extent for further use.

The equation system considers that both in dairy as well as on farm the raw milk used has to be consistent in terms of milk fat and protein with the products obtained:

$$PCRM_M \cdot \delta_{c,M} = \sum_i (NAGR_i - PRCM_i) \cdot \delta_{c,i} \tag{3.5}$$

where

$PRCM$ = processing of raw milk M or dairy product i (e.g. cheese)

$NAGR$ = products obtained in dairies (e.g. MC100, fresh products, from apro_mk_pobta)

c = type of milk content (= FATS, PROT)

i = dairy product (e.g. MC100, fresh products, from apro_mk_pobta)

δ = average content in dairies

In a similar manner we have balances for milk contents in on farm use of raw milk as well as in the products obtained on farm:

$$(INDM_M + HCOM_M) \cdot \delta_{c,M} = \sum_i FARM_i \cdot \phi_{c,i} \tag{3.6}$$

where

$INDM$ = use of raw milk M on farm for farm cheese, farm butter etc (e.g. MF240-UWM)

$HCOM$ = use of raw milk M on farm as drinking milk (MF110-UWM, includes both direct sales as well as home consumption)

$FARM$ = products obtained on farm (e.g. MF110-PRO, MF240-PRO)

ϕ = average content on farm

The content of milk products will typically differ, in particular for the most important product "fresh milk products" (FRMI), as this includes yoghurts etc in dairies but will be dominated by drinking milk on farm. However, to accomodate the important case of drinking milk it is not necessary to have all contents on farm deviating freely from the standard contents in dairies. Instead we require that

$$CORF_{c,i} \cdot \delta_{c,i} = \phi_{c,i} \tag{3.7}$$

where

$CORF$ = ratio of on farm content to the standard content

and CORF is contrained to equal to one except that we permit CORF $\neq 1$ for FRMI.

Production in dairies and on farm may be added to obtain the total production that enters the market balances:

$$MAPR_i = NAGR_i + FARM_i \tag{3.8}$$

$$MAPR_i = HCOM_i + PCRM_i + FEDM_i + NTRD_i \tag{3.9}$$

where

$MAPR$ = Marketable production according to the (discontinued) Eurostat market balances (USAP-FRMI from apro_mk_bal_B4410_12) or in terms of the commercially marketed quantities only:

$$NAGR_i = (HCOM_i - FARM_i) + PCRM_i + FEDM_i + NTRD_i \tag{3.10}$$

The market balance for the raw milk looks as follows:

$$GROF_M = PRCM_M + HCOM_M + INDM_M + FEDM_M + LOSM_M \tag{3.11}$$

where

$FEDM$ = Feed use of raw milk (apro_mk_farm_MF520_UWM)

$LOSM$ = Losses of raw milk (apro_mk_farm_MF600_UWM)

After solving the data consolidation according to the above equations the following rebookings will be useful for subsequent modules:

$$MAPR'_i = NAGR_i \tag{3.12}$$

$$HCOM'_i = HCOM_i - FARM_i \tag{3.13}$$

$$HCOM'_M + FEDM'_M + LOSF'_M = HCOM_M - FDEM_M + LOSM_M + INDM_M \tag{3.14}$$

The first two of the previous equations transform the standard (total) market balances including on farm use and production into "commercial" market balances only which is useful for comparisons with some datasets. The last equation is active for a while already in COCO. It identifies $HCOM'_M$ = raw milk for direct sales (regardless of in terms of drinking milk or on farm products), feed milk and $LOSF'_M$, an aggregate of losses and on farm use of milk by farm households themselves. The original position $INDM_M$ is basically allocated to a part consumed on farm and that part of direct sales which occurs in processed form (farm cheese, butter...). As the form of on farm consumption is not modelled in CAPRI, items FARM, NAGR, INDM are not passed on to subsequent modules, only LOSF is passed on, because this needs to be accounted for when calculating deliveries to dairies ($PRCM_M$).

Related to **land use** data there are also a number of particularities and details. We have various sources reporting data on the same item (LEVL) that evidently contradict each other before the data consolidatuion. During the consolidation the following equation ensures the identity of land use areas among different sources (LEVCLC, LEVFAO etc):

Based on the previous constraint all other land related accounting restrictions only have to be checked for the item "LEVL", while the objective functions minimizes deviation from supports of all sources. Accounting restrictions ensure consistency of crop activities with land use classes and their aggregates.

Complications in the consolidation of land use data are related to the use of UNFCCC data for 6 land use classes (set "LUclass": CROP, FORE, ARTIF, GRSLND, WETLND, RESLND), because three of the UNFCCC land use classes (GRSLND, WETLND, RESLND) differ conceptually from "related" categories

```
*
*  ------ ensures identical values among different sources
*
   IdentLandUseAgg_("%MS%000",EstC,Source,T) ..

           EstimY("%MS%000",EstC,Source,T)
                               =E= EstimY("%MS%000",EstC,"LEVL",T);
```

from other data sets. Thus it is only possible to specify some inequnalities and an aggregation condition as constriants:

```
*  ------ Consistency of land use or aggregates with UNFCCC land use information
*
   e_ConsisWet("%MS%000",T) $ (estC("WETLND") $ estC("INLW") )..
*
           v_EstimY("%MS%000","INLW","LEVL",T) =L= .95*v_EstimY("%MS%000","WETLND","LEVL",T);

   e_ConsisGrs("%MS%000",T) $ (estC("GRSLND") $ estC("GRAS") )..
*
           v_EstimY("%MS%000","GRAS","LEVL",T) =L= .95*v_EstimY("%MS%000","GRSLND","LEVL",T);

   e_ConsisRes("%MS%000",T) $ (estC("RESLND") $ estC("OLND") )..
*
           v_EstimY("%MS%000","OLND","LEVL",T) =G= 1.05*v_EstimY("%MS%000","RESLND","LEVL",T);
*
   e_ConsisGio("%MS%000",T) $ (estC("GIOLND") )..
*
           v_EstimY("%MS%000","GRAS","LEVL",T)+v_EstimY("%MS%000","INLW","LEVL",T)+v_EstimY("%MS%000","OLND","LEVL",T)
            =E= v_EstimY("%MS%000","GIOLND","LEVL",T);
*
*  ------ Adding up of land use change to a LU class (including the non-changing area) to the final (end of year) area
*
   e_LUCtoClass("%MS%000",LUclass,T) ..
*
          SUM(LUC_TO_LUclass(LUCpos,LUclass) $ estC(LUCpos),v_EstimY("%MS%000",LUCpos,"LEVL",T))
              =E= v_EstimY("%MS%000",LUclass,"LEVL",T);
```

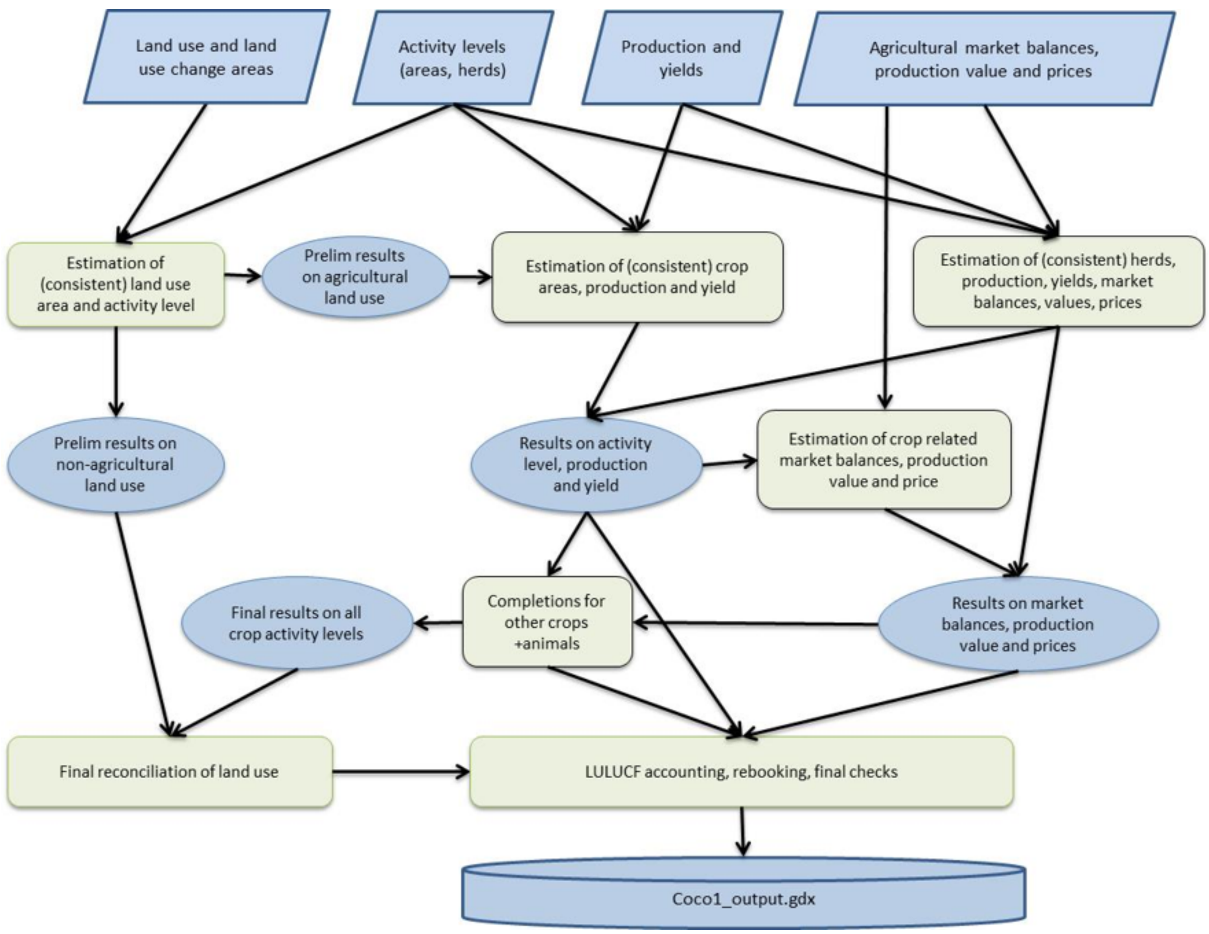
The last equation illustrates that the land use accounting based on UNFCCC data (introduced in 2015) also involves the land use *changes* (LUCpos) into the 6 LU classes (and a corresponding condition for changes *from* those LU classes).

It should also be explained that Equation 1 is not applied simultaneously to the whole dataset because the optimisation would take too long. Instead it is applied to subsets of closely related variables:

1. Land use and land balance (Estimation step 1 for preliminary LU results).
2. Crop production (land balance + yields) for all crops simultaneously (Estimation step 2).
3. Production, yields, EAA, market balances for groups of animals like "cattle" (Estimation step 3).
4. Crop EAA + market balances for groups of crops, taking production from (2.) as given (Estimation step 4).
5. As the crop level estimation or the other crop completions may have slightly changed aggregate areas, the land use estimation has to be repeated (Estimation step 5).

This procedure has developed as a path dependent compromise between computation time and presumed quality. It starts with an estimation of land use in combination with agricultural land balance, including the land transition between LU classes. This determines the utilisable agricultural area (UAA) and non-agricultural land use. Step 2 distributes crop areas within the fixed UAA from step 1 and estimates crop production and yields. Step 3 only tackles the complete animal sector data (activities, markets, EAA). The crop production is taken as given, when market balance and EAA are estimated for the crops and derived processed products (step 4). However, with all steps completed some final checks may modify the results (e.g. delete tiny activity levels or estimate another crop area from another crop output value and thus change the UAAR). Furthermore the crop estimation may have slightly changed the ratio of cropland to productive grassland. Therefore the accounting identities ensured in steps 1 are not necessarily fulfilled in a strict sence anymore. Hence a final reconciliation of land use is added for full consistency:

Figure 3: Overview on main estimations in for the consolidation of national data in Europe (in coco1.gms)



Results are not always fully satisfactory (perhaps impossible given some raw data). For example the resulting prices (unit values) are far from a priori expectations for a number of series, in particular less important ones. This is because, apart from some additional security checks, unit values are by and large considered a free balancing variable calculated to preserve the identity between largely fixed EAA values and fixed production (in coco1_estimb). The priority for EAA values has been reduced somewhat in recent years but a more thorough revision would require to estimate production, market balances and EAA simultaneously rather than consecutively (first (*a*), then (*c*) for crops). As this is infeasible for all crops at the same time the whole estimation would need to be split up differently in the crop sector, perhaps first for the aggregates and then within those.

Furthermore it should be mentioned that the main parts of COCO are handled in a program ('coco1.gms') looping over MS because there are no direct linkages between them. However, for practical reasons it will be useful to run COCO in country groups that have the same coverage of years. The longest series (as off 1984) can be established for EU15[11] countries except Germany. For the New MS it turned out that data before 1989 are often very unreliable and create considerable burden in the data maintenance. These countries (and Germany) are only completed for years from 1989 onwards therefore. Norway also offers reliable series as of 1984. In the case of the Western Balkan countries it is rather hopeless to provide very recent data as key data are still missing such that the series can only be completed from 1995 onwards. Furthermore for the Western Balkan counties it was necessary to transfer certain coefficients and shares from (previously consolidated) neighbouring countries to the Western Balkan, such that a certain sequence is necessary for a reasonable application of COCO1:

- Run COCO1 for EU28 countriesand Norway, either in one batch from the GUI or one by one (always with sub-steps 1 to 5).
- Run COCO1 for the set of candidate countries (Western Balkan and Turkey) on the reduced time span with given data (1995 – 2009). Because these use some shares and ratios from an average of selected EU28 countries the latter have to be consolidated first.

3.2.5 COCO2: Data Preparation

The data consolidation in COCO2 only covers a few special topics:

- producer prices of dairy products and vegetable oils
- consumer prices
- consumer losses and nutrient intake after losses
- feed stuff quantities without market balances (by-product, fish emal)
- loss rates of fodder for preliminary balancing of animal nutrients
- corrections of certain LULUCF coefficients based on UNFCCC

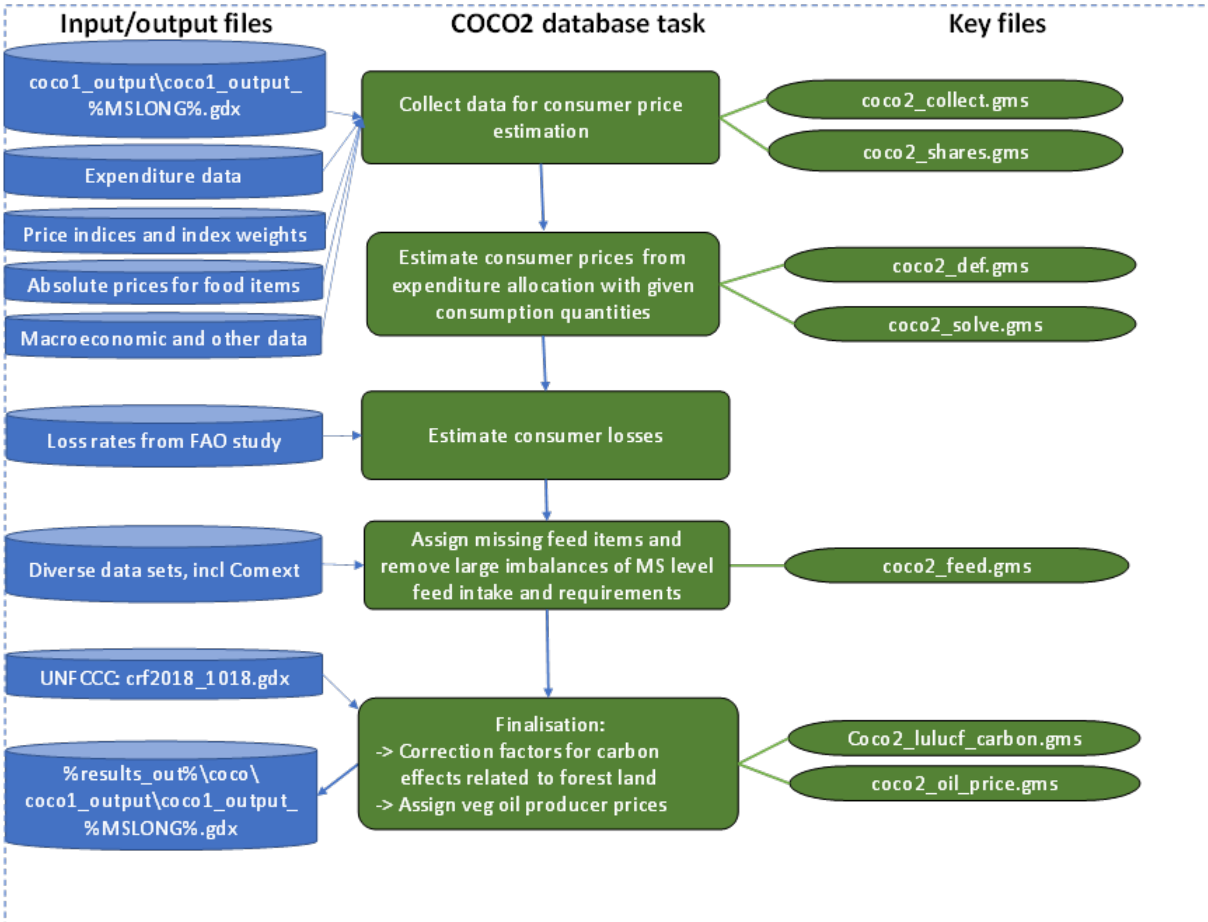
An overview is given in the following figure.

In spite of only limited subtasks tackled in coco2.gms, the multitude of different data inputs is comparable to that in COCO1.

Include file *'coco2_collect.gms'*

[11] Belgium and Luxembourg are aggregated in COCO for reasons of data availability.

Figure 4: Overview on main elements in the finalisation step for the consolidation of national data in Europe (in coco2.gms)



Various input files are collected with some adjustments to match to CAPRI definitions and with some gap filling. As the consumer prices follow from a top down expenditure allocation problem, the input data range from macroeconomic information to very detailed prices of food items.

- Consolidated data from COCO1
- Macroeconomic information from Eurostat and UNSTATS: Exchange rates, population, GDP deflator, private consumption of households in current prices.
- Price index information: Aggregate food price index, relative (to EU) food price index, harmonised indices of consumer prices (HICPs) with item weights all from Eurostat
- Expenditure by product groups (from Eurostat and national sources)

- Auxiliary data for special cases (Prices for some milk products in selected countries, fish meal information etc)
- Country Sheets of the Western Balkan and Turkey: Exchange rate, inhabitants, inflation rate, food expenditure shares
- Disaggregate absolute consumer prices for selected narrowly defined food items (ILO and Eurostat)

Where available, producer prices for milk products were already included from Eurostat statistics (Agricultural prices and price indices) in COCO1. Completeness was not achieved in COCO1, however, because processed dairy products are not part of the EAA. Here we complete some gaps using price information for some Member States and (partly assumed) relationships among dairy product prices and their fat and protein contents. Data on total consumer expenditures as well as expentitures by food groups are included from various sources as described in Chapter 2.2.2.5 FIXME, partly extended using general price index information.

Consumer price index weights and price indices for food aggregates (2005=100) are coming from Eurostat tables on HICP. Supplementary information for Albania, Bosnia and Croatia comes from national agencies. The price index weights are used to extend older series on food expenditure by product groups (say “meat”) which have been discontinued (see below under file coco2_shares.gms).

Finally we use very narrowly defined absolute consumer prices (e.g. for spaghetti) and price indices. The earlier years (before 2008) had been provided by ILO which has discontinued this activity. For a subset of those Eurostat offers matching information as “detailed average prices (table prc_dapYY) that has been used to extend the ILO series. These prices are mapped to CAPRI regions, products and units (*‘coco2_ilo_addup.gms’*).

Price indices for food and non-alcoholic beverages from HICP as well as the general food price index are used to complete the disaggregate ILO prices for single typical food items. (like “Wheat bread white unsliced not wrapped”) using a Hodrick-Prescott filter and the expectation that their changes should follow the price index informaiton collected.

Finally another HPD estimator is used to adjust the dissagregate prices to be (somewhat) in line with Eurostat information on relative food price levels across Europe.

Include file *‘coco2_shares.gms’*

Expenditure shares are defined and completed top-down using simple OLS estimates against related statistical expenditure information or, as a last fall back option, based on a trend.

The food expenditure share completions start with data from COICOP level 3 giving results on food and non-alcoholic beverages. Further disaggregation relies on historical Eurostat data (HIST), on the above mentioned index weights from HICP and partly national data (Germany and Spain).

A conveninent expenditure group is potatoes as these expenditure shares may be extrapolated based on COCO1 human consumption multiplied by producer price as regressors for OLS.

3.2.6 COCO2: Estimation procedure

Include file *‘coco2_def.gms’*

The approach to determine consumer prices is to distribute food expenditure on groups with consumption quantities given from COCO1 results such that endogenous consumer prices link endogenous expenditure with exogenous quantities. Deviations of estimated expenditure and consumer prices from their supports is penalised in an entropy framework. Estimation is done year by year, starting with the most recent year where hard data are usually available to a greater extent than for the oldest years in the database. Including consumer price changes (always relative to the previously solved year) serves to stabilise the results to some extent such that the objective does not only have supports for the consumer prices, but also for their changes. The entropy problem is solved by maximizing:

$$
\begin{aligned}
max_t - \sum_{m,j,k} & CPS_{m,j,2} \cdot HCOM_{m,j,k}/1000/TOFO_{m,t} \\
& \cdot PE_{m,j,k} \cdot log(PE_{m,j,k}/PQ_k) \\
& - \sum_{m,j,k} CPS_{m,j,2} \cdot HCOM_{m,j,k}/1000/TOFO_{m,t} \\
& \cdot PED_{m,j,k} \cdot log(PED_{m,j,k}/PQ_k) \\
& - \sum_{m,FOPOS,k} EXS_{m,FOPOS,2}/TOFO_{m,t} \\
& \cdot PEX_{m,FOPOS,k} \cdot log(PEX_{m,FOPOS,k}/PQ_k) \\
& - \sum_{m,j,k} PFAC_{m,k} \cdot LOG(PFAC_{m,,k}/PQ_k) \cdot 1000
\end{aligned}
\tag{3.15}
$$

where m represents the region, j the food item with consumer price, FOPOS the food group, t stands for the current estimation year, t_1 for the year estimated before and k for the number of support points (=3).

Parameters are

$HCOM_{m,j,t}$	Human consumption, result from COCO1
$UVAD_{m,j,t\_1}$	Consumer price from last simulation of year t+1
$CPS_{m,j,k}$	Support points for consumer prices
$DCPS_{m,j,k}$	Support points for consumer price changes
$EXS_{m,FOPOS,k}$	Support points for group expenditures
$TOFACS_{m,k}$	Support points for total food expenditure slack
PQ_k	A priori probabilities for support points
$TOFO_{m,t}$	Total food expenditure
and entropy variables	
$PE_{m,j,t}$	Probability of support points for consumer prices
$PED_{m,j,t}$	Probability of support points for consumer price changes
$CP_{m,j}$	Consumer prices
$DCP_{m,j}$	Consumer price changes
$PEX_{m,FOPOS,t}$	Probability of support points for group expenditure
$PFAC_{m,k}$	Probability of support points for food expenditure slack
EX_{mFOPOS}	Group expenditures
$TOFAC_m$	Food expenditure slack

Constraints are as follows: Summing up probabilities for support points

$$\sum_{k\forall_{m,j}(CP.L_{m,j}\geq 0\wedge HCOM_{m,j,i}\geq 0)} PE_{m,j,k} = 1 \tag{3.16}$$

$$\sum_{k\forall_{m,j}(DCPS_{m,j}\geq 0\wedge HCOM_{m,j,i}\geq 0)} PE_{m,j,k} = 1 \tag{3.17}$$

$$\sum_{k\forall_{m,j}(EX.L_{m,FOPOS}\geq 0)} PE_{m,FOPOS,k} = 1 \tag{3.18}$$

$$\sum_{k\forall_{m}(TOFAC.LO_m\geq TOFAC.UP_m)} PFAC_{m,k} = 1 \tag{3.19}$$

Define consumer price changes from support points

$$DCP_{m,j} = \sum_{k\forall_{m,j}(CP.L_{m,j}\geq 0\wedge HCOM_{m,j,i}\geq 0\wedge DCPS_{m,j,2}\geq 0)} PED_{m,j,k}\cdot DCPS_{m,j,k} \tag{3.20}$$

Of course consumer prices changes are also related to the last simulation result (which is for T+1 due to backward looping)

$$DCP_{m,j} = UVAD_{m,j,t\_1} - CP_{m,j} \tag{3.21}$$

Define consumer prices from support points and probabilities

$$CP_{m,j} = \sum_{k\forall_{m,j}(CP.L_{m,j}\geq 0 \wedge HCOM_{m,j,i}\geq 0)} PE_{m,j,k} \cdot CPS_{m,j,k} \tag{3.22}$$

Define group expenditure from support points and probabilities

$$EX_{m,FOPOS} = \sum_{k\forall_{m,j}(EX_{m,FOPOS}\geq 0)} PEX_{m,FOPOS,k} \cdot EXS_{m,FOPOS,k} \tag{3.23}$$

Define total expenditure slack from support points and probabilities

$$TOFAC_m = \sum_{k\forall_m(TOFAC.LO_m \geq TOFAC.UP_m)} PFAC_{m,k} \cdot TOFACS_m \tag{3.24}$$

Exhaustion of food expenditure may be relaxed with a slack factor different from one. However, this "last resort" to achieve feasibility in the expenditure allocation problem is limited to years and countries with precarious data and subject to strong penalties.

$$\sum_{FOPOS} EX_{m,FOPOS} = TOFO_{m,t} \cdot TOFAC_{m,k} \tag{3.25}$$

Consistency of group expenditure

$$EX_{m,FOPOS} = \sum_{j\forall_{m,FOPOS}(j\in FOPOS \wedge HCOM_{m,j}\geq 0)} CP_{m,j} \cdot HCOM_{m,j}/1000 \tag{3.26}$$

For most countries the exhaustion of total expenditure is the only evident hard constraint (and even this is relaxed in problem cases). However, as the penalties for group expenditure are set high, and furthermore as the range of expenditure supports defines additional implicit hard constraints, the problem may turn out infeasible (typically solved by additional leeway). To meet the expenditure constraints the solver would tend to concentrate deviations from supports on the most important expenditure items while setting the less important items close to their supports. A more balanced distribution of deviations from supports was achieved in practice by weighting all contributons to the overall objective (except the last one for the total expenditure slack) with expected expenditure shares. The weights may be interpreted as expected expenditure shares because supports are specified in a symmetric way such that the central, second (of three) supports, which is used in the objective function, is equal to the expectation.

Include file *'coco2_solve.gms'*

The initialisation, solving, reporting and storage is organised in the next include files with a few elements worth mentioning

- The initialisation tries to ensure positive consumer margins by the assignments of expected values and by specifying bounds on estimated consumer prices. The reference point for these margins is an average of EU and national prices that reflects the importance of domestic sales vs. imports.

- Bounds and spread of supports around expected consumer prices are set high for items without ILO style prices (say “table olives” TABO) or where the fit of available price information is questionable (e.g. cabbage prices for “OVEG”).
- A checking parameter (“p_checks”) permits to check the iniitalisation in case of infeasibilites. The most frequent case observed in the last years is that lower bounds on oils expenditure become binding, suggesting the need for some systematic mismatch of price and expenditure information for this group.

3.2.7 COCO2: Final completions

At this point it may be motivated why there is at all a need for a COCO2 module instead of handling all further topics in COCO1, that is MS by MS. There are basially two motives:

- In some cases it is convenient to have the completed COCO1 results of all countries at hand for comparison purposes and in order to achieve a balanced picture across MS. This is the main motive for the assignments of consumer loss rates (Section 3.2.7.1).
- Whenever averages of consolidated data (from COCO1) across several or all MS are involved, a solution in a loop requires certain sequence (such as first solving for non-candidate countries to form the averages that are input to candidate countries) or is better solved in a new module like COCO2. This applies to the expenditure allocation problem (Section 3.2.5), to completions for certain feedstuffs (Section 3.2.7.2, EU averages used due to the scarcity of data), and to corrections of LULUCF coefficients (Section 3.2.7.3). FIXME

Assignment of consumer loss rates and nutrient intake per head Since a number of years diet shift scenarios have increase in importance and therefore the plausibility of per capita consumption projectios and hence their starting values, per capita consumption in the data base. A common yardstick to assess plausibility is nutrient (e.g. calorie) consumption per head where the nutrition literature offers guidance in terms of recommendable as well as “observed” consumption. For nutrition issues it is intake, so consumption after losses, which matters, such that the assignment of these loss rates becomes a critical element of the database. The starting values are due to an FAO study and stored in the /dat folder

```
*  --- Nutrient consumption per head and day should be net of losses in distribution and households
*      => sets and data included from xls
*
   Set lossReg(*) "Regions for loss rates (from FAO study Gustavsson et al 2011)";

*$CALL "gdxxrw %datdir%\arm\food_waste_fao2011.xlsx O=%datdir%\arm\lossReg.gdx  DSet=lossReg  rdim=1
$GDXIN %datdir%\arm\lossReg.gdx
$LOAD  lossReg
$GDXIN
```

The aggregate food share (= 1-loss shares) links intake (INHA(i)) to total consumption (sum(i, HCOM(i)*foodSh(i)) / INHA(levl) and is therefore stored in the database as well.

```
*  foodSh = share of HCOM used as food after losses in household + distribution sectors:
   DATA(MSact,"foodSh",prods,allT) $ DATA(MSact,"INHA","LEVL",ALLT)
     = sum((MS_to_lossReg(MSact,lossReg),rows_to_lossAgg(prods,lossAgg)),(1-p_lossrates_FAO2011(lossReg,"distribution",lossAgg))
                                                                         *(1-p_lossrates_FAO2011(lossReg,"household",lossAgg)));
```

In spite of the FAO study the real loss rates are highly uncertain. Therefore they are reduced if the estimate of calorie intake based on the FAO loss rates strongly falls short of recommendations (most strongly in a set of "low calory regions"). Conversely loss rates are increased, if the estimate of calorie intake based on the FAO loss rates strongly exceeds recommendations (e.g. in Turkey).

```
*       -- reduce/increase FAO loss rates if aggregate calory consumption deviates strongly from plausible values
*          Note1: foodSh(cor)=1-lossSh(cor)=1-lossSh(ori)*corFac=1-(1-foodSh(ori))*corFac
*                 where we have designed an ad hoc formula for corfac below.
*          Note2: fishEtcShar is the share of missing Coco items OTHO,COCO,fish that are only added in capmod (data_prep)
*                 The shares are from a preliminary baseline for BAS=2008. In the future they should be included in COCO.
*
   set msLoKcal(msCon) /KO000, BA000,AL000/;
   set msHiKcal(msCon) /TR000/;
```

Completion of feed related data in coco2_feed The first sections of coco2_feed handle completions for certain by-products and other product so far ignored in coco1. These are by-products of the milling and the brewing industry and for corn gluten feed, sugarbeet pulp, manioc and fish meal where the database is completed for market balance positions production, imports, exports and feed. This relies on discontinued Eurostat tables (collected on p_feedAgri) which are extended using national data and external trade data from Comext. After completion the detailed by-products are aggregated to the CAPRI rows FENI (Rich energy fodder imported or industrial) and FPRI (Rich protein fodder imported or industrial). Based on completed data for all feedingstuffs nutrient contents for the CAPRI feed "bulks" (cereal feed FCER, protein feed FPRO etc) are assigned as an aggregate of their components.

These completions are useful as such but they also permit a balancing of (preliminary) total nutrient supply and demand in the animal sector that ultimately serves to adjust loss rates for fodder with the help of a number of include files:

Include files *'feed_decl.gms'* and *'req_or_man_fcn.gms'*

These files are not only active in COCO2, but also in CAPREG, and in the baseline calibration of CAPMOD. This "reuse" of the same files in different modules is efficient and ensures consistency, but usually also requires some adaptations of set definitions:

```
*   --- some sets used in the requirement functions:
*
    Alias(T,YEARS);
    SET RS(MS);
    SET R_RAGG(MS,MS);
    RS(MS) = YES;
    R_RAGG(MS,MS) = YES;
    SET A/ T /;
```

The previous snippet from coco2_feed gives an example that some sets (RS, R_RAGG) are assigned specifically to ensure functionality in different modules (here COCO2).

As the name should signal file *'feed_decl.gms'* mainly collects a number of declarations but it also specifies some bounds for process length DAYS and daily growth DAILY that are imposed throughout of CAPRI (example: maximum daily growth for male cattle = 1.5kg/day). The second include file (*'req_or_man_fnc.gms'*) specifies the requirement functions (with the argument "req" passed on) for animal activities of CAPRI.

Requirement functions are specified that determine:

- ENNE Net energy for ruminants as sum of
  - NEL net energy for lactation (cows, ewes, goats)
  - NEM net energy for maintenance (cows, calves, bulls, heifers, ewes, goats)
  - NEA net energy for activity (cows, calves, bulls, heifers, ewes, goats)
  - NEP net energy for pregnancy (cows)
  - NEG net energy for growth (calves, bulls, heifers)
- ENMC Net energy chicken
- ENMP Net energy pigs
- CRPR crude protein (all categories) and LISI lysine aminoacid (sows, poultry)
- DRMA dry matter (all categories with min and max requirements)
- Various fiber measures (irrelevant for COCO2)

There are three main sources for these functions:

- IPCC 2006 guidelines for the estimation of emissions (`http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_10_Ch10_Livestock.pdf`)
- Kirchgessner Tierernährng, 7th edition, 1987
- CAPRI working paper 97-12 (`http://www.ilr.uni-bonn.de/agpo/publ/workpap/pap97-12.pdf`)

These functions are one the one hand quite complex. They are composed of various parts that finally give the requirements, for example for energy, as a function of various parameters that may be specific to the region (often the final weights, process length, daily growth) or uniform across regions (carcass ratio). In spite of several components these are typically linked in a straightforward fashion as will be illustrated with a relatively easy example (energy for maintenance of heifers for fattening).

As a starting point, the daily growth from COCO is forced into the range defined in *'feed_decl.gms'*. At the same time regions with a stocking rate above the MS average are assumed to rely on more intensive technologies, such that their daily growth is also above average (but within the range $[DAILY_{lo}, DAILY_{up}]$). This is irrelevant in COCO (r=MS, no subnational regions) but relevant for CAPREG and CAPMOD calling the same *'req_or_man_fnc.gms'*:

$$\begin{aligned} &dailyIncrease_r^{HEIF} \\ &= min\left[DAILY_{up}^{HEIF}, max\left(DAILY_{lo}^{HEIF}, \frac{stockingrate_r}{stockingrate_{MS}} DAILY_{MS}^{HEIF}\right)\right] \end{aligned} \tag{3.27}$$

The daily increase is then used to determine the process length (rearrangement of equation below with empty days EDAYS = 0)

$$fatngday_r^{HEIF} = min\left[DAYS_{up}^{HEIF}, max\left\{DAYS_{lo}^{HEIF}, \frac{BEEF_r^{HEIF}/carcassSh_{HEIF} - startWgt_{HEIF}}{dailyIncrease_r^{HEIF}}\right\}\right] \quad (3.28)$$

The daily increase and process length may be conbined to estimate the mean live weight,

$$meanWgt_r^{HEIF} = startWgt_{HEIF} + \frac{dailyIncrease_r^{HEIF} \cdot fatngdays_r^{HEIF}}{2} \quad (3.29)$$

which in turn is the last information to estimate energy requirements for maintenance according to the IPCC guidelines:

$$NEM_r^{HEIF} = (meanWgt_{HEIF})^{0.75} \cdot 0.322 \cdot fatngdays_r^{HEIF} \quad (3.30)$$

Other energy requirements (for growth and activity) are calculated in a similar fashion as well as those for other animals. Important aspects to note are

- Fixed bounds for DAYS and DAILY ensure reasonable requirements, but require that the same constraints are anticipated in COCO and CAPREG to avoid inconsistencies.
- Regional coefficients are derived from the MS level information

Include file *'coco2_gras.gms'*

With animal requirements specified the results of COCO1 for grass, other fodder and as a last resort cereals might be revised in terms of losses on farm to achieve an acceptable relationship of energy and protein requirements of total herds compared to the intake with feed. For gras and other fodder on arable land the contents may be adjusted in certain limits as well. The corrections do not eliminate the typical oversupply of nutrients compared to the requirements based on the literature, but they should give reasonable starting values for the feed allocation addressed in module CAPREG.

Compare COCO1 results with UNFCCC and compute correction factors in coco2_lulufc_carbon In COCO1, an assignment of LULUCF effects (totals and per ha) has taken place, mostly relying on IPCC coefficients. These assignments are compared in coco2_lulucf_carbon with the reportings from EU MS to UNFCCC. For forestry and any transitions involving forestry, the standard IPCC reporting appears rather coarse, as it implies, for example, that management of forest land remaining forest has zero carbon effects. By contrast most EU countries report that there is still a considerable gain in biomass from forest management because the forests have not yet achieved a stable state (as implied by IPCC standard methodology).

To pick up the detailed knowledge of management practices, disturbances, age and species structure embededed in the country level UNFCCC reporting the forest management coefficients per ha for the remaining class (FORFOR) have been already adopted in COCO1. Here we also compute correction factors for the default per ha effects from transitions involving forestry. These are ultimately stored on the data(.) array unloaded in the main result file to be used in LULUCF accounting of CAPMOD.

Complete prices for vegetable oil in coco2_oil_price The EU prices for vegetable oils relevant for biofuel processing functions are assigned using prices from a USDA source. These assignments refer to prices at the wholesale level (relevant for the processing industry), not to consumer prices which have been determined previously.

After this last include file the completions in module COCO2 are finished and the main output file (coco2_output.gdx) is unloaded. This file is loaded in subsequent modules (main use in CAPREG, but also in CAPTRD for nowcasting and in CAPMOD for update of LULUCF coefficients).

3.2.8 Annex: Code lists for the COCO database

This section includes detailed code lists, which are in use in the COCO database.

Table: Codes used for storing the original REGIO tables in the database and their description (rows)

Codes used in CAPRI REGIO tables	Original REGIO description
TOTL	Territorial area
FORE	Forest land
AGRI	Utilized agricultural area
GARD	Private gardens
GRAS	Permanent grassland
PERM	Permanent crops
VINE	Vineyards
OLIV	Olive plantations
ARAB	Arable land
GREF	Green fodder on arable land
CERE	Cereals (including rice)
WHEA	Soft and durum wheat and spelt
BARL	Barley
MAIZ	Grain maize
RICE	Rice
POTA	Potatoes
SUGA	Sugar beet
OILS	Oilseeds (total)
RAPE	Rape
SUNF	Sunflower
TOBA	Tobacco
MAIF	Fodder maize
CATT	Cattle (total)
COWT	Cows (total)
DCOW	Dairy cows
CALV	Other cows
CAT1	Total cattle under one year
CALF	Slaughter calves
CABM	Male breeding calves (lt;1 year)
CABF	Female breeding calves (lt;1 year)
BUL2	Male cattle (1-2 years)
H2SL	Slaughter heifers (1-2 years)
H2BR	Female cattle (1-2 years)
BUL3	Male cattle (2 years and above)
H3SL	Slaughter heifers (2 years and above)
H3BR	Breeding heifers
BUFF	Total buffaloes
PIGS	Total pigs (total)
PIG1	Piglets under 20 kg
PIG2	Piglets under 50 kg and over 20 kg
PIG3	Fattening pigs over 50 kg
BOAR	Breeding boars
SOW2	Total breeding sows
SOW1	Sows having farrowed
GILT	Gilts having farrowed for the first time
SOWM	Maiden sows
GILM	Maiden gilts
SHEP	Sheep total)
GOAT	Goats (total)
EUQI	Equidae (total)
POUL	Poultry (total)
OUTP	Final production

Table: Codes used for storing the original REGIO tables in the data base and their description (columns)

Codes used in CAPRI REGIO tables	Original REGIO description
LEVL	Herd size / Area / # of persons
LSUN	Live stock units
PROP	Physical production
YILD	Yield
VALE	EAA position in ECU
VALN	EAA position in NC

Table: Connection between CAPRI and REGIO crop areas, crop production and herd sizes

SPEL-code	REGIO-code	REGIO-code	REGIO-code	REGIO-code	Description of SPEL activity
SWHE	WHEA	CERE	ARAB		Soft wheat
DWHE	WHEA	CERE	ARAB		Durum wheat
RYE		CERE	ARAB		Rye
BARL	BARL	CERE	ARAB		Barley
OATS		CERE	ARAB		Oats
MAIZ	MAIZ	CERE	ARAB		Maize
OCER		CERE	ARAB		Other cereals (excl. rice)
PARI	RICE	CERE	ARAB		Paddy rice
PULS			ARAB		Pulses
POTA	POTA		ARAB		Potatoes
SUGB	SUGA		ARAB		Sugar beet
RAPE	RAPE	OILS	ARAB		Rape and turnip rape
SUNF	SUNF	OILS	ARAB		Sunflower seed
SOYA		OILS	ARAB		Soya beans
OLIV		OLIV	PERM		Olives for oil
OOIL		OILS	ARAB		Other oil seeds and oleaginous fruits
FLAX			ARAB		Flax and hemp (faser)
TOBA	TOBA		ARAB		Tobacco, unmanufactured, incl. dried
OIND			ARAB		Other industrial crops
CAUL			ARAB		Cauliflowers
TOMA			ARAB		Tomatoes
OVEG			ARAB		Other vegetables
APPL			PERM		Apples, pears and peaches
OFRU			PERM		Other fresh fruits
CITR			PERM		Citrus fruits
TAGR		VINE	PERM		Table grapes
TABO		OLIV	PERM		Table olives
TWIN		VINE	PERM		Table wine
OWIN		VINE	PERM		Other wine
NURS			PERM		Nursery plants
FLOW			ARAB		Flowers,ornamental plants, etc.
OCRO			ARAB		Other final crop products
MILK	DCOW				Dairy cows
BEEF	BUL2	BUL3			Bulls fattening
CALF	CALF				Calves fattening (old VEAL)
PORK	PIG3	PIG2	PIG1		Pig fattening
MUTM	GOAT	SHEP			Ewes and goats
MUTT	GOAT	SHEP			Sheep and goat fattening
EGGS	POUL				Laying hens
POUL	POUL				Poultry fattening
OANI					Other animals
OROO			ARAB		Other root crops
GRAS	GRAS				Green fodder
SILA	GREF		ARAB		Silage
CALV	CALV				Suckler cows
RCAL	CABM	CABF			Calves, raising
HEIF	H2SL	H2BR	H3SL	H3BR	Heifers
PIGL	SOW2				Pig breeding
FALL			FALL		Fallow land

Tables: Codes of the input allocation estimation

FADN inputs (FI)	Label
TOIN	total inputs
COSA	animal specific inputs
FEDG	self grown feedings
ANIO	other animal inputs
FEDP	purchased feedings
COSC	crop specific inputs
SEED	seeds
PLAP	plant protection
FERT	fertilisers
TOIX	other inputs (overheads)

CAPRI inputs (CI) used in the reconciliation	label
TOIN	total inputs
FEED	feedings
IPHA	other animal inputs
COSC	crop specific inputs
SEED	seeds
PLAP	plant protection
FERT	fertilisers
REPA	repairs
ENER	energy
SERI	agricultural services input
INPO	other inputs

1. The set of *Other* activities that had been omitted from the econometric estimation:
   - OTHER={OCER, OFRU, OVEG, OCRO, OWIN, OIND, OOIL, OFAR, OANI}
2. The set of activity groups, and their elements, used in the replacement or missing/negative coefficients
   - GROUPS = {YOUNG, VEGE, SETT, PULS, PIG, OILS, MILK, MEAT, INDS, HORSE, GOAT, FRU, FOD, FLOWER, DENNY, COW, CHICK1, CHICK2, CHICK3, CERE, ARAB}
   - YOUNG={YBUL, YCOW},
   - VEGE={TOMA},
   - SETT={SETA, NONF, FALL, GRAS},
   - PULS=PULS
   - PIG={PIGF, SOWS},
   - OILS={RAPE, SOYA, SUNF, PARI, OLIV},
   - INDS={TOBA, TEXT, TABO},

- GOAT={SHGM, SHGF},
- FRU={APPL, CITR, TAGR, TWIN},
- FOD={ROOF, MAIF},
- FLOWER={FLOW, NURS},
- DENNY={PORK, SOWS},
- COW={DCOW, SCOW, HEIF, HEIR, CAMF, CAFF, BULF, CAMR, CAFR},
- CHICK1={HENS, POUF},
- CERE={SWHE, DWHE, BARL, OATS, RYEM, MAIZ},
- ARAB={POTA, SUGB}

3. The sets of Northern European, Southern European countries:

- NEUR={NL000, UK000, AT000, BL000, DE000, DK000, FI000, FR000, SE000}
- SEUR={El000, ES000, PT000, IT000, IR000}

Table: Codes of land use classes (Set LandUse)

Code	Label
OART	artificial
ARAO	(other) arable crops - all arable crops excluding rice and fallow (see also definition of ARAC below)
PARI	paddy rice (already defined)
GRAT	temporary grassland (alternative code used for CORINE data, definition identical to TGRA
FRCT	fruit and citrus
OLIVGR	Olive Groves
VINY	vineyard (already defined)
NUPC	nursery and permanent crops (Note: the aggregate PERM also includes flowers and other vegetables
BLWO	board leaved wood
COWO	coniferous wood
MIWO	mixed wood
POEU	plantations (wood) and eucalyptus
SHRUNTC	shrub land - no tree cover
SHRUTC	shrub land - tree cover
GRANTC	Grassland - no tree cover
GRATC	Grassland - tree cover
FALL	fallow land (already defined)
OSPA	other sparsely vegetated or bare
INLW	inland waters
MARW	marine waters
KITC	kitchen garden

Table: Codes of land use aggregates (Set LandUseAgg)

Code	Label
OLND	other land - shrub, sparsely vegetated or bare
ARAC	arable crops
FRUN	fruits, nursery and (other) permanent crops
WATER	inland or marine waters
ARTIF	artificial - buildings or roads
OWL	other wooded land - shrub or grassland with tree cover (definition to be discussed)
TWL	total wooded land - forest + other wooded land
SHRU	shrub land
FORE	forest (already defined)
GRAS	grassland (already defined)
ARAB	arable (already defined)
PERM	permanent crops (already defined)
UAAR	utilizable agricultural area (already defined)
ARTO	total area - total land and inland waters
ARTM	total area including marine waters
CROP	crop area - arable and permanent

Table: Codes of mutually exclusive subset adding up to total area - ARTO (Set LandUse-ARTO)

Code	Label
OLND	other land - shrub, sparsely vegetated or bare
ARTIF	artificial - buildings or roads
FORE	forest
UAAR	utilizable agricultural area
INLW	Inland waters

3.2.9 Annex: Detailed description of Eurostat data processing in COCO (coco1_eurostat.gms)

The program starts by importing pre-processed data from Eurostat. The pre-processing includes simple data selection routines and also manual checks. The Eurostat domains are processed one by one, and the corrections are done for each Member State [12]

Below we discuss the specific data-processing tasks related to Eurostat table groups. The first Eurostat Table Group is “p_AgriProd” covering market balances and activity levels.

Corrections and complements for all MS:

[12] Eurostat offers data for Belgium and Luxembourg separately, whereas the database combines both countries to the model region “BL000” (Belgium and Luxembourg). The key reason is that Eurostat offers data mainly for the aggregate Belgium and Luxembourg up to the year 1999, especially for all market balances. Furthermore, Luxembourg has a rather small agricultural sector (2004 total output was about EUR 250 million) with some similarities to Belgium.

- The following data are not anymore available form Eurostat, starting with the 2010 data extractionBeginning with Eurostat selection 2010 some data are missing from the Eurostat website:
  - DWH1, RAP1, POT1, POT2, ROO1 and ROO2 are not longer supported
  - data for slaughter heads and slaughter tons for calves are only available for recent years
  - deliveries to dairy of RMLK missing for earlier years in selection starting with February 2018

For an Interim solution, data for the missing data points are collected from an earlier Eurostat selection (March 2010).

- UNFCCC data is included, here sheep and goats population, to prolong data of some countries where Eurostat data collection stopped 2008/2009.
- Recent dairy sector data from Eurostat via DG supplements the ordinary dairy data downloaded from the website of Eurostat.
- Sugar trade data from the market balances of Eurostat is extended with Comext (Eurostat) data.
- For the milk products WMIO, SMIP, FRMI and COCM some market balance positionpositions are corrected: "industrial use" is added to "feed on market and "processing" is added to "human consumption.
- COCO code "FRUI" is aggregated from auxiliary data for fruit trees, plus soft fruits, plus strawberries.
- All activities for the aggregate ILAM are added up from SHEP and GOAT.
- The units for wine balance sheets are converted from 1000hl to 10000hl=1000000l
- A rice milled equivalent balance without paddy rice (separate product) is constructed.
- Survey data on buffaloes are used to increase the bovine stock data to cover the whole cattle herd.

Corrections and complements for specific MS:

Due to years of database updates, a number of corrections on input data are carried out. For special cases in some MS, data are read in from additional data sources:

- Belgium-Luxemburg: trade for potatoes (Eurostat: EU trade since 1988 by HS2-HS4 [DS-016894])
- France: market balances for cereal products (Agreste, Direction générale des douanes et droits indirects (DGDDI))
- Denmark: market balances for some cereal products (StatBank Denmark)
- Finland: market balances for some cereal products (Natural Resources Institute Finland, Balance sheet for food commodities)
- Germany: activity levels for textile crops (BMELF)
- Ireland: trade for citrus fruits and some milk products (Eurostat: EU trade since 1988 by HS2-HS4 [DS-016894]) and activity levels for grass land (StatBank Ireland)
- Austria: production of cow milk, fruit products and potatoes (Statistisches Amt Österreich)

- Czechia: trade of life animals (Eurostat: EU trade since 1988 by CN8 [DS-016890])
- Lithuania: human consumption cereal products (calculated from data from statistical yearbook 2018)
- Slovenia: slaughtering (SiStat Slovenia)
- Romania: data for the meat and in the milk sectors (Romanian experts)
- Trade data for sugar are collected from Eurostat COMEXT data.

The remaining domains/table groups only require a few case-by-case corrections:

- The second Eurostat Table Group is "p_ExchRate" covering exchange rates
- The third Eurostat Table Group is "p_EcoAct" covering the economic accounts for agriculture.
- The fourth Eurostat Table Group is "p_AgriPri" covering agricultural producer prices.

3.2.10 Annex: Testing procedure and checking intermediate steps in COCO (biofuels)

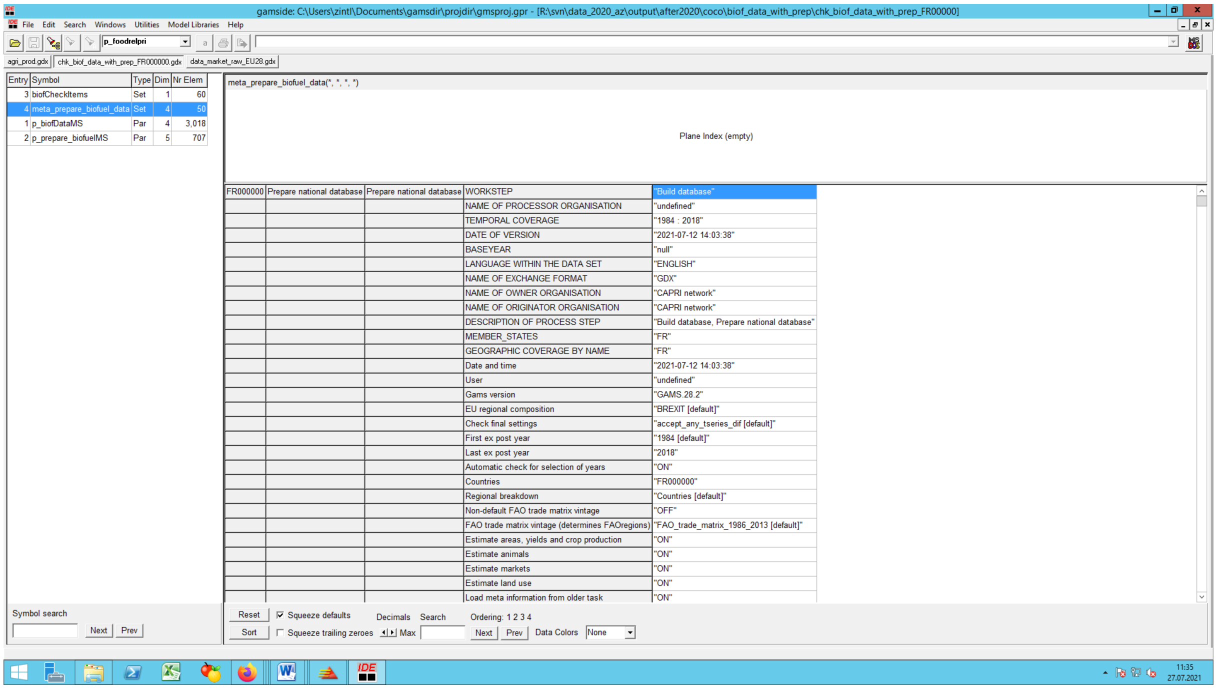
The COCO module produces various reporting files on the intermediate data processing steps. These files can be used to trace back potential errors in the COCO database to their origin. These debugging files also contain meta-information on the input data and settings used for producing the COCO database.

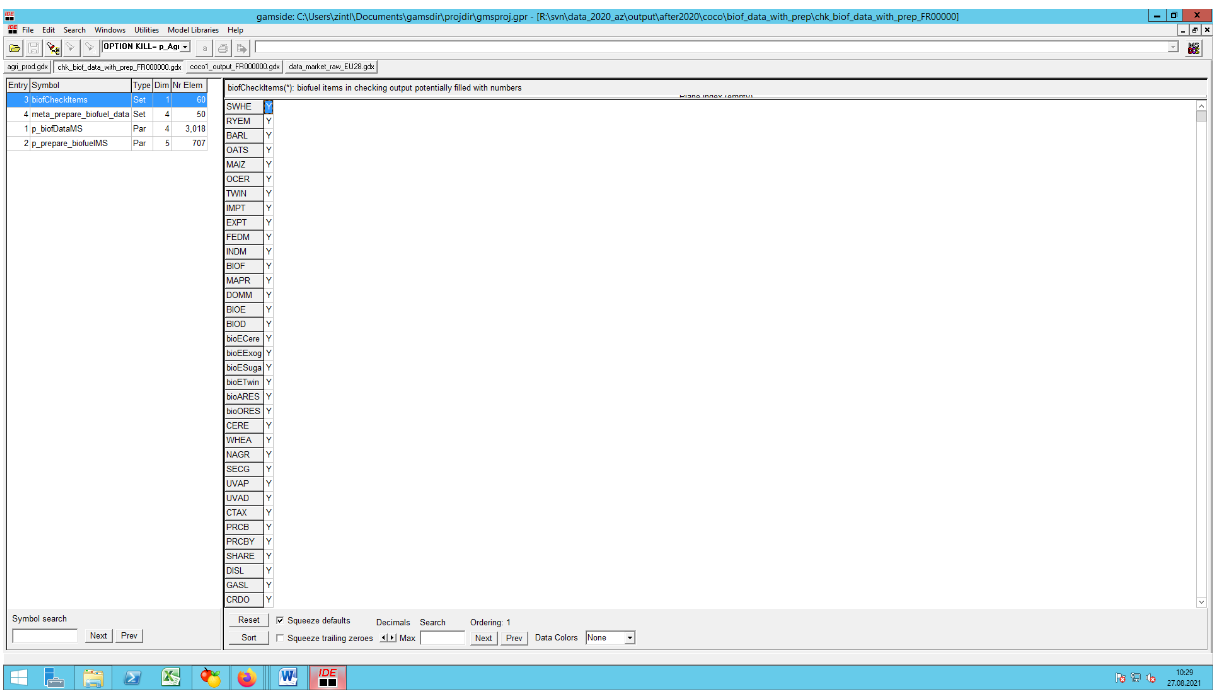
The following example is a walk-through on the typical data processing steps, covring biofuels data preparation in France.

The reporting file 'output/results/coco/biof_data_with_prep/chk_biof_data_with_prep_FR000000.gdx' reports on the data preparation for biofuels for France (FR000) in COCO1. The file includes the set 'meta_prepare_biofuel_data', with meta-information on the recent coco1 run (e.g. creation date of file, GAMS version used).

The set *biofCheckItems* in the same reporting .gdx file shows all biofuel items potentially filled with numbers.

The complete list of the biofuel items in *biofCheckItems* includes codes which are additional to the CAPRI activity codes (see Annex on code lists above). The full code list includes the following items:


gamside: C:\Users\zintl\Documents\gamsdir\projdir\gmsproj.gpr - [R:\svn\data_2020_az\output\after2020\coco\biof_data_with_prep\chk_biof_data_with_prep_FR00000]
File Edit Search Windows Utilities Model Libraries Help
p_foodrelpri
agri_prod.gdx
chk_biof_data_with_prep_FR000000.gdx
data_market_raw_EU28.gdx
Entry Symbol Type Dim Nr Elem
3 biofCheckItems Set 1 60
4 meta_prepare_biofuel_data Set 4 50
1 p_biofDataMS Par 4 3,018
2 p_prepare_biofuelMS Par 5 707
meta_prepare_biofuel_data(*, *, *, *)
Plane Index (empty)
FR000000 Prepare national database Prepare national database WORKSTEP "Build database"
NAME OF PROCESSOR ORGANISATION "undefined"
TEMPORAL COVERAGE "1984 : 2018"
DATE OF VERSION "2021-07-12 14:03:38"
BASEYEAR "null"
LANGUAGE WITHIN THE DATA SET "ENGLISH"
NAME OF EXCHANGE FORMAT "GDX"
NAME OF OWNER ORGANISATION "CAPRI network"
NAME OF ORIGINATOR ORGANISATION "CAPRI network"
DESCRIPTION OF PROCESS STEP "Build database, Prepare national database"
MEMBER_STATES "FR"
GEOGRAPHIC COVERAGE BY NAME "FR"
Date and time "2021-07-12 14:03:38"
User "undefined"
Gams version "GAMS 28.2"
EU regional composition "BREXIT [default]"
Check final settings "accept_any_tseries_dif [default]"
First ex post year "1984 [default]"
Last ex post year "2018"
Automatic check for selection of years "ON"
Countries "FR000000"
Regional breakdown "Countries [default]"
Non-default FAO trade matrix vintage "OFF"
FAO trade matrix vintage (determines FAOregions) "FAO_trade_matrix_1986_2013 [default]"
Estimate areas, yields and crop production "ON"
Estimate animals "ON"
Estimate markets "ON"
Estimate land use "ON"
Load meta information from older task "ON"
Symbol search
Next Prev
Reset Squeeze defaults Decimals Search Ordering: 1 2 3 4
Sort Squeeze trailing zeroes Max Next Prev Data Colors None
11:35
27.07.2021


gamside: C:\Users\zintl\Documents\gamsdir\projdir\gmsproj.gpr - [R:\svn\data_2020_az\output\after2020\coco\biof_data_with_prep\chk_biof_data_with_prep_FR00000]
File Edit Search Windows Utilities Model Libraries Help
OPTION KILL= p_Agi
agri_prod.gdx
chk_biof_data_with_prep_FR000000.gdx
coco1_output_FR000000.gdx
data_market_raw_EU28.gdx
Entry Symbol Type Dim Nr Elem
3 biofCheckItems Set 1 60
4 meta_prepare_biofuel_data Set 4 50
1 p_biofDataMS Par 4 3,018
2 p_prepare_biofuelMS Par 5 707
biofCheckItems(*): biofuel items in checking output potentially filled with numbers
SWHE Y
RYEM Y
BARL Y
OATS Y
MAIZ Y
OCER Y
TWIN Y
IMPT Y
EXPT Y
FEDM Y
INDM Y
BIOF Y
MAPR Y
DOMM Y
BIOE Y
BIOD Y
bioECere Y
bioEExog Y
bioESuga Y
bioETwin Y
bioARES Y
bioORES Y
CERE Y
WHEA Y
NAGR Y
SECG Y
UVAP Y
UVAD Y
CTAX Y
PRCB Y
PRCBY Y
SHARE Y
DISL Y
GASL Y
CRDO Y
Symbol search
Next Prev
Reset Squeeze defaults Decimals Search Ordering: 1
Sort Squeeze trailing zeroes Max Next Prev Data Colors None
10:29
27.08.2021

bioECere	Ethanol processed from cereals
bioESuga	Ethanol processed from sugar beets
bioETwin	Ethanol processed from wine
bioEFrui	Ethanol processed from fruits
bioEOcro	Ethanol processed from other agricultural crops
bioEExog	Ethanol processed from crops not explicit in biofuel modelling (fruits, potatoes, other crops)
bioARES	Biofuels processed from crops residues
bioORES	Biofuels processed from forest residues and waste material (municipal waste, waste oil, other waste)
SECG	Biofuel quantities from second generation
MAPRagr	Ethanol production from agricultural sources
EloBio	Biofuel production and demand data from DG Energy project EloBio
DG_Agri	Ethanol data from DGAgri website and supplementary files
ProdCom	Eurostat: PRODCOM ANNUAL SOLD (NACE Rev. 2.) [DS-066341]
EIA	Independent Statistics & Analysis, US Energy Information Administration
comext	Eurostat: Comext
Energy_bal	Eurostat: Supply, transformation, consumption - renewable energies - annual data [nrg_107a]
Energy_dem	Eurostat: Supply, transformation, consumption - renewable energies - annual data [nrg_102a, nrg_1073a]
final	results of the calculations
ODOM	other domestic use (activity from biostock calculations
APRagr	Ethanol production from agricultural sources
INDt	Sum of model results for BIOF and INDM
BIOi, INDi, DOMi	intermediate activities to save data from model initialisation for later documentation.

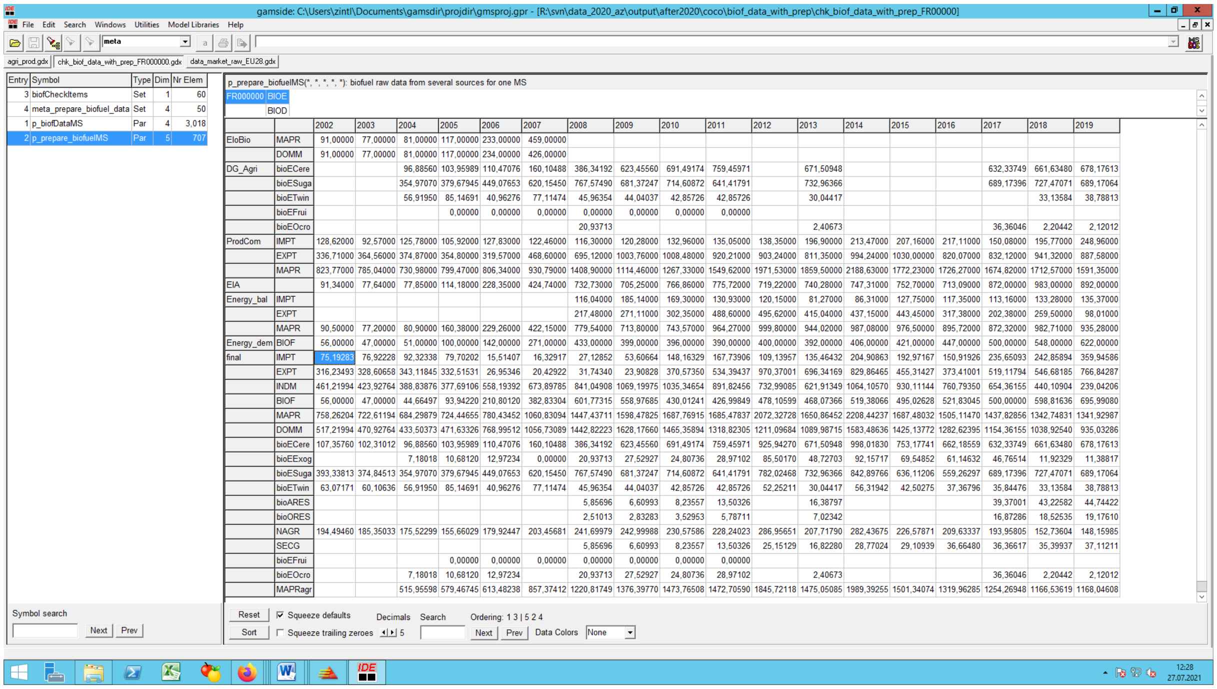
Biofuels production (levels) are calculated for biodiesel (BIOD) and bioethanol (BIOE). Input data and final initialization values before the consistency models are run are documented on the parameter *p_prepare_biofuelsMS* (see examples below). The results of the consistency models m_bioFitD (BIOD) and m_bioFitE (BIOE) are documented on the parameter *p_biofDatatMS* (see examples below).

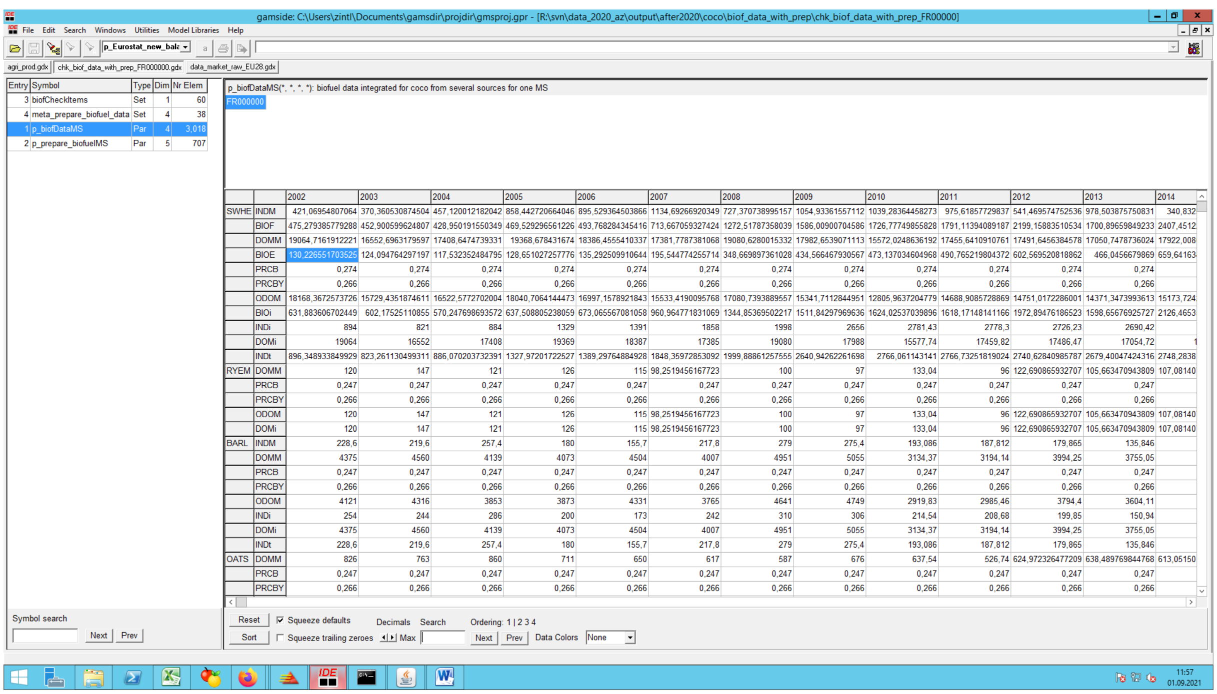
Example 1: Bioethanol

The screenshot demonstrates the input data and final initialization values collected on parameter *p_prepare_biofuelsMS*. The first column of the table indicates the data source, respectively the processing status of the data. Data sources for bioethanol (BIOE) include data from EloBio, DG_Agri, ProdCom, EIA, Engergy_bal and Energy_dem. The second column of the table shows the activity.

The results of the model m_bioFitE (BIOE) are documented on the parameter p_biofDatatMS.

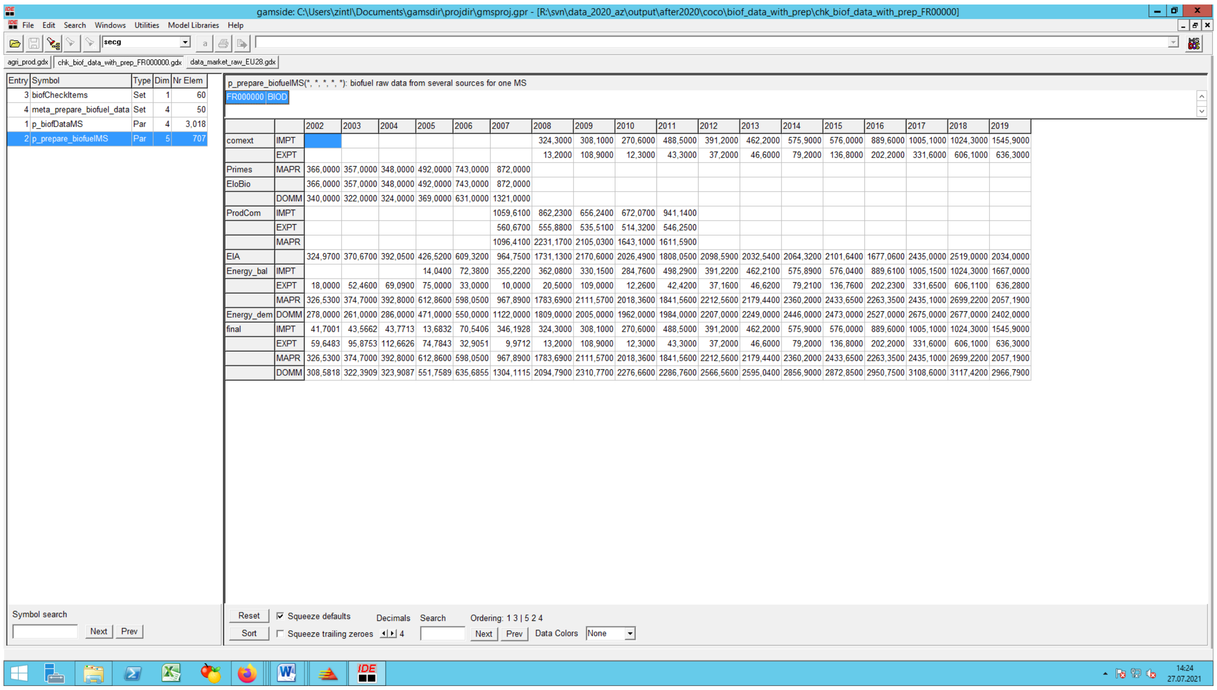
We take soft wheat (SWHE) as an example for biofuel feedstock, and walk through the initialization and consistency model results. From data input (Eurostat and FAO) we received in 2002 an industrial use of 894 1000t, saved on INDi. For production of bio-ethanol 631 1000t were initialized, saved on BIOi. The results of the breakdown by use for bio-ethanol and others industrial use, are saved on BIOF and INDM. BIOE shows the yield of soft wheat for bio-ethanol.


gamside: C:\Users\zintl\Documents\gamsdir\projdir\gmsproj.gpr - [R:\svn\data_2020_az\output\after2020\coco\biof_data_with_prep\chk_biof_data_with_prep_FR00000]
p_prepare_biofuelMS(*, *, *, *, *): biofuel raw data from several sources for one MS


gamside: C:\Users\zintl\Documents\gamsdir\projdir\gmsproj.gpr - [R:\svn\data_2020_az\output\after2020\coco\biof_data_with_prep\chk_biof_data_with_prep_FR00000]
p_biofDataMS(*, *, *, *): biofuel data integrated for coco from several sources for one MS

Example 2: Biodiesel

The first dimension of the reporting parameter *p_prepare_biofuels* shows the data source (processing status). The second dimension of the parameter shows the activity.



For Bio-diesels, PRIMES model results are used as an additional data source.

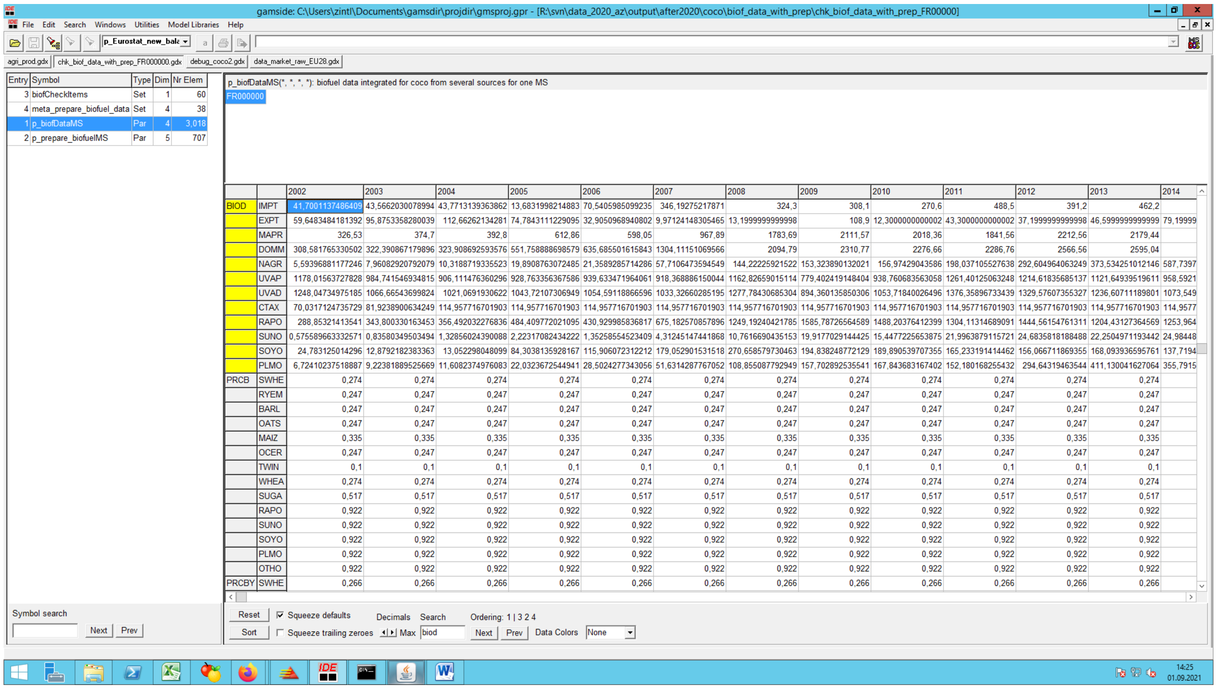
Data source code	Data source description
Primes	PRIMES MODEL, EC3MLAB of ICCS, National University of Athens

The parameter *p_biofDataMS* reports on production (MAPR), trade (import:IMPT, export:EXPT), production from non-agricultural sources (NAGR), prices (UVAD, UVAP) and consumer taxes (CTAX). The distribiutio of total biodiesel processing to the feedstock is also reported, for rapeseed oil (RAPO), sunflower oil (SUNO), soya oil (SOYO) and palm oil (PLMO).

3.2.11 Annex: Testing procedure and checking intermediate steps in COCO (dairy)

The following three examples show how to use the intermediate reporting files to trace the data preparation steps. Screenshots demonstrate the arrangement of the reporting parameters by using the CAPRI Graphical User Interface. COCO automatically produces the reporting files in the folder *results/coco/res_estima/*

Example 1: Production of cow (COMI) and sheep (SGMI) milk



In order to document the procedure of data consolidation and rebooking, we look at the reporting file for France “chk_estima_FR000.gdx”.

The codes in the rows show the activity code, the product code and its status. For activity codes see Annex 1: Code list.

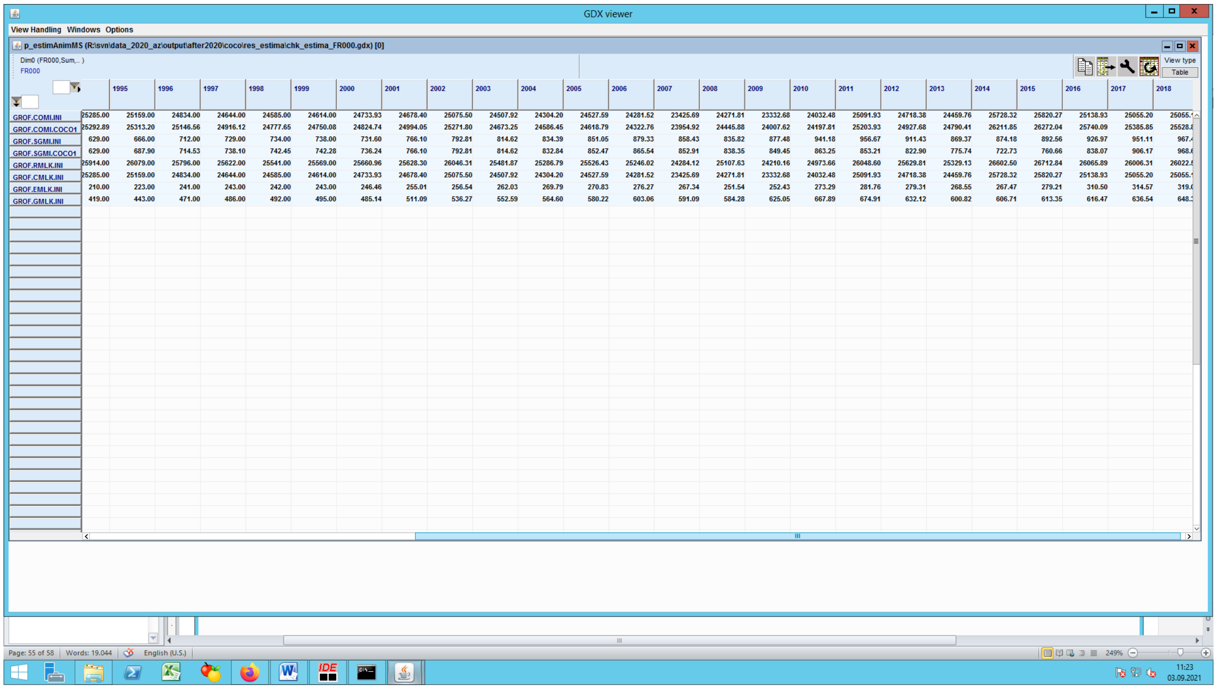
Status codes:

- INI: initial value
- COCO1: estimation value

The initialization of the production of COMI and SGMI is done in the module *coco1_milk.gms* (see section 3.1.3). Additional remarks to better understand the example:

- COMI: Milk from cows (CMLK) and buffaloes (BMLK) is added up.
- SGMI: Milk from ewes (EMLK) and goats (GMLK) is added up.
- If data on cow or sheep and goat milk is not available separately, but total milk production (RMLK) is available, then production of COMI is set equal to total milk production.
- Only COMI and SGMI are included in the estimation in *coco1_estima.gms*
- The production of RMLK and its components CMLK, BMLK, EMLK and GMLK are only copied from raw data tables into this check parameter for documentation purposes.

Example 2: data consolidation for cow milk



GDX viewer

View Handling Windows Options

p_estimAnimMS (R:\svn\data_2020_az\output\after2020\coco\res_estima\chk_estima_FR000.gdx) [0]

Dim0 (FR000,Sum,..)
FR000

View type
Table

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
GROF.COMI.INI	25285.00	25159.00	24834.00	24644.00	24585.00	24614.00	24733.93	24678.40	25075.50	24507.92	24304.20
GROF.COMI.COCO1	25292.89	25313.20	25146.56	24916.12	24777.65	24750.08	24824.74	24994.05	25271.80	24673.25	24586.45
GROF.SGMI.INI	629.00	666.00	712.00	729.00	734.00	738.00	731.60	766.10	792.81	814.62	834.39
GROF.SGMI.COCO1	629.00	687.90	714.53	738.10	742.45	742.28	736.24	766.10	792.81	814.62	832.84
GROF.RMLK.INI	25914.00	26079.00	25796.00	25622.00	25541.00	25569.00	25660.96	25628.30	26046.31	25481.87	25286.79
GROF.CMLK.INI	25285.00	25159.00	24834.00	24644.00	24585.00	24614.00	24733.93	24678.40	25075.50	24507.92	24304.20
GROF.EMLK.INI	210.00	223.00	241.00	243.00	242.00	243.00	246.46	255.01	256.54	262.03	269.79
GROF.GMLK.INI	419.00	443.00	471.00	486.00	492.00	495.00	485.14	511.09	536.27	552.59	564.60

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
GROF.COMI.INI	24527.59	24281.52	23425.69	24271.81	23332.68	24032.48	25091.93	24718.38	24459.76	25728.32	25820.27	25138.93	25055.20
GROF.COMI.COCO1	24618.79	24322.76	23954.92	24445.88	24007.62	24197.81	25203.93	24927.68	24790.41	26211.85	26272.04	25740.09	25385.85
GROF.SGMI.INI	851.05	879.33	858.43	835.82	877.48	941.18	956.67	911.43	869.37	874.18	892.56	926.97	951.11
GROF.SGMI.COCO1	852.47	865.54	852.91	838.35	849.45	863.25	853.21	822.90	775.74	722.73	760.66	838.07	906.17
GROF.RMLK.INI	25526.43	25246.02	24284.12	25107.63	24210.16	24973.66	26048.60	25629.81	25329.13	26602.50	26712.84	26065.89	26006.31
GROF.CMLK.INI	24527.59	24281.52	23425.69	24271.81	23332.68	24032.48	25091.93	24718.38	24459.76	25728.32	25820.27	25138.93	25055.20
GROF.EMLK.INI	270.83	276.27	267.34	251.54	252.43	273.29	281.76	279.31	268.55	267.47	279.21	310.50	314.57
GROF.GMLK.INI	580.22	603.06	591.09	584.28	625.05	667.89	674.91	632.12	600.82	606.71	613.35	616.47	636.54

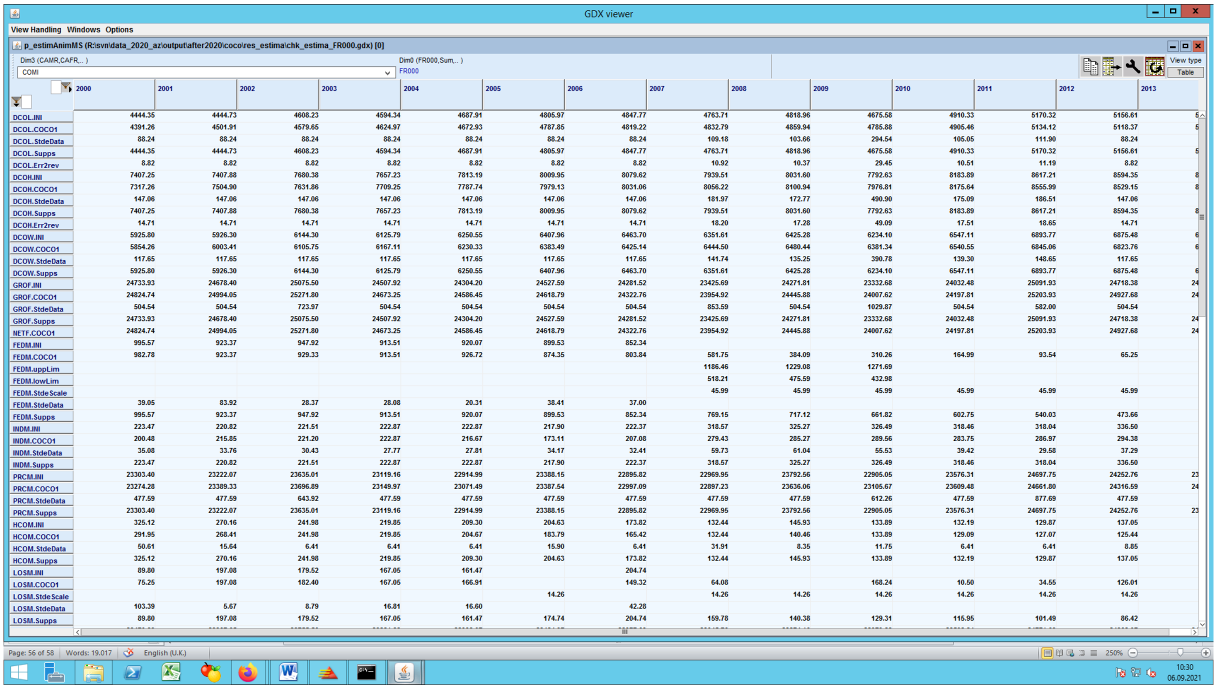
The procedure of data consolidation and rebooking of all activities for the CAPRI product "COMI" (cow milk) is shown in the following screenshot (only part of the reporting parameter p_estimAnimMS is shown, but the full scope of the table is visible in the GUI).

The codes in the rows show the activity code and its status. For activity codes see Annex 1: Code list. Additional codes for status include the following.

Status code	Status code description
StdeData	Final (small) Stde (standard deviation) attached to priors from raw data
StdeScale	Final (large) Stde attached to priors from trends but not from raw data
Upplim	Soft upper limits triggering extra penalties if violated
Lowlim	Soft lower limits triggering extra penalties if violated
Supps	Prior value = support: comes from raw data or trends plus HP filter
Err2rev	Original error term from preest: to steer speed of bound opening

Under activity dairy cows (DCOW) the following items are reported: yield, total production (GROF), feed use (FEDM) and losses on market (LOSM). Eurostat's *National Accounts of Agriculture (EAA)* only supply data for the aggregate milk (MILK). The equation *e_EAAMLK* in the consolidation model *AnimNSSQ* ensures the consistency of EAA values for MILK, as they are split up between cow milk (COMI) and sheep and goat milk (SGMI).

```
e_EAAMLK("%MS%000",T)
```



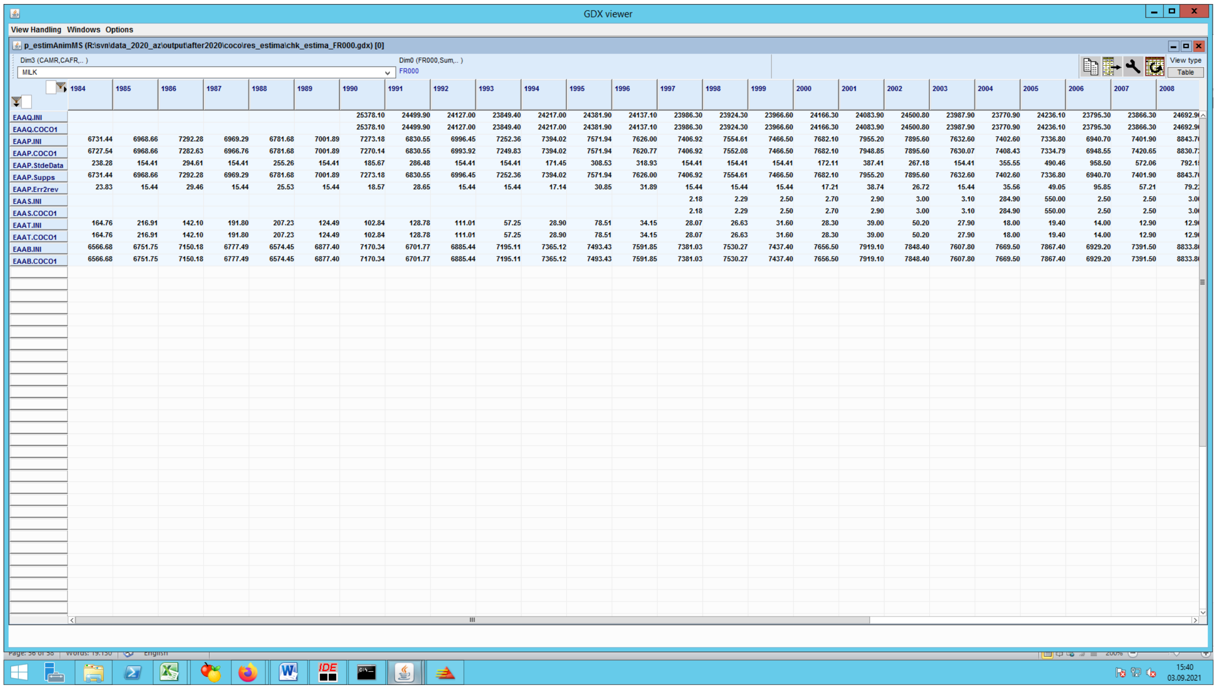

```
$ (p_NobsP("%MS%000","EAAP","MILK") AND ESTR("MILK") and
   (p_NobsP("%MS%000","EAAP","COMI") or p_NobsP("%MS%000","EAAP","SGMI"))) ..
*
     v_EstimY("%MS%000","EAAP","MILK",T) =E=
                    v_EstimY("%MS%000","EAAP","COMI",T) $ p_NobsP("%MS%000","EAAP","
                        COMI")
                  + v_EstimY("%MS%000","EAAP","SGMI",T) $ p_NobsP("%MS%000","EAAP","
                        SGMI") ;
```

Finally. the producer prices (UVAP) are calculated directly from the monetary EEA values and production. The following picture shows the data processing steps (states) for the EAA values for milk.

From the example for COMI above you can also understand the influence of the standard deviation from raw data (e.g. FEDM.StdeData), and standard deviation from trends (e.g. FEDM.StdeScale) Standard deviations are calculated both for raw data and the trends. For years where FEDM.StdeData is given, the results are very close to the prior values FEDM.Supps, whereas they are deviating sizeably for years where only FEDM.StdeScale is available.

The first initialisation of *StdeData* and *StdeScale* is done in module coco1_preest.gms, which is a pre-step for the data consolidation models (crops, animals, market balances), using a Hodrick-Prescott filter to smooth the combination of given values and trend line. Both standard deviations enter the objective function (see chapter 3.1.4).

Example 3: data consolidation for cow dairy cow activity (DCOW)



	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
EAAQ.INI							25378.10	24499.90	24127.00	23849.40	24217.00	24381.90	24137.10
EAAQ.COCO1							25378.10	24499.90	24127.00	23849.40	24217.00	24381.90	24137.10
EAAP.INI	6731.44	6968.66	7292.28	6969.29	6781.68	7001.89	7273.18	6830.55	6996.45	7252.36	7394.02	7571.94	7626.00
EAAP.COCO1	6727.54	6968.66	7282.63	6966.76	6781.68	7001.89	7270.14	6830.55	6993.92	7249.83	7394.02	7571.94	7620.77
EAAP.StdeData	238.28	154.41	294.61	154.41	255.26	154.41	185.67	286.48	154.41	154.41	171.45	308.53	318.93
EAAP.Supps	6731.44	6968.66	7292.28	6969.29	6781.68	7001.89	7273.18	6830.55	6996.45	7252.36	7394.02	7571.94	7626.00
EAAP.Err2rev	23.83	15.44	29.46	15.44	25.53	15.44	18.57	28.65	15.44	15.44	17.14	30.85	31.89
EAAS.INI													
EAAS.COCO1													
EAAT.INI	164.76	216.91	142.10	191.80	207.23	124.49	102.84	128.78	111.01	57.25	28.90	78.51	34.15
EAAT.COCO1	164.76	216.91	142.10	191.80	207.23	124.49	102.84	128.78	111.01	57.25	28.90	78.51	34.15
EAAB.INI	6566.68	6751.75	7150.18	6777.49	6574.45	6877.40	7170.34	6701.77	6885.44	7195.11	7365.12	7493.43	7591.85
EAAB.COCO1	6566.68	6751.75	7150.18	6777.49	6574.45	6877.40	7170.34	6701.77	6885.44	7195.11	7365.12	7493.43	7591.85

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
EAAQ.INI	23986.30	23924.30	23966.60	24166.30	24083.90	24500.80	23987.90	23770.90	24236.10	23795.30	23866.30	24692.9
EAAQ.COCO1	23986.30	23924.30	23966.60	24166.30	24083.90	24500.80	23987.90	23770.90	24236.10	23795.30	23866.30	24692.9
EAAP.INI	7406.92	7554.61	7466.50	7682.10	7955.20	7895.60	7632.60	7402.60	7336.80	6940.70	7401.90	8843.7
EAAP.COCO1	7406.92	7552.08	7466.50	7682.10	7948.85	7895.60	7630.07	7408.43	7334.79	6948.55	7420.65	8830.7
EAAP.StdeData	154.41	154.41	154.41	172.11	387.41	267.18	154.41	355.55	490.46	958.50	572.06	792.1
EAAP.Supps	7406.92	7554.61	7466.50	7682.10	7955.20	7895.60	7632.60	7402.60	7336.80	6940.70	7401.90	8843.7
EAAP.Err2rev	15.44	15.44	15.44	17.21	38.74	26.72	15.44	35.56	49.05	95.85	57.21	79.2
EAAS.INI	2.18	2.29	2.50	2.70	2.90	3.00	3.10	284.90	550.00	2.50	2.50	3.0
EAAS.COCO1	2.18	2.29	2.50	2.70	2.90	3.00	3.10	284.90	550.00	2.50	2.50	3.0
EAAT.INI	28.07	26.63	31.60	28.30	39.00	50.20	27.90	18.00	19.40	14.00	12.90	12.9
EAAT.COCO1	28.07	26.63	31.60	28.30	39.00	50.20	27.90	18.00	19.40	14.00	12.90	12.9
EAAB.INI	7381.03	7530.27	7437.40	7656.50	7919.10	7848.40	7607.80	7669.50	7867.40	6929.20	7391.50	8833.8
EAAB.COCO1	7381.03	7530.27	7437.40	7656.50	7919.10	7848.40	7607.80	7669.50	7867.40	6929.20	7391.50	8833.8

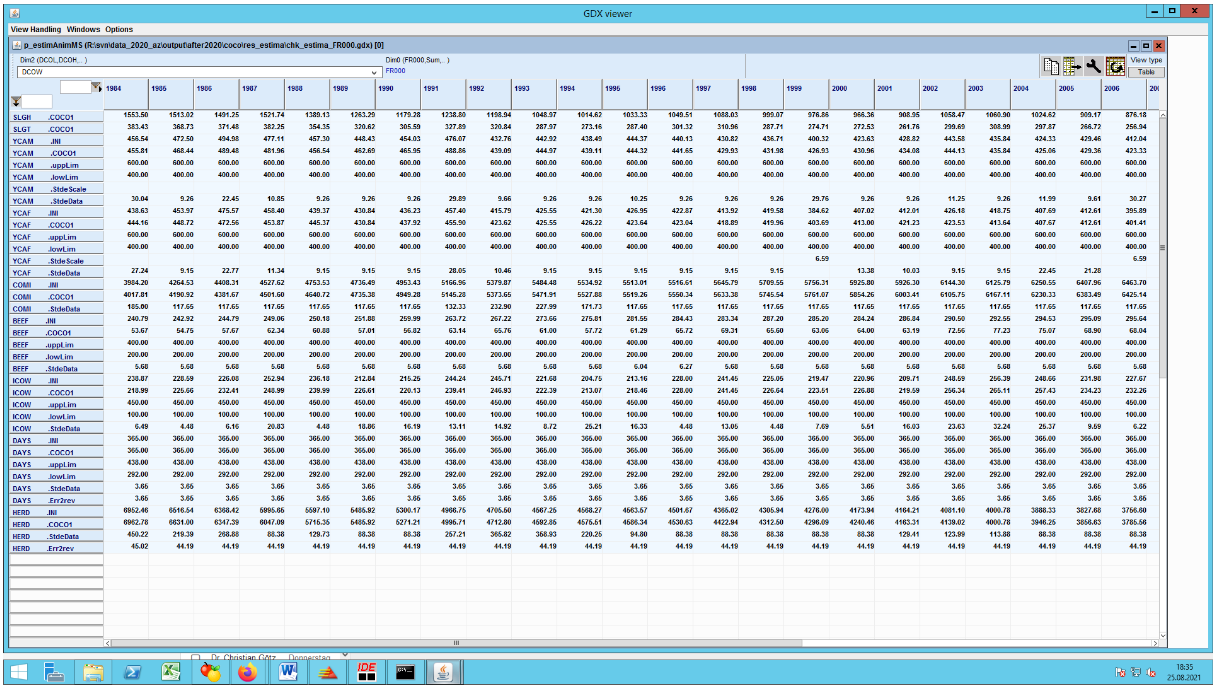
The procedure of data consolidation and booking intermediate data processing results for the dairy cow activity (DCOW) is demonstrated in the following screenshot.

The rows of the table show the product item code for the production activity DCOW, and the data processing steps (status). The first two lines show the coco1 results for slaughtering. The items starting with Y and I stand for the output and input of calves. The initialization, the estimation steps and the final results are all documented on the reporting parameter *p_estimAnimMS*. Items COMI and BEEF show the yields for cow milk and beef. Item DAYS is the process length, initialized by 365 days (equals one year). Finally, the item HERD models the herd size of dairy cows.

3.3 The Regionalised Data Base (CAPREG)

3.3.1 Data requirements and sources at the regional level

CAPRI aims at building up a Policy Information System of the EU's agricultural sector, regionalised at NUTS 2 level or farm types inside NUTS 2 regions with an emphasis on the impact of the CAP. The core of the system consists of a regionalized or farm type agricultural sector model using an activity based non-linear programming approach. One feature of such a highly disaggregated, activity based agricultural sector model is the detailed information resulting from *ex ante* simulations of policy scenarios concerning the output and input of specific agricultural production activities and their relationships. This information is also a pre condition to judge possible impacts of agricultural production on the environment. However, these systems require as well this kind of information (data) *ex-post*, at least partially. It is especially necessary to define for each region in the model, at least for the basis year,



		1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
SLGH	.COCO1	1553.50	1513.02	1491.25	1521.74	1389.13	1263.29	1179.28	1238.80	1198.94	1048.97	1014.62	1033.33
SLGT	.COCO1	383.43	368.73	371.48	382.25	354.35	320.62	305.59	327.89	320.84	287.97	273.16	287.40
YCAM	.INI	456.54	472.50	494.98	477.11	457.30	448.43	454.03	476.07	432.76	442.92	438.49	444.37
YCAM	.COCO1	455.81	468.44	489.48	481.96	456.54	462.69	465.95	488.86	439.09	444.97	439.11	444.32
YCAM	.uppLim	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00
YCAM	.lowLim	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00
YCAM	.StdeScale												
YCAM	.StdeData	30.04	9.26	22.45	10.85	9.26	9.26	9.26	29.89	9.66	9.26	9.26	10.25
YCAF	.INI	438.63	453.97	475.57	458.40	439.37	430.84	436.23	457.40	415.79	425.55	421.30	426.95
YCAF	.COCO1	444.16	448.72	472.56	453.87	445.37	430.84	437.92	455.90	423.62	425.55	426.22	423.64
YCAF	.uppLim	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00
YCAF	.lowLim	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00
YCAF	.StdeScale												
YCAF	.StdeData	27.24	9.15	22.77	11.34	9.15	9.15	9.15	28.05	10.46	9.15	9.15	9.15
COMI	.INI	3984.20	4264.53	4408.31	4527.62	4753.53	4736.49	4953.43	5166.96	5379.87	5484.48	5534.92	5513.01
COMI	.COCO1	4017.81	4190.92	4381.67	4501.60	4640.72	4735.38	4949.28	5145.28	5373.65	5471.91	5527.88	5519.26
COMI	.StdeData	185.00	117.65	117.65	117.65	117.65	117.65	117.65	132.33	232.90	227.99	171.73	117.65
BEEF	.INI	240.79	242.92	244.79	249.06	250.18	251.88	259.99	263.72	267.22	273.66	275.81	281.55
BEEF	.COCO1	53.67	54.75	57.67	62.34	60.88	57.01	56.82	63.14	65.76	61.00	57.72	61.29
BEEF	.uppLim	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00
BEEF	.lowLim	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00
BEEF	.StdeData	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	6.04
ICOW	.INI	238.87	228.59	226.08	252.94	236.18	212.84	215.25	244.24	245.71	221.68	204.75	213.16
ICOW	.COCO1	218.99	225.66	232.41	248.99	239.99	226.61	220.13	239.41	246.93	222.39	213.07	218.46
ICOW	.uppLim	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00
ICOW	.lowLim	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
ICOW	.StdeData	6.49	4.48	6.16	20.83	4.48	18.86	16.19	13.11	14.92	8.72	25.21	16.33
DAYS	.INI	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00
DAYS	.COCO1	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00
DAYS	.uppLim	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00
DAYS	.lowLim	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00
DAYS	.StdeData	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65
DAYS	.Err2rev	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65
HERD	.INI	6952.46	6516.54	6368.42	5995.65	5597.10	5485.92	5300.17	4966.75	4705.50	4567.25	4568.27	4563.57
HERD	.COCO1	6962.78	6631.00	6347.39	6047.09	5715.35	5485.92	5271.21	4995.71	4712.80	4592.85	4575.51	4586.34
HERD	.StdeData	450.22	219.39	268.88	88.38	129.73	88.38	88.38	257.21	365.82	358.93	220.25	94.80
HERD	.Err2rev	45.02	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19

		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
SLGH	.COCO1	1049.51	1088.03	999.07	976.86	966.36	908.95	1058.47	1060.90	1024.62	909.17	876.18
SLGT	.COCO1	301.32	310.96	287.71	274.71	272.53	261.76	299.69	308.99	297.87	266.72	256.94
YCAM	.INI	440.13	430.82	436.71	400.32	423.63	428.82	443.58	435.84	424.33	429.46	412.04
YCAM	.COCO1	441.65	429.93	431.98	426.93	430.96	434.08	444.13	435.84	425.06	429.36	423.33
YCAM	.uppLim	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00
YCAM	.lowLim	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00
YCAM	.StdeScale											
YCAM	.StdeData	9.26	9.26	9.26	29.76	9.26	9.26	11.25	9.26	11.99	9.61	30.27
YCAF	.INI	422.87	413.92	419.58	384.62	407.02	412.01	426.18	418.75	407.69	412.61	395.89
YCAF	.COCO1	423.04	418.89	419.96	403.69	413.00	421.23	423.53	413.64	407.67	412.61	401.41
YCAF	.uppLim	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00	600.00
YCAF	.lowLim	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00
YCAF	.StdeScale				6.59							6.59
YCAF	.StdeData	9.15	9.15	9.15		13.38	10.03	9.15	9.15	22.45	21.28	
COMI	.INI	5516.61	5645.79	5709.55	5756.31	5925.80	5926.30	6144.30	6125.79	6250.55	6407.96	6463.70
COMI	.COCO1	5550.34	5633.38	5745.54	5761.07	5854.26	6003.41	6105.75	6167.11	6230.33	6383.49	6425.14
COMI	.StdeData	117.65	117.65	117.65	117.65	117.65	117.65	117.65	117.65	117.65	117.65	117.65
BEEF	.INI	284.43	283.34	287.20	285.20	284.24	286.84	290.50	292.55	294.53	295.09	295.64
BEEF	.COCO1	65.72	69.31	65.60	63.06	64.00	63.19	72.56	77.23	75.07	68.90	68.04
BEEF	.uppLim	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00
BEEF	.lowLim	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00	200.00
BEEF	.StdeData	6.27	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68
ICOW	.INI	228.00	241.45	225.05	219.47	220.96	209.71	248.59	256.39	248.66	231.98	227.67
ICOW	.COCO1	228.00	241.45	226.64	223.51	226.88	219.59	256.34	265.11	257.43	234.23	232.26
ICOW	.uppLim	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00
ICOW	.lowLim	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
ICOW	.StdeData	4.48	13.05	4.48	7.69	5.51	16.03	23.63	32.24	25.37	9.59	6.22
DAYS	.INI	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00
DAYS	.COCO1	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00	365.00
DAYS	.uppLim	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00	438.00
DAYS	.lowLim	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00	292.00
DAYS	.StdeData	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65
DAYS	.Err2rev	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65	3.65
HERD	.INI	4501.67	4365.02	4305.94	4276.00	4173.94	4164.21	4081.10	4000.78	3888.33	3827.68	3756.60
HERD	.COCO1	4530.63	4422.94	4312.50	4296.09	4240.46	4163.31	4139.02	4000.78	3946.25	3856.63	3785.56
HERD	.StdeData	88.38	88.38	88.38	88.38	88.38	129.41	123.99	113.88	88.38	88.38	88.38
HERD	.Err2rev	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19	44.19

the **matrix of I/O-coefficients** for the different production activities together with **prices** for these outputs and inputs. Moreover, for calibration and validation purposes information concerning **land use and livestock numbers** is necessary.

Already from the beginning of the development of the CAPRI model, the regional agricultural statistics (EUROSTAT table group reg_agr) was judged as the only harmonized data source available on regionalized agricultural data in the EU. Other regional Eurostat data are suplementing the regional agricultural statistics such that we are currently using the following:

- Land use from regional landuse statistics [agr_r_landuse, discontinued table]
- Land cover from LUCAS [lan_lcv_ovw, currently only used in COCO1]
- Crop production - harvested areas, production and yields [table agr_r_crops]
- Animal production - livestock numbers [table agr_r_animal]
- Milk production [agr_r_milkpr]
- Agricultural accounts on regional level [table agr_r_accts]
- Structure of agricultural holdings including labour force [ef_ls_ovlsureg, ef_olslsureg, ef_oluaareg, ef_oluaareg, ef_r_nuts]

Although the content of the regional datasets has remained in time, the naming and classification within EUROSTAT is undergoing continuous modifications. Tables considered of low interest are discontinued (and may be still used in CAPRI some time after this point, such as table agr_r_landuse). And new topics are covered providing useful data in some areas, for example from agri-environmental indicators (table reg_aei):

- Estimated soil erosion by water, by NUTS 3 regions (aei_pr_soiler)
- Manure storage facilities by NUTS 3 regions (aei_fm_ms)

The following table shows the availability of the different regional tables as they have been used in the current database (with series completed up to 2014). However, the current coverage concerning time and sub-regions differs dramatically between the tables and within the tables between the Member States. A second problem consists in the relatively high aggregation level especially in the field of crop production. Hence, additional sources, assumptions and econometric procedures must be applied to close data gaps and to break down aggregated data.

