CAPRI model documentation 2008: Version 2

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

Table of Contents

---

[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. We hope for more contributions, revisions and thus sections authors or editors from the CAPRI network and thereby for an improved Version 2 in the next weeks.

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.

In the short run the editors of this updated documentation have simply taken the last comparable effort (Britz, Heckelei, Kempen 2007) as a starting point and added, based on their personal involvement, knowledge and assessment of urgency, modifications or simply brief assessments on the topicality of the respective sections ("Update notes"). This is thus not a full update but a kind of annotated new edition of the existing documentation amended with some new material. All changes are therefore marked to highlight what has been changed in recent times.

End of 2008, a user manual for CAPRI was generated. The section on the CAPRI Graphical User interface was therefore removed.

1 Introduction

1.1 Structure of the documentation

The documentation is structured as follows. The short introduction in chapter 0 first gives an overview of the CAPRI activities followed by a short description of the system. The rest of the document follows the project workflow: the different steps of building up the national and regional data base (chapter 2), the allocation of different inputs (chapter 2.5) and the projection tools needed to establish a baseline (chapter 3) are discussed. Chapter4 deals with the scenario impact analysis: description of the different modules of the economic model and their relationships. In the last three chapters (chapters 5 - 8) the farm type approach, features for post model analysis and the exploitation tools used in CAPRI are briefly 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

---

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

these to a simple market model distinguishing the EU and rest-of-the-world. In parallel, a team at the FAL 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) added a dis-aggregation below administrative regions in form of farm type models, refined the existing environmental indicators and added new ones. A new project with the original network (QLTR-2000-00394: CAP-STRAT 2001-2004) refined many of the approaches of the first phase, and linked a complex spatial global multi-commodity model into the system. The application of CAPRI for sugar market reform options in the context of another project improved the way the complex ABC sugar quota system is handled in the model.

In 2004, again a larger project (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, a project for IPTS/JRC started to update and improve 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 SEAMLESS the farm type layer of CAPRI will be refined and updated, and a module for endogenous structural change is foreseen. In parallel, the team in LEI, The Hague, The Netherlands, will apply 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. Two related smaller projects (“Baltic Stern” and “KLIMMZUG”) have also been launched in 2007 to investigate agricultural contributions and abatement options related to emissions into the Baltic sea. With Kick-off in June 2008 CAPRI is 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 will be extended to 30 years and in turn linkages to forestry and bioenergy sectors of EUFASOM will be developed.

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.

---

[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

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 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. If not explicitly mentioned in the following, the documented features had been co-financed by DG-RSRCH. The documentation as it stands now captures the state of the system in spring 2007 at the end of the CAPRI Dynaspat project. It is planned to update the documentation on a regular basis if the need arises but for the time being only a selective update on important recent changes is possible.

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)[7]. 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 0 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

---

[7] 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.

different aspects not covered by CAPRI or help to learn about possibility aggregation errors in CAPRI.

The economic model 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 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, 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.

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 40 primary and processed agricultural products, covering about 60 countries or country blocks in 28 trading blocks (0 on page 89). 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 Product Support Equivalents and Consumer Support Equivalents (PSE/CSE) from the OECD, (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.

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 model may be used in a comparative-static or recursive-dynamic fashion.

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*. These exploitation tools are further explained in the last chapter.

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 three main 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 2.3).
- 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. The input allocation step allows the calculation of regional and activity specific economic indicators such as revenues, costs and gross margins per hectare or head. The regionalisation step introduces supply oriented CAP instruments like premiums and quotas (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 2.6).

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

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.

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

2.2 Linking production activities and the market

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

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

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

Equation 1 $$GROF_{io} = \sum_j LEVL_j IO_j$$

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

$$\mathrm{GROF}_{io} - \mathrm{SEDF}_{io} - \mathrm{LOSF}_{io} - \mathrm{INTF}_{io} = \mathrm{NETF}_{io}$$

Equation 2

$$\begin{aligned}\mathrm{NETF} + \mathrm{IMPT}_{io} = {} & \mathrm{EXPT}_{io} + \mathrm{STCM}_{io} \\ & + \mathrm{FEDM}_{io} + \mathrm{LOSM}_{io} \\ & + \mathrm{SEDM}_{io} + \mathrm{HCOM}_{io} \\ & + \mathrm{INDM}_{io} + \mathrm{PRCM}_{io} \\ & + \mathrm{BIOF}_{io}\end{aligned}$$

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 2.3.4).

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.

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	

2.3 The Complete and Consistent Data Base (COCO) for the national scale (Peter Witzke)

2.3.1 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[8] 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 into the database additional national sources were used which became available as part of CAPSIM projects[9].

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

Data items and their main sources

Data items	Source
Activity levels	Land use statistics, herd size statistics, slaughtering statistics, statistics on import and export of live animals
Production	Farm and market balance statistics, crop production statistics, slaughtering statistics, statistics on import and export of live animals
Farm and market balance positions	Farm and market balance statistics
Sectoral revenues and costs	Economic Accounts for Agriculture (EAA)
Producer prices	Derived from production and EAA
Consumer prices	Derived from macroeconomic expenditure data and International Labour Office data on food prices
Output coefficients	Derived from production and activity levels, engineering knowledge

Source: Eurostat (http://epp.eurostat.cec.eu.int), several bio-physical econometric studies and European Commission (http://publications.eu.int/general/oj_en.html).

2.3.2 COCO Step 1: Initialisation as an overlay from various sources

The COCO module is basically divided into two main parts: (1) including and combine input data according to some hierarchical criteria, and (2) calculating complete and consistent time series while remaining close to the raw data. The first part, closely related to the collection of raw data, forms a bridge between raw data and data consolidation to impose completeness and consistency described in the next section. This first 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

---

[8] The ‘Common Agricultural Policy Simulation Model’ (CAPSIM) was developed by Dr. Heinz-Peter Witzke, EuroCare, Bonn (http://www.eurocare-bonn.de/profrec/capsim/capsim_e.htm).

[9] 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).

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. A quite detailed exposition of this initialisation step close to the GAMS code is given in Witzke, Zintl, Tonini 2008. At this point we only need to list the recent changes:

New products casein and whey powder

Because 'ready to use' market balances are not offered by Eurostat these market balances had to be estimated based on some hard data and assumptions. The "hard" data for the extension of the product list were (1) production data from Eurostat and (2) consolidated external trade data based on the Food and Agriculture Organization (FAOSTAT) integrated core database. This allowed to calculate total demand on the EU level. The disaggregation of demand was 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. Moving from the EU level to the MS level required, for example to apply the estimated EU shares for demand components to all EU countries. MS particularities are covered, however, in terms of the significance of casein and whey powder for national dairy markets, because total domestic use may be calculated from Eurostat data.

New region: Turkey

Relying on merged Eurostat and expert data from ASA, a dataset (1995-2005) has also been compiled for Turkey which corresponds to CAPRI definitions. The quality is comparable to Western Balkan data but checks are still ongoing.

Revised methodology for initialisation

Use of FAO data for trade in MS15 countries

Earlier versions of COCO left the task of completiong the market balances in case of missing data entirely to the second estimation step (next section). Occasionally this a led to funny forecasting or backcasting with huge trade quantities if only production was known. Therefore it was deemed useful to use the statistical data from FAO as a fall back option with an appropriate conversion factor to Eurostat to estimate missing trade of MS15 and thus to stabilise market balance completions.

Merging of Eurostat with expert data in selected countries

So far a merging procedure existed for MS10 countries to combine national data on New MS from an earlier CAPSIM project with existing Eurostat data. After a former partner in this project these are called the 'Ariane data'.

The solution in the 2006 Balkan study had been different as only very few Eurostat data were given. In this situation we used only national data and any Eurostat data had been ignored. In 2007 with more Eurostat data becoming available it appeared useful to introduce some merging methodology also on selected Balkan countries, in particular Croatia, Macedonia, and Turkey. Supplementary national data were also introduced on the dairy and meat sectors in BG and RO, using information collected by our partner ASA institute in the CAPSIM study.

Revision of milk content estimation

The existing methodology basically fixed the processing of raw milk and marketable production of derived dairy products as well as the contents of raw milk and butter. The balancing thus had to occur mainly through an adjustment of 'other' milk contents leading to implausible results in a number of cases, for example fresh milk products containing only 0.5% of fat on average.

The revision involved two modifications: Use of milk production statistics to get statistical data on the fat content of cheese, cream etc. (rather than relying exclusively on assumed technical coefficients) and second some leeway for marketable production and raw milk contents to increase the plausibility of milk product contents.

Implausible zero and nonzero yields

It is ensured now with a probability close to one that the coco results will not have areas without yields which occurred occasionally in the case of small areas, in particular for activities OCRO and OIND. The price to be paid for this increase in plausibility is that the EAA may need to be changed a bit (inventing small yields if necessary). More rarely the former COCO results may have included some production of OIND without a producing area this should be fixed as well.

Double counting of cotton area

There is a clear double counting in the land use statistics at Eurostat. This holds for Spain, Greece, Bulgaria, and Italy where data on cotton area are available. Fixing this so far undetected error implies that the aggregate activity textile crops will produce cotton seeds and cotton lint and that the production of cotton seeds will not be an output of an oilseeds activity anymore. This required to correct the yields. The yields of textile crops in terms of other oilseeds are calculated by dividing the production of cotton seeds by the area for textile crops and the yields of the other oilseeds activity is calculated by dividing the (remaining) production of other oilseeds by the corresponding area. 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.

Processing yields of oilseeds

Thanks to complaints by Torbjoern Jansson, LEI, additional security checks have been included on the processing yields of oilseeds. This includes an equation ensuring that the sum of crushing coefficients is bounded between 0.96 and 1.00 in the estimation procedure described in the following.

2.3.3 COCO Step 2: 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.

COCO included data ranging from 1985 to 2002 for the 14 member states of the EU[10] at that time, from the national data found in NEWCRONOS[11]. 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.

---

[10] In CAPRI Luxembourg is aggregated to Belgium as a NUTS 2 region. The 10 new Member States were included in 2004, Bulgaria, Romania, and the Western Balkan countries in 2006.

[11] Data for Norway are processed by COCO as well, but naturally, stem from different sources.

(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. Given ex-post data, we naturally want the estimates to fit the given data as close as possible, but simultaneously require the estimates to be in line with economic theory. The latter point is typically ensured by two approaches: (1) the estimation equations are in line with some optimisation problem in the background (for example profit maximisation, i.e. the supplied outputs are regressed on a function of prices whose functional form is derived from first order conditions of a profit maximisation problem) and (2) appropriate restrictions on the parameters ensure that the resulting system is in line with first and second order conditions of a profit maximisation problem. The ultimate aim is the combination of a functional form and parameter restrictions which allows for both a good fit and conformity with micro-economic theory. 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:

Equation 4

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

s.t.

(1.2) $y_{i,t}^{LO} \le y_{i,t} \le y_{i,t}^{UP}$

(1.3) Accounting identities defined on $y_{i,t}$

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 $y_{i,t}^{lo}, y_{i,t}^{up}$. The ‘hard’ bounds $y_{i,t}^{LO}, y_{i,t}^{UP}$ are constraining the feasible space for a number of solution attempts. However, if it turns out that certain constraints would persistently preclude feasibility of the data consolidation problem, they are relaxed in a stepwise fashion, but this widening of bounds is monitored on a parameter to check.

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

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

Equation (1.3) 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.

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:

a) Crop production (land balance + yields) for all crops simultaneously

b) Production, yields, EAA, market balances for groups of animals (e.g. 'cattle sector')

c) Crop EAA + market balances for groups of crops, taking production from (a) as given.

This procedure has developed as a path dependent compromise between computation time and presumed quality. 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 the recent (2007/08) update 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 in the case of the Western Balkan countries it was necessary to transfer certain coefficients and shares from (previously consolidated) neighbouring countries to the Western Balkan, such that a certain sequence in necessary for a reasonable application of COCO.

