CAPRI model documentation 2014:

Editors[1]: W. Britz, P. Witzke

Table of Contents

CAPRI model documentation 2014: 1

Table of Contents 1

0 Preface 5

1 Introduction (W. Britz) 5

1.1 Structure of the documentation 5

1.2 History of CAPRI 5

1.3 Overview on CAPRI 7

2 The CAPRI Data Base 10

2.1 Production Activities as the core (W. Britz) 11
2.1.1 Technology variants for production activities (W. Britz, M Adenaeuer) 12

2.2 The Complete and Consistent Data Base (COCO) for the national scale (P. Witzke, A. Zintl, M. Kempen) 14
2.2.1 Overview and data requirements for the national scale 14
2.2.2 Data Import 16
2.2.3 COCO1: Overlay from various sources 24
2.2.4 COCO1 Estimation procedure 39
2.2.5 COCO2: Data Preparation 45
2.2.6 COCO2: Estimation procedure 47
2.2.7 COCO2: Final completions 49

2.3 The Regionalised Data Base (CAPREG, W. Britz) 51
2.3.1 Data requirements at regional level 51
2.3.2 Data sources at regional level 52
2.3.3 Data availability at regional level 52
2.3.4 Reading and storing the original REGIO data 52
2.3.5 Methodological proceeding 53
2.3.6 Prices for outputs and inputs 53
2.3.7 Filling gaps in REGIO 54
2.3.8 Mapping crop areas and herd sizes from REGIO to COCO definitions 54
2.3.9 Perfect aggregation between regional and national data for activity levels 55
2.3.10 Outlier detection for yields 56
2.3.11 Estimating expected yields with a Hodrick-Prescott filter 56

2.4 Input Allocation (W. Britz) 56
2.4.1 Input allocation excluding young animals, fertiliser and feed 57
2.4.2 Input allocation for young animals and the herd flow model 63

[1] The editors and sections authors mentioned so far only indicate the contributors to this update. They are not necessarily the original authors and developers of this material.

2.4.3 Input allocation for feed 66
2.4.4 Input allocation for fertilisers and nutrient balances 68
2.4.5 Input allocation for labour (Markus Kempen, Eoghan Garvey) 79

2.5 The world Data Base (M. Adenäuer) 93
2.5.1 Biofuels in the global database 94

2.6 Policy data 94
2.6.1 Policy data linked to production activities in Europe (P. Witzke, M. Adenäuer, T. Jansson) 94
2.6.2 Policy data linked to European and international markets (W. Britz, P. Witzke, B. van Doorslaer) 95

3 Baseline Generation (W. Britz, P. Witzke) 97

3.1 Overview beyond CAPTRD 97

3.2 CAPTRD Step 1: Independent, weighted nonlinear least squares 99

3.3 CAPTRD Step 2: Consistency constraints in the trend projection tool 100
3.3.1 Constraints relating to market balances and yields 102
3.3.2 Constraints relating to land use and cropping area 103
3.3.3 Constraints relating to agricultural production 103
3.3.4 Constraints relating to prices, production values and revenues 105
3.3.5 Constraints relating to consumer behaviour 105
3.3.6 Constraints relating to processed products 106
3.3.7 Constraints relating to bio-fuel production 106
3.3.8 Constraints relating to policy 107
3.3.9 Constraints relating to growth rates 107

3.4 CAPTRD Step 2: Integration of specific expert support (Member State level or lower) 108
3.4.1 Integration of bio-fuel information from PRIMES 109

3.5 CAPTRD Step 3: Adding comprehensive sets of supports from AGLINK or other agencies 111
3.5.1 Integration of expert information for long run projections beyond 2020 112

3.6 CAPTRD: Breaking down results from Member State to regional and farm type level 114

3.7 Calibrating the model to the CAPTRD projection 115
3.7.1 Calibrating the regional supply models 115
3.7.2 Calibrating the global trade model 116

4 Simulation Scenario Model (CAPMOD) 119

4.1 Overview of the system (W. Britz) 119

4.2 Module for agricultural supply at regional level (W. Britz) 121
4.2.1 Basic interactions between activities in the supply model 121
4.2.2 Detailed discussion of the equations in the supply model 123
4.2.3 Calibration of the regional programming models 133
4.2.4 Estimating the supply response of the regional programming models 134
4.2.5 Price depending crop yields and input coefficients (W. Britz) 134

4.3 Premium module (T. Jansson, W. Britz, J. Helming) 135
4.3.1 Overview 135
4.3.2 Basic concept 135
4.3.3 A didactic example and further explanations 139
4.3.4 Mc Sharry and Agenda reform premiums 140
4.3.5 The MTR-reform and the health check 142
4.3.6 The October 2013 proposal by the Commission: Greening top-up 149
4.3.7 Tradable Single Premium Scheme entitlements 150
4.3.8 Pillar II 153
4.3.9 Co-financing rates, assignment of premiums to pillars, WTO boxes and PSE-types 153

4.4 Market module for young animals (W. Britz) 156

4.5 Market module for agricultural outputs (W. Britz, M. Adenäuer) *157*
4.5.1 Overview on the market model 157
4.5.2 The approach of the CAPRI market module 159
4.5.3 Behavioural equations for supply, feed demand and land markets 160
4.5.4 Behavioural equations for final demand 162
4.5.5 Behavioural equations for the processing industry 163
4.5.6 Trade flows and the Armington assumption 165
4.5.7 Market clearing conditions 167
4.5.8 Price linkages 167
4.5.9 Bio-fuel module (A. Becker and M. Adenäuer) 169
4.5.10 Endogenous policy instruments in the market model 178
4.5.11 Endogenous tariffs under Tariff Rate Quotas, flexible levies and the minimum import price regime for fruits and vegetables of the EU 180
4.5.12 Overview on a regional module inside the market model 186
4.5.13 Basic interaction inside the market module during simulations 187

4.6 Solving the market model (W. Britz) *188*
4.6.1 Calibration of the system of supply functions 188
4.6.2 Calibration of the final demand systems 189
4.6.3 Overview on the calibration mechanism 190

4.7 Linking the different modules – the price mechanism (W. Britz) *192*
4.7.1 Price linkage in the sugar sector 193

4.8 Sensitivity of the CAPRI model to the Armington substitution elasticities (W. Britz) *193*

5 Farm Type Programming Model: a FADN-based approach (A.Gocht, W. Britz) 199

5.1 The CAPRI farm type approach *199*

5.2 Linkage to a SEAMLESS Farm Type Models *205*

6 A feasibility study for the now abandoned recursive-dynamic version (W. Britz) 206

7 Post model analysis 209

7.1 Dual analysis (T. Jansson, W. Britz) *209*

7.2 Decomposition of changes in aggregates yields and activity related income indicators (W. Britz) *209*

7.3 Post model sensitivity analysis with features in the supply model (W. Britz) *211*
7.3.1 Methodology 211
7.3.2 Technical implementation 211

7.4 Welfare analysis (W. Britz, T. Jansson, P. Witzke) *213*

7.5 Energy Use in Agriculture (T. Kränzlein, M. Kempen) *214*
7.5.1 Introduction and basics 214
7.5.2 Energy assessment in CAPRI 215
7.5.3 Energy output assessment 221
7.5.4 Analysis of CAPRI energy module results 222

8 Spatial dis-aggregation and link to the biophysical model DNDC 225

8.1 A spatial land use map (M. Kempen and R. Köbl) *225*
8.1.1 Spatial calculation unit (HSMU) 225
8.1.2 Consistent disaggregation of land use 227

8.2 Yields, irrigation shares and stocking densities (W. Britz) *234*

8.3 Linkage to process-based modelling (DNDC) (A. Leip and G. Mulligan) *236*
8.3.1 Introduction 236
8.3.2 Input data 238
8.3.3 Model setup 245

8.4 Landscape indicators (M.L. Paracchini) *246*
8.4.1 Background on indicators for agrarian landscapes 247
8.4.2 Indicators on configuration 249
8.4.3 Indicators on farming orientation/management intensity 254
8.4.4 Indicators on farming and ecosystems 255

References **263**

Annex: Code lists **264**

0 Preface

Documentation is a notoriously neglected area in model development because it evidently competes with model application and model improvement which are usually considered more important by model developers, the people best equipped to report on their work. This neglect is reinforced by outside factors such as given deadlines for the completion of some deliverables. A further impediment is given by division of labour and modular development of systems which implies that the respective specialists from the whole CAPRI network should participate in the documentation activity.

The lack of ongoing and up to date documentation has been identified recently as a serious impediment for knowledge dissemination and a conclusion has been reached to set up an own website "CAPRI_MODEL.ORG" for web based documentation. An attractive example meeting many but not all needs of CAPRI is given by the website of the MIRAGE model (http://mirage.cepii.free.fr/miragewiki/index.php?title=Accueil). However the development of this web based documentation will take a while.

End of 2008, a user manual for CAPRI was generated. The section on the CAPRI Graphical User interface was therefore removed. This version covers updates up to the end of 2011, as far as the developers managed to document their modifications.

1 Introduction (W. Britz)

1.1 Structure of the documentation

The documentation is structured as follows. The short introduction in Section 1.2 first gives an overview of past CAPRI activities followed by a short description of the system (Section 1.3). The rest of the document largely follows the project workflow: the different steps of building up the national, regional and global data base provide the foundations on which the system rests (Chapter 2). Subsequently the procedure needed to establish a baseline (Chapter 3) is discussed. Chapter 4 deals with the scenario impact analysis: description of the different modules of the economic model and their relationships, also addressing the calibration approach of CAPRI (even where this is technically a part of the database modules). In the last four chapters (Chapters 5 - 8) the farm type approach, a feasibility study on a recursive-dynamic CAPRI version, features for post model analysis, and options for spatial downscaling are presented.

1.2 History of CAPRI

CAPRI stands for '**C**ommon **A**gricultural **P**olicy **R**egionalised **I**mpact analysis' and is both the acronym for an EU-wide quantitative agricultural sector modelling system and of the first project centred around it[2]. The name hints at the main objective of the system: assessing the effect of CAP policy instruments not only at the EU or Member State level but at sub-national level as well.

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 vTI) 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

[2] Web Site: http://www.agp.uni-bonn.de/agpo/rsrch/capri/capri_e.htm.

market reform options in the context of another project improved the way the complex ABC sugar quota system is handled in the model.

In 2004, again a larger project (EU research FP VI, Nr. 501981: CAPRI-Dynaspat) started 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. At the same time, a project began to apply CAPRI to analyse the effects of bi-lateral trade liberalisation with Mediterranean countries (FP VI, Nr. 502457: EU-MedAgPol). In 2005/2006, a project for IPTS/JRC updated and improved the farm type model layer and to include Bulgaria and Romania. At the same time, the SEAMLESS project (FP VI: 2005-2009) started, with CAPRI used to link results with a complex layer of farm type models and from there to national, EU and global markets. In parallel, the team in LEI, The Hague, The Netherlands, applied CAPRI in the integrated project SENSOR (2005-2008). In 2006-2008 JRC ISPRA has taken over initiatives to improve the linkages to biophysical model DNDC and to achieve a first biofuels coverage in CAPRI during an interim stay of Wolfgang Britz at ISPRA. In 2006-07 CAPRI made contributions to study "Integrated measures in Agriculture to reduce Ammonia emission"[3] together with MITERRA-Europe (Alterra, Wageningen) and GAINS (IASSA, Laxenburg) which led to an update of the N-cycle description in CAPRI. Since 2006 CAPRI is part of LIFE funded EC4MACS[4], the "European Consortium for Modelling of Air Pollution and Climate Strategies" which will strengthen earlier linkages to GAINS and other models in this network. Since 2007 CAPRI is also contributing to CCAT – EU Cross compliance tool[5] an FP6 project coordinated by Wageningen University for an integrated assessment of cross compliance impacts. Also in 2007 work has started on CAPRI FARM[6] aiming at an analysis of farming sustainability. For this purpose the NUTS II modelling regions of CAPRI will be split up to 10 typical farms together with a certain overhaul of the current farm type layer in CAPRI. Since June 2008 CAPRI was part of FP7 project CC-TAME (Climate Change - Terrestrial Adaptation and Mitigation in Europe) combining complementary climate and land use models with different focus where the base line horizon of CAPRI was be extended to 50 years. In turn linkages to forestry and bioenergy sectors of EUFASOM/GLOBIOM have been developed, a direction of model development also supported by several service contracts for DG CLIMA (recently: project "EUCLIMIT"). GHG abatement options have also been investigated in two studies by the JRC (IPTS, Seville[7] and IES, Ispra[8]) that involved CAPRI.

In 2009/2010 projects financed by JRC-IPTS Sevilla allowed a complete rehaul of the CAPRI farm type layer and to integrate a bio-fuel component of CAPRI.

Until summer 2013, again a EU framework project co-ordinated by the team in Bonn called "CAPRI-RD" ensured system maintenance and updates, while working on the integrating of CAP pillar 2 measures in the system, adding a layer of regional CGE and further improvements of the system. In parallel, a project called GLUE-S, financed by the German research department, will extend the market model by integrating a land use components.

During the years, the system was applied to a wide range of different scenarios. 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 biological production systems. WTO scenarios were run by the team in Bonn in 2002 and 2005. Moreover, CAPRI was applied to

---

3 See http://ec.europa.eu/environment/air/cafe/activities/ammonia_en.htm

4 See http://www.ec4macs.eu/home/index.html

5 See http://www.ccat.wur.nl/UK/General+Information/

6 See http://agrilife.jrc.ec.europa.eu/s_study3.html

7 See http://ftp.jrc.es/EURdoc/JRC69817.pdf

8 See http://ec.europa.eu/agriculture/analysis/external/livestock-gas/full_text_en.pdf

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) and tradable permits for greenhouse gas emission from agriculture analysed (Pérez 2005). The team in Louvain-La-Neuve, together with the group in Bonn, analysed sugar market reform options, applying the market module linked to the regional supply models (Adenaeuer et al. 2004). In 2004 followed an analysis of a compulsory insurance paid by farm against Food and Mouth disease by SLI and runs dealing with methane emission by the team in Galway, Ireland. In the same year, CAPRI was installed by DG-AGRI in Brussels and a baseline generated in order to match DG-AGRI's outlook projections which has become a regular activity. The ammonia study involved scenarios to investigate technical abatement options for ammonia and nitrates emissions with scenario assumptions coordinated with GAINS. Several studies have been launched in 2007 on particular aspects of the ongoing CAP reform, in particular a decoupling project by LEI for DEFRA, UK, a modulation study by LEI for DG Agri and a study coordinated by EuroCARE Bonn on the impacts of the expiry of the milk quota system in 2015 for JRC, IPTS, Seville.

Three teams should be mentioned, as they provided their own funds to share the network and contribute to the system: the teams at FAT, Tänikon in Switzerland, the team at NILF, Oslo in Norway, and the team at SLI, Lund in Sweden. In addition the EU Commission has provided support in various forms. If not explicitly mentioned in the following, the documented features had been co-financed by DG-RSRCH through the series of past FP projects. Furthermore the DG-JRC (IPTS, Seville and IES Ispra) has actively contributed to improvements and extensions in various components of the system. The documentation as it stands now captures the state of the system in November 2012, at least for those sections that have been updated recently. However it has to be admitted that each update leaves several sections in its current form even though the CAPRI GAMS code has already moved ahead. In spite of this moving target problem it seems preferable to release improved documentation in shorter time intervals than to strive for a "comprehensive" update that almost requires a kind of (undesirable) standstill.

1.3 Overview on CAPRI

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.

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)[9]. Specific modules ensure that the data used in CAPRI are mutually compatible and complete in time and space. They cover about 50 agricultural primary and processed products for the EU (see Table 43 in 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 large modelling network such as SEAMLESS. 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 possibility aggregation errors in CAPRI.

The economic model, comparative-static in nature is split into two major modules. 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 programming models are a kind of hybrid approach, as they combine a Leontief-technology for

[9] 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.

variable costs covering a low and high yield variant for the different production activities 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. 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 -, set-aside obligations and milk quotas. The complex sugar quota regime is captured by a component maximising expected utility from stochastic revenues. 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 let total area use shrink and expand depending on returns to land.

The market module consists of two sub-modules. The sub-module *for marketable agricultural outputs is a spatial, non-stochastic global multi-commodity* model for about 50 primary and processed agricultural products, covering about 70 countries or country blocks in 40 trading blocks (see Section 4.5.1). Bi-lateral trade flows and attached prices are modelled based on the Armington assumptions (Armington 1969). The behavioural functions for supply, feed, processing and human consumption 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. Policy instruments cover (bi-lateral) tariffs, the Tariff Rate Quota (TRQ) mechanism and, for the EU, intervention stocks and subsidized exports. This sub-module delivers prices used in the supply module and allows for market analysis at global, EU and national scale, including a welfare analysis. A second sub-module deals with prices for young animals.

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.

The regional CGEs are also linked based on sequential iteration in the system: the response of the agricultural sector from the supply modules and price information from the market model is used to update the production nest for agriculture, while the CGE deliver back information on primary factor prices for labor and capital.

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. The farm type model layer may be switched ON or OFF. Equally, the regional CGEs and the global market model can be run in stand-alione mode.

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 high detail. Environmental indicators cover NPK balances and output of climate relevant gases according the guidelines of the Intergovernmental Panel on Climate Change (IPCC). 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.

An important recent extension is the *spatial down-scaling part* to clusters of 1x1 km grid cells, covering crop shares, crop yields, animal stocking densities and fertilizer application rates and allows for linkage with the bio-physical model DNDC.

The *technical solution* of CAPRI is centred on the modelling language GAMS which is applied for most of the data base work and CONOPT applied as solver for the different constrained (optimisation) problems. The different modules are steered by a Graphical User Interface based on JAVA which also allows exploitation of results as tables, graphs and maps. The different data are either stored in GAMS readable text format or in the GAMS specific binary GDX format. The GAMS code, data base and the Java code underlying the GUI are maintained via a software version system.

Methodological development, updating, maintenance and application of CAPRI are based on a *network approach* with is currently centred in Bonn, but with distributed responsibilities. The CAPRI modelling system may be defined as a 'club good': there are no fees attached to its use but the entry in the network is controlled by the current club members. The members contribute by acquiring new projects, by quality control of data, new methodological approaches, model results and technical solutions, and by organising events such as project meetings or training sessions. So far, the network approach worked quite successfully but it might need revision if the club exceeds a certain size.

2 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 algorithm 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 EU27 plus Norway and Western Balkan countries - need to fit to data at regional level -administrative units at the so-called NUTS 2 level, about 300 regions for EU25- and data at global level, currently 23 non-EU regions. A further layer consists of georeferenced information at the level of clusters of 1x1 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 2.2).
- 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 2.3).
- 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 2.4.
- 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) plus data domestic market support instruments (market interventions, subsidies to consumption) (Section 2.5).
- 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 2.6.

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. The concept was developed end of seventies building on similar approaches at the farm level at the Institute for Agricultural Policy in Bonn and first applied in the so-called SPEL/EU data base.

2.1 Production Activities as the core (W. Britz)

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, the problem of subsidies outside of CAP for the EU Member States remains so far unsolved, but is on the agenda for future ameliorations.

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	%

Source: CAPRI data base, Denmark, three year average 2000-2002

2.1.1 Technology variants for production activities (W. Britz, M Adenaeuer)

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*').

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):

Equation 1 $$GROF_{io} = \sum_{j} LEVL_{j} IO_{j}$$

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

$$GROF_{io} - SEDF_{io} - LOSF_{io} - INTF_{io} = NETF_{io}$$

$$NETF + IMPT_{io} = EXPT_{io} + STCM_{io} + FEDM_{io} + LOSM_{io} + SEDM_{io} + HCOM_{io} + INDM_{io} + PRCM_{io} + BIOF_{io}$$

Equation 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), which has been introduced recently (Section 1.2).

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):

Equation 3 $$EAAP_{io} = UVAP_{io} NETF_{io}$$

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**	**Positions 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	

2.2 The Complete and Consistent Data Base (COCO) for the national scale (P. Witzke, A. Zintl, M. Kempen)

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 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.

2.2.1 Overview and data requirements for the national scale

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 order to use this information directly in the model, the CAPRI and CAPSIM[10] teams developed out of EUROSTAT data a complete and consistent data base (COCO) at Member State level (Britz et al. 2002). In the attempt to include first the New Member States (NMS) and subsequently the Western Balkan Countries and Turkey into the database additional national sources were used which became available as part of CAPSIM projects[11]. For special issues like trade, biofuels, sugar and milk data from further sources were implemented.

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

[10] See http://www.eurocare-bonn.de/profrec/capsim/capsim_e.htm.

[11] See Witzke, Zintl, Tonini 2008 for details on the Western Balkan countries. The extension to the NMS occurred under an earlier Eurostat project in 2005 (Ref. 2004/S 42-036276/EN).

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 New Member States: Eurostat supplemented with FAOstat and a few data from the questionnaires received by Eurostat contractor ARIANE[12] For Western Balkan Countries and Turkey: Eurostat supplemented with national statistical yearbooks, data from national ministries, results from AgriPolicy[13], FAOstat production statistics
Production	Eurostat: Farm and market balance statistics, crop production statistics, slaughtering statistics, statistics on import and export of live animals For New Member States: Eurostat supplemented with FAOstat and a few ARIANE data For Western Balkan Countries and Turkey: Eurostat supplemented with national statistical yearbooks, data from national ministries, results from AgriPolicy, FAOstat production statistics
Farm and market balance positions	Eurostat:Farm and market balance statistics FAOSTAT2: Trade and Food Balance Sheets[14] For New Member States: Eurostat supplemented with a few data from the questionnaires received by Eurostat contractor Ariane
Sectoral revenues and costs	Eurostat: Economic Accounts for Agriculture (EAA) For New Member States: Eurostat supplemented with a few data from the questionnaires received by Eurostat contractor Ariane
Producer prices	Derived from production and EAA 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 International Labour Office data on food prices
Output coefficients	Derived from production and activity levels, engineering knowledge

[12] The consulting firm "ARIANE II" was responsible for the year 2006 Eurostat project (related to CAPSIM): Extension of the agricultural sector model to Candidate Countries and establishment of a dataset for use in agricultural sector modelling for Candidate Countries , lot 2:" Establishment of a dataset for use in agriculture sector modelling for Acceding countries ". The dataset compiled during this project will be called "ARIANE data" in the following.

[13] AgriPolicy" supported by the European Commission under the 7th Framework Programme (**FP7**).(http://www.europartnersearch.net/agripolicy/)

[14] Resulting from two EuroCARE projects for FAOstat, Consistency and Completeness of SUA and Trade Matrices Based on Entropy Estimators, Division: ESSD, 2004, 2005 (http://www.eurocare-bonn.de).

2.2.2 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 States	Range
EU15 Member States without Germany	1984 – 2010
Germany and (12) New Member States	1989 – 2010
Western Balkan (WB) Countries and Turkey	1995 – 2009
Norway	1984 – 2010

2.2.2.1 Eurostat data

First step: Data download

Data are downloaded in TSV-format, as offered by Eurostat for bulk data users. The TSV-format is a flat file format for time series, identified by New Cronos codes. Data can be selected for all the EU-27 MS and some Candidate Countries. Availability differs by country, of course, with most severe limitations for some WBs (literally nothing for the Kosovo, Montenegro, Bosnia & Herzegonina). The following themes and table groups of Eurostat are accessed:

Agriculture, forestry and fisheries

- Agriculture ("agri")
    - Economic Accounts for Agriculture (Table Group "aact", saved on CAPRI parameter "COSA", after the historical Eurostat domain name for this data group)
    - Agricultural prices and price indices (Table Group "apri", saved on CAPRI parameter "ZPA1", after the historical Eurostat domain name for this data group)
    - Agricultural products (Table Group "apro", saved on CAPRI parameter "PRAG", after the historical Eurostat domain name for this data group)
- Food: From farm to fork statistics ("food")
    - From production to distribution – which quality label at which price, Relative price level indices of food products (Table "food_pd_prc1", parameter FOOD)
    - Inputs to the food chain - Availability of feedingstuffs (Table "food_in_afeed1", saved on parameter FEEDAGRI)
    - Production of Feedingsstuffs (only available until 2005 under the old Eurostat name "BILARES", saved on parameter FEEDAGRI)

Economy and Finance

- National accounts ("na")
    - Annual national accounts - National Accounts detailed breakdowns (by industry, by product, by consumption purpose) - Final consumption expenditure of households by consumption purpose (COICOP) – Aggregates at current prices (saved on parameter BRKDOWNS) 2 digits and 3 digits
    - Auxiliary indicators to National Accounts - Population and employment (parameter AUXIND)
    - GDP and main components - Current prices, volumes, price indices (parameter AGGS)

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

Second step: Format conversion

The second step of data processing is the conversion of the TSV-files to CSV-files. This step will keep all selected Eurostat data only removing the data flags. The Excel macro “ImportTsvFiles” embedded in the Excel file “template_input.xlsm” is working for each Table Group (“apro”, “aact”, etc.) collecting data from all TSV-files of this domain. The results of the macro runs are particular CSV-files, which will enter a GAMS program (‘*coco_input.gms’*) for further data processing.

Third step: data selection and code mapping

The third step is data selection and code mapping performed by the GAMS program ‘*coco_input.gms’*. Cross sets linking New Cronos 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’*:

- *‘New Cronos_agriculture_mapping.gms’* for the domains from Eurostat’s “Agriculture and Fisheries” Statistics
- *‘New Cronos_econfinc_mapping.gms’* for the domains from Eurostat’s “Economy and Finance” Statistics

Example from file ‘*New Cronos_agriculture_mapping.gms’*

```
SET ZPA1_codes /
 apro_cpp_crop_C1050_AR          "CEREALS-EXCLUDING RICE-"
 apro_cpp_crop_C1120_AR          "COMMON WHEAT AND SPELT"
 apro_cpp_crop_C1123_AR          "COMMON WINTER WHEAT AND SPELT"
 apro_cpp_crop_C1124_AR          "COMMON SUMMER WHEAT AND SPELT"
apro_cpp_crop_C1160_AR          "BARLEY"
.
SET ZPA1MAP(ASS_COLS,ASS_ROWS,ZPA1_CODES)/
CERE.LEVL. apro_cpp_crop_C1050_AR
SWHE.LEVL. apro_cpp_crop_C1120_AR
SWH1.LEVL. apro_cpp_crop_C1123_AR
SWH2.LEVL. apro_cpp_crop_C1124_AR
BARL.LEVL. apro_cpp_crop_C1160_AR
```

The results of the program run are gdx-files for each New Cronos domain.

2.2.2.2 Western Balkan Countries and Turkey

For those countries Eurostat data need completion in almost every area. The structure of these supplementary Excel country sheets and the definitions of the data are tailored to COCO. The resulting database tables 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.

The Excel country files are structured, as shown, in the following table for Croatia:

COUNTRY: Croatia	
Content	
	Economic situation of the country and the agricultural sector
Table 1	Macroeconomic indicators
Table 2	Gross agricultural output, prices and inputs
	Crop and livestock production
Table 3	Crops production
Table 4	Livestock and livestock products: Herd size and production
	Agricultural markets and prices
Table 5	Market balance sheets for main agricultural products
Table 6	Producer prices for agricultural products
	Budget expenditure and policy measures
Table 7	Budgetary expenditures in agriculture
Table 8	Characteristic measures for individual agricultural markets
	Agricultural and food trade
Table 9	Breakdown of agricultural and food exports into food aggregates (e.g. SITC or HS4)
Table 10	Breakdown of agricultural and food imports into food aggregates (e.g. SITC or HS4)
	Farm structure and rural development
Table 11	Indicators for rural development and environment
Table 12	Farm structures
	Data from different sources
Table 13	FAO areas and production, selection 2010
Table 14	FAO livestock primary, selection 2010
Table 15	FAO commodity balance, selection 2010
Table 16	FAO producer prices, selection 2010
Table 17	Eurostat domain ZPA1, selection 2010
	Additional information
Table 18	HS6 and model codes
	Data selection and use fordatabase
Table 19	COCO-Data

The key data, directly connected to the database (COCO) via the coding system is prepared in Tables 1 to 6. As time goes by, all input data for the Western Balkan countries and Turkey will be selected from Eurostat (hopefully). Currently the available Eurostat data are entered like any other source (and marked with its own colour). The manual data selection follows some rules:

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 from data collection in 2006[15] (mainly available for early years).
AgriPolicy data including further updates by CAPRI-RD partner Ales Kuhar are used for the last years, because they are mainly model results.
FAO data provides fall-back solutions for all remaining missing time series.

[15] Resulting from the Arcotrass study: Study on the State of Agriculture in Five Applicant Countries, 2006. see http://ec.europa.eu/agriculture/analysis/external/applicant/index_en.htm

As an example for a sheet with key data we show a part of Table 3: Crops: Sown area and production of Croatia.

Crops: Sown areas		1000 ha	946.00	934.00	952.00	1241.53	1212.89	1256.39	1274.91	1279.50	1259.74	1160.28	848.60	856.00				
sum			821.40	743.40	761.42	850.83	828.05	731.29	747.02	752.51	743.05	733.34	752.43	763.14	741.00	646.77	648.90	
Total cereal		1000 ha	632.00	611.00	633.00	690.00	627.00	696.00	722.00	716.00	690.00		543.00	565.00				
		1000 ha						569.96	588.90	582.90	574.90	571.10	557.20	570.10	558.60			Statistical yearbook 2009, p. 246
	CERE.LEVL	1000 ha	630.25	613.61	636.47	689.36	638.95	561.60	582.80	577.70	557.90	566.00	543.20	559.60	555.70	560.90	562.70	2000-2007 ZPA1 data fit to sum of components (incl. triticale and buckwheat), earlier years sum of components
		1000 ha														562.50		
sum		1000 ha	630.25	613.61	636.47	689.36	638.95	558.90	580.60	574.90	555.20	562.80	538.60	557.90	552.30	557.44	558.78	
Wheat		1000 ha	227.00	201.00	208.00	242.00	169.00	236.00	240.00	234.00	206.00		146.00	176.00				
		1000 ha	227.04	200.85	208.38	241.73	169.28	235.94	239.86	233.70	206.00	214.51	146.25	175.55	175.05	156.54	180.38	
	WHEA.LEVL	1000 ha	196.49	173.82	180.34	209.21	146.50	182.30	184.30	179.20	157.20	162.60	146.30	175.60	175.10	156.54	180.38	2001-2007 ZPA1 data fit to statistical yearbook for SWHE, earlier years corrected FAO data
														176.0	175.0	156.5		Statistical yearbook 2009, p. 247
Soft Wheat	SWHE.LEVL	1000 ha						182.30	184.30	179.20	157.20	162.60	146.30	175.60	175.00	154.70	179.20	
Durum wheat	DWHE.LEVL	1000 ha	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1% assumption according to reply_croatia_data_rev1180606.doc
Buchwheat	BWHE.LEVL	1000 ha						2.50	2.40	2.80	2.70	3.10	4.60	0.20	0.70	0.40	0.80	
Rye and Meslin		1000 ha	2.00	2.00	2.00	2.00	2.00	3.00	3.00	3.00	3.00		2.00	2.00				
		1000 ha	1.93	2.04	1.96	2.15	2.45	2.93	3.19	3.47	3.19	2.00	1.85	2.01	1.73	1.37	1.00	
		1000 ha						2.70	3.00	3.20	3.00	2.90	1.90	2.00	1.70			Statistical yearbook 2009, p. 247
	RYEM.LEVL	1000 ha	1.92	2.04	1.95	2.14	2.44	2.70	3.00	3.20	3.00	2.90	1.80	2.00	1.70	1.40	1.00	2001-2007 ZPA1 data fit to statistical yearbook, earlier years corrected FAO data
Corn/Maize		1000 ha	354.00	361.00	371.00	378.00	384.00	389.00	406.00	407.00	406.00		319.00	296.00				
		1000 ha	354.06	361.27	371.27	377.82	384.18	388.64	406.15	408.88	405.95	413.76	318.97	296.20	288.50			
		1000 ha										306.3			288.5	314.1		
		1000 ha						292.40	305.90	306.80	304.70	306.40	318.90	296.20	288.60	314.06	296.91	Statistical yearbook 2009, p. 247
	MAIZ.LEVL		366.61	374.07	384.43	391.21	397.80	292.40	305.90	306.80	304.70	306.30	319.00	296.20	288.50	314.10	296.90	2000-2007 ZPA1 data fit to statistical yearbook, earlier years corrected FAO data
Barley		1000 ha	33.00	31.00	34.00	43.00	45.00	46.00	51.00	51.00	54.00		50.00	59.00				
		1000 ha	32.52	31.03	33.76	42.74	44.52	46.36	51.17	50.65	53.83	46.00	50.34	59.16	59.00			
		1000 ha						55.50	61.30	61.20	65.00	67.50	50.30	59.20	59.00	65.54	59.58	Statistical yearbook 2009, p. 247
	BARL.LEVL	1000 ha	47.16	45.01	48.96	61.98	64.57	55.50	61.30	61.20	65.00	67.50	50.30	59.20	59.00	65.50	59.60	2000-2007 ZPA1 data fit to statistical yearbook, earlier years corrected FAO data
Oats		1000 ha	16.00	16.00	18.00	22.00	24.00	20.00	20.00	19.00	20.00	...	21.00	25.00				
		1000 ha	15.76	16.29	18.14	21.67	24.12	20.38	20.53	19.22	19.94	20.00	21.19	24.91	27.97			
		1000 ha						26.00	26.10	24.50	25.30	23.50	21.20	24.90	27.97			Statistical yearbook 2009, p. 247
	OATS.LEVL	1000 ha	18.06	18.66	20.79	24.83	27.64	26.00	26.10	24.50	25.30	23.50	21.20	24.90	28.00	19.90	20.90	2000-2007 ZPA1 data fit to statistical yearbook, earlier years corrected FAO data
Other cereals (triticale, sorghum...)		1000 ha	0.00	0.00	0.00	3.00	3.00	2.00	2.00	2.00	1.00	...	5.00	7.00				
		1000 ha	0.13	0.11	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13			
	OCER.LEVL	1000 ha	0.00	0.00	0.00	0.00	0.00	9.31	6.55	6.25	17.95	6.45	16.85	10.45	4.49	0.00	0.00	
- Triticale	TRIT.LEVL	1000 ha												1.60	2.70	3.20	3.10	

Source: CROSTAT
FAOSTAT 2010 selection
Statitical yearbooks
AgriPolicy statistiics
Eurostat 2010
Calculated data

A number of additional data tables (Table 7 to Table 12) are added with specific country data. This information may help to view agriculture in its entire economic and environmental context. These data is not directly input for the COCO database but may help to compute other values. Some tables (like Table 12) are not used at all at the moment.

Table 19 is the interface to the GAMS programing world of COCO. The Excel macro "SELECT_data_all" collects all time-series from Table 1 to Table 6, which are flagged with a COCO code. In other words alternative series for the same item that have often been included for comparison purposes in Tables 1 to 6 are ignored, provided they do not carry a code.

2.2.2.3 Supplementary data for Romania and Bulgaria

Country level data were available in Excel files that help in particular to complete the meat and milk sectors, see Section 2.2.3.2.

2.2.2.4 FAO data selection

For all Regions covered missing trade data are completed by FAO data, because trade data offer the option to compute total demand if production is given and thus to establish the key data for complete market balances in the frequent case where only production data are given. For the New Member States data on production and levels are as well taken from FAO if missing otherwise.

Two FAO data sources are combined:

- Results of the WATSIM model (see Chapter 1.2), where the most recent year is 2002. These data have lost in importance in the last years but may be still in use for selected series in some New Member States.
- Results of the FAOSTAT2[16] project, where trade series have been completed to 2005.

2.2.2.5 Other additional input data

COCO1: Biofuels

- Production quantities for biodiesel and bioethanol are collected from:
  - EBB, European Biodiesel Board, Brussels (http://www.ebb-eu.org/stats.php)
  - ePURE, European renewable ethanol, Brussels (http://www.epure.org/statistics/info/Productiondata)
  - PRIMES model[17] database
- Consumption quantities and demand for diesel, gasoline, biodiesel and bioethanol are given from Eurostat
- Market balance positions for ethanol are collected from FO Licht's World Ethanol and Biofuels Report.
- 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).
- Taxes for diesel, gasoline, biodiesel and bioethanol are collected from 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.

[16] Resulting from two EuroCARE projects for FAOstat, Consistency and Completeness of SUA and Trade Matrices Based on Entropy Estimators, Division: ESSD, 2004, 2005 (http://www.eurocare-bonn.de).

[17] 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.
- 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.
- For the New Member States additional information on levels of fresh milk products are still taken from the ARIANE data.

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). All relevant (raw) information is stored in a gdx file. The data base comprises information on land use classes from various sources:

- REGIO - Eurostat, land use, REGIO domain(NUTS2 level - yearly, 1984-2010)
- ENVIO - Eurostat, land use, env_la_luc1.xls (MS level - 1985, 1990,1995, 2000)
- LANDCOVER - Eurostat, land cover(MS level - 2009)
- FSS - Eurostat, FSS(NUTS2 level - 1990, 1993, ..., 2007)
- Corine Land Cover (CLC), 44clc_nuts2.xls (NUTS2 level - 1990, 2000, 2005)
- FAO - area.xls(MS level - yearly, 1984-2007)
- MCPFE - MCPFE_quantitative_tables_SoEF2007.xls (MS level - 1990, 2000, 2005)

During preparation obviously false data was deleted (e.g. the 2006 CLC data only covered parts of Greece, hence we cannot calculate reasonable totals at MS level, see '*\coco_input\land_use\ReadAndPrepForDatCoco.gms*'), but no further plausibility check is done.

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, AUXIND: population. To complete early years data from Eurostat: Old domain AGRIS, Population. This domain does no longer exist, only until 2004.
- Country Sheets for the Western Balkan Countries and Turkey: Data from other national statistics on exchange rates, number of inhabitants and inflation rate
- The GDP price index expressed in Euros with 2000 = 100 is selected from Eurostat, AGGS.

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, BRKDOWNS
- United Nations Statistics Division (UNSTATS): Household consumption expenditure in USD
- Eurostat, PRICE: Consumer expenditure weights
- 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.
- The Institute of Statistics Albania (INSTAT): Consumer price indices and shares
- Household budget survey Bosnia: Consumer price shares
- Croatian Bureau of Statistics: Consumer price shares
- 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, consumer prices for Bosnia.

COCO2: By-products

- FAO: Food Balance Sheets, Commodity Balances, Livestock and Fish Primary Equivalence: Imports and exports quantities for fish meal, dried cassava, gluten deed and meal, as well as feed quantities for fish meal.
- Eurostat, PRAG: Purchase prices for fish meal, dried sugar beet pulp and wheat bran

- FAO: Food Balance Sheets, Commodity Balances, Crop Primary Equivalence: 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
- Eurostat, PRAG: Purchase prices for soya cake.

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
- FAPRI: Recent collection

2.2.3 COCO1: Overlay from various sources

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, it will use the second or even third best source if useful.

2.2.3.1 Including data from New Cronos ('coco1_eurostat.gms')

The program starts by importing data from Eurostat prepared beforehand, by the data selection routines and manual checking. The different domains are processed step by step and corrections made on selected data for all MS[18].

The first Eurostat Table Group is "apro" (historical name ZPA1) covering market balances and activity levels. After specific corrections for a particular MS, some aggregations applicable to all MS are performed. The COCO item "APPL" is aggregated from auxiliary data for apples, pears and peaches and COCO code "FRUI" is aggregated from auxiliary data for fruit trees, plus soft fruits, plus strawberries. The example below (aggregation of "APPL") illustrates a general rule in COCO, not always followed in the raw data. Series are only aggregated if all components are given or if the component series is entirely empty:

[18] 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.

```
ZPA1(MS,"APPL","LEVL",TT1) $
  ( (ZPA1(MS,"APPS","LEVL",TT1) OR (SUM(TT2,ZPA1(MS,"APPS","LEVL",TT2)) EQ 0))
   AND (ZPA1(MS,"PEAR","LEVL",TT1) OR (SUM(TT2,ZPA1(MS,"PEAR","LEVL",TT2)) EQ 0))
   AND (ZPA1(MS,"PEAC","LEVL",TT1) OR (SUM(TT2,ZPA1(MS,"PEAC","LEVL",TT2)) EQ 0)))
= ZPA1(MS,"APPS","LEVL",TT1) +
  ZPA1(MS,"PEAR","LEVL",TT1) +
  ZPA1(MS,"PEAC","LEVL",TT1);
```

Other corrections on "ZPA1" data applicable to all MS deal with specific problems on wine (scaling, changing codes over time), paddy rice (included in total rice balance data) and sheep and goat slaughtering (aggregated in some MS, disaggregated in others).

To complete the COCO database in early 2011 for years up to 2010 specific solutions were needed:

- Bulgaria and Belgium: Area levels for most crops were deleted from Eurostat for the year 2000 to 2007. As an interim solution data for missing levels from Eurostat selection of March 2010 is used
- For animal slaughtering: For the recent years 2007 -2010 slaughtering data from the monthly slaughtering statistics are added up and included.
- Data for human consumption of soybeans from FAO SUA is included, because Eurostat data is mixed up with processing and industrial use and FAO should be one of the most competent organisations for food consumption
- Data on production and area levels for olives in France are selected from Ministère de l'agriculture, de l'alimentation, de la pêche, de la ruralité et de l'aménagement du territoire (AGRESTE), Paris (http://agriculture.gouv.fr/).
- Trade data for sugar are collected from Eurostat COMEXT data.

Other Eurostat domains covered in '*coco1_eurostat.gms'* are EXINT (exchange rates), COSA (Economic Accounts) and PRAG (prices). These domains only require a few case-by-case corrections.

2.2.3.2 Data from additional sources for the NMS ('coco1_non_ms15.gms')

For the ten NMS additional data are included 'ARIANE data'). This fall back information is still used if no data is available from Eurostat.

A data check for Bulgaria suggested to revise the data for the meat sector (slaughtered tons and heads) based on FAO, because FAO reasonably fits to the country data in 2004 and more years are available than from New Chronos.

National data are used for Romania both in the meat and in the milk sectors. National information is available for 2004 and 2005 in the meat sector for all cattle components. FAO data reasonably match national data for 2004 and 2005 and fit to Eurostat data from 1995 to 1998 but cover also more recent years. The FAO and national data are combined using a correction factor equal to a ratio from national information to FAO data for 2004 and 2005. More importantly are the corrections in the Eurostat raw milk balance ("RMLK") based on national data, which sometimes appeared to be more complete and plausible than Eurostat. Essentially, national information is used to disaggregate those Eurostat series considered plausible.

2.2.3.3 Data from additional sources for the Western Balkan Countries and Turkey ('coco1_candi_zpa1_prag.gms')

Specific data files are included for the Western Balkan Countries and Turkey as explained in Section 2.2.2.2. The module *'coco1_candi_zpa1_prag.gms'* further prepares the WB data to match the COCO definitions:

1. Similar to EU-27 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:
   - Trade data are frequently missing in the WBs, such that FAO data are included where available.
   - Production of oilcakes and sugar is estimated from raw products, if missing, using the sister country aggregate processing coefficients;
   - 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.

2.2.3.4 Final completions and revisions for all Member States ('coco1_finish_zpa1.gms')

- Completion of production data from the Eurostat market balance statistics (model code "USAP") has also been tried with quantity information given from the agricultural account statistics (model code "EAAQ") using a correction factor calculated from overlapping years.
- As regards trade, FAO data are included as fall back information for all MS. Trade is often a crucial point in data selection, because it is very volatile and difficult to estimate with smooth trends. The advantage of FAO is that trade is usually well covered in FAO. Trade series have been completed to 2005 based on a FAOSTAT2 project. If trade data are not available from FAO, they are estimated by a weighted moving average (weights decreasing when distance to the gaps increases) to obtain at least a mechanical completion of trade information.
- Domestic use can be calculated[19] 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 "ZPA1" 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, total demand is calculated from the production data from Eurostat and the above mentioned FAO based trade data. 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.
- Additional information on meat production for Austria is as well selected from FAO, because all market balances before 1993 are missing from Eurostat.
- 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")

[19] Extraordinary jumps are also deleted and completion of time series is left to subsequent trend estimations. To avoid inconsistencies in aggregates, the groups for exports, import and domestic use are recalculated.

Finally the biofuel sector is prepared (Section 2.2.3.5).

2.2.3.5 EU biofuel sector data ('coco1_finish_zpa1.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 (see Section 2.2.2.5). The production information used by CAPRI is in detail in the following table (compare *biofuel\prepare_biofuel_data.gms*):

Table 5 Overview on data sources utilized for biofuel production

Source	Variable	Period	Regions
ePURE	Fuel ethanol	2002-2011	EU MS
EBB	Biodiesel	2002-2011*	EU MS
PRIMES	Fuel ethanol	2000-2007	EU MS
PRIMES	Biodiesel	2000-2007	EU MS
AGLINK-COSIMO	Share of ethanol from non-agriculture	2001-2018	EU27
F.O.Licht	Denatured ethanol	2003-2008	EU MS
F.O.Licht	Total (denatured or not) ethanol	2003-2008	EU MS

* 2011 data are still estimates in November 2012.

The overlay considers data availability and consistency across sources:

- Fuel ethanol is mainly from PRIMES (consistent with www.epure.org/statistics/info/Productiondata)
- Total ethanol production is mainly from FO Licht (found in line with AGLINK-COSIMO).
- Biodiesel is from PRIMES for early years (found well in line with EBB) and has been extended with EBB after 2006

A final problem related to the biofuel market balances is the share of non-agricultural production (NAGR). For ethanol the share given by AGLINK is adopted. For biodiesel there is very scattered information only such that for most countries (except UK and DE) a share of 10% is assumed.

In addition to being the min source for total ethanol quantities, F.O.Licht production data have also been used in part for some non-European countries as part of the global database (Section 2.5).

Biofuel consumption

COMEXT trade data were used to calculate domestic use, given the earlier assignments of production data. This is more complicated in the case of biodiesel, as only the 2008-2010 data correctly identify biodiesel. Therefore this information had to be extended backwards using PRIMES or Eurostat consumption data.

The consumption information used by CAPRI is in detail in the following table (compare *biofuel\prepare_biofuel_data.gms*):

Table 6 Overview on data sources utilized for biofuel consumption

Source	Variable	Period	Regions
COMEXT	Ethanol trade (HS220710+220720)	2000-2010	EU MS
COMEXT	Biodiesel trade (HS38249091)	2008-2010	EU MS
Eurostat	Fossil fuel demand quantities (Table nrg_102a)	2001-2008	EU MS
Eurostat	Biofuel demand quantities (Table nrg_1073a)	2005-2008	EU MS
PRIMES	Biodiesel demand	2000-2007	EU MS
F.O.Licht	Fuel ethanol demand	2003-2008	EU MS

Again, F.O.Licht data have also been used in part for some non-European countries as part of the global database (Section 2.5).

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 which are not recorded statistically. A few data exist for some countries (France, Germany, Sweden, Austria[20], Poland) and the EU[21]. From those sources it got obvious that in addition to the feedstock stated by EBIO also table wine is used for ethanol production, in particular in Spain and Sweden. However, after collecting different national data it was not possible to merge a complete ex post feedstock demand dataset such that the MS data had to be estimated. 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 *'pepare_biofuel_data.gms'*) which is assumed to be the main source to "cut out" the required biofuel processing quantities (BIOF) from market balances that so far did not include BIOF. The estimation tries to stay close to the initial feedstock composition while complying with upper bounds on palm oil use (where some data are given), technical conversion coefficients

[20] Österreichischer Biomasse-Verband (http://www.biomasseverband.at/biomasse?cid=4)

[21] http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Biofuels%20Annual_The%20Hague_EU-27_6-11-2010.pdf, http://www.ebio.org/product.php, extracted on 18.07.2009. Also AGLINK/COSIMO offers feedstock data on the EU level.

and minimising the corrections in the given market balance data. Thereby corrections of the total industrial use (for biofuels and for other uses) are penalised less that corrections of total demand which should only be the last resort for the solver to close the balances.

A special case was palm oil as the CAPRI database (COCO) doesn't cover an industrial use position for this product. 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.

- Import quantities - export quantities are equal to domestic consumption of palm oil as domestic production in European Member states can be neglected.
- 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 the quantities used for industrial processing could be derived which was assumed to be equal to processing.

Technology parameters

Conversion coefficients for 1st generation biofuels were collected from different sources. The PRIMES database includes conversion coefficients but only for the feedstock aggregates which are covert in the PRIMES - Biomass module (vegetable oils, sugar crops, starchy crops and corn)[22]. As CAPRI needs coefficients for individual agricultural crops or products these parameter values were used as a starting point but further sources had to be consulted. The AgLink-Cosimo model (version 2010 includes a set of conversion coefficients which are in line with the CAPRI product definition. To check these values further publications were taken into account[23] and the different sources were consolidated. Because the parameter values differ only in a small range the average values were calculated and implemented in the CAPRI data base. The table below displays the used set of conversion coefficients for 1st generation biofuels and corresponding by-products.

[22] Compare ANNEX2: PRIMES questionnaire

[23] i.a. "Market penetration of biodiesel and ethanol" Szulczyk, K. (2007). Available at http://agecon.tamu.edu/graduate/pdf/Szulczyk07a.pdf (18.07.2009)

Table 7 Table: 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**
Vegetable 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

* considered as molasses (1t vinasses=0,1 t molasses equivalent) depending on the reduced sugar content

Source: Own compilation base on AgLink database, PRIMES questionnaire and Szulczyk, K. (2007)

Prices

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 4.5), 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[24] 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[25] 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[26]. 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 recalculated. As the envisaged biofuel demand function will be a function of (among other variables) the

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

[25] 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)

[26] http://ec.europa.eu/energy/renewables/biofuels/ms_reports_dir_2003_30_en.htm

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).

2.2.3.6 Assigning data to database array

So far data processing has focussed on the key Eurostat Table Group “apro”. The next parts of COCO involve moving data computed for all Table Groups (“apro”, “apri”, “aact”) to a single GAMS array "data" for subsequent completion steps and consistency calculations.

Sub-module *‘coco1_crops.gms’*

This sub-module assigns the areas, crop production data and most market balance positions from Eurostat’s Table Group “apro” . This involves another overlay problem within “apro”: Usually there are two candidates for the production quantity, the value from production statistics and the value from the market balances (USAP), the latter receiving preference as it is consistent with the rest of the market balances.

However, before tackling the main tasks it is necessary to deal in this sub-module 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 and corresponding yield. For the EAA position the value of the production of cotton seeds had to be extracted from textile crops to preserve the usual identity that the production value is price times production quantity.

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 the given sub-components. This principle applies to "GRAS" (permanent grass land = PMEA+PPAS), "TAGR" (table grapes = VINY-TWIN), "ROOF" (fodder root crops = ROO1+ROO2[27]), and "OFAR" (other fodder on arable land = TGRA+FCLV+FLUC+FPGO+ FAGO). In the fodder sector there is often second best information on some aggregates, fodder on arable land ("FARA"), annual green fodder ("FANG") and perennial green fodder ("FPEG") which is only used if none of their components is available.

In terms of gross production (GROF) it has been mentioned already that preference is given to the market balance information ZPA1(USAP) over the production statistics ZPA1(GROF), where possible. Thus we set:

For products with market balance	DATA(GROF,t) = ZPA1(USAP,t+1)
Remaining products	DATA(GROF,t) = ZPA1(GROF,t)

[27] See Annex 1 for further detailed information on model codes.

Under some conditions gaps in the USAP series have been filled already (in coco1_finishZPA1) using the information on "GROF" from production statistics with a correction factor. Some special cases are the availability of 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.

Furthermore it is assumed for all products (except fodder) that all production is used for human consumption, where only domestic use is available from the balance sheets. On the other hand, domestic use is added up from its components, if missing. If still missing, it is calculated backwards from usable production, imports, exports and stock changes. Conversely, if production is still missing, an attempt is made to calculate this from domestic use and trade data.

In several cases upper or lower limits are assigned, 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) on a moving average over given observations.

Sub-module *'coco1_milk.gms'*

This sub-module 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), farm cheese, butter and other processing products (INDM.COMI) and finally losses and home consumption of liquid milk (LOSM.COMI). For WB countries, this disaggregation often had to be estimated with shares taken from the sister country aggregate (see Section 2.2.3.3).

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.

Sub-module *'coco1_anim.gms'*

Assigning herd size, process length, activity level, yield and production data often involves significant reaggregations from the slaughtering statistics.

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 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_zpa1.gms'* (Section 2.2.3.4), 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 (Section 2.2.6), 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.

Sub-module *'coco1_eaa.gms'*

In this sub-module EAA data from Eurostat's COSA domain are assigned, including unit values and, as a fall back option, selling prices from the Eurostat PRAG domain. For a number of aggregates special assignments are needed to obtain monetary values matching with the aggregates used elsewhere in COCO:

CERE CERE – PARI

INDU INDC – SUGB – PULS + OLIV

OILS OILC + OLIV

APPL APPS + PEAR + PEAC

FRUI FFRU + TROP

Kidney beans are a special case. This product has to be extracted from vegetables and added to pulses to be in line with the grouping in the market balances.

For all missing data at producer price, the values at basic price minus subsidies, plus taxes are taken. The rule, vice versa, is applied for missing data at basic price. The COSA domain offers, in addition, some values unrelated to the market balances, such as maintenance of machinery and buildings. They are copied to the "data" array without any adjustments.

Unit values at producer price are preferably calculated as a quotient from the value at producer price and the quantity as selected from the COSA domain. However some checks are used to discard grossly implausible unit values.

To serve as a fall back option for the EAA unit values, the prices from the PRAG domain are used (with a correction factor to acknowledge the typical differences between producer prices and selling prices). To take care of gaps, the time series are filled with the moving average approach. 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

Sub-module *'coco1_resid.gms'*

This sub-module 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.

Sub-module *'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 (± 100t 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[28].

Sub-module *'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[29] on typical grass yields in dry matter for 2002 in all EU-27 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 may thus be approximate. The yields for pasture, meadows and other fodder on arable land are adjusted accordingly.

Sub-module *'coco1_landuse.gms'*

This module 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 information, new rows are defined, indicating from which source the information on land use area is coming:

- **LEVZPA1** - Eurostat national land use data (Eurostat Table: "apro_cpp_luse"). 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

[28] 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.

[29] 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 , N° 070501/2005/422822/MAR/C1) with the participation of EuroCARE.

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
  i. Original Corine Land Cover (44 classes, aggregated to the NUTS2 level[30] obtained from JRC, Ispra for 1990, 2000, 2006. To link the Corine information to the CAPRI land use classes we used as an interim step a contingency table[31] from CLC to LUCAS categories prepared by Javier Gallego, JRC Ispra for the EU dataset in 2006. 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, see Table 46). 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.
  ii. 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. Currently the Corine data have been used by CAPRI in this transformed form. This resolved some problems but has created others. For example the transformed Corine data for Austria show some "marine waters", because in the EU contingency table a small percentage (7.4%) of the Corine class "watercourses" may be mapped to the LUCAS class of "Coastal water bodies". Misclassifications like this one may be avoided by using regionalised contingency tables which are under preparation. 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"). 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 LEVZPA1. Furthermore a few categories are missing (no inland waters, no other wooded land). However there are few alternative annual series avaialble to regionalise the national data in CAPREG.
- **LEVFAO** - Land use data from the resource FAOSTAT domain[32] 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.

---

30 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.

31 From http://agrienv.jrc.ec.europa.eu/publications/pdfs/LUCAS_CORINE.pdf. See also Javier Gallego, Catharina Bamps (2008) Using CORINE land cover and the point survey LUCAS for area estimation, International Journal of Applied Earth Observation and Geoinformation 10 pp 467–475.

32 See http://faostat3.fao.org/home/index.html#DOWNLOAD.

- **LEVLandCov** - Eurostat land cover data for 2009 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"[33]). 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.
- **LEVMcpfe** - Year 2007 version of 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 and 2005[34].
- **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). 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.

These sources each provide information on some "land use classes" (Table 45) 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 46).

Sub-module *'coco1_finish_raw.gms'*

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

- To avoid that seed quantities are estimated unbelievable high (often observed for residual items like other cereals), all seed quantities so far available are checked and limited to reasonable quantities, based on two criteria:
  - For those crops supplying seeds a list of standard values for seed quantities of kg/ha is included. This list was prepared from standard values published in "Faustzahlen für die Landwirtschaft"35 and average values of seed quantities of kg/ha from Eurostat and FAO statistics. A limit is calculated by multiplying the standard value with the area level.
  - A second (generous) limit is calculated assuming that seeds should not exceed 50% of production.

  From both criteria and the quantity of seeds coming from statistics the minimum is selected and used as initial value for the subsequent estimations.

---

33 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/.

34 http://www.foresteurope.org/filestore/foresteurope/Publications/pdf/state_of_europes_forests_2007.pdf. In November 2011 the year 2011 report has beome available, including year 2010 data, but the updated numbers have not yet been used in CAPRI.

35 KTBL, Faustzahlen für die Landwirtschaft, 13. Auflage

- For all non crop products producer prices are assigned from EAAP or PRIC
- For all products without area or herd size level, production values are deleted.
- For FEDM, HCOM, SEDF and SEDM lower and upper limits are introduced to limit yearly change in the subsequent estimation routines.

2.2.4 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, 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. Minimal corrections of given statistical data, if not in line with the accounting identities, had to be made.

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, Mittelhammer, Britz 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 ‘supports’ 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.

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:

$$
\begin{aligned}
\min_{y_{i,t}} \; & \sum_{i,t \in obs} wgt^{dat}\left((y_{i,t}-y_{i,t}^{dat})/abs(y_{i,t}^{trd}-y_{i,t}^{dat})\right)^2 \\
& + \sum_{i,t \notin obs} wgt^{ini}\left((y_{i,t}-y_{i,t}^{ini})/s_{i,t}\right)^2 \\
& + \sum_{i,t} wgt^{hp}\left(\left((y_{i,t+1}-y_{i,t})-(y_{i,t}-y_{i,t-1})\right)/s_{i,t}\right)^2 \\
& + \sum_{i,t} wgt^{up}\left(\left(\max(y_{i,t}^{up},y_{i,t})-y_{i,t}^{up}\right)\right)/abs(y_{i,t}^{up})\Big)^2 \\
& + \sum_{i,t} wgt^{lo}\left(\left(\min(y_{i,t}^{lo},y_{i,t})-y_{i,t}^{lo}\right)\right)/abs(y_{i,t}^{lo})\Big)^2 \\
s.t. & \\
& y_{i,t}^{LO} \le y_{i,t} \le y_{i,t}^{UP} \\
& \textit{Accounting identities defined on } y_{i,t} \\
& \textit{Identity of land use from different sources}
\end{aligned}
$$

Equation 4

where *i* represents the index of the elements to estimate (crop production activities or groups, herd sizes etc.), *t* stands for the year, wgt^x 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 = $\{(i,t) \mid y_{i,t}^{dat} \neq 0\}$, the set of data points with nonzero data

$y_{i,t}^{trd}$ = the trend value of an initial trend 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 $y^{lo}_{i,t}, y^{up}_{i,t}$. The 'hard' bounds $y^{LO}_{i,t}, y^{UP}_{i,t}$ 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 $y^{lo}_{i,t}, y^{up}_{i,t}$?)

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, NEIL, 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 avaialable 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:

$$PRCM_M \cdot \delta_{c,M} = \sum_i \left(NAGR_i - PRCM_i\right) \cdot \delta_{c,i}$$

Equation 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

For on farm use we have information on the rw milk used as well as (partly) on the product obtained on farm:

$$\left(INDM_M + HCOM_M\right)\cdot \delta_{c,M} = \sum_i FARM_i \cdot \phi_{c,i} \quad \text{Equation 6}$$

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

$$CORF_{c,i} \cdot \delta_{c,i} = \phi_{c,i} \quad \text{Equation 7}$$

where

CORF = ratio of on farm content to the standard content

and either CORF will be contrained to be the same for all i or we permit CORF $\neq$ 1 only for FRMI.

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

$$MAPR_i = NAGR_i + FARM_i \quad \text{Equation 8}$$

$$MAPR_i = HCOM_i + PRCM_i + FEDM_i + NTRD_i \quad \text{Equation 9}$$

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:

$$NAGR_i = \left(HCOM_i - FARM_i\right) + PRCM_i + FEDM_i + NTRD_i \quad \text{Equation 10}$$

The market balance for the raw milk looks as follows:

$$GROF_M = PRCM_M + HCOM_M + INDM_M + FEDM_M + LOSM_M \quad \text{Equation 11}$$

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:

Equation 12 $$MAPR_i' = NAGR_i$$

Equation 13 $$HCOM_i' = HCOM_i - FARM_i$$

Equation 14 $$HCOM_M' + FEDM_M' + LOSF_M' = HCOM_M + FEDM_M + LOSM_M + INDM_M$$

The first two transform the standard 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 contradic each other before the data consolidatuion. During the consoidation the following equation ensures the identity of land use areas among different sources (LEVCLC, LEVFAO etc):

```
*
*  ------ 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);
```

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 (analogous to Equation 59).

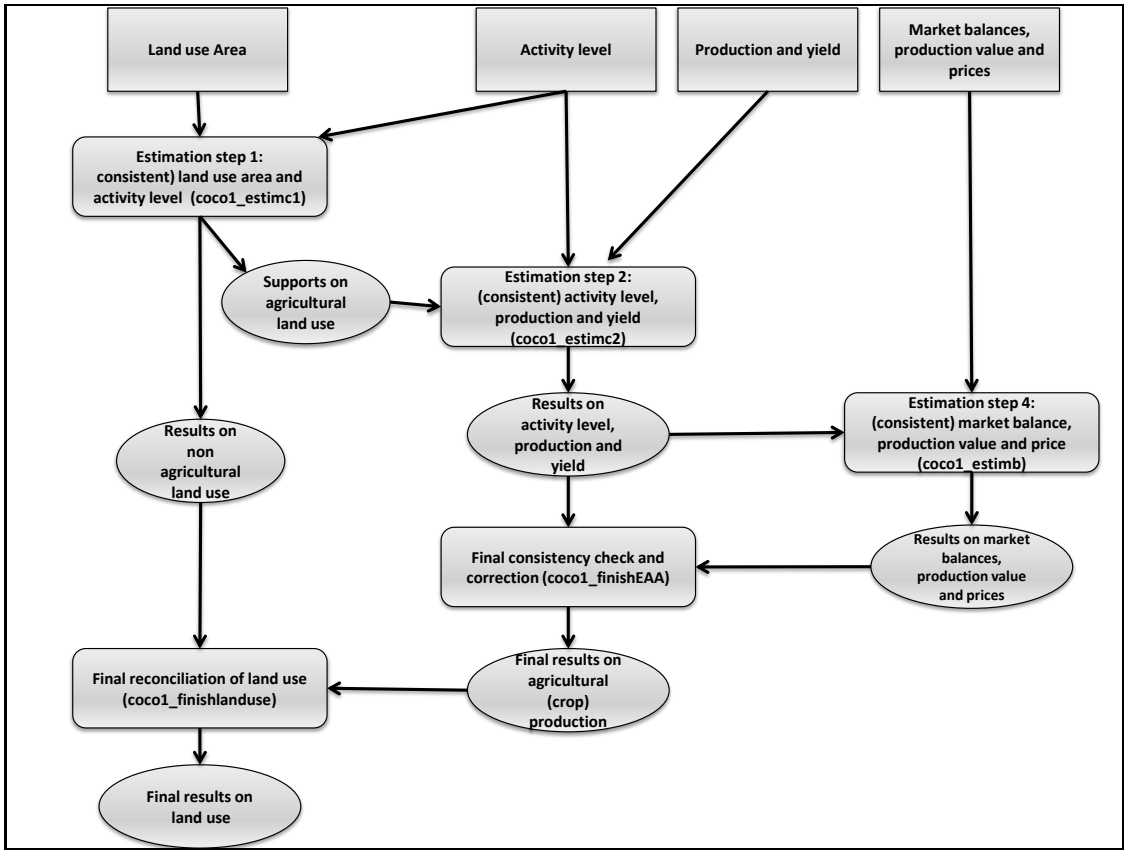
It should also be explained that Equation 1 is not applied simultaneously to the whole dataset because the optimisation would take too long. Instead it is applied to subsets of closely related variables:

1. Land use and land balance (Estimation step 1)
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 (1.) as given (Estimation step 4).