Table 6 Availability of regional datain current database after 1983

Table	Official availability
Land use	from 1974 yearly
Crop production (harvested areas, production and yields)	from 1975 yearly
Animal production (livestock numbers)	from 1977 yearly
Agricultural accounts on regional level	from 1980 yearly
Structure of agricultural holdings and labour force	2000, 2003, 2005, 2007, 2010, 2013

Source: capri/dat/capreg/regio_data_all.gdx

3.3.2 Methodology applied in the regional data consolidation

In the last major update of 2015 the original data had been first stored in the TSV format designed by EUROSTAT:

- Unordered List ItemIn a first step, these files had been converted by an excel macro into csv format and an overall set with all items including their long text has been created to prepare further processing.
- In a second step these alredy GAMS readable files are stored in GDX format in folder “dat/capreg” and under version control. Meta data are added in the process as well.

The results of these two steps is a single large tables, which comprise time series of all data retrieved from Eurostat for all tables: land use, crop production, animal populations, cow’s milk collection and agricultural accounts.

The starting point of the methodological approach is the decision to use the consistent and complete national data base (COCO) as a frame or reference point for any regionalization. In other words, any aggregation of the main data items (areas, herd sizes, gross production and intermediate use, unit value prices and EAA-positions) of the regionalized data over regions must match the national values. This is the general rule with some exceptions.

Given that starting position, the following approaches are generally applied:

- Unordered List ItemData as loaded from the regional statistics are subject to some manual consistency checks (in gams/capreg/check_and_cor_regio.gms) as well as checks for regional consistency. The latter is mainly true for animal herd sizes where we have data at the same or even more disaggregated level as found in COCO.
- Gaps in regional data are completed and data only given at a higher aggregation level as required in CAPRI are broken down by using existing national information.
- Fall back and other rules for assignments are structurally and (often) numerically identical for all regional units and groups of activities and inputs/outputs.
- Econometric analysis or additional data sources are used to close gaps.

All the approaches described in the following sub sections are only thought as a first crude estimate. Wherever additional data sources are available, their content should be checked and is often used to overcome the list of these 'easy to use' estimates presented in here. Examples are (some) data for Norway, Sweden or Luxembourg that have been collected from national sources. The procedures described in here can be thought as a 'safety net' to ensure that regionalized data are technically available but not as an adequate substitute for collecting these data from additional sources.

Prices The agricultural domain of REGIO does not cover regionalized prices. For simplicity, the regional prices are therefore assumed to be identical to sectoral ones[13]:

$$UVAG_r = UVAG_s \tag{3.31}$$

Young animal prices are a special case since they are not included in the COCO data base (the current methodology of the EAA does not value intermediate use of animals) but are necessary to calculate income indicators for intermediate activities (e.g. raising calves). Only exported or imported live animals are implicitly accounted for by valuing the connected meat imports and exports.

Young animals are valued based on the 'meat value' and assumed relationships between live and carcass weights. Male calves (ICAM, YCAM) are assumed to have a final weight of 55 kg, of which 60 % are valued at veal prices. Female calves (ICAF, YCAF) are assumed to have a final weight of 60 kg, of which 60 % are valued at veal prices. Young heifers (IHEI, YHEI) are assumed to have a final weight of 300 kg, of which 54 % are valued at beef. Young bulls (IBUL, YBUL) are assumed to have a final weight of 335 kg, of which 54 % are valued at beef. Young cows (ICOW, YCOW) are assumed to have a final weight of 575 kg, of which 54 % are valued at beef. For piglets (IPIG, YPIG), price notations were regressed on pig meat prices and are assumed to have a final weight of 20 kg of which 78 % are valued at pig meat prices. Lambs (ILAM, YLAM) are assumed to weight 4 kg and are valued at 80 % of sheep and goat meat prices. Chicken (ICHI, YCHI) are assumed to weight 0.1 kg and are valued at 80 % of poultry prices.

Another special case are sugar beet prices. They are still determined in a program (*'sugar/price_est.gms'*) inherited from the 2003 EuroCARE sugar study (Henrichsmeyer et al. 2003). It determines sugar beet prices according to the sugar prices, levies and partial survey results in the 90ies. The estimation results are subsequently used to determine the beet price differentiation also in subsequent years. It is

[13] There is no easy way to relax this assumption if no further data sources are available.

noteworthy that the same program is applied in CAPREG (via quotasprices.gms) and in CAPMOD (via data_prep.gms) to determine base year beet prices.

Activity Levels In cases where data on regional activity levels are missing, a linear trend line is estimated for regional and Member State time series in the definition of the regional database. The gap is then filled with a weighted average between the trend line – using a weight of R^2 - and a weighted average of the available observations around the gap, using a weight of 1-R^2. The specific formulation has the following properties. In cases of a strong trend in a time series, the back-casted and forecasted numbers will be dominated by the trend as the weight of R^2 will be high. With decreasing R^2, the estimated values will be pulled towards known values.

Apart from gap filling another problem is that in annual cropland statistics at the regional level only cover a few crop activities (cereals with wheat, barley, grain maize, rice; potatoes, sugar beet, oil seeds with rape and sunflower; tobacco, fodder maize; grassland, permanent crops with vineyards and olive plantations). The COCO data base, however, covers some 30 different crop activities. In order to break these aggregates down to COCO definitions, the national shares of the aggregate are used.

As an example, this approach is explained for cereals. Data on the production activities WHEA (wheat = SWHE+DWHE), BARL (barley), MAIZ (grain maize) and PARI (paddy rice) as found in COCO match directly the level of disaggregation in the regional data. Therefore, the mapped regionalized data are directly set equal to the corresponding values in the regional "raw" data. The difference between the sum of these 4 activities and the aggregate data on cereals in the regional raw data must be equal to the sum of the remaining activities in cereals as shown in COCO, namely RYE (rye and meslin), OATS (oats) and OCER (other cereals). As long as no other regional information is available, this difference from the regional raw data is hence broken down applying national shares.

The approach is shown for OATS in the following equations, where the suffix r stands for regional data:

$$
\begin{aligned}
LEVL_{OATS,r} = (&CEREAL_r \\
&- WHEAT_r - BARLEY_r - MAIZEGR_r - RICE_r)\cdot \\
&\frac{LEVL_{OATS,COCO}}{(LEVL_{OATS,COCO} + LEVL_{RYE,COCO} + LEVL_{OCER,COCO})}
\end{aligned}
\tag{3.32}
$$

Similar equations are used to break down other aggregates and residual areas in the regional data [14]. The Farm Structure Survey (FSS) provides crop areas for a larger number of crops but this survey is usually conducted only every three years. Data from FSS, when available, is also used to aproximate crop areas at regional level.

One important advantage of the approach is the fact that the resulting areas are automatically consistent to the national data if the ingoing information from REGIO was consistent to national level. Fortunately, the regional information on herd sizes covers most of the data needed to give nice proxies for all animal activities in COCO definition. The regional data break down for herd sizes is often more detailed than COCO at least for the important sectors. Regional estimates for the activity levels are therefore the result of an aggregation approach, in opposite to crop production.

[14] If no data at all are found, the share on the utilisable agricultural area is used.

In order to generate good starting points for the following steps of data processing and to avoid systematic deviations between regional and national levels in the following consistency steps, all regional level in REGIO are first scaled with the relation between the (national) results in COCO and the regional results when aggregated to the national level (key file is gams/capreg/map_from_regio.gms).

Besides technological plausibility and a good match with existing regional statistics, the regionalized data for the CAPRI model must be also consistent to the national level. The minimum requirement for this consistency includes activity levels and gross production. The "initialisation" of the regional database has been undertaken already to meet this requirement as good as possble but cannot guarantee it. Consistency for activity levels is therefore based on Highest Posterior Density Estimator which ensures (in gams/capreg/cons_levls.gms):

1. Adding up of activity levels from lower regional level (NUTS II, NUTS I) to higher ones (NUTS I, NUTS 0)
2. Adding up of crop areas to UAA at regional level.

The objective function minimizes in case of animal herds simple squared relative deviations from the herds. In case of crops, a 25% weight for absolute squared difference of the crop shares on UAA plus 75% deviation of relative squared differences is introduced. In the crop sector consistency is also imposed to regional transition matrices for 6 UNFCCC land use categories relevant for carbon accounting (forest land, cropland, grassland, settlements, wetlands, residual land) which are initialised from the national transition matrix estimated in the COCO1 module.

A specific problem is the fact that land use statistics do not report a break down of idling land into obligatory set aside, voluntary set aside and fallow land[15]. Equally, the share of oilseeds grown as energy crops on set aside needs to be determined. An Highest Posterior density estimator is used (in gams/capreg/cal_seta.gms) to 'distribute' the national information on the different types of idling land to regional level, with the following restrictions:

- Obligatory set-aside areas must be equal to the set-aside obligations derived from areas and set-aside rates for Grandes Cultures (which may differ at regional level according to the share of small producers). For these crops, activity levels are partially endogenous in the estimation in order to allow a split up of oilseeds into those grown under the set-aside obligations and those grown as non-fo-od crops on set-aside.
- Obligatory and voluntary set-aside cannot exceed certain shares of crops subjects to set-aside (at least before Agenda 2000 policy)
- Fallow land must equalise the sum of obligatory set-aside, voluntary set-aside and other idling land.
- Total utilisable area must stay constant.

In some cases, areas reported as fallow land are smaller than set-aside obligations. In these cases, parts of grassland areas and 'other crops' are allowed to be reduced.

Production and yields The proceedure for gross output (GROF) is similar to the one for activity levels, as correction factors are applied to line up regional yields with given national production:

[15] The necessary additional information on non-food production on set-aside, obligatory and voluntary set-aside areas can be found on the DG-AGRI web server.

$$
\begin{aligned}
CORR_{GROF,o} &= \sum_{j,r} Levl_{j,r} O_{j,r} / GROF_{o,n} \\
O^*_{j,r} &= O_{j,r} \cdot CORR_{GROF,o}
\end{aligned}
\tag{3.33}
$$

In case of missing statistical information for regional yields, national yields are used. A special rule is used for fodder maize yields, where regional yields are derived from national fodder maize yields, and the relation between regional and national average cereal yields.

For grassland and fodder from arable land, missing yields are derived from national ones using the relation between regional and national stocking densities of ruminants, in combination with assumed share of concentrates in terms of a weighted sum of energy and protein per ruminant activity in CAPRI. Those shares are then scaled with a uniform factor to exhaust on average the available energy and protein from concentrates at the national level. Accordingly, higher fodder yields are expected where ruminant stocking densities are high, acknowledging differences in concentrate shares. If e.g. the stocking densities solely stem from sheep and goat, the assumed impacts on yields is higher. In order to avoid unrealistic low or high yields, those are bounded to a 25%-400% range compared to the regional aggregate.

The input allocation in any given year should not be linked to realised, but to expected yields. Expected yields are constructed using the following modified Hodrick-Prescott filter:

$$
\min \quad hp = 1000 \sum_{1<t<T-1} (y^*_{t+1} - y^*_{t-1})^2 + \sum_t (y^*_t - y_t)^2 \tag{3.34}
$$

where y covers all output coefficients in the data base. The Hodrick-Prescott filter is applied both at the national and regional level after any gaps in the time series had been closed.

3.3.3 Final steps of regional data completion

The regional database modules also cover some aspects which are discussed in other parts of this documentation.

- For policy data at the regional level (mostly premium related data) see Section 3.6. These policy related assignments require a good part of the CAPREG module
- For the fertiliser and feed allocations and environmental indicators, also important elements of the regional database, see the next Section 3.4
- Towards the end of the regional data base consolidation supply side PMP parameters are calibrated as a final test of consistency and sometimes to serve as starting values for the subsequent baseline calibration (in *gams/capreg/pmp.gms*)

3.3.4 Build and compare time series of GHG inventories

The regionalised data base module CAPREG runs in two steps:

- The first steps prepares regional time series covering activities, production, land use and the fertiliser allocation

- The second step involves more time consuming processing steps which are therefore only executed for the selected base year: feed allocation, computation of GHG results, and the final calibration test

To assess the reliability of the CAPRI database in terms of GHG results against official UNFCCC notifications, results from the first step (time series) were insufficient, as the GHG accounting also requires information on the feed allocation. This problem was addressed within the scope of the IDEAg (Improving the quantification of GHG emissions and flows of reactive nitrogen) project[16], where an option has been introduced to allow for a consistent accounting of GHG emissions over time. This is able to combine input information from CAPREG time series runs as well as (short run, nowcasting-style) CAPMOD simulation results. Furthermore, an R-based tool was introduced to the CAPRI GUI that maps GHG emissions data from CAPRI to the GHG emission balances contained in the National Inventory Reports (NIRs) that are submitted annually by countries in compliance with UNFCCC GHG reporting obligations.

3.4 Input Allocation

The term input allocation describes how aggregate input demand (e.g. total anorganic N fertiliser use in Denmark) is 'distributed' to production activities. The resulting activity specific data are called input coefficients. They may either be measured in value (€/ha) or physical terms (kg/ha). The CAPRI data base uses physical terms and, where not available, input coefficient measured in constant prices.

Micro-economic theory of a profit maximising producer requires revenue exhaustion, i.e. marginal revenues must be equal to marginal costs simultaneously for all realised activities. The marginal physical input demand multiplied with the input price exhausts marginal revenues, leading to zero marginal profits. Marginal input demands per activity can only be used to define aggregate input demand if they are equal to average input demands. The latter is the case for the Leontief production function.

The advantage of assuming a Leontief technology in agricultural production analysis is the fact that an explicit link between production activities and total physical input use is introduced (e.g. environmental indicators can be linked directly to individual activities or activity specific income indicators, since gross margins can be calculated). The disadvantage is the rather rigid technology assumption. We would for example expect that increasing a crop share in a region will change the average soil quality the crop uses, which in turn should change yields and nutrient requirements. It should hence be understood that the Leontief assumption is an abstraction and simplification of the 'real' agricultural technology in a region. The assumption is somewhat relaxed in CAPRI as two 'production intensities' are introduced.

Input coefficients for different inputs are constructed in different ways which will be discussed in more detail in the following sections:

- *For nitrate, phosphate and potash*, nutrient balances are constructed so to take into account crop and manure nutrient content and observed fertiliser use, combined with gaseous losses. These balances ex post determine the effective input coefficients and regional availability of manure and overfertilisation parameters.

[16]The IDEAg project was commissioned by the JRC-IES in Ispra in 2015 and was carried out by the Thünen Institute in cooperation with the JRC-IES (August 2015 – August 2016). A more detailed explanation of the CAPRI task "Build GHG inventories" and its use has been prepared by the Thünen contributors at the time, Sandra Marquardt and Alexander Gocht, see capri/doc/GHG_inventory_module.docx.

- *For feed*, the input calculation is rooted in a mix of engineering knowledge (requirement functions for animal activities, nutrient content of feeding stuff, recommendations on feed mix), observed data ex post (total national feed use, national feed costs), combined within a Highest Posterior Density (HPD) estimation framework.
- *For the remaining inputs*, estimation results from a FADN sample in the context of the CAPSTRAT project (2000-03) are combined with current aggregate national input demand reported in the EAA and standard gross margin estimations, again using a HPD estimation framework.

3.4.1 Input allocation excluding young animals, fertiliser and feed

There is a long history of allocating inputs to production activities in agricultural sector analysis, dating back to the days where I/O models and aggregate farm LPs where the only quantitative instruments available. In these models, the input coefficients represented a Leontief technology, which was put to work in the quantitative tools as well. However, input coefficients per activity do not necessary imply a Leontief technology. The allocated input demands can be seen as marginal ones (which are identical to average ones in the Leontief case) and are then compatible with flexible technologies as well.

Input coefficients can be put to work in a number of interesting fields. First of all, activity specific income indicators may be derived, which may facilitate analyzing results and may be used in turn to define sectoral income. Similarly, important environmental indicators are linked to input use and can hence be linked to activities as well with the help of input coefficients.

Given the importance or the input allocation, the CAP STRAT project (2000-2003) comprised an own work package to estimate input coefficients. On a first step, input coefficients were estimated using standard econometrics from single farm record as found in FADN. Additionally, tests for a more complex estimation framework building upon entropy techniques and integrating restrictions derived from cost minimization were run in parallel. The need to accommodate the estimation results with data from the EAA in order to ensure mutual compatibility between income indicators and input demand per activity and region on the one hand, and sectoral income indicators as well as sectoral input use on the other, requires deviating from the estimated mean of the coefficients estimated from single farm records. Further on, in some cases estimates revealed zero or negative input coefficients, which cannot be taken over. Accordingly, it was decided to set up a second stage estimation framework building upon the unrestricted estimates from FADN. The framework can be applied to years where no FADN data are available, and thus ensures that the results will be continuously used for the years ahead, before an update of the labor-intensive estimations is again necessary and feasible.

As a result of the unrestricted estimation based on FADN [17]a matrix of input coefficients for 11 input categories (Total Inputs, Crop Only Inputs, Animal Only Inputs, Seeds, Plant Protection, Fertilizer, Other Crop Inputs, Purchased and Non-Purchased Feeds and Other Animal Only Inputs) and their estimated standard errors is available. Some of those coefficients are related to the output of a certain activity (e.g. how much money is spend on a certain input to produce one unit of a product), some of them are related to the acreage of on activity (input costs per activity level).

All of the econometric coefficients were required to be transformed into an 'activity level' form, due to

[17] More details on the FADN estimation were reported in older versions of the CAPRI documentation, accessible in the /doc folder of any stable release of the CAPRI system up to star 2.4 from `https://www.capri-model.org/dokuwiki/doku.php?id=capri:get-capri`.

the fact that this is the definition used in the CAPRI model. Before this could be done, it seemed necessary to fill up the matrix of estimated coefficients because some estimates were missing and others were negative. In order to this we constructed a number of coefficients that were weighted averages among certain groups. These mean coefficients were the following.

1. *Mean coefficients of activity groups.* Each activity was allocated to a certain group (e.g. soft wheat belonged to cereals). For each group we built weighted averages among the positive estimates within a group using the estimated t statistics as weights. This coefficient only existed if there was at least one positive estimate inside that group and was then used to replace the gaps inside the coefficient matrix. If that mean coefficient was not available, due to no positive estimate inside a group at all, the next type of mean coefficients became relevant:

2. *Mean coefficients for an activity among European regions.* This second type of mean coefficients calculates weighted averages among three types of regional clusters. These clusters are Northern European States, Southern European states and all European regions. Again, the estimated t statistics were used as aggregation weights. Unfortunately, this type of averages did not fill all gaps in the coefficient matrix as there were some activities that had no positive estimate over the entire EU. For those the third type of mean coefficients was calculated.

3. *Mean coefficients for activity groups among regional clusters.* Here we calculated for the three regional clusters the averages of the first type of mean coefficients. As even the latter are synthetic, we gave each mean of them the same weight. Fortunately there was only a small probability that this coefficient did not exist for one of the groups as this was only the case if no coefficient inside a group over the entire EU had a positive estimate, which was not the case.

Following these rules we finally got a matrix of estimated and synthetic calculated input coefficients for both, the 'per activity level' and the 'per production' unit definition[18]. For the synthetic one there was no estimated standard error available but we wanted to use those later on. So we assumed them –to reflect that these coefficients have only weak foundation– to have a t statistic of 0.5.

The 'per level' definition was only taken over if the coefficient was really estimated or if no per production unit definition did exist. To transfer the latter into per activity level definition, we multiplied them with the average yield (1985 2001) of the respective activity. The resulting coefficients and their standard errors were then used a HPD approach as a *first set of priors*[19].

Missing econometric estimates and compatibility with EAA figures were not the only reasons that made a reconciliation of estimated inputs coefficients necessary. Moreover, the economic sense of the estimates could not be guaranteed and the definition of inputs in the estimation differed from the one used in CAPRI. Therefore we decided to include further prior information on input coefficients in agriculture. The *second set of priors* in the input reconciliation was therefore based on data from the EAA. Total costs of a certain input within an activity in a European Member State was calculated by multiplying the total expenditures on that input with the proportion of the total expected revenue of that activity

[18] In addition, a similar procedure (using slightly different groups) was applied to constructing coefficients for the 'Other' activities (e.g. OCER, OFRU, OVEG), which had been omitted from the econometric estimations. They are given the average group coefficient, unless there is none; then they are given the average northern or southern European coefficient as appropriate.

[19] The previously described completions are implemented in file gams/input/fill_inp_matrix.gms. Adjustments were made for scaling issues with regard to eggs for certain countries, and grass for Finland. In addition, when 'CAFR','CAFF' and 'HEIR' did not have econometric data, they assumed the coefficients and standard errors of 'CAMR', 'CAMF' and 'HEIF' respectively (CAPRI activity code definitions in the Annex).

to that of all activities using the input. Total expected revenue in this case was the production value (including market value and premiums) of the respective activity. If this resulted in a certain coefficient being calculated as zero due to missing data, then this coefficient would be replaced by one from a similar activity e.g. a zero coefficient for 'MAIF' would be replaced by the coefficient for 'GRAS'

This kind of prior information tries to give the results a kind of economic sense. For the same reason the *third type of priors* was created based on standard gross margins for agricultural activities received from EUROSTAT. Those existed for nearly all activities. The set from 1994 was used, since this was the most complete available. Relative rather than absolute differences were important, given the requirement to conform to EAA values[20].

Given the three types of prior information explained above –estimated input coefficients, data from EAA and standard gross margins , a HPD estimator has been used to reconcile the prior information on input coefficients. Accounting constraints ensure (see in "dist_input.gms") first that gross margins for an activity is the difference between expected revenue per activity level of that activity and the sum over all inputs used in that activity and second that the sum over all activities of their activity levels multiplied with an input gives the total expenditures on that input given by the EAA. The estimation is carried out in GAMS within and run for each year in the database. Some bounds are further set to avoid estimates running into implausible ranges.

The Highest Posterior Density estimation yields monetary input coefficients for the fertiliser types (Nitrate, Phosphate, Potassium), seeds, plant protection, feeds, pharmaceutical inputs, repairs, agricultural service input, energy and other inputs. While some of these can be directly used in the CAPRI model, we need special treatments for others –e.g. fertilisers, because they are used in physical units inside the model, and feeds, since they are much more disaggregated.

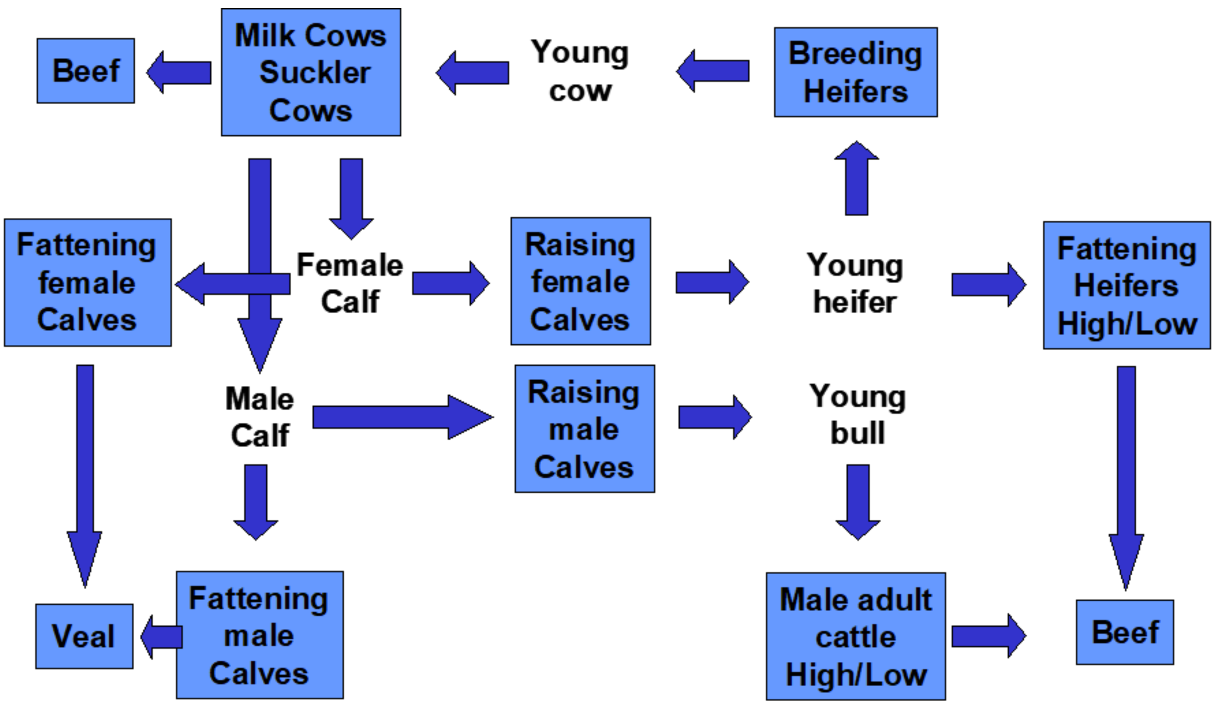
3.4.2 Input allocation for young animals and the herd flow model

Figure below shows the different cattle activities and the related young animal products used in the model. Milk cows (DCOL, DCOH) and suckler cows (SCOW) produce male and female calves (YCAM, YCAF). The relation between male and female calves is estimated ex post in the COCO framework. These calves are assumed to weigh 50 kg at birth (see gams/feed/feed_decl.gms) and to be born on the 1st of January. They enter immediately the raising processes for male and female calves (CAMR, CAFR) which produce young heifers (YHEI, 300 kg live weight) and young bulls (YBUL, 335 kg). The raising processing are assumed to take one year, so that calves born in t enter the processes for male adult fattening (BULL, BULH), heifers fattening (HEIL, HEIH) or heifers raising (HEIR) on the 1st January of the next year t+1. The heifers raising process produces then the young cows which can be used for replacement or herd size increasing on the first of January of t+2. The table below the diagram shows a numerical example (for DK, 1999-2001) for these relationships.

Accordingly, each raising and fattening process takes exactly one young animal on the input side. The raising processes produce exactly one animal on the output side which is one year older. The output

[20] Contrary to the econometric estimated priors, the two other types were different in different years, since the reconciliation had to be done for each year in the database. The second prior type is year specific by nature, as the EAA values differ between years. In case of standard gross margins, unfortunately, we had them only for one year (1994). So we decided to 'drive them over time' using the proportion of expected revenue of an activity in a certain year to that in the year 1994. Furthermore it may be mentioned that for plant protection coefficients a fourth set of priors from an industry source has been used and that energy inputs also received a special treatment in the key file gams/input/dist_inputs.gms.

Figure 5: The cattle chain



Source: CAPRI Modelling System

of calves per cow, piglets per sow, lambs per mother sheep or mother goat is derived ex post, e.g. simultaneously from the number of cows in t-1, the number of slaughtered bulls and heifers and replaced in t+1 which determine the level of the raising processes in t and number of slaughtered calves in t. The herd flow models for pig, sheep and goat and poultry are similar, but less complex, as all interactions happen in the same year, and no specific raising processes are introduced.

Table 7: Example for the relation inside the cattle chain (Denmark, 1999-2001)

		1999	2000
Male calves used in t and born in t			
DCOWLEVL	Number of dairy cows	667,03	654,08
DCOWYCAM	Number of male calves born per 1000 dairy cows	420,72	438,62
Number of males calves born from dairy cows		280,63	286,89
SCOWLEVL	Number of suckler cows	127,36	126,91
SCOWYCAM	Number of male calves born per 1000 suckler cows	420,72	411,83
Number of male calves born from suckler cows		53,58	52,27
Number of all male calves born		334,22	339,16
GROFYCAM	Number of male calves produced	334,21	339,16
CAMFLEVL	Number of male calves fattened	81,32	72,57
CAMRLEVL	Activity level of the male calves raising process	252,89	266,59
Sum of processes using male calves		334,21	339,16
GROFYCAM	Number of male calves used	334,21	339,16
Female calves used in t and born in t			
DCOWLEVL	Number of dairy cows	667,03	654,08
DCOWYCAF	Number of female calves born per 1000 dairy cows	404,15	421,58
Number of female calves born from dairy cows		269,58	275,75
SCOWLEVL	Number of suckler cows	127,36	126,91
SCOWYCAF	Number of male calves born per 1000 suckler cows	404,15	398,04
Number of female calves born from suckler cows		51,47	50,52
Number of all female calves born		321,05	326,26
GROFYCAF	Number of female calves produced	321,05	326,27
CAFFLEVL	Number of female calves fattened	26,64	28,74
CAFRLEVL	Activity level of the female calves raising process	294,41	297,53
Female calves used in t and born in t		321,05	326,27
GROFYCAF	Number of female calves used	321,05	326,27
Young bulls used in t and young bulls produced in t			
BULFLEVL	Activity level of the bull fattening process	262,94	252,89
GROFIBUL	Number of young bulls used	262,94	252,89
GROFYBUL	Number of young bulls raised from calvs	252,89	266,59
CAMRLEVL	Activity level of the male calves raising process	252,89	266,59
Heifers used in t and heifers produced in t			
HEIFLEVL	Activity level of the heifers fattening process	64,36	67,25
HEIRLEVL	Activity level of the heifers raising process	235,45	227,16
Sum of heifer processes		299,81	294,41
GROFIHEI	Number of heifers used	299,81	294,41
GROFYHEI	Number of heifers raised from calves	294,41	297,53
CAFRLEVL	Activity level of the female calves raising process	294,41	297,53
Cows used in t and heifers produced in t			
DCOWLEVL	Number of dairy cows	667,03	654,08
DCOWICOW	Number of young cows needed per 1000 dairy cows	332,01	332,5
Sum of young cows needed for the dairy cow herd		221,46	217,48
DCOWSLGH	Slaugthered dairy cows	221,47	217,48
SCOWLEVL	Number of suckler cows	127,36	126,91
SCOWICOW	Number of young cows needed per 1000 suckler cows	332,01	332,48
Sum of young cows needed for the suckler cow herd		42,28	42,20
SCOWSLGH	Slaugthered suckler cows	42,29	42,19
Sum of slaughtered cows		263,76	259,67
GROFICOW	Number of young cows used	263,75	259,67
Stock change in dairy cows	(DCOWLEVL(t+1)-DCOWLEVL(t)	-12,95	-22,16
Stock change in suckler cows	(SCOWLEVL(t+1)-SCOWLEVL(t)	-0,45	-2,06
Sum of stock changes in cows		-13,4	-24,22

The table above is taken from the COCO data base. In some cases, regional statistical data or estimates for number of young animals per adult are available, but in most cases, all input and output coefficients relating to young animals are identical at regional and national level. Nevertheless, experiences with simulations during the first CAPRI project phase revealed that a fixed relationship between meat output and young animal need as expressed with on bull fattening process overestimates the rigidity of the technology in the cattle chain, where producers may react with changes in final weights to relative changes in output prices (meat) in relation to input prices (feed, young animals). A higher price for young animals will tend to increase final weights, as feed has become comparatively cheaper and vice versa. In order to introduce more flexibility in the system, the dairy cow, heifer and bull fattening processes are split up each in two processed as shown in the following table.

Table 8: Split up of cattle chain processes in different intensities

	Low intensity/final weight	High intensity/final weight
Dairy cows (DCOW)	DCOL: 60% milk yield of average, variable inputs besides feed an young animals at 60% of average	DCOH: 140% milk yield of average, variable inputs besides feed an young animals at 140% of average
Bull fattening (BULF)	BULL: 20% lower meat output, variable inputs besides feed an young animals at 80% of average	BULH: 20% higher meat output, variable inputs besides feed an young animals at 120% of average
Heifers fattening (HEIF)	HEIL: 20% lower meat output, variable inputs besides feed an young animals at 80% of average	HEIH: 20% higher meat output, variable inputs besides feed an young animals at 120% of average

3.4.3 Input allocation for feed

The input allocation for feed describes how much kg of certain feed categories (cereals, rich protein, rich energy, feed based on dairy products, other feed) or single feeding stuff (fodder maize, grass, fodder from arable land, straw, milk for feeding) are used per animal activity level[21].

The input allocation for feed takes into account nutrient requirements of animals, building upon requirement functions. The input coefficients for feeding stuff shall hence ensure that energy, protein requirements, etc. cover the nutrient needs of the animals. Further on, ex post, they should be in line with regional fodder production and total feed demand statistics at national level, the latter stemming from market balances. And last but not least, the input coefficients together with feed prices should lead to reasonable feed cost for the activities.

Estimation of fodder prices Since the last revision of the EAA, own produced fodder (grass, silage etc.) is valued in the EAA. Individual estimates are given for fodder maize and fodder root crops, but no break down is given for fodder on arable land and fodder produced as grassland as presented in the CAPRI data base. The difference between grass and arable land is introduced, as conversion of grass to arable land is forbidden under cross compliance conditions so that marginal values of grassland and arable land may be different.

[21]The reader should notice again that the activity definition for fattening processes are slaughtered plus exported minus imported animals and not stable places.

The price attached to fodder should reflect both its nutritional content and the production costs at regional level. The entropy based estimation process tries to integrate both aspects.

The following equations are integrated in the estimator. Firstly, the regional prices for ‘grass’, ‘fodder on arable land’ and ‘straw’ (fint) multiplied with the fed quantities at regional level must exhaust the vale reported in the economic accounts, so that the EAA revenues attached to fodder are kept unchanged:

$$\sum_{r,fint} \overline{FEDUSE}_{r,fint} PFOD_{r,fint} = \overline{EAAP}_{OFAR,MS} + \overline{EAAP}_{GRAS,MS} \tag{3.35}$$

Secondly, the Gross Value Added of the fodder activities is defined as the difference between main revenues (from main fodder yield), other revenues, and total input costs based on the input allocation for crops described above.

$$GVAM_{r,fint} = \overline{YIELD}_{r,fint} PFOD_{r,fint} + \overline{OREV}_{r,fint} - \overline{TOIN}_{r,fint} \tag{3.36}$$

Other revenues may be from the nutrient value in crop residues. Next, an HDP objective is added which penalises deviations from the a priori mode.

The a priori mode for the prices of ‘grass’ and ‘other fodder on arable land’ are the EAAP values divided by total production volume which is by definition equal to feed use. The price of straw for feed use is expected to be at 1 % of the grass price.

Supports for Gross Value Added per activity are centred around 150 % of the value of total inputs as allocated by the rules and algorithm described above, with wide bounds.

Wide supports for the Gross Value Added of the fodder activities mirror the problem of finding good internal prices but also the dubious data quality both of fodder output as reported in statistics and the value attached to it in the EAA. The wide supports allow for negative Gross Value Added, which may certainly occur in certain years depending on realised yields. In order to exclude such estimation outcomes as far as possible an additional constraint is introduced:

$$GVAM_{r,fint} \geq \overline{TOIN}_{r,fint} \overline{gvafac} \tag{3.37}$$

The parameter $gvafac$ is initialised with zero so that first a solution is tried where all activities have positive GVAs. If infeasibilities arise, the factor is stepwise increased until feasibility is achieved, to ensure that estimated fodder prices are giving the minimal number of activities with negative Gross Value Addeds.

Calibration of the feed allocation The allocation of feed to animal activities has been changed several times (like the fertiliser allocation). The most recent version has been developed [22] in the Stable

[22] This section draws upon a corresponding Star 2 deliverable and coding which are due in major parts to CAPRI expert Markus Kempen. As Markus was not involved in this documentation, he is released from any responsibility for remaining errors. A more detailed version of this section is offered as `https://www.capri-model.org/dokuwiki/lib/exe/fetch.php?media=docu_feed_calib.pdf`.

Release 2 (in the following: "Star2") project which will become also the standard version in the CAPRI trunk at the next opportunity.

General concept

In the "pre-Star2"[23] implementation, based on the CAPRI model procedures, the objective in the data consolidation in tasks "build regional database" (capreg base year) and "baseline calibration supply" (capmod, baseline mode) is to cover the daily needs per animal with the available feed stuff (considering the daily feed intake capacity). In CAPRI most parameters determining the actual requirements of animals can be derived from statistics, e.g. milk yield, final live weight, daily gain, Apart from the uncertainty of statistical data, the calculated requirements can be seen as the "true" requirements in a country or region, as the differences between different animal nutrition literature sources are usually small. Nonetheless uncertainty in the data derived parameters can often lead to an over- or underestimation of the requirements in a range of 5-20% from the computed average need. This uncertainty may be taken into account when specifying the objective function for the required allocation model in a high posterior density (hpd) approach where the uncertainty on feeding requirements is expressed in terms of a standard deviation. This basic approach also underlies the "pre-star2" feed allocation. The pre-star2 feed calibration approach also considered two economic indicators that depend on the feed allocation:

- Feed costs and
- Gross margins, in particular the avoidance of negative gross margins [24]

These two criteria have been abandoned because technical plausibility was considered more important for the feed allocation than the derived value items. It may be argued that uncertainty in feed prices should not be transferred to the physical coefficients which is a consequence when considering both in the objective. Furthermore, the pmp approach of CAPRI has proven able to cope with negative margins even though it is admitted that they may not be entirely plausible.

In the pre-star2 CAPRI approach minimum and maximum bounds on specified feeding stuffs are specified to ensure technical plausibility, but to prevent infeasibilities they left considerable degrees of freedom. Additional hard constraints were for lysin and fiber contents of feed. However, a detailed analysis revealed that the purpose of these restrictions to ensure plausible feed ratios, for example regarding the relation of concentrate feed and roughage, was often missed. It has been decided therefore to skip these constraints.

The revised feed allocation methodology includes several new additional terms in its objective to capture technical plausibility beyond the animal requirements in terms of energy and protein and technical reproducibility of the calibration approach. These will be explained in more detail in the following sections.

Equations An overview of the equations used in the old and new feed allocation procedure is given in Table below. The objective function has changed significantly and more details on this will be discussed below. The equations ensuring consistency among production and consumption of feed, as well consistency across regional levels are unchanged.

[23] It has to be acknowledged that the specificaiton described in this section is not activated by default in CAPRI task "build regional database" whereas it is active in CAPRI task "Calibrate supply models". This setting will be changed shortly.

[24] Note that this refers to gross margins of animal activties, not to the gross margins of fodder activities which have been addressed in the previous section.

Table 9: Equations used in old and new feed allocation routine

equation			
old	new	description	comment
hpdFeed	hpdFeed_	objective function	changed significantly (see following section)
FEDUSE_	FEDUSE_	Balance for feeding stuff regional	needed to achieve consistency between produced feed and feed input to
FEDUSEA_	FEDUSEA_	Aggregation to regional feed input coefficient to aggregate one	
FEDUSES_	FEDUSES_	Fixation for feeding stuff regional in calibration	
REQSE_	REQSE_	Requirements of animals written as equality	for energy ENNE and crude protein CRPR
REQSN_		Requirements of animals written as in-equality	other requirements (lysine, dry matter and fibre)
MINSHR_		Maximum feed shares	Constraints on single feed stuff not used as hard bounds in new version
MAXSHR_		Minimum feed shares	Constraints on single feed stuff not used as hard bounds in new version
CST_	CST_	Definition of feed cost from feed input coefficients and prices	Feed cost in new version only for monitoring, not in objective or constraints
MEANDEV_		Definition of average deviation from requirements for all herds	oversupply by animal type was pulled against the mean oversupply.
	NutContFeed_	Nutrition content in the feed aggregates supplied to an animal category	nutrient content (per kg dry matter) is part of the objective
	FEDAGGR_	aggregate to roughage, concentarte feed, etc	Defines feed aggregates from single bulks FEED
	FeedAggrShare_	Calculate share of feed aggregates (roughage, concentrates, other)	shares of roughage and concentrate feed enter objective
	MeanFeedTotal_	Calculates total feed intake in DM per animal	Part of revised objective function

The four additional equations developed in the new feed allocation procedure are described in more detail in the following.

NutContFeed_

For nutrient content (energy, crude protein) in the total feed mix or in concentrate feed recommendations are frequently given in the animal nutrition literature. The equation NutContFeed_ calculates this based on the estimated feed input coefficients and the data on nutrient content and dry matter per feeding stuff.

```
*
*   --- nutrition content in total feed of animal category (energy and protein per kg dry matter: for calibration)
*
     NutContFeed_(RUNR,MAACT,A,REQMSE,FeedAggr) $ (    p_NutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MIN")
                                                   or  p_NutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MAX")
                                                   or  p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MEAN")) ..
*
     v_NutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr) =E=
*           nutrition in feed diveded by dry matter of feed
            SUM((Feed_FeedAggr(FEED,FeedAggr)),
                     v_feedInpCoeff(RUNR,MAACT,A,FEED)
                          * SUM(R_RAGG(RUNR,MSACT),%data%(MSACT,REQMSE,FEED,"Y")))

            / (SUM((Feed_FeedAggr(FEED,FeedAggr)),
                     v_feedInpCoeff(RUNR,MAACT,A,FEED)
                          * SUM(R_RAGG(RUNR,MSACT),%data%(MSACT,"DRMA",FEED,"Y"))) + 0.0001);
```

A small number is added to the denominator to avoid division by zero (e.g. while gams is searching for a feasible solution)

FedAggr_

```
* ---------- aggregate to roughage, concentrate feed, etc (in DRMA, needed in calibration step) -----
*
    FEDAGGR_(RUNR,MAACT,A,FeedAggr) $ p_technFact(RUNR,MAACT,"LEVL",A) ..
*
      v_feedInpCoeff(RUNR,MAACT,A,FeedAggr)
        =E= SUM(Feed_FeedAggr(Feed,FeedAggr), v_feedInpCoeff(RUNR,MAACT,A,FEED)
                                            * SUM(R_RAGG(RUNR,MSACT),%data%(MSACT,"DRMA",FEED,"Y")));
*
```

An aggregation of specific feeding stuff to aggregates (roughage, concentrates) is done since prior shares as well as minimum and maximum shares are more often found in the literature for aggregates than for single feedstuffs. The mapping is shown in Table below. It has been specified basically by putting into the "other" category all "special" items. Therefore, straw is a component of this "other" category rather than "roughage".

Table 10: Mapping feeding stuff to feed aggregates

	FGRA	FMAI	FOFA	FROO	FCOM	FSGM	FSTR	FCER	FPRO	FENE	FMIL	FOTH
FeedRough	X	X	X	X								
FeedCons								X	X	X	X	
FeedOth					X	X	X					X
FeedTotal	X	X	X	X	X	X	X	X	X	X	X	X

FeedAggrShare_

MeanFeedTotal_

One of the aggregates calculated is the total feed intake per animal. It is expected that, inspite of regional differences in fodder supply, this total feed intake is mostly a genetic characteristic of animals and hence should not vary markedly across regions. To influence this distribution in the objective, the average across regions needs to be computed.

```
*
*       --- calculate share of concentrate and roughage
*
    FeedAggrShare_(RUNR,MAACT,A,FeedAggr) $ (  p_animReq(RUNR,MAACT,A,"DRMA")) ..
*
            v_feedInpCoeff(RUNR,MAACT,A,FeedAggr) / (v_feedInpCoeff(RUNR,MAACT,A,"FeedTotal") + 1E-12)

                             =E= v_FeedAggrShare(RUNR,MAACT,A,FeedAggr) ;
*
```

```
*       *** definition of mean total feed intake across sub regions
*
    MeanFeedTotal_  (RUNAGG,MAACT,A) $  SUM(RUNR, p_AnimReq(RUNR,MAACT,A,"ENNE")) ..
*
        SUM(RUNR , v_actLevl(RUNR,MAACT,A) * v_feedInpCoeff(RUNR,MAACT,A,"FeedTotal"))
      / (SUM(RUNR , v_actLevl(RUNR,MAACT,A)) + 1E-6)

                 =E= v_feedInpCoeff(RUNAGG,MAACT,A,"FeedTotalSubRegionAvg");
```

Objective function

The objective function is extensively revised compared to the pre-star2 versions. The criteria to be optimised are now:

1. coverage of animal requirements with feed
2. regional variation of certain feed input coefficients
3. concentration of energy and protein in feed mix
4. shares of feed aggregates (roughage, concentrates, other) in total feed mix
5. feed input coefficients of all FEED bulks receive prior expectations

The parameters in the objective function are partly means and imputed standard deviations AND so-called "soft" upper and lower limits. The "soft" limits increase the penalty significant when the solver picks values close to or even beyond them.

Coverage of animal requirements with feed

```
*         ---- All relative deviations per animal are weighted with the "importance" of that animal type for the total sector
*              Otherwise the solver may try to fix any balancing needs with modifications on 1-2 'large' activities
*              because this is 'cheaper' in terms of the penalties than fiddling around with all activities

*
*         ---- deviation of requirements from mean of a priori expectation
*
           (
           - 1E5 * SUM( (RUNR,MAACT,A,REQMSE) $ p_AnimReq(RUNR,MAACT,A,REQMSE),
                        SQR( (v_animReq(RUNR,MAACT,A,REQMSE) - p_AnimReq(RUNR,MAACT,A,REQMSE))
                           /  p_AnimReq(RUNR,MAACT,A,REQMSE))
*              weighting:
                      * [(v_actLevl(RUNR,MAACT,A)+.01)*p_AnimReq(RUNR,MAACT,A,REQMSE)*p_animProdDays(RUNR,MAACT,A)]**.1)
```

This part of the objective functions tries to minimize the difference between the requirements calculated from the feed input coefficients (v_animReq) and the expected (mean) requirements (p_animReq) coming from literature. Due to the weighting with number of animals (v_actLevl) and expected requirements (p_animReq) the optimal solution tends to distribute over or under supply of nutrients relatively even over all activities and regions. It has been decided to attach an exponent smaller one to these weights which strongly pulls them towards unity (see: [...] FIXME (section? .1). This tends to give more weight to "less important" animal types compared with untransformed weights.

Deviation of sub regional total feed intake from regional average

```
---- deviation of sub regional total feed intake from sub regional average (dry intake equal across regions)

  - 1E4 * SUM( (RUNAGG,RUNR,MAACT,A) $ p_AnimReq(RUNR,MAACT,A,"ENNE"),
               SQR( (v_feedInpCoeff(RUNR,MAACT,A,"FeedTotal")-v_feedInpCoeff(RUNAGG,MAACT,A,"FeedTotalSubRegionAvg"))
                  / (v_feedInpCoeff(RUNAGG,MAACT,A,"FeedTotalSubRegionAvg") + 1E-3)))
```

As argued above, we expect that total feed intake in DRMA is mostly a genetic characteristic of animals and hence should not vary markedly across regions. Deviations of (sub-)regional feed intake from the associated regional average (NUTS1 or MS) are therefore penalised.

Deviations of sub regional feed input coefficients of non-ruminants from regional average

```
---- deviation of regional feed intake from aggregate intake for non-ruminants
     (non-ruminants do not depend on regional fodder production, should be similar across regions))

  - 1E4 * SUM( (RUNAGG,RUNR,MNRUMI,A,FEED) $ p_maxFeedShare(RUNR,MNRUMI,A,Feed),
               SQR( (v_feedInpCoeff(RUNR,MNRUMI,A,FEED) - v_feedInpCoeff(RUNAGG,MNRUMI,A,FEED))
                  / (v_feedInpCoeff(RUNAGG,MNRUMI,A,FEED) + 1E-4)))
```

As the comment explains, non-ruminants should have a rather standardised diet across regions.

Concentration of energy and protein in feed aggregates

```
---- deviation of nutrition content from expected feeding practices (typical energy/protein density for aggregates)

   - 1E2 * SUM( (RUNR,MAACT,A,REQMSE,FeedAggr) $ p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MEAN"),
               SQR( (p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MEAN")-v_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr))
                  / (p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MEAN")))
     weighting:
               * [(v_actLevl(RUNR,MAACT,A)+.01)*p_AnimReq(RUNR,MAACT,A,REQMSE)*p_animProdDays(RUNR,MAACT,A)]**.1)

---- nutrient content in feed: high penalty for the part of the estimates over or under the lower/upper limits

  - 1E3 * SUM( (RUNR,MAACT,A,REQMSE,FeedAggr) $ ( p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MAX")
                                               $ p_aprDistAnimReq(RUNR,MAACT,A,"ENNE","MEAN")),
              SQR( (-ncpcm(-v_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr),
                          -p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MAX"),
                          max(.01,.05*p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MAX")))
                    -p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MAX"))
                  / MAX(.1,ABS(p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MAX"))))
     weighting:
              * [(v_actLevl(RUNR,MAACT,A)+.01)*p_AnimReq(RUNR,MAACT,A,REQMSE)*p_animProdDays(RUNR,MAACT,A)]**.1)

   - 1E3 * SUM( (RUNR,MAACT,A,REQMSE,FeedAggr) $ ( p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MIN")
                                                $ p_aprDistAnimReq(RUNR,MAACT,A,"ENNE","MEAN")),
              SQR( ( ncpcm( v_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr),
                           p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MIN"),
                           max(.01,.05*p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MIN")))
                    -p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MIN"))
                 / MAX(.1,ABS(p_nutContFeed(RUNR,MAACT,A,REQMSE,FeedAggr,"MIN"))))
     weighting:
              * [(v_actLevl(RUNR,MAACT,A)+.01)*p_AnimReq(RUNR,MAACT,A,REQMSE)*p_animProdDays(RUNR,MAACT,A)]**.1)
```

This part of the objective functions tries to minimize the difference between the nutrient content of feed aggregates (v_nutContFeed) and the expected nutrient (p_nutContFeed(…"MEAN")) coming from literature or IFM-CAP. To avoid unreasonably large deviations from MEAN, lower and upper limits are introduced (MIN, MAX), where the penalty in the objective function increases significantly. The extra penalties rely on the GAMS built-in smooth approximation of the min operator (Chen-Mangasarian smoothing function ncpcm). The values for mean and upper and lower limits are presented in the table below.

Table 11: Expected nutrient content in total feed per animal category

	Energy			Crude protein		
	MEAN	MIN	MAX	MEAN	MIN	MAX
DCOL	6.7	6.4	7	0.155	0.14	0.17
DCOH	6.8	6.6	7.2	0.155	0.14	0.17
BULL	6.7	6.2	7	0.155	0.14	0.17
BULH	6.8	6.4	7.2	0.155	0.14	0.17
HEIL	6.3	5.8	7	0.155	0.14	0.17
HEIH	6.8	6.2	7.2	0.155	0.14	0.17
SCOW	6.4	6	7	0.155	0.14	0.17
HEIR	6.4	6	7	0.155	0.14	0.17
CAMF	6.6	6.6	7.2	0.155	0.14	0.17
CAFF	6.6	6.6	7.2	0.155	0.14	0.17
CAMR	6.6	6.6	7.2	0.155	0.14	0.17
CAFR	6.6	6.6	7.2	0.155	0.14	0.17
PIGF	8	7.8	8.2	0.155	0.14	0.17
SOWS	8	7.8	8.2	0.155	0.14	0.17
SHGM	6.3	5.8	7	0.155	0.14	0.17
SHGF	6.3	5.8	7	0.155	0.14	0.17
HENS	8	7.8	8.2	0.18	0.14	0.2
POUF	8	7.8	8.2	0.18	0.14	0.2

Shares of feed aggregates in total feed intake in DRMA

```
---- shares of feed aggregates: high penalty for the part of the estimates over or under the lower/upper limits

  - 1E4 * SUM( (RUNR,MAACT,A,FeedAggr) $ ( p_maxFeedShare(RUNR,MAACT,A,FeedAggr)
                                           $ p_aprDistAnimReq(RUNR,MAACT,A,"ENNE","MEAN")),
             SQR( (-ncpcm(-v_FeedAggrShare(RUNR,MAACT,A,FeedAggr),
                          -p_maxFeedShare(RUNR,MAACT,A,FeedAggr),
                          max(.02,.05*p_maxFeedShare(RUNR,MAACT,A,FeedAggr)))
                  -p_maxFeedShare(RUNR,MAACT,A,FeedAggr))
                / MAX(.1,ABS(p_maxFeedShare(RUNR,MAACT,A,FeedAggr))))
     weighting:
                 * [  max(SUM(R_RAGG(RUNR,MSACT),p_feedInpCoeffDRMA(MSACT,maact,A,FeedAggr,"Adjusted")),.1)
                  * (v_actLevl(RUNR,MAACT,A)+.01)*p_animProdDays(RUNR,MAACT,A)]**.1)

  - 1E4 * SUM( (RUNR,MAACT,A,FeedAggr) $ ( p_minFeedShare(RUNR,MAACT,A,FeedAggr)
                                           $ p_aprDistAnimReq(RUNR,MAACT,A,"ENNE","MEAN")),
             SQR( ( ncpcm( v_FeedAggrShare(RUNR,MAACT,A,FeedAggr),
                           p_minFeedShare(RUNR,MAACT,A,FeedAggr),
                           max(.02,.05*p_minFeedShare(RUNR,MAACT,A,FeedAggr)))
                  -p_minFeedShare(RUNR,MAACT,A,FeedAggr))
                / MAX(.1,ABS(p_minFeedShare(RUNR,MAACT,A,FeedAggr))))
     weighting:
                 * [  max(SUM(R_RAGG(RUNR,MSACT),p_feedInpCoeffDRMA(MSACT,maact,A,FeedAggr,"Adjusted")),.1)
                  * (v_actLevl(RUNR,MAACT,A)+.01) *   p_animProdDays(RUNR,MAACT,A)]**.1)
```

The shares of roughage and concentrate feed are only controlled by upper (p_maxFeedShare) and lower (p_minFeedShare) limits. The literature suggests that ruminants can digest at most 40% of concentrate feed (or at least 60% roughage), and perhaps 45% for activity DCOH. The upper and lower limits are partially taken from IFM-CAP, literature and expert knowledge of Markus Kempen (Assumed values in table 12).

Table 12: Maximum and minimum shares of feed aggregates

	Maximum shares		Minimum shares	
	FeedRough	FeedCons	FeedRough	FeedCons
DCOL	0.85	0.4	0.75	0.1
DCOH	0.7	0.45	0.6	0.1
BULL	0.8	0.4	0.65	0.1
BULH	0.8	0.4	0.65	0.1
HEIL	0.9	0.3	0.65	0.1
HEIH	0.9	0.3	0.7	0.1
SCOW	0.95	0.3	0.7	0.05
HEIR	0.9	0.3	0.7	0.05
CAMF		0.3		0.15
CAFF		0.3		0.15
CAMR		0.3		0.1
CAFR		0.3		0.1
PIGF		1		0.95
SOWS		1		0.9
SHGM		0.3		0.05
SHGF		0.3		0.05
HENS				0.99
POUF				0.99

For „other feed“ there are no lower bounds but rather low upper bounds: 10% for adult cattle, 5% for calves and sheep, 1% for pigs and 1E-6 (so near zero) for poultry.

Feed input coefficients for single feed bulks

Apart from plausibility of the results a second objective of the revision has been reproducability. The previous specification essentially gave random results within the feasible set because no prior expectations had been specified. This has been revised with penalties for deviations of feed input coefficients from their assumed MEAN (specification to be explained below). However, just like is the case for the nutrient content of feed aggregates or their shares in the total, this prior information has to be considered quite imprecise which is reflected in rather low factors (1E2) attached to these terms. The penalties are increased if the solver tries to approach or exceed "soft" lower or upper limits. As the lower limits also turned out useful to prevent the solver from ending up in infeasible corners a higher factor has been attached to them (1E5).

It should also be reported that in many cases of infeasible solutions encountered in the extensive testing of this and previous specifications the last iteration result reported from the solver had often all feed input coefficients for some animal type zero or near zero. To avoid these cases the solution attempt starts with hard lower bounds:

```
---- penalty for deviation from expected feed input

  - 1E2 * SUM( (RUNR,MAACT,A,FEED) $ (p_feedInpCoeff(RUNR,MAACT,A,FEED,"STDE") $ p_maxFeedShare(RUNR,MAACT,A,Feed)),
               SQR( (v_feedInpCoeff(RUNR,MAACT,A,FEED) - p_feedInpCoeff(RUNR,MAACT,A,FEED,"Mean"))
                  /  p_feedInpCoeff(RUNR,MAACT,A,FEED,"STDE"))
     weighting:
                 * [(v_actLevl(RUNR,MAACT,A)+.01) *  p_animProdDays(RUNR,MAACT,A)]**.1)

---- high penalty for the part of the estimates over or under the lower/upper limits

  - 1E4 * SUM( (RUNR,MAACT,A,Feed) $ (p_FeedInpCoeff(RUNR,MAACT,A,Feed,"UppLim")),
                SQR((-ncpcm(-v_FeedInpCoeff(RUNR,MAACT,A,Feed),
                     -p_FeedInpCoeff(RUNR,MAACT,A,Feed,"UppLim"),
                      max(2,0.02*p_FeedInpCoeff(RUNR,MAACT,A,Feed,"UppLim")))
                     -p_FeedInpCoeff(RUNR,MAACT,A,Feed,"UppLim"))
                  / MAX(.1,ABS(p_FeedInpCoeff(RUNR,MAACT,A,Feed,"UppLim"))))
     weighting:
                  * [(v_actLevl(RUNR,MAACT,A)+.01) * p_animProdDays(RUNR,MAACT,A)]**.1)

   - 1E5  * SUM( (RUNR,MAACT,A,Feed) $ ( p_FeedInpCoeff(RUNR,MAACT,A,Feed,"LowLim")
                                        $ p_aprDistAnimReq(RUNR,MAACT,A,"ENNE","MEAN")),
                  SQR( ( ncpcm(v_FeedInpCoeff(RUNR,MAACT,A,Feed),
                         p_FeedInpCoeff(RUNR,MAACT,A,Feed,"LowLim"),
                         max(2,0.05*p_FeedInpCoeff(RUNR,MAACT,A,Feed,"LowLim")))
                        -p_FeedInpCoeff(RUNR,MAACT,A,Feed,"LowLim"))
                      / MAX(.1,ABS(p_FeedInpCoeff(RUNR,MAACT,A,Feed,"LowLim"))))
     weighting:
                    * [(v_actLevl(RUNR,MAACT,A)+.01) * p_animProdDays(RUNR,MAACT,A)]**.1)

     Lower bounds close to zero are maintained as this turned out to help the solver
v_feedInpCoeff.lo(RUNR,maact,A,FEED) $ p_feedInpCoeff(RUNR,MAACT,A,FEED,"lowLim")
    = p_feedInpCoeff(RUNR,MAACT,A,FEED,"lowLim")
    * ( 1.E-1 $ ( (not sameas(feed,"fstr")) $ SAMEAS(R_LEVL,"TOP"))
      + 1.E-3 $ ( (not sameas(feed,"fstr")) $ SAMEAS(R_LEVL,"MS")));
```

In case of infeasibilities after x trials these are removed:

```
    After x trials we further relax the lower bounds here, but independent from marginals and for all coefficient:
v_feedInpCoeff.lo(RUNR,maact,A,FEED)  $ (p_feedInpCoeff(RUNR,MAACT,A,FEED,"lowLim") $ (p_nSolved(RAGG) gt 5))
  = v_feedInpCoeff.lo(RUNR,maact,A,FEED)  / 10;
v_feedInpCoeff.lo(RUNR,maact,A,FEED)  $ (p_feedInpCoeff(RUNR,MAACT,A,FEED,"lowLim") $ (p_nSolved(RAGG) gt 10))
  = 0;
```

This procedure led to an acceptable or at least considerably improved stability of the feed calibration in tasks “build regional database” as well as “baseline calibration supply models”.

Priors for feed input coefficients

The priors for feed input coefficients are specified in a new include file capri/gams/feed/fedtrm_prior.gms:

```
*  calculate share of single feed in feed aggregate
p_FeedShareInAggr(RallInMS,MAACT,"T",Feed,FeedAggr,"RegMean") $ ( Feed_FeedAggr(FEED,FeedAggr)
                                                                 $ p_feedQuantDRMA(RallInMS,FeedAggr,"FEDM"))
  = p_feedQuantDRMA(RallInMS,FEED,"FEDM") / p_feedQuantDRMA(RallInMS,FeedAggr,"FEDM");
---------------------------------------------------------------------------------- 4 line(s) not displayed --------------------
*  prior expectation for annual feed intake in DM per animal type and feed stuff:
*  = adjusted DRMA from feed aggregate  times regional share of feed stuff in aggregate
*  ATTENTION: does not account for animal specific diets
*             Example 1: pigs should have about 20% FPRO and 80% FCER in their concentrate, independent of national mean mix
*             Example 2: calves for raising should receive a higher share of liquid raw milk FCOM than calves for fattening
p_feedInpCoeffDRMA(RallInMS,maact,"T",Feed,"PRIOR") $ p_maxFeedShare(RallInMS,MAACT,"T",FEED)
  = sum((MAP_RR(MS,RallInMS),Feed_FeedAggr(FEED,FeedAggr)), p_feedInpCoeffDRMA(MS,maact,"T",FeedAggr,"Adjusted")
                                                           * p_FeedShareInAggr(RallInMS,MAACT,"T",Feed,FeedAggr,"RegMean"));
```

The shares of feed aggregates in the diets of animal types may build upon recommendations from the literature (see the previous section). They are adjusted to be in line with the statistical ex post data or the baseline projections, giving the “adjusted” aggregate feed input coefficients shown in the code snippet above.

However, feed recommendations do snot exist for *single* feedstuffs because these are easily substitutable. Stability of the feed calibration requires however some priors. A simple default assumption made has been therefore: the composition of feed aggregates in terms of their components is the same for all animals (corresponding to the regional average). This is evidently a simplification such that the penalties for deviations from these priors have been set rather low to achieve both the desired stabilization effect while not competing too strongly with other components of the objective.

Nutrient contens and requirements

For the nutrient contents and requirement functions comparisons with IFM-CAP showed a good consistency such that the pre_star2 specifications were retained.

Calibration of PMP terms

The calibration of pmp terms for feeding coefficients is unchanged. But the constraints of minimum and maximum shares of feeding stuffs and some contents (fibre, lysin, etc) have been removed. The pmp terms have therefore a considerably increased role in simulations: Whereas the feed mix was so far steered by technical constraints, at least to a significant extent, all of these are gone except the equality constraints on feed energy and protein. The feed mix in simulation is therefore critically determined by the feed related pmp terms. In case of undesirable simulation behaviour it might be considered to include at least bounds for the total feed intake in terms of dry matter where feed recommendations apparently provide some bounds for plausible values.

3.4.4 Input allocation for fertilisers and nutrient balances

In the following section, the existing environmental indicators in CAPRI, planned and already achieved improvements, and possible further extensions are briefly discussed. It should be noted that CAPRI is basically a regionalised agricultural sector model, thus concentrating on the modelling of aggregated reactions of agricultural producers and consumers to changes in long term shifters as technical progress, income changes and CAP programs. Most indicators are rather robust pressure indicators and can be calculated easily based on fixed parameters approaches from the endogenous variables of the regional aggregate supply models. Accordingly, economic (dis)-incentives can be linked to the pressure indicators or further passive indicators can be introduced or the current ones changed easily.

Currently, CAPRI estimates the following environmental indicators:

1. Greenhouse gas emissions from enteric fermentation (CH4), manure management (CH4, N2O), manure and mineral fertilizer application to soils (N2O, CO2), grazing animals (N2O), crop residdues (N2O), cultivation of histosols (N2O, CO2), indirect emissions from the volatilization of ammonia (N2O), indirect emissions from leaching and runoff (N2O), land use change emissions from carbon stock changes in above and below ground biomass (CO2), soils carbon stock changes (CO2,N2O), the burning of biomass (CH4,N2O). For details see (Pérez 2005) and Leip et al. (2010).
2. Ammonia emissions from manure management, manure and mineral fertilizer application (Leip et al (2010).
3. Nitrate Leaching and Runoff (Leip et.al. (2010)
4. Soil erosion

Moreover, CAPRI provides the complete nutrient cycle for nitrogen and carbon, while for phosphate and potassium only the separate nutrient balances for crops and feed are considered. An important limitation of phosphate and potassium balancing is that output at tail is unrelated to feed intake because fixed coefficeints are used.

Nutrient balances for NPK and Nitrates Leaching Nutrient balances in CAPRI are built around the following elements:

- Export of nutrient by harvested material per crop –depending on regional crop patterns and yields, and livestock products, and crop residues.
- Output of manure at tail –depending on animal type, regional animal population and animal yields, as final weights or milk yields (see section on Output at tail).
- Manure imports and exports (to the region)
- Input of mineral fertiliser –as given from national statistics at sectoral level.
- Input of crop residues, biological fixation, atmospheric deposition
- Emissions (NH3, NOx, N2, N2O, CO2, CH4, NO3, C from soil erosion) only for nitrogen and carbon, and removals (carbon sequestration) only for carbon

The numbers in the following table are based on older methodology and coefficients but nonetheless provide a useful illustration of the accounting. Details on the emissions are provided in the respective sections on ammonia and greenhouse gases. Details on the inputs in the sections on NPK output at tail and NPK input distribution.

Table 13: Nitrogen balance (EU 15, year 2001)

INPUT			OUTPUT		
Import of nitrogen by anorganic fertiliser	a	68.2	Export of nitrogen with harvested material	f	80.95
Import of nitrogen by organic fertiliser (in manure)	b	77.31	Nitrogen in ammonia, NOx, N2O and runoff losses from manure fallen on grazings	g	2.08
Nitrogen from biological fixation*	c	2.89	Nitrogen in ammonia, NOx and N2O losses from manure in stable	h	7.13
Nitrogen from atmospheric deposition	d	14.36	Nitrogen in ammonia, NOx, N2O,N2 and runoff losses from manure storage	i	2.53
			Nitrogen in ammonia, NOx, N2O and runoff losses from manure application on the field	j	8.34
			Nitrogen in ammonia, NOx, N2O and runoff losses from organic fertiliser	k=g+h+i+j	20.08
			Nitrogen in ammonia, NOx, N2O and runoff losses from mineral fertiliser	l	2.89
TOTAL INPUT	**e=a+b+c+d**	**162.768**	**TOTAL OUTPUT**	**n=f+k+l+m**	**103.92**
			Nutrient losses at soil level (SURPLUS)	**m=e-f-k-l**	**58.85**

The difference between nutrient inputs and outputs corresponds to the soil surplus. For nitrates the leaching is calculated as a fraction of the soil surplus, which is based on estimates from the MITERRA project, and depends on the soil type, the land use (grassland or cropland), the precipitation surplus, the average temperature and the carbon content in soils. For details see Velthof et al. 2007 "Development and application of the integrated nitrogen model MITERRA-EUROPE". Alternatively, a version was developed which uses the leaching fractions from the official Greenhouse gas inventories of the member states. For phosphate, currently emissions (mainly superficial runoff) are not quantified.

NPK output at tail

The output of P and K at tail is estimated based on typical nutrient contents of manure:

Table 14: Nutrient content in manure in kg pure nutrient/m³

	P	**K**
Cattle	2.0	5.5
Swine	3.3	3.3
Poultry	6.3	5.1

These data are converted into typical pure nutrient emission at tail per day and kg live weight in order to apply them for the different type of animals. For cattle, it is assumed that one live stock unit (=500 kg) produces 18 m³ manure per year, so that the numbers in the table above are multiplied with 18 m³ and divided by (500 kg *365 days).

For the different types of cattle activities, it is hence necessary to determine the average live weight and the length of the production process.

For calves fattening (CAMF, CAFF), the carcass weight is divided by 60 % in order to arrive at final weight and a start weight of 50 kg is assumed. Daily weight increases are between 0.8 kg/day and 1.2 kg/day and depend proportionally on average stocking densities of cattle in relation to the average EU stocking density for which a daily weight increase of 1 kg/day is assumed. Total emissions per animal hence increase with final weights but decrease per kg of meat produced for intensive production systems with high daily weight increases. The same relationship holds for all other animal categories discussed in the following paragraphs.

For calves raising (CAMR, CAFR), two periods are distinguished. From 50 to 150 kg, a daily increase of 0.8 kg/day is assumed. The remaining period captures the growth from 151 to 335 kg for male and 330 kg for female calves, where the daily increase is between 1 kg/day and 1.4 kg/day, again depending on stocking densities.

The bull fattening process captures the period from 335 kg live weight to final weight. Daily increases are between 0.8 kg/day up to 1.4 kg/day, depending on final weights and stocking densities. Carcass weights as reported in the data base are re-converted into live weight assuming a factor of 54% for low and 57% for higher final weights.

The heifers fattening process captures the period from 300 kg live weight to final weight, assuming a daily increase of 0.8 kg/day. Carcass weights, as reported in the data base, are re converted into live weight assuming a factor of 54 % for low and 57 % for higher final weights.

Suckler cows are assumed to be whole year long in production and weight 550 kg, whereas milk cows are assumed to have a weight of 600 kg and are again for 365 days in production. Additional data relate to the additional NPK output per kg milk produced by cows and are taken from the RAUMIS model:

Table 15: Additional emission of NPK per kg of milk produced

N	0.0084
P	0.004
K	0.0047

FIXME

The factors shown above for pigs are converted into a per day and live weight factor for sows by assuming a production of 5 m^3 of manure per sow (200 kg sow) and 15 piglets at 10 kg over a period of 42 days. Consequently, the manure output of sows varies in the model with the number of piglets produced.

For pig fattening processes, it is assumed that 1.9 m^3 are produced per 'standard' pig with a final carcass weight of 90 kg at 78 % meat content, a starting weight of the fattening period of 20 kg (weight of the piglet), a production period of 143 days and 2.3 rounds per year. The actual factors used depend on tables relating the final weight to typical daily weight increases.

For poultry, it is assumed that 8 m^3 of manure are produced by 100 laying hens, which are assumed to weigh 1.9 kg and stay for 365 days in production. For poultry fattening processes, a fattening period of 49 days to reach 1.9 kg is assumed.

For sheep and goat used for milk production or as mother animals, the cattle factors are applied by assuming a live weight of 57.5 kg and 365 days in production. For fattening processes, a daily increase of 200 kg and a meat content of 60 % of the carcass weight are assumed.

The nitrogen emission factors from animal activities are coupled to crude protein intake (IPCC 2006), and hence the requirement functions for animal activities according to a *farm gate approach*. According to the literature (Udersander et al. 1993), there is a relation of 1 to 6 between crude protein and N in feeding. By combining this information with N retention rates per animal activity (IPCC 2000, Table 4.15), manure production rates can be estimated (N intake minus N retention). A specific advantage of that approach is the fact that gross nutrient surplus is not longer depending on assumption on fodder yields and manure emissions factors. Changing the fodder yields in the combined farm-gate and soil-balance approach in CAPRI will change both nutrient retention in crops and nutrient deliveries from manure by the same values, leaving the balance unchanged.

Table 16: Crude protein intake, manure production and nitrogen retention per head (EU 15, year 2001)

	Crude protein	Nitrogen in manure	Nitrogen retention
BULH	1.7	83.8	0.07
BULL	1.4	31.7	0.07
CAFF	0.8	21.5	0.07
CAFR	0.9	38.4	0.07
CAMF	0.8	20.2	0.07
CAMR	0.9	38.6	0.07
DCOH	4.3	210.1	0.20
DCOL	2.7	129.4	0.20
HEIH	1.5	64.4	0.07
HEIL	1.2	20.6	0.07
HEIR	1.7	95.9	0.07
HENS (1000 units)	21.2	900.9	0.30
PIGF	0.4	7.0	0.30
POUF (1000 units)	7.6	52.9	0.30
SHGM	0.2	13.7	0.10
SHGF	0.1	2.0	0.10
SOWS	0.9	36.4	0.30
SCOW	1.5	87.2	0.07

Calibration of the input allocation of organic and inorganic NPK

The input allocation of organic and inorganic fertilizer determines how much NPK organic and inorganic fertiliser is applied per ha of a crop, simultaneously estimating the NPK availability in manure as well as parameters describing the degree of overfertilisation. Firstly, nutrient export by the harvested material is determined, based on the following factors:

Table 17: Exports of nutrients in kg per ton of yield or constant Euro revenues

	N	P	K
Soft wheat	20	8	6
Durum wheat	23	8	7
Rye	15	8	6
Barley	15	8	6
Oats	15.5	8	6
Grain maize	14	8	5
Other cereals	18	8	6
Paddy rice	22	7	24
Straw	6	3	18
Potatoes	3.5	1.4	6
Sugar beet	1.8	1.0	2.5
Fodder root crops	1.5	0.09	5.0
Pulses	4.1	1.2	1.4
Rape seed	33	18	10
Sunflower seed	28	16	24
Soya	58	16	24
Other oil seeds	30	16	16
Textile crops	3	8	15
Gras	5	1.5	3.5
Fodder maize	3.2	2.0	4.4
Other fodder from arable land	5.5	1.75	3.75
Tomatoes	2.0	0.7	0.6
Other vegetables	2.0	0.7	0.6
Apples, pear and peaches	1.1	0.3	1.6
Citrus fruit	2.0	0.4	1.6
Other fruits	2.0	0.4	1.7
Nurseries, flowers, other crops, other industrial crops	65	22	20
Olive oil	4.5	1.0	0.5
Table olives	22.5	5.0	2.5
Table grapes	1.9	1.0	3.1
Table wine, other wine	1.9/0.65	1.0/0.65	3.1/0.65
Tobacco	30.0	4.0	45.0

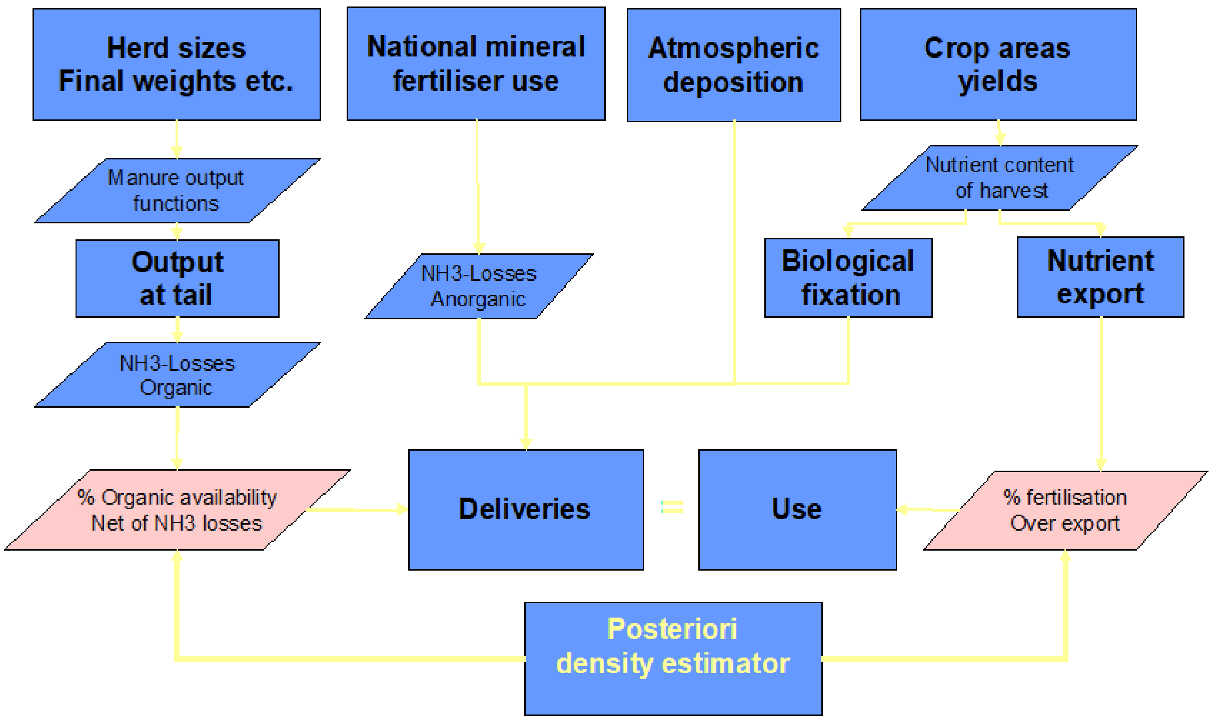
The factors above are applied to the expected yields for the different crops constructed with the Hodrick Prescott filter explained above. Multiplied with crop areas, they provide an estimate of total nutrient export at national and regional level (right hand side of the figure below). The maximum exports per ha allowed are 200 kg of N, 160 kg of P and 140 kg of K per ha.

Ex post, the amount of nutrients found as input in the national nutrient balance is hence 'known' as the sum of the estimated nutrient content in manure plus the amount of inorganic fertiliser applied, which is based on data of the European Fertiliser Manufacturer's Association as published by FAOSTAT. In order to reduce the effect of yearly changes in fertilizer stocks, three year averages are defined for the NPK quantities demanded by agriculture.

For the nitrogen balance, losses of NH3, N2O, NOx, N2 are handled as in MITERRA-Europe. The remaining loss to the soil, after acknowledging surface run-off, is disaggregated with leaching fractions into leaching or denitrification in soil. Atmospheric sources of N are taken into account as well (for details see section on nutrient balances).

Figure below offers a graphical representation of these relationships.

Figure 6. Ex-post calibration of NPK balances and the ammonia module



Source: CAPRI modelling system

The following equations comprise together the cross-entropy estimator for the NPK (Fnut=N, P or K) balancing problem. Firstly, the purchases (NETTRD) of anorganic fertiliser for the regions must add up to the given inorganic fertiliser purchases at Member State level:

$$\overline{Nettrd}_{MS}^{Fnut} = \sum_{r} Nettrd_{r}^{Fnut} \tag{3.38}$$

The crop need –minus biological fixation for pulses– multiplied with a factor describing fertilisation beyond exports must be covered by:

1. inorganic fertiliser, corrected by ammonia losses during application in case of N,
2. atmospheric deposition, taking into account a crop specific loss factor in form of ammonia, and

3. nutrient content in manure, corrected by ammonia losses in case of N, and a specific availability factor.

FIXME

$$
\begin{aligned}
&\sum_{cact} Levl_{r,cact} Fnut_{r,cact} (1 - NFact^{biofix}_{Fnut,cact}) \\
&NutFac_{r,fnut} (1 + NutFacG_{r,fnut} \wedge cact \in ofar, grae, grai) \\
&= NETTRD^{Fnut}_{r} (1 - NH3Loss^{Anorg}_{Fnut,r}) \\
&+ NBal^{AtmDep}_{r} NFact^{AtmDep}_{Cact} \\
&\sum_{aact} Levl_{r,aact} Fnut_{r,aact} (1 - NH3Loss^{Manure}_{Fnut,r})(1 - NavFac_{r,fnut})
\end{aligned}
\tag{3.39}
$$

The factor for biological fixation ($NFact^{biofix}$) is defined relative to nutrient export, assuming deliveries of 75 % for pulses ($PULS$), 10 % for other fodder from arable land (*OFAR*) and 5 % for grassland (*GRAE, GRAI*).

The factor describing 'luxury' consumption of fertiliser (*NutFac*) and the availability factors for nutrient in manure (*NavFac*) are estimated based on the HPD Estimator:

$$
\begin{aligned}
min\ HDP - &\sum_{r,fnut} \left(\frac{NutFac_{r,fnut} - \mu^{NutFac}_{r,fnut}}{\sigma^{NutFac}_{r,fnut}} \right)^2 \\
- &\sum_{r,fnut} \left(\frac{NavFac_{r,fnut} - \mu^{NavFac}_{r,fnut}}{\sigma^{NavFac}_{r,fnut}} \right)^2 \\
- &\sum_{r,fnut} \left(\frac{NutFacG_{r,fnut} - \mu^{NutFacG}_{r,fnut}}{\sigma^{NutFac}_{r,fnut}} \right)^2 \\
- &\sum_{r,ngrp} \left(\frac{Nitmr, ngrp - \mu^{Nitm}_{r,ngrp}}{\sigma^{NavFac}_{r,fnut}} \right)^2 \frac{\overline{LEVL}_{r,UAAR}}{\overline{LEVL}_{r,ngrp}}
\end{aligned}
\tag{3.40}
$$

The expected means γ for the availability for P and K in manure (*Navfac*) are centred around 50 %, for N at 50 %*40 %+25 %*86%, since 50 % are assumed to be released immediately, of which 60 % are lost as ammonia and 25 % are released slowly, with a crop availability of 86 %. These expected means at national level are multiplied with the regional output of the nutrient per hectare divided by the national output of nutrient per hectare so that the a priori expectation are higher losses with higher stocking densities. The lower limits are almost at zero and the upper limits consequently at the unity. The standard deviation σ is calculated assuming a probability of 1% for a zero availability and 1% for an availability of 100%.

The expected mean γ for the factor describing over fertilisation practices (*Nutfac*) is centred around 120 %, with a 1% probability for 160 % and a 1 % probability for 80 % (support points) with define the standard deviation σ. Upper and lower limits are at 500% and 5%, respectively. A second factor

(*Nutfacg*) is only applied for grassland and other fodder from arable land and centred around zero, with expected mean of +10% and a 10% with probabilities of 1%. Bounds for the factor *Nutfacg* are at 0.5 and 2.5.

The last term relates to the distribution of organic N to the different group of crops. The distribution is needed for simulation runs with the biophysical model DNDC (Joint Research Center, Ispra, Italy) linked to CAPRI results in the context of the CAPRI-Dynaspat project.

It is important to note that the CAPRI approach leads to nutrient output coefficient at tail taking into account regional specifics of the production systems as final weight and even daily weight increase as well as stocking densities. Further on, an important difference compared to many detailed farm models is the fact that the nutrient input coefficients of the crops are at national level consistent with observed mineral fertiliser use.

The nutrient balances are constraints in the regional optimisation models, where all the manure must be spread, but mineral fertiliser can be bought at fixed prices in unlimited quantities. Losses can exceed the magnitude of the base year but are not allowed to fall below the base year value. The latter assumption could be replaced by a positive correlation between costs and nutrient availability of the manure spread. There is hence an endogenous cross effect between crops and animals via the nutrient balances.

The factors above together with the regional distribution of the national given inorganic fertiliser use are estimated over a time series. Trend lines are regressed though the resulting time series of manure availability factors of NPK and crop nutrient factors for NPK, and the resulting yearly rates of change are used in simulation to capture technical progress in fertiliser application. The following table shows a summary by highlighting which elements of the NPK are endogenous and exogenous during the allocation mechanism and during model simulations:

Table 18: Elements entering the of NPK balance ex-post and ex-ante

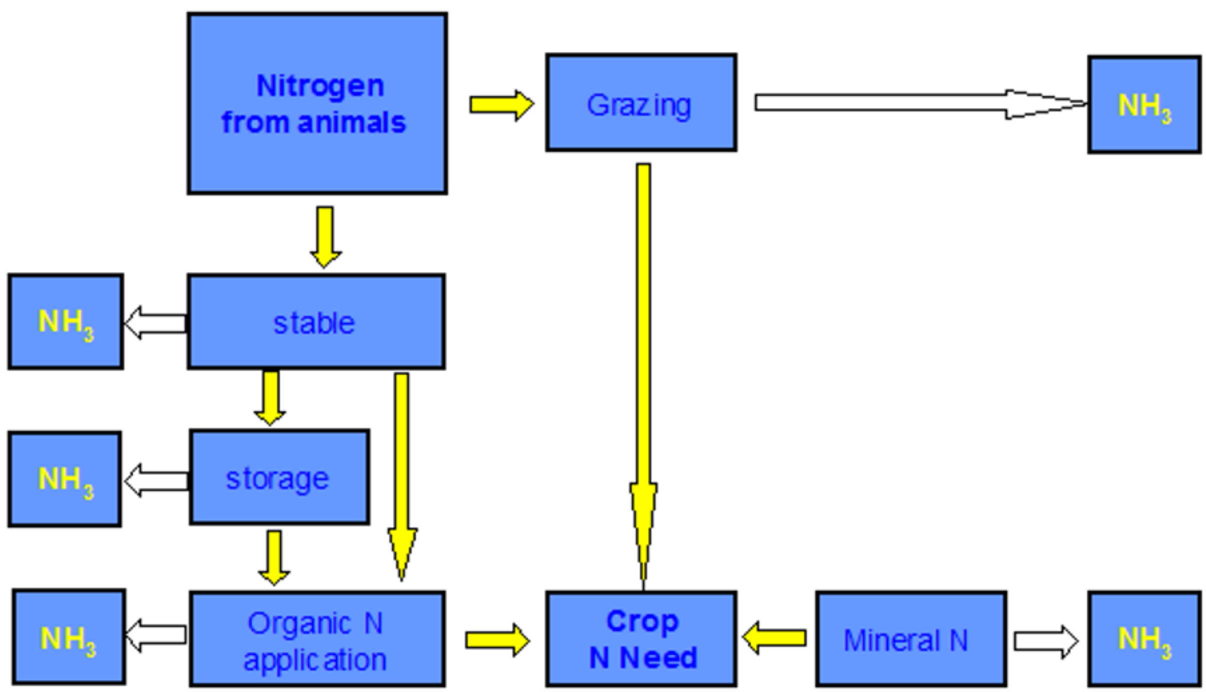
Ex-post	**Ex-ante**
Given:	**Model result:**
-Herd sizes	-Herd sizes
=> Manure output	=> manure output
-Crop areas and yields	-Crop areas and yields
=> Export with harvest	=> Export with harvest
-National anorganic application	-National and Regional anorganic application
Estimated:	**Given:**
-Regional anorganic application	-Factor for Fertilization beyond export (trended)
-Factor for Fertilization beyond N export	-Manure availability (trended)
-Manure availability	

A good overview on how the Nitrogen balances are constructed and can be used for analysis can be found in: Leip A., Britz W., de Vries W. and Weiss F. (2011): Farm, land, and soil nitrogen budgets for agriculture in Europe calculated with CAPRI, Environmental Pollution 159(11), 3243-3253 and Leip, A., Weiss, F. and Britz, W. (2011): Agri-Environmental Nitrogen Indicators for EU27, in: Flichman G. (ed.), Bio-Economic Models applied to Agricultural Systems, p. 109-124, Springer, Netherlands.

Update note The overall N Balance calibration problem has been revised several times. For example, since 2007 it delivers estimates of the shares of different sources of N (mineral fertiliser, excretions, crop residues) distinguished by crop groups. As of Stable Release 2.1, the calibration problem is augmented by an explicit maximization of the probability density functions described in the section on fertilization in the supply model chapter of this documentation [25].

The ammonia module The ammonia (NH3) and nitrous oxide (NOx) output module takes the nitrogen output per animal from the existing CAPRI module and replaces the current fixed coefficient approach with uniform European factors per animal type by Member State specific ones, taking into account differences in application, storage and housing systems between the Member States. The general approach follows the work at IIASA and has been updated under the Ammonia project in 2006/07. The following diagram shows the NH3 sinks taken into account by coefficients.

Figure 7: Ammonia sinks in the Ammonia emission module



Source: CAPRI modelling system

In the figure above, white arrows represent ammonia losses and are based on uniform or Member State specific coefficients. A first Member State specific coefficient characterises for each animal type the share

[25] A rather self contained presentation with a focus on the fertiliser calibration methodology (rather than environmental indicators or data sources) is given in Deliverable 4a: "Revision of the fertilizer module in CAPRI" in the context of specific contract 154208.X39 "IMPROVEMENT OF THE STABLE RELEASE OF THE CAPRI MODEL: FERTILIZER AND FEED ALLOCATION ROUTINES" (Star2).

of time spent on grassland and spent in the stable. For dairy cows, for example, the factors are between 41 % spent in the stable in Ireland and 93 % in Switzerland. During grazing about 8% of the excreted N is assumed lost as ammonia.

The time spent in the stable is then split up in liquid and solid housing systems. To give an example, 100 % of the Dutch cows are assumed to use liquid manure systems, whereas in Finland 55 % of the cows are in solid systems. Ammonia losses in both systems are assumed to be identical per animal types but differ between animals. 10 % ammonia losses are assumed for sheep and goat, 12 % for cattle, 17 % for pigs and 20 % for poultry, if no abatement measures are taken.

The remaining nitrate is then either put into storage or directly applied to the ground. No storage is assumed for sheep and goats and in all remaining cases not-covered systems are assumed with loss factors of 4-20 % of the N brought initially into storage.

After storage, the remaining N is applied to the soil, either spread to the surface –losses at 8 40%% or using application techniques with lower (20-40% saving) or high (80% saving) emission reductions. According to IIASA data most farmers work still with the standard techniques.

The update of this calculation during the Ammonia project in 2006/07 has included new coefficients from IIASA through the project partner Alterra. Furthermore, it has been acknowledged that in addition to NH3 there are losses of N as N2O, NOx and N2. The loss factors depend on the application of abatement techniques the penetration of which may be varied in scenarios. Technically, the underlying calculations are embedded as GAMS code in an own module both called during updates of the data base and model runs. This module in turn includes GAMS code borrowed from the MITERRA-Europe model of our former partner.

Recently ammonia mitigation technologies have been implemented as endogenous farm practices (see section on greenhouse gases) and environmental constraints related to important environmental directives like the Nitrates Directive (ND), the National Emisssions Ceiling (NEC), and the Industrial Emissions Directice (IED) have been implemented directly to the supply model. For the ND we consider upper limits for the application of manure and total nitrogen, for the NEC the upper limits member states committed to until 2030, and for the IED minimum reqirements for the implementation of manure storage measures.

Carbon balance The carbon cycle model quantifies relevant carbon flows in the agricultural production process related to both livestock and crop production (see Figure 6). Carbon flows and CO2 emissions from land use changes (LUC) are not considered meaning that the quantified balance applies to cropland remaining cropland and pasture/meadow land remaining in use. Default IPCC coefficients are used to quantify the carbon effects of LUC.

In CAPRI, so far the following carbon flows are taken into account, starting with animal production and ending with crop production (Weiss and Leip, 2016):

- Feed intake in livestock production (C)
- Carbon retention in livestock and animal products (C)
- Methane emissions from enteric fermentation in livestock production (CH4)
- Animal respiration in livestock production (CO2)
- Carbon excretion by livestock (C)

- Manure imports and exports to the region (C)
- Methane emissions from manure management in livestock production (CH4)
- Carbon dioxide emissions from manure management in livestock production (CO2)
- Runoff from housing and storage in livestock production (C)
- Manure input to soils from grazing animals and manure application (C)
- Carbon input from crop residues (C)
- Carbon export by crop products (C)
- Carbon dioxide emissions from the cultivation of organic soils (CO2)
- Carbon dioxide emissions from liming (CO2)
- Runoff from soils (C)
- Methane emissions from rice production (CH4)
- Carbon sequestration in soils (C)
- Carbon losses from soil erosion (C)
- Carbon dioxide emissions from soil and root respiration (CO2)

Accordingly, CAPRI does not consider the following carbon flows:

- Volatile organic carbon (VOC) losses from manure management (C)
- Carbon losses from leaching (C)
- Carbon dioxide emissions from urea application (CO2)

The VOC losses (non-CH4) from manure management are small and can be neglected. Carbon losses from leaching can be a substantial part of carbon losses from agricultural soils (see e.g. Kindler et al. 2011). Although they are not yet specifically quantified in the CAPRI approach, they are not neglected but put together with soil respiration in one residual value in the CAPRI carbon balance. CO2 emissions from urea application account for about 1% of total GHG emissions in the agriculture sector, but are not yet included in the CAPRI carbon cycle model.

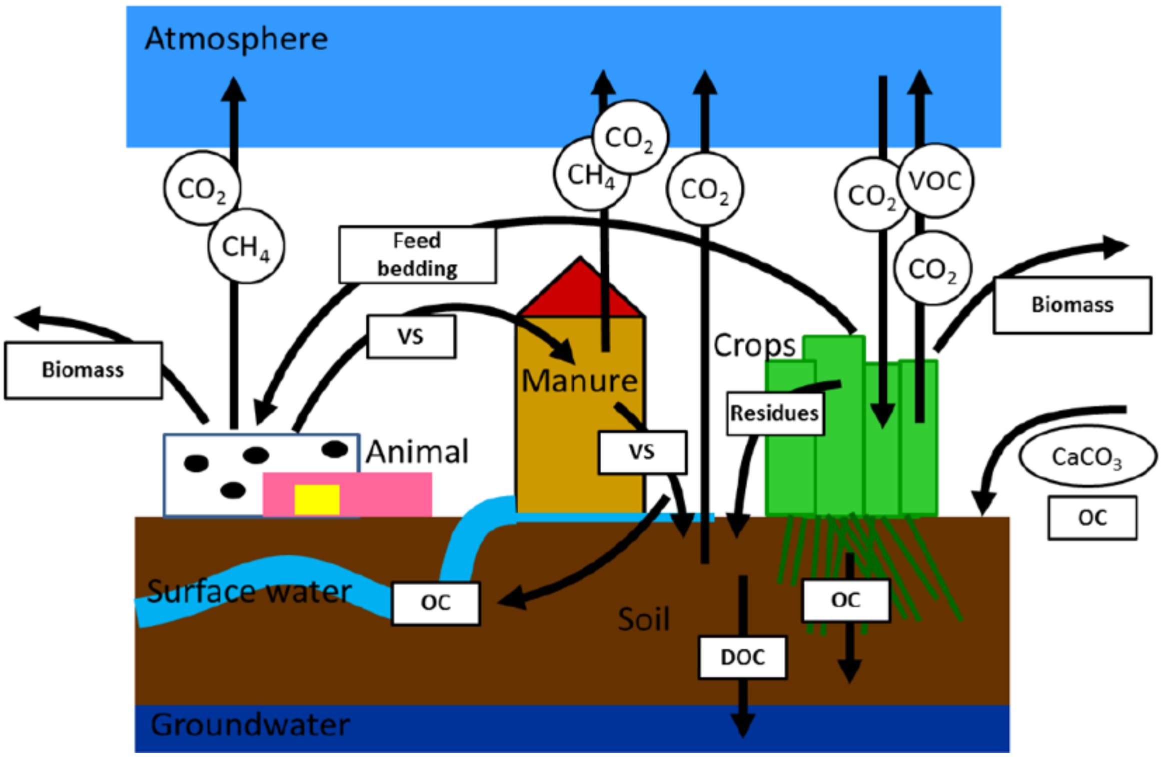
In the following, we briefly describe the general methodology for the quantification of the carbon flows that are taken into account in the CAPRI approach.

Subsequently, some details on the quantification of carbon flows (emissions and removals) are presented:

Feed intake in livestock production
Feed intake is determined endogenously in CAPRI based on nutrient and energy needs of livestock. The carbon content of feedstuff is derived from the combined information on carbon contents of amino acids and fatty acids, the shares of amino acids and fatty acids in crude protein and fats of different feedstuffs, and the respective shares of crude protein, fats and carbohydrates. For carbohydrates we assume a carbon content of 44%. Data was taken from Sauvant et al. (2004) and from NRC (2001).

Carbon retention in livestock and animal products
Similar to feed intake, we can quantify the carbon stored in living animals using the above mentioned

Figure 8: Carbon flows in the agricultural production process



Source: Weiss and Leip (2016)

data for animal products. At the end the values from meat are multiplied with the animal specific relation of live weight to carcass. For simplification, the fact that bones or skins etc. may have different carbon contents than meat is ignored.

Methane emissions from enteric fermentation
Methane emissions from enteric fermentation are calculated endogenously in CAPRI based on a Tier2 approach following the IPCC guidelines.

Animal respiration in livestock production
Intake of carbon is a source of energy for the animals. CAPRI calculates the gross energy intake on the basis of feed intake as described above. However, not all carbon is 'digestible' and hence can be transformed into biomass or respired. Digestibility of feed (for cattle activities) is calculated on the basis of the NRC (2001) methodology. Non-digestible energy (or carbon) is excreted in manure (see next point 5), while the 'net energy intake' refers to the equivalent to the energy stored in body tissue and products plus losses through respiration and methane.

According to Madsen et al. (2010) the heat production per litre of CO2 is 28 kJ for fat, 24 kJ for protein and 21 kJ for carbohydrates. Using a factor of 1.98 kg/m3 for CO2 (under normal pressure) or

505.82 l/kg we get 14.16 MJ/kg CO2 for fat, 12.14 MJ/kg CO2 for protein and 10.62 MJ/kg CO2 for carbohydrates, which translates into 0.071, 0.082 and 0.094 kg CO2 per MJ, respectively. These values are used to get the carbon directly from net energy intake (for each feedstuff), which is an endogenous variable in CAPRI depending on the feed intake. From this we subtract the carbon retained in living animals and in animal products and the methane emissions from enteric fermentation in order to compute the carbon respiration from livestock.

Carbon excretion by livestock
Carbon excretion is defined as the difference between the carbon intake via feed, the retention in livestock and the emissions as carbon dioxide (respiration) and methane (enteric fermentation):

$$Excretion = Feed\ intake–retention–emissions(CO_2, CH_4) \tag{3.41}$$

Carbon excretion can, therefore, be determined as the balance between the positions 1-4. As Carbon retention plus emissions by default gives the net energy intake (see 4), this is equivalent to

$$Excretion = C\ from\ gross\ energy\ intake–C\ in\ net\ energy\ intake \tag{3.42}$$

Manure imports and exports to the region
Manure available in a region may not just come from animal's excretion in the region but could also be imported from other regions, while, conversely, manure excreted may be exported to another region. CAPRI calculates the net manure trade within regions of the same EU member state, and this has to be accounted in the carbon balance as a separate position. For simplification, the model assigns the emissions of all manure excreted to the exporting region, while the carbon and nutrients are assigned to the importing region.

Methane emissions from manure management in livestock production
Once the carbon is excreted in form of manure (faeces or urine), it will either end up in a storage system or it is directly deposited on soils by grazing animals. Depending on temperature and the type of storage, part of the carbon is emitted as methane. These emissions are quantified in CAPRI following a Tier 2 approach, using shares of grazing and storage systems from the GAINS database (for more explanation see also Leip et al. 2010).

Carbon dioxide emissions from manure management in livestock production
During storage or grazing, carbon is not only emitted in form of methane, but part of the organic material is mineralized and carbon released as carbon dioxide. Following the FarmAC model[26], we assume a constant relation between carbon emitted as methane and total carbon emissions (methane plus carbon dioxide) of 63%. Therefore, the carbon loss through carbon dioxide emissions can be quantified as:

$$C(CO_2) = C(CH_4) * 0.37/0.63 \tag{3.43}$$

Runoff from housing and storage in livestock production
Part of the carbon excreted by animals is lost via runoff during the phase of housing and storage. We

[26]The FarmAC model simulates the flows of carbon and nitrogen on arable and livestock farms, enabling the quantification of GHG emissions, soil C sequestration and N losses to the environment (for more information see: `http://farmac.dk`).

assume the share to be equivalent to the share of nitrogen lost via runoff. In CAPRI we use the shares from the Miterra-Europe project, which are differentiated by NUTS 2 regions (for more information see Leip et al. 2010).

Manure input to soils from grazing animals and manure application
Carbon from manure excretion minus the emissions from manure management and runoff during housing and storage, corrected by the net import of manure to the region, is applied to soils or deposited by grazing animals. Other uses related to manure (e.g. trading, burning, etc.) are so far not considered in CAPRI. Moreover, we add here the carbon from straw from cereal production not fed to animals, assuming that all harvested straw (endogenous in CAPRI) not used as feedstuff is used for bedding in housing systems. The carbon content from straw is quantified in the same way as for feedstuff (see position 1). By contrast, other cop residues are treated under the position "carbon inputs from crop residues". Bedding materials coming from other sectors are currently ignored.

Carbon input from crop residues
The dry matter from crop residues is quantified endogenously in CAPRI following the IPCC 2006 guidelines (crop specific factors for above and below ground residues related to the crop yield). For the carbon content, a unique factor of 40% is applied as the information used in position 1 (feed input) is generally only available for the commercially used part of the plants, but not specified for crop residues.

Carbon export by crop products
Carbon exports by crop products are calculated as described under position 1, using the composition of fat and proteins by fatty and amino acids and the respective shares of these basic nutrients in the dry matter of crops.

Carbon fixation via photosynthesis of plants
Photosynthesis is the major source of carbon for a farm. Carbon is incorporated in plant biomass as sugar and derived molecules to store solar energy. Some of these molecules are 'exudated' by the roots into the soil. They provide an energy source for the soil microorganism – in exchange to nutrients. In the current version of CARPI, we assume that 100% of the photosynthetic carbon not stored in harvested plant material or crop residues, returns 'immediately' to the atmosphere as CO2 (root respiration) and has therefore no climate relevance. Accordingly, the effective fixation of carbon via photosynthesis is assumed to equal the exported carbon with crop products plus the carbon from crop residues. It is, therefore, not calculated as an explicit term.

Carbon dioxide emissions from the cultivation of organic soils
Carbon dioxide emissions from the cultivation of organic soils are calculated by using shares of organic soils derived from agricultural land use maps for the year 2000. For details see Leip et al. (2010).

Carbon inputs from liming
Agricultural lime is a soil additive made from pulverised limestone or chalk, and it is applied on soils mainly to ameliorate soil acidity. Total liming application on agricultural land as well as the related emission factor is taken from past UNFCCC notifications. A coefficient per ha is computed dividing the UNFCCC total amount by the UAA in the CAPRI database. For projection purposes this coefficient per ha, computed from the most recent data, is maintained in simulations. In the context of the carbon balance the CO2 emissions are converted into C and become carbon input into the system.

Carbon runoff from soils
Similar to position 9 (runoff from housing and storage in livestock production) we assume that the share

of carbon lost via runoff from soils is equivalent to the respective share of nitrogen lost. The respective shares are provided by the Miterra-Europe project (see Leip et al. 2010).

Methane emissions from rice production
Methane emissions from rice production are relevant only in a few European regions and they are quantified in CAPRI via a Tier 1 approach following IPCC 2006 guidelines.

Carbon sequestration in soils
Finally, we quantify the sequestered material after 20 years. The carbon change is based on simulations with the CENTURY agroecosystem model (Lugato et al. 2014) (aggregated from 1 km2 to NUTS2 level), and calculated from the difference in the manure and crop residue input to soils between the simulation year and the base year. This is done because carbon sequestration is only achieved from an increased carbon input, assuming that the carbon balance in the base year is already in equilibrium. The total cumulative carbon increase is divided by 20, in order to spread the effect over a standardised number of years (consistent with the 2006 IPCC guidelines).[27]

Carbon losses from soil erosion
Carbon losses from soil erosion are calculated on the basis of the RUSLE equation (see the setion on soil erosion). In order to get the carbon loss we have to multiply with the carbon content of the soil. As approximation we assume a 3% humus share for arable land and a 6% humus share for grassland. The carbon share in humus is around 2/3.

Carbon dioxide emissions from respiration of carbon inputs to soils
Carbon losses from soil are quantified as the residual between all carbon inputs to soils, the emissions and the carbon sequestered in the soils:

$$\begin{aligned}
&Carbon\ losses\ via\ soil\ and\ root\ respiration = \\
&Manure\ input\ from\ grazing\ and\ manure\ application \\
&+ input\ from\ crop\ residues \\
&- carbon\ losses\ (CH4)\ from\ rice\ production \\
&- carbon\ losses\ (CO2)\ from\ the\ cultivation\ of\ organic\ soils \\
&- carbon\ losses\ from\ runoff\ from\ soils \\
&- carbon\ losses\ from\ soil\ erosion \\
&- carbon\ sequestration\ in\ soils
\end{aligned} \tag{3.44}$$

Carbon losses from leaching should also be subtracted, but they are not specifically quantified in the CAPRI carbon cycle model so far. Therefore, the share of soil respiration is currently overestimated by the model.

Greenhouse Gases For the purpose of modelling GHG emissions from agriculture, a *multi strategy approach* is followed. It is important to take into account that agriculture is an important emitter of several climate relevant gases other than carbon dioxide. Therefore, three types of pollutants are modelled:

[27]The simulations with the CENTURY model were carried out by Emanuele Lugato from JRC.D3 in Ispra (for more details see Lugato et al. 2014).

methane (CH4) ,nitrous oxide (N2O), and carbon dioxide (CO2) emissions. The sources considered are: *CH4 emissions from animal production, manure management and rice cultivation, N2O from agricultural soils and manure management, and CO2 emissions from agricultural soils.* Moreover, carbon removals and emissions from land use change are quantified, and translated into CO2.

In CAPRI consistent GHG emission inventories for the European agricultural sector are constructed. As already mentioned, *land use* and *nitrogen flows* are estimated at a regional level. This is the main information needed to calculate the parameters included in the IPCC Good Practice Guidance (IPCC, 2006). The following table lists the emission sources modelled:

Table 19: Agricultural greenhouse gas emission sources included in the model

Greenhouse Gas	Emission source	Code
Methane	Enteric fermentation	CH4Ent
	Manure management	CH4Man
	Rice production	CH4Ric
	Land use change emissions from biomass burning	CH4bur
Nitrous Oxide	Manure management	N2OMan
	Manure excretion on grazings	N2OGra
	Application of synthetic fertiliser	N2OSyn
	Application of manure	N2OApp
	Crop residues	N2OCro
	Indirect emissions from ammonia losses	N2OAmm
	Indirect emissions from leaching and runoff	N2OLea
	Cultivation of histosols	N2Ohis
	Land use change emissions from the burning of biomass	N2Obur
Carbon dioxide	Cultivation of histosols	CO2his
	Applicaton of ureum	CO2urea
	Liming	CO2lim
	Land use change emissions from above and below ground biomass	CO2bio
	Land use change emissions from soil carbon changes	CO2soi

For a detailed analysis of these single emission sources refer to Pérez 2006: Greenhouse Gases: Inventories, Abatement Costs and Markets for Emission Permits in European Agriculture -A Modelling Approach and Leip et al 2010: Evaluation of livestock sector's contribution to the RU greenhouse gas emissions (GGELS).