A final step (COCO2.gms) estimates consumer prices and by-products used as feedstuffs. Both tasks run simultaneously for all countries and build on intermediate results (e.g. human consumption and processing quantities). A full application of COCO thus requires three steps in technical terms:

1. Run COCO1 for the full time series on EU27 countries, either in one batch from the GUI or one by one (always with sub-steps a to c)
2. Run COCO1 for the set of candidate countries (Western Balkan and Turkey) on the reduced time span with given data (1995 – current). Because these use some shares and ratios from an average of selected EU27 countries step 1 must be completed first.
3. Run COCO2 for all countries with time span as in step 1 to obtain the full results.

2.3.4 Update note: ex post data on biofuels in the EU

EU biofuels data are currently (April 2008) introduced via a subprogram of CAPREG, the CAPRI regionalisation tool which has incorporated a number of other tasks in the meantime (Section 2.4). However because it is logically a part of the national ex post data preparation and probably will become a part of COCO soon, the following information is given as a supplement to Section 2.3.

The ex-post data on biofuel production ex-post are taken from the European Biodiesel Board and the European Bioethanol Fuel Associations. The data had been edited in the file “biofuel\bio_fuel_prod_data.gms”. Given the short nature of the time series, and especially the fact that the current base year period 2001-2003 is not fully, or not fully covered, the data were backcasted by simply using the last known value. That will most probably somewhat overestimate the production data in the base year – a there year average around 2001 -, but that was not deemed as a major problem.

In order to integrate the data in the CAPRI data base, the production number for the biofuels must be converted into processing demand for single products. The conversion factors were based on OECD study on biofuel based on the AGLink model (V. LAMPE, M., 2006). Those data are edited in “biofuel\coeff.gms” along with the conversion factors for gluten feed from bioethanol production, equally taken from FAO.

In order to estimate the share of the raw products on the production of biodiesel and bioethanol, some data could be located for France, Germany, Sweden and Poland. In order to estimate a share for the other Member States, the data on industrial use (INDM) from the CoCo data base were used, and if no industrial use was shown, the data on human consumption (HCOM) taken into account to define starting values for the shares. From those shares, expected quantities used for biofuel production (BIOF) were derived ex-post. Finally, the estimates were reduced if necessary where the production data for a biofuel exceeded the sum of human consumption plus industrial use converted into fuel of the products it could be produced from.

In order to get a consistent data set where the production of biofuels is equals to the sum of the inputs multiplied with the respective conversion factors, and the market balances are closed, a very simple Highest Posterior Density estimator was applied, with the following constraints:

- The sum of industrial use (INDM) and human consumption (HCOM) as found in the CAPRI data base must be equal to the corrected estimates for industrial use (INDM) and human consumption (HCOM). Plus the newly introduced position “use for bio-fuel production (BIOF)”.
- The production of biofuels must be equal to the sum of the processing input (BIOF) for the different products times their conversion rates.

For details, see “biofuel\trim_expost.gms”. The objective function minimizes the relative deviations between the estimated use for biofuel production derived from the shares plus the relative deviation from the given human consumption, the latter weighted with thousand to give preference for adjustments of the shares over correction of the human consumption position. Human consumption will hence only be sizably changed if the industrial use shown is not sufficient to provide the input for production data on biofuels.

2.4 The Regionalised Data Base (CAPREG)

2.4.1 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 with an emphasis on the impact of the CAP. The core of the system consists of a regionalized 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.

2.4.2 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]

2.4.3 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.

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)

2.4.4 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.

2.4.5 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[12].

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

---

[12] 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.

- *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.

2.4.6 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[13]:

Equation 5 $UVAG_r = UVAG_s$

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

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

2.4.7 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-

---

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

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.

2.4.8 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:

Equation 6

$$\begin{aligned}\mathrm{LEVL}_{\mathrm{OATS,r}} = (&\mathrm{CEREAL}_{\mathrm{r}} \\ &- \mathrm{WHEAT}_{\mathrm{r}} - \mathrm{BARLEY}_{\mathrm{r}} - \mathrm{MAIZEGR}_{\mathrm{r}} - \mathrm{RICE}_{\mathrm{r}}) \cdot \\ &\frac{\mathrm{LEVL}_{\mathrm{OATS,COCO}}}{(\mathrm{LEVL}_{\mathrm{OATS,COCO}} + \mathrm{LEVL}_{\mathrm{RYE,COCO}} + \mathrm{LEVL}_{\mathrm{OCER,COCO}})}\end{aligned}$$

Similar equations are used to break down other aggregates and residual areas in REGIO[14].

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.

2.4.9 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.

---

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

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[15]. 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 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 7

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

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

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

---

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

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.

2.4.10 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 8

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

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.

2.5 Input Allocation

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

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

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

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

- *For nitrate, phosphate and potash*, nutrient balances are constructed so to take into account crop and manure nutrient content and observed fertiliser use, combined with 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.

2.5.1 Input allocation excluding young animals, fertiliser and feed

2.5.1.1 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.

2.5.1.2 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.

2.5.1.3 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.

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 0, 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.[16] 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. [17]

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.[18]

2.5.1.4 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.[19] 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 (0).

---

[16] 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.

[17] 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 0 or the appendix).

[18] 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.

[19] 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:

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

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

$$\text{s.t.}$$

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

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

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

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

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

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

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

Equation 9

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.

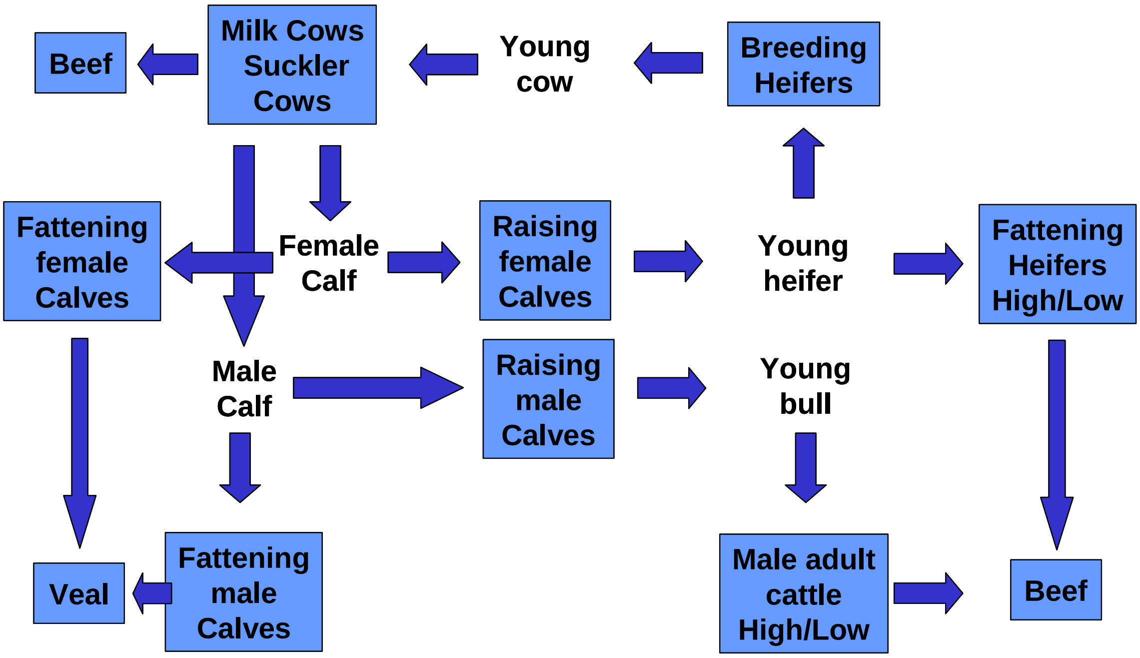
2.5.1.5 How are the results used in CAPRI?

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.

2.5.2 Input allocation for young animals and the herd flow model

Figure 1 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 weight 50 kg (female) and 55 kg (male) 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 1. 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.

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.

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

2.5.3 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[20].

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.

2.5.3.1 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 10 $$\sum_{r,\text{fint}} \overline{FEDUSE}_{r,\text{fint}} PFOD_{r,\text{fint}} = \overline{EAAP}_{OFAR,MS} + \overline{EAAP}_{GRAS,MS}$$

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

---

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

Equation 11 $$GVAM_{r,\text{fint}} = \overline{YIELD}_{r,\text{fint}} PFOD_{r,\text{fint}} - \overline{TOIN}_{r,\text{fint}}$$

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

Equation 12
$$
\begin{aligned}
&\sum_k \sup_{r,\text{fint},gvam,k} p_{r,\text{fint},gvam,k} = GVAM_{r,\text{fint}} \\
&\sum_k \sup_{r,\text{fint},price,k} p_{r,\text{fint},price,k} = PFOD_{r,\text{fint}} \\
&\sum_k p_{r,\text{fint},gvam,k} = 1 \\
&\sum_k p_{r,\text{fint},price,k} = 1 \\
H(PROB) = &- \sum_{r,\text{fint},price,k} p_{r,\text{fint},price,k} \log(p_{r,\text{fint},price,k} / pq_k) \\
&- \sum_{r,\text{fint},price,k} p_{r,\text{fint},gvam,k} \log(p_{r,\text{fint},gvam,k} / pq_k)
\end{aligned}
$$

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 13 $$\overline{YIELD}_{r,\text{fint}} PFOD_{r,\text{fint}} \geq \overline{TOIN}_{r,\text{fint}}\, \overline{gvafac}$$

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.

2.5.3.2 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).
- Feeding ratios of animals are quite unstable.

2.5.4 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.

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

2.5.4.1 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 sub-section 2.5.4.3)

2.5.4.2 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 m^3 manure per year, so that the numbers in the table above are multiplied with 18 m^3 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:

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 m^3 of manure per sow (200 kg sow) and 15 piglets at 10 kg over a period of 42 days. Consequently, the manure output of sows varies in the model with the number of piglets produced.

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

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

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

The nitrogen emission factors from animal activities are coupled to crude protein intake (IPCC 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.

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*

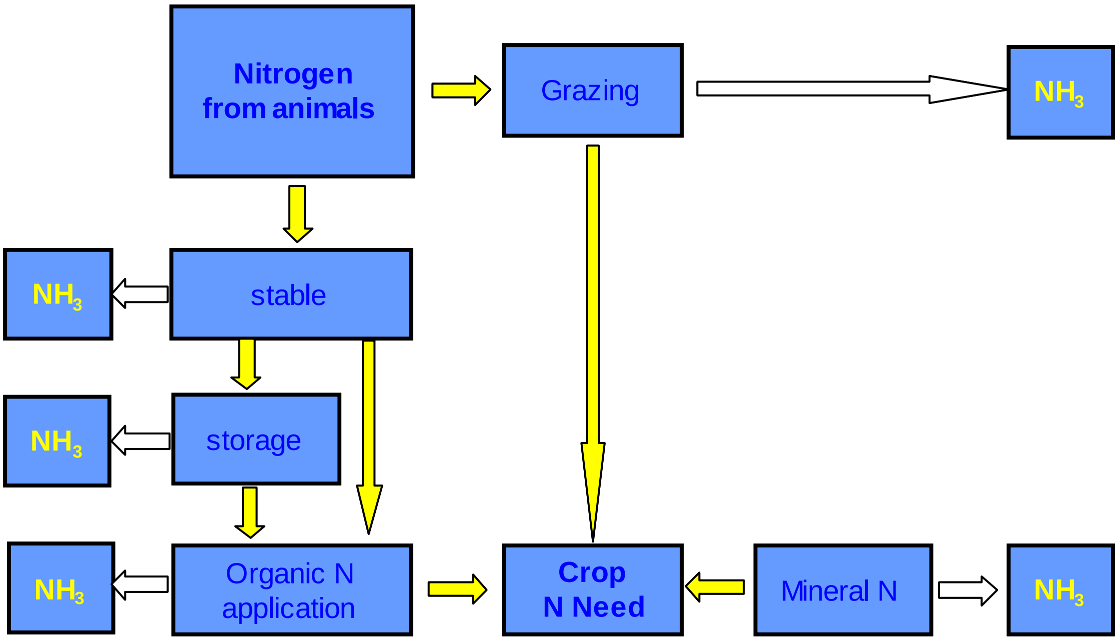
Source: CAPRI Modelling System

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.