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. 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, set activity level zero if there is no EAA). This may slightly change the UAA as well and and 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. The crop sector data processing may be summarised as follows:

Figure 1. Overview on statistical estimators in COCO



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 (see Section 2.2.2). The longest series (1984-2010) can be established for EU15[36] 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 1989-2010 therefore. Norway offers reliable series as of 1984 but at the moment only up to 2007. 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 1995-2007 at this time (Mid 2010). Furthermore for the Western Balkan counties it was necessary to transfer certain coefficients and shares from (previously consolidated) neighbouring countries to the Western Balkan (see Section 2.2.3.3), such that a certain sequence is necessary for a reasonable application of COCO1:

- Run COCO1 for subsets of EU27 countries with the same coverage of years, either in one batch from the GUI or one by one (always with sub-steps 1 to 4).
- EU15 except Germany (1984-2010)
- EU12 plus Germany (1989-2010)
- Norway (1984-2007)
- 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 EU27 countries step 1 must be completed first.

2.2.5 COCO2: Data Preparation

Sub-module *'coco2_collect.gms'*

Data from various input files are collected and prepared to match to CAPRI definitions.

In a first step data collected Excel sheets are included:

- Eurostat: Exchange rates - National currency per Euro" (=1 if ILO data in Euro)
- Aggregate food price index with old Eurostat methodology and base 1985
- Eurostat: Private consumption of households in current prices
- Eurostat: population in 1000 hds
- Prices for milk products in selected countries
- Country Sheets of the Western Balkan and Turkey: Exchange rate, inhabitants, inflation rate, food expenditure shares

The market balance for cow and sheep milk is corrected for direct sales. DG Agri data on the quantity of direct sales of milk as required for the application of the quota system permits to disaggregate the Eurostat data on farm processing collected on industrial use (INDM) until May 2008 into a part that goes into direct sales and another part that is consumed on farm (and thus part of subsistence consumption). We apply this share also to sheep milk as direct sales of sheep milk are unknown and the degree of market integration could be similar.

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 derived from ZMP and AMI, using a conversion factor, where possible. Missing price data of all EU15 Member States are completed with the average ratio (available price information to EU15 price), or average difference if ratio > 1. For totally empty time series the EU15 price is used instead.

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

For non EU15 we assume that prices for butter and cheese change as prices for cow milk if at least one price is observed and for the other milk product, besides whole milk powder only, whey powder and casein, it is assumed that unknown national producer prices are related to raw milk prices as in EU15. Prices for whole milk powder are completed by prices of skimmed milk powder and butter, prices for why powder by prices f whole milk powder and prices for casein by prices for skimmed milk powder.

Data on expenditures are included from various sources as described in Chapter 2.2.2.5. UNSTAT data requires conversion to Euro using exchange rates mainly coming from Eurostat. Where missing, the exchange rates for the Western Balkan Countries and Turkey are completed from the Country Sheets. For still missing values, a last source is the UNSTAT site which sometimes gives longer series into the past as NC/USD.

The GDP price index (2000=100) is included from Eurostat domain AGGS. If missing for a Balkan country we use the inflation rates from the Country Sheets. If still missing the UNSTAT index (calculated to real expenditure per head in Euro of year 2000) is used.

Consumer expenditure weights and price indices for food aggregates (2005=100) are coming from Eurostat domain PRICE. Additional information for Albania, Bosnia and Croatia are from national agencies (see Section 2.2.2.5)

The most recent data on total private consumption of households in current prices are selected from Eurostat (see Section 2.2.2.5). If necessary, these data are backcasted by OLS on household consumption expenditure data from UNSTAT.

The first priority information on the number of inhabitants is selected from Eurostat domain, AUXIND. Missing data are calculated using conversion factors in the following sequence:

- Data for the Western Balkan and Turkey from the Country Sheets
- UNSTAT
- Eurostat, old domain AGRIS

Finally ILO consumer prices and price indices are prepared. The original selection from ILO is pre-processed in Excel. ILO prices are matched to CAPRI regions, products and units (*'coco2_ilo_addup.gms'*). The ILO food index is matched to CAPRI regions.

Price indices for food and non-alcoholic beverages from PRICE are completed first by the ILO price indices and for MS15 by price indices from old Eurostat domain FOOD (1985=100).

The consumer price indices (Eurostat PRICE) are converted from national currency into Euro. Furthermore they are rescaled to have 2001 = 100, because for year 2001 all price index data are available.

ILO prices for single typical food items (like "Wheat bread white unsliced not wrapped") are completed using a Hodrick-Prescott filter and the average regional EU015 and EU012 price with 1999-2003 total consumer expenditure (BRKDOWNS) as weights is calculated to replace missing observations.

Finally another HPD estimator is used to adjust ILOP relative price levels to those from Eurostat domain FOOD.

Sub-module *'coco2_shares.gms'*

Expenditure shares are defined and completed using simple OLS estimates.

For expenditure shares data are collected, starting with data from BRKDOWNS (2 and 3 digits) giving results on food and non-alcoholic beverages, food and non-alcoholic beverages. Domain HIST provides consumer expenditure on food items from old domain SEC2 up to 1997 and data for most recent years are coming from domain PRICE (Harmonized indices of consumer prices). For Germany (RWI and SBA) and Spain (AEA) national data are included.

The time series on food shares from BRKDOWNS are completed by OLS on data from the additional domains (HIST, PRICE) or finally only by trend, for Germany and Spain also with national data.

The food shares are computed using a conversion factor from domain PRICE to domain HIST on share ratios calculated as geometric mean of available overlapping years (1-3, depending on MS).

For potatoes the shares are extrapolated from domain HIST based on human consumption multiplied by producer price by OLS.

2.2.6 COCO2: Estimation procedure

The main problem for consumer prices is to distribute data for food expenditure on groups.

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.

The entropy problem is solved by maximizing

$$\max_{t}$$

$$-\sum_{m,j,k} CPS_{m,j,2} * HCOM_{m,j,t} / 1000 / TOFO_{m,t} * PE_{m,j,k} * LOG(PE_{m,j,k} / PQ_k)$$

$$-\sum_{m,j,k} CPS_{m,j,2} * HCOM_{m,j,t} / 1000 / TOFO_{m,t} * PED_{m,j,k} * LOG(PED_{m,j,k} / PQ_k)$$

$$-\sum_{m,FOPOS,k} EXS_{m,FOPOS,2} / TOFO_{m,t} * PEX_{m,FOPOS,k} * LOG(PEX_{m,FOPOS,k} / PQ_k)$$

$$-\sum_{m,j,k} PFAC_{m,k} * LOG(PFAC_{m,k} / PQ_k) * 1000$$

Equation 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 food expenditures 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

Respecting the equations:

Summing up probabilities for support points

Equation 16
$$\sum_{k\,\forall_{m,j}(CP.L_{m,j}\geq 0\,\wedge\, HCOM_{m,j,t}\geq o)} PE_{m,j,k} = 1$$

Equation 17
$$\sum_{k\,\forall_{m,j}(DCPS_{m,j}\geq 0\,\wedge\, HCOM_{m,j,t}\geq o)} PED_{m,j,k} = 1$$

Equation 18
$$\sum_{k\,\forall_{m,j} EX.L_{m,FOPOS}\geq 0} PEX_{m,FOPOS,k} = 1$$

Equation 19
$$\sum_{k\,\forall_{m,} TOFAC.LO_m \geq TOFAC.UP_m} PFAC_{m,k} = 1$$

Define consumer price changes from support points

Equation 20
$$DCP_{m,j} = \sum_{k\,\forall_{m,j}(CP.L_{m,j}\geq 0\,\wedge\, HCOM_{m,j,t}\geq o\,\wedge\, \mathrm{DCPS}_{m,j,2}\geq 0)} PED_{m,j,k} * DCPS_{m,j,k}$$

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

Equation 21
$$DCP_{m,j} = UVAD_{m,j,t_1} - CP_{m,j}$$

Define consumer prices from support points and probabilities

$$CP_{m,j} = \sum_{k \forall_{m,j}(CP.L_{m,j} \geq 0 \wedge HCOM_{m,j,t} \geq o)} PE_{m,j,k} * CPS_{m,j,k}$$

Equation 22

Define group expenditure from support points and probabilities

$$EX_{m,FOPOS} = \sum_{k \forall_{m,FOPOS}(EX_{m,FOPOS} \geq 0)} PEX_{m,FOPOS,k} * EXS_{m,FOPOS,k}$$

Equation 23

Define total expenditure slack from support points and probabilities

$$TOFAC_{m,} = \sum_{k \forall_{m,}(TOFAC.LO_m \geq TOFAC.UP_m)} PFAC_{m,k} * TOFACS_{m,}$$

Equation 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.

$$\sum_{FOPOS} EX_{m,FOPOS} = TOFO_{m,t} * TOFAC_{m,k}$$

Equation 25

Consistency of group expenditure

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

Equation 26

After solving this entropy problem the database is complete in terms of consumer prices and expenditures for all food items.

2.2.7 COCO2: Final completions

2.2.7.1 Completion of feed related data

For by-products of the milling and the brewing industry and for corn gluten feed, sugarbeet pulp, manioc and fish meal the database is completed for market balance positions production, imports, exports and feed. Available data from Eurostat FEEDAGRI is completed by additional input data from different sources, as described in Chapter 2.2.2.5. Finally completed data for all detailed by-products are aggregated to the CAPRI rows FENI (Rich energy fodder imported) and FPRI (Rich protein fodder imported). 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.

Sub-modules *'feed_declnew.gms'* and *'reqfncnew.gms'*

All of the included modules mentioned so far were only active when the national database is being established, i.e. when running the COCO parts. By contrast we are now discussing modules that are active in COCO2, in CAPREG, and in the baseline calibration of CAPMOD. As the name should signal *'feed_declnew.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 (*'reqfncnew.gms'*) specifies the requirement functions for animal activities of CAPRI. The file name reqfnc**new** refers to the adjustments in 2011 when animal herds (survey information on stocks of male cattle in the December counting, for example) were included in COCO1. This allowed to *estimate* the process length as

DAYS = HERD/LEVL*365 (see coco1_anim.gms) based on statistical data. Earlier versions of CAPRI considered the process length entirely engineering data and thus were likely (or known) to be in conflict with statistical data.

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

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_declnew.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 *'reqfncnew.gms'*:

Equation 27

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

The daily increase is then used to determine the process length (rearrangement of Equation 64 below with empty days EDAYS = 0)

Equation 28

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

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

Equation 29

$$meanWgt_r^{HEIF} = startWgt_{HEIF} + \frac{dailyIncrease_r^{HEIF} \cdot fatngdays_r^{HEIF}}{2}$$

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

Equation 30

$$NEM_r^{HEIF} = \left(meanWgt_r^{HEIF}\right)^{0.75} \cdot 0.322 \cdot fatngdays_r^{HEIF}$$

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

Sub-modules *'coco2_gras.gms'*

With animal requirements specified the results of COCO1 for grass and fodder from arable land net yields and contents of energy and protein depending on requirements of total herd sizes are revised to ensure reasonable starting values for feed allocation in CAPREG.

2.2.7.2 Complete prices for vegetable oil

The prices for vegetable oils relevant for biofuel processing functions are calculated using prices from FAPRI. These assignments refer to prices at the wholesale level (relevant for the processing industry), not to consumer prices which have been determined previously.

Data on total fuel demand are copied to the COCO database. These data results from data of total fuel demand from Primes scaled by Aglink

2.3 The Regionalised Data Base (CAPREG, W. Britz)

2.3.1 Data requirements at 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, 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.

2.3.2 Data sources at regional level

Already during the first CAPRI meeting, the REGIO domain of EUROSTAT was judged as the only harmonized data source available on regionalized agricultural data in the EU. REGIO is one of several parts of NEWCRONOS and is itself broken down in domains, one of which covers agricultural and forestry statistics.

In the agricultural and forestry domain [AGRI] the following tables are available:

- Land use [A2LAND]
- Crop production - harvested areas, production and yields [A2CROPS]
- Animal production - livestock numbers [A2ANIMAL]
- Cows's milk collection - deliveries to dairies, % fat content [A2MILK]
- Agricultural accounts on regional level [A2ACCT]
- Structure of agricultural holdings [A2STRUC, A3STRUC]
- Labour force of agricultural holdings [A2WORK]

2.3.3 Data availability at regional level

The following table shows the official availability of the different tables of REGIO. 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 8 Official data availability in REGIO

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
Cows's milk collection (deliveries to dairies, % fat content)	from 1977 yearly
Agricultural accounts on regional level	from 1980 yearly
Structure of agricultural holdings	1983, 1985, 1987, 1989/91, 1993
Labour force of agricultural holdings	from 1983 yearly

Source: Eurostat (http://epp.eurostat.cec.eu.int)

2.3.4 Reading and storing the original REGIO data

The original REGIO data are stored in an ASCII-format designed by EUROSTAT for NEWCRONOS and used in connection with the CUB-X, EUROSTAT's data browser. The data can be browsed and extracted to several formats directly with CUB-X (one table each time). However, in the case of the CAPRI-project, data from several tables must be merged

together, adding up to some million numbers. CUB-X was never designed for such quantities. Therefore, the group in Bonn designed a tool called DFTCON which converts these files into a rather simple format:

– In a first step, these files are sorted by region, year and original code, so that they can be easily accessed by other software to perform extraction from the original NEWCRONOS data base.

– In a second step these files are converted into GAMS tables which are then stored in GDX format. The input files are stored in "dat\capreg" and under version control. Meta data are added currently still manually to those files.

The results of these two steps are tables, typically per Member States, which comprise time series of all data retrieved from the REGIO tables: land use, crop production, animal populations, cow's milk collection and agricultural accounts.

2.3.5 Methodological proceeding

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[37].

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

- *Data enter the consistency checks as found in REGIO.* This is mainly true for animal herd sizes where REGIO offers data at the same or even more disaggregated level as found in COCO.
- *Gaps in REGIO are filled out and data found in REGIO at a higher aggregation level as required in CAPRI are broken down by using existing national information.*
- *Functions used 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 made available to overcome the list of these 'easy-to-use' estimates presented in here. 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.

2.3.6 Prices for outputs and inputs

The agricultural domain of REGIO does not cover regionalized prices. For simplicity, the regional prices are therefore assumed to be identical to sectoral ones[38]:

Equation 31 $$UVAG_r = UVAG_s$$

[37] Numbers such as grassland areas and yields which are considered quite uncertain also at the national level, are sometimes permitted to change to prevent infeasibility of the data consolidation task, for example in the set-aside and feed trimming problems, see below.

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

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.

2.3.6.1 Young animal prices

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.

2.3.6.2 Sugar beet prices (Marcel Adenaeuer)

Sugar beet prices 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 noteworthy that the same program is applied in CAPREG and in CAPMOD to determine base year beet prices.

2.3.7 Filling gaps in REGIO

In cases where data in REGIO on regional activity levels are missing, a linear trend line is estimated for regional and Member State time series in REGIO definition. 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.

2.3.8 Mapping crop areas and herd sizes from REGIO to COCO definitions

Only some few crop activities are available in REGIO (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 REGIO. Therefore, the regionalized data are directly set to the values in REGIO. The difference between the sum of these 4 activities and the aggregate data on cereals in REGIO 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, the difference from REGIO is broken down applying national shares.

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

$$\text{Equation 32} \qquad \begin{aligned} \text{LEVL}_{\text{OATS,r}} = &\left(\text{CEREAL}_{\text{r}}\right. \\ &\left. - \text{WHEAT}_{\text{r}} - \text{BARLEY}_{\text{r}} - \text{MAIZEGR}_{\text{r}} - \text{RICE}_{\text{r}}\right) \cdot \\ &\frac{\text{LEVL}_{\text{OATS,COCO}}}{\left(\text{LEVL}_{\text{OATS,COCO}} + \text{LEVL}_{\text{RYE,COCO}} + \text{LEVL}_{\text{OCER,COCO}}\right)} \end{aligned}$$

Similar equations are used to break down other aggregates and residual areas in REGIO[39].

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. REGIOs break down for herd sizes is 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.

In order to generate good starting points and avoid systematic deviation between regional and national levels and following consistency steps, all regional level in REGIO are first multiplied with the relation between the results in COCO and the REGIO results at national level.

2.3.9 Perfect aggregation between regional and national data for activity levels

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.

Consistency for activity levels is based on Highest Posterior Density Estimator which ensures:

1. Adding up of activity levels from lower regional level (NUTS II, NUTSI) to higher ones (NUTSI, NUTS0)
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. Deviations from the given UAA receive a very high weight.

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[40]. Equally, the share of oilseeds grown as energy crops on set-aside needs to be determined.

An Highest Posterior density estimator is used 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

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

[40] 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.

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-food 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.

The proceeding 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:

Equation 33

$$CORR_{GROF,o} = \sum_{r,j} Levl_{j,r} O_{j,r} \Big/ GROF_{o,n}$$

$$O^{*}_{j,r} = O_{j,r} * CORR_{GROF.o}$$

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.

2.3.10 Outlier detection for yields

....

2.3.11 Estimating expected yields with a Hodrick-Prescott filter

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:

Equation 34

$$\min \quad hp = 1000 \sum_{1<t<T-1} \left(y^{*}_{t+1} - y^{*}_{t-1} \right)^2 + \sum_{t} \left(y^{*}_{t} - y_{t} \right)^2$$

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.

2.4 Input Allocation (W. Britz)

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 a simple fixed coefficient approach for ammonia losses. These balances ex-post determine the effective input coefficients based on a cross-entropy estimation framework.
- *For feed*, the input calculation is rooted in a mix of engineering knowledge (requirement functions for animal activities, nutrient content of feeding stuff), observed data ex-post (total national feed use, national feed costs) and estimated feed costs from a FADN sample, combined within a Highest Posterior Density (HPD) estimation framework.
- *For the remaining inputs*, estimation results from a FADN sample are combined with aggregate national input demand reported in the EAA and standard gross margin estimations, again using a HPD estimation framework.

2.4.1 Input allocation excluding young animals, fertiliser and feed

2.4.1.1 Background

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.

2.4.1.2 Econometric Estimation

Standard econometric methods were employed to calculate input coefficients from single farm records found in FADN (within a consistent aggregation framework, as explained in Chapter 5). Raw data were transformed into CAPRI compatible categories. Fixed-Effects, Random Effects, Weighted Fixed-Effects, and Weighted Random-Effects as well as OLS and WLS models were tested with varying degrees of success. After finding heteroskedasticity problems, deciding to neglect from using an intercept (in order to conform to the Leontief technology assumed by the model) and after comparing results for plausibility, it was decided that a straightforward WLS model was the most suitable form if a consistent estimation technique was to be used for all estimations. The main reason for choosing such a simple WLS estimator over a weighted random effects model with no 'fixed effect' intercept was the question of plausibility of results. Specification tests suggested, in fact, that fixed effects estimators might have been used in every regression, but apart from the problem of distributing farm specific fixed effect intercepts across crop and animal activities, there were two (related) reasons not to use these results. Firstly, the results of the fixed effects specifications –on the whole- were implausible, with a large number of negative coefficients. Secondly, it was felt that any possible endogeneity in the estimations would probably have a greater proportionate effect in the fixed effects results. The weight actually used in the final WLS versions was total output.

Initial experiments also revealed a high degree of multicollinearity if activity levels and outputs were both used on the right hand side. Accordingly, it was decided to use output on the right hand side if possible (so that regional variations could be incorporated into the model). Where sufficient output values were not available, activity levels were used, using the criterion described below. Furthermore, because of a clearly deleterious effect on results, the equivalents of the CAPRI residual activity categories OCRO (other crops), OFRU (other fruits), OCER (other cereals), OVEG (other vegetables), etc. were all dropped from the estimations.

All regressions were run using STATA 7.0. Price indices were taken from the COCO database in order to calculate input costs in real terms. The starting sample sizes were, as follows, all multiplied by 10 (for the years 1990-1999) unless otherwise stated:

•AT - Austria - 2451 farms
•BL - Belgium, 2601 farms
•DE - Germany, 15110 farms --> price data from '91-'99
•DK - Denmark, 6625 farms
•EL - Greece, 11877 farms --> price data from '95-'99

•FI - Finland, 1324 farms
•IR - Ireland, 3409 farms --> no price data prior to 1995
•IT - Italy, 57264 farms
•PT - Portugal, 6379 farms
•SE - Sweden, 1191 farms
•UK - United Kingdom, 6668 farms
•ES - Spain, 22609 farms
•NL – Netherlands, 3565 farms
•FR – France, 17332 farms

The following data cleaning procedures were used:

- The regressors with less than or equal to 100 observations for both activity levels and output were excluded.
- The data were truncated at zero in order to eliminate reported negative level and output values and also reported negative real input costs.
- All non-zero values were counted and a choice made between either activity level or output, as the appropriate right-hand side variable (only one could be use to avoid multicollinearity).
- An activity's *output* value was used if the number of non-zero output values associated with that activity was greater than the number of the activity's non-zero levels minus 500. Thus, output was always the preferred option unless levels were reported for at least 500 more observations than outputs. This procedure was necessary because of a number of cases in the data when only output or activity level values but not both.

Several regressions were run to yield estimates for coefficients in each of 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.

2.4.1.3 Reconciliation of Inputs, using Highest Posterior Density Estimators

As a result of the unrestricted estimation based on FADN, a matrix of input coefficients 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). The table below presents a sample of the results from the econometric regressions. These are the output (GROF) coefficients of 2 activities, soft wheat and barley, for 4 input categories; total inputs (TOIN), total other inputs (TOIX), crop only inputs (COSC), and fertiliser (FERT). All coefficients are statistically significant except those in red.

Table 9 Sample of soft wheat and barley production coefficients for 4 inputs (1995 prices)

GROF Coef.	AT	BL	DE	DK	EL	ES	FI	FR	IR	IT	NL	PT	SE	UK
S. Wheat														
TOIN	214.22	152.79	135.37	192.92	197.24	104.67	231.66	138.88	136.94	194.47	154.15	341.65	0.00	140.91
TOIX	160.85	86.18	90.92	148.00	116.05	60.15	162.51	80.28	83.70	125.86	100.27	238.65	0.00	86.86
COSC	49.27	49.60	61.61	40.69	78.05	49.06	61.05	63.04	51.58	60.76	50.65	109.21	0.00	54.33

FERT	21.00	17.71	21.59	19.45	35.98	25.03	41.74	26.49	20.58	26.36	14.37	57.24	0.00	19.39
Barley														
TOIN	184.03	184.74	204.03	0.00	210.21	113.49	183.27	173.23	131.03	266.92	179.64	168.95	158.99	205.53
TOIX	131.26	110.17	133.50	0.00	121.68	67.88	106.87	80.16	63.38	178.87	128.77	109.11	92.98	107.86
COSC	52.49	73.53	74.00	0.00	54.13	48.57	68.96	78.81	73.80	65.94	60.24	52.04	48.08	82.59
FERT	23.49	36.69	32.42	0.00	30.99	29.40	45.62	42.99	33.36	30.11	17.12	29.32	20.36	42.85

Source: input estimation, CAPRI modelling system

For example, the 'TOIN' coefficient for soft wheat in Austria reveals that on average it costs an Austrian farmer 214.22 € to produce an extra tonne of wheat. These coefficients should reveal a reasonable sense of cross-country comparative advantage among activities.

In table 9, the coefficients of variation for soft wheat for 'TOIN', 'TOIX', 'COSC', and 'FERT' were 34 %, 41 %, 29 % and 44 % respectively. Those for barley were 21 %, 29 %, 19 %, and 27 % respectively. Thus, a high degree of variation for 'TOIX' and 'FERT' is clear in this sample. This gives an indication of the general variability underlying the estimated coefficients.

All of the econometric coefficients were required to be transformed into an 'activity level' form, due to 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.[41] 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 in the cross entropy approach described below. [42]

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 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.[43]

2.4.1.4 Highest Posterior Density estimation framework

Given the three types of prior information explained above –estimated input coefficients, data from EAA and standard gross margins-, the choice of a HPD Estimator to reconcile estimated input coefficients seemed to be convenient.[44] The estimation was carried out for all CAPRI activities (z) -excluding activities that where split up like DCOW into DCOL and DCOH-, and a number of inputs in CAPRI (denoted by $X_{CI,z}$) and FADN ($X_{FI,z}$) definition. The list of input definitions can be found in the annex (table 44).

---

[41] 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.

[42] 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 table 42 or the appendix).

[43] 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.

[44] The advantage of cross entropy is that one can define the support space rather wide and give the edges a very low prior probability.

For each prior we defined 4 support points (k) centred on the value of the priors defined as above. The support range was defined as follows:

- For the econometric estimates:

  $S_{XFi,z,k}\ P_{XFi,z} + [-100; -1; 1; 100]\ \sigma_{XFi,z}$,

where $S_{XF,I,z,k}$ gives the support points for the FADN input $X_{FI,z}$ that has a standard error of σ_{XFi}.

- For the EAA priors: prior *(1+ [-10; -0.1; 0.1; 10]).

  $S_{XCi,z,k}\ P_{XCi,z}\ (1+ [-10; -0.1; 0.1; 10])$,

where $S_{XCi,z,k}$ gives the support points for the CAPRI input $X_{CI,z}$. A special treatment was chosen for the total input coefficient. Here the support range was half that from above.

- For the standard gross margins:

  $S_{GM,z,k}\ P_{GM,z}\ (1+ [-10; -0.1; 0.1; 10])$,

where $S_{XCi,k}$ gives the support points for the standard gross margin of activity z.

We define the a priori probability for each support point to b

e:

$AP_k = [0.002; 0.49; 0.49; 0.002]$,

in order to give the outermost support points less weight. Posterior probabilities are denoted by PP.

The model setup is then given by:

$$\max H(PP_{CI,Z,K}, PP_{FI,Z,K}, PP_{GM,Z,K}) =$$

$$-\left(\sum_{CI,FI,Z,K}\left[PP_{CI,Z,K} * \ln\left[\frac{PP_{CI,Z,K}}{AP_K}\right] + PP_{FI,Z,K} * \ln\left(\frac{PP_{FI,Z,K}}{AP_K}\right) + PP_{GM,Z,K} * \ln\left(\frac{PP_{FI,Z,K}}{AP_K}\right)\right]\right)$$

$$s.t.$$

$$\sum_k PP_{CI,Z,K} = 1,\ \sum_k PP_{FI,Z,K} = 1,\ \sum_k PP_{GM,Z,K} = 1$$

$$X_{CI,Z} = \sum_k PP_{CI,Z,K} S_{CI,Z,K},\ X_{FI,Z} = \sum_k PP_{FI,Z,K} S_{FI,Z,K},\ GM_Z = \sum_k PP_{GM,Z,K} S_{GM,Z,K}$$

$$GM_Z = EREV_Z - \sum_{CI \in G1(CI,Z)} X_{CI,Z} - X_{exo,Z}$$

$$EAA_{CI} = \sum_{Z \in G1(CI,Z)} X_{CI,Z} LEVL_Z$$

$$\sum_{CI \in G2(CI,FI)} X_{CI,Z} = X_{FI,Z}$$

$$\sum_{FI \in G3(CI,FI)} X_{FI,Z} = X_{CI,Z}$$

$$\sum_{FI1 \in G4(FI,FI1)} X_{FI1,Z} = X_{FI,Z}$$

Equation 35

The first two rows of the equation shown above are subject to maximize cross entropy, while the third row guaranties that all probabilities sum up to unity. In the fourth row, the estimates for input coefficients and gross margins are re-parameterized from the posterior probabilities and the support points. The fifth row defines gross margins for an activity z as the difference between expected revenue per activity level (EREV) of that activity and the sum over all

inputs used in that activity. The Set G1(CI,Z) allocates the inputs used to each activity and $X_{exo,Z}$ are inputs, that are not estimated here, but cannot be neglected in defining gross margins (like young animal inputs). In the sixth row, we find a statement which guarantees 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 seventh and eighth rows link the inputs in the CAPRI definition to those in FADN definition. The first of those two are used when the FADN inputs are an aggregate of CAPRI inputs (defined in the set G2(CI,FI)) or they have the same definition and the second one when CAPRI inputs are an aggregate of FADN inputs. Since estimated inputs in the FADN definition exist for aggregates and components of them, we ensure in the last line that the sum over FADN inputs that belong to an aggregated FADN input (defined in the set G4 (FI,FI1)) sum up to the latter.

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.

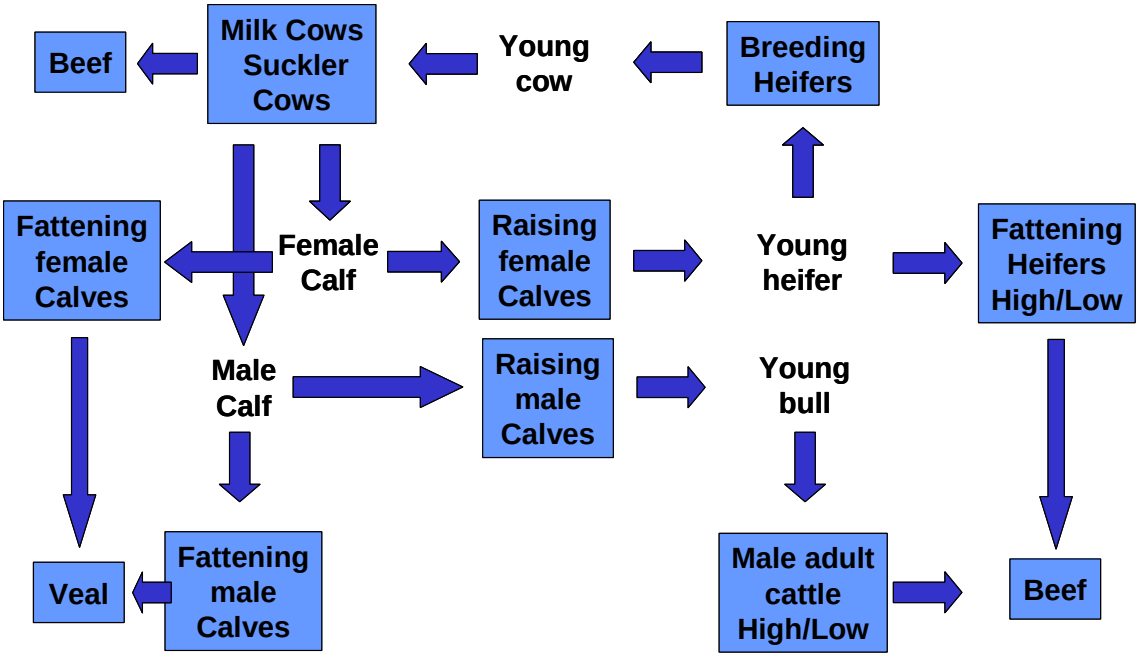
2.4.1.5 How are the results used in CAPRI?

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 the latter four types can directly be used in the CAPRI model, we need special treatments for the other types –e.g. fertilisers, because they are used in physical units inside the model, and feeds, since they are much more disaggregated. Therefore, the estimated results will go to other parts in the regionalisation. The costs for feeds go into the feed trimming, where animal requirements are brought into equilibrium with the contents of the feeding stuff as supports. A similar thing could be done with the fertiliser costs in the fertiliser trimming.

2.4.2 Input allocation for young animals and the herd flow model

Figure 2 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 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 the relationships.

Figure 2. The cattle chain



Source: CAPRI Modelling System

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 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.

Figure 3. Example for the relation inside the cattle chain (Denmark, 1999-2001)

		1999	2000	2001
Male calves used in t and born in t				
DCOWLEVL	Number of dairy cows	667.03	654.08	631.92
DCOWYCAM	Number of male calves born per 1000 dairy cows	420.72	438.62	438.26
Number of males calves born from dairy cows		*280.63*	*286.89*	*276.95*
SCOWLEVL	Number of suckler cows	127.36	126.91	124.85
SCOWYCAM	Number of male calves born per 1000 suckler cows	420.72	411.83	401.61
Number of male calves born from suckler cows		*53.58*	*52.27*	*50.14*
Number of all male calves born		334.22	339.16	327.09
GROFYCAM	Number of male calves produced	334.21	339.16	327.09
CAMFLEVL	Number of male calves fattened	81.32	72.57	49.18
CAMRLEVL	Activity level of the male calves raising process	252.89	266.59	277.91
Sum of processes using male calves		334.21	339.16	327.09
GROFYCAM	Number of male calves used	334.21	339.16	327.09
Female calves used in t and born in t				
DCOWLEVL	Number of dairy cows	667.03	654.08	631.92
DCOWYCAF	Number of female calves born per 1000 dairy cows	404.15	421.58	412.86
Number of female calves born from dairy cows		*269.58*	*275.75*	*260.89*
SCOWLEVL	Number of suckler cows	127.36	126.91	124.85
SCOWYCAF	Number of male calves born per 1000 suckler cows	404.15	398.04	387.21
Number of female calves born from suckler cows		*51.47*	*50.52*	*48.34*
Number of all female calves born		321.05	326.26	309.24
GROFYCAF	Number of female calves produced	321.05	326.27	309.24
CAFFLEVL	Number of female calves fattened	26.64	28.74	18.39
CAFRLEVL	Activity level of the female calves raising process	294.41	297.53	290.85
Female calves used in t and born in t		321.05	326.27	309.24
GROFYCAF	Number of female calves used	321.05	326.27	309.24
Young bulls used in t and young bulls produced in t				
BULFLEVL	Activity level of the bull fattening process	262.94	252.89	266.59
GROFIBUL	Number of young bulls used	262.94	252.89	266.59
GROFYBUL	Number of young bulls raised from calvs	252.89	266.59	277.91
CAMRLEVL	Activity level of the male calves raising process	252.89	266.59	277.91
Heifers used in t and heifers produced in t				
HEIFLEVL	Activity level of the heifers fattening process	64.36	67.25	68.12
HEIRLEVL	Activity level of the heifers raising process	235.45	227.16	229.4
Sum of heifer processes		299.81	294.41	297.52
GROFIHEI	Number of heifers used	299.81	294.41	297.53
GROFYHEI	Number of heifers raised from calves	294.41	297.53	290.85
CAFRLEVL	Activity level of the female calves raising process	294.41	297.53	290.85
Cows used in t and heifers produced in t				
DCOWLEVL	Number of dairy cows	667.03	654.08	631.92
DCOWICOW	Number of young cows needed per 1000 dairy cows	332.01	332.5	327.52
Sum of young cows needed for the dairy cow herd		*221.46*	*217.48*	*206.97*
DCOWSLGH	Slaugthered dairy cows	221.47	217.48	206.11
SCOWLEVL	Number of suckler cows	127.36	126.91	124.85
SCOWICOW	Number of young cows needed per 1000 suckler cows	332.01	332.48	327.52
Sum of young cows needed for the suckler cow herd		*42.28*	*42.20*	*40.89*
SCOWSLGH	Slaugthered suckler cows	42.29	42.19	40.72
Sum of slaughtered cows		*263.76*	*259.67*	*246.83*
GROFICOW	Number of young cows used	263.75	259.67	247.86
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*	
Sum of slaughtered cows and stock change			235.45	
GROFYCOW	Numer of heifers raised to young cows	235.45	227.16	229.4
HEIRLEVL	Activity level of the heifers raising process	235.45	227.16	229.4

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 10 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

2.4.3 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[45].

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.

2.4.3.1 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.

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:

Equation 36
$$\sum_{r,\text{fint}} \overline{FEDUSE}_{r,\text{fint}} PFOD_{r,\text{fint}} = \overline{EAAP}_{OFAR,MS} + \overline{EAAP}_{GRAS,MS}$$

Secondly, the Gross Value Added of the fodder activities is defined as the difference between revenues and total input costs based on the input allocation for crops described above

Equation 37
$$GVAM_{r,\text{fint}} = \overline{YIELD}_{r,\text{fint}} PFOD_{r,\text{fint}} - \overline{TOIN}_{r,\text{fint}}$$

Next, the standard ingredients of a cross entropy estimator are added: definition of the estimated values from supports and the posterior probabilities, summing up of the posterior probabilities to unity and the definition of the cross entropy itself

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

Equation 38

$$\sum_k \sup_{r,\text{fint},gvam,k} p_{r,\text{fint},gvam,k} = GVAM_{r,\text{fint}}$$

$$\sum_k \sup_{r,\text{fint},price,k} p_{r,\text{fint},price,k} = PFOD_{r,\text{fint}}$$

$$\sum_k p_{r,\text{fint},gvam,k} = 1$$

$$\sum_k p_{r,\text{fint},price,k} = 1$$

$$H(PROB) = -\sum_{r,\text{fint},price,k} p_{r,\text{fint},price,k} \log(p_{r,\text{fint},price,k} / pq_k) - \sum_{r,\text{fint},price,k} p_{r,\text{fint},gvam,k} \log(p_{r,\text{fint},gvam,k} / pq_k)$$

The a priori mean 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. The outer supports are set so that the higher support is at four times the a priori mean.

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 rather wide bounds. The a priori probabilities for the three supports are set at 1 %, 98 % and 1 %.

The 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:

Equation 39

$$\overline{YIELD}_{r,\text{fint}} PFOD_{r,\text{fint}} \geq \overline{TOIN}_{r,\text{fint}} \overline{gvafac}$$

The parameter *gvafac* is initialised with unity so that first a solution is tried where all activities have revenues exceed costs. If infeasibilities arise, the factor is stepwise reduced until feasibility is achieved, to ensure that the minimal number of activities with negative Gross Value Addeds is estimated.

2.4.3.2 Update note

Whereas the estimation of fodder prices is explained in some detail the allocation of feed to activities is only covered very briefly above. In the 2006/07 Ammonia study CAPRI was used to investigate the impacts of low protein feed which led to some overhaul of the feed allocation. The problem was that nutrient intake was sometimes implausibly exceeding the requirements from the literature. A certain luxury consumption is perfectly plausible, just reflecting that observed data usually do not meet the high efficiency laboratory situations in the literature. Nonetheless a measured excess of 50% and more was considered troublesome. A number of remedies have been introduced therefore in the Ammonia study to reduce the number of odd cases:

- Grass and other fodder yields have been estimated (in COCO already) as a compromise of statistical and expert information (from Alterra, O. Oenema, G. Velthof)
- Losses of straw have been permitted to vary according to the surplus situation in the region
- A luxury consumption embedded in the sectoral data on feed input and animal products has been steered mainly towards the less intensive (sheep, cattle) activities

Remaining problems

- In some countries there is still a surprisingly high national excess consumption of protein which is presumably due to erroneous raw data. However some of these data constellations persist even after communication with statistical offices (*very* high fish meal consumption in Denmark).

Additional revisions have occurred in the feed module after the ammonia study:

- Allowing for settings better tailored to animal types rendered the feeding ratios more stable.
- The objective function in the last calibration step (after determining feed prices) has been transformed from an Entropy to a HPD estimator.

2.4.4 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.

So far, the link between instruments of agri-environmental instruments and pressure indicators had been explored for the case of greenhouse gas emissions (Pérez 2005). During the first phase of CAPRI (1996-1999), NPK balances and output of greenhouse gases had been introduced, and an energy use indicator was explored for Switzerland. The project for DG-ENV (2001-2002) then led to (1) the improvement of the current state indicators -especially ammonia output and nitrate leaching, (2) the introduction of new ones as a water balances and chemical indicators, (3) feasibility studies for the application of the Nutrient Flow Model for the Netherlands and the bio-physical model CropSyst for regions in France, and (4) improving the interpretation of environmental indicators by contrasting them with soil and land-use maps. The following table shows an overview of the indicators embedded in the CAPRI system after the finalisation of the DG-ENV project.

Table 11 Indicators in the CAPRI system

Indicator	Linked to	Fixed at	Source/Comment
NPK output at tail	Regional animal population and yields (final weights, milk yield, length of production period)	Animal type	Farm management literature, operationally embedded in system
Ammonia emissions	Animal population, housing & storage type, crop level & yields	Member state level	IASSA, prototype embedded; Nutrient Flow Model (LEI, Netherlands)
NPK losses by leaching and soil storage	NPK output at tail and ammonia emission, N-crop need	EU level	Operational, currently with old emission factors
Output of greenhouse gases (nitrous oxide, methane)	Animal herds, mineral fertiliser	Uniform coefficients per animal type and pure mineral nutrient for EU	Counter-check with European Environmental Agency, IPCC rules
Water balances	Meteorology, management, irrigation, soil	Regional coefficients per crop activity	CropWat model, partial counter-check with CropSyst model
Nitrate concentrations in ground water	soil type, ground water level, nitrogen surpluses	Region, crops and farm types	Case studies for the Netherlands and France
Chemical emissions	crop production	Regional coefficients per crop activity	Case studies for the Netherlands and France

Source: CAPRI modelling system

2.4.4.1 Nutrient balances for NPK

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.
- Output of manure at tail –depending on animal type, regional animal population and animal yields, as final weights or milk yields.
- Input of mineral fertiliser –as given from national statistics at sectoral level.
- The Ammonia emission model (see Section 2.4.4.3)

2.4.4.2 NPK output at tail

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

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

Source: Lufa von Weser-Ems, Stand April 1990, Naehrstoffanfall.

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 12 Additional emission of NPK per kg of milk produced

N	0.0084
P	0.004
K	0.0047

Source: RAUMIS Model (http://www.agp.uni-bonn.de/agpo/rsrch/raumis_e.htm).

The factors shown above for pigs are converted into a per day and live weight factor for sows by assuming a production of 5 m³ 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 m³ 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 m³ 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 1997), 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 13 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*

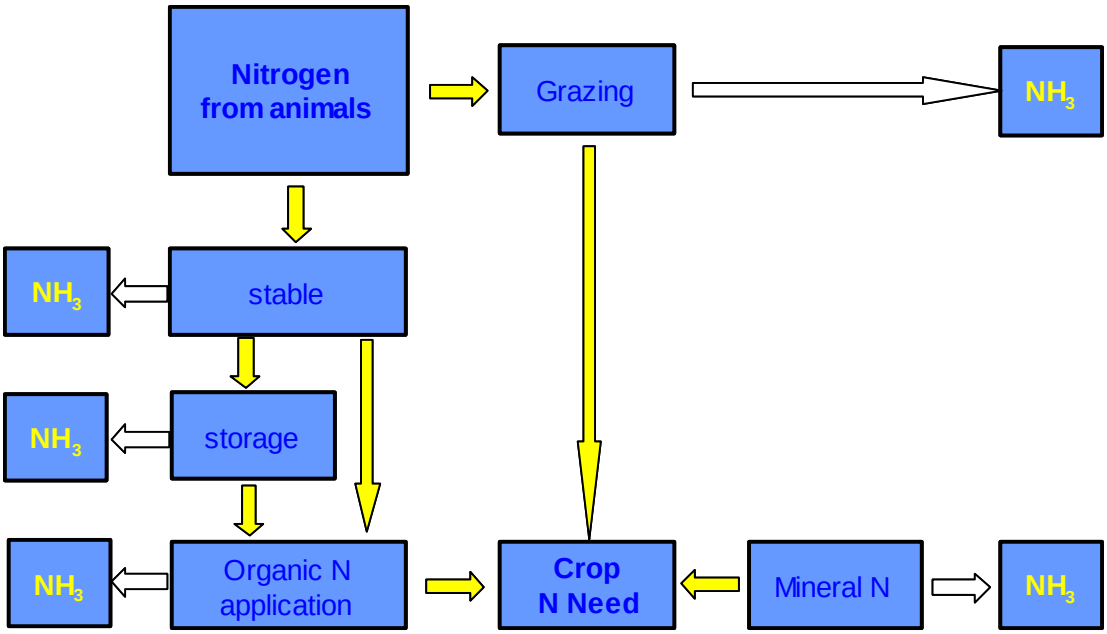
Source: CAPRI Modelling System

The coefficients in the previous table would be different for the new database and Gams code. There have been important updates of the underlying requirement functions for crude protein in the meantime but the basic ideas are still valid.

2.4.4.3 The ammonia module

The ammonia (NH_3) 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 IASSA and has been updated under the Ammonia project in 2006/07. The following diagram shows the NH_3 sinks taken into account by coefficients.

Figure 4. Ammonia sinks in the Ammonia emission module



Source: CAPRI modelling system

In Figure 4, 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 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 IASSA 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 IASSA 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.The following table is still based on the older methodology and coefficients but nonetheless provides a useful illustration of the accounting.

Table 14 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 losses from manure fallen on grazings	g	2.08
Nitrogen from biological fixation*	c	2.89	Nitrogen in ammonia losses from manure in stable	h	7.13
Nitrogen from atmospheric deposition	d	14.36	Nitrogen in ammonia losses from manure storage	i	2.53
			Nitrogen in ammonia losses from manure application on the field	j	8.34
			Nitrogen in ammonia losses from organic fertiliser	k=g+h+i+j	20.08
			Nitrogen in ammonia 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**

Source: CAPRI modelling system

2.4.4.4 Input allocation of organic and inorganic NPK and the nutrient balance

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. Firstly, nutrient export by the harvested material is determined, based on the following factors:

Table 15 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

Source: CAPRI modelling system

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:

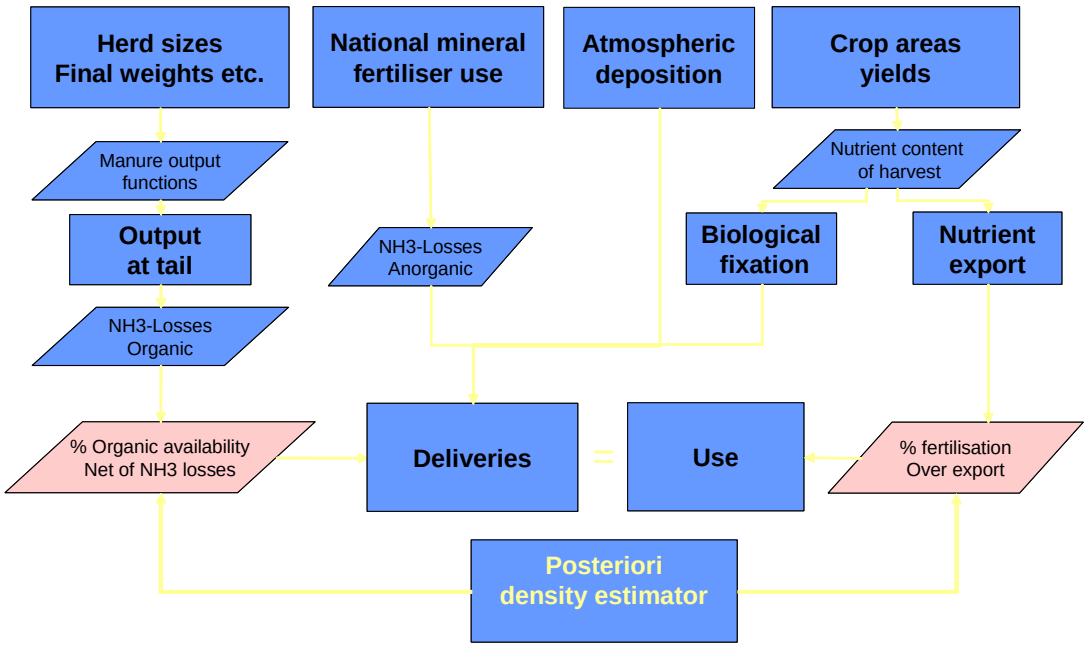
Table 16 Atmospheric deposition of N per kg and year

Austria	20
Belgium	32
Denmark	18
Finland	5
France	16
Germany	29
United Kingdom	15
Greece	7
Irland	10
Italy	12
Netherlands	36
Norway	5
Portugal	3
Spain	6
Sweden	5
Switzerland	18

Source: CAPRI modelling system

Figure 5 offers a graphical representation of these relationships.

Figure 5. 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:

Equation 40
$$\overline{Nettrd}_{MS}^{Fnut} = \sum_{r} Nettrd_{r}^{Fnut}$$

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.

Equation 41
$$\begin{aligned}
&\sum_{cact} Levl_{r,cact} Fnut_{r,cact} \left(1 - NFact_{Fnut,cact}^{biofix}\right) \\
&NutFac_{r,fnut} \left(1 + NutFacG_{r,fnut} \wedge cact \in ofar, grae, grai\right) \\
&= NETTRD_{r}^{Fnut} (1 - NH3Loss_{Fnut,r}^{Anog}) \\
&+ NBal_{r}^{AtmDep} NFact_{Cact}^{AtmDep} \\
&+ \sum_{aact} Levl_{r,aact} Fnut_{r,aact} (1 - NH3Loss_{Fnut,r}^{Manure}) \left(1 - NavFac_{r,fnut}\right)
\end{aligned}$$

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:

Equation 42
$$\begin{aligned}
\min \quad HDP = &- \sum_{r,fnut} \left(\frac{NutFac_{r,fnut} - \mu_{r,fnut}^{NutFac}}{\sigma_{r,fnut}^{NutFac}} \right)^2 \\
&- \sum_{r,fnut} \left(\frac{NavFac_{r,fnut} - \mu_{r,fnut}^{NavFac}}{\sigma_{r,fnut}^{NavFac}} \right)^2 \\
&- \sum_{r,fnut} \left(\frac{NutFacG_{r,fnut} - \mu_{r,fnut}^{NutFacG}}{\sigma_{r,fnut}^{NutFac}} \right)^2 \\
&- \sum_{r,ngrp} \left(\frac{Nitm_{r,ngrp} - \mu_{r,ngrp}^{Nitm}}{\sigma_{r,ngrp}^{NavFac}} \right)^2 \frac{\overline{LEVL}_{r,UAAR}}{\overline{LEVL}_{r,ngrp}}
\end{aligned}$$

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 17 Elements entering the of NPK balance ex-post and ex-ante

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

Source: CAPRI modelling system

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.

2.4.4.4.1 Update note

The overall N Balance calibration problem has been revised several times, the last time in 2007. Among other improvements it now delivers estimates of the shares of different sources of N (mineral fertiliser, excretions, crop residues) distinguished by crop groups.

2.4.4.5 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, two types of pollutants are modelled: methane (CH_4) and nitrous oxide (N_2O). The sources considered are: *CH_4 emissions from animal productio*n, *manure management and rice cultivation* and *N_2O from agricultural soils and manure management*[46].

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, 2000). The following table lists the emission sources modelled:

Table 18 Agricultural greenhouse gas emission sources included in the model

Greenhouse Gas	Emission source	Code
Methane	Enteric fermentation	CH4Ent
	Manure management	CH4Man
	Rice production	CH4Ric
Nitrous Oxide	Manure management	N2OMan
	Manure excretion on grazings	N2OGra
	Emissions from synthetic fertiliser	N2OSyn
	Emissions from organic animal waste	N2OWas
	Emissions from fertiliser application	N2OApp
	Emissions from crop residues	N2OCro
	Emissions from nitrogen-fixing crops	N2OFix
	Indirect emissions from ammonia losses	N2OAmm
	Emissions from atmospheric deposition	N2ODep

Source: CAPRI Modelling System

[46] Carbon sinks are not included since the measurement of carbon dioxide absorption through agricultural biomass is highly complex (high uncertainty involved, especially in agricultural soils) and has strong linkages with other economic activities not considered in this analysis, such as bio-diesel production and forestry management.

For a detailed analysis of these single emission sources refer to Pérez 2005.

The model code also comprises a life-cycle assessment for GHGs from livestock, for details see: Weiss, F. and Leip. A.,: Greenhouse gas emissions from the EU livestock sector: A life cycle assessment carried out with the CAPRI model, Agriculture, Ecosystems and Environment, Vol. 149 (March 2012), pp. 124-134, doi:10.1016/j.agee.2011.12.015 and Leip A., Weiss F., Wassenaar T., Perez I., Fellmann T., Loudjani P., Tubiello F., Grandgirard D., Monni S. and Biala K. (2010). Evaluation of the livestock sector's contribution to the EU greenhouse gas emissions (GGELS) -final report, European Commission, Joint Research Centre (http://ec.europa.eu/agriculture/analysis/external/livestock-gas/index_en.htm).

2.4.4.5.1 Update note

Accounting for gaseous emissions has been updated several times (but the brief remarks from above still apply). It now relies on more recent IPCC recommendations. Further changes are possible because N2O accounting in the framework of the nitrogen balance (derived from Miterra which in turn relied on IPCC) and accounting of greenhouse gases (directly derived from IPCC) may benefit from an alignment.

2.4.5 Input allocation for labour (Markus Kempen, Eoghan Garvey)

With the decline of the importance of agriculture in all Member States, there is concern as to the consequences for on-farm employment. With farming populations falling steadily, the increase in agricultural income per unit of labour is primarily a result of labour productivity. Within the EU, there is a marked difference in farm structures between Northern and Southern countries, with the average size of holdings much smaller in the latter than in the former. Economic factors will also play a role in the future structure of farms, as demographics will. To this end, we utilise a Cohort Analysis approach which allows for a separate, complementary analysis of both the demographic and economic trends as they affect the number of farm holders. This is useful for CAPRI in a number of ways – providing baseline figures for the number of farmers in future years, helping in the calculation of income per capita figures and linking CAPRI with the wider economic changes in EU regions.

Also labour input demands are estimated for CAPRI. The term input allocation describes how aggregate input demand (e.g. total family or paid labour) is ‘distributed’ to production activities. The resulting activity specific data are called input coefficients. In general, they may either be measured in value (€) or physical terms (hours). The CAPRI data base generally uses physical terms and, where not available, input coefficients are measured in constant prices. In our estimations we have estimated input coefficients for labour in hours (both paid labour and family labour) and we have estimated wage payments in constant (1995) euro.

Labour (and other inputs) in CAPRI are estimated from a Farm Accounting Data Network (FADN) sample 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.

The scientific relevance is that for the first time there will be available a set of EU wide labour coefficients for family and paid farm labour, using a standardised database. This has not been available heretofore. It is important that these results are plausible, and bear some relation to the known engineering coefficients (usually calculated on the basis of ‘best practice’). Hence, the constant revision of results is important.

The societal relevance is the existence of plausible labour coefficients enables calculates of employment effect within the sector following on from policy changes or from the simple passage of time. The work on cohort analysis also enables a link to be made between on and

off-farm regional changes. This is extremely important if we are to have idea of the time allocation effects on farm households of policy changes.

2.4.5.1 Labour Input Allocation

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 were the only quantitative instruments available. 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 analysing results and may be used in turn to define sectoral income. Similarly, important environmental indicators are linked to some input uses and can hence be linked to activities as well with the help of input coefficients. Important income, employment and social indicators can be linked to the coefficients reported on this deliverable.

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. Additionally, tests for a more complex estimation framework building upon entropy techniques and Bayesian and integrating restrictions derived from cost minimization were run in parallel.

In some cases estimates revealed zero or negative labour input coefficients, which cannot be taken over into CAPRI. Accordingly, it was decided to set up a second stage estimation framework building upon the unrestricted estimates from FADN. This is described below.

Econometric Estimation

Standard econometric methods are employed to calculate labour input coefficients from single farm records found in FADN. At a first stage, raw data were transformed into CAPRI compatible categories. Different kind of panel models, such as Fixed-Effects, Random Effects, Weighted Fixed-Effects, and Weighted Random-Effects as well as OLS and WLS models were tested with varying degrees of success.

The starting point for the building of our statistical model is to treat the unobserved variable as "unobserved heterogeneity" or individual effect that varies only across farms and not over time. As a result, it follows that all behavioral differences between individual farms are captured by the intercept. Examples of this heterogeneity, in our case, could be the average quality of land depending highly on soil quality, the managerial quality of family running the farm and other unobserved time-constant factors.

In our models the unit specific component is initially included in the error term. Furthermore, by adopting the fixed effect model (which all the statistical tests suggest is the correct model, for the weighted data), we allow for the unobserved fixed effect to be correlated with the explanatory variables, level (ha) and the interaction variable level multiplied by maximum yield or herd size (ha or heads*tones/ha). Hence, we regard that for example management ability or soil quality may be correlated with the maximum yield of the farm or the decision of how many hectares will be attributed to every production activity.

Main model:

$$\text{Equation 43} \quad Input_{it} = \beta_{1i} + \sum_{k=1}^{53} \beta_{2kr} Level_{iktr} + \sum_{k=1}^{53} \beta_{3k} \max(yield)_{ikt} * Level_{ikt} + u_{it}$$

Benchmark model:

Equation 44 $$Input_{it} = \beta_{1i} + \sum_{k=1}^{53} \beta_{2kr} Level_{iktr} + u_{it}$$

Two types of specification were considered, as reported above. One with the level variable and the interaction term and a second one with one regressor (level) which is used as a benchmark model. We should note that maximum yield or herd sizes is chosen as part of the interaction term because it is considered a reliable proxy for the expected yield, as this is anticipated in the decision making of the farmer to use any particular input. Regional variations are incorporated by using activity level on the right hand side at the NUTS I, NUTS II levels accordingly with the compatibility of FADN and NUTS administrative regions. In addition, we should remark that the interaction term is included at the national level apart from the case of Italy, Spain, France, and Germany where it is at NUTS II and NUTS I level, for the last one, respectively.
Furthermore, because of a clearly deleterious effect on results, the equivalents of the CAPRI residual activity categories OCRO (other crops), OFRU (other fruits), OCER (other cereals), OVEG (other vegetables), etc. were all dropped from the estimations.

As previously mentioned, the data for the input demand estimations is the FADN dataset for the EU 15 from 1989 to 2001. Sample sizes vary from country to country (Italy, for example, has over 200,000 observations, while most countries have about 15,000-50,000). On average each particular farm appears 5 times in the 13 year panel.

Several regressions were run to yield estimates for coefficients in each of 24 input categories available (not just labour: these other input coefficients MAY also be useful for CAPRI and it was felt to be worthwhile to combine them with the labour estimations) : Total Inputs, Crop Specific Inputs, Animal Specific Inputs, Seeds, Plant Protection, Fertilizer, Repair, Energy, Agricultural Services, Depreciation, Compensation of Employees, Other Taxes on Production, Other Inputs, Other Crop Inputs, Purchased and Non-Purchased Feeds, Other Animal Inputs, Water, Rent, Interest Paid, Electricity, Fuels, Wages, hours of Paid and Family Labour.

Regional Reconciliation

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 designed by the Bonn team has to be undertaken to correct figures for labour input by adjusting the labour input coefficients for both total labour and family labour. 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 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.

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

0 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.

2.4.5.2 Projecting Labour Supply using Cohort Analysis

In terms of CAPRI, regional projections of the number of holders facilitate a more accurate welfare analysis in terms of regional income per 'capita' (i.e. farmer). Holder projections can be linked to the Galway team's work on labour input estimations in terms of adding plausibility to forecasts of changes in labour requirements. Therefore, the cohort analysis will broaden CAPRI's range of policy analysis and enhance its existing capabilities.

Cohort Model

Changes in farm structure over time can be separated in to 2 components: (1) an autonomous component, which comprises of 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. Thus, cohort analysis tires to dichotomise the effects of the components so as to simplify the econometric analysis of changes in the structure of farm holdings.

As mentioned, Cohort Analysis is used to separate autonomous changes in the structure of land holdings from non-autonomous changes. The cohorts are the holders of land divided in to different age groups. Steel and gallney (1998) offer the following example:

Equation 45 $C_{25}(1993)=C_{24}(1992)_{n24\text{-}25}\text{-}NA_{24\text{-}25}$

Where C_{25}= Cohort aged 25

C_{24}= Cohort aged 24

$N_{24\text{-}25}$= Probability of survival for 24 year olds in 1992

$NA_{24\text{-}25}$= Net non-autonomous change of cohort size 1992-1993

Net non-autonomous changes are those arising from farmers' decisions to leave farming and join other labour markets, or vice versa. These factors will vary from region to region and from year to year. The autonomous factors comprise of demographic changes such as the death rate or the probability of permanent disability.

Cohort analysis will estimate the expected size of a particular size state, given these autonomous factors. The difference between the estimated and observed size of a cohort, the 'residual', is the net non-autonomous change of the cohort size. The second stage of this approach involves the analysis of this residual econometrically. Using the number of holders as an approximation for total agricultural labour can be problematic. Thus, we examine the relationship between the number of holders and total agricultural labour input. It is examined for 11 member states, yielding the following correlations. It can be seen that the approximation is reasonable acceptable for all countries except Italy, where the large number of very small farms means that the number of holders diverges considerably from the amount of agricultural labour.

Table 19 Correlations between Holder Numbers and Agricultural Labour

Country	Correlation Coefficient
Germany	0.929342
Italy	0.486949
Ireland	0.947002
France	0.991825

Spain	0.941472
Greece	0.895167
Denmark	0.979908
Belgium	0.993277
United Kingdom	0.917235
Netherlands	0.932555
Luxembourg	0.993772

While accepting that this is a less than ideal approximation, we proceed pending the availability of more appropriate data. The probability of survival in any year in a region is calculated using crude death rates and population statistics obtained from Eurostat. The data is then divided in to 5-year cohorts. Linear interpolation is used to disaggregate the age distribution for holders at each age. The midpoints of each age group for which data is available is found and then connected with straight-line segments. These lines give a particular value for the number of holders at each age.

Autonomous changes are forecast by multiplying the number of holders for each age by the probability of survival for each age. New entrants were allowed for by creating a cohort of holders aged between 12 and 24. We base our ex-post analysis of the cohort results on the following assumptions:

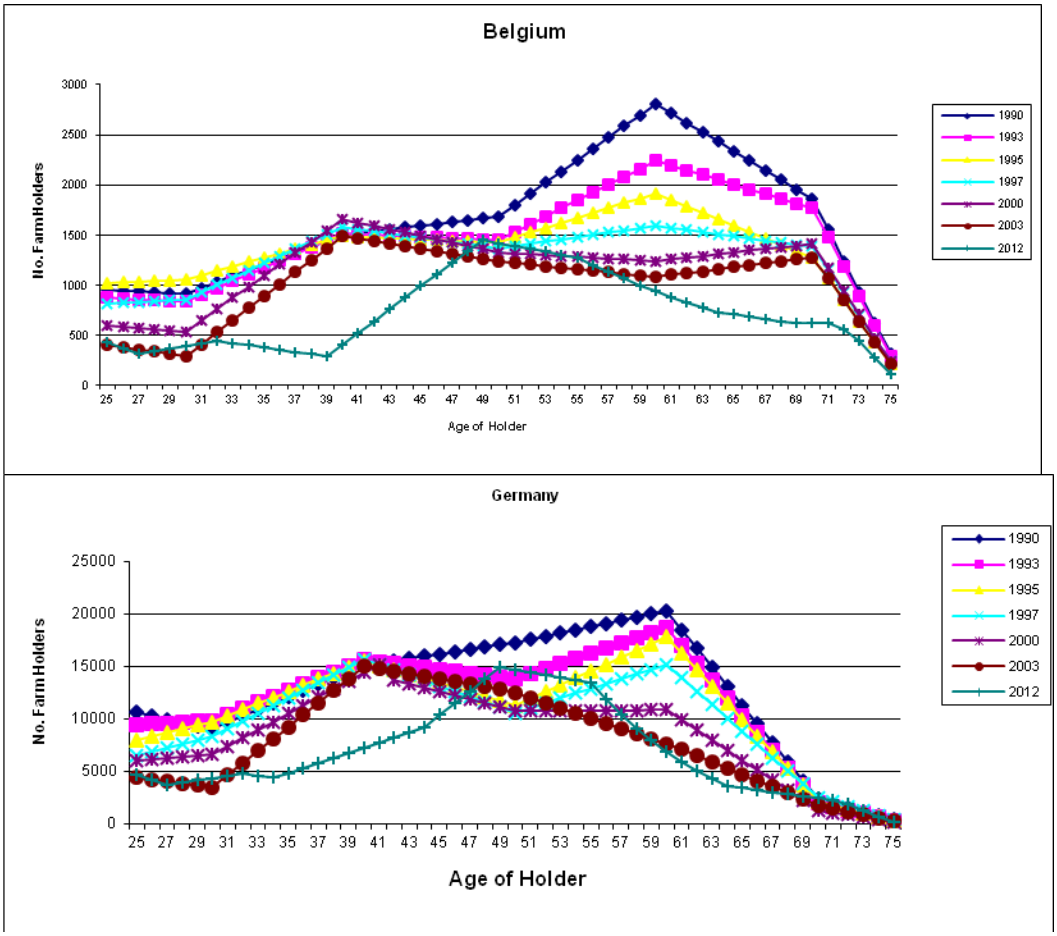
- Death and disability: disability occurs through physical invalidity or by reaching the age of 75 years.
- Occupational mobility is approximated by the sum net changes in cohorts 25-55, when these changes were not due to death or disability. Those, who release their agricultural holdings, are assumed to do so in order to undertake alternative employment. We accept that they are not allowing for the possibility of interregional movements of holders.
- Holders in the 55-75 cohort who release their holdings are assumed to retire, as people in this age group are less likely to seek alternative employment.