The model code also comprises a life-cycle assessment for GHGs (first approach explained in Leip et al, 2010, but newer approach not yet documented in an official publication), and a module to estimate emission leakage in Non-European world regions (for details see e.g. Jansson et al.,2010: Estimation of Greenhouse Gas coefficients per commodity and world region to capture emission leakage in European Agriculture; Pérez Dominguez et al., 2012: Agricultural GHG emissions in the EU: An Exploratory Economic Assessment of Mitigation Policy Options., Van Doorslaer et al, 2015: An economic assessment of greenhouse gas mitigation options for EU agriculture). Moreover, in recent projects (Ecampa1-3) mitigation technologies and farm practices have been introduced to the supply model, which directly impact on the emissions. Currently, the following mitigation technologies can be activated:

- Anaerobic digestion
- Feed additives to reduce methane emissions from ruminants (lineseed, nitrate)
- Precision farming
- Variable Rate Technology
- Nitrification Inhibitors
- Better timing of fertilizer application
- Winter cover crops
- No Tillage
- Conservation Tillage
- Buffer strips
- Fallowing of histosols
- Measures to reduce methane emissions in rice production
- Increased legume share on temporary grassland
- Genetic measures to increase milk yields and feed efficiency
- Urea Substitution
- Manure application measures to reduce ammonia emissions (high and low efficiency)
- Manure storage measures to reduce ammonia emissions (high and low efficiency)
- Stable design measures to reduce ammonia emissions
- Low Nitrogen Feed
- Manure storage basins in concrete to reduce nitrate leaching
- Flexible limits for nitrogen application to soils
- Flexible limits for livestock density
- Vaccination against methanogenic bacteria

For details see Van Doorslaer et al. 2015, and Perez et.al 2016 (Most recent developments not yet published).

Soil erosion Soil erosion is calculated on the basis of the RUSLE equation. The equation has the following form:

$$A = R \cdot K \cdot L \cdot S \cdot C \cdot P \tag{3.45}$$

where
A = soil loss in ton per ha/acre per year

R = rainfall-runoff erosivity factor
K = soil erodibility factor
L = slope length factor
S = slope steepness factor
C = cover management factor
P = support practice factor

For more details on the factors used see Panagos et al. (2015).

3.4.5 Input allocation for labour

Labour (and other inputs) in CAPRI are estimated from a Farm Accounting Data Network (FADN) sample [28] and then these estimation results are combined with total labour requirements within a region (or aggregate national input demand reported in the EAA), using a Highest Posterior Density (HPD) estimation framework.

Labour Input Allocation Input coefficients (family labour and paid labour, both in hours, as well as wage regressions for paid labour) were estimated using standard econometrics from single farm records as found in FADN. While many of results from this process are plausible a number of CAPRI estimates of labour input are inaccurate and untrustworthy, not least when fitted values for labour using the econometric coefficients are compared with total regional labour inputs recoverable from FADN data survey weights. To remedy this, a reconciliation process is undertaken to correct figures for labour input by adjusting the labour input coefficients for both total labour and family labour, handled in file gams/inputs/labour_calc.gms.

The reconciliation process has two components. The first component is to fix on a set of plausible estimates for the labour input coefficients (based on the econometric results) while the second involves a final reconciliation, where further adjustments are made to bring the estimates into line with the FADN values for labour inputs. Implementing these two steps involves the following procedures.

Step one involves preparing the econometric estimates in order to remove unreliable entries. This process removes specific unsuitable estimates for particular regions and crop types. In addition, this process also involves adjusting certain agricultural activities labour input coefficients (such as the estimates for triticale) so as to bring them into line with similar activities (such as for soft wheat). Furthermore, a Bayesian probability density function is used where EU averages are used as priors, and a number of bounds are added, in order to generate realistic labour input coefficients.

While the procedure described above help to ensure plausible estimates, the labour input values generated will still not be such as to reconcile total fitted labour with total actual labour at a regional or national level (as estimated by FADN). Step 2 in this process is to implement a final reconciliation, where the labour input coefficients are adjusted in order to bring estimates of labour input closer to the total labour used in the region/country. However, this adjustment process has to be balanced with a recognition that many of the labour input coefficient estimates are relatively reliable and that we don't need or want

[28] More details on the FADN estimation were reported older versions of this section (originally drafted by Markus Kempen and Eoghan Garvey) the CAPRI documentation, accessible in the /doc folder of any stable release of the CAPRI system up to star 2.4 from `https://www.capri-model.org/dokuwiki/doku.php?id=capri:get-capri`.

to radically adjust all of them. Therefore the final reconciliation has to specify which input coefficients have to be adjusted most. The main way in which this is achieved is through the consideration of the coefficients' standard errors in a second Bayesian posterior density function.

As well as the reconciliation process, two other procedures have to be carried out. The first results from the fact that a number of activities don't have labour input coefficient estimates. In order to estimate them, the revenue shares for the relevant activities are used as a proxy for the amount of labour they require. Labour input for the different activities is then calculated based on these shares. The second procedure is due to the presence of infeasibilities in this model. In order to try and eliminate them, a number of courses of action can be followed from excluding outlying estimates to dropping regional estimates.

It should be noted that the reconciliation process has to be divided into these two steps because it is highly computationally burdensome. For the model to run properly (or even at all), it is necessary to divide it into two parts, with the one part obtaining plausible elements and the other implementing the final reconciliation.

Table 20: Total labour input coefficients from different econometric estimations and steps in reconciliation procedure (selected regions and crops)

Region	crop or aggregate	Econometric estimation			HPD solution including		
		regional	national- including yield	national - without yield	regional, national, crop aggregates	+ expert assumption	+ regional labour supply
Belgium (BL24)	Soft wheat	31.49	31.26	31.49	24.99	32.73	53.88
	Sugar beet	76.25	77.39	76.25	62.19	48.27	68.36
	Cereals	28.23	32.89	28.23	32.78	28.16	32.66
	Root crops	58.75	65.43	58.75	58.8	64.52	105.89
Germany (DEA1)	Soft wheat	36.78	35.32	36.78	36.98	38.62	34.46
	Sugar beet	82.01	58.99	82.01	55.06	39.61	43.58
	Cereals	40.13	32.63	40.13	39.94	41.65	35.12
	Root crops	28.83	14.23	28.83	38.32	41.26	0.01
France (FR24)	Soft wheat	14.65	23.3	23.68	14.71	16.5	13.22
	Sugar beet	-7.42	2.24	-1.68	11.08	19.72	18.5
	Cereals	10.48	35.9	22.7	15.61	15.43	12.7
	Root crops	11.68	29.78	19.42	17.05	24.64	18.43

The Table visualizes the adjustments regarding an implausible labour input coefficient for sugar beet in a French region. The econometric estimation come up with very low or negative values. The HPD solution combining crop specific estimates with corresponding averages of crop aggregates corrects this untrustworthy value to 11.08 h/ha. This value is in an acceptable range but it strikes that in opposite to many other regions the labour input for sugar beet is still less than for soft wheat. After adding equations in the reconciliation procedure that ensure that the relation of labour input coefficients among crops follows an similar “European” pattern the labour input is supposed to be 19.72 h/ha. There is up to now no theoretical or empirical evidence for this similar pattern regarding relation of input coefficients but the results seem to be more plausible when checked with expert knowledge. In the last column bounds on regional labour supply derived from FADN are added which “scales” the regional value. This final result is and is now part of the CAPRI model.

Projecting Labour Use For typical applications of CAPRI, regional projections of labour use are needed. Such projections have been prepared as well in the CAPSTRAT project, using a cohort analysis to separate 2 components of changes over time: (1) an autonomous component, which comprises structural changes due to demographic factors such as ageing, death, disability and early retirement, and (2) a non-autonomous component, which incorporates all other factors that influence changes in farm structure and has been analysed econometrically.

The results of this analysis are loaded in the context of CAPRI task “Generate trend projection” in file baseline/labour_ageline.gms, but only to serve as one type of bounds for labour use in the contrained trends for European regions. Other bounds are derived from engineering knowledge (or assumptions) on plausible labur use per activity which is based on the initial estimation of labour allocation by activity.

3.5 The global database components

3.5.1 Task: Prepare FAOSTAT database

This task prepares and partially combines FAO data originally contained in separate tables from the FAO-stat webpage to finally store them in gdx files for further use. This refers to: Commodity balances, production and landuse statistics (all stored in faodata.gdx), special balances for dairy products (fao_milk-data.gdx), population (fao_population.gdx), as well as the bilateral trade matirix (fao_trade_for_global.gdx).

The FAOSTAT task consists of two independent consolidation routines, the (A) Country data, and the (B) Trade flow related consolidation part. Part (A) imposes consistency rules on market balances, yields, activity levels, land use data, and population at the country level. Part (B) consolidates bilateral trade at the level of CAPRI trading blocks comprising quantities, values, unit values (UVAL) and the world price index (PRII).

The task requires input data stemming from an external preparation routine which is not a CAPRI module or sub-module. It is executed only on an intermittend basis depending on the availability of new raw data from FAOSTAT and the requirement for an update of the corresponding input data.

The resulting output from the external preparation routine are six gdx-files that have to be present in the /dat-folder of the CAPRI working directory: (1) commodityBalances, (2) population, (3) ProductionAndRessources, (4) fao_trade_matrix. Input data files (1) to (3) are required for the country related part (A), the trade matrix (4) is required for consolidation part (B).

Consolidation of country level data In this step (1) activity levels, yields and production quantities are checked for completeness and heuristic rules are applied for gap filling. The (2) information on production statistics on crops is mapped to the commodity balances for the primary product equivalent to produce consistent data on yield and area. The (3) land use data is consolidated such that nested land categories add up to their totals. As the milk sector in FAOSTAT is organized differently from the CAPRI concept the (4) products' mapping of the dairy sector is adjusted accordingly. (5) Gaps in population data for Serbia and Montenegro and for Belgium and Luxembourg are filled as well.

The head section of the sub-module comprises (a) initialization of FAOSTAT-related and mapping sets which are used in all futher consolidation sections, (b) loading union sets from the CommodityBalances and ProductionAndRessources data files, (c) introducing the land categories relevant for the land use consolidation, (d) introduction of multiplication factors for the mapping of units between FAOSTAT and CAPRI items, and (e) initialization of parameters. The (c) land categories relevant for the land use consolidation are as follows:

The first consolidation section is on "Production and Ressources". After loading the raw data at the beginning, the FAOSTAT units are mapped to CAPRI units via the "unit_map" set and corresponding multiplication factors as provided under (d) in the head section of the program to harmonise the units. After that the data is checked for completeness and various heuristic rules are applied to fill gaps in the data:

After aggregating data for China and some reporting on missing data the consolidated production data is written to the /fao folder in the restart-directory for usage in the following consolidation steps.

The next stept consolidates "Commodity Balances" and introduces the sets for the main balance components and demand positions as well as the mapping between the original FAOSTAT item codes and the commodity balance codes. This is another example that any data consolidation combining different data sets (even when coming from the same agency like FAO) needs to consider different coding systems used in those data sets:

In addition to the item code and unit matching and the removal of flags, negative observations are removed (except for stock changes) from the data. Gap filling is based on weighted averages and smoothed interpolation. Total demand is added up from single demand positions if missing and single demand positions are scaled to given total demand in case they do not sum up consistently. Finally, stock changes are adjusted to ensure that market balances are closed. The consolidated commodity balance data is written to the /fao-folder in the restart directory for further usage inside the fao_balance_consolidation.

The next stept combines production and ressources with the data on commodity balances in order to consolidate the land use data. The consolidation procedure for land use categories is a separate sub-routine included under this section:

The land use consolidation step takes care of the mapping between FAOSTAT and CAPRI land use categories, imposes gap filling routines, introduces auxiliary data from UNFCCC and UNSTATS and ensures that nested land use categories consistently sum up to their totals.

The land use consistency is solved as an optimization problem ensuring (a) adding up of single crop areas to land use aggegates and (b) imposes constraints stemming from transition probabilities between

```
*   --- additional elements of prodStat_rows:for more convenient coding given in CAPRI style codes

 Set landCatUnfcccIni  "UNFCCC based land categories"/

*    compare luClass(COLS):
    CRPLND  "Same meaning as CROP but different code to indicate that source = UNFCCC"
    FORLND  "Same meaning as FORE but different code to indicate that source = UNFCCC"
    ARTIF
    GRSLND
    WETLND
    RESLND
    /;

Set landCatConsIni  "land categories to be consolidated"/
    ARTO      "Country area"
    LAND      "Land area"
    UAAR      "Agricultural area"
    CROP      "Arable (annual) and permanent crops"
    ANNC      "Any arable (temporary) crops: temp crops + temp meadows + temp pastures + fallow"
    TCRP      "Temporary crops"
    TCRPsum   "Temporary crops from summation of single crops"
    TGRS      "Temporary meadows and pastures incl fallow land"
    FRUN      "Permanent crops (mostly fruits and nuts)"
    FRUNsum   "Permanent crops (mostly fruits and nuts) from summation of single crops"
    GRAS      "Permanent meadows and pastures"
    FORE      "Forest area"
    OLNDARTIF "Other land incl artificial areas (settlements)"
    INLW      "Inland water"
    SET.LUCini
    SET.noLUCIni
    Set.landCatUnfcccIni
    GIOART "GIOART = GRAS+INLW+OLNDARTIF = ARTO-FORE-CROP = GRSLND+RESLND+WETLND+ARTIF"
    /;

 Set landCatAddIni  "additional land categories in database assingments"/
    GIOLND "GIOLND = GRAS+INLW+OLND = ARTO-FORE-CROP-ARTIF = GRSLND+RESLND+WETLND"
    OLND   "OLND = OLNDARTIF-ARTIF"
 /;
*
Set prodStat_rows(*) /
    set.prodStat_rows_ori
    Set.landCatConsIni
    Set.landCatAddIni
/;
```

different UNFCCC land use categories:

Finally, crop area levels are rescaled based on the solution from the optimization problem and yields are recalculated accordingly. The consolidated land use data is written to the /fao-folder in the restart directory.

The next step consolidates data for the milk sector. The FAOSTAT market balances differ from CAPRI in four aspects that require special adjustment in addition to the mapping and gap filling routines. (1) Farm household production is not included in output from CAPRI COCO module but in the data from FAOSTAT, (2) Liquid whey and (3) liquid skimmed milk are considered in FAOSTAT but not in COCO, (4) Raw milk is not disaggregated into a category for final consumption as required by COCO. At the end of the consolidation section the result is written to /fao-folder in the results-directory. This file is

```
--------------------------------------------------------------------------
--- FILL GAPS AND MAKE DATA CONSISTENT
--------------------------------------------------------------------------

--- in case only production or only levels are given, but yield are available for some years, use average yield

p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years)
$  ((p_observ(prodStat_reg,prodStat_crops,"onlyLevel",prodStat_years) or p_observ(prodStat_reg,prodStat_crops,"o
    $ SUM(prodStat_Years1, p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years1)) )
= SUM(prodStat_Years1, p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years1))/SUM(prodStat_Years1 $ p_

--- in case Data is complete, or yield is missing, adjust yield, so that levl*yield=prod

p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years)
$  (p_observ(prodStat_reg,prodStat_crops,"complete",prodStat_years) or p_observ(prodStat_reg,prodStat_crops,"noy
=  p_prodStatAdj(prodStat_reg,prodStat_crops,"51",prodStat_years)*1000/ p_prodStatAdj(prodStat_reg,prodStat_crop

--- in case production is missing, calculate levl*yield=prod

p_prodStatAdj(prodStat_reg,prodStat_crops,"51",prodStat_years)
$  (p_observ(prodStat_reg,prodStat_crops,"noProduction",prodStat_years) or p_observ(prodStat_reg,prodStat_crops,
=  p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years)* p_prodStatAdj(prodStat_reg,prodStat_crops,"31

--- in case activity levels are  missing, calculate levl=prod/yield

p_prodStatAdj(prodStat_reg,prodStat_crops,"31",prodStat_years)
$  ((p_observ(prodStat_reg,prodStat_crops,"noLevel",prodStat_years) or p_observ(prodStat_reg,prodStat_crops,"onl
    $ p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years))
=  p_prodStatAdj(prodStat_reg,prodStat_crops,"51",prodStat_years)*1000/ p_prodStatAdj(prodStat_reg,prodStat_crop

--- in case only yields are given (even after completions), we cannot do anything (should not be too often...)

p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years)
$  ( (     p_prodStatAdj(prodStat_reg,prodStat_crops,"41",prodStat_years))
   $ (not p_prodStatAdj(prodStat_reg,prodStat_crops,"31",prodStat_years))
   $ (not p_prodStatAdj(prodStat_reg,prodStat_crops,"51",prodStat_years))) = 0;
```

also a major input for the CAPRI GLOBAL module (/fao/fao_milkdata_...gdx).

Data on population only requires adjustments for Serbia, Montenegro, and China which is taken care of in the following step. The aggregated population time series for Serbia and Montenegro from before 2006 is prolonged to the time after whereas the respective disaggregated time series are back-casted to the period before. Data for China is aggregated. The result is written to the /fao-folder in the results-directory which is a major input for the CAPRI task "Build global database".

Consolidation of the trade flow matrix The consolidation of trade flows is split up across product specific groups to keep the task feasible in terms of computational complexity. The task is split up among 29 groups in total:

The whole procedure for creating a consistent data base as a starting point for the CAPRI task "Build global database" consists of two major tasks that are called the "groupSpecific" and "nongroupSpecific" tasks. The first one is the actual consolidation part that is done for each commodity group separately but executed in parallel. The second one is necessary for exporting the results such that they may be exploited via the GUI or be used as major input for the GLOBAL module.

The group specific task starts 29 separate consolidation processes in parallel where the actual consolidation processes are defined in the separate include file "/fao/do_trade_consolidation_for_one_group.gms".

```
  faoItemIds, as used in gams code


 SET faoItemIds_map(faoItemIds_ori,faoCombal_cols)/
     5610.61
     5910.91
     5300.100
     5153.151
     5510.51
     5120.121
     5141.141
     5071.71
     5520.101
     5525.111
     5130.131
 /;

 Set faoComBal_main(faoComBal_cols)  "main balance components are production + trade + demand" /
     51       "Production Quantity"
     61       "Import Quantity"
     91       "Export Quantity"
     100      "Domestic supply quantity"
 /;

 SET dempos(faoComBal_cols) 'domestic demand positions in food_balances' /
 101     "Feed"
 111     "Seed"
 121     "Waste"
 131     "Processed"
 141     "Food"
 151     "Other Util"
 /;
 Alias (dempos,dempos1);

* --- land use consolidation

    $$batinclude "util\title.gms" "'Step %STEP%  ----- consolidate landuse data '"
*
$include 'fao\fao_land_consolidation.gms'
*

*   --- output

    Set fao_cols / set.faoComBal_cols, 31 "area harvested", 41 "yield" ,999 "cropping index", set.landCat /
    Set fao_rows / set.faoComBal_rows ,levl /;
    Set fao_final_years(prod_yr) "years selected for FAOSTAT output file";
    fao_final_years(prod_yr) $ (p_calyea(prod_yr) GE p_calyea("1980")) =yes ;
```

The trade consolidation part requires specific FAOSTAT trade data related sets that are loaded at the beginning of the include file. There are 18 different types of output reported in the result array.

There are also 25 different statistics reported for the time series that are important intermediate indicators for the trade consolidation process.

The trade consolidation consists of eight steps in sequence that are dependent on each other, i.e. each step produces an intermediate output file that is written to the /fao folder in the restart directory for

```
*************************************************************************
*
* Set up a optimisation problem in order to :
* - impose adding up of crops (considering the double cropping index) to LU aggregates and
* - incorporate LUC matrix between UNFCCC categories
*
*************************************************************************
*
  unfcccReg(prodStat_reg) $ sum(prodStat_years, p_prodStatAdj(prodStat_reg,"ARTART","11",prodStat_years)) = yes;

  parameter  p_weight(prodStat_reg,prodStat_Rows,prodstat_years) "weight for deviations from start values";

* --- default weight
  p_weight(prodStat_reg,landcatcons,prodstat_years) = 1;
*
* --- higher weight for some areas
*     CROP is also represented by ANNC(+FRUN):
  p_weight(prodStat_reg,"CROP",prodstat_years) = .5;
  p_weight(prodStat_reg,"TCRPsum",prodstat_years) = 5;
  p_weight(prodStat_reg,"FRUNsum",prodstat_years) = 3;
* To counteract somewhat the compensatory movement of TGRS and stability of ANNC:
  p_weight(prodStat_reg,"TGRS",prodstat_years) = 3;
*
* --- lower weight for constant values before/after fstObs/lstObs
*     Note: THese need to be recomputed after completions above
  option kill = lulucFaoPos;
  lulucFaoPos(lucFaoPos)=yes;
  lulucFaoPos(landCatUnfccc)=yes;
  lulucFaoPos("GIOART")=yes;
$batinclude 'fao\end_avg_fao.gms' 'p_prodStatAdj'  'prodStat_reg' 'lulucFaoPos' '"11"' 'prodStat_years'
*
```

usage in the follow-up steps.

The process starts with the (1) SELECT step loading the raw trade data from the file "/dat/fao/-FAO_trade_matrix_...gdx" which was produced by an external data preparation routine as described under the head section of this chapter. The raw data are just unloaded without any modifications in smaller files containing only the trade flows for one of the 29 product groups which facilitates subsequent processing. The next step is (2) AGGTRADE taking care of cutting off trade below a threshold of 1.E-5 and assigning a dummy variable for the case that trade was above this threshold.

The following step (3) UVATRADE filters trade flows computes unit values after some filtering procedures and fills gaps of their national times series based on linear interpolation. Time series of the producer price index are also completed based on averaging over different time horizons, on group averages, and on unit values.

In the following step (4) STATRADE a trust indicator is computed that allows to assign a trade flow value in case of conflicting notifications between trade partners. It is based on the sum of absolute differences to partner notifications relative to total notified trade.

The next step (5) TRDTRADE calculates national linear trend lines for quantities, values, unit values and price indices.

Step (6) INITRADE prepares the trade data for the final consolidation procedure by calculating expected means of imports, exports and unit values, and by computing the trust indicator, standard errors and expected standard errors for trade quantity and units, and unit values. The trust indicator is used for adjusting the standard errors in the estimation of trade flows between partners. Higher trust indicators result in lower standard errors and lower standard errors lead to smaller deviations from reported trade,

```
*-----------------------------------------------------------------------
*   Generate group child gams processes for FAOSTAT trade consolidation
*-----------------------------------------------------------------------

Set Group /group1*group29/;
$setglobal MAXPROCDIR -maxProcDir=255
Set dumyx /Exportgui, Exportglobal/;
** map definition of groups from the GUI to those used in const.gms


SET GUItoGroup(*,group)/
     agrInputs.group1,
     lifeAnim.group2,
     nonAlcBeverage.group3,
     cereFeed.group4,
     cereFood.group5,
     cereGrain.group6,
     animFats.group7,
     feedstuff.group8,
     fruitProc.group9,
     fruitTemp.group10,
     fruitTrop.group11,
     induCrop.group12,
     meats.group13,
     meatProc.group14,
     milkProd.group15,
     nuts.group16,
     oilcakes.group17,
     vegoils.group18,
     oilseeds.group19,
     pulses.group20,
     roots.group21,
     skinhide.group22,
     skinProd.group23,
     sugaProd.group24,
     vegeProc.group25,
     vegetabl.group26,
     alcBeverage.group27,
     textiles.group28,
     cereFlour.group29
           /;

Set activeGroup(group);

 activeGroup(group) $ sum(GuiToGroup(groups,group),1) = Yes;
```

i.e. the outcome from the estimation will deviate less from the reportings for more trustworthy partners, and vice versa.

The computations are accomplished for each commodity separately.

Step (7) MODTRADE solves the trade consolidation problem by a Highest posterior density approach under constraints of (a) minimizing deviations from the expected means as computed in step (6), (b)

```
LOOP(activeGroup,
    count=count+1;


    --- generate a scratch directory for each group
    put_utility batch1 'shell' / "mkdir %scrdir%\"activeGroup.tl;

    --- Assume error level to "3" meaning "failure" for each thread.
        Write this in a gdx file, that is replaced by each process if successful
    p_errorLevl = 3;
    put_utility batch1 'gdxout' /'%scrdir%\'activegroup.tl:0 '\errorlevl'  ;
    execute_unload p_errorLevl;

    --- Report to GUI that process was started
    $$batinclude 'util\title.gms' "'Starting trade consolidation for '" activeGroup.tl;

    --- create two flag files:
        One group specific, used to determine if any groups are running after this loop
        One general (main) flag file that is deleted by any process as it finishes, to
        indicate that a new one can be submitted.

    put_utility batch1 'shell' / "echo test > %scrdir%\main.flag";
    put_utility batch1 'shell' / "echo test > %scrdir%\"activeGroup.tl".flag";

    --- Start trade consolidation as child process for this group via a batch command file
    put runit  '%GAMSpath%\gams fao\do_trade_consolidation_for_one_group.gms %MAXPROCDIR% scrdir=%scrdir%\'activeGroup.tl'
     '  --start="%START%" -lo=3 --curGrp='activeGroup.tl' --FLAG_FILE=%scrdir%\'activeGroup.tl'.flag --MAIN_FLAG_FILE=%scrd
    put /;
    put 'exit';
    putclose;

    putclose batch1 '%start%  %scrdir%\execute.bat';
    execute '%scrdir%\batch1.bat';
```

```
$include "fao\fao_trade_sets.gms"

SET Type "Type of results reported on result array"

          /
            Data          "Given notifications"
            Data_QV       "Given notifications where both data and value are non-zero"
            ORDNEXT       "Next notification year after a gap"
            ORDPRIOR      "Prior notification year before a gap"
            DATANEXT      "Next notification data after a gap"
            DATAPRIOR     "Prior notification data before a gap"
            Trend         "Linear trend line"
            Partner       "Results of linear regression on partner notification"
            WAInput       "Input in the temporal weighted average"
            WA            "Results of temporal weighted average"
            MOD           "Model results"
            Notified      "Indicator for non-zero notification"
            ExpMean       "Expected mean of a priori distribution"
            ExpStdDev     "Expected standard deviation of a priori distribution"
            MOD_T_s2      "Modified variance of error of trend line"
            BFPeriod      "Observation is in back or forecasting period"
            cur           "Indicator for current region"

            PreQuant       "Predefined quantity weights for country average"
          /;
```

minimizing dispersion around yearly country averages, (c) binding country level to world level unit values, and by (d) tying relative changes in country level unit values to relative changes in world unit values.

Finally, (8) SHOWTRADE stores the consolidated trade flow quantities in a gdx-file for exploitation and

```
SET Stat "Statistics reported for time series"
              /
                Trust   "Trust indicator derived from absolute error of notifications between partners"
                Nobs    "NuCloseLogr of observations with data or both notifications being zero"
                Mean    "Mean of notifications or both notifications being zero"
                Var     "Variance of notifications or both notifications being zero"
                CofStd  "Coefficient of variation"
                FstObs  "First notified year"
                LstObs  "Last notified year"

                T_R2    "R squared of trend line"
                T_s2      "Variance of error of trend line"
                T_Mean  "Mean trend value"
                T_Cov   "Covariance with trend variable"
                T_Var   "Variance of trend variable"
                T_2     "Sum of squared deviations of trends from mean"
                at      "param a of trend line"
                bt      "param b of trend line"

                P_R2    "R squared of regression on partner notification"
                P_S2    "Variance of error of regression on partner notification"
                P_Nobs  "Common observations with partner notification"
                P_Mean  "Mean partner notification"
                P_Var   "Variance of partner notification against trend"
                P_Cov   "Covariance with partner notification"
                P_2     "Sum of squared deviations of partner notification from mean"
                ap      "param a of regression on partner notification"
                bp      "param b of regression on partner notification"

                SUM_T_Report "Sum over time and reports, weighted with trust indicator"
```

```
================================================================================================
--- Finally the 'Trust indicator' influencing where the compromise between two conflicting notifications will be
    Key statistic = sum of abs differences to partner notifications relative to total notified trade
    (NOTE: This initial trust indicator is transformed below)

  $$batinclude "fao\set_title_child_process_window.gms" "'Define trust indicator '"

  Stats_Trade(RT,"imports",PTE_S,"imports","Trust",QV) $ Stats_Trade(RT,"imports",PTE_S,"imports","Nobs",QV)
        = SUM( (RT1,T), ABS(    Result_Trade(RT1,RT,PTE_S,"imports","Data",T,QV)
                              - Result_Trade(RT1,RT,PTE_S,"exports","Data",T,QV)))
            normalising by the total gives smaller measure than normalising each difference
          / SUM( T,  Result_Trade(RT,"imports",PTE_S,"imports","Data",T,QV));

  Stats_Trade(RT,"exports",PTE_S,"exports","Trust",QV) $ Stats_Trade(RT,"exports",PTE_S,"exports","Nobs",QV)
        = SUM( (RT1,T), ABS(    Result_Trade(RT,RT1,PTE_S,"exports","Data",T,QV)
                              - Result_Trade(RT,RT1,PTE_S,"imports","Data",T,QV)))
        / SUM( T,  Result_Trade(RT,"exports",PTE_S,"exports","Data",T,QV));


  Several reasons to define trust indicator for UVAL as sum of VALUE+QUANT:
  1) UVAL is derived from QUANT+VALUE
  2) Normalisation (sum of UVALs?) is less natural than for QUANT+VALUE (sum of QUANT or VALUE in denominator)
  3) Zero UVAL notifications are not usable (and probably frequent) but zero notifications for QUANT+VALUE may enter the index
  =>
  Stats_Trade(RT,Report,PTE_S,Report,"Trust","UVAL") $ (     Stats_Trade(RT,Report,PTE_S,Report,"Trust","QUANT")
                                                         and Stats_Trade(RT,Report,PTE_S,Report,"Trust","VALUE")
                                                         and Stats_Trade(RT,Report,PTE_S,Report,"Nobs","UVAL"))
                                                       = Stats_Trade(RT,Report,PTE_S,Report,"Trust","QUANT")
                                                       + Stats_Trade(RT,Report,PTE_S,Report,"Trust","VALUE");
```

inspection.

The second “nongroupSpecific” task in the trade consolidation part takes care of exporting the consolidated trade data to the /fao-folder in the results directory. This output is a major input for the CAPRI task “Build global database” (“fao_trade_for_global…gdx”). The trade data is complemented with data on conversion coefficients, on extraction rates, mappings between product equivalent and product codes, and between raw and processed goods, production data on the animal sector, and caseinTrade. The

```
--- define relative trust indicator
    (StDev will be scaled by relative trust indicator
     => decreases estimated StDev for partner with higher trust
     => tends to reduce deviations here and increase deviations for less 'trustworthy' partner)

$$batinclude "fao\set_title_child_process_window.gms" "'Set trust indicator up to product '" i "' of '" CARD(PTE_s) "' :

                                         exporter or importer report some trade (some year) for the flow
Stats_Trade(RT,RT1,PEM,Report,"Trust",QVU) $ (    Stats_Trade(RT,RT1,PEM,"Imports","Nobs",QVU)
                                                or Stats_Trade(RT,RT1,PEM,"Exports","Nobs",QVU))

           =      MAX(0.05,((Stats_Trade(RT1,"imports",PEM,"Imports","Trust",QVU)
                           use 'export trust' of same region if desired 'import trust' is missing
                          +Stats_Trade(RT1,"exports",PEM,"Exports","Trust",QVU)
                          $ (NOT Stats_Trade(RT1,"imports",PEM,"Imports","Trust",QVU))) $ SAMEAS(report,"Imports")
                         +(Stats_Trade(RT,"Exports",PEM,"Exports","Trust",QVU)
                           use 'import trust' of same region if desired 'export trust' is missing
                          +Stats_Trade(RT,"Imports",PEM,"Imports","Trust",QVU)
                          $ (NOT Stats_Trade(RT,"Exports",PEM,"Exports","Trust",QVU))) $ SAMEAS(report,"Exports")))
                   if not even the surrogate is there we may have only one indicator in denominator
                   => counting is safer than dividing by (Trust(RT)+Trust(RT1))/2
                 * (1. $ ( Stats_Trade(RT1,"imports",PEM,"Imports","Trust",QVU)
                         +Stats_Trade(RT1,"exports",PEM,"Exports","Trust",QVU)
                        $ (NOT Stats_Trade(RT1,"imports",PEM,"Imports","Trust",QVU)))

                 +  1. $ ( Stats_Trade(RT,"Exports",PEM,"Exports","Trust",QVU)
                         +Stats_Trade(RT,"Imports",PEM,"Imports","Trust",QVU)
                          $ (NOT Stats_Trade(RT,"Exports",PEM,"Exports","Trust",QVU))))

                   /((Stats_Trade(RT1,"imports",PEM,"Imports","Trust",QVU)
                     +Stats_Trade(RT1,"exports",PEM,"Exports","Trust",QVU)
                    $ (NOT Stats_Trade(RT1,"imports",PEM,"Imports","Trust",QVU)))

                   +(Stats_Trade(RT,"Exports",PEM,"Exports","Trust",QVU)
                     +Stats_Trade(RT,"Imports",PEM,"Imports","Trust",QVU)
                    $ (NOT Stats_Trade(RT,"Exports",PEM,"Exports","Trust",QVU))));
```

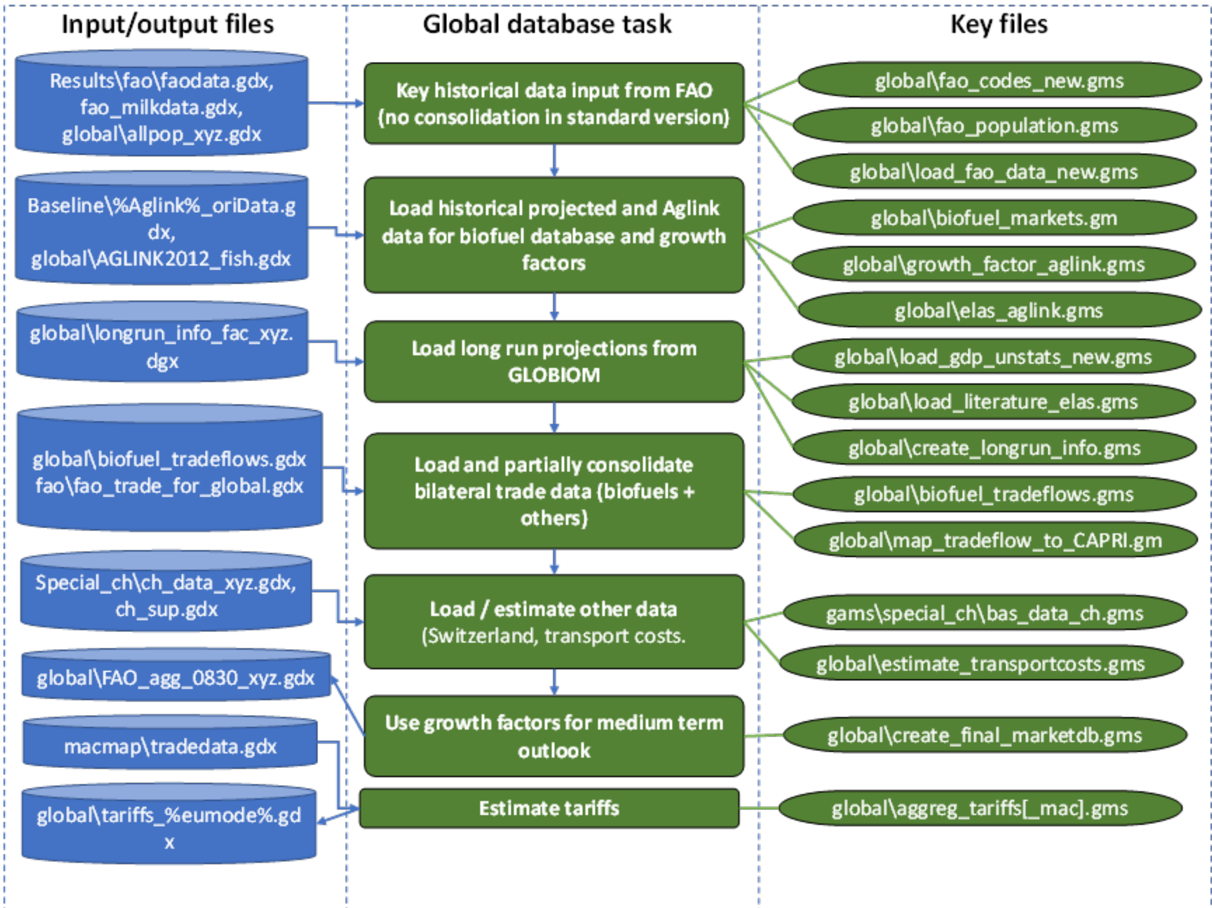
export job is included as a separate program under the nongroupSpecific task.

```
*--- now export results to GUI and Global
*

$batinclude 'util\title.gms' "'Create unified trade database for GLOBAL'" ;

if (p_stepType("nongroupSpecific"),
    LOOP(consSteps $ ((sameas(consSteps,"ExportGui")) or (sameas(consSteps,"ExportGlobal"))),

        $$batinclude 'util\title.gms' "'Starting step 'conssteps.tl' ... lasting several minutes!'" ;

*       --- generate a scratch directory for each step
        put_utility batch1 'shell' / "mkdir %scrdir%\"consSteps.tl;

*       --- Assume error level to "3" meaning "failure" for each thread.
*           Write this in a gdx file, that is replaced by each process if successful
        p_errorLevl = 3;
        put_utility batch1 'gdxout' /'%scrdir%\'conssteps.tl:0'\errorlevl'  ;
        execute_unload p_errorLevl;

*       --- Generate a flag to indicate that it is still running
        put_utility batch1 'shell' / "echo test > %scrdir%\"conssteps.tl".flag";

*       --- Start the gams process

        put runit  '%GAMSpath%\gams' ' fao\export_trade.gms %maxprocdir% o=%scrdir%\'conssteps.tl'.lst gdxc
                    ' --start "%start%" scrdir=%scrdir%\'conssteps.tl' --datdir=%datdir% --FLAG_FILE=%scrdi
        put /;
```

3.5.2 Task: Build global database

The main program [29] for this task is capri/gams/global_database.gms which collects a number of included files for separate sub-tasks, some of which being trivial, others more complex.

Figure 9: Overview on key elements in the consolidation of global data (in global_database.gms)



Source: own illustration

The program starts with including three general programs also present (possibly in task specific form) in other main programms plus the steering file (runglobal.gms) with more precise settings for the current run which may come from the GUI or from a batch file[30]:

[29] A "program" refers in this section to a file with CAPRI code for performing certain task or sub-task and which may in turn include other "code files" or "programs".

[30] A batch file is a steering file to execute a CAPRI task with all settings that are usually made in the GUI (say which simulation years) expressed equivalently in a certain language in a text file.

```
********************************************************************
$ontext

   CAPRI project

   GAMS file : GLOBAL.GMS

   @purpose  : Generate the raw data for the international market
               model
--------------------------------------------------------------------
$offtext
********************************************************************
--------------------------------------------------------------------
$include 'util\global_Settings.gms'

--------------------------------------------------------------------
$include 'runGlobal.gms'
*
--------------------------------------------------------------------
$include "util\ensure_that_optional_result_directories_are_set.gms"

--------------------------------------------------------------------
$include "global\ensure_existence_of_result_folders.gms"
```

After these general settings the programm continues in a rather standard manner with a section collecting various declarations of sets and parameters. Among these are the general sets of CAPRI (sets.gms), and the sets specific to the market model (arm_sets.gms) because the purpose of the task is to compile the data needed for the market model at the global level of CAPRI:

```
--------------------------------------------------------------------
$include 'sets.gms'
$include 'arm\arm_sets.gms'
```

The most important data source for task “Build global database” is FAOstat which involves a fairly long file (FAO_codes_new.gms) with sets and cross-sets to map from FAO regions, items, and products into the CAPRI world (defined by the code system in the annex). This serves to map some key data from FAO compiled in the previous task: population (fao_population.gms), commodity balances combined with production and landuse statistics. Furthermore special balances for dairy products are loaded (all in load_fao_data_new.gms).

The second most important group of data, both historical as well as projections, for the global market model of CAPRI come from the Aglink-Cosimo model[31], including its ex post database.

- The first $include file (load_%aglink%_new.gms[32]) includes the relevant sets to handle the Aglink data, including the cross-sets to map to CAPRI. In addition it also merges a special data set on

[31]This model is also used by DG Agri for its own outlook and provides important inputs to the CAPRI baseline.

[32]A string like %textname% is a placeholder in GAMS code for some other text to be substituted for %textname% during the program execution. In this example it holds the name for the specific Aglink-Cosimo version that should be loaded.

```
*  --- sets for FAO data, UNSTATS, WFM and cross sets
*
$include 'global\FAO_codes_new.gms'
------------------------------------------------------------
*-----------------------------------------------------------
*
*  Load population data and map to CAPRI country codes
*
*-----------------------------------------------------------
------------------------------------------------------------
$include 'global\fao_population.gms'
*
*-----------------------------------------------------------
*
*  Load FAO SUAs and map to CAPRI
*
*-----------------------------------------------------------
------------------------------------------------------------
$include "global\load_FAO_data_new.gms"

*-----------------------------------------------------------
*
*  Use AGLINK data for some positions ..
*
*-----------------------------------------------------------
------------------------------------------------------------
$include 'global\load_%aglink%_new.gms'
* --- Build international market balances for biofuels based or

$include "global\biofuel_markets.gms"
------------------------------------------------------------
* --- Calculate growth factor
------------------------------------------------------------
$include "global\growth_factor_aglink.gms"
------------------------------------------------------------
*  --- import AGLINK elasticities
*
$INCLUDE "global\elas_aglink.gms"
------------------------------------------------------------
```

fish markets with other original Aglink data.

- The next file (biofuel_markets.gms) data set builds biofuel market balances for non-EU regions, as FAOstat do not include biofuels and their demand for agricultural goods, but biofuel markets for EU countries are covered in COCO1 and the special projection tool for European regions (captrd.gms). This is an example for special treatments in the biofuel sector which are often unavoidable.

- The third $include file (growth_factor_aglink.gms) performs the first steps of data processing for the bulk of market balance data from Aglink-Cosimo: Items are mapped to the CAPRI codes, prices are converted from national currency into Euros, and projections from Aglink-Cosimo are converted into growth factors such that they may be conveniently combined with historical data from FAOstat (in "create_final_marketdb.gms).
- Finally the fourth file collects and maps various elasticities to the CAPRI coding system in order to serve as prior values in the CAPRI elasticity trimming code (trim_par, called during market calibration). Of these currently only the supply elasticities are used, such that this program gives a typical example of historicall grown coding: Obsolete elements are more often overwritten or ignored rather than deleted.

The next three $include files cover additional macroeconomic data from UNstats (load_gdp_unstats_new.gms), include and map long run projections beyond the Aglink horizon from the GLOBIOM[33] model (create_longrun_info.gms, comment "merge FAO and IMPACT 2050 projections is obsolete), and collect prior values for demand elasticities from the literature (collect_literature_elas.gms, whereas demand elasticties from Aglink-Cosimo are ignored).

```
*---------------------------------------------------------------------------------
*
*   Select GDP and EXPD from UNSTATS
*
*---------------------------------------------------------------------------------
------------------------------------------------------------------ 2 line(s) not displayed
$include 'global\load_gdp_unstats_new.gms'
------------------------------------------------------------------ 2 line(s) not displayed
*---------------------------------------------------------------------------------
*
*  Merge FAO and IMPACT 2050 projections
*
*---------------------------------------------------------------------------------
*
$ifi %updatelongrun%==on $include 'global\create_longrun_info.gms'
------------------------------------------------------------------ 3 line(s) not displayed
*--------------------------------------------------------------------------------
*
* Reading income elasticities from demel 2008
*
*--------------------------------------------------------------------------------
------------------------------------------------------------------ 2 line(s) not displayed
$include 'global\load_literature_elas.gms'
*
```

It may be seen that "create_longrun_info.gms" is active or not depending on a setting from the GUI or a batch file. Similar to the code processing Aglink information it includes sets and mappings to handle the GLOBIOM information. Another similarity with the Aglink related files is that this code basically needs annual adjustments, because some definitions are changing from year to year and there are two GLOBIOM versions to distinguish, one with a certain EU focus, the other one with a perfectly global

[33]The GLOBIOM model is the second model providing key inputs to the CAPRI baseline. It is mainly developed and operated at IIASA.

orientation. Finally, it may be mentioned that the projections are introduced into the CAPRI world mostly in the form of growth factors.

The CAPRI market model is spatial and therefore requires data on bilateral trade flows. These are covered in two include files, the first one dealing with the special case of biofuel trade flows, the second one with the general case.

```
*-------------------------------------------------------------
*
*  Calculate trade flow from consolidated FAO data
*
*-------------------------------------------------------------
------------------------------------------------------------
$include 'global\biofuel_tradeflows.gms'
------------------------------------------------------------
$include "global\map_tradeflow_to_CAPRI.gms"
```

Biofuel trade requires a special treatment again because FAOstat does not cover these. Instead, bilateral trade flows are constructed using total exports and imports from AGLINK and trade data from COMEXT, USDA and FO-Licht. By contrast the data for the trade matrix for other commodities is from FAOstat.

Both the biofuel trade matrix as well as non-biofuel trade are rendered "approximately" consistent with the totals from the previously collected market balance data with a small optimisation model that tries to minimise deviations from the prior data. File "map_tradeflow_to_capri.gms also tackles the problem of bilateral trade data entirely missing. In this case (relevant for fish, for example) default trade flows are introduced where commodities are mostly supplied by the largest exporters or imported by the most important importers.

After consolidating the trade flows two special data sets need to be considered. The first is a special data set on Switzerland checked in detail by the Swiss Federal Office on Agriculture (FOAG) and including trade flows involving Switzerland (hence included *after* the previous consolidation such that these data overwrite the trade flow information but also the market balance information from FAOstat).

The second is a transport cost matrix estimation using the original FAOstat trade matrix (so before gap filling and consolidation) and distance related information from CEPII. Together with price information the transport costs are estimated to provide a link between CIF and FOB prices for bilateral tradeflows.

The next $include file extends the Aglink-Cosimo projections to 2030, if needed, with a trend estimation involving a number of pragmatic modifications (such as the trend line passing trough the last observation). Then the the growth factors computed previously or the default trends are used to estimate a medium term outlook projections for global market balances, prices or GDP. These projections do however not include any consictency checks on closed market balances or similar properties. This is achieved in the baseline calibration only.

Finally, data on trade policy variables such as applied and scheduled tariffs, tariff rate quotas or bilateral trade agreements are collected from the Agricultural Market Access Database (AMAD, obsolete current

```
*--------------------------------------------------------------------------
*
*  Specific expost data for Switzerland
*
*--------------------------------------------------------------------------
*
$include '..\gams\special_ch\bas_data_ch.gms'
*
*--------------------------------------------------------------------------
*
*  Estimate bilateral transport cost matrix
*
*--------------------------------------------------------------------------
------------------------------------------------------------ 2 line(s) not displayed -
$if %updateTransportcosts%==ON $include "global\estimate_transportcosts.gms"
------------------------------------------------------------ 1 line(s) not displayed -

*--------------------------------------------------------------------------
*
*  Merge Faostat database with Aglink projections and simple trends for products not covered in AGLINK
*
*--------------------------------------------------------------------------
------------------------------------------------------------ 5 line(s) not displayed ---------------
$include global\create_final_marketdb.gms
------------------------------------------------------------ 2 line(s) not displayed ---------------
*--------------------------------------------------------------------------
*
*  Store data in gdx format
*
*--------------------------------------------------------------------------
```

version) or from the MacMaps database (%macMap%)[34]==on, but not yet activated under Star2.4).

The very last include file is probably also the least important one: FAPRI projections had a more important role several years ago, are not updated anymore and presumable affect less than a dozen numbers (if any at all) in the global database compiled in this task:

3.6 Policy data

3.6.1 Policy data linked to European and international markets

Data on trade policies on the global agri-food markets first appear in the global database of CAPRI. More specifically, the original tariff data are aggregated to the commodity definitions of CAPRI in the tariff aggregation module. The tariff aggregation procedures in CAPRI require data not only on the tariffs themselves but also on traded quantities and import prices. two tariff/trade databases are supported currently:

1. AMAD database, which is unfortunately discontinued by OECD and is expected to be phased out from the CAPRI system as database updates will be no longer available.
2. ITC-MacMap and ITC-TradeMap database. MacMap includes ad valorem equivalent tariff rates at the 6-digit level of the Harmonized System (HS6), while TradeMap supplies the necessary trade statistics (quantities and prices) for the aggregation.

[34]See GAMS Documentation on The GAMS Call and Command Line Parameters (`https://www.gams.com/latest/docs/UG_GamsCall.html`

```
*-------------------------------------------------------------
*
*  Aggregate tariffs from amad
*
*-------------------------------------------------------------
-------------------------------------------------------------
$ifi     %macmap%==on $include "global\aggreg_tariffs_mac.gms"
$ifi not %macmap%==on $include "global\aggreg_tariffs.gms"
-------------------------------------------------------------
*-------------------------------------------------------------
*
*  Map results from FAPRI baseline
*
*-------------------------------------------------------------
-------------------------------------------------------------
$include "global\convert_fapri.gms"
*
$exit
```

The tariff aggregation results are part of the .gdx output of the global module, and can be found in results/global/tariffs.gdx.

Although the tariffs in the tariff databases should already reflect the tariff schedules of the implemented Free Trade Agreements (FTA) on global agricultural markets, CAPRI nevertheless explicitly includes data on a number of FTAs. That FTA-specific policy information enters the CAPRI system in the market model calibration workstep (gams/arm/def_tariff.gms, see Table below for the list of implemented FTAs).

Specific trade policy data on Switzerland enters CAPRI both in the tariff aggregation module (in the global) part and also during market model calibration, and often overwrites tariff data from the above sources. The Switzerland-specific datasets in CAPRI are managed by the team at the Federal Office for Agriculture, and the data are based on national trade statistics: Swiss-Impex database and the databases of the TRIMAG tariff aggregation tool[35]. The relevant model code is collected under the subfolder gams/special_ch.

Tariffs and Tariff Rate Quotas Data on trade policy instruments other than tariffs (Tariff Rate Quotas, export subsidies, entry price system and flexible levies) enter CAPRI directly in the market model calibration workstep. Note that the ad valorem equivalent tariff rates in MacMap already include an estimated equivalent tariff rate for TRQs. Nevertheless, the CAPRI market model separates TRQs from fixed tariff rates by using a sigmoid function-representation of the TRQ regime switch mechanism[36].

[35] For more information on TRIMAG please refer to Himics, M., Listorti, G., Tonini, A., 2019. Simulated economic impacts in applied trade modelling: A comparison of tariff aggregation approaches. Economic Modelling. doi:10.1016/j.econmod.2019.08.007

[36] Tariff rates under TRQ vary between the lower in-quota and the higher out-of-quota rates, depending on the quota fill rates. For more details on the methodological approach please visit section 5.4.12

The TRQ system of the EU is included in great detail, based on DG AGRI.information. Data on TRQ orders are aggregated to the geographical and commodity definitions of CAPRI in dat/arm/TRQ_orderds.gms. Specific GAMS routines convert some of the compound TRQs into ad valorem TRQs if necessary[37](gams/arm/convert_compound_trqs.gms).

Export subsidies Data on (EU) export subsidies (e.g. maximum commitments) enter the system in the market model calibration workstep, under gams/arm/calc_feoga.gms. Current WTO negotiations aim at the full phase-out of export subsidies, and accordingly, the EU does not grant export subsidies to agricultural products currently. Nevertheless, the possibility to introduce export subsidies in policy scenarios is kept in CAPRI (e.g. Border Carbon Adjustment policies may take the form of export subsidies, for which the availability of the export subsidy mechanism is valuable).

Producer subsidies Producer Subsidy Estimates (PSE) are formally part of the price transmission equations in the market model from the (equilibrium) market prices to the producer prices. However, a complete and up-to-date PSE dataset is not part of CAPRI at the moment, and therefore PSE support is not considered in the standard version of CAPRI (was only available in some specific model applications only).

Consumer subsidies Consumer Subsidy Estimates (CSE) are formally part of the price transmission equations in the market model from the (equilibrium) market prices to the consumer prices. However, a complete and up-to-date CSE dataset is not part of CAPRI at the moment, and therefore CSE support is not considered in the standard version of CAPRI (was only available in some specific model applications only).

Public Intervention purchases and sales Data on public intervention (stocks, buy-ins, releases, administrative prices etc.) enter the system in the market model calibration workstep, under gams/arm/calc_feoga.gms. Once one of the most impactful measure of the Common Agricultural Policy (CAP), public intervention has been reduced regarding its scope and is currently only available for EU farmers as an emergency measure (in crisis situations, e.g. under exceptionally high price fluctuations). Therefore, its use in CAPRI is also limited to scenario analysis.

Further update of this section is pending

[37]Compound TRQs are TRQs applying a compound tariff (combination of specific and ad valorem) on the in-quota or out-of-quota imports. For methodological reasons, the compound tariffs might need to be converted into their ad valorem equivalent rates.

Chapter 4

Baseline generation

4.1 On forecasts in simulation models

The purpose of a baseline is to serve as a comparison point or comparison time series for counterfactual analysis. The baseline is interpreted as a projection in time covering the most probable future development of the European agricultural sector under the status-quo policy and including all future changes already foreseen in the current legislation.

Conceptually, the baseline should capture the complex interrelations between technological, structural and preference changes for agricultural products world-wide in combination with changes in policies, population and non-agricultural markets. Given the complexity of these highly interrelated developments, baselines are in most cases not a straight outcome from a model but developed using a combination of trend analysis, model runs and expert consultations. In this process, model parameters such as elasticities and exogenous assumptions, e.g., technological progress captured in yield growth, are adjusted in order to achieve plausible results (as regarded by experts, e.g. European Commission projections). It is almost unavoidable that the process is somewhat intransparent.

The kind of baseline process described above is not specific to CAPRI, but is found also in other large scale modelling projects. Two typical examples are discussed here.

- In the case of the Aglink modelling system of the OECD, questionnaires are sent out to the OECD Member States covering all endogenous and exogenous variables of Aglink. The Member States fill in time series regarding the future developments for their respective countries. The projections reported by the member states may themselves stem from country specific model baselines, expert consultations, trend analyses or other sources – in many cases, their provenience is not known in detail. The OECD then sets the constant terms in all behavioural equations of Aglink so that the country modules would exactly recover the values for the endogenous variables for that country found in the questionnaires at the values assigned to the exogenous variables. Clearly, as the countries fill their questionnaires without knowing about the future expectations of other OECD Members, the expectations of the different teams e.g. regarding imports/exports or world market prices may differ and lead to values at country level which are mutually not compatible when linked globally together in the modelling framework. To eliminate such differences, the OECD will repeatedly start Aglink to generate technically compatible results and receive comments on these runs which will lead to updated data in the questionnaires and thus new shift terms in the behavioural equations.

- The second example is that of FAPRI model, where a so-called melting down meeting is organised where the modellers responsible for specific parts of the system come together with market experts. Results are discussed, parameters and assumptions changed until there is consensus. Little is known about how the process works exactly, but both examples underline the interaction between model mechanisms and ex-ante expectations of market experts.

As is the case in other agencies, the CAPRI baseline is also fed by external ("expert") forecasts, as well as by trend forecasts using data from the national 'COCO' and regionalized CAPREG databases (sections 3.2 and 3.3). The purpose of these trend estimates is, on the one hand, to compare expert forecasts with a purely technical extrapolation of time series and, on the other hand, to provide a 'safety net' position in case no values from external projection are available. Usually the projections for a CAPRI baseline are a combination of expert data (e.g. from FAO, European Commission, World Bank, other research teams and even private entreprises) and simple statistical trends of data contained in the CAPRI database.

4.2 Overview of CAPRI baseline processes

The creation of a baseline in CAPRI is considered a "workstep" in the GUI, and it consists of three distinct tasks. In addition to those three tasks, the modeller will usually want to perform a simulation of "no change" to reproduce the calibrated baseline with the model. Figure below illustrates the principal data flows involved in the baseline process. Each step is further described in a separate section of this chapter.

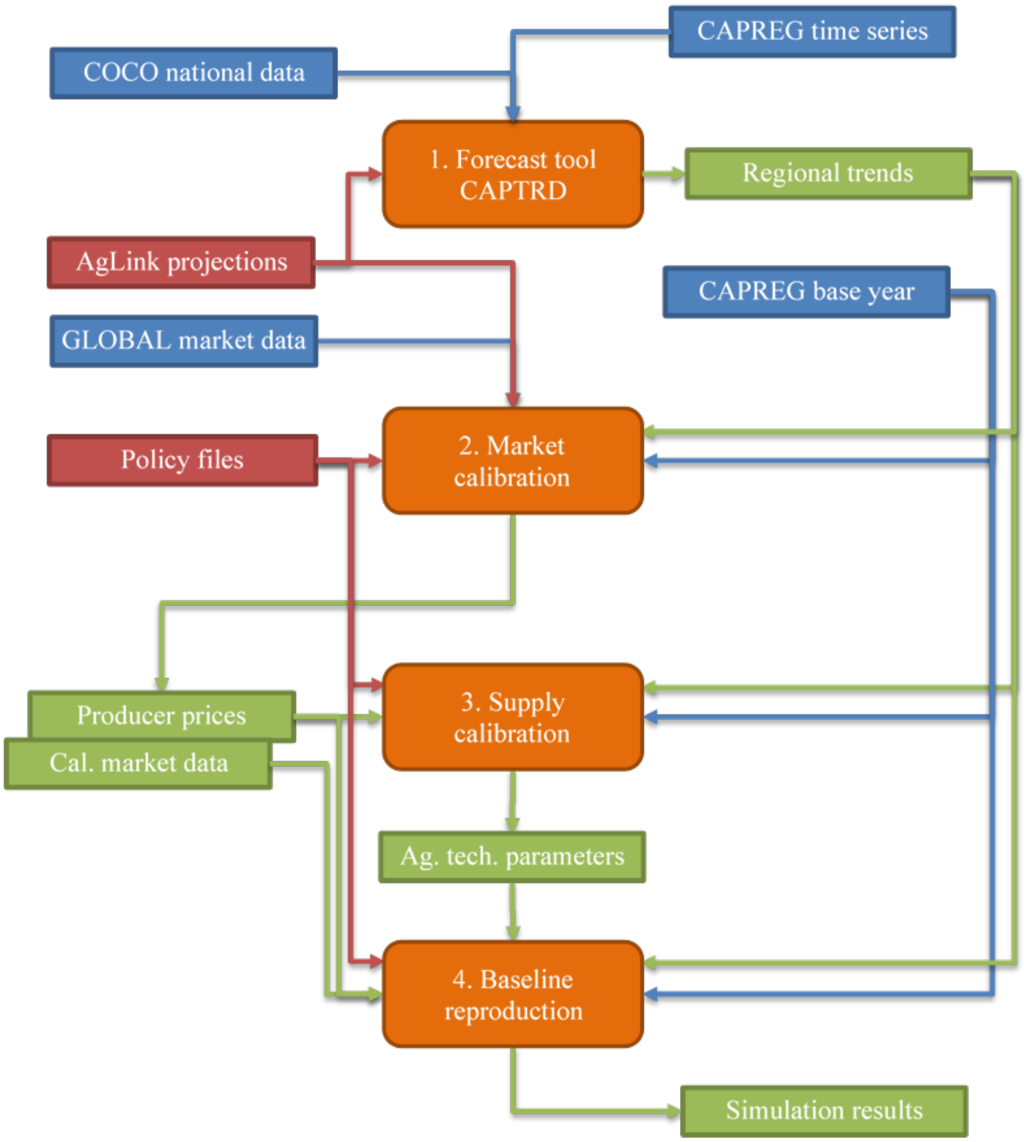
The forecast tool CAPTRD uses the consolidated national and regional time series from COCO and CAPREG together with external projections from the AgLink model. The result is a projection for the key variables in the agricultural sector (activity levels and market balances) of all regions in the supply models (EU+) that is consistent with the supply model equations.

1. Next task is the market model calibration. That task uses the same AgLink projections, complemented with the harmonized trade database GLOBAL (see section 3.5), the baseline policy files, the regional data for the base year (CAPREG) and the regional trends coming from CAPTRD. The output includes a market data set that is consistent with the regional trends, with calibrated parameters to steer behavioural functions, and adds producer prices to be used by the supply models.

2. The third task is the calibration of the supply models. This step also uses the regional data base, regional trends, and policy files, and calibrates various technical and behavioural economic parameters of the supply models so that the projected regional production is the optimal production at the producer prices coming from the market model calibration.

3. Finally, the modeller typically wants to perform a simulation using all the calibrated parameters and projected data. The purpose is twofold: to verify that the calibration of the baseline worked as intended and to generate all reports for inspection in the GUI.

Figure 10: Overview of CAPRI baseline process



Source: own illustration

4.3 Forecast tool CAPTRD

The tool providing projections for the European regions (the EU as of 2019, Turkey, Norway, Albania, Northern Macedonia, Montenegro, Bosnia and Herzegovina, Kosovo and Serbia) in CAPRI is called CAPTRD. It operates in several steps:

- Step 1 involves independent trends on all series, providing initial forecasts and statistics on the goodness of fit or indirectly on the variability of the series.
- Step 2 imposes constraints like identities (e.g. production = area * yield) or technical bounds (like non-negativity or maximum yields) and introduces specific expert information given on the MS level.
- Step 3 includes expert information on aggregate EU markets, typically coming from the AgLink model or GLOBIOM. This external data is not available for all individual countries in CAPRI, but for larger regions. Therefore, several countries must be simultaneously estimated in order to ensure proper use of this important prior information.
- Step 4 Depending on the aggregation level chosen the MS result may be disaggregated in subsequent steps to the regional level (NUTS2) or even to the level of farm types.

The trends estimated in CAPTRD are subject to consistency restrictions in steps 2 and 3. Hence they are not independent forecasts for each time series and the resulting estimator is hence a system estimator under constraints (e.g. closed area and market balances). Nonetheless, the trends are mechanical in the sense that they respect technological relationships but do not include any information about behavioural functions or policy developments[1].

CAPTRD results are in turn only the first of several steps before a full CAPRI baseline is ready to use. The rest of this chapter focuses on CAPTRD.

4.3.1 Step 1: Independent, weighted nonlinear least squares

Before entering into the details it should be stated that ultimately almost any projection may be reduced to a particular type of trend projections, at least if the exogenous inputs, such as population, prices or household expenditure are also projected (usually by other research teams) as functions of time. In this sense trend projection may provide a firm ground on which to build projections and this is exactly their purpose in our work.

The first ingredient in the estimator is the trend curve itself which is defined as:

$$X_{r,i,t}^{j,Trend} = a_{r,i,j} + b_{r,i,j} t^{c_{r,i,j}} \tag{4.1}$$

where the parameters a, b and c are to be estimated so that the squared deviation between given and estimated data are minimized. The X stands for the data and represents a five dimensional array, spanning up products i and items j (as feed use or production), regions r, points in time t and different data status such as *'Trend'* or *'Observed'*. The trend curve itself is a kind of Box-Cox transformation, as parameter c is used as the exponent of the trend. For *c* equal unity, the resulting curve is a straight line, for c

[1]The only exceptions are the quota regimes on the milk and sugar markets which are recognised in the trend projections.

between 0 and 1, the curve is concave from below, i.e. increasing but with decreasing rates, whereas for c > 1, the curve is convex from below, i.e. increasing with increasing rates. In order to prevent differences between time points to increase sharply over the projection period, the parameters *c* are restricted to be below 1.2.

This form has the advantage of ensuring monotonic developments whereas quadratic trends often gave increasing yields for the first part of the projection period and afterwards a decrease. Another conclusion from the early explorations was that it is useful to define the trend variable $t_{1984} = 0.1, t_{1985} = 0.2, t_{1986} = 0.3$ etc., giving a potentially strong nonlinearity in the early years of the database (where the frequency of high changes, possibly due to data weaknesses was high) and a rather low nonlinearity in the projection period.

The ex-post period usually covers the period from 1985 towards the end of the underlying CAPREG output (file *res_time_series.gdx*) which is typically 4-6 years before the current year. The national level COCO data may have somewhat longer series than the regional CAPREG data. To account for this different availability of ex post data the following sets should be distinguished:

- *Expost*: defined from the length of the series in CAPREG output *res_time_series.gdx*
- *Exante*: covering any sequence of intermediate result years up to the user specified final year[2].
- *ExanteD*: Ex ante years with additional COCO1 data (assigned in '*captrd/load_coco1_data.gms*')
- *ExpostT*: Union of Expost and ExanteD = years with data for trend estimation

The estimator minimises the weighted sum of squares of errors using the trend variable as weights:

$$wSSE_{r,i,j} = \sum_{expost} \left(X^{j,\text{"Data"}}_{r,i,expost} - a_{r,i,j} + b_{r,i,j} t^{c_{r,i,j}}_{expost} \right)^2 t_{expost} \tag{4.2}$$

The weighting with the trend was introduced in the exploration phase based on the following considerations and experience. First of all, it reflects the fact that statistics from the early years (mid eighties) are often less reliable then those from later years. Secondly, even if they are reliable, older data will tend to contribute less useful information than more recent ones due to ongoing structural change. For this reason we have discarded any years before 1992 for the New MS, for example, but the data from the mid 90ies may nonetheless represent a situation of transition that should count less than the recent past. In technical terms the step 1 estimates are found by a grid search over selected values of parameter c with analytical OLS estimates for parameters a and b (see *'captrd/estimate_trends.gms'*) that have been found identical to those of the econometric package Eviews for a given value of c (holds also for wSSE).

4.3.2 Step 2.1: Consistency constraints in the trend projection tool

Step 2 adds the consistency conditions and thus transforms the naïve independent trends into a system estimator. In almost all cases, the unrestricted trend estimates from the first step would violate one or several of the consistency conditions. We want to find estimates that both fit into the consistency constraints and exploit the information comprised in the ex-post development in a technical feasible way.

[2] For technical reasons some years are "obligatory" result years, for example the year immediately following after the last ex post year.

Consider the identity that defines production as hectares/herd sizes times yield. Running independent trend estimates for barley area, barley yield and barley production will almost certainly produce estimates where production is not equal to yield times area. One solution would be to drop one of the three estimates, say yield, and replace it instead by the division of forecasted production by forecasted area. However, by doing so, we throw away the information incorporated in the development of barley yield over time. Adding relations between time series hence helps us to exploit more information than is contained in single series.

When consolidating simultaneously the different Step1 estimates, we will minimise the squared deviations from values computed in Step 1, in the following called "supports", while complying with all constraints. A risk is that shaky trends may give a forecast line with an end point far away from ex-post observations. Hence we need safeguards pulling our estimates to a 'reasonable' value in such cases.

The confidence interval from the Step 1 trend estimation will not help, as it will be centred around the last projection value and as it will simply be quite large in case of a bad R^2. However, we may use the idea underlying the usual test statistics for the parameters related to the trend (*a,b,c*). These statistics test the probability of (*a,b,c*) being significantly different from zero. It can be shown that these tests are directly related to R^2 of the regression. If the zero hypotheses would be true, i.e. if the estimated parameters would have a high probability of being zero, we would not use the trend line, but the mean of the series instead.

This reasoning is the basis for the supports derived from the Step 1 estimates in CAPTRD (*'captrd/define_stats_and_supports.gms'*), after some modifications. First of all, we used a three-year average based on the last known values as the fallback position and not the mean of the series. Secondly, in typical econometric analysis, test statistics would only be reported for the final estimation layout, some variables would have been dropped from the regression beforehand if certain probability thresholds are undercut. For our applications, we opted for a continuous rule as the choice of threshold values is arbitrary. The smaller the weighted R^2 the stronger the estimates are drawn towards our H0 – the value is equal to the recent three year average:

$$X_{r,i,exante}^{j,\text{"Support"}} = wR_{r,i,j}^2 \left(a_{r,i,j} + br,i,jt_{exante}^{c_{r,i,j}}\right) + \left(1 - wR_{r,i,j}^2\right) X_{r,i,bas}^{j,\text{"Data"}} \tag{4.3}$$

where

$$wR_{r,i,j}^2 = 1 - wSSE_{r,i,j}/wSST_{r,i,j} \tag{4.4}$$

and the weighted total sum of squares is defined analogous to equation below:

$$wSST_{r,i,j} = \sum_{expost} \left(X_{r,i,expost}^{j,\text{"Data"}} - X_{r,i,wAve}^{j,\text{"Data"}}\right)^2 t_{expost} \tag{4.5}$$

with a trend weighted average

$$X_{r,i,wAve}^{j,\text{"Data"}} = \sum_{expost} X_{r,i,expost}^{j,\text{"Data"}} \cdot t_{expost} / \sum_{expost} t_{expost} \tag{4.6}$$

How is this rule motivated? If R^2 for a certain time series is 100%, in other words: for a perfect fit, the restricted trend estimate is fully drawn towards the unrestricted Step 1 estimate. If R^2 is zero, the trend curve does not explain any of the weighted variance of the series. Consequently, the support is equal to the 'base data'. The 'base data' represent a three-year average around the last three known years. For all cases in between, the supports are the weighted average of the unrestricted trend estimate weighted with R^2 and the three-year average weighted with ($1-R^2$). Generally, all trend estimates are restricted to the non-negative domain.

The above definition of supports works for series with *expost* data from CAPREG only as well as for those series with an extended set of observations (*expostT*, see above). The only difference is whether the three year average denoted above simply with "bas" is calculated using the three last years from set *expost* or from set *expostT* (BASM or BAST in *'captrd/define_stats_and_supports.gms'*).

Our objective function for Step 2 will be the sum of squared deviations from the supports defined above, weighted with the variance of the error terms from the first step:

$$Penalty = \sum_{r,i,j,expost} \left(\frac{X^{j,"Trend"}_{r,i,exante} - X^{j,"Support"}_{r,i,exante}}{\sqrt{X}^{j,"Step1"}_{r,i,verErr}} \right)^2 \tag{4.7}$$

where the weighted variance of errors is

$$X^{j,"Step1"}_{r,i,verErr} = wSSE_{r,i,j} / \left(\sum_{expost} t_{expost} - 1 \right) \tag{4.8}$$

The variance of the error term is used to normalise the squared deviations from all series which serves two purposes. First the weighted error variance is decreasing with the mean of the explanatory variable. Normalizing with it will hence ensure that the penalty targets relative rather than absolute deviations. Otherwise the solver would only tackle the deviations from "large" crops, say soft wheat, and more or less ignore the deviations of oats, for example. Secondly the deviations from the support are penalized stronger where the Step 1 trend had a high explanatory power and therefore a low variance of the error term.

The constraints in the trend projection enforce mutual compatibility between baseline forecasts for individual series in the light of relations between these series, either based on definitions as 'production equals yield times area' or on technical relations between series as the balance between energy deliveries from feed use and energy requirements from the animal herds. The set of constraints is deemed to be exhaustive in the sense as any further restriction would either not add information or require data beyond those available. The underlying data set takes into account all agricultural activities and products according to the definition of the Economic Accounts for Agriculture.

The constraints discussed in the following (from *'captrd/equations.gms'*) can be seen as a minimum set of consistency conditions necessary for a projection of agricultural variables. The full projection tool features further constraints especially relating to price feedbacks on supply and demand.

Constraints relating to market balances and yields Closed market balances (CAPTRD eq. MBAL_) define the first set of constraints and state that the sum of imports (IMPT) and production (GROF) must be equal to the sum of feed (FEDM) and seed (SEDM) use, human consumption (HCOM), processing (INDM,PRCM,BIOF), losses (LOSM) and exports (EXPT):

$$
\begin{aligned}
X_{r,i,t}^{IMPT,Trend} + X_{r,i,t}^{GROF,Trend} = & X_{r,i,t}^{FEDM,Trend} + X_{r,i,t}^{SEDM,Trend} + X_{r,i,t}^{PRCM,Trend} \\
& + X_{r,i,t}^{INDM,Trend} + X_{r,i,t}^{BIOF,Trend} + X_{r,i,t}^{LOSM,Trend} \\
& + X_{r,i,t}^{HCOM,Trend} + X_{r,i,t}^{EXPT,Trend}
\end{aligned}
\tag{4.9}
$$

Where r are the Member States of the EU, i are the products, t the different forecasting years, corresponding to the equation. In the case of secondary products (dairy products, oils and oilcakes, for example) production is given on item *MAPR*. Domestic use *DOMM* (sum of the right hand side without exports) and net trade *NTRD* are defined in separate equations (*DOMM_*, *NTRD_*) not reproduced here. They do not act as constraints but permit a link to expert projections for EU markets in Step 3.

Secondly, production of agricultural raw products (*GROF*) is equal to yield times area/herd size (*LEVL*) where acts are all production activities (eq. *GROF_*):

$$
X_{r,i,t}^{GROF,Trend} = \sum_{acts} X_{r,i,t}^{acts,Trend} X_{r,LEVL,t}^{acts,Trend} \tag{4.10}
$$

The market balance positions for certain products enter adding up equations for groups of products (cereals, oilseeds, industrial crops, vegetables, fresh fruits, fodder production, meat, eq. *MBALGRP_*). As an example, total cereal production is equal to the sum over the produced quantities of the individual cereals.

$$
X_{r,pro\_grp,t}^{MrkBal,Trend} = \sum_{i \in pro\_grp} X_{r,i,t}^{MrkBal,Trend} \tag{4.11}
$$

Constraints relating to land use and cropping area Adding up over the individual crop areas defines the total utilizable agricultural area (*UAAR,LEVL, AREAB_*):

$$
X_{r,LEVL,t}^{UAAR,Trend} = \sum_{crops} X_{r,LEVL,t}^{crops,Trend} \tag{4.12}
$$

Adding up over the individual crop areas defines (in *GRPLEVL_*) the level of groups (set GrpC = {cereals, oilseeds, industrial crops, vegetables, fresh fruits, fodder production on arable land}):

$$
X_{r,LEVL,t}^{GrpC,Trend} = \sum_{crops \in GrpC} X_{r,LEVL,t}^{crops,Trend} \tag{4.13}
$$

Adding up over mutually exclusive land use (in *LANDUSEB_*, for set LandUseARTO, see Annex: Tables 7-9) defines the total area (*ARTO,LEVL*):

$$
X_{r,LEVL,t}^{ARTO,Trend} = \sum_{LandUseARTO} X_{r,LEVL,t}^{LandUseARTO,Trend} \tag{4.14}
$$

Constraints relating to agricultural production Another Equation (*OYANI_*) links the different animal activities over young animal markets:

$$X^{GROF,Trend}_{r,oyani,t} - X^{STCM,Trend}_{r,oyani,t} = \sum_{iyani \leftrightarrow oynai} X^{GROF,Trend}_{r,iyani,t} \tag{4.15}$$

Where *oyani* stands for the different young animals defined as outputs (young cows, young bulls, young heifers, male/female calves, piglets, lambs and chicken). These outputs are produced by raising processes, and apart from stock changes *STCM* (defined in Equation *SOYANI_*, not reproduced here), they are completely used as inputs in the other animal processes (fattening, raising or milk producing).

For those activites that have been split up in the database into a high and low yielding variant (DCOW, BULF, HEIF, GRAS) with 50% for each, this split is maintained (*SPLITFIX_*)

$$X^{splitactlo,Trend}_{r,LEVL,t} = X^{splitacthi,Trend}_{r,LEVL,t} \tag{4.16}$$

The purpose of this split has been to permit an endogenous variation of yields also for animal activites, but so far no statistical information on the distribution of intensities has been available. Hence "intensive" has been *defined* to represent the upper 50% of the total distribution and it makes sense to maintain this split also in the baseline.

Animal herds (HERD) are related to animal activity levels through the process length in days (DAYS) via *HERD_*.

$$X^{maact,Trend}_{r,HERD,t} = X^{maact,Trend}_{r,LEVL,t} \cdot X^{maact,Trend}_{r,DAZS,t}/365 \tag{4.17}$$

The process length is fixed to 365 days for female breeding animals (activities DCOL, DCOH, SCOW, SOWS, SHGM, HENS) such that the activity level is equal to the herd size[3]. For fattening activites the process length, net of any empty days (relevant for seasonal sheep fattening in Ireland, for example) times the daily growth should give the final weight after conversion into live weight with the carcass share *carcassSh* and consideration of any starting weight *startWgt* in *FinalWgt_*

$$\begin{aligned} X^{maact,Trend}_{r,yield,t}/carcassSh_{maact} =& startWgt_{maact} + X^{maact,Trend}_{r,DAILY,t} \\ & \cdot (X^{maact,Trend}_{r,DAYS,t} - X^{maact,data}_{r,EDAYS,BAS}) \end{aligned} \tag{4.18}$$

As the daily growth is an important input into the livestock sector requirement functions it turned out useful to explicitely link it to the yields in terms of meat, both in the expost data (accounting identities in COCO) and here in the projections. Heavier animals require in this way a higher daily growth and/or a longer fattening period. For all inputs into the requirement functions hard constraints have been imposed

[3]The wording for animal numbers is a continuous source of confusion that may also affect older parts of this documentation or table headings from the CAPRI GUI. It is therefore recommendable to reserve the term "herd" strictly to stock variables (animals countable at a particular day) whereas the flow variable "produced heads per year" is the activity level for fattening activities.

(without the possibility to relax them in the solution process) to ensure that projected variables are fully in line with these contraints, mostly over bounds in *'estimate_MS.gms'*, but also through a specific (ad hoc) equation for male adult cattle that permits at most a daily growth of 0.4+500*0.0016 = 1.2 kg per day for a 500 kg final live weight, but more for heavier animals (*DAILYUP_*).

$$X^{bulf,Trend}_{r,DAILY,t} < 0.4 + X^{bulf,Trend}_{r,meat,t} / carcassSh_{maact} \cdot 0.0016 \tag{4.19}$$

While all information for the requirement functions of CAPRI is projected consistently, they are not active in their detailed form in CAPTRD due to the complexity of the respective calculations. Instead these requirement functions are included in a simplified form as part of the balances for energy and protein requirements (*REQS_*) for each animal type *maact*:

$$\sum_{feed} X^{maact,Trend}_{r,feed,t} X^{Cont,Trend}_{r,feed,t} = 0.998^t (a^{Const}_{maact} + a^{Slope}_{maact} X^{maact,Trend}_{r,yield,t}) \tag{4.20}$$

where *Cont* are the contents in terms of energy and crude protein. The left hand side of the equation defines total delivery of energy or protein from the current feeding practise per animal activity in region r, whereas the right hand side the need per animal derived from requirement functions depending on the main output (meat, milk, eggs, piglets born). The parameters a and b of the requirement functions are estimated from engineering functions as implemented in the CAPRI modelling system, and scaled so that the balance holds for the base year. The factor in front of the requirements introduces some input saving technical progress of -0.2% per annum.

The feeding coefficients multiplied with the herd sizes define total feed use for the different feeding stuffs 'bulks' (cereals, protein rich, energy rich, dairy based, other) and single nontradable feed items (grass, maize silage, fodder root crops, straw, milk for feeding, other fodder from arable land), technically in the same (*GROF_*) equation as equation below:

$$X^{GROF,Trend}_{r,feed,t} = \sum_{maact} X^{maact,Trend}_{r,feed,t} X^{maact,Trend}_{r,levl,t} \tag{4.21}$$

Feed use of individual products must add up to the feed use of the 'bulks' mentioned above (in *FEED_*):

$$X^{FEDM,Trend}_{r,feed,t} = \sum_{o \rightarrow feed} X^{FEDM,Trend}_{r,o,t} \tag{4.22}$$

Additional equations impose that certain stable relationships of agricultural technology are also maintained in projections:

- Equation *EFED_* ensures that feed use of non-tradable fodder items must be equal to production after accounting for losses.
- Other equations (*PosLo_*, *PosUp_*) force the relation of seed use or losses to production (plus imports for losses) into a +-20% range around the base year value.
- Production has to exceed seed use and losses (*SEED_*)

- The ratio of straw to cereal yields is maintained at base year values (*STRA_*)
- Livestock units per hectare are calculated (*LU_*) and may thus be subject to constraints (limiting their deviations from the supports, for example).

Finally there is an Equation (*LABO_*) ensuring that projections of family (*LABH*) and hired labour (*LABN*) in agriculture add up to total labour (*LABO*):

$$X_{r,LABO,t}^{GROF,Trend} = X_{r,LABH,t}^{GROF,Trend} X_{r,LABN,t}^{GROF,Trend} \tag{4.23}$$

In the first place projections of family and hired labour follow from input coefficients combined with the activity levels, but the previous equation permits to apply bounds to the total.

Constraints relating to prices, production values and revenues The check of external forecasts revealed that for some products, external price projections are not available. It was decided to include prices, value and revenues per activity in the constrained estimation process. The first Equation (*EAAG_*) defines the value (*EAAG*, position from the Economic Accounts for Agriculture) of each product and product group as the product of production (*GROF*) times the unit value prices (*UVAG*):

$$X_{r,i,t}^{EAAG,Trend} = X_{r,i,t}^{GROF,Trend} X_{r,i,t}^{UVAG,Trend} \tag{4.24}$$

The revenues of the activities (*TOOU*, total output) for each activity and group of activities *acts* are defined in Equation *REVE_* as:

$$X_{r,TOOU,t}^{acts,Trend} = \sum_{o} X_{r,o,t}^{acts,Trend} X_{r,o,t}^{UVAG,Trend} \tag{4.25}$$

Consumer prices (*UVAD*) are equal to producer prices (*UVAG*) plus a margin (*CSSP*) according to Equation *UVAD_*: [4] FIXME (fußnote61)

$$X_{r,i,t}^{UVAD,Trend} = X_{r,i,t}^{UVAG,Trend} X_{r,i,t}^{CMRG,Trend} \tag{4.26}$$

Constraints relating to consumer behaviour Human consumption (*HCOM*) is defined as per head consumption multiplied with population (*HCOM_*):

$$X_{r,i,t}^{HCOM,Trend} = X_{r,i,t}^{INHA,Trend} X_{r,LEVL,t}^{INHA,Trend} \tag{4.27}$$

Consumer expenditures per caput (*EXPE*) are equal (via *EXPE_*) to human consumption per caput (*INHA*) times consumer prices (*UVAD*):

$$X_{r,i,t}^{EXPE,Trend} = X_{r,i,t}^{INHA,Trend} X_{r,LEVL,t}^{UVAD,Trend} \tag{4.28}$$

[4] The symbol CSSP (initially for "consumer surplus") is usually used for the welfare effects related to final consumers (currently expressed as equivalent variation). Consumer margins are stored on CMRG in the market model. This misuse of code CSSP in CAPTRD is due to historical reasons.

Total per caput expenditure (*EXPE.LEVL*) must add up (in Equation *EXPETOT_*):

$$X_{r,LEVL,t}^{EXPE,Trend} = \sum_{i} X_{r,i,t}^{ESPE,Trend} \tag{4.29}$$

Constraints relating to processed products Marketable production (*MAPR*) of secondary products (*sec*) - cakes and oils from oilseeds, molasses and sugar, rice and starch - is linked in Equation *MAPR_* to processing of primary products (*PRCM*) by processing yields (*PRCY*):

$$X_{r,sec,t}^{MAPR,Trend} = \sum_{i \wedge sec \leftarrow i} X_{r,i,t}^{PRCM,Trend} X_{r,sec,t}^{PRCY,Trend} \tag{4.30}$$

In case of products from derived milk (*mlkseco*) – butter, skimmed milk powder, cheese, fresh milk products, cream, concentrated milk, whole milk powder whey powder, and casein – eq. *MLKCNT_* requires that fat and protein content (*MLKCNT*) of the processed raw milk (MILK[5]) be equal to the content of the derived products, after acknowlegding that small quantities of dairy products are themselves transformed to other dairy products (most relevant for processed cheese):

$$X_{r,MILK,t}^{PRCM,Trend} X_{r,MILK,t}^{MLKCNT,Trend} = \sum_{mlk\,sec\,o} \left(X_{r,mlk\,sec\,o,t}^{MAPR,Trend} X_{r,mlk\,sec\,o,t}^{PRCM,Trend} \right) X_{r,mlk\,sec\,o,t}^{MLKCNT,Trend} \tag{4.31}$$

Marketable production of by-products from the brewery, milling and sugar industry (set *RESIMP* = { *FENI*, *FPRI*}) are derived from corresponding uses of related products (cereals and sugar, Equation *MaprByFeed_*):

$$\begin{aligned} X_{r,resimp,t}^{MAPR,Trend} = & \sum_{o \rightarrow resimp} \left(X_{r,o,t}^{HCOM,Trend} + X_{r,o,t}^{PRCM,Trend} + X_{r,o,t}^{INDM,Trend} + X_{r,o,t}^{BIOF,Trend} \right) \\ & \cdot \frac{X_{r,resimp,t}^{MAPR,bas}}{\sum_{o \rightarrow resimp} \left(X_{r,o,t}^{HCOM,bas} + X_{r,o,t}^{PRCM,bas} + X_{r,o,t}^{INDM,bas} + X_{r,o,t}^{BIOF,bas} \right)} \end{aligned} \tag{4.32}$$

Constraints relating to bio-fuel production Marketable production (*MAPR*) of biofuels (*seco_biof*) derives (according to Equation *BIOF_*) from non agricultural production NAGR (e.g biodiesel from waste oil), from second generation production *SECG* , or through processing yields in terms of biofuels [6] (*PRCB*) from biofuel use of first generation feedstocks (*BIOF*):

[5]This is somewhat indirectly related to processing of cow milk and sheep & goat milk over *MAPR.MILK* = *PRCM.COMI* + *PRCM.SHGM* with a processing yield *PRCY.MILK* = 1 and over the market balance for product *MILK* which ensures that, with minimal trade of raw *MILK*, most of *MAPR.MILK* will end up as *PRCM.MILK*.

[6]Note that the processing yields PRCY (say X tons of rape oil per ton of rape) are associated with the outputs, because there is just one possible input for the given output (say PRCY.RAPO = yield of rape in terms of rape oil). But in the case of bio-ethanol, for example, there are several feedstocks (wheat, barley etc) producing one output (ethanol). Hence the output coefficients PRCB are associated with the inputs (say PRCB.BARL = yield of barley in terms of ethanol) and we need different types of coefficients.

$$X^{MAPR,Trend}_{r,seco\_biof,t} = \sum_{stocks \rightarrow seco\_biof} X^{BIOF,Trend}_{r,stocks,t} X^{PRCB,Trend}_{r,stocks,t} \tag{4.33}$$

In case of ethanol there is another by-product, *DDGS*, which is usable as a feedstuff and produced according to by-product coefficients from cereals (*DDGS_*):

$$X^{MAPR,Trend}_{r,DDGS,t} = \sum_{stocks \rightarrow DDGS} X^{BIOF,Trend}_{r,stocks,t} X^{PRCBY,Trend}_{r,stocks,t} \tag{4.34}$$

Constraints relating to policy There are only a few constraints directly taken from an EU regulation: firstly, the acreage under compulsory set-aside (abolished in the CAP Health Check of 2008) must be equal to the set-aside obligations of the individual crops (*OSET_*):

$$X^{"OSET",Trend}_{r,"levl",t} = \sum_{cact} X^{cact,Trend}_{r,"levl",t} \frac{0.01 X^{cact,Trend}_{r,"setr",t}}{\left(1 - 0.01 X^{cact,Trend}_{r,"setr",t}\right)} \tag{4.35}$$

Secondly, we have the quota products milk and sugar. The milk quotas on deliveries are acknowledged with a fixing on processing of cow milk without an explicit equation, taking into account that there are countries with persistent under- or over-deliveries. Given the expiry of milk quotas after 2015 this is largely irrelevant for current applications of CAPTRD. The sugar quotas, by contrast, are included as an upper bound (*SugaQuot_*) that may be relaxed (see Regulation 318/2006, Article 12) through industrial or biofuel use of sugar (and losses of sugar):

$$X^{MAPR,Trend}_{r,SUGA,t} \leq X^{QUTS,Trend}_{r,SUGA,t} + X^{INDM,Trend}_{r,SUGA,t} + X^{BIOF,Trend}_{r,SUGA,t} + X^{LOSM,Trend}_{r,SUGA,t} \tag{4.36}$$

Finally, there are upper bounds on new plantings of vineyards according to the CMO for wine from Regulation 1493/1999

Constraints relating to growth rates During estimation, a number of safeguards regarding the size of the implicit growth rates had been introduced in the course of various past CAPRI projects (bounds mainly found in *'captrd/fix_est.gms'*):

- In general, input or output coefficients (yields) are not allowed to change by more than +/- 2.5 % per annum, with a higher ranges for feed input coefficients (+/- 10 % and +/ 5 % for non-marketable fodder).
- The number of calves born per cow is may only change up to +/- 10 % around the base period value until the last projection year.
- The number of young cows (or sows) needed for replacement may only change up to +/ 20 % around the base period value until the last projection year.
- Final fattening weights must fall into a corridor of +/- 20% around the base period value.

- Milk yields are assumed to increase at least by 0.25% and at most by 1.25% near the EU average with some correction for below or above average initial yields (in *'captrd/comibounds.gms'*).
- Crop yields (except those of very hererogeneous crops like "other fruits" or "other fodder on arable land) should have a minimum yield growth of 0.5%.
- Specific (and quite generous) upper limits are applied to prevent unrealistic crop yields (for example: 15 tons/ha for cereals)
- Technical coefficients like contents of milk products or processing yields are also subject to plausible bounds.
- Strong increases in pork and poultry production in the past are restricted by environmental legislation in force, notably the nitrate directive. Accordingly, yearly increases were restricted to +1% for pork in EU15 Member States (even more stringent for Denmark and The Netherlands) and to 1.5% for poultry. In the new MS these maximum growth rates are assumed to be half a percentage point larger, in line with a weaker implementation of environmental legislation. The same bounds are also applied to the corresponding activity levels.
- A strong decrease of animal activity levels (below 20% of the base year) is not allowed.
- Total agricultural area is not allowed to decline at a rate exceeding -0.2 % per annum.
- Shares of arable crop on total arable area are bounded by a formula which allows small shares to expand or shrink more compared to crops with a high share. A crop with a base year share of 0.1% is allowed to expand to 2.5%, one of 10% only to 25%, and one of 50% to only 70%:

$$X^{arab,Trend}_{r,"levl",t}.up/lo = X^{arab,Trend}_{r,"levl",bas} \pm 1/4 \left(\frac{X^{arab,Trend}_{r,"levl",bas}}{X^{"arab",Trend}_{r,"levl",bas}} \right)^{1/4} X^{"arab",Trend}_{r,"levl",bas}\ max\left(0.2, \frac{t-bas}{last-bas}\right) \tag{4.37}$$

- However, in line with cross-compliance constraints from the CAP, permanent grass land must not decrease by more than 10% compared to the base year.
- An upper bound of 1% applies to the yearly growth of the area of "other oils" (for unclear reasons)
- Total labour must not deviate by more than 5% from forecasts based on coefficients estimated in an earlier study ("CAPRI-DYNASPAT").
- Changes in human consumption per caput for each of the products cannot exceed a growth rate of +/- 2% per annum. Due to some strong and rather implausible trends for total meat and total cereals consumption, the growth rate was restricted to +/- 0.8 % per annum for meat and +/- 0.4% per annum for cereals assuming that trend shifts between single items are more likely than strong trends in aggregate food groups.
- A downward sloping corridor is defined for subsistence consumption of raw milk (in 'captrd/comibounds.gms').
- Changes in prices are not allowed to exceed a growth rate of +/- 2% per annum, usually.

- Expert supports for biofuel related variables are given high priority with mostly tight corridors around these supports (in *'captrd/biobounds.gms'*).
- If a variable has dropped to zero according to recent COCO data it will be fixed to zero.

4.3.3 Step 2.2: Integration of specific expert support (Member State level or lower)

The definition of expert "supports" allows for provision of a mean and a standard deviation for all elements, and it is particularly useful for items for which the AgLink forcasts in step 3 are missing, or where there are other reasons for stability problems, such as missing historical data or very short time series

The expert supports are dealt with in *'captrd/expert_support.gms'*. Currently, mainly three sources can be distinguished:

- Support for the development of the sugar and sugar beet sectors, evolved from a small study with the seed production company KWS
- Expert on the development of bio-fuel production (bio-ethanol, bio-diesel), and the input demand for the related feedstocks, mainly based on results from the PRIMES model
- Expert supports for some key time series impacting on GHG emission for some Member States provided by the EC4MACS projects

The standard deviation is expressed by a "trust level" between 1 and 10.

The following table presents selected results related to the EU27 biomass feedstock for bioenergy production from the PRIMES[7]) biomass component (also given for each MS):

[7]PRIMES is a modelling tool for the EU energy system projections and impact assessment of the respective policies (see `https://ec.europa.eu/clima/policies/strategies/analysis/models_en`

Table 22: Selected results related to the EU27 biomass feedstock for bioenergy production from the PRIMES biomass component

Unit: ktoe (unless specified otherwise)	2000	2005	2010
Domestic Production of Biomass Feedstock	69,087	87,595	101,303
Crops	1,228	5,419	12,500
- Wheat	0	601	2,462
- Sugarbeet	0	1,291	4,518
- Sunflower/Rapeseed	1,228	3,527	5,520
- Lign. Crops	0	0	0
Agricultural Residues	4,194	6,428	7,200
Waste	19,990	26,002	28,054
Net imports of Biomass Feedstock	239	1,598	4,289
Pure Vegetable Oil as feedstock for bioenergy production	239	1,598	4,289
Cultivated Land (Kha)	896	3,022	5,422
Starch crops	0	320	1,218
Oil crops	896	2,654	4,031
Sugar Crops	0	48	172
Lignocellulosic crops	0	0	0

Source: own compilation. Comments: SWHE in Product code column indicates soft wheat commodity. SWHE in Activity code indicates yield of soft wheat. The CAPRI model used for this example was calibrated to the projections of Aglink-Cosimo model.

The above information on the biomass production is NOT used as the immediate input for CAPRI for several reasons. Converting from ktoe to 1000 tons (using 0.37 ktoe/1000t for cereals, 0.05 ktoe/1000t for sugar beet, 0.52 ktoe/1000t for rape seed) gives the production *for the bio-fuel sector* which matches with the market position “BIOF” = processing to biofuels. For cereals we have indeed 6.7 million tons from PRIMES in 2010 and 7.0 million tons according to CAPRI. For oilseeds we have to convert the PRIMES information in terms of oilseeds into a quantity of vegetable oil, giving approximately 5.5 mtoe / 0.52 ktoe/1000t * 0.4 [rape oil/ rape seed] = 4.2 million tons which is considerably larger than the results from CAPRI[8] 1.8 million tons. A similar comparison for the sugar sector may point at conversion problems with the units. The PRIMES sugar beet production should correspond to a sugar quantity of 4.5 mtoe / 0.05 ktoe/1000t * 0.15 [sugar/sugar beet] = 13.5 million tons of sugar equivalents which is close to the *total* sugar production in CAPRI of 15.7 million tons. Apart from these unresolved differences in the ex post data the main reason for NOT using these biomass production quantities from PRIMES is conceptual: They are given from supply functions specific to the bio-fuel sector whereas CAPRI covers the whole production (mostly for food purposes) such that the use of exogenous information for parts of the total may create problems for the CAPRI market balances.

A similar consideration also applies to the area information from PRIMES which refers to the specific areas used for biofuel purposes, except for the area for lignocellulosic crops.

Basically, the information “close” to agriculture (feed stock use and required areas) has not been taken

[8]It appears that the CAPRI bio-fuel results of August 2011 are affected by reporting errors in the oilseeds and sugar sectors.

from PRIMES assuming that it is preferable to estimate those in the context of the agricultural sector model CAPRI. On the other hand, the information on the production of bioenergy, including its main technologies and pathways, was supposed to be given reliably from the PRIMES biomass component exactly because it covers beyond agriculture also forestry and various forms of waste. The next table focuses on those results that will be used as the immediate inputs for CAPRI (thus omitting bio-energy from forestry, for example).

First of all PRIMES offers net imports, production and demand quantities for the biofuels itself. Production of biodiesel is split up according to the technology in first generation and second generation technologies (FT diesel, HTU diesel, pyrolysis diesel). For ethanol such a breakdown is not given in terms of production volumes, but the PRIMES output includes among the installed capacities also those for fermentation of sugar crops, starchy crops and lignocellulosic crops, the latter identifying the share for second generation production of ethanol. The input for first generation production of biodiesel (through esterification) is “bioheavy” which includes pure vegetable oil from domestic production, but also from various forms of waste oil (recovered oils, biocrude, pyrolysis oil). In addition the market balance for bioheavy includes imports (pure vegetable oil, the larger part according to the previous table for biodiesel production, a smaller part for direct use as fuel) and demand quantities of bioheavy. These are the key inputs for CAPRI, plus the area of lignocellulosic crops that is also a direct input to CAPRI.

In addition, there is more information that may be used in the future. Biogas production is mainly based on sewage systems but in part it also relies on animal manure (whereas the German particularity of biogas from green maize is not yet included). Biogas production from manure might be coordinated between PRIMES and CAPRI in the future. Equally the PRIMES assumptions on the amount of crop residues usable for bio-energy are not yet cross-checked with CAPRI. Finally, it should be mentioned that the use of waste in the PRIMES tables refers to other sources of bioenergy (like municipal waste).

Table 23: Results on biofules of PRIMES model

Unit: ktoe (unless specified otherwise)	2000	2005	2010
Net imports of Bioenergy	400	1,731	5,820
Biodiesel	0	0	1,948
Bioethanol	0	20	1,130
Pure Vegetable Oil	8	390	505
Bioenergy Production	67,971	84,554	95,430
Biodiesel	610	2,548	6,578
- Biodiesel (1st gen.)	610	2,548	6,578
- FT diesel	0	0	0
- HTU diesel	0	0	0
- Pyrolysis diesel	0	0	0
Bioethanol	0	561	2,193
BioHeavy	1	83	605
- Recovered Oils	0	43	589
- Pure Vegetable Oil	1	40	15
- BioCrude	0	0	0
- Pyrolysis oil	0	0	0
BioGas	352	871	2,049
- Bio-gas	352	871	2,049
- Synthetic Natural Gas	0	0	0
Waste Solid	12,353	13,985	14,654
Waste Gas	1,898	3,537	4,538
Demand	68,372	86,285	101,250
Biodiesel	610	2,548	8,526
Bioethanol	0	581	3,234
BioKerosene	0	0	0
BioHydrogen	0	0	0
BioHeavy	9	473	1,110
BioGas	352	871	2,049
Waste Solid	12,353	13,985	14,654
Waste Gas	1,898	3,537	4,538
Capacities (Ktoe/yr)	10,440	16,067	26,754
Fermentation	134	1,127	4,104
- Sugar	0	551	2,103
- Starch	134	576	2,001
- Lignocellulosic	0	0	0
Esterification	1,141	4,170	9,021

In technical terms the PRIMES results are given as a set of Excel tables that is usually amended with each release in some detail. To extract these data a small GAMS program (*‘merge.gms’*) prepares strings that, when saved and reload with Excel, are interpreted as external links to the PRIMES files using the “Vlookup” function of Excel. The relevant data are written to a parameter p_PRIMESresults, including the following:

P_PRIMESresults(MS,BIOEshare,SECG,year)
= capacity, lignocellulosic / capacity fermentation

Otherwise the selection addresses directly certain lines of the PRIMES output.

4.3.4 Step 3: Adding comprehensive sets of supports from AGLINK or other agencies

In Step 3, results from external projections on market balance positions (production, consumption, net trade etc.) and on activity levels for EU aggregates (EU15, EU12) are added. Currently, these projections are provided by Aglink-COSIMO model projections. The baseline of Aglink-COSIMO integrates the market outlook results from DG-AGRI, but is also globally harmonised, so that it also enters the baseline generation for the market model of CAPRI.

Integration of results from another modelling system is a challenging exercise as neither data nor definitions of products and market balance positions are fully harmonized. That holds especially for Aglink-COSIMO, where at least in the past the mnemonics had even not been harmonized across equations of the model itself. After a restructuring exercise in 2010, that had somewhat been improved. The ingredients in the mapping process are first a list of the codes for the regions, products and items used in Aglink-COSIMO (‘*baseline/aglink*_sets.gms*’, where * can be 2009 or 2010 to differentiate the versions before and after the restructuring). A second program, (*‘baseline/aglink*_mappings.gms*’) links the CAPRI regions, products and items to the mnemonics and Aglink-COSIMO, and a larger program (‘*baseline/loag_aglink*.gms*’) then uses the mapping to assign them to the CAPRI code world.

Aglink-COSIMO currently features results at EU15 and EU12 level. It is hence not possible to funnel the Aglink-COSIMO results into Step 2 above without an assumption of the share of the individual Member States.

As DG-AGRI is often the main client of the CAPRI projections for the EU, it was deemed sensible to pull the projections towards the DG-AGRI baseline wherever the constraints of the estimation problem and potentially conflicting other expert sources allow for it. That is achieved by two assignments related to the objective function:

1. Step 2 results (except those steered by other expert supports) are scaled proportionally to give MS level supports for step 3 that are consistent with the Aglink-COSIMO baseline (after adjusting for different definitions in the respective databases).
2. The standard errors from the default trends are replaced with a special formula reflecting a high confidence in the Aglink-COSIMO derived supports.

More precisely, the weighted variance is replaced with the following setting for external supports (*“XSupport”* = AGLINK or expert supports):

$$X_{r,i,"varErr"}^{j,"XSupport"} = \left(X_{r,i,"exante"}^{j,"XSupport"} \cdot 0.05/3 \cdot \left(10/X_{r,i,"trustlevl"}^{j,"XSupport"} \right) \right)^2 \tag{4.38}$$

The “trust level” in the last denominator is a scaling factor for the implied coefficient of variation. A higher trust level translates into a lower error variance of the external information. With a normal distribution we would have

- at "trust level" = 10: X [-0.055*Mean, +0.055*Mean] with probability 99.9%
- at "trust level" = 5: X [-0.275*Mean, +0.275*Mean] with probability 99.9%
- at "trust level" = 1: X [-0.55*Mean, +0.55*Mean] with probability 99.9%

The default setting for "DGAgri" supports is a "trust level" of 5, which is a moderately high value to leave some distance for special cases that should be pulled very tightly towards their supports.

The Aglink-COSIMO projections currently run to 2020 or a few years beyond. For climate related applications CAPRI has to tackle projections up to 2030 or even 2050. CAPRI projections up to 2030 have been prepared in the context of EC4MACS project (`http://www.ec4macs.eu`). The methodology was quite simple: The year 2020 projection (usually prepared in the same run of CAPTRD) has been extrapolated in a nonlinear dampened (logistic) fashion (in *'define_eu_supports.gms'*) with some additional bounds to prevent unreasonable increases of certain variables (nonnegativity already provided a good lower bound). Together with the information in the time series database this has been an ad hoc but operational procedure to address the 2030 horizon, but it would have been inappropriate for a move to the long run up to 2050 as required for a recent study on behalf of DG CLIMA[9].

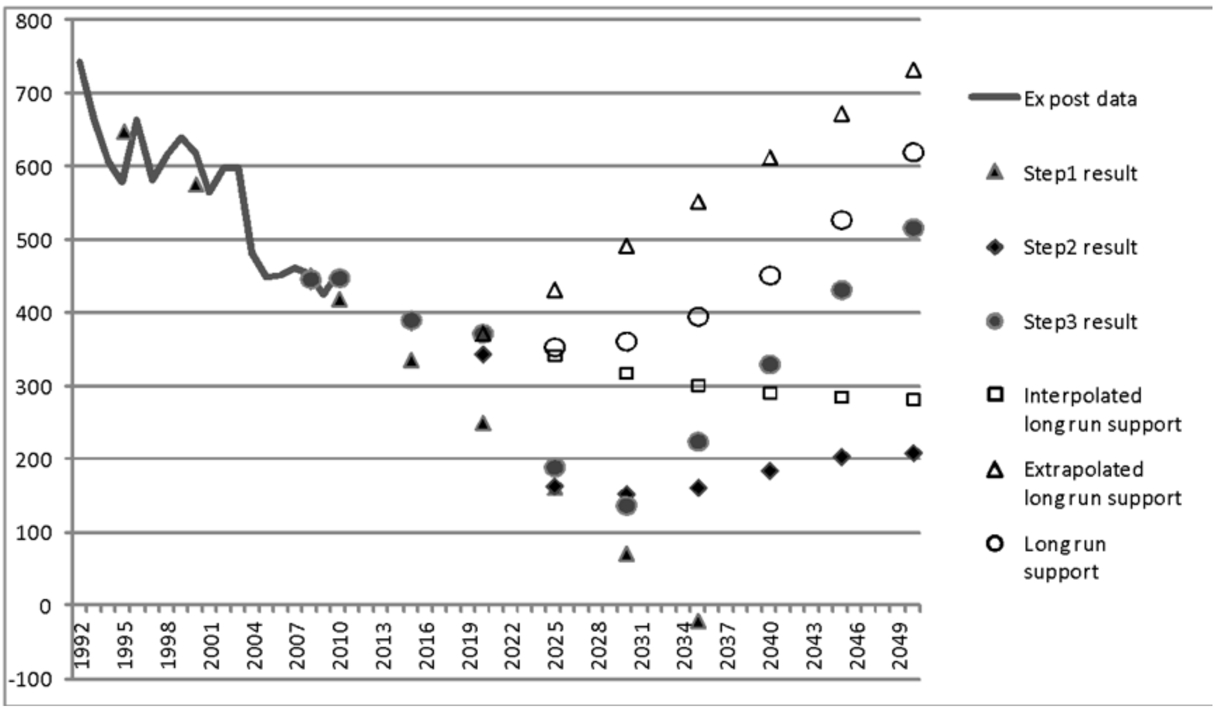
For the long run evolution of food production a link has been established to long run projections from two major agencies (FAO 2006 and the IMPACT projections in Rosegrant et al 2009, see also Rosegrant et al 2008). This linkage required mappings to bridge differences in definitions (see *'gams/global/f2050_impact.gms'* called when running *'gams/global.gms'*).

Furthermore, methodology was needed to avoid a break in the projections at the transition of medium run expert information (Aglink-COSIMO, up to 2020) and long run information (FAO/IFPRI for 2050). For this purpose a variable weighting scheme is introduced (in *'gams/captrd/expert_support.gms'*) that gives an increasing weight to our "long run" sources (FAO/IFPRI) as the projection horizon approaches 2050. This tends to give projections that gradually approach the long run sources, for example as in the case of pork production in Hungary (taken from a baseline established in November 2011).