2.5.4.3 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 2. Ammonia sinks in the Ammonia emission module



Source: CAPRI modelling system

In Figure 2, 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.

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

2.5.4.4 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:

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:

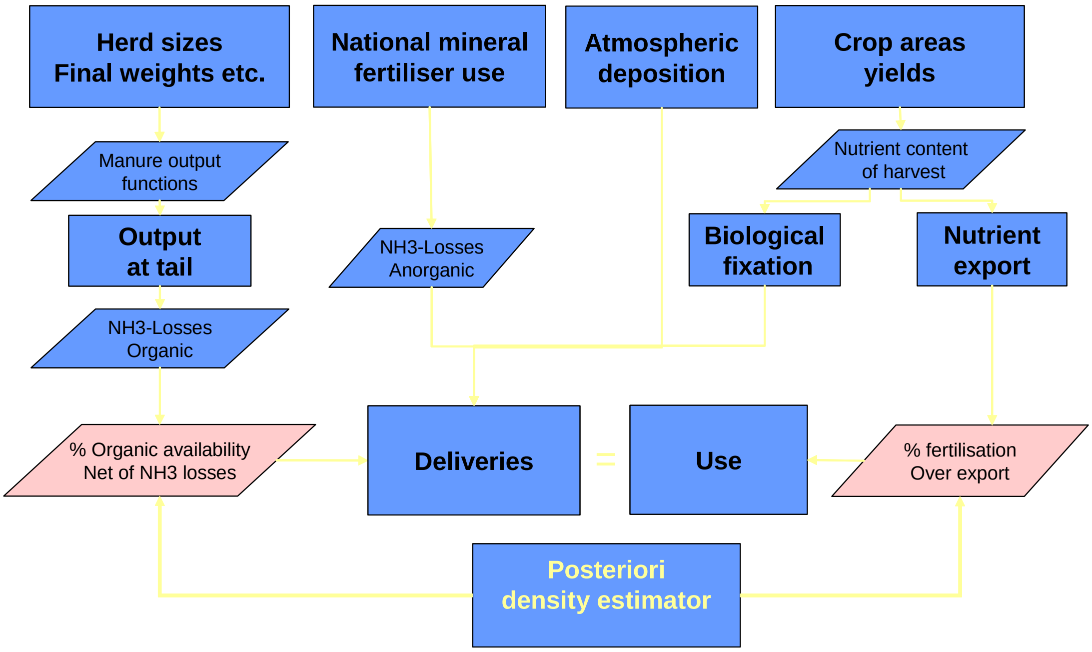
Atmospheric deposition of N per kg and year

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 3 offers a graphical representation of these relationships.

Figure 3. 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 14 $$\overline{Nettrd}_{MS}^{Fnut} = \sum_{r} Nettrd_{r}^{Fnut}$$

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 15
$$\begin{aligned}
&\sum_{cact} \begin{array}{l} Levl_{r,cact} Fnut_{r,cact}\left(1 - NFact_{Fnut,cact}^{biofix}\right) \\ NutFac_{r,fnut}\left(1 + NutFacG_{r,fnut} \wedge cact \in ofar, grae, grai\right) \end{array} \\
&= NETTRD_{r}^{Fnut}\left(1 - NH3Loss_{Fnut,r}^{Anog}\right) \\
&+ NBal_{r}^{AtmDep} NFact_{Cact}^{AtmDep} \\
&+ \sum_{aact} Levl_{r,aact} Fnut_{r,aact}\left(1 - NH3Loss_{Fnut,r}^{Manure}\right)\left(1 - NavFac_{r,fnut}\right)
\end{aligned}$$

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 16
$$\begin{aligned}
\min \quad HDP = &-\sum_{r,fnut}\left(\frac{NutFac_{r,fnut} - \mu_{r,fnut}^{NutFac}}{\sigma_{r,fnut}^{NutFac}}\right)^{2} \\
&-\sum_{r,fnut}\left(\frac{NavFac_{r,fnut} - \mu_{r,fnut}^{NavFac}}{\sigma_{r,fnut}^{NavFac}}\right)^{2} \\
&-\sum_{r,fnut}\left(\frac{NutFacG_{r,fnut} - \mu_{r,fnut}^{NutFacG}}{\sigma_{r,fnut}^{NutFac}}\right)^{2} \\
&-\sum_{r,ngrp}\left(\frac{Nitm_{r,ngrp} - \mu_{r,ngrp}^{Nitm}}{\sigma_{r,ngrp}^{NavFac}}\right)^{2} \frac{\overline{LEVL}_{r,UAAR}}{\overline{LEVL}_{r,ngrp}}
\end{aligned}$$

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:

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

Ex-post

- **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

Ex_ante

- **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

2.5.4.4.1 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.

2.5.4.5 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 production, manure management and rice cultivation* and *N_2O from agricultural soils and manure management*[21].

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:

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

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

---

[21] 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.

2.5.4.5.1 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.

2.5.5 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.

2.5.5.1 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:

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

Benchmark model:

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

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.

2.5.5.2 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:

$C_{25}(1993)=C_{24}(1992)_{n24-25}-NA_{24-25}$

Where C_{25}= Cohort aged 25
C_{24}= Cohort aged 24
N_{24-25}= Probability of survival for 24 year olds in 1992
NA_{24-25}= 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.

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.

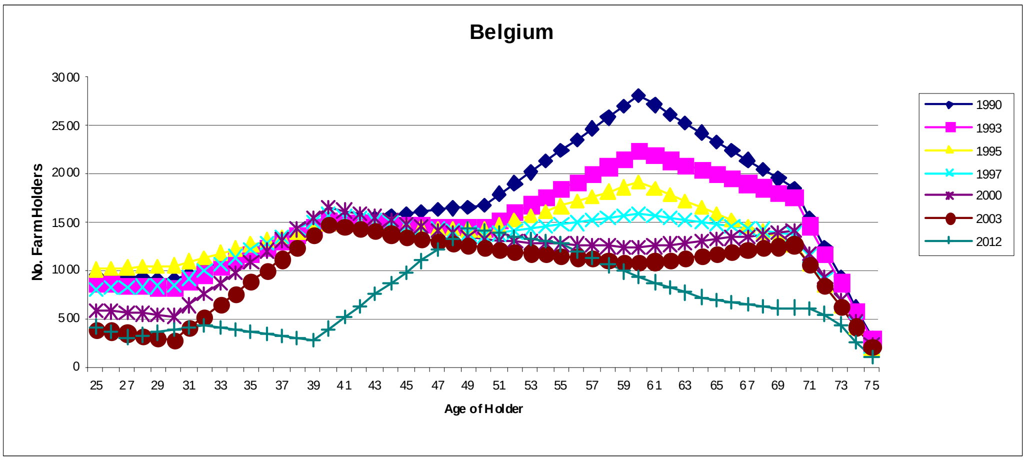
Autonomous changes are forecast by multiplying the number of holders for each age by the probability of survival for each age. New entrants were allowed for by creating a cohort of holders aged between 12 and 24. We base our ex-post analysis of the cohort results on the following assumptions:

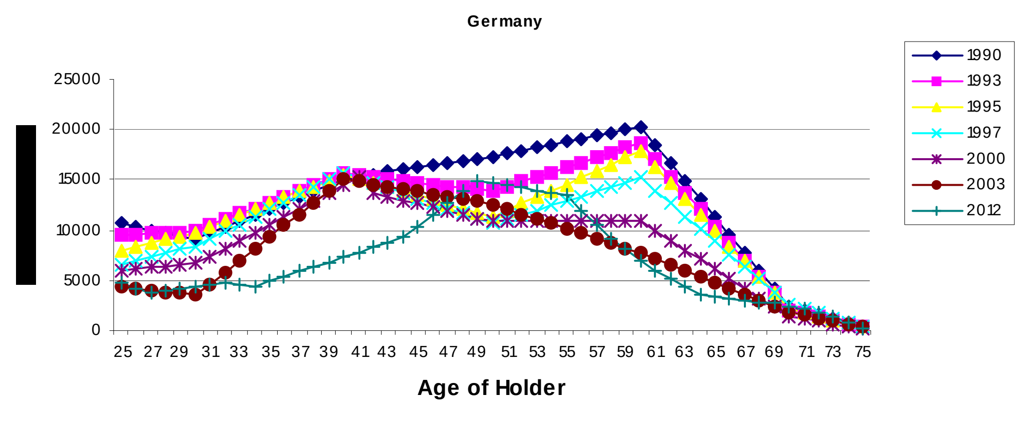
- Death and disability: disability occurs through physical invalidity or by reaching the age of 75 years.
- Occupational mobility is approximated by the sum net changes in cohorts 25-55, when these changes were not due to death or disability. Those, who release their agricultural holdings, are assumed 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.

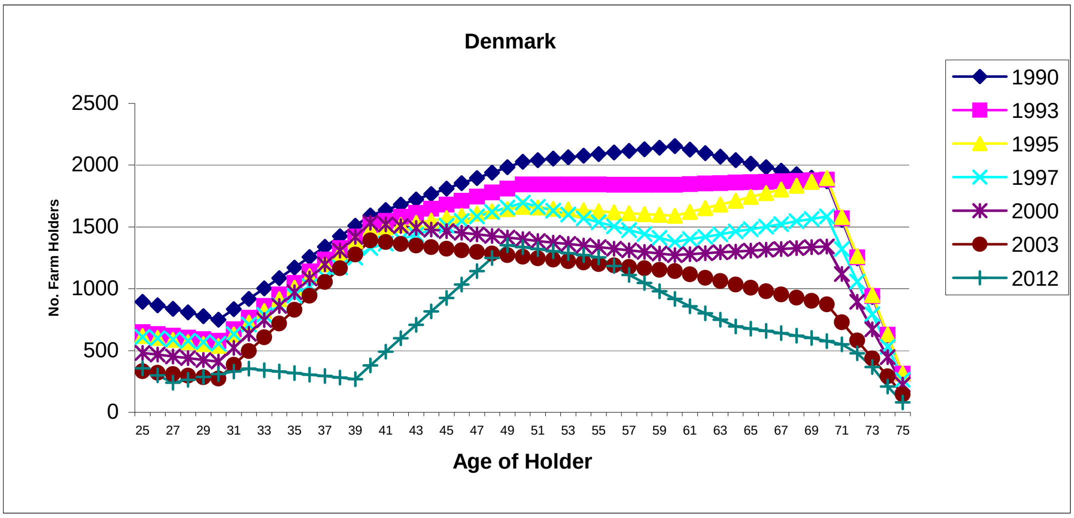
To supplement the analysis, we construct ‘age-lines’ that illustrate the estimated number of holders in a country at each age. We incorporate two additional procedures to Steele and Gaffney (1998); (1) in order to allow for new entrants, the ratio of holders from 10-14, 15-19, and 20-24 to the total 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.