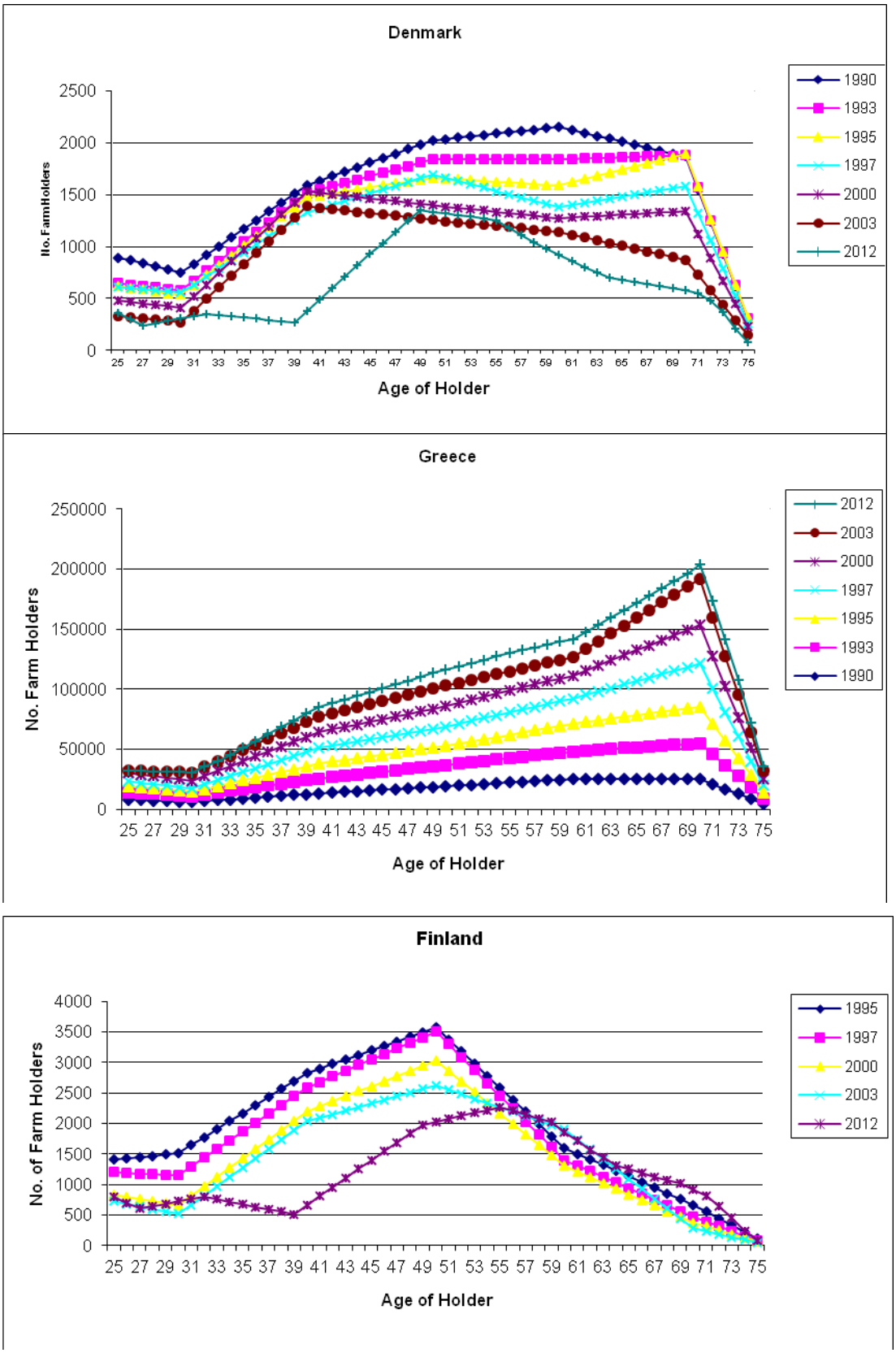
To supplement the analysis, we construct 'age-lines' that illustrate the estimated number of holders in a country at each age. We incorporate two additional procedures to Steele and Gaffney (1998); (1) in order to allow for new entrants, the ratio of holders from 10-14, 15-19, and 20-24 to the total population in those age groups in each region was assumed to equal the ratio of farm holders of 25-34 year olds and total population in that age group for each region., (2) age-lines are in some cases adjusted so that from 10-34 year olds, no age-lines sloped downwards. This was achieved by constructing support points (0.1, 0.1, 0.1, 3.7) multiplied by the frequency at each age (the linear interpolation method described above. The entropy function was maximised subject to the restriction that the frequency at each age must be lower than that for the group one year older.

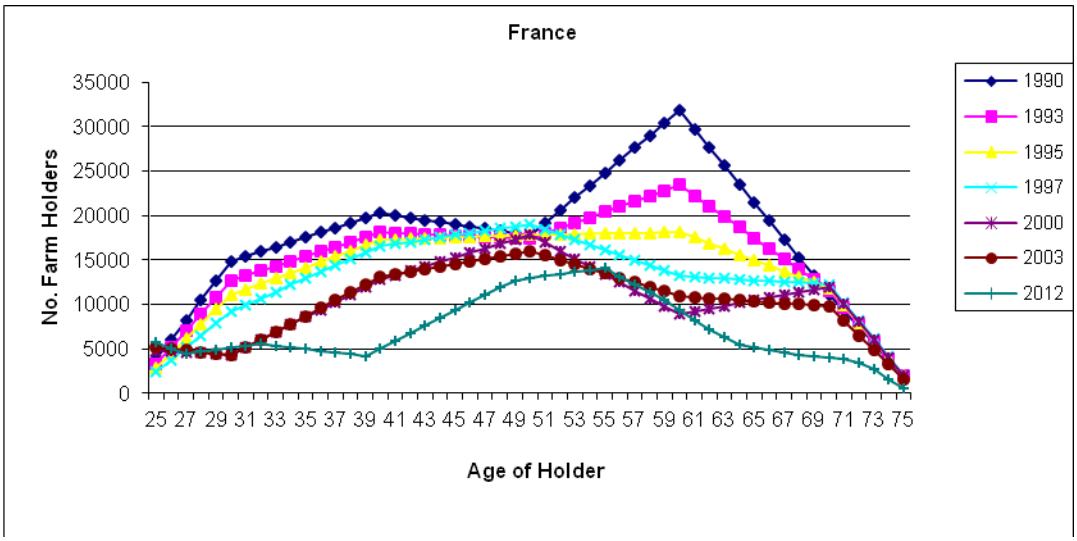
In making projections for 2012, we are required to make additional assumptions. Trends for 10-54 year olds are calculated based on demographic changes using 1993 death rates. For the 55-75 cohort it is assumed that the annual rate of retirement (or exit for non-demographic reasons) is the same as the annualised average of rates for 55-75 age group between 1987 and 1993. This exit rate is then spread across the individual years by a factor multiplied by the death rate for each year. These factors range from 0.995 for 55 year olds to 0.9 to 70 year

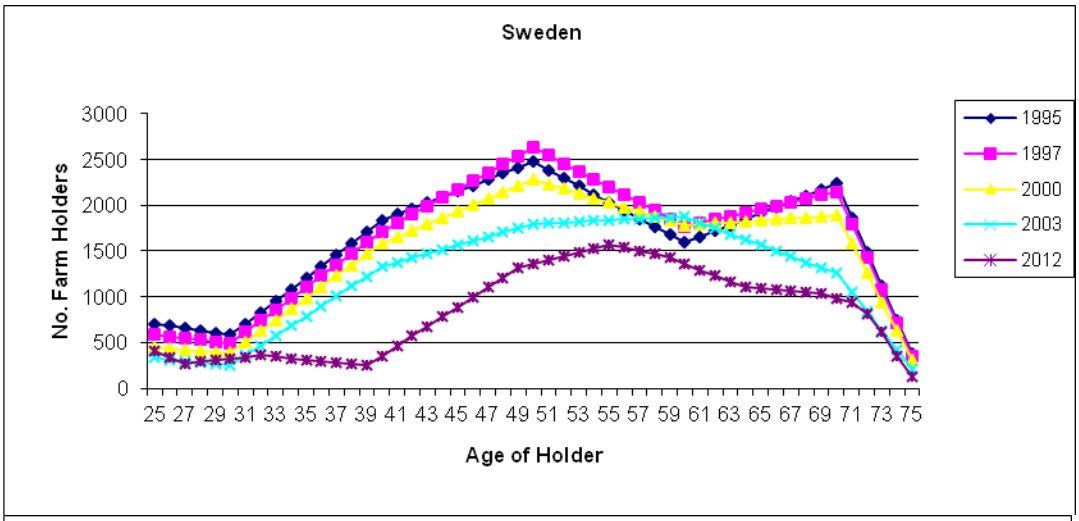
olds. The projection for individual years is adjusted across the 21 years to ensure that the sum of the projection for the individual years is equal to the projection for the cohort as a whole. Figure 6 shows composite agelines for 10 EU member states (excluding Germany and Luxembourg) for 7 years. The movement towards a more normal distribution of holders relative to their age is clear from the Figure 6.

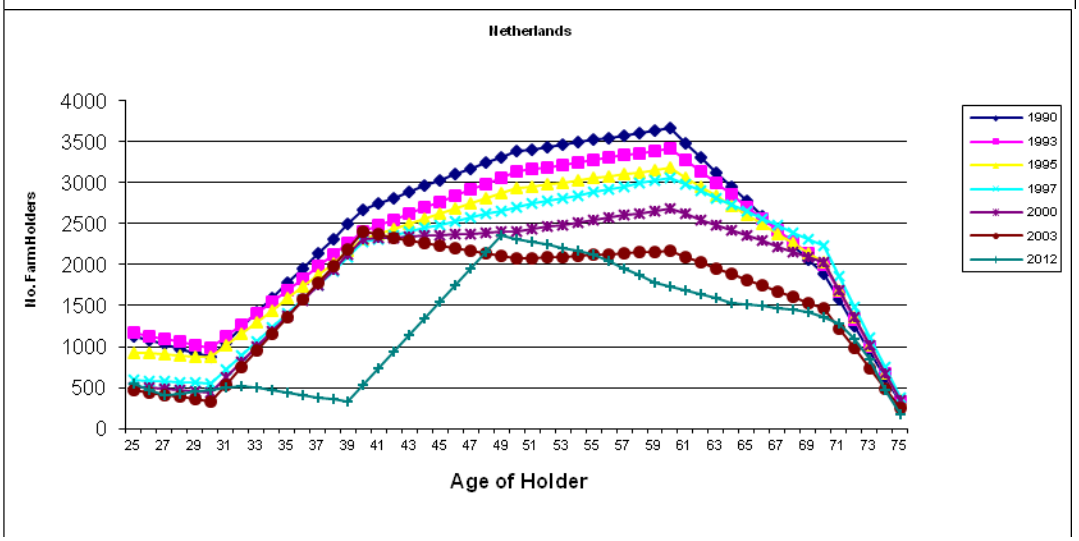
Figure 6. Agelines: Farm Stuctural Change 1990-2003 and Forecasted Change in 2012

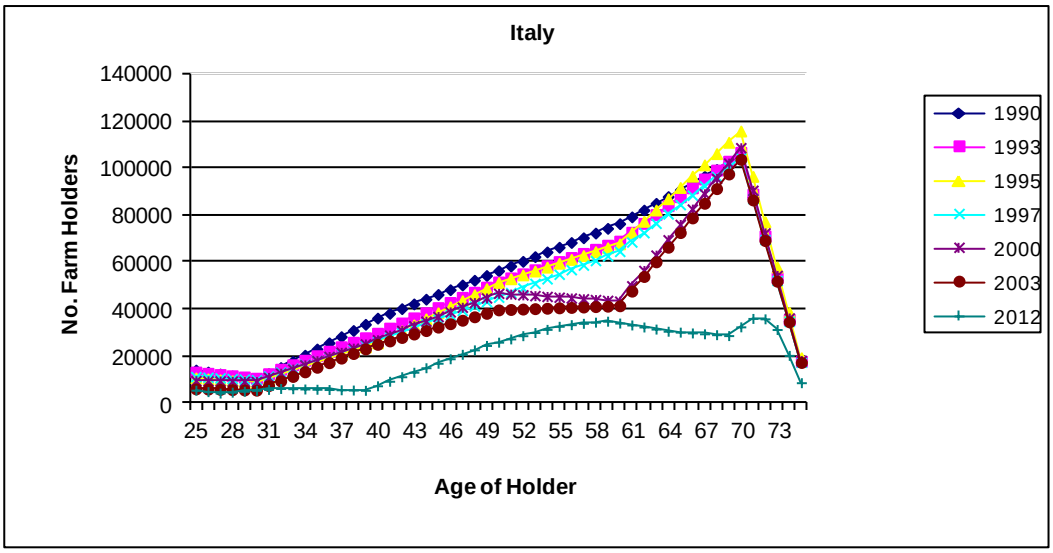


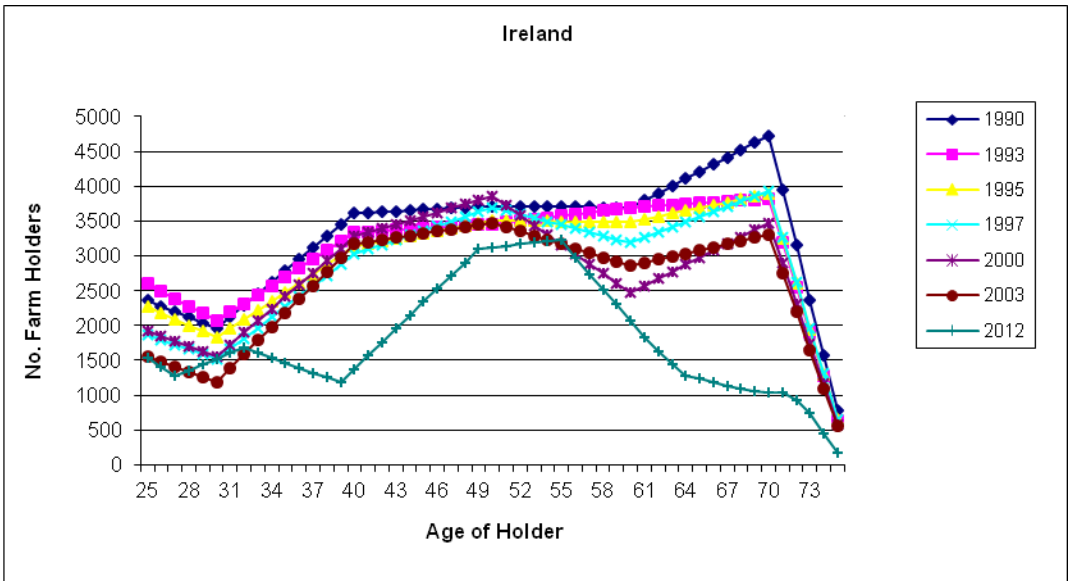

Denmark
No. Farm Holders
2500
2000
1500
1000
500
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
1990
1993
1995
1997
2000
2003
2012
Greece
No. Farm Holders
250000
200000
150000
100000
50000
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
2012
2003
2000
1997
1995
1993
1990
Finland
No. of Farm Holders
4000
3500
3000
2500
2000
1500
1000
500
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
1995
1997
2000
2003
2012


France
No. Farm Holders
35000
30000
25000
20000
15000
10000
5000
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
1990
1993
1995
1997
2000
2003
2012


Sweden
No. Farm Holders
3000
2500
2000
1500
1000
500
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
1995
1997
2000
2003
2012


Netherlands
No. Farm Holders
4000
3500
3000
2500
2000
1500
1000
500
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
1990
1993
1995
1997
2000
2003
2012





Detailed projections for 2012 and 2015 at a reasonable level have been made for all EU15 NUTS 1 regions.

As far as we know, De Haen and von Braun (1978) provides the only other attempt at analysing structural change using Cohort Analysis. They examine farm entry and exit for the *Laender* of West Germany from 1962-1973. They include in their econometric analysis variables that describe the labour market situation (ratio of vacancies to the number of unemployed, and a manufacturing wage index), the returns to agriculture (farm income) and the general macroeconomic environment pertaining to the region (regional GDP). Their results show clear interregional differences in the impact of regional labour markets, income differentials and age structures on level and structure of the total rate of change.

In the next section, we describe the data used in the calculation of the autonomous change variables and in the econometric analysis.

Residual and Econometric Data

Data on holders is taken from Eurostat's Farm Structure Surveys. Unfortunately, comprehensive holder data at NUTS2 level is a limited, particularly until the year 2003, our

last year of observation. In addition, regions that only enter the dataset in 2003 have to be excluded, as it is not possible to calculate forecasted holders for this year. We are further constrained by the availability of NUTS2 data on some macroeconomic variables and therefore include data from the 1995, 1997, 2000 and 2003 surveys only. This leaves 433 observations, from 14 member states, 63 NUTS1 and 87 NUTS2 regions.

Farm Structure Surveys decompose holders in to several cohorts: those under 35 years old, those between 35 and 44, between 45 and 54, between 55 and 64, and those over 65. We assume that all holders 'under 35' in this data are older that 25. In order to make projections for future 25-34 year old holders we construct new cohorts between 10 and 14 years old, between 15 and 19, and between 20 and 24. These are constructed by simply assuming that the ratio of holders in the 25-34 cohort to the total male population in this age group in the region is the same for the three youngest cohorts. Observations are further restricted by the lack of NUTS2 data for some regions and years as is described below. Each holder group, or cohort, is then linearly interpolated in order to ensure a smooth distribution of holders across the cohort for each year of age. This gives the number of holders in a region for each age between 10 and 75[47].

In order to calculate the probability of survival we obtain data on regional death rates and population for each age from the Regio database. The probability of survival is one minus the former divided by the latter[48]. For Germany, regional deaths rates were taken from the National Statistics Agency[49] and were only available for ages in 5 year intervals. The intervals were linearly interpolated to give the number of deaths at each age. The probability of survival is then calculated as before.

Data on regional retirement rates of farmers at each age were unavailable and so crude assumptions were necessarily made. Specifically, we assume a linear retirement rate for holders between 55 and 75 years of age. A more accurate, region-specific retirement function is clearly desirable and this is one aspect of our current approach which we intend to modify.

Our residuals for the econometric analysis of farm entry and exit are computed as the difference between forecasted holders for a particular cohort in a particular year and the actual number of holder in that cohort for that year divided by the actual number of holders in that cohort in the previous year of observation. Thus a positive residual indicates net entry to agriculture in that region.

As mentioned, the econometric analysis comprises, on the left hand side, the 'residuals' from above, and, on the right hand side, information on the contemporaneous macroeconomic climate. Following De Haen and von Braun we choose variables to reflect the labour market situation, income and to the national growth rate of the economy, in this case over the 1995-2000 period.

To capture the general economic climate we include Regional GDP per Capita, which is only available from 1995 from Eurostat. The actual measure available from Eurostat is 'GDP at current market prices'. This is divided by regional population to get a per capita measure and then deflated using the GDP deflator, also taken from Eurostat. Including GDP per capita reduces the number of observations to 409.

In order to measure the economic significance of the agricultural sector in a region, we calculate the share of total employment attributable to agriculture in the region during the

[47] We assume that there are no holders over 75

[48] more specifically, as deaths are reported as the total at the end of the year, this is 1-(no. of deaths at age j)/(population of age j+(deaths at age j/2))

[49] www.regionalstatistik.de

1995-2000 period. The Eurostat variable we use is "full-time/part-time male employment by economic activity at NUTS level 2". Including the share of agricultural employment in the model with the residuals reduces the number of observations to 333.

As a complementary measure, we also calculate the share of Gross Value added in a region that is derived from Agriculture. The Eurostat variable used is 'Gross Value Added at Basic Prices'. Including this variable with the residuals reduces the number of observations to 237. Both of these variables enable us to control for possible differences in the availability of employment outside of agriculture.

In addition, we include a measure for farm income constructed using data on family farm income in currency units from the Farm Accountancy Data Network (FADN)[50]. This measure is deflated using the national consumer price index from Eurostat (base year=1995). We calculate farm income at NUTS2 level where possible: where it is possible to discern the NUTS2 region form the FADN farm code. The number of observations falls to 305 when this income measure is included with the residuals.

Finally, we attempt to control for the labour market environment by including the total unemployment rate in the region, also taken from the Regio database. This differs from the measure used by De Haen and Von Braun in that they use the ratio of job vacancies to the number of unemployed in the German *Laender* to measure slackness in the market for labour. Unfortunately, NUTS2 data is not available on job vacancies. The number of observations falls to 377 when the unemployment rate is included with the residuals.

Having described the data used to construct our variables for the first and second stage of the Cohort Analysis we now describe our econometric model and results.

The inclusion of the above variables in the models reduces the number of observations to less than 200, and the periods modelled to the last three dates for which we have data: 1995, 1997, 2000.

Econometric Analysis

The equation estimated, for both young (25-35) and slightly older (35-55) farmers, is:

Equation 46

$$Net \quad non-autonomous \quad entry \quad (\%) = \sum_{j=1}^{k} \beta_j X_{jit} + \sum_{c=1}^{m} \alpha_c Country_c + \sum_{t=1}^{T} \lambda Year_t + e_{it}$$

The X variables are Regional GDP per capita, Unemployment rate, Farm Family Returns per AWU, Share of Total Output from Agriculure, Share of Total Workforce in Agriculture. The i subscripts indicate region and the t subscripts year. The m country dummies are dummies for the 14 countries, to control for nation-specific factors. After a two-way stepwise procedure only those variables with p-values less than .25 are retained. A "small" model does not include Farm Family Returns per AWU, Share of Total Output from Agriculure, Share of Total Workforce in Agriculture and has a sample size of 303. A "full model" includes the above variables as well as regional GDP and unemployment.

The results of the econometric analysis are presented below. The reported results only contain those variables with a p-value less than .2. The two most satisfactory models (for entry/exit of 35-55 year olds and entry/exit for older farm holders) only have one substantive significant (or close to significant) right hand side variable: regional unemployment at the beginning of

[50] The FADN code is SE425

each period (although regional GDP is also significant in one regression). The other variables are all insignificant once appropriate national and annual dummies are included in the regression. In one way this is unfortunate, and may be due to the small sample size. But perhaps, in the present context, it is sufficient to have one variable that links on and off-farm employment. The fact that this variable is unemployment and that the signs and coefficient sizes for the two age groups concerned are quite plausible means that linking the results with versions of CAPRI may turn out to be less problematic than might otherwise have been the case.

Table 20 Net Entry for Non-Demographic reasons: Small Model

	Net Entry for non-demographic reasons:		
	25-34 year olds	**35-54 year olds**	**55-75 year olds**
lag Male Unemployment	-	0.76	0.65
	-	[3.83]***	[2.98]***
1997	7.40	-	-6.97
	[1.73]*	-	[3.96]***
2000	27.13	3.07	-11.08
	[6.09]***	[2.08]**	[5.98]***
United Kingdom	39.63	19.90	27.06
	[5.28]***	[7.00]***	[8.81]***
Italy	7.93	7.17	9.18
	[1.58]	[3.63]***	[4.34]***
Germany	-	17.26	-13.93
	-	[7.02]***	[5.22]***
Spain	16.09	5.72	10.57
	[3.09]***	[2.68]***	[4.57]***
Portugal	-	7.90	17.31
	-	[1.61]	[3.24]***
Greece	20.57	12.14	16.37
	[3.63]***	[5.34]***	[6.73]***
Ireland	-	-	16.57
	-	-	[2.26]**
Netherlands	9.11	7.59	-
	[0.85]	[1.86]*	-
constant	-18.58	-7.59	-4.52
	[5.13]***	[4.08]***	[2.23]**
Nobs	303	303	303
Adj R^2	0.179	0.254	0.418

Net Entry for Non-Demographic reasons: Full Model

	Net Entry for non-demographic reasons:		
	25-34 year olds	**35-54 year olds**	**55-75 year olds**
lag GDP	-	-	-0.0004
	-	-	[2.37]**
lag Male Unemployment	-	0.850	0.539
	-	[5.01]***	[2.51]**
Portugal*1997	-20.401	-	-
	[1.47]	-	-
Portugal*2000	-	7.332	-
	-	[1.38]	-

Austria*1997	-16.864	-	-
	[1.21]	-	-
Sweden*1997	-11.511	-	11.958
	[1.14]	-	[2.59]***
Sweden*2000	-11.871	-9.981	-
	[1.18]	[2.58]*	-
Sweden*2003	-24.997	-14.547	-9.883
	[2.81]***	[4.39]***	[2.36]**
Luxembourg*2003	-22.781	-	-
	[0.96]	-	-
Germany*1997	-13.372	-	-10.568
	[1.5]	-	[2.6]***
Germany*2000	-	-	-28.493
	-	-	[7.00]***
Germany*2003	-28.524	20.445	-15.088
	[3.93]***	[7.59]***	[4.58]***
Spain*1997	-16.583	-6.972	-
	[2.48]**	[2.57]*	-
Spain*2000	58.623	15.799	-
	[8.96]***	[6.2]***	-
Spain*2003	-43.960	-12.411	8.416
	[6.56]***	[5.12]***	[2.88]***
Netherlands*1997	-29.493	-	-
	[2.43]**	-	-
Netherlands*2003	-	-	-14.053
	-	-	[2.52]**
Denmark*2003	-20.841	-	-
	0.88	-	-
Greece*2003	-25.287	-	9.935
	[3.48]***	-	[2.8]***
United Kingdom*1997	-	6.545	-
	-	[1.96]*	-
United Kingdom*2003	33.177	19.173	45.316
	[4.1]***	[6.41]***	[12.61]***
Italy*1997	16.272	-	-
	[2.54]**	-	-
Italy*2000	-12.222	5.343	-
	[1.91]*	[2.33]**	-
Italy*2003	-27.780	-	10.134
	[4.34]**	-	[3.69]***
constant	8.761	-0.893	3.099
	[2.7]***	[0.68]	[0.79]
Nobs	231	231	231
Adj R^2	0.526	0.536	0.591

Dependent variable is net entry of holders net of demographic factors for cohorts aged between 25 and 34, 35 and 54, and 55 and 75 years old respectively. Regressors in the full cohort model: Regional GDP per capita, Unemployment rate, Farm Family Returns per AWU, Share of Total Output from Agriculture, Share of Total Workforce in Agriculture and country-year interactions. All money variables are in Euro, 2000. All independent variables measured at the beginning of the relevant period.

2.5 The world Data Base (M. Adenäuer)

Since 2013 the global database is populated in global_database.gms (former global.gms). It mainly uses data from FAOSTAT, which was preprocessed and stored in the results\fao folder. This contains: Commodity balances, production and landuse statistics (all stored in faodata.gdx), special balances for dairy products (fao_milkdata.gdx), population (fao_population.gdx), producer prices (fao_prices.gdx) as well as the bilateral trade matirix (fao_trade_for_global.gdx).

These elements are directly mapped to CAPRI codes (mappings can be found in global\fao_codes_new.gms). Some of them receive, however adjustments and amendments.

Market balances and trade flows from FAOSTAT do not include biofuels and their demand for agricultural goods. Market balances and feedstock demand for biofuels are constructed from the AGLINK-COSIMO database. This happens in global\biofuel_markets.gms. Bilateral trade flows are also constructed using total exports and imports from AGLINK and trade data from COMEXT, USDA and FO-Licht. This happens in global\biofuel_tradeflows.gms.

Both, market balances and bilateral trade flows are made consistent for the selected base year in global\map_tradeflow_to_CAPRI.gms, where a small optimisation model adjusts trade flows such that trade positions from the market balances are approximated.

AGLINK data is also used for market prices in the base year.

Other data sources can be mentioned:

- Data on policy variables such as applied and scheduled tariffs, tariff rate quotas or bilateral trade agreements are obtained from the AGLINK Model (OECD) and the Agricultural Market Access Database (AMAD).
- The UNSTATS database (United Nation statistical Division)[51] is used for GDP and expenditure data.
- Federal Office for Agriculture (FOAG, Switzerland)) data for balances prices and trade flows of Switzerland
- Elasticities calculated from AGLINK-COSIMO experiments – only used in parts currently (global\aglink_elas.gms)
- Medium terms outlooks for quantity and price developments are mainly taken from from AGLINK-COSIMO.
- Long run projections for market balances in world regions are derived from earlier projections developed on behalf of the FAO with the @2030 model (http://www.ilr1.uni-bonn.de/agpo/rsrch/at2030/at2030_e.htm) and in cooperation with the GLOBIOM modelling system

[51] Se http://unstats.un.org/unsd/snaama/dnllist.asp

A new feature compared to earlier versions is the construction of medium term outlook projections for global market balances, prices or GDP in global\create_final_marketdb.gms. So far only one growth rate was calculated for 2020 based on AGLINK projections meaning that a simulation for 2015 would use a linear interpolation with this growth rate disregarding the path contained in the OECD or EU outlook. Now, projections are calculated for each year up to 2030 using AGLINK data until the final available outlook year and linear extensions until 2030. These projections do however not include any consictency checks on closed market balances or similar things. This is still achieved in the baseline calibration only.

2.5.1 Biofuels in the global database

2.6 Policy data

2.6.1 Policy data linked to production activities in Europe (P. Witzke, M. Adenäuer, T. Jansson)

2.6.1.1 Sugar and milk quotas

Sugar quotas (Marcel Adenaeuer, Torbjoern Jansson)

After decades of stability the EU sugar CMO has been reformed in 2006 through Regulations EU 318 to 320/2006. Both the pre-reform quotas (‘*sugar\load_quotas.gms’*) as well as modifications to the former quotas through renouncements, withdrawal and buy-back steps (sugar\reformsugarquotas.gms) are represented in detail, including a conversion of sugar quotas on the MS level to beet quotas for NUTS2 regions (‘*sugar\ breakdown_quots.gms’*). The sugar (and beet) quota data are loaded in the regionalisation workstep (CAPREG, this Chapter 2.3, here in ‘*sugar\techf.gms’*), in the trend projections (CAPTRD, Chapter 3), and in the simulation step (CAPMOD, Chapter 4)

A particularity of the sugar market CMO that has survived the reform in modified form, is a kinked marginal revenue curve that results from high “in-quota” prices (in the pre reform regime sometimes differentiated between “A” and “B” quotas) and lower “over-quota” prices (former “C” beets, now for bio-energy and industrial uses). Considering yield uncertainty this kinked demand curve gives rise to a specific form of expected revenue maximisation (see Adenaeuer 2010) that builds on a set of parameters inherited from an in depth EuroCARE sugar study in 2003 (reference). These parameters determining the supply behaviour of the sugar sector may also be considered supplementary policy variables (‘*sugar\sugar_dat.prn’*).

Milk quotas (Peter Witzke, Marcel Adenaeuer, Torbjoern Jansson)

A reform of the EU dairy sector CMO has been decided in the context of the 2009 Health Check (Regulation 72/2009). It was the subject of a specific milk study (Witzke et al 2009, edited by T Fellmann) that led to some updates. This historical structure is still visible in the current code:

- The milk quota data before the reform (according to Regulation 1256/1999 are included in ‘*pol_input\dairy_quotas_agenda.gms’* and therefore read both in the context of the regionalisation (CAPREG) and in each simulation (in CAPMOD).
- Milk quota data as of 2004, supplemented with data on observed deliveries and direct sales are incorporated (in CAPTRD and CAPMOD) via *‘pol_input\mquota_2008update.gms’*. The observed deliveries are used to correct the “model internal” quotas for typical over- or underdelivery as observed in the past.

Both in CAPREG as well as in CAPMOD milk quotas are assigned to regions according to the production shares in the base period (in ‘*pol_input\agenda.gms’* and ‘*supply\prep_milk_quotas.gms’*, respectively).

2.6.1.1 Premiums

An exposition of the various premium types included in CAPRI is given in Section 4.3.

2.6.2 Policy data linked to European and international markets (W. Britz, P. Witzke, B. van Doorslaer)

2.6.2.1 Tariffs and Tariff Rate Quotas

Update pending.

2.6.2.2 Export subsidies

Update pending.

2.6.2.3 Producer subsidies

Update pending.

2.6.2.4 Consumer subsidies

Update pending.

2.6.2.5 Public Intervention purchases and sales

Update pending.

3 Baseline Generation (W. Britz, P. Witzke)

3.1 Overview beyond CAPTRD

The purpose of a baseline is to serve as a comparison point or comparison time series for counterfactual analysis. The baseline may be interpreted as a projection in time covering the most probable future development or 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 in conjunction of trend analysis, model runs and expert consultations. In this process, model parameters such as e.g. elasticities and exogenous assumptions such as 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. 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 values inputted may stem themselves 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 inputted for the exogenous variables. Clearly, as the countries will fill in their questionnaire 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 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 mechanism 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 by trend forecasts using data from the 'COCO' database (Chapter 2.2). 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 data base.

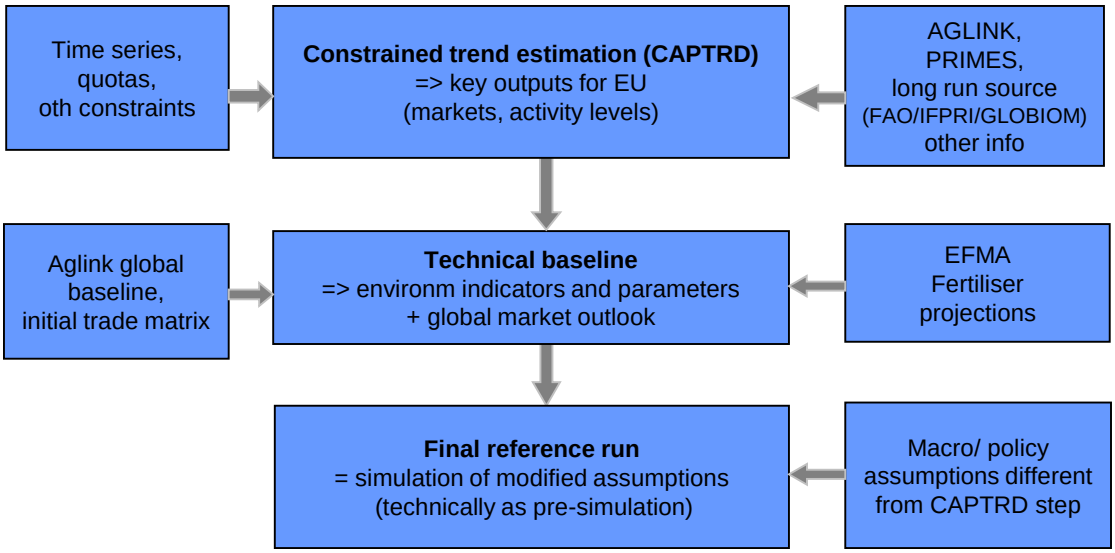
The tool providing projections for European regions (CAPTRD) operates in several steps:

- 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. Because this requires some disaggregation to single MS but also because it often the key information steering the outcome, it is treated in a step distinct from (2).
- 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 constrained trends from CAPTRD are simultaneously subject to the 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 it is to be acknowledged here that the trend remain mechanical in that they try to respect technological relationships but remain ignorant about behavioural functions or policy developments[52].

However it should be explained that the CAPTRD results are in turn only the first of several steps before a full CAPRI baseline is ready to use. There are at least one and often two steps following:

Figure 7. Overview on CAPRI baseline process



1. The constrained trend estimation merges the information in the ex post time series with external information. The result of this first step is a first projection for the key variables in the agricultural sector (activity levels and market balances) of Europe.
2. The "technical baseline" calibrates missing parameters and in this context also calculates missing variables that are related to the key variables, in particular complete nutrient balances in the crop and livestock sectors and all non EU market balances and the bilateral trade matrix.

[52] The only exceptions are the quota regimes on the milk and sugar markets which are recognised in the trend projections.

3. A third step may give the final reference run if some assumptions made in steps one or two need to be revised to obtain the desired starting point for further analysis. In some studies it turned out useful, for example to modify the macro assumptions of the "agricultural" expert sources (AGLINK, FAO, IFPRI).

The rest of this chapter focuses on CAPTRD. The next Section 3.2 presents some details on the CAPTRD step 1 and its preparation. Key ingredients of the next Step 2 are the objective function and constraints, detailed in Section 3.3. Various types of MS level expert information also introduced as part of step 2 are explained in Section 3.4. Section 3.5 discusses the procedure and important details related to the CAPTRD Step 3. Disaggregation to the regional or farm type level is briefly explained in Section 3.6 and the final Section 3.7 gives a brief overview on the subsequent use of the CAPTRD results in the simulation step (CAPMOD). A more complete description presupposes an understanding of the simulation model itself such that this is postponed to a later section.

3.2 CAPTRD 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:

Equation 47 $$X_{r,i,t}^{j,Trend} = a_{r,i,j} + b_{r,i,j}\, t^{c_{r,i,j}}$$

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* 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. Coverage of more recent years in CAPREG is often precluded by missing values on several variables. However, in 2011, the COCO1 output has been completed up to 2010 in a "nowcasting" exercise, relying on the availability of the key data at the MS level. To make use of this information, these recent COCO1 data can be used to extend the CAPREG time series on the MS level (and only for EU27 countries). However, the length of regionalised time series remains limited by the most recent CAPREG output. To account for this different availability of ex post data requires to distinguish the following sets:

- *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[53].
- *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:

Equation 48 $$wSSE_{r,i,j} = \sum_{expost} \left(X_{r,i,expost}^{j,"Data"} - a_{r,i,j} + b_{r,i,j} t_{expost}^{c_{r,i,j}} \right)^2 t_{expost}$$

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 that 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).

3.3 CAPTRD Step 2: 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 which both fit into the consistency constraints and exploit the information comprised in the ex-post development in a technical feasible way. Consider the identity which 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 Step 1 derived values, 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.

[53] For technical reasons some years are "obligatory" result years, for example the year immediately following after the last ex post year.

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 H_0 – the value is equal to the recent three year average:

Equation 49
$$X^{j,"Support"}_{r,i,exante} = wR_{r,i,j}{}^2 \left(a_{r,i,j} + b_{r,i,j} t^{c_{r,i,j}}_{exante}\right) + \left(1 - wR_{r,i,j}{}^2\right) X^{j,"Data"}_{r,i,bas}$$

where

Equation 50
$$wR_{r,i,j}{}^2 = 1 - wSSE_{r,i,j} / wSST_{r,i,j}$$

and the weighted total sum of squares is defined analogous to Equation 48:

Equation 51
$$wSST_{r,i,j} = \sum_{expost} \left(X^{j,"Data"}_{r,i,expost} - X^{j,"Data"}_{r,i,wAve}\right)^2 t_{expost}$$

with a trend weighted average

Equation 52
$$X^{j,"Data"}_{r,i,wAve} = \sum_{expost} X^{j,"Data"}_{r,i,expost} \cdot t_{expost} \Big/ \sum_{expost} t_{expost}$$

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:

Equation 53
$$Penalty = \sum_{r,i,j,exante} \left(\frac{X^{j,"Trend"}_{r,j,exante} - X^{j,"Support"}_{r,i,exante}}{\sqrt{X^{j,"Step1"}_{r,i,"varErr"}}} \right)^2$$

where the weighted variance of errors is

Equation 54
$$X^{j,"Step1"}_{r,i,"varErr"} = wSSE_{r,i,j} \Big/ \left(\sum_{expost} t_{expost} - 1 \right)$$

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. As discussed in Section 3.7, the full projection tool features further constraints especially relating to price feedbacks on supply and demand.

3.3.1 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*):

Equation 55
$$\begin{aligned} X^{,IMPT,Trend}_{r,i,t} + X^{GROF,Trend}_{r,i,t} = \\ X^{FEDM,Trend}_{r,i,t} + X^{SEDM,Trend}_{r,i,t} + X^{PRCM,Trend}_{r,i,t} + X^{INDM,Trend}_{r,i,t} \\ + X^{BIOF,Trend}_{r,i,t} X^{LOSM,Trend}_{r,i,t} + X^{HCOM,Trend}_{r,i,t} + X^{,EXPT,Trend}_{r,i,t} \end{aligned}$$

Where *r* are the Member States of the EU, *i* are the products, *t* the different forecasting years, corresponding to Equation 2 above. 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 of Equation 55 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 (see Section 3.5).

Secondly, production of agricultural raw products (*GROF*) is equal to yield times area/herd size (*LEVL*) where *acts* are all production activities (eq. *GROF_*):

Equation 56
$$X^{GROF,Trend}_{r,i,t} = \sum_{acts} X^{acts,Trend}_{r,i,t} X^{acts,Trend}_{r,LEVL,t}$$

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.

Equation 57 $$X^{MrkBal,Trend}_{r,pro_grp,t} = \sum_{i \in pro_grp} X^{MrkBal,Trend}_{r,i,t}$$

3.3.2 Constraints relating to land use and cropping area

Adding up over the individual crop areas defines the total utilizable agricultural area (*UAAR,LEVL, AREAB_*):

Equation 58 $$X^{UAAR,Trend}_{r,LEVL,t} = \sum_{crops} X^{crops,Trend}_{r,LEVL,t}$$

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}):

Equation 59 $$X^{GrpC,Trend}_{r,LEVL,t} = \sum_{crops \in GrpC} X^{crops,Trend}_{r,LEVL,t}$$

Adding up over mutually exclusive land use (in *LANDUSEB_*, for set LandUseARTO, see Annex) defines the total area (*ARTO,LEVL*):

Equation 60 $$X^{ARTO,Trend}_{r,LEVL,t} = \sum_{LandUseARTO} X^{LandUseARTO,Trend}_{r,LEVL,t}$$

3.3.3 Constraints relating to agricultural production

Another Equation (*OYANI_*) links the different animal activities over young animal markets:

Equation 61 $$X^{GROF,Trend}_{r,oyani,t} - X^{STCM,Trend}_{r,oyani,t} = \sum_{iyani \leftrightarrow oyani} X^{GROF,Trend}_{r,iyani,t}$$

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_*)

Equation 62 $$X^{splitactlo,Trend}_{r,LEVL,t} = X^{splitacthi,Trend}_{r,LEVL,t}$$

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_*.

Equation 63 $$X^{maact,Trend}_{r,HERD,t} = X^{maact,Trend}_{r,LEVL,t} \cdot X^{maact,Trend}_{r,DAYS,t} / 365$$

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[54].

[54] 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.

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_*

Equation 64

$$X^{maact,Trend}_{r,yield,t} / carcassSh_{maact} = startWgt_{maact} + X^{maact,Trend}_{r,DAILY,t} \cdot \left(X^{maact,Trend}_{r,DAYS,t} - X^{maact,data}_{r,EDAYS,BAS} \right)$$

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 (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_*)

Equation 65

$$X^{bulf,Trend}_{r,DAILY,t} < 0.4 + X^{bulf,Trend}_{r,meat,t} / carcassSh_{maact} \cdot 0.0016$$

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* :

Equation 66

$$\sum_{feed} X^{maact,Trend}_{r,feed,t} X^{Cont,Trend}_{r,feed,t} = 0.998^{t} \left(a^{Const}_{maact} + a^{Slope}_{maact} X^{maact,Trend}_{r,yield,t} \right)$$

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 non-tradable 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 56 above:

Equation 67

$$X^{GROF,Trend}_{r,feed,t} = \sum_{maact} X^{maact,Trend}_{r,feed,t} X^{maact,Trend}_{r,levl,t}$$

Feed use of individual products must add up to the feed use of the 'bulks' mentioned above (in *FEED_*):

Equation 68

$$X^{FEDM,Trend}_{r,feed,t} = \sum_{o \rightarrow feed} X^{FEDM,Trend}_{r,o,t}$$

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*):

Equation 69 $$X_{r,LABO,t}^{GROF,Trend} = X_{r,LABH,t}^{GROF,Trend} + X_{r,LABN,t}^{GROF,Trend}$$

In the first place projections of family and hired labour follow from input coefficients combined with the activity levels (like Equation 56), but the previous equation permits to apply bounds to the total.

3.3.4 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*):

Equation 70 $$X_{r,i,t}^{EAAG,Trend} = X_{r,i,t}^{GROF,Trend} \; X_{r,i,t}^{UVAG,Trend}$$

The revenues of the activities (*TOOU*, total output) for each activity and group of activities *acts* are defined in Equation *REVE_* as:

Equation 71 $$X_{r,TOOU,t}^{acts,Trend} = \sum_{o} X_{r,o,t}^{acts,Trend} \; X_{r,o,t}^{UVAG,Trend}$$

Consumer prices (*UVAD*) are equal to producer prices (*UVAG*) plus a margin (*CSSP*[55]) according to Equation *UVAD_*:

Equation 72 $$X_{r,i,t}^{UVAD,Trend} = X_{r,i,t}^{UVAG,Trend} + X_{r,i,t}^{CMRG,Trend}$$

3.3.5 Constraints relating to consumer behaviour

Human consumption (*HCOM*) is defined as per head consumption multiplied with population (*HCOM_*):

Equation 73 $$X_{r,i,t}^{HCOM,Trend} = X_{r,i,t}^{INHA,Trend} \; X_{r,LEVL,t}^{INHA,Trend}$$

Consumer expenditures per caput (*EXPE*) are equal (via *EXPE_*) to human consumption per caput (*INHA*) times consumer prices (*UVAD*):

Equation 74 $$X_{r,i,t}^{EXPE,Trend} = X_{r,i,t}^{INHA,Trend} \; X_{r,LEVL,t}^{UVAD,Trend}$$

Total per caput expenditure (*EXPE.LEVL*) must add up (in Equation *EXPETOT_*):

---

55 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.

Equation 75
$$X^{EXPE,Trend}_{r,LEVL,t} = \sum_{i} X^{EXPE,Trend}_{r,i,t}$$

3.3.6 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*):

Equation 76
$$X^{MAPR,Trend}_{r,sec,t} = \sum_{i \wedge sec \leftarrow i} X^{PRCM,Trend}_{r,i,t} \; X^{PRCY,Trend}_{r,sec,t}$$

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[56]) 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):

Equation 77
$$X^{PRCM,Trend}_{r,MILK,t} X^{MLKCNT,Trend}_{r,MILK,t} = \sum_{mlk\,seco} \left(X^{MAPR,Trend}_{r,mlk\,seco,t} - X^{PRCM,Trend}_{r,mlk\,seco,t} \right) X^{MLKCNT,Trend}_{r,mlkseco,t}$$

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_*):

Equation 78
$$X^{MAPR,Trend}_{r,resimp,t} = \sum_{o \rightarrow resimp} \left(X^{HCOM,Trend}_{r,o,t} + X^{PRCM,Trend}_{r,o,t} + X^{INDM,Trend}_{r,o,t} + X^{BIOF,Trend}_{r,o,t} \right) \cdot \frac{X^{MAPR,bas}_{r,resimp,t}}{\sum_{o \rightarrow resimp} \left(X^{HCOM,bas}_{r,o,t} + X^{PRCM,bas}_{r,o,t} + X^{INDM,bas}_{r,o,t} + X^{BIOF,bas}_{r,o,t} \right)}$$

3.3.7 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[57] (*PRCB*) from biofuel use of first generation feedstocks (*BIOF*):

Equation 79
$$X^{MAPR,Trend}_{r,seco_biof,t} = \sum_{stocks \rightarrow seco_biof} X^{BIOF,Trend}_{r,stocks,t} \; X^{PRCB,Trend}_{r,stocks,t}$$

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_*):

---

[56] This is somewhat indirectly related to processing of cow milk and sheep & goat milk over *MAPR.MILK = PRCM.COMI + PRCM.SHGM* according to Equation 76 with a processing yield *PRCY.MILK* = 1 and over the market balance Equation 55 for product *MILK* which ensures that, with minimal trade of raw *MILK*, most of *MAPR.MILK* will end up as *PRCM.MILK*.

[57] 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.

Equation 80 $$X^{MAPR,Trend}_{r,DDGS,t} = \sum_{stocks \rightarrow DDGS} X^{BIOF,Trend}_{r,stocks,t} \; X^{PRCBY,Trend}_{r,stocks,t}$$

3.3.8 Constraints relating to policy

There are only a few constraints directly taken from an EU regulation: firstly, the acreage under compulsatory set-aside (abolished in the CAP Health Check of 2008) must be equal to the set-aside obligations of the individual crops (*OSET_*):

Equation 81 $$X^{"OSET",Trend}_{r,"levl",t} = \sum_{cact} X^{cact,Trend}_{r,"levl",t} \; \frac{0.01 \; X^{cact,Trend}_{r,"setr",t}}{\left(1 - 0.01 \; X^{cact,Trend}_{r,"setr",t}\right)}$$

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):

Equation 82 $$X^{MAPR,Trend}_{r,SUGA,t} \leq X^{QUTS,Trend}_{r,SUGA,t} + X^{INDM,Trend}_{r,SUGA,t} + X^{BIOF,Trend}_{r,SUGA,t} + X^{LOSM,Trend}_{r,SUGA,t}$$

Finally there are upper bounds on new plantings of vineyards according to the CMO for wine from Regulation 1493/1999

3.3.9 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%:

Equation 83

$$X^{arab,Trend}_{r,"level",t}.up/lo = X^{arab,Trend}_{r,"level",bas} \pm \tfrac{1}{4}\left(\frac{X^{arab,Trend}_{r,"level",bas}}{X^{"arab",Trend}_{r,"level",bas}}\right)^{\tfrac{1}{4}} X^{"arab",Trend}_{r,"level",bas} \max\left(0.2, \frac{t-bas}{last-bas}\right)$$

- 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.

3.4 CAPTRD Step 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. The expert supports are e.g. extensively used in the bio-fuel sector were often historical are not existing or time series are very short. Equally, expert supports are used for

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 (see below, Equation 84).

3.4.1 Integration of bio-fuel information from PRIMES

The following table presents selected results related to the EU27 biomass feedstock for bioenergy production from the PRIMES biomass component (also given for each MS):

Unit: ktoe (unless specified otherwise)	**2000**	**2005**	**2010**
Domestic Production of Biomass Feedstock	**69087**	**87595**	**101303**
Crops	1228	5419	12500
- Wheat	0	601	2462
- Sugarbeet	0	1291	4518
- Sunflower/Rapeseed	1228	3527	5520
- Lign. Crops	0	0	0
Agricultural Residues	4194	6428	7200
Waste	19990	26002	28054
Net imports of Biomass Feedstock	**239**	**1598**	**4289**
Pure Vegetable Oil as feedstock for bioenergy production	239	1598	4289
Cultivated Land (Kha)	**896**	**3022**	**5422**
Starch crops	0	320	1218
Oil crops	896	2654	4031
Sugar Crops	0	48	172
Lignocellulosic crops	0	0	0

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[58] 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

[58] It appears that the CAPRI bio-fuel results of August 2011 are affected by reporting errors in the oilseeds and sugar sectors.

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 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)

Unit: ktoe (unless specified otherwise)	**2000**	**2005**	**2010**
Net imports of Bioenergy	**400**	**1731**	**5820**
Biodiesel	0	0	1948
Bioethanol	0	20	1130
Pure Vegetable Oil	8	390	505
Bioenergy Production	**67971**	**84554**	**95430**
Biodiesel	610	2548	6578
- Biodiesel (1st gen.)	610	2548	6578
- FT diesel	0	0	0
- HTU diesel	0	0	0
- Pyrolysis diesel	0	0	0
Bioethanol	0	561	2193
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	2049

- Bio-gas	352	871	2049
- Synthetic Natural Gas	0	0	0
Waste Solid	12353	13985	14654
Waste Gas	1898	3537	4538
Demand	**68372**	**86285**	**101250**
Biodiesel	610	2548	8526
Bioethanol	0	581	3234
BioKerosene	0	0	0
BioHydrogen	0	0	0
BioHeavy	9	473	1110
BioGas	352	871	2049
Waste Solid	12353	13985	14654
Waste Gas	1898	3537	4538
Capacities (Ktoe/yr)	**10440**	**16067**	**26754**
Fermentation	134	1127	4104
- Sugar	0	551	2103
- Starch	134	576	2001
- Lignocellulosic	0	0	0
Esterification	1141	4170	9021

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.

3.5 CAPTRD 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. The baseline of AGLINK 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, where at least in the past the mnemonics had even not be harmonized across equations of the AGLINK model. After a restructuring exercise in 2010, that has somewhat improved. The ingredients in the mapping process are first a list of the codes for the regions, products and items used in AGLINK (*‘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, and a larger program

(*'baseline\loag_aglink*.gms'*) then uses the mapping to assign them to the CAPRI code world.

AGLINK currently features results at EU15 and EU12 level. It is hence not possible to funnel the AGLINK results into Step 2 above without an assumption of the share of the individual Member States.

As DG-AGRI is often the main client, 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 baseline (after adjusting for different definitions in the respective databases).
2. The standard errors from the default tends are replaced with a special formula reflecting a high confidence in the AGLINK derived supports.

More precisely, the weighted variance from Equation 54 is replaced with the following setting for external supports ("*XSupport*" = AGLINK or expert supports) :

Equation 84 $$X^{j,"XSupport",t}_{r,i,"varErr"} = \left(X^{j,"XSupport",t}_{r,i,"exante"} \cdot 0.05/3 \cdot \left(10/X^{j,"XSupport",t}_{r,i,"trustlevl"}\right)\right)^2$$

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 DG Agri 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.

3.5.1 Integration of expert information for long run projections beyond 2020

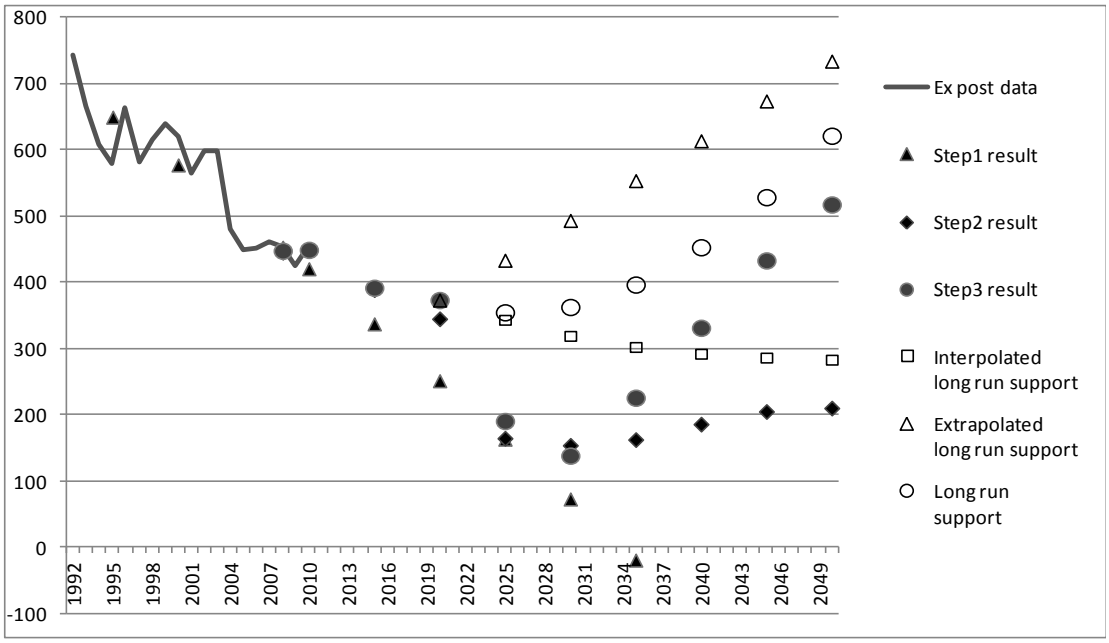
The AGLINK projections currently run to 2020 or a few years beyond. For climate related applications CAPRI has to tackle projections up to 2030 or even 2050. CAPRI projections up to 2030 have been prepared in the context of EC4MACS project (www.ec4macs.eu). The methodology was quite simple: The year 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[59].

[59] 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.

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, Fernandez, Sinha 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, 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):

Figure 8. Pork production in Hungary as an example for merging medium run and long run a priori information in the CAPRI baseline approach



The example has been chosen because historical trends (and AGLINK 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 (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 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.

3.6 CAPTRD: 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:

Equation 85
$$X^{GROF,Trend}_{MS,i,t} = \sum_{r \in MS} X^{GROF,Trend}_{r,i,t}$$

Equation 86
$$X^{j,Trend}_{MS,"levl",t} = \sum_{r \in MS} X^{j,Trend}_{r,"levl",t}$$

Equation 87
$$X^{j,Trend}_{MS,"levl",t} \cdot \left(X^{j,Trend}_{MS,"feed",t} + 10\right) = \sum_{r \in MS} X^{j,Trend}_{r,"levl",t} \cdot \left(X^{j,Trend}_{r,"feed",t} + 10\right)$$

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 58), obligatory set aside (Equation 81), aggregation to groups like cereals (0).
- Related to yields: linkage of production, activity levels and yields (Equation 56), stabilisation of straw yields (*STRA_*)
- Related to animals: Nutrient balances (Equation 66), 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.

3.7 Calibrating the model to the CAPTRD projection

3.7.1 Calibrating the regional supply models

The supply side models of the CAPRI simulation tool are programming models with an objective function. A calibration to the results of the projection tools thus requires that first order optimality conditions (marginal revenues equal to marginal costs, all constraints feasible) hold in the calibration point for each of the NUTS 2 or farm type models. The consequences regarding the calibration are twofold:

1. Elements not projected so far but entering the constraints of the supply models must be defined in such way that constraints are feasible,
2. The cost function of the models must be shifted such that marginal costs and marginal revenues are equal in the calibration point.

As explained above, the requirement functions used in the projection tools are a linear approximation for the ones used in the simulation tool; additional constraints restrict on top 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 (see Section 2.4.3). The resulting factors are stored in external files and reloaded by counterfactual runs.

Secondly, methods borrowed from Positive Mathematical Programming are applied to define the difference between marginal revenues and marginal costs in the calibration point, and these differences are added to the activity specific constant terms of the non-linear cost function (see also below, Section 4.2.3). The resulting parameters are as well stored in external files to be reloaded in case of counterfactual runs.

3.7.1.1 Update note on feed demand calibration

During the months before the biofuels work in 2007 started, simulation experiments revealed often convergence problems with feed demand. An analysis showed that a probable cause is the overestimation of the implicit feed demand elasticities inside the supply models. There are cost terms incorporated in the objective function of the supply model which characterize additional costs linked of changing the feed mix. Those had been defined in the past from rather large elasticities (-10%) and a quantity equivalent to at least 10% of the dry matter. That calculation rule is now changed, and the 10% are reduced if the total dry matter of a feedingstuff in the regional feed use is below 10%. That dampens the changes of feed ingredient used in small amounts and helps in achieving convergence.

3.7.1.2 Update note on fertiliser calibration

The past versions of CAPRI derive fertiliser projections from trend based forecasts of basically two "fudging" parameters introduced in Equation 14 of Section 2.2.6 which describe changes in the fertiliser management of farmers:

- Nutrient availability factor ("NavFac"): This reflects the partial availability of nutrients from manure relative to mineral fertiliser. In the ex-post data consolidation process this is the key variable to accommodate the difference between nutrient retention by crops and nutrient supply from mineral and organic fertiliser and other sources (see Figure 5 above). In many EU15 countries this factor was trending

upwards, reflecting efficiency improvements in fertiliser management, be it autonomous or enforced through more stringent environmental legislation.

- Nutrient overfertilisation factor ("NutFac"): This reflects the fact that farmers tend to apply more fertiliser than needed, even after accounting for partial availability of nutrients from manure (Figure 5). This is strongly pulled towards the expected mean (125%) in the ex-post data consolidation such that any trends or fluctuations are often weaker than for the NavFacs. Nonetheless we may often observe in MS15 that their NutFacs are slightly trending downwards whereas those from MS12 are often more irregular and sloping upwards, mirroring some catching up in fertiliser application after the turmoil of the transition phase.

The purely trend based projections of these parameters has some drawbacks. First of all the trends are sometimes very weak or estimated with a high standard error. Secondly it ignores the existing projections of the European Fertiliser Manufacturers Association (EFMA) which has been the source of ex post information on fertilisers in CAPRI for several years (through the International Fertiliser Association's website http://www.fertilizer.org/). The new solution introduced in 2010 relies essentially on a similar methodology as used thoughout the CAPRI system, namely to solve an optimisation problem with penalties for deviations from target values (see *'gams\capmod\def_fert_and_requirement.gms'*). In the fertiliser context the specification of fertiliser applications per ha had to be consistent with the total fertiliser use quantities and in the ex post period there were taken to be hard data, given without (or minimal) error. For the ex ante period the CAPRI team could use selected data from the EFMA projections which incorporate considerable expert knowledge. Nonetheless experts may be wrong, such that the information has been used in the form of a priori information only with penalties for deviations.

The other source of information for projections comes, as before, from the current trends on fertiliser parameters (NutFacs/NavFacs) that are now treated as a priori information as well, rather than being imposed strictly. The compromise solution depends on the STDEs specified for the objective function which is formally a Highest Posterior Density (HPD) estimator (see Heckelei, Mittelhammer, Jansson, 2008).

The advantage of the revised procedure is that the weights for the EFMA information may be varied according to its presumed reliability:

- In EU15 countries the data situation is usually better and the national experts have longer experience than in the case of EU12 countries.
- The EFMA information currently in use only lasts to 2018. It has been extrapolated statistically but evidently this implies an increasing margin of error as the projection year approaches 2050.

Both aspects are taken into account by appropriate variations of the weights.

3.7.2 Calibrating the global trade model

The projection results at EU25 level plus Norway, Turkey, Western Balkans, Bulgaria and Romania are taken as given when the global trade model is calibrated. That calibration step on the one hand defines bi-lateral import and export flows from these countries or country blocks to other trade blocks, as well as development in production, feed use, processing and human consumption for the different regions of the world not covered by the projection tool.

Wherever possible, the current market outlook from AGLINK-COSIMO is used to define target values for the simulation years. It is complemented by results by the latest available FAPRI outlook. And finally, for some countries where no other information is available, results for a FAO long-term outlook are used. For long term projections (currently beyond

2020), the market outlook are complemented respectively replaced with baseline results from the IFPRI Impact model and the FAO long term outlook.

The results from these outlooks are expressed as growth rates against the base in order to overcome probably deviations in the underlying data.

4 Simulation Scenario Model (CAPMOD)

4.1 Overview of the system (W. Britz)

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 (data from EUROSTAT) 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)[60] is assumed to be non-tradable, and hence links animal processes to the crops and regional land availability. All other outputs and inputs can be sold and purchased at fixed prices. Selling of milk cannot exceed the related quota, the sugar beet quota regime is modelled by a specific risk component. The use of a mathematical programming approach has the advantage to directly embed compensation payments, set-aside obligations, voluntary set-aside 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).

[60] A detailed description can be found in: Wolfgang Britz & Thomas Heckelei (1999): Calibration of Feed Requirements and Price determination of feed in CAPRI, CAPRI working paper 99-06, available on the project web site. (http://www.agp.uni-bonn.de/agpo/rsrch/capri/capri_e.htm)

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[61]. Tariff rate quotas (TRQs), intervention purchases and subsidised exports under the World Trade Organisation (WTO) commitment restrictions are explicitly modelled for the EU.

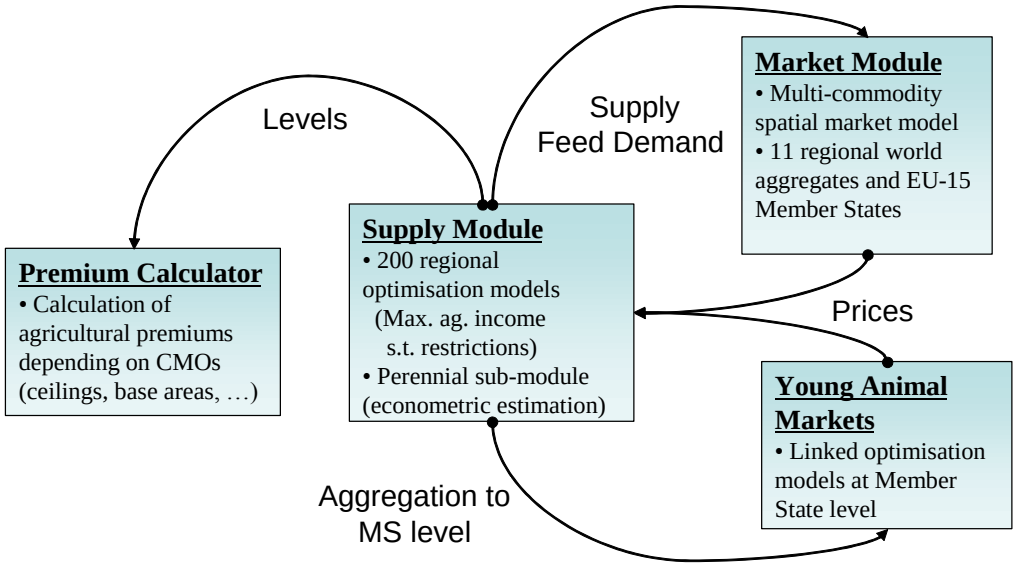
In the market module, special attention is given to the processing of dairy products. First, balancing equations for fat and protein ensure that these make use of the exact amount of fat and protein contained in the raw milk. The production of processed dairy products is based on a normalised quadratic 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 need, set-aside, 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 Models for young animal trade, separately for EU15, EU10, Bulgaria & Romania, Norway, Western Balkans and Turkey., as shown in Figure 9. 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.