The example has been chosen because historical trends (and Aglink-COSIMO projections) on the one hand and long run expectations differ markedly. This is not unusual because medium run forecasts often give a stronger weight to recent production trends, often indicating a stagnating or declining production in the EU, whereas the long run studies tend to focus on the global growth of food demand in the coming decades. The simple trends (filled triangles) would evidently give unreasonable, even negative forecasts after 2030. Already the imposition of constraints from relationships to other series would stabilise the projections and imply some recovery after 2030 (filled squares). The year 2020 supports from Aglink-COSIMO (not shown) produces some upward correction of the step 2 results for 2020, giving a final projection (filled circles) of about 375 ktons for pork production in Hungary. This is also the starting point for the specification of the long run support (empty circles) which is a weighted average of two components. The first is a linear interpolation to the external projection from FAO/IFPRI for 2050 (empty triangles). The second is a nonlinear damped extrapolation of the medium run projection beyond 2020 (empty squares). Changing the weight for the first component (FAO/IFPRI support) with increasing projection horizon creates a long run target value (empty circles) that gives a smooth transition from the

[9] Service contract on "Model based assessment of EU energy and climate change policies for post-2012 regime" (Tender DG ENV.C.5/SER/2009/0036), coordinated by the Energy-Economy-Environment Modelling Laboratory (E3MLab), National Technical University of Athens with the International Institute for Applied Systems Analysis (IIASA) and EuroCARE as subcontractors.

Figure 11: Pork production in Hungary as an example for merging medium run and long run a priori information in the CAPRI baseline approach



Source: own elaboration

medium to the long run. As the final projections (filled circles) tend to follow these target values, they show a turning point in the future evolution of pork production in Hungary that ultimately reflects the consideration of increasing global demand underlying the FAO/IFPRI projections.

Evidently this approach is quite removed from economic modelling and it is not intended to be. Instead it tries to synthesize the existing projections from various agencies, each specialised in particular fields and time horizons, in a technically consistent and plausible manner. The specification of a constraint set and penalties of the objective function translates plausibility in an operational form. Technical consistency is imposed through the system of constraints active during the estimation.

4.3.5 Step 4: Breaking down results from Member State to regional and farm type level

Even if it would be preferable to add the regional dimension already during the estimation of the variables discussed above, the dimensionality of the problem renders such an approach infeasible. Instead, the step 3 projection results regarding activity levels and production quantities are taken as fixed and given, and are distributed to the regions minimizing deviation from regional supports. The aggregation conditions for this step (and correspondingly for the disaggregation of NUTS2 regions to farm types) are:

- Adding up of regional production to Member State production (*MSGROF_*)
- Adding up of regional agricultural and non-agricultural areas to Member State areas (eqs. *MSLEVL_* and *MSLANDUSE_*)
- Adding up of regional feed use by animal types to Member State values (*MSFEEDI_*).

The results at Member State level are thus broken down to regional level, ensuring adding up of production, areas and feed use:

$$X_{MS,i,t}^{GROF,Trend} = \sum_{r \in MS} X_{r,i,t}^{GROF,Trend} \tag{4.39}$$

$$X_{MS,"levl",t}^{j,Trend} = \sum_{r \in MS} X_{r,"levl",t}^{j,Trend} \tag{4.40}$$

$$X_{MS,"levl",t}^{j,Trend} \cdot \left(X_{MS,"feed",t}^{j,Trend} + 10 \right) = \sum_{r \in MS} X_{r,"levl",t}^{j,Trend} \cdot \left(X_{r,"feed",t}^{j,Trend} + 10 \right) \tag{4.41}$$

The addition of the “10” (kg/animal) considerably improves the scaling in case of very small quantities (say 1 gram per animal). This is an example of a technical detail that may be crucial for numerical stability but usually cannot be reported fully in this documentation.

In addition to the above aggregation conditions, the lower level (NUTS2 or farm type) models only require the following constraints (as the market variables are already determined at the MS level):

- Related to areas: area balance (Equation 57 FIXME), obligatory set aside (Equation 80 FIXME), aggregation to groups like cereals (0).
- Related to yields: linkage of production, activity levels and yields (Equation 55 FIXME), stabilisation of straw yields (*STRA_*)
- Related to animals: Nutrient balances (Equation 65 FIXME), local use of fodder (*EFED_*), definition of livestock density (*LU_*).

In order to keep developments at regional and national level comparable, relative changes in activity levels are not allowed to deviate very far from the national development. These bounds are widened in cases of infeasibilities.

Table below contains an example of the final output of the trends estimation task (C:/....CAPRI/STAR/star_2.4/output/results/baseline/results_BBYY.gdx), where BB stands for base year and YY for simulation year). Its main purpose is to provide with explanations on the variables of this output and, thus, a possibility to review the results in a step-by-step manner.

4.4 Calibrating the global trade model

After the Task on Trends generation have been successfully completed, meaning that the projections for the defined (in GUI or a batch file) future years (currently, 2015, 2020, 2025 and 2030 are available) have been produced, the next step in the Baseline generation process (“Generate baseline” workstep in CAPRI

GUI) is to calibrate the CAPRI global trade model. In the CAPRI GUI this refers to the task “Baseline calibration of market model”.

The calibration of the market model is steered by the C:/.../CAPRI/gams/capmod.gms file. The relevant parts of the code are activated by setting the setglobal ’BASELINE’ to ON.

4.4.1 Stage I: Data preparation and balancing

The CAPRI database is composed of many different data sources, and requires data processing before the market model equations can be calibrated against the data set. Sources of potential problems include missing data and price-quantity framework that is inconsistent with the behavioural assumptions (e.g. profit maximizing producers, utility maximizing consumers).

Stage I of the market model calibration makes the CAPRI database consistent, and creates a dataset for the global agri-food markets against which the market model can be calibrated. As CAPRI is a comparative static model, the market model is calibrated only against the simulation year. But technically the CAPRI dataset is first made consistent to the model structure in the base year, and then shifted to the simulation year. More specifically the main steps in this stage include:

1. Prepare the necessary data by
   (a) loading them from various intermediate data files;
   (b) mapping them to correct code lists;
   (c) adjusting if necessary, often by applying security bounds;
2. Ensure the consistency of the dataset to the market model structure for the base year (BAS)
3. Shift the consistent dataset from the base year to the simulation year
4. Ensure the consistency of the dataset to the market model structure for the simulation year (SIMY)

Data preparation Before actually performing the calibration of the market model parameters, CAPRI first loads the necessary sets, parameters and data. These refer to periods (years), regions, activities, commodities, agricultural policies (e.g., premiums, quotas, rural development payments, set-aside requirements), environmental indicators, feed and fertilizer requirements, nutrient content of the commodities, global warming potentials, and other necessary input. The data loaded also includes two very important for this calibration step files: C:/.../CAPRI/output/results/capreg/res_BBCC.gdx and C:/.../CAPRI/output/results/baseline/trends_BBYY.gdx. The first file, res_BBCC.gdx, includes the results of generation of data for the base year (BB, currently 2012) for European countries and Turkey (CC) at NUTS0, NUTS1 and NUTS2 aggregation levels (GUI workstep “Build database”, task “Build regional database”). The second file, trends_BBYY.gdx, includes the results of trends generation task (see sections above) for all of the European countries and Turkey at NUTS0, NUTS1 and NUTS2 aggregation levels for the target simulation year (currently, 2030).

Constraints, requirements, policies and other data loaded including base year and trends data (i.e., of res_BBCC.gdx and trends_BBYY.gdx files) are subject to certain (mainly non-major) adjustments, additional calculations and assumptions that serve the purposes of data balancing, checking and provision of necessary for the calibration information. These include, for example, deleting positions not needed during the calibration run, (re-)assigning parameter names, deleting tiny quantities, checks for production without activity levels, possible empty projections and negative inland waters, setting the output coefficients for young animals equal to the ones at country (EU MSs, as young animals are not presented in the non-EU countries) level if missing at regional level, correcting fat and protein content of raw milk, assumption that second generation biofuels are produced 50/50 by agricultural residuals and new energy crops, etc.

Next, FAO data on the non-European countries as well as the trade flows among all of the countries (country trade blocks) accounted for in CAPRI are loaded. These FAO data together with the European data, which has already been subjected to certain adjustments as described in the previous paragraph, undergo the, so-called, data preparation step. This process is controlled by C:/.../CAPRI/gams/arm/-market1.gms file which calls the C:/.../CAPRI/gams/arm/data_prep.gms file - specifically for this step. The data preparation step mostly refers to the base year and includes: among else, modification of GDP to fit the sum of final household expenditure, final government expenditure, gross capital formation and current account balance; import and export flows to be in line with net trade from production minus demand; scaling of demand side to fit production plus net trade; estimation of consumer prices for some countries, if missing; calculation of nutrient consumption per head and day as net of losses in distribution and households; scaling of outliers in prices etc. This step as well provides with estimation of yearly change factors beyond the base year: for prices, GDP, population, quantities and areas. Additionally, i) substitution elasticities (i.e., p_rhoX, where X indicates continuation of the parameter name) for bio-fuel feedstocks, feed, dairy products, sugar, table grapes, tobacco, cheese, fresh milk products, fruits, vegetables, distilled dried grains and rice for the CAPRI demand system[10], and ii) transformation elasticity for oil seed processing and land supply elasticities are assigned.

Together with the data, equations of the CAPRI market module are loaded. They are described in detail in section 5.4. These equations include behavioural functions for market demand including expenditure function, feed demand, blocks for dairy products, oilseeds processing and biofuels, netput functions, trade equations and balances, equations for prices and price transmission, functions for trade policies and for intervention stocks. There are additionally two crucial for data calibration functions: minimization of deviation of estimated values from the observed data. These two functions are described in detail later in this section.

Data balancing After data preparation, data calibration for the base (currently, 2012) and simulation years (currently, 2030) take place. The main file steering the data balancing process is C:/.../CAPRI/gams/arm/data_cal.gms, which in turn is included in arm/market1.gms.

Data balancing for the base year

Data calibration for the base year aims at modifying the base year data to fit the system of equations of the market module. Some of the parameters defined in Stage I (e.g., p_rhoX) as well as parameter values and bounds defined at this stage are used. For example, starting points and corridors for quantity variables are set (e.g., calculating of world production to define correction corridor for calibration of production/demand/trade flows globally), global TRQ data are converted into ad valorem tariffs and

[10]See section 5.4.1 on overview of the market model "capri comprises a two stage armington system: on the top level, the composition of total demand from imports and domestic sales is determined, as a function of the relation between the domestic price and the average import price. the lower stage determines the import shares from different origins and defines the average import price."

checked for consistency and completeness, policy variables for the EU market model such as e.g., intervention stocks, are loaded. Also, starting values for prices of dairy products are estimated. In particular, a non-linear programming model is used, where the objective function is formulated as a Highest Posterior Density function. The value of this objective function equals sum of squared deviations of fat and protein prices, fat and protein content of milk products and processing margins of milk products from the respective means, weighted with the a priori variances. The means are defined as parameters based on the prices and fat and protein content of milk in the base year. The objective function is restricted by the balance: fat and protein of raw milk delivered to dairies shall equal fat and protein content of dairy products. The model is solved by minimizing the value of highest posterior density, hence minimizing the differences between the variables and their means. Prices of milk products are then defined as: product of fat and protein content and of far and protein prices plus processing margin. Furthermore, administrative prices for cereals and dairy products, and minimal import prices for cereals are constructed.

With the file C:/.../CAPRI/gams/arm/cal_models.gms, the so-called, models, used in calibration of data base are defined and solved. These models represent collection of equations, solutions of which provide with parameter values used for data calibration. The first model (MODEL m_trimSubsExports) calibrates the parameters of the function which defines the values of subsidized exports with and without the increase of market price above the administrative price. The second model (MODEL m_trimInterv) defines parameters of equations for intervention stock changes. It includes an objective function defined as a sum of: squared scaled difference of estimated and observed intervention stock changes and squared scaled parameters for behavioural function of intervention stock changes. This objective function is minimized subject to constraints represented by equations for intervention sales, probability for an undercut of administrative price, release from intervention stock, intervention stock changes and value of the intervention stock. The constraints are equations of the market model (see section 5.4).

The model that calibrates base year data (MODEL m_calMarketBas) is defined in cal_models.gms file as well and includes almost all equations of the market model. In particular: equations for processing margin for dairy products (ProcMargM_), fat and protein balance between raw milk and dairy products (FatsProtBal_), processing margin for oilseeds ProcMargO_, processing yields of oilseeds (procYield_), 1st generation output of biofuels (prodBiof_) and total output of biofuels (MaprBiof_); <u>balancing and adding up equations</u>: equations which add production, processing demand, human consumption, feed demand quantities and quantities for processing from single countries (or block of countries) to trade blocks (ProdA_, ProcA_, HconA_, FeedUseA_, Proca_), adding up inside of the Armginton aggregate (total domestic consumption) (ArmBal1_), supply balance (SupBalM_) and imports and exports added up to bilateral trade flows (excluding diagonal element) (impQuant_); <u>price equations</u>: 1st stage Armington quantity aggregate (ArmFit1_), 2nd stage Armington quantity aggregate (ArmFit2_), import price relation to producer price (impPrice_), consumer price as average of domestic and import prices (arm1Price_), average price as average of different import prices (arm2Price_), average import price (arm2Val_), consumer price (Cpri_), producer price (PPri_), market price (PMrk_), average market price (MarketPriceAgg_); <u>trade and tariff equations</u>: aggregated trade flows (TradeFlowsAgg_), average transportation costs (TransportCostsAgg_), sum of imports under a non-allocated TRQ (TRQImports_), share of the tariff applied for the EU entry price system (EntryPriceDriver_), tariff specific entry price (tarSpecIfEntryPrice_), Cif price (cifPrice_), equation for defining levy (replaces tariff) in case of minimal border prices (FlexLevyNotCut_), cuting flexible levy by specific tariff if it exceeds the bound rate (FlexLevy_), tariffs under bi-lateral TRQs (trqSigmoidFunc_), specific tariffs as function of import quantities, if TRQ is present (tarSpec_, prefTriggerPrice_), tariffs under globally open (not bilaterally allocated) TRQs (tarSpecW_), ad valorem tariffs, if TRQ is present (tarAdval_), ad valorem tariff under not bilaterally allocated TRQs (tarAdValW_), export quantities from bi-lateral trade flows (expQuant_), exports included in the calculation of the export unit values excluding flows under double-zero agreements (nonDoubleZeroExports_), unit value exports (unitValueExports_, valSubsExports_), subsidised export values (EXPs_); equations for intervention stocks: probability weight for an undercut of administrative price (probMarketPriceUnderSafetyNet_), intervention sales (buyingToIntervStock_), intervention stock end size (intervStockLevel_), intervention stock changes (intervStockChange_), release from intervention stocks (releaseFromIntervStock_), aggregators for intervention purchases; equation for world market price (wldPrice_), and equation for minimization of deviation from given base year data and estimated data (NSSQ_). The model is solved by minimizing the SSQ value of NSSQ equation which is constrained by all of the rest of the equations included in the model.

The NSSQ equation is crucial to the data calibration as it, in its essence, minimizes the difference between the estimated and the observed (already adjusted at the previous stage) data of the base year. Its logic is analogues to the one of equation below:

$$SSQ \cdot \sum_{RMS} \sum_{XXX} p\_weight^{i}_{RMS} = \sum_{RMS} \sum_{XXX} \left(\frac{v^{i}_{RMS,XXX} - DATA^{i}_{RSM,XXX,BAS}}{max(DATA^{i}_{RSM,XXX,BAS}, 0.1) \cdot p\_weight^{i}_{RMS}} \right)^{2} \tag{4.42}$$

where SSQ is an artificial variable to be minimized, indices RMS, XXX, BAS and i indicate, respectively, regions, commodities, base year and activities (e.g., production, processing, imports etc.), and p_weight is a parameter of weights between 1 and 100 assigned to regions and activities. These weights are necessary to achieve plausible calibrated values and their specification is the outcome of a trial and error process, inspecting results from data calibration and retrying. They depend on the results of global database and trends generation. On the right hand-side of the equation v stands for a variable to be estimated and DATA – for base year data already adjusted at the data preparation and balancing stage. Hence with this equation squared sum over regions and commodities of differences between estimated and observed values (and or quantities), these differences being scaled by the observed data times the weight parameters, is minimized. Respectively, calibrated base year data fits the system of the market equations, given certain parameter values, and resembles the observed data as closely as possible. The activities implied under the i index include quantities of production, human consumption, feed, processing, processed to biofuels, import and export, producer, consumer and market prices, difference between market prices and import prices to reduce differences between physical and Armington aggregation, consolidated gap between producer and market prices, processing margin, trade flows and transport costs.

The process of model solving is navigated with C:/.../CAPRI/gams/arm/data_fit.gms file. Its main function is to assure model solving by keeping the market balances closed and price system consistent. Because of the very large number of equations with the exact similar number of variables (36 thsds) that makes the system of equations square, as well as non-linear formulation of some of the equations, it is very likely that infeasibilities will occur during the model solving. To ensure the feasibility as far as possible, code elements such as widening of variable bounds, once they become binding, reducing non-smoothness of the functional forms and introduction of slack variables are introduced. More detailed information on this process can be found in a technical document by Wolfgang Britz and Heinz-Peter Witzke *Infeasibilities in the market model of CAPRI – how they are dealt with* at `https://www.capri-model.org/docs/infes.pdf`.

After solving the MODEL m_calMarketBas, the calibrated data are stored, new producer prices for agricultural outputs are set, sugar beet prices as a function of – sugar market price – sugar export price (pre-reform) or ethanol market price (post-reform) – processing yield (specific to CUR to calibrate to any set of projected beet prices) – levying model for A- and B- sugar (pre-reform) are calculated, share and shift parameters of CES-functions used in the Armington approach to determine import shares as a function of import prices are defined (file C:/.../CAPRI/gams/arm/cal_armington.gms). Furthermore, energy conversion factors for animal products are defined with MODEL m_fitFeedConv (in file C:/.../CAPRI/gams/arm/feed_conv_decl.gms).

Data balancing for the simulation year

Aim of data calibration for the simulation year aims at generating such quantity, price and other market values (see list below) for the simulation year that they fit the system of equations of the market module and variable and parameter lower and upper bounds, as well as remain as close as possible to the values to which they are calibrated (e.g., trends, estimated with growth rates from the base year, Aglink-COSIMO values, GLOBIOM values etc.). Thus process, basically, follows similar approach as for the base year. There are, however, a few differences. The main is that the model used for calibration is MODEL m_calMarketFin. As the model for base year calibration (MODEL m_calMarketBas), it is defined in cal_models.gms file and includes similar equations of the market model with the exception of NSSQ_ equation. The latter equation is replaced by NSSQ1_. Its major difference from NSSQ_ is that DATA parameter includes not values of the base year, but values projected in trend generation step for some of the factors and values shifted to the simulation year based on assumptions or growth rates for the other factors. Thus, it is used for minimizing the differences between estimated and projected (with trend generation step or growth rates) values of the variables in question. Another difference of NSSQ1_ with NSSQ_ is that it includes the differences in intervention stock changes and excludes the differences in consumer prices and gaps between producer and market prices.

Before MODEL m_calMarketFin is solved, values of DATA parameter for the simulation year are defined. For example, administrative prices for dairy products and cereals and minimum import prices for cereals (in C:/.../CAPRI/gams/arm/prep_pol.gms) and policy data are defined, market prices, quantity variables are shifted with growth rates (C:/.../CAPRI/gams/arm/shift_quantities.gms) and tariffs are defined. Bounds for tariff variables, market prices, milk fat and protein as well as upper and lower limits on quantity variables are assigned as well. At this point, models to calibrate TRQs and entry price equations (MODEL m_fitTrq) and parameters of equations for the intervention stock changes (MODEL m_trimInterv) are solved as well (now for the simulation year, as before it was solved for base year values).

As m_calMarketBas model, m_calMarketFin model is solved by minimizing SSQ value by applying the approach of assuring feasibility via data_fit.gms file. After the solution is found and energy conversion factors for animal products are defined with MODEL m_fitFeedConv, the results are stored in C:/.../CAPRI/ output/results/baseline/data_market_1230.gdx.

4.4.2 Stage II: Elasticity trimming

Elasticity trimming in CAPRI aims at adjusting prior estimations of elasticities so that

- the behavioural functions can be parameterized/calibrated to the given prices/quantity framework with the elasticities;

- the calibrated elasticities satisfy regulatory conditions (homogeneity, additivity) and correct curvature in line with microeconomic theory;
- the calibrated elasticities are as close as possible to prior elasticities (minimize deviation).

At first, parameters for land use market are calculated based on data from FAO world food market model. Among them are land use classes, crop yields, land demand of non-crop activities, areas used for fodder and average land price, total energy use for feeding and producer price of feed. Next, starting elasticity values, as well as their lower and upper bounds are loaded (e.g., demand elasticities used in SPEL/MFSS). Finally, elasticities are trimmed.

Elasticities trimming is controlled by C:/.../CAPRI/gams/arm/trim_par.gms file. The elasticity groups are: for calibration of demand and supply systems, feed demand system, oilseeds crush, oil processing and dairy industry. Elasticities of supply system, oilseeds crushing, oil processing and dairy industries, as well as for feed demand, are estimated with MODEL m_trimElas. It is solved by minimising absolute squares between given and calibrated elasticities including land elasticities (FitElas_) subject to the following constraints: marginal effects from price and quantity for current elasticity estimate (Hess_), homogeneity of degree zero for elasticities in prices (HomogN_), Cholesky decomposition of marginal effects to ensure correct curvature (Chol_), Ensure that own price elasticity exceeds (yield elasticity * 1.5) (YieldElas_) and elasticities for total energy and protein intake from feeding (ReqsElas_).

Human consumption elasticities are estimated with MODEL m_trimDem by minimizing absolute squares between given and calibrated elasticities (FitElas_). Apart from the objective function the model includes several equations related to the definition of the demand system as Generalized Leontief, homogeniety of degree zero for elasticities in prices, additivity of income elasticities weighted with budget shares and elasticities for total calorie intake.

4.4.3 Stage III: Feed and fertilizer calibration

In this stage, the feed system is calibrated against the primary product prices of the market model (both marketable and non-marketable feed). The nutrient requirements of the crops are calculated together with the nutrient and energy requirement of the animal production activities.

The fertilizer flows are also calibrated here. The prior parameters for the fertilizer flows are defined based on the *posterior* mode of the base year, by modifying them with land use changes: the fertilization per ha is computed in the base year situation and then multiplied with the areas in the calibration point. The fertilizer flows are calibrated with the same calibration model as used for the base year in the database tasks.

The file C:/.../CAPRI/gams/capmod/def_fert_and_requirements.gms defines animal nutrient requirements and the nutrient requirements of the crops given trend forecasted yields. In particular, feed input coefficients are defined and calibrated, days in production process of fattening are defined, and manure output is taken into consideration as an input for fertilizer calibration. Fertilizer calibration is basically a merge of trend based forecasts from the ex-post CAPREG results. The fertilizer need is calculated as a function of yield, and adjusted according to the exogenous assumptions. Furthermore, crop nutrient need factors from trends are scaled and logistic function is used to calculate average growth rate of fertilizer use. The calculations must as well comply with the fertilizer equations of the supply model.

4.4.4 Stage IV: Initialization and test run

After the behavioural blocks of the market model are calibrated (one-by-one), the whole model should be also tested for being correctly calibrated. In essence, the test initializes the model with the data against the model was calibrated, and then executes/solves the market model. In theory, a perfectly calibrated model can be solved in one single iteration, without adjustments in the values of the model variables. That is why the iteration limit is technically set to zero (i.e. not allowing for adjustment in the model variables) for the test solve. In practice, a number of infeasibilities might exist due to the accuracy of the numerical solution. But infeasibilities stemming from rounding errors must be small, so the sum of all infeasibilities gives a good indication on the quality of the model calibration.

At the final stage, some of the starting values and bounds for the market model are set, and agricultural policy data are loaded, adjusted and extended to the simulation year. The policy data include single area payment scheme, set-aside regulations, differentiation between old and new MSs payments, special national envelopes, Nordic schemes, changes in administrative prices, rural development policy and other major CAP post-2014 instruments. Policy files used for the baseline are located in C:/.../CAPRI/gams/s-cen/base_scenarios folder. Their loading into the baseline process is controlled by CAP_2014_2020.gms file. With the data mentioned, the outcome of calibration of the CAPRI market module can be tested. In particular, the market model is solved at “trend values” and, thus, the calibration outcome is checked for fitting to the square system of market model equations. This is controlled by C:/.../CAPRI/.gam-s/arm/prep_market.gms file.

4.4.5 Technical remarks

Note that the task “Baseline calibration of market model” deletes the sim_ini.gdx file, but does not create a new one at the end of the calibration process. The new sim_ini.gdx file will be only created at the first simulation run after the calibration. That is also the reason why a specific GUI option 'Kill simini file' is provided for the simulation tasks. The simini file can be deleted upon request at the beginning of any scenario run, forcing CAPRI to re-create it before the scenario shock is introduced.

Technically, the calibration of the biofuel demand system and the Armington bilateral trade system is not directly linked to the BASELINE mode, but also executed every time when the simini file is missing (by create_sim_ini_gdx module).

4.5 Calibrating the supply models to the CAPTRD projection

4.5.1 Introduction

The supply side models of the CAPRI simulation tool are programming models with an objective function. If we want the optimal solution to coincide with the forecast produced by the projection tools of CAPTRD, we need to ensure that first and second order optimality conditions (marginal revenues equal to marginal costs, all constraints feasible, and the solution is a maximum point) hold in the calibration point for each of the NUTS 2 or farm type models. The consequences regarding the calibration are threefold:

1. Elements not projected so far but entering the constraints of the supply models (e.g. feed, fertilization) must be defined in such way that constraints are feasible,

2. The cost function of the models must be shifted so that marginal costs and marginal revenues are equal in the calibration point.
3. The curvature of the functions must be such that the solution obtained is a maximum, not a minimum or a saddle point.

4.5.2 Calibrating feed and fertilizer restrictions

The calibration of feed and fertilization restrictions happens in the file *gams/capmod/def_fert_and_requirement.gms.* As explained above, the requirement functions used in the projection tools are linear approximations for the ones used in the simulation tool; additional constraints restrict the feed mix in the supply modules.

It is hence necessary to find a *feed mix* in the projected point which exhausts the projected production of non-tradable feed and the projected feed mix of marketable bulk feeds (cereals, protein feed, …), fits in the requirement constraints and leads to plausible feed cost. In order to do so, the feed allocation framework used to construct the base year allocation of feedstuff to animals is re-used. The resulting factors are stored in external files and reloaded by counterfactual runs.

Similar to animal feed balance, the crop nutrient needs must be consistent with available projected nutrients from various sources. To find such a feasible point, the distribution of various fertilizer sources (manure, mineral fertilizers and crop residues) to crops estimated in the database (CAPREG), is shifted with changes in crop areas to make a first best guess (prior) of the allocation to crops in the baseline. This prior is used as the modal value of a probability density function of a Bayesian estimation, similar to the CAPREG procedure described in a previous section of the documentation. Thus, a crop nutrient allocation is sought that is in some sense “as similar” to the base year estimate as possible. The result of the fertilizer calibration for the baseline is stored in a GDX file for each country, found in the directory “results/fert”, from where it is loaded in simulations (by the file *gams/capmod/load_fert_baseline.gms*).

4.5.3 Calibrating the marginal cost functions

Since the very first CAPRI version, ideas based on Positive Mathematical Programming were used to achieve perfect calibration to observed behaviour – namely regional statistics on cropping pattern, herds and yield – and data base results as the input or feed distribution. The basic idea is to interpret the ‘observed’ situation as a profit maximising choice of the agent, assuming that all constraints and coefficients are correctly specified with the exemption of costs or revenues not included in the model. Any difference between the marginal revenues and the marginal costs found at the base year situation is then mapped into a non-linear cost function, so that marginal revenues and costs are equal for all activities. In order to find the difference between marginal costs and revenues in the model without the non-linear cost function, calibration bounds around the choice variables are introduced.

The reader is now reminded that marginal costs in a programming model without non-linear terms comprise the accounting cost found in the objective and opportunity costs linked to binding resources. The opportunity costs in turn are a function of the accounting costs found in the objective. It is therefore not astonishing that a model where marginal revenues are not equal to marginal revenues at observed activity levels will most probably not produce reliable estimates of opportunity costs. The CAPRI team responded to that problem by defining exogenously the opportunity costs of two major restrictions: for the land balance and for milk quotas. The remaining shadow prices mostly relate to the feed block, and are less critical as they have a clear connection to prices of marketable feed as cereals which are not subject to the problems discussed above.

The development, test and validation of econometric approaches to estimate supply responses at the regional level in the context of regional programming models form an important task for the CAPRI team. Up to now, there is still no fully satisfactory solution of the problem, but some of the approaches are discussed in here.

The two possible competitors are standard duality based approaches with a following calibration step or estimates based directly on the Kuhn-Tucker conditions of the programming models. Both may or may not require a priori information to overcome missing degrees of freedom or reduce second or higher moments of estimated parameters. The duality based system estimation approach has the advantage to be well established. Less data are required for the estimation, typically prices and premiums and production quantities. That may be seen as advantage to reduce the amount of more or less constructed information entering the estimation, as input coefficients. However, the calibration process is cumbersome, and the resulting elasticities in simulation experiments will differ from the results of the econometric analysis.

The second approach – estimating parameters using the Kuhn-Tucker-conditions of the model – leads clearly to consistency between the estimation and simulation framework. However, for a model with as many choice variables as CAPRI that straightforward approach may require modifications as well, e.g. by defining the opportunity costs from the feed requirements exogenously.

The dissertation work of Torbjoern Jansson (Jansson 2007) focussed on estimating the CAPRI supply side parameters. The results have been incorporated in the current version. The milk study (2007/08) contributed additional empirical evidence on marginal costs related to milk production, see also Kempen, M., Witzke. P., Pérez-Dominguez. I., Jansson, T. and Sckokai, P. (2011): Economic and environmental impacts of milk quota reform in Europe, Journal of Policy Modeling, 33(1), pp 29-52.

4.5.4 Calibration tests with supply models

After calibrating the various functions of the supply models, a test for successful calibration is carried out. The purpose of the test is to ensure that the models are really properly calibrated, and to avoid that a disequilibrium in the baseline is misinterpreted as the effect of some policy change in a scenario.

To test for successful calibration, all supply models are solved directly after the calibration, and the solutions are compared to the target values to which the models should have been calibrated. If the solutions deviate more than some tiny amount, an error message is produced and the execution terminated. The calibration test checks for deviations of activity levels and allocation of fertilizers to crops.

4.5.5 Sensitivity experiments with the supply models

The market model of CAPRI is solved with a simplified representation of the supply model behaviour (see model overview). Even in countries where we do have a detailed supply model representation of agriculture, the market model contains, for technical / numerical reasons, a simpler linearized supply model that is iteratively re-calibrated to reflect the results of the underlying supply models in the current iteration between supply and demand.

If the linearized supply models would replicate the behaviour of the supply models exactly, then no iterations would be needed. In fact, no programming models of supply would be needed either. However, the approximation is not perfect, and hence the model needs to iterate between supply and demand. Since these iterations with re-calibrations are time consuming, it is desirable to have as good an approximation as possible.

The functional form of the approximation is derived from a ”normalized quadratic profit function”, meaning that the supply of any commodity is a linear function of all prices divided by a price index. Hence, the slope of those supply functions is a square matrix equivalent to the Hessian matrix of the normalized quadratic profit function itself. In order to find out how the supply models, including all policies and constraints, respond to changes in market prices, the calibration procedures of the CAPRI system contains a suite of structured and automated simulation experiments. The GAMS scenario solver is used to vary prices one by one and evaluate changes in supply. The results are summarized in the matrix of second-order derivatives used in the supply approximation in the market model.

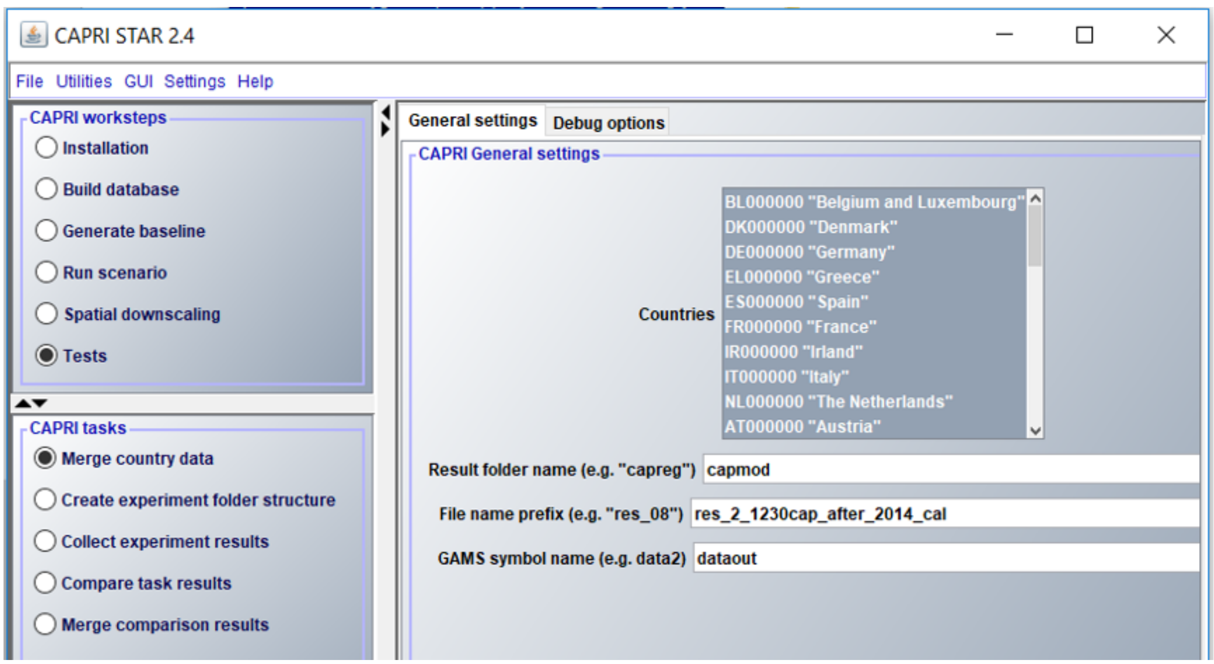
4.6 Baseline reproduction run

Not formally a component of the baseline calibration procedure, it has become an established habit to validate the calibration of supply using the simulation models themselves. There are many conceivable circumstances where the build-in calibration tests would pass, but a normal simulation nevertheless would not replicate the calibration point, for instance if some necessary and calibrated data is not properly loaded. Furthermore, the calibration does not produce the full report output, but only a limited set of variables. Therefore, a baseline reproduction is generally also needed in any applied project to establish the equilibrium comparison point.

In order to facilitate the evaluation of the calibration point, we run the same scenario as the one used to calibrate the model, but under a different name. “cal” and “ref” are frequently used name suffixes. Since the calibration is done country-wise, the result files of the calibration are found in one gdx-file per country. In order to be able to load all of them into the GUI and compare them to the outcome of the reproduction run, the utility “Merge country data” found under the work step “Tests” can be used. In the figure below the settings are shown that can read in all the country specific gdx files from the results directory (capmod), starting with the string “res_2_1230cap_after_2014_cal”, load a specified symbol (dataout), and store it back into a gdx file with the same name but without country suffix.

Then, the GUI can be used in a standard fashion to manually compare the activity levels reported after calibration with those computed in a baseline reproduction run.

Figure 12: CAPRI settings: read in all the country specific gdx files from the results directory (capmod); load a specified symbol (dataout); store the data back into a gdx file with the same name but without country suffix



Source: own illustration

Chapter 5

Scenario simulation

5.1 Overview of the system

The CAPRI simulation tool is composed of a supply and market modules, interlinked with each other.

In the *supply module*, regional or farm type agricultural supply of crops and animal outputs is modelled by an aggregated profit function approach under a limited number of constraints: the land supply curve, policy restrictions such as sales quotas and set aside obligations and feeding restrictions based on requirement functions. The underlying methodology assumes a *two stage decision process*.

In the *first stage*, producers determine *optimal variable input coefficients* per hectare or head (nutrient needs for crops and animals, seed, plant protection, energy, pharmaceutical inputs, etc.) for given yields, which are determined exogenously by trend analysis (CAPRI reference scenario) and updated depending on price changes against the baseline. Nutrient requirements enter the supply models as constraints and all other variable inputs, together with their prices, define the accounting cost matrix.

In the *second stage, the profit maximising mix of crop and animal activities* is determined simultaneously with cost minimising feed and fertiliser in the supply models. Availability of grass and arable land and the presence of quotas impose a restriction on acreage or production possibilities. Moreover, crop production is influenced by set aside obligations and animal requirements (e.g. gross energy and crude protein) are covered by a cost minimised feeding combination. Fertiliser needs of crops have to be met by either organic nutrients found in manure (output from animals) or in purchased fertiliser (traded good).

A cost function covering the effect of all factors not explicitly handled by restrictions or the accounting costs –as additional binding resources or risk ensures calibration of activity levels and feeding habits in the base year and plausible reactions of the system. These cost function terms are estimated from ex post data or calibrated to exogenous elasticities.

Fodder (grass, straw, fodder maize, root crops, silage, milk from suckler cows or mother goat and sheep) is assumed to be non-tradable, and hence links animal processes to the crops and regional land availability. A detailed description can be found in Britz and Heckelei (1999). All other outputs and inputs can be sold and purchased at fixed prices. The use of a mathematical programming approach has the advantage to directly embed compensation payments, set-aside obligations, and sales quotas, as well as to capture important relations between agricultural production activities. Not at least, environmental indicators as NPK balances and output of gases linked to global warming are directly inputted in the system.

The market module breaks down the world into 40 country aggregates or trading partners, each one (and sometimes regional components within these) featuring systems of supply, human consumption, feed and processing functions. The parameters of these functions are derived from elasticities borrowed from other studies and modelling systems and calibrated to projected quantities and prices in the simulation year. Regularity is ensured through the choice of the functional form (a normalised quadratic function for feed, processing and supply and a generalised Leontief expenditure function for human consumption) and some further restrictions (homogeneity of degree zero in prices, symmetry and correct curvature). Accordingly, the demand system allows for the calculation of welfare changes for consumers, processing industry and public sector. Policy instruments in the market module include bilateral trade flows[1]. Tariff rate quotas (TRQs), intervention purchases and subsidised exports under the World Trade Organisation (WTO) commitment restrictions are explicitly modelled for the EU.

In the market module, special attention is given to the processing of dairy products. First, balancing equations for fat and protein ensure that these make use of the exact amount of fat and protein contained in the raw milk. The production of processed dairy products is based on a normalised quadratic function driven by the regional differences between the market price and the value of its fat and protein content. Then, for consistency, prices of raw milk are also derived from their fat and protein content valued with fat and protein prices.

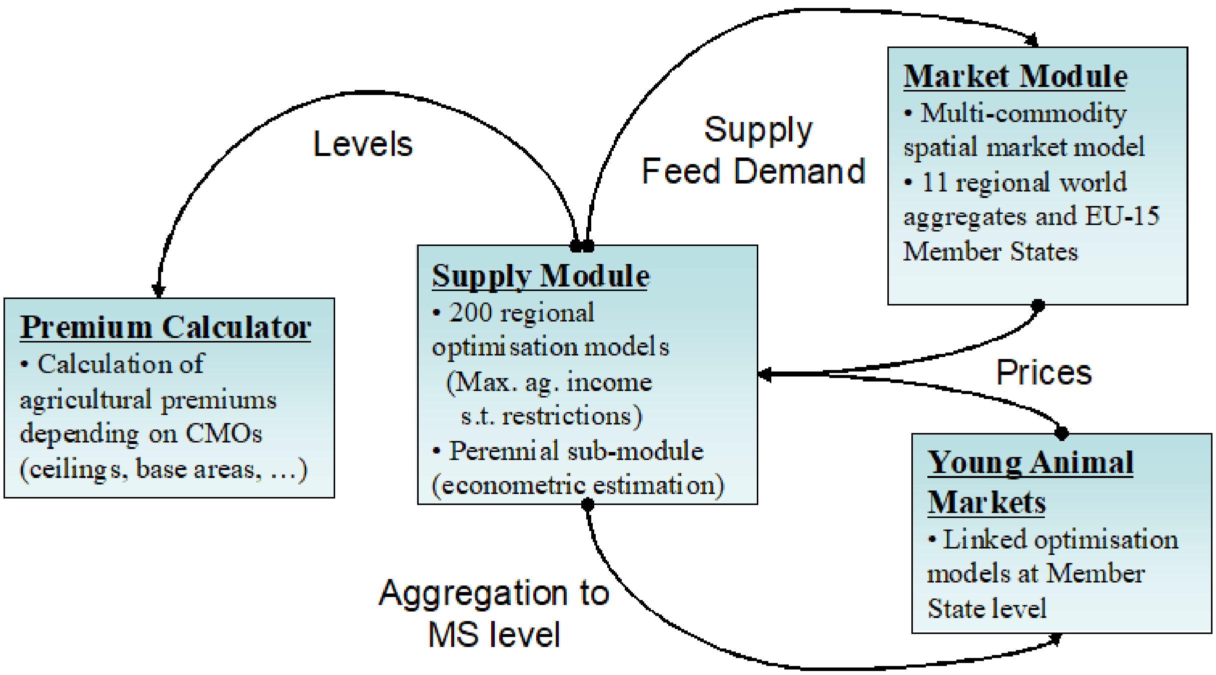
The market module comprises of a bilateral world trade model based on the Armington assumption (Armington, 1969). According to Armington's theory, the composition of demand from domestic sales and different import origins depends on price relationships according to bilateral trade flows. This allows the model to reflect trade preferences for certain regions (e.g. Parma or Manchego cheese) that cannot be observed in a net trade model.

The *equilibrium* in CAPRI is obtained by letting the *supply and market modules* iterate with each other. In the first iteration, the regional aggregate programming models (one for each NUTS 2 region or farm type) are solved with exogenous prices. Regional agricultural income is therefore maximised subject to several restrictions (land, fertiliser allocation, feed requirements, etc). After being solved, the regional results of these models (crop areas, herd sizes, input/output coefficients, etc.) are aggregated and enter a small, non-spatial multi-commodity module for young animal trade, as shown in Figure 12. In the second iteration, supply and feed demand functions of the market module are first calibrated to the results from the supply module on feed use and production obtained in the previous iteration. The market module is then solved at this stage (constrained equation system) and the resulting producer prices at Member State level transmitted to the supply models for the following iteration. At the same time, in between iterations, premiums for activities are adjusted if ceilings defined in the Common Market Organisations (CMOs) are overshot.

The implementation in CAPRI is based on a core module file *gams/capmod.gms*, which calls the different components of the system. The main input comes from the CAPRI database (COCO, CAPREG, GLOBAL), trends (CAPTRD) and baseline calibration parameters. The output of a scenario run is stored in a GDX file in folder output/results/capmod.

[1]Currently, no PSE/CSE data are used, and CSE are only introduced for EU dairy products as derived from FEOGA budget position.

Figure 13: Link of modules in CAPRI



Source: CAPRI Modelling System. Note: the number of regions is outdated in December 2011

5.2 Module for agricultural supply at regional level

5.2.1 Basic interactions between activities in the supply model

There are two sources for interactions between activities in simulation experiments: the objective function and constraints. In the current version of CAPRI, the objective function does solve inter-activity terms for groups of arable crops, so that the major interplay is due to constraints. The interaction is best understood by looking at the first order conditions of a programming model including PMP terms:

$$Rev_j = Cost_j + ac_j + \sum_k bc_{j,k} Levl_k + \sum_i^m \lambda_i a_{ij} \tag{5.1}$$

The left hand side (Rev) shows the marginal revenues, which are typically equal to the fixed prices times the fixed yields plus premiums. The right hand side shows the different elements of the marginal costs. Firstly, the variable or accounting costs ($Cost$) which are fix as they are based on the Leontief assumption. The term $(ac_j + \sum_k bc_{j,k} Levl_k)$ shows the marginal non-linear costs, which are increasing with the activity levels. The cross effects are only introduced to let major arable crop groups interact, whereas for fruits & vegetables, permanent crops, grassland and the animal sectors, only diagonal terms are introduced. The methodology for the estimation of these terms is described in Jansson and Heckelei

(2011).

The remaining term ($\sum_i^m \lambda_i a_{ij}$) captures the marginal costs linked to the use of exhausted resources and is equal to the sum of the shadow prices λmultiplied the per unit demand of resource i for activity j; the matrix $\boldsymbol{A}$ being again based on Leontief technology. The shadow values of binding resources hence are the drivers linking the activities.

The land balance plays a central role in the CAPRI supply model. The land shadow price appears as a cost in all crop activities including fodder producing ones, so that animals are indirectly affected as well. The second major link is the availability of not-marketable feeding stuff, and finally, less important, organic fertiliser.

The basic effects are best discussed with a simple example. Assume an increase of a per hectare premium for soft wheat, all other things unchanged.

- *What will happen in the model*? The increased premium will lead to an imbalance between marginal revenues (= yield times prices plus premium) and marginal costs (=accounting costs, 'resource use cost', non-linear costs). In order to close the gap, as marginal revenues are fixed, the area under soft wheat will be increased until marginal costs of producing soft wheat have increased to a point where they are again equal to marginal revenues. As the marginal costs linked to the non-linear cost function ($ac_j + \sum_k bc_{j,k} Levl_k$) are increasing in activity levels, increasing the area under soft wheat will hence reduce that gap. At the same time, as the land balance must be kept closed, other crop activities must be reduced. The non-linear cost function will for these crops now provoke a countervailing effect: reducing the activity levels of competing crops will lead to lower costs for these crops. With marginal revenues (*Rev*) and accounting costs (*Cost*) fixed, that will require the shadow price   of the land balance to increase.
- *What will be the impact on animal activities*? Again, the shadow price of the land balance will be crucial. For activities producing non-marketable feed, marginal revenues are not defined as prices times yields, but as internal feed value times prices. The internal feed value is determined as the substitution value of non-marketable fodder against other feeding stuff, and depends on their nutrient content and further feed restrictions. Increasing the shadow price of land will hence either require decreasing other costs in producing fodder or increasing the internal marginal revenues. In other words, a high shadow price of land renders non-marketable fodder less competitive compared to other feeding stuff. As feed costs are – however very slightly – increasing in quantities fed per head, feed costs for animals will increase. But as there are several requirement constraints involved, some feeding stuff may increase and other decrease. Clearly, the higher the share of non-marketable fodder in the mix for a certain animal type, the higher the effect. As marginal feed costs will increase, and marginal revenues for the animal process are not changing, other marginal costs in animal production need to be reduced, and again the non-linear cost function will be the crucial part, as the marginal cost related to it will decrease if herd sizes drop.

To summarize the supply response, increasing premiums for a crop will hence increase the cropping share of that crop, reduce the share of other crops, increase the shadow price of land, lead to less fodder production, higher fodder costs and thus reduced herd size of animals.

- *What will be the impacts covered by the market*? The changes in hectares will lead to increased supply of the crop with the higher premium and less supply of all other crops at given prices, i.e. one upward and many downward shifts of the supply curves. Equally, supply curves for animal products will shift downwards. On the other hand, some feed demand curve will shift as well, some upward, other downward. These shifts will move the market module away from the former fixed points where market balances were closed. For the crop product with the increased premiums, increased supply plus some changes in feed will most probably lead to lower prices, whereas prices of other crops will most probably increase. That will require new adjustments during the next iteration where the supply models are solved, with to a certain extent countervailing effects.

Table 25: Overview on a regional aggregate programming model

	Crop Activities	Animal Activities	Feed Use	Net Trade	Constraints
Objective function – Acc.Costs – variable cost function terms – Acc.Costs – variable cost function terms terms for feeding	+ Premium + Premium - variable cost function + Price				
Output	+	+	-	-	= 0
Area	-				< = land supply
Set aside	+/-				= 0
Quotas	-	-			< = Ref. Quantity
Fertilizer needs	-	+		+	= 0
Feed requirements		-	+	+	= 0

5.2.2 Detailed discussion of the equations in the supply model

The definition of the supply model can be found in *'supply/supply_model.gms'*

Feed block The feed block ensures that the requirements of the animal processes in terms of feed energy and protein are met and links these to the markets and crop production decisions.

$$\overline{AREQ}_{r,act,req}\overline{DAYS}_{r,act,req} = \sum_{feed} FEDNG_{r,act,feed}\overline{REQCNT}_{r,act,feed} \tag{5.2}$$

The left hand side captures the daily animal requirements ($AREQ$) for each region r, animal activity act and requirement $AREQ$ multiplied with the days ($DAYS$) the animal is in the production process. Both are parameters fixed during the solution of the modelling system. The right hand side ensures that the requirement content of the actual feed mix represented by the feeding ($FEDNG$) of certain type of feed to the animals multiplied with the requirement content ($REQCNT$) in the regions covers these nutritional demands. Requirements and contents are specified in the feed calibration while production days are determined in the "COCO1" module. Total feed use (*FEDUSE*) in a region is defined as the feeding per head multiplied with the activity level (*LEVL*) for the animal activities:

$$FEDUSE_{r,feed} = \sum_{aact} LEVL_{r,aact} FEDNG_{r,aact,feed} \tag{5.3}$$

Total feed use might be either produced regionally in the case of fodder assumed not tradable (grass, fodder root crops, silage maize, other fodder from arable land), or bought from the market at fixed prices.

Land balances and set-aside restrictions The model distinguishes arable and grassland and comprises thus two land balances:

$$\overline{LEVL}_{r,\text{"}arab\text{"}} \leq \sum_{arab} LEVL_{r,arab} \tag{5.4}$$

$$\overline{LEVL}_{r,\text{"}gras\text{"}} \leq LEVL_{r,\text{"}grae\text{"}} + LEVL_{r,\text{"}grai\text{"}} \tag{5.5}$$

Both land balances might become slack if marginal returns to land drops to zero. For arable land, idling land not in set-aside (activity *FALL*) is a further explicit activity. For the grassland, the model distinguishes two types with different yields (*GRAE*: grassland extensive, *GRAI*: grassland intensive) so that idling grassland can be expressed of an average lower production intensity of grassland by changing the mix between the two intensities.

The model comprises a land use module with two major components:

1. Imperfect substitution between arable and grass lands depending on returns to the two types of agricultural land uses.
2. A land supply curve which determines the land available to agriculture as a function to the returns to land.

There are hence two further equations:

$$\overline{LEVL}_{r,\text{"}uaar\text{"}} = \overline{LEVL}_{r,\text{"}arab\text{"}} + \overline{LEVL}_{r,\text{"}gras\text{"}} \tag{5.6}$$

And a further one which prevents numerical problems with the terms relating to land supply in the objective function

$$\overline{LEVL}_{r,\text{"}uaar\text{"}} = 0.999\overline{LEVL}_{r,\text{"}asym\text{"}} \tag{5.7}$$

Where "asym" is the land asymptote, i.e. the maximal amount of economically usable agricultural area in a region when the agricultural land rent goes towards infinity. For an application where the land market is used see Renwick et al. (2013).

Set aside policies have changed frequently during CAP reforms. The recent specification is covered in the context of the premium modelling in Section 5.3. The obligatory set-aside restriction introduced by the

McSharry reform 1992 and valid until the implementation of the Luxembourg compromise of June 2003 has been explicitly modelled through this equation:

$$
\begin{aligned}
& LEVL_{r,\text{"}iset\text{"}} + LEVL_{r,\text{"}gset\text{"}} + LEVL_{r,\text{"}tset\text{"}} \\
& = \sum_{arab} LEVL_{r,arab} \left(1 - NONS_{r,arab}\right) \frac{1/100 SETR_{r,arab}}{1 - 1/100 SETR_{r,arab}}
\end{aligned} \tag{5.8}
$$

LEVL_{r,“iset”}As seen from above, the model distinguishes between three types of obligatory set-aside: idling (ISET), for grass land use (GSET) and for forestation purposes (TSET). The share of so-called non-food production exempt from set-aside (NONS) for each activity and region is fixed and given.

The equation above is replaced for years where the Luxembourg compromise of June 2003 is implemented by a Member State, where the level of obligatory set-aside is fixed instead to the historical obligations.

For certain years of the McSharry reform, the total share of set-aside – be it obligatory or voluntary – on a list of certain crops was not allowed to exceed a certain ceiling. That restriction is captured by the following equation:

$$
\begin{aligned}
& LEVL_{r,\text{"}iset\text{"}} + LEVL_{r,\text{"}gset\text{"}} + LEVL_{r,\text{"}tset\text{"}} + LEVL_{r,\text{"}vset\text{"}} \\
& \leq \sum_{arab \wedge SETF_{r,arab}} LEVL_{r,arab} / \overline{MXSETA}
\end{aligned} \tag{5.9}
$$

Fertilising block As of CAPRI Stable Release 2.1, the fertilizer allocation was modified, and this section of the documentation updated. Notation has changed compared with previous versions of the model and documentation. Here, we represent the equations in more general mathematical notation, avoiding the long GAMS code names of the source code, in order to save space.

We distinguish the three macro-nutrients N, P and K. The supply and uptake of those nutrients are modelled in a uniform way, save for the fact that there is fixation and atmospheric deposition only of N.

Each crop has a requirement per hectare, calculated based on the yield. Yields are exogenous from the vantage point of the producer, but there are alternative technologies available for each cropping activity, and a separable, i.e. handled outside of the optimization model, relation between prices and optimal yields.

From the basic nutrient requirement we first deduct the rate of biological fixation (only for nitrogen and selected crops). The remainder is inflated by a (calibrated) factor and additive term of over-fertilization, and then scaled with a soil-specific factor (only for nitrogen), to arrive at the total amount of nutrients that need to be supplied to the crop. This is the left hand side of Equation 96 FIXME.

Nutrient supply, shown on the right handside, comes from mineral fertilizer, manure, crop residues and atmospheric deposition. Mineral fertilizer may have ammonia losses during application. For manure, there are both losses and inefficiencies. When manure is applied to crops, there is an efficiency factor applied to the nutrient content (denoted by _(r,“excr”, n)), corresponding to the Fertilizer Value (FV) of manure relative to mineral fertilizer. The efficiency factor is a key parameter of interest in simulations carried out in some studies. Crop residues can be re-distributed among crop groups for annual arable crops but not for grassland and permanent crops, where it stays with the crop that produced it. For crop residues there is both a loss rate and a fertilizer value.

$$
\begin{aligned}
&\sum_{i \in I\, j,k} \left[levl_{rik} \left(ret_{rni}(1 - biofix_{rni})\lambda^{prop}_{rnik} + \lambda^{const}_{rni} \right) soil_{rn}\, yf_{rnik} \right] \\
&= fmine_{rni}(1 - loss_{rn}) + fexcr_{rni}\phi_{r,excr,n} + (1 - isPerm_j) fcres_{rni}(1 - loss_{rn})\phi_{r,cres,n} \\
&+ isPerm_j \sum_{i \in I\, j,k} levl_{rik} res_{rni}(techf_{rink} + 1)(1 - loss_{rn})\phi_r, cres, n \\
&\forall r, n, j
\end{aligned}
\tag{5.10}
$$

FIXME

Indices:

r = region
i = crop
j = crop group
k = technological crop option (high/low yield)
n = nutrient (N/P/K)
$isPerm_j$ = indicates that crop group j contains permanent crops

Endogenous choice variables:

$levl_{rik}$ = Area (ha) of each crop i and technology k in region r.
$fmine_{rnj}$ = Application of mineral fertilizer n to crop group j in region r.
$fexcr_{rnj}$ = Application of manure n to crop group j in region r.
$fcrex_{rnj}$ = Allocation of crop residue n to crop group j in region r.

Parameters: ret_{rni} = Retention (uptake) of nutrients by the crop
res_{rni} = Crop residues output
$biofix_{rni}$ = Biological fixation, share (only for N and selected crops)
$^{prop}_{rnik}$ = Over-fertilization factor, calibrated
$^{const}_{rni}$ = Over-fertilization term, calibrated
$soil_{rn}$ = Soil factor
yf_{rnik} = Yield factor for technologies
$loss_{rn}$ = Loss rate
$_{r",excr",n}$ = Nutrient availability ratio for manure
$_{r,"cres",n}$ = Nutrient availability ratio for crop residues

The reader may have noted that there is no loss rate for manure in the Equation 96 FIXME . CAPRI does contain such loss rates, but they are specific for each animal type and therefore happens on the manure supply side of the regional manure balance (see section on input allocation).

The model contains three types of manure: N-manure, P-manure and K-manure. From an agricultural point of view this may seem odd. It might be more intuitive to think of one type of manure per animal category. The motivation is to keep the system simple and flexible. With the present representation, where each animal category supplies N, P, and K-manure, the number of manure classes can be limited and yet the unique mix of nutrients from each animal category can be defined.

The supply of each manure type is collected in a "pool" for each regional farm model, i.e. for each NUTS2 region. Regions within a member state may trade manure, subject to a cost. The supply in the pool plus the traded quantities has to be distributed to the crops in the region, i.e. there is an equality-restriction in place. This is handled in the equations "FertDistExcr_" and "ManureNPK_". Note that fertilizer flows are measured in *tons*, for the sake of scaling, whereas other total quantities in CAPRI are measured in 1000 tons. Hence the factors 1000 and 0.001.

$$\sum_j fexcr_{rnj} = 1000 v\_ManureNPK_{rn} \tag{5.11}$$

$$v\_ManureNPK_{rn} + \sum_s T_{rs} nutshr_{rn} = 0.001 \sum_{i \in Anim_j, k} levl_{rik} o_{rnik}(1 - loss_{rin}) \quad \forall r, n \tag{5.12}$$

where
o_{rnik} is the output of manure nutrient n from animal type i using technology k in region r,
$nutshr_{rn}$ is the average content of each nutrient in the regional manure pool,
T_{rs} is the quantity of manure traded from r to s,
$isAnim_i$ indicates that activity i is an animal production activity

Equation "FertDistMine_" allocates total mineral fertilizer sales to the crops / group of crops.

$$\sum_j fmine_{rnj} = -netPutQuant_{rn} \tag{5.13}$$

Finally, crop residues and atmospheric definition are distributed in equation "FertDistCres_".

$$\sum_j fcers_{rnj} = \sum_{i \notin isPerm_j, k} levl_{rik} res_{rni}(techf_{rink} + 1) \tag{5.14}$$

One flow from a source s={"mine" ,"cres" ,"excr" } to a sink j={"crop groups"} can in general be anything from zero and upwards. The nutrient balance equations above do not uniquely determine each flow of nutrients from sources to sinks, but it is indeed possible that in one simulation, say, a particular crop group gets much crop residues and little manure, whereas the opposite holds in the next simulation. The total balances will hold equally well in either situation, and the profits will not be affected since the same total amount of mineral fertilizer is purchased, but we do have a stability problem for the model. Furthermore, the different nutrient flows may influence the greenhouse gas emission coefficients of crops (if e.g. the emissions of enteric fermentation follows the manure to the crops). The problem is under-determined, or ill-posed.

To resolve the ill-posedness of the fertilizer distribution, we propose a probabilistic approach. This means that we do not introduce any additional economic model for the allocation that somehow makes increasing fertilizer flows more expensive. Instead, we assume that whatever the reasons the farmers have for choosing a particular distribution, those reasons are similar in two simulations, and therefore the fertilizer flows are also similar. Thus, a larger deviation from some reference flows is deemed improbable, albeit not costlier than the situation with the reference flows.

To develop this probabilistic model, we assume that the decisions of the farmer are separable and taken in two steps: first, the farmer decides about the cropping plan and just ensures that the total amount of fertilizer available is sufficient. This is called the outer model. Then, a statistical model is solved that finds the most probable fertilizer flows out of the continuum of possible ones. This is called the inner model. The structure with outer and inner models makes the problem a bi-level programming one.

To implement the bi-level programming problem in a way that does not change the present structure of the model (with just one optimization solve of the representative farm model) we implement the inner model by its optimality conditions. By carefully choosing the proper probability density functions we ensure that no complementary slackness conditions are needed, so that the inner model is simple to solve. For this the gamma density function is very suitable, as it has a support from zero to infinity, with a probability that goes towards zero as the random variable goes to zero.

The parameters of the gamma function are determined in the calibration step, described further below, and then kept constant in simulation. The gamma density function for some random variable x has the form

$$p(x|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{5.15}$$

where $\Gamma(\ )$ is the gamma function, and and are parameters that determine the shape of the density function. The gamma density is nonlinear, and the joint density, being the product of the densities of all nutrient flows, is even more so. In order to reduce nonlinearity we note that are interested in finding the highest posterior density, i.e. maximizing a joint density function, and since the maximum is invariant to monotonous positive transformations we compute the logarithm of the joint density, which will be the sum of terms like the following (the constant term has been omitted since it also does not influence the optimal solution for x):

$$log\, p(x|\alpha,\beta) \propto (\alpha-1) log\, x - \beta x \tag{5.16}$$

Maximization of the logged density under the constraints that the nutrient balance restrictions of the supply modes have to be met gives a set of equations that define explicit and unique fertilizer flows, where v are the Lagrange multipliers of the source-pool restrictions and u the Lagrange multipliers of the nutrient balance equation.

$$\text{FOC w.r.t. manure use:} \frac{\alpha_{r,excr,nj}-1}{fexcr_{rnj}} - \beta_{r,excr,nj} - v_{r,excr,n} + \phi_{r,excr,n} u_{rnj} = 0 \tag{5.17}$$

$$\text{FOC w.r.t. crop residues use:} \frac{\alpha_{r,cres,nj}-1}{fcres_{rnj}} - \beta_{r,cres,nj} - v_{r,cres,n} + (1 - loss_{rn} \phi_{r,excr,n} u_{rnj} = 0 \forall j \notin isPerm_j \tag{5.18}$$

$$\text{FOC w.r.t. mineral feritilzer use:} \frac{\alpha_{r,mine,nj}-1}{fmine_{rnj}} - \beta_{r,mine,nj} - v_{r,mine,n} + (1 - loss_{rn}u_{rnj} = 0 \quad (5.19)$$

The system of FOC contains expressions of the type $1/fmine$ which is likely to impair performance as the second derivatives are not constant (CONOPT computes second derivatives). Therefore, the first term in each FOC was turned into a new variable z defined as $z_{r,"excr",nj} fexcr_{r,"excr",nj} = \alpha_{r,"excr",nj} - 1$, and similar for each source, which is a quadratic expression.

Balancing equations for outputs Outputs produced must be sold – if they are tradable across regions – or used internally, as in the case of young animals or feed.

$$\sum_{act} Levl_{r,act} OUTP_{r,act,o} = NETTRD_r^{o \notin fodder} + YANUSE_r^{o \notin oyani} + FEDUSE_r^{o \in fodder} \quad (5.20)$$

In the case of quotas (milk, for sugar beet) the sales to the market may be bounded (noting that NETTRD = v_netPutQuant in the code):

```
*
* --- upper limit on sales are quotas
*
  v_netPutQuant.UP(RU,OM) $ ((DATA(RU,"QUTS",OM,"Y") ne eps)
                              $ DATA(RU,"QUTS",OM,"Y") $ (NOT SAMEAS(OM,"SUGB"))) =
     DATA(RU,"QUTS",OM,"Y");
```

As described in the data base chapter, the concept of the EAA requires a distinction between young animals as inputs and outputs, where only the net trade is valued in the EAA on the output side. Consequently, the remonte expressed as demand for young animals on the input side must be mapped into equivalent 'net import' of young animals on the output side:

$$\sum_{aact} Levl_{r,aact} I_{r,aact,yani} = YANUSE_r^{oyani \leftrightarrow iyani} \quad (5.21)$$

In combination with the standard balancing equation shown above, the NETTRD variable for young animals on the output side becomes negative if the YANUSE variable for a certain type of young animals exceeds the production inside the region.

The objective function The objective function is split up into the linear part, the one related to the quadratic cost function for activities, and the quadratic cost function related to the feed mix costs:

$$OBJE = \sum_r LINEAR_r + QUADRA_r + QUADRF_r \quad (5.22)$$

The linear part comprises the revenues from sales and the costs of purchases, minus the costs of allocated inputs not explicitly covered by constraints (i.e. all inputs with the exemptions of fertilisers, feed and young animals) plus premiums:

$$LINEAR_r = \sum_{io} NETTRD_{r,io}\overline{PRICE}_{io} + \sum_{a} ctLEVL_{r,act}\left(\overline{PRME}_{r,act} - \overline{COST}_{r,act}\right) \tag{5.23}$$

The quadratic cost function relating to feed is defined as follows:

$$QUDRAF_r = \sum_{aact,feed}\left[\begin{array}{c} LEVL_{r,aact}FEDNG_{r,aact,feed} \\ (a_{r,aact,feed} + 1/2 b_{r,aact,feed}FEDNG_{r,aact,feed}) \end{array}\right] \tag{5.24}$$

The marginal feed costs per animal increase hence linearly with an increase in the feed input coefficients per animal. It should be emphasised that this is the main mechanism that "stabilises" the feed allocation by animals. The two balances on feed energy and protein alone would otherwise leave the feed allocation indeterminate and give a rather "jumpy" simulation behaviour.

There is another more complex PMP term (equation quadra_ in supply_model.gms, not reproduced in this section) quadratic in activity levels and differentiated by the two technologies that "stabilises" the composition of activites according to previous econometric estimates or default assumptions.

A final term relates to the entitlements introduced with the 2003 Mid Term reviews. If those entitlements are overshot, a penalty term equal to the premium paid under the respective scheme (regional, historical etc.) is subtracted to the objective. Accordingly, the marginal premium for an additional ha above the entitlement ceiling is zero.

Sugar beet (M. Adenäuer, P. Witzke) The Common Market Organisation (CMO) for sugar regulates European sugar beet supply with a system of production quotas, even after the significant reforms of 2006, up to year 2017 when the quota system expired. Before that reform, two different quotas had been established subject to different price guarantee (A and B quotas, qA and qB). Beet prices were depending on intervention prices and levies to finance the subsidised export of a part of the quota production to third countries. Sugar beets produced beyond those quotas (so called C beets) were sold as sugar on the world market at prevailing prices, i.e. formally without subsidies. However, a WTO panel initiated by Australia and Brazil concluded that the former sugar CMO involved a cross-subsidisation of C-sugar from quota sugar such that all exports of C sugar was also counted in terms of the EU's limits on subsidised exports. As a consequence, this outlet for EU surplus production was closed. The reformed CMO therefore does not allow any exports beyond the Uruguay round limits. Instead, processing of beets to ethanol emerged as a new outlet that economically plays a similar role as former C beet production: It offers an outlet for high production quantities that exceed the quota limits of farmers, but at a reduced price. Basically, farmers face a kinked beet demand curve that potentially involved three price levels:

- A-beets receiving the highest price derived from high sugar prices (and before the 2006 reform less a small levy amount)
- B-beets receiving a lower price as the applicable levies were higher before the reform. However, the 2006 reform eliminated the distinction of A and B quotas. Furthermore, the sugar industry applied a pooling price system in many MS that also eliminated the distinction between A and B beets.
- C-beets receiving the lowest price, formerly derived from world market sugar prices, now derived from ethanol prices.

The high price sector covers for farmers at least the farm level quota endowment. However, the sugar industry may grant high prices also for a limited, “desirable” over-quota production, for example to avoid bottlenecks in sugar or ethanol production. This has been the case in some EU countries before the reform (so-called “C1 beets”) and it is also current practice (see, for example `http://www.liz-online.de`).

Considering a kinked demand curve and in addition yield uncertainty renders the standard profit maximisation hypothesis inappropriate for the sugar sector (at least). The CAPRI system therefore applies an expected profit maximisation framework that takes care for yield uncertainty (see Adenäuer 2005). The idea behind this is that observed C sugar productions in the past are unlikely to be an outcome of competitiveness at C beet prices rather than being the result of farmers' aspirations to fulfil their quota rights even in case of a bad harvest. This approach essentially assumes that the “behavioural quotas” of farmers may exceed the “legal quotas” (derived from the sugar CMO) by some percentage. This percentage reflects in part the pricing behaviour of the regional sugar industry, but it may also depend on farmers expectations on the consequences of an incomplete quota fill. These aspects may be captured with the following specification of expected sugar beet revenues that substitute for the expression $NETTRD_{r,io}\ PRICE_{io}$ (if $io = SUGB$) in equation below:

$$\begin{aligned} SegbREV_r = {} & p^A NETTRD_{r,SUGB} \\ & - \left(p^A - p^B\right) \begin{bmatrix} (1 - CDFSugb(q^A))(NETTRD_{r,SUGB} - q^A) \\ +(\sigma^S)^2 PDFSugb(q^A) \end{bmatrix} \\ & - \left(p^B - p^C\right) \begin{bmatrix} (1 - CDFSugb(q^A + B))(NETTRD_{r,SUGB} - q^A + B) \\ +(\sigma^S)^2 PDFSugb(q^A + B) \end{bmatrix} \end{aligned} \tag{5.25}$$

Where $PDFSugb_r$ and $CDFSugb_r$ are the probability res. cumulated density functions of the NETTRD variable with the standard deviation σ^S. σ^S is defined as $NETTRD_{r,SUGB} * VCOF_r$, where the latter is the regional coefficient of yield variation estimated from FADN. $p^A BC$ are the prices for the three different types of sugar beet which are exogenous and linked to the EU and world market prices for sugar. The quotas q^A and q^{A+B} used in Equation 111 FIXME are the “behavioural quotas, currently specified as follows:

$$\begin{aligned} p^A &= legalqout^A \cdot scalefac \\ &= legalqout^A \cdot \left(\frac{NETTRD^{cal}_{SUGB}}{legalqout^A}\right)^{0.8} \end{aligned} \tag{5.26}$$

The scaling factor to map from the legal quota legalquotA (as the B quota has been eliminated in the sugar reform, it holds that $q^A = q^{A+B}$)to the behavioural quota qA depends on the projected sugar beet sales quantity in the calibration point $NETTRD^{cal}_{SUGB}$: For a country with a high over quota production (say 40%) we would obtain a scaling factor of 1.31, such that this producer will behave like a moderate C-sugar producer: responsive to both the C-beet prices as well as to the quota beet price (and the legal quotas). Without this scaling factor, producers with significant over quota p roduction, like France and Germany, would not show any sizeable response to a 10% cut of either the legal quotas or the quota price (at empirically observed coefficients of variation). As it is likely that the profitability of ethanol beets benefit from cross-subsidisation from the quota beets such a zero responsiveness was considered implausible.

Update note A number of recent developments are not covered in the previous exposition of supply model equations

1. A series of projects have added a distinction of rainfed and irrigated varieties of most crop activities which is the core of the so-called "CAPRI-water" version of the system[2].
2. Several projects have added endogenous GHG mitigation options[3]
3. Several new equations serve to explicitly represent environmental constraints deriving from the Nitrates Directive and the NEC directive[4].
4. A complete area balance monitoring the land use changes according to the six UNFCCC land use types (cropland, grassland, forest land, wetland, settlements, residual land) has been introduced for carbon accounting

5.2.3 Calibration of the regional programming models

Since the very first CAPRI version, ideas based on Positive Mathematical Programming were used to achieve perfect calibration to observed behaviour – namely regional statistics on cropping pattern, herds and yield – and data base results as the input or feed distribution. The basic idea is to interpret the 'observed' situation as a profit maximising choice of the agent, assuming that all constraints and coefficients are correctly specified with the exemption of costs or revenues not included in the model. Any difference between the marginal revenues and the marginal costs found at the base year situation is then mapped into a non-linear cost function, so that marginal revenues and costs are equal for all activities. In order to find the difference between marginal costs and revenues in the model without the non-linear cost function, calibration bounds around the choice variables are introduced.

The reader is now reminded that marginal costs in a programming model without non-linear terms comprise the accounting cost found in the objective and opportunity costs linked to binding resources. The opportunity costs in turn are a function of the accounting costs found in the objective. It is therefore not astonishing that a model where marginal revenues are not equal to marginal revenues at observed activity levels will most probably not produce reliable estimates of opportunity costs. The CAPRI team responded to that problem by defining exogenously the opportunity costs of two major restrictions: for the land balance and for milk quotas. The remaining shadow prices mostly relate to the feed block, and are less critical as they have a clear connection to prices of marketable feed as cereals which are not subject to the problems discussed above.

5.2.4 Estimating the supply response of the regional programming models

The development, test and validation of econometric approaches to estimate supply responses at the regional level in the context of regional programming models form an important task for the CAPRI

[2] A more complete presentation is given in `https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/extension-capri-model-irrigation-sub-module`.

[3] These are most completely included in the "trunk" version of the CAPRI system. For details, see, for example, `http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101396/jrc101396_ecampa2_final_report.pdf`.

[4] These are most completely included in the "trunk" version of the CAPRI system but developments are still ongoing.

team. Up to now, there is still no fully satisfactory solution of the problem, but some of the approaches are discussed in here.

The two possible competitors are standard duality based approaches with a following calibration step or estimates based directly on the Kuhn-Tucker conditions of the programming models. Both may or may not require a priori information to overcome missing degrees of freedom or reduce second or higher moments of estimated parameters. The duality based system estimation approach has the advantage to be well established. Less data is required for the estimation, typically prices and premiums and production quantities. That may be seen as advantage to reduce the amount of more or less constructed information entering the estimation, as input coefficients. However, the calibration process is cumbersome, and the resulting elasticities in simulation experiments will differ from the results of the econometric analysis.

The second approach – estimating parameters using the Kuhn-Tucker-conditions of the model – leads clearly to consistency between the estimation and simulation framework. However, for a model with as many choice variables as CAPRI that straightforward approach may require modifications as well, e.g. by defining the opportunity costs from the feed requirements exogenously.

The dissertation work of Torbjoern Jansson (Jansson 2007) focussed on estimating the CAPRI supply side parameters. The results have been incorporated in the current version. The milk study (2007/08) contributed additional empirical evidence on marginal costs related to milk production (Kempen et al. 2011)

5.2.5 Price depending crop yields and input coefficients

Let Y denote yields and j production activities. Yield react via iso-elastic functions to changes in output prices

$$log(Y_j) = \alpha_j + \epsilon_j \, log(p_o) \tag{5.27}$$

The current implementation features yield elasticities for cereals chosen as 0.3, and for oilseeds and potatoes chosen as 0.2. These estimates might be somewhat conservative when compared e.g. with Keeney and Hertel (2008). However, in CAPRI they relate to small scale regional units and single crops, and to European conditions which might be characterized by a combination of higher incentive for extensive management practises and dominance of rainfed agriculture where water might be a yield limiting factor.

Currently, the code is set up as to only capture the effect of output prices. However, in order to spare calculation of the constant terms , the actual code implemented in '*endog_yields.gms*' change the yields iteratively in between iterations t, using relative changes:

$$Y_{j,t} = Y_{j,t-1}^{[\epsilon_j log \frac{p_{o,t-1}}{p_t}]} \tag{5.28}$$

5.2.6 Annex: Land supply and land transitions in the supply part of CAPRI

Introduction

This technical paper explains how the most aggregate level of the CAPRI area allocation in the context of the supply models has been re-specified in the TRUSTEE project[5] and subsequently adopted in the CAPRI trunk. The former specification for land supply and transformation functions focused on agricultural land use and the transformation of agricultural land between arable land and grass land[6].

During the subsequent period, CAPRI was increasingly adapted to analyses of greenhouse gas (GHG) emission studies. Examples include CAPRI-ECC, GGELS, ECAMPA-X, AgCLim50-X, (European Commission, Joint Research Centre), ClipByFood (Swedish Energy Board), SUPREMA (H2020). This vein of research is very likely to gain in importance in the future.

In order to improve land related climate gas modelling within CAPRI, it was deemed appropriate to (1) extend the land use modelled to *all* available land in the EU (i.e. not only agriculture), and (2) to explicitly model *transitions* between land use classes. The pioneering work was carried out within the TRUSTEE project[7], but as always, an operational version emerged only after integrating efforts by researchers in several projects working at various institutions.

This paper focusses on the theory applied while data and technical implementation are only briefly covered.

A simple theory of land supply

Recall the dual methodological changes attempted in this paper:

1. Extend land use modelling to the entire land area, and
2. Explicitly model transitions between each pair of land uses

In order to keep things as simple as possible, we opted for a theory where the decision of how much land to allocate to each use is independent of the explicit transitions between classes. This separation of decisions is simplifying the theoretical derivations, but also seem to have some support in theory: land use transitions show a good deal of stability over time. We would like to remind sceptics of this assumption that the converse is not implied: land transitions are certainly strongly depending on the land use requirements.

The land supply and transformation model developed here is a bilevel optimization model. At the higher level (sometimes termed the *outer problem*), the land owner decides how much land to allocate to each aggregate land use based on the rents earned in each use and a set of parameters capturing the costs required in order to ensure that the land is available to the intended use. At the lower level (sometimes termed the *inner problem*), the transitions between land classes are modelled, with the condition that the total land needs of the outer problem are satisfied. The inner problem is modelled as a stochastic process involving no explicit economic model.

For the outer problem, i.e. the land owner's problem, we propose a quadratic objective function that maximizes the sum of land rents minus a dual cost function. The parameters of the dual cost function were specified in two steps:

[5] `https://www.trustee-project.eu/`

[6] See `https://svn1.agp.uni-bonn.de/svn/capri/trunk/doc/landSupplyCAPRI_v5.pdf` (Torbjörn Jansson, Wolfgang Britz, Alan Renwick and Peter Verburg (2010) Modelling CAP reform and land abandonment in the European Union with a focus on Germany.)

[7] `https://www.trustee-project.eu/`

1. A matrix of land supply elasticities was estimated (by TRUSTEE partner Jean Saveur Ay, CESEAR, Dijon (JSA). This estimation might be updated in future work or replaced with other sources for elasticities.
2. The parameters of the dual cost function are specified so that the supply behaviour replicates the estimated elasticities as closely as possible while exactly replicating observed/estimated land use and land rents.

The model is somewhat complicated by the fact that land use classes in CAPRI are defined somewhat differently compared to the UNFCCC accounting and also in the land transition data set. Therefore, some of the land classes used in the land transitions are different from the ones used in the land supply model. In particular, “Other land”, “Wetlands” and “Pasture” are differently defined. To reconcile the differences, we assumed constant shares of the intersections of the different sets, as explained below.

Inner model – land transitions

A vector of supply of land of various types could result from a wide range of different transitions. The inner model determines the matrix of land transitions that is “most likely”. The concept of “most likely” is formalized by assuming a joint density function for the land transitions, based on the historically observed transitions. The model then is to find the transition matrix that maximizes the joint density function.

Since each transition is non-negative, but in principle unlimited upwards, we opted for a gamma density function, that has the support $[0, \infty]$. For those that cannot immediately recall what the gamma density function looks like, and as entertainment for those that can, Figure 1 shows the graph of the density function for different parameters, all derived from an assumed mode of “1” and different assumed ratios “mode/standard deviations” (that we called “acc” for “accuracy” in the figure).

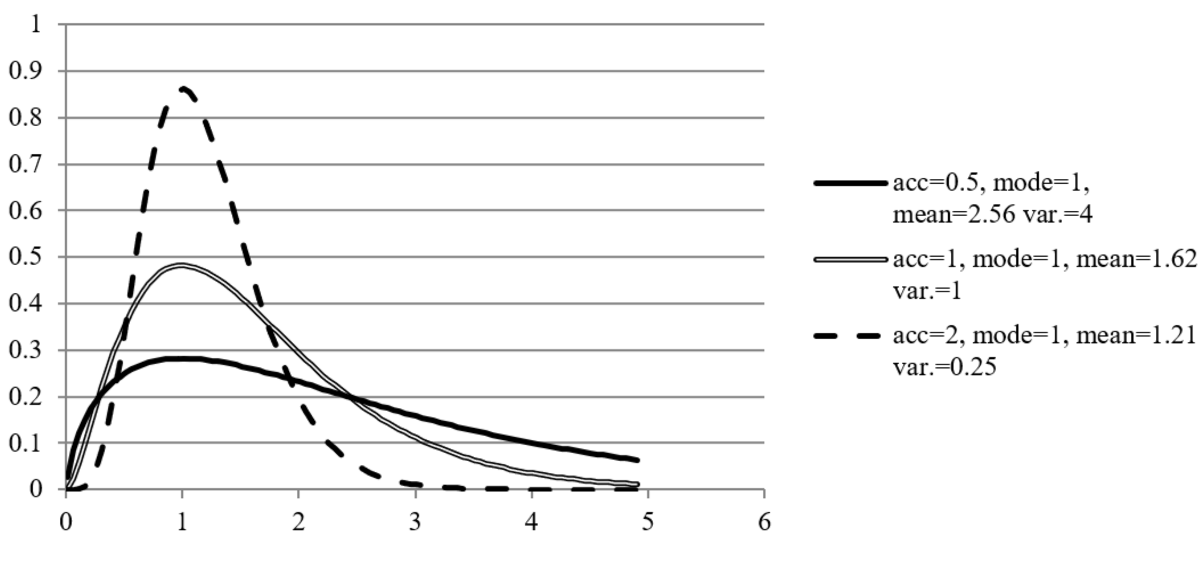


Figure 1: Gamma density graph for mode=1 and various standard deviations. “acc”=“mode/standard deviation”.

Let i denote land use classes in CAPRI definition, whereas l and k are land uses in UNFCCC classification. Let LU_k be total land use after transitions and $\mathrm{LU}_l^{\mathrm{initial}}$ be land use before transitions. Furthermore, let T_{lk} denote the transition of land from use l to use k. Noting that it is simpler and fully equivalent to maximize a sum of logged densities than a product of densities, the likelihood maximization problem can be written (with f being the gamma density function)

$$\max_{T_{\mathrm{lk}}} \log \prod_{\mathrm{lk}} f\left(T_{\mathrm{lk}}|\alpha_{\mathrm{lk}}, \beta_{\mathrm{lk}}\right) = \max_{T_{\mathrm{lk}}} \sum_{\mathrm{lk}} \log f\left(T_{\mathrm{lk}}|\alpha_{\mathrm{lk}}, \beta_{\mathrm{lk}}\right) \tag{5.29}$$

$$\Rightarrow \max_{T_{\mathrm{lk}}} \sum_{\mathrm{lk}} \left[\left(\alpha_{\mathrm{lk}} - 1\right) \log T_{\mathrm{lk}} - \beta_{\mathrm{lk}} T_{\mathrm{lk}}\right] \tag{5.30}$$

subject to

$$\mathrm{LU}_k - \sum_l T_{\mathrm{lk}} = 0 \ \ [\tau_k] \tag{5.31}$$

$$\mathrm{LU}_l^{\mathrm{initial}} - \sum_k T_{\mathrm{lk}} = 0 \ \ \left[\tau_l^{\mathrm{initial}}\right] \tag{5.32}$$

$$\mathrm{LU}_k - \sum_i \mathrm{share}_{\mathrm{ki}} \mathrm{LEV} L_i = 0 \tag{5.33}$$

The last equation is needed to convert land use in UNFCCC classification to land use in CAPRI classification, using a fixed linear transformation matrix $\mathrm{share}_{\mathrm{ki}}$. This discrepancy between land class accounts will be expanded on in a subsequent section. Forming the Lagrangian function and taking the derivatives with respect to land transitions gives the following first-order optimality conditions:

$$\left(\alpha_{\mathrm{lk}} - 1\right) T_{\mathrm{lk}}^{-1} - \beta_{\mathrm{lk}} + \tau_k + \tau_l^{\mathrm{initial}} = 0 \tag{5.34}$$

The parameters α and β of the gamma density function were computed by assuming that (i) the observed transitions are the mode of the density, and (ii) the standard deviation equals the mode. Then the parameters are obtained by solving the following quadratic system:

$$\mathrm{mode} = \frac{\alpha - 1}{\beta} \tag{5.35}$$

$$\mathrm{variance} = \frac{\alpha}{\beta^2} \tag{5.36}$$

Land use transitions as implemented in CAPRI

The implementation in CAPRI differs from the above general framework in that it explicitly identifies the *annual* transitions in year t T_{lk}^t from the initial $\mathrm{LU}_l^{\mathrm{initial}}$ land use to the final land use LU_k. This is necessary to identify the annual carbon effects occurring only in the final year in order to add them to the current GHG emissions, say from mineral fertiliser application in the final simulation year. If the initial year is the base year = 2008 and projection is for 2030, then the carbon effects related to the change from the 2008 $\mathrm{LU}_l^{\mathrm{initial}}$ to the final land use LU_k (=T_{lk}in the above notation, without time index) refer to a period of 22 years that cannot reasonably be aggregated with the "running" non-CO2 effects from the final year 2030. Furthermore the historical time series used to determine the mode of the gamma density for the transitions also refer to annual transitions.

Initially the problem to link total to annual transitions has been solved by assuming a linear time path from the initial to the final period, but this was criticised as being an inconsistent time path (by FW). Ultimately the time path has been computed therefore in the supply model in line with a static Markov chain with constant probabilities P_{lk} such that both land use LU_l^t as well as transitions T_{lk}^t in absolute ha require a time index (e_luOverTime in supply_model.gms).

$$\mathrm{LU}_k^t - \sum_l P_{\mathrm{lk}} \mathrm{LU}_l^{t-1} = 0\ ,\ t = \{1, \dots s\} \tag{5.37}$$

Where LU_k^s is the final land use in the simulation year s and $\mathrm{LU}_k^0 = \mathrm{LU}_k^{\mathrm{iniital}}$ is the initial land use. The transitions in ha in any year may be recovered from previous years land use and the annual (and constant) transition probabilities (e_LUCfromMatrix in supply_model.gms).

$$T_{\mathrm{lk}}^t = P_{\mathrm{lk}} * \mathrm{LU}_l^{t-1} \tag{5.38}$$

The absolute transitions may enter the carbon accounting (ignored here) and if we substitute the last period's transitions we are back to the condition for consistent land balancing in the final period from above:

$$\mathrm{LU}_k^s = \sum_l P_{\mathrm{lk}} \mathrm{LU}_l^{s-1} = \sum_l T_{\mathrm{lk}}^s \tag{5.39}$$

When using the transition probabilities in the consistency condition for initial land use we obtain

$$\mathrm{LU}_l^{\mathrm{initial}} - \sum_k T_{\mathrm{lk}}^1 = 0 \tag{5.40}$$

$$\Longleftrightarrow \mathrm{LU}_l^{\mathrm{initial}} = \sum_k P_{\mathrm{lk}} \mathrm{LU}_l^{\mathrm{iniital}} \tag{5.41}$$

$$\Leftrightarrow 1 = \sum_k P_{\mathrm{lk}} \tag{5.42}$$

So the simple condition is that probabilities have to add up to one (e_addUpTransMatrix in supply_model.gms). In this form the model is currently implemented in CAPRI.

Outer model – land supply

The outer problem is defined as a maximization of the sum of land rents minus a quadratic cost term, subject to the first order optimality conditions of the inner problem:

$$\max \sum_i \mathrm{LEV}L_i r_i - \sum_i \mathrm{LEV}L_i c_i - \frac{1}{2}\sum_{\mathrm{ij}} \mathrm{LEV}L_i D_{\mathrm{ij}} \mathrm{LEV}L_j \tag{5.43}$$

subject to,

$$\mathrm{LU}_k - \sum_i \mathrm{share}_{\mathrm{ki}} \mathrm{LEV}L_i = 0 \tag{5.44}$$

$$\mathrm{LU}_k - \sum_l T_{\mathrm{lk}} = 0 \ \ [\tau_k] \tag{5.45}$$

$$\mathrm{LU}_l^{\mathrm{initial}} - \sum_k T_{\mathrm{lk}} = 0 \ \ \left[\tau_l^{\mathrm{initial}}\right] \tag{5.46}$$

$$(\alpha_{\mathrm{lk}} - 1)\, T_{\mathrm{lk}}^{-1} - \beta_{\mathrm{lk}} + \tau_k + \tau_l^{\mathrm{initial}} = 0 \tag{5.47}$$

The parameters of the inner model and may be determined as explained in the previous sections. For the outer model, we need to define the parameters **c** and **D**. We have a single data point of land use and land rent for each land use class. Since we have, for N land classes, $N + N(N-1)/2$ parameters, but only N price-quantity pairs (one data point for each land class). This means that without any additional information, we could e.g. calibrate the model exactly by computing the **c** parameter, but have no information left for defining **D**. However, we have at our disposal prior estimates of the regional matrices of land supply elasticities that may be used to define prior densities for the elasticity matrix implied jointly by the **c** and **D** parameters and the inner problem. Another way of expressing this is that we compute a meta parameter matrix $\eta\left(\mathbf{c}, \mathbf{D}, \mathbf{LU}^{\mathrm{initial}}\right)$ that is a function of the real parameters, and use the prior elasticity matrix as a prior for this meta parameter. If cast in this way, the problem becomes a Bayesian econometric estimation.

There are a few methodological and numerical challenges to overcome. In particular, we need to (i) analytically derive $\eta\left(\mathbf{c}, \mathbf{D}, \mathbf{LU}^{\mathrm{initial}}\right)$, and (ii) ensure that the resulting model has the appropriate curvature to ensure a unique interior solution – anything else would result in a rather useless model. We start by simplifying the problem by observing that all the constraints (the first order conditions of the inner problem) can be replaced with an ordinary land constraint:

$$\sum_i \mathrm{LEV}L_i - \sum_l LU_l^{\mathrm{initial}} = 0 \tag{5.48}$$

Note that the second sum is a constant. This simplification is based on the observation that the land transitions don't appear in the objective function of the outer problem, so that all solutions to the inner problems are equivalent from the perspective of the outer problem, and that any land use vector that preserves the initial land endowment is a feasible solution to the inner problem.

Next, we formulate the first order condition (FOC) of the modified outer problem to obtain land use as an implicit function of the parameters, $F\left(LEVL, c, D, LU^{\text{initial}}, r\right) = 0$. We can then use the implicit function theorem to compute the derivative of land supply LEVL_i with respect to land rent r_j, which in turn can be used to define the elasticity matrix η.

The first order conditions, and the implicit function, become

$$F\left(LEVL, \lambda, c, D, LU^{\text{initial}}, r\right) = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial LEVL_i} = & r_i - c_i - \sum_j D_{\text{ij}}\text{LEV}L_j - \lambda & = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = & \sum_i \text{LEV}L_i - \sum_l LU_l^{\text{initial}} & = 0 \end{bmatrix} \tag{5.49}$$

In order to apply the implicit function theorem[8] we need to differentiate the FOC once w.r.t. the variables LEVL_i and λ and once with respect to the parameter of interest, r_j, invert the former and take the negative of the matrix product. If (currently) irrelevant parameter are omitted, the following matrix of $(N+1)\times(N+1)$ is obtained (the "+1" is the uninteresting derivative of total land rent λ with respect to individual land class rent r_i)

$$\left[\frac{\partial LEVL}{\partial r}\right] = -\left[D_{LEVL,\lambda}F(LEVL,\lambda,r)\right]^{-1} D_r F(LEVL,\lambda,r) \tag{5.50}$$

$$\begin{bmatrix} \frac{\partial LEVL}{\partial r} \\ \frac{\partial \lambda}{\partial r} \end{bmatrix} = -\begin{bmatrix} \frac{\partial F}{\partial LEVL} & \frac{\partial F}{\partial \lambda} \end{bmatrix} \begin{bmatrix} \frac{\partial F}{\partial r} \end{bmatrix} \tag{5.51}$$

Carrying out the differentiation specifically for land rent r_j, we obtain:

$$\begin{bmatrix} \frac{\partial LEVL_i}{\partial r_j} \\ \frac{\partial \lambda}{\partial r_j} \end{bmatrix} = -\begin{bmatrix} [-D_{\text{ij}}] & -1 \\ -1' & 0 \end{bmatrix}^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{5.52}$$

Discarding the last row of the resulting $(N+1)\times N$ matrix finally lets us compute the elasticity as

$$[\eta_{\text{ij}}] = \left[\frac{\partial LEVL_i}{\partial r_j}\right]\left[\frac{r_j}{\text{LEV}L_i}\right] \tag{5.53}$$

In the estimation, we assumed that the prior elasticity matrix is the mode of a density where each entry were independently distributed. Furthermore, the off-diagonal or any diagonal elements with negative priors were normally distributed, whereas the diagonal elements with positive priors (as required for a well-behaved curvature) were gamma distributed. For the standard deviation of elasticities we used either information from the prior estimates or some fall-back assumptions on standard deviations relative to the mode of elasticities. Denoting the prior elasticities with e_{ij}, we solved the following optimization problem, where parameters α and β were already estimates as explained in the sections on the inner problem.

[8] Recall that the implicit function theorem states that if F(x,p) = 0, then dx/dp = -[dF/dx]$^{-1}$[dF/dp]

$$\max_{\eta,c,D} \sum_{ij\in normal(i,j)} -\frac{1}{s_{\text{ij}}^2}\left(\eta_{\text{ij}} - e_{\text{ij}}^{\text{jsa}}\right)^2 + \sum_{ij\in gamma(i,j)} \left[(\alpha_{\text{ij}}-1)\log\eta_{\text{ij}} - \beta_{\text{ij}}\eta_{\text{ij}}\right] \tag{5.54}$$

subject to

$$\left[\frac{\partial LEVL_i}{\partial r_j}\right] = -\begin{bmatrix}[-D_{\text{ij}}] & -1\\ -1' & 0\end{bmatrix}^{-1}\begin{bmatrix}I\\ 0\end{bmatrix} \tag{5.55}$$

$$[\eta_{\text{ij}}] = \left[\frac{\partial LEVL_i}{\partial r_j}\right]\left[\frac{r_j}{\text{LEV}L_i}\right] \tag{5.56}$$

$$\begin{aligned} r_i - c_i - \textstyle\sum_j D_{\text{ij}}\text{LEV}L_j - \lambda &= 0\\ \textstyle\sum_i \text{LEV}L_i - \sum_l LU_l^{\text{initial}} &= 0 \end{aligned} \tag{5.57}$$

and the curvature constraint using a stricter variant of the Cholesky factorization

$$D_{\text{ij}}\left(1-\delta I_{\text{ij}}\right) = \sum_k U_{\text{ki}}U_{\text{kj}} \tag{5.58}$$

where δ is a small positive number and I_{ij} entries of the identity matrix such that the factor $(1-\delta I_{\text{ij}})$ shrinks the diagonal of the D-matrix, ensuring *strict* positive definiteness instead of *semi*-definiteness. We used $\delta = 0.05$. Furthermore, the Lagrange multiplier of the total land constraint, λ, was fixed at the weighted average of the rents r_i, i.e. $\lambda = \frac{\sum_i \text{LEV}L_i r_i}{\sum_i \text{LEV}L_i}$. Without the latter assumption, the parameters c and D are not uniquely identified.

Prior elasticities and area mappings

The empirical evidence obtained in the TRUSTEE project applied to prior elasticities for land categories based on Corine Land Cover (CLC) data. These categories are also covered in the CAPRI database based on various sources (see the database section in the CAPRI documentation):

The introduction has mentioned already three systems of area categories that need to be distinguished. The first one is the set of area aggregates with good coverage in statistics that has been investigated recently by JS Ay (2016), in the following "JSA":

$$\text{LEVL} = \{\text{ARAC},\ \text{FRUN},\ \text{GRAS},\ \text{FORE},\ \text{ARTIF}, \text{OLND}\} \tag{5.59}$$

Where

ARAC = arable crops

FRUN = perennial crops

GRAS = permanent grassland

FORE = forest

ARTIF = artificial surfaces (settlements, traffic or industrial)

OLND = other land

The above categories are matching reasonably well with the definitions in JSA. A mismatch exists in the classification of paddy (part of ARAC in CAPRI but in the perennial group in JSA) and terrestrial wetlands (part of OLND in CAPRI and a separate category in JSA). Inland waters are considered exogenous in CAPRI and hence not included in the above set LEVL.

For carbon accounting we need to identify the six LU classes from IPCC recommendations and official UNFCCC reporting:

$$LU = \{\mathrm{CROP},\ \mathrm{GRSLND},\ \mathrm{FORE},\ \mathrm{ARTIF}, WETLND, RESLND\} \tag{5.60}$$

which is typically indexed below with “l” or “k”   LU and where

CROP = crop land (= sum of arable crops and perennial crops)

GRSLND = grassland in IPCC definition (includes some shrub land and other “nature land”, hence GRSLND>GRAS)

WETLND = wetland (includes inland waters but also terrestrial wetlands)

RESLND = residual land is that part of OLND not allocated to grassland or wetland, hence RESLND<OLND

FORE = forest

ARTIF = artificial surfaces

In the CAPRI database, in particular for its technical base year, we have estimated an allocation of other land OLND into its components attributable to the UNFCCC classes GRSLND,WETLND, and RESLND:

$$\mathrm{OLND}^0 = \mathrm{OLND}G^0 + \mathrm{OLND}W^0 + \mathrm{OLND}R^0 \tag{5.61}$$

Lacking better options to make the link between sets LEVL (activity level aggregates) and LU (UNFCCC classes, technically in CAPRI code: set “LUclass”) we will assume that these shares are fixed and may estimate the “mixed” LU areas from activity level aggregates as follows

GRSLND	=	*GRAS + OLND OLNDG^0/OLND^0*
WETLND	=	*INLW + OLND OLNDW^0/OLND^0*
RESLND	=	*OLND OLNDR^0/OLND^0*

which means that the mapping from set LEVL to set LU only uses some fixed shares of LEVL areas that are mapped to a certain LU:

$$LU_k = \sum_i \mathrm{share}_{\mathrm{i,k}} \mathrm{LEVL}_i \tag{5.62}$$

where 0 $share_{i,k}$ 1.

Technical implementation

The key equations corresponding to the approach explained above are collected in file supply_model.gms or the included files supply/declare_calibration_models_for_luc.gms and supply/declare_calibration_models_for_land_supply.gms. The declarations of parameters, variables, equations, models and even some sets only used in the calibration given in these files are included by the "supply_model.gms" only if "BASELINE==ON" or if it was a CAPREG base year task that was carried out. Loading of priors, initialisation of parameters and variables for the calibration as well as the organisation of solve attempts are handled in new sections of file "cal_land_nests.gms", in turn called by the gams file "prep_cal.gms". This implies that the land supply and land use change calibrations were inserted before the ordinary calibration of the supply models.

The new land supply specification is only activated if the global variable %trustee_land%==on which may be set via the CAPRI GUI. In order to store the results of the calibration in a compact way that is compatible with the existing code, the existing parameter files "pmppar_XX.gdx" was used. The parameters of the land supply functions, called "c" and "D" above, were stored on two parameters "p_pmpCnstLandTypes" and "p_pmpQuadLandTypes". As a new symbol (p_pmpCnstLandTypes) is introduced in an existing file, the first run of CAPRI after setting %trustee_land%==on may give errors if the file exists already but has been used with the previous land supply specification before. In this case it helps to delete or rename the old pmppar files.

At this point, it should also be explained that rents for non-agricultural land types were entirely based on assumptions (a certain ratio to agricultural rents). As there were no plans to run scenarios with modified non-agricultural rents, these land rents r used in calibration for those land types were subtracted from the "c-paramter", so that it is implicitly stored in p_pmpCnstLandTypes and enters the objective function through the PMP terms. This requires changes if the rents shall be modified or if non-agricultural production shall be included in some simplified form.

Furthermore, the class Inland Waters (INLW) was given a special treatment: it is supposed to be entirely exogenous. For this purpose the special acronym "exogenousLandSupply" was introduced, and stored on the p_pmpCnstLandTypes and used to trigger an equation "e_exogenousLand " in the supply model setting the variable to a constant. In that way, the fixity of INLW (or any land type, should it happen) is stored in the pmp terms and cannot be "forgotten".

More detailed explanations on the technical implementation are covered elsewhere, for example in the "Training material" included in the EcAMPA-4 deliverable D5.

5.3 Premium module

5.3.1 Overview

For the European Union, the CAPRI programming models cover in rich detail the different coupled and de-coupled subsidies of the so-called first Pillar 1 of the CAP, as well as major ones from Pillar 2 (i.e., Less Favoured Area support, agri-environmental measures, Natura 2000 support). The interaction between premium entitlements and eligible hectares for the Single Farm Payment (SFP) of the CAP is explicitly considered, as are the different national SFP implementations, possibly remaining coupled payments (previously under article 68 of Council Regulation (EC) No. 73/2009, from 2014 as “Voluntary Coupled Support”).

Decoupled payments – as with other premium schemes of the CAP from the present and past – are simulated in CAPRI relatively closely to their definition in existing legislation. The rather high disaggregation of the model template regarding production activities and the resolution by farm types inside of NUTS 2 regions clearly eases that task. Currently, 260 different voluntary coupled support schemes are implemented, in addition to decoupled income support (Basic Payment Scheme, BPS).

The payments – both in reality and in the model – tend to be defined in a cumulative manner. In 1992, the direct payments were introduced based on the previous price support levels multiplied by regional historic reference yields. The payments therefore became regionally differentiated. In the subsequent decoupling under Agenda 2000 and the following reforms, the single farm payments were defined based on the payments that each farm had previously received. The payments got very different expressed per hectare for farms in high yield regions versus low yield regions and for farms that had held animal payments versus arable farms. Then, the 2013 reforms introduced convergence of payment rates, both across farms (internal convergence) and between member states (external convergence). CAPRI reflects these incremental reforms, so that the payment rates in the most recent reform are based on coupled payments from MacSharry times.

Given that each reform has added complexity to the previous system, so too has the premium module of CAPRI grown to maintain the capacity to model previous reforms while allowing novel features such as greening to be introduced. Nevertheless, two generic features can be distinguished, that form the basis for many payments: the basic concept of a premium payment, and the idea of payment entitlements. Those are treated in separate sections here. Then, we proceed to describe the incremental reforms of the first pillar of the CAP, starting with the most recent, and the implementation of selected second pillar payments. Finally, we discuss certain elements relating to the reporting of premium payments, most notably the financing over different budgets.

5.3.2 Basic concept

In the CAPRI supply module, premiums are always paid per activity level (per hectare or per animal) basis. They can be differentiated by the low and high yield variant of each crop activity. The premiums are calculated in the premium module from different premium schemes.

A premium scheme (such as DPGRCU for the Grandes Cultures premiums after the Fischler reform) is a logical entity which encompasses:

1. A specific application type (defining the basis for the payment amount)
2. a region or regional aggregate to which it is applied,
3. Possible ceilings in entitlements (CEILLEV) and in value (CEILVAL)
4. Payment rates for possibly several lists of activities (such as PGGRCU for all types of Grandes Cultures or PGPROT for protein crops).
5. Optionally an indication of the marginal payment when a ceiling is reached
6. Optionally a modifier for different amounts per technology

The schemes provide many-to-many mappings between policy instruments and agricultural activities: each scheme can apply to many different activities – with possibly differentiated rates – and each activity can draw support from different schemes.

The application type defines how the nominal amount (called PRMR) is applied. Currently, the following application types are supported:

- perLevl = per ha or head
- perSlgtHd = per slaughtered head
- perYield = per unit of main output
- perHistY = per historic yield
- perLiveStockUnit = per livestock unit
- noDirPay = Norwegian direct payment
- noPriceSup = Norwegian price support

The application type points to a factor by which the nominal amount PRMR (for PRemiuM in Regulation) is converted to a declared value per hectare or head (PRMD). For perLevl, the factor is unity (it is already per hectare), but for instance for perYield, the amount is interpreted as a payment per unit of main output. That is used for the Nordic Aid Scheme for dairy cows in the northmost parts of Europe and for coupled payments in Norway.

Each payment is defined for one or seveal *groups* of activities, functioning as lists of eligible activities. Additionally, the payment can be applied in different rates to the high and low yield variant, to model e.g. an extensification premium.