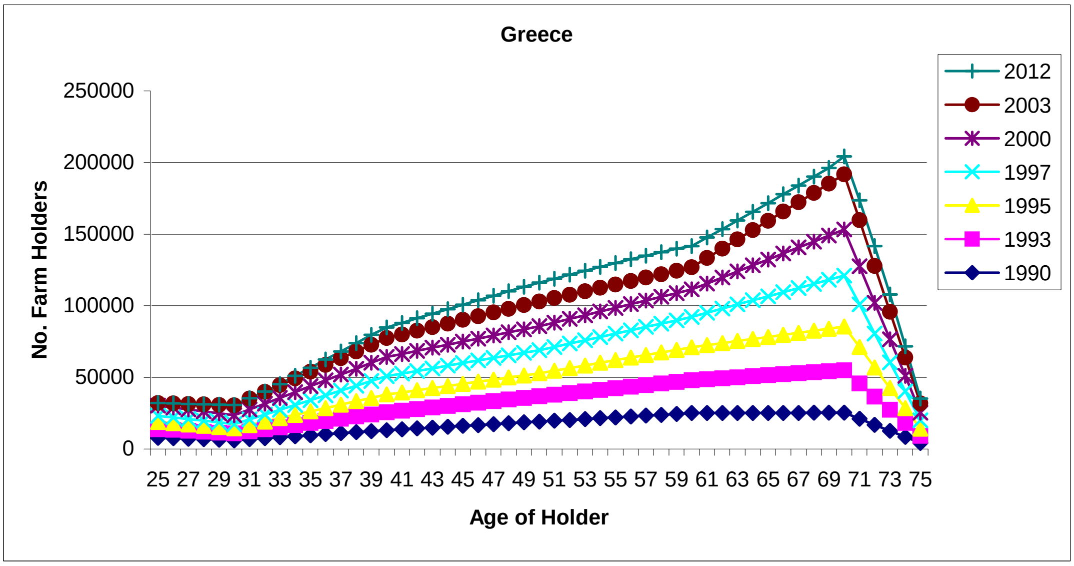
In making projections for 2012, we are required to make additional assumptions. Trends for 10-54 year olds are calculated based on demographic changes using 1993 death rates. For the 55-75 cohort it is assumed that the annual rate of retirement (or exit for non-demographic reasons) is the same as the 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 4 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 4.

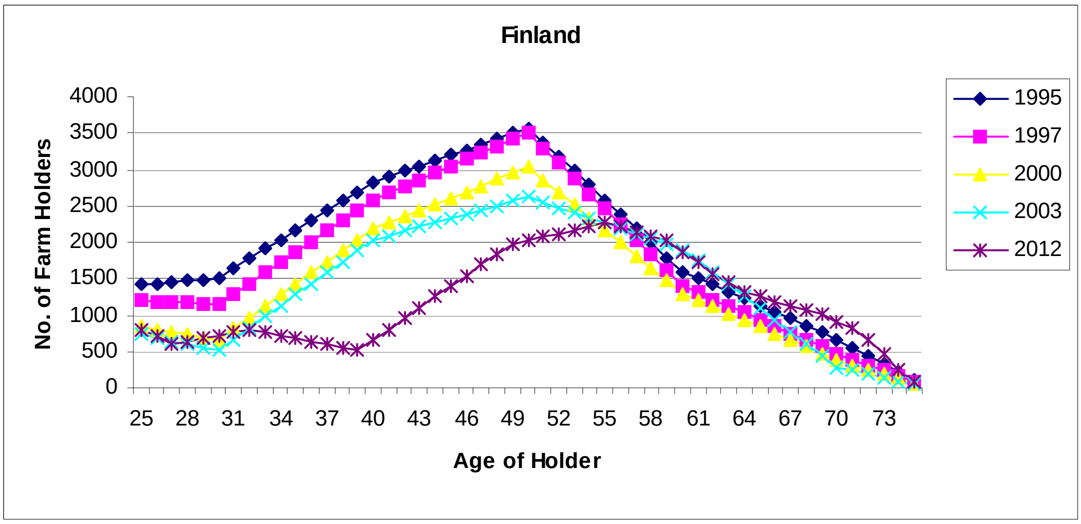
Figure 4. Agelines: Farm Stuctural Change 1990-2003 and Forecasted Change in 2012

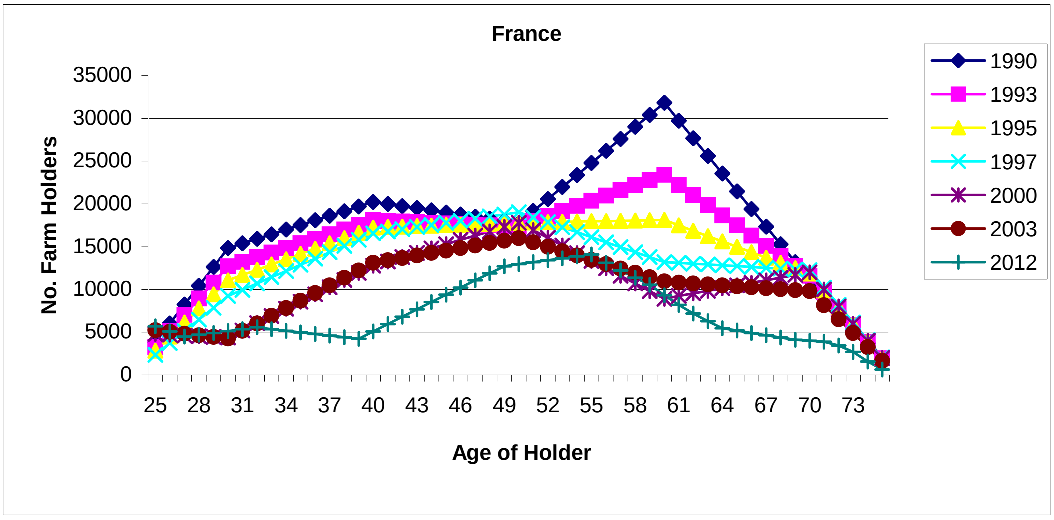


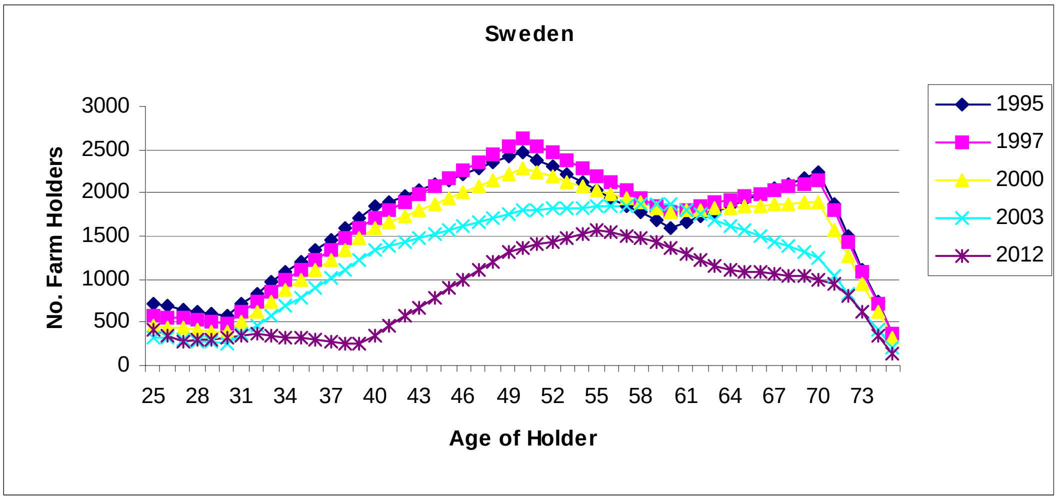


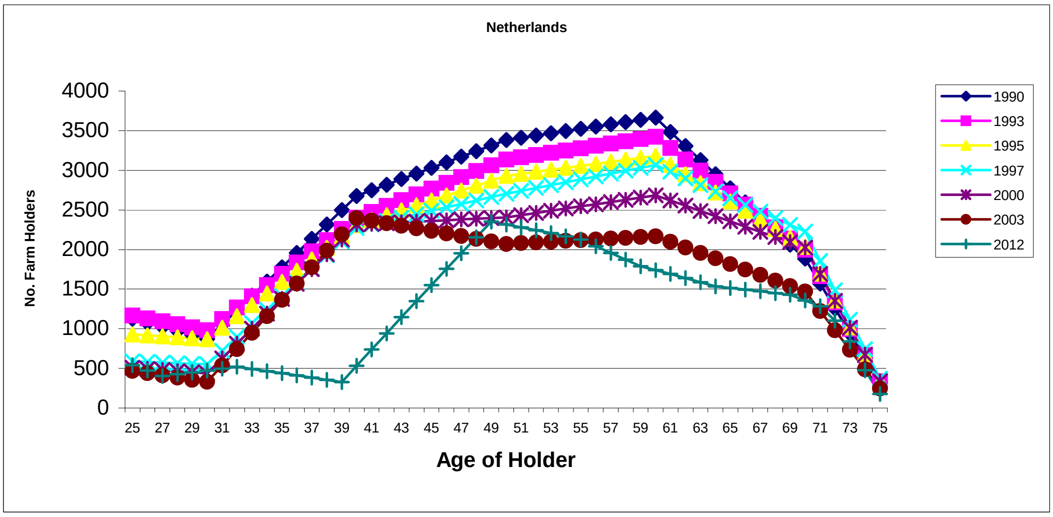

Denmark
No. Farm Holders
2500
2000
1500
1000
500
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
1990
1993
1995
1997
2000
2003
2012

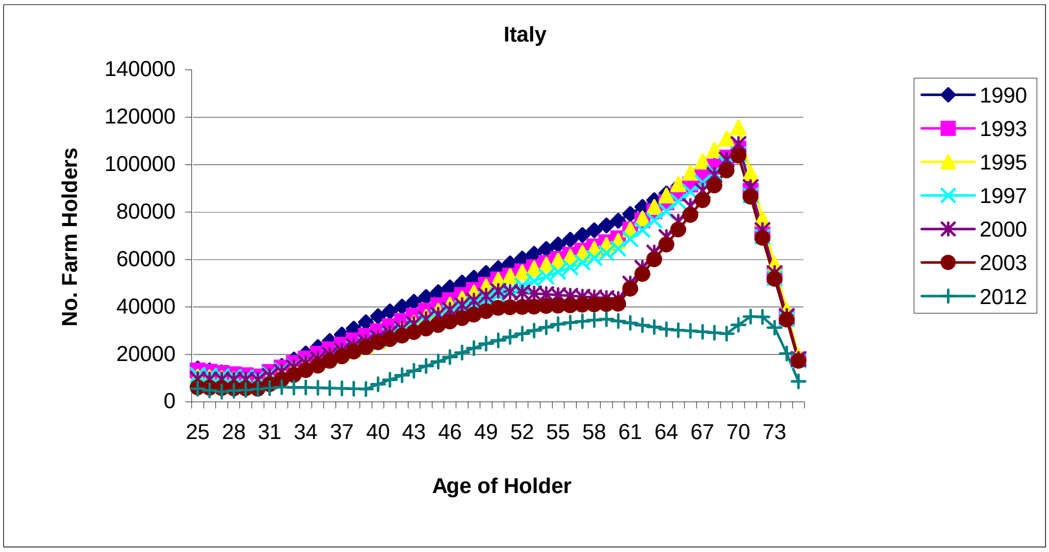

Greece
No. Farm Holders
250000
200000
150000
100000
50000
0
25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75
Age of Holder
2012
2003
2000
1997
1995
1993
1990