[61] Currently, no PSE/CSE data are used, and CSE are only introduced for EU dairy products as derived from FEOGA budget position.

Figure 9. Link of modules in CAPRI



Source: CAPRI Modelling System. Note: the number of regions is outdated in December 2011

4.2 Module for agricultural supply at regional level (W. Britz)

4.2.1 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:

Equation 88 $$Rev_j = Cost_j + ac_j + \sum_k bc_{j,k} Levl_k + \sum_{i=1}^{m} \lambda_i a_{ij}$$

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 ($ac_j + \sum_k bc_{jk} Levl_k$) shows the marginal non-linear costs, these marginal costs are increasing in the activity levels. The cross effects are only introduced to let major arable crop groups interacts, where as for fruits & vegs, permanent crops, grass land and the animal sectors, only diagonal terms are introduced. For the estimation of these terms, see Jansson, T., und Heckelei, T. (2011): Estimating a Primal Model of Regional Crop Supply in the European Union, Journal of Agricultural Economics, 62 (1), pp 137-152.

The remaining term $\sum_{i=1}^{m} \lambda_i a_{ij}$ captures the marginal costs linked to the use of exhausted resources and the equal to the sum of the shadow prices λ multiplied the per unit demand of that activity j for resource i, the matrix $\boldsymbol{A}$ being again based on Leontief technology. The shadow values of binding resources hence are the drivers linking the activities.

A central role in the CAPRI supply model plays the land-balance. Its 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 ha 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 ($ac_j + \sum_k bc_{jk} Levl_k$) 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 λ 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 to decrease other costs in producing fodder or to increase the internal marginal revenues. Stating it the other way around 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 their 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.

Figure 10. Overview on a regional aggregate programming model

	Crop Activities	Animal Activities	Feed Use	Net Trade	Constraints
Objective function	+ Premium – Acc.Costs – variable cost function terms	+ Premium – Acc.Costs – variable cost function terms	- variable cost function terms for feeding	+ Price	
Output	**+**	**+**	**-**	**-**	= 0
Area	**-**				<= land supply
Set aside	**+/-**				= 0
Quotas	**-**	**-**			<= Ref. Quantity
Fertilizer needs	**-**	**+**		**+**	= 0
Feed requirements		**-**	**+**	**+**	= 0

Source: CAPRI modelling system

4.2.2 Detailed discussion of the equations in the supply model

The definition of the supply model can be found in *'supply\supply_model.gms'*

4.2.2.1 Feed block

The feed block ensures that the requirements of the animal processes are met, and links these to the markets and crop production decisions. The first type of equation ensures that requirements (energy, protein, lysine, minimum and maximum dry matter, different fibre requirements for ruminants) are met:

Equation 89
$$\overline{AREQ}_{r,acct,req}\ \overline{DAYS}_{r,aact} \leq \sum_{feed} FEDNG_{r,acct,feed}\ \overline{REQCNT}_{r,feed,req}$$

The left hand side captures the daily animal requirements (*AREQ*) for each region *r*, animal activity *acct* 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. For energy and protein, the less than is replaced by an equal sign to ensure a more plausible substitution inside the feed mix.

Two additional restrictions ensure that the content of a certain type of feed in the mix measured in dry matter is in between pre-defined upper and lower limits (*MAXSHR*, *MINSHR*):

Equation 90
$$\overline{\mathrm{AREQ}}_{\mathrm{r,acct,"DRMA"}}\ \overline{\mathrm{DAYS}}_{\mathrm{r,aact}}\ \overline{\mathrm{MAXSHR}}_{\mathrm{r,acct,feed}} \geq \mathrm{FEDNG}_{\mathrm{r,acct,feed}}\ \overline{\mathrm{REQCNT}}_{\mathrm{r,feed,"DRMA"}}$$

$$\overline{\mathrm{AREQ}}_{\mathrm{r,acct,"DRMA"}}\ \overline{\mathrm{DAYS}}_{\mathrm{r,aact}}\ \overline{\mathrm{MINSHR}}_{\mathrm{r,acct,feed}}$$

Equation 91
$$\leq \mathrm{FEDNG}_{\mathrm{r,acct,feed}}\,\overline{\mathrm{REQCNT}}_{\mathrm{r,feed,"DRMA"}}$$

Total feed use (*FEDUSE*) in a region is defined as the feeding per head multiplied with the activity level (*LEVL*) for the animal activities:

Equation 92
$$FEDUSE_{r,\,feed} = \sum_{aact} LEVL_{r,aact} FEDNG_{r,aact,\,feed}$$

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.

4.2.2.2 Land balances and set-aside restrictions

The model distinguishes arable and grassland and comprises thus two land balances:

Equation 93
$$\overline{LEVL}_{r,"arab"} \leq \sum_{arab} LEVL_{r,arab}$$

Equation 94
$$\overline{LEVL}_{r,"gras"} \leq LEVL_{r,"grae"} + LEVL_{r,"grai"}$$

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:

Equation 95
$$\overline{LEVL}_{r,"uaar"} = \overline{LEVL}_{r,"arab"} + \overline{LEVL}_{r,"gras"}$$

And a further one which prevents numerical problems with the terms relating to land supply in the objective function

Equation 96
$$\overline{LEVL}_{r,"uaar"} \leq 0.999\,\overline{LEVL}_{r,"asym"}$$

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, A., Jansson, T., Verburg, P. H., Revoredo-Giha, C., Britz, W., Gocht, A. and McCracken, D.,: Policy reform and agricultural land abandonment in the EU, Land Use Policy, 30 (2013) 446 - 457, doi:10.1016/j.landusepol.2012.04.005.

The obligatory set-aside restrictions introduced by the McSharry reform 1992 and valid until the implementation of the Luxembourg compromise of June 2003 is an explicit restriction in the model:

Equation 97

$$LEVL_{r,"iset"} + LEVL_{r,"gset"} + LEVL_{r,"tset"} = \sum_{arab} LEVL_{r,arab}\left(1 - NONS_{r,arab}\right)\frac{1/100\, SETR_{r,arab}}{1 - 1/100\, SETR_{r,arab}}$$

The somewhat astonishing way the set-aside rate is introduced mirrors the legislation. A set-aside rate of 10% does not imply that for one ha of the crop with the set-aside obligation 0.1 ha of land must be put into set-aside, but that 0.9 ha of the crop must be combined with 0.1 ha of idling land.

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:

Equation 98

$$LEVL_{r,"iset"} + LEVL_{r,"gset"} + LEVL_{r,"tset"} + LEVL_{r,"vset"} \leq \sum_{arab \wedge SETR_{r,arab}} LEVL_{r,arab} \Big/ \overline{MXSETA}$$

4.2.2.3 Fertilising block

The equations below are discussed in the input allocation chapter in more detail. Sufficient to say here that the first line covers nutrient crop needs minus biological fixation of leguminosae, and must be equal to purchases of inorganic fertiliser, reduced by ammonia losses in the case of N, the plant available part of atmospheric deposition in the case of N, and the available nutrients in manure and losses.

For certain crops respectively group of crops, balancing between crop nutrient needs and deliveries from three source (mine – mineral fertilizers, excr – animal excrements and cres – crop residues and atmospheric deposition) is introduced:

Equation 99

$$\begin{aligned}
&\left[\sum_{cact \in ngrp} Levl_{r,cact}\left(Fnut_{r,cact}\right)\left(1 - NFact^{biofix}_{Fnut,cact}\right) NutFac_{r,fnut} + NutFacc_{r,fnut}\right] \\
&= FertDist_{r,ngrp,fnut,"\text{mine}"}\left(1 - EmiLoss^{Netf}_{tot}\right) \\
&+ FertDist_{r,ngrp,fnut,"excr"} NavFac_{r,fnut} \\
&+ FertDist_{r,ngrp,fnut,"cres"} NavFacc_{r,fnut}\left(1 - EmiLoss^{Netf}_{tot}\right) \\
&+ Losses_{r,fnut}
\end{aligned}$$

The crop nutrient retention (FNUT) is corrected for biological fixation in the case of nitrogen (NFACT) and then corrected by two "management factor": a proportional one called NutFac and a constant per ha rate NutFacc. The two can be interpreted as the security margin farmers apply to account for unavoidable losses or in case of higher then expected mean yields.

The distribution activities for the three sources of N, P and K are defined in separate equations. The first one defines nutrient stemming from animal excrements, corrected for losses during storage and application.

Equation 100

$$TotMannNPK_{r,fnut} = \sum_{aact} Levl_{r,aact} Fnut_{r,aact} (1 - EmislossN_{Fnut,r}^{Manure})$$

A separate equation distributes it to the different crops / group of crops:

Equation 101

$$\sum_{ngrp} FertDist_{r,fnut,ngrp,"excr"} = TotMannNPK_{r,fnut}$$

The second one distributes total mineral fertilizer sales to the crops / group of crops:

Equation 102

$$\sum_{ngrp} FertDist_{r,fnut,ngrp,"mine"} = -NETTRD_{r,fnut}$$

Finally, crop residues and atmospheric definition are distributed:

Equation 103

$$\sum_{ngrp} FertDist_{r,fnut,ngrp,"cres"} = \sum_{cact} Levl_{r,cact} Fnut_{r,cact}$$

A second equation ensures that a certain minimum share of the crop need is covered by inorganic fertiliser:

Equation 104

$$\sum_{cact \in ngrp} \left[Levl_{r,cact} \left(Fnut_{r,cact} \right) NutFac_{r,fnut} + NutFacc_{r,fnut} \right]$$
$$MINAN_{r,cact,fnut} \leq FertDist_{r,fnut,ngrp,"mine"}$$

With the latter equation, it would not make any sense to have the "FertDist" variables differentiated by crop or groups of crops, as the objective function does not change if e.g. manure and mineral fertilizer are shifted between crop groups while the nutrient need equation remains binding. However, the very same variable are used in the ex-post fertilizer distribution where external information at least at country level about the per ha rates of certain crops is known.

4.2.2.4 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.

Equation 105

$$\sum_{act} \mathrm{Levl}_{r,act} \mathrm{OUTP}_{r,act,o} = \mathrm{NETTRD}_{r}^{o \notin fodder} + \mathrm{YANUSE}_{r}^{o \in oyani} + \mathrm{FEDUSE}_{r}^{o \in fodder}$$

In the case of quotas (milk, for sugar beet: see section 4.2.2.8) 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:

Equation 106

$$\sum_{aact} Levl_{r,act} I_{r,aact,yani} = YANUSE_{r}^{oyani \Leftrightarrow iyani}$$

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.

4.2.2.5 The objective function

The objective function is split up into the linear part, the one relating to the quadratic cost function for activities and the quadratic cost function relating to the feed mix costs:

Equation 107

$$OBJE = \sum_{r} LINEAR_{r} + QUADRA_{r} + QUADRF_{r}$$

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:

Equation 108

$$\begin{aligned} &\mathrm{LINEAR_r} \\ &= \sum_{\mathrm{io}} \mathrm{NETTRD_{r,io}}\, \overline{\mathrm{PRICE}}_{,\mathrm{io}} + \sum_{\mathrm{act}} \mathrm{LEVL_{r,act}} \left(\overline{\mathrm{PRME}}_{\mathrm{r,act}} - \overline{\mathrm{COST}}_{\mathrm{r,act}} \right) \end{aligned}$$

The quadratic cost function relating to feed is defined as follows:

Equation 109

$$\mathrm{QUADRF_r} = \sum_{\mathrm{aact,feed}} \left[\begin{array}{l} \mathrm{LEVL_{r,aact}\, FEDNG_{r,aact,feed}} \\ \left(\mathrm{a_{r,acct,feed}} + \tfrac{1}{2} \mathrm{b_{r,aact,feed}\, FEDNG_{r,aact,feed}} \right) \end{array} \right]$$

The marginal feed costs per animal increase hence linear with the amount of feed.

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.

4.2.2.6 Premium entitlements in the supply model (W. Britz)

In order to obtain the full premium under the Single Premium, a farmer must not only possess one ha of land in good agricultural and environmental condition, but also a (tradable) payment right. In CAPRI these payment rights are linked to eligible activities. The farmer may expand his activities and produce beyond the level of the initial amount of payment rights, by taking new land into agricultural production. In this case the marginal premium payment per hectare of this new land will be zero. This is because new land has no payment rights attached to it. Technically this is achieved in CAPRI by including in the supply model of CAPRI an equation that calculates the overshooting of the available premium rights:

```
*
*  --- define sum of entitlements
*
   sumEntl_(RUNR,PSDPAY_cutEndog,DDTarget) $ (p_entlLimit(RUNR,PSDPAY_cutEndog,DDTarget,"Limit")) ..

       v_sumEntl(RUNR,PSDPAY_cutEndog,DDTarget)
            =E= sum( (DDReg_MPACT(DDTarget,MPACT),A) $ p_technFact(RUNR,mpact,"LEVL",A),
                          v_actLevl(RUNR,mpact,A));

   overShotEntl_(RUNR,PSDPAY_cutEndog,DDTarget)   $ p_entlLimit(RUNR,PSDPAY_cutEndog,DDTarget,"Limit") ..
      v_overShotEntl(RUNR,PSDPAY_cutEndog,DDTarget) =G= v_sumEntl(RUNR,PSDPAY_cutEndog,DDTarget) - p_entlLimit(RUNR,PSDPAY_cutEndog,DDTarget,"Limit");
```

Basically this means summing up number of hectares of the eligible activities for which each the SPS premium applies and subtracting the available premium entitlements in the initial situation. Next, in the objective function the overshoot amount per SPS payment right is multiplied with the premium per payment right and subtracted from the total profits:

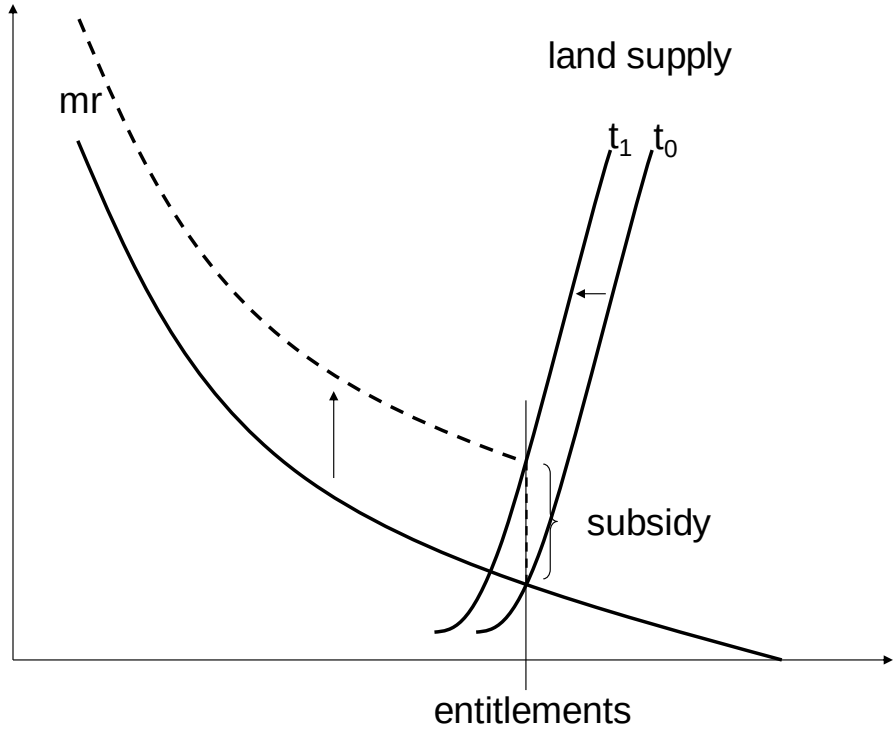
```
*
*      --- endogenous cut of premiums in case of overshot
*
     -  SUM(  (PSDPAY_cutEndog,DDTarget) $ p_entlLimit(RUNR,PSDPAY_cutEndog,DDTarget,"Limit"),
              v_overshotEntl(RUNR,PSDPAY_cutEndog,DDTarget)  * p_entlLimit(RUNR,PSDPAY_cutEndog,DDTarget,"Cut")  );
*
```

That mechanism differs from the McSharry and Agenda 2000 payments (see the Chapter 4.3) as now the overshot is farm specific.

The interrelationship between subsidies and marginal returns in relation to land supplies is shown in the figure below nnd helps in understanding the reactions to changes in the SPS. For simplicity, we do not capture graphically changes in other premium schemes. First the amount of land used in agriculture is determined by the intersection of the marginal returns from agriculture (mr) and the land supply (to) from other sectors. The SPS as a subsidy to agriculture has the effect of shifting the mr curve upward in proportion to the size of the subsidy (dashed line in graphic). Without further restrictions, agricultural land use would be expanded to a new intersection between the two curves. However, due to the introduction of the entitlements, such an expansion does not happen as the old and the new mr curve are identical beyond the entitlement point where the subsidy cannot be claimed and thus has no continued effect on changing land use. Accordingly, immediately after its introduction, the subsidy is capitalized into the premium rights. However, urbanisation and other factors that make land unavailable for production, shift the land supply curve over time to the left (reducing the area available). Depending on the slope of the land supply curve and speed and size of the shift, the economic rent on the entitlements will be eroded over time until the new intersection between the marginal returns and the land supply curves is less than (to the left of) the original entitlements. At that point, the intersection between the dashed line and (t1), the subsidy will be fully capitalised in the land rent. Given that the land supply curves are rather steep in many countries according to our parameterisation and that the continuous decline in agricultural land cover can be observed in almost any EU Member States capitalisation only takes place on land and not on payment rights.

From the other hand, a reduction of the SPS as a subsidy to agriculture, would lead to a reduction of the marginal returns from agriculture. In this case, for example, the dashed line is the old curve mr and the other (closed) line is the new mr curve. Due to the reduction of the subsidy, part of the land ceases to be economically viable. The amount of land in agriculture reduces along the line t1, until the intersection with the new mr curve and part of the entitlements will be unused.

Figure 11. Figure: Effects of SPS on land markets



4.2.2.7 October 2011 CAP reform proposal: Greening constraints (W. Britz)

Without further changes, the new payments would be paid out always exactly as the former different SFP premiums and nothing would have been gained. We need to construct is a mechanism which does not pay out the top-ups in case of non-compliance. As we are facing an opt-in measure, farmers may choose to not comply if the marginal costs of compliance exceed the value of the greening top-up.

The toal cut is the maximum of the violation for the individual measures which are discussed below:

```
   greenOverShot_(RUNR) $ p_entlLimit(RUNR,"DPGREEN","DDGreen","Limit") ..

       v_overShotEntl(RUNR,"DPGreen","DDGreen") =E=

             sum(greeningMeasures $ p_dpGreeningLimits(runr,greeningMeasures),
                    v_greenViol(runr,greeningMeasures));
*
```

We hence need an endogenous mechanism in the model which either pays out the premiums or not, depending on the farm program.

We first define suitable constraints describing the greening requirements as shown below.

Permanent grasslands

```
*
*  --- keep permanment grass lands
*
   permGrasGreening_(RUNR) $ (  p_entlLimit(RUNR,"DPGREEN","DDGreen","Limit")
                              $ p_DPGreeningLimits(runr,"permGras")) ..

     v_greenViol(runr,"permGras")
          =G=  (p_DPGreeningLimits(runr,"permGras") - v_actLevl(runr,"gras","t"))
               /p_DPGreeningLimits(runr,"permGras") * %data%(runr,"UAAR","LEVL","Y");
```

The constraint is triggered if (1) there is entitlement limit for the greening payment in that region (thus e.g. excluding Norway) and (2) if the permanent grassland maintenance requirement is switched on. If the current farm program exceeds the limits, the LHS of that constraint (v_greenVoil) will need to take on a positive value.

Assume the farmers in the region or in the group must have at least 100 ha to fulfill the requirement. If the level drops e.g. to 50, the first term in the bracket of the RHS ( ) take on the value 50 ha. It is then scaled with the limit, i.e. divided by 100 ha and multiplied with total UUA in ha, say 200. We interpret that as if 50% of the farmers do not comply so that 50% of the area does not receive the top-up. We hence deduct for 100 out of 200 ha the top-up from the objective function. "v_overShotEntl" hence counts the non-complying hectares (non-compliance share of the measures times total area).

Crop diversity

The Shannon index depicting crop diversity is only applicable to annual crops which are now defined in *'sets.gms'*:

```
cropDivGreening_(RUNR) $  (  p_entlLimit(RUNR,"DPGREEN","DDGreen","Limit")
                           $ p_DPGreeningLimits(runr,"cropDiv")) ..

  v_greenViol(runr,"cropDiv")
        =G=  (p_DPGreeningLimits(runr,"cropDiv")

           + sum( annualCrops $ sum(a, p_technFact(runr,annualCrops,"levl",a)),

                 sum(a $ p_technFact(runr,annualCrops,"levl",a),
                          v_actLevl(runr,annualCrops,a) + 1.E-10 $ (v_actLevl.lo(runr,annualCrops,a) eq 0))
                         /v_actLevl(runr,"annC","t")
            * log(sum(a $ p_technFact(runr,annualCrops,"levl",a),
                          v_actLevl(runr,annualCrops,a) + 1.E-10 $ (v_actLevl.lo(runr,annualCrops,a) eq 0))
                         /v_actLevl(runr,"annC","t"))))

                       /p_DPGreeningLimits(runr,"cropDiv") * %data%(runr,"UAAR","LEVL","Y");
```

The Shannon index threshold implemented in the analysis is based on single farm record analysis using FADN data.

Ecological set-aside

The third measure is ecological set-aside:

```
*
*  --- ecological set aside, can be fullfilled by fallow land and idling set-aside
*
   ecoSetAside_(runr) $  (  p_entlLimit(runr,"DPGREEN","DDGreen","Limit")
                            $ p_DPGreeningLimits(runr,"ecoSetAside")
                            $ %data%(runr,"ARAB","LEVL","Y")) ..
*
*                                    --- arable crops * relative set aside
*                                    --- obligation: minimum set aside needed
*
     v_greenViol(runr,"ecoSetAside") =G=

        (SUM( (arbAct,A) $ (p_technFact(runr,arbAct,"LEVL",A)
                            and (NOT SETAPACT(arbAct)) and (NOT SAMEAS(arbAct,"FALL"))),
                  v_actLevl(runr,arbAct,A) * 0.01 * p_DPGreeningLimits(runr,"ecoSetAside")/
                                              (1.-0.01 * p_DPGreeningLimits(runr,"ecoSetAside") ))

                - SUM( A $ p_technFact(runr,"ISET","LEVL",A),
                           v_actLevl(runr,"ISET",A))
                - SUM( A $ p_technFact(runr,"USET","LEVL",A),
                           v_actLevl(runr,"USET",A))
                - SUM( A $ p_technFact(runr,"FALL","LEVL",A),
                           v_actLevl(runr,"FALL",A)))
                     /

               SUM( (arbAct,A) $ (p_technFact(runr,arbAct,"LEVL",A)
                         and (NOT SETAPACT(arbAct)) and (NOT SAMEAS(arbAct,"FALL"))),
                             v_actLevl(runr,arbAct,A) * 0.01 * p_DPGreeningLimits(runr,"ecoSetAside")/
                                                 (1.-0.01 * p_DPGreeningLimits(runr,"ecoSetAside") ) )

                    * %data%(runr,"UAAR","LEVL","Y");
```

Winter cover

A fourth discussed measure was winter cover, integrated as a constraint as follows:

```
*
*  --- winter cover
*
   winterCover_(runr) $  (  p_entlLimit(runr,"DPGREEN","DDGreen","Limit")
                              $ p_DPGreeningLimits(runr,"winterCover") ) ..

     v_greenViol(runr,"winterCover") =G=
*
*       --- required cover
*
        %data%(runr,"UAAR","LEVL","Y") * p_DPGreeningLimits(runr,"winterCover")/100
*
*       --- winter cover provided by regular crops
*
      -  SUM( (mcact,A) $ p_technFact(runr,mcact,"LEVL",A),
                     v_actLevl(runr,mcact,A) * p_winterCoverShare(runr,mcact))
*
*       --- winter cover provided by additional catch crop
*
                   -  v_actLevl(runr,"CATC","T");
```

As seen from above, each crop has a certain share to which is provides winter cover (p_winterCoverShare). That allows, e.g. for soft wheat, to define the share cropped as summer or winter cereal, or in the case of olives, to account the share grown in traditional systems where green cover is provided. That share is current very roughly estimated and stored in *'policy\define_greening_limits.gms'*:

```
PARAMETER p_winterCoverShare(RALL,mcact) "Share of crop in a region which provides winter cover" /

     EU027000.SWHE 0.9
     EU027000.DWHE 0.9
     EU027000.BARL 0.9
     EU027000.RYEM 0.7
     EU027000.OATS eps
     EU027000.MAIZ eps
     EU027000.OCER 0.7

     EU027000.RAPE 0.9
     EU027000.SUNF eps
     EU027000.SOYA eps

     EU027000.OOIL eps
     EU027000.OIND 0.5


     EU027000.PULS 0.2
     EU027000.SUGB eps
     EU027000.POTA eps
     EU027000.TEXT eps
     EU027000.TOBA eps

     EU027000.OLIV 0.2
     EU027000.TOMA eps
     EU027000.OVEG eps
     EU027000.APPL 0.2
     EU027000.CITR 0.2
     EU027000.TAGR 0.2
     EU027000.TABO 0.2
     EU027000.TWIN 0.2

     EU027000.NURS eps
     EU027000.FLOW eps
     EU027000.OCRO eps
     EU027000.NECR 1.0



     EU027000.FALL 1.0
     EU027000.ISET 1.0
     EU027000.GSET 1.0
     EU027000.TSET 1.0
     EU027000.VSET 1.0



     EU027000.ROOF eps
     EU027000.OFAR 0.9
     EU027000.GRAI 1.0
     EU027000.GRAE 1.0

/;
```

4.2.2.8 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. 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 forme 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 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 108:

Equation 110

$$
\begin{aligned}
&SugbREV_r \\
&= p^A NETTRD_{r,SUGB} \\
&-\left(p^A - p^B\right)\begin{bmatrix}\left(1-CDFSugb\left(q^A\right)\right)\left(NETTRD_{r,SUGB}-q^A\right)\\ +\left(\sigma^S\right)^2 PDFSugb\left(q^A\right)\end{bmatrix} \\
&-\left(p^B - p^C\right)\begin{bmatrix}\left(1-CDFSugb\left(q^{A+B}\right)\right)\left(NETTRD_{r,SUGB}-q^{A+B}\right)\\ +\left(\sigma^S\right)^2 PDFSugb\left(q^{A+B}\right)\end{bmatrix}
\end{aligned}
$$

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^{ABC} 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 110 are the "behavioural quotas", currently specified as follows:

Equation 111

$$q^A = legalquot^A \cdot scalefac$$
$$= legalquot^A \cdot \left(\frac{NETTRD^{cal}_{SUGB}}{legalquot^A} \right)^{0.8}$$

The scaling factor to map from the legal quota $legalquot^A$ (as the B quota has been eliminated in the sugar reform, it holds that $q^A = q^{A+B}$) to the behavioural quota q^A 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 production, 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.

4.2.2.9 Update note

Furthermore at the time of the biofuel incorporation, simulation experiments revealed often convergence problems with feed demand. An analysis showed that a probable cause is the overestimation of the implicit feed demand elasticities inside the supply models. There are cost terms incorporated in the objective function of the supply model which characterize additional costs linked of changing the feed mix. Those had been defined in the past from rather large elasticities (-10%) and a quantity equivalent to at least 10% of the dry matter. That calculation rule is now changed, and the 10% are reduced if the total dry matter of a feedingstuff in the regional feed use is below 10%. That dampens the changes of feed ingredient used in small amounts and helps in achieving corvergence.

4.2.3 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.

4.2.4 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 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.

4.2.5 Price depending crop yields and input coefficients (W. Britz)

Let Y denote yields and j production activities. Yield react via iso-elastic functions to changes in output prices

$$\log(Y_j) = \alpha_j + \varepsilon_j \log(p_o)$$

Equation 112

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 & Hertel 2008(a,b). 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 α, the actual code implemented in *'endog_yields.gms'* change the yields iteratively in between iterations t , using relative changes:

$$Y_{j,t} = Y_{j,t-1}^{\left[\varepsilon_j \log \frac{Po,t-1}{p_t}\right]}$$

Equation 113

4.3 Premium module (T. Jansson, W. Britz, J. Helming)

4.3.1 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 SFP of the CAP is explicitly considered, as are the different national SFP implementations, possibly remaining coupled payments for suckler cows and sheep & goat, and national "specific support programs" under article 68 of Council Regulation (EC) No. 73/2009.

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 dis-aggregation of the model template regarding production activities and the resolution by farm types inside of NUTS 2 regions clearly eases that task. Currently, more than 30 coupled premium schemes are differentiated, in addition to decoupled income support (SFP). The schemes cover almost all payments within the first pillar of the CAP, such as premiums to arable land ("Grandes Cultures"), durum wheat premiums in traditional regions, durum wheat premiums in established regions, rice premiums, premiums to pulses, premiums to energy crops, silage premiums in Sweden and Finland, suckler cow premiums, direct premiums to dairy cows, extensification premiums to bulls, steers and suckler cows, payments to sheep and goats, and supplementary payments to sheep and goats, to olives, to fruits and vegetables, to the wine sector, to tobacco, to cotton, to starch potatoes; different type of Nordic aid premiums in Northern Sweden and Finland as well as complementary national premiums in the new Member States during the transition period. Important payments of the second pillar of the CAP (Less Favourite Area support, agri-environmental schemes, Natura 2000 support) are covered.

4.3.2 Basic concept

In the CAPRI supply module, premiums are always paid per activity level (per hectare or per ha) 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
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).

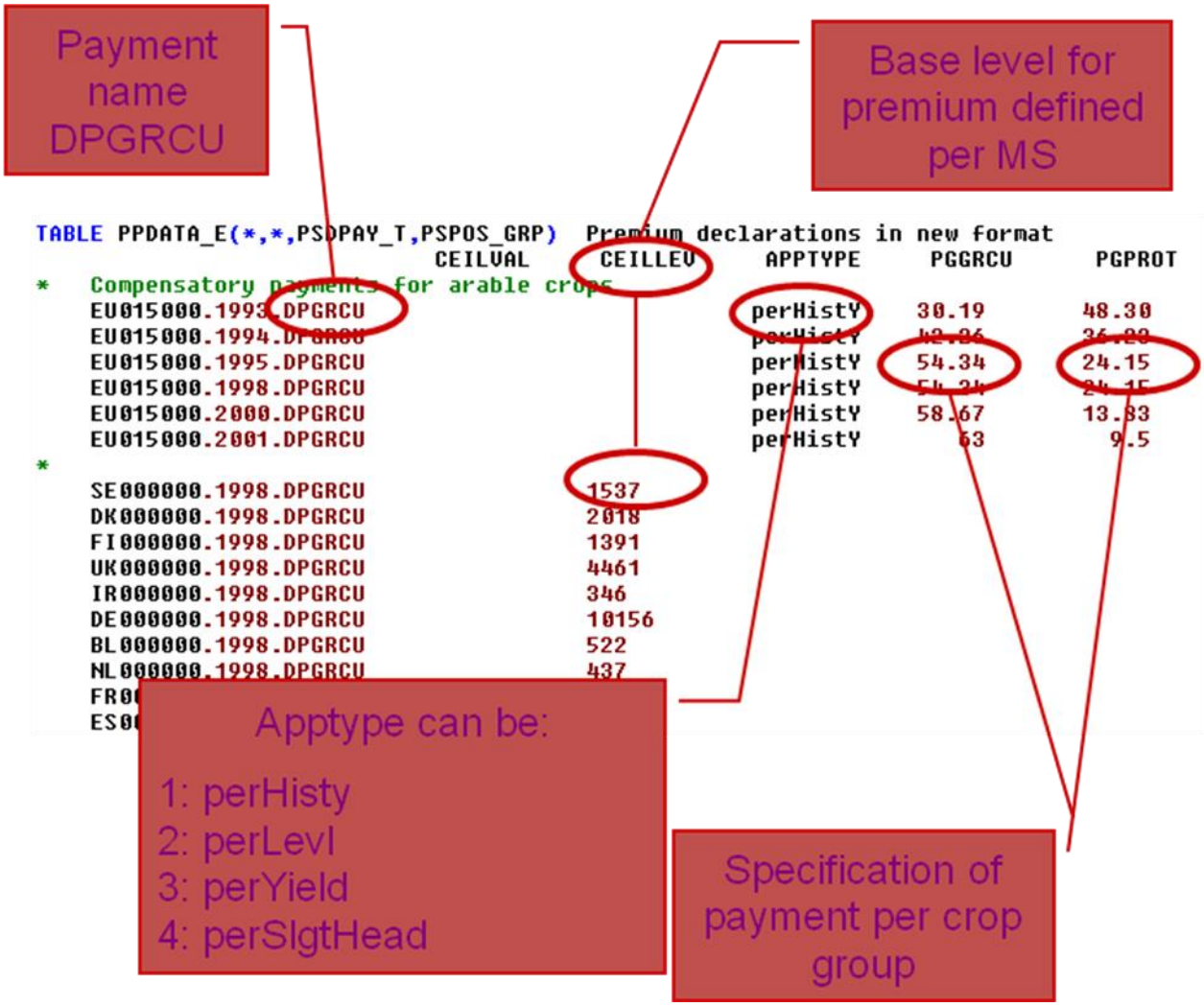
The scheme provide hence many to many mappings between policy instruments and agricultural activities: each scheme can apply to many different activities – with possibly differentiated rates – and each activity can draw support from different schemes.

Currently, the following application types are supported:

- perLevl define per ha or head
- perYield per unit of main output
- perHisty per historic yield
- perSlgtHd per slaughtered head

- perLiveStockUnit per livestock unit

Additionally, the payment can be applied in different rates to the high and low yield variant, to model e.g. an extensification premium. The following scheme shows the technical implementation at an example:



Note: The parameter is now called p_premDataE as shown below.

And overview on the payment scheme operating for pillar I is shown below

```
*
*       --- premiums McSharry and Agenda 2000
*
        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 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"
*
*       -- and a new one!
*
        DPEUFLAT        "EU wide single flat rate premium per ha of land"
*
*       -- 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",
```

During the phasing in of the CAPI in the new Member States, so called complementary national direct payments are implemented by some of the new Member States:

```
*
*       -- national top-ups in new Member States during the intermediate period after accession
*
        DPCN_ARAB       "Complementary National Direct Payment for arable crops"
        DPCN_CATT       "Complementary National Direct Payment for cattle"
        DPCN_EWE        "Complementary National Direct Payment for ewes"
        DPCN_PARI       "Complementary National Direct Payment for rice"
        DPCN_TOBA       "Complementary National Direct Payment for tobacco"
        DPCN_BEEF       "Complementary National Direct Payment for beef"
        DPCN_ADCT       "Complementary National Direct Payment for adult cattle"
        DPCN_SCOW       "Complementary National Direct Payment for suckler cows"
        DPCN_DCOW       "Complementary National Direct Payment for dairy cows"
        DPCN_FODD       "Complementary National Direct Payment for fodder area"
```

And finally, a first representation of pillar 2 payments is included:

```
SET s_PSDPAY_AE(*) "Agri-environmental payments" /
     DPRD_AE1       "Agri-environmental payments for field crops"
     DPRD_AE2       "Agri-environmental payments for horticulture"
     DPRD_AE3       "Agri-environmental payments for wine"
     DPRD_AE4       "Agri-environmental payments for permanent crops"
     DPRD_AE5       "Agri-environmental payments for dairy farms"
     DPRD_AE6       "Agri-environmental payments for grassland farms"
     DPRD_AE7       "Agri-environmental payments for pig farms"
     DPRD_AE8       "Agri-environmental payments for mixed farms"
     /;


SET s_PSDPAY_RD(*) "Rural development payments" /
     DPRD_LFA       "Less favoured area payment"
     DPRD_N2K       "Support in Natura 2000 area",
     SET.s_PSDPAY_AE
     /;
```

Additional, remaining coupled payments under the so-called article 68 are covered:

```
SET s_PSDPAY_68(*) "Direct aid under article 68 of Healt Check regulation CR73-2009" /
     dp68milk       "Article 68 support to milk production"
     dp68graz       "Article 68 support to grassland"
     dp68shgt       "Article 68 support to sheep and goats, various animals"
     dp68oliv       "Article 68 support to oliv production"
     dp68dwhe       "Article 68 support to durum wheat"
     dp68sugb       "Article 68 support to sugar production"
     dp68scow       "Article 68 support to suckler cows"
     dp68lfa        "Article 68 support to LFA land use"
     dp68toba       "Article 68 support to tobacco"
     dp68proi       "Article 68 support to protein crops and oil seeds"
     dp68beef       "Article 68 support to beef and veal meat production"
     dp68hort       "Article 68 support to horticultural sectors"
     dp63gras       "Article 63 support to grassland farming and forage crops in France"
     dp63pota       "Article 63 support to potatoes in France"
     dp63vege       "Article 63 support to vegetables in France"
     dp63fodc       "Article 63 support to fodder crops in France"

  /;
```

We refrain from showing the payment schemes for Norway and for Nordic aid explicitly which are also covered by the model.

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 perLev1)
            + 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);
```

Generally, all attributes for a premium scheme are mapped down in space, e.g. from EU27 to EU 27 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:

```
*
*   --- calculate application factor
*
    p_premPaym(RSEU,PPPACT,PSDPAY,"Ap_premModifFactT") $ ( (p_premPaym(RSEU,PPPACT,PSDPAY,"PRMR") GT eps) $ PSDPAY_T_A(PPPACT,PSDPAY))
        =
*     --- apptype perLev1: 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_LVUnits(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");
```

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).

```
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 schemes 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"));
```

4.3.3 A didactic example and further explanations

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.

(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 (e.g. EU15). In the example, this means that it is 80 EURO for dairy cows, suckler cows, male adult cattle and fattened heifers in all regions of EU15 in the the model and null elsewhere. 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. Thses numbers are taken 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 in each Agenda 2000 member state. 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[62], 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.

4.3.4 Mc Sharry and Agenda reform premiums

Coupled direct payments were introduced by the 1992 Reform of the Common Agricultural Policy (CAP) of the European Union (EU), also known as the Mac Sharry reform. It affected market regulations for cereals, oilseeds, tobacco, milk, beef and lamb. Administrative minimum prices were decreased and as compensation producers received direct payments in fixed amounts per hectare or head. Regional reference acreages and herd sizes where installed, which, if exceeded reduced the number of hectares or animals per producer that receive payment. That implementation, along with obligatory set-aside, let the premiums fit to the blue box definition of the WTO Agreement on Agriculture.

[62] There are exemptions for that rule, see below for the section on entitlements.

In March 1999, the EU Heads of Governments and States accepted a further reform of the CAP known as the Agenda 2000 CAP reform. To comply with the principle of budgetary discipline, the Agenda 2000 CAP reform included a further reduction of minimum prices in the cereals, beef and milk sector. Another important element was the alignment of the special regime of oilseeds to that of cereals. In doing so the area payment for cereals increased from € 58.67 per ton to € 63 per ton, while the area payment for oilseeds decreased from € 94.24 per ton to €63 per ton. In the beef sector direct payments increased for both male animals and suckler cows, subject to national ceilings, to compensate a drop in the administered price for the beef price cuts for both male animals and suckler cows, subject to national ceilings. A new slaughter premium was introduced of € 80 for all the animals over the age of 8 months. The price cuts in the dairy sector were compensated by a subsidy to be paid per ton of milk quota, based on the quota held by each producer in the 1999/2000 year. Milk quotas were maintained and increased amounting to an overall increase of 2.4% in the EU by the end of the implementation of the reform in 2008.

The premiums and policy changes discussed above are defined in the file *'pol_input\agenda.gms'*. The reader should note that the edited premiums as a time series are stored on the parameter p_premDataE. The definition of the premium rates of the Grandes Cultures premiums is uniform for all EU Member states, e.g.:

```
*
TABLE p_premDataE(*,*,PSDPAY_T,PSPOS_GRP)  "Premium declarations"
                              CEILVAL     CEILLEV     APPTYPE     PGGRCU      PGPROT
*   Compensatory payments for arable crops
    EU015000.1993.DPGRCU                                perHistY    30.19       48.30
    EU015000.1994.DPGRCU                                perHistY    42.26       36.23
    EU015000.1995.DPGRCU                                perHistY    54.34       24.15
    EU015000.1998.DPGRCU                                perHistY    54.34       24.15
    EU015000.2000.DPGRCU                                perHistY    58.67       13.83
    EU015000.2001.DPGRCU                                perHistY      63          9.5
```

However, there are country specific ceiling in levels:

```
*
TABLE p_premDataE(*,*,PSDPAY_T,PSPOS_GRP)  "Premium
                    CEILVAL     CEILLEV
*   Compensatory payments for arable crops
-------------------------------------------------------
*
*    data multiplied with 1.20755 to change green p
*    beware: changes in agri-monetary system requir
*
*
    SE000000.1993.DPGRCU                1537
    DK000000.1993.DPGRCU                2018
    FI000000.1993.DPGRCU                1391
    UK000000.1993.DPGRCU                4461
    IR000000.1993.DPGRCU                346
    DE000000.1993.DPGRCU                10156
    BL000000.1993.DPGRCU                522
    NL000000.1993.DPGRCU                437
    FR000000.1993.DPGRCU                13526
    ES000000.1993.DPGRCU                9220
    PT000000.1993.DPGRCU                1035
    AT000000.1993.DPGRCU                1203
    EL000000.1993.DPGRCU                1492
    IT000000.1993.DPGRCU                5801
```

And the history yields vary by regions and are found in the file *'data\policy\hsty_setr.gms'*. The file comprises data for many other premium schemes of which the names are shown above.

A specific problem with the implementation in an aggregated model such as CAPRI is to deal with farm specific condition impacting the payment rates. There were some schemes with farm specific ceilings and others in the beef sector where premium rates depended on stocking rates. The latter created an incentive to expand fodder areas which were otherwise no eligible for premiums under the McSharry reform. Switching to e.g. alternative implementation such as the decoupled payments under the MTR would hence overestimate the allocation effect. To account for that, a share of certain ruminant payments is paid to fodder areas as shown below for the example of the suckler cow premium.

First of all, the reader should now that heifers were up to a certain share eligible for suckler cow premiums. That is accounted for by the additional "p_entlWght1" parameter: where all

suckler cows (=1) are eligigle, heifers receive in average only 20% (0.2) of the premium. At farm level, a heifer is either eligible or not, but in an aggregated agent model, such binary decisions at agent level cannot be modelled.

In a second step, the resulting entitlement (= number of animals in the base period weighted with the eligibility factor) are corrected with a factor which shifts part of the incentives from the animals to the permanent grass lands.

```
*
*   ----- 50% of suckler cow premiums are mapped to grass land
*
    p_entlWght1("EU015000","SCOW","DPSCOW") = 1.0;
    p_entlWght1("EU015000","HEIR","DPSCOW") = 0.2;
    p_entlWght1("EU015000","HEIH","DPSCOW") = 0.2;
    p_entlWght1("EU015000","HEIL","DPSCOW") = 0.2;
*
    p_premModifFact(RUNR,SIMY,"SCOW","DPSCOW") $ (CALYEA(SIMY) GE 1992) = 0.5;
    p_premModifFact(RUNR,SIMY,"HEIR","DPSCOW") $ (CALYEA(SIMY) GE 1992) = 0.1;
    p_premModifFact(RUNR,SIMY,"HEIH","DPSCOW") $ (CALYEA(SIMY) GE 1992) = 0.1;
    p_premModifFact(RUNR,SIMY,"HEIL","DPSCOW") $ (CALYEA(SIMY) GE 1992) = 0.1;
*
    p_premModifFact(RUNR,SIMY,"GRAI","DPSCOW") $ ((DATA(RUNR,"GRAE","LEVL","BAS")+DATA(RUNR,"GRAI","LEVL","BAS"))
                                        $ (CALYEA(SIMY) GE 1992))
        = (    (1.0-p_premModifFact(RUNR,SIMY,"SCOW","DPSCOW"))  * DATA(RUNR,"SCOW","LEVL","BAS")
             + (0.2-p_premModifFact(RUNR,SIMY,"HEIR","DPSCOW"))  * DATA(RUNR,"HEIR","LEVL","BAS")
             + (0.2-p_premModifFact(RUNR,SIMY,"HEIH","DPSCOW"))  * DATA(RUNR,"HEIH","LEVL","BAS")
             + (0.2-p_premModifFact(RUNR,SIMY,"HEIL","DPSCOW"))  * DATA(RUNR,"HEIL","LEVL","BAS"))
            /(DATA(RUNR,"GRAE","LEVL","BAS")+DATA(RUNR,"GRAI","LEVL","BAS"));
    p_premModifFact(RUNR,SIMY,"GRAE","DPSCOW") = p_premModifFact(RUNR,SIMY,"GRAI","DPSCOW");
```

It is one of several examples where a direct translation of the legal text into a policy instrument in CAPRI is not possible.

4.3.5 The MTR-reform and the health check

4.3.5.1 Overview on reform process

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.

- 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:
    a. 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,
    b. reduction of the monthly increments in the cereals sector by half, the current intervention price will be maintained,
    c. 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.

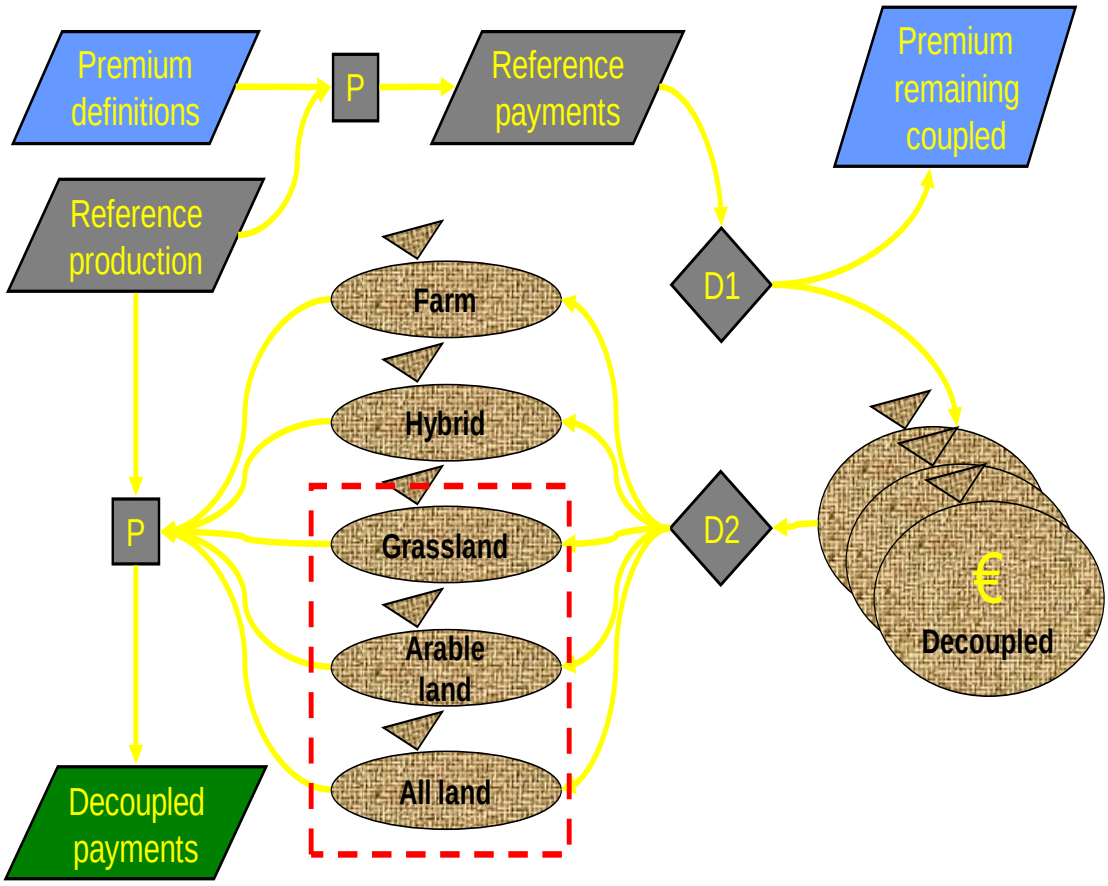
Under the HC of the CAP it was decided that remaining coupled payments should be decoupled and moved into the Single Payment Scheme (SPS), with the exception of suckler cow, where Member States may maintain current levels of coupled support. Moreover, member states are allowed to review the decision 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.

4.3.5.2 Technical implementation

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.

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.

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:

```
*
*   --- 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;
```

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:

```
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                                                                 17.90
    LV000000.2004.DPCN_FODD     7.075     perLevl                                                                                80.00
    LV000000.2004.DPCN_SCOW     2.684     perLevl          138.57
    LV000000.2004.DPCN_ADCT     9.946     perSlgtHd
    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 /;
```

These top-ups have to be reduced towards the end of the period where the the Pillar I premiums are phased in:

```
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;
```

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:

```
*
   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
```

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)

```
*
      PGARAB.(SWHE,DWHE,BARL,RYEM,OATS,OCER,MAIZ,PARI,RAPE,SUNF,SOYA,PULS,ISET,GSET,TSET,USET,MAIF,
              POTA,SUGB,TEXT,TOBA,OIND,OOIL,ROOF,OFAR,FALL,TOMA,OVEG,APPL,CITR,OFRU,OLIV,NURS,FLOW,TWIN,OCRO,TAGR,TABO,NECR)
```

- And permanent grass land (PGGRAS) is eligble

```
      PGGRAS.(GRAE,GRAI)
```

- The historic implementation

```
*
*       --- the single farm payment scheme (after inclusion of med. crops)
*
        PGFRM.(SWHE,DWHE,BARL,RYEM,OATS,OCER,MAIZ,PARI,RAPE,SUNF,SOYA,PULS,USET,ISET,GSET,TSET,MAIF,
               SUGB,TEXT,OIND,OOIL,ROOF,OFAR,GRAE,GRAI,FALL,OLIV,TOBA)
```

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"

```
    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
```

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.

```
*
   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;
```

There two other factors:

- A possible greening share according to the October 2011 proposal by the Commission
- 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.

The part which is not longer coupled goes into the decoupled schemes:

```
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);
```

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:

```
--- times the decoupling factor (1 minus the coupling factor)
                        *  (1 - p_couplPercent(PSDPAY_MTR_EL,SIMY,MS)/100
```

After that, a possible share going into the greening payment is deducted:

```
* (1.-p_dpGreeningShare(MS))
```

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:

```
* (1+p_nationalCeilCut(MS,SIMY))
```

That sum if then distributed to the relevant MTR implementation scheme according to the distribution 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'*.

4.3.6 The October 2013 proposal by the Commission: Greening top-up

The available measures are defined in *'sets.gms'*:

```
set set_greeningMeasures /
                      permGras      "Maintainance of permanent grass land"
                      cropDiv       "Crop diversity"
                      winterCover   "Winter cover in crop land"
                      ecoSetAside   "Ecological set-aside"
                     /;
```

The active measures in a run can be activated over the following set defined as:

```
activeGreeningMeasures("permGras") = YES;
activeGreeningMeasures("cropDiv")  = YES;
```

The top-up share, i.e. the amount of the single farm payment which paid out conditioned on compliance with the greening measures, is defined via the following parameter (see *'pol_input\mtr_green.gms'*)

```
p_DPGreeningShare("EU027000") = 0.3;
```

That share can be set at EU, MS and regional level to simulate different top up percentages for different regions. It will "deroute" part of the MTR payments (SFP, remaining coupled premiums) into the top-up (see *'policy\calc_mtr.gms'*):

```
p_premDataE(RSEU,"%SIMY%","DPGreen",PSGROUP) $ p_dpGreeningShare(RSEU)
   = sum(PSDPAY_MTR $ (NOT SAMEAS("DPGreen",PSDPAY_MTR)),
      (                                                    p_premDataE(RSEU,"%SIMY%",PSDPAY_MTR,PSGROUP)
       + sum(R_RAGG(RSEU,SRNUTS1) $ (Not sameas(RSEU,SRNUTS1)), p_premDataE(SRNUTS1,"%SIMY%",PSDPAY_MTR,PSGROUP) ))

                          * p_dpGreeningShare(RSEU));
```

And, accordingly, a maximum value ceiling for the top up is defined:

```
p_premDataE(RSEU,"%SIMY%","DPGreen","CEILVAL") $ p_dpGreeningShare(RSEU)
   = sum(PSDPAY_MTR $ (NOT SAMEAS("DPGreen",PSDPAY_MTR)),
      (                                                    p_premDataE(RSEU,"%SIMY%",PSDPAY_MTR,"CEILVAL")
       + sum(R_RAGG(RSEU,SRNUTS1) $ (Not sameas(RSEU,SRNUTS1)), p_premDataE(SRNUTS1,"%SIMY%",PSDPAY_MTR,"CEILVAL")))

                          * p_dpGreeningShare(RSEU));
```

Payments and value ceilings (not shown as GAMS code) of the original premium schemes needs now be lowered:

```
p_premDataE(RSEU,"%SIMY%",PSDPAY_MTR,PSGROUP)    $ (   p_premDataE(RSEU,"%SIMY%",PSDPAY_MTR,PSGROUP)
                                                     $ (NOT SAMEAS("DPGreen",PSDPAY_MTR))
                                                     $  p_dpGreeningShare(RSEU))
   = p_premDataE(RSEU,"%SIMY%",PSDPAY_MTR,PSGROUP)   * (1. - p_dpGreeningShare(RSEU));
```

4.3.7 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 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.

4.3.7.1 Implementation in the code

Switching on the entitlement

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';
*
```

Policy files such as *'mtr_conv.gms'* can switch on the module:

```
$setglobal entl_trade on
```

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:

```
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);
```

From these a maximum of 10% is defined as the demand in each iteration:

```
*
*    --- max 10 % of existing entitlements
*
  p_entlTrade(RU,"demand",step) =  p_entlTrade(RU,"demand",step) * 0.1;
```

In order to take differences in the marginal returns into account, an indicator based on the squared value is used:

```
p_entlTrade(RU,"valueSqr",step) $ p_entlTrade(RU,"demand",step)
    = p_entlTrade(RU,"demand",step) * sqr(p_entlTrade(RU,"price",step));
```

It serves as the distribution key of unused entitlements, which are determined as follows:

```
*
* --- 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:

```
*
* --- 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;
```

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.

4.3.8 Pillar II

The pillar II instruments are grouped in two categories:

(1) Payments which are similar in the intervention logic to

4.3.9 Co-financing rates, assignment of premiums to pillars, WTO boxes and PSE-types

EU and national budget contribution

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:

```
    ,
    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;
```

Currently, the following budget categories are supported (see *'sets.gms'* and *'policy\policy_sets.gms'*):

```
   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
    /;
```

In *'reports\feoga.gms'*, these categories are first aggregated for each activities from actual schemes ("PRME" = actual payment rate, p_budToPsdpay: distribution key):

```
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));
```

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:

```
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 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 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)

4.4 Market module for young animals (W. Britz)

The market module for young animals ensures closed balances for piglets, calves etc. at European level. The individual regional models may sell or buy young animals in unlimited quantities at fixed prices during each iteration. The market module must hence generate prices which lead to an equilibration of regions with excess demand and such with excess supply of young animals.

The first trials were based on a simple algorithm which was changing prices as a function of excess demand or supply at European level. However, especially due to the high interdependencies inside the cattle chain, there are important cross-price effects, which could not be sorted out with a simple approach. That left the team with two possible competitors: a kind of multi-commodity model for young animals, where the parameters would need to be estimated from simulation experiments with the regional supply models, or a framework building directly on the regional programming models. For years, the latter one was active: it required computationally expensive aggregation of the regional constraint matrices to national ones and a calibration of the national models based on PMP. These were then linked together and solved, and the marginal values of the market clearing equations for young animals were exploited to derive price changes.

Finally, the original idea of using a square system of equations where price changes would drive young animal demand and supply in behavioural equations was revitalised again. The difference to the situation in 1996-1999 is the fact that since Heckelei 2002 the functional relation between the supply responses, the quadratic cost function terms used in the regional programming models and the coefficients relating to binding constraints is known:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{gm}} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{A}'\left(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\right)^{-1}\mathbf{A}\mathbf{Q}^{-1}$$

Equation 114

Where **x** denotes the decision variables, **gm** the vector of gross margins, **A** the constraint coefficient matrix and **Q** the quadratic terms of the variable costs function of the regional programming models. Accordingly, the supply effects can be calculated from the parameters of the regional programming models, and the animals market cleared as:

$$\sum_i \left(x_i^0 + \sum_{nj} \Delta p_n IO_{jn} \frac{\partial x_i}{\partial gm_j} IO_{im} \right) = 0 \wedge m$$

Equation 115

Where *m,n* are the young animals, *i* the production activities, and **IO** the input output coefficients for young animals. 0 denotes the current solution returned from the regional programming models. The term $\Delta p_n IO_{jn}$ captures the absolute change in the gross margin of the activity j if the price of the young animal type *n* is changed, and $\frac{\partial x_i}{\partial gm_n} IO_{in}$ converts the change in the activity level of activity *i* into a change of supply or demand for young animal animals.

The implementation in CAPRI is based on two relatively small programs. *'supply\simu_yani_model.gms'* defines the square system as a NLP problem with a dummy objective , and the parameters and sets used. *'supply\simu_yani_new.gms'* first calculated a simplified **A** matrix from the coefficient matrix of the regional models, and aggregates them to Member State level. The simplification consists especially in using fixed feed input coefficients. Arable, grass land, the set-aside constraint, fodder balances and the milk quota are taken into account. During the first iteration, the regional **Q** matrices are inverted, and then summed up to Member State level. Due to updates of the **A** matrix, the term $-\mathbf{Q}^{-1}\mathbf{A}'\left(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\right)^{-1}\mathbf{A}\mathbf{Q}^{-1}$ is recalculated for the Member States in each iteration, which requires an additional inversion.

As the Δp_n are only variables in linear system, and currently still defined at the level of country blocks as EU15, the solver defines a solution in less then a second. The calculation of (2) are more demanding, especially the inversion of the **Q** for all regions requires close to one minute for all 252 regions currently covered, but is only necessary once. The remaining calculations take around 5 seconds, so that animal markets can now be solved in about 10%-20% of the previously required time on a powerful four processor machine. Due to the modifications, a typically iteration in CAPRI now requires around 90 seconds if the market model does not require a pre-solve.

Value added for the global market model

The behavioural functions in the global market model are derived from a non-symmetric normalised quadratic cost function with the price index for non-agricultural products used the normalisation price. Accordingly, the supply functions are linear in the prices of the global market, and the supply effects $\frac{\partial S_x}{\partial p_y}$ for the products in the market model *x*,*y* can be defined by $\sum_{yij} IO_{jy} \frac{\partial x_i}{\partial gm_j} IO_{ix}$. The first term IO_{jy} captures again the effect of a change in price *y* on the gross margin of activity *j*, where as the second term $\frac{\partial x_i}{\partial gm_j} IO_{ix}$ captures the marginal increase of output *x* due to the change in activity level *i*. It is hence possible to update parameters of the supply functions in the market model in each iteration based on the terms used in young animal market models, which promises to speed up and ease solution of the market model.

4.5 Market module for agricultural outputs (W. Britz, M. Adenäuer)

4.5.1 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 for agricultural world markets are available as OECD's AgLink, the FAPRI system at the University of Missouri or the WATSIM[63] system at Bonn University, it was still considered necessary to have an independent market module for CAPRI.

[63] 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.

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 50 commodities. 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 bi-lateral 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 75 countries or country aggregates, each featuring systems of supply, human consumption, feed and processing functions. The regional-break down is a compromise between full coverage of all individual countries (there are close to 200 UN member countries) and computational respectively data quality considerations. Most countries which a high population number are covered independently, sometimes they aggregated with smaller countries. Given the EU focus of CAPRI, agricultural trading partners of the EU are captured in some detail.

The processing can be differentiated by processing of oilseeds to cakes and oils, of bio-fuel feedstocks to bio-fuels, and of raw milk to dairy products and a 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 parameter 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 70 countries are blocked to country aggregates with a uniform border protection, and bilateral trade flows are modelled solely between these blocks. Such blocks are the EU15, EU10, 'other Mediterranean' countries, Uruguay and Paraguay, Bolivia, Chile and Venezuela, Western Balkan countries, and an aggregate of Bulgaria and Romania. All other countries or country aggregates are identical to one of the 40 trade blocks in the model.

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 current no data are introduced. Tariff Rate Quotas (TRQs), bil-laterally 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.

The three EU blocks (EU15, EU10, Bulgaria & Romina) interact via trade flows with the remaining 37 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. Alternatives are currently analysed in the context of the CAPRI-RD project.

The market model in its current layout comprises about 70.000 endogenous variables and the identical number of equations.

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.

4.5.2 The approach of the CAPRI market module

Multi-commodity models are as already mentioned above a wide-spread 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 – as 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 FAPRI 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. Besides FAPRI, other examples of such models are AGLink-COSIMO at the OECD_FAO 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 50 products and 75 countries or country blocks world wide with the resource 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.

4.5.3 Behavioural equations for supply, feed demand and land markets

4.5.3.1 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:

Equation 116

$$v_prodQuant_{i,r} = as_{i,r} + \sum_j bs_{i,j,r} \frac{v_prodPrice_{j,r}}{p_{index.r}}$$

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 above).

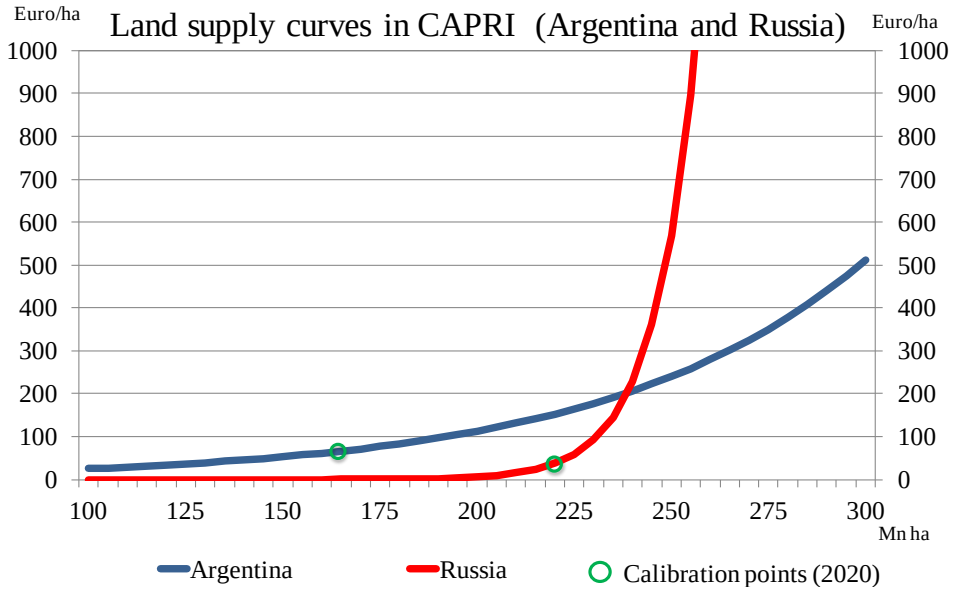
4.5.3.2 Land supply and demand

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.

Equation 117

$$v_prodQuant_{r,"Land"} = p_cnstLandSupply_r + p_landSupplyElas_r \log\left(v_prodPrice_{r,"Land"}\right)$$

Figure 12. Land supply curve examples



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:

$$dLand/dp_i = \sum_j dQ_j/dp_i \; 1/yield_j - ela_i \, Q_i/p_i \; 1/yield_i$$

Equation 118

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.

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 (AMER AHMED, S., HERTEL, T.W. AND R. LUBOWSKI (2008). Calibration of a Land Cover Supply Function Using Transition Probabilities. GTAP Research Memorandum No. 14 October 2008.). Generally, land supply is rather inelastic so that price increase of agricultural commodities rather increase 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 is however used in a post model analysis 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.

4.5.3.3 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*:

Equation 119 $$v_feedQuant_{i,r} = \left(af_{i,r} + \sum_j bf_{i,j,r} \frac{v_arm1Prices_{j,r}}{p_{index.r}} \right)$$

One of the price 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:

Equation 120 $$\sum_i v_feedQuant_{i,r} \; p_calContent_r = \sum_i v_prodQuant_{i,r} \; pv_feedConv_{i,r}$$

The 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, Bulgaria and Romania 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 dis-advantage 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.