Each premium scheme also has up to two ceiling values:

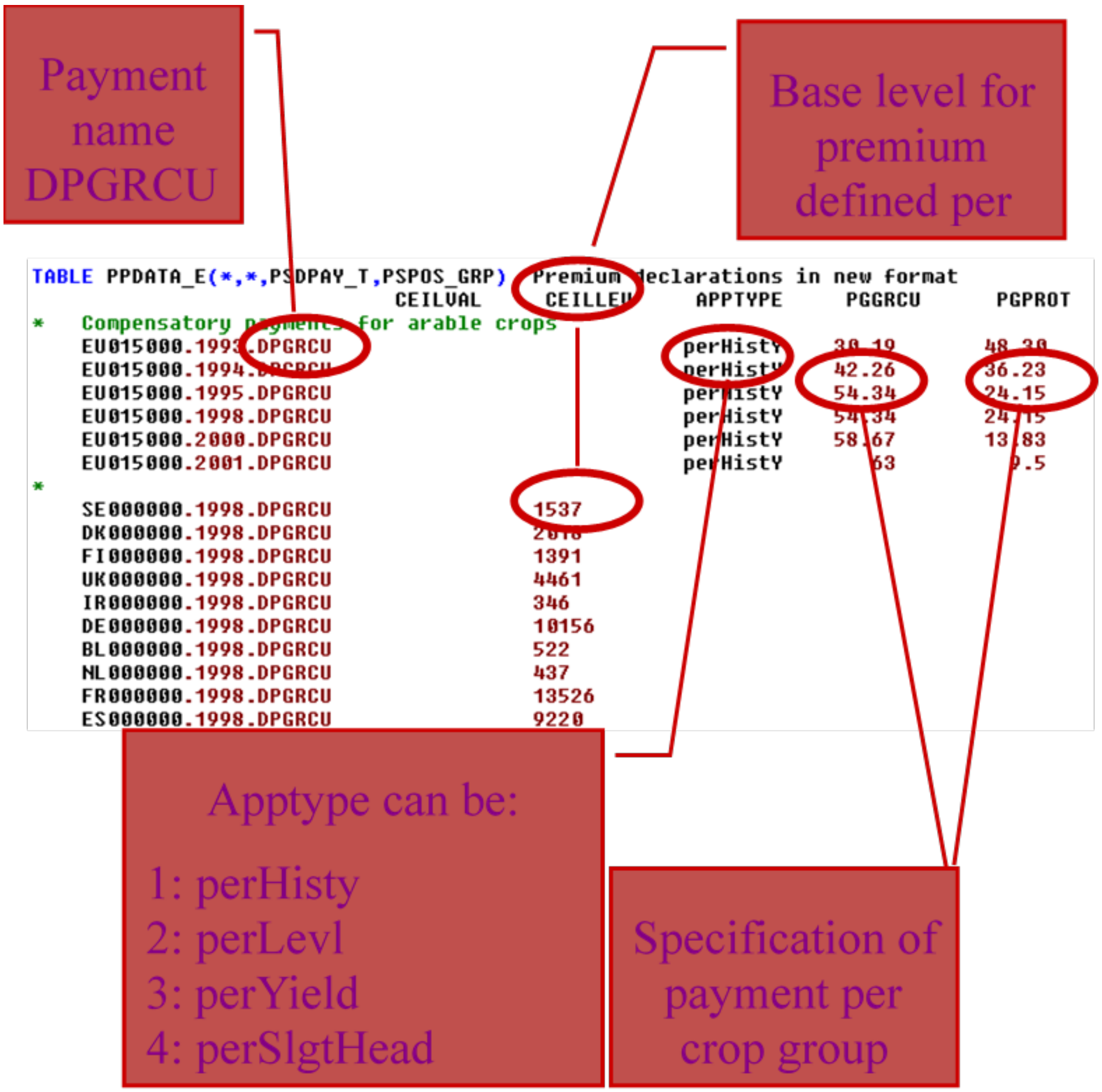
- ceilLev = Ceiling on LEVL, i.e. the number of hectares or heads
- ceilVal = Ceiling on the total budget (envelope) spent on the scheme

In the basic setting, the ceilings work as the old Grandes Cultures payment: if the total quantity (hectares or amount) exceeds the ceiling, then the payment to each farmer is reduced so that the ceilings are respected. This means that the marginal payment is somewhat reduced but does not become zero. For some other schemes, such as the Basic Payment Scheme of the CAP 2014-2020, there is a hard limit on the number of payment entitlements, so that the marginal payment becomes zero if the ceiling is overshot. That behaviour can be triggered by including the payment scheme set element in a special set *PSDPAY_cutEndog*, as in the following example from "pol_input/mtr_until2013.gms".

PSDPAY_cutEndog("DPSAPS") = YES;
PSDPAY_cutEndog("DPREG") = YES;
PSDPAY_cutEndog("DPFRMS") = YES;
PSDPAY_cutEndog("DPFRMF") = YES;
PSDPAY_cutEndog("DPGREEN") = YES;

The following figure shows the technical implementation at an example:

Figure 14: Example of technical implementation of a premium scheme



Source: CAPRI Modelling System. Note: The parameter PPDATA_E is now called p_premDataE.

The sets of payments, exemplified by DPGRCU in the figure, and the activity groups, exemplified by PGGRCU and PGPROT are defined in the file policy/policy_sets.gms. Since this is a "static" GAMS file used in any simulation, it contains the gross list of all policies that currently can be simulated, including legacy ones. In order to work efficiently with the acronyms which define the application types, these are converted to numerical attributes as shown below (*'policy/policy.gms')*:

```
*
*   --- convert acronyms temp. to numerical values so that the SUM() operator can be used
*       (the p_premPaym "APPTYPE" position is deleted later on, so we use in here the numerical values
*        The acronyms are not replaced all over the code as they ease editing the tables)
*
*
    p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") $ (   (p_premPaym(RSEU,PPPACT,PSDPAY,"PRMR") gt eps) $ PSDPAY_T_A(PPPACT,PSDPAY))
          =   1 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq perLevl)
            + 2 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq perSlgthd)
            + 3 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq perYield)
            + 4 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq perHistY)
            + 5 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq perLiveStockUnit)
            + 6 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq noDirPay)
            + 7 $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq noPriceSup);
```

CAPRI also provides the possibility to incentivise extensification or intensification via the payments. Most production activities come in technological variants, by default one higher yielding and one lower yielding one, and those variants can be eligible to different rates of premium payments. This is used for instance in the implementation of agri-environmental schemes in the file policy/rd_logic.gms as shown in the figure below. The parameter p_technFact is the standard coefficient that modifies the technology of the production activities in CAPRI. In the figure below, the two statements change the rate of premium payments for the set of currently active regions (rs), for all model activities (MPACT), for all agri-environmental schemes (psdpay_ae) with different rates for technology T1 (high yield) and T2 (low yield) in the case where T2 exists. +0.5 for T2 means that the premium payment in the model becomes the nominal rate times (1 + 0.5), i.e. 50% higher, whereas the -0.5 for T1 means that the premium payment in the model becomes the nominal rate times (1 – 0.5), i.e. 50% lower. This approximates the stylized fact that agri-environmental schemes, which in reality consist of a wide range of measures, in general favour extensive technologies (see section on Pillar II payments below).

```
*   2) For crops with alt. technologies: increasing the premium for T2 and reduce it for T1.
*   It will no longer fit the ceiling exactly, because different payments for T1, T2 are not
*   foreseen in prmcut elsewhere.


 p_technFact(rs,MPACT,psdpay_ae,"T2") $ p_technFact(rs,MPACT,"LEVL","T2") =  0.5;
 p_technFact(rs,MPACT,psdpay_ae,"T1") $ p_technFact(rs,MPACT,"LEVL","T2") = -0.5;
```

The general flow of logic inside of CAPRI (inside the model file capmod.gms) as regards premiums is shown in the following figure. The process starts by loading baseline data, including calibrated behavioural parameters. That data set represents an equilibrium situation for the policy (premiums) that were used in the baseline generation process.

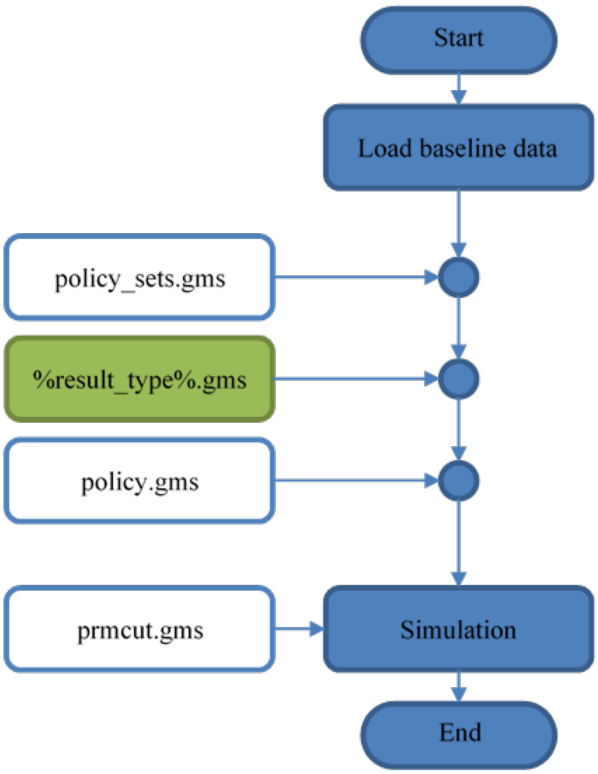
After loading data, the file with declarations of all available premium schemes et cetera (policy_sets.gms) is loaded. The particular policy to use in the present simulation is contained in the *policy file* with the name defined by the placeholder (environment/macro variable) %result_type%. This string will also be part of the file name used for the simulation results.

The premiums defined in the policy file is processed by the file policy.gms. That processing implies a translation of the regulation-like definitions used in the policy file to parameters useful for CAPRI.

Within the simulation algorithm itself, a special file called prmcut.gms is called repeatedly, with the purpose to cut effective premium rates paid in case any ceiling is overshot, so that budgets are respected.

Generally, all attributes for a premium scheme are mapped down in space, e.g. from EU27 to EU 27

Figure 15: General flow of logic of CAPRI model as regards premiums



Source: own illustration

member states, from countries to NUTS1 regions inside the country, from there to the NUTS2 regions inside the NUTS1, and from NUTS2 regions to the farm types in a NUTS2 region (see *'policy/policy.gms'*), e.g.

```
p_premPaym(MS,PPPACT,PSDPAY,"PRMR") $ ((      (NOT p_premPaym(MS,PPPACT,PSDPAY,"PRMR"))
                                          AND p_dataCopy(MS,PPPACT,"LEVL","Y")) $ PSDPAY_T_A(PPPACT,PSDPAY))
     = SUM(R_RAGG(MS,REGEU1015), p_premPaym(REGEU1015,PPPACT,PSDPAY,"PRMR"));
```

In order to map the premium rate as defined in a legal text into one paid out on a per-activity basis, the relevant activity based attribute matching the application type is set to a premium modification factor ("Ap_premModfFactT") as shown below:

The actually declared premium per activity unit (ha, [1000] [slaughtered] heads) is then the multiplication of the premium rate and that modification factor. For crops, the unit of the resulting entries are current € per ha, for animal, it depends on the exact definition of the activity level (per [1000] [slaughtered] heads).

```
*
*   --- calculate application factor
*
    p_premPaym(RSEU,PPPACT,PSDPAY,"Ap_premModifFactT") $ ( (p_premPaym(RSEU,PPPACT,PSDPAY,"PRMR") GT eps) $ PSDPAY_T_A(PPPACT,PSDPAY))
      =
*     --- apptype perLevl: fixed premium per ha or head
           1                      $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 1)
*     --- apptype perSlgtHd: slaughterings
         + 1                      $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 2)

*     --- apptype perYield: main yield
         + (sum(PACT_TO_Y(PPPACT,O), p_dataCopy(RSEU,PPPACT,O,"Y")) * 0.001)
                                  $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 3)
*     --- apptype perHisty: historic yield
         + (p_dataCopy(RSEU,PPPACT,"HSTY","Y") + p_dataCopy(RSEU,"CERE","HSTY","Y") $ (NOT p_dataCopy(RSEU,PPPACT,"HSTY","Y")))
                                  $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 4)

*     --- apptype perLiveStockUnit: Live stock unit
         + p_LUUnits(pppact)       $ (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 5)

$iftheni %MS_NO% == ON

*     --- apptype NODirPay: factor for Norwegian direct payments
         + p_norwPremPact(RSEU,%1,PSDPAY,PPPACT)     $ (R_RAGG(RSEU,"NO000000") AND (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 6))

*     --- apptype NOPriceSup: factor for Norwegian price support * yield
         + sum(PACT_TO_Y(PPPACT,O), p_dataCopy(RSEU,PPPACT,O,"Y") * 0.001 *
             p_norwPremOutput(RSEU,%1,PSDPAY,O))    $ (R_RAGG(RSEU,"NO000000") AND (p_premPaym(RSEU,PPPACT,PSDPAY,"APPTYPE") eq 7))

$endif
*
*     --- expost policy for new Member States (subsidies per unit of output derived from Economic Accounts)
*
         + sum(PACT_TO_Y(PPPACT,O), p_dataCopy(RSEU,PPPACT,O,"Y") * 0.001 * p_nonCAPPremOutput(RSEU,O,%1)) $ SAMEAS(PSDPAY,"DPNONCAP");
```

```
  p_premPaym(RSEU,PPPACT,PSDPAY,"PRMD") $ PSDPAY_T_A(PPPACT,PSDPAY)
      =    p_premPaym(RSEU,PPPACT,PSDPAY,"PRMR")
         * p_premPaym(RSEU,PPPACT,PSDPAY,"Ap_premModifFactT") $ (p_premPaym(RSEU,PPPACT,PSDPAY,"PRMR") GT eps);
```

These declared rates can hence be aggregated to higher regional units using the activity levels as weights, e.g. from farm types to NUTS2:

```
 p_premPaym(NUTS2Agg,PPPACT,PSDPAY,"PRMD") $ (( (p_dataCopy(NUTS2Agg,PPPACT,"LEVL","Y") gt eps)) $  PSDPAY_T_A(PPPACT,PSDPAY))
  = sum( Types_to_R(NUTS2Agg,TYPES) $ p_premPaym(Types,PPPACT,PSDPAY,"PRMD"),
          p_premPaym(Types,PPPACT,PSDPAY,"PRMD")*p_dataCopy(Types,PPPACT,"LEVL","Y"))/p_dataCopy(NUTS2Agg,PPPACT,"LEVL","Y");
```

Before the supply module is started between iterations, the current activity levels and premiums paid out are summed up for each scheme and regional level where ceilings in levels or value are defined. If one of the aggregated sums exceeds the ceilings, all premium rates for the scheme are cut proportionally to fit under the tighter of the two envelops:

```
 p_premData(RSEU,PSDPAY,"CEILCUT") $ ( ((p_premData(RSEU,PSDPAY,"CEILLEV") GT eps) OR (p_premData(RSEU,PSDPAY,"CEILVAL") GT eps)) $ PSDPAY_T_R(RSEU,PSDPAY))
     = 1/MAX(1, MAX( (p_premData(RSEU,PSDPAY,"CEILACL") / p_premData(RSEU,PSDPAY,"CEILLEV")) $ (p_premData(RSEU,PSDPAY,"CEILLEV") gt eps)
                   ,(p_premData(RSEU,PSDPAY,"CEILACV") / p_premData(RSEU,PSDPAY,"CEILVAL")) $ (p_premData(RSEU,PSDPAY,"CEILVAL") gt eps)));
```

From the declared rates and these cut factors, the actually paid premiums are defined:

```
 p_premPaym(RUNR,PPPACT,PSDPAY,"PRME") $ ( (p_dataCopy(RUNR,PPPACT,"LEVL","Y") $ p_premPaym(RUNR,PPPACT,PSDPAY,"PRMD")) $ PSDPAY_T_A(PPPACT,PSDPAY))
    = p_premPaym(RUNR,PPPACT,PSDPAY,"PRMD") * p_premPaym(RUNR,"CUTT",PSDPAY,"CEILCUT")
```

The indivudal premiums from each premium scheme are then added up to arrive at one average rate for each activity which enters the objective function of the supply model, the data base and post-model reporting:

```
*
*   --- Aggregate payments to one PRME and PRMD for each activity
*
    DATA(RUNR,PPPACT,"PRME","Y") $ p_dataCopy(RUNR,PPPACT,"LEVL","Y")
       = sum(PSDPAY_T_R(RUNR,PSDPAY), p_premPaym(RUNR,PPPACT,PSDPAY,"PRME"));
```

5.3.3 An example of a payment with a ceiling

We explain the different elements and steps in the following based on an example of the slaughter premium for adult cattle of 80 EURO per slaughtered head in Latvia, defined in 2004. The following screen shot comes from the policy file gams/pol_input/mtr_until2013.gms, with some lines hidden.

	CEILVAL	APPTYPE	PGARAB	PGSCOS	PGBULF	PGHEIF	PGCALV	PGSHGM	PGPARI	PGTOBA	PGDCOW	PGGRAS	PGMEAT
* (...)													
LV000000.2004.DPCN_ADCT	9.946	perSlgtHd											80.00

1. The application type defines the criterion upon which the payment depends, in the case of the slaughter premium it is defined per slaughtered head.
2. The regulation premium rate (PRMR) is the default (maximum, uncut) amount of the premium according to regulatory texts, for all activities coverd by the premium group (here PGMEAT) and regions for which the premium is defined. In the example, this means that it is 80 EURO for the group of activities PGMEAT, which is dairy cows, suckler cows, male adult cattle and fattened heifers, in Latvia (LV000000). This is defined in a hierarchical way: if it is set to 80 EURO for the EU and not set at all at lower regional level, the 80 EURO are mapped down to all sub regions by the program. The program also lets you define groups of activities that are linked to the premium. In this case a group PGMEAT has been defined which contains the relevant animals (set s_PSGRP(*) in the file *'policy_sets.gms'*).
3. The declared amount in the activity definition of CAPRI (per ha, per head, per 1000 heads) way that the amount PRMR should be applied or declared in CAPRI is called declared premium (PRMD) and applies per head or hectare. In our example, the regulation says that 80 EURO should be paid when the animal is slaughtered. That means that in order to get the amount per living animal and year, the 80 EUROs have to be multiplied by the frequency with which the animal is slaughtered. For male beef cattle it is 1/year whereas it for dairy cows is something like 1/5 years. These numbers come from the CAPRI database.
4. Regional ceiling, expressed in maximum number of premiums paid and/or total payment in EURO. In the example with the slaughter premiums, this is used to set a national ceiling limiting the total amount spent on slaughter premiums to 9.946 million euro. There can be additional ceilings at other regional levels, and the most strongly binding is always the one that limits payments.

Those four pieces of information are generally easily accessible without further processing from the regulatory texts. Starting with PRMR and APPTYPE (information pieces 1 and 2 above), it is possible to calculate (3), PRMD, the amount of premium per head or hectare that would be paid if there were no (active) ceiling. These preparatory calculations, e.g. the hierarchical break down from higher to lower regional level and from activity groups to individual activities, as well as the calculations of PRMD from PRMR (using APPTYPE) is carried out in a file called '*policy/policy.gms*' as shown above.

For most premiums in CAP there are ceilings, which if they are binding decrease the average amount of premiums actually paid (effective premium, PRME) per head or hectare. As discussed, due to the different kind of ceilings, the reduction of premiums and the treatment of PRME can only be done endogenously during the simulations depending on the simuled production patterns.

How is this problem solved in CAPRI? The effective premium (PRME) is exogenous during the optimisation of the supply model[9], but adjusted iteratively between the main model iterations. So, for most premium schemes, the premium level is constant in the objective function and hence the model does not realise that the marginal premium payment is zero as soon as the ceiling is reached. Technically, the iterative adjustment of the effective premiums PRME is handled in a file called '*policy/premcut.gms*' for "premium cut". That reasoning is correct as long as the ceiling is not farm specific.

In each iteration, once all regional model are solved, the program adds up total number of premium units (hectares or heads for which it is paid) that belong to each ceiling. In most cases this simply means summing up number of animals or hectares of the activities for which each premium applies. This is also multiplied with the declared amount PRMD to get the total payment which would be paid if it would not be cut. For each premium this is compared to the ceilings defined (total level with the level ceiling and total amount with the value ceiling) and a "cut factor" is calculated, which defines how much the premium has to be reduced in order to fit under all ceilings. Then PRMD is multiplied by this factor to get the effective premium (PRME) for the next iteration.

5.3.4 Pillar I

The MTR-reform and the health check On 26 June 2003, EU farm ministers adopted a further fundamental reform of the Common Agricultural Policy (CAP). The central element of the 2003 CAP reform was the introduction of the so-called single payment scheme (SPS). The SPS is based on payments entitlements linked to eligible land, but decoupled from production. However, to avoid abandonment of production, Member States could still choose to maintain a limited link between subsidy and production under well defined conditions and within clear limits. Moreover, these new "single farm payments" would be linked to environmental, food safety and animal welfare standards and obligations.

Key elements of the 2003 CAP reform were:

- A single farm payment for EU farmers, independent from production; limited coupled elements may be maintained to avoid abandonment of production;
- land receiving payments should be kept in good agricultural and environmental condition (G.A.E.C). "Good agricultural condition is generally interpreted to mean that the land will not be abandoned and environmental problems such as erosion will be avoided" this requirement could be interpreted as re-establishing the link between the payment and the factors of production employed (land management practices) and ultimately current production; some form of management of the land should be maintained;
- entitlements are tradable within the EU member states (not among them) but certain limitations are imposed (Ciaian, Kancs and Swinnen, 2010). For example, in the Netherlands, entitlements can be transferred among farmers only when the farmer has land without entitlements;
- areas already under permanent pasture should must remain so; in practise, certain reductions at regional level were accepted before Member States would be forced to interact.

[9] There are exemptions for that rule, see below for the section on entitlements.

- a strengthened rural development policy based on expanded EU budget outlayswith more EU money, new measures to promote the environment, quality and animal welfare and to help farmers to meet EU production standards starting in 2005,
- a reduction in direct payments ("modulation") for bigger farms to contribute to finance the new rural development policy,
- a mechanism for financial discipline to ensure that the farm budget fixed until 2013 is not overshot,
- revisions to the market policy of the CAP:
  - asymmetric price cuts in the milk sector: The intervention price for butter will be reduced by 25% over four years, which is an additional price cut of 10% compared to Agenda 2000, for skimmed milk powder a 15% reduction over three years, as agreed in Agenda 2000, is retained,
  - reduction of the monthly increments in the cereals sector by half, the current intervention price will be maintained,
  - reforms in the rice, durum wheat, nuts, starch potatoes and dried fodder sectors.

In implementing the SPS, member states (MS) could opt for a historical model (payment entitlements based on individual historical reference amounts per farmer), a regional model (flat rate payment entitlements based on amounts received by farmers in a region in the reference period) or a hybrid model (mix of the two approaches, either in a static or in a dynamic manner). An overview of the implementation of direct payments under the CAP in the different MS can be found at `http://ec.europa.eu/agriculture/markets/sfp/ms_en.pdf`.

Denmark, Germany, Luxembourg, Finland, Sweden, England and Northern Ireland applied a hybrid model. The remaining MS implemented the historical model. From 2007 onwards, dairy payments will be decoupled from production and included in the single payment scheme in all MS.

Although the intention of the CAP reform 2003 is to decouple payments, some payments were not included. In particular the crop specific payment for protein crops, 60% of the payment for starch potatoes, 42% of the payment for rice, the quality premium for wheat and the area payment for nuts. Market organizations for commodities not included in the reform also remained in place. For sugar this changed by the end of 2005 as a reform of the sugar market was decided upon by the EU ministers of Agriculture. The reform included a reduction of the administrative price levels of sugar and sugar beet by with 36%, the introduction of a compensation payment for sugar beet farmers, a premium scheme for the termination of sugar production at factory level (what is referred to as the 'restructuring scheme') and the opportunity to purchase quota sugar. The compensation for sugar beet farmers will be included in the single payment scheme.

Until the end of 2007 for several fresh and processed fruit and vegetables coupled payments were given. Since 2008 fruit and vegetables are decoupled and land covered by fruit and vegetables is eligible for payment entitlements under the decoupled aid scheme which applies in other farm sectors (EC, 2007). All existing support for processed fruit and vegetables will be decoupled and the national budgetary ceilings for the SPS will be increased accordingly.

The last step of the EU CAP reform dates from 20 November 2008 when EU agriculture ministers reached a political agreement on the so-called Health Check (HC) of the CAP. Among a range of measures, the agreement abolishes arable set-aside, increases milk quotas gradually leading up to their abolition in

2015, and converts market intervention into a genuine safety net (EC, 2009). Ministers also agreed to increase modulation, whereby direct payments to farmers under the SPS are reduced and the money transferred to the Rural Development Fund. This should allow a better response to the new challenges and opportunities faced by European agriculture, including climate change, the need for better water management, the protection of biodiversity, and the production of green energy. Member States will also be able to assist dairy farmers in sensitive regions to adjust to the new market situation.

Under the HC of the CAP it was decided that remaining coupled payments should be decoupled and moved into the Single Payment Scheme (SPS), with the exception of suckler cow, where Member States may maintain current levels of coupled support. Moreover, member states are allowed to review the decision taken on the decoupling of fruit and vegetables in 2007, provided that it results in lower coupled payments. For soft fruits transitional support will continue until 31st December 2011 and be converted into decoupled payment as of 2012 (EC, 2009). Before the HC Member States could retain by sector 10 percent of their national budget ceilings for direct payments for use for environmental measures or improving the quality and marketing of products in that sector (Article 68/69' measures: Assistance to sectors with special problems). Under the HC this possibility will become more flexible. The money will no longer have to be used in the same sector; it may be used to help farmers producing milk, beef, goat and sheep meat and rice in disadvantaged regions or vulnerable types of farming; it may also be used to support risk management measures such as insurance schemes for natural disasters and mutual funds for animal diseases; and countries operating the Single Area Payment Scheme (SAPS) system will become eligible for the scheme (EC, 2009).

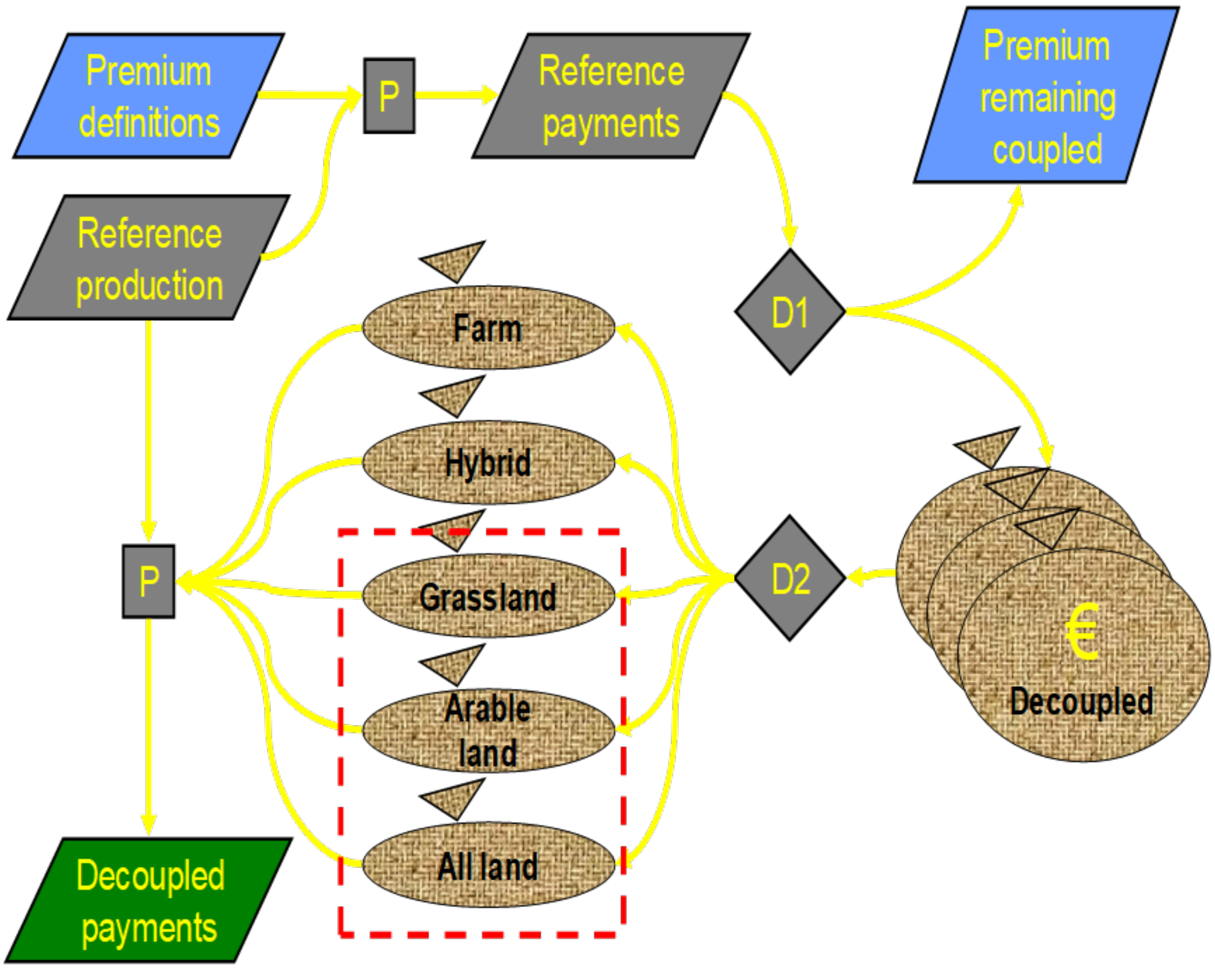
Money for Article 68/69' measures increases as currently unspent money can be used for these measures as well. The Rural Development budget money increases as money shifted away from direct aid based on expanding (modulation) increases to 10% of the direct aid. The additional funding for Rural Development obtained this way may be used by Member States to reinforce programs in the fields of climate change, renewable energy, water management, biodiversity, innovation linked to the previous four points and for accompanying measures in the dairy sector. This transferred money will be co-financed by the EU at a rate of 75 percent and 90 percent in convergence regions where average GDP is lower. Finally for our purposes it is important to mention that under the HC of the CAP a series of small support schemes will be decoupled and shifted to the SPS from 2012. The energy crop premium will be abolished.

In CAPRI, The different implementations of the single farm premium (SFP) introduced with the so-called Mid Term Review of the CAP apply the same logic as for the payment schemes. To give an example, the regional implementation is called DPREG, and might have different payment rates for arable crops (PGARAB) and grass lands (PGGRAS).

The general way these premiums are introduced in CAPRI is shown below. From a reference situation (expost statistical data) and the premiums valid at that time, it is first determined (Decision D1) how much of each existing payments in each scheme are continued to be payed as a coupled scheme, and how much is going into envelops for different types of decoupled payments (part of decision D2). That envelop can be at farm, regional or Member State level and can be implemented in different ways (e.g. historic implementation, regional, regional with different payment rates to arable or grass lands).

In opposite to the reforms until Agenda 2000, there are hence in most cases not longer premium rates or individual ceilings in hectares found in legal texts. Rather, these are calculated by the model itself from the decoupled part of the "old" Mac Sharry and Agenda 2000 premiums which introduces additional complexity in the model code.

Figure 16: General way of SFP implementation in CAPRI



Source: own illustration

Only an overall budget envelop is given covering all pillar I premiums of the EU CAP ("old" MacSharry and Agenda 2000 premiums, SPS premiums, article 63/68/69 premiums, etc.) per Member State nad per year on the position p_premDataE(MS,SIMY,"DPMTR","CEILVAL") in '*pol_input/mtr_hc.gms*'. Here MS refers to member states and SIMY to a certain year.

Single area payment scheme (SAPS)

The MS who joined the EU since 2004 could choose to apply the single area payment scheme (SAPS), a simplified area payment system, for a transitory period until end 2010 or to apply the same system as in the EU-15 immediately. The most important difference between the SPS and the SAPS is that the entitlements under the SPS can be transferred between farms.

From a technical viewpoint, the single-area premium scheme (SAPS) is the easiest to implement:

```
TABLE p_premDataE(*,*,PSDPAY_T,PSPOS_GRP)  Premium declarations in new format
*
                              CEILLEV    CEILVAL     APPTYPE   PGSAPS
*
    CY000000.2004.DPMTR        140        9.96       perLevl      71
    CZ000000.2004.DPMTR       3469      198.94       perLevl      57.35
    EE000000.2004.DPMTR        800       21.40       perLevl      26.75
    HU000000.2004.DPMTR       4355      305.81       perLevl      70.22
    LV000000.2004.DPMTR       1475       30.48       perLevl      20.66
    LT000000.2004.DPMTR       2288       82.07       perLevl      35.87
    PL000000.2004.DPMTR      14843      659.95       perLevl      44.46
    SK000000.2004.DPMTR       1955       85.72       perLevl      43.85
*
*   Reg. 583/2004
*

    CY000000.2005.DPMTR        140        8.90       perLevl      74.29
    CZ000000.2005.DPMTR       3469      228.80       perLevl      65.70
    EE000000.2005.DPMTR        800       23.40       perLevl      29.25
    HU000000.2005.DPMTR       4355      350.80       perLevl      80.55
    MT000000.2005.DPMTR                   0.67       perLevl
    LV000000.2005.DPMTR       1475       33.90       perLevl      22.98
    LT000000.2005.DPMTR       2288       92.00       perLevl      40.21
    PL000000.2005.DPMTR      14843      724.60       perLevl      48.80
    SI000000.2005.DPMTR                  35.50       perLevl
    SK000000.2005.DPMTR       1955       97.70       perLevl      49.92
*
*   Commission Regulation (EC) No 552/2007
*
    BG000000.2007.DPMTR       5237      202.097      perLevl      46.29
    RO000000.2007.DPMTR      14606      440.635      perLevl      36.15
    ;
```

As it defines a flat rate premiums per ha of agricultural land. The ceilings in values and thus the application rates per ha are step wise increased over time:

To reach their full level in 2013 (EU 10) or 2016 (Bulgaria and Romania).

During that transition period where not yet the full EU premiums were paid out, the Member States had the right to paid up to certain limits to so-called complementary national direct payments (the list of schemes used in CAPRI was shown above). They also edited in a tabular format:

These top-ups have to be reduced towards the end of the period where the the Pillar I premiums are phased in:

Non-SAPS implementation

The non-SAPS implementation of the Mid-Term Review package is far more demanding. First of all, the countries could, at least in the earlier years of the reform, keep certain percentages of specific premium scheme still coupled to production. These coupling factors are stored on the parameter p_couplPercent_E:

The amount of payments which is not kept coupled is then paid out to different implementations of the MTR:

- Regional implementation where all arable crops (PGARAB)

```
*
*   --- assume a linear increase of SAPS over time
*
    p_premDataE(MS,"2005","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  30;
    p_premDataE(MS,"2006","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  35;
    p_premDataE(MS,"2007","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  40;
    p_premDataE(MS,"2008","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  50;
    p_premDataE(MS,"2009","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  60;
    p_premDataE(MS,"2010","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  70;
    p_premDataE(MS,"2011","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  80;
    p_premDataE(MS,"2012","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 *  90;
    p_premDataE(MS,"2013","DPMTR","PGSAPS")   $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","PGSAPS") /30 * 100;

    p_premDataE(MS,"2005","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  30;
    p_premDataE(MS,"2006","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  35;
    p_premDataE(MS,"2007","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  40;
    p_premDataE(MS,"2008","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  50;
    p_premDataE(MS,"2009","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  60;
    p_premDataE(MS,"2010","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  70;
    p_premDataE(MS,"2011","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  80;
    p_premDataE(MS,"2012","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 *  90;
    p_premDataE(MS,"2013","DPMTR","CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2005","DPMTR","CEILVAL") /30 * 100;
*
    p_premDataE(MS,"2007","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  25;
    p_premDataE(MS,"2008","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  30;
    p_premDataE(MS,"2009","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  35;
    p_premDataE(MS,"2010","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  40;
    p_premDataE(MS,"2011","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  50;
    p_premDataE(MS,"2012","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  60;
    p_premDataE(MS,"2013","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  70;
    p_premDataE(MS,"2014","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  80;
    p_premDataE(MS,"2015","DPMTR","CEILVAL")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","CEILVAL") /25 *  90;


    p_premDataE(MS,"2007","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  25;
    p_premDataE(MS,"2008","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  30;
    p_premDataE(MS,"2009","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  35;
    p_premDataE(MS,"2010","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  40;
    p_premDataE(MS,"2011","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  50;
    p_premDataE(MS,"2012","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  60;
    p_premDataE(MS,"2013","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  70;
    p_premDataE(MS,"2014","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  80;
    p_premDataE(MS,"2015","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  90;
    p_premDataE(MS,"2016","DPMTR","PGSAPS")  $ MAP_RR("BUR",MS)      = p_premDataE(MS,"2007","DPMTR","PGSAPS") /25 *  100;
```

- And permanent grass land (PGGRAS) is eligble

- The historic implementation

The exact set member ship depends on the year. The distribution shares which map the decoupled part of the premiums received under the Agenda package (see above) to these implementation schemes are edited on the Table "p_premToDDTarget_E"

That information is the basis to define regional premium envelops (= CEILVAL) for the different Member states. That is a rather complex program ('*policy/calc_mtr.gms*'). A first key statement defines the *remaining budget envelops for the still coupled payments*. It takes the minimum of the existing ceiling values for that scheme (CEILVAL) or the total payments paid out times the modulation factors and multiplies it with the coupling degree.

There two other factors:

- A possible greening share according to the October 2011 proposal by the Commission, see the section on CAP 2014-2020 for more details

- A national ceiling cut factor which aligns the envelops calculated from the past payments with he total MTR ceiling as defined in the legal texts.

```
TABLE p_premDataE(*,*,PSDPAY_T,PSPOS_GRP)  Premium declarations in new format
*
                              CEILVAL     APPTYPE    PGARAB PGSCOS PGBULF PGHEIF PGCALV PGSHGM PGPARI PGTOBA PGDCOW PGGRAS PGMEAT
*
    EE000000.2004.DPCN_ARAB    20.656     perLevl    56.81
    EE000000.2004.DPCN_CATT     7.228     perLevl            69.29  49.49  34.65   9.90
    EE000000.2004.DPCN_EWE      0.673     perLevl                                        14.02
*
    HU000000.2004.DPCN_ARAB   327.554     perLevl    93.03
    HU000000.2004.DPCN_PARI     0.746     perLevl                                              231.56
    HU000000.2004.DPCN_TOBA    16.336     perLevl                                                      2850
    HU000000.2004.DPCN_BEEF    12.887     perLevl                  136.09 136.09
    HU000000.2004.DPCN_SCOW    14.537     perLevl           124.25
    HU000000.2004.DPCN_CATT     9.702     perLevl            45.85  45.85  45.85
    HU000000.2004.DPCN_EWE      8.352     perLevl                                        10.13
    HU000000.2004.DPCN_DCOW    16.970     perYield                                                           8.71
*
    LV000000.2004.DPCN_ARAB    29.259     perLevl    65.96
    LV000000.2004.DPCN_FODD     7.075     perLevl                                                                  17.90
    LV000000.2004.DPCN_SCOW     2.684     perLevl           138.57
    LV000000.2004.DPCN_ADCT     9.946     perSlgtHd                                                                        80.00
    LV000000.2004.DPCN_EWE      0.244     perLevl                                         13.22
    LV000000.2004.DPCN_DCOW     4.048     perYield                                                           5.82
*
    LT000000.2004.DPCN_ARAB    65.880     perLevl    56.81
    LT000000.2004.DPCN_SCOW     3.480     perLevl           144.81
    LT000000.2004.DPCN_BEEF    14.770     perLevl                  147.71
    LT000000.2004.DPCN_ADCT     5.670     perSlgtHd                                                                        25.78
    LT000000.2004.DPCN_EWE      0.100     perLevl                                         11.56
*
    PL000000.2004.DPCN_ARAB   804.509     perLevl    61.83                                                         61.83
    PL000000.2004.DPCN_TOBA    41.230     perYield                                                     1087
*
    SK000000.2004.DPCN_ARAB    38.120     perLevl    96.78

    SET DPCN(PSDPAY_T) / DPCN_ARAB,DPCN_CATT,DPCN_EWE ,DPCN_PARI,DPCN_TOBA,DPCN_BEEF,DPCN_SCOW,
                         DPCN_DCOW,DPCN_FODD,DPCN_ADCT /;


    p_premDataE(MS,"2005",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL");
    p_premDataE(MS,"2006",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL");
    p_premDataE(MS,"2007",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL");
    p_premDataE(MS,"2008",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL");
    p_premDataE(MS,"2009",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL");
    p_premDataE(MS,"2010",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL");
    p_premDataE(MS,"2011",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL")/30 * 20;
    p_premDataE(MS,"2012",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL")/30 * 10;
    p_premDataE(MS,"2013",DPCN,"CEILVAL")  $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,"CEILVAL") * eps;

    p_premDataE(MS,"2005",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP);
    p_premDataE(MS,"2006",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP);
    p_premDataE(MS,"2007",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP);
    p_premDataE(MS,"2008",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP);
    p_premDataE(MS,"2009",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP);
    p_premDataE(MS,"2010",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP);
    p_premDataE(MS,"2011",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP)/30 * 20;
    p_premDataE(MS,"2012",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP)/30 * 10;
    p_premDataE(MS,"2013",DPCN,PSGROUP)    $ MAP_RR("EU010000",MS) = p_premDataE(MS,"2004",DPCN,PSGROUP) * eps;
```

The part which is not longer coupled goes into the decoupled schemes:

The total budget for the new MTR schemes is derived from the summation of all the old Agenda premiums. The total payments under a scheme such as the Grandes Cultures schemes are corrected for any possible remaining coupled payments:

After that, a possible share going into the greening payment (from 2014) is deducted:

And, finally, a factor is applied which lines up the total historic payments as defined from the CAPRI data and premium schemes in that Member State with the total MTR envelop:

That sum if then distributed to the relevant MTR implementation scheme according to the distribution

```
*
  TABLE p_couplPercent_E(*,*,*) "Coupling degree for each payment and member state"
*
                                BL000000 DK000000 DE000000 EL000000 ES000000 FR000000 IR000000

    DPGRCU      .2006           eps      eps      eps      eps       25        25      eps
    DPPULS      .2006           100      100      100      100      100       100      100
    DPDWHETR    .2006           eps      eps      eps      eps      eps       eps      eps
    DPDWHEES    .2006           100      100      100      100      100       100      100
    DPPARI      .2006           100      100      100      100      100       100      100
    DPSILA      .2006           eps      eps      eps      eps      eps       eps      eps
    DPPARI_fa   .2006           eps      eps      eps      eps      eps       eps      eps
    DPSCOW      .2006           eps      eps      eps      100      100       100      eps
    DPBULF      .2006           eps       75      eps      eps      eps       eps      eps
    DPDCOW      .2006           eps      eps      eps      eps      eps       eps      eps
    DPSHGM      .2006           eps       50      eps      eps       50        50      eps
    DPEXTENS    .2006           eps      eps      eps      eps      eps       eps      eps
    DPPOTA      .2006            60       60       60       60       60        60       60
    DPNE_SHGM   .2006           eps      eps      eps      eps      eps       eps      eps
    DPNE_DCOW   .2006           eps      eps      eps      eps      eps       eps      eps
    DPNE_MEAT   .2006           eps      eps      eps      eps      eps       eps      eps
    DPSL_ADCT   .2006           100      eps      eps      eps       40        40      eps
    DPSL_CALV   .2006           100      eps      eps      eps      100       100      eps
    DPNATMILK   .2006           100      100      100      100      100       100      100
    DPENERCRP   .2006           100      100      100      100      100       100      100

*
      PGARAB.(SWHE,DWHE,BARL,RYEM,OATS,OCER,MAIZ,PARI,RAPE,SUNF,SOYA,PULS,ISET,GSET,TSET,VSET,MAIF,
             POTA,SUGB,TEXT,TOBA,OIND,OOIL,ROOF,OFAR,FALL,TOMA,OVEG,APPL,CITR,OFRU,OLIV,NURS,FLOW,TWIN,OCRO,TAGR,TABO,NECR)


   PGGRAS.(GRAE,GRAI)


*
*      --- the single farm payment scheme (after inclusion of med. crops)
*
      PGFRM.(SWHE,DWHE,BARL,RYEM,OATS,OCER,MAIZ,PARI,RAPE,SUNF,SOYA,PULS,VSET,ISET,GSET,TSET,MAIF,
                 SUGB,TEXT,OIND,OOIL,ROOF,OFAR,GRAE,GRAI,FALL,OLIV,TOBA)


    TABLE p_premToDDTarget_E(*,SIMYY,PSDPAY,DDTarget)
*
*                                 all      arable   grass    farm   single
*                                 land     land     land     land   area payment scheme

                                  DDRARE   DDRARA   DDRGRA   DDFRMF DDSAPS
*
*       -- dynamic hybrid model, in 2013 90% all land, 10% to arable/grass land
*

       DE000000.2006.DPGRCU        30       70
       DE000000.2006.DPPOTA        30       70
       DE000000.2006.DPDWHETR      30       70
       DE000000.2006.DPSILA        30                         70
       DE000000.2006.DPPARI_fa     30                         70
       DE000000.2006.DPSCOW        30                         70
       DE000000.2006.DPBULF        30                         70
       DE000000.2006.DPDCOW        30                         70
       DE000000.2006.DPSHGM        30                         70
       DE000000.2006.DPNE_SHGM     30                         70
       DE000000.2006.DPNE_DCOW     30                         70
       DE000000.2006.DPNE_MEAT     30                70
       DE000000.2006.DPSL_ADCT     30                70
       DE000000.2006.DPSL_CALV     30                         70
```

```
*
  p_premDataE(RS,SIMY,PSDPAY_MTR_EL,"CEILVAL")   $ ((NOT REGAGGP(RS)) $ (p_premDataE(RS,SIMY,PSDPAY_MTR_EL,"CEILVAL") GT eps)
                                                                           $ p_premPaym(RS,"TOTAL",PSDPAY_MTR_EL,"PRME"))
                                         =
*
*        --- smaller of the existing value ceiling, and the actual payments minus modulation
*             => payments cannot not increase beyond the national ceiling
*
                 MIN(p_premDataE(RS,SIMY,PSDPAY_MTR_EL,"CEILVAL"),
                     p_premPaym(RS,"TOTAL",PSDPAY_MTR_EL,"PRME")
                         * SUM(R_RAGG(RS,MS), (p_premPaym(MS,"TOTAL",PSDPAY_MTR_EL,"PRME")-p_premDataE(MS,SIMY,PSDPAY_MTR_EL,"MODU"))
                                                                  /p_premPaym(MS,"TOTAL",PSDPAY_MTR_EL,"PRME")
*
*          --- coupling degree
*
                        * p_couplPercent(PSDPAY_MTR_EL,SIMY,MS)/100 * (1.-p_dpGreeningShare(MS)) * (1+p_nationalCeilCut(MS,SIMY)))) + eps;


 p_DDPartialBudget(RU,DDTARGET,SIMY)

  = sum(PSDPAY_MTR_EL $ (p_premPaym(RU,"Total",PSDPAY_MTR_EL,"PRME") gt eps),

             --- total payments in that scheme in the reference period

                   p_premPaym(RU,"Total",PSDPAY_MTR_EL,"PRME")

             --- times modulated premiums in relation to not modulated ones

                     * SUM(R_RAGG(RU,MS), (p_premPaym(MS,"TOTAL",PSDPAY_MTR_EL,"PRME")-p_premDataE(MS,SIMY,PSDPAY_MTR_EL,"MODU"))
                                                                          /p_premPaym(MS,"TOTAL",PSDPAY_MTR_EL,"PRME")

             --- times the decoupling factor (1 minus the coupling factor)

                        *  (1 - p_couplPercent(PSDPAY_MTR_EL,SIMY,MS)/100 * (1.-p_dpGreeningShare(MS)) ) * (1+p_nationalCeilCut(MS,SIMY)))

             --- times the distribution factors

                                * p_premToDDTarget(RU,SIMY,PSDPay_MTR_EL,DDTARGET)/100);
```

```
--- times the decoupling factor (1 minus the coupling factor)

              *  (1 - p_couplPercent(PSDPAY_MTR_EL,SIMY,MS)/100
```

```
* (1.-p_dpGreeningShare(MS))
```

```
* (1+p_nationalCeilCut(MS,SIMY))
```

keys defined above:

```
*
*              --- times the distribution factors
*
                                           * p_premToDDTarget(RU,SIMY,PSDPay_MTR_EL,DDTARGET)/100);
```

These calculation require that first the total premiums received in the history period are calculated which is done in '*policy/calc_mtr_top.gms*'.

CAP 2014-2020 From 2014 onwards, a new agricultural policy entered into force. The key elements of the policy were (i) convergence of payment rates between member states and farmers within member states, (ii) the expansion of the option to use coupled support beyond the previous articles 68/69, and (iii) the introduction of three "greening requirements". These elements were introduced into CAPRI, and their use can be inspected in the commonly used baseline policy file "gams/pol_input/cap_after_2014/ref.gms", the entire content of which is shown below:

```
* Steer the "CAP after 2014" base scenario file by selecting the proper include files to use.
* The root directory of the include files is "gams\scen"
* If component should not be used at all, use the empty file "void.gms"

$setGlobal basicPaymentSchemeFile premiums\bps_convergence.gms

$setGlobal voluntaryCoupledSupportFile premiums\coupling\CAP_2013_2020_vcs.gms

$setGlobal greeningScenarioFile premiums\greening\cap_2013_2020_greening.gms

$include "scen\base_scenarios\CAP_2014_2020.gms"
```

Since the mechanisms behind each of the three elements is somewhat complex, the file relies on include files to define each of the three components. The include files are stored in the scenario directory (gams/scen) of the CAPRI system, and which particular include files to use is indicated by the string variables ($setGlobal) in the first three code lines. The actual logic of the policy file, also the inclusion of the indicated three files, takes place in the file included in the final line, referred to as the base scenario file.

Convergence between member states is set by adjusting the total budget of the CAP first pillar. Regarding the convergence of payment values per entitlement (IUVs, for Individual Unit Values) inside countries, the regulation allows ample room for national customization. Countries define the regions within which convergence occurs, the end year by which convergence shall be achieved, any remaining maximum span for the IUVs after convergence, and the mathematical formula to use for reducing high IUVs and increasing low ones. The file gams/scen/premiums/bps_convergence.gms defines the options chosen by member states in 2014.

Two different uses of the convergence mechanism are illustrated by Austria and Greece, which apply very different models. Austria applies the full convergence using a linear model over time, with the same target payment rate in all of Austria. The convergence should be complete in 2019. This is obtained by assigning all Austrian regions to one generic "BPS-region", for convenience the first one, called "rbps1". Since the convergence mechanism later on works per member state, it is no problem that rbps1 is also used for e.g. the Netherlands. Then, the convergence option is set to "bps_linear" and the target year

```
set bps_convergence_option(*) "Some options for internal convergence of Basic Payment Scheme" /
    bps_linear              "Linear reduction of IUV above average"
    bps_proportional        "Proportional reductio of IUV above average"
    bps_historical          "BPS with historically decided convergence"
    bps_30_percent_rule     "BPS max 30 percent reduction of payment rule" /;


*   Choices made by member countries
parameter p_bps_final_year(rall,bps_region)         "Final year of the BPS internal convergence";
parameter p_bps_tunnel_target_share(rall,bps_region) "Target value of BPS tunnel, share of average IUV";
parameter p_bps_tunnel_gap_closure(rall,bps_region) "Share of gap to target covered in tunnel model";
set bps_region_to_option(msall,bps_region,bps_convergence_option) "BPS internal convergence options chosen";
set bps_region_to_ru(msall,bps_region,rall)         "Mapping of capri regions to areas with joint BPS convergence model";
```

to 2019. Finally, the two parameters defining the rate of the final convergence are set, or, if you like, the width at the end of the convergence funnel and the handling of payments outside of that funnel. For Austria, the parameters are both set to "1", which means that all farms will get exactly the same payments per hectare after convergence is complete in 2019.

```
*   Case AT: flat rate for all of AT from 2019
*   --------------------------------------------------

*   Create BPS-region
    bps_region_to_ru("AT000000","rbps1",rall) $ map_rr("AT000000",rall) = yes;
*   Activate linear reduction option
    bps_region_to_option("AT000000","rbps1","bps_linear") = yes;
    p_bps_final_year("AT000000","rbps1") = 2019;
*   Target convergence rate, lower corridor value as share of average
    p_bps_tunnel_target_share("AT000000","rbps1") = 1.00;
*   Rate of final closure of gap to target corridor
    p_bps_tunnel_gap_closure("AT000000","rbps1") = 1.00;
```

Greece applies different models for different types of regions, depending on the character of agriculture in the region. We approximate this in CAPRI by classifying the NUTS2-regions according to the shares of arable land, grass land and permanent crops in a historical year (2008). Based on those shares, three BPS-regions are created, within each of which the same convergence model is applied. The convergence is linear, but with the additional 30-percent-rule applied, defining that no farm (supply model region) should get more than 30 percent higher payments per hectare than the average of the BPS-region. Convergence proceeds up to the year 2019, and in each year, the lower limit for convergence, expressed as a share of the averge of the BPS-region, is set to 90%. The lower limit defines whether a farm needs convergence or not. Farms above the lower limit will get the same payments per unit as before, but for farms below the limit, the final option "p_bps_tunnel_gap_closure" kicks in, and defines what share of the gap to the lower convergence limit should be closed. For Greece, this value is set to 1/3, implying that for a farm receiving less than 90% of the average payment in the BPS-region, 1/3 of the gap shall be closed. The increased premiums are financed by a linear reduction of the payments to all farms with payments above the average payment, while also capping the highest premiums to be no more than 30% higher than the regional average.

The code implementing the logic behind these various settings is generic and found in the file "gams/policy/implement_bps.gms". The result is a payment per region, defined using the general premium mechanism of CAPRI, that is called "dp_bps" and with the eligible activity list "pgsaps". The application type is "perLevl" and the budget is set on national level in the base scenario file "gams/scen/base_sce-

```
*   Case Greece: Three tunnel regions dominated by arable, grass or permanent crop land
*   ----------------------------------------------------------------------------------
*   Regions with more than average share (28%) of permanent crops in 2008:
*   EL250000, EL300000, EL410000, EL420000, EL430000
*   Regions with more than average share (25%) of grassland in 2008
*   EL110000, EL120000, EL210000, EL220000, EL240000
*   Remaining regions are classified as arable, then:
*   EL130000, EL140000, EL230000

bps_region_to_ru("EL000000","rbps2",ru) = yes $ [map_rr("EL110000",ru) or sameas("EL110000",ru)];
bps_region_to_ru("EL000000","rbps2",ru) = yes $ [map_rr("EL120000",ru) or sameas("EL120000",ru)];
bps_region_to_ru("EL000000","rbps3",ru) = yes $ [map_rr("EL130000",ru) or sameas("EL130000",ru)];
bps_region_to_ru("EL000000","rbps3",ru) = yes $ [map_rr("EL140000",ru) or sameas("EL140000",ru)];
bps_region_to_ru("EL000000","rbps2",ru) = yes $ [map_rr("EL210000",ru) or sameas("EL210000",ru)];
bps_region_to_ru("EL000000","rbps2",ru) = yes $ [map_rr("EL220000",ru) or sameas("EL220000",ru)];
bps_region_to_ru("EL000000","rbps3",ru) = yes $ [map_rr("EL230000",ru) or sameas("EL230000",ru)];
bps_region_to_ru("EL000000","rbps2",ru) = yes $ [map_rr("EL240000",ru) or sameas("EL240000",ru)];
bps_region_to_ru("EL000000","rbps1",ru) = yes $ [map_rr("EL250000",ru) or sameas("EL250000",ru)];
bps_region_to_ru("EL000000","rbps1",ru) = yes $ [map_rr("EL300000",ru) or sameas("EL300000",ru)];
bps_region_to_ru("EL000000","rbps1",ru) = yes $ [map_rr("EL410000",ru) or sameas("EL410000",ru)];
bps_region_to_ru("EL000000","rbps1",ru) = yes $ [map_rr("EL420000",ru) or sameas("EL420000",ru)];
bps_region_to_ru("EL000000","rbps1",ru) = yes $ [map_rr("EL430000",ru) or sameas("EL430000",ru)];
currbps("rbps1") = yes;
currbps("rbps2") = yes;
currbps("rbps3") = yes;

*   Activate linear reduction option and 30% rule
bps_region_to_option("EL000000",currbps,"bps_linear") = yes;
bps_region_to_option("EL000000",currbps,"bps_30_percent_rule") = yes;
*   Set final year to 2019
p_bps_final_year("EL000000",currbps) = 2019;
*   Target convergence rate, lower corridor value as share of average
p_bps_tunnel_target_share("EL000000",currbps) = 0.90;
*   Rate of final closure of gap to target corridor is 100 percent
p_bps_tunnel_gap_closure("EL000000",currbps) = 1/3;
```

narios/cap_2014_2020.gms".

Voluntary Coupled Support is defined using the standard premium mechanisms of CAPRI, based on notifications received from the European Commission. We have interpreted the notified target activities in terms of CAPRI activities, and set budget ceilings and nominal amounts in the file "gams/scen/premiums/coupling/cap_2013_2020_vcs.gms".

The *Greening Measures* can be steered by the modeller. Even though the greening in itself is complex in implementation, the choices open to the CAPRI modeller are limited. The standard greening policy switches can be inspected in the file "gams/scen/premiums/greening/cap_2013_2020_greening.gms":

The first statement defines the share of the national pillar 1 envelope that is dedicated to the "greening top-up". By default, this is 30%. Then, a set of active greening measures is populated. There are three options available, and by default, they are all active:

- The share of permanent grass land to arable land cannot decline relative to the base year.

```
* --- The share of the BPS-envelope that is dedicated to greening top-ups
*     This comes from the parent file.

   p_DPGreeningShare(MS) =  p_dpGreeningShareCAP14(MS);

* --- activate greening conditions

   activeGreeningMeasures("permGras")    = YES;
   activeGreeningMeasures("cropDiv")     = YES;
   activeGreeningMeasures("ecoSetAside") = YES;

*  --- define the weighting factors how much a hectare of the crops below
*      can be accounted for ecological setaside

   p_DPgreeningWeight(runr,"ISET") = 0;
   p_DPgreeningWeight(runr,"OFAR") = 0.7;
   p_DPgreeningWeight(runr,"VSET") = 1;
   p_DPgreeningWeight(runr,"FALL") = 1;
   p_DPgreeningWeight(runr,"PULS") = 0.7;
   p_DPgreeningWeight(runr,"CATC") = 0.3;


$Setglobal greening_setasiderate 5

$include 'policy\define_greening_limits.gms'
```

- A minimum measure of crop-diversity must be maintained.
- A share of land must be allocated to certain activities counting as "ecological set-aside".

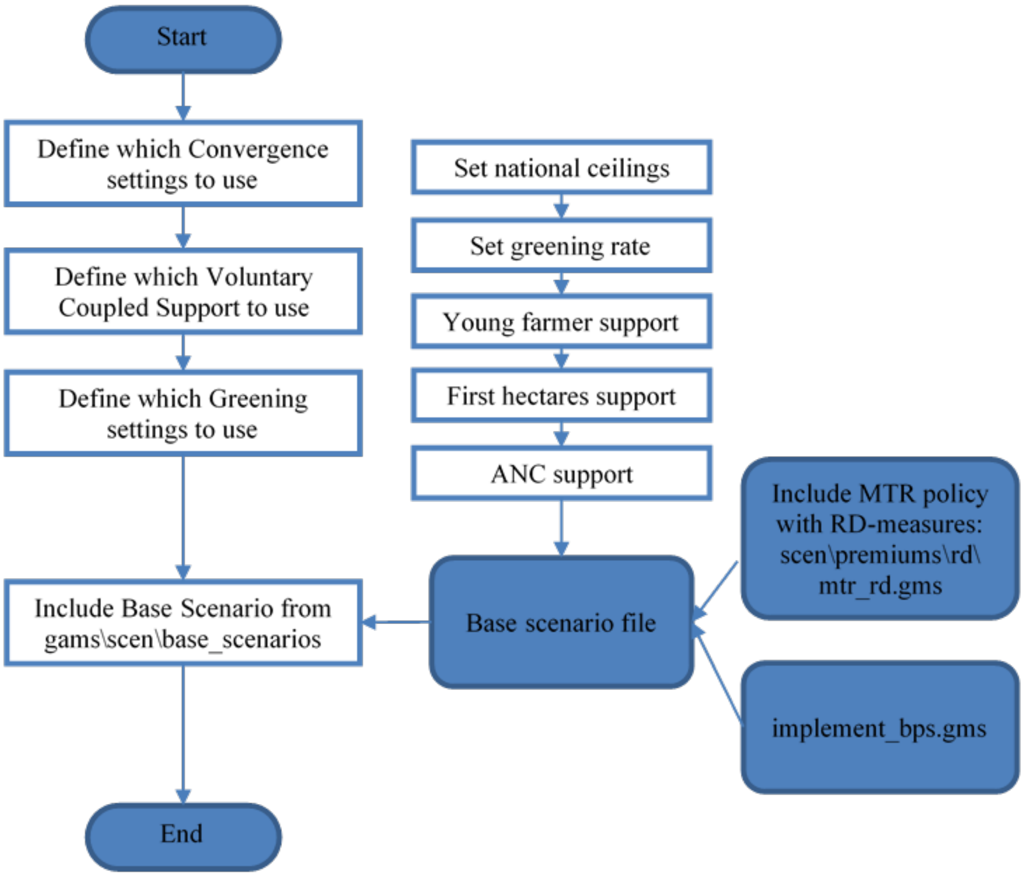
The shares of activities eligible as ecological set-aside is then defined in the concluding parameter definition in the file. The set-aside rate itself is defined as a string variable "$setglobal greening_setasiderate 5", defining it to be 5% by default. The three greening restrictions are implemented as constraints in the supply models. The greening top-up is implemented as a standard CAPRI premium called DPGREEN. The logic behind the greening restrictions is activated in the include file "*policy/define_greening_limits.gms*".

The CAP 2014-2020 also contains three more payment schemes: Support to young farmers, support to smaller farms (first hectares) and support to areas with natural constraints (ANC). These payment schemes, with their associated budgets, are defined in the base scenario file.

The following figure summarizes the logic of the CAP 2014-2020 reference policy as implemented in the CAPRI policy module in the policy file *pol_input/cap_after_20147ref.gms*.

Tradable Single Premium Scheme entitlements With the so-called Mid Term Review of the Common Agricultural Policy, the so-called Single Farm Premium (SFP) as a decoupled payment was introduced which is implemented as a subsidy which does not require production, is subject to cross-compliance and paid per ha up to a number of entitlements. The original entitlements, defined on a hectare basis, had been distributed to farmers operating the land and not the land owners. Both land and entitlements can be traded independently from each other. After a sequence of reform steps, basically all crop production sectors are now included in the subsidy program, so that farmers can be assumed to have received entitlements for all hectares they cropped historically. The same was true from the

Figure 17: The logic of the CAP 2014-2020 reference policy as implemented in the CAPRI policy module



Source: own illustration

beginning for the so-called regional implementation. If the land available to agriculture decreases, e.g. by urbanization, some entitlements cannot not longer be matched with a hectare of eligible land. Such unused entitlements are removed from the markets after a number of years.

In CAPRI, the assumption in the baseline is that all hectares used by agriculture are able to claim the SFP and that any unused entitlements had been removed so that the SFP becomes fully capitalized into land. Subsequent changes in the premiums including the SFP, prices or other policy instruments in a counterfactual run could decrease the marginal returns to agricultural land. Based on the land supply curve implemented in CAPRI, agricultural land use would shrink and some entitlements become unused. Vice versa, if changes let the marginal return to land increase, the entitlements become the limiting factor to claim the subsidy. The increase is thus mapped into an economic rent to the entitlement. If changes generate rents on entitlements in some farm types and not in others, one would assume that trade in entitlements will occur. A simple algorithm to trade the entitlement is now included in CAPRI and described below.

Switching on the entitlement trade

The trade module is implemented in the file '*policy/prem_entl_trade.gms*' which is included on demand in capmod and called in each iteration

```
*
$ifi %entl_trade%==on $include 'policy\prem_entl_trade.gms';
*
```

By default, the entitlement trade is switched OFF in the general settings file of CAPMOD, called gams/capmod/set_global_variables.gms

```
* --- distribute unsued entitlements to other regions/farm types in MS
$setGlobal entl_trade off
```

The basic idea of the module is very simple: shift entitlements from farm type or regions which unused entitlements to other farm types or regions which have an economic rent on their entitlements. The trading entities should receive the very same premium on the entitlement for the current implementation in the code. One should hence set the trade level according to the regional level for which flat rate premiums are implemented as shown below in an example:

```
    Parameter p_premToDDTargetNuts(*) /
$ifi     %farm_m% == on     EU015000 2
$ifi not %farm_m% == on     EU015000 1
    /;
*
    p_premToDDTargetNuts(RMSSUP) = p_premToDDTargetNuts("EU015000");
    p_premToDDTargetNuts(MS) $ (SUM(PSDPAy_MTR_EL, p_premToDDTarget(MS,"%SIMY%",PSDPAY_MTR_EL,"DDSaps"))) = eps;

$setglobal entl_trade on

$ifi %farm_m% == on  $setglobal entl_trade_level NUTS2
$ifi %farm_m% == off $setglobal entl_trade_level NUTS1
```

How the entitlement trade works

The following code pieces are taken from '*policy/prem_entl_trade.gms*'. In a first step, the demand of entitlements is determined. The dual value does only provide an indication that entitlements are scarce, but not how many additional entitlements are needed. Accordingly, first, the average marginal value of the different type of entitlements is determined:

From these a maximum of 10% is defined as the demand in each iteration:

In order to take differences in the marginal returns into account, an indicator based on the squared value is used:

It serves as the distribution key of unused entitlements, which are determined as follows:

```
 p_entlTrade(RU,"demand",step)
     =  sum( (PSDPAY_cutEndog,DDTarget) $ (     (not sameas(DDTarget,"DPGreen"))
                                                                  and (overShotEntl_.m(RU,PSDPAY_cutEndog,DDTarget) lt 0)),
              p_entlLimit(RU,PSDPAY_cutEndog,DDTarget,"Limit"));

 p_entlTrade(RU,"value",step) $ p_entlTrade(RU,"demand",step)
     = -sum( (PSDPAY_cutEndog,DDTarget) $ (     (not sameas(DDTarget,"DPGreen"))
                                                                  and (overShotEntl_.m(RU,PSDPAY_cutEndog,DDTarget) lt 0)),
              p_entlLimit(RU,PSDPAY_cutEndog,DDTarget,"Limit") * overShotEntl_.m(RU,PSDPAY_cutEndog,DDTarget));

 p_entlTrade(RU,"price",step) $ p_entlTrade(RU,"demand",step)
     = p_entlTrade(RU,"value",step)/p_entlTrade(RU,"demand",step);


*
*    --- max 10 % of existing entitlements
*
  p_entlTrade(RU,"demand",step) =  p_entlTrade(RU,"demand",step) * 0.1;



 p_entlTrade(RU,"valueSqr",step) $ p_entlTrade(RU,"demand",step)
    = p_entlTrade(RU,"demand",step) * sqr(p_entlTrade(RU,"price",step));


*
* --- supply : 50% of unused entitlement
*
  p_entlTrade(RU,"supply",step)
      = sum( (PSDPAY_cutEndog,DDTarget) $ p_entlLimit(RU,PSDPAY_cutEndog,DDTarget,"Limit"),
            -MIN(0,v_sumEntl.l(RU,PSDPAY_cutEndog,DDTarget) - p_entlLimit(RU,PSDPAY_cutEndog,DDTarget,"Limit"))) * 0.50;
```



```
*
* --- supply : 50% of unused entitlement
*
  p_entlTrade(RU,"supply",step)
      = sum( (PSDPAY_cutEndog,DDTarget) $ p_entlLimit(RU,PSDPAY_cutEndog,DDTarget,"Limit"),
            -MIN(0,v_sumEntl.l(RU,PSDPAY_cutEndog,DDTarget) - p_entlLimit(RU,PSDPAY_cutEndog,DDTarget,"Limit"))) * 0.50;
```


Next, the number of unused entitlements is stored:

As seen, only 50% of the unused entitlements are released in any iteration. We next determine the size of the markets, i.e. total demand and supply:

```
*
* --- Aggregate demand,supply, value of farm types to NUTS2 level
*
  p_entlTrade(NUTS2Agg,"demand",step) =  sum( Types_to_r(NUTS2Agg,Types), p_entlTrade(Types,"demand",step));
  p_entlTrade(NUTS2Agg,"supply",step) =  sum( Types_to_r(NUTS2Agg,Types), p_entlTrade(Types,"supply",step));
  p_entlTrade(NUTS2Agg,"value",step)  =  sum( Types_to_r(NUTS2Agg,Types), p_entlTrade(Types,"value",step));
  p_entlTrade(NUTS2Agg,"valueSqr",step)  =  sum( Types_to_r(NUTS2Agg,Types), p_entlTrade(Types,"valueSqr",step));
```

The supply is then distributed according to the squared value of the individual demanders

```
  p_entlLimit(Types,PSDPAY_cutEndog,DDTarget,"Limit") $ (     (overShotEntl_.m(Types,PSDPAY_cutEndog,DDTarget) lt 0)
                                                              and  p_entlLimit(Types,PSDPAY_cutEndog,DDTarget,"Limit"))
    = p_entlLimit(Types,PSDPAY_cutEndog,DDTarget,"Limit")
          + sum(types_to_r(Types,NUTS2Agg) $  p_entlTrade(NUTS2Agg,"value",step),
                  p_entlTrade(Types,"valueSqr",step)
                     * p_entlTrade(NUTS2Agg,"supply",step) / p_entlTrade(NUTS2Agg,"valueSqr",step));
```

An example printout

The following code snippet shows an example for a NUTS2 regions and the related farm types for a test run for Greece without the market module:

	S1	S2	S3	S4	S5	S99
EL110000.Demand	32.075	32.439	32.620	32.711	32.756	32.779
EL110000.Value	48766.234	24791.974	13600.485	7543.924	4360.840	2730.454
EL110000.SUPPLY	3.639	1.810	0.906	0.456	0.230	0.116
EL110000.valueSqr	788765.873	202072.288	59256.730	17969.995	5967.575	2332.077
EL110016.Demand	13.441	13.554	13.661	13.714	13.740	13.753
EL110016.Price	134.681	93.743	50.647	27.389	15.621	9.709
EL110016.Value	18102.989	12705.794	6918.624	3756.173	2146.387	1335.326
EL110016.valueSqr	243812.335	119107.489	35040.478	10287.771	3352.898	1296.493
EL110026.SUPPLY	2.300	1.144	0.573	0.288	0.146	0.073
EL110027.Demand	7.982	8.157	8.172	8.182	8.188	8.191
EL110027.Price	218.325	44.225	28.427	16.935	10.066	6.398
EL110027.Value	17426.013	3607.525	2322.885	1385.571	824.122	524.038
EL110027.valueSqr	380452.941	15954.221	6603.166	2346.486	829.521	335.274
EL110056.SUPPLY	0.510	0.253	0.127	0.064	0.032	0.016
EL110057.SUPPLY	0.327	0.163	0.082	0.041	0.021	0.010
EL110086.SUPPLY	0.502	0.250	0.125	0.063	0.032	0.016
EL110999.Demand	10.652	10.728	10.788	10.815	10.828	10.835
EL110999.Price	124.271	79.034	40.406	22.212	12.840	8.039
EL110999.Value	13237.233	8478.655	4358.976	2402.180	1390.331	871.090
EL110999.valueSqr	164500.597	67010.579	17613.085	5335.738	1785.156	700.309

As seen from above, we have two farm types in the starting situation which acts as demanders, i.e. have a marginal value on their entitlements (016 and 999). Their marginal value on the entitlement is quite high in the starting situation with > 125 € / entitlement. We have also a total of 3639 ha after the first round of unused entitlements which can be sold to the demanders. Distributing half of them (ca. 1800 ha) to the two demanders reduces the marginal value of the entitlements already below 95€, the next round distributed ca. 900 ha and brings the price down to 50€ until in the last round almost nothing is left for distribution and the value of the entitlements has dropped below 10€. The reader should note the trade is not yet taking into account in the income calculation of the farm types.

Finally, we come to the main point which motivated the introduction of that module. As indicated above, we interpret the SFP as a subsidy to agricultural land use which at the margin is capitalized in the land rent. It thus increases the marginal returns to land use in agriculture. In our baseline, we start with a situation with an assumed equilibrium in land markets, i.e. marginal returns in agriculture including any subsidies are equal to marginal returns of alternative uses.

Reducing the SFP will render agricultural land use less competitive so that land owner will rent out less to agriculture and put the land into other uses. That effect can be clearly seen below in the first iteration: in the farm types where the SFP drops due to uniform SFP at NUTS2 in Greece, land use is reduced. Total land use in Greece drops by 1.2%. But if we re-distribute the subsidy between farm types, farms which were competing before with below average subsidies against alternative land use possibilities now would like to expand land use. Without additional entitlements, they cannot: the marginal return on the next ha drops by the SFP rate. But once they buy entitlements, they offset a larger part of the land loss: in step twp, the reduction is only about 0.6%. And towards the end, the basically a no-change in land use, as we would have assumed at the aggregated level if the same type of subsidy is paid on average with the same rate.

5.3.5 Pillar II

Overview Modelling of Pillar II for an EU-wide assessment provides a specific challenge given data availability regarding measures which are programmed and implemented at Member State or even regional level. Official reporting of measures under Pillar II in standardized data bases uses a rather rough categorisation, so are e.g. all agri-environmental measures grouped into one category. But even at that rather aggregated level, there is no ready to use data base available to researchers, especially with a regional resolution. A first task therefore consisted in building up a suitable data base on funds for Rural Development measures, a task undertaken by a team which had already compiled such data in the past (BALDOCK et.al. 2002).

The most important measures from a budgetary view point are the Agri-Environmental Schemes and the Less Favourite Area Payments. Therefore some care must be taken to model these measures accurately. The project draws here on the work of a study by LEI and IEEP (LEI and IEEP 2009) for DG-AGRI, which use in the case of the agri-environmental payments analysis based on FADN and for LFA based on FADN and CLUE (Verburg et.al. 2010) results.

Table 26: Overview of pillar II measures modelled in CAPRI

Measure type	Measure codes EU	Modelling approach in CAPRI
LFAs	211-212	Regional direct support. Distribution over sectors and regions based on FADN data and CLUE results.
Natura 2000	213,224	Regional direct support. Distribution over sectors and regions based on FADN data and CLUE results. Conditional on extensive technology being used.
Agri-environment	214-215	Regional direct support. Distribution over sectors and regions based on FADN data. 50% of the support directed towards TF8 farm types 1, 2, 3, 4 and 8 is conditional on extensive technology being used, for remaining amounts extensive as well as intensive technology is eligible.

Modelling of LFA The LFA measure was implemented as a direct payment to cropping and grassland, with the same amount for all cropping activities except for fallow land. The first challenge encountered when implementing the LFA premiums is that the regions do not coincide with the administrative regions used in CAPRI. – Recall that CAPRI only has one single representative firm in each nuts 2 region (or up to nine farm types). In reality, thus, only a share, generally much less than 100%, of the land in a nuts region is eligible for LFA payments, and it may very well be the case that the agricultural production on that land is different from the regional average. For example, one may expect that a mountainous LFA area contains more grass land than the surrounding flat land agricultural areas in the same nuts 2 region. In order to capture a possible bias of this nature, a GIS tool (CLUE-model) was used to compute the shares S_{ij} of LFA in different broad land use classes $j \in$ {non-irrigated arable land, irrigated arable land, pasture, permanent crops} in each region i. Those shares were used to compute a (potentially) different nominal premium amount for crops belonging to each class j. The so computed different amounts were taken to reflect the biased distribution of crops inside and outside of LFA regions. Since CLUE does not distinguish "Mountainous" and "Other" LFA, the nominal amount A to which the shares S were applied was assumed the same everywhere: 250 euro, the maximum amount in mountainous LFA regions.

$$P_{ij} = AS_{ij} \tag{5.63}$$

where
P = Premium per hectare
i = Region
j = Group of crops
A = Maximum amount per hectare, 250 euro
S = Share of LFA in all land of class j

A value ceiling for the premium was computed by adding the budgets for the component measures. Recall that the premium module of CAPRI will apply a cut factor to the amount P such that the ceiling is not overshot.

In economic terms, the potentially different premium rates for different groups of crops has a production effect, so that the type of production in CAPRI that receives the higher rates may expand on the expense of other activities. The interpretation would be that more farmers in the LFA areas comply with the LFA eligibility rules and modify their production plans to comply with the criteria there. Nevertheless, this is a simplification, because in CAPRI, no special technical restrictions are required in order to comply with the payment.

Modelling of N2k Each production activity in CAPRI is split into a low and high yield variant with adjusted input coefficients, and thus own costs and revenues, such that their weighted average recovers regional resp. farm type means. The N2k premiums are modelled in pretty much the same way as the LFA premiums, but now with the additional assumption that the payments are conditional on extensification. This was implemented using the two alternative technologies. Thus, only the technological alternative with yield 20% below nuts2 average and lower input requirements (following a yield function) was made eligible for the payment. This is based on no empirical investigation, but is a pure assumption based on the frequently stated fact that the N2k payments are conditional on extensive management practices.

The interpretation is the following: If more money is spent on the measure, more farmers within the designated areas may switch to extensive agriculture. Then the average payment per hectare of the nuts 2 region would increase, reflecting that a larger share of the farmers now participate in the measure. It is today indeed the case that not all farms within an N2k area receive support.

Modelling agri-environmental payments The agri-environmental payments is a very diverse set of measures, which accounts for the largest share of the second pillar. In the modelling approach opted for here, with one aggregated measure "05 agri-environment", a uniform implementation across member states cannot be used, which is in contrast to the LFA and N2k measures. Instead, a way of capturing the national or even regional preferences within the agri-environmental schemes must be sought.

The method for “nationalization” of the AE scheme employed here is to use the distribution of the sum of the AE measures, i.e. the measure allocated to class 05, to agricultural sectors using the receipts by farm types according to FADN in 2005 as key. This implies linking the support to production. Whether this corresponds to reality is an empirical question. It is doubtless the case for some measures in some regions, but certainly not so for all AE measures in all regions. Refining the implementation would thus involve conditioning the support on technical constraints. Nevertheless, the implementation described above has the merit that it allocates the correct budget, resulting from the LEI budget model, to approximately the right group of farmers.

Table 27: Mapping from aggregated farm types in FADN (TF8) to activity groups in CAPRI

TF8 type	Group of activities in CAPRI
1	Grandes Cultures
2	Vegetables
3	Wine
4	Permanent crops
5	Dairy cows including pastures
6	Suckler cows, sheep and goats, including pastures
7	Pigs and poultry
8	All agricultural activities

The agri-environmental (AE) payments are implemented as extensification subsidies, and they are distributed to regions based on the distribution of less-favoured areas across regions. However, the distribution is modulated by the national level shares of farms in or out of LFA receiving AE support. We needed the following two probabilities:

p_landPartition(ru,lcAgri,”LFA”): The share of land of each land class that is LFA (computed based on extrations from the DYNA-CLUE database).

p_aeLfa(ms,tf8): The probability of a farm of type tf8 having agri-environmental support conditional on the farm being in an LFA region or not, computed based on the FADN sample.

Then, the payment rate for each region is set in proportion to the weighted share of farms likely to have some AE support, predicted by the regional share of each land class (grass, arable) being classified as LFA in each region times the share of farms in/out of LFA having AE-support in the national FADN sample. The computation takes place in policy/rd_logic.gms:

Note that the code does not know how high the absolute level of payments shall be for each region, but allocates the relative levels. Then, the national ceiling for AE payments are applied to adjust all regional payments until the ceiling is respected.

Finally, an extensification effect to the AE payments is introduced using the possibility to make technological variants differently eligible.

```
* --- Refine premium rates by considering the share of different land classes in/out LFA

 LOOP((ms,psdpay_ae,tf8) $ ((p_aeLfa(Ms,tf8,"LFA") or p_aeLfa(Ms,tf8,"noLFA")) $ rdMS(ms) $ dpae_tf8(psdpay_ae,tf8)),

    p_premDataE(runr,rdYear,psdpay_ae,psgroup) $ (p_premDataE(runr,rdyear,psdpay_ae,"ceilVal")
                                              $ map_rr(ms,runr) $ psgroup_tf8(psgroup,tf8))

        = min(2000,

            SUM(tf8_lc(tf8,lcAgri),      p_LandPartition(runr,lcAgri,"LFA")  * p_aeLfa(ms,tf8,"LFA")
                               + (1 - p_LandPartition(runr,lcAgri,"LFA")) * p_aeLfa(ms,tf8,"NoLFA")));
 );

*
 p_premDataE(runr,rdYear,psdpay_ae,"AppType") $ p_premDataE(runr,rdyear,psdpay_ae,"ceilVal")
    = perLevl;

*   1) Modify the entitlement rates for the AE payments where there are NOT different technologies
*      implemented in the p_technFact parameter but as separate activities.


*   Animals for which AE payments are defined are eligible in proportion to livestock units

 p_entlWght1(runr,maact,psdpay_ae)
    $ SUM((rdYear,psGroup) $ psg_to_a(psGroup,maact), p_premDataE(runr,rdYear,psdpay_ae,psGroup))
    = p_luUnits(maact);

*   High weight versions of animals receive less (50%) payments to reflect "extensification"

 p_entlWght1(runr,"DCOH",psdpay_ae) $ p_entlWght1(runr,"DCOH",psdpay_ae)
    = 0.50*p_entlWght1(runr,"DCOH",psdpay_ae);

 p_entlWght1(runr,"BULH",psdpay_ae) $ p_entlWght1(runr,"BULH",psdpay_ae)
    = 0.50*p_entlWght1(runr,"BULH",psdpay_ae);

 p_entlWght1(runr,"HEIH",psdpay_ae) $ p_entlWght1(runr,"HEIH",psdpay_ae)
    = 0.50*p_entlWght1(runr,"HEIH",psdpay_ae);

*   2) For crops with alt. technologies: increasing the premium for T2 and reduce it for T1.
*   It will no longer fit the ceiling exactly, because different payments for T1, T2 are not
*   foreseen in prmcut elsewhere.


 p_technFact(rs,MPACT,psdpay_ae,"T2") $ p_technFact(rs,MPACT,"LEVL","T2") =  0.5;
 p_technFact(rs,MPACT,psdpay_ae,"T1") $ p_technFact(rs,MPACT,"LEVL","T2") = -0.5;

 p_technFact(rs,"ISET",psdpay_ae,T_TACT) $ p_technFact(RS,"ISET","LEVL",T_TACT) = 0;
 p_technFact(rs,"VSET",psdpay_ae,T_TACT) $ p_technFact(RS,"VSET","LEVL",T_TACT) = 0;
```

5.3.6 Co-financing rates, assignment of premiums to pillars, WTO boxes and PSE-types

EU and national budget contribution

The reporting part of the system was expanded to account for (co-)financing rates of the different schemes, so that contributions from EU and national budgets can be differentiated. The underlying factors are currently defined in *'policy/policy_sets.gms'*:

```
PARAMETER P_budToPsdPay(psBud,psdPay) "Aggregation from single payments to aggregates of payments";

p_budToPsdPay("budPil1",PSDPAY) $ p_aggToPsdPay("DPPIL1",PSDPAY) = 1;
p_budToPsdPay("budPil2",PSDPAY) $ p_aggToPsdPay("DPPIL2",PSDPAY) = 0.5;
p_budToPsdPay("budNAT",PSDPAY)  $ p_aggToPsdPay("DPPIL2",PSDPAY) = 0.5;
p_budToPsdPay("budNAT",PSDPAY)  $ p_aggToPsdPay("DPNAT",PSDPAY)  = 1.0;


p_budToPsdPay("budRestEU",PSDPAY) = 1. - sum(psBud, p_budToPsdPay(psBud,PSDPAY));

p_budToPsdPay("budEU",PSDPAY) = p_budToPsdPay("budPil1",PSDPAY)
                              + p_budToPsdPay("budPil2",PSDPAY)
                              + p_budToPsdPay("budRestEU",PSDPAY);
```

PSEs

The mapping to the PSE-types is defined in *'policy/policy_sets.gms'*:

```
*
* ----------------------------------------------------------------------------
*
*   Mapping of CAPRI premium schemes to PSE categories
*
* ----------------------------------------------------------------------------

*
*   --- agri-env schemes (C2) / N2K mapped not non-commodity output (F2)
*
    p_budToPsdPay("PSE_C2",PSDPAY_AE)  = 1;
    p_budToPsdPay("PSE_F2","DPRD_N2K") = 1;
*
*   --- non current, production not required: SFP
*
*
    p_budToPsdPay("PSE_E2",psdPayGreen) = 1;
*
*   --- and less favorite area payments
*
    p_budToPsdPay("PSE_C2","DPRD_LFA")  = 1;
*
*   --- based on current area/animal numbers
*
*       --- MCSharry / fruits and vegs
*
    p_budToPsdPay("PSE_C2",psdPayBlue)  = 1;
*
*       --- Article 68, nordic aid
*
    p_budToPsdPay("PSE_C2",psdpay_68)  = 1;
    p_budToPsdPay("PSE_C2",psdpay_noa) = 1;
```

WTO boxes

In a similar fashion, the premiums are allocated to the WTO boxes. The following payments are allocated to the green box (‘*policy/policy_sets.gms’*):

```
* ------------------------------------------------------------------------------
*
*   WTO Boxes
*
* ------------------------------------------------------------------------------
*
*   --- all pillar II payments are categorized as green box
*
    p_budToPsdPay("greenBox",PSDPAY) $ p_aggToPsdPay("DPPIL2",PSDPAY) = 1;

*
*   --- Single farm payments in different implementation: green box
*
    set psdPayGreen(PSDPAY) /
*
*       -- and a new one!
*
        DPEUFLAT       "EU wide single flat rate premium per ha of land"
*
*       -- environmental top-up
*
        DPGREEN        "Part of SFP which is conditioned on greening"
*
*       -- premiums established with the "Luxembourg compromise" following the Mid-Term Review of Agenda 2004
*
        DPFRMS         "Farm premium specific"
        DPFRMF         "Farm premium flat rate"
        DPREG          "Regional flat rate premium"
        DPSAPS         "Single Area Payment scheme",


*
*       --- Nordic aid
*
        set.psdpay_noa

*
*       -- Norwegian green box payments
*
        DPNOR4         "Welfare payments animal sectors green box"
        DPNOR5         "Miscellaneous payments animal sectors green box"

    /;
    p_budToPsdPay("greenBox",psdPayGreen) = 1;
```

The blue box, i.e.g payments under supply control or only paid up to certain upper limits, is defined as along with remaining amber box payments in Norway:

Currently, the following budget categories are supported (see ‘*sets.gms*’ and ‘*policy/policy_sets.gms*’):

In ‘*reports/feoga.gms*’, these categories are first aggregated for each activities from actual schemes (“PRME” = actual payment rate, p_budToPsdpay: distribution key):

In order to come to a product based accounting scheme as used by the PSEs and WTO, these payments are assigned to the main outputs of the activities. Payments to obligatory set-aside are allocated to the activities according to set-aside rates:

For a discussion about the WTO and PSE Boxes and their implementation in CAPRI see: Mittenzwei, K., Britz, W. und Wieck, C. (2012): Studying the effects of domestic support provisions on global agricultural

```
. ,
p_budToPsdPay("greenBox",psdPayGreen) = 1;

set psdPayBlue(PSDPAY) /

    DPGRCU        "Direct payment to cereals, oilseeds and pulses"
    DPDWHETR      "Direct payments to durum wheat in so-called traditional regions"
    DPDWHEES      "Direct payments to durum wheat in so-called established regions"
    DPPARI        "Specific rice premium remaining after MTR policy"
    DPPARI_fa     "Farm income rice premiums"
*
    DPPULS         "Specific payment for pulses"
    DPENERCRP      "Direct payments to energy crops"
    DPSILA         "Silage premiums for Sweden and Finland"
*
    DPSCOW         "Suckler cow premium, Reg. 1254/1999, Art. 6 and Annex II"
    DPBULF         "Special premium to bulls and steers, all assumed to be bulls, Reg. 1254/1999, Art. 4 and Annex I"
    DPDCOW         "Direct payments to dairy cows, Reg. 1255/1999, Art. 16"
    DPEXTENS       "Extensification payment for bulls, steers and suckler cows, Reg. 1254/1999, Art. 13"
    DPSHGM         "Direct payment for sheep and goat"
    DPSHGM_SUP     "Supplementary payment for sheep and goat, Reg. 2529/2001, Art. 5"
    DPADDSHGM      "Additional payments for sheep and goat"
    DPNE_SHGM      "Additional payments to sheep and goat, Reg. 2529/2001, Art. 11 Annex II"
    DPNE_DCOW      "Additional payments to dairy cows, Reg. 1255/1999, Art. 17, Annex III"
    DPNE_MEAT      "Additional payments to bovine meat cattle, Art. 1255/1999, Art. 14 and Annex IV"
    DPSL_ADCT      "Slaughter premium for adult cattle, Reg. 1254/1999, Art. 11"
    DPSL_CALV      "Slaughter premium for calves, Reg. 1254/1999, Art. 11"
*
*   --- premiums constructed partially from FEOGA data
*
    DPOLIV         "Olive and olive oil sector"
    DPFRVG         "Fruits and vegetables"
    DPWINE         "Wine sector"
    DPTOBA         "Tobacco"
    DPTEXT         "Textile crops"
    DPPOTA         "Starch potatoes",
*
*   --- Article 68
*
    set.s_psdpay_68,
*
*   --- complementary payments during acession period
*
    set.s_psdpay_cn
*
*   -- Norwegian blue box payments
*
    DPNOR1         "Deficiency payment blue box factors of production"
    DPNOR3         "Deficiency payment blue box current production"
    DPNOR7         "Regionalised price support blue box"
/;

p_budToPsdPay("blueBox",psdPayBlue)  = 1;

*
*   -- Norwegian amber box payments
*
p_budToPsdPay("AmberBoxOther","DPNOR2") = 1;
p_budToPsdPay("AmberBoxOther","DPNOR6") = 1;
```

trade: WTO and OECD policy indicators in the CAPRI model, selected paper presented at the 15th Annual Conference on Global Economic Analysis, "New Challenges for Global Trade and Sustainable Development", June 27-29,2012, Geneva (Switzerland).

5.4 Market module for agricultural outputs

5.4.1 Overview on the market model

Whereas the outlay of the supply module has not changed fundamentally since the CAPRI project ended in 1999, the market module was completely revised. Even if several independent simulation systems

```
SET s_PS_budget /

      budEU     "EU contribution total"
      budPil1   "Pillar I contribution"
      budPil2   "Pillar II contribution"
      budRestEU "Other EU contribution"
      budNat    "National contribution"
      budNatEU  "National contribution to EU budget",

*
*     --- WTO categories
*
      set.s_PS_WTO,
*
*     --- PSE categories
*
      set.s_PS_PSE
 /;
```

```
DATA(RUNR,MPACT,psBudRows,"%SIMY%") $ DATA(RUNR,MPACT,"PRME","%SIMY%")
    = sum( (PSDPAY,psBud) $ ( p_budToPsdPay(psBud,PSDPAY) $ sameas(psBudRows,psBud) $ PSDPAY_T_A(MPACT,PSDPAY)),
                                              p_premPaym(RUNR,MPACT,PSDPAY,"PRME") * p_budToPsdPay(psBud,PSDPAY));
```

```
DATA(MS,psBudCols,0,"%SIMY%") $ DATA(MS,"FEOP",0,"%SIMY%")
   = SUM( (PACT_TO_Y(MPACT,0),psBudRows) $ ( (DATA(MS,MPACT,psBudRows,"%SIMY%") or DATA(MS,"OSET",psBudRows,"%SIMY%")) $ sameas(psBudCols,psBudRows)),
          ( DATA(MS,MPACT,psBudRows,"%SIMY%")
          + (DATA(MS,MPACT,"SETR","%SIMY%") * 0.01 * DATA(MS,"OSET",psBudRows,"%SIMY%"))  $ (DATA(MS,MPACT,"SETR","%SIMY%") gt eps)
          )
            * DATA(MS,MPACT,"LEVL","%SIMY%"))/1000.;
```

for agricultural world markets are available as OECD's AgLink, the FAPRI system at the University of Missouri or the WATSIM[10] system at Bonn University, it was still considered necessary to have an independent market module for CAPRI.

The CAPRI market module can be characterised as a comparative-static, deterministic, partial, spatial, global equilibrium model for most agricultural primary and some secondary products, in total about 65 commodities, including the young animals that only matter for market clearing in European regions. The list of commodities is chosen to cover as far as possible all products used for food and feed.

It is deterministic as stochastic effects are not covered and partial as it excludes factor (labour and capital) markets, non-agricultural products and some agricultural products as flowers. It is spatial as it includes bilateral trade flows and the related trade policy instruments between the trade blocks in the model.

The term partial equilibrium model or multi-commodity model stands for a class of models written in physical and valued terms. Demand and supply quantities are endogenous in that model type and driven by behavioural functions depending on endogenous prices. Prices in different regions are linked via a price transmission function, which captures e.g. the effect of import tariffs or export subsidies. Prices in different markets (beef meat and pork meat) in any one region are linked via cross-price terms in the behavioural functions. These models do not require an objective function; instead their solution is a fix point to a square system of equations which comprises the same number of endogenous variables as equations.

The CAPRI market module breaks down the world in about 80 countries, each featuring systems of supply, human consumption, feed and processing functions. The regional breakdown is a compromise between

[10] In the beginning, the CAPRI market part draw on the data base from the WATSIM modelling system. As the latter is not longer active, the CAPRI market part has become an independent world trade model for agricultural products.

full coverage of all individual countries (there are close to 200 UN member countries) and computational data quality considerations. Most countries which a high population number are covered independently, sometimes aggregated with smaller countries. Given the EU focus of CAPRI, agricultural trading partners of the EU are captured in some detail and the regional break down is occasionally updated in view of new trade agreements.

The processing can be differentiated by processing of (a) oilseeds to cakes and oils; (b) biofuel feedstocks to biofuels; (c) raw milk to dairy products; and (d) any other type of industrial processing. The parameters of all types of behavioural functions are derived from elasticities, borrowed from other studies and modelling systems, and calibrated to projected quantities and prices in the simulation year. The choice of *flexible functional forms* (normalised quadratic for feed and processing demand as well as for supply, Generalised Leontief Expenditure function for human consumption) and imposition of *restrictions* (homogeneity of degree zero in prices, symmetry, correct curvature, additivity) on the parameters used ensure *regularity* as discussed below. Accordingly, the system allows for the calculation of welfare changes for the different agents represented in the market model.

Some of the about 80 countries with behavioural functions (CAPRI set "RMS") are aggregated for trade policy modelling into in 44 trade blocks or country aggregates (CAPRI set "RM") with a uniform border protection, and bilateral trade flows are modelled solely between these blocks. Such blocks are EU_WEST, EU_EAST, 'other Mediterranean' countries, Uruguay and Paraguay, Bolivia, Chile and Venezuela, and Western Balkan countries. All other countries or country aggregates are identical to one of the 40 trade blocks in the model.

The two EU blocks (EU_WEST, EU_EAST) interact via trade flows with the remaining trade blocks in the model, but each of the EU Member States features an own system of behavioural functions. The prices linkage between the EU Member States and the EU pool is currently simply one of equal relative changes, not at least ease the analysis of results. If regional competitiveness and hence net exports change significantly it may be expected (and has been observed in Hungary since 2004) that prices in 'surplus' regions would decrease relative to the EU average, contrary to the assumption of proportional linkage. Alternative specifications have been analysed in the context of the CAPRI-RD project and an option of a flexible definition of EU subaggregates is currently the topic of an ongoing project.

The current break down of the CAPRI market model can be found at `http://www.capri-model.org/dokuwiki/doku.php?id=capri:concept:regMarket`.

The market model in its current layout comprises about 70.000 endogenous variables and the identical number of equations.

Policy instruments in the market module include (bi)lateral ad-valorem and specific tariffs and, possibly, Producer/Consumer Subsidy Equivalent price wedges (PSE/CSE) where no current data is entered. However, negative entries on the "PSEi" position (for indirect support) have been used in several appications to reflect a global carbon price. Tariff Rate Quotas (TRQs), billaterally allocation or globally open are integrated in the modelling system, as are intervention stock changes and subsidised exports under WTO commitment restrictions for the EU. Subsidies to agricultural producers in the EU are not covered in the market model, but integrated in a very detailed manner in the supply model. For the EU, flexible tariffs for cereals and the minimum import price regime for fruits & vegetables are introduced. Flexible tariffs related to minimum import prices are also present for Switzerland.

5.4.2 The approach of the CAPRI market module

Multi-commodity models are as already mentioned above a widespread type of agricultural sector models. There are two types of such models, with a somewhat different history. The first type could be labelled 'template models', and its first example is Swopsim. Template models use structurally identical equations for each product and region, so that differences between markets are expressed in parameters. Typically, these parameters are either based on literature research, borrowed from other models or simply set by the researcher, and are friendly termed as being 'synthetic'. Domestic policies in template models are typically expressed by price wedges between market and producer respectively consumer prices, often using the PSE/CSE concept of the OECD. Whereas template models applied in the beginning rather simple functional forms (i.e, constant elasticity double-logs in Swopsim or WATSIM), since some years flexible functional forms are in vogue, often combined with a calibration algorithm which ensures that the parameter sets are in line with microeconomic theory. The flexible functional forms combined with the calibration algorithm allow for a set of parameters with identical point elasticities to any observed theory consistent behaviour which at the same time recovers quantities at one point of observed prices and income. Ensuring that parameters are in line with profit respectively utility maximisation allows for a welfare analysis of results.

Even if using a different methodology (explicit technology, inclusion of factor markets etc.), it should be mentioned that Computable General Equilibrium models are template models as well in the sense that they use an identical equation structure for all products and regions. Equally, they are in line with microeconomic theory.

The second type of model is older and did emerge from econometrically estimated single-market models linked together, the most prominent example being the AGLink-COSIMO modelling system. The obvious advantages of that approach are firstly the flexibility to use any functional relation allowing for a good fit ex-post, secondly that the econometrically estimated parameters are rooted in observed behaviour, and thirdly, that the functional form used in simulations is identically to the one used in parameter estimation. The downside is the fact that parameters are typically not estimated subject to regularity conditions and will likely violate some conditions from micro-theory. Consequently, these models are typically not used for welfare analysis. Examples of such models are AGLink-COSIMO at the OECD_FAO, the FAPRI set of market models, or the set of models emerging from AgMemod. However, AGLink-COSIMO is currently in the process to move closer to a template model.

The CAPRI market module is a template model using flexible functional forms. The reason is obvious: it is simply impossible to estimate the behavioural equations for about 65 products and 80 countries or country blocks worldwide with the resources available to the CAPRI team. Instead, the template approach ensures that the same reasoning is applied across the board, and the flexible functional forms allow for capturing to a large degree region and product specificities. As such, the results from econometric analysis or even complete parameters sets from other models could be mapped into the CAPRI market model.

5.4.3 Behavioural equations for supply, feed demand and land markets

The definition of the market model can be found in *'arm/market_model.gms'*

Agricultural supply *Supply* for each agricultural output i and region r (EU Member States or regional aggregate) is modelled by a supply function derived from a normalised quadratic profit function via the envelope theorem. Supply depends on producer prices *v_prodPrice* normalised with a price index. The price index relates to all those goods – either inputs or outputs – which are not explicitly modelled in the system:

$$v\_prodQuant_{i,r} = as_{i,r} + \sum_j bs_{i,j,r} \frac{v\_prodPrice_{j,r}}{p_{index,r}} \tag{5.64}$$

Supply curves for the EU Member States, Norway, Turkey, Western Balkans are calibrated in each iteration to the last output price vector used in the supply model and the aggregated supply results at Member State level, by shifting the constant terms *as*. The slope terms *bs* which capture own and cross-price effects are set in line with profit maximisation, as discussed below. The calibration of the price dependent parameters *bs* is discussed below.

For the countries which matching regional models on the supply side, the bs parameters are derived directly from the costs function terms and elements of the constraints matrix (see Chapter 4.3 FIXME above).

Land supply and demand The reader should note that land is one of the products in the above system which is an input into agriculture. A land supply curve defines the demand side of the land balance which together with the land demand according to the equation above define the land rent clearing the balance. Equally, the price for feed energy is an input price entering the equations.

$$\begin{aligned} v\_prodQuant_{r,"Land"} = {} & p\_cnstLandSpply_r \\ & + p\_cnstLandElas_r \; log(v\_prodPrice_{r,"Land"}) \end{aligned} \tag{5.65}$$

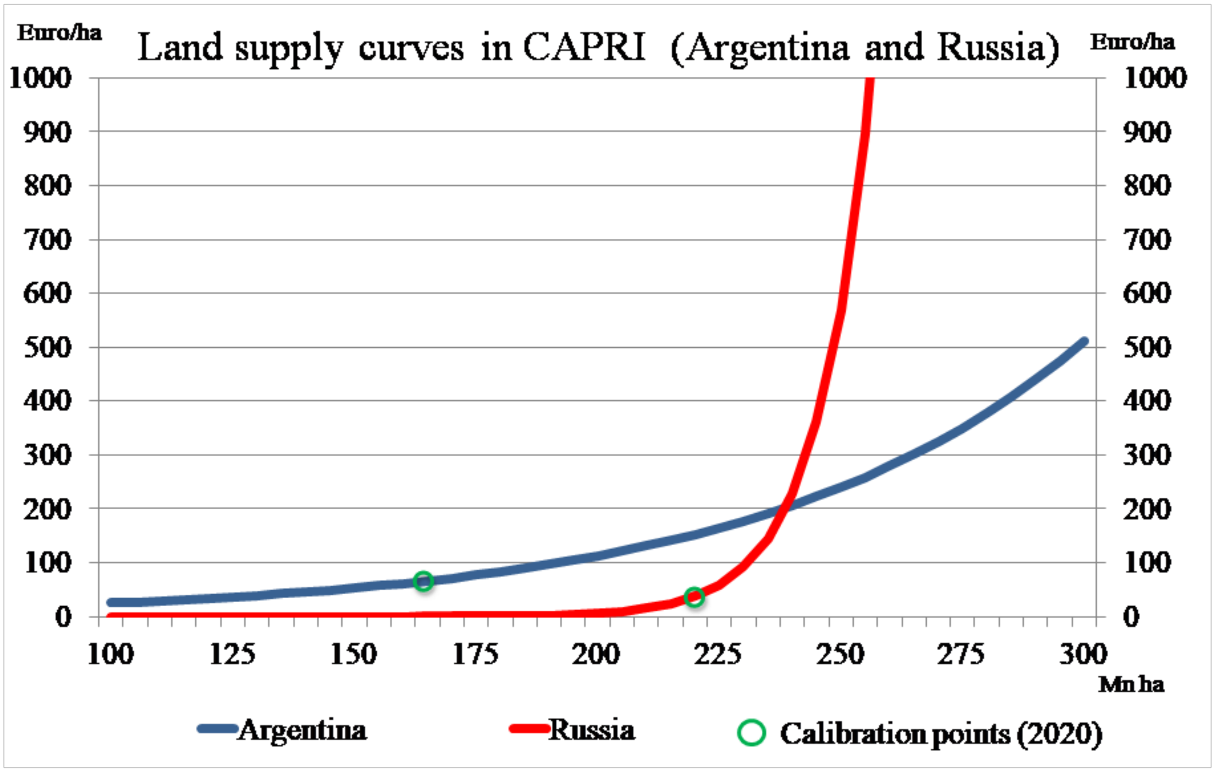
In order to parameterize the land demand function, information about yield and supply elasticities is used. The marginal reaction of land to a marginal change in one of the prices is defined as the total supply effect minus the yield effect:

$$\frac{dLand}{dp_i} = \sum_j \frac{dQ_j}{dp_i} 1/yield_j - ela_i \frac{Q_i}{p_i} 1/yield_i \tag{5.66}$$

where Q denotes quantities and p prices. The calibration requires prices for land which is set to 0.3 of the crop revenues. For fodder area demand from animals it was set lower (cp. page 56 in Golub et.al. 2006). The summation term over all products j defines the "land demand change" if the price i changes. The quantity change of each product j is translated into a land demand change based on its yield. The reciprocal of the yield can be interpreted as the land demand per unit produced.

The last term translates first the yield elasticity (an endogenous variable) into a marginal quantity effect and, then again, based on the yield, in a changed land demand. The land demand change is subtracted as the land demand for product i per unit decreases if yields increase due to a positive own price yield elasticity. This formulation assumes, as conventionally done in multi-commodity models, that cross-price yield effects are zero.

Figure 18: Land supply curve examples



Source: own calculations

The CAPRI market model does not handle explicitly non-tradable crop outputs such as grass, silage or fodder maize. However, the land demand for these products is also taken into account in order to allow closing an overall land balance. This is achieved by deriving per unit land demands for animal products.

The link to land supply is straightforward: if the land supply curve is known, a market balance equilibrates the simulated land demand based on the behavioral equation derived from the NQ profit function with that land supply curve. The land price is the equation multiplier attached to the land market clearing equation. If the land supply curve collapses to a constant, any output price induced change in land demand is leveled out by a change in the land price. That solution is equivalent to the behavior of programming model with a fix land endowment.

If the market balance equation is dropped and the land price fixed, totally elastic land supply to agriculture is assumed. This assumption should be of little empirical value as agriculture is typically the major demander for land suitable to agriculture. The current solution incorporates a land supply curve with exogenous given elasticities (similar to Tabeau et.al. 2006). The parameterization of the land supply curve is currently chosen similar to the land elasticities e.g. reported in GTAP (Ahmed et al. 2008). Generally, land supply is rather inelastic so that price increase of agricultural commodities rather increases land prices and not so much land use.

Although the profit function approach does not allocate areas to individual crops, it captures in a dual formulation how product and land price changes impact on *total* agricultural land use. The estimated yield elasticity may be used however to derive an allocation which is consistent with the assumptions used during parameter calibration. Basically, for each crop product, the yield elasticity is used to derive from the simulated quantity change the implied land change. The resulting land demands are then scaled to match the simulated change in total land. This allocation is currently active at solution time (with explicit equations) in the "trunk" version of CAPRI but not yet in all branches (in particular not in the "star2.4" version underlying the CAPRI Training 2019).

Feed demand The system for *feed demand* for countries not covered by the supply part is structured identically to the supply system. However, not producer prices, but raw product prices *v_arm1Price* determined by the Armington top level aggregator drive feed demand *v_feedQuant*, combined with changes in the supply of animal products weighed with feed use factors w:

$$v\_feedQuant_{i,r} = \left(af_{i,r} + \sum_j bf_{i,j,r} \frac{v\_arm1Prices_{j,r}}{p_{index,r}} \right) \tag{5.67}$$

One of the prices in the equations above is the price for feed energy which is conceptually the output sold by the feed producing industry where products used for feed are its input. A balance between feed energy demanded, derived from the supply quantities and feed energy need for tradeables per unit produced and feed energy delivered from current feed demand quantities drives the price for feed energy:

$$\sum_i v\_feedQuant_{i,r}\ p\_calContent_r = \sum_i v\_prodQuant_{i,r}\ pv\_feedConv_{i,r} \tag{5.68}$$

For countries which matching modules on the supply side, a different system is set-up. In the supply part, individual commodities are aggregated to categories as cereals. The composition of these aggregates is determined by a CES function, whereas total demand for each category depends on the average price aggregated from the ingredients.

As for supply, feed demand for the aggregate EU Member States, Norway, Turkey and Western Balcans, are calibrated in each iteration to the last output price vector used in the supply model and the aggregated feed demand at Member State level.