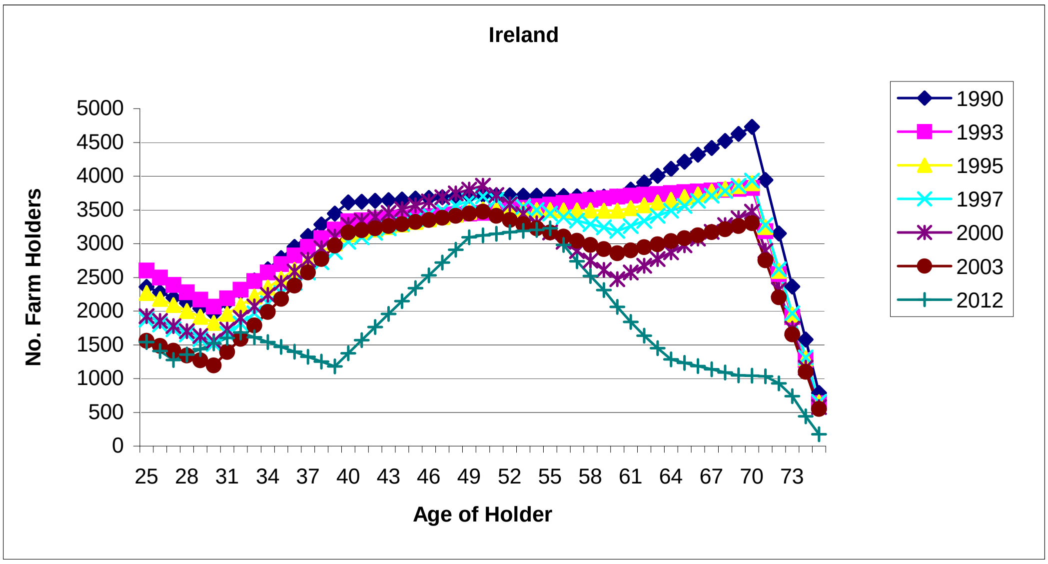

Finland
No. of Farm Holders
4000
3500
3000
2500
2000
1500
1000
500
0
25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73
Age of Holder
1995
1997
2000
2003
2012


France
No. Farm Holders
35000
30000
25000
20000
15000
10000
5000
0
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
Age of Holder
1990
1993
1995
1997
2000
2003
2012


Sweden
No. Farm Holders
3000
2500
2000
1500
1000
500
0
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
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[22].

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[23]. For Germany, regional deaths rates were taken from the National Statistics Agency[24] 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 1995-2000 period. The Eurostat variable we use is "full-time/part-time male employment by

---

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

[23] 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))

[24] www.regionalstatistik.de

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)[25]. 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:

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

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 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

---

[25] The FADN code is SE425

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.

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.

2.6 The world Data Base (Andrea Zintl)

The global data base of CAPRI comprehends macro-economic data for different world regions, policy data and global agricultural production data. Several data sources can be mentioned:

- Data on bilateral trade between the CAPRI world regional aggregates are mainly relying on FAOSTAT.
- 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).
- 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).
- International future price developments are currently adopted from FAPRI or ESIM projections.

In 2007 the international data base program (global.gms) was expanded to include palm oil (PLMO). Data on palm oil were not added to COCO/CAPREG, but rather introduced ad hoc in the data preparation step for the market model. That required some code changes, as so far, the data on market balance positions for regions covered by CoCo/CAPREG were not taken directly from the FAO data base. This solution is likely to change.

2.6.1 Update note on the 2008 global database

While updating the CAPRI base year to 2004 (2003-05) some problems were encountered. The last FAOSTAT1 dataset for years 1986 to 2003 offered a list of products which ensures a complete mapping to all CAPRI products, including new products whey powder and casein after the following adaptations:

- Production of cakes from oilseeds is estimated from data on production of oils.
- Whey powder is estimated from data on total fresh whey
- Missing production of casein is calculated as demand minus net imports.

However, for years 2004 to 2005 alternative datasets had to be used as FAOSTAT1 is discontinued. For this purpose the relative changes to 2002 are calculated in the alternative datasets. The old FAO selection is then extrapolated using these changes. Year 2003 from FAOSTAT1 has also been re-estimated because it was likely to include many preliminary and missing values.

The first alternative dataset came from the new FAOSTAT2 website[26]. Basically the problem with this data set was that FAOSTAT2 data do not offer processed products anymore, in

---

[26] Due to internal revisions the FAOSTAT2 database is not available anymore from the website, but fortunately the data have been downloaded by IPTS already.

particular vegetable oils, oil cakes, and dairy products. For all other products the data for the years 2003 to 2005 were updated as described above.

A second alternative database had to be used therefore which is the AGLINK modelling database. Unfortunately AGLINK does not offer the same level of detail (single country data) as FAOSTAT2 for a number of relevant regions. Hence an additional regional mapping was necessary before the database update could be completed but some CAPRI regions could be mapped only approximately.

Furthermore AGLINK does not offer complete market balances for all products, which also applies to those needed. In case of missing demand side data these had to be extrapolated for years 2003 to 2005 from the changes of corresponding production quantities therefore. The data situation for oilseeds in AGLINK is even more troublesome. The complete balance is only available for the aggregate oilseeds. For the single seeds only production data can be directly used, whereas all other balance positions have to be estimated from the development of the aggregate oilseeds. Even for total oilseed oils only production data are available. For the cakes there are at least the aggregate data on total "oilseed meal" imports, exports and consumption, which is assumed to be for feed.

Less problematic were other updates: For the data on GDP and consumer expenditures we switched to the UNSTATS database (United Nation statistical Division)[27] rather than using FAO as previously. The new bilateral trade matrix (including data up to 2005) was estimated based on an FAOSTAT2 dataset in an earlier study and offers data for the whole time horizon required.

2.6.2 Update note on international policy variables

Description and current status of international policy variables needs an update.

---

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

3 Baseline Generation Model (CAPTRD) (Wolfgang Britz)

The aim of the CAPRI projection tool is to provide a baseline used as comparison points 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.

This section explains in detail the methodology used in CAPRI to construct a baseline. Before entering into these 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. These trends are supplemented in the CAPRI baseline tools with results from other baselines, especially from DG-AGRI.

The projection tool is fed both by forecasts from different experts or modelling tools, as well by trend forecasts using data from the 'COCO' database[28] as ex-post information. The purpose of these trend estimates is, on the one hand, to compare expert forecasts with a purely technical prolongation of time series and, on the other hand, to provide a 'safety net' position in case no values from external projection are available. Therefore, trend variables for baseline generation in the model are mainly constructed out of expert data on projections (e.g. FAO, European Commission or World Bank) and linear trends of data contained in the CAPRI data base. These trend variables are simultaneously subject to the consistency restrictions imposed by the mathematical programming model and not made as independent forecasts for each time series (e.g. closed area and market balances). The resulting estimator is hence a system estimator under constraints whose properties are discussed in the following section. 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[29].

3.1 Trend curve

The first ingredient in the estimator is the trend curve itself which is defined as:

Equation 17 $$X_{r,i,t}^{j,Trend} = a_{r,i,j} + b_{r,i,j}\ t^{c_{r,i,j}}$$

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 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.

In a first prototype of the module, a polynomial trend curve of degree two was evaluated. However, a quadratic function is not necessarily monotone on the forecast interval so that a trend curve may for example show increasing yields for the first part of the projection period and afterwards a decrease. As such outcomes are purely technical and not motivated by a priori knowledge, it was deemed more plausible to switch to the formulation shown above with the same number of free parameters as a quadratic trend curve, but with monotony guaranteed.

The ex-post period covers the period from 1985 towards 2000. In order to cut down the size of the resulting problem, the ex-ante period is defined in ten years steps (2003, 2010, 2020, 2030), as intermediate years can be simply calculated once the estimated parameters are known.

---

[28] Britz, W., Wieck, C., Jansson, T. (2002): National framework of the CAPRI-data base - the COCO – Module, CAPRI Working Paper 02-04, Institute of Agricultural Policy, Bonn.

[29] The only exception is the quota regime on the milk market which has been recognised in the trend projections in that the milk production has been derived from the quota endowments (where current quotas are assumed to persist until 2025).

3.2 Consistency constraints in the trend projection tool

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 can be seen as a minimum set of consistency conditions necessary for a projection of agricultural variables. As discussed above in detail, the full projection tool features further constraints especially relating to price feedbacks on supply and demand.

3.2.1 Constraints relating to market balances and yields

Closed market balances 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 18

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

Where *r* are the Member States of the EU, *i* are the products, *t* the different forecasting years. All elements of the market balances are expressed as primary product equivalents according to the concept of ‘supply utilization accounts’. Human consumption of wheat does hence include floor, bread, pasta etc. recalculated into what equivalent based on conversion factors. The only expectations are oilseeds, where processing to cakes and oils is explicitly covered, and raw milk, where again, processing to the different dairy products is included explicitly.

Secondly, production (*GROF*) is equal to yield times area/herd size (*LEVL*) where *acts* are all production activities:

Equation 19

$$X_{r,i,t}^{GROF,Trend} = \sum_{acts} X_{r,i,t}^{acts,Trend} X_{r,LEVL,t}^{acts,Trend}$$

A set of equations relates to the hectares for groups of crop activities (cereals, oilseeds, industrial crops, vegetables, fresh fruits, total vineyards, fodder production on arable land). It defines e.g. that the total hectares of cereals is equal to the sum of hectares for the individual cereals as soft wheat, durum wheat, barley and so forth.

Equation 20

$$X_{r,LEVL,t}^{crop\_grp,Trend} = \sum_{j \in crop\_grp} X_{r,LEVL,t}^{j,Trend}$$

Equally, the market balance positions for certain products enter adding up equations for groups of products (cereals, oilseeds, industrial crops, vegetables, fresh fruits, total vineyards, fodder production, meat). As an example, total cereal production is equal to the sum over the produced quantities of the individual cereals.

Equation 21

$$X_{r,pro\_grp,t}^{MrkBal,Trend} = \sum_{i \in pro\_grp} X_{r,i,t}^{MrkBal,Trend}$$

3.2.2 Constraints relating to agricultural production

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

Equation 22 $$X^{UAAR,Trend}_{r,LEVL,t} = \sum_{crops} X^{crops,Trend}_{r,LEVL,t}$$

Further constraints link the different animal activities over young animal markets:

Equation 23 $$X^{GROF,Trend}_{r,oyani,t} = \sum_{iyani \leftrightarrow oyani} X^{GROF,Trend}_{r,iyani,t}$$

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 used as inputs in the other animal processes (fattening, raising or milk producing).