4.5.4 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*[64] where *G* and *F* are functions of degree zero in prices (RYAN & WALES 1996) which will be defined below[65]:

Equation 121 $$U(cpri, y) = \frac{-G}{(y - F)}$$

Using Roy's identity, the following per capita Marshallian demands *PerCap* are derived:

Equation 122 $$PerCap_i = F_i + \frac{G_i}{G}\left(y - F\right)$$

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:

Equation 123 $$F_r = \sum_i d_i \; cpri_i$$

---

[64] 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.

[65] 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.

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 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:

Equation 124 $$G = \sum_i \sum_j bd_{ij} \sqrt{cpri_i \; cpri_j}$$

with the derivative of *G* versus the own price is labelled G_i and defined as:

Equation 125 $$G_j = \sum_i bd_{ij} \sqrt{cpri_i / cpri_j}$$

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:

Equation 126
$$\sum_i PerCap_i \; cpri_i = \frac{\sum_i G_i cpri_i}{G}(y - F) + \sum_i d_i cpri_i ,$$
$$= \frac{G}{G}(y - F) + F = y$$

and homogeneity is guaranteed by the functional forms as well. The expenditure function follows from rearranging Equation 121:

Equation 127 $$y = e(U, cpri) = F - \frac{G}{U}$$

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_{ij} $(cpri_i \; cpri_j)^{0.5}$ 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:

Equation 128 $$hcom_{i,r} = pop_r \; perCap_{i,r}$$

4.5.5 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:

Equation 129
$$v_procMarg_{i,r} = ac_{i,r} + \sum_j bc_{i,j,r} \frac{v_procMarg_{j,r}}{p_{index.r}}$$

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*:

Equation 130
$$\begin{aligned} v_procMarg_{seed,r} = & -v_arm1Price_{seed,r} \\ & + v_prodPrice_{seed \to cak,r} v_procYield_{cak,r} \\ & + v_prodPrice_{seed \to oil,r} v_procYield_{oild,r} \end{aligned}$$

Finally, output of oils and cakes *supply* depends on the processed quantities *proc* of the oilseeds and the crushing coefficients:

Equation 131
$$\begin{aligned} supply_{cake,r} &= proc_{seed,r} v_procYield_{cak,r} \\ supply_{oil,r} &= proc_{seed,r} v_procYield_{oil,r} \end{aligned}$$

The processing yields in the base year are defined as:

Equation 132
$$v_procYield_{cak,r} = \frac{v_prodQuant^{bas}_{seed \to cak,r}}{v_prodQuant^{bas}_{seed,r}}$$

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** *cont* of raw milk and milk products *mlk* is based on statistical and engineering information, and kept constant at calibrated base year levels.

Equation 133
$$v_prodQuant_{"milk",r}\ cont_{"milk",fp} = \sum_{mlk} v_prodQuant_{mlk,r}\ cont_{mlk,fp}$$

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.

Equation 134
$$\begin{aligned} \mathrm{v_prodQuant}_{\mathrm{mlk,r}} &= \mathrm{am}_{\mathrm{mlk,r}} \\ &+ \sum_{\mathrm{j}} \mathrm{bm}_{\mathrm{mlk,j,r}} \left(\mathrm{v_prodPrice}_{\mathrm{j}} - \mathrm{cont}_{\mathrm{j,"fat"}} \mathrm{pFatProt}_{\mathrm{fat,r}} - \mathrm{cont}_{\mathrm{j,"prot"}} \mathrm{pFatProt}_{\mathrm{prot,r}} \right) \Big/ \mathrm{p}_{\mathrm{index,r}} \end{aligned}$$

And lastly, prices of raw milk are derived from its fat and protein content valued with fat and protein prices and a processing margin.

4.5.6 Trade flows and the Armington assumption

The *Armington*[66] *assumption* drives the composition of demand from domestic sales and the different import origins depending on price relations and thus determines *bilateral trade flows*. 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:

Equation 135 $$U_{i,r} = \alpha_{i,r}\left[\sum_{r1}\delta_{i,r,r1}M_{i,r,r1}^{\ -\rho_{r,i}}\right]^{-1/\rho_{r,i}}$$

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 so-called share parameters, α is called the shift-parameter, and ρ is a parameter derived from the substitution elasticity. Deriving the first order conditions for utility maximisation under budget constraints leads after some re-arrangements to the following relation between imported quantities *M*:

Equation 136 $$\frac{M_{i,r,r1}}{M_{i,r,r2}} = \left[\frac{\delta_{i,r,r1}}{\delta_{i,r,r2}}\frac{P_{i,r,r2}}{P_{i,r,r1}}\right]^{1/(1+\rho_{r,i})}$$

where the term $1/(1+\rho)$ denotes the substitution elasticity. 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 price *r2*. 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. 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.

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. 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.

[66] Armington, Paul S. (1969), A Theory of Demand for Products Distinguished by Place of Production, *IMF Staff Papers* 16, pp. 159-178

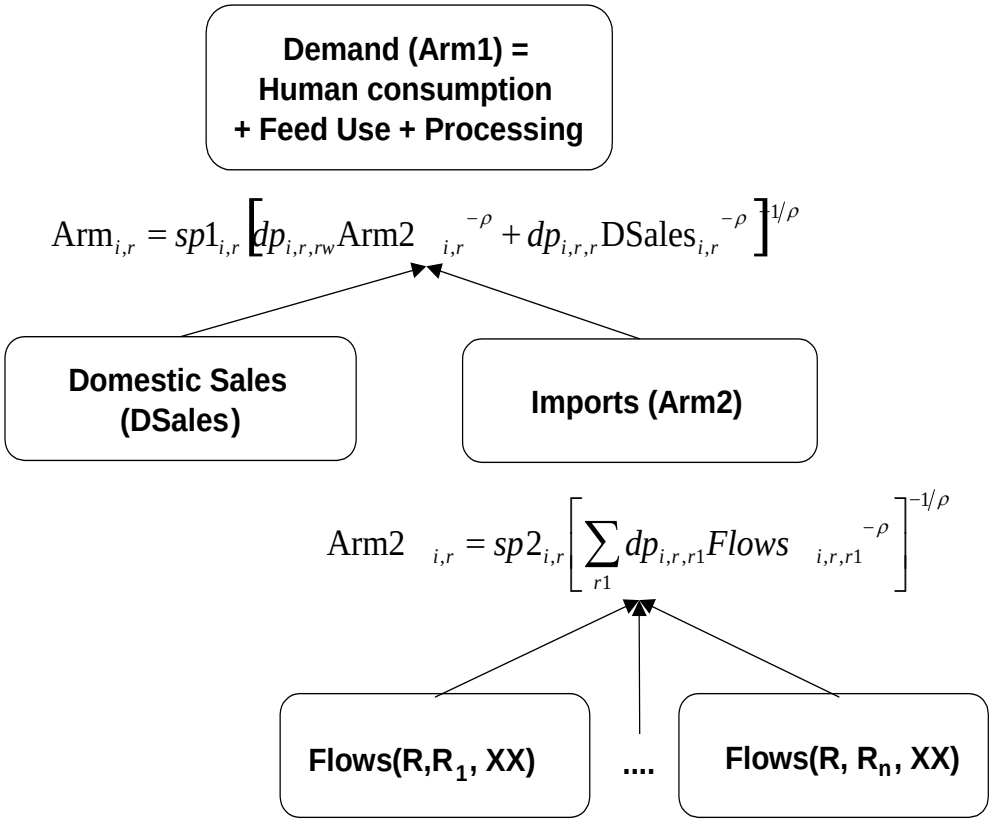
Substitution elasticities for the Armington CES utility aggregators[67]

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

Source: Own calculations

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.

Figure 13. Two-stage Armington System



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 zero level. For most simulation runs, that shortcoming should not be serious. It is planned to overcome that problem by introducing constant terms in the CES utility function, and consequently the share equations.

[67] A sensitivity analysis on those elasticities is given in section 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.

4.5.7 Market clearing conditions

All quantities in the model are measured in 1000 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:

Equation 137

$$\begin{aligned} v_domSales_{i,r} = {} & v_prodQuant_{i,r} \\ & - v_\exp Quant_{i,r} \\ & + v_intervSotckChange_{i,r} \end{aligned} \qquad \text{(SupBalM_)}$$

Further on, **imports and exports** are defined from bilateral trade flows as:

Equation 138

$$v_impQuant_{i,r} = \sum_{r1 \neq r} v_tradeFlows_{i,r,r1} \quad \text{(impQuant_)}$$

Equation 139

$$v_\exp Quant_{i,r} = \sum_{r1 \neq r} v_tradeFlows_{i,r1,r} \quad \text{(expQuant_)}$$

Finally, the **Armington first stage aggregate** *arm1*, shown in the diagram above, is equal to the domestic consumption elements feed, human consumption and processing:

Equation 140

$$\begin{aligned} v_arm1Quant_{i,r} = {} & v_feedQaunt_{i,r} \\ & + v_consQuant_{i,r} \\ & + v_procQuant_{i,r} \\ & + v_biofprocQuant_{i,r} \end{aligned} \qquad \text{(armBal1_)}$$

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.

4.5.8 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*):

Equation 141

$$impp_{i,r,r1} = \begin{bmatrix} v_marketPrice_{i,r1} \; p_exchgRateChangeFactor_{r,r1} \\ -expsub_{i,r1} + v_transpCost_{i,r,r1} \end{bmatrix} \left(1 + v_tarAdVal_{i,r,r1}/100\right) + v_tarSpec_{i,r,r1} + DiffLevies_{i,r,r1}$$

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 EU15, EU10 and Bulgaria and Romania. The reader is reminded that for the EU27, 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.

Equation 142 $$v_prodPrice_{i,r} = \left[v_marketPrice_{i,r} + PSEd_{i,r} + PSEi_{i,r} \right] Pmrg_{i,r}$$

```
*
* --- 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.

The **average prices of imports** derived from the Armington second stage aggregate are labelled v_*arm2Price* and defined as total import value divided by the Armington second stage utility aggregate v_*arm2Quant*:

Equation 143 $$v_arm2Price_{i,r} = \frac{\sum_{r1 \neq r} v_tradeFlows_{i,r,r1}\ impp_{i,r,r1}}{v_arm2Quant_{i,r}} \quad \text{(arm2Price_)}$$

Similarly, the **average prices for goods consumed domestically** *arm1p* are a weighted average of the domestic market price *pmrk* weighted with domestic sales *dsales* and the Armington second stage utilitiy aggregate *arm2* weighted with the average import price *arm2p*:

Equation 144 $$v_arm1Price_{i,r} = \frac{v_arm2Qaunt_{i,r}\ v_arm2Price_{i,r} + v_domSales_{i,r}\ v_pmrk_{i,r}}{v_arm1Quant_{i,r}}$$

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:

Equation 145 $$v_consPrice_{i,r} = v_arm1Price_{i,r} - CSEd_{i,r} - CSEi_{i,r} + cmrg_{i,.r}$$

Unit value exports net of border protection are defined as average market prices in the export destination minus tariffs as:

Equation 146 $$v_unitValueExports_{i,r} = \frac{\sum_{r1 \neq r} \left(pmrk_{i,r1} - tariffs_{r1,r,i} \right) \Big/ \left(1 - tariffa_{i,r,r1} \right) v_tradeFlows_{r1,r,i}}{v_nonDoubleZeroEports_{r,i}}$$

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.

Equation 147 $$expsub_{i,r} = \frac{exps_{i,r}}{exports_{i,r}} \left(pmrk_{r,i} - uvae_{r,i} + c\exp s_{r,i} \right)$$

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:

Equation 148 $$\frac{v_arm2Quant_{i.r}}{v_domSales_{i,r}} = \left(\frac{dp_{i,rw,r}}{dp_{i,r,r}} \frac{pmrk_{i,r}}{arm2\Pr icep_{i,r}} \right)^{\frac{1}{1+\varphi_1}}$$

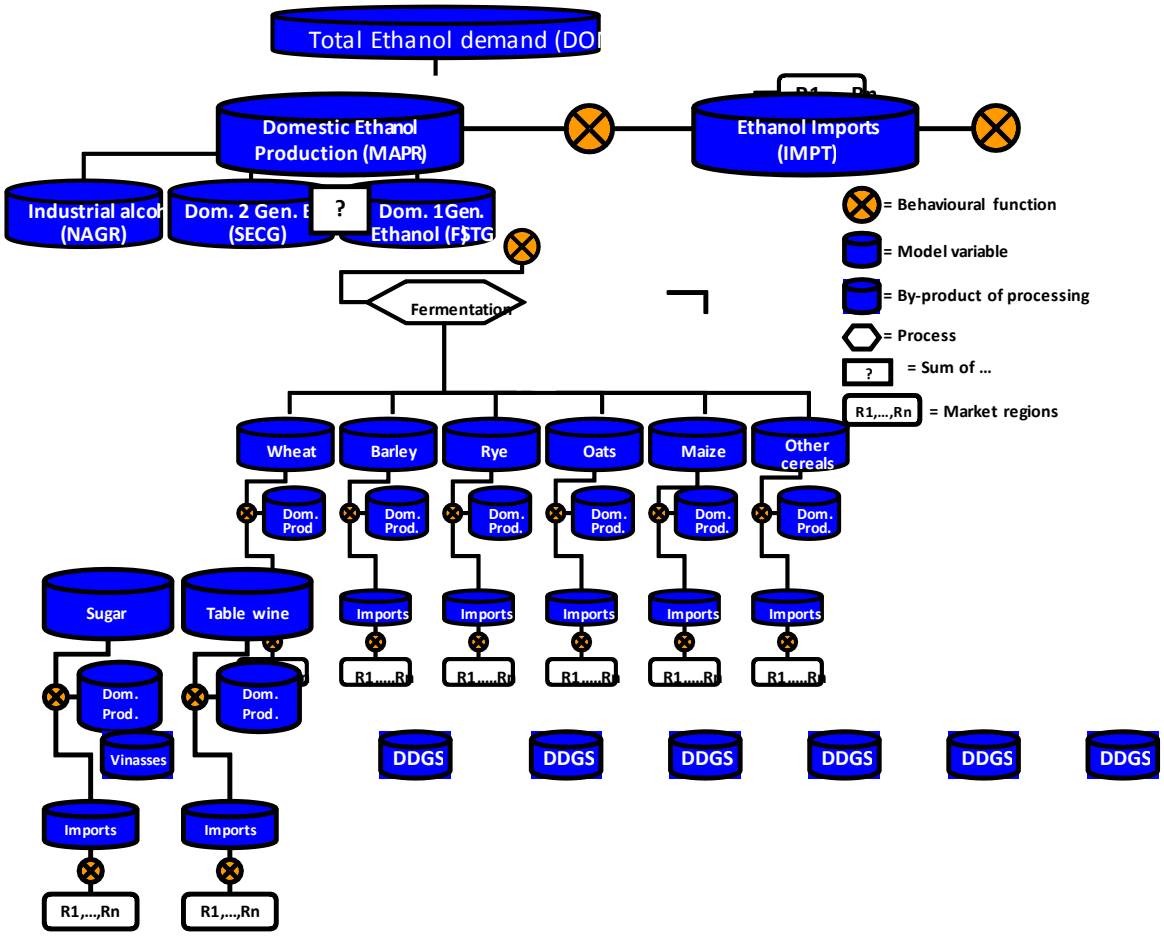
Similarly, relations between import shares are determined by:

Equation 149 $$\frac{v_tradeFlows_{i.r,r1}}{v_tradeFlows_{i,r,r2}} = \left(\frac{dp_{i,r,r1}}{dp_{i,r,r2}} \frac{impp_{i,r,r2}}{impp_{i,r,r1}} \right)^{\frac{1}{1+\varphi_2}}$$

4.5.9 Bio-fuel module (A. Becker and M. Adenäuer)

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 flexibly to biofuel prices and further relevant drivers.

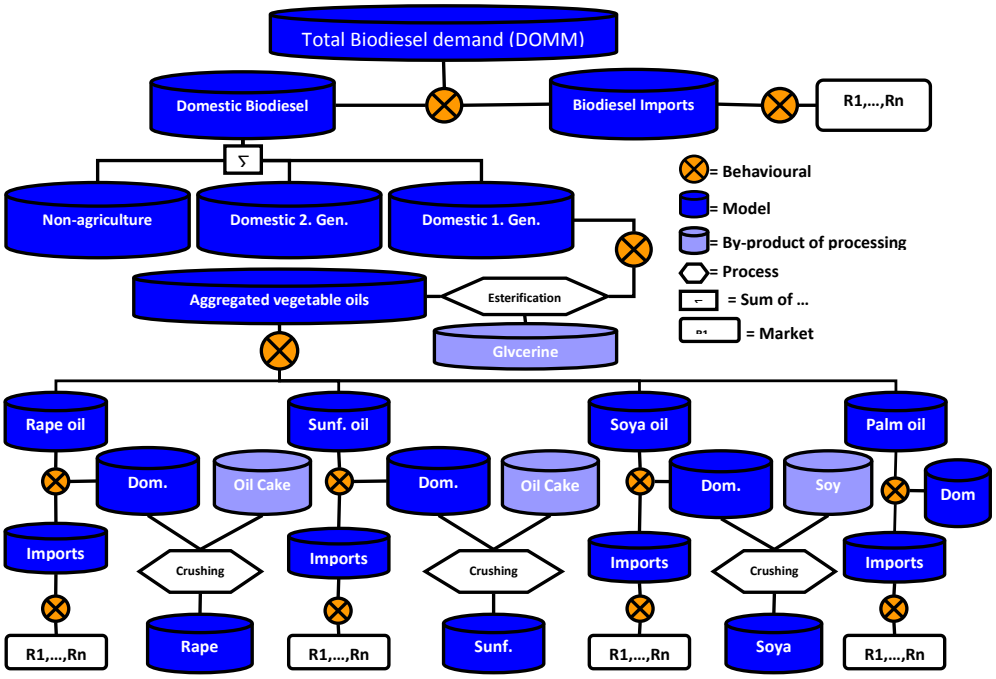
Figure 14. Construction of the ethanol market implemented in CAPRI



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 as shown in Figure 13; 1st generation ethanol ($BIOF_E$) - 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 15. The three production pathways for biodiesel are; 1st generation biodiesel ($BIOF_D$) produced from vegetable oils (rape oil, sunflower oil, soya oil, and palm oil), 2nd generation biodiesel (SECG); and non-agricultural biodiesel (NAGR).

Figure 15. Construction of the biodiesel market implemented in CAPRI

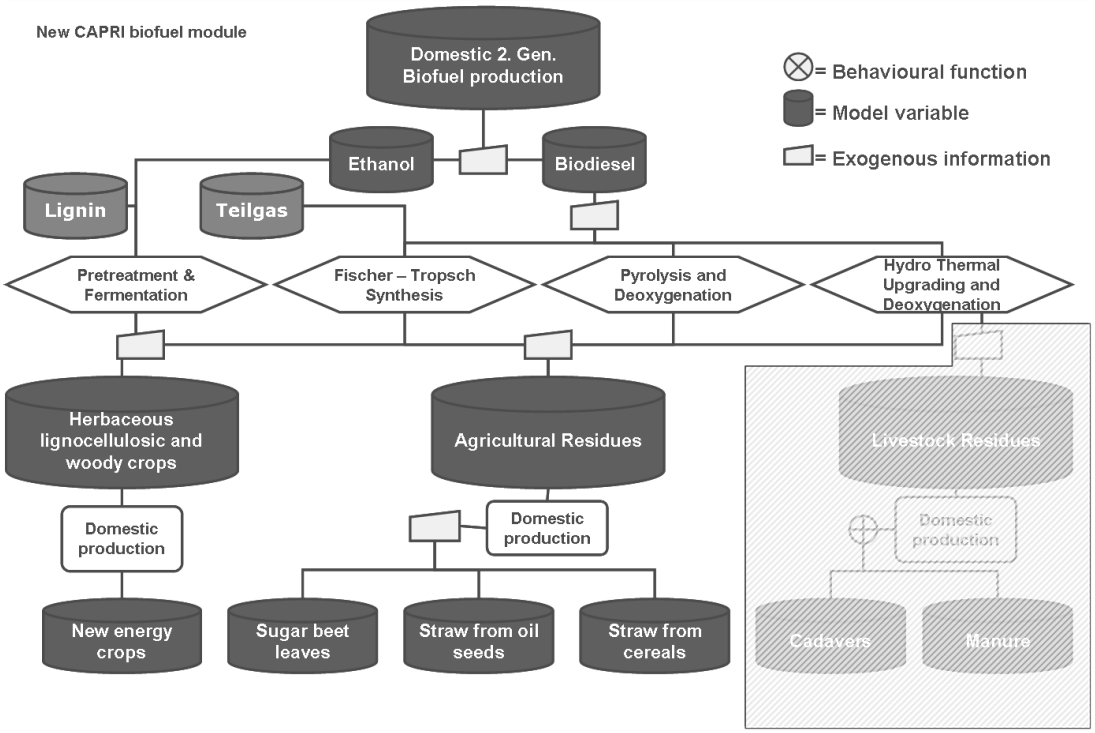


The figure below provides a schematic diagram of the process of 2nd generation biofuel production in CAPRI. Two different product aggregates are introduced in the CAPRI product list to cover feedstock for 2nd generation biofuel processing:

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.

Figure 16. Consideration of 2nd generation biofuel production and related feedstock



4.5.9.1 Behavioural biofuel module

4.5.9.2 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 feedstock price minus by product revenues per ton of input.

Equation 150 $$\mu_{r,xf} = \mathrm{p}_{r.xf} - \sum_{xbp} \mathrm{p}_{r.xbp} \alpha_{r,xf,xbp}$$

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 are then aggregated to an average cost for the respective biofuel. The two biofuel types ethanol and biodiesel in xb:

Equation 151 $$\mu_{r,xb} = \frac{\sum_{xf} \mu_{r,xf}\, fd_{xf}}{\sum_{xf} fd_{xf}}$$

The feedstock demand of each xf is denoted by fd.

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 a cost minimisation problem with a CES cost function for a given biofuel output:

Equation 152

$$fd_{r,xf} = \phi_{r,xf} x^{1st}_{r,xb} \left[\frac{\mu_{r,xb}}{\mu_{r,xf}} \right]^{\rho_{r,xb}} \quad \forall xf \neq num$$

$$s.t. \quad \rho_{r,xb} = \frac{1}{\sigma_{r,xb}} - 1$$

The subsition elasticity of the CES function is given by σ and their share parameters by Φ. First generation biofuel procuction (x^{1st}) is then derived by the product sum of feedstock demand and their biofuel processing coefficinents:

Equation 153

$$x^{1st}_{r,xb} = \sum_{xf} fd_{r,xf} \alpha_{r,xf,xb}$$

Dropping the numerraire from input demand Equation 152 and recalculating it residually assuming Leontief technology in Equation 153 is inconsistent as it mixes a CES technology with a Leontief one. Another option might be to adjust the processing coefficients endogeneously without dropping a nummeraire. Based on the CES technology the average per unit input costs from Equation 151 can be rewritten as:

Equation 154

$$\mu_{r,xb} = \mu^{c}_{r,xb} \left[\sum_{xf} \phi_{r,xf} x^{1st}_{r,xb} \left[\frac{\mu_{r,xf}}{\mu^{c}_{r,xb}} \right]^{(1-\rho_{r,xb})} \right]^{\left(\frac{1}{1-\rho_{r,xb}}\right)}$$

The superscript c indicates the average input costs in the calibration point. This formular is on a first sight more complicated than Equation 151, but since it does not need any quantity variable it eases the solution finding behaviour of the solver, which is why it is used as model equation.

Simultaneously x^{1st} is a function of the relation of the respective biofuel price and the corresponding average feedstock per unit costs.. 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.

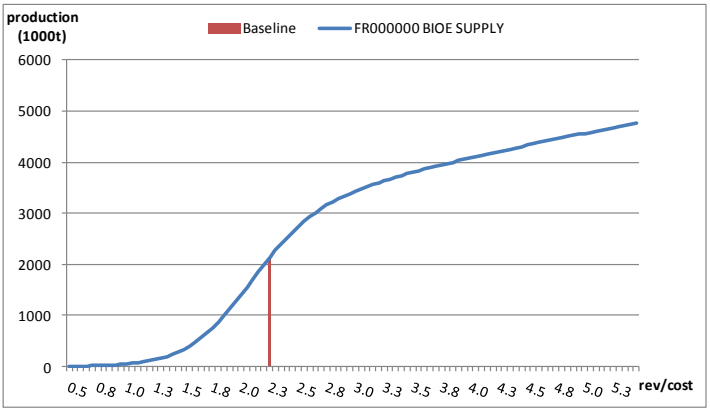
Equation 155

$$x^{1st}_{r,xb} = \left[\partial_{r,xb} \frac{p_{r,xb}}{\mu_{r,xb}} + \exp\left(\beta^{1}_{r,xb} + \beta^{2}_{r,xb} \ln\left(\frac{p_{r,xb}}{\mu_{r,xb}} \right) \right) \frac{1}{1+\exp\left(\left(\frac{p_{r,xb}}{\mu_{r,xb}} - \delta^{1}_{r,xb} \right) \delta^{2}_{r,xb} \right)} \right]$$

This function consists of three parts on the RHS: the first part is linear (a small positive value for δ), 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 β and δ are behavioural parameters in these functions. All biofuel supply equations are generally of the style presented below with an example of bioethanol in France.

Figure 17. Biofuel supply function in France



The supply of by products is directly linked to the first generation biofuel output:

Equation 156 $$x_{r,xbp} = fd_{r,xf}\alpha_{r,xf,xbp}$$

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):

Equation 157 $$x^{tot}_{r,xb} = x^{1st}_{r,xb} + x^{secg}_{r,xb} + x^{nagr}_{r,xb} + x^{exo}_{r,xb}$$

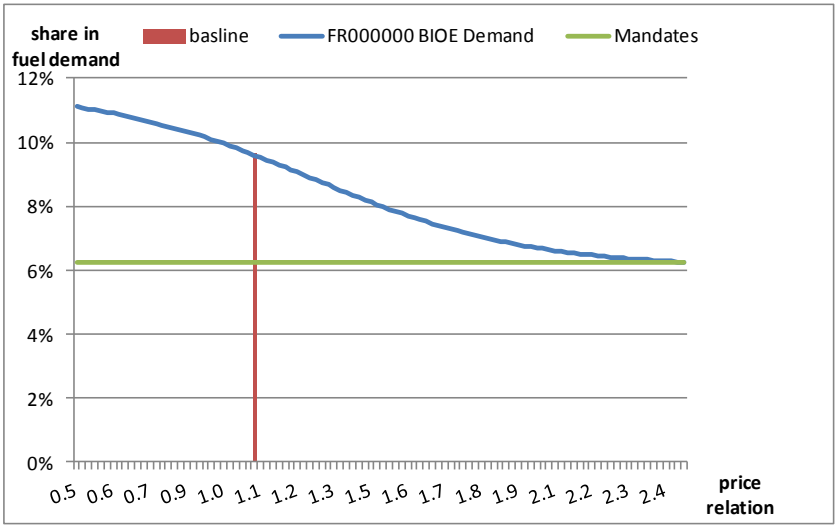
4.5.9.3 Biofuel demand

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:

Equation 158 $$bsh_{r,xb} = bsh^{q}_{r,xb} + \frac{bsh^{\max}_{r,xb}}{1+\exp\left(\left(\frac{p_{r,xb}}{p_{r,f}} - \chi^{1}_{r,xb}\right)\chi^{2}_{r,xb}\right)}$$

Again the coefficients χ are used to specify the exact slope of these functions. The first term (bsh^{q}) defines the part of the biofuel demand which is enforced by any kind of obligation quota or mandate, 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:

Figure 18. Biofuel demand share function in France



Total biofuel demand ($d_{r,xb}$) is then derived by multiplying this share to the exogenous total fuel demand ($d_{r,f}$):

Equation 159 $$d_{r,xb} = bsh_{r,xb} d_{r,f}$$

4.5.9.4 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:

Equation 160

$$\log\left(y_{i,j,s,t}\right) = \delta_{i,j} + \alpha_{i,j} \cdot \log\left(p_{i,j,s,t}\right) + \beta_{i,j} \cdot \log\left(gdp_{j,s,t}\right) + \gamma_{i,j} \cdot \log\left(trend_t\right) + \varepsilon_{i,j,s,t}$$

where:

i	= Fuel Type	trend	= Trend variable
j	= Region	ε	= Error term of the regression
s	= Scenario	δ	= Intercept
t	= Year	α	= Price elasticity of demand
y	= Fuel demand	β	= GDP elasticity of demand
p	= Fuel price including tax rates	γ	= Trend elasticity of demand
gdp	= Gross Domestic Product		

The results of the regression analysis (differentiated into biodiesel and ethanol for every EU MS) cover estimates for α, β, γ and the intercept (δ). 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.

Table 21 Assumed elasticities for total fuel demand after filling with average values

		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

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 (Section 4.5.6). In contrast to the common CAPRI variable notation within the demand system biofuel demand for fuel use is set on BIOF (also used for feedstock demand) to ensure that biofuels do not enter the human consumption demand system. The non fuel demand for biofuels (e.g. ethanol demand of the chemical industry) is consequently set on INDM or PROC (industrial use). The other demand components in the common CAPRI version feed demand and human consumption are neglected for biofuels.

4.5.9.5 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 χ^1 and χ^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).[68]

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 [69]. For the two δ parameters the following rules were applied: δ^1, representing the turning point of the sigmoid function was defined to be left to the calibration point at 90% of the 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.

4.5.10 Endogenous policy instruments in the market model

4.5.10.1 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:

Equation 161 $$Sigmoid(x) = \exp(\min(x,0) / (1 + \exp(-abs(x))))$$

[68] 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.

[69] An elasticity below 1 turned out to produce more reasonable supply responses as above 1. Again an empirical basis for this is still missing.

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.

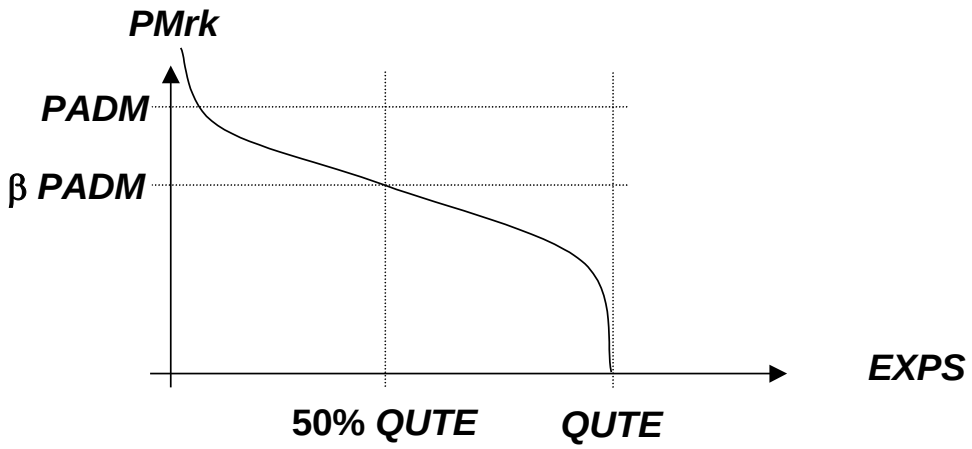
Equation 162 $$expsVa1_{i,r} = QutE_{i,r}\left[sigmoid\left(\frac{\alpha_{i,r}}{PADM_{ir}}\left(v_marketPrices_{i,r} - \beta^{E}_{i,r} PADM_{i}\right)\right)\right]$$

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:

Equation 163 $$expSub_{i,r} = 1000\, expsVal_{i.r} / v_nonDoubleZeroExports_{r,i}$$

The relation is shown in the figure below.

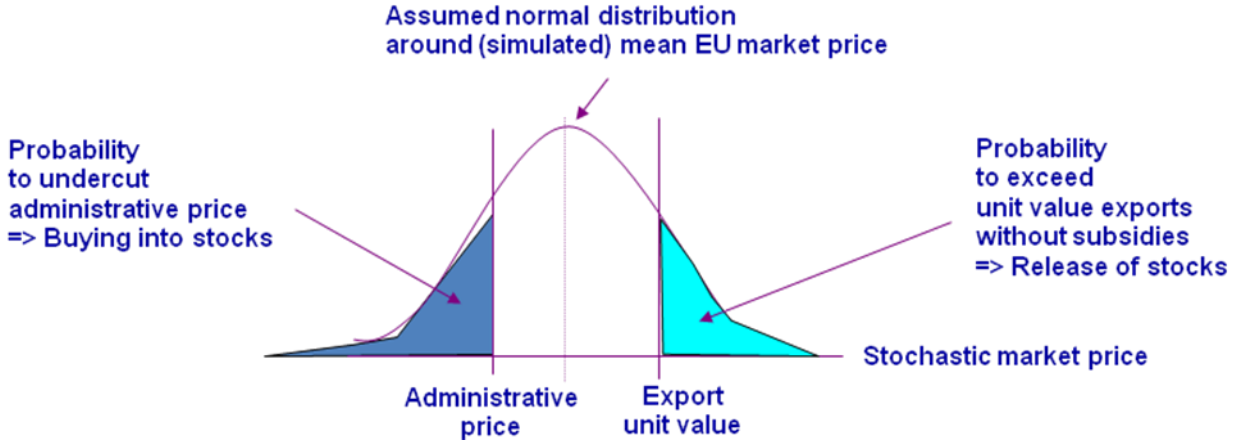
Figure 19. Modelling of subsidised export costs by a logistic function



4.5.10.2 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.

Figure 20. Endogenous administrative stocks in CAPRI



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*:

Equation 164

$$v_buyingToIntervStocks_{i,r} = IntM_{i,r} errf\left(\left(padm_{i,r} - v_marketPrice_{i,r} + \gamma^{p}_{i,r}\right)/stddev_{i,r}\right)$$

A decrease of the administrative price or an increase of the market price will hence decrease purchases to intervention stocks.

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^d, ,multiplied with the current intervention stock size being equal to starting size *intk* plus intervention purchases *intp*:

Equation 165

$$intd_{i,r} = \left(intk_{i,r} + intp_{i,r}\right) errf\left(\left(uvae_{i,r} - pmrk_{i,r} + \gamma^{d}_{i,r}\right)/stddev_{i,r}\right)$$

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*:

Equation 166

$$ints_{i,r} = intp_{i,r} - intd_{i,r}$$

4.5.11 Endogenous tariffs under Tariff Rate Quotas, flexible levies and the minimum import price regime for fruits and vegetables of the EU

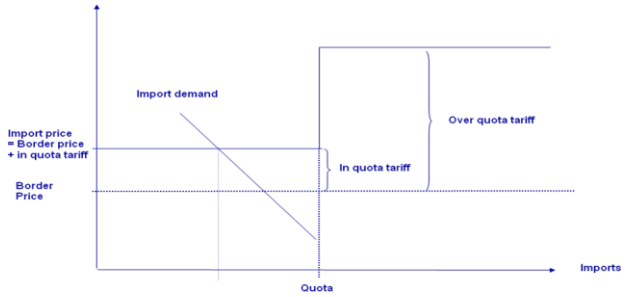
4.5.11.1 Tariff Rate Quotas

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:

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.

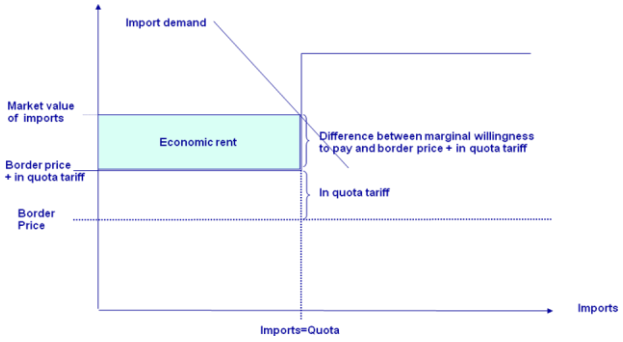
Figure 21. Quota underfill regime



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

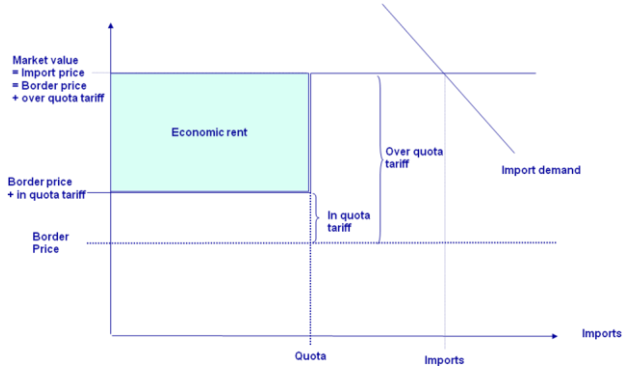
the border price plus the in-quota tariff establishes a quota rent. Depending on property rights on the quota and the allocation mechanism, the quota rent is shared in different portions by the producers, importing agencies, the domestic marketing chain or the administration. Typically, the quota rent can neither be observed nor is their knowledge about distribution of the rent.

Figure 22. Quota binding regime



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.

Figure 23. Quota overfill regime



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.

```
*
* --- 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);
```

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.

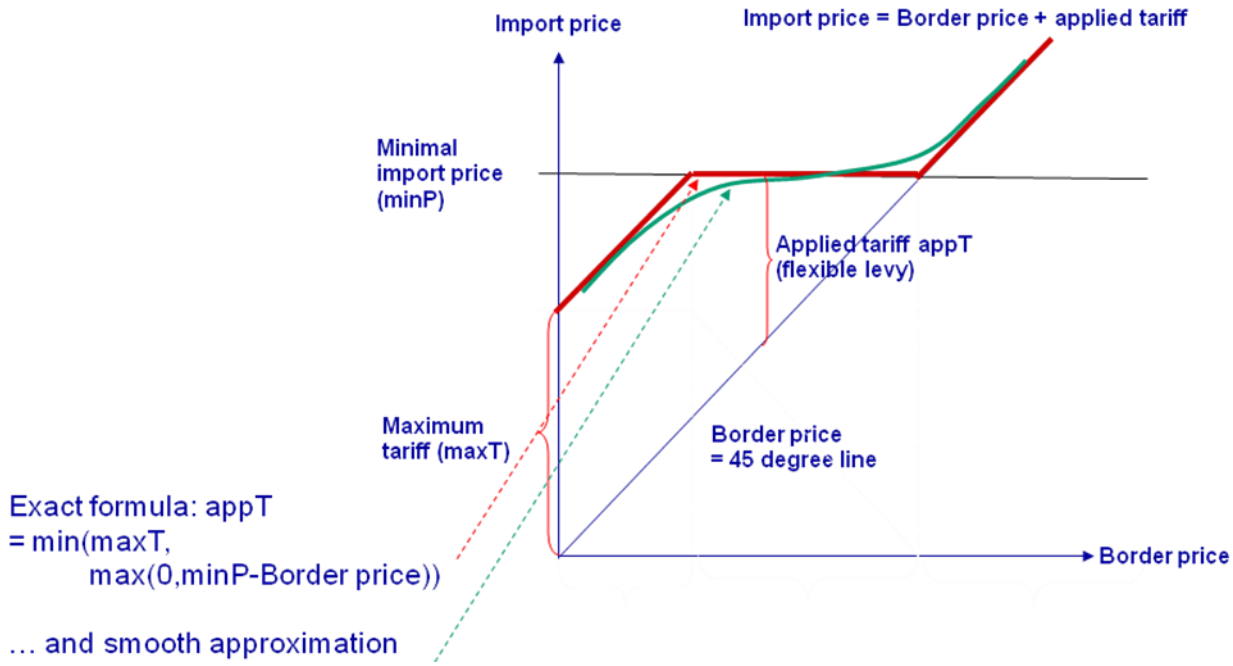
Besides the problem of defining the regime ex-post, the relation between the import quantity and the tariff is not differentiable but kinked. Therefore, again a sigmoid function (Figure 19) is applied in the CAPRI market part:

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.

4.5.11.2 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 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."

Figure 24. Variable levies



In CAPRI, the system is implemented as follows:

Equation 167 $$v_flexLevy_{i,r} = \min\left[v_tarSpec_{i,r}, \max\left(0, minBordP_{i,r} - cif_{i,r}\right)\right]$$

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 an 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:

```
*
* --- 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");
*
```

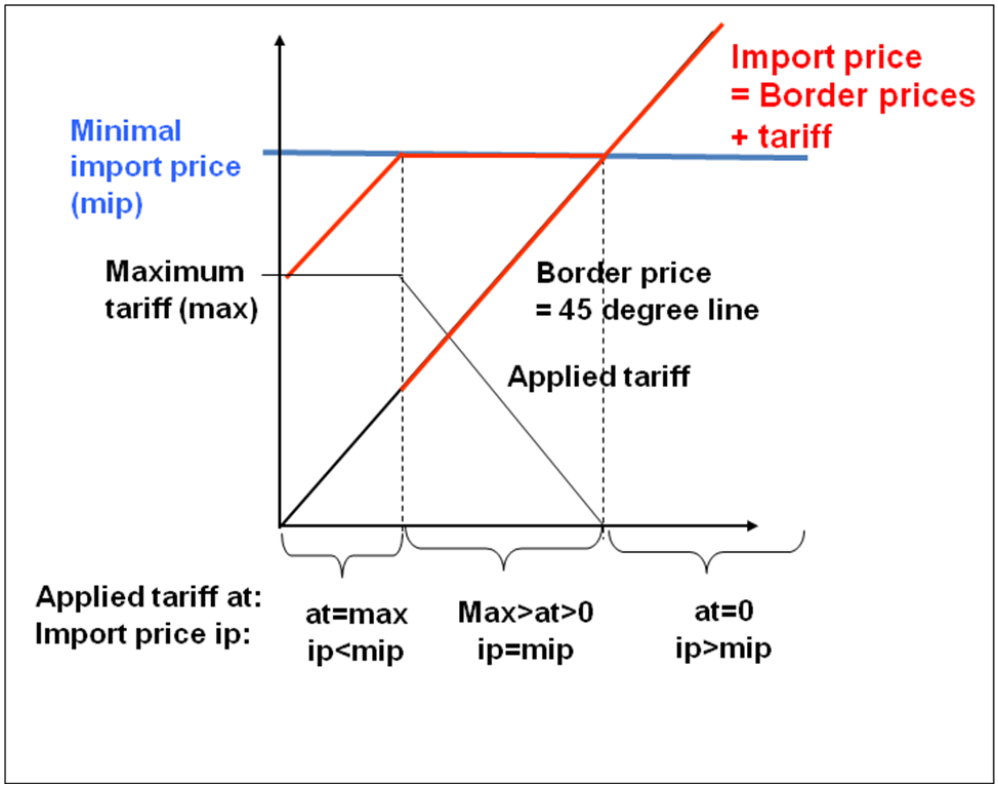
The second equation defines the actual tariff applied:

```
*
*
* --- 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");
```

4.5.11.3 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 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.

Figure 25. EU entry price system for fruits and vegetables



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.

```
*
* --- 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");
```

That factor is the fed into a modified sigmoid function which as a result approximates the relations in the graphic shown above:

```
*
* --- 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));
```

4.5.11.4 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.

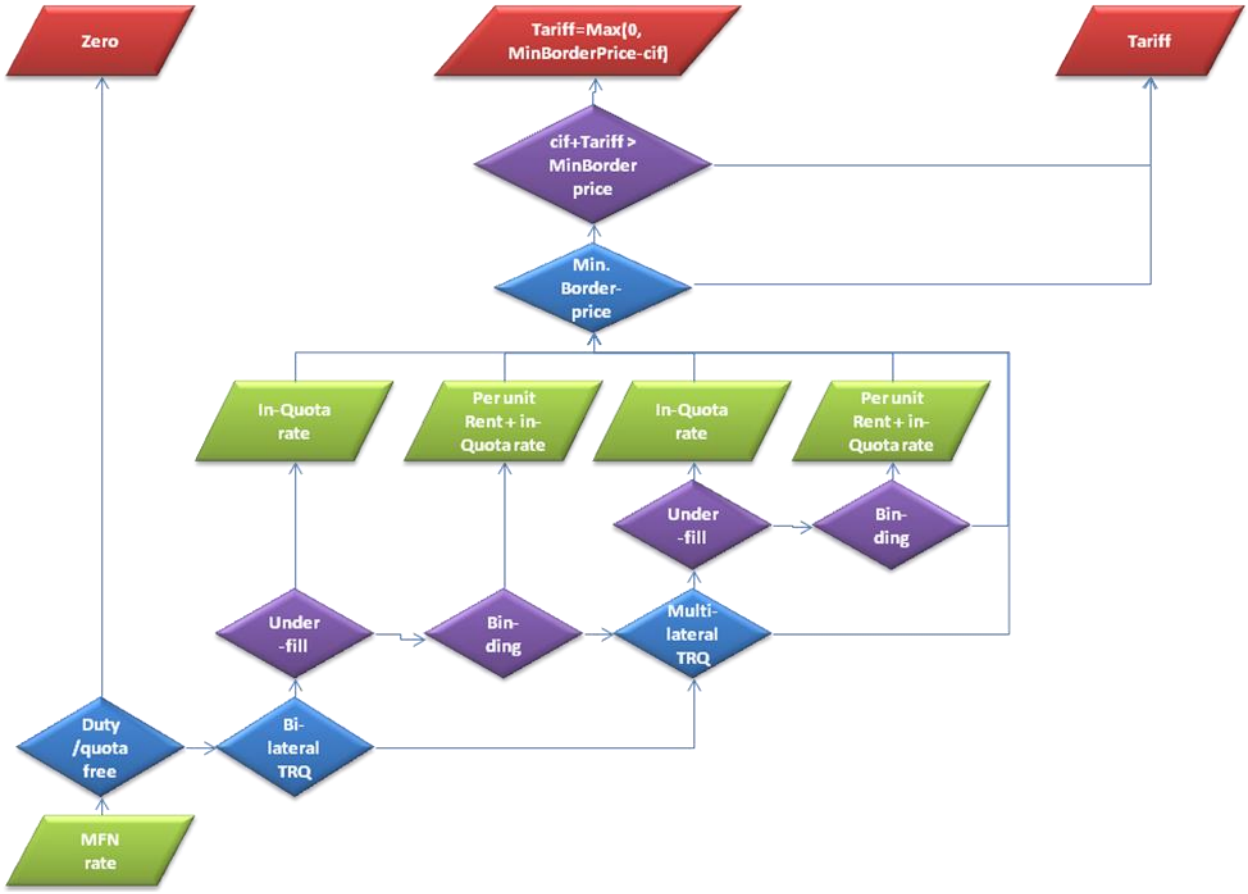
The simplest decision is in the left lower corner: a check if the importer benefits from duty and quota free accesss. Examples are intro-EU trade or import from LDCs into the EU under the “Everything But Arms”-Agreement. If that is the case, the applied tariff is zero.

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.

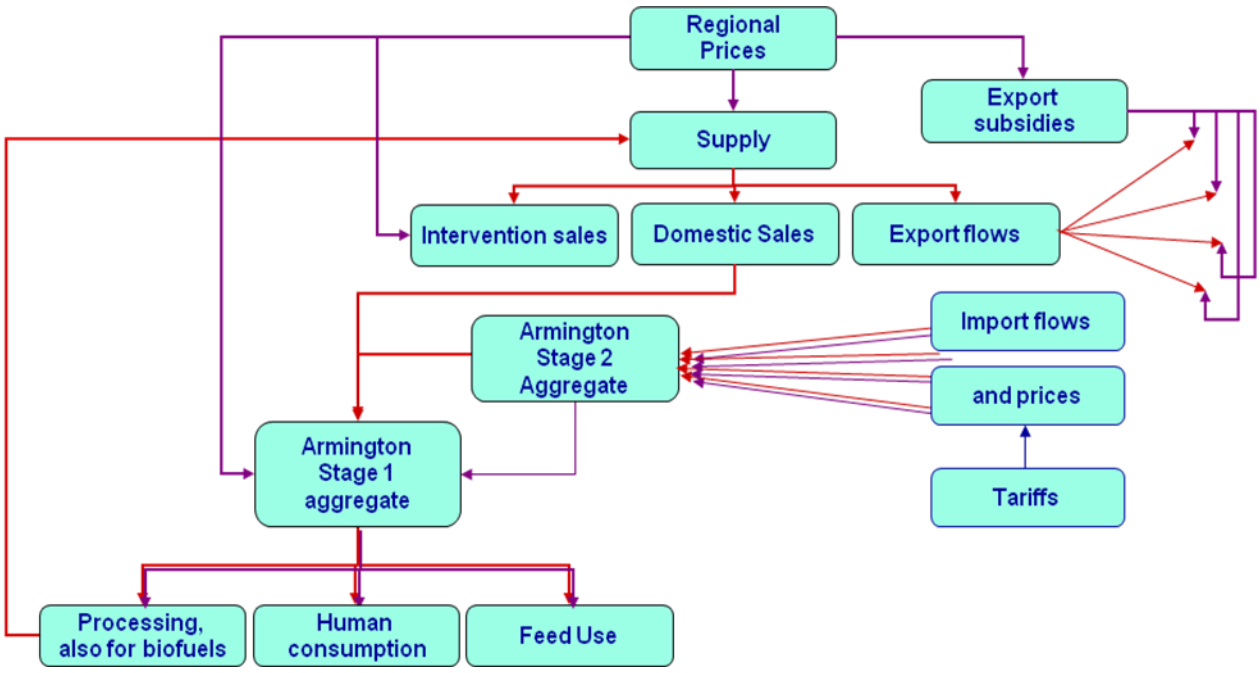
Figure 26. Tariff computation in the model



4.5.12 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.

Figure 27. Graphical presentation for one region of a spatial market system



Source: CAPRI modelling system

4.5.13 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 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.

4.6 Solving the market model (W. Britz)

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 is 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.

Parameter calibration and sources for the behavioural equations

4.6.1 Calibration of the system of supply functions

The supply equation was already introduced in Equation 116. The matrix *bs* is equal to the Hessian matrix of second derivatives of the normalised profit function with respect to

normalised prices and must hence be symmetric by definition. As *bs* is equal to the first derivative of the supply function against normalised prices, the supply elasticities at the calibration point are defined as:

$$\text{Equation 168} \qquad \varepsilon_{i,j} = bs_{i,j} \frac{\overline{ppri}_{j,r} / \overline{p}_{index,r}}{\overline{supply}_i}$$

Homogeneity of supply functions of degree zero is given due to the normalisation with a price index: if all prices and the price index are raised by the same percentage, the supply quantity does not change.

Remains the question of curvature, which is guaranteed if *bs* is positive definite, ensured by a Cholesky decomposition during the calibration process. The curvature ensures that marginal profits are increased if one or several of the prices are increased, and is one of properties of a profit function derived from micro-theory. The calibration searches for minimal squared deviations between the consistent elasticities and given ones.

The uncalibrated elasticities for the non-EU regions are taken from the World Food Model of the FAO, status 1995. Missing own-supply elasticities are set to 0.5. It is assumed that the elasticity to all remaining products including the inputs is -0.25, if not given.

There are some further restrictions introduced:

- Absolute elasticities are not allowed to be larger than 10.
- Reactions in between cereals and between cereals and meats must be substitutive.

4.6.2 Calibration of the final demand systems

According to the concept of the Supply Utilization Accounts, all processing demand by the food industry is counted as human consumption. Equally, imports of food products are re-converted in primary product equivalents. Human consumption of a primary product in the market model does hence include all processed food products derived from it as pasta, muesli, bread etc. rooting in bread.

As discussed above, the demand system discussed above is homogenous of degree zero in prices and income, and symmetric if *bd* is symmetric. The somewhat more cumbersome proof that utility is decreasing in prices and increasing in income as long as the matrix *bd* has only positive off-diagonal elements is left out in here. The down-side of the restriction on the sign of the elements of *Pbd* is that fact that the function then allows for Hicksian substitutes, only. The function is then clearly not longer flexible which may be seen as a disadvantage in econometric applications. Given the product list of the CAPRI market model, the limitation was even judged as a safeguard against curious price effects[70] as complementarities for the compensated demands are not easy to argue for.

The symmetry and non-negativity conditions are imposed during the calibration of the parameters to the price and income elasticities borrowed from the WFM. The calibration necessitates derivatives of Marshallian demands versus prices and income from the expenditure system above which are determined as follows:

---

[70] As an alternative, a normalized quadratic expenditure system was tested. According to the family of indirect utility functions discussed above, the function G is then replaced by a quadratic form in normalized prices. However, a Cholesky decomposition is then necessary to ensure correct curvature during the calibration process, which renders the solution more cumbersome. An advantage of the NQ system is the fact that it allows formally for complementarity in the Hicksian effects. In practice, that would mean that the Marshallian elasticities created by the calibration of the NQ system have to be carefully checked for such complementarities to ensure a plausible behaviour of the demand system in simulations.

Equation 169

$$\frac{\partial PerCap_{r,i}}{\partial y} = \frac{G_i}{G_r}$$

$$\frac{\partial PerCap_{r,i}}{\partial cpri_{r,j}} = \left(\frac{G_{ij}}{G} - \frac{G_i G_j}{G_r^{\,2}} \right) \left(y - F \right) \wedge i \neq j$$

where:

$$G_{ij} = \frac{\partial G_i}{\partial cpri_j} = \tfrac{1}{2} bd_{ij} \Big/ \sqrt{cpri_i cpri_j} \quad \wedge i \neq j$$

The terms for the own price effects are somewhat more complicated, and therefore determined indirectly via the homogeneity condition for elasticities during calibration. The objective function minimizes squared differences between given and consistent elasticities, simultaneously for the base year and the last year of the projection period. The parameters *di* are chosen so that the functions calibrate to quantities and prices in the calibration point.

In the milk study a new calibration approach was developed which is now the standard: Elasticities are initialised as a combination of group elasticities from Seale, Regmi, Bernstein 2003, the FAPRI elasticity database ((http://www.fapri.iastate.edu/tools/elasticity.aspx) and an assumed set of Allen elasticities of substitution to obtain a full matrix of initial cross price elasticities (Similar to Witzke, Zintl 2005, Section 2.2).

A certain disadvantage of the Seale, Regmi, Bernstein 2004 elasticities is that this source reports Frisch price elasticities, holding constant marginal utility rather than utility. Frisch price elasticities will lie between Marshallian and Hicksian elasticities for food items (Seale, Regmi, Bernstein 2004, p 29) but without knowledge of all parameters and data of the underlying study they cannot be converted exactly. To resolve this problem it has been assumed that the Frisch price elasticities given are *exactly* half way between the compensated (Hicks-Slutsky) and uncompensated (Cournot-Marshall) price elasticity which is sufficient to convert them either into Hicksian or Marshallian elasticities, using the expenditure elasticities given in the same source

4.6.3 Overview on the calibration mechanism

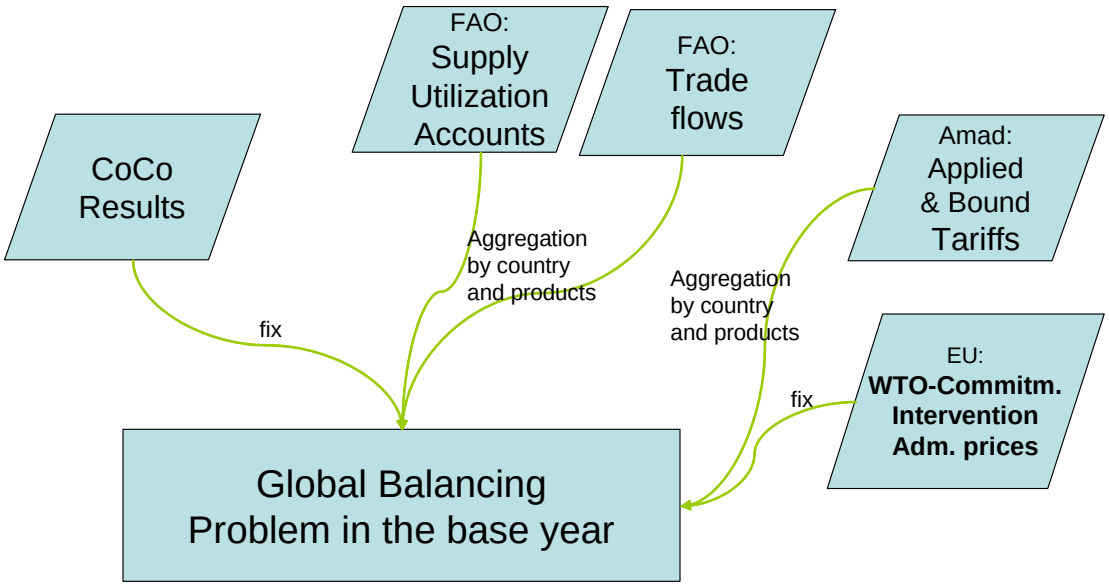
The calibration mechanism requires a consistent data set, so that after introducing the parameters in the behavioural equations, the model is self calibrating. That requires a pre-step where raw statistical input data are corrected so that market balances are closed. Equally, prices for the different agents and regions must be set so that at given transport costs, tariffs and further policy instrument as tariff rate quotas all price transmission equations are in equilibrium.

The process is shown in the figure below. In order to allow for independent calibration of the supply and the market models, it is important that the balancing process for the market model does not change the data the supply part is calibrated, too. As the market part works on the level of single countries or group of countries, the relevant data set for the countries covered in the supply part are hence the market balances and prices generated by CoCo. Those are kept fix during the balancing process. The input data from FAO – the market balances expressed in primary product equivalents termed “supply utilization accounts” and the trade flows in similar definitions, are first aggregated to the product and regional definitions of the CAPRI global market model. The same holds for the tariff information, which stems mainly from the Amad data base, additional information e.g. relating to Tariff Rate Quotas stems from the Official Journal of the EU and further data sources.

The balancing process attacks the different products independently from each other, with the exemption of oilseeds, oils and cakes which need to be solved simultaneously, as well as

dairy products. The balancing problem is solved as a Highest Posteriori Density estimator which takes the raw statistical data as expected means. It comprises the equations from the market model with the exemptions of behavioural equations as constraints, plus additionally bounds which prevent rather sharp changes in the raw statistical data. Those bounds are iteratively relaxed in case of infeasibilities.

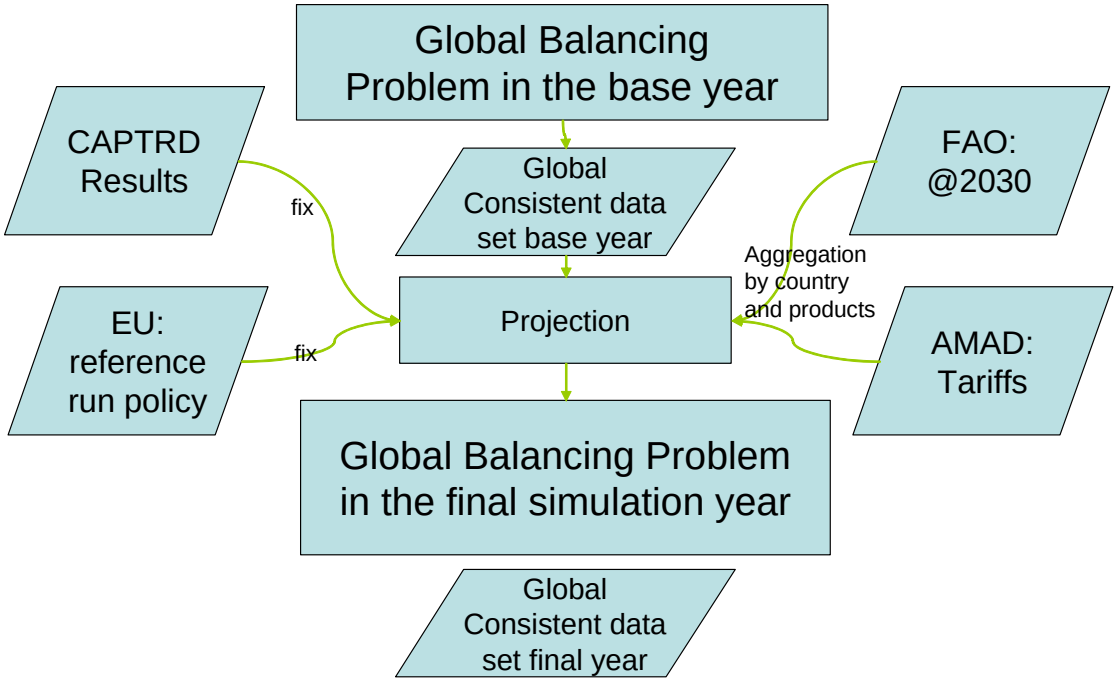
Figure 28. Balancing the market model for the base year



That process need to he repeated for the final simulation year. Before, all the market balance elements and prices need to be projected to the final simulation year, and the set of policy parameters – administrative prices, tariff, tariff rate quotas etc. – be updated as to reflect the policy for the base line. As in the balancing in the base year, it is important to find a solution which does not deviate for the countries covered by the supply part from the solution generated by CAPTRD, as otherwise, the total system will not calibrate if the different modules are linked together.

The projection for the rest of the world draws to a large extent in the medium term outlook published once a year by OECD and FAO building on the AGLINK-COSIMO model, and, to a lesser extent on FAPRI medium term projections. Indeed, there is ready to use GAMS code to convert the AGLINK-COSIMO and FAPRI baselines into a GAMS data table in CAPRI definitions.

Figure 29. Balancing the market model in the final simulation year.



4.7 Linking the different modules – the price mechanism (W. Britz)

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 point elasticities of the aggregated response from the supply module differ from the ones in the market modules which necessitate an iterative update. In order to speed up convergence, the supply side uses a weighted average of prices of the last iterations.

The first version of CAPRI fixed supply of EU Member States in the market module during iterations. It turned out however, that convergence is achieved faster if supply is price responsive even with differing point elasticities. One of the options discussed is to generate a set of price elasticities from the regional programming models and to calibrate the parameters of the market module to it. However, given the large amount of commodities and regional or even farm type models, these sensitivity analysis would take quite some time.

The interaction between the regional programming models and the young animal module was already explained above. Basically, it is again an iterative update of parameters in a more aggregate model; however, the young animal module comprises models at Member State level which are structurally identical to the regional models. The update thus requires both the definition of a weighted average of the I/O coefficients as well as the application of ideas borrowed from Positive Mathematical Programming to achieve a point calibration. As for marketable outputs, prices for young animals used in any iteration are a weighted average of previous iterations.

4.7.1 Price linkage in the sugar sector

Whereas the linkage of prices from the market model to the supply model is usually a proportional one, possibly with some lag structure across the solution steps, the price linkage in the sugar sector involves additional parameters (only the post 2006 reform specification is shown here):

Equation 170
$$p^{A} = share^{A} \cdot \left(p^{SUGA} \cdot prcy_{SUGB}^{SUGA} + add^{MOLA} \cdot prcy_{SUGB}^{MOLA} \right)$$

Equation 171
$$p^{C} = share^{C} \cdot \left(p^{BIOE} \cdot prcy_{SUGA}^{BIOE} \cdot prcy_{SUGB}^{SUGA} + add^{MOLA} \cdot prcy_{SUGB}^{MOLA} \right)$$

The share parameters share^{A} and share^{C} reflect the pricing behaviour on the part of the sugar industry. This is heavily influenced by the sugar CMO for the quota prices (where minimum prices for beets are specified), but more dependent on market conditions for C-beets. These parameters are calibrated in the baseline process to be consistent with projections on beet prices and they are held constant in subsequent policy simulations. The additional molasse payment is only small part of the total beet price.

The price differentiation between quota beets (mainly used for the food sector) and C-beets for the ethanol industry implies that the ethanol industry receives raw material from the sugar sector at a discount price relative to the food industry. This is reflected in the biofuel demand equation with a correction factor for the feedstock price:

Equation 172
$$\begin{aligned} p_{BIOE}^{SUGA} &= p^{SUGA} \cdot corfac^{SUGA} \\ &= p^{SUGA} \cdot \frac{p^{C}}{p^{A}} \end{aligned}$$

Where p^{C} and p^{A} are the beet prices for A-beets and C-beets, respectively. It should be remembered that beets are not part of the market module and that the feedstock quantities of beets used for ethanol are only expressed in terms of sugar such that the sugar price is the relevant driver for the feedstock demand. The above correction factor (set = 1 for all other ethanol feed stocks like cereals) ensures that the ethanol industry faces the correct price, benefiting from the iterations between supply and market models. It also ensures that the discount disappears if the quota system is abandoned, thereby reducing the profitability of using beets for ethanol.

4.8 Sensitivity of the CAPRI model to the Armington substitution elasticities (W. Britz)

Note: the sensitivity analysis was conducted in the context of the CAPRI-Dynspat project and does not reflect in the current regional and commodity breakdown of the model.

A conventional sensitivity analysis consists to run the model using initial Armington elasticities to obtain the baseline, then to rerun it under various elasticity values, all other things held constant, and finally to compare the reference and simulation results. In our sensitivity study, the implementation of this type of analysis shows very small numerical variations on every variable level at less than 0.002 percent. This is the reason why we chose to associate exogenous shocks to the sensitivity analysis.

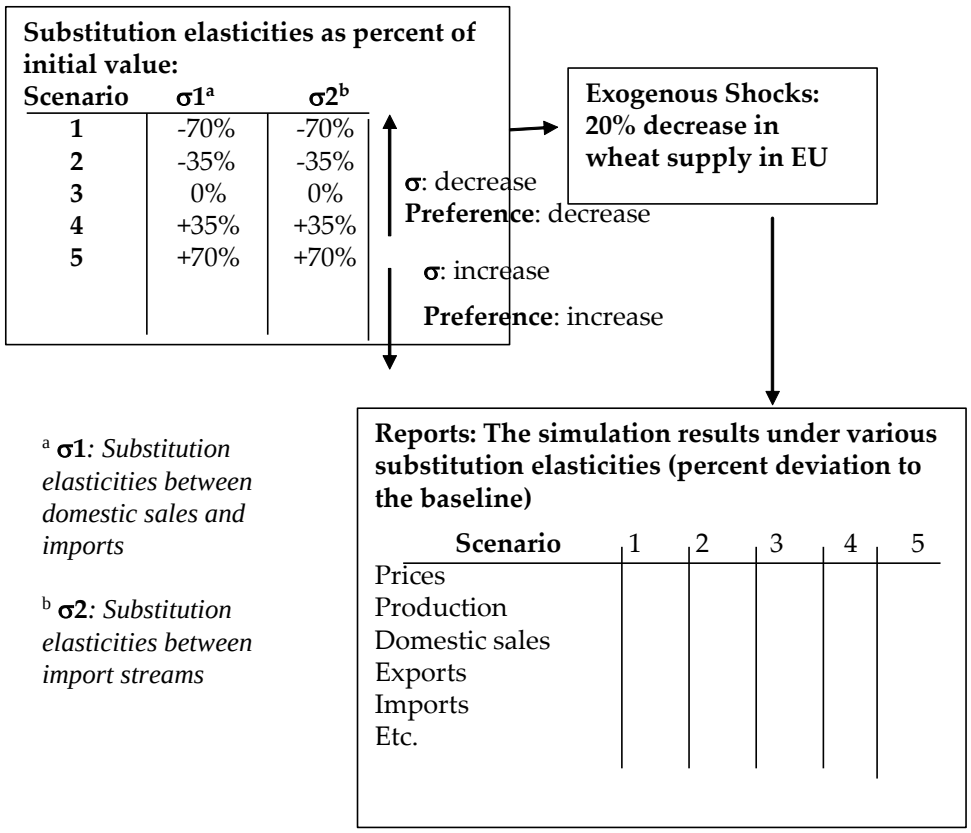
To perform the sensitivity analysis, we first introduce different sets of Armington substitution elasticities in the model. Then, we introduce an exogenous shock by changing, for example, the policy parameters or the shift factors in the supply equations. Finally, we compare the reactions of endogenous variables (price, production, domestic sales, imports, exports) for different sets of elasticities as show in Table 22.

Three exogenous shocks, associated to the sensitivity analysis, are thus implemented: (i) a 20% decrease in supply, (ii) a 10% decrease in subsidized exports, (iii) an increase in tariff

rate quotas. For each shock, the simulation related to the initial Armington elasticities, i.e., scenario 3, is used as the baseline. Its results are compared to those of the sensitivity runs.

The sets of substitution elasticities are obtained by shifting the initial value of these elasticities to more or less 70 percent. The use of the same percentage change - between the baseline and the other sensitivity runs - allows to evaluate the degree of symmetry in the sensitivity. Lower values for elasticities imply a decrease of preference and thus a greater difficulty in substituting between demand origins, whereas higher values for elasticities imply an increase of preference and, thus, a greater ease in substituting between demand origins.

Table 22 Illustration of Sensitivity Analysis on the CAPRI Market Module



Source: Own calculations

To keep the discussion readable, we only present the results associated to the large variation of the elasticities (i.e. scenarios 1 and 5: ± 70%). 'High' and 'low' values are specified to represent 70 percent more or less than the initial values used in the baseline. We strict ourselves also to results for the European Union (EU) and to some key commodities which present a variation higher than 0.1 per thousand. However, we point out important findings for other markets where necessary.

As one would expect, the results of sensitivity depend strongly on the exogenous shock associate to the sensitivity analysis. When performing a 20% decrease in supply (Table 23), changes in production levels are insensitive to the Armington elasticities, except for 'other meat' and 'sugar' productions which show a change exceeding 2%. The same observation applies to changes in producer and consumer prices. All the price changes show little

reactions with less than or around 2% in either direction, except for change in the producer prices of 'other meat' and 'sugar' which increase to 3% and 10% respectively. Like changes in production and prices, changes in domestic sales are practically invariant with respect to changes in the Armington elasticities, except for change in the 'rice' domestic sales which shows a reaction exceeding 11%.

Table 23 Deviation of the simulation results to the baseline under high and low substitution elasticities with a 20% decrease in supply

	Elasticity of substitution[a]	WHEAT	BARLY	SUGA	RICE	MEAO
Producer price	Low	1,5%	0,3%	10,5%	1,6%	3,2%
	High	-0,6%	-0,1%	-4,1%	-0,9%	-1,7%
Consumer price	Low	0,2%	0,0%	1,9%	0,2%	1,4%
	High	-0,1%	0,0%	-0,8%	-0,1%	-0,8%
Production	Low	1,1%	0,1%	2,7%	0,8%	2,3%
	High	-0,4%	0,0%	-1,1%	-0,4%	-1,2%
Domestic sales	Low	0,4%	-0,2%	0,0%	11,4%	2,1%
	High	-1,1%	0,0%	0,0%	-5,5%	-1,2%
Exports	Low	7,0%	2,8%	12,7%	-11,8%	0,3%
	High	4,8%	-0,4%	-5,1%	4,3%	-0,1%
Imports	Low	-24,5%	-7,9%	0,6%	-10,1%	-12,2%
	High	25,1%	0,2%	-0,3%	4,8%	6,7%

Source: CAPRI results

[a] *Low elasticity of substitution: -70% of the initial value*
High elasticity of substitution: +70% of the initial value

The same sensitivity results, pertaining to changes in prices, production and domestic sales are obtained with the two other exogenous shocks which consist in a 10% decrease in subsidized exports and an increase in tariff rate quotas. As shown in Table 24 and Table 25, change in all these variables do not exceed 2% except for changes in the domestic sales of 'skim milk powder' and 'rice' which vary by 5 to 7% under a 10% decrease in subsidized exports, and changes in producer and consumer prices of 'cheese' under an increase in tariff rate quotas.

Table 24 Deviation of the simulation results to the baseline under high and low substitution elasticities with a 10% decrease in subsidized exports

	Elasticity of substitution	WHEAT	BARLY	MILS	CHES	RICE	BEFM
Producer price	Low	0,9%	0,6%	0,2%	0,3%	0,8%	1,6%
	High	-0,7%	-0,5%	0,1%	-2,4%	-0,3%	-1,4%
Consumer price	Low	0,1%	0,1%	-0,2%	0,2%	0,1%	0,8%
	High	-0,1%	0,0%	0,1%	-1,5%	-0,1%	-0,7%
Production	Low	0,5%	0,4%	0,1%	0,2%	0,3%	0,9%
	High	-0,4%	-0,3%	-0,1%	-0,7%	-0,1%	-0,8%
Domestic sales	Low	-1,2%	-1,0%	-7,1%	-0,2%	-5,4%	-0,6%
	High	1,3%	0,8%	6,7%	-0,5%	6,3%	1,0%
Exports	Low	14,4%	18,5%	20,1%	7,2%	22,4%	14,6%
	High	-14,4%	-16,4%	-18,7%	-4,1%	-18,7%	-16,5%
Imports	Low	21,4%	16,1%	14,0%	13,8%	8,1%	10,2%
	High	-29,4%	-24,4%	-13,1%	-0,9%	-9,3%	-15,4%

Source: CAPRI results

[a] *Low elasticity of substitution: -70% of the initial value*
High elasticity of substitution: +70% of the initial value

As expected, the main changes in variables that are affected by the Armington elasticities are those of trade flows. Independently of the shock and market types, the largest changes concern import and export quantities and, hence, are the more sensitive to elasticities. Export changes are sensitive to changes in the Armington elasticities. Of course, import changes are even more affected. The largest effects on trade changes are observed for most commodities whose trade is large and characterised by high initial Armington elasticities such as in the case of 'cereals'.

As shown in Table 24, the largest effects on trade changes are observed when performing a 10% decrease in subsidized exports. For some markets, such as the wheat market, the effect on import changes can reach 30%. Most of the large effects on export changes are found in markets characterized by little trade such as the 'rice' market. Under this shock, markets with higher elasticities show lower effects on export and import changes and larger effects on domestic sales changes, and conversely for markets with lower elasticities, larger effects on export and import changes and lower effects on domestic sales changes. This means that, under a shock of 10% decrease in subsidized exports, higher values of Armington elasticities imply an increase of preference in domestic sales against imports, which results in a decrease in exports.

Table 25 Deviation of the simulation results to the baseline under high and low substitution elasticities with an increase in tariff rate quotas[a]

	Elasticity of substitution[b]	BARLY	MILS	CHES	BTCR	SUGA
Producer price	Low	0,0%	0,1%	-5,3%	2,8%	0,0%
	High	0,0%	0,0%	0,8%	-1,5%	0,0%
Consumer price	Low	0,0%	-0,6%	-3,2%	1,6%	0,0%
	High	0,0%	0,1%	0,5%	-0,9%	0,0%
Production	Low	0,0%	0,1%	-0,6%	0,5%	0,0%
	High	0,0%	-0,1%	0,1%	-0,2%	0,0%
Domestic sales	Low	-0,1%	0,5%	-1,1%	0,6%	0,0%
	High	0,0%	-0,1%	0,2%	-0,2%	0,0%
Exports	Low	0,1%	-0,7%	6,7%	19,9%	0,0%
	High	0,0%	0,0%	-0,5%	-23,7%	0,0%
Imports	Low	0,8%	-0,4%	-1,2%	-15,7%	0,0%
	High	-0,3%	0,1%	0,3%	5,0%	0,0%

Source: CAPRI results

[a] *The percentage of the increase in the TRQ applied for each commodity depends on the imports and the tariff rate quotas in the base years*
[b] *Low elasticity of substitution: -70% of the initial value*
High elasticity of substitution: +70% of the initial value

As shown in Table 25, when performing an increase on tariff rate quotas (TRQ), effects on the changes in most of the variables are not sensitive to the Armington elasticities. It means that effects of the TRQ on model outcomes under different sets of Armington elasticities are marginal.

With respect to symmetry in the opposite change in Armington elasticities, we observe that the percentages of change in variable levels versus their initial values do not show much symmetry. For most of the variables and commodities, changes are larger in the lower substitution elasticities (-70%) than in the higher substitution elasticities (+70%), as expected since the relative change in parameters is larger in the former than in the latter. Exceptions appear on changes in imports under the assumption of a decrease in subsidized exports, which react conversely, i.e. changes are less in the lower substitution elasticities than in the higher substitution elasticities (Table 24).

In sum, all the effects on the changes in variable levels remain low compared to the changes applied on the Armington elasticities (± 70%). The model outcomes are thus comparatively insensitive to the actual magnitude of the Armington elasticities.

5 Farm Type Programming Model: a FADN-based approach (A.Gocht, W. Britz)

5.1 The CAPRI farm type approach

The main aims linked to the introduction of farm types in the system is to ameliorate the analysis of agricultural policies linked to structural variables as farm size or stocking density, improve the reliability of environmental indicators and allow for income analysis at farm type level. In other words, the introduction of data for single farms from the FADN data base reduces the aggregation bias of the model at regional level.

The farm group models could be classified by a number of indicators like the economic importance (farms with high agricultural income against those with lower ones), environmental impact (classic against ecological farming) and many others. The standard grouping in FADN is based on *specialisation* (e.g. specialised in pig production), which might be supported on the following arguments:

- First of all, the resulting groups are already clearly defined according to official European documents (Commission Decision 2003/369/EC) and results obtained can be easily compared to other studies,
- secondly, the grouping is based on standard gross margins, reducing the stochastic impact of weather or price changes on the grouping for single years, and
- as a third point, it can be argued that environmental impacts are often linked to farm specialisation.

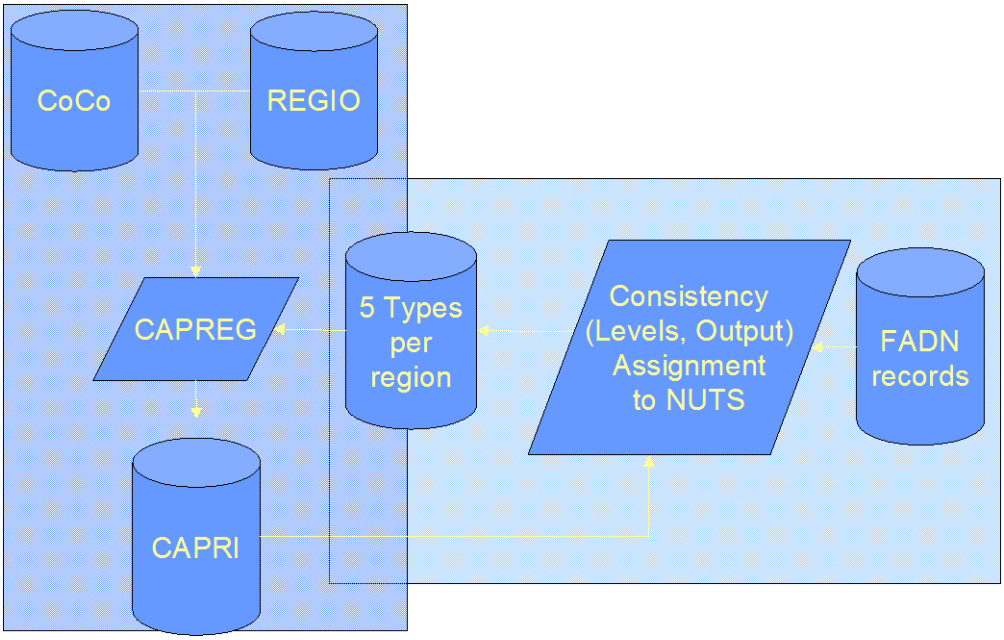
But even with the farm typology according to European standards applied, a number of issues need to be addressed for its application in CAPRI:

(1) *Number of farm groups defined for each region*. Clearly, the amount of detail increases with the number of farm types, in line with computing time and management costs to handle the additional information. Due to such resource and technical restrictions, in CAPRI it was decided to choose not more than five farm types (the most representative) plus a mixed remaining group representing all other farms for the modelling system (and allowing consistent aggregation of regional data).

(2) *Level of typology*: For simplicity and a better comparison to FADN, we use the same three digit typology as defined in FADN. Consequently, 50 different types of specialisation can be found in CAPRI (see table 26).

The following diagram shows the relation between the FADN data base and the elements of the CAPRI data processor.

Figure 30. Integration of farm types in the CAPRI data base



Source: Own calculations

In a first integration step, ex-post data on NUTS 2 level from the CAPRI data on activity levels and output were selected for about 50 production activities. Further on, an extraction program provided the necessary data from the FADN data base.

The second integration step consisted in a non-linear optimisation program which ensured matching activity levels (hectares, herd sizes) and production quantities between CAPRI and FADN. Part of the problem at this stage related to the different regional breakdown of CAPRI and FADN: whereas the CAPRI data base refers to administrative NUTS regions, the FADN data base has its own set of FADN-regions. In order to increase the number of farms available per type and region and, at the same time, preventing problems with confidentiality limits, the algorithm used in CAPRI 'distributed' the aggregation weights for each farm over several FADN-regions. A specific farm in the network may easily represent farms not only in the FADN-region where the farm is situated but in other regions as well (within the boundaries of a NUTS 2 region).

In order to match the CAPRI data base –which is in major elements derived from the REGIO data base at EUROSTAT– it was necessary to change the aggregation weights and activity data of single FADN records. Minimising squared differences ensured that the changes were not bigger then necessary. After that step, the single farm records were aggregated to specialised farms per region (see table 26) and the five most frequent farm types were selected, with the frequency relating to the aggregation weights. This step is necessary only once for a given base year. Afterwards, an additional algorithm ensures that input use aggregated over the farm types matches the input use at NUTS 2 level. These algorithms are integrated in the so-called regionalisation step in CAPRI, which combines the COCO data base (with its time series at national level) with information from REGIO and other sources at regional level.

Table 26 Farm types found in the system

131	Specialist COP (other than rice)
132	Specialist rice
133	COP and rice combined
141	Specialist root crops
142	Cereals and root crops combined
143	Specialist field vegetables
144	Various field crops
201	Specialist market garden vegetables
202	Specialist flowers and ornamentals
203	General market garden cropping
311	Quality wine
312	Wine other than quality
313	Quality & other wine combined
314	Vineyards for various types of production
321	Specialist fruit (other than citrus)
322	Citrus fruits
323	Fruits & citrus fruits combined
330	Olives
340	Various permanent crops combined
411	Milk
412	Milk & cattle rearing
421	Cattle rearing
422	Cattle fattening
431	Dairying with rearing & fattening
432	Rearing & fattening with dairying
441	Sheep
442	Sheep & cattle combined
443	Goats
444	Various grazing livestock
501	Specialist pigs
502	Specialist poultry
503	Various garnitures combined
601	Market gardening & permanent crops
602	Field crops & market gardening
603	Field crops & vineyards
604	Field crops & permanent crops
605	Mixed cropping-mainly field crops
606	Mixed cropping-mainly market gardening or permanent crops
711	Mixed livestock-mainly dairying
712	Mixed livestock-mainly non-dairy grazing
721	Mixed livestock-granivores & dairying
722	Mixed livestock-granivores & non-dairy grazing
723	Mixed livestock-granivores with various livestock
811	Field crops & dairying
812	Dairying & field crops
813	Field crops & non-dairy grazing
814	Non-dairy grazing & field crops
821	Field crops & granivores
822	Permanent crops & grazing livestock
823	Various mixed crops and livestock
999	Rest

Source: FADN (http://europa.eu.int/comm/agriculture/rica/index_en.cfm).

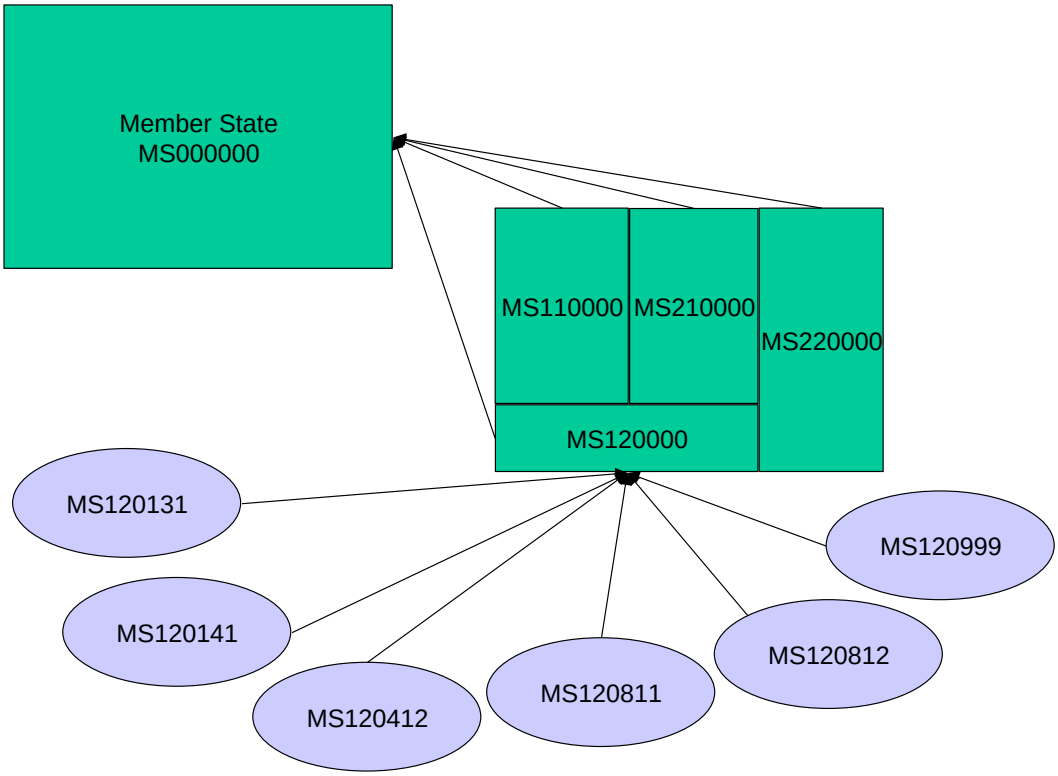
In the CAPRI modelling system, farm types are treated technically as a further breakdown inside NUTS 2 regions (pseudo-regions): the activity levels in each farm type feature own

input and output coefficients and are independently optimised for maximal profits (template approach of the CAPRI supply module). After a model run, the farm type results are aggregated to NUTS 2, Member State and EU level.

It should be noted that the relation between NUTS 2 and Member States is geographical; the disaggregation thus provides localised effects in space. Farm type data however cannot be linked to specific locations in the NUTS 2 regions, even if they break down consistently output, in physical and valued terms, activity levels, and economic and environmental indicators. An improvement in that respect would require a complete link with a Geographical Information System plus intensive economic analysis to create mapping algorithms between spatial specifics (soil types, local climate, slope, altitude ..), production program and farm specialisation. Some work in this direction is being undertaken in CAPRI-Dynaspat and, possibly, in SEAMLESS.

Figure 31 shows the coding scheme. Member States are labelled with two character codes according to EUROSTAT standards (AT for Austria, BL for Belgium and Luxembourg, DK for Denmark, DE for Germany, ...). Regions inside a Member State receive a 3-digit code (first position: NUTS 1 level, second: NUTS 2 level, third: NUS III level) following the EUROSTAT NUTS classification scheme. The farm types are labelled with alphanumerical three-digits code as well, where the '000' refers to the regional level.

Figure 31. Aggregation from farm types to NUTS 2 and Member State



Source: CAPRI Modelling System

Moreover, the system aggregates across regions all farms of the same specialisation, allowing for the analysis of effects for farms of a certain specialisation across Europe. In order to add additional layers of information, the specialised farm types can be also aggregated, as shown in table 27.

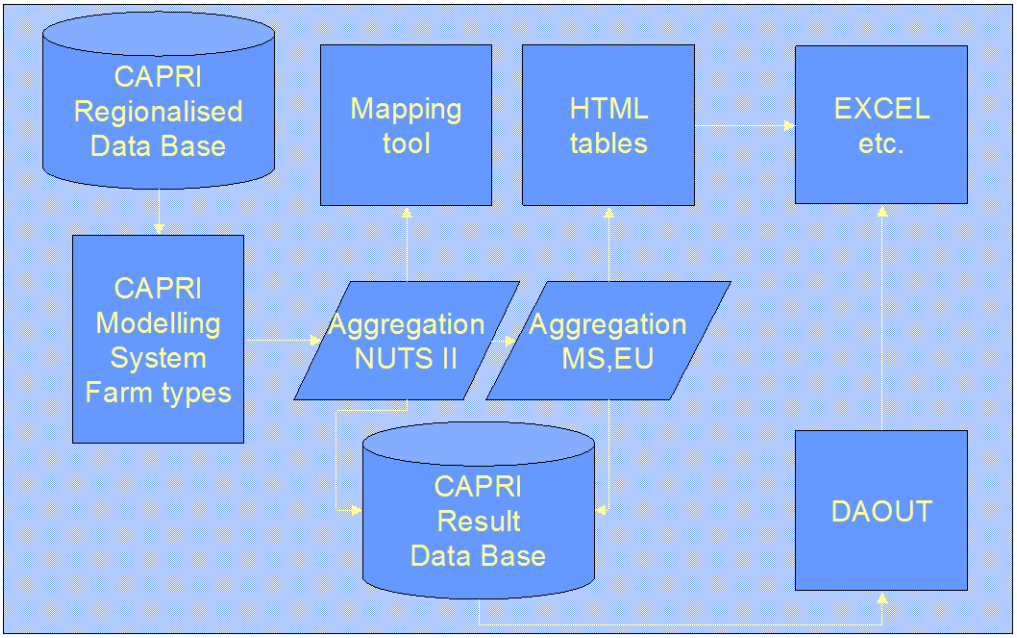
Table 27 Aggregated farm types used for impact assessment

Code	Description	Farm type included
A10	Specialist COP (other than rice) or various field crops	133,144
A13	Specialist Rice or Rice & COP	132,133
A14	Root crops	141,142
A23	Permanent crops & vegetables	143,201,202,203,311,312,313,314,321,322,323,330,340
A41	Dairy	411,412,431
A42	Cattle fattening & rairing	421,422,432
A44	Sheep & goats	441,442,443,444
501	Specialist pigs	501
A52	Specialist poultry	502,503
A60	Field crops diversified	601,602,603,604,605,606
A70	Livestock diversified	711,712,721,722,723
A80	Livestock & crops diversified	811,812,813,814,821,822,823
999	Various	

Source: CAPRI modelling system

Figure 32 shows the relation between the farm types and other elements of the modelling system. Inside the system, farm types are aggregated to NUTS 2 and Member States, to allow a link to the policy and market module. These aggregations allow exploiting the results from farm types in maps and tables relating to geographical units. All results are stored in the data base management system as well and can be easily accessed.

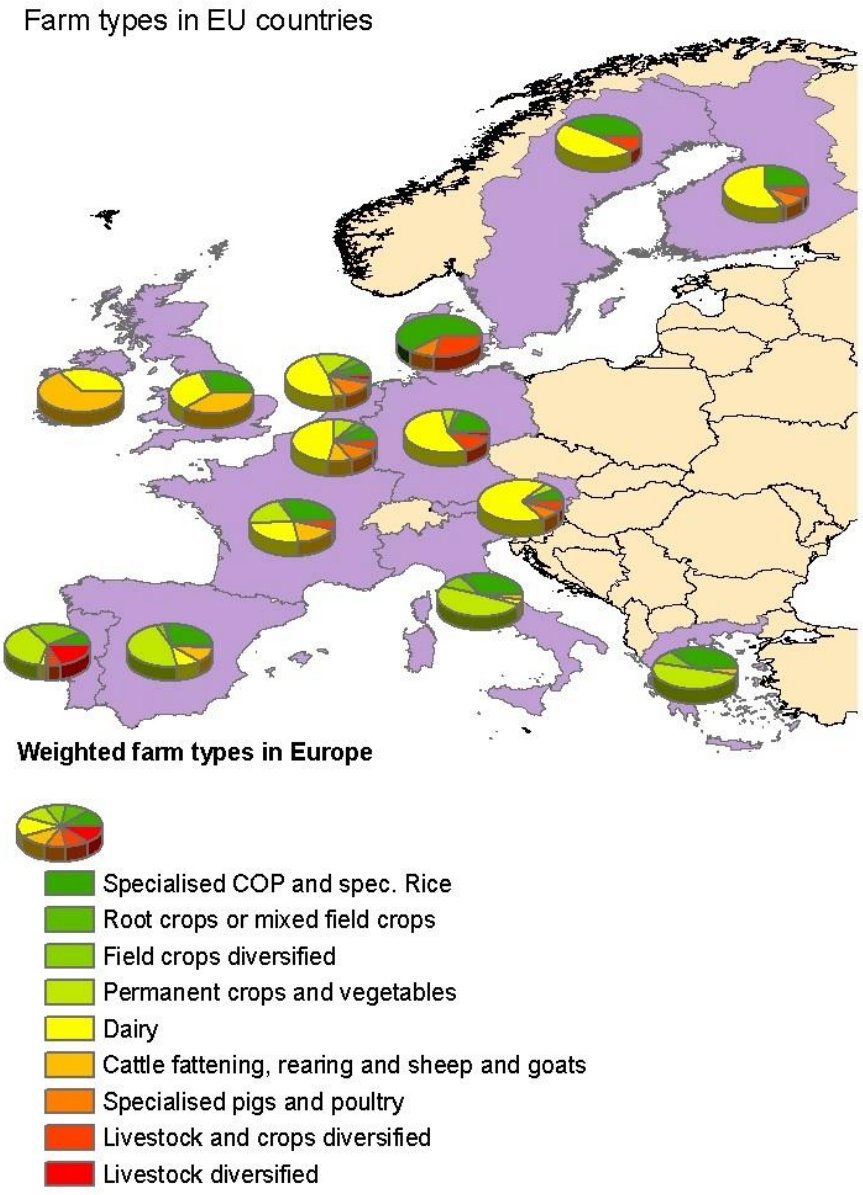
Figure 32. Integration of farm types in the CAPRI modelling system



Source: Own calculations

Figure 33 shows the dominant farm types per country. For reasons of survey research the farm types mentioned in table 27 are further combined. It clearly shows that dairy is a dominant farm type in north of Europe. An exception is Denmark where specialised COP, livestock and crops diversified and specialised pigs and poultry are the dominant farm types. Cattle fattening, rearing, sheep and goats are the dominant farm types in Ireland and United Kingdom. In the south of Europe, Portugal, Spain, Italy and Greece, permanent crops and vegetables is the dominant farm type. Also in France and to a lesser extent in Belgium/Luxembourg and the Netherlands, this farm type is relatively important. The heterogeneity of farm types seems to be quite big in France (different farm types have about the same weight) and small in Ireland.

Figure 33. Farm types in EU15 countries



Source: Own calculations

5.2 Linkage to a SEAMLESS Farm Type Models

(Extract from the DOW:)

Currently, four approaches for that linkage with farm type models can be found in agricultural sector modelling:

1. Simultaneous solution of all farm models in a region with endogenous prices – computationally not feasible in SEAMLESS as world market price feedback is needed.
2. Loose, iterative exchange of results between models at different scale (examples: the model family at FAL, Braunschweig, Germany)
3. Iterative link between farm type or regional models and market scale models – a possible solution within SEAMLESS-IF.
4. Deriving farm level supply responses from the farm level models after preliminary runs of the market models.

CAPRI already comprises a link of the third type, working with fixed aggregation weights. In order to motivate the possible solution for SEAMLESS-IF, the current link in CAPRI is briefly described: Basically, it is an iterative process between the market module and programming models. These individual farm type models inside NUTS 2 regions or regional models are solved independently from each other with fixed prices and premiums. Afterwards, their results regarding supply and feed demand are aggregated to the Member State level, and the supply and feed demand functions of the market model are re-calibrated so that they would generate identical quantities at the given prices for the commodities. Then, the market model is solved, which will return both changes in aggregate supply and feed quantities as well as in prices. The supply models are then solved again at the new prices, and the process of calibrating the market model and solving the supply models with new prices is repeated until expected and realized prices become (almost) identical, and there are no longer any sizeable differences in quantities obtained from the market model and the farm type or regional models. Equally, in between iterations, premiums are re-calculated where they depend on ceilings. The iteration process works better if the aggregate supply and feed demand functions in the market part mimic well the behaviour of the aggregate supply models.

The fourth, and theoretically the most interesting, approach is to derive supply responses from the farm level models. These responses will be not only responses to price and subsidies changes, but also to all other environmental policies that will be simulated at farm level (i.e., cross-compliance policies). The benefits of such an approach are fewer iterations, and theoretically clearer linkage between farm and market level models. There is a risk related to the fact that for obtaining a rigorous linkage between models heterogeneous from the methodological point of view, time may be limited, especially for the first versions of SEAMLESS-IF. Indirect advantages of the latter include that consistency between the farm and market level model outcomes is less likely to be forced by some ad-hoc adjustment parameters, and that it opens up for closer integration with other market level models than CAPRI.

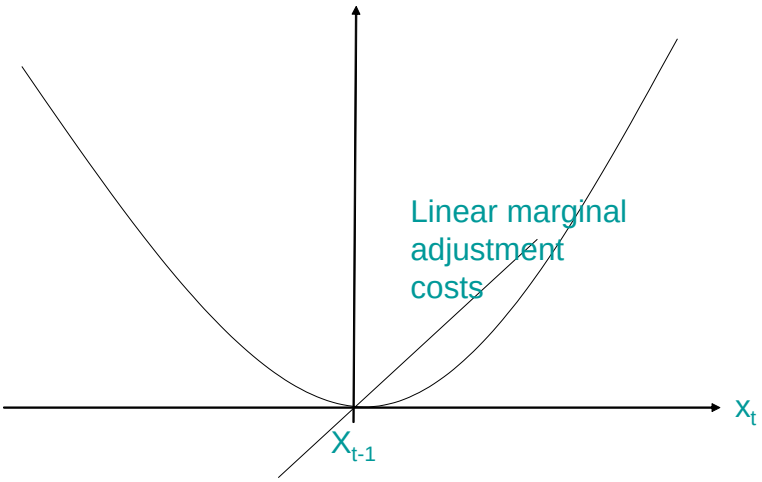
Due to this risk, we propose to pursue both approaches 3 and 4, having the third approach as a backup.

6 A feasibility study for the now abandoned recursive-dynamic version (W. Britz)

During the CAPRI-Dynaspat project (2004-2007), an extension to a recursive dynamic version was tested. It was thought that the iterations between supply and market parts in the comparative static version could be replaced by loop over years, which would allow to see the adjustment path to changes in policies. At the start of the exercise, the CAPRI code was set up so that there were basically different policy sets for the final simulation year available, and one set of data for the base year and another for the final simulation year. It was not possible to host in the same CAPRI installation different base and simulation years. Equally, all policy parameters did not carry a time dimension.

A first important step towards a recursive-dynamic version was hence a rather thorough revision of the code: many parameters, especially those relating to policies, carry now a time dimension; and files which host parameters without a time dimension comprise the year they refer to in the file name. Accordingly, the same CAPRI installation and even the same scenario files can be use for simulation for different base and simulation years. Secondly, in order to allow for looping through the years, the structure of the code had to be rather thoroughly revised. An important outcome of that exercise is the ability to run simulations for the base year, including generation of the so-called "policy shifts" for the baseline generation process.

Figure 34. Linear adjustment costs



After some literature study, it was decided to use quadratic adjustment costs both in the supply and market part of the model to capture adjustment processes as additional quadratic terms in the objective function of the supply are likely to have a limited impact on solution time and fit well to the remaining overall outlay.

Equation 173

$$\frac{C_t(x)}{p_I} = \underbrace{a + \sum_i b_i x_i^t + \tfrac{1}{2}\sum_{ij} q_{ij} x_i^t x_j^t}_{\text{current approach}} + \underbrace{\tfrac{1}{2}\sum_i ac_i \left(x_i^t - x_i^{t-1}\right)^2}_{\text{adjustment cost}}$$

Taken the derivative of the above the cost function which carries as last term quadratic adjustment costs gives the following linear cost function whose last term let costs increase

linearly in the difference between activity levels at the current point in time t and the last time in point t-1.

Equation 174 $$\frac{\partial(C_t / P_I)}{\partial x_i} = b_i + \sum_j q_{ij} x_j^t + ac_i\left(x_i^t - x_i^{t-1}\right)$$

On the market side, adding quadratic adjustment costs to the normalized quadratic profit function leads to partial adjustment approach which can be found in other global modelling system as FAO's World Food Model for the @2030 model of FAO.

Equation 175

$$\frac{\Pi}{p_I} = \sum_i (1. - PartAdj_i)\, a_i \tilde{p}_i + \tfrac{1}{2} \sum_i \sum_j (1. - PartAdj_i)\, B_{ij} \tilde{p}_i \tilde{p}_j + \sum_i PartAdj_i\, X_i^{t-1} \tilde{p}_i$$

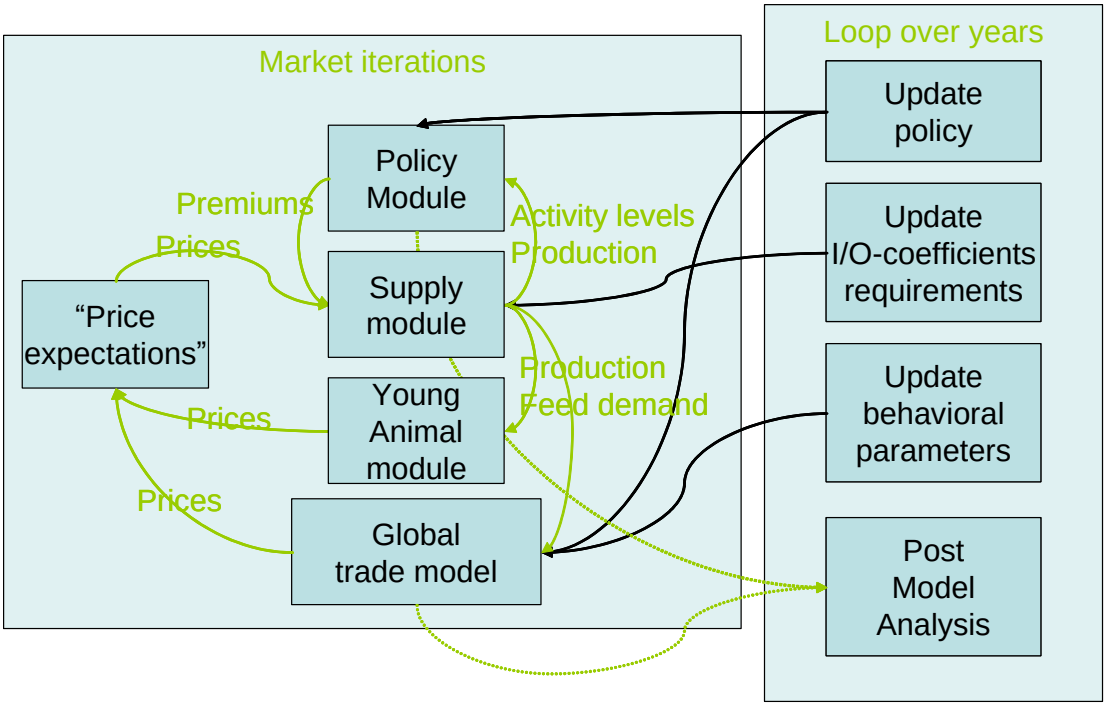
Standard profit function

Adjustment part: A certain part of profits depends on last year's production

$$X_i = \frac{\partial \tilde{\Pi}}{\partial \tilde{p}_i} = (1. - PartAdj_i)\left[a_i + \sum_j \tilde{p}_j B_{ij}\right] + PartAdj_i\, X_i^{t-1}$$

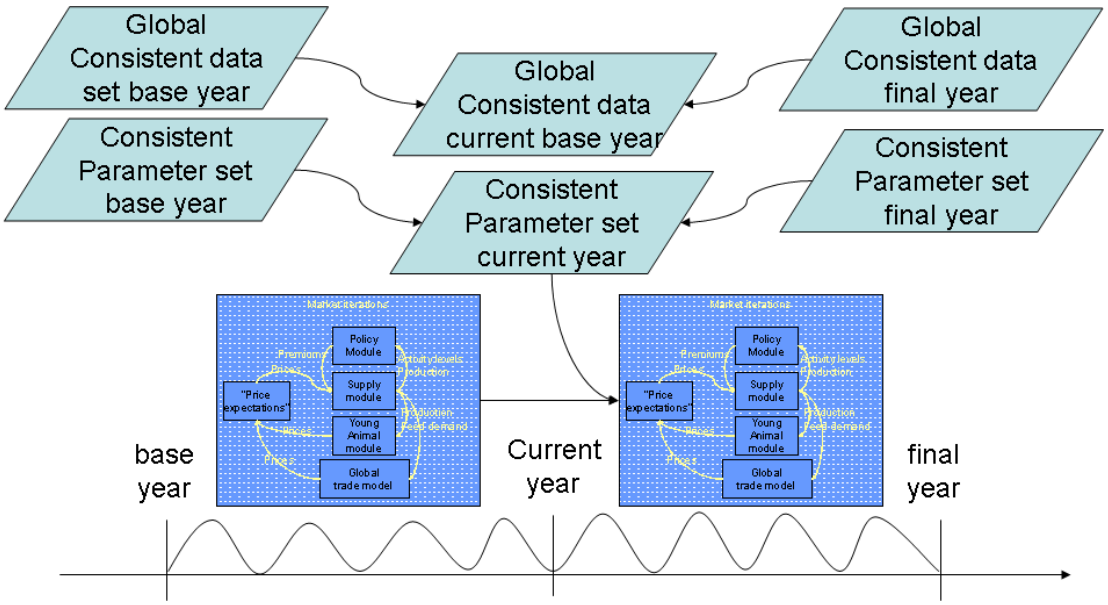
Trials to estimate simultaneously the adjustment cost parameters and the parameters of the long term cost function from a combined time-series cross-sectional estimation using regional time series on crop hectares, yield and the allocated costs from CAPREG as data set did not yield stable regression coefficients. It became therefore obvious that an econometrically based estimation of the adjustment cost terms was not feasible, one major argument to not include the recursive version in future CAPRI master releases.

Figure 35. General layout of the recursive dynamic version



Another important point was the baseline generation process. Albeit technically certainly feasible, computing time restrictions exclude running the projection tool CAPTRD for each single year between the base and the final simulation year. Accordingly, it was decided to calibrate the system to the two end points of the time series where coherent data sets – closed market balances, production equal to output coefficient times activity levels etc. – are available. All behavioural parameters for the time points in between were linearly interpolated. Should the policy instruments also follow a linear curve between the base and the final simulation year, the model should solve rather smoothly in between the two points.

Figure 36. Calibration process of the recursive-dynamic version



However, some traps need to be kept in mind. Firstly, the feed restriction part of the model comprises non-equalities. If those switch from binding to non-binding or vice versa between the base and the final simulation year, a linear combination of the two parameter sets fitting

to an optimal solution between the two points may generate even infeasible solution. Accordingly, the code for solving the supply model was revised as to automatically relax restrictions in the feed distribution part if infeasibilities occur. Despite that mechanism, the feed restriction parts due to its LP based character typically will provoke jumps if the regional supply model is solved along the linear interpolation between the base and final simulation year. The second traps roots in the fact that policy instruments do not develop linear over time, and some, as Tariff Rate Quotas, are highly non-linear in their character.

7 Post model analysis

7.1 Dual analysis (T. Jansson, W. Britz)

According to theory of (non)-linear optimization under constraints, there is a symmetric presentation of the primal model (the one we solve) by a dual one. The dual one would minimize the costs of "using" the constraints under the condition that marginal revenues of the endogenous variables cover marginal costs (including those for resource use) as long as the endogenous variables are not bouncing against lower limits. The equality of marginal revenues and costs provides important insights in the functioning of a programming model. Imagine e.g. that the model solution shows an increase in soft wheat area by 10%. What was driving that expansion? By comparing marginal revenues and costs between the two equilibrium points – in the sense that in both solution marginal revenues and costs are balances – we can decompose that expansion. What we could e.g. see is that an increase in premiums has driven up marginal revenues per ha, that revenues were reduced somewhat by lower prices, and that the revenue increase was exhausted by an increase in the land rent and an increase of the implicitly modelled costs of labour and capital as expressed by the quadratic cost function.

Note: see *'reports\margcr.gms'* and the dual analysis tables in the exploitation tools

7.2 Decomposition of changes in aggregates yields and activity related income indicators (W. Britz)

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 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.

It should be mentioned that we do not have information about the premiums paid in the baseline, so that the results shown for the modified Gross Value Added need to be interpreted keeping in mind that premiums might have changed. The same holds for the case where the scenario change yields or inputs per ha as part of the scenario description.

		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%

The code is implemented in „reports\yield_change_decomp". The table can be found in the GUI under "farm\decomposition".

Keeney R. & Hertel T. (2008a), YIELD RESPONSE TO PRICES: IMPLICATIONS FOR POLICY MODELING, Working Paper #08-13, August 2008, Dept. of Agricultural Economics, Purdue University

Keeney R. & Hertel T. (2008b), The Indirect Land Use Impacts of U.S. Biofuel Policies: The Importance of Acreage, Yield, and Bilateral Trade ResponsesGTAP Working Paper No. 52, 2008

7.3 Post model sensitivity analysis with features in the supply model (W. Britz)

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. Recently, the template of the supply model was extended by features such as the land supply curve, substitution between grass and arable land or the introduction of price dependent inputs and output coefficients for production activities

Result interpretation in explicit optimisation models can be cumbersome as no closed form representation of the behavioural response (such as supply quantities as a function of prices and policy instruments) is available. Even if the exploitation tool offer a dual analysis of the results, 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 is the risk of mis-interpretations which might then also lead to wrong conclusions about the impacts of policy instruments and other driver. The analysis can be improved and might draw less on assumptions by systematically analysing the contribution of the different allocative mechanisms in the model. Therefore, a relatively small and easy part for was added to the CAPRI simulation model.

7.3.1 Methodology

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. Less rich model variants are generated by changing the template, and then used to simulate the impacts. The results of these simpler variants can then be compared to the version used for the real production run.

In order to do so, the full model is first solved to convergence, i.e. the normal iterations between the supply and market module are not changed. Afterwards, the supply models are solved several times again at the final prices, each time switching specific features off, namely:

1. *Price dependent input and output coefficients*, by re-introducing the I/O coefficient from the baseline, i.e. those underlying the calibration point of the model.
2. The *land supply curve*, by treating arable and grass land as fixed endowment.
3. Substitution between *low and high yield technology variants*, by forcing the model to use equal shares of the available technology.
4. The *PMP cross terms between crop groups*, only leaving the diagonal terms while recalibrating the linear term such that the model would still calibrate at baseline price and policies.

7.3.2 Technical implementation

The implementation is highly modular in the sense that it does not really interact with any other code in the system. It consists of four extensions of the existing code basis:

- A definition file where the necessary parameters and sets are declared (*'sens\declare_sens.gms'*)

```
set  sensExperiments /

                        curResults
                        withoutEndogYields
                        withoutTechfShares
                        withoutLandSupply
                        withoutPMPCrossTerms

                      /;



scalar old_standard_land_market;
scalar old_fix_techf_shares;

Parameter p_sensResults(RALL,COLS,ROWS,sensExperiments);
```

- The file defining the sensitivity experiments and re-running the supply models (*'sens\sens.gms'*), which first defines for each sensitivity instruments what to change (an example is given below)

```
if ( sameas(sensExperiments,"withoutLandSupply"),

   --- store the current setting for endogenous land supply

   old_standard_land_market = standard_land_market;

   --- switch off land supply curve and substitution between
       gras and arable land

   standard_land_market = 1;

   --- available hectares in land types as in baseline

   DATA(RU,MCACT,"LEVL","Y") = DATA(RU,MCACT,"LEVL","TRD");
```

Starts the supply models:

```
         ',
*
*        --- run supply model
*
$include 'supply\simu_supply_grid.gms'
```

Stores the activity levels and quantities sold or purchased in a parameters:

```
*
*       --- store major results (currently only activity levels)
*
        p_sensResults(RU,MPACT,"LEVL",sensExperiments)
            = SUM(A $ TECHF(RU,MPACT,"LEVL",A), LEVL.L(RU,MPACT,A));
*
        p_sensResults(RU,"NETF",OMS,sensExperiments)
            = NETTRD.L(RU,OMS) * (1 - 2 $ I(OMS));
```

And finally, aggregates they to activity and product groups, such as in:

```
    ',
*
*   --- aggregation from regional units in supply model to higher regional levels
*
    p_sensResults(NUTS2AGG,MPACT,"LEVL",sensExperiments)
     = SUM(Types_to_R(NUTS2Agg,Types), p_sensResults(Types,MPACT,"LEVL",sensExperiments));

    p_sensResults(NUTS2AGG,"NETF",OMS,sensExperiments)
     = SUM(Types_to_R(NUTS2Agg,Types), p_sensResults(Types,"NETF",OMS,sensExperiments));
```

Note: the snippet above show only part of the code.

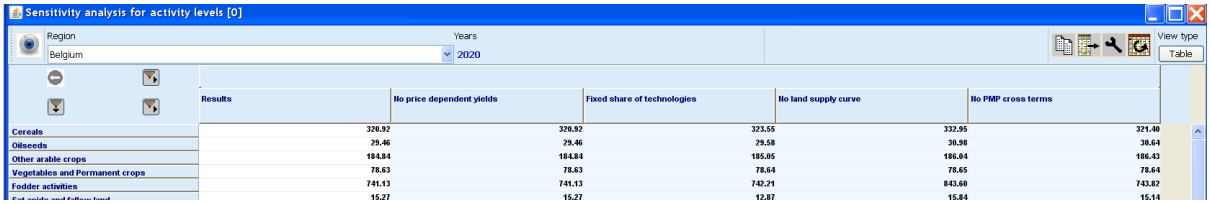
- Copying the sensitivity results on the parameter carrying all the results exported to GDX (see *'capmod\set_and_store_dataout.gms'*):

```
*
*  --- load and store results from sensitivity analysis
*
   execute_load "%scrdir%\sens.gdx" p_sensResults;

   DATAOUT(RSS,sensExperiments,COLS,"LEVL","%SIMY%") = p_sensResults(RSS,COLS,"LEVL",sensExperiments);
   DATAOUT(RSS,sensExperiments,"NETF",ROWS,"%SIMY%") = p_sensResults(RSS,"NETF",ROWS,sensExperiments);