The disadvantage of the behavioural functions above is the fact that they might generate non-positive values. That situation might be interpreted as a combination of prices where the marginal costs exceed marginal revenues. Accordingly, a fudging function is applied for supply, feed and (see below) processing demand which ensures strictly positive quantities. That fudging function is highly non-linear, and therefore only switched on on demand.

5.4.4 Behavioural equations for final demand

The final demand functions are based on the following family of indirect utility functions depending on consumer prices *cpri* and per capita income y[11] where G and F are functions of degree zero in prices (RYAN & WALES 1996) which will be defined below[12]:

$$U(cpri, y) = \frac{-G}{y - F} \tag{5.69}$$

Using Roy's identity, the following per capita Marshallian demands *PerCap* are derived:

$$PerCap_i = F_i + \frac{G_i}{G}(y - F) \tag{5.70}$$

where the F_i and G_i are the first derivative of F and G versus own prices. The function F is defined as follows:

$$F_r = \sum_i d_i \; cpri_i \tag{5.71}$$

where the d_i have a similar role as constant terms in the Marshallian demands and can be interpreted as 'minimum commitment levels' or consumption quantities independent of prices and income. The term in brackets in the per capita demands in Equation 122 FIXME above hence captures the expenditure remaining after the value F of price and income independent commitments d ('committed income') has been subtracted from available income y to give so called 'non-committed' income. The function G is based on the Generalised Leontief formulation and must be positive to have indirect utility increasing in income:

$$G = \sum_i \sum_j bd_{ij} \sqrt{cpri_i \; cpri_j} \tag{5.72}$$

with the derivative of G versus the own price is labelled G_i and defined as:

$$G_j = \sum_i bd_{ij} \sqrt{cpri_i / cpri_j} \tag{5.73}$$

Symmetry is guaranteed by a symmetric *bd* matrix describing the price dependent terms, correct curvature by non-negative the off-diagonal elements of *bd*, adding up is automatically given, as Euler's Law for a homogenous function of degree one $\left(a(x) = \sum_i \frac{\partial a(x)}{\partial x_i} x_i\right)$, leads to:

[11] Per capita income and total expenditure are used as synonyms in the following as the demand is cover all goods and thus exhaust available income.

[12] Note that indirect utility must be increasing in income. At the same time Y must be larger than F, so called 'committed' income. Hence function G must be positive and utility is a negative number approaching zero as income increases to infinity.

$$\sum_i PerCap_i\ cpri_i = \frac{\sum_i G_i cpri_i}{G}(y-F) + \sum_i d_i cpri_i$$
$$= \frac{G}{G}(y-F) + F = \ y \tag{5.74}$$

and homogeneity is guaranteed by the functional forms as well. The expenditure function follows from rearranging Equation 121 FIXME :

$$y = e(U, cpri) = F - \frac{G}{U} \tag{5.75}$$

The function is flexible to reflect all conceivable own price and expenditure elasticities but the non-negativity imposed on the off-diagonal elements ensuring excludes Hicksian complementarity, a restriction not deemed important in the light of the product list covered. Note that concavity of e is given if G is concave, as U < 0 and F is linear. Concavity of G in turn follows from nonnegative off-diagonal bd_{ij} without further restrictions, because G is then a sum of concave elementary functions $bd_{i,j}(cpri_i cpri_j)^{0.5}$ with the linear terms on the diagonal being both concave and convex regardless of signs of bd_{ii}.

Human consumption hcom is simply the sum of population *pop* multiplied with the per capita demands:

$$hcom_{i,r} = pop_r\ PerCap_{i,r} \tag{5.76}$$

5.4.5 Behavioural equations for the processing industry

Processing demand for oilseeds is modelled by using behavioural functions derived from a normalised quadratic profit function under the assumption of a fixed I/O relation between seeds, cakes and oils. Consequently, the processing demand *proc* depends on processing margins *procMarg* which are differences between the value of the outputs (oil and cake) per unit of oilseed processed and the value of the oilseed inputted:

$$v\_procMarg_{i,r} = ac_{i,r} + \sum_j bc_{i,j,r} \frac{v\_procMarg_{j,r}}{p_{index,r}} \tag{5.77}$$

The processing margins are replaced by producer prices times -1 for all products besides oilseed. For the latter, the processing margin is defined from the producer prices *v_prodPrice* for the cakes and oils time the respective crushing coefficients minus the buying prices (average of domestically sold and imported quantities) *v_arm1Price*:

FIXME

$$\begin{aligned} v\_prodMarg_{seed,r} =& - v\_arm1Price_{seed,r} \\ &+ v\_prodPrice_{seed\rightarrow cak,r} v\_procYield_{cak,r} \\ &+ v\_prodPrice_{seed\rightarrow oil,r} v\_procYield_{oild,r} \end{aligned} \tag{5.78}$$

Finally, output of oils and cakes *supply* depends on the processed quantities *proc* of the oilseeds and the crushing coefficients: FIXME

$$\begin{aligned} supply_{cake,r} &= proc_{seed,r} v\_procYield_{cak,r} \\ supply_{oil,r} &= proc_{seed,r} v\_procYield_{oil,r} \end{aligned} \tag{5.79}$$

The processing yields in the base year are defined as:

$$v\_procYield_{cak,r} = \frac{v\_prodQuant^{bas}_{seed \to cak,r}}{v\_prodQuant^{bas}_{seed,r}} \tag{5.80}$$

The processing yields are however not fixed during simulation, but depend on a CES function: the share of oil increases very slightly if the oil price increases compared to the cake price, and vice versa.

Special attention is given to the processing stage of *dairy products*. First of all, balancing equations for fat and protein ensure that the processed products use up exactly the amount of fat and protein comprised in the raw milk. The *fat and protein content* of raw milk *cont* and milk products *mlk* is based on statistical and engineering information, and kept constant at calibrated base year levels.

$$v\_prodQuant_{"milk",r} cont_{"milk",fp} = \sum_{mlk} v\_prodQuant_{mlk,r} cont_{mlk,fp} \tag{5.81}$$

Production of *processed dairy products* is based on a normalised quadratic function driven by the difference between the dairy product's market price and the value of its fat and protein content.

$$\begin{aligned} v\_prodQuant_{mlk,r} =& am_{mlk,r} \\ &+ \sum_j bm_{mlk,j,r} \frac{v\_prodPrice_j - cont_{j,"fat"} pFatProt_{fat,r} - cont_{j,"prot"} pFatProt_{prot,r}}{p_{index,r}} \end{aligned} \tag{5.82}$$

And lastly, prices of raw milk are derived from its fat and protein content valued with fat and protein prices and a processing margin.

5.4.6 Trade flows and the Armington assumption

The *Armington assumption* drives the composition of demand from domestic sales and the different import origins depending on price relations and thus determines *bilateral trade flows* (Armington 1969). The Armington assumption is frequently used in that context, and e.g. applied in most Computable General Equilibrium models to describe the substitution between domestic sales and imports.

The underlying reasoning is that of a two-stage demand system. At the upper level, demand for products such as wheat, pork etc. is determined as a function of prices and income – see above. These prices are a weighted average of products from different regional origins. At the lower level, the composition of demand per product i in region r stemming from different origins r1 is determined based on a CES utility function:

$$U_{i,r} = \alpha_{i,r} \left[\sum_{r1} \delta_{i,r,r1} M_{i,r,r1}^{1/\rho_{r,i}} \right]^{1/\rho_{r,i}} \tag{5.83}$$

where U denotes utility in region r and for product i due to consumption of the import quantities M stemming from the different origins *r1*. If r is equal *r1*, M denotes domestic sales. δ are the share parameters, α is called the shift-parameter, and ρ is a parameter linked to the substitution elasticity $\sigma = 1/(1+\rho)$. Deriving the first order conditions for utility maximisation under budget constraints leads after lengthy re-arrangements to the following relation between imported quantity $M_{i,r,r1}$, utility $U_{i,r}$ (quantity aggregator), import price $P_{i,r,r1}$, and the aggregate import price $P_{i,r}$:

$$M_{i,r,r1} = U_{i,r} \delta_{i,r,r1}^{\sigma_{r,i}} \left[\frac{P_{i,r}}{P_{i,r,r1}} \right]^{\sigma_{r,i}} \quad \text{(importShares\_)} \tag{5.84}$$

As seen from the equation, imports from region *r1* will increase if its competitiveness increases – either because of a lower price in *r1* or a higher average import price. The resulting changes in the compositions of imports increase with the size of the related share parameter $\delta_{i,r,r1}$ and with the size of the substitution elasticity. The CES utility function is rather restrictive as it has solely one parameter δ per import flow. The substitution elasticity $1/(1+\rho)$ is set exogenously, see below. The δ parameters are determined when calibrating the model to known import flows, whereas α is used to meet the known quantities in the calibration point (and effectively is in the δ parameters in the implementation in the code).

The model comprises a two stage Armington system (see below): on the top level, the composition of total demand from imports and domestic sales is determined, as a function of the relation between the domestic price and the average import price. The lower stage determines the import shares from different origins and defines the average import price. This also means that the implementation in the code has to distinguish between a top and lower level level quantity (or price) aggregate. The substitution elasticity on the top level stage is smaller than for the second one, i.e. we assume that consumers will be less responsive regarding substitution between domestic and imported goods compared to changes in between imported goods.

The following table shows the substitution elasticities used for the different product groups. Compared to most other studies, we opted for a rather elastic substitution between products from different origins, as agricultural products are generally more uniform then aggregated product groups, as they can be found e.g. in CGE models.

There are some specific settings, such as a value of 2 for rice and the EU15, 2.5 respectively 5 for Japan to account for its specific tariff system, as well as some lower values for EU's Mediterrean partner countries.

The above "primal" formulation of the Armington approach in terms of quantity aggregators turned out numerically less stable in the implementaiotn than the dual representation in terms of price aggregators. The Armington approach suffers from two important shortcomings. First of all, a calibration to a zero flow is impossible so that only observed import flows react to policy changes while all others are fixed at

Table 28: Substitution elasticities for the Armington CES utility aggregators

Product (group)	Substitution elasticity between domestic sales and imports	Substitution elasticity between import flows
Cheese, fresh milk products	2	4
Other vegetables	1.5	1.5
Other fruits	3	3
Sugar	12	12
All other products	8	10

A sensitivity analysis on those elasticities is given in section 5.7

zero level. For most simulation runs, that shortcoming should not be serious. If it is relevant, it may be overcome using the modified Armington approach as explained in Section 5.4.8.

Secondly, the Armington aggregator defines a utility aggregate and not a physical quantity. That second problem is healed by re-correcting in the post model part to physical quantities. Little empirical work can be found regarding the estimation of the functional parameters of Armington systems. Hence, substitution elasticities were chosen as to reflect product properties as shown above.

5.4.7 Market clearing conditions

All quantities in the model are measured in thousand metric tons. The *quantity balances* for the trade blocks first state that production must be equal to domestic sales plus export flows plus changes in intervention stocks:

$$\begin{aligned} v\_domSales_{i,r} =& v\_prodQuant_{i,r} \\ & - v\_expQuant_{i,r} \qquad \text{(SupBalM\_)} \\ & + v\_intervStockChange_{i,r} \end{aligned} \tag{5.85}$$

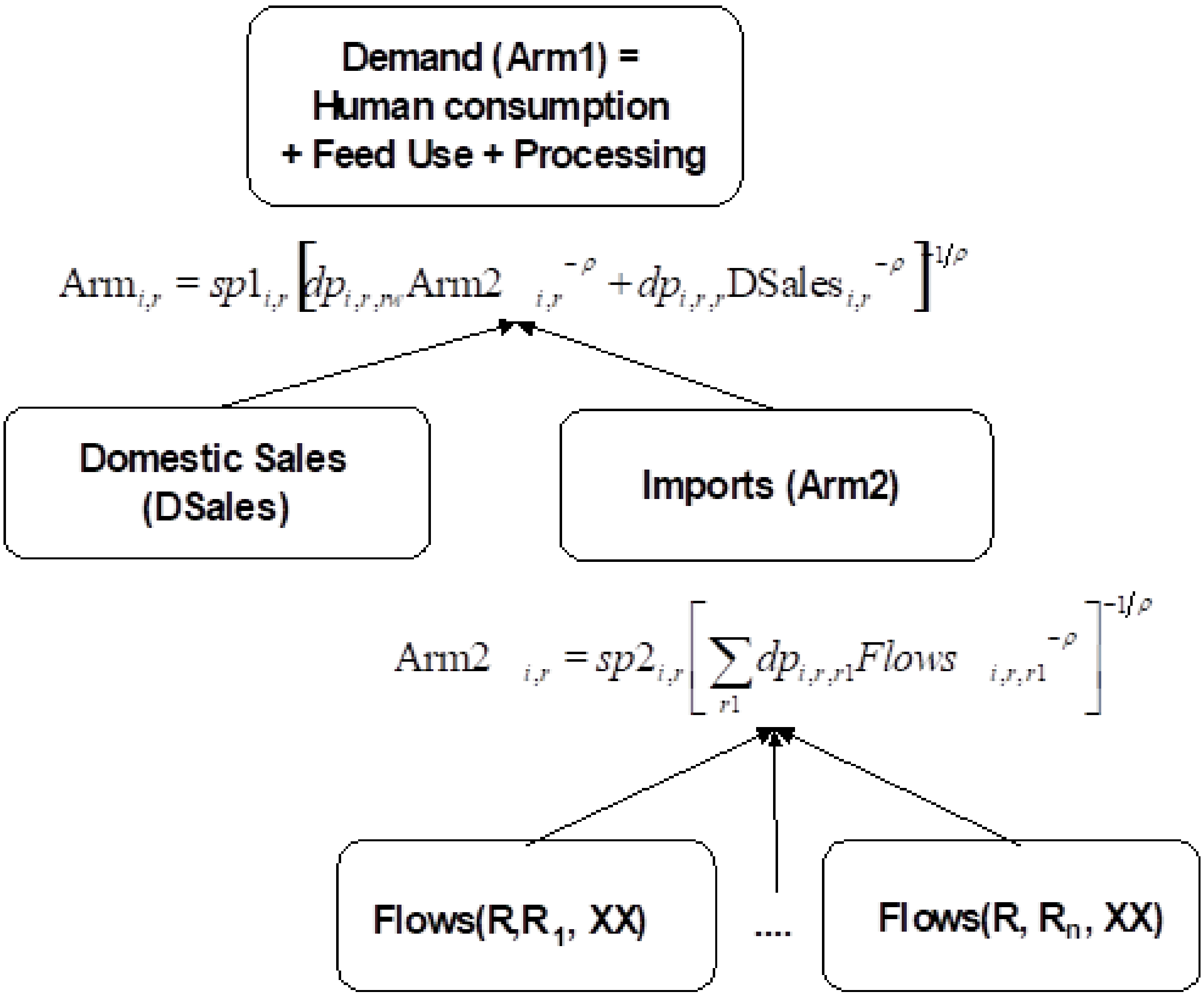
Further on, imports and exports are defined from bilateral trade flows as:

$$v\_impQuant_{i,r} = \sum_{r \neq r1} v\_tradeFlows_{i,r,r1} \quad \text{(impQuant\_)} \tag{5.86}$$

$$v\_expQuant_{i,r} = \sum_{r \neq r1} v\_tradeFlows_{i,r,r1} \quad \text{(expQuant\_)} \tag{5.87}$$

Finally, the *Armington first stage aggregate - arm1*, shown in the diagram above, is equal to the domestic consumption elements feed, human consumption and processing:

Figure 19: Two-stage Armington System



Source: Capri Modelling System

$$
\begin{aligned}
v\_arm1Quant_{i,r} = & v\_feedQuant_{i,r} \\
& + v\_consQuant_{i,r} \\
& + v\_procQuant_{i,r} \qquad \text{(armBall\_)} \\
& + v\_biofprocQuant_{i,r}
\end{aligned} \tag{5.88}
$$

The reader should note that for those trade blocks comprising several countries such as the EU, the right hand side quantities are an aggregate over individual countries.

5.4.8 Price linkages

All prices in model are expressed as € per metric ton. *Import prices* - $impp_{i,r,r1}$ from region *r1* into region r of product i are determined from market prices *v_marketPrice* times a possible exogenous change in exchange rates compared to the baseline minus export subsidies *expsub* plus bi-lateral transport costs aking into account bilateral ad valorem (*v_tarAdval*) and specific (*v_tarSpec*) tariffs respectively flexible levies (*DiffLevies*):

$$impp_{i,r,r1} = \begin{bmatrix} v\_marketPrice_{i,r1}\, p\_exchgRateChangeFaktor_{r,r1} \\ -expsub_{i,r1} + v\_transpCost_{i,r,r1} \end{bmatrix} \\ (1 + v\_tarAdVal_{i,r,r1}/100) + v\_tarSpec_{i,r,r1} + DiffLevies_{i,r,r1} \tag{5.89}$$

Bilateral tariffs may be endogenous variables if they are determined by a tariff rate quota (TRQ), see below. Equally, export subsidies are endogenous variables.

Producer prices are derived from market prices formally using direct and indirect PSEs price wedges – albeit these are all zero at currently -, except for EU_WEST, EU_EAST and Bulgaria and Romania. The reader is reminded that for the EU, the supply model includes a rather detailed description of the different premium schemes of the CAP, so that the EU premiums need not to be modelled as price wedges in the market part. In cases where several countries are comprised in a trade block, the market price refers to the trade block. The "Pmrg" is a factor which is defined such that observed producer prices are recovered.

$$v\_prodPrice_{i,r} = [v\_marketPrice_{i,r} + PSEd_{i,r} + PSEi_{i,r}]Pmrg_{i,r} \tag{5.90}$$

```
*
* --- Producer price
*
 PPri_(RMS,XXX) $ (DATA(RMS,"Prod",XXX,"CUR") or RMSSUP(RMS))  ..

      v_prodPrice(RMS,XXX)/(DATA(RMS,"PPRI",XXX,"CUR")+1)
         =E= (SUM(RMS_TO_RM(RMS,RM),

*     --- Internal market prices from OECD calculations
         (v_marketPrice(RM,XXX) + DATA(RM,"PSEi",XXX,"CUR") + DATA(RM,"PSEd",XXX,"CUR"))
*     --- difference in percentage terms from base year
            * pv_prodPriceMarg(RMS,XXX)
*     --- plus area / per head payment
            + (DATA(RMS,"AREP",XXX,"CUR")/DATA(RMS,"Yild",XXX,"CUR")) $ DATA(RMS,"Yild",XXX,"CUR"))/(DATA(RMS,"PPRI",XXX,"CUR")+1);
*
```

The reader is reminded that currently, the PSE data are not introduced in the system with two exceptions: carbon price scenarios involve negative PSEi amounts and Swiss agricultural policies are involving land subsidies entered.

The *average prices of imports* derived from the Armington second stage aggregate are labelled *v_arm2Price*:

$$v\_arm2P_{i,r} = \left[\sum_{r1} \delta_{i,r,r1}^{\sigma_{2,r,i}} v\_impP_{i,r,r1}^{1-\sigma_{2,r,i}}\right]^{1/(1-\sigma_{2,r,i})} \qquad \text{(arm2PriceAgg\_)} \tag{5.91}$$

Similarly, the *average prices for goods consumed domestically* - *arm1p* are a weighted average of the domestic market price $v_marketPrice_{i,r}$ and the average import price $v_arm2P_{i,r}$:

$$v\_arm1P_{i,r} = \left[\sum_{r1} \delta_{i,m}^{\sigma_{r,i}} P_{i,r,D}^{1-\sigma_{r,i}}\right]^{1/(1-\sigma_{r,i})} \tag{5.92}$$

where $D \in \{M, S\}, P_{i,r,M} = v\_arm2P_{i,r}$ and $P_{i,r,S} = v\_marketPrice_{i,r}$. Consumer prices - *v_consPrice* are derived from the composite good price index *v_arm1Price* taken into account policy introduced price wedges as direct and indirect consumer subsidy equivalents (currently not available) plus a fix margin covering transport, processing and all other marketing costs:

$$v\_consPrice_{i,r} = v\_arm1Price_{i,r} - CSEd_{i,r} - CSEi_{i,r} + cmrg_{i,r} \tag{5.93}$$

Unit value exports net of border protection are defined as average market prices in the export destination minus tariffs as:

$$v\_unitValueExports_{i,r} = \frac{\sum_{r1 \neq r}(pmrk_{i,r1} - tariffs_{i,r,r1})/(1 - tariffa_{i,r,r1})v\_tradeFlows_{i,r,r1}}{v\_nonDoubleZeroEports_{i,r}} \tag{5.94}$$

The unit values exports are used to define the per unit *export subsidies* - *expsub* as shown in the equation below. The parameter *cexps* is used to line up the market equation with the subsidies observed ex-post. Per unit export subsidies hence increase, if market prices *pmrk* increase or export unit values *uvae* drop, or if the share of subsidised exports *exps* on total exports increase. How the amount of subsidised exports is determined is discussed below.

$$expsub_{i,r} = \frac{exps_{i,r}}{exports_{i,r}}(pmrk_{i,r} - uvae_{i,r} + cexps_{i,r}) \tag{5.95}$$

The Armington aggregator functions are already shown in the diagram above. The compositions inside of the Armington composite goods can be derived from first order conditions of utility maximisation under budget constraints and lead to the following conditions: FIXME

$$\frac{v\_arm2Quant_{i,r}}{v\_domSales_{i,r}} = \left(\frac{dp_{i,rw,r}}{dp_{i,r,r}} \frac{pmrk_{i,r}}{arm2pricep_{i,r}}\right)^{\frac{1}{1+\phi_1}} \tag{5.96}$$

Similarly, relations between import shares are determined by:

$$\frac{v\_tradeFlows_{i,r,r1}}{v\_tradeFlows_{i,r,r2}} = \left(\frac{dp_{i,r,r1}}{dp_{i,r,r2}} \frac{impp_{i,r,r2}}{impp_{i,r,r1}}\right)^{\frac{1}{1+\phi_2}} \tag{5.97}$$

5.4.9 Modified Armington import demand system

The standard Armington import demand system in CAPRI has been extended in a study on behalf of JRC-IPTS (154208.X1) in order to tackle the zero trade-flow issue. The zero-trade issue concerns the inability to model emerging bilateral trade relations in the current version of the CAPRI model. Emerging trade flows in this context are those that are zero in the corresponding baseline, but that are expected to be significant under specific scenario assumptions (e.g. trade liberalization scenarios removing prohibitive trade barriers).

Methodology

The traditional approach in Armington trade models applies the CES functional form in a nested import demand structure. The CES functional form, however, cannot be calibrated against zero observations. The easiest illustration probably is to recall the import demand function and import price aggregator from above. In case of zero flows the share parameter $\delta_{i,r}$ becomes zero and thus the trade flows will have no contribution to the price index. Furthermore, the import demand equation cannot assign zero values when prices, share parameters, the quantity aggregate and the substitution elasticity are all positive.

To overcome the inability of calibrating the import-demand system to zero observations the approach of Witzke et al. (2005) has been implemented in CAPRI. This modifies the standard CES form by shifting them with an additional commitment term. That additional (commitment) term requires not only the observed price/quantity points but additional information. The additional information in this case covers expected imports under a hypothetical set of relative import prices that differ from the observed ones. The approach is illustrated in the simple two-goods case, where good 2 is not imported in the calibration point, but is expected to be positive if its relative price to good 1 decreases.

The additional commitment parameter involves another degree of freedom that needs to be eliminated with additional information. During the calibration this is provided by the expected imports from region 2 at the second hypothetical set of relative prices. Following the dual approach, the lower Armington nest is represented with Armington share-equations and with equations for the composite price indexes:

$$M_{i,r,r1} = U_{i,r}\delta_{i,r,r1}^{\sigma_{r,i}} \left[\frac{P_{i,r}}{P_{i,r,r1}}\right]^{\sigma_{r,i}} + \mu_{i,r,r1} \tag{5.98}$$

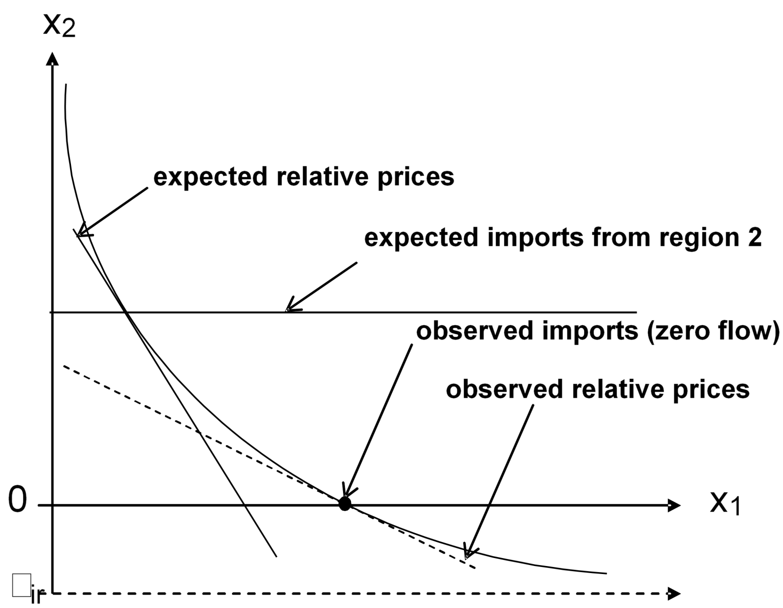
The functional form is almost identical to the standard CES formulation, except the introduction of the commitment parameter μ_i . Geometrically, the isoquant is shifted by the commitment term in the goods space.

Code implementation in the CAPRI market model

The Witzke et al. approach is implemented in a modular fashion in the CAPRI market model. The calibration of the modified Armington lower nest can be switched on or off through a designated button on the CAPRI GUI (see the next Figure). In order not to interfere with the work of other CAPRI users, a specific GUI is created for the project that can be started by running GUI/capri64modarm.bat.

The calibration of the non-homothetic Armington demand system does not require a full re-calibration of the complete CAPRI modelling system; it can be found under the workstep “Run scenario”, task: “Run scenario with market model”. Technically, the calibration modifies the simini/sim_ini.gdx file that contains the necessary starting parameters for a CAPRI simulation: if the above GUI option is selected

Figure 20: Witzke et al. calibration, two-goods case



Source: Witzhe et al 2005

then the existing sim_ini.gdx file is automatically deleted and the create_sim_ini CAPRI module is called. This is all steered in the main capmod.gms file by a specific GAMS setglobal variable called modArmington:

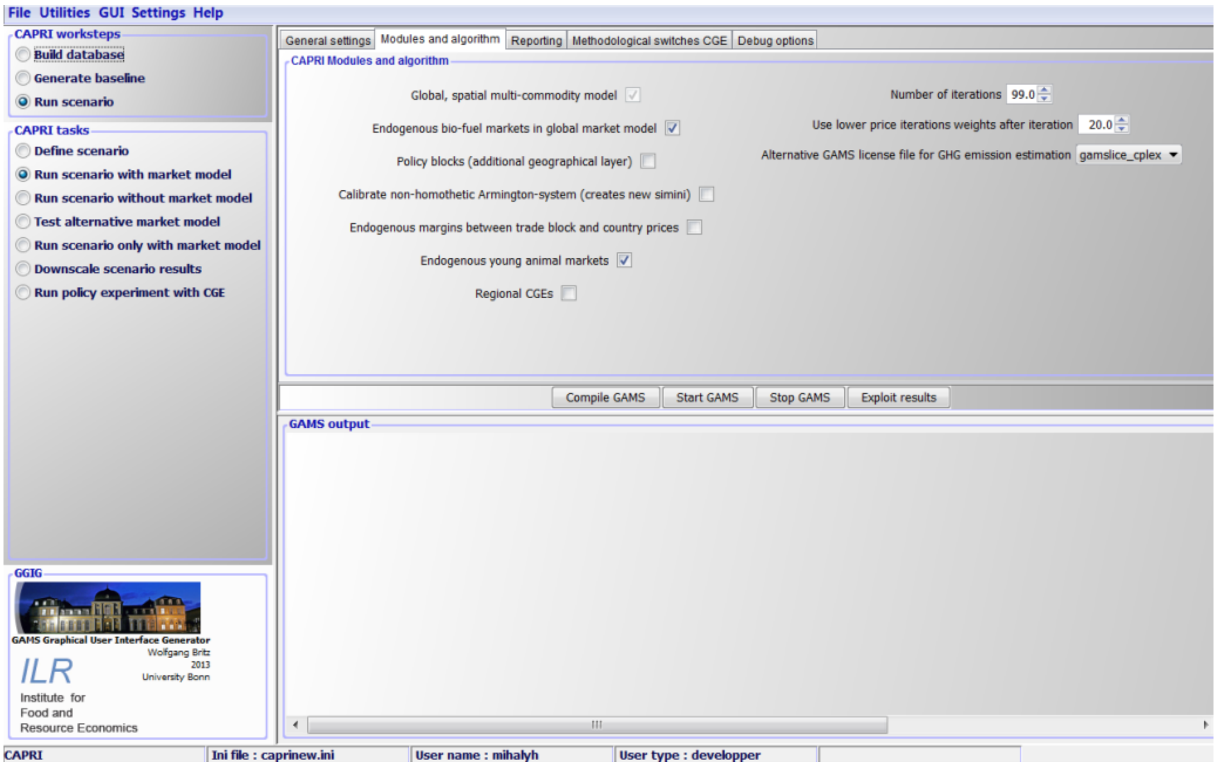
The calibration model itself is called directly by the arm/market1.gms file:

The file arm/modArmington.gms file contains the definition of the calibration model and executes the calibration itself. The calibration model simply consists of Armington share equations (importShares_) and price index equations (arm2PriceAgg __), following the approach presented above.

The share- and price index equations of the calibration model are similar to those in the CAPRI market model, but extended with an additional dimension called 'cal_points'. The additional dimension indicates whether the equations correspond to the observed or the expected calibration points. The arm/modArmington.gms file calculates the expected price/quantity framework necessary for the Witzke et al. approach. The data input for this calculation is currently implemented in the scenario file in order to allow for a compact definition of baseline and scenario assumptions. The test scenario for the Witzke et al. demand system assumes the following expected EU poultry imports from the US under a hypothetical import price:

The calibrated share equations and the commitment terms are then stored in the appropriate parameters and later picked up by the market model. The relevant equations of the market model, therefore, also had to be modified. For example, the Armington share equations of the CAPRI market model are extended

Figure 21: GUI Option for the non-homothetic Armington system




```
===== * * * Top of File * * *
===== --------------------------------------------------------------------------- 133 line(s) not displayed ---
===== *       modArmington: re-calibrates the Armington nest and runs simulation (new simini file generated)
===== ---------------------------------------------------------------------------- 1 line(s) not displayed ----
===== $setglobal modArmington off
===== --------------------------------------------------------------------------- 460 line(s) not displayed ---
===== $ifi %modArmington% == on $setglobal declAmmoSymbols NO
===== --------------------------------------------------------------------------- 195 line(s) not displayed ---
===== $ifi %modArmington% == on execute 'del %results_out%\simini\sim_ini_%NTSLVL%%BAS%%SIM%.gdx';
===== $ifi %modArmington% == on $setGlobal generateSimIni yes
===== --------------------------------------------------------------------------- 903 line(s) not displayed ---
===== * * * End of File * * *
```

with the commitment term (p_arm2Commit = μ):

The calibration of the full market model is tested by solving the model at trend values in the CAPRI module arm/prep_market.gms. This module also had to be modified in order to initialize the modified Armington system appropriately. The modifications mostly affect the trade flows, import prices and trade policy instruments.

If properly calibrated, the modified Armington system with the test reference scenario should replicate the standard baseline results. It means, for example, that emerging trade flows being zero in the baseline will remain zero in the reference run and only become positive under specific scenario assumptions.

```
$iftheni %modArmington% == on

* -- read in the expected price/quantity calibration point
*    for the non-homothetic Armington nest
$setglobal read_modArm on
$include 'pol_input\%result_type%.gms'
$setglobal read_modArm off

$batinclude 'util\title.gms' "'Calibrate the modified Armington demand system with commitment terms'"
$batinclude 'arm\modArmington.gms' %system.fn%
$endif
```

```
   imp_demand_(RM_MODARM,RM1,XXX,cal_points)
     $ ( (not SAMEAS(RM_MODARM,RM1))
*  --- the imp. demand equations are paired with:
*      1) share parameters for the observed points
*      2) commitment parameters for the fix expected demands (import flows with commitment parameters)
*      3) simply with import demand in other cases
*      These equation/variable pairs are reflected in the conditional statment below
     $ (pv_arm2Share.range(rm_modarm,rm1,xxx) or pv_arm2Commit.range(rm_modarm,rm1,xxx) or v_tradeFlows_cal.range(rm_modarm,rm1,xxx,cal_points))
     ) ..


     v_tradeFlows_cal(RM_MODARM,RM1,XXX,cal_points)/(p_tradeFlows_cal(RM_MODARM,RM1,XXX,cal_points)+1.)

              =E= [v_arm2Quant_cal(RM_MODARM,XXX,cal_points)

                 * pv_arm2Share(RM_MODARM,RM1,XXX)

                 * (  v_arm2Price_cal(RM_MODARM,XXX,cal_points)
                   / v_impPrice_cal(RM_MODARM,RM1,XXX,cal_points)) ** p_rhoArm2(RM_MODARM,XXX)
                 + pv_arm2Commit(RM_MODARM,RM1,XXX)]

                 /(p_tradeFlows_cal(RM_MODARM,RM1,XXX,cal_points)+1.)
    ;


* --- The price index equations are simply paired with the price indexes (so simple as that)...

    price_index_(RM_MODARM,XXX,cal_points) $ ( sum(rm1, p_tradeFlows_cal(rm_modArm,rm1,xxx,cal_points))
                                           $   v_arm2Price_cal.range(rm_modArm,xxx,cal_points)
                                           ) ..

     v_arm2Price_cal(RM_MODARM,XXX,cal_points) / (data(RM_MODARM,"Arm2P",XXX,"CUR")+1) =E=

     sum(RM1 $ ( sum(cal_points1, p_tradeFlows_cal(RM_MODARM,RM1,XXX,cal_points1))
             $ (not SAMEAS(RM_MODARM,RM1))),

        pv_arm2Share(RM_MODARM,RM1,XXX) * (   v_impPrice_cal(RM_MODARM,RM1,XXX,cal_points)
                                        /(data(RM_MODARM,"Arm2P",XXX,"CUR")+1) ) ** (1-p_rhoArm2(RM_MODARM,XXX))
       ) ** (1/(1-p_rhoArm2(RM_MODARM,XXX)));
```

```
$label datainput
*
*   ---  Data input: expected price/quantity pairs for bilateral trade
*

table  p_expectedTrade(rm,rm,xx,price_quant) "expected calibration point (price and quantity)"
                        POUM.price       POUM.quantity
EU015000.USA              1600               50
;
```

5.4.10 Biofuel module

An achievement of the CAPRI biofuel module is that biofuel supply and feedstock demand react flexibly to the price ratio of biofuel and feedstock prices as well as biofuel demand and bilateral trade flows react

```
importShares_(RM,RM1,XXX) $ ( (p_tradeFlows(RM,RM1,XXX,"Cur") or p_arm2Commit(rm,rm1,xxx)) $ (not SAMEAS(RM,RM1))
                              $  (v_tradeFlows.range(RM,RM1,XXX) or NonZeroFixedImportFlow(RM,RM1,XXX))) ..

    v_tradeFlows(RM,RM1,XXX)/(p_tradeFlows(RM,RM1,XXX,"Cur")+1.)

            =E= [v_arm2Quant(RM,XXX)

               * p_dpCESTrade(RM,RM1,XXX)

               * (  v_arm2Price(RM,XXX)
                 / (    v_impPrice(RM,RM1,XXX)       $ (NOT NonZeroFixedImportFlow(RM,RM1,XXX))
                   +    v_impShadowPrice(RM,RM1,XXX) $ (    NonZeroFixedImportFlow(RM,RM1,XXX)) )) ** p_rhoArm2(RM,XXX)
               + p_arm2Commit(RM,RM1,XXX)]
                /(p_tradeFlows(RM,RM1,XXX,"Cur")+1.)
    ;
```

flexibly to biofuel prices and further relevant drivers.

Basically two biofuel product markets are covered in the model; Ethanol (BIOE) and Biodiesel (BIOD). For total domestic ethanol production, three technology pathways are covered; 1st generation ethanol (BIOFE) - differentiated in wheat, barley, rye, oats, maize, other cereals, sugar and table wine, 2nd generation ethanol (SECG), and non-agricultural ethanol (NAGR).

A similar technological pathway for biodiesel production has been implemented as shown in Figure 22. The three production pathways for biodiesel are; 1st generation biodiesel (BIOFD) produced from vegetable oils (rape oil, sunflower oil, soya oil, and palm oil), 2nd generation biodiesel (SECG); and non-agricultural biodiesel (NAGR).

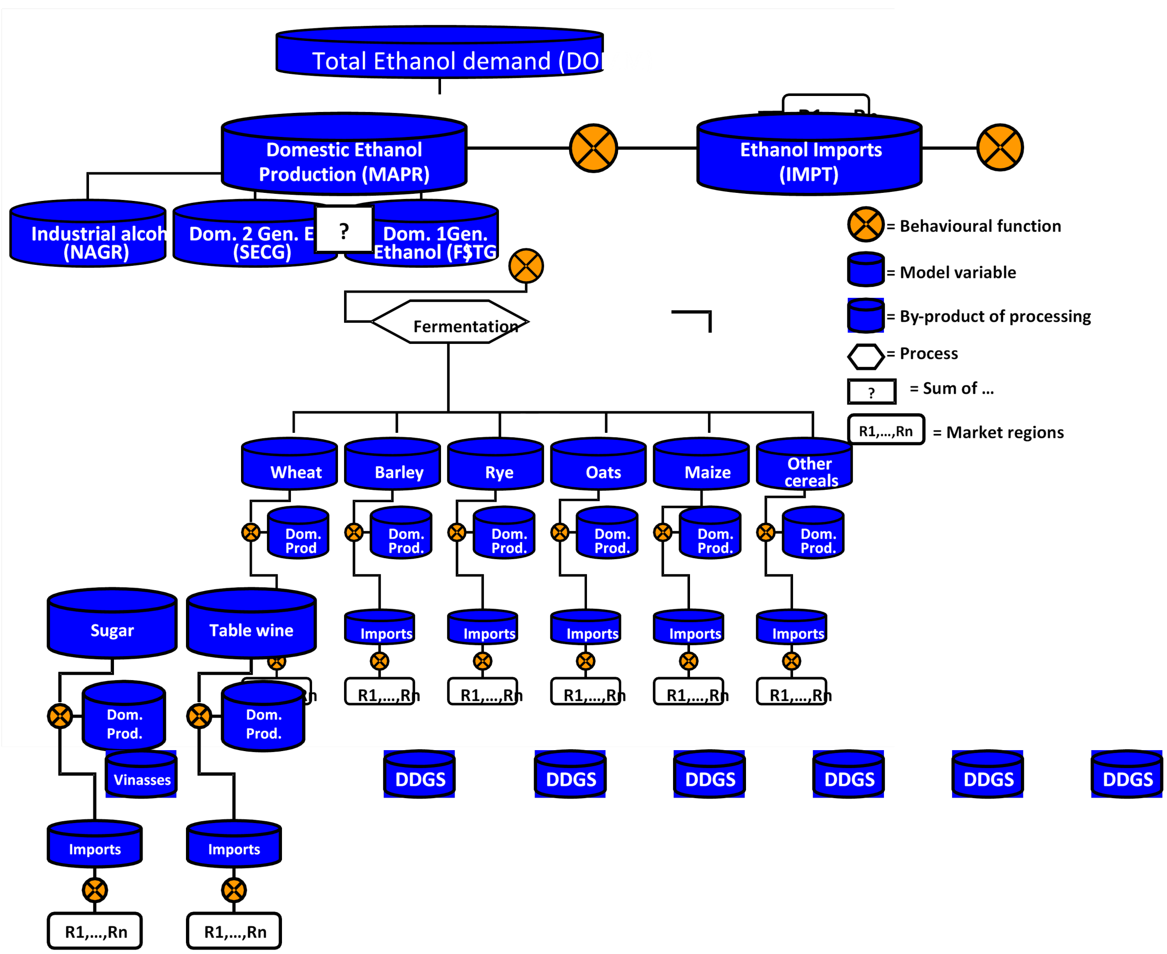
The figure below provides a schematic diagram of the process of 2nd generation biofuel production in CAPRI. Two different product aggregates are introduced in the CAPRI product list to cover feedstock for 2nd generation biofuel processing:

1. A product aggregate for agricultural residues (ARES) which covers straw from cereals and oil seed production and sugar beet leaves and
2. A product aggregate for new energy crops (NECR) which cover herbaceous and woody crops like poplar, willow and miscanthus.

The use of residues from livestock production, which covers manure and cadavers, is not included in the second generation processing as this source is assumed to have only a marginal importance for biofuel processing, Furthermore, the demand shares for the single agricultural residues are provided exogenously in the model. This assumption relies on the observation that the potential of ARES resulting from the activity levels of cereals, oilseeds and sugar beet production in the base and projection year is so high that even in a high second generation scenario (50% of biofuel demand in EU should be stemming from 2nd generation biofuel processing) could only generate a demand of up to max. 10% of the actual potential. The demand share for new energy crops in the second generation production quantities is also provided exogenously in the model. However, as the production of new energy crops require agricultural land the available agricultural land for the production of other agricultural products is reduced accordingly with the yield information collected for NECR.

Biofuel supply and feedstock demand Biofuel feedstock demand is driven by per unit net input costs (μ). These are defined per ton of input used and are calculated for each feedstock in a country by

Figure 22: Construction of the ethanol market implemented in CAPRI FIXME



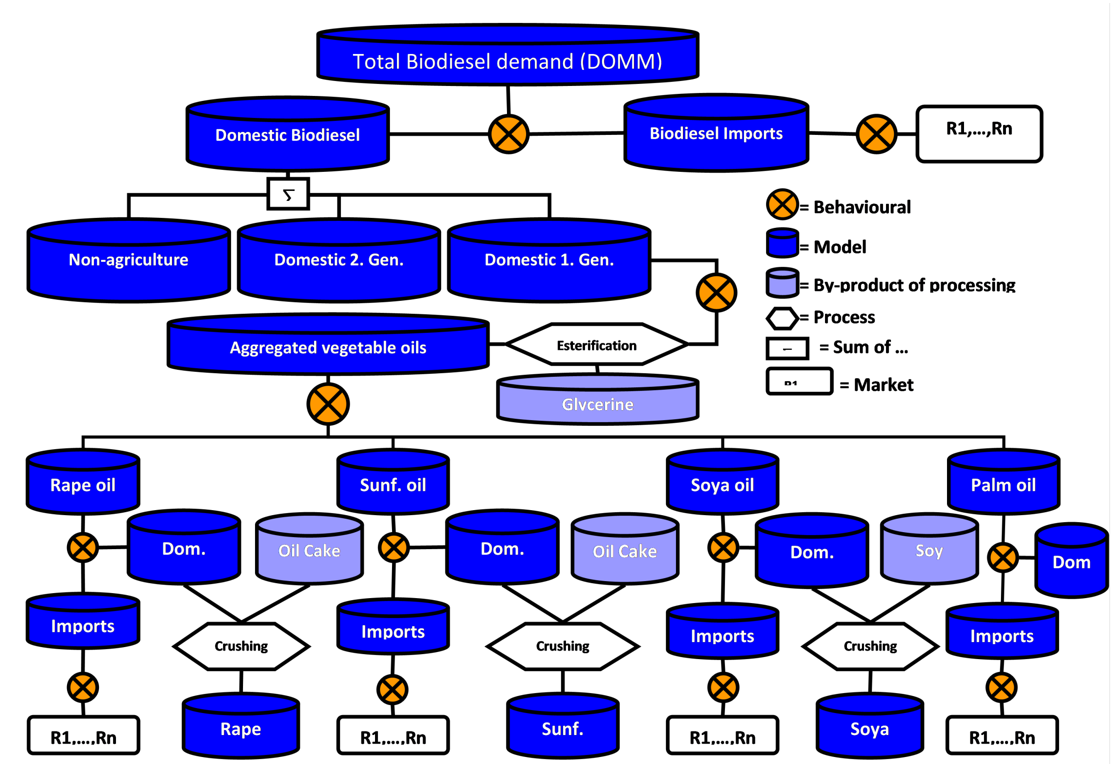
Source: Capri Modelling System

feedstock price minus by product revenues per ton of input.

$$\mu_{r,xf} = p_{r,xf} - \sum_{xbp} p_{r,xbp}\alpha_{r,xf,xbp} \tag{5.99}$$

The index r contains all regions in the market module that have biofuel production. All feedstocks that can be used to produce first generation biofuels are stored in the index xf and the by products Glycerine, DDGs and Vinasses in xbp. Prices are denoted by p. One speciality exists in the case of sugar prices in the EU, where a specific ethanol sugar price is assumed in case of the existence of production quotas. This is due to the fact that ethanol beet in the EU purchased at a lower price than beets processed to

Figure 23: Construction of the biodiesel market implemented in CAPRI



Source: Capri Modelling System

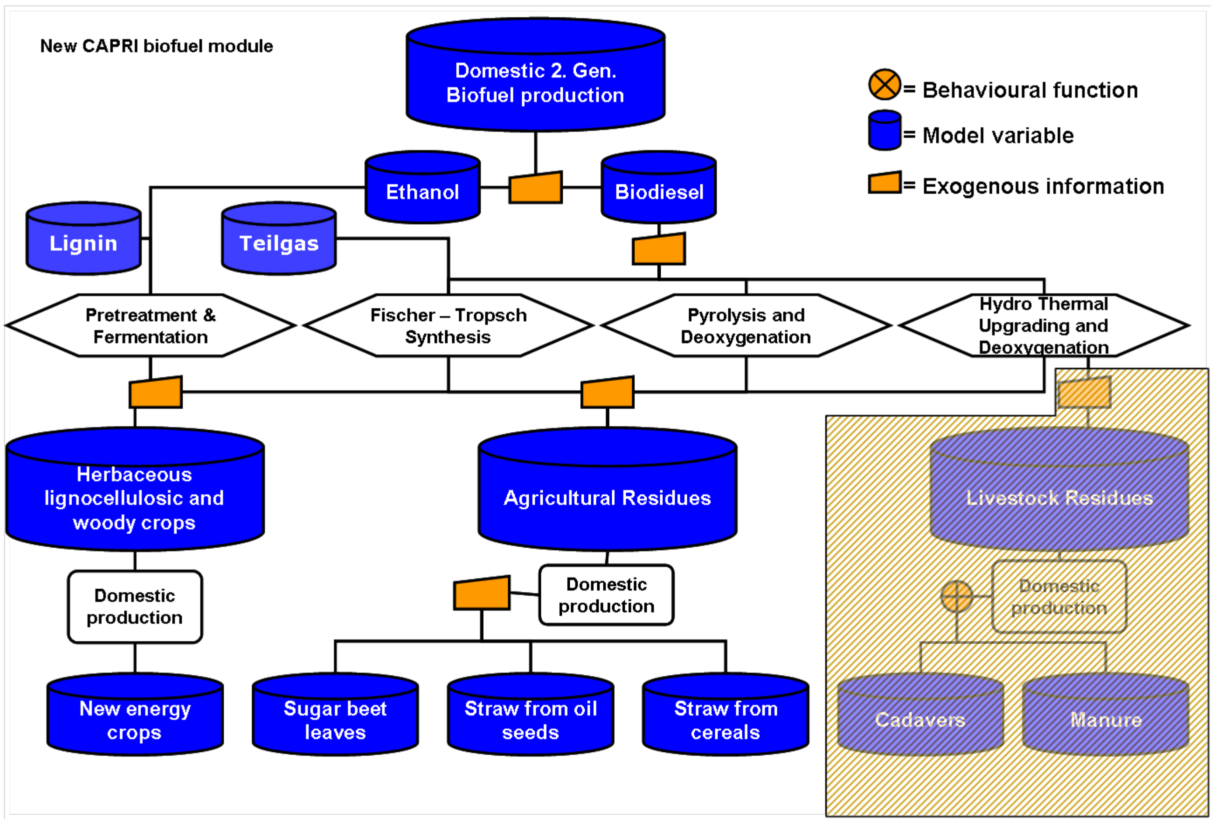
sugar. These single feedstock costs then form a CES aggregate to give the average cost for the respective biofuel:

The index r contains all regions in the market module that have biofuel production. All feedstocks that can be used to produce first generation biofuels are stored in the index xf and the by products Glycerine, DDGs and Vinasses in xbp. Prices are denoted by p. One speciality exists in the case of sugar prices in the EU, where a specific ethanol sugar price is assumed in case of the existence of production quotas. This is due to the fact that ethanol beet in the EU purchased at a lower price than beets processed to sugar. These single feedstock costs then form a CES aggregate to give the average cost for the respective biofuel: FIXME

$$\mu_{r,xb} = \mu^c_{r,xb} - \left[\sum_{xf} \phi_{r,xf} \left[\frac{\mu_{r,xf}}{\mu^c_{r,xb}} \right]^{(1-\rho_{r,xb})} \right]^{\frac{1}{(1-\rho_{r,xb})}} \tag{5.100}$$

The superscript c indicates the average input costs in the calibration point which are only introduced to have price variables “around one” in the expression to help the solver.

Figure 24: Consideration of 2nd generation biofuel production and related feedstock



Source: own illustration

The decision on the total biofuel production happens simultaneously with the decision on the optimal feedstock mix. The latter is based on the FOC of cost minimisation with a CES cost function for a given biofuel output:

$$
\begin{aligned}
fd_{r,xf} &= \phi_{r,xf} x^{1st}_{r,xb} \left[\frac{\mu_{r,xb}}{\mu_{r,xf}} \right]^{\rho_{r,xb}} \qquad \forall xf \neq num \\
s.t. \quad \rho_{r,xb} &= \frac{1}{(1-\sigma_{r,xb})} - 1
\end{aligned} \tag{5.101}
$$

The subsitution elasticity of the CES function is given by and their share parameters by ϕ. First generation biofuel procuction (x^{1st}) is then derived by the product sum of feedstock demand fd and their biofuel processing coefficinents:

$$x_{r,xb}^{1st} = \sum_{xf} fd_{r,xf}\alpha_{r,xf,xb} \tag{5.102}$$

For biofuel supply from first generation technologies (x^{1st}) a function of the relation of the respective biofuel price and the corresponding average feedstock per unit costs has been specified. A synthetic supply function was chosen that satisfied some plausibility considerations, which were 1) that supply strongly decreases when the the price relation gets below a certain "trigger" value and that this strong slope is not maintined throughout the whole function.

$$x_{r,xb}^{1st} = \begin{bmatrix} \partial_{r,xb}\frac{p_{r,xb}}{\mu_{r,xb}} \\ +exp\left(\beta_{r,xb}^{1} + \beta_{r,xb}^{2} ln\left(\frac{p_{r,xb}}{\mu_{r,xb}}\right)\right)\frac{1}{1+exp\left(\left(\frac{p_{r,xb}}{\mu_{r,xb}} - \delta_{r,xb}^{1}\right)\delta_{r,xb}^{2}\right)} \end{bmatrix} \tag{5.103}$$

This function consists of three parts on the RHS: the first part is linear (a small positive value for), the second part is semi-log and the third is sigmoid. The linear term guarantees a minimal slope, where the sigmoid function would return a slope of almost 0. The semi-log term is active at processing margins considerably higher than in the baseline point and the sigmoid function guarantees a steeper slope in a range where processing starts and production is close to zero when feedstock costs exceed output values. The coefficients and are behavioural parameters in these functions. All biofuel supply equations are generally of the style presented below with an example of bioethanol in France.

The supply of by products is directly linked to the first generation biofuel output:

$$x_{r,xbp} = fd_{r,xf}\alpha_{r,xf,sbp} \tag{5.104}$$

Total biofuel output is then defined as the sum over first generation, second generation (secg), non agricultural (nagr) and some exogenous production (exo) from products not mapped to the feedstocks in CAPRI (only relevant in extra EU countries):

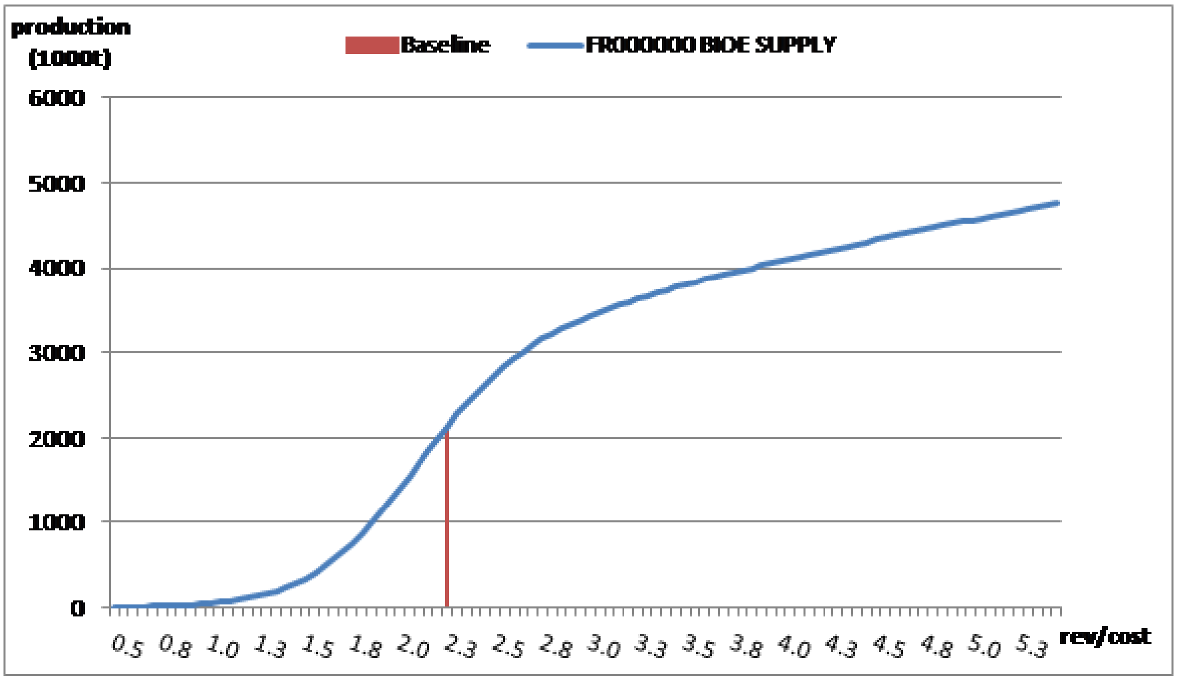
$$x_{r,xb}^{tot} = x_{r,xb}^{1st} + x_{r,xb}^{secg} + x_{r,xb}^{nagr} + x_{r,xb}^{exo} \tag{5.105}$$

Biofuel demand The representation of biofuel demand was simplified compared to the approach chosen first and applied in Becker (2011). There the Aglink demand system was more or less reproduced using a different functional form but keeping the three types of biofuel demand, the use as additive, as low blends and in flexible fuel vehicles. The actual biofuel demand equations consist of only one sigmoid function instead of stacking three of them. The share of biofuel in total fuel demand (bsh) is hereby defined as:

$$bsh_{r,xb} = bsh_{r,xb}^{q} + \frac{bsh_{r,xb}^{max}}{1 + exp\left(\left(\frac{p_{r,xb}}{p_{r,f}} - X_{r,xb}^{1}\right) X_{r,xb}^{2}\right)} \tag{5.106}$$

Again the coefficients X are used to specify the exact slope of these functions. The first term (bshq) defines the part of the biofuel demand which is enforced by any kind of obligation quota or mandate,

Figure 25: Biofuel supply function in France



Source: own calculations

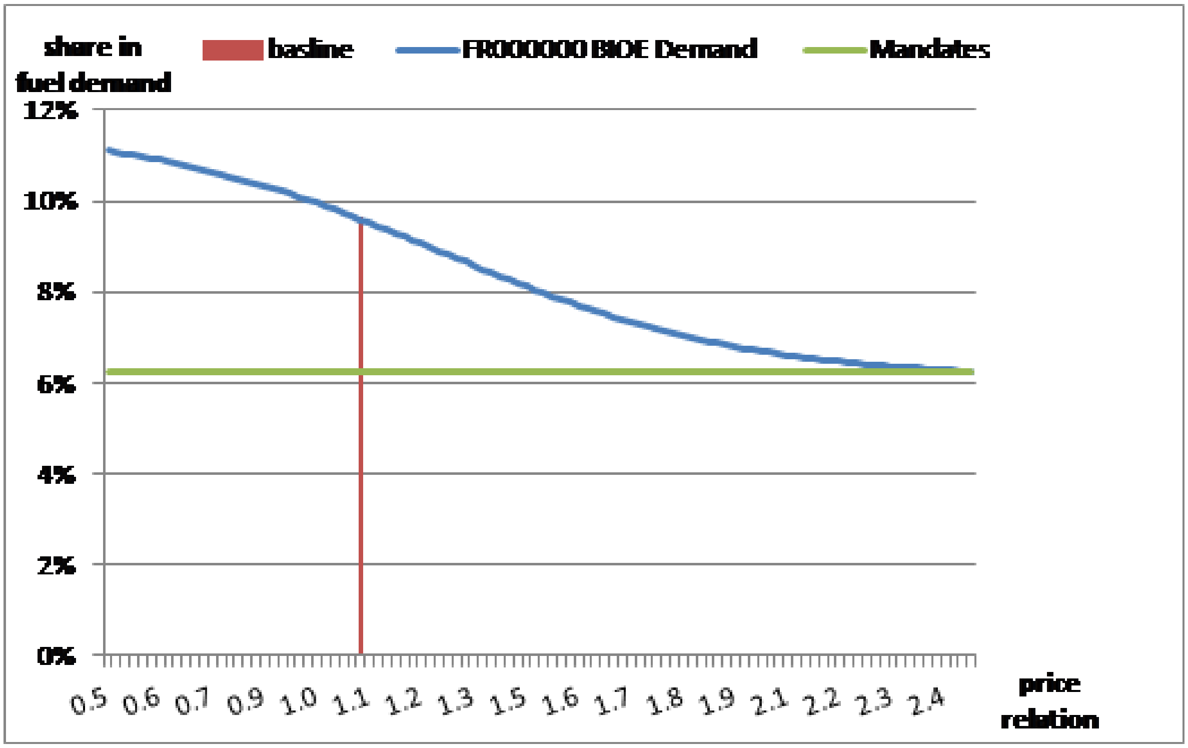
while the second part defines an "endogenous" part of the demand. This term has the upper linmit bsh_{max} which represents the maximum biofuel share on top of the quota obligation that is deemed reachable in a certain country. The endogenous demand component is driven by the price relation of a biofuel ($p_{r,xb}$) to the respective fossil fuel substitute ($p_{r,f}$). These demand share functions are of the type represented by the example of France below:

Total biofuel demand ($d_{r,xb}$) is then derived by multiplying this share to the exogenous total fuel demand ($d_{r,f}$):

$$d_{r,xb} = bsh_{r,xb} + d_{r,f} \tag{5.107}$$

Total fuel demand *Total fuel demand* is exogenous to the CAPRI model. However, an econometric estimation was undertaken to receive a demand reaction on exogenous drivers like the oil price and GDP. This function can then be used in Scenarios to adjust total fuel demand, if these drivers are altered. A response surface estimation on the basis of available PRIMES scenarios from 2008 was undertaken. The PRIMES output files at hand allow for estimating the relation between total fuel demand, GDP and fossil fuel prices. For the estimation an ordinary least square estimator is used. A double log demand function is chosen where the estimation coefficients can directly be interpreted as elasticities. The regression function and thereby the total fuel demand function is defined by:

Figure 25: Biofuel demand share function in France



Source: own calculations

$$\begin{aligned} log(y_{i,j,s,t}) = \delta_{i,j} + \alpha_{i,j}\, log(p_{i,j,s,t}) + \beta_{i,j}\, log(gdp_{j,s,t}) \\ + \gamma\, log(trent_t) + \epsilon_{i,j,s,t} \end{aligned} \tag{5.108}$$

where,
i = Fuel type
r = Region
s = Scenario
t = Year
y = Fuel demand
p = Fuelprice including tax
gdp = Gross Domestic Product
$trend$ = Trend variable
ϵ = Error term for regression
δ = Intercept
α = price elasticity of demand

β = GDP elasicity of demand
γ = Trend elasicity of demand

The results of the regression analysis (differentiated into biodiesel and ethanol for every EU MS) cover estimates for , , and the intercept ( ). The significant estimates are used directly in the respective fuel demand function. If no significance was observed for a coefficient in a respective country, the estimated value is replaced by an average value which is derived from the weighted average of significant coefficients over all EU MS. The resulting matrix of regression coefficients (elasticities) in the fossil fuel demand function are displayed in table below. As the PRIMES data only covers values for European countries but also estimates for the non-European CAPRI regions are required it was assumed that the coefficient estimates for the aggregated EU27 are also applicable for those regions.

Assumed elasticities for total fuel demand after filling with average values are demonstrated in the table below.

Biofuel Trade

Behavioural functions for global bilateral trade of biodiesel and ethanol are intrinsically tied to the final biofuel demand functions. The general methodology is that of a two stage demand system relying on the Armington assumption as already applied for other agricultural commodities in the standard CAPRI version. Biofuel demand for fuel use is considered a derived demand of refineries and responsive to the price ratio of biofuels to fossil fuels. The non fuel demand for biofuels (e.g. ethanol demand of the chemical industry) is consequently set on INDM or PROC (industrial use).

Calibration of the biofuel system So far, only the general form of the biofuel supply and demand functions where derived, but without any adjustments, they won't reproduce the biofuel price-quantity framework of our baseline. Therefore both behavioural functions are due to a calibration process which takes place in '*gams/biofuel/def_biofuel_params.gms*'.

Firstly, the demand system is calibrated. We here assume that only the part of the observed biofuel demand share in total fuel demand that is above the quota obligations is the result of a consumer decision and thus a result of the flexible parts on the demand equations. To calibrate the demand functions to the observed combination of the price ratio bio- to fossil fuel and demand share in total fuel consumption, we chose the two parameters X^1 and X^2 such that:

- It recovers the baseline combination of price and quantity relations
- It reaches 90% of the max share (bsh_{max}) at a certain price relation (currently 0.5 for ethanol and 0.3 for biodiesel)[13].

The maximum biofuel demand share of a region is chosen 2% above the observed baseline share.

The parameters β^2, representing the supply elasticities of the double log part of the biofuel supply equation were chosen at 0.5[14]. For the two parameters the following rules were applied: 1, representing the turning point of the sigmoid function was defined to be left to the calibration point at 90% of the

[13] These values were chosen by trial and error to achieve a reasonable demand response in certain scenarios. However a more empirically based representation of the demand response would greatly improve the system.

[14] An elasticity below 1 turned out to produce more reasonable supply responses as above 1. Again an empirical basis for this is still missing.

Table 29: Overview of pillar II measures modelled in CAPRI FIXME

		β (GDP)	α (price)
EU027000	GASL	0.515	-0.420
EU027000	DISL	0.538	-0.750
AT000000	GASL	0.515	-0.356
AT000000	DISL	0.538	-0.679
BE000000	GASL	0.515	-0.230
BE000000	DISL	0.538	-0.679
LU000000	GASL	0.515	-0.180
LU000000	DISL	0.538	-0.570
NL000000	GASL	0.290	-0.400
NL000000	DISL	0.538	-0.750
DE000000	GASL	0.515	-0.356
DE000000	DISL	0.538	-0.790
FR000000	GASL	0.515	-0.250
FR000000	DISL	0.538	-0.679
ES000000	GASL	0.360	-0.220
ES000000	DISL	0.538	-0.679
PT000000	GASL	0.515	-0.270
PT000000	DISL	0.538	-0.740
UK000000	GASL	0.460	-0.540
UK000000	DISL	0.538	-0.679
IR000000	GASL	0.260	-0.480
IR000000	DISL	0.530	-0.710
IT000000	GASL	0.500	-0.250
IT000000	DISL	0.538	-0.620
DK000000	GASL	0.515	-0.356
DK000000	DISL	0.538	-0.679
FI000000	GASL	0.515	-0.356
FI000000	DISL	0.538	-0.620
SE000000	GASL	0.515	-0.356
SE000000	DISL	0.538	-0.679
EL000000	GASL	0.515	-0.510
EL000000	DISL	0.538	-0.550
PL000000	GASL	0.450	-0.490
PL000000	DISL	0.538	-0.720
HU000000	GASL	0.470	-0.520
HU000000	DISL	0.310	-0.679
CZ000000	GASL	0.515	-0.356
CZ000000	DISL	0.538	-0.730
SK000000	GASL	0.515	-0.356
SK000000	DISL	0.538	-0.800
SI000000	GASL	0.515	-0.356
SI000000	DISL	0.538	-0.550
LT000000	GASL	0.680	-0.260
LT000000	DISL	0.460	-0.650
LV000000	GASL	0.790	-0.390
LV000000	DISL	0.690	-0.770
EE000000	GASL	0.730	-0.440
EE000000	DISL	0.710	-0.750
RO000000	GASL	0.680	-0.450
RO000000	DISL	0.530	-0.679
BG000000	GASL	0.510	-0.190
BG000000	DISL	0.538	-0.720
CY000000	GASL	0.515	-0.270
CY000000	DISL	0.538	-0.679
MT000000	GASL	0.515	-0.356
MT000000	DISL	0.538	-0.430

Source: Own calculation based on PRIMES 2009

processing margin of the baseline. The slope parameter δ^2 defines in which range the sigmoid function increases from 0 to 1. A higher value corresponds then to a steeper slope. Assuming that countries with higher processing margin are more competitive, we assume a higher slope for lower processing margins. Furthermore in non-EU countries we assume the functions less steep. Finally the parameter β^1 is chosen such that the baseline is reproduced.

5.4.11 Endogenous policy instruments in the market model

Subsidised exports On the market side, the amount of subsidised exports (exps) are modelled by a sigmoid function, driven by the difference between EU market (*pmrk*) and administrative price (*padm*), see equation below. The sigmoid function used looks like:

$$Sigmoid(x) = exp\frac{min(x,0)}{1 + exp(-abs(x))} \tag{5.109}$$

where x is replaced by the expression shown below in the equations.

The response was chosen as steep as technically possible by setting a high value for α, i.e. intervention prices are undercut solely if WTO commitment (QUTE) and the maximum quantity of stock changes are reached.

$$expsVal_{i,r} = QutE_{i,r}\left[sigmoid\left(\frac{\alpha_{i,r}}{PADM_{i,r}}(v\_marketPrices_{i,r} - \beta^E_{i,r}PADM_i)\right)\right] \tag{5.110}$$

The parameters α, β are determined based on observed price and quantities of subsidised exports. The per unit subsidy is defined from non-preferential exports and the value of the subsidies:

$$expSub_{i,r} = 1000expsVal_{i,r}/v\_nonDoubleZeroExports_{i,r} \tag{5.111}$$

The relation is shown in the figure below.

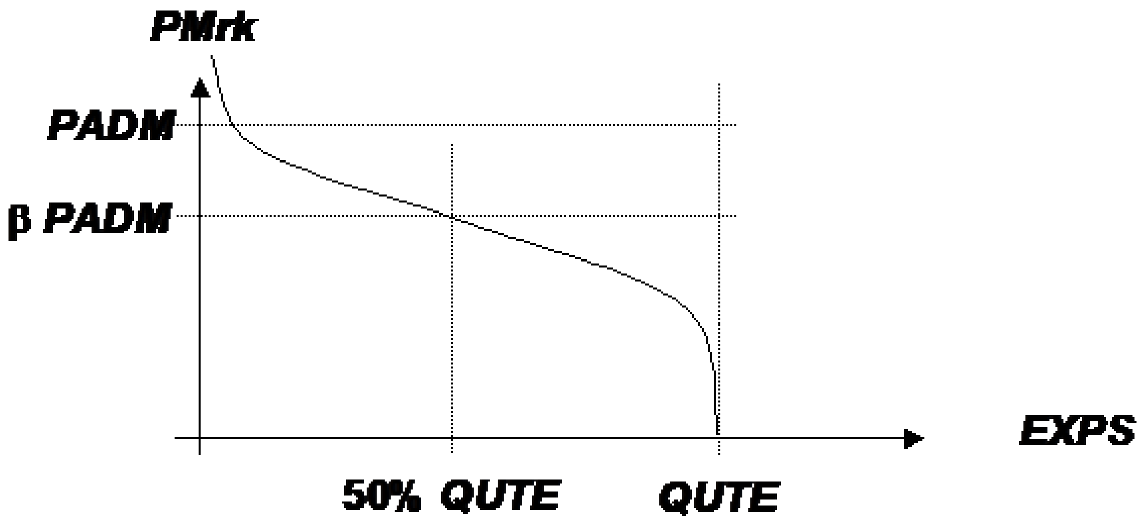
Endogenous administrative stocks For years, the CAP defended administrative prices in key markets such as cereals, beef, butter and skim milk poweder by direct interventions into markets which where out into public stocks. The basic functioning of that mechanism in CAPRI in shown in the figure below.

Purchases to intervention stocks *v_buyingToIntervStock* depend on the probability of the current market price *v_marketPrice* to undercut the administrative price *padm* and a calibration parameter γ^p, assuming a normally distributed market price with standard deviation *stddev* and maximal amounts of purchases *INTM*:

$$v\_buyingToIntervStocks_{i,r} = INTM_{i,r}errf\left(\left(padm_{i,r} - v\_marketPrice_{i,r} + \gamma^p_{i,r}\right)/stddev_{i,r}\right) \tag{5.112}$$

A decrease of the administrative price or an increase of the market price will hence decrease purchases to intervention stocks.

Figure 27: Modelling of subsidised export costs by a logistic function



Releases from intervention stocks *v_releaseFromIntervStock* depend on the probability of market prices *v_marketPrice* to undercut unit value exports *uvae* and a calibration parameter , γ^p, multiplied with the current intervention stock size being equal to starting size *intk* plus intervention purchases *intp*:

$$intd_{i,r} = (intk_{i,r} + intp_{i,r})errf\left(\left(uvae_{i,r} - pmrk_{i,r} + \gamma^p_{i,r}\right)/stddev_{i,r}\right) \tag{5.113}$$

Releases will hence increase if world market price increases or the EU market price drops, and if the size of the intervention stock increases. The parameters γ are determined from ex-post data on prices and intervention stock levels. The change in intervention stocks *ints* entering the market balance is hence the difference between intervention purchases *intp* and intervention stock releases *intd*:

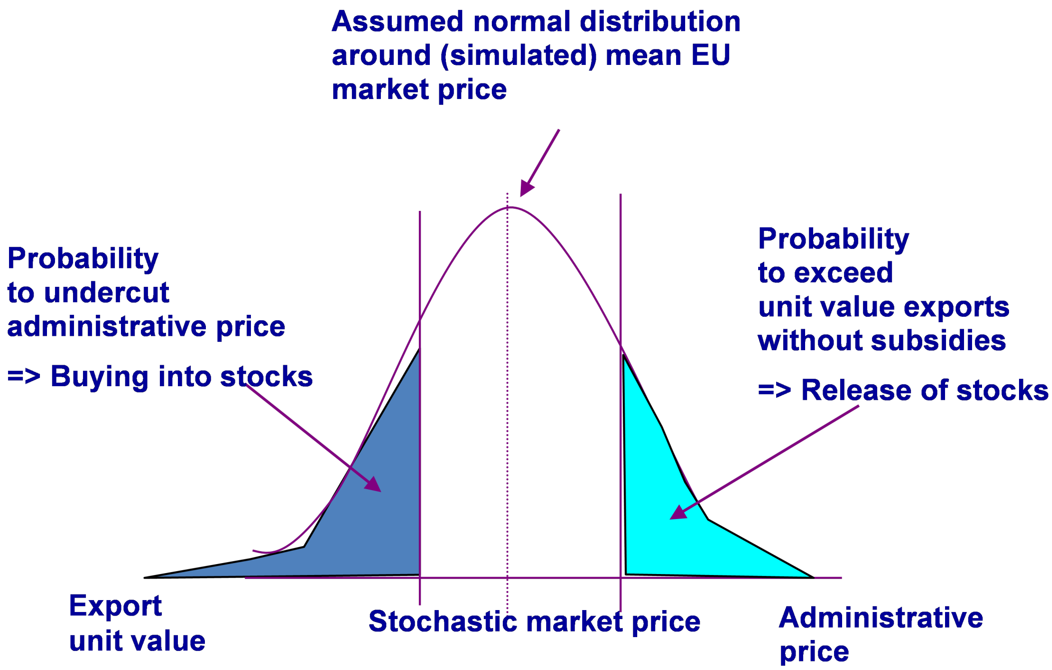
$$ints_{i,r} = intp_{i,r} - intd_{i,r} \tag{5.114}$$

5.4.12 Endogenous tariffs under Tariff Rate Quotas, flexible levies and the minimum import price regime for fruits and vegetables of the EU

Tariff Rate Quotas Tariff Rate Quotas (TRQs) establish a two-tier tariff regime: as long as import quantities do not exceed the import quota, the low in-quota tariff is applied. Quantities above the quota are charged with the higher Most-Favoured-Nation (MFN) tariff. CAPRI distinguishes two types of TRQs: such open to all trading partners, and bi-laterally allocated TRQs. As a rule, bi-lateral allocated quotas are filled first. Equally, as for all tariffs, TRQs may define ad valorem and/or specific tariffs.

A market under a TRQ mechanism may be in one of the following regimes:

Figure 28: Endogenous administrative stocks in CAPRI



Source: own illustration

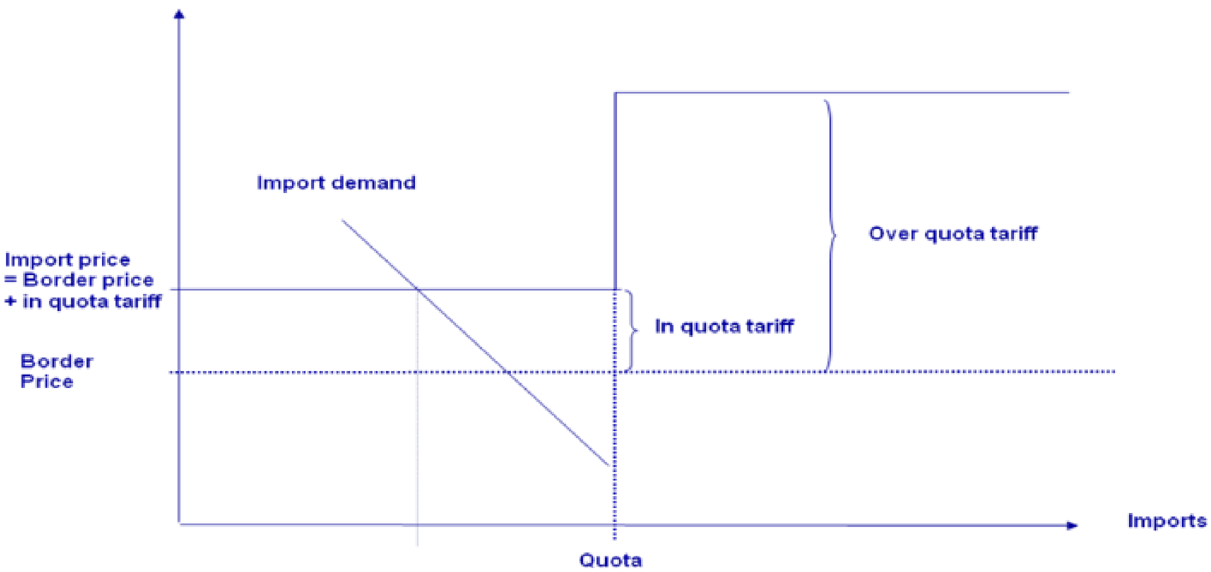
Quota underfill: the in-quota tariff is applied. The willingness to pay of the consumers is equal to the border price plus the in-quota tariff.

Quota binding, i.e. exactly filled: the in-quota tariff is applied. The willingness to pay of consumers and thus the price paid is somewhere between the border plus the in-quota tariff and the border price plus the MFN tariff. The difference between the price in the market and the border price plus the in-quota tariff establishes a quota rent. Depending on property rights on the quota and the allocation mechanism, the quota rent is shared in different portions by the producers, importing agencies, the domestic marketing chain or the administration. Typically, the quota rent can neither be observed nor is their knowledge about distribution of the rent.

Quota overfill: the higher MFN-tariff is applied. The quota rent is equal to the difference between the MFN and the in-quota tariff. Again, how the quota rent is distributed to agents is typically not known.

The fill rate for global TRQs is defined in the code as follows, adding all imports which are not under no duty/not quota access (p_doubleZero), not from the same trade block and not prohibited. A special case provides a bi-lateral quota, here, only import quantities beyond the allocated quota quantity enter the global one.

Figure 29: Quota underfill regime



Source: own illustration

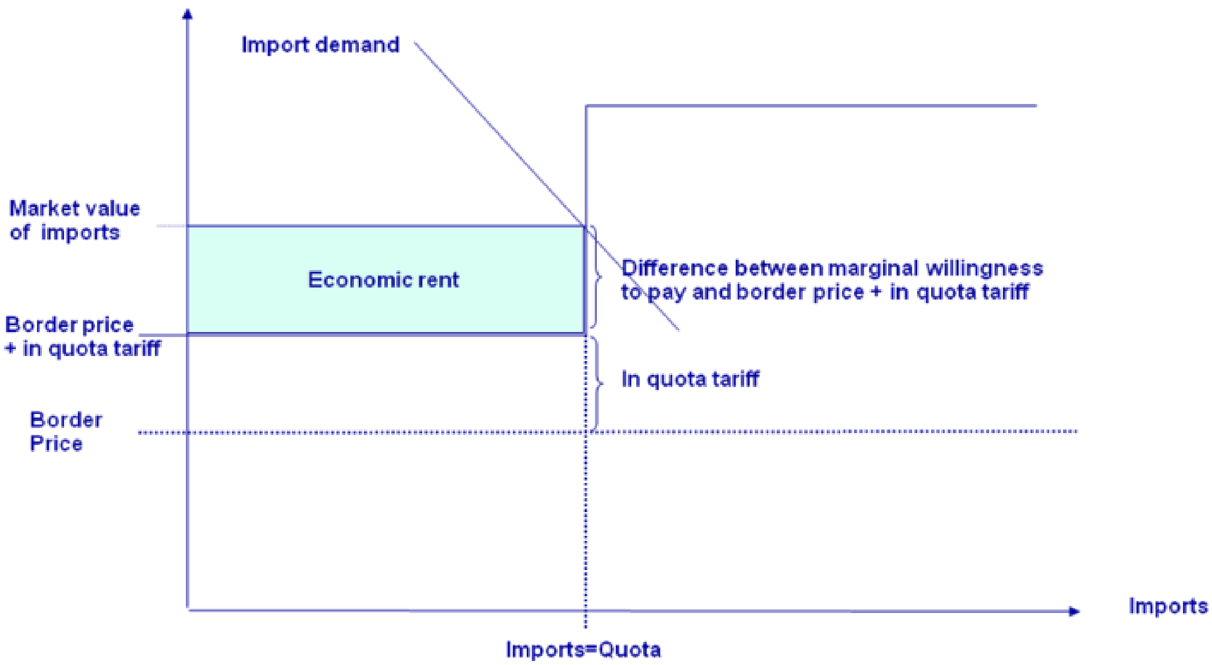
There are a couple of further complications, linked to spatial and commodity aggregation problems. In many cases, TRQs are defined for very specific data qualities, which are more dis-aggregated as the product definition of the model. TRQs for beef may refer e.g. to specific cuts, races or even feeding practises. That typically leads to a situation where both imports covered and not covered by a TRQ mechanism are aggregated in the data base of the model. Consequently, it is not clear, which regime governs the market. Further on, TRQs may be defined for individual countries where the model works on a country block.

Besides the problem of defining the regime ex-post, the relation between the import quantity and the tariff is not differentiable but kinked. Therefore, again a sigmoid function is applied in the CAPRI market part:

In many cases, the EU features for the very same market so-called bi-lateral quotas and market access quotas from the URA round which must be open to all imports ("erga omnes"). As the allocation shares for the latter are currently not know to the CAPRI team, any importer is allowed to import under these global TRQs. Importers have bi-lateral quotas might import under global TRQs once the bi-lateral TRQs are overshot.

Flexible tariffs Geneally, the WTO rules only set upper bounds on the tariffs (so-called Most Favorite Nature or MFN for short rates), but allows its members to reduce the tariffs as long as the same tariff is applied for all WTO members. Exemption from MFN rates which are implemented in CAPRI are preferential rates for Developing countries (Everything But Arms agreement of the EU), Free Trade Agreements and bi-lateral concessions e.g. results from minium market access obligation from the Uruguay rounds under TRQ. The EU generally uses MFN rates, but operates in the cereal markets a specific form of a

Figure 30: Quota binding regime



Source: own illustration

variable tariff called the "levy" system. The last WTO EU trade policy review described the operation of the CAP import regime for cereals as follows: "In response to fluctuations in world prices, the EU has, within the limits of its bound tariffs, changed its MFN applied tariffs. It reduced tariffs on cereals to zero in January 2008 in response to high world prices, and reintroduced them at the end of October 2008. For wheat, the tariff is based on the difference between world prices and 155% of the intervention price, up to the bound rate of €95 per tonne for high quality wheat and €148 per tonne for high quality durum wheat with similar systems for other cereals."

In CAPRI, the system is implemented as follows:

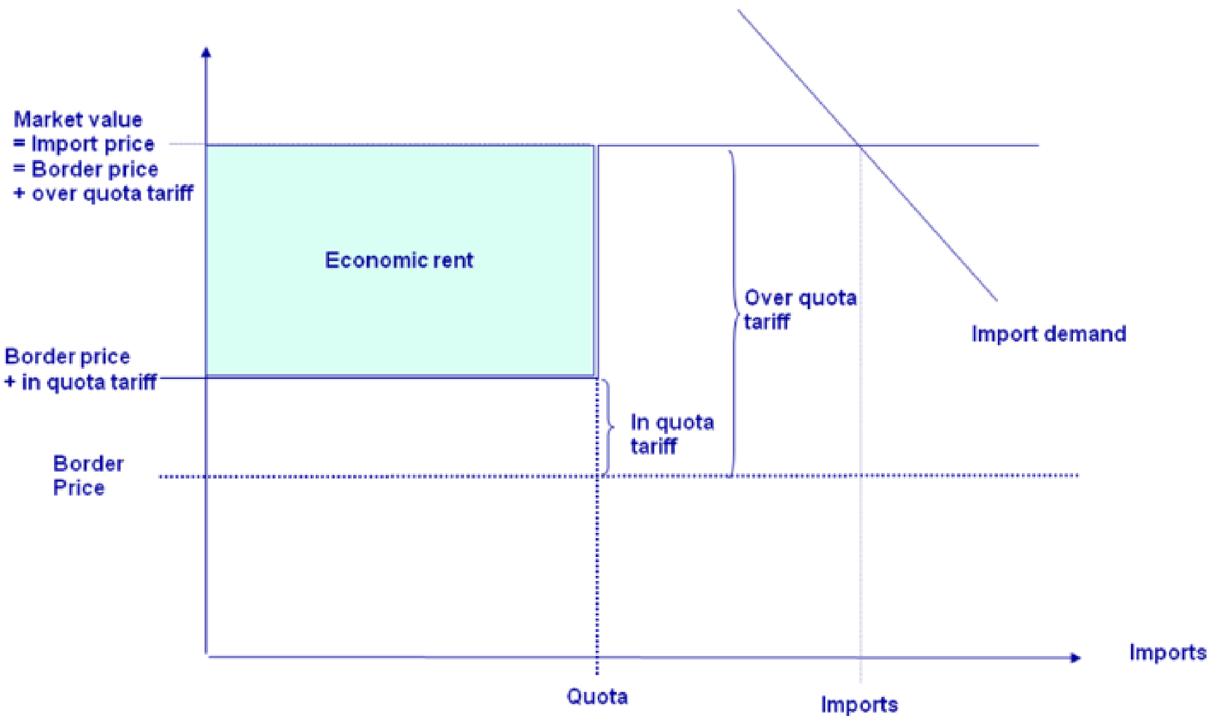
$$v\_flexLevy_{i,r} = min\left[v\_tarSpec_{i,r}, max(0, minBordP_{i,r} - cif_{i,r})\right] \tag{5.115}$$

The actual implementation in the code differs somewhat, as the min and max operators are replaced by "fudging function", and the cif price is expressed as a term from several variables and constants instead of being a separate variable. The first expression is equivalent to the max(0,minBord-cif) expression above:

The second equation defines the actual tariff applied:

Entry price system for fruits and vegetables A somewhat similar instrument is the entry price system used in the fruit and vegetable sectors of the EU. The entry price relates the applied tariff to a

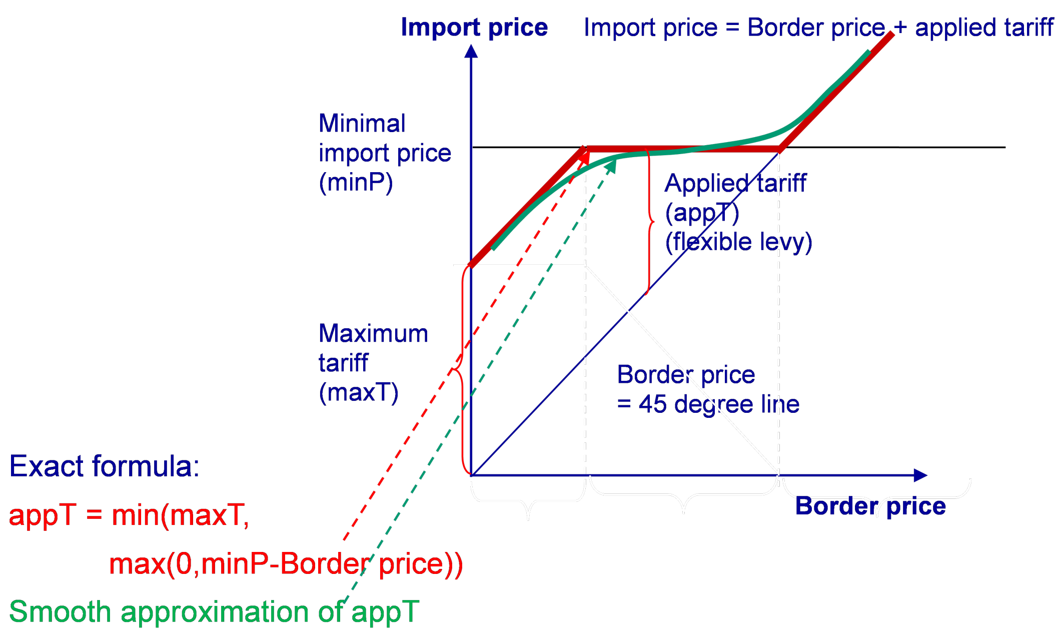
Figure 31: Quota overfill regime



Source: own illustration

```
*
* --- Sum of imports under a non allocated TRQ
*
 TRQImports_(RM,XXX) $ ( p_trqGlobl(RM,XXX,"TrqNT","Cur") gt eps) ..

    v_TRQImports(RM,XXX)/(p_trqGlobl(RM,XXX,"TrqNT","Cur")+1)
        =E=  SUM(RM1 $ (      (NOT p_doubleZero(RM,RM1,XXX,"CUR"))
                            $ (NOT SAMEAS(RM,RM1))
                            $ p_tradeFlows(RM,RM1,XXX,"CUR")
                            $ (NOT p_trqBilat(RM,RM1,XXX,"TrqNt","CUR") eq prohibitive)),

*
*          --- trade flows in case there is not bi-lateral TRQ
*
               v_tradeFlows(RM,RM1,XXX)
                $ ( (p_trqBilat(RM,RM1,XXX,"TrqNt","CUR") le eps) $ p_tradeFlows(RM,RM1,XXX,"CUR"))
*
*          --- overfill of bilateral imports: trade flows exceeding the bilateral quota
*
           + (-ncpcm(-(v_tradeFlows(RM,RM1,XXX)-p_trqBilat(RM,RM1,XXX,"TrqNt","CUR")),
                                    0,1.E-3*p_trqBilat(RM,RM1,XXX,"TrqNt","CUR")))
                $ (p_trqBilat(RM,RM1,XXX,"TrqNt","CUR") gt eps)
                                                           )/(p_trqGlobl(RM,XXX,"TrqNT","Cur")+1);
*
```

Figure 32: Variable levies



Source: own illustration

```
*
* --- Define levy in case of minimal border prices, ensure that it does not get negative
*     (may be turned into a "classical levy" without a bound rate by fixing "tariffs" to a high number)
*
 FlexLevyNotCut_(RM,RM1,XXX) $ ((DATA(RM,"MinBordP",XXX,"CUR") gt eps)
                                $ ( v_flexLevy.lo(RM,RM1,XXX) ne v_flexLevy.up(RM,RM1,XXX))
                                $ p_tradeFlows(RM,RM1,XXX,"CUR")
                                $ (NOT p_doubleZero(RM,RM1,XXX,"CUR")) $ (NOT SAMEAS(RM,RM1))
                                ) ..
*
*    MAX(0,Min. Border price - CIF)
*
   v_flexLevyNotCut(RM,RM1,XXX)/DATA(RM,"MinBordP",XXX,"CUR") =E=

          -ncpcm( -(  DATA(RM,"MinBordP",XXX,"CUR")
                     -(v_marketPrice(RM1,XXX)*p_exchgRateChangeFactor(RM,RM1)+v_transpCost(RM,RM1,XXX))),
                          0,1.E-3*DATA(RM,"MinBordP",XXX,"CUR"))
                                /DATA(RM,"MinBordP",XXX,"CUR");
*
```

specified trigger price in a way that encourages imports at a price (CIF plus tariffs) that is between 92% and 98% of the trigger price.

In order to implement the system, first the difference beween 96% of the entry price and the cif in relation to the triggerprice is defined, times a possible factor to ease solution.

That factor is the fed into a modified sigmoid function which as a result approximates the relations in

```
*
*
* --- Cut flexible levy by specific tariff if it exceeds the bound rate
*     (may be turned into a "classical levy" without a bound rate by fixing "tariffs" to a high number)
*
 FlexLevy_(RM,RM1,XXX) $ ( (DATA(RM,"MinBordP",XXX,"CUR") gt eps)
                                    $ ( v_flexLevy.lo(RM,RM1,XXX) ne v_flexLevy.up(RM,RM1,XXX))
                                    $ p_tradeFlows(RM,RM1,XXX,"CUR")
                                    $ (NOT p_doubleZero(RM,RM1,XXX,"CUR")) $ (NOT SAMEAS(RM,RM1))
                                    ) ..
*
*   MIN( DiffLevies1 =  MAX(0, Min. Border price - CIF), Bound Rate resp. tariff under TRQ)
*
        v_flexLevy(RM,RM1,XXX)/DATA(RM,"MinBordP",XXX,"CUR")
           =E= ncpcm(v_flexLevyNotCut(RM,RM1,XXX),v_tarSpec(RM,RM1,XXX),1.E-3*DATA(RM,"MinBordP",XXX,"CUR"))
                              /DATA(RM,"MinBordP",XXX,"CUR");
```

the graphic shown above:

Tariff computation in the model The figure below depicts the interaction of the various elements of the tariff calculation discussed above. The blue boxes are policy instruments depicted in the model; the purple ones describe endogenous switches and the green ones intermediate model variables which can be interpreted as intermediate results for the applied tariffs. The red boxes show the rate applied to derive the import price. Arrows upwards from a decision box mean yes, to the left no.

The simplest decision is in the left lower corner: a check if the importer benefits from duty and quota free accesss. Examples are intro-EU trade or import from LDCs into the EU under the "Everything But Arms"-Agreement. If that is the case, the applied tariff is zero.

Next we check for a bi-lateral TRQ. If we find one, we check if it is underfilled in which case we apply the in-quota rate. Next we check if the quota is just binding, in which case the applied rate represents the sum of the in-quota rate and an endogenous per unit quota rent. The remaining case is that of a quota overfill where we are left with the MFN rate.

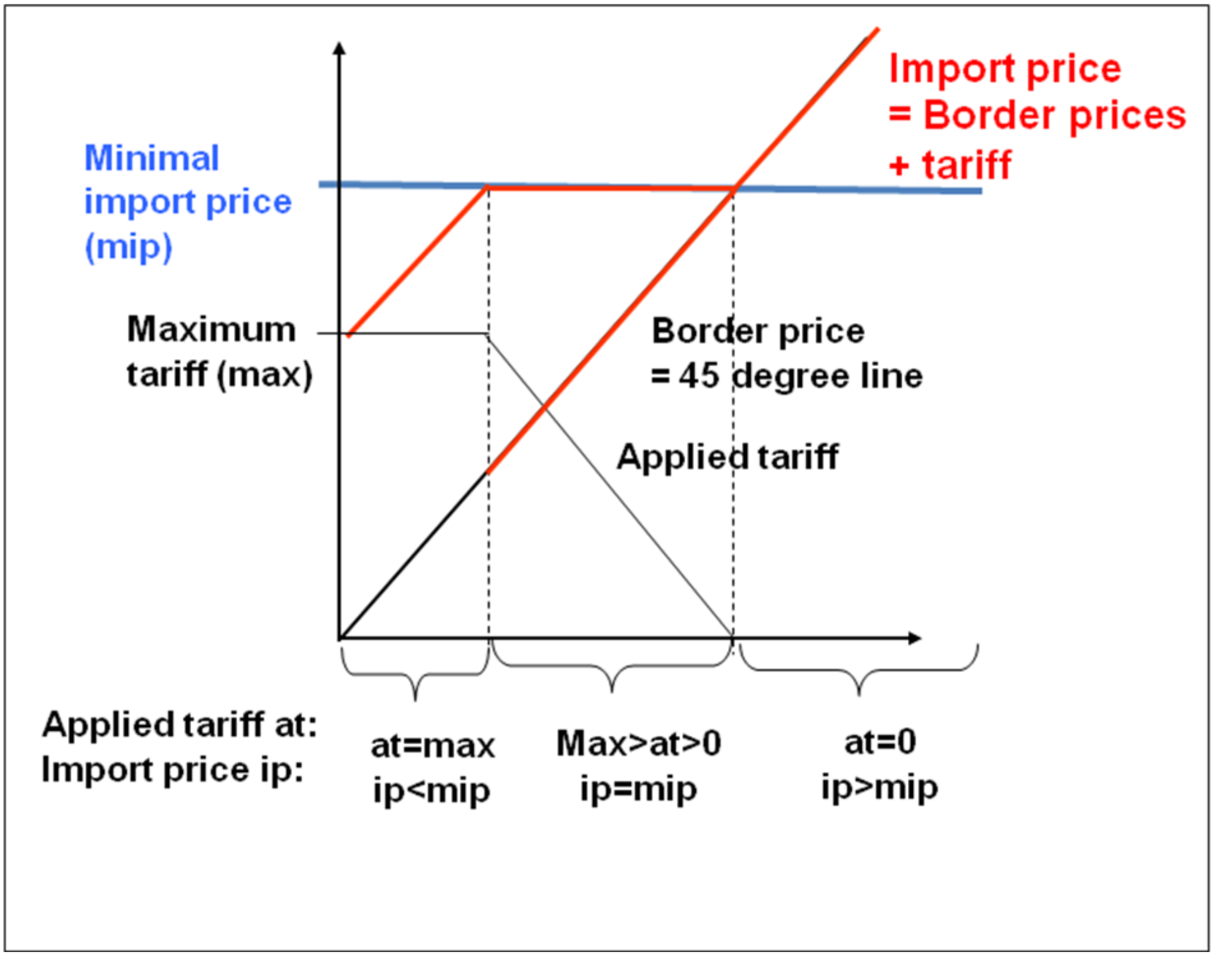
Next we check for a multi-lateral TRQ, also in case we have an overfilled bi-lateral TRQ. If we find one, we check if it is underfilled in which case we apply the in-quota rate. Next we check if the quota is just binding, in which case the applied rate represents the sum of the in-quota rate and an endogenous per unit quota rent. The remaining case is that of a quota overfill in which case the MFN rate is applied.

In all cases for specific tariffs, the results applied rates are checked against the existence of a minimum border price system. In that case, the import price resulting from applying the tariff to the cif price is compared to the minimum price. If it is higher than the minimum price, the tariff is cut such that the import price becomes equal to the minimum border price as long as the resulting tariff does not become negative.

5.4.13 Welfare-consistent tariff aggregation module

The heterogeneity of trade policies across different traded goods generates a serious index number problem for large-scale applied equilibrium modelling. Trade policies must be described with aggregate indices in order to incorporate them in the aggregate commodity structure of applied equilibrium models. Even though the CAPRI commodity coverage is quite detailed in comparison to other applied equilibrium models of global agri-food trade, tariff policies are usually defined and applied at an even finer level of

Figure 33: EU entry price system for fruits and vegetables



Source: own illustration

details, i.e. at the tariff line level. In order to improve on the state-of-the-art in CAPRI, an advanced tariff aggregation module has been developed recently in a JRC-IPTS study (154208.X1) than may be switched on demand (and it the required data are avilable). The stand-alone tariff aggregation module aggregates tariffs and Tariff Rate Quotas (TRQ) defined at the HS tariff-lines into an ad valorem equivalent tariff rate at the level of commodities in the CAPRI market module.

Review of current tariff aggregation approach in CAPRI

The tariff aggregation in CAPRI is based on the weighted average method, but applies a combination of different weights in order to (1) overcome the endogeneity bias and to (2) correct for outliers that are frequently created by statistical errors in the trade data. Technically, the tariff aggregation is performed

```
*
* --- Complex sigmoid function which determines the share of the tariff applied for the EU entry
*     price system for fruits and vegetables
*
 EntryPriceDriver_(RM,RM1,XXX) $ (  (DATA(RM,"TriggerP",XXX,"CUR") gt eps)
                                    $ ((v_entryPriceDriver.LO(RM,RM1,XXX) NE v_entryPriceDriver.up(RM,RM1,XXX)) or p_Trim)
                                    $ p_tradeFlows(RM,RM1,XXX,"CUR")
                                    $ (NOT p_doubleZero(RM,RM1,XXX,"CUR"))
                                    $ (NOT SAMEAS(RM,RM1))) ..
*
*     Entry price system of the EU for fruits and vegetables:
*
*        Tariff is zero at 98% of trigger prices, reduces linear along CIF+Tariff
*         = Triggerprice until CIF=92% of trigger price,
*        and then jumps to MFN bound rate.
*        That is smoothed by using a non-symmetric sigmoid function
*
  v_entryPriceDriver(RM,RM1,XXX) =E=
*
            (v_entryPrice(RM,RM1,XXX) *(0.98+0.92)/2
                - (v_marketPrice(RM1,XXX)*p_exchgRateChangeFactor(RM,RM1)+v_transpCost(RM,RM1,XXX)))
                       / DATA(RM,"TriggerP",XXX,"CUR") * p_entryPriceFac(RM,Rm1,XXX,"CUR");


*
* --- Apply the driver defined above via a second sigmoid function to the bound rate
*
 tarSpecIfEntryPrice_(RM,RM1,XXX) $ (    (DATA(RM,"TriggerP",XXX,"CUR") gt eps)
                              $ ((v_tarSpec.LO(RM,RM1,XXX) NE v_tarSpec.up(RM,RM1,XXX)) or p_Trim)
                              $ p_tradeFlows(RM,RM1,XXX,"CUR")
                              $ (NOT p_doubleZero(RM,RM1,XXX,"CUR")) $ (NOT SAMEAS(RM,RM1))) ..
*
*     Entry price system of the EU for fruits and vegetables:
*
*        Tariff is zero at 98% of trigger prices, reduces linear along CIF+Tariff
*         = Triggerprice until CIF=92% of trigger price,
*        and then jumps to MFN bound rate. That is smoothed by using a non-symmetric sigmoid function
*
   v_tarSpec(RM,RM1,XXX)/DATA(RM,"TARS",XXX,"CUR") =E=
*
       (EXP( MIN(0,v_entryPriceDriver(RM,RM1,XXX)))/(1+20*EXP(-ABS(v_entryPriceDriver(RM,RM1,XXX))))
                        /DATA(RM,"TARS",XXX,"CUR")

             * (  DATA(RM,"TARS",XXX,"CUR")              $    (  P_trqBilat(RM,RM1,XXX,"TsPref","CUR") le eps)
                + p_trqBilat(RM,RM1,XXX,"TsPref","CUR")  $    (  p_trqBilat(RM,RM1,XXX,"TsPref","CUR") gt eps));
```

in the module responsible for creating the global database for the market model. The main information source is the AMAD[15] database providing applied and bound tariffs as well as import statistics of agricultural commodities at the HS6 level. Unfortunately, the AMAD database does not contain bilateral trade statistics. Therefore additional trade statistics had to be requested and added to the CAPRI system (see below).

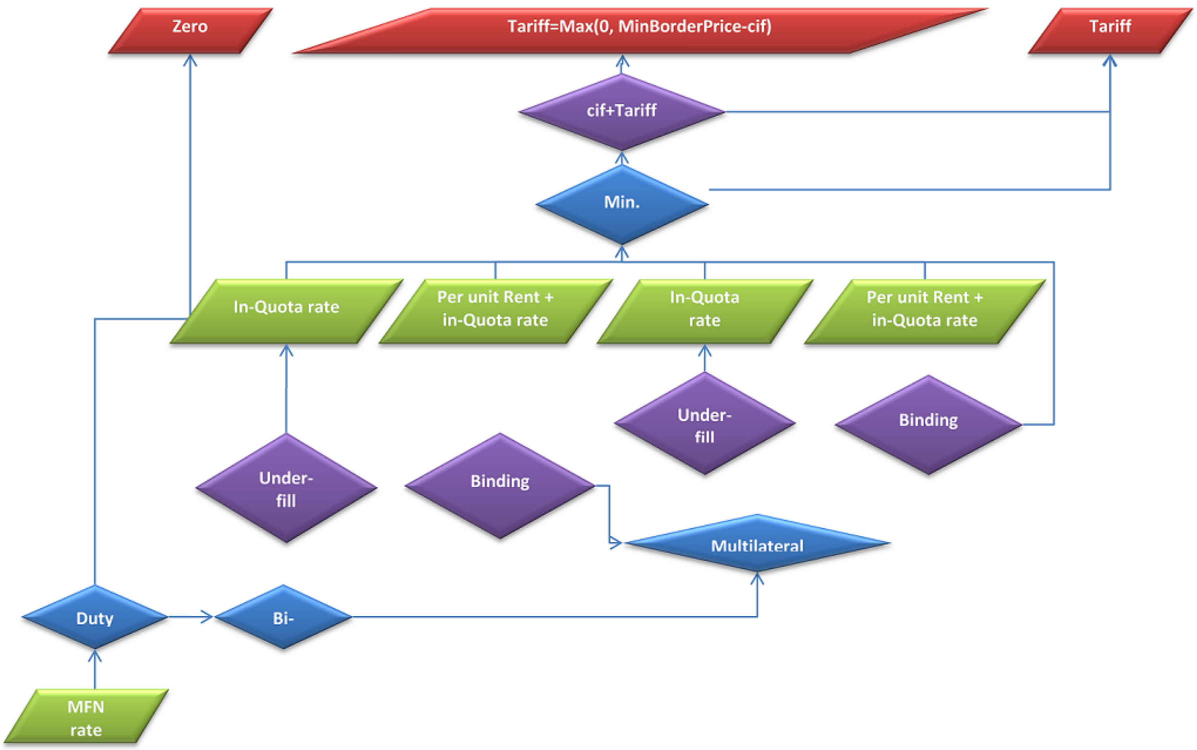
Although Tariff Rate Quotas (TRQ) are typically defined in the legal texts over tariff lines, the CAPRI database does not contain data on TRQs at that level of product aggregation. In fact, TRQs are defined in CAPRI at a much more aggregated commodity level. As a consequence, tariffs under TRQ must be aggregated on an ad-hoc basis to the CAPRI commodity nomenclature before they enter the system. The advanced tariff aggregation module provides an alternative by calculating aggregate tariff equivalents of TRQs systematically, taking into account a more detailed tariff and TRQ structure at the HS6 level.

Advanced tariff aggregation techniques considered

The two fundamental obstacles to aggregate tariffs and other border protection measures are

[15] Agricultural Market Access Database, http://www.oecd.org/site/amad/

Figure 34: Tariff computation in the model



Source: own illustration

- The conversion problem, i.e. different policy instruments need to be expressed in a common metric before they can be aggregated
- The index number problem, i.e. individual trade restrictions must be appropriately aggregated (weighted)

A large number of tariff aggregation techniques are available in the literature, each having its specific objective, drawbacks and merits. Cipollina and Salvatici 2008 provide a typology for the aggregate measures of border protection:

1. Incidence measures are based on the intensities of the policy measures, and are derived only from direct observations on policies. They do not consider the distortive effects of the trade policies on the economy. Typical incidence measures are tariff dispersion or the frequency of various types of Non-Tariff Measures.
2. Outcome measures incorporate other variables than policy variables in order to take into account the distortive impacts of policies on the economy. Typical outcome measures include trade weighted average tariffs. Outcome measures remain 'a-theoretic' in a sense that they do not meet any

equivalence criteria.

3. Equivalence measures provide aggregates that are equivalent to the original data in terms of selected economic variables. Welfare-consistent measures, for example, provide aggregates that are equivalent in their impact on selected indicators of the the economy's welfare (e.g. real income). In order to derive equivalence measures, explicit model structures (supply/demand structures) must be assumed and parameterized.