Finally, balances for energy and protein requirements for each animal type *maact* are introduced as:

Equation 24 $$\sum_{feed} X^{maact,Trend}_{r,feed,t} X^{Cont,Trend}_{r,feed,t} = 0.996^{t} \left(a^{Cont}_{maact} + a^{Cont}_{maact} X^{maact,Trend}_{r,yield,t} \right)$$

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 basis period. The factor in front of the requirements introduces some input saving technical progress of -0.4% 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 (grass, maize silage, fodder root crops, straw, milk for feeding, other fodder from arable land):

Equation 25 $$X^{FEDM,Trend}_{r,feed,t} = \sum_{maact} X^{maact,Trend}_{r,feed,t} X^{maact,Trend}_{r,"levl",t}$$

Finally, the feed use of individual products must add up to the feed use of the 'bulks' mentioned above:

Equation 26 $$X^{FEDM,Trend}_{r,feed,t} = \sum_{o \to fed} X^{FEDMt,Trend}_{r,o,t}$$

3.2.3 Constraints relating to prices, production values and revenues

The check of external forecasts revealed that for some products, price projections are not available. It was decided to include prices, value and revenues per activity in the constrained estimation process. The first equation 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 27 $$X^{EAAG,Trend}_{r,i,t} = X^{GROF,Trend}_{r,i,t} X^{UVAG,Trend}_{r,i,t}$$

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

Equation 28
$$X^{acts,Trend}_{r,TOOU,t} = \sum_{o} X^{acts,Trend}_{r,o,t} \; X^{UVAG,Trend}_{r,o,t}$$

As for the market balances, the values for certain aggregate product groups must add up:

Equation 29
$$X^{EAAG,Trend}_{r,pro\_grp,t} = \sum_{i \in pro\_grp} X^{EAAG,Trend}_{r,i,t}$$

Consumer prices (UVAD) are equal to producer prices (UVAG) plus a margin (CMRG):

Equation 30
$$X^{UVAD,Trend}_{r,i,t} = X^{UVAG,Trend}_{r,i,t} + X^{CMRG,Trend}_{r,i,t}$$

3.2.4 Constraints relating to consumer behaviour

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

Equation 31
$$X^{HCOM,Trend}_{r,i,t} = X^{INHA,Trend}_{r,i,t} \; X^{INHA,Trend}_{r,LEVL,t}$$

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

Equation 32
$$X^{EXPE,Trend}_{r,i,t} = X^{INHA,Trend}_{r,i,t} \; X^{UVAD,Trend}_{r,LEVL,t}$$

As for the market balances, the per caput expenditure (EXPE) for certain aggregate product groups – including an aggregation over all products - must add up:

Equation 33
$$X^{EXPE,Trend}_{r,pro\_grp,t} = \sum_{i \in pro\_grp} X^{EXPE,Trend}_{r,i,t}$$

3.2.5 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 to processing of primary products (PRCM) by processing yields (PRCY):

Equation 34
$$X^{MAPR,Trend}_{r,sec,t} = \sum_{i \vee sec \leftarrow i} X^{PRCM,Trend}_{r,i,t} \; X^{PRCY,Trend}_{r,sec,t}$$

In case of products from derived milk (*mlkseco*) – butter, skimmed milk powder, cheese, fresh milk products, cream, concentrated milk and whole milk powder -, fat and protein content (MLKCNT) of the processed milk (COMI – cow milk, SHGM – sheep & goat milk) must be equal to the content of the derived products:

Equation 35
$$\begin{aligned} & X^{PRCM,Trend}_{r,COMI,t} \; X^{MLKCNT,Trend}_{r,COMI,t} + X^{PRCM,Trend}_{r,SHGM,t} \; X^{MLKCNT,Trend}_{r,SHGM,t} \\ & = \sum_{mlkseco} X^{MAPR,Trend}_{r,mlkseco,t} \; X^{MLKCNT,Trend}_{r,mlkseco,t} \end{aligned}$$

3.2.6 Constraints relating to policy

There are two constraints: firstly, the acreage under compulsory set-aside must be equal to the set-aside obligations of the individual crops:

Equation 36
$$X^{"OSET",Trend}_{r,"levl",t} = \sum_{cact} X^{cact,Trend}_{r,"levl",t} \; \frac{0.01 \; X^{cact,Trend}_{r,"setr",t}}{\left(1 - 0.01 \; X^{cact,Trend}_{r,"setr",t}\right)}$$

Secondly, milk production is fixed to the milk quota, modified by eventual under- or over-deliveries in the base year.

3.2.7 Constraints relating to growth rates

During estimation, some safeguards regarding the size of the implicit growth rates had been introduced:

- Total agricultural area is not allowed to decline at a rate exceeding -0.5 % per annum.
- 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 cereals consumption, the growth rate here 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.
- Changes in prices are not allowed to exceed a growth rate of +/- 2% per annum.
- The number of calves born per cow is not allowed to exceed a range of +/- 10 % 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.
- Strong increases in pork production in the past are restricted by environmental legislation in force, notably the nitrate directive. Accordingly, increases were restricted to +1% for EU15 Member States (+0.5% for Denmark and The Netherlands) per annum.
- Milk yields per dairy cows were restricted by an upper bound of 12.000 litres per cow and year.
- Shares of arable crop on total arable area are bounded by the formula which allows small shares to expand or shrink more compared to crops with a high share. A crop with a base year share of 0.1% is allowed to expand to 2.5%, one of 10% only to 25%, and one of 50% to only 70%:

$$X_{r,\text{"level"},t}^{arab,Trend}.up/lo = X_{r,\text{"level"},bas}^{arab,Trend}$$

$$\pm 1/4 \left(\frac{X_{r,\text{"level"},bas}^{arab,Trend}}{X_{r,\text{"level"},bas}^{\text{"arab"},Trend}} \right)^{1/4} \; X_{r,\text{"level"},bas}^{\text{"arab"},Trend} \; \max\left(0.2, \frac{t-bas}{last-bas}\right)$$

Equation 37

3.3 Three-stage procedure for trends

The estimation process is a two-stage procedure, where results from previous steps feed into the current on.

3.3.1 Step 1: Unrestricted trends

The first stage estimates unrestricted trend curves. The optimal values of the estimated trend parameters *a*, *b* and *c* are defined by minimizing squared errors normalized with the mean of the time series (for technical reasons, solely), using the trend as weights:

Equation 38

$$SSQ = \sum_{r,i,j,expost} \left(\frac{X^{j,\text{"Data"}}_{r,j,expost} - a_{r,i,j} + b_{r,i,j} t^{c_{r,i,j}}_{expost}}{X^{j,\text{"Data"}}_{r,i,mean}} \right)^2 t_{expost}$$

The weighting with the trend was introduced after a careful analysis of the results of the first step. 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, is moves the centre of gravity of the estimation in direction of the base period which is used as a kind of fallback position the worse the fit of the above equation.

The resulting parameters provide firstly a starting point for the constrained estimations. Secondly, the variance of the resulting error terms defines the weights for the next two steps. And thirdly, the trend estimate together with R^2 from that first step is used to define the 'support point' for the next steps:

Equation 39

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

The support point is hence the weighted average of the trend forecast and the base year values, defined as a five year average around 1998 -2002.

3.3.2 Step 2: Constrained trends at Member State level

The second step adds the consistency conditions discussed above. In almost all cases, the unrestricted trend estimates from the first step would violate one or several of the consistency conditions. We need hence now to find estimates which both fit into the consistency constraints and exploit in a technical feasible way the information comprised in the ex-post development. Take the second type of consistency constraints as an example, which defines production as hectares/herd sizes times yield. Clearly, we would like our ex-ante trend estimates to fulfil that condition. However, 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 acreage. However, by doing so, we deliberately throw away the information comprised in the development of barley yield over time. Adding the kind of definitional relations between the time series does hence help us to exploit more information than is comprised in single series, and refrains from throwing away ex-ante parts of the information available.

However, when estimating simultaneously the different trends, we need to reflect if the sum of squares (SSQ) as a penalty function still works reasonable. A nice property is the fact that strong trends – i.e. such with a high explanatory power – will dominate weak ones. However, as our last forecasted point is far away from the mean, changing slightly the parameters could lead to drastic differences in the estimates without a sizeable effect especially on the SSQ when it is already small. Especially shaky trends will show values at the tails which can be far away from those observed ex-post. We need hence a safeguard which draws our estimates to a 'reasonable' value in such cases.

The confidence interval from the trend estimate will not help, as it will be centred around the tail value and simply be quite large for bad R^2. However, we may use the argumentation 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 strongly 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.

The reasoning behind the test statistics is the basis for the supports defined above. We modified it however to match the problem at hand. 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 it would simply be impossible to analyze manually each and every trend line and decide upon an alternative estimation. The continuous rule draws the estimates stronger in direction of our H0 – the value is equal to the three year average around the last known points – the shakier the estimated parameters are.

The resulting penalty function is defined as minimization of the squared deviations from the supports defined above, weighted with the variance of the error terms from the first step:

Equation 40

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

The value used by that penalty function for each time point consists hence of two elements:

(1) the difference between the trend estimate fitting into the consistency conditions and the supports derived from the unrestricted trends, and

(2) the variance of the error terms from the trend estimates.

For all unrestricted trend lines, the mean error will be zero so that it cannot be used as a criterion. Instead, the variance of the error term is used as a measurement for the magnitude of the error terms. It is decreasing with the mean of the explanatory variable and with a better fit of the trend curve. Normalizing with the variance of the error terms will hence ensure that relative rather than absolute deviations are penalized, and that deviations from the support are penalized stronger where the trend had a high explanatory power.

How is the first element of the term motivated, i.e. the squared difference between the restricted trend estimates and the supports? If R^2 for a certain time series is 100%, the penalty is defined as the squared difference between the restricted trend estimate and the unrestricted one (see definition of the support above). In other words: for a perfect fit, the restricted trend estimate is drawn towards the unrestricted trend estimate.

If R^2 is zero, and the trend curve does not explain any of the variance and the probability for (a,b,c) being equal to zero becomes maximal. Consequently, we let the solver find the minimal squared difference between the 'base data' points and the restricted trend estimate as the support becomes equal to the 'base data'. The 'base data' represent a three-year average around the last three known years.

For all cases in between, we minimize squared difference from the weighted average of the unrestricted trend estimate weighted with R^2 and the three-year average weighted with $(1-R^2)$. The weights ensure that deviations for lines with a secure unrestricted fit are smaller than for time series with more shaky trends. Generally, all trend estimates are restricted to the non-negative domain.

For selected variables, instead of using solely the mechanistic corridors shown above, additional estimations corridors had been introduced as discussed above.

Originally, it was foreseen to add a third step where aggregation to EU level should be added as an additional layer of information, with some elements as net trade and imports/exports not planned to be included in the estimation step at Member State level. However, during the development of the tool, the number of simultaneously estimated items and their relations captured by the constraints increased so that an integration of the individual Member State modules into one framework with additional adding up constraints to EU level became technically not longer feasible. Instead, the elements planned to be solely included in the EU aggregation step, namely the positions relating to net trade, where added to the individual Member State modules.

3.3.3 Step 3: Adding supports based on external results and breaking down to regional level

In the final estimation step, results from external projections on market balance positions (production, consumption, net trade etc.) and on activity levels are added. Currently, these projections are provided by DG-AGRI. As DG-AGRI is the main client, it is deemed sensible to force the projections to comply with the DG-AGRI baseline wherever the constraints of the estimation problem allow for it. That is achieved by two changes to the objective function:

1. Supports are replaced by the results of DG-AGRI baseline, the latter proportionally scaled so that results from the DG-AGRI baseline and the CAPRI data base are identical.
2. Deviations against DG-AGRI results are weighted 100 times higher as trend based supports.

Accordingly, the Step 3 objective function is defined as:

Equation 41

$$Penalty = \sum_{r,i,j,exante \wedge \neg X_{r,i,exante}^{j,"DG-AGRIt"}} \left(\frac{X_{r,j,exante}^{j,"Trend"} - X_{r,i,exante}^{j,"Support"}}{\sqrt{X_{r,i,"varErr"}^{j,"Step1"}}} \right)^2 + \sum_{r,i,j,exante \wedge X_{r,i,exante}^{j,"DG-AGRIt"}} \left(\frac{X_{r,j,exante}^{j,"Trend"} - X_{r,i,exante}^{j,"DG-AGRIt"}}{\sqrt{X_{r,i,"varErr"}^{j,"Step1"}} * 100} \right)$$

The results at Member State level are then broken down to regional level, ensuring adding up of areas and production:

Equation 42

$$X_{MS,i,t}^{GROF,Trend} = \sum_{r \in MS} X_{r,i,t}^{GROF,Trend}$$

Equation 43

$$X_{MS,"levl",t}^{j,Trend} = \sum_{r \in MS} X_{r,"levl",t}^{j,Trend}$$

3.3.4 Breaking down results from Member State to regional 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 unfeasible. 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. There are only four restrictions active:

- The set-aside obligations at regional levels
- Adding up of regional areas to Member State areas
- Adding up of regional production to Member State production
- Adding up crop activities to utilisable agricultural area.

In order to keep developments at regional and national level comparable, relative changes in activity levels are not allowed to deviate more then 50% from the national development, in case of yields, development is bounded to a +/-20% range relative to the national one. These bounds are softened in cases of infeasibilities.