   option kill=p_sensResults;
```

- Two new tables in the exploitation tools which report the changes (farm->Sensitivity Analysis for Activity Levels, farm->Sensitivity Analysis for Sales and Purchases):



Each column reports the results which were obtained if one of the features would have been switched off.

7.4 Welfare analysis (W. Britz, T. Jansson, P. Witzke)

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 linked to the so-called indirect utility function.

On the producer side, CAPRI uses changes in the gross value added (GVA) plus premiums as the main indicator. The gross value added is the different 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 enumerate labour, capital and land, independent on property rights of these factors. It should not be confused with "farm profits", which depend on ownership and some accounting rules for depreciation of capital, and where also non-family labour costs are already remunerated. Equally, the household income will clearly differ 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.

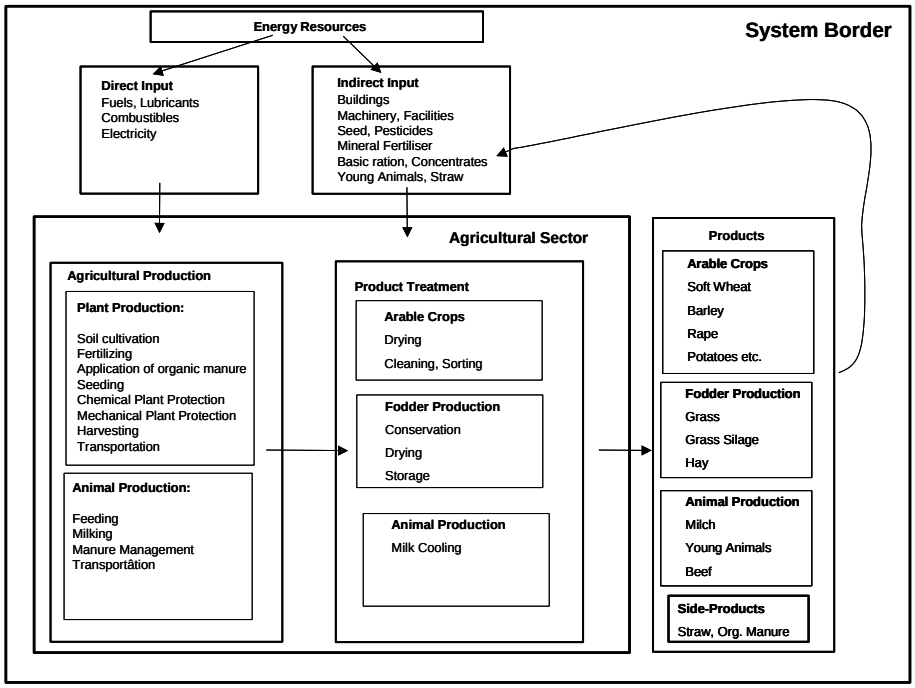
Tax payer costs in CAPRI refer to those policy instruments explicitly covered, i.e. premiums paid to farmers, cost of public market interventions and export subsidization and some subsidies paid for demanders of agricultural goods, minus revenues from import tariffs. For premium schemes, financing rates for the EU and national budgets as well as for pillar I and II of the CAP can be defined, which allow to allocate costs to those different budgets.

7.5 Energy Use in Agriculture (T. Kränzlein, M. Kempen)

7.5.1 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 37 shows the relevant borders and integrated processes for CAPRI energy module. The term “Agricultural Production” in Figure 37 is the main interface between LCA and CAPRI production activities.

Figure 37. System border and processes considered for CAPRI energy input assessment



Source: Based on ecoinvent (2003)

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 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.

7.5.2 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.

7.5.2.1 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 28 shows the main direct energy sources used in CAPRI.

Table 28 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: ecoinvent (2003)

7.5.2.2 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.

7.5.2.3 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 Switzerland (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. 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.

7.5.2.4 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.

7.5.2.5 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 25. The following chapters give an overview on the methodology to estimate indirect components.

Table 29 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

Source: Ecoinvent (2003) and SALCA061 (2006)

7.5.2.6 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 29. Those are compiled using average registered consumption quantities of the single fertilisers on the market, which are broken down to their active substance content.

7.5.2.7 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 29. 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.).

7.5.2.8 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. Machinery 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.

7.5.2.9 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 29, 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.

7.5.2.10 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.

7.5.2.11 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.

7.5.2.12 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 7.5.2.7. 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.

7.5.2.13 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.

7.5.2.14 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 calculation 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 29).

7.5.3 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)

7.5.3.1 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 30. The procedure follows literature data.

7.5.3.2 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 30.

Table 30 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	-

Source: CAPRI Modelling System

7.5.4 Analysis of CAPRI energy module results

The results of the CAPRI energy module can be displayed in various ways and on different levels. Table 31 gives an overview. Further down, each parameter is shown in more detail.

Table 31 Energy module results structure

Parameter	Parameter Unit	Description	Availability
Energy per CAPRI activity unit	MJ/ha; MJ/head	Covers all energy requirements necessary for one CAPRI activity unit per year	Region-specific and activity-specific; weighted averages on NUTS-0 and EU level
Energy per CAPRI output unit	MJ/kg	All energy requirements for one CAPRI activity unit are divided by the output level; allocation between main product and by-products is carried out for a number of activities	Region-specific and activity-specific; weighted averages on NUTS-0 and EU level
Energy efficiency – Type "energy"	MJ/MJ	The output level of a CAPRI activity is assessed by its energy content (See Chapter 7.5.3) whereas allocation between main product and side-products is done for some activities. The result is divided by all energy requirements of the CAPRI activity unit. In short: Energy output (per kg) divided by energy input (per kg)	Region-specific and activity-specific; weighted averages on NUTS-0 and EU level
Energy efficiency – Type "Finance"	MJ/€	The output level of a CAPRI activity is assessed by its energy content (See Chapter 7.5.3) whereas allocation between main product and side-products is done for some activities. The result is divided by the income for the activity.	Aggregated on NUTS-II, NUTS-0 and EU level
Energy balance	MJ	The output level of all CAPRI activities of a region are assessed by its energy contents (See Chapter 7.5.3) whereas allocation between main product and side-products is done for some activities and then sum up over the region. The input energy requirements for all CAPRI activities are multiplied with the relevant activity levels and then sum up over the region. The result shows energy requirements (INPUT) and energy output (OUTPUT). Imports and exports of energy can be shown separately.	Aggregated on EU level
Energy requirements-overview	MJ/ha; MJ/head	On an activity-based, regional level, the composition of total energy requirements can be shown on an aggregated level.	Region-specific and activity-specific; weighted averages on NUTS-0 level

Energy requirements-detail	MJ/ha; MJ/head	On an activity-based, regional level, the composition of total energy requirements can be show n on in detail.	Region-specific and activity-specific; w eighted averages on NUTS-0 level
Energy input units	Input unit/ha; Input unit/head	On an activity-based, regional level, the composition of input units driving the energy needs can be show n in detail..	Region-specific and activity-specific; w eighted averages on NUTS-0 level
Energy content products	MJ/kg product	On an activity-based level, the energy content for products can be show n; energy assessment of output is based on this parameter; Energy content is assumed being equal throughout all NUTS-II regions.	Activity-specific

Source: CAPRI Modelling System

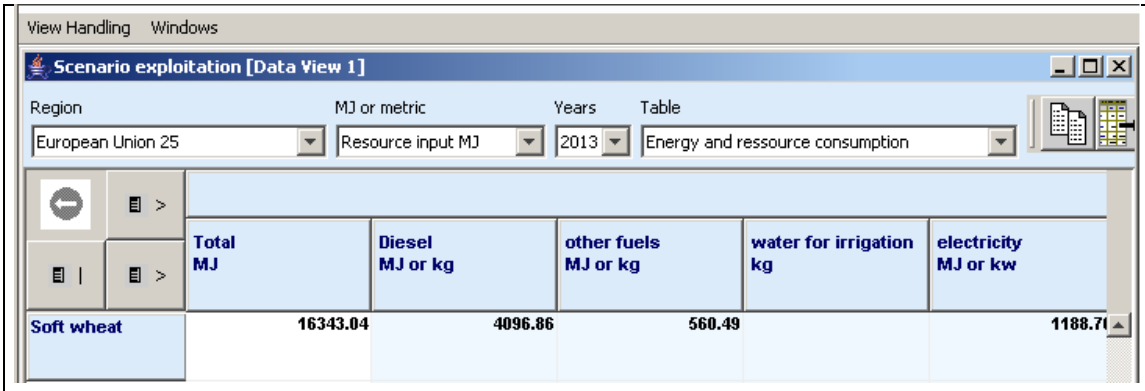
7.5.4.1 Application notice

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.

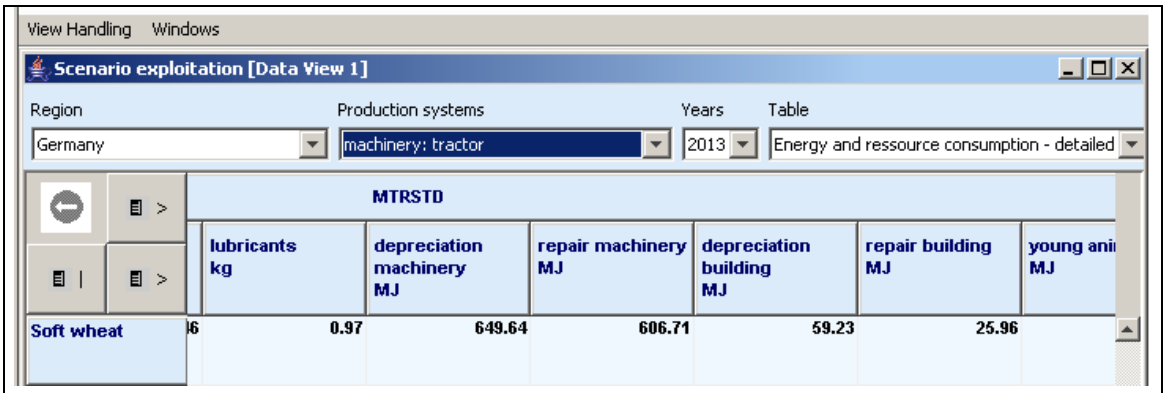
7.5.4.2 Structure of output tables

A broad range of output tables permits to display results of the energy indicator. Some of those are presented in this chapter. Figure 38 shows some examples for display modes of the energy indicator.

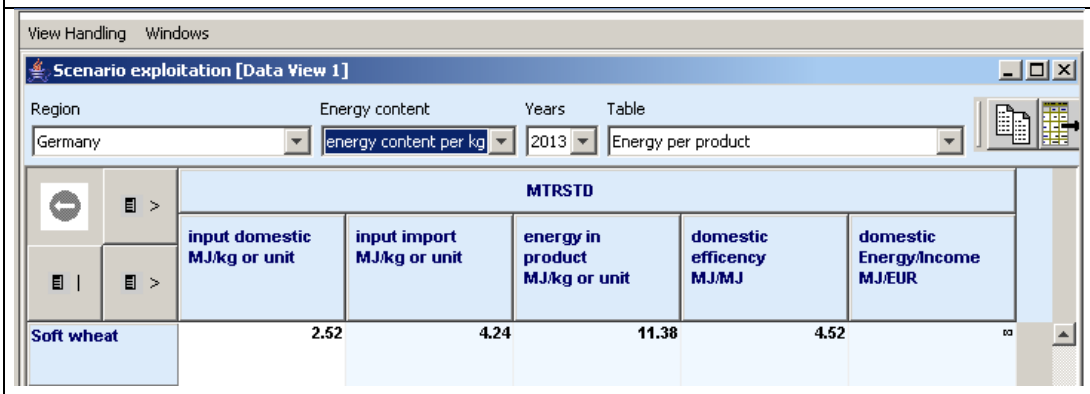
Figure 38. Energy parameters: examples for results displaying



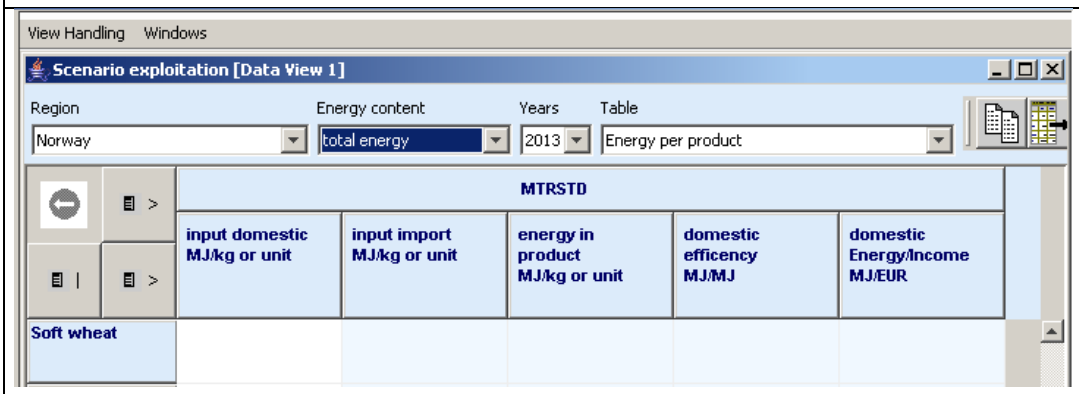
Example 1: Energy consumption - overview



Example 2: Energy consumption - detailed



Example 3: Energy parameters w ith 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 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.

8 Spatial dis-aggregation and link to the biophysical model DNDC

8.1 A spatial land use map (M. Kempen and R. Köbl)

Not at least due to the so-called multi-functional model of European agriculture, there is growing interest in modelling environmental effects of the agricultural sector in the EU. In many cases, results beyond rather crude passive indicators can only be obtained linking biophysical models to economic models for policy impact analysis. An important methodological problem in this context is "bridging" the scales: whereas most bio-physical models work on field scale, comprehensive EU-wide economic models generally work on large administrative regions.

Within these administrative boundaries the natural conditions of soil, relief and climate usually differ in such a manner, that the assumption of identical cropping pattern, yields or input use cannot be maintained. Simulations with bio-physical models thus require breaking down results from the economic models into a smaller regional scale. This paper proposes a statistical approach combining a logit model with a Bayesian highest posterior density estimator to break down production data of 30 crops in about 150 European administrative regions for EU15 (NUTS 2) to, so called, Homogeneous Spatial Mapping Units (HSMUs).

The approach is based on two steps. The first step regresses cropping decisions in each HSMU on geographic factors (soil, climate etc.), using results of the Land Use / Cover Area Frame Statistical Survey (LUCAS) providing observations on agricultural crops at approximately 40.000 sampling points all over the EU territory. Spatial statistical techniques are used to allow for spatial heterogeneity of the coefficients using a locally weighted logit model. In the second step of the disaggregation procedure, simulated or given data for the administrative Nuts II regions are broken down to HSMU level by Bayesian methods. Two possible ways to introduce prior information from the logit regression step are discussed: (1) using means and variances of the predicted shares in each HSMU, or (2) using the estimated coefficients and their covariance matrix in the Nuts II region. In the first case, we search for shares at HSMU level consistent with Nuts II results maximizing the posterior density of the predicted shares. The second approach selects the most probable set of regression coefficients producing data consistent shares over all HSMU maximizing the posterior density of the coefficients.

The basic approach – estimating prior information and achieving consistency between scales afterwards is in line with previously suggested disaggregation procedures (Howitt and Reynaud, 2003). While different estimation procedures are motivated by data availability, the proposed method contributes to the literature in the following respects: (1) Lower level units are defined by homogeneous production conditions rather than administrative boundaries; (2) Functional relationships between location factors and land use are identified explicitly using spatial statistical techniques. This allows to discern prior information on crop shares even under scarce data information for some lower level units; (3) The applied Bayesian method fully and transparently accounts for the available prior information – prior distributions – when searching for consistency between the scales.

8.1.1 Spatial calculation unit (HSMU)

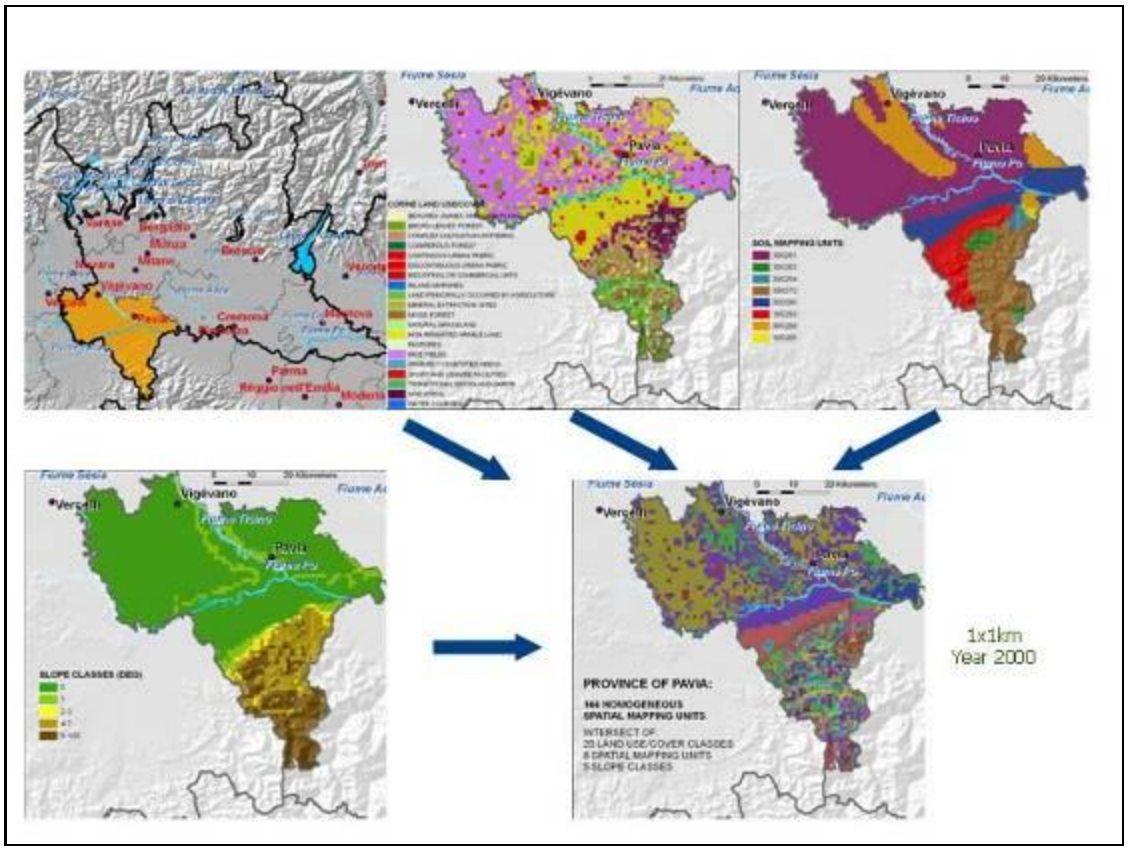
The aim of building HMSU is the definition of areas inside an administrative region where approximate homogeneity according location factors may be assumed. The HMSU serve then as simulation units for the bio-physical models and are constructed by overlaying different maps (land cover, soil map, climatic factors etc.). In order to allow for a manageable number of HSMUs, the most important factors must be selected, and continuous parameters must be grouped in classes.

We chose four delimiters to define a spatial calculation unit, which in the following is also denoted as "Homogeneous Spatial Mapping Unit (HSMU), i. e. soil, slope, land cover and administrative boundaries. The HSMU is regarded as similar both in terms of agronomic practices and the natural environment, embracing conditions that lead to similar emissions of greenhouse gases or other pollutants.

The HSMUs are built from four major data sources, which were available for the area of the European Union i. e. the European Soil Database V2.0 (European Commission, 2004) with about 900 Soil Mapping Units, the CORINE Landcover map (European Topic Centre on Terrestrial Environment, 2000), and a Digital Elevation Model (CCM DEM 250, 2004). Prior to further processing all maps were re-sampled to a 1 km raster map (ETRS89 Lambert Azimuthal Equal Area 52N 10E, Annoni, 2005) geographically consistent with the European Reference Grid and Coordinate Reference System proposed under INSPIRE (Infrastructure for Spatial Information in the European Community, Commission of the European Communities, 2004).

One HSMU is defined as the intersection of a soil mapping unit, one of 44 Corine land cover classes, administrative boundaries at the NUTS 3 level (EC, 2003; Statistical Office of the European Communities (EUROSTAT), 2003), and the slope according to the classification 0 degree, 1 degree, 2-3 degrees, 4-7 degrees and 8 or more degrees. As the HSMU of at least two single pixel of one square kilometer are not necessarily contiguous, we can speak from the HSMU as of "pixel cluster".

Figure 39. HSMU creation



The HSMUs cover a wide range of sizes from a minimum area of 1 km2 but some reach very large areas (up to 9,723 km2) in regions with a homogeneous landscape in terms of land cover and soil. The mean area of a homogeneous spatial mapping unit, indicates the range of environmental diversity with regard to land cover, administrative, data, soil and slope, and ranges from 7 km^2 for Slovenia to 94 km^2 for Finland with an European average around 21

km^2. In total, they cover a total of 206,000 km^2 for the almost 4,300,000 km^2 in Europe. Small discrepancies in the the surface area of countries (ranging from 2,597 to 546,745km2 still for Luxembourg and France respectively) stems from rounding errors during the re-sampling procedure and higher in areas with a high geographical fragmentation (i. e., small islands, complex coastlines or borders).

For EU27 we obtained in total about 138,000 HSMUs in which agricultural activities (arable land and grassland) are allowed to occur, occupying about 77% of the European landscape.

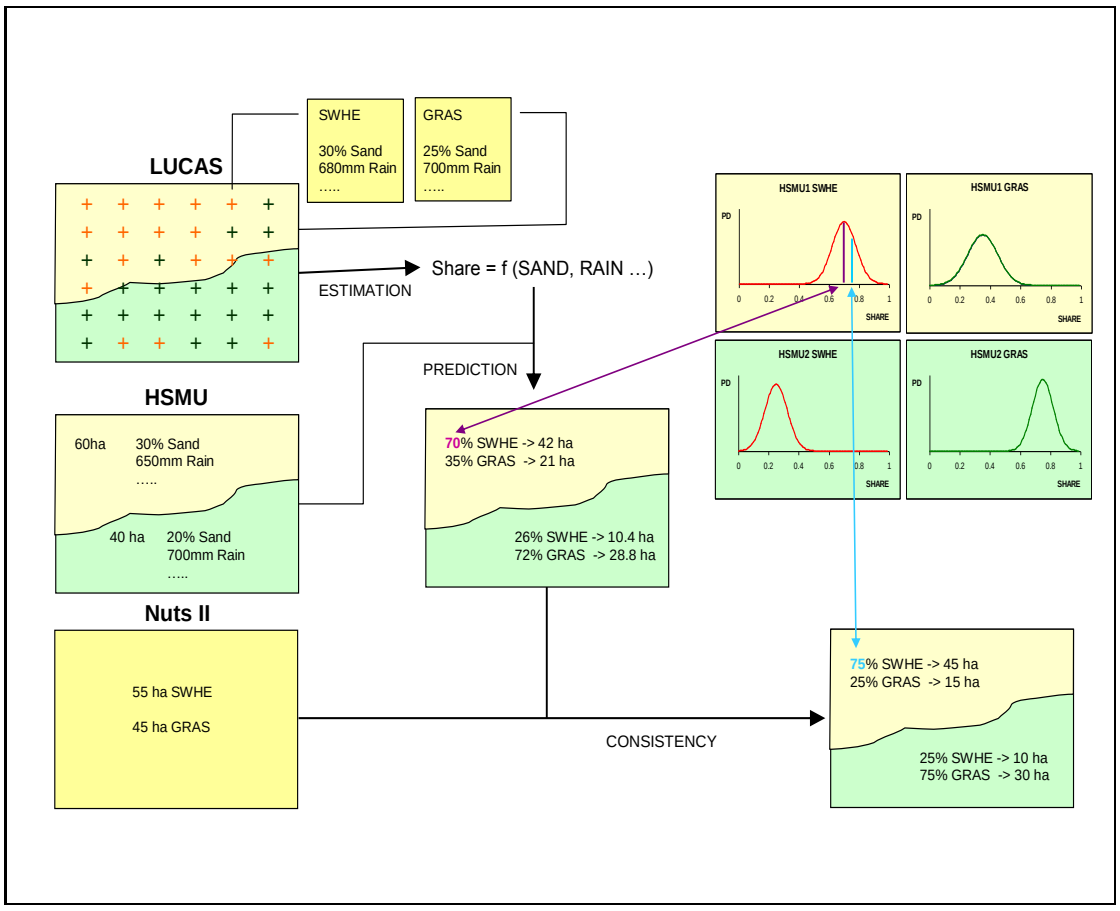
8.1.2 Consistent disaggregation of land use

The calculation of a spatial explicit land use map makes use of the LUCAS survey. In opposite to mapping approaches, area frame surveys based on a common statistical sampling method gather land cover and land use data at specific sample points, only, and extrapolate from these to the entire area under investigation. LUCAS covers the territory of all EU Member States and all kinds of land uses, and is based on a two-stage sampling design: at the first level, so-called Primary Sampling Units (PSUs) are defined as cells of a regular grid with a size of 18 × 18 km, while the Secondary Sampling Units (SSUs) are 10 points regularly distributed (in a rectangular of 1500 × 600 m side length) around the centre of each PSU resulting in approximately 10.000 PSUs for the whole EU.

Due to possible measurement errors regarding the geo-references in the CORINE maps (Gallego 2002), about 30% of the LUCAS points closer then 100 m to the border of a CORINE class were not considered in here. The 38 agricultural classes found in LUCAS (36 crop land, 2 permanent grassland classes) were re-grouped according to the crops found in CAPRI. All other classes (artificial areas, woodland, water, etc.) are aggregated in a residual classed termed "OTHER".

Before describing the crucial steps in detail the general approach of the disaggregation procedure is illustrated in Figure 40. Suppose there is a Nuts II region divided in only two HSMUs each comprising two crops – grassland (GRAS) and soft wheat (SWHE). Combining the LUCAS survey with digital maps provides us with several observations of crops grown at a defined point with a set of natural conditions. Using an adequate estimation model we can regress the probabilities of finding a crop at a certain location on the natural conditions. As this probability can be interpreted as the share of the crop in a homogeneous region, applying these estimated coefficients to the average natural conditions in a certain HSMU yields normally distributed predictions of crop shares for this HSMU under corresponding assumptions on the stochastic processes governing crop choice. These a priori information on cropping shares are generally not consistent with the "known" cropping area in the Nuts II region. The "best" set of data-consistent shares given the prior information is identified by a Bayesian *highest posterior density* approach. The concept of the HPD estimator allows the direct inclusion of the uncertainty of the prior mean. The variance can be derived from asymptotic properties or bootstrapping procedures.

Figure 40. Scheme of Disaggregation Procedure



Locally Weighted Binomial Logit Estimation

Generally, shares for each crop $\hat{Y}_C$ are regressed on the following explanatory variables describing natural conditions:

- Set of soil code
- Drainage
- Presence of stones
- Slope
- Elevation
- Rainfall
- Sum of temperature in vegetation period

The regressions were estimated independently for each crop *c* in each CORINE class *clc*:

The arguments for using specific coefficients for each CORINE class are as follows. Assume grass land parcels are found in the LUCAS survey in the "non-irrigated land" CORINE class. We would assume that slope has a positive effect on the probability to find grass. In the "pasture" class of CORINE, we would eventually find the opposite effect: with increasing slope, grass land could be replaced by forest. For convenience the indices *c* and *clc* are omitted in the following.

The LUCAS survey reports one point in time observations and hence does not deliver cropping shares (or rotations), but requires a binary choice model. Both logit and probit models (see e.g. Green 2000) were originally tested, with the logit approach giving slightly

better results. The likelihood function of finding crop *c* at a specific LUCAS point *i* for the binomial logit model is defined as:

$$\Lambda(\boldsymbol{\beta}'\mathbf{x}_i) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_i}}$$

$$\log L = \sum_{i=1}^{n} [y_i \log \Lambda(\boldsymbol{\beta}'\mathbf{x}_i) + (1-y_i)(1-\Lambda(\boldsymbol{\beta}'\mathbf{x}_i))]$$

Equation 176

where $\boldsymbol{Y}$ is a dummy vector indicating whether a certain crop was observed at a location *i* (y_i=1), $\mathbf{x_i}$ is the design matrix containing data on natural conditions and $\Lambda(\boldsymbol{\beta}'\mathbf{x_i})$ is the probability that a specific crop is grown at location *i*.

Applying the estimated $\hat{\boldsymbol{\beta}}$ to the average natural conditions in a HSMU ($\mathbf{x}_h$) give us a prior estimate for the share of a specific crop in a certain HSMU:

$$\hat{\mathbf{Y}} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_h) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_h}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_h}}$$

Equation 177

Binomial versus Multinomial Regression

The approach discussed above examines the crops independently from each other and thus neglects the information that crops compete for the available land, with two possible effects. Firstly, the error terms for the different crops are probably correlated, and secondly, the individual estimated shares don't add up to unity. The multinomial probit model would be ideal as it allows for an unrestricted variance covariance structure of the error terms and satisfies the additivity condition, but is computationally infeasible for 30 crops and 10.000 points. The assumption of an identity matrix for the variance covariance matrix underlying the multinomial logit model was deemed as too inflexible (Nelson et al. 2004), albeit it is easier to solve. The way out might be a nested logit model, a possible expansion in further analysis.

However, both problems were not deemed crucial for the application at hand. Given the large number of observations, the possible gain of taking correlations between the error terms across crops into account is most probably small. Furthermore, the violation of the adding up condition for the shares is explicitly accommodated in the second step of the disaggregation procedure, where the estimated shares serve as prior information, only.

Local versus Global Regressions

The assumption of European wide invariant relationships between the share of each crop and a limited number of location factors describing natural conditions may be problematic if other omitted explanatory factors are not randomly distributed in space, but "clustered". Suppose, for example, two HSMUs with identical natural conditions, the first one close to a sugar refinery, and the second one far way from the next sugar plant. The share of sugar beets in the first unit will be probably much higher, an effect not linked to the natural conditions. Clearly, omitted variables as the effect of sugar refineries could lead to seriously biased parameter estimates. Adding more explanatory variables would certainly help, but it is simply impossible to collect information on all probably relevant factors (market points, transport infrastructure, environmental legislation, etc.). Instead, spatial econometric techniques are applied to overcome the problem of omitted variables that are correlated over space.

The basic idea behind Locally Weighted Regression, which was proposed by Cleveland and Devlin (1988), is to produce site specific coefficient estimates using Weighted Least Squares to give nearby observation more influence than those far away. Further on, the estimation for any specific site is limited to a number of observations within a certain bandwidth around the site. Locally Weighted Regression are mostly found combined with Least Squares estimators,

but application to Maximum Likelihood Estimation as needed in the case of discrete dependent variables are described as well (Anselin et al. 2004).

The weight given to any observation *i* in constructing the estimate for site *j* is given by ω_{ij}. The tri-cube is a commonly used weighting function:

$$\varpi_{ij} = \left[1 - \left(\frac{\delta_{ij}}{d_j}\right)^3\right]^3 I(\delta_{ij} < d_j)$$

Equation 178

Where δ_{ij} is the distance between site *i* and observation *j*. d_j is the bandwidth and $I(.)$ is an indicator function that equals one when the condition is true. The effect of any one location in space on near points thus falls depending on the distance and becomes zero once the distance exceeds the bandwidth. There are other common weighting schemes like the Gaussian function or several Kernel weighting functions (see: Anselin et al. 2004 or Fotheringham et al. 2002). But it has been shown that opting for a proper bandwidth is more significant than choosing a certain spatial weighting function.

When there is no prior justification for applying a particular bandwidth, an appropriate bandwidth can be found by the minimising either the cross-validation score (CV), the Akaike Information Criterion (AIC) or the Schwartz Criterion (SC). The AIC and the SC are offered by most software packages. The CV is calculated as:

$$CV = \sum_{i=1}^{n} \left(y_i - \hat{y}_{i \neq i}\right)^2$$

Equation 179

where *n* is the number of data points and the prediction for the *i*th data point $\hat{y}_{i \neq i}$ is obtained with the weight for that observation set to zero. Each of the criteria can be minimised by a golden section search (see Press et al. 1989). In our study all criteria led to similar results. We opted to minimise the Schwartz Criterion, because according to Boots et al. (2002) it seems to have better large sample properties.

In typical applications, sites and observations would be identical. In our context, that would require estimates per crop and CORINE class for each LUCAS point, which is computational impossible. Instead, the NUTS II regions were chosen as sites. When estimating for a particular NUTS II region, all LUCAS point inside that NUTS II region received uniform unity weight, and points in neighbouring NUTS II regions weights equal or smaller unity according to (4). That still leads to a large number of possible estimations: 150 Nuts II regions times 10 agricultural CORINE classes times 30 crops, but fortunately, many of the combinations do not comprise any observations. Weighting each likelihood contribution with ω_{ij} gives (Fotheringham et al. 2002):

$$\log L = \sum_{i=1}^{n} \omega_{ij} [y_i \log \Lambda(\boldsymbol{\beta}_\mathbf{j}^{'} \mathbf{x}_\mathbf{i}) + (1 - y_i)(1 - \Lambda(\boldsymbol{\beta}_\mathbf{j}^{'} \mathbf{x}_\mathbf{i}))]$$

Equation 180

Attaining Variance of Land Use Shares

Given the non-linear character of the estimations, the variance-covariance matrices offered by the statistical packages are not analytically calculated, but are instead numerically approximated which proved to be not suitable. Quite small predicted mean values in combination incredibly high variances led to shaky final results. Consequently the estimation of the prior variance attracts our attention. Statistical formulas can be used to derive the variance of a predicted mean. Furthermore results from bootstraps from the finite sample can be taken to calculate the variance.

The prior variance $\hat{\mathbf{Y}}$ is based on the asymptotic covariance matrixes for the coefficients. A robust covariance matrix can be calculated analytically (see White (1982)) as:

$$V_\beta = Cov\left[\hat{\boldsymbol{\beta}}\right] = \hat{\mathbf{H}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{H}}^{-1}$$

Equation 181

where for the weighted logit model the elements of Hessian ***H*** and the Brendt, Hall, Hall and Hausman matrix ***B*** are given by (Green 2000):

$$\mathbf{H} = \frac{\partial^2 LogL}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = -\sum_i \omega_i \Lambda_i (1-\Lambda_i)\mathbf{x_i}\mathbf{x_i}'$$

Equation 182

$$\mathbf{B} = \sum_i \omega_i (y_i - \Lambda_i)^2 \mathbf{x_i}\mathbf{x_i}'$$

Equation 183

As insignificant parameter estimates might influence the efficient calculation of a robust covariance matrix although they do not influence the forecasted value, insignificant variables were removed from the estimations. The variance of $\hat{\mathbf{Y}}$ builds upon the calculated covariance matrix $\mathbf{V}_\beta$.

$$\mathbf{V_Y} = Var\left[\hat{\mathbf{Y}}\right] = \Lambda_i(1-\Lambda_i)\mathbf{x}'\mathbf{V}_\beta\mathbf{x}$$

Equation 184

Using specific x_{HSMU} yields variances of the predicted land use share in each HSMU (Green 2000).

Bootstrap resampling procedures can alternatively be used to derive the variance of $\hat{\mathbf{Y}}$. Once the bandwidth is defined, we draw randomly with replacement *n* values from available dataset of $\mathbf{Y}_i$ and X_i and reestimate the model. After repeating this process B times and calculating $\hat{\mathbf{Y}}_b = \Lambda(\boldsymbol{\beta}_b{}'\mathbf{x}_h)$ the variance of $\hat{\mathbf{Y}}$ is simply the variance of the B $\hat{\mathbf{Y}}_b$ values:

$$V_Y = VAR(\hat{Y}) = \frac{1}{B-1}\sum_{b=1}^{n}\left(\hat{Y}_b - \hat{Y}\right)^2$$

Equation 185

As there is anyway a large number of models to be estimated it would be impractical to apply these procedure to all possible combinations. Therefore we applied it only to a limited number of Nuts II regions in France and Spain and opted for B=100.

Data-consistent Disaggregation

The second step of the disaggregation procedure identifies crop shares in each HSMU using the prior information on the estimated crop shares from the first estimation step under two data constraints: Firstly, adding up the areas per crop in each HSMUs must recover the cropping areas *CA* for that crop at NUTS II level. Secondly, the posterior shares in each HSMU must add to unity, including all non-agricultural land use from the LUCAS survey aggregated to the category "OTHER". In opposite to the first step this requires simultaneous accounting for all crops *c* in all relevant HSMUs *h*. The notation is therefore extended, e.g. from *Y* to $Y_{c,h}$.

The crop areas in each HSMU are defined by multiplying the posterior shares $Y_{c,h}^{con}$ with the entire area A_h thus

$$\sum_{h\in N2} Y_{c,h}^{con} A_h = CA_{c,N2}$$

Equation 186

and the adding up to unity

$$\sum_{C} Y_{c,h}^{con} = 1$$

Equation 187

must be imposed.

As the predicted unrestricted shares will typically violate the constraints, a penalty function is necessary to define the optimal deviations from the predictions. Generalized Maximum Entropy (GME) techniques (Golan, Judge and Miller 1996) have often been used for this type of data balancing exercises in recent times. Here, however, a *Bayesian highest posterior density (HPD) estimator* is applied allowing for a direct and transparent formulation of prior information and considerably reducing the computational complexity compared to the GME approach (Heckelei et al. 2005). The prior information is expressed as normal densities of predicted shares, with mean vector $\hat{\mathbf{Y}}_{c,h}$ and variance derived by the methods described before. After taking logs, the prior density function for the consistent shares $Y_{C,HSMU}^{con}$ is:

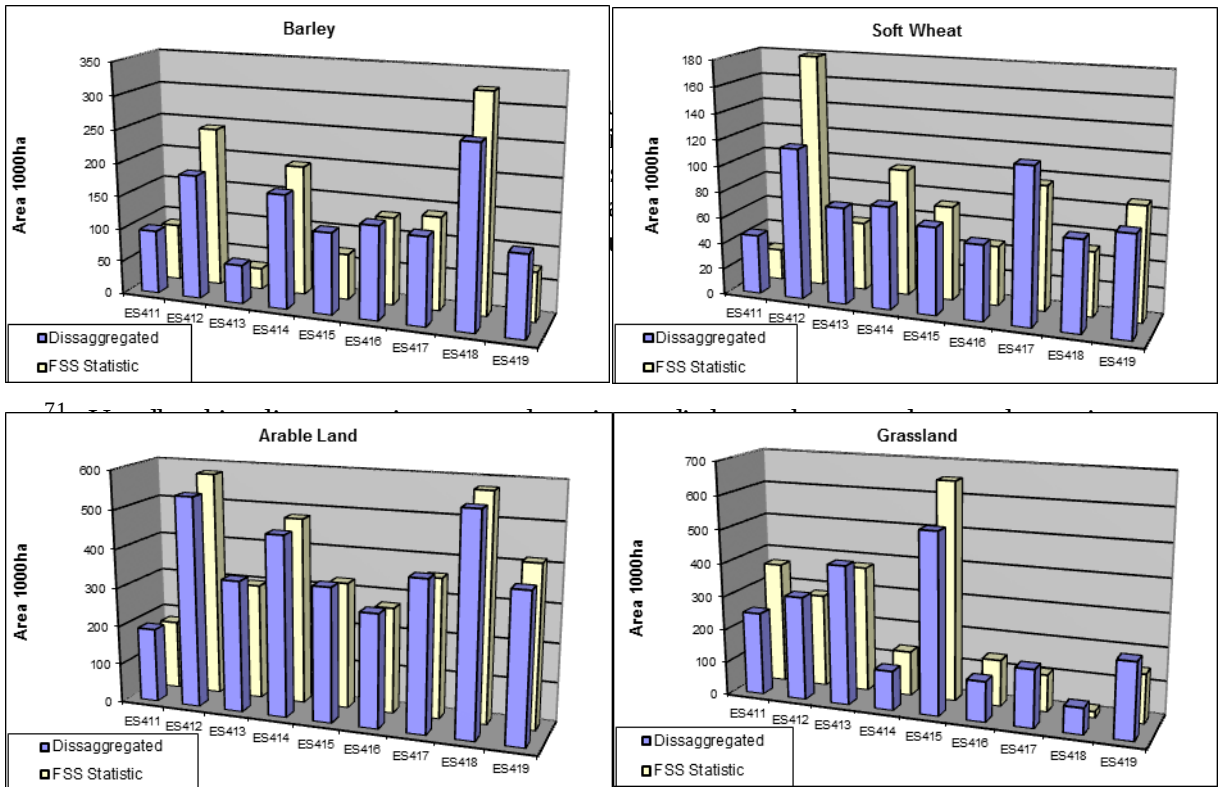
$$-\sum_{c}\sum_{h}\left[\log\left(\sqrt{2\pi}\mathbf{V}_{\mathbf{Y_{c,h}}}\right)+\frac{\left(\mathbf{Y}_{c,h}^{con}-\hat{\mathbf{Y}}_{c,h}\right)^{2}}{2\mathbf{V}_{\mathbf{Y_{c,h}}}^{\ 2}}\right]$$

Validation of Results based on asymptotic Variance

As it is hardly possible to present here the actual outcome of the disaggregation procedure – 30 land use shares in 100.000 HSMU. In order to enable the use of the data for further calculation the outcome is translated to a GIS map. In this section we want to focus on validation of the results

For some European regions, land use statistics at a lower administrative level, called Nuts III, are available from the farm structure survey (FSS; EUROSTAT, 2002). This information is used as out-of-sample observation to validate the results of the disaggregation algorithm, which predicts cropping areas for the HSMUs consistent to NUTS II[71]. Adding up over the corresponding HSMU yields crop areas at NUTS III level that can be compared to the observed data. Figure 41 exemplary contrasts actual and predicted cropping areas for selected crops in the nine Nuts III regions in Castilla-Leon, Spain. Although the disaggregation reflects the principal pattern quite well there are sometimes large differences.

Figure 41. Comparison of estimated and observed shares in NUTS III region for different crops (Castilla Leon ES410)



[71]

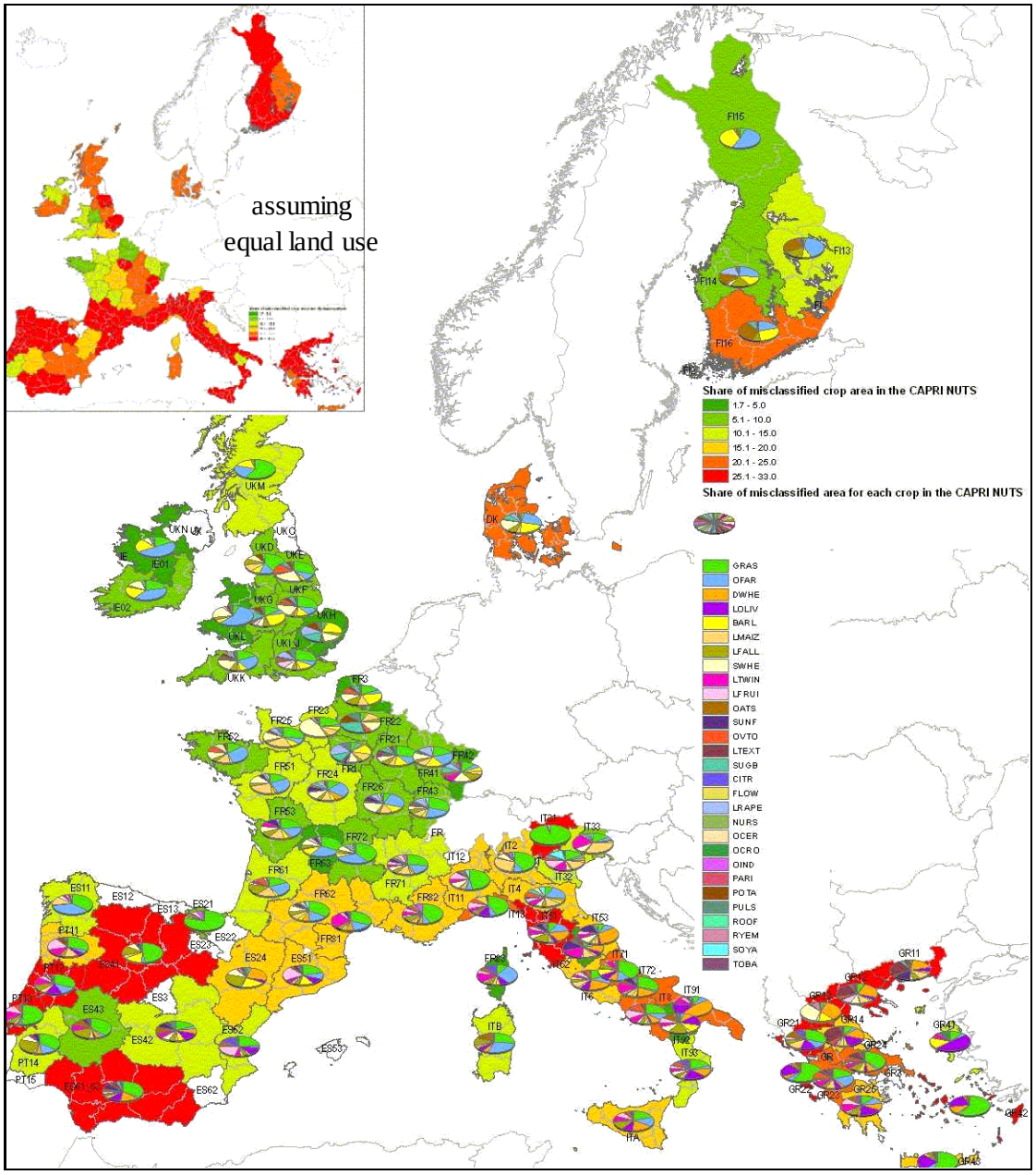
Table 32 Precentage of misclassified area for different crops (Castilla Leon ES410)

Crop	Missclassified Area in Nuts III (% of UAA in Nuts II)			
	Single Crop	Groups		
Soft Wheat	3.39	6.44	Arable Land 3.96	UAA 8.43
Durum Wheat	0.32			
Barley	5.02			
Rye	0.87			
Oats	1.46			
Maize	1.24			
Other Cereal	0.06			
Fallow Land	2.95	2.95		
Rice	0.00	1.09		
Sunflower	1.19			
Soya	0.00			
Texture Crops	0.59			
Pulses	0.34			
Other Crops	0.00			
Potatoes	0.24	0.83		
Sugar Beet	0.60			
Root Crops	0.01			
Rape	0.02			
Tobacco	0.01	0.18		
Other Industrial	0.04			
Tomatoes	0.00			
Other Vegetable	0.17			
Flowers	0.00			
Other Fodder	1.88	Fodder Production 10.15		
Grassland	9.27			
Nursery	0.01	Permanent Crops 0.4		
Fruits	0.10			
Citrus	0.00			
Olive	0.11			
Vine	0.40			
Nuts II	**30.29**	**22.39**	**14.51**	**8.43**

Figure 42 illustrates the misclassified areas in Europe where out of sample data is available from the FSS statistics. In regions with a high percentage of misclassified area often grassland accounts for a significant part of the errors. This is astonishing since grassland has its "own" Corine land cover class and indicates that misclassification might not only be a consequence of a poor disaggregation procedure but also a result of contradictious data sources[72]. Nonetheless the dissaggregation is a significant improvement compared to the assumption of identical cropping pattern within each Nuts II region.

[72]The Corine land cover map reports indeed about 2 Mio ha "Pasture" and "Natural Grassland" in Spain while in the FSS statistic about 9 Mio ha Grassland are declared.

Figure 42. Precentage of misclassified areas in validated Nuts II Regions after disaggregation



8.2 Yields, irrigation shares and stocking densities (W. Britz)

The crop yield estimation combines three different types of a priori information in a HPD estimation framework to derive simultaneously spatially explicit yield estimates and irrigation shares per crop. A first input data set in the estimation process is the irrigation map from FAO used to provide per HSMU an estimate of the share of irrigated agriculture. Secondly, the FSS delivers data for irrigated areas for certain crops at administrative level and, thirdly, MARS offered potential yields for rainfed and fully irrigated agriculture. The FSS data about irrigated hectares at regional scale had been used via regressions to find some basic relations between soil properties and climatic parameters and the irrigated share per crop or crop group. From those regression models, forecasts are derived at the level

of single HSMUs about the irrigated share per crop. The HDP framework minimizes simultaneously deviations from the estimated crop specific irrigation shares per HSMU, from the irrigation shares per HSMU derived from the FAO map and from the potential yields. Constraints ensure that firstly the area weighed average of the yields per HSMU is equal to the one found in regional statistics, and secondly that the irrigated area per HSMU exhaust the irrigated area at regional level found in the FSS.

The crop yields are used as explained below as explanatory factor in the estimation of animal stocking densities and drive as well the estimate of crop specific fertilizer application rates. Using simple linear input demand function per crop activity for the different inputs (plant protection, repair costs etc.) and assuming uniform prices for output and inputs insides the administrative units, the crop yields at HSMU level are also used to derive economic indicators per crop (revenues, variable costs, gross value added, gross value added plus CAP pillar I premiums). It is planned to add soon estimates about CAP pillar II payments.

Unfortunately, in opposite to the LUCAS sample for crops, no high resolution observation sample for animal stocking densities at Pan-European level is available. Additionally, especially for area independent animal production activities as pigs and poultry, a weak relation between local natural factors as soil and climate and stocking densities can be expected. Therefore, the estimation of stocking densities builds on a cross-sectional estimation from the Farm Structure Survey for a mix of NUTS II and NUTS III administrative units with overall about 500 observations for EU27 per animal 7 category. Regression models for the different animal activities in CAPRI as well as aggregates for ruminant and non-ruminants were estimated, using crop and land cover shares (forest, shrubs, total UAA, non-agricultural land cover, cereals, grassland, fodder maize, all type of fodder production), fodder maize and cereals yields as well as revenues and GVA plus premiums per ha for Grandes Cultures and cereals, altitude and slope along with climate data (annual rain fall, temperature sum, length of the vegetation period) as explanatory variables. All variables were offered untransformed, as squares and square roots to the estimator. The estimators then used a backward elimination, removing explanatory variables as long as the adjusted R squared was increasing or a variable was not significantly different from zero at the 5% level. In order to account for specific national legislation and market conditions, either the FSS regions of a country were estimated separately (France, Italy) or national dummies we used in the estimation for group of countries (Group 1: Germany, The Netherlands, Belgium; Group 2: Spain, Portual and Greece; Group 3: Denmark, Sweden, Finland, UK, Irland and Austria; Group 4: EU12). Such grouping ensured sufficient degrees of freedom during estimation. Not surprisingly, the explained variance for the ruminants was general high in the range of 80% and above, whereas for pigs and poultry, R^2 were in some instances as low as 40%. As own produced fodder and organic fertilizer may be transported easily even over several kilometres, it was decided to base the estimation of local stocking densities not on the explanatory variables per HSMU, but rather on a distance and area weighted average of the area around each pixel cluster.

Those locally weighted averages per HMSU were then used to estimate the expected mean and its forecast error for each animal category, and livestock unit aggregates for ruminants, non-ruminants and all types of animals, providing a priori distribution for the stocking densities per HSMU. A HPD estimator chooses then those combinations of stocking densities per HSMU which exhaust the regional herd sizes. During estimation, bounds prevent the generation of very large stocking densities. In order to stabilize the results, the estimation included also the mentioned aggregates for ruminants, nonruminants and all type of animals expressed in livestock units. The resulting data set was evaluated against out-of-sample from France showing stocking densities for 35.000 single communes based on the FSS. The comparison revealed that the estimation was doing significantly better compared to a solution assigning average regional stocking densities per fodder area for ruminants and average stocking densities per ha for the non-ruminants (Leip et.al. 2007). The stocking densities

allow is also to include economic performance indicators for animal activities in the calculation at sub-regional level.