The advanced tariff aggregation module provides three welfare-consistent aggregators: Bach-Martin approach, Anderson's optimal tariff combination and the Trade Restrictiveness Index (TRI). The procedure developed by Anderson (2009) combines two indicies in a modified balance-of-trade condition for welfare-consistent tariff aggregation. The trade weighted average tariff is combined with the 'true average tariff'. The true average tariff is the tariff aggregator that consistently generates total aggregate trade volume and aggregate prices. The true average tariff alone does not allow for welfare consistent aggregation in the general equilibrium framework. It needs to be combined with the trade weighted average tariff in order to calculate tariff revenues and hence real income correctly. This inconsistency issue has been first raised and addressed by Bach and Martin (2001). They used a revenue constant uniform tariff together with a trade-expenditure constant uniform tariff, as the basis of a compensating variation approach for tariff aggregation. Their approach, however, is only consistent in the one-country case. In the multi-country setup their welfare-analysis fails, because global payments do not balance. One of the key contributions of Anderson (2009) is exactly the solution he provides for the global balance problem. The TRI of Anderson and Neary (1994) preserves real income in shifting from the tariff line structure of trade policies to a single uniform ad-valorem tariff. The TRI is a commonly used measure of aggregate protection, specifically useful for comparing protection level across countries. The implementation of the welfare-consistent tariff aggregators closely follow Himics and Britz, 2014 which includes an extension of the original methods to the case of TRQs. As a result, the tariff aggregation module is able to handle ad-valorem and specific tariffs as well as TRQs at the tariff line level.

A traditional outcome measure, called the MacMap-type aggregator, is also available in the tariff aggregation module. The aggregator is named after the conversion rule for TRQs, which is the same as the one underlying the MacMap database: TRQs are converted to an ad-valorem equivalent based on the fill rate of the TRQ. This approach takes into account the quota rent generated by the TRQ, but defines its level arbitrarily (the unit quota rent is set to half of the difference between out-of-quota and in-quota rates).

Additional data requirement and its integration in CAPRI

An extraction from the UN-COMTRADE database has been integrated in the global module of CAPRI. The dataset comprises of import values and calculated unit prices for 326 tariff lines (mainly agricultural commodities), covering 79 reporting and 99 partner countries, for the years 2007-2013.

A major difficulty arises when we use raw UN-COMTRADE data for modelling purposes, due to their lack of symmetry. Country A's import of a given product from country B is not the same as country B's export of that product to country A. The literature identifies three main causes for this discrepancy (McCleery and DePaolis, 2014):

- An obvious wedge between import and export values is created by the valuation of exports at point of origin (usually f.o.b. prices) and the valuation of imports at destination (mostly c.i.f. prices).

- Border disputes, export bans or prohibitive trade restrictions may lead to only one half of the trade transactions being recorded.
- Border frictions may also lead to distorted trade statistics, e.g. large discrepancies in the US-China trade statistics can be observed due to recording trade with Hong-Kong differently

This problem is currently solved by using UN-COMTRADE data only on imports, applying the assumption that countries tax and regulate imports more thoroughly than exports.

The only source of trade policy data at the tariff line level in the current CAPRI system is the AMAD database. AMAD is not anymore updated by OECD, and in many respect contains outdated policy information. According to the Technical Specification, the current study relies on the AMAD data, and does not include and update of trade policies at the tariff line level. This is a clear limitation that should be improved on in the future. It is very likely that a combination of additional trade policy databases need to be added to CAPRI for a correct and comprehensive representation of global trade policies at the tariff line level. MAcMap is one of the candidates to be included, containing global trade protection measures at HS6 level. MAcMap, however, contains TRQs already in a converted tariff-equivalent form, and is therefore insufficient to provide policy data for the explicit TRQ mechanism in the new tariff aggregation module.

The code implementation of the UN-COMTRADE data processing is modular, i.e. it can be switched on and off upon demand. A dedicated option in the GUI activates the data processing algorithms in the global part of CAPRI (see below). Technically, the extended GAMS routines create an additional intermediate dataset (' */global/tariff_aggregation.gdx*') that is a direct input of the tariff aggregation module in subsequent steps. The .gdx file contains bilateral trade and tariff information at the tariff line level, already mapped into the CAPRI regional nomenclature.

The different tasks implemented in the aggreg_tariffs.gms tariff aggregation module includes:

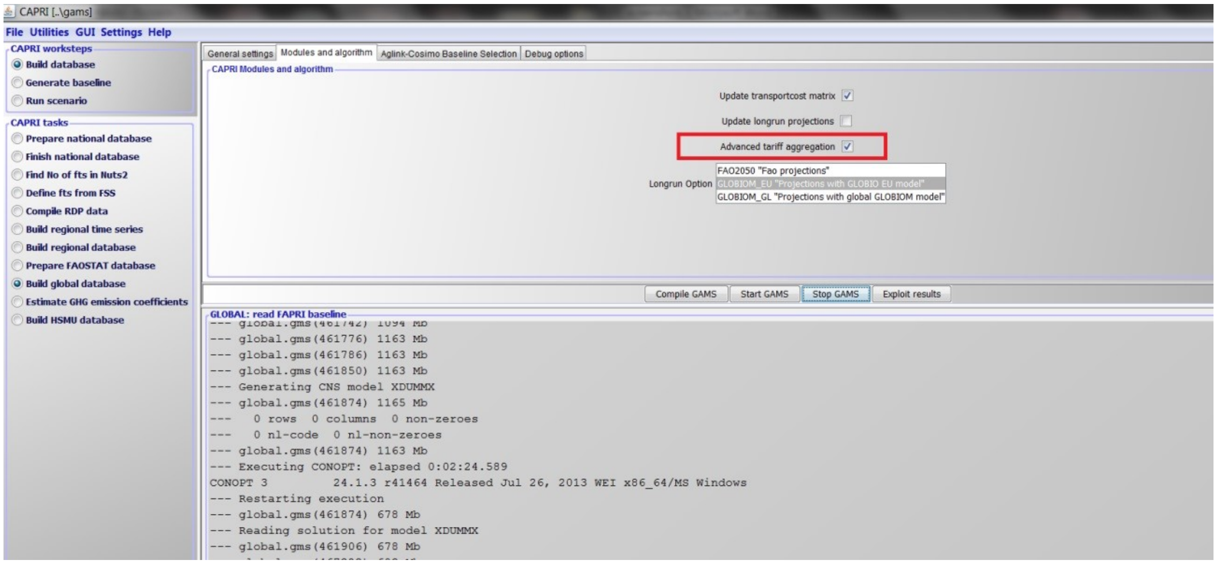
- Defining nomenclatures and sets for the UN-COMTRADE dataset ('global/comtrade_sets.gms')
- Processing, filtering and mapping UN-COMTRADE data in order to align it with the CAPRI database
- Aggregate tariffs to the CAPRI regional nomenclature. The aggregation follows the standard CAPRI approach; the only difference is that tariffs are not aggregated over tariff lines.

Unit values in the UN-COMTRADE dataset have been found to be subject to significant statistical errors. An outlier-detection algorithm has been therefore implemented in order to tackle this problem. Using a simple and robust approach, observations outside a given range around the mean are identified as outliers and replaced with the mean. The outlier detection is implemented in R[16], and can be easily combined with other outlier detection algorithms. Improvements of outlier detection algorithms (subject to future research) can further increase the accuracy of the tariff aggregation. Technically, the GAMS code automatically communicates with the outlier detection algorithms in R, and performs the data checks on the raw UN-COMTRADE data.

Defining Tariff cut scenarios at the tariff line level

[16] As a consequence, the current implementation requires R to be installed on the computer running the tariff aggregation module.

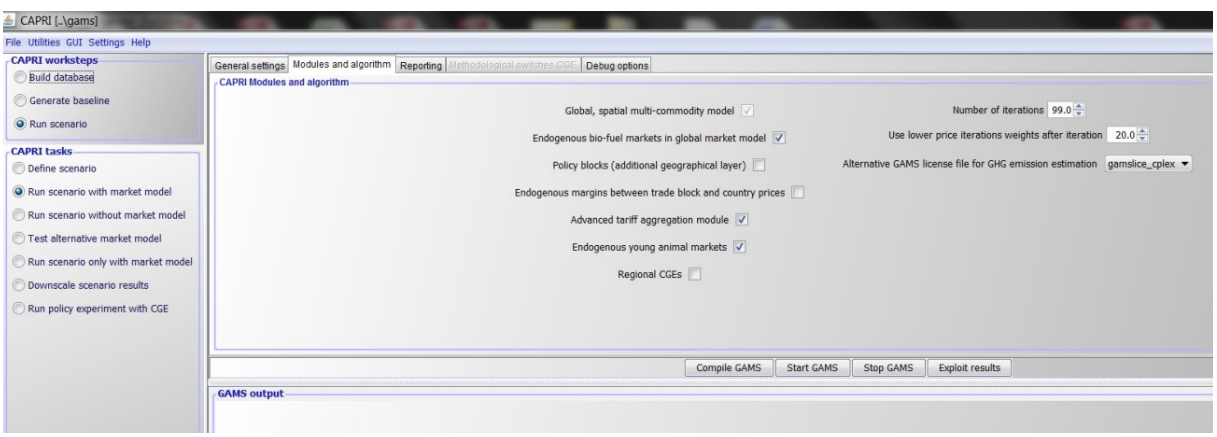
Figure 35: Tariff computation in the model



Source: CAPRI Modelling System

The tariff aggregation module is loosely linked to the CAPRI modelling system. The feature can be activated by a GUI option (see below), assuming that the intermediate database has been already created by the global module.

Figure 36: Activation of the tariff aggregation module on the GUI



Source: CAPRI Modelling System

The tariff aggregation module takes over the appropriate tariff cuts from the scenario file and applies them at the tariff line level. The module then feeds back an aggregate tariff equivalent of the resulting (cut) tariffs.

CES demand structure

With specific assumptions on the demand side, it is possible to take into account the changes in the consumption bundle, as a response to relative price changes induced by the tariff cut scenario. Loosely speaking, the tariff aggregation module takes into account the substitution between goods within an aggregated CAPRI commodity, under specific assumptions on the demand structure. We assume a nested CES import demand structure on the lines of the usual Armington approach to model bilateral imports with cross-hauling. The current implementation is a “small country” approach: tariffs are aggregated for one importer region after the other, assuming fix border (c.i.f.) prices.

Dropping the small country assumption would require a full partial equilibrium model at the tariff line level for the commodity considered, as done in e.g. Grant et al. (2007). This, however, requires a substantial extension of the CAPRI database by including tariff line-specific trade and trade policy information at the global scale. The complexity of the market model would increase rapidly by adding trade flows at the tariff line level. A simultaneous solve for all commodity markets would be technically impossible. No surprise that similar examples in the literature only focus on selected markets and do not implement a full-fledged market model with many interacting markets at the tariff line level. Grant et al. (2007) focus on the dairy market only, and implements a sequential model linkage in order to reduce the computational requirements of solving the complete system whereas Narayanan et al. (2010) extend the standard GTAP model with a partial equilibrium component that only covers the automobile industry.

Technical details of implementing tariff aggregation scenarios

Generally, tariff cuts have to be defined in a specific format in the CAPRI scenario file. A simple example is illustrated in hte following, where ad-valorem and specific tariffs are cut relative to their initial level and TRQ thresholds are increased.

Reporting (GUI tables)

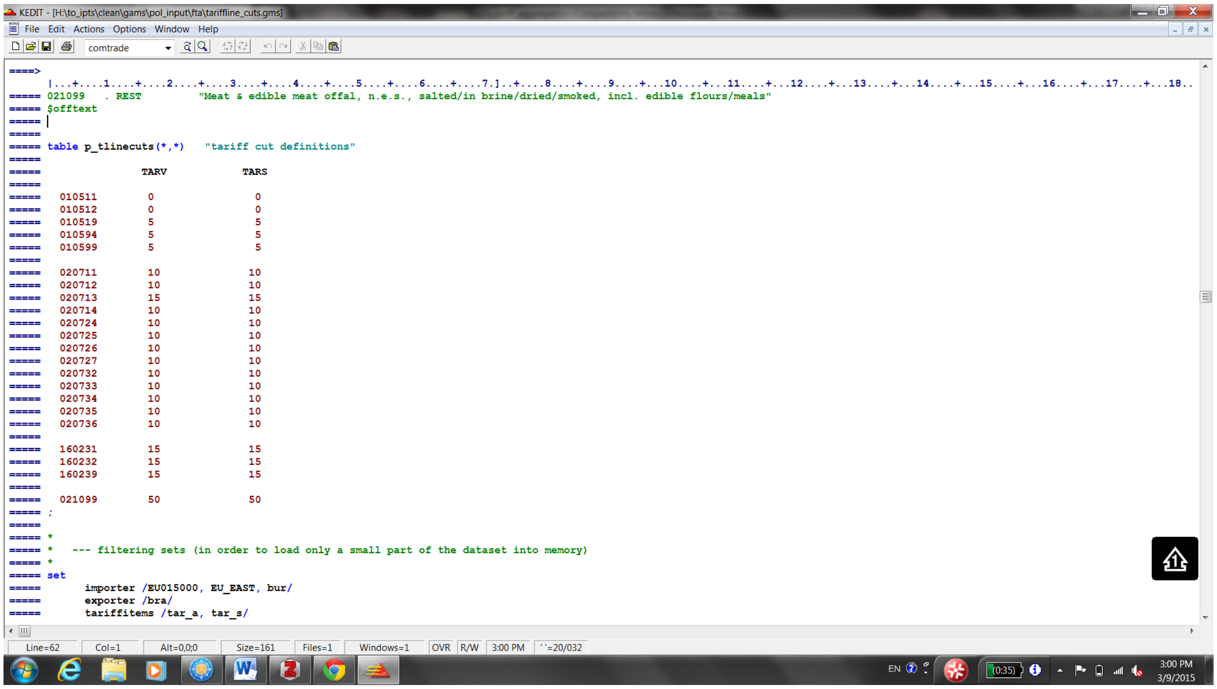
The GUI has been extended with tables under ’Trade|Advanced tariff aggregators’ that collect the results of the tariff aggregation module:

The MacMap-type aggregators are calculated both with respect to bilateral trade relations and with respect to a total (’from World’) measure of tariff/TRQ restrictions. In the above calculation of total aggregate measures, imports from other countries are neglected; an assumption that can be easily relaxed if relevant policy and trade data will be available in future applications. MacMap-type trade weighted aggregators give the following results.

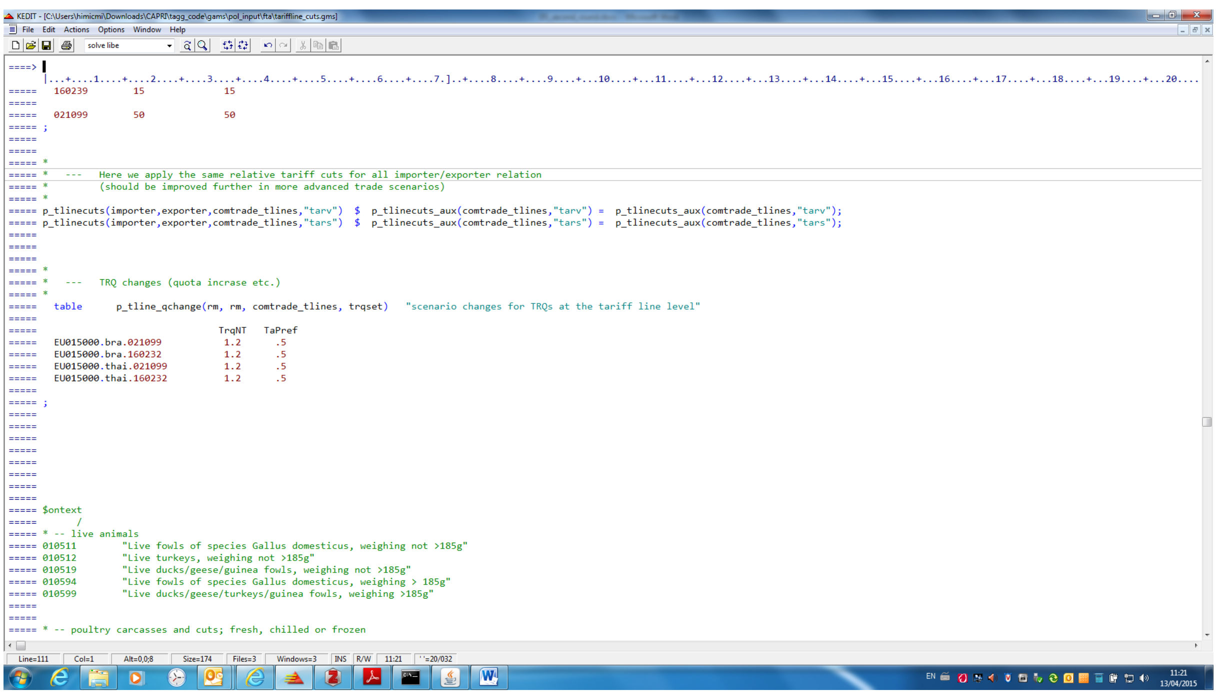
TRI estimates are also reported in a specific GUI table. By definition the TRI indicies are defined for all trade relations only (not a bilateral index):

The Anderson tariff combination is presented next. The current implementation is an extension of the original approach, including correction factors for TRQs (Himics and Britz, 2014):

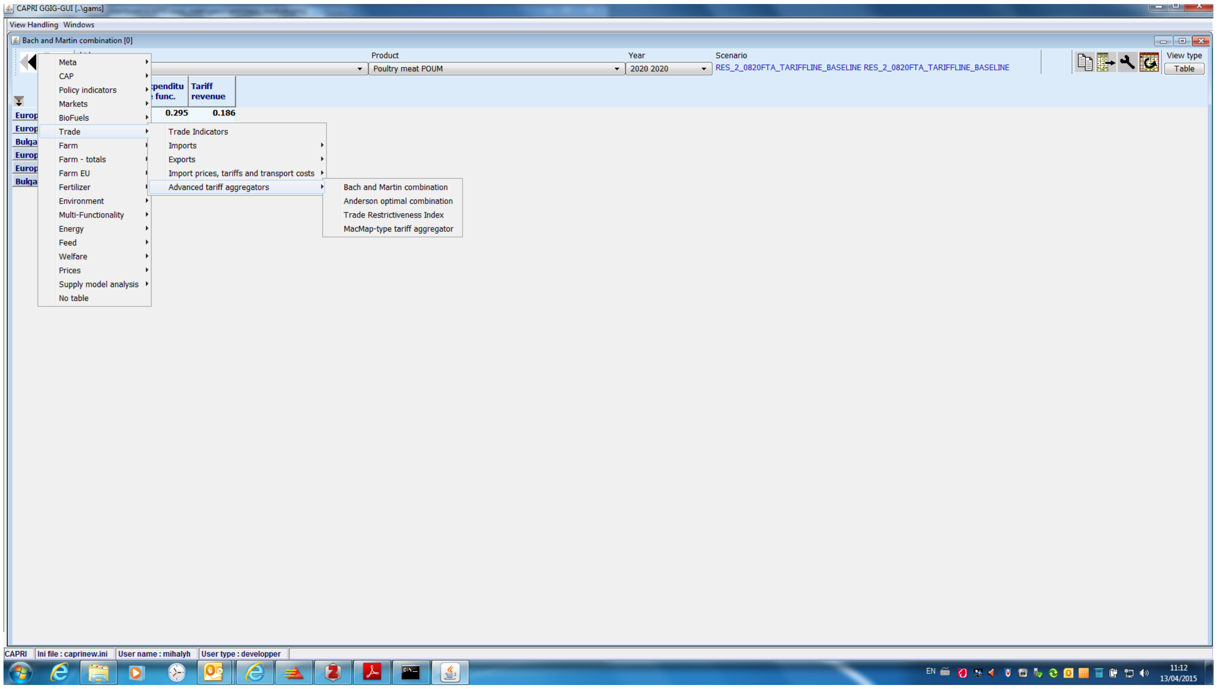
The Bach and Martin (2001) approach, i.e. a combination of an aggregator for the expenditures and another one for the tariff revenues, is also implemented and reported in a designated GUI table:



```
021099   . REST          "Meat & edible meat offal, n.e.s., salted/in brine/dried/smoked, incl. edible flours/meals"
$offtext


table p_tlinecuts(*,*)   "tariff cut definitions"

                TARV            TARS

  010511         0               0
  010512         0               0
  010519         5               5
  010594         5               5
  010599         5               5

  020711         10             10
  020712         10             10
  020713         15             15
  020714         10             10
  020724         10             10
  020725         10             10
  020726         10             10
  020727         10             10
  020732         10             10
  020733         10             10
  020734         10             10
  020735         10             10
  020736         10             10

  160231         15             15
  160232         15             15
  160239         15             15

  021099         50             50
;

*
*   --- filtering sets (in order to load only a small part of the dataset into memory)
*
set
      importer /EU015000, EU_EAST, bur/
      exporter /bra/
      tariffitems /tar_a, tar_s/
```



```
  160239        15              15

  021099        50              50
;


*
*   ---   Here we apply the same relative tariff cuts for all importer/exporter relation
*         (should be improved further in more advanced trade scenarios)
*
p_tlinecuts(importer,exporter,comtrade_tlines,"tarv")  $  p_tlinecuts_aux(comtrade_tlines,"tarv") =  p_tlinecuts_aux(comtrade_tlines,"tarv");
p_tlinecuts(importer,exporter,comtrade_tlines,"tars")  $  p_tlinecuts_aux(comtrade_tlines,"tars") =  p_tlinecuts_aux(comtrade_tlines,"tars");



*
*   ---   TRQ changes (quota incrase etc.)
*
  table      p_tline_qchange(rm, rm, comtrade_tlines, trqset)   "scenario changes for TRQs at the tariff line level"

                               TrqNT   TaPref
  EU015000.bra.021099           1.2      .5
  EU015000.bra.160232           1.2      .5
  EU015000.thai.021099          1.2      .5
  EU015000.thai.160232          1.2      .5

;







$ontext
      /
* -- live animals
010511        "Live fowls of species Gallus domesticus, weighing not >185g"
010512        "Live turkeys, weighing not >185g"
010519        "Live ducks/geese/guinea fowls, weighing not >185g"
010594        "Live fowls of species Gallus domesticus, weighing > 185g"
010599        "Live ducks/geese/turkeys/guinea fowls, weighing >185g"


* -- poultry carcasses and cuts; fresh, chilled or frozen
```



Source: CAPRI Modelling System

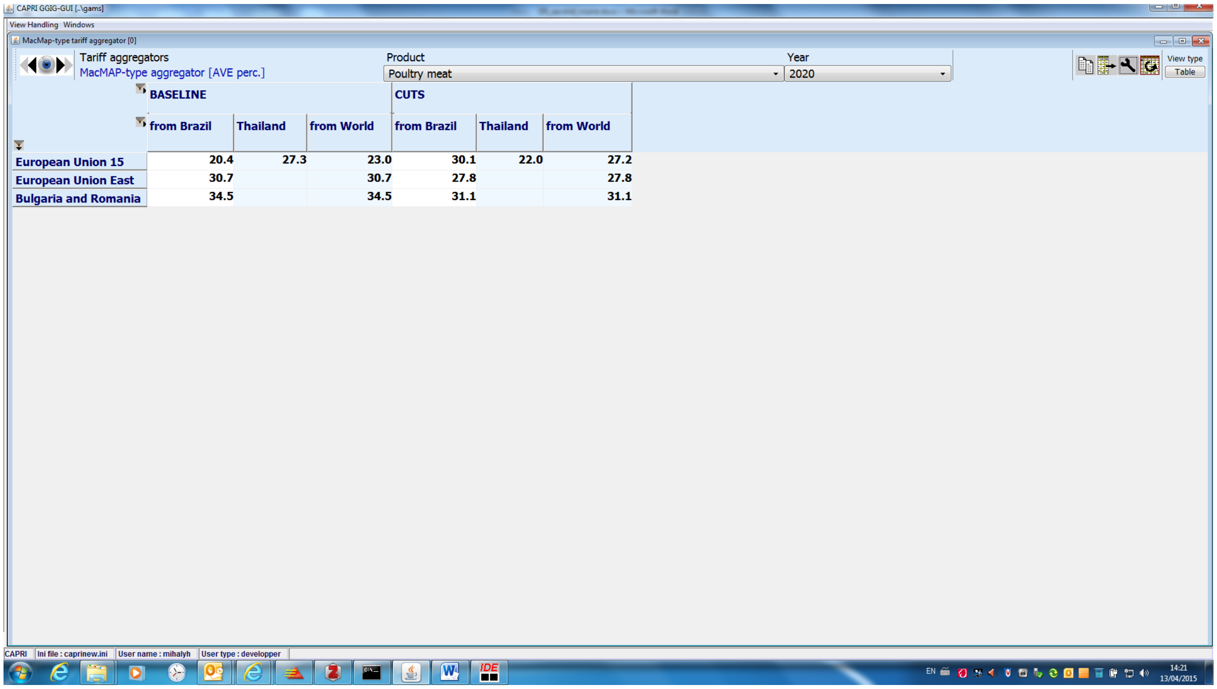
5.4.14 Overview on a regional module inside the market model

The resulting layout of a market for a country (aggregate) in the market module is shown in the following diagram. Due to the Armington assumption, product markets for different regions are linked by import flows and import prices if observed in the base year. Accordingly, no uniform world market price is found in the system.

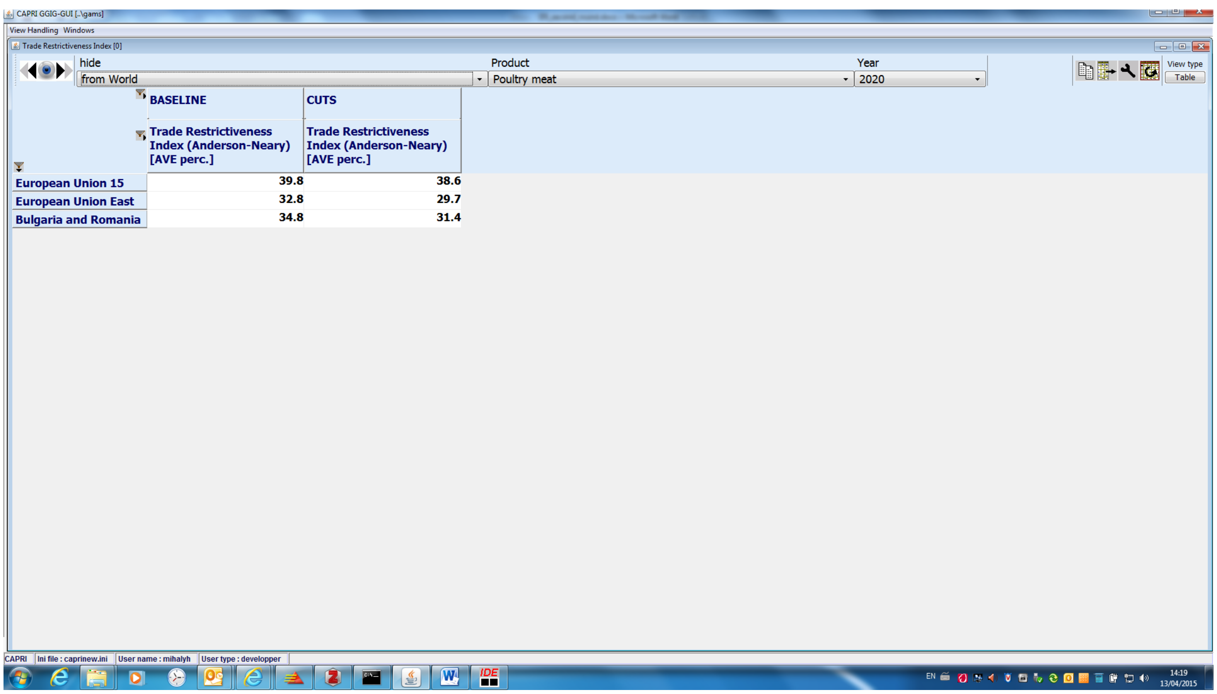
5.4.15 Basic interaction inside the market module during simulations

As with the supply module, the main difficulty in understanding model reactions is based on the simultaneity of changes occurring after a shock to the model. Cross-price effects and trade relations interlink basically all product markets for all regions. Whereas in the supply model, interactions between products are mostly based on explicit representation of technology (land balances, feed restrictions), such interactions are captured in multi-commodity models in the parameters of the behavioural functions.

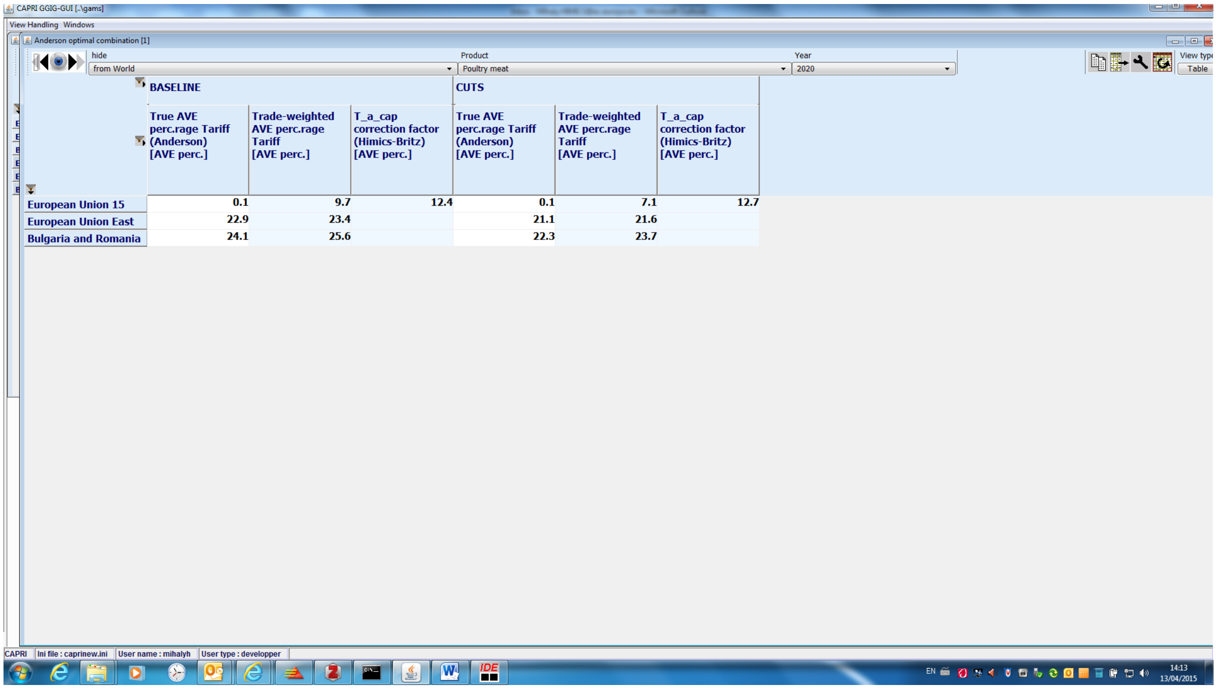
Even if the following narrative is simplifying and describing reactions as if they would appear in a kind of natural sequence where they are appear simultaneously in the model, we will nevertheless 'analyse' the effect of an increased supply at given prices for one product and one region. Such a shift could e.g. result from the introduction of a subsidy for production of that product. The increased supply will lead to imbalances in the market clearing equation for that product and that region. These imbalances can only be equilibrated again if supply and demand adjust, which requires price changes. In our example, the price in that region will have to drop to reduce supply. That drop will stimulate feed demand, and to



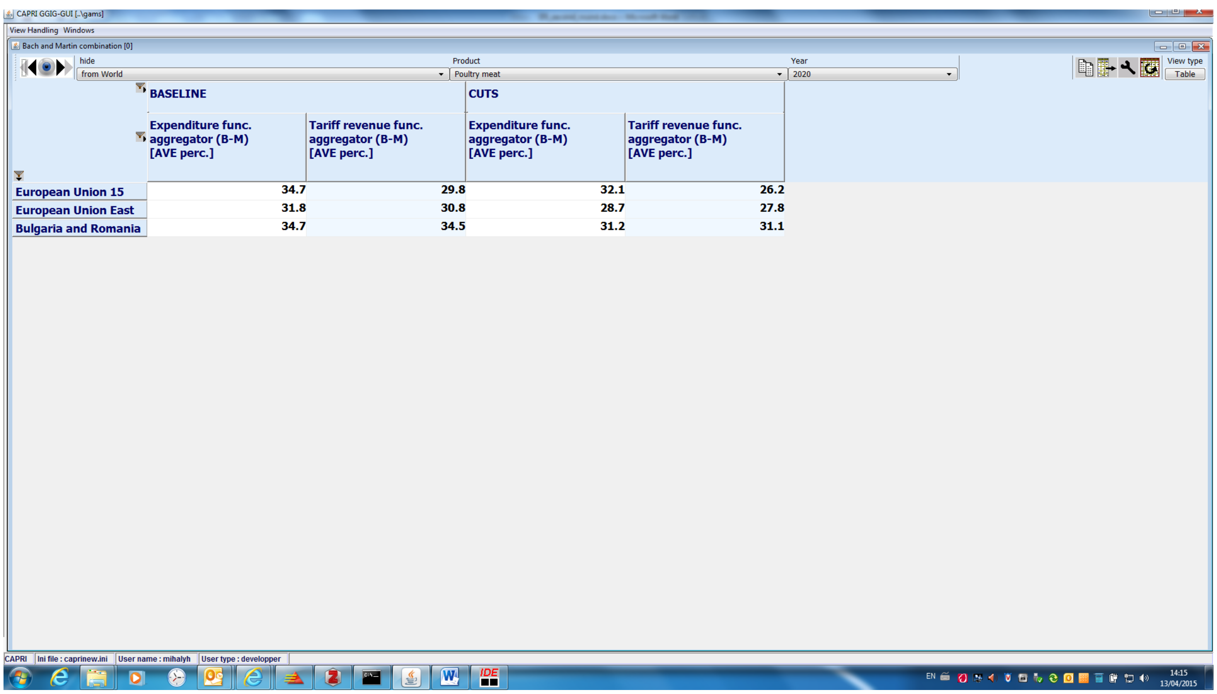
	BASELINE			CUTS		
	from Brazil	Thailand	from World	from Brazil	Thailand	from World
European Union 15	20.4	27.3	23.0	30.1	22.0	27.2
European Union East	30.7		30.7	27.8		27.8
Bulgaria and Romania	34.5		34.5	31.1		31.1



	BASELINE	CUTS
	Trade Restrictiveness Index (Anderson-Neary) [AVE perc.]	Trade Restrictiveness Index (Anderson-Neary) [AVE perc.]
European Union 15	39.8	38.6
European Union East	32.8	29.7
Bulgaria and Romania	34.8	31.4

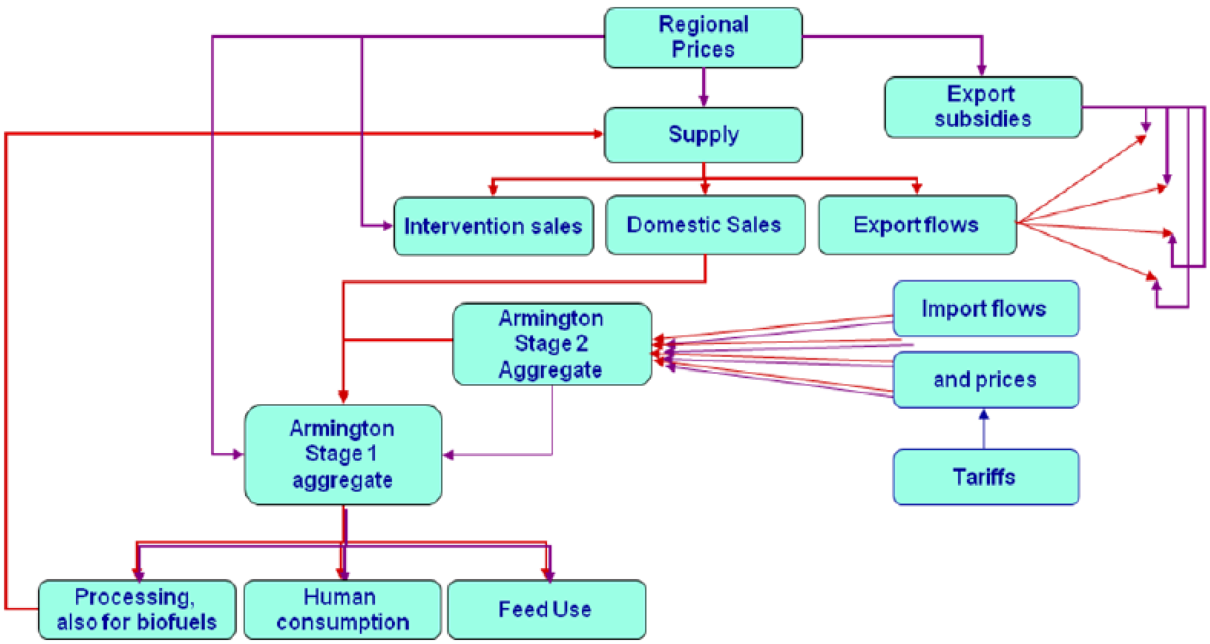


	BASELINE			CUTS		
	True AVE perc.rage Tariff (Anderson) [AVE perc.]	Trade-weighted AVE perc.rage Tariff [AVE perc.]	T_a_cap correction factor (Himics-Britz) [AVE perc.]	True AVE perc.rage Tariff (Anderson) [AVE perc.]	Trade-weighted AVE perc.rage Tariff [AVE perc.]	T_a_cap correction factor (Himics-Britz) [AVE perc.]
European Union 15	0.1	9.7	12.4	0.1	7.1	12.7
European Union East	22.9	23.4		21.1	21.6	
Bulgaria and Romania	24.1	25.6		22.3	23.7	



	BASELINE		CUTS	
	Expenditure func. aggregator (B-M) [AVE perc.]	Tariff revenue func. aggregator (B-M) [AVE perc.]	Expenditure func. aggregator (B-M) [AVE perc.]	Tariff revenue func. aggregator (B-M) [AVE perc.]
European Union 15	34.7	29.8	32.1	26.2
European Union East	31.8	30.8	28.7	27.8
Bulgaria and Romania	34.7	34.5	31.2	31.1

Figure 37: Graphical presentation for one region of a spatial market system



Source: CAPRI modelling system

a lesser extent, human consumption. The smaller effect on human consumption has two reasons: firstly, price elasticities for feed demand are typically higher, and secondly, consumer prices are linked with rather high margins to farm gate prices.

The resulting lower price at farm gate increases international competitiveness. Due to the Armington mechanism, consumers around the world will now increase the share of that region in their consumption of that product, and lower their demand from other origins. That will put price pressure in all other regional markets. The pressure will be the higher, the higher the import share of the region with the exogenous increase of supply on the demand of that product. The resulting price pressure will in turn reduce supply and stimulate demand and feed everywhere, and, with reduced prices, offset partially the increased competitiveness of the region where the shock was introduced.

Simultaneously, impacts on market for others products will occur. Depending on the size of the cross price elasticities, demand for other products will drop with falling prices for a substitute. At the same time, reduced prices will stimulate supply of other products. The resulting imbalances will hence force downwards price adjustments in other markets as well.

5.5 Solving the market model

The solution of the market model with its close to 750.000 equations of which some are highly non-linear poses a serious challenge for any non-linear solver. CAPRI applies CONOPT which has proven quite stable and fast to solve both constrained system and optimization problems. However, even CONOPT would spend quite some time when trying to solve the full market model in one block after a larger shock is introduced.

Therefore, a sequence of pre-solves is introduced (see *'arm/simu_prestep.gms'*). Firstly, single commodity models are defined by only allowing changes of the endogenous variables of one commodity in the equation template. Cross prices and their effects on quantities for variables related to the current market are hence fixed. These relatively small models can typically be solved rapidly by the solver, and are solved in parallel based on the so-called "grid solve option" in GAMS. That process is repeated a few times, each time updating the cross prices, to let differences between the single models and the full system decrease.

As a next step, the single products are clustered to groups where larger cross price effects can be expected, such as all cereals or all oilseeds. Again, these groups are solved repeatedly, in each round with updated cross-prices, to close in to the final solution. The full system is only solved at the very end.

Heuristics track the time needed for these solves and determine if it looks promising to skip solving single commodity and start with solving the groups or even the full model directly. The solution time of the model clearly depends on the hardware platform the models runs, but the heuristics do not take that into account. Accordingly, it cannot be guaranteed that the model finds exactly the same solution in a scenario on different machines.

Another problem possible problem beside long solution times is the occurrence of infeasibilities. Bounds are generally introduced for all endogenous variables to avoid numerical errors such as a division by zero. Bounds also help the solver in the solution process. However, they might also restrict the solution space so that no feasible solution exists. The CES functions for the Armington might as a response to a larger price shocks – e.g. provoked by removal of very large tariffs – drive trade flows almost to zero towards their lower bounds. Once that bounds are hit, the equation system is not longer symmetric as a new constraint becomes binding, and typically, the system will become infeasibility. If one would have the time to inspect the solution, one might perhaps accept that if the infeasibility is small and found only for that CES share equation. It is however generally impossible to leave it up to the model user to decide if she accepts infeasibility solutions or not, simply as there is simply not enough time to check these infeasibilities.

Fortunately, CONOPT helps us with in that case as it uses a gradient approach to reduce the sum of infeasibilities. It therefore introduces an objective into our problem, does also adding dual values to the constraints. We hence can inspect automatically the solution to find out which bounds carry a shadow values – removing these bounds will reduce the sum of infeasibilities. There is hence code ('*arm/widen_bounds.gms*') which in case of a infeasible solution will check which bounds carry dual values and will expand those stepwise. That proceeding generally guarantees that for most shocks, the market model finds a feasible solution.

5.6 Linking the different modules – the price mechanism

5.6.1 Iterative solution method

As hinted at above several times, the market modules and the regional programming models interact with each other in an iterative way. Basically, the market modules deliver prices to the supply module, and the supply module information to update the supply and feed demand response from the market models.

For the market module for agricultural outputs, the update of the supply and feed demand response is put to work by changing the constant terms in the behavioural equations such that supply and demand quantities simulated at prices used during the last iteration in the supply module would be identical to the quantities obtained from the market module at that prices. However, the "functional form" of the regional programming models is unknown but certainly differs from the one in the market model which necessitates an iterative update. In order to speed up convergence, the supply side uses a weighted average of prices of the last iterations.

Convergence is achieved faster if supply has the same price responsiveness in the market model as in the regional programming models. To achieve this at least approximately, a set of price elasticities is generated from the regional programming models and is used to calibrate the parameters of the market model. This also applies to young animals which are "netputs" that are only traded between European regions with regional supply models.

Price linkage in the sugar sector Whereas the linkage of prices from the market model to the supply model is usually a proportional one, the price linkage in the sugar sector is specific for ethanol beets. Sugar and ethanol beets are considered two products with independent proportional linkage applied to each of them. Their prices may move independently therefore and European beet producers respond to both prices.

5.7 Sensitivity analysis

The CAPRI model results depend on a large number of parameters, some of which are more uncertain than others. In order to analyze how model results depend on uncertain parameters, a set of sensitivity analyses were carried out in the context of an analysis of the climate impact of EU coupled support[17] . In that study, a baseline scenario for 2030 representing the CAP 2014-2020 was compared to a scenario where the voluntary coupled support was removed. This resulted in less production of e.g. beef in the EU, and less GHG emissions there. Due to inelastic demand for food, trade flows changed, so that less beef was exported from the EU and more imported, and thus carbon leakage arose. We carried out sensitivity analyses to investigate how carbon leakage depend on the model parameters.

We selected four types of parameters that were assumed to be most critical to emissions leakage, and varied those in three levels: "low" (lo), "high" (hi) and "most likely" (ML). The groups of parameters subjected to the sensitivity analyses are as follows:

[17] Blanco, G., R. Gerlagh, J. Barrett, S. Suh, H.C. de Coninck, C.F. Diaz Morejon, R. Mathur, N. Nakicenovic, A. Ofosu Ahenkora, H. Pathak J. Pan, J. Rice, R. Richels, S.J. Smith, D.I. Stern, F.L. Toth, and P. Zhou. 2014. Drivers, Trends and Mitigation. In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report,* eds. Edenhofer et al. Cambridge and New York: Cambridge University Press.

- The elasticities of supply (SupElas) of ruminants in the EU are influenced by the slope of the marginal cost function[18]. Higher slope means lower supply elasticity and vice versa. The slope was varied +/- 50% to create the lo and hi scenario variants.
- The elasticities of demand (DemElas) for meat and dairy products. We recalibrated the demand systems for all countries so that the own-price demand elasticities would be as close as possible to +/- 50% of the standard value, while observing relevant regularity conditions for demand systems.
- Substitution elasticities (CES) between imports and domestic products and between different import sources were also set to +/- 50% of the standard values. The standard values differ per product, ranging from 2 to 10.
- GHG emission factors (EF) per commodity outside of the EU. Emissions leakage depends more on the relationship between EF in the EU to those outside the EU than on the absolute level. Therefore, we chose to vary only the factors outside of the EU. Since, in general, N_2O factors are considered less certain than emissions of CH_4, which in turn are less certain than CO_2, we chose to apply the uncertainty ranges of the IPCC (Blanco et al. 2014) to construct the hi and lo scenarios. The ranges used were +/- 60% for N_2O and +/- 20% for CH_4.

We do not know the covariance of the uncertain parameters across countries and products. In order to avoid running a very large number of simulation experiments, we chose to implement only the most extreme variants given by setting all parameters of the same type to lo/ML/hi simultaneously (e.g., elasticities of supply of all ruminants in all countries being hi, ML or lo simultaneously, etc.). We thus obtained $3 \times 3 \times 3 \times 3 = 81$ result sets; this should span the extremes of the result space.

The manuscript was submitted to a journal. Therefore, this section does not yet contain any results from this exercise, but it will be completed as soon as the review process of the manuscript has been completed.

[18]CAPRI contains quadratic cost functions in the tradition of Positive Mathematical Programming (PMP). In the sensitivity analyses, we varied the coefficient of the quadratic term.

Chapter 6

Post model analysis

6.1 Dual analysis

T. Jansson

6.1.1 Constrained optimisation background

The supply model in CAPRI is a constrained optimization problem, maximizing profits plus a nonlinear term subject to multiple constraints. In some simulations, production activities in the model are affected in several ways simultaneously, and it is difficult to understand what the key drivers are behind a specific change. There may be changes both in parameters of the model such as prices or subsidies and also of dual values of resources such as land or fodder. Then a *dual analysis* can help you decompose the impacts by explaining how much the profitability of a production activity changed due to various factors. The key idea is to analyse how each term of the *first order conditions* changed.

For any well-behaved optimization model (one that e.g. maximizes a concave objective function on a convex constraint set) we can formulate the first order condition stating that in the optimal point, the sum of all marginal changes of all variables would provoke no change[1] in the objective function. Let $i = 1...n$ be activities in the model, x_i be the level of activity i, $f : \mathbb{R}_n \rightarrow \mathbb{R}$ be the concave and continuously differentiable objective function (e.g. profits), and $g_j : \mathbb{R}_n \rightarrow \mathbb{R}$ be a function serving as a restriction for each of $j = 1...m$ resources. Finally, assume that all activity levels have to be non-negative[2]. The agent of the model acts as if solving the following maximization problem:

$$max\, f(x_1, ..., x_n)$$

Subject to

[1]Depending on the particular use of the Lagrange function, one might also say that all marginal changes would cause the objective to deteriorate.

[2]With a broader interpretation of "activity" which includes slack variables, this allows us to model equality and inequality constraints.

$$\begin{array}{lll} g_j(x_1, ..., x_n) = 0 & [\lambda_j] & \forall j = 1...m \\ x_1 \geq 0 & [\pi_i] & \forall i = 1...n \end{array} \tag{6.1}$$

Greek letters (and) in square brackets denote the Lagrange multipliers associated with each constraint. The Lagrange function[3] for this problem can be written

$$L(x_1, ..., x_n, \lambda_1, ..., \lambda_m, \pi_1, ..., \pi_n) = f(x_1, ..., x_n) - \sum_j \lambda_j g_j(x_1, ..., x_n) + \sum_i \pi_i x_i \tag{6.2}$$

The necessary conditions for an optimal solution include the first order derivative of the Lagrangean with respect to the activity levels:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_j \lambda_j \frac{\partial g_j}{\partial x_i} + \pi_i = 0 \tag{6.3}$$

The functions f and g depend on some parameters, e.g. prices and technical i/o coefficients respectively that were not shown in the exposition above in order to reduce the size of the expressions. By changing such parameters, we introduce shocks to the model. Assume that in a reference scenario, we have prices p^0 and technology a^0 resulting in the solution $(x_1^0, ..., x_n^0, {}_1^0, ..., {}_m^0, {}_1^0, ..., {}_n^0)$. In another simulation, we have other prices p^* and technology a^* resulting in the alternative solution $(x_1^*, ..., x_n^*, {}_1^*, ..., {}_m^*, {}_1^*, ..., {}_n^*)$.

In this simulation we now would like to know more about why some particular activity x_i reacts as it does, i.e. why x_i^* is different from x_i^0. We then compute each term in the first order conditions, and compare the two simulation. To be slightly more explicit, we can assume that the resources are j={"land" ,"fodder" ,"young animals" }. We can then do a comparison such as the following: FIXME

$$\begin{array}{llllll} ref & \frac{\partial f(p^0)}{\partial x_i^0} & -\lambda_{land}^0 \frac{\partial g_{land}(a^0)}{\partial x_i^0} & -\lambda_{fodder}^0 \frac{\partial g_j(a^0)}{\partial x_i^0} & -\lambda_{young\,animals}^0 \frac{\partial g_j(a^0)}{\partial x_i^0} & +\pi_i^0 = 0 \\ sim & \frac{\partial f(p^*)}{\partial x_i^*} & -\lambda_{land}^* \frac{\partial g_j(a^*)}{\partial x_i *} & -\lambda_{fodder}^* \frac{\partial g_j(a^*)}{\partial x_i^*} & -\lambda_{young\,animals}^* \frac{\partial g_j(a^*)}{\partial x_i^*} & +\pi_i^* = 0 \end{array} \tag{6.4}$$

If we would compute the difference between the two rows above for each corresponding term, the differences would also sum up to zero, and each difference would mean something and have a numerical value: The first term is the change in marginal profits, the second term is the change in marginal cost of land, the third is the marginal cost of fodder, the fourth the marginal cost of young animals, and the final term the possible marginal value of a lower bound on the activity at hand. This can usually be directly translated into a storyline, like "The price increase increases the marginal profit of barley production by q euro per hectare, which leads to an expansion of production until it is matched by a similarly large increase in land rents".

Of course, each partial derivative may be further decomposed into its smallest parts. For instance, profits may contain several items such as revenues, costs and behavioural terms.

[3]Treating the non-negativity conditions as any other constraint, so that the first-order conditions become a set of equalities.

6.1.2 Example – removing greening payments and requirements

The dual analysis of the supply models is computed by a file called "supply/margcr.gms" (where margcr probably means "MARGinal Cost and Revenues"), which is always included after the last iteration. The figures below show the CAPRI GUI comparing the results of two scenarios: one reference continuing CAP after 2014 up to 2030, and a counterfactual where the greening payments and requirements are removed. We look at the results of a single region "Sydsverige" to avoid aggregation issues.

In Figure 39, we can look for the biggest changes (absolute difference between simulations is printed in brackets in the right hand column), and find a 71 euro reduction in payments. This is in line with the scenario, which removed 30% of the farm payment. All other positions adjust to some extent, and as usual with CAPRI, we find a similarly sized change in "PMP terms" (+77 euro income or reduction in costs), i.e. the behavioural terms adjust. As a note, we see that for this region, grassland is highly unprofitable in the reference scenario, and therefore the PMP term is calibrated to become a marginal *income*. The PMP effect is decomposed [4]into a change in the constant term (brought about by a shift between GRAI and GRAE), diagonal terms (the main effect, 65 euro) and cross effects (zero in this case).

There are more adjustments. Also land rents drop by 18 euro, the output of grass shown in the first line increases in value by 25 euro. Both variable costs and the value of the fertilizer restriction decrease, but those effects are comparatively minor.

6.2 Decomposition of changes in aggregates yields and activity related income indicators

The idea behind the decomposition is to analyze which factors drive the change in yields and income indicators using growth rates. Take for example the market income of cereals per ha at EU level. Its change in a simulation against the baseline depends on the change in prices and the change in yields. The change in yields in turns depends on the effect of the yield elasticity, the change in the shares of low and high technology variants, the change in the regional shares and, in weights of low and high yielding regions. And finally, the share of high and low yielding cereals such as soft what and rye in the total aggregate might change. When interpreting the results, it is often useful to understand the contribution of the different factors.

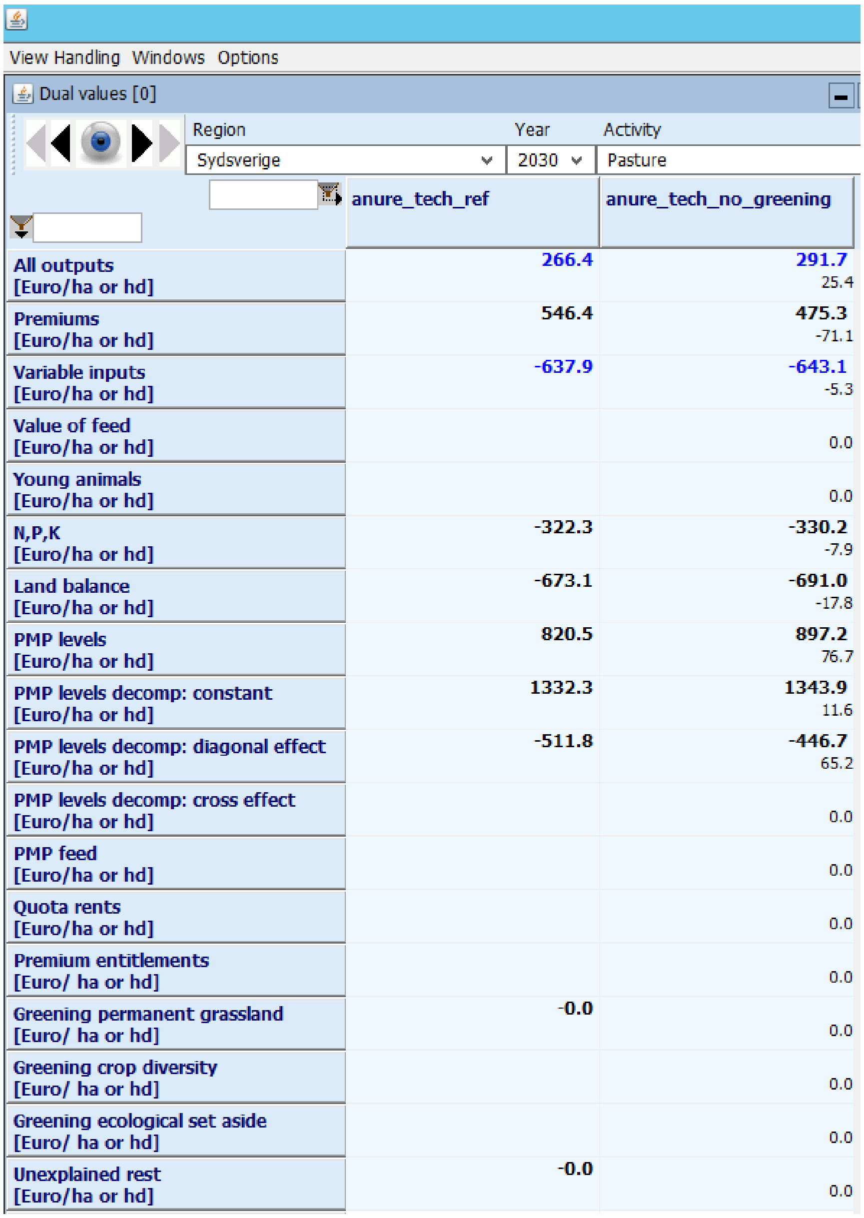
The screenshot below shows an example. All results shown are always expressed in the absolute value and the units used in the result set.

The columns are defined as follows:

- *Result*: Final result from the simulation – in the example below, average EU27 cereals yields increase by 8.57%.
- *Effect of endogenous IO coefficients*: Result calculated by using the final IO coefficients, but keeping all other factors (technology shares, prices, regional weights, activity levels) at trend levels – in the example below, market revenues per ha would have gone up by 1.55% if only the yields had adjusted.
- *Effect of technology shares*: Result calculated by only updating the technology shares, but keeping all other factors (IO coefficients, prices, regional weights, activity levels) at trend levels – in the

[4]This decomposition of changes in the PMP level contributions is not available in all CAPRI versions.

Figure 38: Dual analysis of changes in pasture area (intensive and extensive) in a scenario removing the CAP greening components


View Handling Windows Options

Dual values [0]

Region: Sydsverige | Year: 2030 | Activity: Pasture

	anure_tech_ref	anure_tech_no_greening	
All outputs [Euro/ha or hd]	266.4	291.7	25.4
Premiums [Euro/ha or hd]	546.4	475.3	-71.1
Variable inputs [Euro/ha or hd]	-637.9	-643.1	-5.3
Value of feed [Euro/ha or hd]			0.0
Young animals [Euro/ha or hd]			0.0
N,P,K [Euro/ha or hd]	-322.3	-330.2	-7.9
Land balance [Euro/ha or hd]	-673.1	-691.0	-17.8
PMP levels [Euro/ha or hd]	820.5	897.2	76.7
PMP levels decomp: constant [Euro/ha or hd]	1332.3	1343.9	11.6
PMP levels decomp: diagonal effect [Euro/ha or hd]	-511.8	-446.7	65.2
PMP levels decomp: cross effect [Euro/ha or hd]			0.0
PMP feed [Euro/ha or hd]			0.0
Quota rents [Euro/ha or hd]			0.0
Premium entitlements [Euro/ ha or hd]			0.0
Greening permanent grassland [Euro/ ha or hd]	-0.0		0.0
Greening crop diversity [Euro/ ha or hd]			0.0
Greening ecological set aside [Euro/ ha or hd]			0.0
Unexplained rest [Euro/ha or hd]	-0.0		0.0

Source: CAPRI modelling system

example below, intermediate input costs would have increased by 0.2% of only the share of low and high yielding variants had changed.

- *Effect of prices*: Result calculated by only updating the prices shares, but keeping all other factors (IO coefficients, technology shares, regional weights, activity levels) at trend levels – in the example below, the Gross Value Added at producer prices would have increased by 23.62% if only the input and output prices had changed.
- *Effect of regional composition*: Result calculated by only updating the activity levels shares, but keeping all other factors (IO coefficients, technology shares, activity levels) at trend levels
- *Effect of other factors*: Difference between the start values and result, after all other effects above had been accounted for. Comprises the multiplicative cross-effects of the different effects, the effect of changed premiums in the case of the MGVA, and of change in the activity aggregate composition.

Figure 39: Illustrative scenario comparison for the yield and income indicator decomposition

		MTR_RD						MTR_RDNOPIL1					
		Result	Effect of endogenous IO coefficients	Effect of changed technology shares	Effect of prices	Effect of changed regional composition	Effect of other factors	Result	Effect of endogenous IO coefficients	Effect of changed technology shares	Effect of prices	Effect of changed regional composition	Effect of other factors
Cereals	Yield [kg or 1/1000 head/ha or head]	5510.02 0.00%	5510.01 0.00%	5510.00 0.00%		5510.01 0.00%	5510.00 0.00%	5674.34 2.98%	5605.74 1.74%	5518.89 0.16%		5511.60 0.03%	5568.11 1.05%
	Market revenues [Euro/ha or head]	841.19 0.00%	841.17 0.00%	839.59 0.00%	839.60 0.00%	839.59 0.00%	839.59 0.00%	913.31 8.57%	854.19 1.55%	840.86 0.15%	883.17 5.19%	839.77 0.02%	853.67 1.68%
	Intermediate inputs [Euro/ha or head]	655.18 0.00%	655.18 0.00%	655.18 0.00%	655.18 0.00%	655.18 0.00%	655.18 0.00%	667.84 1.93%	661.95 1.03%	656.50 0.20%	655.18 0.00%	653.69 -0.23%	661.23 0.92%
	Gross Value added at producer prices [Euro/ha or head]	186.01 0.00%	185.99 0.00%	184.41 0.00%	184.43 0.00%	184.41 0.00%	184.41 0.00%	245.47 31.97%	192.24 3.36%	184.36 -0.03%	228.00 23.62%	186.07 0.90%	192.44 4.35%
	Gross Value added at producer prices plus premiums [Euro/ha or head]	465.15 0.00%	465.13 0.00%	463.55 0.00%	463.57 0.00%	463.55 0.00%	463.55 0.00%	296.76 -36.20%	243.93 -47.56%	236.05 -49.08%	279.68 -39.67%	238.00 -48.66%	243.49 -47.47%

Source: CAPRI modelling system Note: The code is implemented in „reports/yield_change_decomp". The table can be found in the GUI under "farm => yield decomposition".

It should be mentioned that the coding does not use information about potential changes of the premiums paid in scenario against the baseline, so that the results shown for the modified Gross Value Added should mainly show the different percentage changes on income once the level effect of the premiums is considered.

The coding also stems from the time before climate related yield shock sceanrios or endogenous mitigation modelling. As the yield decomposition probably has not been regularly checked in this kind of scenarios, it is possible that some changes implemented in the CAPRI supply models need to be transferred to this reporting code to give reliable results in all kind of scenarios.

6.3 Post model sensitivity analysis with features in the supply model

The behaviour of the CAPRI supply model depends on the interaction between endogenous variables such as production and feeding activities both via the objective function and the constraints. Even if the dual analysis of the results may help, understanding the changes in levels of endogenous variables compared to the baseline is far from trivial. As the directions and size of certain impacts on the reported changes is not known, result analysis carries a high risk of mis-interpretations. A specific optional reporting tool has been developed therefore and is accessible though the CAPRI GUI (farm => supply model analysis) that systematically analyses the contributions of the different allocative mechanisms in the model. The basic idea consists in evaluating how the result at given prices would have looked like if certain endogenous model features would not have been used.

However due to various code changes in the last years it cannot guaranteed that the reporting option it still fully operational such that we refer the interested reader to earier versions of this documentation[5].

6.4 Welfare analysis

A key element in analysing policy changes from an economic viewpoint is to look at welfare changes. The "classical" elements of a welfare analysis are changes in consumer and producer rents and for the tax payer. That concept is also followed in CAPRI.

For consumers, CAPRI uses the money metric concept. It can be broadly understood as a measurement for changes in the purchasing power of the consumer. The concept is also linked to the expenditure function as introduced in Section 5.4.4:

$$M(cpri^r, cpri^s, Y^r) = e(cpri^r, U(cpri^s, Y^r)) \tag{6.5}$$

Where $e(.)$ is the expenditure function, Y^r is expenditure in the reference situation, and $cpri^r(cpri^s)$ is the price vector in the reference (scenario) situation. The money metric is thus the expenditure the consumer would need at reference prices to be as well of as if facing the scenario prices at reference income. The difference of money metric for a given scenario to money metric in the reference situation is the equivalent variation. Considering the generalised Leontief form of the indirect utility function used in CAPRI (compare with Section 5.4.4) we get

[5] A more complete version of this section (originally drafted by Wolfgang Britz) in the CAPRI documentation is accessible in the /doc folder of any stable release of the CAPRI system up to star 2.4 from `https://www.capri-model.org/dokuwiki/doku.php?id=capri:get-capri`.

$$e(cpri^r, U(cpri^s, Y^r)) = \left\{ F^r + \frac{G^r}{G^s/(Y^r - F^s)} - Y^r \right\} + Y^r$$
$$= \left\{ \frac{G^r}{G^s}(Y^r - F^s) - (Y^r - F^r) \right\} + Y^r \tag{6.6}$$

The bracket {} is the equivalent variation of the change from the reference to the scenario situation. In the code (in *gams/reports/welfare.gms*) the momey metric is computed in a loop over products such that the contributions of single products are displayed[6]. It is also considered that the per capita demand needs to be multiplied with population.

```
   PS_CAL(RMS,XX1)  = DATA(RMS,"Cpri",XX1,"TRD");
   PS_Y(RMS,XX1)    = PS_CAL(RMS,XX1);

*
   LOOP ( XX1,
*
*      --- set price in current market to price in simulation
*          ( all markets "after" XX are still at reference prices)
*
       PS_Y(RMS,XX1)    = DATA(RMS,"Cpri",XX1,"%SIMY%");
*
*  --- calculate FS for prices at calibration point (used below)
*
      v_GLDemandFS.l(RMS) = SUM(ZZ1 $ p_pdGL(RMS,ZZ1,"CUR"), p_pdGL(RMS,ZZ1,"CUR") * PS_CAL(RMS,ZZ1))/1000.;
*
*      --- equivalent variation in evaluation step XX1 in mio euro:
*
       p_welfareRes(RMS,"CSSP",XX1,"%SIMY%") $ p_pbGL(RMS,XX1,XX1,"CUR")
*
*        --- value of function G = SUMij Bij Pi-0.5 Pj-0.5
*
         =    SUM( (YY1,ZZ1) $ p_pbGL(RMS,YY1,ZZ1,"CUR"),
                 p_pbGL(RMS,YY1,ZZ1,"CUR")
                      * SQRT(PS_CAL(RMS,YY1) *  PS_CAL(RMS,ZZ1) * 1.E-6))
            / SUM( (YY1,ZZ1) $ p_pbGL(RMS,YY1,ZZ1,"CUR"),
                 p_pbGL(RMS,YY1,ZZ1,"CUR")
                      * SQRT(PS_Y(RMS,YY1) * PS_Y(RMS,ZZ1) * 1.E-6))
*
            *  (DATA(RMS,"Ince","LEVL","CUR")/DATA(RMS,"INHA","LEVL","CUR")
                 - SUM(ZZ1 $ p_pdGL(RMS,ZZ1,"CUR"), p_pdGL(RMS,ZZ1,"CUR") * PS_Y(RMS,ZZ1))/1000.)
*
             - (DATA(RMS,"Ince","LEVL","CUR")/DATA(RMS,"INHA","LEVL","CUR") -  v_GLDemandFS.l(RMS));
*

*      --- replace base year price in market just evaluated by price in simulation
*
       PS_CAL(RMS,XX1) = PS_Y(RMS,XX1);
*
   );
*
*   --- from per capita to per total population (plus expenditure in calibration point)
*
   p_welfareRes(RMS,"CSSP",XX1,"%SIMY%") =  DATA(RMS,"INHA","LEVL","CUR")/1000.
                                          * p_welfareRes(RMS,"CSSP",XX1,"%SIMY%")
                                          + DATA(RMS,"HCon",XX1,"TRD") * DATA(RMS,"CPri",XX1,"TRD") * 0.001;
*
```

For some agents the welfare accounting involves particularities for those regions covered by the regional programming models that will be discussed below. In the general case (for “non supply model regions”) producer welfare may be derived in the market model from the normalised profit function underlying the supply functions and the input demands for feed energy and land:

[6]As this is basically the computation of a line integral the contributions of the single products depend on the sequence of products used in the loop, but we do not expect that this path dependency would significantly change the contributions by product. And for the total effect there is no path dependency.

$$\pi(p^s) = {}_0 + \sum_x as_x p_x^s + 0.5 \sum_{x,y} bs_{x,y} p_x^s p_y^s \tag{6.7}$$

where p^s is the price vector in the scenario situation and the arbitrary constant π_0 has been chosen as 50% of reference run revenues. In order to show the *changes* in producer welfare by product we compute producer welfare in the scenario as follows:

$$\pi(p^s) - \pi(p^r) + \pi_0 = \sum_x as_x (p_x^s - p_x^r) + 0.5 \sum_{x,y} bs_{x,y} (p_x^s p_y^s - p_x^r p_y^r) + \pi_0 \tag{6.8}$$

In the code (in *gams/reports/welfare.gms*) we are also considering that it is necessary to revert the normalisaiton of prices with the general price index

```
   LOOP ( XX1,
*
*      --- set price in current market to price in simulation
*          ( all markets "after" XX are still at reference prices)
*
       PS_Y(RMS,XX1)    = (    DATA(RMS,"PPRI",XX1,"%SIMY%")
                           + (DATA(RMS,"PSEd",XX1,"%SIMY%")
                           +  DATA(RMS,"PSEI",XX1,"%SIMY%")
                           +  DATA(RMS,"AREP",XX1,"%SIMY%")*(1-DATA(RMS,"DECP",XX1,"%SIMY%")))$ SAMEAS(XX1,"MILK")
                           -  DATA(RMS,"PSEd",XX1,"%SIMY%") $ SAMEAS(XX1,"LAND")
                          ) / DATA(RMS,"PPri","Inpe","TRD");
*
       p_welfareRes(RMS,"ProfitAgr",XX1,"%SIMY%")
           =         SUM(  ZZ1 $ p_cnstNQSupp(RMS,ZZ1),
                           p_cnstNQSupp(RMS,ZZ1) * (PS_Y(RMS,ZZ1)  - PS_CAL(RMS,ZZ1))) * 0.001

           + 0.5 * SUM( (ZZ1,YY1) $ p_hessNQSupp(RMS,ZZ1,YY1,"CUR"),
                             p_hessNQSupp(RMS,ZZ1,YY1,"CUR")
                           * ( PS_Y  (RMS,ZZ1)*PS_Y  (RMS,YY1)
                              -PS_CAL(RMS,ZZ1)*PS_CAL(RMS,YY1))) * 0.001;
       PS_CAL(RMS,XX1) = PS_Y(RMS,XX1);
   );
*
*  ---- undo normalisation and add 50 % of baseline revenues as profit
*       (for XX1 to also cover the contribution from land to profit change)
*
   p_welfareRes(RMS,"ProfitAgr",XX1,"%SIMY%")
                                 = p_welfareRes(RMS,"ProfitAgr",XX1,"%SIMY%") * DATA(RMS,"PPri","Inpe","TRD")
                                 + DATA(RMS,"Prod",XX1,"TRD")*DATA(RMS,"PPri",XX1,"TRD") * 0.5
                                 * 0.001 $ p_hessNQSupp(RMS,XX1,XX1,"CUR");
*
```

CAPRI uses the normalised quadratic form also for other agents, such that the calculation of welfare changes is very similar for

- the dairy industry (with margins adopting the role of prices)
- the processing industry (again with margins adopting the role of prices)
- the feed industry

Land supply is currently specified as responding to the log of land rents according to a land supply elasticity which also permits an easy calculation of land owner rent.

In the global market model, the main effects on taxpayer welfare comes through tariff revenues. Here it is also necessary to consider the allocation of TRQ rents, and if several bahavioural ("RMS") regions are combined into a single market region ("RM") then the tariff revenues of the market region need to be allocated.

For European (supply model) regions there are a number of particularities. First of all, CAPRI uses changes in the gross value added (GVA) plus premiums as the main indicator of producer (farmer) welfare. The gross value added is the difference between revenues (output quantities valued at farm gate prices) and intermediate input costs (input quantities with the exemption of the primary factors land, capital and labour multiplied with their farm gate prices). The GVA plus premiums is hence the sum the farming sector can spend to remunerate labour, capital and land, independent of property rights of these factors. GVA including subsidies is a traditional measure of "farm income" which has been one of the most important objectives of the CAP in the early years of CAPRI. It was considered useful therefore to monitor farm income defined in this way in the database and to compute it for scenarios. GVA plus premiums differs from some other income indicators

- "Farm profits" may be computed from GVA plus premiums considering ownership and some accounting rules for depreciation of capital, and family as well as non-family labour costs.
- "Economic profit" in the sence of the supply models would be the value of the supply model objective function which depends critically on the specification of unobserved cost function parameters (PMP).
- Farm household income would also consider non-farm income sources and income of all household members and therefore clearly differs from GVA.

Due to the allocation of output and input to individual activities, CAPRI allows to calculate the GVA for single production activities, and to aggregate it from there to groups of activities and over regions.

"Land owner welfare" and "feed industry welfare" are also treated differently as compared to non-European regions, in line with the supply model specification for land supply and feed use. Regarding "land owner welfare" the supply models specify costs for the supply of land to farming that may be interpreted as the opportunity costs of non-agricultural land use. Deducting this opportunity cost of land ensures that scenarios involving changes in land supply are not assessed in a biased way (e.g. by treating land as a "free" input without a price).

Feed demand is also reflected already in the supply models, in fact in a quite sophisticated manner, which balances not only for feed energy (as in the global market model) but also for protein. The GVA plus premiums therefore already includes the effects of changing feed costs and no additional calculation is needed.

The final particularity derives from the detailed representation of the CAP premium system in the supply models. For premium schemes, financing rates for the EU and national budgets as well as for pillar I and II of the CAP are defined, which allow to allocate costs to those different budgets. A few other CAP instruments have become largely irrelevant in the meantime, like the cost of public market "interventions" (purchases for storage at public cost), export subsidization and some subsidies paid for demanders of agricultural goods. These are also covered in the CAPRI welfare accounting (in case they would be revitalised).

6.5 Energy Use in Agriculture

[7] FIXME

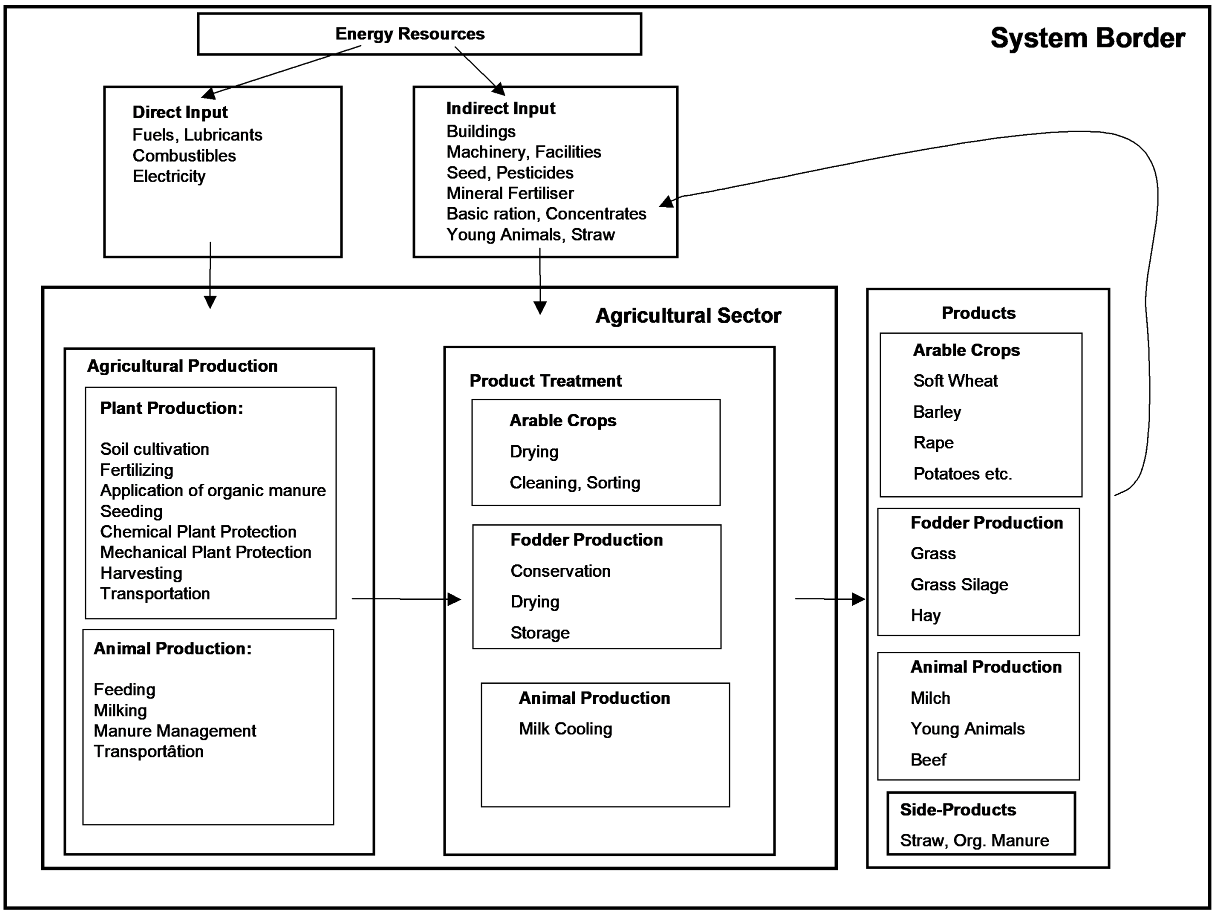
6.5.1 Introduction and basics

The objective of the CAPRI energy indicator is to improve the existing CAPRI model in its capabilities to display environmental effects of agricultural production activities. In order to give a short overview of the structure of the energy indicator, the underlying methodology of Life Cycle Analysis will be introduced covering structure of the energy coefficients used for assessment. In a second part, the assessment methodology of the single direct and indirect components will be described. Finally, the structure of the results being processed by the energy module will be shown and hints for application will be given. Energy input quantification follows process analysis within the methodology description of Cumulative Energy Demand (KEA) guideline N° 4600 (VDI, 1997). Thereby, the KEA states the entire demand of non-renewable energy resources, valued as primary energy, which arises in connection with the production, use and disposal of an economic good (product or service) or which may be attributed respectively to it in a causal relation (VDI, 1997). A precise definition of balancing boundary setting is carried out according to local, temporal and technological criteria and is an important foundation for the KEA. Due to the high complexity and multiplicity of some of the interactions between individual processes, systematic delimitation frequently poses a central problem for energy analysis. A detailed determination of all relevant energy and material flows in the service life of a product requires a separation of the components of the KEA right down to the individual processes. An energy balance in this context registers energy quantities or energy types respectively in Joule or Watt-hours, crossing the defined balance space boundaries during the period of analysis. The energy balance boundaries are identical with the material balance boundaries (VDI, 1997). In the CAPRI context, the input part of the KEA concept is underlying the energy assessment of agricultural production. Life cycle analysis (LCA), by integrating the KEA concept offers a suitable framework for energy assessment of CAPRI production activities. Therefore, guidelines such as ISO 14040 and 14044 (DIN, 2006) are considered in the energy assessment process. Such procedure is based on process analysis, which can be defined as follows: “The network or processes required to make a final product are identified. Each input is assigned an energy requirement so that the total energy requirement can be summed” (Fluck, 1980). Setting the system borders precisely is an essential task in this concern. Figure 40 shows the relevant borders and integrated processes for CAPRI energy module. The term “Agricultural Production” in Figure 40 is the main interface between LCA and CAPRI production activities.

The connecting link between process-based material flows and the energy requirement analysis are energy content factors. Life cycle inventories of agricultural production systems are the necessary tool therefore. The role of inventories such as ecoinvent (2003) is to provide modules for infrastructure and inputs used in agricultural production necessary for modelling production systems. In the case of the CAPRI energy module, several aspects concerning inventories had to be considered. On the one hand, a broad range of different sources provide inventory databases designed for different countries in the agricultural context. On the other hand, to use a uniform methodological basis, a basic decision for inventories analysed by

[7] The energy use linked to agriculture is assessed as a reporting option that may be activated in a normal simulation for any scenario, due to original coding work by Tim Kränzlein and Markus Kempen who also prepared the following sections on its details. As this reporting option has not been used often in recent years, it cannot be guaranteed that the module is still operational and giving results as usual, but there is so far also no evidence to the contrary. Therefore the following sections have been retained without changes from the previous versions of the CAPRI documentation.

Figure 40: System border and processes considered for CAPRI energy input assessment



Source: Based on ecoinvent (2003)

ecoinvent (2003) was taken. Firstly, a great number of single inventories (direct and indirect energy sources as well as agricultural processes such as drying or irrigation) had been analysed. Secondly, the inventories being used are updated regularly and by using SALCA061 (2006) database for CAPRI energy indicator, a most recent version of the inventories was used. Thirdly, special analysis for the CAPRI energy module such as quantifying energy for stables for animal production activities was carried out using the underlying methodology of ecoinvent database in order to consider specifics of CAPRI. Nevertheless, the ecoinvent agricultural inventories have been compiled mainly in a Swiss context using background data of Swiss agriculture. In order to use these inventories for CAPRI, some adjustments have been made. Some minor differences in the energy assessment between CAPRI energy module and other literature sources cannot be avoided. The reason might be in the reference period of the data (most literature data is of some years age) or in the Swiss-based approach of the inventories.

6.5.2 Energy assessment in CAPRI

To integrate the methodology which is described in Chapter 7.3.1 into CAPRI, two parts are required for each single energy input component: an activity-specific, regionalized consumption quantity and an equivalent assessment factor. The following chapters present both parts for each input component integrated into CAPRI.

Direct energy sources Direct energy covers those energy sources that are consumed directly in the production process for the purpose of generation of usable energy (Werschnitzky et al. 1987). For diesel fuel, petrol, heating oil, electricity, gas, coal etc., the input quantities of each animal and plant production activity are calculated and afterwards assessed by energy content factors. The content factors are based on ecoinvent modules using SALCA061 database (2006). Table 30 shows the main direct energy sources used in CAPRI.

Table 30: Energy content factors for direct energy input

Direct energy component	Cumulative energy demand	Unit
Diesel	45.7	MJ/l
Electricity (at grid)	11.7	MJ/kWh
Heating Gas (in industrial furnace)	47.9	MJ/m3
Heating Oil (in industrial furnace)	49.7	MJ/l

Source: CAPRI Modelling System

Diesel fuel Among the direct energy sources in agriculture, diesel fuel is one of the most important. Due to the fact that CAPRI does not include consumption data of diesel fuel, the input quantity is calculated on an activity-based approach using normative data. Therefore, the German KTBL database (KTBL, 2004) is applied. This database offers consumption quantities on a standardized methodological basis for common crop production activities. Different parcel sizes ranging from 1 ha up to 80 ha-parcel and different soil qualities (light, medium, heavy) are additionally considered. To apply this range to CAPRI, the link between the European Soil Map and CAPRI was adapted. Using literature information, a classification of the different soil type classes into light/medium/heavy soils was done and linked to the diesel use database. Parcel size data, which were not available on a member states level, is estimated from EUROSTAT Farm Structure Survey (EU-FSS) data, parameter C-04 displaying numbers of field parcels per farm. To consider work steps in which diesel is used but that are not covered by the KTBL database, such as setup time of machinery, transport processes or feed preparation for animal feeding, additional consumption is charged. Furthermore, processes such as irrigation are considered in the diesel use. In order to link the estimations with national consumption statistics, a correction factor is established. A special focus is on the fuel consumption of grassland. On the one hand, consumption is yield-driven (high grass yield requires more cuttings per season), on the other hand the pasture share has an impact on the total quantity of diesel use. Both aspects are considered in the CAPRI energy module: a stepwise calculation depending on the yield level is carried out for the grassland that is mowed. The pasture share is indicated by national sources or UNFCCC. Minor amounts of diesel fuel are used if irrigation is applied, such consumption is charged to the activity being irrigated.

Electricity Electricity consumption plays a major role in animal production activities and drying cereals. Like diesel, data on electricity use is not included in CAPRI. Therefore a normative approach has been chosen to quantify consumption levels. Concerning housing systems, a distinction between the different activities has been made as well as a grouping of the EU countries in “North”, “Middle” and “South” to reflect the different requirements for heating and cooling. Charges in electricity use then have been set activity-specific and to a minor extent, depending on the herd sizes. Those charges are calculated on an animal-place basis. Electricity requirements for milk cooling are based on the CAPRI milk yield. Consumption quantities are taken from literature sources and calculations of Agroscope ART in Swit-zerland (Project BW04). Electricity used in grain drying is based on a normative approach due to lack of data of drying systems in the EU. Nevertheless, consumption quantities for drying are linked to the harvest moisture content as described in Chapter 7.5.2.11 FIXME. Furthermore, electricity use in greenhouses is considered. This is expressed depending on the lighting and heating efforts, as far as by using electricity. Small quantities of electricity are charged for lighting and ventilation in storage facilities for feeding stuff.

Heating oil and heating gas Main consumption sources for heating oil and heating gas are greenhouses and grain drying. Greenhouses consumption quantity is taken from member states statistics, where available. Alternatively, literature data including national information on heated greenhouses is considered. Grain drying process is displayed, as mentioned for electricity, based on a normative approach.

Indirect energy sources Indirect energy use describes external primary energy expenditures linked to materials utilised in production systems, balanced up to a defined system border (Diepenbrock, 1995; Moerschner, 2000). CAPRI energy indicator covers all relevant indirect energy sources. As for direct energy components, energy content data stem entirely from Ecoinvent modules using SALCA061 database (2006) to ensure a uniform assessment. The database for the most important indirect energy sources can be seen in Table 31. The following chapters give an overview on the methodology to estimate indirect components.

Table 31: Energy content factors for indirect energy input

Indirect energy source	Cumulative energy demand	Unit
Tractor	52.34	MJ/kg machinery weight
Harvester	49.27	MJ/kg machinery weight
Trailed Machinery	36.44	MJ/kg machinery weight
Nitrate fertiliser	58.99	MJ/kg nutrient
Phosphate fertiliser	40.06	MJ/kg nutrient
Potassium fertiliser	9.25	MJ/kg nutrient
Herbicides	218.62	MJ/kg active substance
Insecticides	299.02	MJ/kg active substance
Fungicides	124.38	MJ/kg active substance
Lubricants	79.17	MJ/kg
Minerals	13.52	MJ/kg
Salt	6.62	MJ/kg

Mineral fertiliser Mineral fertiliser energy assessment follows CAPRI-endogenous calculated fertiliser use. Thereby, the assessment is linked to net mineral fertiliser use. Such regionalized and activity-specific input quantities divided into the fertiliser groups (Nitrate, phosphate, potassium) are assessed by the energy content coefficients as shown in Table 31. Those are compiled using average registered consumption quantities of the single fertilisers on the market, which are broken down to their active substance content.

Machinery use and lubricants Machinery energy assessment is sub-divided into different machinery classes such as tractor, harvester and trailed machinery as well as special machinery such as irrigation or drying machinery on the one hand. On the other hand, a distinction between the machinery itself and the efforts for repairing and maintanance is made. In consequence, machinery stock data is required for every region. Such is partially available via EUROSTAT Farm Structure Survey (EU-FSS) parameter K-01 to K-03. The gaps have been filled, if available, with regional and national statistics. Tractor statistics are mostly available divided into different engine power classes, which permits a more detailed assessment mechanism. Energy assessment is carried out related to the physical weight of the machinery. Therefore tractor stock in a region is assessed with an average weight depending on the engine power class and afterwards sum up on NUTS-II level. The distribution of the weight over the useful lifetime of the machine is adequately to economic depreciation mechanism. An average useful lifetime of 20 years is assumed. This calculation step leads to the total machinery weight per NUTS-II region and year. The distribution towards the activities is calculated on a normative approach. Similar to diesel use, KTBL offers a database on machinery use expressed in machinery hours per ha for each activity under defined soil and parcel size conditions. This database divides between tractor-based processes and harvesting. The result of this procedure, expressed in kg machinery weight per ha is assessed with the energy content data shown in Table 31. Repair covering all exchanges of spare parts such as wheels, gearboxes, etc. during the lifetime of the machinery. The coefficient is determined by the energy depreciation factor. An equal approach is chosen for harvesters, whereas combine harvester stock is assumed to be used in the CAPRI CERE aggregate, other harvester stock by SUGB, POTA and ROOF. All activities receive an extra charge of trailed machinery. Such is, due to lack of data, determined by the tractor weight as a basis for activity-based coefficients on trailed machinery use. Trailed machinery receives energy for depreciation and repair. Lubricants‘ use is linked to machinery use time on an activity-based approach. Energy content is an average of different lubricants being used (such as engine oil, hydraulic oil, gearbox oil etc.).

Buildings energy use Quantifying buildings energy use on a regional scale is rather difficult task due to lack of data. None of the common database offers any statistics on the amount, age or structure of agricultural buildings. Those few member states offering such data on a national level do not provide a standardised methodology. In consequence a normative approach has been chosen for CAPRI energy indicator. This approach is based on a life cycle analysis study carried out at AGROSCOPE ART (Project BW04). Standardized building types for different animal production activities have been set up using architectural planning instruments ("ART Preisbaukasten") that permits quantifying the building material used and carrying out energy assessment. This data was broken down on a MegaJoule per square meter and year-term, whereas differentiation between depreciation, repair and maintenance and direct energy requirements was undertaken. Furthermore, several manure management systems are considered which permits using UNFCCC data on manure management (to be found in Table 4, UNFCCC "N2O Emissions from Manure Management") for the different NUTS-0 regions and the most important animal production activities. Finally, to take use of EU-FSS herd size distribution data on NUTS-II, energy for buildings is calculated for different herd sizes. Depreciation of buildings energy is carried out following an economic depreciation approach whereas a useful lifetime of the building of 50 years is assumed. Depreciation covers efforts for building construction and waste disposal. Those parts of the building which have shorter useful lifetime, the exchange of spare parts and facilities as well as the waste disposal for such material is charged in the repair factor. The space charged for animals covers the entire stable area excluding space for feeding stuff. Pure animal space follows Swiss minimal space requirement regulations. To consider different building requirements between the regions of the EU, three region aggre-gates (North, Middle, South) are set up. For the set "South", a typical Italian stable for cattle has been calculated. The charge for "Middle" is calculated based on Swiss stable systems, whereas "North" receives extra charges for heating. Due to lack of data for HENS and POUF, a pig breeding stable containing poultry-specific place requirements is taken for the calculations. Storage facilities for feeding stuff is charged depending on the input quantity of the relevant feeding stuff component, whereas a drive-in silo is the main type of storage facility. Such is charged for MAIF, GRAS and OFAR. Ma-chinery storage in barns is charged depending on the machinery size, derived from the engine power class, building type of the barn and a storage rate. Depreciation and repair distribution is equal to other building types.

Crop Protection To reflect energy input via pesticides, the CAPRI-FADN data on monetary efforts for crop protection is used. Due to the fact that FAOSTAT offers consumption quantities of the different agents on a national level, a mechanism has been chosen to get those two parameters "quantities" and "energy content" together. Data from the EAA database helps to create the link in-between. Multiplying the quantities applied with the energy content data, the total sector energy consumption quantity can be calculated. Beside the pesticide categories shown in Table 31, growth regulatories are included in the calculation. Using the sector expenses, the "energy value" of plant protection application, expressed in MJ/€, is the basis for an activity-based assessment. The last step links the hectar-based CAPRI expenses for plant protection with the "energy value". Certainly, this approach does not consider the shares of the different agents applied per activity, but taking the minor overall role of pesticides, it seems appropriate to follow the way described.

Seed For considering seed in terms of energy, a distinction between certified and non-certified seed is done. A broad range of statistics, both on national and regional level indicate the share of certified and non-certified seed use. Information about total quantities applied is available for most CAPRI activities from literature. Non-certified as well as certified seed contain a "basic value" covering energy efforts for production of the output. Non-certified seed is being assumed to remain in the NUTS-II region for local production. Additionally to the basic value, energy efforts for cleaning, chemical treatment and storage are charged. Certified seed is charged, beside the basic value, with energy requirements for breeding, treatment, cleaning, packaging and transport.

Drying energy efforts Energy required for drying mainly consists of two parameters: Firstly, the difference between harvest moisture content of the cereals and the marketable final moisture content and secondly the direct and indirect energy requirements for the reduction of one unit of moisture content. Estimation of harvest moisture content is carried out with the help of a regression model. To deliver explanatory variables, German harvest statistics are applied. In a first step a linear model is set up for each activity using climate data to find out an interrelationship between climate data and harvest moisture content. In a second step the linear models are applied for other EU countries using EU climate data to project harvest moisture content for regions where no harvest statistics are available.

Three different datasets are used for the generation of the projection module: Harvest statistics of Germany, Climate Data for the EU and Data about cereal cultivation regions in the EU:

- Harvest statistics of Germany: Data stem from a representative statistic survey of the years 2000, 2001 and 2002. Data is shown for 13 NUTS-I regions (excluding the city NUTS-I regions) and gives information about the weighted average moisture content of harvested cereals, divided into the activities wheat, rye, oats and barley.
- Climate data for the EU: Data stem from Climate Research Unit (CRU) of University of East Anglia in the version of CRU TS 2.1. Equally data from the years 2000, 2001 and 2002 is used. Addi-tionally long-term climate data is used displaying a 30-year average from the years 1961-1990.
- Cultivation Data for the EU: a dataset showing 0.5 x 0.5 degree grids with a cereal share lower than 10 % of UAA (based on CAPRI disaggregation crop data) was used to exclude grids being assumed irrelevant for the estimation process.

The harvest moisture statistical data and climate data was linked. The first step of the core statistic model was a principal component analysis (PCA), in which a broad range of variables were summarised into fewer principal components while preserving variability in the original variables. In the next step, the linear model was used to predict the average moisture content for regions, where no harvest moisture content data was available. Therefore, climate data as described above was used. Beside the exclusion of grid cells with a cereal area share lower than 10 percent of the UAA, a number of regions where grain drying is not applied, where not further considered. For the remaining regions, for each grid and production activity, a harvest moisture content estimate was calculated by the use of the linear models described above. In a fourth step, average harvest moisture content estimates are calculated by NUTS-I region and activity. Finally, the energy requirements for the reduction from the estimated moisture content to the marketable final moisture content was calculated.

Irrigation energy Energy requirements for irrigation consist of direct and indirect components. Indirect requirements display machinery depreciation and repairs. EU-FSS (Parameter K-10) as well as national and re-gional sources indicate the share of mobile and fixed irrigation equipment. Furthermore, mainly national sources indicate share of surface and groundwater source of irrigation water. The irrigation machinery type is calculated based on Econinvent inventories. Depreciation and repair efforts are charged as described in Chapter 6.5.2.7 FIXME. Direct energy requirements are largely depending on the water quantity applied. Such data is delivered either by FAOSTAT or by national sources. Partially, mainly for Italy, Greece, France and Spain, activity-based data on irrigation water quantity is available. Due to a lack of statistical data and for plausibility reasons, the main energy source for irrigation was assumed being electricity, partially also diesel.

Energy requirements for greenhouses Energy consumption for greenhouses is determined by direct energy consumption for heating, supplementary illumination, disinfections of soil, substrate and drain water as well as minor efforts for buildings. Due to lack of data, barn energy requirements are charged. Direct energy sources are the main drivers for high total requirements. Concerning the area under glass, EU-FSS (Parameter D/15, D/17, G/07 and I/04) and several national institutions provide data, partially activity-based and mainly on NUTS-II level. For those regions where no model data was available, national institutions offered part of the data, whereas in some cases, no indication about activities was available. To distribute all available information and provide consistency as well as smooth greenhouses shares, a PMP term brings together the single components. Having set the share of each activity level under glass, the major part of energy consumption via greenhouses is calculated. Therefore, on the one hand, national consumption data is considered. This implies information on the activity-specific heated share of greenhouses, on consumption quantities of direct energy and on direct energy sources (heating gas, heating oil, coal, etc.). On the other hand, literature data is considered, where no national consumption statistics are available. Such methodology is only valid for Middle and Northern European countries, as, following literature, most greenhouses located in the Mediterranean basin could be considered as passive systems since they use very little external energy.

Feeding stuff In animal production, feeding stuff plays a major role in energy consumption concerns. Quantification of the requirement is rather complex. The most important database are feeding coefficients implying quantities of different feeding stuff components on an activity- and regionalized basis. Furthermore, additional information on import shares, either on a national or on EU-level are required. Such coefficients are extracted from the CAPRI feeding module. Having those, the energy assessment is carried out. Taking basic ratio feed components (GRAS, MAIF, ROOF, OFAR), the following elements are charged: Firstly, production requirements are considered. Such cover all direct and indirect energy needs during the production process, divided by the yield. Secondly, processing efforts, such as storage and feed preparation are charged. Concerning grassland, pasture share is considered in the cal-culation process. Taking concentrates, things get a bit more complex. Those concentrates‘ components, that are produced in a NUTS-II region and consumed there are charged by the production requirements plus efforts for storage and processing (such as milling, mixing etc.). Those parts that stem from the relevant NUTS-0 region are furthermore charged with transportation needs. Finally, those parts of the ratio that are imported into the EU-27 receive a different treatment. Due to lack of methodological adequate assessed energy requirements of overseas production, an average of EU-27 production needs is assumed. Because data is not available or cannot be extracted in a meaningful way for some components (such as for soybeans), literature data are used. Then, overseas shipment is charged and the remaining processing efforts are considered as described above. Finally, feed supplements such as salt or minerals are charged taking animal-specific consumption quantities and energy content factors (Table 31).

6.5.3 Energy output assessment

In order to calculate energy balances or efficiency parameters, the output generated by agricultural production has to be assessed by its energy content. The CAPRI output level on the one hand is a basis for this assessment. Energy content factors, on the other hand, are used. Those are based on literature research. Basically, the assessment follows a caloric approach designed by FAO. Main coefficients are based on FAOSTAT data, some are taken from Mittenzwei (2006)

Energy allocation For activities producing more than one marketable product (e.g. DCOW: COMI, BEEF, YCAM, YCAF), an allocation between the main output and the side-products has to be carried out. For plant production activities, no such allocation is done, the main product is charged with the complete energy needs. The allocation parameters assumed for animal production activities are shown in Table 32. The procedure follows literature data.

Young animals To achieve a consistency in energy balances for animal activities, young animals assessment is an important item. To achieve such, all energy requirements necessary for a young animal are summed up following the lifelines within the young animal module of CAPRI. Nevertheless, an allocation of the energy content has to be carried out and follows allocation shares shown in Table 32.

Table 32: Allocation of animal products

CAPRI activity	Main product share (%)	Side product N°1 share (%)	Side product N°2 share (%)	Side product N°3 share (%)	Side product N°4 share (%)
DCOW	COMI: 88	BEEF: 8	YCAM: 2	YCAF: 2	COMF: 0
SCOW	YCAM: 44	YCAF: 44	BEEF: 8	COMF: 4	-
SOWS	YPIG: 100	PORK: 0	-	-	-
SHGM	SGMI: 50	YLAM: 30	SGMF: 10	SGMT: 10	-

6.5.4 Analysis of CAPRI energy module results

The results of the CAPRI energy module can be displayed in various ways and on different levels. Table 33 gives an overview. Further down, each parameter is shown in more detail.

Application notice Basically, the energy module is designed as post-model analysis. This implies, that the energy module can be run independently from the CAPRI core model. Nevertheless, a number of energy parameters depend on CAPRI data that changes depending on the scenario and the time under consideration. Consequently, the energy module has to be run each time changes in a scenario result table occur. Having set changes in any of the energy module‘s files, the GAMS-file “enerind_bas” has to be run. To transfer such changes in the energy module to the scenario tables, the GAMS-file ”enerind_calc” has to be run in each scenario mode.

Structure of output tables A broad range of output tables permits to display results of the energy indicator. Some of those are presented in this chapter. Figure 41 shows some examples for display modes of the energy indicator.

Example 1: Energy consumption - overview

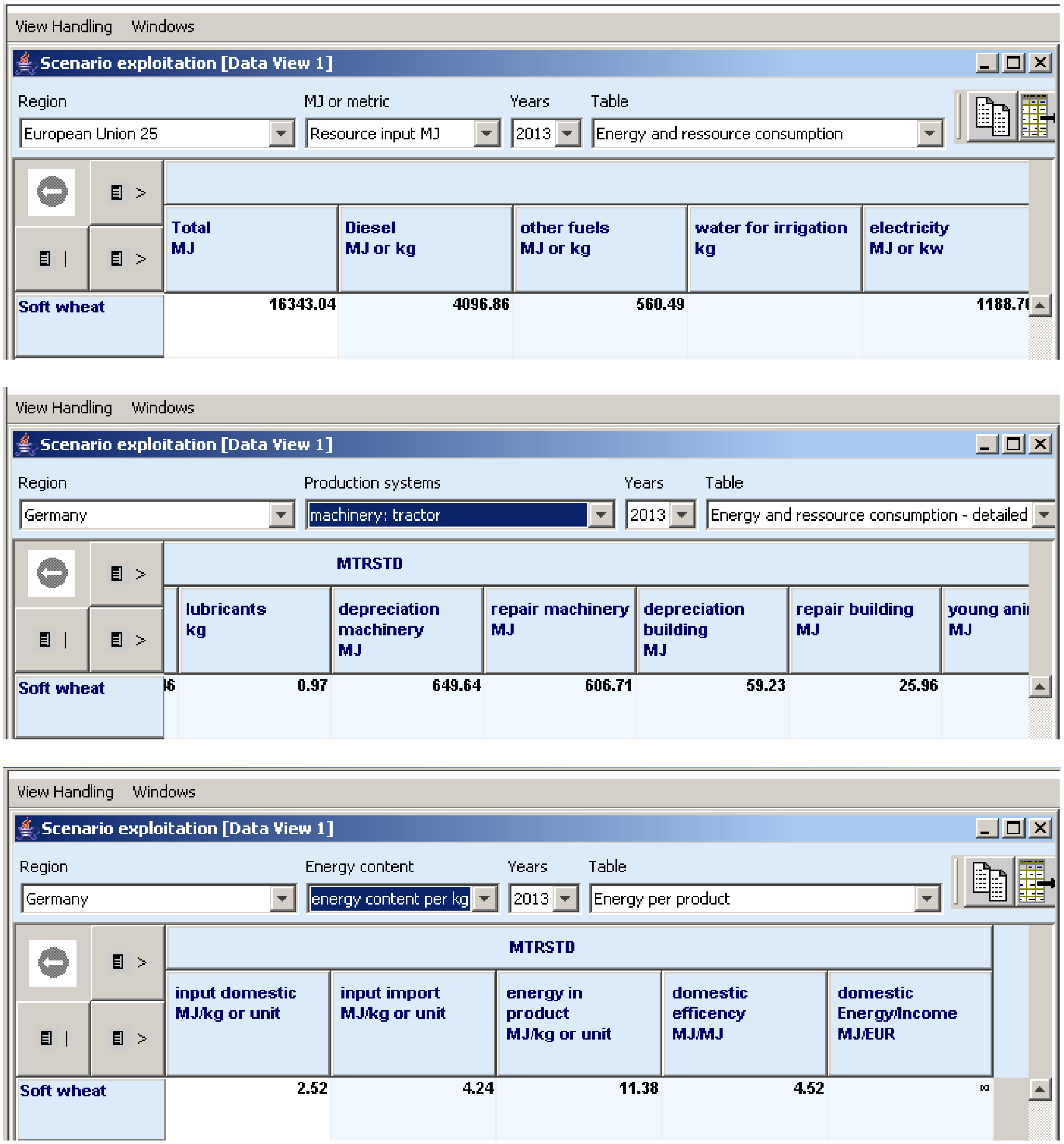
Example 2: Energy consumption - detailed

Example 3: Energy parameters with reference to the product

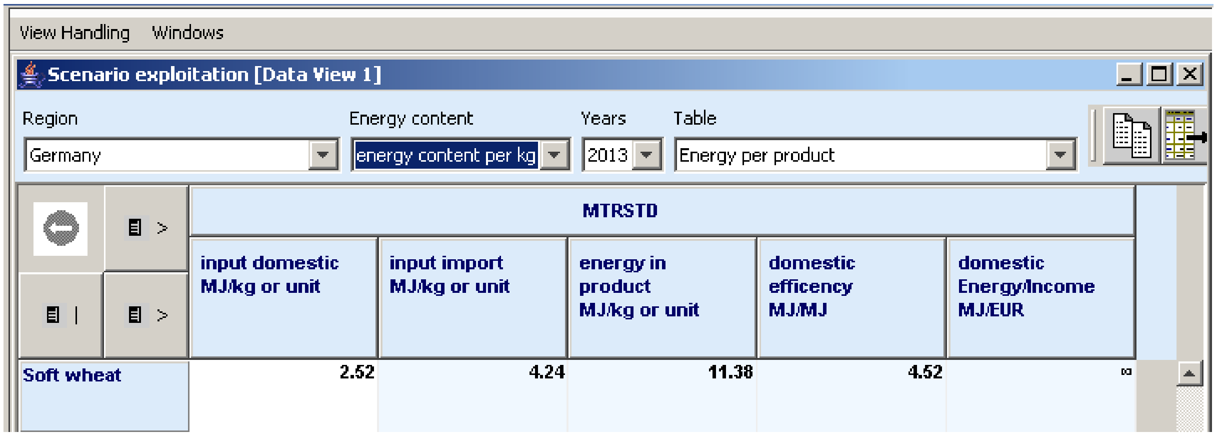
Example 4: Energy parameters: Sectoral balances ;Source: CAPRI Modelling System

Taking Example 1, the results of the energy consumption overview table are shown. This can be explored within the scenario exploitation table. Beside “Total MJ”, which indicates energy consumption per ha or head, a number of energy consumption categories such as diesel, electricity, machinery, fertiliser, young animals, seed and plant protection can be displayed either on a MJ basis or in metric units. Furthermore as can be seen in Example 2, detailed data on energy consumption can be displayed if required. Data on feeding stuff, housing systems, grassland use, tillage systems, machinery use, irrigation, greenhouse use, seed, plant protection, drying etc. can be shown activity- and region specific. Beside regarding the area or the animal, the product can be chosen as a reference point. As shown in Example 3, energy requirements

Figure 41: Energy parameters: examples for results displaying



per kg of product (expressed in MJ/kg) and domestic en-ergy efficiency (expressed in MJ/MJ) can be



shown. On a sectoral basis, efficiency related to the in-come (expressed in MJ/€) is displayed. A sectoral balance can be extracted as shown in Example 4 summing up all energy requirements and all energy output.

Chapter 7

Spatial dis-aggregation capdis module

7.1 Introduction

Environmental impact assessment of farm management decisions is often only possible in a spatial context, taking local conditions as land cover, climatic and soil conditions into account. Equally, farm management decisions depend to a larger extent on those local factors. Therefore, a CAPRI module has been developed that allows monitoring and ex-ante assessment of environmental impacts of agriculture at a 1×1 km spatial scale.