3.3.5 Update note on CAPTRD: biofuels and other issues

The definition of expert "supports" now supports the provision of a mean and a standard deviation, so that it is possible to use also weaker expert supports.

In 2007 CAPTRD has been discussed extensively in the CAPRI network which led to some technical streamlining and debugging. Furthermore several changes were needed to obtain a reasonable biofuels baseline.

The projection engine was extended by three equations:

- An equation defining the output from by-products from milling, brewing and sugar beet processing
- An equation for biofuel production based on the extraction rates given above
- And equation defining output of gluten feed from bioethanol production.

Equally, the program matching the ESIM codes with the CAPRI data world (captrd\esim_map_sets.gms) were expanded to cover bioethanol, biodiesel and palm-oil and to include processing to biofuels. Naturally, the new market balance position for processing to biofuels (BIOF) and the new product Gluten Feed (GLUE) was added to the program as well.

As the definitions in ESIM in CAPRI are not fully harmonized, it is necessary to define scaling factors, and in order to stabilize those, the times series from CAPRI and ESIM should overlap for some years. However, at current state, ESIM provides data only for 2004 onwards, whereas the CAPRI time series end in 2003. Therefore, the time series from ESIM are backcasted. The original idea to use ex-post data from ESIM for 2004 for the backcasting was dropped, as the changes between the two time points are often dramatic, even for position as human consumption which is typically very stable. The large differences between 2004 and 2005 probably hint a definition problem. So could the position be calculated residually, and as stock changes are not reported elsewhere, could comprise those. As an intermediate remedy, the base year results – which seems to be an average 2004-2005 – were copied to the year 2005 (baseline\load_esim.gms) and then used in conjunction with the model results for 2007 to backcast for the years 2001 to 2003, the current base period of CAPRI.

Given the exponential increase of biofuel production in the last years, the rather conservation backcasting algorithm above does not work too well and leads especially for biofuel production and the related demand for raw products to a data constellation where the ESIM base year data are by some 20-30% above the CAPRI base year values, so that also the projections from CAPRI for bioethanol are below the ESIM data. However, that is probably not dramatic, as the estimated share for biofuels in the projection is anyway based on assumptions.

There were some further changes necessary to get acceptable results for biofuel processing. Firstly, as no breakdown of processing is delivered by ESIM, the following code pieces were added: (captrd\scale_DG_AGRI_baseline.gms). A first block ensure that the possible high growth rates for the processing position are not also applied to the position industrial processing (INDM) from the CAPRI data base, by fixing it to base year values. A last block had been added as a very simple remedy to get some sizeable bioethanol production in the baseline. As reported above, without those statements, total output at EU27 level as just 0.5 Mio t, which led to infeasibilities when trying to shift it to target values. The infeasibilities root in the fact that by-product as gluten feed are coupled with a fixed factor to the fixed quantities used for biofuel production. Increasing biofuel processing by factor 20 hence also increases the output of gluten feed by factor 20. As gluten feed can only be used for feed in the current version, and the feed demand function are linear in prices, no acceptable feed demand elasticity would be able to produce a slope for the feed demand to digest an increase of factor 20.

There were several other problems noted linked to usage of external supports which deviate to a large extent from past development. Three major problems have to be fixed before larger processing quantities were found in the projection result set. Firstly, in early versions no changes in the a priori distribution for the production and feed demand of gluten feed were introduced, and a rather stable demand and production ex-post exerted a strong pull towards the base year values. As a remedy, the second moment for the positions relating to gluten feed was drastically increased. Secondly, it turned out favourable to let the output value for the biofuels (bioethanol and biodiesel) drive the agricultural inputs, by reducing their second moments. And thirdly, in case biofuel demand accounted for a larger share of market appearances, the second moment for the imports and exports was relaxed (captrd\define-EU-support).

3.4 Calibrating the model to the projection

3.4.1 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 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, and (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. Further on, the feed mix was only projected at Member State, not at NUTS 2 level.

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 the bulks as cereals, fits in the requirement constraints and leads to plausible feed cost. In order to do so, the feed allocation framework is re-used. 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. The resulting parameters are as well stored in external files to be reloaded in case of counterfactual runs.

3.4.1.1 Update note

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.

3.4.2 Calibrating the global trade model

The projection results at EU25 level plus Norway, 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 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. These developments are currently almost exclusively based on projections by the FAO more precisely the projections from the @2030 model.

4 Simulation Scenario Model (CAPMOD)

4.1 Overview of the system

The CAPRI simulation tool is composed of a supply and market modules, interlinked with each other.

In the *supply module*, regional agricultural supply of annual crops and animal outputs is modelled by an aggregated profit function approach under a limited number of constraints: land, 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). 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) [30] 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 28 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 and supply and a generalised Leontief expenditure function for human consumption) and some further restrictions (homogeneity of degree zero in prices, symmetry and correct curvature). Accordingly, the demand system allows for the calculation of welfare changes for consumers, processing industry and public sector. Policy instruments in the market module include

---

[30] 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)

bilateral tariffs and producer or consumer subsidy equivalent price wedges (PSE/CSE)[31]. Tariff rate quotas (TRQs), intervention purchases and subsidised exports under the World Trade Organisation (WTO) commitment restrictions are explicitly modelled for the EU 15.

In the market module, special attention is given to the processing of dairy products in the EU. 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.

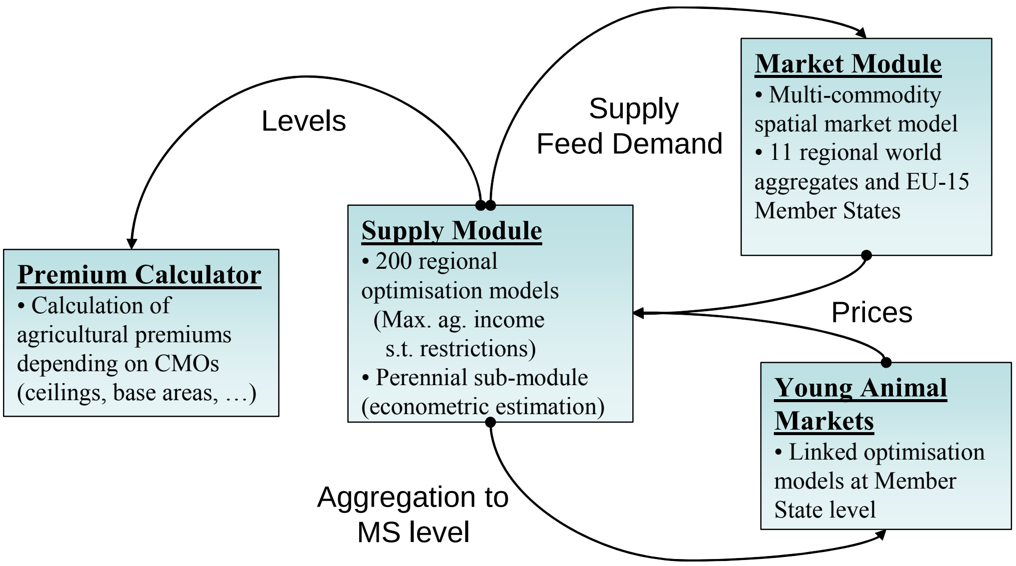
The market module comprises of a bilateral world trade model based on the Armington assumption (Armington, 1969). According to Armington's theory, the composition of demand from domestic sales and different import origins depends on price relationships according to bilateral trade flows. This allows 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) 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 to Member State level models, which are then calibrated using Positive Mathematical Programming (PMP) estimation techniques. Young animal prices are determined by linking these calibrated Member State models into a non-spatial EU trade model with market balances for young animals, as shown in Figure 5. 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.

---

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

Figure 5. Link of modules in CAPRI



Source: CAPRI Modelling System

4.1.1 Update note

The 'premium calculator' has grown in complexity in the last years to reflect recent changes in the CAP (Single Farm Payment, Simplified Area Payment Scheme and Top Ups in NMS, Article 69 payments). It is so far following the historical origin of premiums starting from Agenda 2003 values, then considering updates such as modulation and partial or complete decoupling. This procedure is currently reconsidered.

4.2 Module for agricultural supply at regional level

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 not comprise inter-activity terms, i.e. no marginal cross-cost effects, 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 44
$$Rev_j = Cost_j + ac_j + bc_j Levl_j + \sum_{i=1}^{m} \lambda_i a_{ij}$$

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 ($ac_j + bc_j Levl_j$) shows the marginal non-linear costs, these marginal costs are increasing in the activity levels. 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 ***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 ($ac_j + bc_j Levl_j$) 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.

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	-				<= UAAR
Set aside	+/-				= 0
Quotas	-	-			<= Ref. Quantity
Fertilizer needs	-	+		+	= 0
Feed requirements		-	+	+	= 0

Source: CAPRI modelling system

4.2.1.1 Update note

Some years ago already, the objective function has been generalised from the classical PMP approach to include a full matrix of cross activity effects.

4.2.2 Detailed discussion of the equations in the supply model

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 45 $$\overline{AREQ}_{r,acct,req}\ \overline{DAYS}_{r,aact} \leq \sum_{feed} FEDNG_{r,acct,feed}\ \overline{REQCNT}_{r,feed,req}$$

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 46
$$\overline{\mathrm{AREQ}}_{\mathrm{r,acct,"DRMA"}}\ \overline{\mathrm{DAYS}}_{\mathrm{r,aact}}\ \overline{\mathrm{MAXSHR}}_{\mathrm{r,acct,feed}} \geq \mathrm{FEDNG}_{\mathrm{r,acct,feed}}\overline{\mathrm{REQCNT}}_{\mathrm{r,feed,"DRMA"}}$$

Equation 47
$$\overline{\mathrm{AREQ}}_{\mathrm{r,acct,"DRMA"}}\ \overline{\mathrm{DAYS}}_{\mathrm{r,aact}}\ \overline{\mathrm{MINSHR}}_{\mathrm{r,acct,feed}} \leq \mathrm{FEDNG}_{\mathrm{r,acct,feed}}\overline{\mathrm{REQCNT}}_{\mathrm{r,feed,"DRMA"}}$$

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 48
$$FEDUSE_{r,feed} = \sum_{aact} LEVL_{r,aact} FEDNG_{r,aact,feed}$$

Land balances and set-aside restrictions

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

Equation 49
$$\overline{LEVL}_{r,"arab"} = \sum_{arab} LEVL_{r,arab}$$

Equation 50
$$\overline{LEVL}_{r,"gras"} = LEVL_{r,"grae"} + LEVL_{r,"grai"}$$

Both land balances must be exhausted. For arable land, idling land not in set-aside (activity *FALL*) is an explicit activity which closes the balance. 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 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 51
$$LEVL_{r,"oset"} = \sum_{arab} LEVL_{r,arab} \frac{{}^{1}\!/_{100}\, SETR_{r,arab}}{1 - {}^{1}\!/_{100}\, SETR_{r,arab}}$$

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.

The equation above implies that non-food production on set-aside takes by assumption place on voluntary set-aside, rendering the analysis of model results easier, with no practical consequences for simulation results.

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 52

$$LEVL_{r,"oset"} + LEVL_{r,"vset"} LEVL_{r,"nonf"} \leq \sum_{arab \wedge SETR_{r,arab}} LEVL_{r,arab} \Big/ \overline{MXSETA}$$

Fertilising block

The equation below is 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.