Organic and mineral fertilizer application rates are a highly relevant factor for environmental impacts of agricultural production as they dive realized crop yields and nutrient surpluses, and consequently the whole nutrient and carbon cycle in agriculture. Unfortunately, even at Member State level, data on typical organic and inorganic fertiliser application rates for crops are not available from harmonized European statistics. However, the International Fertilizer Manufacturer Association (IFMA) kindly agreed to let the project team access the results of expert surveys on inorganic application rates for crops or group of crops at Member State level. Those data are used in the process of building the regional data base of CAPRI to define regional fertilizer application rates per crop, taking into account regional yields, manure availability, average regional soil parameters and emission factors lined up with the MITERRA and RAINS models (Oenema et.al. 2007). At sub-regional level, the organic and inorganic application rates per crop are defined as to recover in average the ones at regional level. Firstly, organic application rates per crop and HSMU are estimated by increasing and decreasing the organic application rate for the crop at regional level depending on two factors. The first factor is the estimated local crop nutrient uptake in relation to the regional one, derived from the crop yield. Crop uptakes are derived from yields. A second factor increase or decreases the rate according to the estimated organic nutrient availability derived from stocking densities and manure excretion coefficients. Here again, as in the case of the estimation of the stocking densities, distance and size weighted averages of the organic nutrient availability around the HSMU are used rather than spot observations. The resulting estimated organic application rates per crop are then scaled in order to recover as the area weighted mean the given regional rate per crop. In a similar manner, inputs from crop residues, biological fixation and atmospheric deposition are calculated. Finally, the estimated mineral rate are based on the difference between the crop nutrient need and all non mineral sources, corrected by typical loss rate, and a factor based on soil properties. Those estimates per crop are then again scaled to deliver in average the regional mineral application rates.

8.3 Linkage to process-based modelling (DNDC) (A. Leip and G. Mulligan)

8.3.1 Introduction

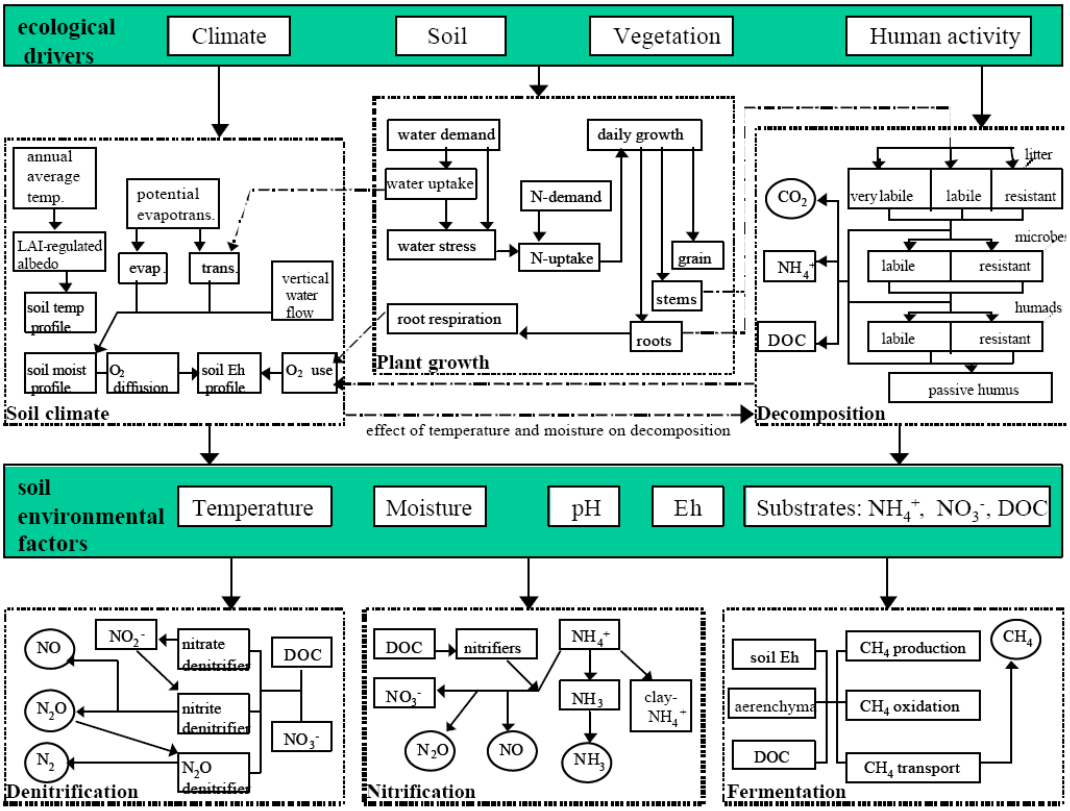
Process-based models are adequate to analyze the impact of changing farming practices, as they are able to cope with the complex interplay of environment and anthropogenic activities. But the accuracy of simulated fluxes with process-based models such as DNDC (Denitrification Decomposition) Model (Li et al., 1992) is largely dependent on the quality of input data. DNDC showed to be especially sensitive to the soil organic matter (SOM) content of the soils and to nitrogen application rates. If no *a priori* information is available, the range of calculated fluxes is determined by the range of SOM occurring in the region, for which statistical information is available. Uncertainties by a factor of 10 or more are common (Mulligan, 2004). Therefore, we paid special attention not only to the link of the CAPRI core model to the process-based model, but also to environmental datasets.

The model chosen to estimate GHG fluxes from agricultural soils was the DNDC model (Denitrification and Decomposition) (Li et al., 1992). DNDC has been developed in 1992 and since then improved continuously (Li, 2000; Li et al., 1992; Li et al., 2006; Li et al., 2004). DNDC is a biogeochemistry model for agro-ecosystems that can be applied both at the plot-scale and at the regional scale. It consists of two components, the first calculating the state of the soil-plant system such as soil chemical and physical status, vegetation growth and organic carbon mineralization, based on environmental and anthropogenic drivers (daily weather, soil properties, farm management) (see Figure 43). The second component uses the information on the soil environment to calculate the major processes involved in the exchange of greenhouse gases with the atmosphere, i. e., nitrification, denitrification, and fermentation.

The model thus is able to track production, consumption and emission of carbon and nitrogen oxides, ammonia, and methane. Main model output for the present purpose is a prediction of NO, N_2O, N_2, CH_4 and NH_3 fluxes as well as nitrogen uptake by the plants and nitrogen losses by nitrate leaching. The model has been tested against numerous field data sets of nitrous oxide (N2O) emissions and soil carbon dynamics (Li et al., 2005).

DNDC has been widely used also for regional modelling studies, under other in the United States of America (e. g., Tonitto et al., 2007), China (Li et al., 2006; Xu-Ri et al., 2003), India (Pathak et al., 2005), and Europe (e. g., Brown et al., 2002; Butterbach-Bahl et al., 2004; Mulligan, 2006; Neufeldt et al., 2006; Sleutel et al., 2006). Our simulations are done using DNDC V.89, however introducing several modifications allowing a more flexible simulation of a large number of pixel-cluster. These modifications enabled us to simulate an un-limited number of agricultural spatial modelling units with individual farm and crop parameterization and with the option to individually select up to 10 different crops to be simulated in a specific calculation unit.

Figure 43. Structure of the DNDC model

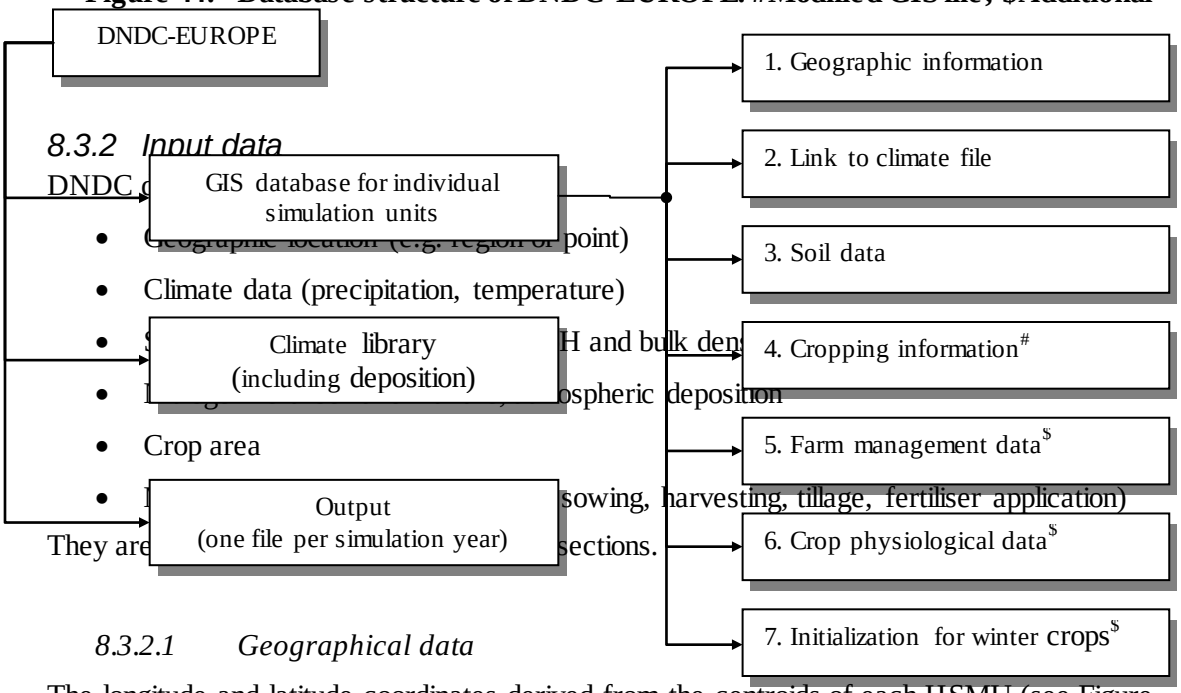


The regional mode of DNDC model uses a pre-defined database containing spatially referenced data. The database is not directly integrated with a GIS, therefore model input data must be previously processed in a GIS and the data imported into the required model database structure. Likewise outputs from the model are provided in text form and must be linked to a GIS for spatial analysis.

Using the default version it was not possible to accommodate the degree of flexibility that was required in our study. Necessary adaptations, however, regarded purely data handling. Scientifically, our study was using the identical approaches as (Li et al., 2004). First, it was necessary to allow for each modelling unit an individual number and selection of crops that are simulated; second, farm data such as fertilizer application rates are calculated individually for each simulation unit. In the default version of DNDC, the farm library is constant at

province level. Third, potential yield is determined for each modelling unit; in the default version of DNDC the crop libraries are constant at national level. Last, for easier post-processing of the data, output files were grouped into single tables for each simulation year.

Figure 44. Database structure of DNDC-EUROPE. #Modified GIS file; $Additional



8.3.2 Input data

DNDC c

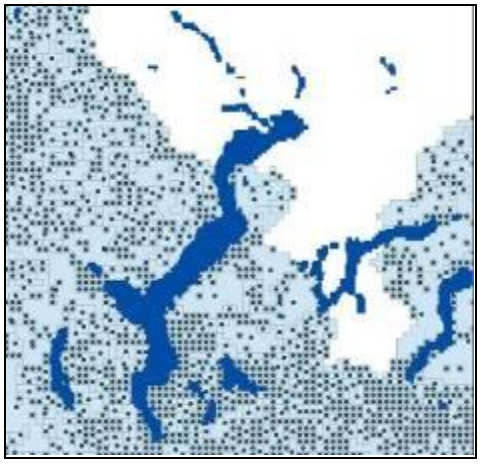
- Geographic location (e.g. region or point)
- Climate data (precipitation, temperature)
- S H and bulk den
- l ospheric deposition
- Crop area
- I sowing, harvesting, tillage, fertiliser application)

They are sections.

8.3.2.1 *Geographical data*

The longitude and latitude coordinates derived from the centroids of each HSMU (see Figure 45) are used within the DNDC model to drive the day length function of the crop growth model.

Figure 45. HSMU centroids used to determine geographic location



The geographic file in the DNDC database provides the link, via the HSMU ID to all the other parameters needed to run the DNDC model at the regional scale (see Table 33).

Table 33 Geographic information data required by DNDC.

Grid Characteristics		
1	HSMU ID	*1001*

2	Name *	*IT201*
3	Region *	*Varese*
4	Longitude	*8.764*
5	Latitude	*45.734*

8.3.2.2 Weather data

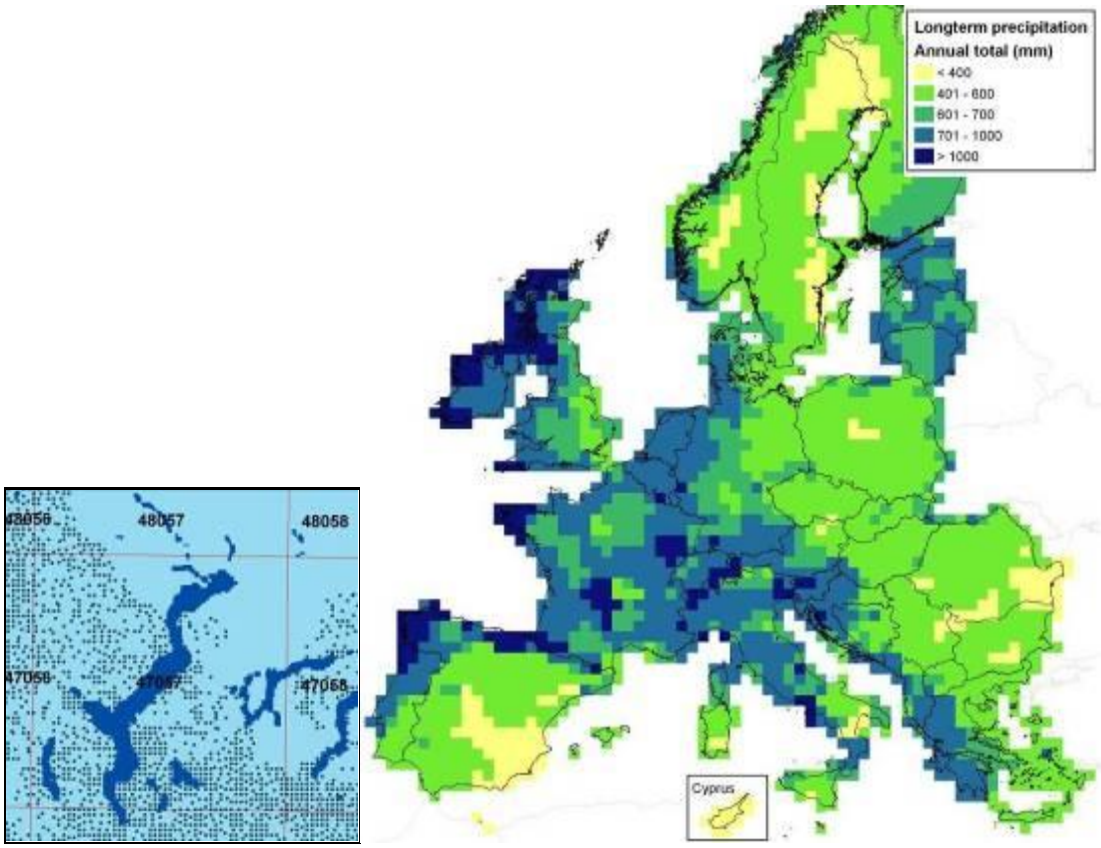
The climate file in the DNDC GIS database contains the atmospheric nitrogen deposition data and provides the link between the modelled (HSMU) units and the meteorological text files containing the daily meteorological data (Table 34)

Table 34 Climate information required by DNDC.

Climate file		
1	HSMU ID	*1001*
2	Climate file (MARS ID)	***45055***
3	N concentration	*0.95*
Climate library file		
1	Climate file (MARS ID)	***45055***
2	Julian day	*1 - 365*
3	Max temp	*4.1*
4	Min temp	*2.3*
5	Rainfall (cm)	*0.5*

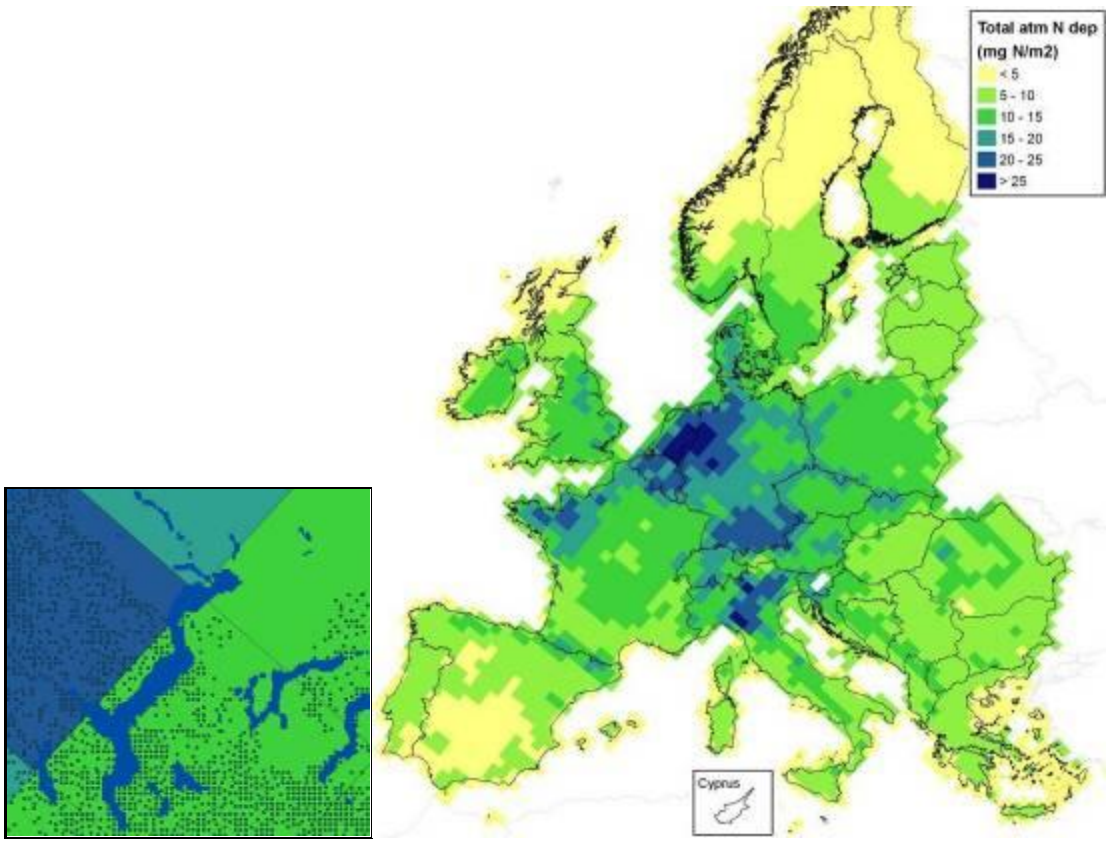
Daily meteorological data for 2000 has been extracted from the Monitoring Agriculture and Regional Information Systems (MARS) database containing data interpolated on a 50 km x 50 km grid (Figure 46) (Orlandi and Van der Goot, 2003). The DNDC model requires minimum and maximum temperature (°C) and precipitation data (cm).

Figure 46. Link between HSMU and meteorological grid (50 km x 50 km)



Annual N (dissolved nitrate and ammonium) concentration in rainfall (mg N/l or ppm) was derived from The Co-operative Programme for the Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe (EMEP) Precipitation Chemistry Database (EMEP, 2001). The data are reported as precipitation weighted arithmetic mean values in mg N L^{-1} as ammonium and nitrate measured at one of the permanent EMEP stations. A European coverage of the data was achieved in representing each measurement station by a theissen polygon and spatially relating the polygon to the spatial calculation units (Mulligan, 2006).

Figure 47. Link between HSMU and N deposition grid (EMEP 50 km x 50 km)



8.3.2.3 Soil data

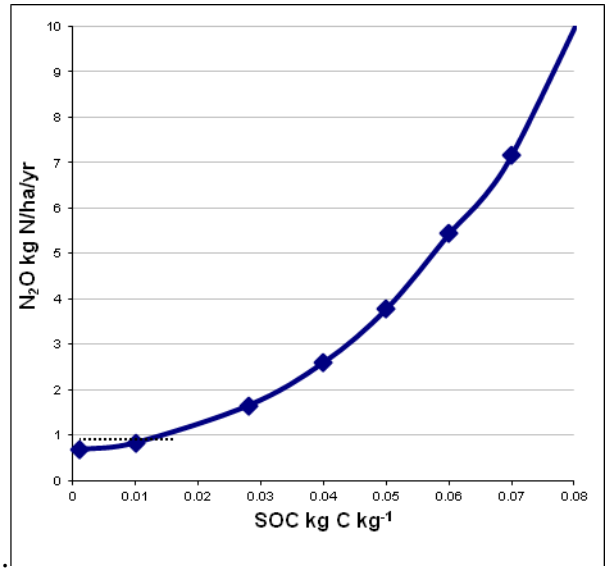
Pan European soil data are available from the Soil and Waste Unit of European Commission's JRC through the activities of the European Soil Bureau Network (ESBN)[73].

The ESBN have created a series of 1 km x 1 km soil rasters including topsoil organic carbon content that have been calculated using a refined pedo-transfer rule derived from the European Soil Database, an extended CORINE land cover dataset), a digital elevation model (DEM) and mean annual temperature data (Jones et al., 2005). Additional 1 km x 1 km topsoil and subsoil rasters provided by the ESBN include clay (content %), base saturation (%), and packing density (g cm^{-3}) (Hiederer et al., 2003). The DNDC model requires initial content of total soil organic carbon data (SOC) in kg C kg^{-1} of soil including litter residue, microbes, humads and passive humus in the topsoil layer (0-5 cm).

Sensitivity analysis has shown that the DNDC model is very sensitive to SOC (Figure 48) (Mulligan, 2006)

[73] Distribution version 2.0, http://eusoils.jrc.it/ESDB_Archive/ESDBv2/fr_intro.htm

Figure 48. DNDC model sensitivity to SOC



Additional 1 km x 1 km topsoil and subsoil rasters provided by the ESBN include:

- Clay (content %)
- Base saturation (%)
- Packing density (g/cm^3)

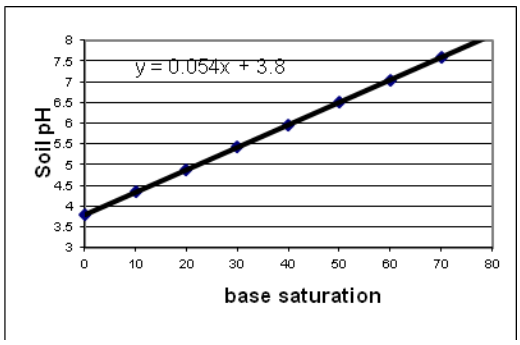
For this project bulk density data (g/cm^3) were derived from packing density (*pers comm.* Jones, 2003) using the pedo-transfer function:

Equation 188 $$\rho_b = P_D - (C_l \times 0.009)$$

Where ρ_b (kg/m^3) is the dry bulk density, P_D is the packing density (g/cm^3) and C_l = clay content (fraction)

Base saturation data, representing the fraction of CEC occupied by base cations, were used to derive soil pH. A linear relationship (Figure 49) between base saturation and soil pH was estimated based on expert knowledge from R.A. Jones (*pers comm.* 2003).

Figure 49. Linear relationship between soil pH and base saturation

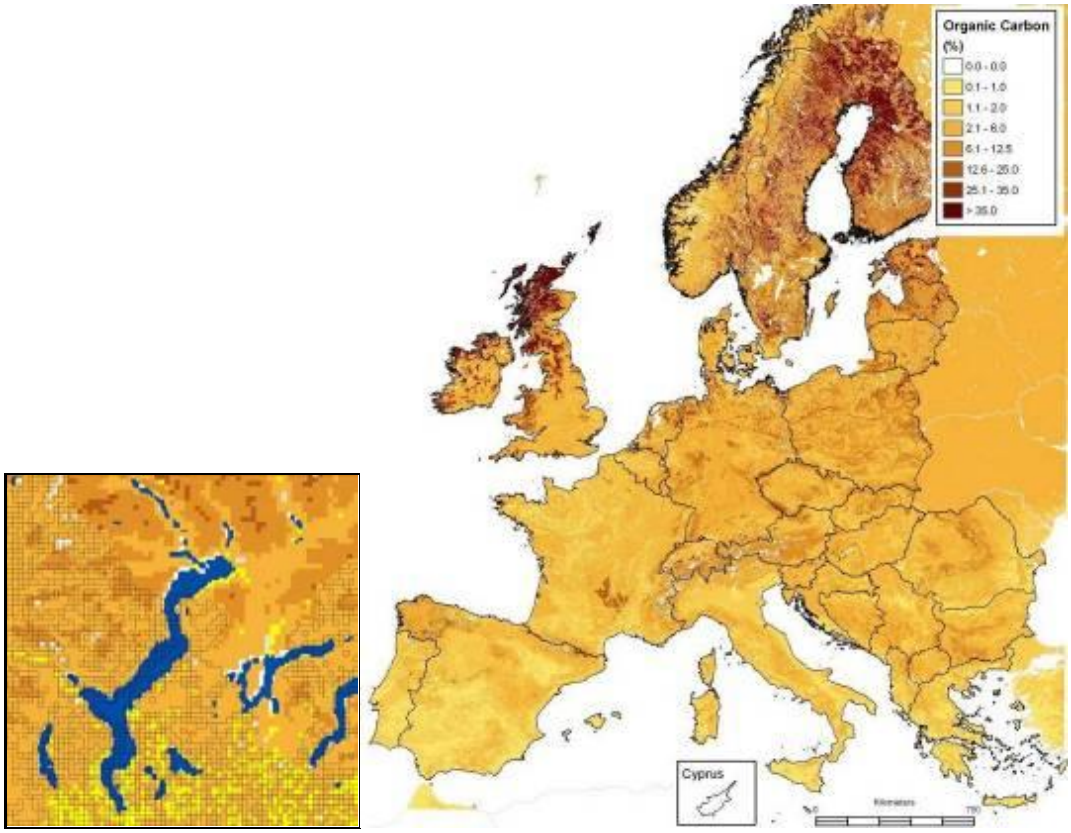


The soil parameters shown in 0 were calculated in ARC GIS zonal statistics using the HSMU shape to define the regions and soil rasters for the values to be summarised (see Figure 50)

Soil information required by the DNDC

Soil Properties		
1	HSMU ID	*45055*
2	SOC (min, max)	*0.01*
4	Clay (min, max)	*0.2*
6	pH (min, max)	*6*
8	BD (min, max)	*1.4*

Figure 50. HSMU shapefile used to summarise soil raster values.



8.3.2.4 Management Data

The farm file database structure of DNDC enables crop management data to be applied for each individual crop types within a chosen region. The farm file structure contains the following information for each modelled crop:

- Planting timing (month/day
- Harvest timing (month/day)
- Fertilisation timing (month/day)
- Fertilisation rate (kg/N/ha)
- Percent residue left
- Manure N rate
- Manure C:N
- Manure timing
- Flooding
- Irrigation

Crop acreages, crop yield and nitrogen application rates are estimated as described above. Crop sowing and harvesting dates are obtained from (Bouraoui and Aloe, 2007).

The percentage crop residue requirement for DNDC is defined by (Li, 2002) as the fraction of above-ground crop residue (leaves and stems) left as stubble or litter in the field. Crop residue

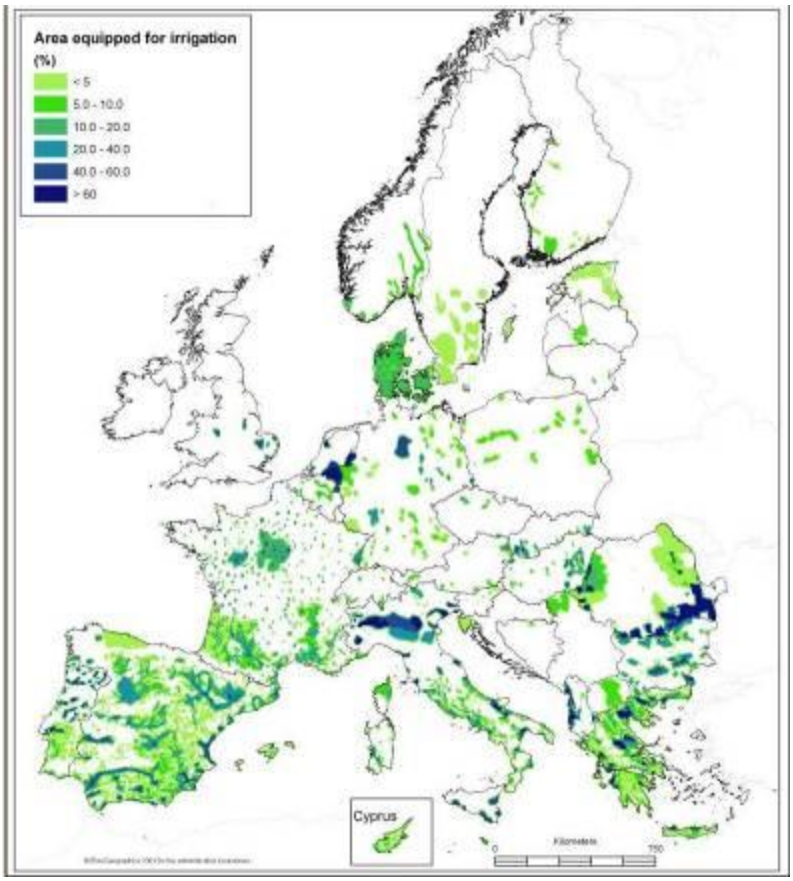
incorporation rates are available from IPPC guidelines for estimating N_2O emissions due to Nitrogen content in crop residue.

The DNDC model treats irrigation such that a calculated water deficit is re-plenished to a pre-defined percentage. Irrigated cultures do not suffer any water deficit, while non-irrigated cultivation will feel water-stress when water demand by the plants exceeds the water supply. Percentage of irrigated area was calculated on the basis of the map of irrigated areas (Siebert et al., 2005), and was taken as fixed for all crops being cultivated within an HSMU (see Figure 51).

Number and timing of fertilizer and tillage applications is taken from the DNDC farm library (Li et al., 2004) taking for good the dates relative to sowing or harvesting and applying these time lags to the actually simulated sowing or harvesting dates, respectively.

All other information needed to describe farm management and crop growth, such as tillage technique, maximum rooting depth and so on are taken from the DNDC default library and used as a constant for each crop for the whole of the simulated area.

Figure 51. Area equipped for irrigation (Siebert et al., 2005)



8.3.3 Model setup

The above-defined HSMU can be regarded as the smallest unit on which simulations can be carried out. This, however, is not always practical, as the high number of units is combined with a number of scenarios or if a multi-year simulation is carried out. Therefore, an intermediate step re-aggregates the HSMUs for each scenario that is simulated by the model, into model simulation units (MSUs) on the basis of both agronomic and environmental

criteria. In this way, the design of the scenario calculations can be best matched with the objectives of the study.

In the following we show exemplarily the procedure and the parameters used for re-aggregating the spatial calculation units (HSMU) to the spatial simulation units (MSU). The reader should however keep in mind that this refers to a possible solution, which was selected in an simulation exercise.

The objective of the exercise was to cover as much variability as possible in order to enable to assess the impact of the environment (represented in the model by daily weather data and soil parameters) and cultivation patterns. Therefore, for each region defined in the economic model (NUTS2), all crops that cover at least 5% of the agricultural area are included in the model. These crops were simulated on MSUs that had a crop share of more than 35% of the agricultural area within an agricultural unit (defined by a minimum of 40% of the area used for agriculture) or the crop share was at least 85% of the maximum share of the crop occurring in the region. Before eliminating single units, however, all units were clustered according to their similarity in the environmental conditions. To this purpose, a tolerance is defined for each parameter that gives the maximum spread allowed within a single cluster. For example topsoil organic matter content was clustered if the values differed less than ±10%. The thresholds and tolerances used in this study are listed in Table 35. These moderate tolerances for soil conditions lead to an average number of more than 68 (up to 266) different soil conditions that were distinguished in each region, with add to 11,438 environmental situations for EU-15, out of which 6,391 MSU were simulated with a total of 11,063 crop-MSU combinations. Each of these simulations runs over 99 years to smooth out unrealistic estimates for topsoil organic carbon.

Table 35 Thresholds and tolerances used to cluster HSMU into MSU and to select the simulated crops

Parameter	Explanation	Value
MINUAAR	Minimum UAAR in a MSU for simulation	0.40
MINSHAR	Minimum share of crop in UAAR of the MSU	0.35
MINPLUS	Minimum share of crop in UAAR not yet considered	0.85
MINMINS	Limitation share to add more crops if not relevant in region	0.05
M-ID	Tolerance for daily weather condition (file-number)	0.05
NDEP	Tolerance for N-deposition values [mg N / ml rain-water]	0.05
OC_MAX	Tolerance for soil organic carbon content	0.10
CL_MAX	Tolerance for clay content	0.20
PH_MAX	Tolerance for topsoil pH	0.20
BD_MAX	Tolerance for topsoil bulk density	0.20

8.4 Landscape indicators (M.L. Paracchini)

The implementation of the spatial layer and the availability of disaggregated information, on crop shares but also relative to other parameters (nitrogen input and surplus, livestock density, premiums etc.), adds relevant potential for improved and new landscape and environmental indicator calculations, and opens wide possibilities for ex-ante impact assessment of the CAP at the regional level when scenarios are taken into consideration.

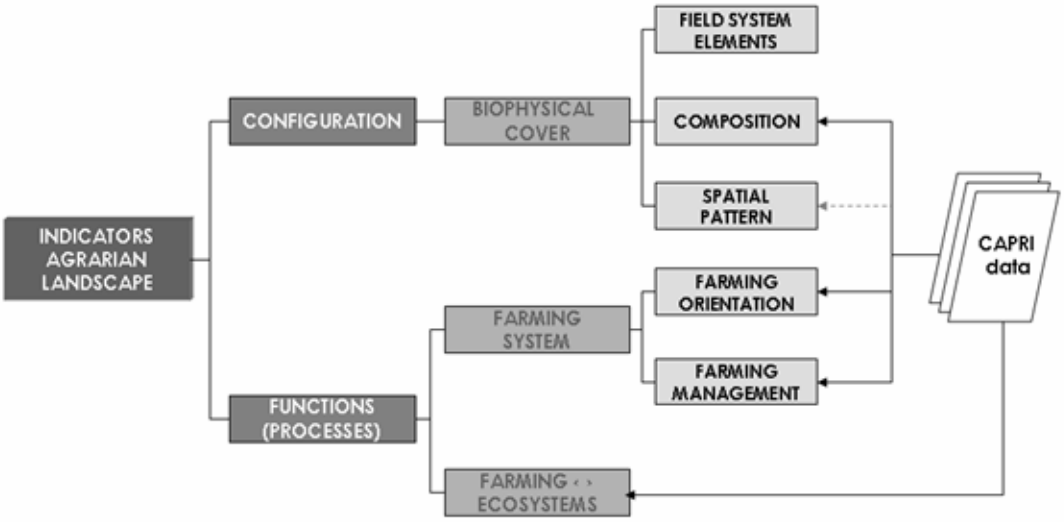
The added value of the new data layers is given by the fact that the information they provide was traditionally available at NUTS2 level; at such scale it is not possible to carry out assessments of the impact of the agricultural policy on the landscape, due to the large size of the administrative regions. Furthermore, from a conceptual point of view frameworks of indicators have been drawn in the past years and also recently (e.g. ELISA, PAIS, OECD, IRENA), that define agri-environmental indicators for sustainable agriculture; but often some of these, though considered of high relevance, could not be calculated because of lack of data with an appropriate level of detail.

8.4.1 Background on indicators for agrarian landscapes

The selection of landscape indicators to be implemented in the CAPRI model is the result of a review of journal articles and reports on the topic. An overview of the issues covered by these indicators review is presented in Figure 52. A first group of indicators is identified by studies which refer to aspects linked to spatial configuration of land use/cover, and is based on information on the biophysical cover. From them, few studies target elements of the field system elements and many address land cover composition and spatial pattern characteristics. A second group of indicators is related to the farming system (farming orientation, farming management); a third group is bridging the interaction between farming and ecosystems, and provides essential information on the influence of farming in agrarian landscapes.

Indicators to be implemented in the CAPRI frame were selected according to their relevance for agrarian landscape, to their frequency in literature, and their potential implementation with CAPRI data.

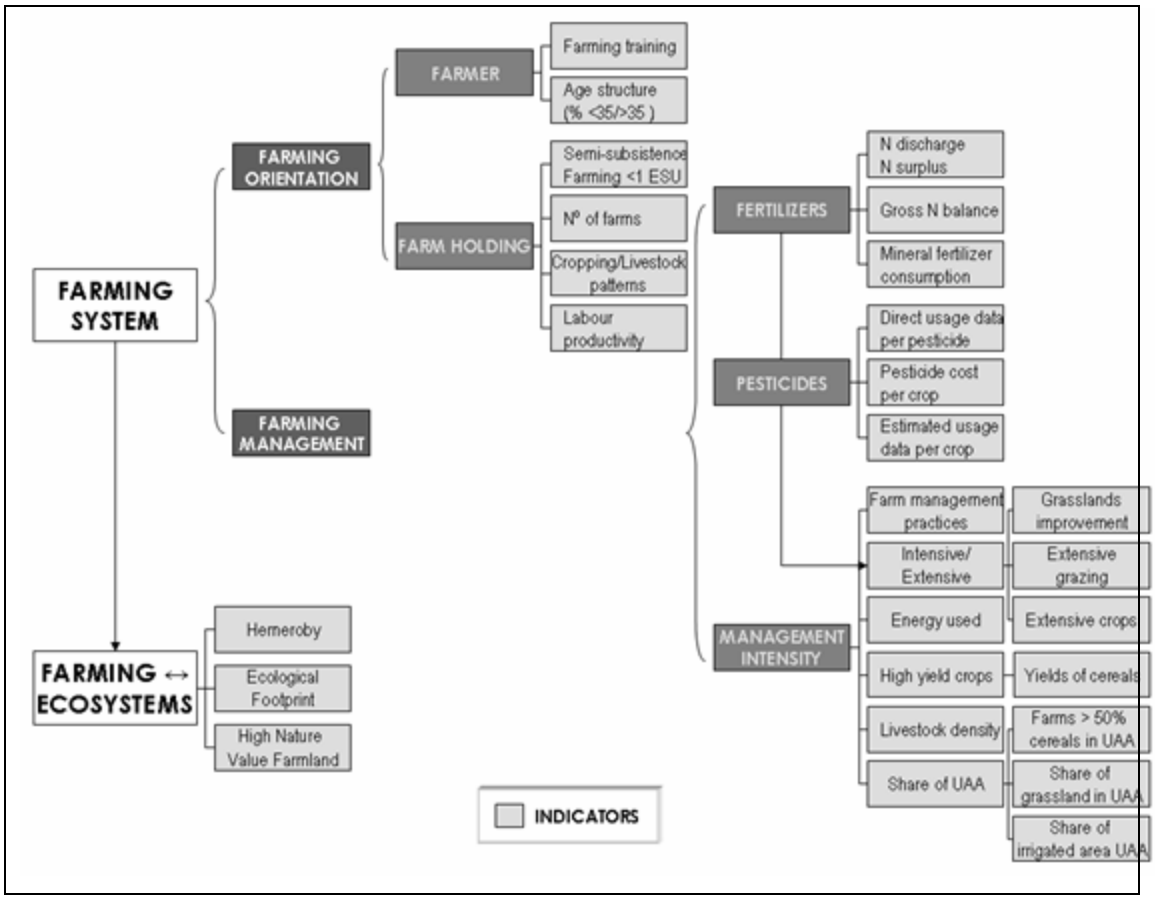
Figure 52. Overview of issues addressed by indicators in relation to the agrarian landscape and their potential implementation with CAPRI disaggregated data



Taking into account the characteristics of CAPRI spatial disaggregated data, the indicators more suitable to be implemented are those related to configuration-composition and function indicators (see Figure 52). Most commonly used composition indicators are crop diversity, crop distribution and the degree of openness; all of them can be easily retrieved from the cropping shares provided by CAPRI. Regarding function indicators, the spatial dissagregation of agricultural statistical data on farming orientation and farming management provides the opportunity to analyse -in combination with other indicators - the relationship between farming activity and the agrarian landscape. As seen in Figure 53 the implementation of the indicator on intensive/extensive use is linked to other proposed management indicators (grassland improvement, extensive grazing and extensive grazing) and to indicators

measuring the usage of fertilizers and pesticides. Although crop shares are already considered as composition indicators, other four indicators based on share of UAA have been proposed, that provide information on management intensity (i.e. share of grassland in UAA). Farming system spatial information could contribute, as well, to the improvement of delineation for High Nature Value Farmland areas, and to the implementation of indicators related to the impact of human activity on the landscape (ecological footprint, human appropriation of net primary production).

Figure 53. Overview of indicators related to the farming system and to the interaction between farming and ecosystems



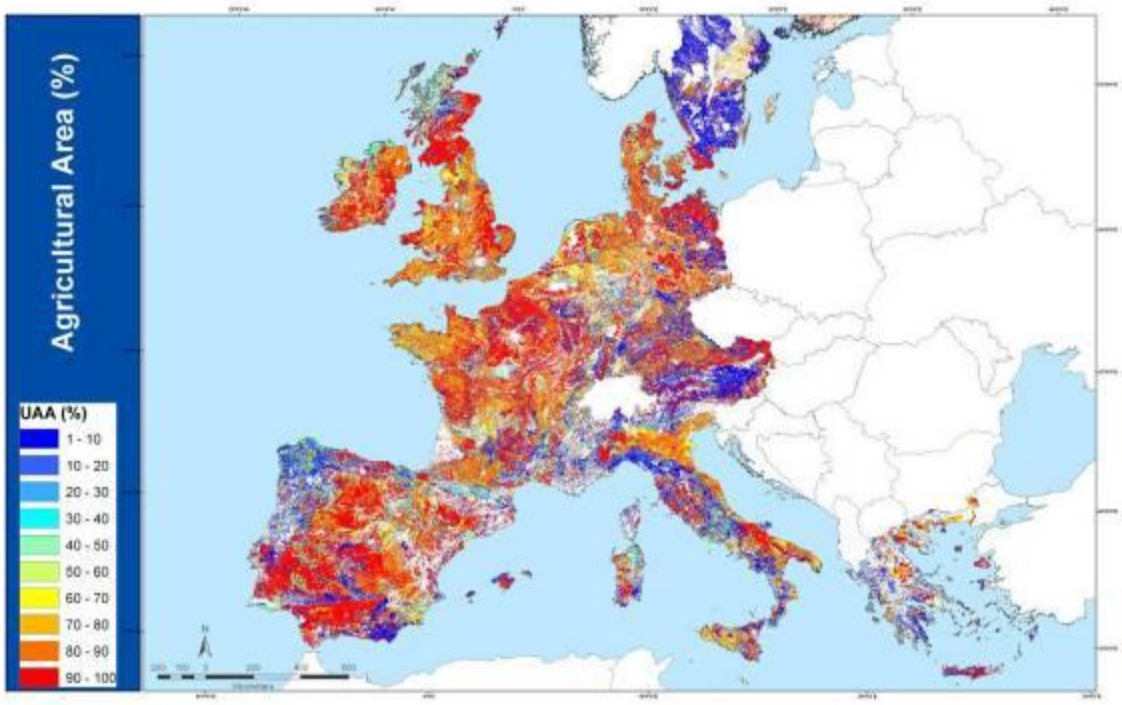
For completeness of information the field system indicators are also reported in fig.1, but the spatial resolution of 1 km and the information on biophysical cover expressed mainly as crop shares makes it not suitable for the retrieval of information regarding linear features or the historical parcel pattern. Furthermore, the use of CAPRI data for the calculation of spatial pattern indicators is also constrained by these characteristics. Indeed many of the metrics proposed as indicators are based on categorical land cover data, so the use of crop shares makes only advisable the calculation of metrics based as well in shares like diversity indices. Therefore, their interpretation remains constrained by the coarse spatial resolution, so given a certain extent low values of diversity are indicating a more homogenous landscape at the scale of observation but there is the possibility that coarse spatial resolution could mask the presence of heterogeneity only observable at more detailed level.

8.4.2 Indicators on configuration

Within this group of indicators the focus is set on different ways to assess how the variety and diversity of crops characterises European agricultural landscapes. This information is relevant since higher varieties of species and genotypes positively influence biodiversity and stability of agro-ecosystems.

A first –basic- indicator is represented by the share of Utilised Agricultural Area (UAA) in the total area of the cell (1 sqkm), obtained by summing up the shares of each single crop. For many indicators this value is used as reference area (i.e. the indicators are calculated on the UAA and not on the total area of the cell).

Figure 54. Share of Utilised Agricultural Area



An interesting and simple information that can be derived from the CAPRI dataset is the share of the prevailing crop in % of UAA (fig.4), high values are indicative of a prevailing crop or land-use, like e.g. the large grazing areas of Ireland and Scotland. This information is complemented by the map on the prevalent crop type (fig.5) that allows to easily identify i.e. the large cereal areas in Europe.

Figure 55. Prevailing crop with reference to the UAA

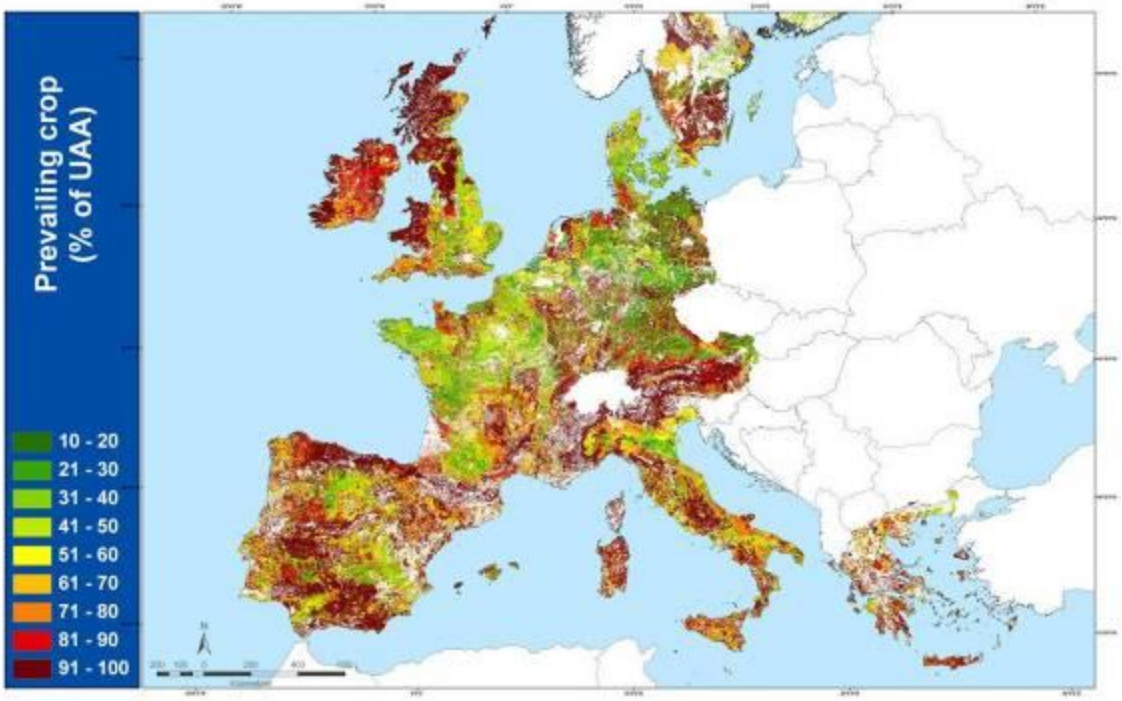
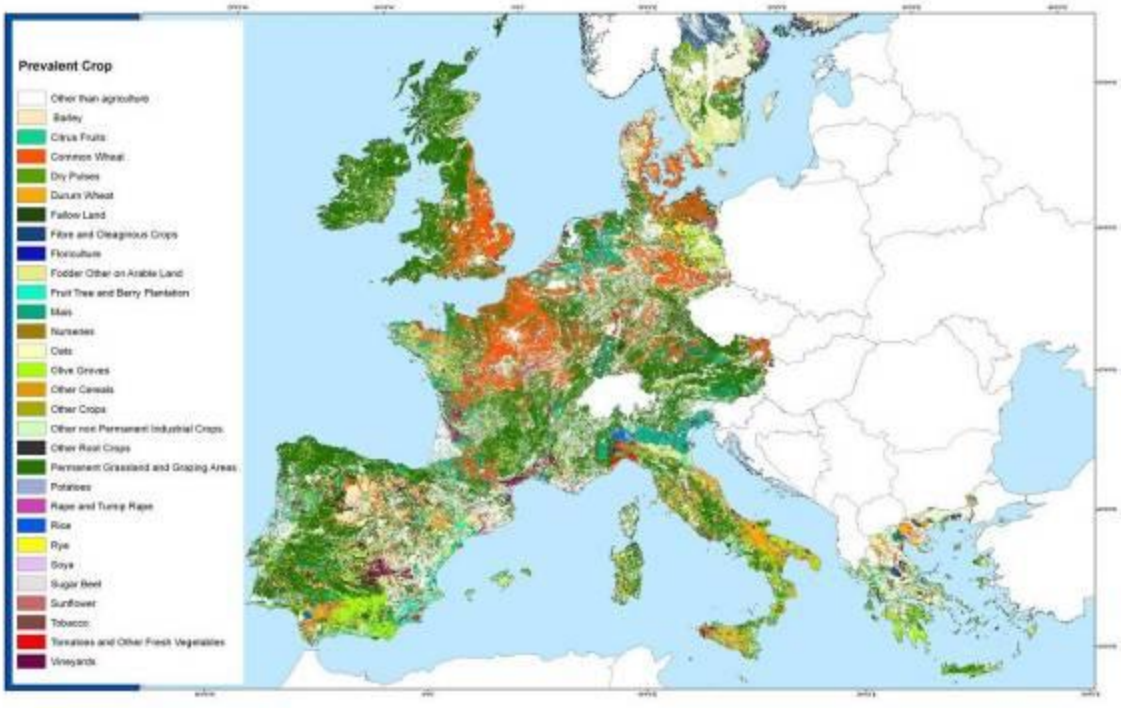


Figure 56. Prevalent crop



Using the same approach targeted indicators can be obtained on specific crop typologies that are particularly relevant both from an economic and environmental point of view, like cereals and fallow land (fig.6 and 7)

Figure 57. % of cereals in UAA

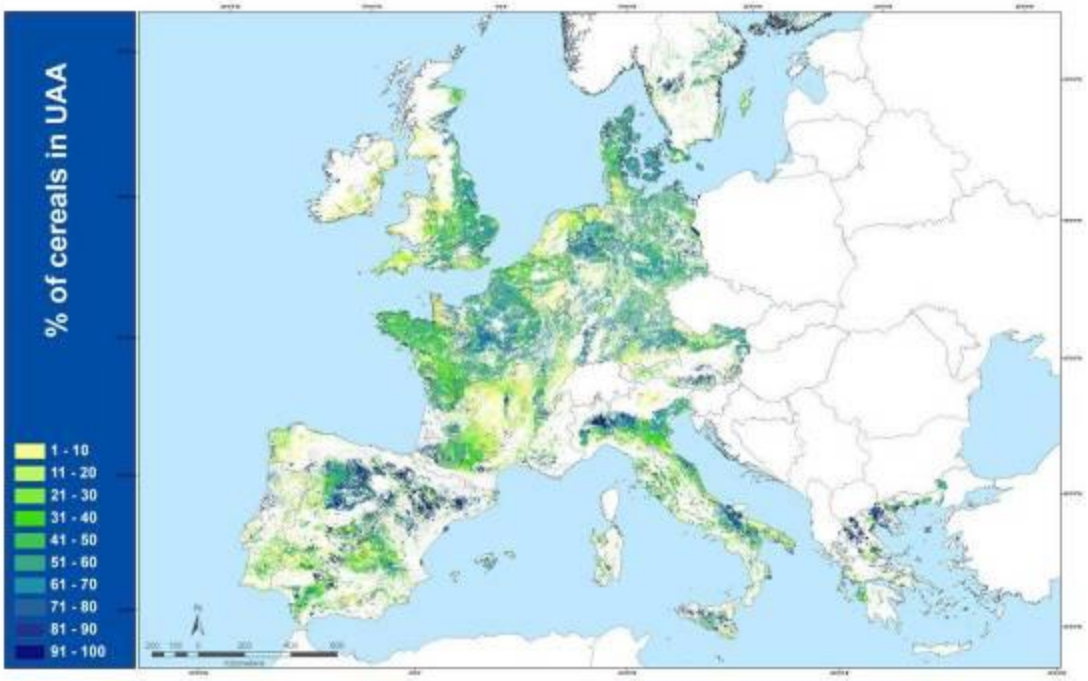
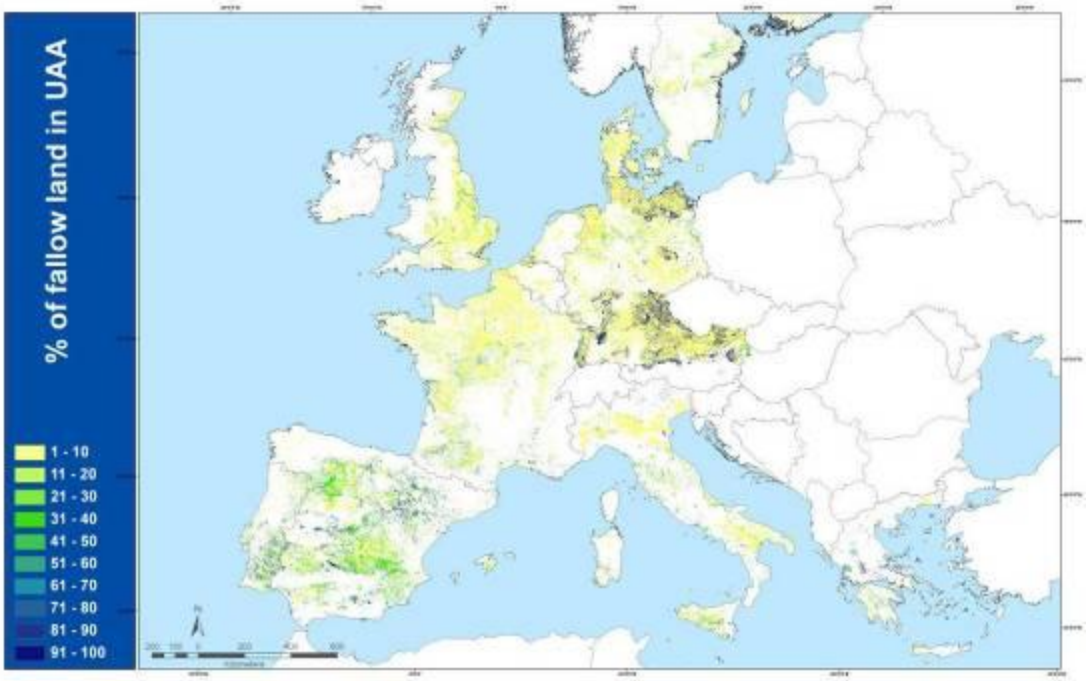


Figure 58. % of fallow land in UAA

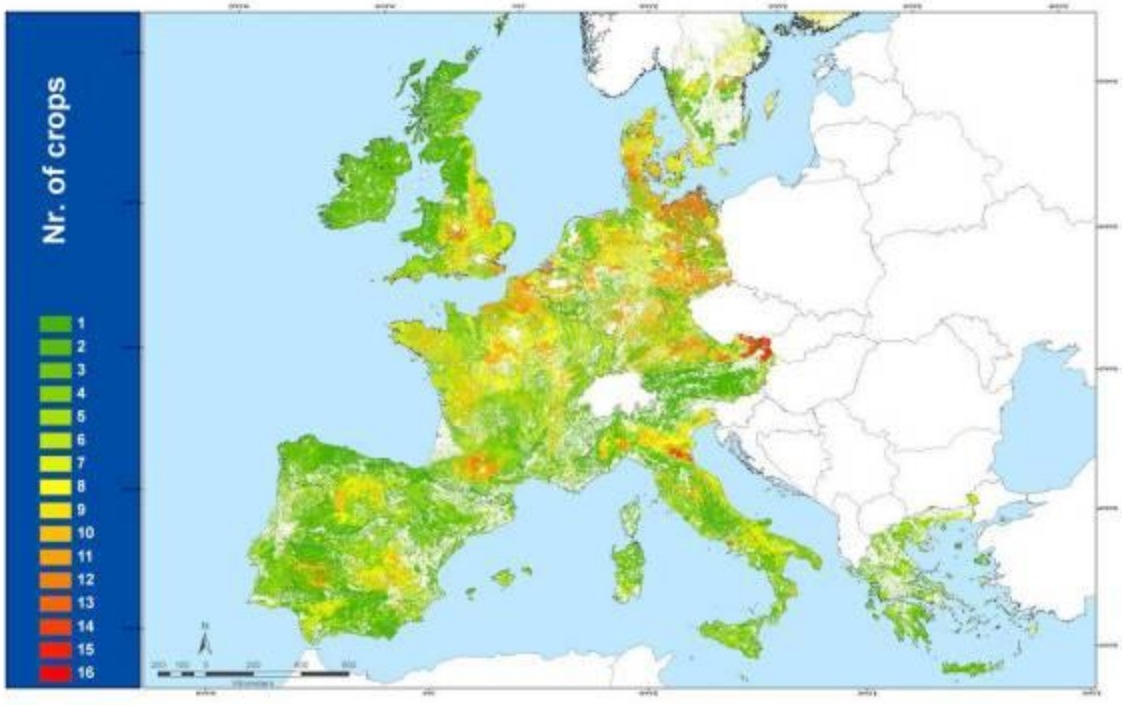


Crop diversity

The diversity of the agricultural landscape is strictly related to the number of crops. A higher variety corresponds to different cultivation methods, that is an indirect indicator of the presence of a multitude of habitats for species (H.-P. Piorr et al., 2005).

There are different ways to calculate crop diversity. A simple indicator is the number of crops per reference unit, that in this case is calculated taking into account the presence of each of the crops listed in tab.1, when they occur at least in 1% of the total reference area:

Figure 59. Crop diversity (see text for explanation)



A more elaborated index is Simpson's diversity index:

Equation 189 $$SIDI = 1 - \sum_{i=1}^{m} P_i^2$$

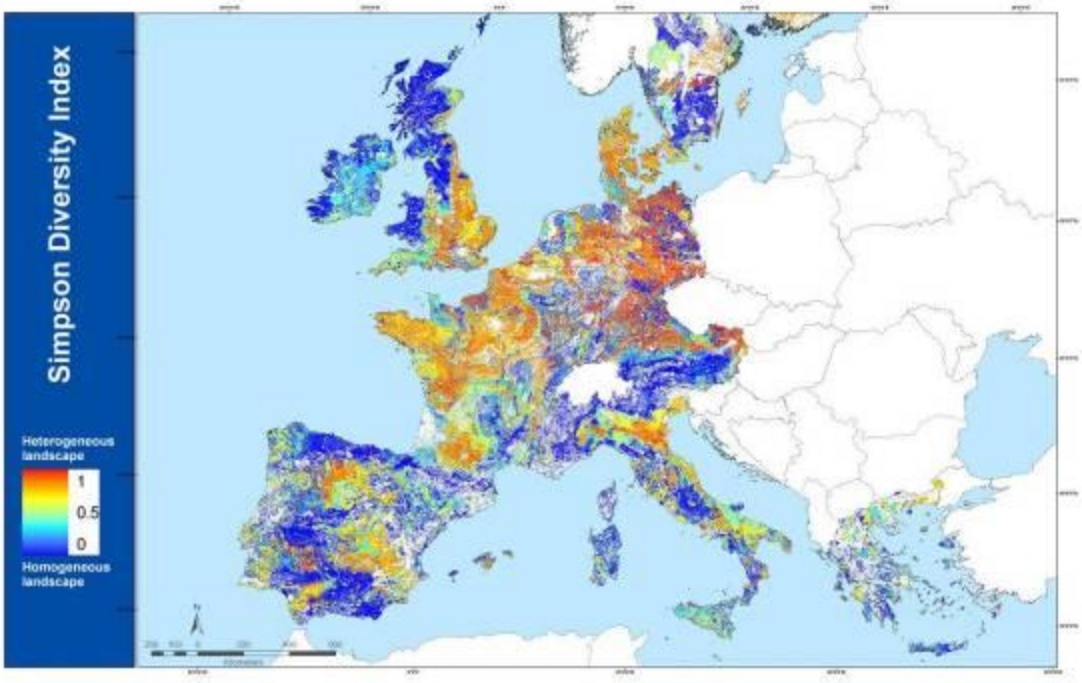
$$SIDI = 1 - \sum_{i=1}^{m} P_i^2$$

P_i = proportion of the landscape occupied by patch type (class) i, based on total landscape area (A) excluding any internal background present.

SIDI = 0 when the landscape contains only 1 patch (i.e., no diversity). SIDI approaches 1 as the number of different patch types (i.e., patch richness, PR) increases and the proportional distribution of area among patch types becomes more equitable.

The value of Simpson's index represents the probability that any 2 points selected at random would be different patch types (FRAGSTATS metrics manual http://www.umass.edu/landeco/research/fragstats/documents/fragstats_documents.html)

Figure 60. Simpson's Diversity Index



In fig.9 areas characterised by a low score of the index are covered by one or two crop types, this can be indicative of a monoculture (e.g. the rice fields in NW Italy) or the presence in high percentage of the UAA of permanent grass and grazing, that explains the low score in the mountainous areas, in Ireland and Scotland.

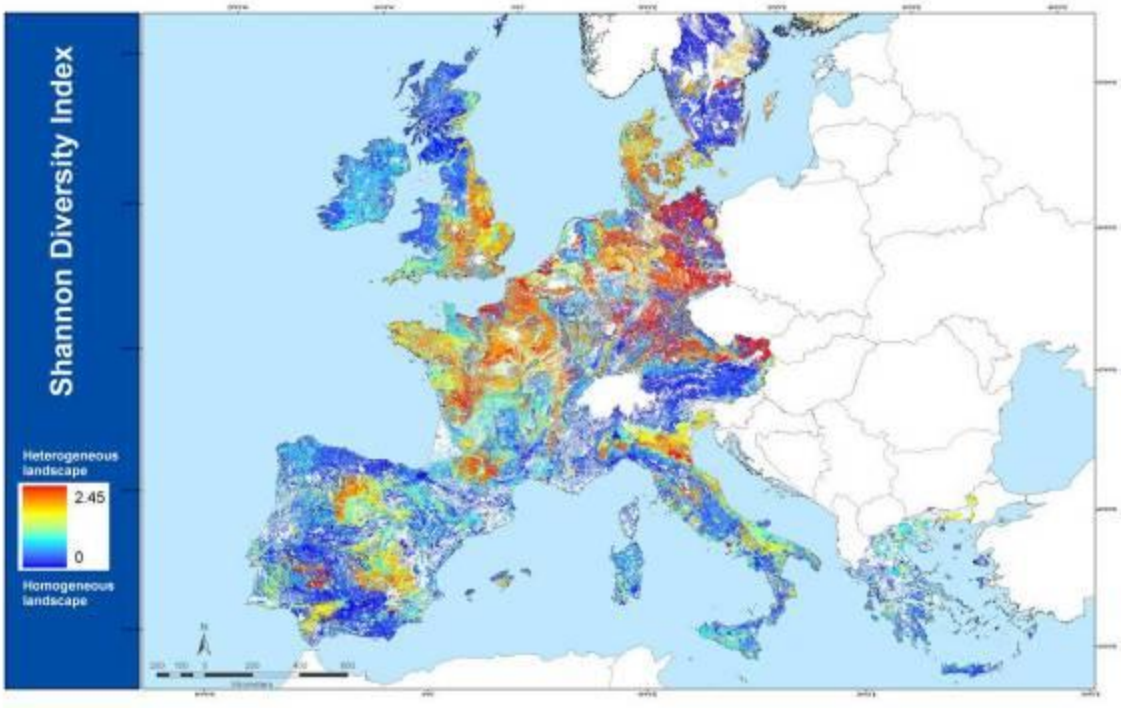
A third, well known but still discussed index is Shannon's Diversity Index

Equation 190 $$SHDI = -\sum_{i=1}^{m}(P_i * \ln P_i)$$

P_i = proportion of the landscape occupied by patch type (class) i, based on total landscape area (A) excluding any internal background present.

The behaviour of this index is similar to Simpson's, but it is more sensitive to rare patch types.

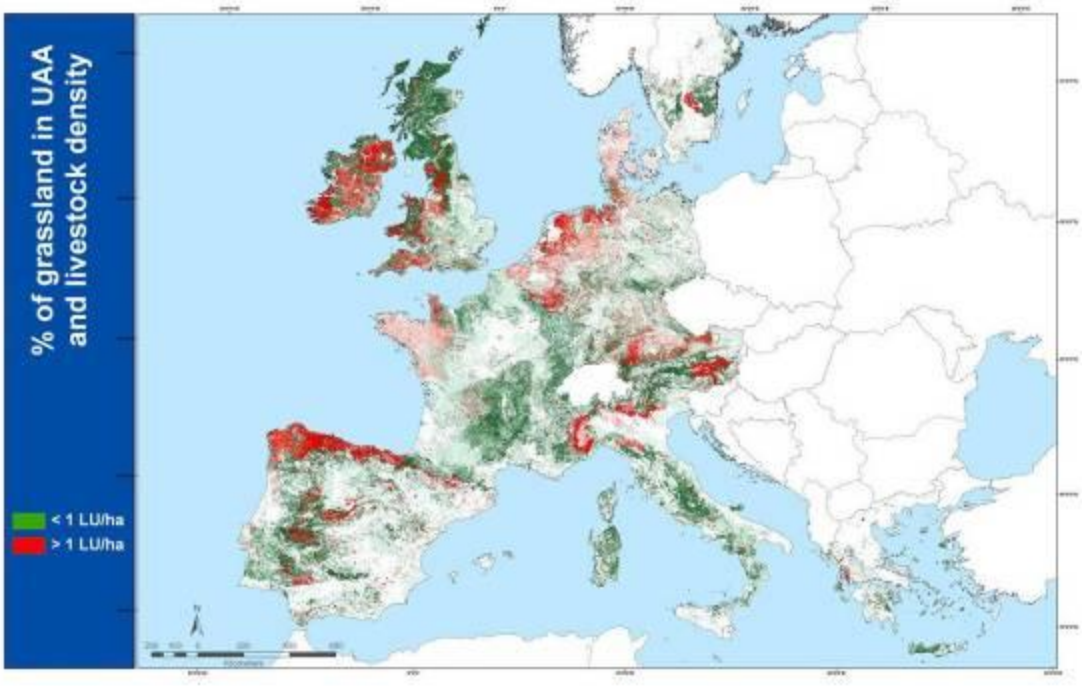
Figure 61. Shannon Diversity Index



8.4.3 Indicators on farming orientation/management intensity

The availability of disaggregated data on inputs offers a substantial improvement in the calculation of the indicators since they provide information on farming orientation. Such data had been available shortly before the time of writing the present report, therefore the example presented illustrates one of the possibilities offered by the new set of data, such as coupling the information on share of grasslands with the information on livestock density. In fig.11 an indicative threshold of 1 LU/ha has been selected.

Figure 62. % of grassland in UAA in areas with different livestock pressure



8.4.4 Indicators on farming and ecosystems

In order to map the influence of farming in agrarian landscapes the information on intensity of management is essential. Given its availability a pilot study was carried out to simulate the presence in Europe of areas of High Nature Value Farmland (HNVF), based solely on CAPRI data.

HNVF is defined as "*areas in Europe where agriculture is a major (usually the dominant) land use and where that agriculture supports, or is associated with, either a high species and habitat diversity or the presence of species of European conservation concern, or both*". Mapping HNVF requires a good availability of land cover data and biodiversity relevees (i.e. semi-natural grasslands surveys), and information on farming systems that are likely to maintain areas of high natural value.

A study carried out using French national data (Biala et. al 2007) demonstrated that the availability of statistics at municipal level concerning crop types, levels of input and presence of linear elements is sufficient to provide a good approximation of HNVF areas at the Country level.

In the specific case the input data were the French Farm Structure Survey (original data per farm aggregated at NUTS5), FSS 2000 "specific regional questions" (Traditional orchards), Agricultural Annual Survey 2000 (Common land), National Forest Survey (IFN) (Forest borders and hedges), Grassland survey (Grassland management of productive grasslands), Regional data (Traditional orchards).

The mapping of HNVF was carried out through the definition of three sub-indicators on: crop rotation and presence of grasslands, intensity of management of grasslands and crops, presence of linear elements. The CAPRI data do not contain information on linear elements, therefore the simulation is carried out on the basis of crop information and levels of input.

The first sub-indicator (crop diversity and share of grassland) is a proxy for the rotation system, and allows a first approach to the diversity of landscape. Longer rotations are indicative of less intensive agriculture and allow a reduction of pesticide use. It assumes that

when there is a high crop diversity and/or a high share of grassland there are favourable conditions for biodiversity.

The score is calculated for each farm (660 000 farms in France) with a weighting (taking into account the UAA surface of the farm) at the scale of the commune.

The equation is as follows:

Equation 191 $Index\ 1 = 10 + ((1-C1/UAA*10))+(1-(C2/UAA*10))+...$

Where C1 is a crop with a surface of more than 10% of UAA, other than temporary and permanent grasslands. The score ranges from 1 to 10.

The index can be simulated with CAPRI crop shares, trying to build similar legends, since the nomenclatures of input data are not exactly the same.

The comparison of the two classifications is illustrated in Table 36

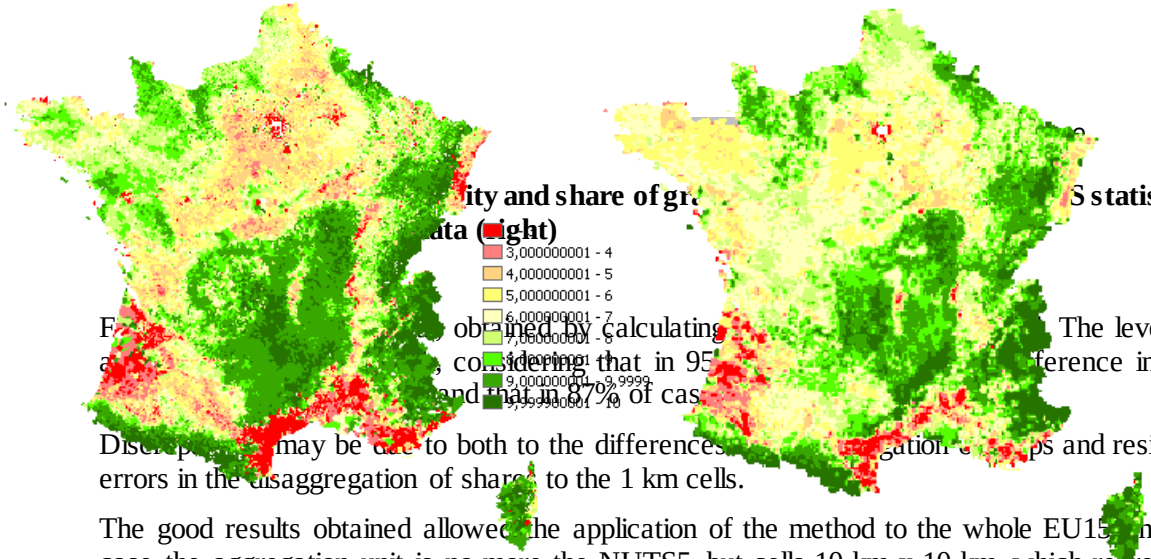
Table 36 FSS and CAPRI nomenclatures

FSS 2000 - France			CAPRI	
Common and durum wheat	*	*	dwhe	Durum Wheat
		*	swhe	Common wheat
Barley			barl	Barley
Maize for grain, maize for seeds and green maize			lmaiz	Maize
Oat			oats	Oat
Triticale			gras	Permanent grass and grazing
Rye			ryem	Rye
Sorghum			puls	Dry pulses
Other cereals			ocer	Other cereals
Sugar beet			sugb	Sugar beet
Rapeseed			lrape	Rape and turnip rape
Sun flower			sunf	Sunflower
Soja beans			soya	Soya
Other industrial plants	**	**	ltext	Fibre and oleaginous crops
		**	oind	Other non permanent industrial crops
		**	toba	Tobacco
Pea			ocro	Other crops
Broad bean			ofar	Fodder other on arable land
Other legumes and dry vegetables			pari	Rice
Other root crops			roof	Other root crops
Potatoes			pota	Potatoes
Fresh vegetables			ovto	Tomatoes and Other fresh Vegetables
Floriculture			flow	Floriculture
Vineyard			ltwin	Vineyards
Fruit production (apple tree, pear tree, plum tree, cherry tree, peach tree, apricot tree only)			lfrui	Fruit tree and berry plantations
Others fruit trees and nurseries	***	***	citr	Citrus fruits
		***	loliv	Olive groves
		***	nurs	Nurseries
Fallow land			lfall	Fallow land
Other annual forage				

same classes

grouped

similar



ity and share of gr S statistics
ata (right)

F obtained by calculating The level of a considering that in 95 ference in the and that in 87% of cas

Discrepancies may be due to both to the differences gation ps and residual errors in the disaggregation of shares to the 1 km cells.

The good results obtained allowed the application of the method to the whole EU15, in this case the aggregation unit is no more the NUTS5, but cells 10 km x 10 km, which represent equal portions of landscape and allow a direct comparison of the value of the index across Europe.

The aggregation used is reported in Table 37, the indicator in Figure 64.

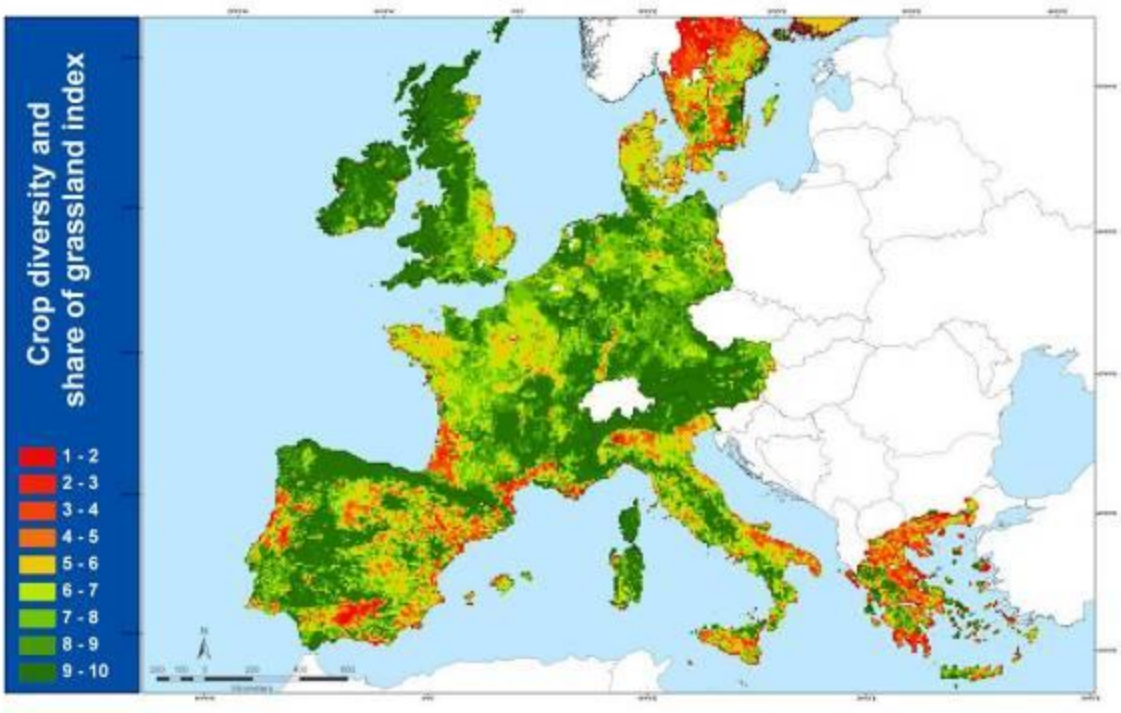
Table 37 Crop classes used in the calculation of the EU15 indicator

barl	Barley
citr	Citrus fruits
dwhe	Durum Wheat
flow	Floriculture
nurs	Nurseries
gras	Permanent gras and grazing
lfall	Fallow land
lfrui	Fruit tree and berry plantations
lmaiz	Maize
loliv	Olive groves
lrape	Rape and turnip rape
ltext	Fibre and oleaginous crops
ltwin	Vineyards
oats	Oats
ocer	Other cereals
ryem	Rye
ocro	Other crops
ofar	Fodder other on arable land
ovto	Tomatoes and Other fresh Vegetables
pari	Rice
pota	Potatoes
puls	Dry pulses
soya	Soya
roof	Other root crops
sugb	Sugar beet

sunf	Sunflower
swhe	Common wheat
oind	Other non permanent industrial crops
toba	Tobacco

	same classes
	grouped

Figure 64. Crop diversity and share of grassland index from CAPRI data



The second sub-indicator (intensity of management of grasslands and crops) can be calculated using as a proxy livestock density of nitrogen surplus. Both are available in CAPRI at a 1 km resolution, and have been rescaled in order to give more weight to areas with low pressure and a negative score to areas with high pressure, according to Table 38.

Table 38 N surplus and Livestock Units Score

	N surplus
-2	>200 (kg/ha)
-1	100 - 200 (kg/ha)
0	50 - 100 (kg/ha)
2	30 -50 (kg/ha)

4	<30	(kg/ha)

Livestock density

-1	>2.0	LU/ha
0	1.5 - 2	LU/ha
1	0.8 - 1.5	LU/ha
3	< 0.8	LU/ha

The sum of the two sub-indicators (with the two options Livestock density or N-surplus) provides the results shown in fig.14 and 15.

Figure 65. Simulation of High Nature Farmland areas with CAPRI data, based on crop diversity and share of grasslands, and livestock density

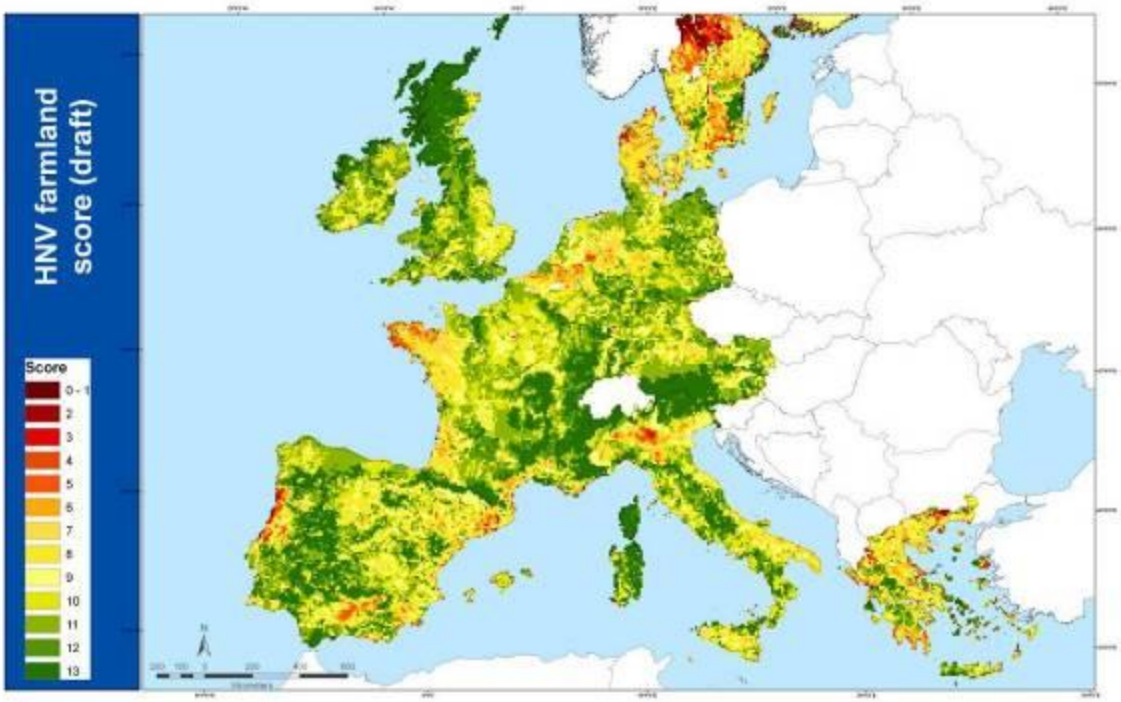
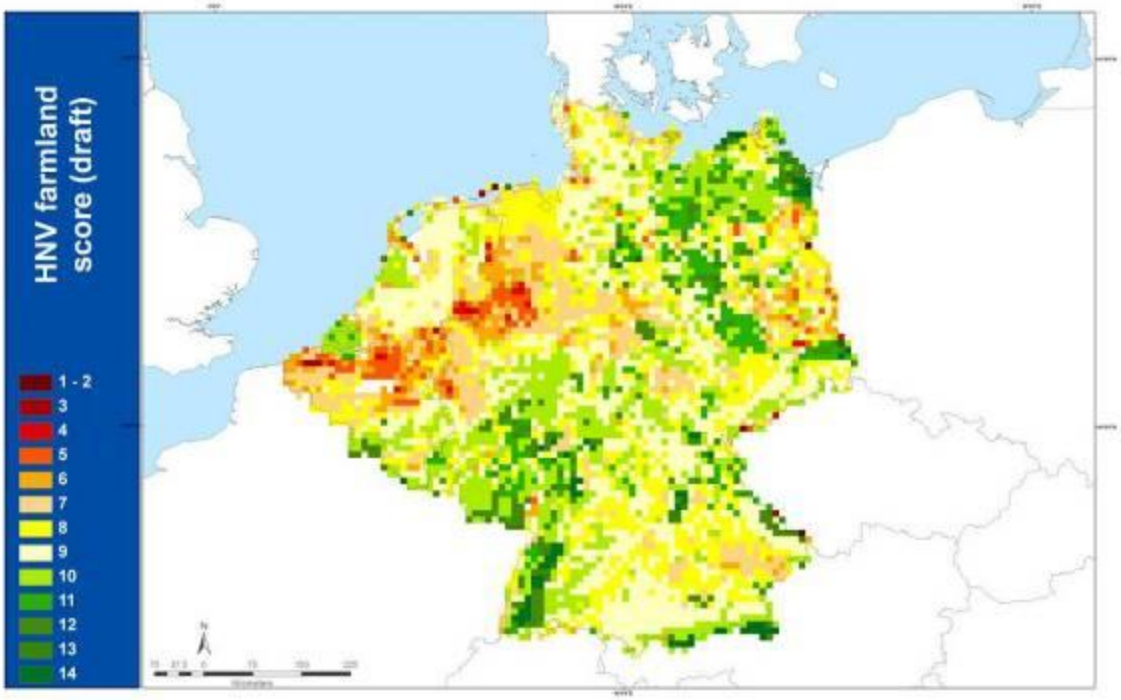


Figure 66. Simulation of High Nature Farmland areas with CAPRI data, based on crop diversity and share of grasslands, and N-surplus



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

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.

Annex: Code lists

Table 39 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)

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
CROP	Total crops production
DWHE	Durum wheat
PULS	Pulses
ROOT	Roots and tubers
INDU	Industrial crops
TEXT	Textile fibre plants
HOPS	Hops
VEGE	Fresh vegetables
TOMA	Tomatoes
CAUL	Cauliflowers
FRUI	Fresh fruit
APPL	Apples
PEAR	Pears
PEAC	Peaches
CITR	Citrus fruit (total)
ORAN	Oranges
LEMN	Lemons
MAND	Mandarins
GRAP	Table grapes
WINE	Wine
TABO	Table olives
OLIO	Olive oil
NURS	Nursery plants
FLOW	Flowers and ornamental plants

OCRO	Other crops
ANIT	Total animal production
ANIM	Animal
SHGO	Sheep and goats
ANIP	Animal products
MILK	Milk
EGGS	Eggs
INPU	Intermediate consumption (total)
FEED	Animal feeding stuffs
FDGR	Animal compounds for grazing livestock
FDPI	Animal compounds for pigs
FDPO	Animal compounds for poultry
FODD	Straight feeding stuffs
FERT	Fertilizers and enrichments
ENER	Energy and lubricants
INPO	Other inputs
GVAM	Gross value added at market prices
SUBS	Subsidies
TAXS	Taxes linked to production (including VAT balance)
GVAF	Gross value added at factor costs
DEPM	Depreciation
LABO	Compensation and social security contributions of employees
RENT	Rent and other payments
INTE	Interests
GFCF	Total of gross fixed capital formation
BUIL	Buildings and other structures
MACH	Transport equipment and machinery
GFCO	Other gross fixed capital formation

Table 40 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 41 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

Table 42 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	Tomatoes	TOMA
Fruits	Other vegetables	OVEG
Other perennials	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	HEIF
	Heifers raising	HEIR
	Fattening of male calves	CAMF
	Fattening of female calves	CAFF
	Raising of male calves	CAMR

Group	Activity	Code
	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
	Other animals	OANI

Table 43 Output, inputs, income indicators, policy variables and processed products in the data base

Group	Item	Code
Outputs		
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	Tomatoes	TOMA
Fruits	Other vegetables	OVEG
Other perennials	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	Gras	GRAS
	Fodder maize	MAIF
	Other fodder from arable land	OFAR
	Fodder root crops	ROOF
	Straw	STRA
Marketable products	Milk from cows	COMI

Group	Item	Code
from animal product	Beef	BEEF
	Pork meat	PORK
	Sheep and goat meat	SGMT
	Sheep and goat milk	SGMI
	Poultry meat	POUM
	Other marketable animal products	OANI
Intermediate products from animal production	Milk from cows for feeding	COMF
	Milk from sheep and goat cows for feeding	SGMF
	Young cows	YCOW
	Young bulls	YBUL
	Young heifers	YHEI
	Young male calves	YCAM
	Young female calves	YCAF
	Piglets	YPIG
	Lambs	YLAM
	Chicken	YCHI
	Nitrogen from manure	MANN
	Phosphate from manure	MANP
	Potassium from manure	MANK
Other Output from EAA	Renting of milk quota	RQUO
	Agricultural services	SERO
Inputs		
Mineral and organic fertiliser Seed and plant protection	Nitrogen fertiliser	NITF
	Phosphate fertiliser	PHOF
	Potassium fertiliser	POTF
	Calcium fertiliser	CAOF
	Seed	SEED
	Plant protection	PLAP
Feedings tuff	Feed cereals	FCER
	Feed rich protein	FPRO
	Feed rich energy	FENE
	Feed based on milk products	FMIL
	Gras	FGRA
	Fodder maize	FMAI
	Other Feed from arable land	FOFA
	Fodder root crops	FROO
	Feed other	FOTH
	Straw	FSTRA
Young animal Other animal specific inputs	Young cow	ICOW
	Young bull	IBUL
	Young heifer	IHEI
	Young male calf	ICAM
	Young female calf	ICAF
	Piglet	IPIG
	Lamb	ILAM
	Chicken	ICHI
	Pharmaceutical inputs	IPHA

Group	Item	Code
General inputs	Maintennce machinery	REPM
	Maintennce buildings	REPB
	Electricity	ELEC
	Heating gas and oil	EGAS
	Fuels	EFUL
	Lubricants	ELUB
	Water	WATR
	Agricultural services input	SERI
	Other inputs	INPO
Income indicators	Production value	TOOU
	Total input costs	TOIN
	Gross value added at producer prices	GVAP
	Gross value added at basic prices	GVAB
	Gross value added at market prices plus CAP premiums	MGVA
Activity level	Cropped area, slaughtered heads or herd size	LEVL
Policy variables Relating to activities	Premium ceiling	PRMC
	Historic yield	HSTY
	Premium per ton historic yield	PRET
	Set-aside rate	SETR
	Premium declared below base area/herd	PRMD
	Premium effectively paid	PRME
	Premium amount in regulation	PRMR
	Type of premium application	APPTYPE
	Factor converting PRMR into PRMD	APPFACT
	Ceiling cut factor	CEILCUT
Processed products	Rice milled	RICE
	Molasse	MOLA
	Starch	STAR
	Sugar	SUGA
	Rape seed oil	RAPO
	Sunflower seed oil	SUNO
	Soya oil	SOYO
	Olive oil	OLIO
	Other oil	OTHO
	Rape seed cake	RAPC
	Sunflower seed cake	SUNC
	Soya cake	SOYC
	Olive cakes	OLIC
	Other cakes	OTHC
	Gluten feed from ethanol production	GLUE
	Biodiesel	BIOD
	Bioethanol	BIOE
	Palm oil	PLMO
	Butter	BUTT
	Skimmed milk powder	SMIP
	Cheese	CHES
	Fresh milk products	FRMI
	Creams	CREM
	Concentrated milk	COCM

Group	Item	Code
	Whole milk powder	WMIO
	Whey powder	WHEP
	Casein and caseinates	CASE
	Feed rich protein imports or byproducts	FPRI
	Feed rich energy imports or byproducts	FENI

Table 44 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 45 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 46 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 47 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*

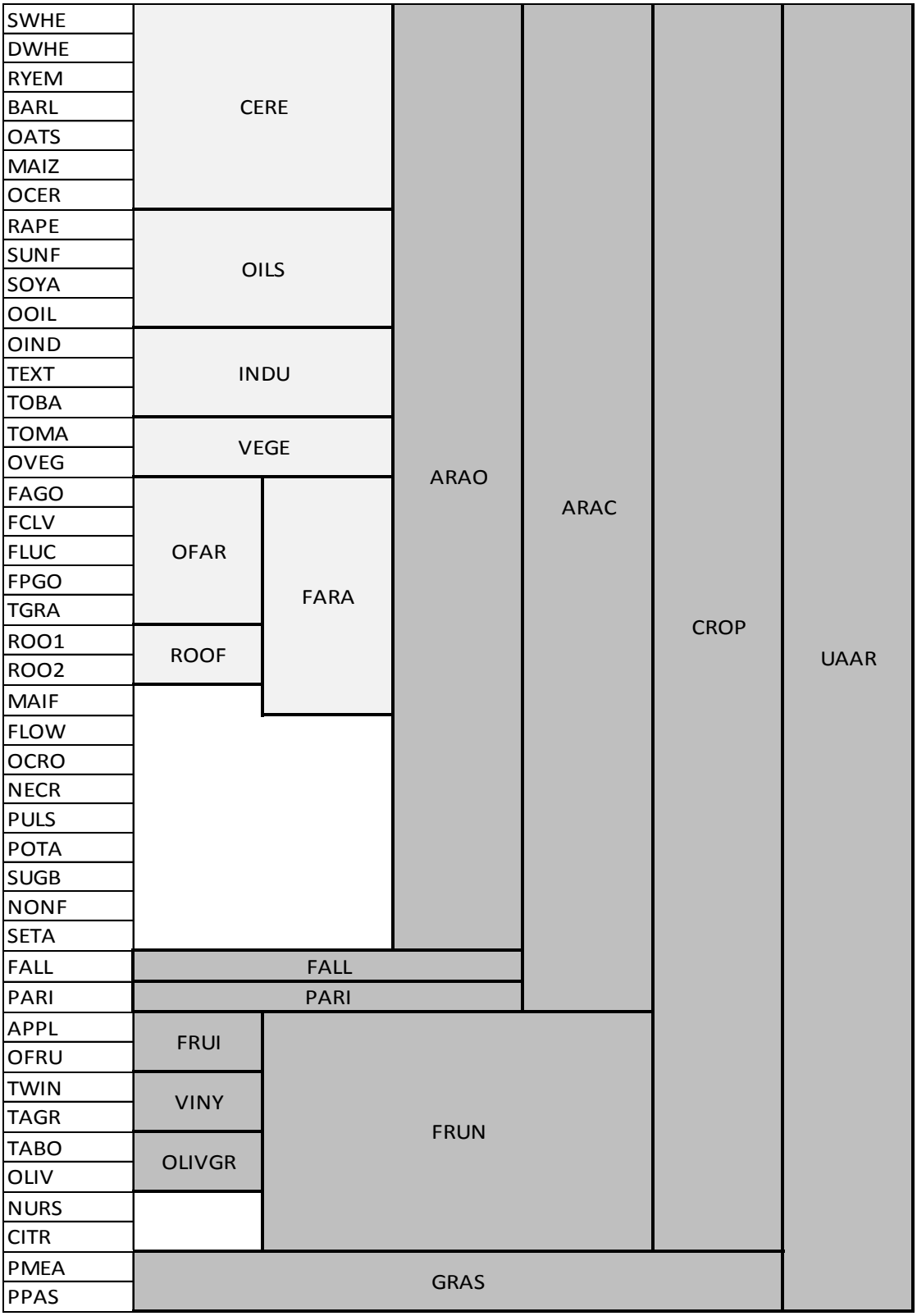



Figure 67. Mapping primary agricultural activities to groups and land use in coco