7.1.1 Purpose of disaggregation of CAPRI results

Update pending

7.1.2 A brief history of the CAPDIS module

Update pending

CAPRI-DynaSpat
HSU

7.1.3 Applications

Update pending

Link to DNDC-Europe
NitroEurope
European Nitrogen Assessment
IDEAg
CAPRI-RD

7.2 CAPDIS module – GUI basics

Files:

```
%curdir%/capdis.gms
%curdir%/capdis/capddis_relevantsets.gms
```

The distribution of crop shares, livestock numbers, yield and nitrogen flows is a sequential process of a number of sequential steps (CAPRI tasks). The distribution of land use and livestock numbers, yield and irrigation shares in each step is based on the results of the previous step which is used as priors.

A priori land use distribution (mode: LAPM): The first step uses statistical information available at the intersection of 10 km x 10 km grid cells and administrative units at the NUTS3 level (*literature source to be pasted* FIXME). This data is disaggregated into the spatial units.

CAPREG disaggregation (mode: CAPREG): the a priori land use distribution is used as prior information to disaggregate statistical information for the CAPRI base year, currently the year 2012.

Timeseries disaggregation (mode: TIMESERIES): results from the CAPREG disaggregation are used as priors for the time series disaggregation.

Baseline and scenarios (mode: CAPMOD): results from the CAPREG disaggregation are used as priors for the ex-ante simulation results

The disaggregation module works at the level of CAPRI NUTS2 regions, which are run in parallel mode. The number of parallel instances depends on the computing capacity (check CAPRI settings, tab 'GAMS and R', Number obtained from 'Get the number of processors …' or written into the combo box). To combine result files at regional level into country-level result files, a CAPRI task 'Collect disaggregation results' is available. This can be done for results of all other CAPRI disaggregation tasks. Optionally, a new country file can be generated (overwriting previous results) or an existing country file updated (overwriting only the data of the 'new' NUTS2 results). Also the NUTS2 files can be deleted.

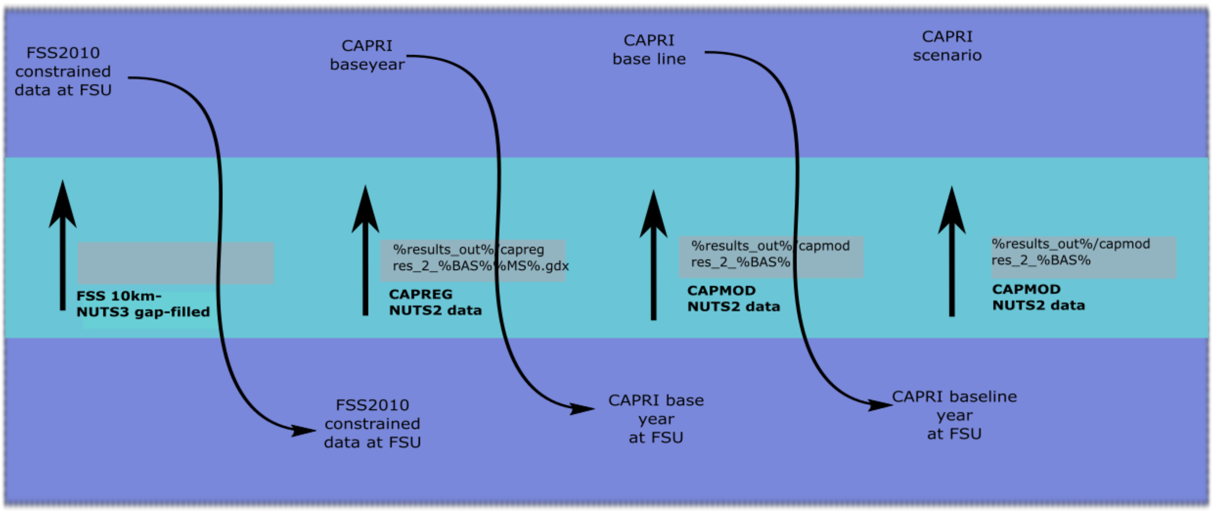
To activate CAPDIS (disaggregation module), select the CAPRI workstep "Disaggregate Results". Then select one of the following CAPRI tasks: A priori land use distribution, Disaggregate CAPREG, Disaggregate timeseries, Disaggregate baseline, Disaggregate scenario, or Collect disaggregation results. All tasks can run in one go, or in separate simulations, if result files of the previous variable are available (e.g. land use already disaggregated for the disaggregation of livestock).

7.3 Farm Structure Units – FSU

Files:

```
%datdir%/capdishsu/s_fsu_srnuts2.gdx
%datdir%/capdishsu/fsunogo.gms
%datdir%/capdishsu/m_grid10n2.gms
%datdir%/capdishsu/p_fsu_grid10n2.gdx
%datdir%/capdishsu/p_fsu_srnuts2.gdx
%datdir%/capdishsu/p_fsu_area.gdx
%datdir%/capdishsu/p_fss2010.gdx
```

Figure 42: Sequential disaggregation of land use



Source: own illustration. Note: this figure needs update and does not yet include the disaggregation of time series.

7.3.1 Delineation

Update pending

Administrative units
Inspire grid
Soil mapping units
NOGO and Forest units (Corine)

7.3.2 Data preparation

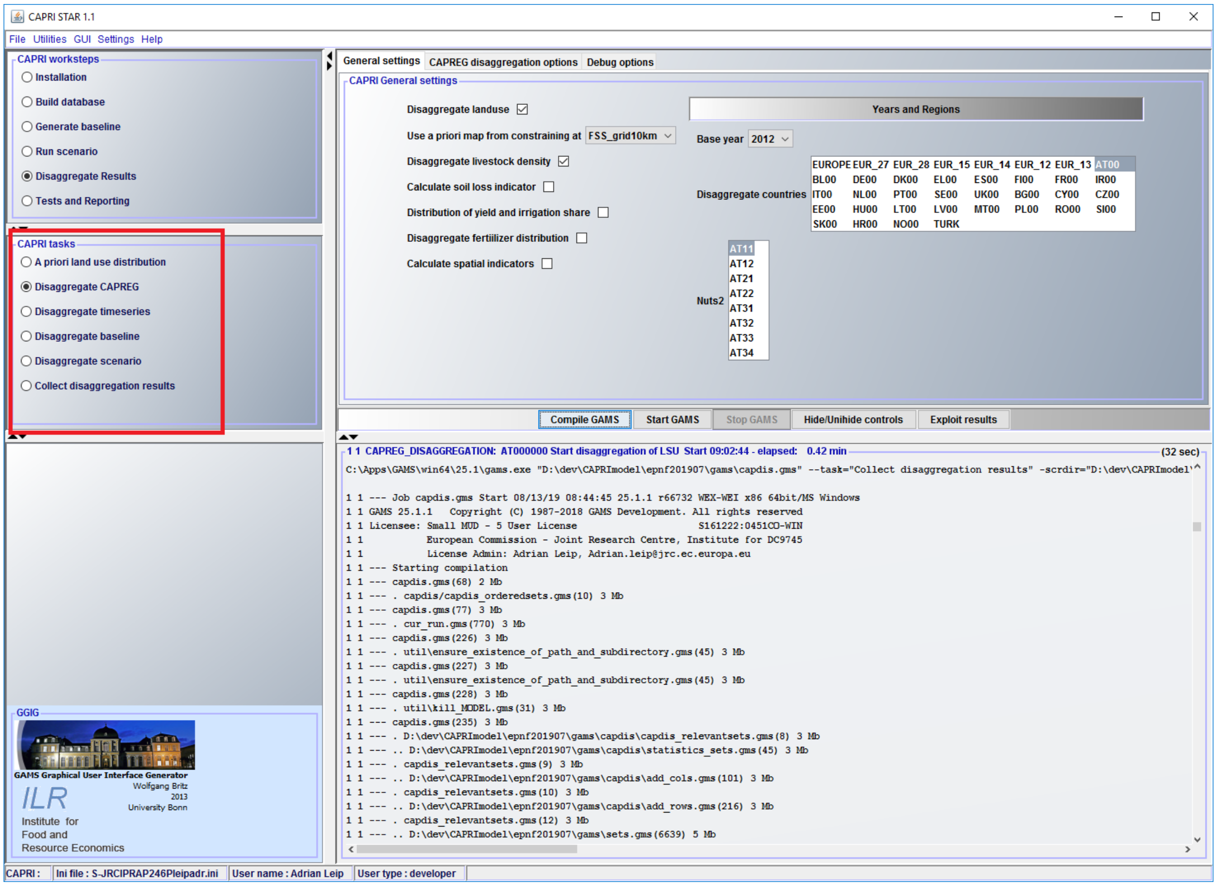
Update pending

Determination of the centre
Corine Shares
DEM
PESETA
EMEP

7.4 Disaggregation of crop Areas

Files:

`%curdir%/capdis/disyield.gms`

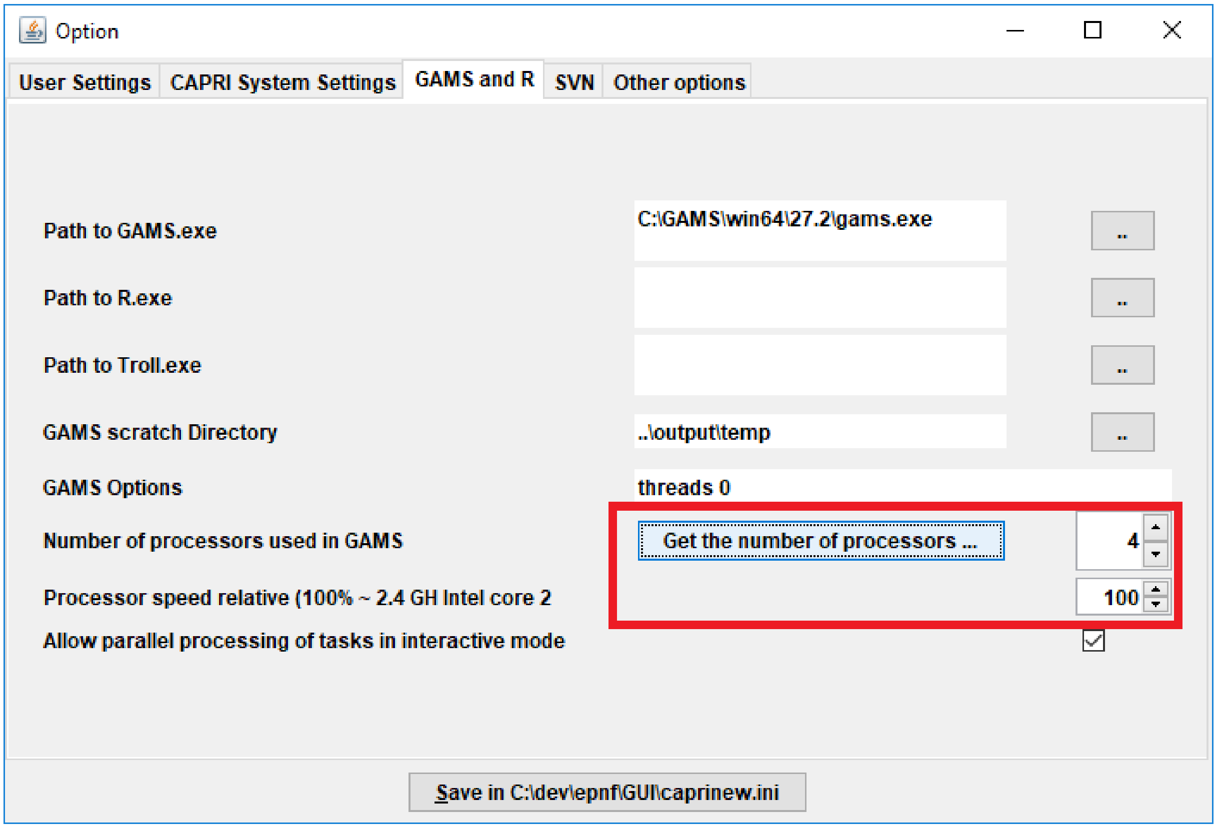



```
%curdir%/capdis/disyield_sets.gms
%curdir%/capdis/m_hpdCropSpat.gms
%datdir%/capdishsu/pesetagrid_fractionfsu.gdx
%datdir%/capdishsu/irriShare2000fsu.gdx
```

The principle of the distribution of crop areas is based on few constraints only: full exhaustion of available ares for each spatial unit, vertical consistency, and primacy of land stability.

Vertical consistency means that the sum of each land use type over all spatial units recovers available land use at the higher spatial level. As available information are not (necessarily) geo-referenced, the allocation of a given statistical information to a spatial unit is associated with uncertainty. For example, a farmer with the residence in region A can have land also in regions B and C, but will declare them together, and they will be allocated to her residence (region A). Accordingly, there is some blurring in particular at boundaries and this is accounted for in the methodology: also at the high disaggregation level, the land uses are principally to be interpreted as 'Land owned by a farmer with residence in this spatial unit'.

Primacy of land stability means that if there is no indication (i. e. new observation, policy restricting



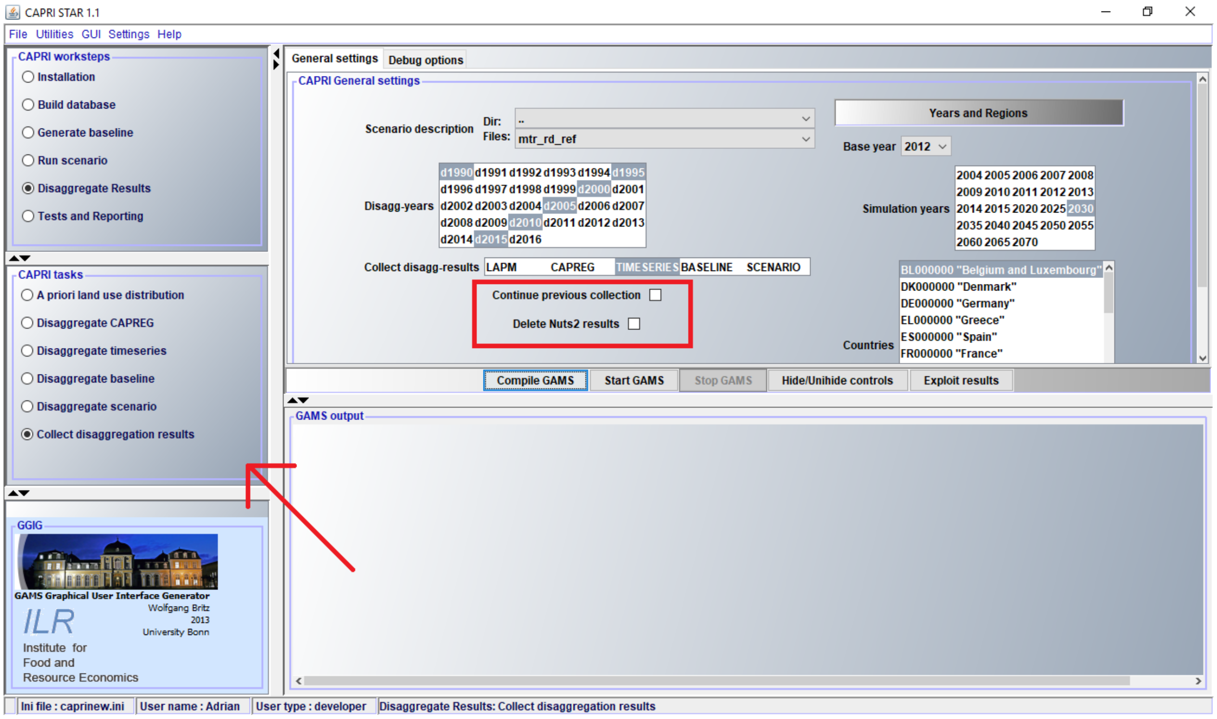
previous land distributions, ...) it is more likely that the spatial pattern remains similar to the previous (prior) pattern. Therefore, once a likely distribution of land and livestock has been determined on the basis of high-resolution FSS statistics, the model tries to stay as close as possible to this distribution. This is achieved with penalty factors that are activated as soon as the estimated land use area deviated from the prior values, assigning a higher penalty for deviations of permanent crops and forests, and very high penalties of a land use is estimated in a spatial unit where it didn't exist in the prior's data base.

The disaggregation model m_hpdCropSpat is described in **??**. **??** describes the required input data, and **??** describes the preparation of the input data for their use in hpdCropSpat.

7.4.1 Simulation model m_hpdCropSpat

Equation 1 RULEVL_

```
     The levels assigend to the FSUs must recover the given NUTS II area
     both NUTS2 level and FSU level in 1000 ha
     rur = grid cell in mode LAPM, rur = NUTS2 in all other modes
RULEVL_(rur,curact) ..
SUM(regmap(rur,cur%spatunit%),v_levlCon(rur,cur%spatunit%,curact))
```




```
=E= p_nutsLevl(rur,curact);
```

$$A_{c^*,r} = \sum_{h} \{a_{c^*,h}\} \tag{7.1}$$

$a_{c^*,h}$ = Area [parameter, km^2] cultivated with crop c or covered by 'other land' use excluding forest in spatial unit h
$A_{c^*,h}$ = Area [parameter, km^2] cultivated with crop c or covered by 'other land' use excluding forest in region r

Equation 2 ADDUPGRID_ Due to several reasons it could be impossible to distribute all agricultural area under the given constraints of total available area in the spatial units (net of the area of 'nogo' units and forest area). In order to enable a feasible solution, an error term is introduced that allows the units to slightly shrink or grow. The reason:

- Statistical data are not (necessarily) geo-referenced; i.e. an area of crops (or a livestock) might be assigned to one unit/grid cell because this is where the farm is registered rather than the physical location of crop/livestock
- Uncertainties in the data
- Inconsistencies of data sources (i.e. FSS agricultural statistics, Corine Land Cover data, CAPRI regional statistics)

1. – The FSU area must be exhausted, but the variable v_%spatunit%SizeChg

allows some flexibility if needed.

```
ADDUPGRID_(rur,cur%spatunit%) $ p_levlunit(rur,cur%spatunit%,"area") ..
SUM(curact,v_levlCon(rur,cur%spatunit%,curact))
=E=
v_%spatunit%SizeChg(rur,cur%spatunit%)*p_levlunit(rur,cur%spatunit%,"area");
```

$$a_h \epsilon_{a,h} = \sum_{c^*} \{a_{c^*,h}\} \tag{7.2}$$

a_h = Area [parameter, km^2] of spatial unit h
$a_{c^*,h}$ =Area [parameter, km^2] cultivated with crop c or covered by 'other land' use excluding forest in spatial unit h
$\epsilon_{a,h}$ = Error term, allowing a spatial unit to shrink or grow slightly in order to enable a feasible disaggregation of statistical data.

Equation 3 PDF_ The most likely solution is obtained with the 'Highest Posterior Density' method. A penalty function calculates deviations from prior information, applying uncertainties

- A random re-allocation of crops should be avoided. Therefore, a penalty is given with increasing deviation from the prior distribution to ensure stability in the time series
- In particular the 'appearance' of crops in spatial units where they have not been observed in the prior data should be restricted (disagg("penelizenewcrops"))
- The error term for the area of the spatial units should be kept at a minimum (disagg("penalize-sizechange"))

```
PDF_ ..
# Scale density value a good couple of magnitudes for numerical reasons.
v_hpd
*[SUM((curact,regmap(rur,cur%spatunit%)) $ p_levlStde(cur%spatunit%,curact),
p_levlunit(rur,cur%spatunit%,"area"))
 +SUM(regmap(rur,cur%spatunit%) $(v_%spatunit%SizeChg.LO(rur,cur%spatunit%) NE
v_%spatunit%SizeChg.UP(rur,cur%spatunit%)),
    p_levlunit(rur,cur%spatunit%,"area"))
 ]
=E=
# hsu area-weighted mean square of the deviation from prior mean area, scaled by its stdev
(SUM((regmap(rur,cur%spatunit%),curact),p_levlunit(rur,cur%spatunit%,"area")
       * SQR( (v_levlCon(rur,cur%spatunit%,curact)-p_levlunit(rur,cur%spatunit%,curact))
               *(                  1 $      p_levlunit(rur,cur%spatunit%,curact)
                  + disagg("penalizenewcrops") $ (not p_levlunit(rur,cur%spatunit%,curact)))
               /max(1e-3,
                         $$ifi     %MODE%==LAPM p_levlStde(cur%spatunit%,curact)
                         $$ifi NOT %MODE%==LAPM 1
                   )
```

```
            ) \\
    )/SUM((regmap(rur,cur%spatunit%),curact)  \\
$p_levlStde(cur%spatunit%,curact),p_levlunit(rur,cur%spatunit%,"area")) \\
)$sum((regmap(rur,cur%spatunit%),curact),p_levlStde(cur%spatunit%,curact)) \\
# penalty for deviation from hsu area \\
+(SUM(regmap(rur,cur%spatunit%),  \\
disagg("penalizesizechange")*p_levlunit(rur,cur%spatunit%,"area")*SQR((v_%spatunit%SizeChg(rur,cur%spatuni
1))) \\
 /SUM(regmap(rur,cur%spatunit%), p_levlunit(rur,cur%spatunit%,"area")) \\
 )$SUM(regmap(rur,cur%spatunit%),p_levlunit(rur,cur%spatunit%,"area")) \\
; \\
```

Model parameters Some model parameters can be set by the user through the CAPRI GUI. They are collected in the parameter 'disagg'.

```
set disaggcontrol /
mincropshare "Minimum allowed cropshare per HSU"
relcropshare "Defines heterogeneity of crop shares for a crop per HSU"
relstdefix   "Relative standard deviation, predefined if land needs 'to be fixed'"
relstdeperm  "Relative standard deviation, predefined for permanent crops"
relstdeothe  "Relative standard deviation, predefined for other land (large to avoid that other land pushe
penalizenewcrops   "multiply deviations for crops predicted where they haven't been before"
penalizesizechange "multiply deviations for total HSU unit size"
* --- scalars controlling livestock disaggregation
weightRUMIfodduaar "Weighting between fodds and uaar to distribute initial RUMI numbers"
weightMONOcereuaar "Weighting between cereals and uaar to distribute initial NRUMI numbers"
minLSUdens          "Minimum density for LSU allowed to not have them everywhere...     "
# managing crop residues
minmcactSurs   #Miniumum surplus as compared to average over all crops
maxmcactSurs   #Maxiumum surplus as compared to average over all crops
rangemcactSurs #Range of sursoi (max/average) below which the high sursoi are not reduced
/;
```

- Stability of forests – disagg("relstdefix")

Forests cannot easily be 'displaced' and are likely to remain rooted as given in the land cover data sets. So far, estimations of changes of forest areas at the regional level are not included in the disaggregation procedure.

The default value used for disagg("relstdefix") = 0.01

The lower the value the higher becomes the penalty if the estimates are deviating from the priors.

- Stability of permanent crops disagg("relstdeperm")

Permanent crops are long-term investments and require time to grow. Displacement of permanent crops is slow.

The default value used for disagg("relstdeperm") = 0.05.

- Coefficient of variation for 'other land uses' disagg("relstdeothe")

'Other' area is a lump of all non-agricultural areas. We consider this area as relatively flexible.

The default value used for disagg("relstdeothe") = 1.

- Penalization for new crops in spatial units disagg("penalizenewcrops")

New crops 'appearing' in spatial units, if they have not been in the priors data set, are penalized. The penalization factor is a multiplicator of the squared deviation from the prior. Thus, the higher the factor the higher becomes the penalty.

The default values used for disagg("penalizenewcrops ")=2.0;

- Penalization of area changes of spatial units disagg("penalizesizechange")

The default values used for disagg("penalizesizechange")=2.0;

- Minimum crop share allowed in the spatial unit disagg("mincropshare")

The minimum crop share which is allowed in the spatial unit _(min) is used to calculate the lowest allowed crop share, in combination with the minimum relative crop share defining the level of spatial heterogeneity for a crop. See section 7.4.3.5 FIXME .

$_{min}$ can be set through the CAPRI GUI (tab CAPREG disaggregation options – "Suppression of crops if the share is very low")

By default, $_{min}$is set to zero.

- Minimum relative crop share disagg("relcropshare")

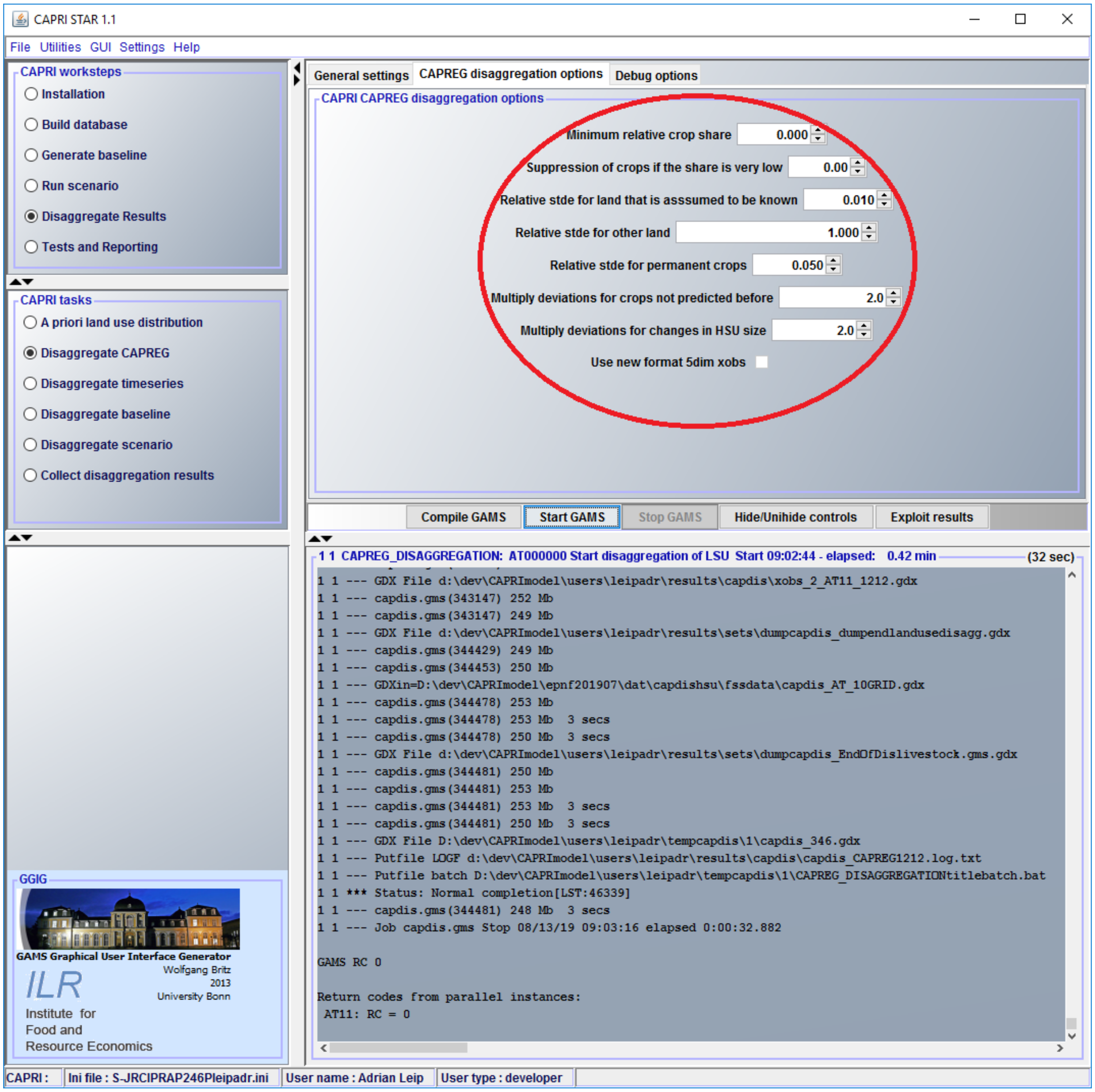
The minimum relative crop share defining the level of spatial heterogeneity for a $_{rel}$ is used to calculate the lowest allowed crop share, in combination with the m crop minimum crop share which is allowed in the spatial unit (see section 7.4.3.5 FIXME).

$_{rel}$ can be set through the CAPRI GUI (tab CAPREG disaggregation options – "Minimum relative crop share")

By default, $_{rel}$ is set to zero.

Defining bounds for the land use distribution model *Bounds for size changes of the total area of the spatial units*

```
v_%spatunit%SizeChg.L (rur,cur%spatunit%) $p_temp3dim(rur,cur%spatunit%,"crops")= 1;
v_%spatunit%SizeChg.UP(rur,cur%spatunit%) $p_temp3dim(rur,cur%spatunit%,"crops")= 1.1;
v_%spatunit%SizeChg.LO(rur,cur%spatunit%) $p_temp3dim(rur,cur%spatunit%,"crops")= 0.9;
v_%spatunit%SizeChg.UP(rur,cur%spatunit%)
$(p_nutslevl(rur,"AREAcorr") and p_temp3dim(rur,cur%spatunit%,"crops"))= 2.0;
v_%spatunit%SizeChg.LO(rur,cur%spatunit%)
$(p_nutslevl(rur,"AREAcorr") and p_temp3dim(rur,cur%spatunit%,"crops"))= 0.5;
$$ifi %MODE%=="LAPM" v_%spatunit%SizeChg.UP(rur,cur%spatunit%) $p_temp3dim(rur,cur%spatunit%,"crops")=
max(1.1,p_nutslevl(rur,"AREAcorr"));
$$ifi %MODE%=="LAPM" v_%spatunit%SizeChg.LO(rur,cur%spatunit%) $p_temp3dim(rur,cur%spatunit%,"crops")=
```




```
1/max(1.1,p_nutslevl(rur,"AREAcorr"));
```

To enable the solver to find feasible solutions even in difficult situations, it is possible to expand or shrink the total area of the spatial units. This is in consistence with the definition of the data compiled by the statistical offices which link the area to residence to the farmer rather than to the geographic location of each field.

Generally, we limit this area-change to plus/minus 10% of the original size.

In cases where inconsistencies between data sets have already been identified (see 0) a higher degree of flexibility is allowed (factor 2) as it is not known in which spatial unit the inconsistency originated.

Only in the task 'A priori land use distribution' the degree of flexibility is calculated as a function of the correction that had to be applied to the regional area.

The bounds for the area-size change are hard-coded and can not be changed by the user.

Setting standard deviations Data do not come with any level of uncertainty attached, and there is no a priori information on what spatial distribution is more likely than any other.

Therefore, the uncertainty in the estimates is 'guessed' based on crop groups.

Other options tested (all standard deviations equal or scaling prior estimates to a plausible range) are currently not used. The standard deviations are only set at the first task. In subsequent tasks, the standard deviations of the priors are used and 'gap-filled' if necessary.

```
$set changelapmstdev bygroups
$iftheni.std %changelapmstdev%=="scalestdevs"
$elseifi.std %changelapmstdev%=="allone"
$elseifi.std %changelapmstdev%=="bygroups"
    p_levlstde(cur%spatunit%, %croptp%)
*        $ p_levlstde(cur%spatunit%, %croptp%)
      = 0.5;
    p_levlstde(cur%spatunit%, %croptp%)
*        $ p_levlstde(cur%spatunit%, %croptp%)
      = 0.001 $sum(fssact2groups(%croptp%, "FORE"), 1)
      + 0.50  $sum(fssact2groups(%croptp%, "CERE"), 1)  # Cereals incl those likely in rotation: Assumptio
      + 0.25  $sum(fssact2groups(%croptp%, "FODD"), 1)  # Fodder crops: roof, ofar, lgras. Assumption: lin
      + 0.25  $sum(fssact2groups(%croptp%, "OILS"), 1)  # Oil crops: rape, sunflower, soya. Assumption: re
      + 0.15  $sum(fssact2groups(%croptp%, "VEGE"), 1)  # Vegetables: flower, pulses, potatoes, sugar beet
      + 0.05  $sum(fssact2groups(%croptp%, "TREE"), 1)  # Permanent crops: olives, nurseries, fruit and nu
      + 0.05  $sum(fssact2groups(%croptp%, "PERM"), 1)  # Permanent crops: olives, nurseries, fruit and nu
      + 0.80  $sum(fssact2groups(%croptp%, "REST"), 1)  # Assumptions: can be easily pushed around
      ;
```

7.4.2 Data sets

Update pending

FSS 2010 data at nested grid levels
FSS 2010 data, gap-filled at 10km-NUTS3 overlay
Forest map

7.4.3 Data preparation

Re-mapping from FSS crops to CAPRI crops

```
# Convert to CAPRI activities (posteact)
#As grids and NUTS are not always consistent - use the grid-HSUs to fill with crops
p_hsufraction(curgrid,%spatunit%_all)=p_hsu_grid10n23(%spatunit%_all,curgrid,"fracHSU");
grid10n23_hsu(curgrid,%spatunit%_all)$p_hsufraction(curgrid,%spatunit%_all)=yes;
cur%spatunit%(%spatunit%_all)=yes$sum(grid10n23_hsu(curgrid,%spatunit%_all),1);
```

Intersecting FSS 2010 10km x NUTS3 to FSU The land use distribution model works at the spatial intersection of the prior and posterior spatial units. Historically (when CAPDIS was based on HSU) this intersection was an area determined by the fraction of the HSU that lies within different FSS-admin grid cells. This fraction $f_{hg} \in [0, 1]$.

However after the update to the FSU, each spatial unit was fully in one FSS-admin grid cell only, thus $f_{hg} \in \{0, 1\}$. The following text is therefore relevant only if CAPDIS is run with 'old' HSU.

```
# Work on the intersection between grid and HSU (Line 585 ff)
p_levlunit(%region%,cur%spatunit%,"AREA")
=p_hsufraction(%region%,cur%spatunit%)*p_hsu(cur%spatunit%,"area");
" Work on intersection of HSU and grid for crops"' '" "'
# Distribute crops over intersected units -and scale to total area (should be already but not always is...
p_levlunit(%region%,cur%spatunit%,%croptp%)
=p_hsufraction(%region%,cur%spatunit%)*p_levlmean(cur%spatunit%,%croptp%);
# Scale all areas such that the sum becomes the AREA of the unit (in case the HSU had to be split)
#    - if LAPM predictions are consistent with total area
p_temp3dim(%region%,cur%spatunit%,"allarea")
=sum(%croptp%$(not sameas(%croptp%,"FORE")),p_levlunit(%region%,cur%spatunit%,%croptp%));
p_levlunit(%region%,cur%spatunit%,%croptp%)
$(p_temp3dim(%region%,cur%spatunit%,"allarea") and not sameas(%croptp%,"FORE"))
=p_levlunit(%region%,cur%spatunit%,%croptp%)*(p_levlunit(%region%,cur%spatunit%,"AREA")-
p_levlunit(%region%,cur%spatunit%,"FORE"))
/p_temp3dim(%region%,cur%spatunit%,"allarea");
```

In order to ensure consistency of total area between the prior and the posterior data sets, all data are re-mapped into their intersection. This is achieved with the fraction of the spatial units of one layer that is part of a unit of the second spatial layer:

$$a_{hg} = \sum_{h,g} \{a_h \cdot f_{hg}\} \tag{7.3}$$

a_{hg} = Area [parameter, km^2] of unit u intersecting spatial unit h and grid cell g.
a_h = Area [parameter, km^2] of spatial unit h
f_{hg} = Fraction of spatial unit h, which is covered by grid cell g

Cultivated crop areas are re-mapped to the intersecting units proportionally to the area fraction, assuming homogeneous distribution of each crop within each spatial unit h.

The LAPM predictions are not constrained to exhaust the total available area. However, this is a required characteristic in CAPRI. Therefore, the land use areas (crops, forest land and 'other' area) are scaled so that their sum matches the total available area. As forest areas are obtained from a data set, which is assumed to be of high precision, forest areas are excluded from scaling.

$$a_{c^o,hg} = \sum_{h,g} \left\{ a_{c,h} \cdot f_{hg} \cdot \frac{a_{hg} - a_{forest,hg}}{\sum_{c^{o\prime}} a_{c^{o\prime}}} \right\}, \tag{7.4}$$

$c^* =$ Land use. $c^* \in \{c, other\ land\}$
$a_{c^o,hg} =$ Area [parameter, km^2] cultivated with crop c or covered by 'other land' use excluding forest in unit u intersecting spatial unit h and grid cell g.
$a_{c^o,h} =$ Area [parameter, km^2] cultivated with crop c or covered by 'other land' use excluding forest in spatial unit h
$f_{hg} =$ Fraction of spatial unit h, which is covered by grid cell g

Note that this re-mapping is done in each step, however it affects only the step 'A priori land use distribution' which is constrained by data for the intersections of the FSS-10km grid cells with NUTS3 regions. These units are not aligned with the spatial units (HSU) for two reasons:

1. HSU are aligned with a regular grid of 0.25° x 0.25° but not to a grid of 10 km x 10 km
2. Even though HSU are aligned with a NUTS3 administrative region layer, changes in the definition of NUTS3 regions over time create shifts in the boundaries

In all other steps, the constraining data set is taken from CAPRI NUTS2 regions to which all spatial units are nested to and $f_{hr} = 1 \forall h, r$.

The same holds if disaggregation is done into the FSU units, which are part of exactly one FSS grid cell.

Dealing with FSS grid cells with too much crops

```
Line 613ff
p_temp3dim(%region%,"AREA","<0")$(p_nutslevl(%region%,"AREA") and (p_nutslevl(%region%,"OTHER")<0))
=(p_nutslevl(%region%,"CROPS")+p_nutslevl(%region%,"FORE"))/p_nutslevl(%region%,"AREA");
p_nutslevl(%region%,"AREAcorr")=p_temp3dim(%region%,"AREA","<0");
# Rescale total and crop area of units if the total area in the %region% had to be changed.
p_levlunit(%region%,cur%spatunit%,"AREA")$(p_temp3dim(%region%,"AREA","<0"))
=p_levlunit(%region%,cur%spatunit%,"AREA")*p_temp3dim(%region%,"AREA","<0");
p_levlunit(%region%,cur%spatunit%,"OTHER")$(not p_temp3dim(%region%,"AREA","<0"))
=max(0,p_levlunit(%region%,cur%spatunit%,"AREA")-
sum(%croptp%,p_levlunit(%region%,cur%spatunit%,%croptp%)));
```

Farm structure surveys collect data on crop areas and allocate them to the geographic location where the farmer resides. Therefore, it is not excluded that a spatial unit (grid cell) has more crop area than the cell is large. It is not possible to 'correct' those allocations.

CAPRI works with an area-consistent approach, thus the total area available must be exactly matching the sum of the areas used for different purposes. As a re-allocation of 'surplus' crop areas is not possible, we inflate the area of spatial units h so that all forest and crop areas can be accommodated. The area of 'other' land uses is adapted to ensure coherence.

$$a_{hg} \leftarrow a_{hg} \cdot \frac{a_{g,forest} + \sum_c a_{g,c}}{a_g} \tag{7.5}$$

$$a_{hg,other} = a_{hg} - a_{hg,forest} - \sum_c a_{hg,c} \tag{7.6}$$

Note that this 'manipulation' of the data is needed to avoid any potential infeasibilities in the land use disaggregation model maintaining the relevant information from the different data sets:

- The total crop areas data collected in the FSS
- The heterogeneity of crop areas ('suitability') as modelled with the LAPModel. This concerns both the spatial heterogeneity within a region across spatial units, as well as the relative abundance of different crops in a single spatial unit.

Adding previously unobserved crops It might well be that crops occur in a grid cell or region which were not predicted or which had not been observed 'before' (e.g. when moving from ex-post to ex-ante simulation). In this case prior estimates of the distribution need to be developed.

This is done on the basis of 'similarity' assuming that similar crops have similar preferences for natural conditions (or available infrastructure) and a similar spatial heterogeneity.

This 'gap-filling' is done in three hierarchical steps:

1. Average crop area of similar crops in the same spatial unit as defined in the set mactgroups:

```
set mactgroups(sgroups,*) "Groups with similar crops - LAPM activities" /
  CERE.(BARL,SWHE,DWHE,LMAIZ,OATS,RYEM,OCER,PARI)
  VEGE.(TOMA,OVEG,SUGB,POTA,PULS,FLOW)
  FODD.(ROOF,OFAR,LGRAS)
  OILS.(SOYA,LRAPE,SUNF)
  FORE.(FORE)
  TREE.(APPL,CITR,OFRU,NURS,NUTS,LOLIV,LVINY)/;
```

2. If there are no 'similar' crops in the region or grid cell, the average area of all available crops is used.
3. If there is still no prediction of the in the spatial unit, the same crop area is given to the prior estimates in all spatial units in the region/grid cell.

Checking availability of standard deviations May become obsolete for the CAPDIS modules of the CAPRI stable release versions following STAR 2.4

The LAPModel provides not only prediction for crop shares for each HSU, but at the same time also an estimate for the standard deviation of each estimate. These standard deviations are 'carried' on throughout the CAPRI disaggregation steps.

The procedures described above however made evident that some crops might appear in spatial units where they have not been observed before; obviously, an estimate of the standard deviation must be provided as well.

If the crops have been observed in other spatial units of the region or grid, the maximum relative standard deviation is used. If the crop has not been observed in the whole region or grid, the maximum standard deviation over all observed crops is used. Still missing standard deviations are assumed to be 100%.

Standard deviation for ‘other land uses’ are also set to a default of 100%, but can be modified by the user (through the CAPRI GUI).

Preparation for specific applications

```
Lines 690ff
p_temp3dim(%region%,"loshare",%croptp%)$ p_nutslevl(%region%,%croptp%)
= max(disagg("mincropshare"),p_temp3dim(%region%,"share",%croptp%)*disagg("relcropshare"));
p_levlunit(%region%,cur%spatunit%,%croptp%)
$((p_levlunit(%region%,cur%spatunit%,%croptp%)/sum(%croptp%1,p_levlunit(%region%,cur%spatunit%,%croptp%1))
<p_temp3dim(%region%,"loshare",%croptp%)) and not (sameas(%croptp%,"other")))
=0;
```

For some applications, the focus is on the analysis of dominant crops in each spatial unit, for example, when linking the result of the disaggregation with process-based crop models. This is done on the basis of two parameters:

$$a_{hc} = 0 \leftarrow \frac{a_{h,c}}{\sum_{c'} a_{h,c'}} < max \left[{}_{min}, \frac{a_{r,c}}{\sum_{c'} a_{r,c'}} \cdot {}_{rel} \right] \tag{7.7}$$

${}_{min}$ = Minimum crop share allowed in the spatial unit
${}_{rel}$ = Minimum relative crop share – defines heterogeneity of crop shares for a crop in a spatial unit

7.5 Downscaling of Livestock numbers

Files:

```
%curdir%/capdis/dislivestock.gms
%datdir%/capdishsu/corine2018classes.gms
%datdir%/capdishsu/s_fracGraz.gms
%datdir%/capdishsu/p_fracGraz.csv
%datdir%/capdishsu/p_grazsharesCorine.gms
%datdir%/capdishsu/p_fsuCorineArea.gdx
```

A different approach is used for the distribution of FSS livestock numbers and for the distribution of CAPRI regional data. This is because livestock require some investment/infrastructure that is not easily given up. Also, if feed is more difficult to grow due to a dry or heat spell (e.g. as observed in 2018 in several countries), feed will be purchased from the market or animal numbers.

For livestock there will also apply the principle of the ‘Primacy of stability’ already described for the distribution of crop areas.

Animal types are distributed proportionally to their shares in different animal classes:

- Grazing cattle
- Grazing sheep and goats
- Non-grazing ruminants (cattle, sheep and goats)
- Non-grazing monogastric animals (pig and poultry)

7.5.1 Distribution of FSS livestock numbers

We use shares of grazing animals from the national submissions of greenhouse gas inventories to the UNFCCC. These shares are calculated from the quantity of manure N managed in various manure management systems, and the quantity of manure N deposited on 'pasture, range and paddock' (Table 3.B(b) of the UNFCCC-Common Reporting Format, CRF, tables).

The distribution of grazing livestock from the FSS grid cells to the FSU is done using areas of various Corine classes as a proxy. This is because grazing can occur also on non-grassland, such as on commune land (outside of the CAPRI UAAR) and mixed land cover classes (agro-forestry, non-agriculural land (shrubland, etc.).

Based on expert information obtained from the EEA (Jan-Erik Petersen, personal communication), the sum of the following Corine class shares is calculated:

```
# Corine classes to be used to calculate grazing shares and areas
#
# Source: EEA (Jan-Erik Petersen) in email from 19/07/2019.
# See kipinca-CLC classes + grazing_draft_+JRC questions_rev 23-07-19.docx
# Animal types distinguished: DairyCattle, NonDairyCattle, SheepGoats
parameter p_CorineShares(*, *, *) 'Shares of CLC area used to distribute grazing animals';
table p_CorineShares
       DairyCattle  NonDairyCattle SheepGoats
all.211          25             25         10
all.223           0              0         25
all.231         100            100        100
all.242          25             25         25
all.243          50             50         50
all.244          50             50         50
all.321         100            100        100
all.322          25             25         50
all.323          25             25         50
all.324           0              0          0
all.333          25             25         25
all.411          50             50         50
all.412           0              0         25
all.421          50             50         50
;
```

For classes 322 ('Moors_and_heathland') and 324 ('Transitional_woodland-shrub') a differentiation by countries is done.

We assume that for grazing animals, the density of v_{lgr} [LU ha^{-1}] is constant within each FSS-admin grid cell and the number of livestock depends on the relevant Corine area.

$$ {}_{lgr,h} = N_{l,r} \cdot {}_{graz,lgr,r} \cdot A_{lcl,lgr,r} \tag{7.8} $$

$$ n_{lgr,h} = {}_{lgr,h} \cdot a_{lcl,lgr,h} \tag{7.9} $$

${}_{lgr,h}$ = Livestock density [parameter, LU ha^{-1}] for animals in livestock group *lgr* in spatial unit h
$N_{l,r}$ = Number of livestock [parameter, LU] of animal type l in region r
$n_{lgr,h}$ = Number of livestock [parameter, LU] of animals in livestock group *lgr* in region r
${}_{graz,lgr,r}$ = Share of grazing animals [parameter, dimensionless] of animals in livestock group *lrg* in region r
$A_{lcl,lgr,r}, a_{lcl,lgr,h}$ = Area [parameter, 1000 ha] of Corine Land Cover Classes *lcl* that are assumed to be available for grazing animals in livestock group *lgr* in spatial unit h.

7.5.2 Distribution of CAPRI livestock numbers

If livestock numbers change at the regional level if compared to the prior data, we assume that this has no influence on the spatial distribution of the animals. Instead, the livestock number in each spatial unit is multiplied with the regional relative change.

$$ n_{l,h} = \hat{n}_{lgr,h} \cdot \frac{N_{l,r}}{\hat{N}_{l,r}} \tag{7.10} $$

$n_{l,h}$ Number of livestock [parameter, LU] of animal type l in spatial unit h.
$\hat{n}_{l,h}$ Number of livestock [parameter, LU] of animal type l in spatial unit h in the prior data set
$N_{l,r}$ Number of livestock [parameter, LU] of animal type l in region r
$\hat{N}_{lhr}$ Number of livestock [parameter, LU] of animal type l in spatial unit h in the prior data set

7.6 Downscaling of Livestock numbers

Files:

```
%curdir%/capdis/dislivestock.gms
%datdir%/capdishsu/corine2018classes.gms
%datdir%/capdishsu/s_fracGraz.gms
%datdir%/capdishsu/p_fracGraz.csv
%datdir%/capdishsu/p_grazsharesCorine.gms
%datdir%/capdishsu/p_fsuCorineArea.gdx
```

A different approach is used for the distribution of FSS livestock numbers and for the distribution of CAPRI regional data. This is because livestock require some investment/infrastructure that is not easily given up. Also, if feed is more difficult to grow due to a dry or heat spell (e.g. as observed in 2018 in several countries), feed will be purchased from the market or animal numbers.

For livestock there will also apply the principle of the 'Primacy of stability' already described for the distribution of crop areas.

Animal types are distributed proportionally to their shares in different animal classes:

- Grazing cattle
- Grazing sheep and goats
- Non-grazing ruminants (cattle, sheep and goats)
- Non-grazing monogastric animals (pig and poultry)

7.6.1 Distribution of FSS livestock numbers

We use shares of grazing animals from the national submissions of greenhouse gas inventories to the UNFCCC. These shares are calculated from the quantity of manure N managed in various manure management systems, and the quantity of manure N deposited on 'pasture, range and paddock' (Table 3.B(b) of the UNFCCC-Common Reporting Format, CRF, tables).

The distribution of grazing livestock from the FSS grid cells to the FSU is done using areas of various Corine classes as a proxy. This is because grazing can occur also on non-grassland, such as on commune land (outside of the CAPRI UAAR) and mixed land cover classes (agro-forestry, non-agricululral land (shrubland, etc.).

Based on expert information obtained from the EEA (Jan-Erik Petersen, personal communication), the sum of the following Corine class shares is calculated:

```
# Corine classes to be used to calculate grazing shares and areas
#
# Source: EEA (Jan-Erik Petersen) in email from 19/07/2019.
# See kipinca-CLC classes + grazing_draft_+JRC questions_rev 23-07-19.docx
# Animal types distinguished: DairyCattle, NonDairyCattle, SheepGoats
parameter p_CorineShares(*, *, *) 'Shares of CLC area used to distribute grazing animals';
table p_CorineShares
       DairyCattle  NonDairyCattle SheepGoats
all.211          25             25          10
all.223           0              0          25
all.231         100            100         100
all.242          25             25          25
all.243          50             50          50
all.244          50             50          50
all.321         100            100         100
all.322          25             25          50
all.323          25             25          50
all.324           0              0           0
all.333          25             25          25
```

```
all.411            50               50           50
all.412             0                0           25
all.421            50               50           50
;
```

For classes 322 ('Moors_and_heathland') and 324 ('Transitional_woodland-shrub') a differentiation by countries is done.

We assume that for grazing animals, the density of v_{lgr} [LU ha^{-1}] is constant within each FSS-admin grid cell and the number of livestock depends on the relevant Corine area.

$$ _{lgr,h} = N_{l,r} \cdot {}_{graz,lgr,r} \cdot A_{lcl,lgr,r} \tag{7.11}$$

$$n_{lgr,h} = {}_{lgr,h} \cdot a_{lcl,lgr,h} \tag{7.12}$$

$_{lgr,h}$ = Livestock density [parameter, LU ha^{-1}] for animals in livestock group *lgr* in spatial unit h
$N_{l,r}$ = Number of livestock [parameter, LU] of animal type l in region r
$n_{lgr,h}$ = Number of livestock [parameter, LU] of animals in livestock group *lgr* in region r
$_{graz,lgr,r}$ = Share of grazing animals [parameter, dimensionless] of animals in livestock group *lrg* in region r
$A_{lcl,lgr,r}, a_{lcl,lgr,h}$ = Area [parameter, 1000 ha] of Corine Land Cover Classes *lcl* that are assumed to be available for grazing animals in livestock group *lgr* in spatial unit *h*.

7.6.2 Distribution of CAPRI livestock numbers

If livestock numbers change at the regional level if compared to the prior data, we assume that this has no influence on the spatial distribution of the animals. Instead, the livestock number in each spatial unit is multiplied with the regional relative change.

$$n_{l,h} = \hat{n}_{lgr,h} \cdot \frac{N_{l,r}}{\hat{N}_{l,r}} \tag{7.13}$$

$n_{l,h}$ Number of livestock [parameter, LU] of animal type l in spatial unit *h*.
$\hat{n}_{l,h}$ Number of livestock [parameter, LU] of animal type l in spatial unit h in the prior data set
$N_{l,r}$ Number of livestock [parameter, LU] of animal type l in region r
$\hat{N}_{lhr}$ Number of livestock [parameter, LU] of animal type l in spatial unit h in the prior data set

7.7 Disaggregation of Nitrogen Input

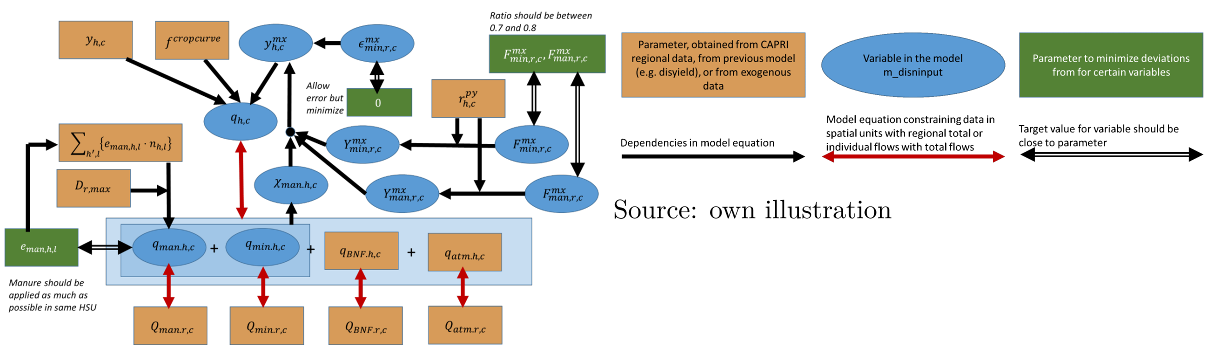
Nitrogen input to fields needs to be distributed over the spatial units in a region where the crop is cultivated. For each crop, the regional nitrogen balance is given with total N inputs per N input type, an N outputs or losses per output and loss type.

We base the distribution model on a generic crop-fertilizer-response curve. Such curves have the characteristics that are desirable for the disaggregation of nitrogen input:

1. higher N input leads to higher yields;
2. with increasing levels of N input the yield increment decreases;
3. the N 'uptake' is always less of N input;
4. saturation, i.e. attaining about 80-90% of the maximum yield at about 100-200 kg N ha^{-1} yr^{-1}

Such crop response curves 'work' with a maximum yield which are approached at high levels of N input. These maximum yield values are unknown. Figure below gives an overview of the approach indicating parameters, variables, and optimization rules.

Figure 43: Schematic overview of the Nitrogen disaggregation model



Source: own illustration

7.7.1 Crop response curve

Different crop response curves are proposed (Bodirsky and Müller, 2014; Godard et al., 2008). We base our response curve on the proposal of (Godard et al., 2008) in particular for the 'saturation' velocity[1].

Crop growth model (Godard et al., 2008)

$$Y_{r,c} = Y_{r,c}^{mx} - (Y_{r,c}^{mx} - Y_{r,c}^{mn}) \cdot exp\{-f^{cropcurve} \cdot Q_{r,c}\} \tag{7.14}$$

Crop growth model (Godard et al., 2008) without 'minimum yield'

$$Y_{r,c} = Y_{r,c}^{mx} - (1 - exp\{-f^{cropcurve} \cdot Q_{r,c}\}) \cdot @Y_{r,c}^{mn} = 0 \tag{7.15}$$

$Y_{r,c}$ = Yield [parameter, kg N ha^{-1} yr^{-1}] for crop c in region r.
$Y_{r,c}^{mx}$ = Maximum yield [parameter, kg N ha^{-1} yr^{-1}] according to the crop response curve (Godard et al., 2008) for crop c in region r.

[1]The model proposed by Bodirsky and Mueller (2014) 'saturates' only at very high N input levels > 1000 kg N ha^{-1} yr^{-1}

$Y^{mn}_{r,c}$ = Minimum yield [parameter, kg N ha^{-1} yr^{-1}] according to the crop response curve (Godard et al., 2008) for crop c in region r. This parameter is set to zero in our model.
$f^{cropcurve}$ = Scaling factor [parameter, dimensionless] used in the crop response curve (Godard et al., 2008). We use a uniform value of $f^{cropcurve} = 0.008$.
$Q_{r,c}$ = Total N input [parameter, kg N ha^{-1} yr^{-1}] for region r and crop c.

7.7.2 Crop growth scaling factor

We use a constant factor $f^{cropcurve}$ for all regions/spatial units and crops in order to not leave too many degrees of freedom. However, if infeasibilities occur, 'opening' this factor to differ between crop types could be a first test. However, the range of possible values for the crop growth scaling factor is narrow:

- For $f^{cropcurve} > 0.010$ a N uptake is larger than N input until an application rate of more than 100 kg N ha^{-1} yr^{-1}. For a value of 0.010 this is the case for an application rate of about 100 kg N ha^{-1} yr^{-1}
- For $f^{cropcurve} < 0.008$ the N input rate at which a yield of 80% of the maximum yield is attained is very high. For a value of 0.0064 this happens at Q=250 kg N ha^{-1} yr^{-1}; and for a value of 0.0054 at Q=300 kg N ha^{-1} yr^{-1}.

Therefore, only a narrow range around a value of 0.008 seems plausible.

Lower efficiency for manure application

We assume that manure is applied with less efficiency than mineral fertilizer. First, because we take into account lower nutrient availability in manure with respect to mineral fertilizer (due to reduced opportunity to target release of nutrients to crop demand, thus increasing the chance of nutrient releases in periods with enhanced risks of losses to the environment). Second, due to the fact that higher availability of manure often goes ahead with increased lack of surface where the manure can be applied in a reasonable manner.

Therefore, we assume a decrease of the NUE the higher the share of manure in the fertilizer mix.

We account for this fact by using a different crop response curve for mineral fertilizer and manure. This is realized by varying the theoretical crop curve's maximum yield.

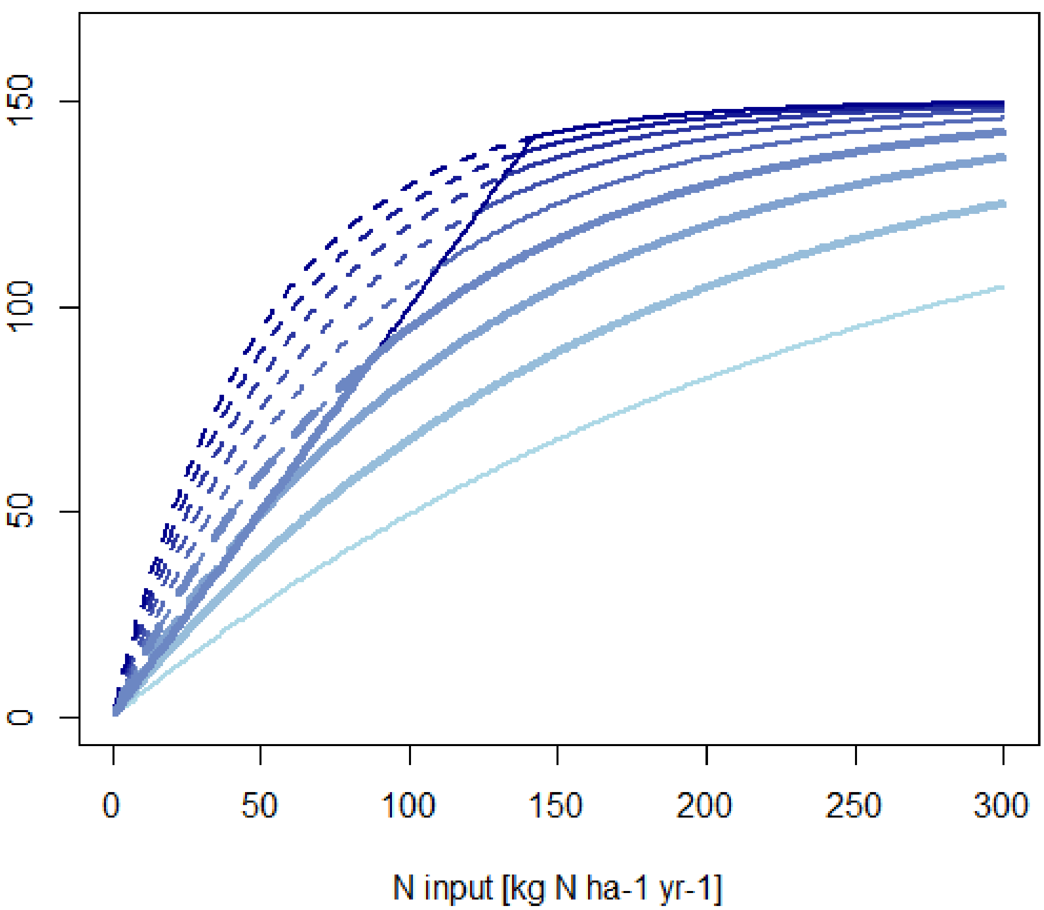
This is shown in figure below.

We introduce a dependency of y^{mx} on the share of mineral fertilizer and manure in the mix of the nitrogen source.

$$\begin{aligned} y^{mx}_{h,c} &= y^{mx}_{man,r,c} + y_{man,h,c} \cdot (y^{mx}_{min,r,c} - y^{mx}_{man,r,c}) \\ {}_{man,h,c} &= \frac{q_{man,h,c}}{q_{man,h,c} + q_{min,h,c}} \end{aligned} \tag{7.16}$$

$y^{mx}_{h,c}$ = Maximum yield [variable, kg N ha^{-1} yr^{-1}] according to the crop response curve (Godard et al., 2008) for crop c in the spatial unit h.
$y^{mx}_{man,r,c}$ = Maximum yield for manure [parameter, kg N ha^{-1} yr^{-1}] according to the crop response curve (Godard et al., 2008) for crop c in spatial unit h.

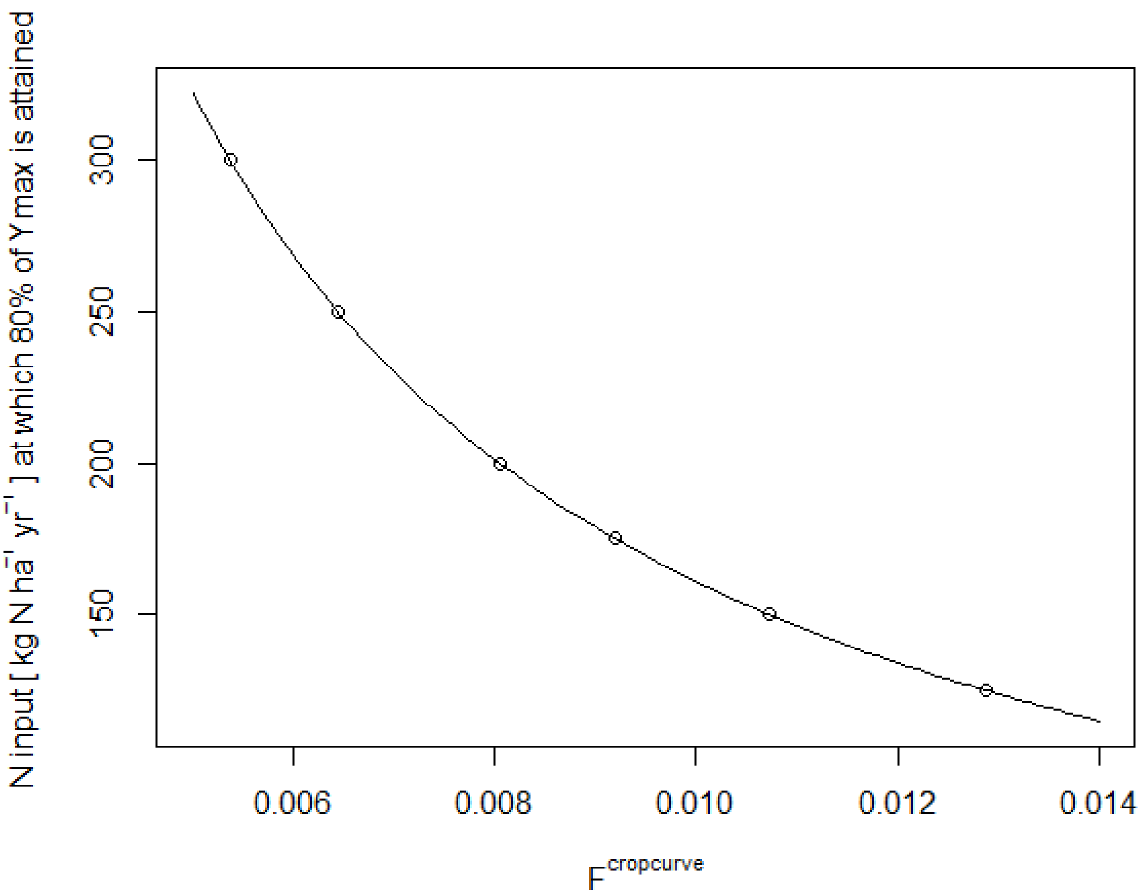
Figure 44: Crop growth curves according to Godard et al. (2008) for different crop growth scaling factors.



Parameters from light blue to dark blue curves: $0.004 \leq f\{cropcurve\} \leq 0.020$. The curves with $f\{cropcurve\} \in \{0.006, 0.008, 0.010\}$ are plotted in bold. Other parameter: Y^{mx}=150; Y^{mn}=0.

$y^{mx}_{min,r,c}$ = Maximum yield for mineral fertilizer [parameter, kg N ha^{-1} yr^{-1}] according to the crop response curve (Godard et al., 2008) for crop c in spatial unit h.
$_{man,h,c}$ = Share of manure [variable, dimensionless] in the application of nitrogen from manure and

Figure 45: N input rates that give a yield of 80% of the maximum yield for different crop growth scaling factors according to Godard et al. (2008)

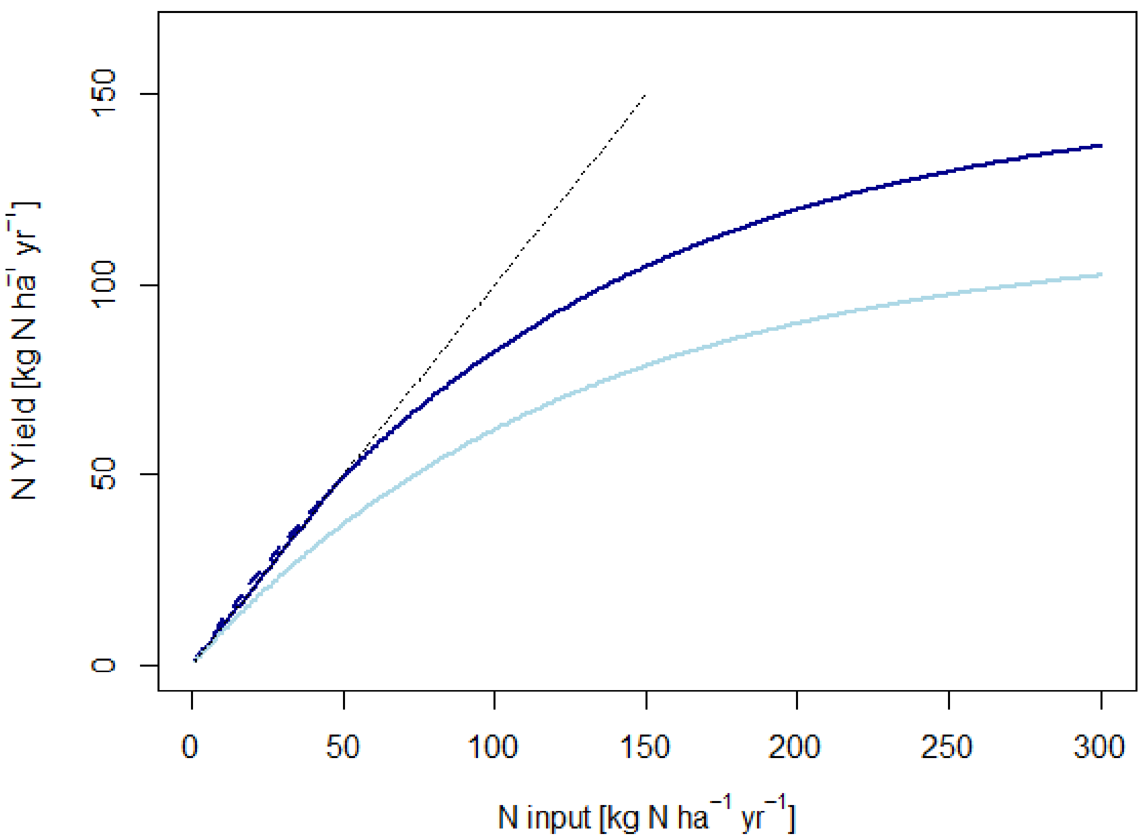


mineral fertilizer.

Manure availability Manure can be traded between individual spatial units. Manure trade between regions (or even countries) is covered by the regional model of CAPRI and does not need to be considered here.

The availability of manure is obtained therefore from each spatial unit plus neighboring spatial units within the same region. The range of spatial units from which manure can be used is assumed to a region-specific variable and

Figure 46: Examples of crop response curves according to Godard et al. (2008).



Parameters used: f=0.008. Light blue curve for manure: y^{mx}=130; dark blue curve for mineral fertilizer: y^{mx}=150.

$$\sum_{c}\{q_{man,h,c} \cdot a_{h,c}\} \leq \sum_{h',l}\{e_{man,h,l} \cdot n_{h,l}\}$$
$$d_{h,h'} \leq D_r^{mx} \tag{7.17}$$

$q_{man,h,c}$ = Manure application rate [variable, kg N/ha] to crop c in spatial unit h
$a_{h,c}$ = Area [parameter, 1000 ha] cultivated with crop c

$e_{man,h,l}$ = Manure excretion [parameter, kg N/head] by animal species l in spatial unit h– net of losses in livestock housing and manure storage and management systems. No heterogeneity is assumed for nitrogen excretion rate within one NUTS2 region.
$n_{h,l}$ = Livestock number [parameter, 1000 heads]
$d_{h,h'}$ = Distance [parameter, km] between spatial unit h and spatial unit $h^{\acute{}}$
D_r^{mx} = Maximum distance [variable, km] for which transport of manure is allowed in region r.

Obviously, the total manure available for application must be exhausted:

$$\sum_{h,c}\{q_{man,h,c} \cdot a_{h,c}\} = \sum_{l}\{E_{man,r,l} \cdot N_{h,l}\} \tag{7.18}$$

7.7.3 Fertilization distribution model

Recover regional N flows For each flow of nitrogen and crop, the sum of flows over all spatial units must recover the total flow at regional level for each crop.

This holds both for input flows and output flows (i.e. harvest, surplus).

Potential yield The maximum (potential) yield is proportional to the relative potential yield (without water limitation).

$$y_{h,c}^{mx} = F_{r,c}^{ymx} \cdot r_{h,c}^{py} \tag{7.19}$$

$y_{h,c}^{mx}$ = Maximum yield [variable, kg/ha] determining the shape of the crop growth curve in each spatial unit for each crop.
$r_{h,c}^{py}$ = Relative potential yield [parameter, dimensionless] of crop c in spatial unit h.
$F_{r,c}^{ymx}$ = Scaling factor [variable, kg/ha] adjusting the relative potential yield so that it gives the maximum yield in the crop growth curve for each spatial unit h and crop c.

Crop growth curve Total input of nitrogen is obtained from the observed yield for the crop in the spatial unit (parameter, calculated in the yield and irrigation module) and the maximum yield obtainable in the crop in the spatial unit (variable).

$$q_{h,c} = -\frac{1}{x} \cdot ln\left\{\frac{y_{h,c}}{y_{h,c}^{mx}}\right\} \tag{7.20}$$

Nitrogen source Once the total N input per crop and spatial unit is determined, the individual N sources need to be calculated. We have:

- Biological N fixation: this is directly calculated from the crop type and yield and is 'fixed'
- Atmospheric deposition: this is obtained from external data and cannot be modified

- Mineralization of soil organic matter. We have no data yet for calculating mineralization of soil organic matter at the regional level, thus it is not possible to include this term in the disaggregation. If there were data on soil organic mineralization, the following assumptions would need be taken:
  - Mineralization of soil organic matter occurs in extensive fields, thus at low application rates of mineral fertilizer and irrigation rates
  - Manure is able to replenish soil organic matter. It is thus unlikely that mineralization of soil organic matter occurs where manure is applied or deposited by grazing animals.

7.7.4 Data preparation

Collecting information At Nuts2 level, y and f are known and $\mathrm{y_m}$ can be calculated

$$y^{mx}_{r,c} = \frac{y_{r,c}}{1 - exp\{-f^{cropcurve} \cdot Q_{r,c}\}} \tag{7.21}$$

For each spatial unit, the yield is given from the distribution of irrigation shares and yield.

We can assume that the potential yield $\mathrm{y_m}$ follows the pattern of the irrigated yield obtained from PESETA.

$$y^{mx}_{h,c} \propto r^{py}_{h,c} \tag{7.22}$$

$r^{py}_{h,c}$ = Relative potential yield [parameter, dimensionless] of crop c in spatial unit h.

Calculation of relative potential yield per spatial unit

$$r^{py}_{h,c} = y^{py}_{h,c} / \overline{y^{py}_{h,c}} \tag{7.23}$$

$r^{py}_{h,c}$ = Relative potential yield [parameter, dimensionless] of crop c in spatial unit h.
$\overline{y^{py}_{h,c}}$ = Average potential yield [parameter, kg/ha] of crop c in region r.

$$\overline{y^{py}_{h,c}} = \frac{\sum_h \{y^{py}_{h,c} \cdot a_{h,c}\}}{\sum_h \{a_{h,c}\}} \tag{7.24}$$

Calculation of distances between HSUs Update pending

Calculation of manure availability Excretion net of all volatilization must be back-calculated so that emissions from applications are not subtracted.

Update pending

Consideration of mitigation options Update pending

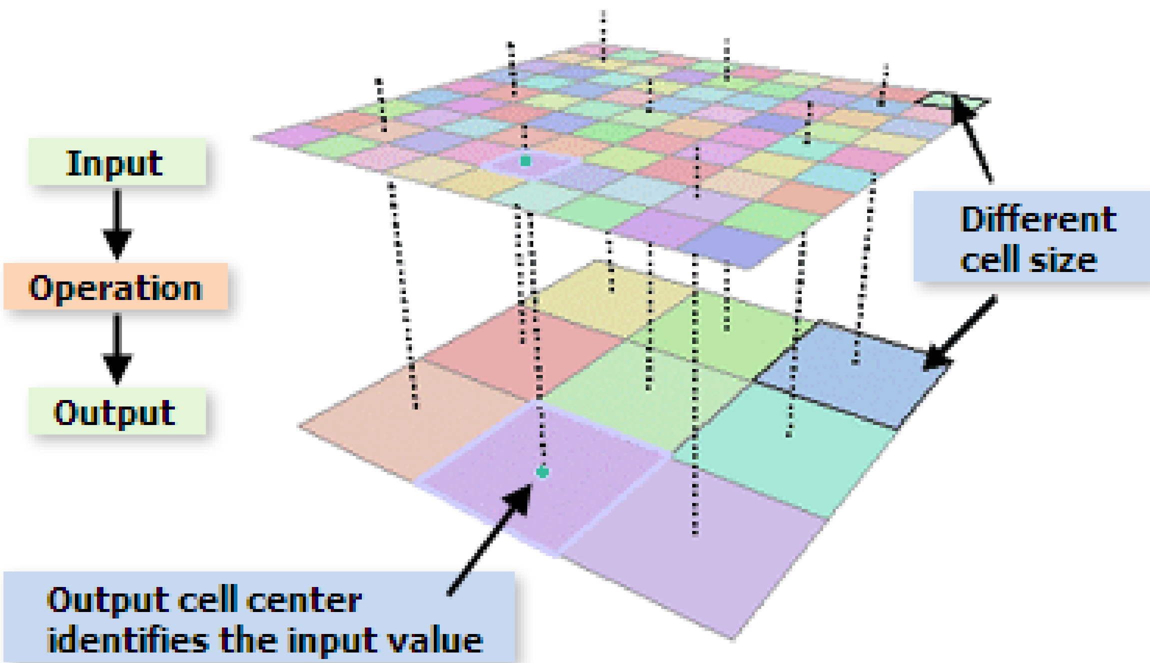
7.8 Data Distribution Procedures

The terms spatial “resampling”, “re-mapping, “downscaling”, “disaggregation” and “distribution” are frequently used in a synonymous way. Admittedly, a very sharp distinction is sometimes difficult and there are overlaps in their meaning, this holds especially for “downscaling” and “disaggregation”. To avoid confusion of the reader we lay down our interpretation of these terms.

7.8.1 Resampling

By resampling we intend the process of interpolating from one grid resolution to a different grid resolution. Quantitative evaluation of data contained on different grids requires resampling to a common grid. There are many resampling methods. Classic interpolation methods include: bilinear, nearest neighbor, inverse distance. The consistency with the original dataset is not necessarily maintained.

Example: Resampling of land use information at a regular grid of 100 x 100m from remote sensing to a 1km x 1km grid using the nearest neighbor method. The 1 x 1km grid receives the value of the 100 x 100 m grid which spatially coincides with the center of the 1 x 1km grid.



Sketch of the nearest neighbor interpolation method applied to a regular grid. (Source: ESRI ArcGIS documentation of resampling methods)

7.8.2 Re-mapping

In the context of this work we use the term re-mapping if the spatial reference unit of a given parameter has to be changed. The aim is to keep changes of the parameters/values at a minimum during the change of the spatial reference. Re-mapping is often required for pre- and post processing of data input/output of the downscaling/disaggregation/distribution procedures.

In the special case of nested data (i.e. a spatial unit at low hierarchical level is member of exactly one unit at a higher hierarchical level s. definition below) the re-mapping of values given at low hierarchical level to a higher hierarchical level can be obtained by simple aggregation (summing up or averaging). A typical example is the re-mapping of information given for administrative regions (NUTS3 - > NUTS2 -> NUTS1 -> NUTS0).

Definition of nested spatial reference data sets

$$if\ 0 < a_{ij} \leq 1\ for\ i \in I\ and\ j \in J\ then\ a_{ik} = 0\ for\ i \in I\ and\ k \neq j\ and\ k \in J \tag{7.25}$$

More frequently data is stemming from various sources having different spatial reference units. For example meteo data which usually comes at a grid level (e.g. 50 x 50 km) and has to be re-mapped to an administrative unit. In these cases a spatial overlay of the data is performed and new spatial units are created at the intersection of the spatial units. In the meteo grid/administrative region case a meteo grid might be split by the border between 2 different administrative units.

To avoid creating very small spatial units during the re-mapping procedure we defined a minimum spatial unit of 1 x 1 km2 (i.e. the 1 by 1km EEA reference grid) as common denominator. All input data is rasterized or resampled to this unit before the re-mapping. For categorical data as e.g. land use/cover classes, the nearest neighbor interpolation method is applied. In other cases the share of the parameter in the 1 x 1 km^2 grid is calculated (e.g forest area). After rasterizing/resampling of all data sets of interest, re-mapping of all parameters to any spatial reference unit present in the input data sets is possible.

Example: meteo data at 50 x 50 km grid level re-mapped to NUTS2 regions

Consistency of data between the rasterized versions of both spatial references is maintained:

$$x_i = \frac{\sum_j x_j \cdot a_{ji}}{a_i} \tag{7.26}$$

I = units of the first layer (e.g. 50 km x 50 km grid)
J = units of the second layer (e.g. NUTS2 regions)
x = Area-based variable (e.g. average annual rainfall mm/m^2; kg N/ha emissions; persons/ha; etc.)
a = Area. a_i is the area of a unit of the layer the variable is re-mapped to. a_{ij} are the areas of the intersections between unit i and all units j that have common surface.

In case that $\sum_j a_{ij} \leq a_i$, that is there is a part of unit i which is not covered, assumptions on 'gap-filling' have to be made. Possible options are:

1. Assuming same area-based variable thus giving higher total quantity
2. Maintaining total quantity thus 'diluting' the area-based variable.

7.8.3 Increasing the spatial resolution of data for nested spatial references

We differentiate approaches for increasing the spatial resolution of data that are applicable to nested spatial data sets in view of the complexity of the approach. The complexity increases from simple distribution over downscaling to disaggregation.

Downscaling As downscaling we understand a procedure to infer high-resolution information from low-resolution variables using *simple proxy information* that is available at the high resolution. Downscaling works only with nested spatial units. The consistency with the original dataset is maintained.

For example:

- downscaling population density in rural areas available at country level to a grid taking into account land cover information (rural areas/urban areas)
- downscaling of fertilizer input from national fertilizer use statistics based on distribution of crop yields available at higher spatial level (e.g. sub-national regions).

Disaggregation Spatial disaggregation is the process by which information at a coarse spatial scale is translated to finer scales using weighting. The weights are based on more or less complex regression or other (optimization) models derived from observations, ancillary data, or previous downscaling/disaggregation steps. The consistency with the original dataset is maintained. One example is the use of LUCAS land use observations, environmental and management parameters to predict the probability of a crop to be cultivated at a certain location.

Simple Distribution This procedure is applied if a parameter is available at high hierarchical level (e.g. country) and no information is available to enhance the spatial pattern for the lower hierarchical level (by proxies/regression/models) In this case, the spatial distribution of a parameter when changing the spatial reference unit from a higher to a lower hierarchical level is kept constant. Example: average nitrogen excretion rates for different livestock available at country level. A homogeneous distribution within all sub-units is assumed. Due to lacking information the effects of sub-national differences due to e.g. specific feeding strategies, are not taken into account.

Chapter 8

Stability testing tools for model tasks

During the Stable Release project, tasks were added to the CAPRI system that are aiding in testing the coherence and stability of the entire system. Those tasks are accessible from the GUI in the work step “Tests”. Those tasks help developers running the model with different settings or with different starting values and systematically comparing the results.

The central task is a somewhat refined version of the GDXDIFF facility built into GAMS. It is called “Compare task results”. It loads the results from two executions of the same task, e.g. two simulations with somewhat different settings, and compares all the results of the main result parameter. Differences are put in relation to the absolute number in the first file, and are then sorted into classes of a cumulative distribution, counting the number of differences in each class. The results are stored in the sub directory “test” in the standard result directory, with the name pattern “stability_”<task name><additional result type identifier>.gdx. It is advisable to use the Additional result type identifier (on the debug options tab) to give a hint to the nature of the experiment (e.g. “_restart”).

For instance, running the task “Build regional database” twice, first without starting value, then with the results of the first run as starting values, resulted in the differences shown in Table 34.

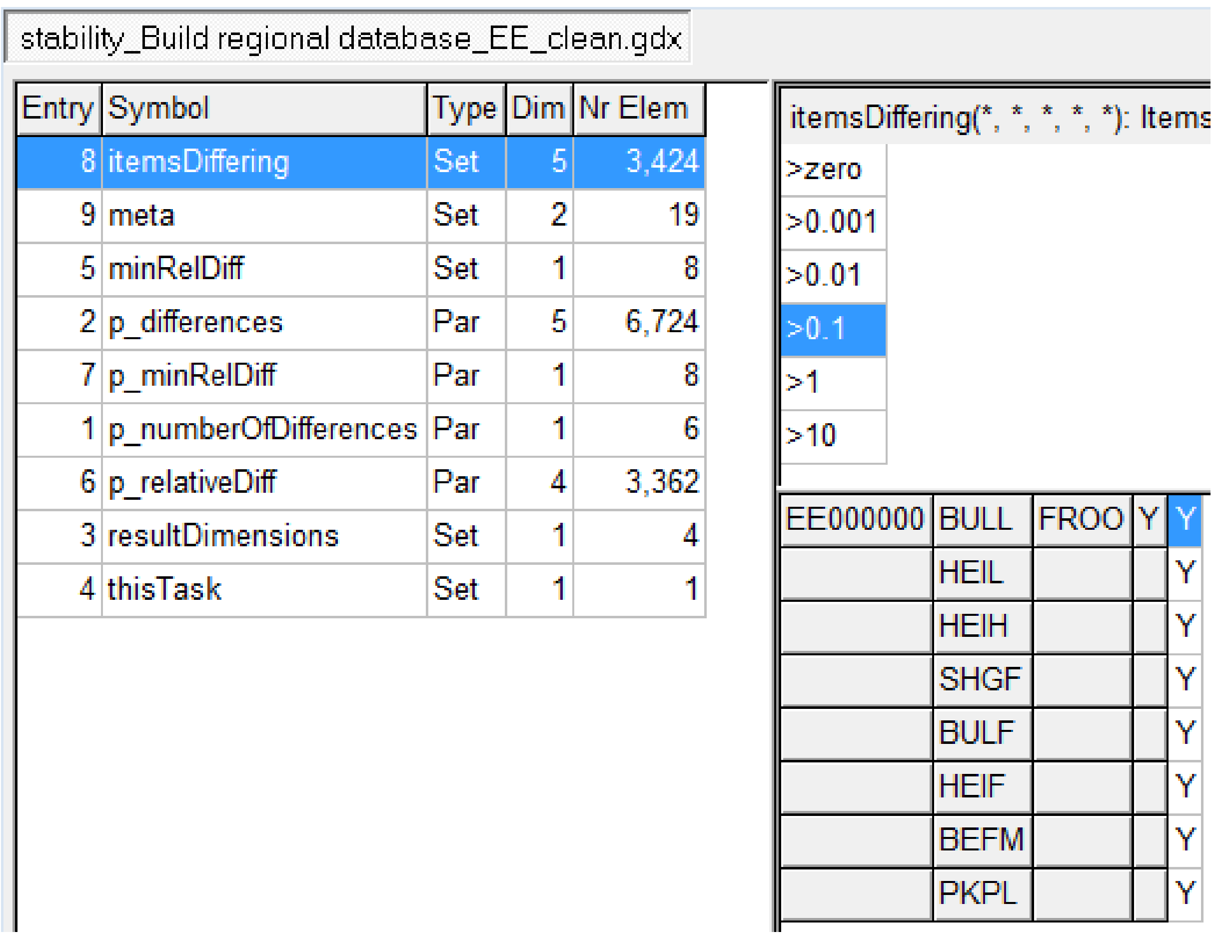
Table 34: Starting value sensitivity for Estonia in the task Build regional database

Class	Number of deviations
>zero	3362
>0.001	34
>0.01	16
>0.1	8
>1	3
>10	1
>100	0
infinite	0

The task also makes a GAMS set mapping linking the items that differ to each of the classes. If the developer would like to investigate particular differences further, (s) he needs to go into the full result file stored in the subdirectory "test" in the main result directory. Doing that for this particular run reveals that all the differences of 10% or more are due to the distribution of FROO, Fodder Root Crops, to various animals (see 0).

If a more comprehensive suite of test should be carried out, there will be a great number of result files. The task "Merge comparison results" can be used to collect the cumulative distribution of differences from many comparison files and build a summary. In order to do so, the merging task needs to know which result files to merge. In order to add the results of a comparison to the list of files to merge, check the box "Add comparison output to cumulative list of comparisons" and make sure you enter a sensible name for the list of comparisons (all on the general settings tab in the GUI).

Figure 47: Items differing for Estonia when restarting Build regional database



Source: CAPRI Modelling System

If many experiments are executed using the batch execution facility, then it is useful to keep result files apart by using different directories for results, restart points and result input. There are two tasks in the "Tests" workstep that are intended to help there:

- Create experiment folder structure: creates an empty tree of result folders, with one sub directory for each "experiment".
- Collect experiment results: copies selected results from the experiment directory tree and stores them in the standard capri directory with a file suffix "_"<experiment>, as if the user had run the task with the property "Additional result type identifier" set to "_"<experiment> but with additional control over starting values and data input.

Chapter 9

References

Adenaeuer, M.; Louhichi, K.; de Frahan, B.H., Witzke, H.P. (2004). Impacts of the 'Everything but Arms' initiative on the EU sugar sub-sector. Selected paper prepared for presentation at the International Conference on Policy Modelling (EcoMod2004), June 30 - July 2, Paris

Ahmed, S.A., Hertel, T., Lubowski, R. (2009). Calibration of a land cover supply function using transition probabilities. GTAP Research Memorandum No. 14 October 2008.

Anselin, L., Florax, R. J. G. M. , and Rey, S. J.: Advances in Spatial Econometrics, Berlin: Springer Verlag, 2004.

Armington, P. S., 1969. A Theory of Demand for Products Distinguished by Place of Production. IMF Staff Papers 16, 159-178.EN.REFLISTWolf, W., 1995. 'SPEL system, Methodological documentation (Rev. 1), Vol. 1: Basics, BS, SFSS', Eurostat Luxembourg.

Blanco M., Ramos F., Van Doorslaer B., Martínez P., Fumagalli D., Ceglar A., Fernández F.J. (2017b). Climate change impacts on EU agriculture: A regionalized perspective taking into account market-driven adjustments. Agricultural Systems 156: 52–66, doi: `https://doi.org/10.1016/j.agsy.2017.05.013`

Britz W., Heckelei T. (1999): Calibration of Feed Requirements and Price determination of feed in CAPRI, CAPRI working paper 99-06, available online: `http://www.ilr.uni-bonn.de/agpo/publ/workpap/Pap99-06.pdf`

Britz, W., Pérez, I. and Wieck, C. (2003). Mid-Term review Proposal Impact Anaysis woth the CAPRI modelling System, in European Commission, Directorate-General for Agriculture Mid-Term Review of the Common Agricultural Policy. Brussels: pp.111-140.

Britz, W. and R. Delzeit (2013). "The impact of German biogas production on European and global agricultural markets, land use and the environment." Energy Policy 62(0): 1268-1275.

Britz W., Witzke H.P. (2014). CAPRI Model Documentation 2014. University of Bonn. Available online: `http://www.capri-model.org/docs/capri_documentation.pdf`

Bodirsky, B.L., Müller, C., 2014. Robust relationship between yields and nitrogen inputs indicates three ways to reduce nitrogen pollution. Environ. Res. Lett. 9, 0–4. `https://doi.org/10.1088/1748-9326/`

9/11/111005

Delincé J, Ciaian P, Witzke P (2015): Economic impacts of climate change on agriculture: the AgMIP approach, Journal of Applied Remote Sensing, Vol. 9, doi: 10.1117/1.JRS.9.097099.

European Commission (2017). EU Agricultural Outlook for the Agricultural Markets and Income 2017–2030; European Commission, DG Agriculture and Rural Development: Brussels, Belgium.

European Commission (2018). Impact Assessment Accompanying the proposals on the CAP post 2020. European Commission, Brussels, SWD (2018) 301 final.

Fellmann, T., Witzke, P., Weiss, F., Van Doorslaer, B., Drabik, D., Huck, I., Salputra, G.,Jansson, T., Leip, A., (2018). Major challenges of integrating agriculture into climatechange mitigation policy frameworks. Mitig. Adapt. Strat. Glob. Change 23, 451–468 doi: http://dx.doi.org/10.1007/s11027-017-9743-2

Gocht A., Britz W., Ciaian P., Gomez y Paloma S. (2013). Farm Type Effects of an EU-wide Direct Payment Harmonisation, Journal of Agricultural Economics 64(1): 1-32.

Gocht A., Espinosa M., Leip A., Lugato E., Schroeder L.A., Van Doorslaer B., Gomez y Paloma S. (2016). A grassland strategy for farming systems in Europe to mitigate GHG emissions—An integrated spatially differentiated modelling approach, Land Use Policy 58: 318–34. doi: https://doi.org/10.1016/j.landusepol.2016.07.024

Gocht, A., Ciaian, P., Bielza, M., Terres, J. M., Röder, N., Himics, M., & Salputra, G. (2017). EU-wide economic and environmental impacts of CAP greening with high spatial and farm-type detail. Journal of Agricultural Economics, 68(3), 651-681, doi: https://doi.org/10.1111/1477-9552.12217

Godard, C., Roger-Estrade, J., Jayet, P.A., Brisson, N., Le Bas, C. (2008). Use of available information at a European level to construct crop nitrogen response curves for the regions of the EU. Agric. Syst. 97, 68-82.

Himics, M., Fellmann, T., Barreiro-Hurlé, J., Witzke, H. P., Domínguez, I. P., Jansson, T., & Weiss, F. (2018). Does the current trade liberalization agenda contribute to greenhouse gas emission mitigation in agriculture? Food policy, 76, 120-129, doi: https://doi.org/10.1016/j.foodpol.2018.01.011

IPCC (2000) Good Practice Guidance and Uncertainty Management in National Greenhouse gas Inventories, Japan.

Jansson T. (2007). Econometric specification of constrained optimization models. PhD Dissertation. http://hss.ulb.uni-bonn.de/2007/1157/1157.htm

Jansson, T. and T. Heckelei (2011). Estimating a primal model of regional crop supply in the European Union. Journal of Agricultural Economics 62(1): 137-152.

Keeney R., Hertel, T.W. (2008). Yield Response to Prices: Implications for Policy Modeling. GTAP Working Paper 08 – 13, doi: https://doi.org/10.22004/ag.econ.45969

Kempen, M., Witzke. P., Pérez-Dominguez. I., Jansson, T. and Sckokai, P. (2011). Economic and environmental impacts of milk quota reform in Europe, Journal of Policy Modeling, 33(1), pp 29-52.

Leip A., Weiss F., Lesschen J. P., Westhoek, H. (2014). The nitrogen footprint of food products in the European Union. The Journal of Agricultural Science 152(S1): 20-33.

Leip A., Billen G., Garnier J., Grizzetti B., Lassaletta L., Reis S., Simpson D., Sutton M.A, de Vries W., Weiss F., Westhoek H. (2015). Impacts of European livestock production: nitrogen, sulphur, phosphorus and greenhouse gas emissions, land-use, water eutrophication and biodiversity. Environmental Research Letters 10: 115004.

Panagos, P., P. Borrelli, K. Meusburger, C. Alewell, E. Lugato, L. Montanarella (2015). Estimating the soil erosion cover-management factor at the European scale. Land Use Policy 48: 38-50.

Pérez Dominguez I. (2005). Greenhouse Gases: Inventories, Abatement Costs and Markets for Emission Permits in European Agriculture - A Modelling Approach, Doctoral Thesis, 165 pp.

Pérez Dominguez I. Fellmann, T. (2018): PESETA III: Agro-economic analysis of climate change impacts in Europe, JRC Technical Reports, EUR 29431 EN, Publications Office of the European Union, Luxembourg, doi: https://doi.org/10.2760/179780

Renwick A., Jansson T., Verburg P. H., Revoredo-Giha C., Britz W., Gocht A., McCracken D. (2013). Policy reform and agricultural land abandonment in the EU. Land use policy 30(1): 446-457.

Schroeder L.A., Gocht A., Britz W. (2015). The impact of Pillar II Funding: validation of a modelling and evaluation perspective, Journal of Agricultural Economics 66(2): 415-441.

Undersander, D., Mertens, D. and Thiex, N. (1993) Forage Analyses Procedures, National Forage Testing Association (NFTA), Omaha.

Witzke, P., Adenäuer, M., Britz, W., Heckelei, T., (2005). Modelling EU sugar market scenarios with a modified Armington appoach, in: IATRC Symposium. Presented at the Pressures for Agricultural Reform, WTO Panels and the Dona Round Negotiations, Sevilla, Spain.

Wolf, W., 1995. 'SPEL system, Methodological documentation (Rev. 1), Vol. 1: Basics, BS, SFSS', Eurostat Luxembourg.

CAPRI working papers: http://www.ilr.uni-bonn.de/agpo/rsrch/capri/wrkpap_e.htm

Chapter 10

Annex code lists

Table 1: Codes used for storing the original REGIO tables in the data base and their description, rows

Codes used in CAPRI's REGIO tables	Original REGIO description
TOTL	Territorial area
FORE	Forest land
AGRI	Utilized agricultural area
GARD	Private gardens
GRAS	Permanent grassland
PERM	Permanent crops
VINE	Vineyards
OLIV	Olive plantations
ARAB	Arable land
GREF	Green fodder on arable land
CERE	Cereals (including rice)
WHEA	Soft and durum wheat and spelt
BARL	Barley
MAIZ	Grain maize
RICE	Rice
POTA	Potatoes
SUGA	Sugar beet
OILS	Oilseeds (total)
RAPE	Rape
SUNF	Sunflower
TOBA	Tobacco
MAIF	Fodder maize
CATT	Cattle (total)
COWT	Cows (total)
DCOW	Dairy cows
CALV	Other cows
CAT1	Total cattle under one year
CALF	Slaughter calves
CABM	Male breeding calves (<1 year)
CABF	Female breeding calves (<1 year)
BUL2	Male cattle (1-2 years)
H2SL	Slaughter heifers (1-2 years)
H2BR	Female cattle (1-2 years)
BUL3	Male cattle (2 years and above)
H3SL	Slaughter heifers (2 years and above)

Table 2: Codes used for storing the original REGIO tables in the data base and their description, columns

Codes used in CAPRI's REGIO tables	Original REGIO description
LEVL	Herd size / Area / # of persons
LSUN	Live stock units
PROP	Physical production
YILD	Yield
VALE	EAA position in ECU
VALN	EAA position in NC

Table 3: Connection between CAPRI and REGIO crop areas, crop production and herd sizes

SPEL-code	REGIO-code	REGIO-code	REGIO-code	REGIO-code	Description of SPEL activity
SWHE	WHEA	CERE	ARAB		Soft wheat
DWHE	WHEA	CERE	ARAB		Durum wheat
RYE		CERE	ARAB		Rye
BARL	BARL	CERE	ARAB		Barley
OATS		CERE	ARAB		Oats
MAIZ	MAIZ	CERE	ARAB		Maize
OCER		CERE	ARAB		Other cereals (excl. rice)
PARI	RICE	CERE	ARAB		Paddy rice
PULS			ARAB		Pulses
POTA	POTA		ARAB		Potatoes
SUGB	SUGA		ARAB		Sugar beet
RAPE	RAPE	OILS	ARAB		Rape and turnip rape
SUNF	SUNF	OILS	ARAB		Sunflower seed
SOYA		OILS	ARAB		Soya beans
OLIV		OLIV	PERM		Olives for oil
OOIL		OILS	ARAB		Other oil seeds and oleaginous fruits
FLAX			ARAB		Flax and hemp (faser)
TOBA	TOBA		ARAB		Tobacco, unmanufactured, incl. dried
OIND			ARAB		Other industrial crops
CAUL			ARAB		Cauliflowers
TOMA			ARAB		Tomatoes
OVEG			ARAB		Other vegetables
APPL			PERM		Apples, pears and peaches
OFRU			PERM		Other fresh fruits
CITR			PERM		Citrus fruits
TAGR		VINE	PERM		Table grapes
TABO		OLIV	PERM		Table olives
TWIN		VINE	PERM		Table wine
OWIN		VINE	PERM		Other wine
NURS			PERM		Nursery plants
FLOW			ARAB		Flowers,ornamental plants, etc.
OCRO			ARAB		Other final crop products
MILK	DCOW				Dairy cows
BEEF	BUL2	BUL3			Bulls fattening
CALF	CALF				Calves fattening (old VEAL)
PORK	PIG3	PIG2	PIG1		Pig fattening
MUTM	GOAT	HEP			Ewes and goats
MUTT	GOAT	SHEP			Sheep and goat fattening
EGGS	POUL				Laying hens
POUL	POUL				Poultry fattening
OANI					Other animals
OROO			ARAB		Other root crops
GRAS	GRAS				Green fodder
SILA	GREF		ARAB		Silage
CALV	CALV				Suckler cows
RCAL	CABM	CABF			Calves, raising
HEIF	H2SL	H2BR	H3SL	H3BR	Heifers
PIGL	SOW2				Pig breeding
FALL			FALL		Fallow land

Table 4: List of activities in the supply model

Group	Activity	Code
Cereals	Soft wheat	SWHE
	Durum wheat	DWHE
	Rye and Meslin	RYEM
	Barley	BARL
	Oats	OATS
	Paddy rice	PARI
	Maize	MAIZ
	Other cereals	OCER
Oilseeds	Rape	RAPE
	Sunflower	SUNF
	Soya	SOYA
	Olives for oil	OLIV
	Other oilseeds	OOIL
Other annual crops	Pulses	PULS
	Potatoes	POTA
	Sugar beet	SUGB
	Flax and hemp	TEXT
	Tobacco	TOBA
	Other industrial crops	OIND
Vegetables & Fruits & Other perennials	Tomatoes	TOMA
	Other vegetables	OVEG
	Apples, pear & peaches	APPL
	Citrus fruits	CITR
	Other fruits	OFRU
	Table grapes	TAGR
	Table olives	TABO
	Table wine	TWIN
	Nurseries	NURS
	Flowers	FLOW
	Other marketable crops	OCRO
Fodder production	Fodder maize	MAIF
	Fodder root crops	ROOF
	Other fodder on arable land	OFAR
	Graze and grazing	GRAS
Fallow land and set-aside	Set-aside idling	SETA
	Non food production on set-aside	NONF
	Fallow land	FALL
Cattle	Dairy cows	DCOW
	Sucker cows	SCOW
	Male adult cattle fattening	BULF
	Heifers fattening	HEIR
	Heifers raising	HEIF
	Fattening of male calves	CAMF
	Fattening of female calves	CAFF
	Raising of male calves	CAMR
	Raising of female calves	CAFR
Pigs, poultry and other animals	Pig fattening	PIGF
	Pig breeding	SOWS
	Poultry fattening	POUF
	Laying hens	HENS
	Sheep and goat fattening	SHGF
	Sheep and goat for milk	SHGM

Table 5: Output, inputs, income indicators, policy variables and processed products in the data base

Group	Item
Outputs	
Cereals	Soft wheat
	Durum wheat
	Rye and Meslin
	Barley
	Oats
	Paddy rice
	Maize
	Other cereals
Oilseeds	Rape
	Sunflower
	Soya
	Olives for oil
	Other oilseeds
Other annual crops	Pulses
	Potatoes
	Sugar beet
	Flax and hemp
	Tobacco
	Other industrial crops
Vegetables & Fruits & Other perennials	Tomatoes
	Other vegetables
	Apples, pear & peaches
	Citrus fruits
	Other fruits
	Table grapes
	Table olives
	Table wine
	Nurseries
	Flowers
	Other marketable crops
Fodder	Gras
	Fodder maize
	Other fodder from arable land
	Fodder root crops
	Straw
Marketable products from animal product	Milk from cows
	Beef
	Pork meat
	Sheep and goat meat
	Sheep and goat milk
	Poultry meat
	Other marketable animal products
Intermediate products from animal production	Milk from cows for feeding
	Milk from sheep and goat cows for feeding
	Young cows
	Young bulls
	Young heifers
	Young male calves
	Young female calves
	Piglets

Table 6: Codes of the input allocation estimation

The set of FADN inputs (FI)	
TOIN	total inputs
COSA	animal specific inputs
FEDG	self grown feedings
ANIO	other animal inputs
FEDP	purchased feedings
COSC	crop specific inputs
SEED	seeds
PLAP	plant protection
FERT	fertilisers
TOIX	other inputs (overheads)
The set of CAPRI inputs (CI) used in the reconcilation	
TOIN	total inputs
FEED	feedings
IPHA	other animal inputs
COSC	crop specific inputs
SEED	seeds
PLAP	plant protection
FERT	fertilisers
REPA	repairs
ENER	energy
SERI	agricultural services input
INPO	other inputs

1. The set of 'Other' activities that had been omitted from the econometric estimation:
OTHER={OCER, OFRU, OVEG, OCRO, OWIN, OIND, OOIL, OFAR, OANI}
2. The set of activity groups, and their elements, used in the replacement or missing/negative coefficients
'GROUPS'= {YOUNG, VEGE, SETT, PULS, PIG, OILS, MILK, MEAT, INDS, HORSE, GOAT, FRU, FOD, FLOWER, DENNY, COW,
CHICK1, CHICK2, CHICK3, CERE, ARAB}
YOUNG={YBUL, YCOW},
VEGE={TOMA},
SETT={SETA, NONF, FALL, GRAS},
PULS=PULS
PIG={PIGF, SOWS},
OILS={RAPE, SOYA, SUNF, PARI, OLIV},
INDS={TOBA, TEXT, TABO},
GOAT={SHGM, SHGF},
FRU={APPL, CITR, TAGR, TWIN},
FOD={ROOF, MAIF},
FLOWER={FLOW, NURS},
DENNY={PORK, SOWS},
COW={DCOW, SCOW, HEIF, HEIR, CAMF, CAFF, BULF, CAMR, CAFR},

CHICK1={HENS, POUF},
CERE={SWHE, DWHE, BARL, OATS, RYEM, MAIZ},
ARAB={POTA, SUGB}
3. The sets of Northern European, Southern European countries:
'NEUR'={NL000, UK000, AT000, BL000, DE000, DK000, FI000, FR000, SE000}
'SEUR'={El000, ES000, PT000, IT000, IR000}

Table 7: Codes of land use classes (Set LandUse)

OART	artificial
ARAO	(other) arable crops - all arable crops excluding rice and fallow (see also definition of ARAC below)
PARI	paddy rice (already defined)
GRAT	temporary grassland (alternative code used for CORINE data, definition identical to TGRA
FRCT	fruit and citrus
OLIVGR	Olive Groves
VINY	vineyard (already defined)
NUPC	nursery and permanent crops (Note: the aggregate PERM also includes flowers and other vegetables
BLWO	board leaved wood
COWO	coniferous wood
MIWO	mixed wood
POEU	plantations (wood) and eucalyptus
SHRUNTC	shrub land - no tree cover
SHRUTC	shrub land - tree cover
GRANTC	Grassland - no tree cover
GRATC	Grassland - tree cover
FALL	fallow land (already defined)
OSPA	other sparsely vegetated or bare
INLW	inland waters
MARW	marine waters
KITC	kitchen garden

Table 8: Codes of land use aggregates (Set LandUseAgg)

OLND	other land - shrub, sparsely vegetated or bare
ARAC	arable crops
FRUN	fruits, nursery and (other) permanent crops
WATER	inland or marine waters
ARTIF	artificial - buildings or roads
OWL	other wooded land - shrub or grassland with tree cover (definition to be discussed)
TWL	total wooded land - forest + other wooded land
SHRU	shrub land
FORE	forest (already defined)
GRAS	grassland (already defined)
ARAB	arable (already defined)
PERM	permanent crops (already defined)
UAAR	utilizable agricultural area (already defined)
ARTO	total area - total land and inland waters
ARTM	total area including marine waters
CROP	crop area - arable and permanent

Table 9: Codes of mutually exclusive subset adding up to total area - ARTO (Set LandUseARTO)

OLND	other land - shrub, sparsely vegetated or bare
ARTIF	artificial - buildings or roads
FORE	forest
UAAR	utilizable agricultural area
INLW	Inland waters

Figure 48: Mapping primary agricultural activities to groups and land use in coco