Equation 53

$$\begin{aligned}
&\sum_{cact} Levl_{r,cact} \left(Fnut_{r,cact} \right) \left(1 - NFact^{biofix}_{Fnut,cact} \right) NutFac_{r,fnut} \\
&= -NETTRD^{Fnut}_{r} (1 - NH3Loss^{Anog}_{Fnut,r}) \\
&+ NBal^{AtmDep}_{r} NFact^{AtmDep}_{Cact} \\
&+ \sum_{aact} Levl_{r,aact} Fnut_{r,aact} (1 - NH3Loss^{Manure}_{Fnut,r}) \left(1 - NavFac_{r,fnut} \right) \\
&+ Losses_{r,fnut}
\end{aligned}$$

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

Equation 54

$$\sum_{cact} Levl_{r,cact} \left(Fnut_{r,cact} \right) NutFac_{r,fnut} MINAN_{r,cact,fnut} \leq NETTRD^{Fnut}_{r}$$

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 55

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

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 56

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

In combination with the standard balancing equation shown above, the NETTRD variable for young animals on the output side becomes negative if the YANUSE variable for a certain type of young animals exceeds the production inside the region.

The objective function

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

Equation 57 $$OBJE = \sum_{r} LINEAR_{r} + QUADRA_{r} + QUADRF_{r}$$

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 58 $$\begin{aligned} &\mathrm{LINEAR}_{\mathrm{r}} \\ &= \sum_{\mathrm{io}} \mathrm{NETTRD}_{\mathrm{r,io}} \overline{\mathrm{PRICE}}_{\mathrm{,io}} + \sum_{\mathrm{act}} \mathrm{LEVL}_{\mathrm{r,act}} \left(\overline{\mathrm{PRME}}_{\mathrm{r,act}} - \overline{\mathrm{COST}}_{\mathrm{r,act}} \right) \end{aligned}$$

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

Equation 59 $$\mathrm{QUADRF}_{\mathrm{r}} = \sum_{\mathrm{aact,feed}} \begin{bmatrix} \mathrm{LEVL}_{\mathrm{r,aact}} \mathrm{FEDNG}_{\mathrm{r,aact,feed}} \\ \left(\mathrm{a}_{\mathrm{r,acct,feed}} + \tfrac{1}{2} \mathrm{b}_{\mathrm{r,aact,feed}} \mathrm{FEDNG}_{\mathrm{r,aact,feed}} \right) \end{bmatrix}$$

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

Sugar beet

The current Common Market Organisation (CMO) for sugar regulates European sugar beet supply with a system of production quotas. Two different quotas are established subject to different price guarantee (A and B quotas, qA and qB). Sugar beets produced beyond those quotas (so called C beets) are sold as sugar on the world market at prevailing prices. The CAPRI system features an expected profit maximisation framework that cares for yield uncertainty as developed by 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 dependant on the farmers' incentive to fulfil their quota rights even in case of a bad harvest.

Regional sugar beet quotas are defined based on a FADN analysis. Expected profit of sugar beet production is then represented by:

Equation 60 $$\begin{aligned} &\mathrm{SugbREV}_{\mathrm{r}} \\ &= \mathrm{p}^{\mathrm{A}} \mathrm{NETTRD}_{\mathrm{r,SUGB}} \\ &-\left(\mathrm{p}^{\mathrm{A}} - \mathrm{p}^{\mathrm{B}}\right) \begin{bmatrix} \left(1 - \mathrm{CDFSugb}\left(\mathrm{q}^{\mathrm{A}}\right)\right)\left(\mathrm{NETTRD}_{\mathrm{r,SUGB}} - \mathrm{q}^{\mathrm{A}}\right) \\ +\left(\sigma^{\mathrm{S}}\right)^{2} \mathrm{PDFSugb}\left(\mathrm{q}^{\mathrm{A}}\right) \end{bmatrix} \\ &-\left(\mathrm{p}^{\mathrm{B}} - \mathrm{p}^{\mathrm{C}}\right) \begin{bmatrix} \left(1 - \mathrm{CDFSugb}\left(\mathrm{q}^{\mathrm{A+B}}\right)\right)\left(\mathrm{NETTRD}_{\mathrm{r,SUGB}} - \mathrm{q}^{\mathrm{A+B}}\right) \\ +\left(\sigma^{\mathrm{S}}\right)^{2} \mathrm{PDFSugb}\left(\mathrm{q}^{\mathrm{A+B}}\right) \end{bmatrix} \end{aligned}$$

Where $\mathrm{PDFSugb}_{\mathrm{r}}$ and $\mathrm{CDFSugb}_{\mathrm{r}}$ are the probability res. cumulated density functions of the NETTRD variable with the standard deviation σ^{S}. σ^{S} is defined as $\mathrm{NETTRD}_{\mathrm{r,SUGB}}$ * $\mathrm{VCOF}_{\mathrm{r}}$, where the latter is the regional coefficient of yield variation estimated from FADN. $\mathrm{p}^{\mathrm{ABC}}$ 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 variable SugbREV_r substitutes for the expression $\text{NETTRD}_{r,io}\text{PRICE}_{io}$ (if io=SUGB) in Equation 58.

4.2.2.1 Update note

In 2007, econometric estimation for the cost functions were introduced which lead to cross-activity terms in the objective function.

Due to some changes, for example those referring to nutrient balancing, the above representation reflects the key elements but not the details of the current code.

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.

4.2.4.1 Update note

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 ongoing milk study (2007/08) will yield additional empirical evidence on marginal costs related to milk production which will be incorporated as well.

4.3 Market module for young animals

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. The latter seemed more promising, despite the fact it is computationally infeasible to link all regional models simultaneously.

Instead, the Input/Output coefficients and all other coefficients appearing in the constraints of the regional programming models are aggregated to Member State level using activity levels as weights.

The resulting models are hence structurally identical to the regional models and comprise a technology equal to the weighted average over all regions in that Member States. Due to the typical aggregation bias, these Member State models will however perform differently in a simulation from solving all regional models and then aggregating the results. More specifically, they will even not reproduce the solution obtained from the regional models at current prices.

In order to overcome the aggregation problem, the Member State models are calibrated using ideas borrowed from Positive Mathematical Programming to the current results from the regional models in any iteration. In order to do so, calibration bounds are introduced around the aggregated results for the activity levels and the feeding activities. Equally, a regionally weighted average for shadow prices of grassland, arable land and the milk quotas is calculated and added to the costs of the related production activities. Land balances and milk quotas are then removed from the model. The model is then solved

Afterwards, they are stacked together with a set of new equations representing market clearing conditions for young animals. The shadow prices of these constraints at the optimal solution then define the prices for young animals.

4.4 Market module for agricultural outputs

4.4.1 Overview on the market model

Whereas the outlay of the supply module has not changed a lot 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[32] system at Bonn University, it was still considered necessary to have an independent market module for CAPRI.

The CAPRI market module can be characterised as a recursive-dynamic, deterministic, partial, spatial, global equilibrium model for most agricultural primary and some secondary products, in total about 50 commodities. The recursive-dynamic aspect is currently only captured in a partial adjustment approach on the supply side. 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 into 60 countries or country aggregates, each 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. 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) 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 60 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, Western Balkan countries, and an aggregate of Bulgaria and Romania. All other countries or country aggregates are identical to a trade block in the model.

Policy instruments in the market module include (bi)lateral tariffs and Producer/Consumer Subsidy Equivalent price wedges (PSE/CSE). Tariff Rate Quotas (TRQs) are integrated in the

---

[32] 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.

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.

The EU interacts via trade flows with the remaining 25 regions 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 to render the analysis of results more easy. 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. As this is also likely to happen after a milk market liberalisation alternative solutions are currently tested in the CAPRI milk study of 2007/08.

The market model in its current layout comprises about 25.000 endogenous variables and the identical number of equations.

Regional disaggregation of the market module[33]

	Country/Country aggregate	Code	Components with own behavioural functions		In supply module ?
1.	European Union 15, broken down into Member States (Luxembourg aggregated with Belgium)	EU015000	AT000000 BL000000 DK000000 DE000000 EL000000 ES000000 FI000000 FR000000 IR000000 IT000000 NL000000 PT000000 SE000000 UK000000	Austria Belgium/Lux Denmark Germany Greece Spain Finland France Irland Italy Netherlands Protugal Sweden United Kingdom	Yes
2.	European Union 10, broken down into Member States	EU010000	CY000000 CZ000000 EE000000 HU000000 LT000000 LV000000 MT000000 SI000000 SK000000 PL000000	Cyprus Czech Republic Estonia Hungary Lithuania Latvia Malta Slovenia Slovakia Poland	Yes
3.	Norway	NO000000			Yes
4.	Bulgaria & Romania	BUR	BG000000 RO000000	Bulgaria Romania	Yes
5.	Turkey	TUR			No
6.	Morocco	MOR			No
7.	Other mediterranean countries	MED	TUN ALG EGY ISR	Tunisia Algeria Eqypt Israel	No
8.	Western Balkan countries	WBA	HR000000 CS000000 MO000000 KO000000 AL000000 BA000000 MK000000	Croatia Serbia Montenegro Kosovo Albania Bosnia & Herzegov. TFYR Macedonia	Yes
9.	Rest of Europe	REU			No
10.	Russia, Belarus & Ukraine	RBU			No
11.	United States of America	USA			No
12.	Canada	CAN			No
13.	Mexico	MEX			No
14.	Venezuela	VEN			No

---

[33] A detailed description can be found in: C. Tritten, B. Henry de Frahan, W.Britz (2001): Regionalisation of the Rest of the World Aggregate, CAPRI working paper 01-01, available on the project web site: http://www.agp.uni-bonn.de/agpo/rsrch/capstr/pap01-01.doc

	Country/Country aggregate	Code	Components with own behavioural functions		In supply module ?
15.	Argentina	ARG			No
16.	Brazil	BRA			No
17.	Chile	CHL			No
18.	Uruguay	URU			No
19.	Paraguay	PAR			No
20.	Bolivia	BOL			No
21.	Rest of South America	RSA			No
22.	Australia & New Zealand	ANZ			No
23.	China	CHN			No
24.	India	IND			No
25.	Japan	JAP			No
26.	Least developed countries	LDC			No
27.	ACP countries which are not least developed	ACP			No
28.	Rest of the world	ROW			No

Source: CAPRI modelling system

4.4.2 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 at the OECD or the set of models emerging from AgMemod.

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 60 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.

4.4.3 Behavioural equations for supply and feed demand

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 *ppri* 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 61

$$supply_{i,r} = as_{i,r} + \sum_{j} bs_{i,j,r} \frac{ppri_{j,r}}{p_{index.r}}$$

Supply curves for the EU Member States, Norway, Bulgaria and Romania 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.

The system for **feed demand** is structured identically. However, not producer prices, but raw product prices *arm1p* determined by the Armington top level aggregator drive feed demand *feed*, combined with changes in the supply of animal products weighed with feed use factors *w*:

Equation 62

$$feed_{i,r} = \left(af_{i,r} + \sum_{j} bf_{i,j,r} \frac{arm1p_{j,r}}{p_{index.r}} \right) \sum_{i} w_i \frac{supply_i}{supply_i^{cal}}$$

Feed use does hence proportionally increase if the supply of meat or milk is increased, and price changes drive substitution inside of the feed mix. It is planned to replace that system in the near future by explicit energy and protein requirement balances linked to energy and protein ‘shadow’ prices which will define then ‘feed incentive’ prices, as it is already realised for the fat and protein balances for dairy products.

As for supply, feed demand curves for the EU Member States, Norway, 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, by shifting the constant terms *af*.

4.4.4 Behavioural equations for final demand

The final demand functions are based on the following family of indirect utility functions depending on consumer prices *cpri* and per capita income y^{34} where *G* and *F* are functions of degree zero in prices (RYAN & WALES 1996) which will be defined below[35]:

Equation 63 $$U(cpri, y) = \frac{-G}{(y-F)}$$

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

Equation 64 $$PerCap_i = F_i + \frac{G_i}{G}\left(y - F\right)$$

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 65 $$F_r = \sum_i d_i \; cpri_i$$

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 64 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 (see fn.):

Equation 66 $$G = \sum_i \sum_j bd_{ij} \sqrt{cpri_i \; cpri_j}$$

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

Equation 67 $$G_j = \sum_i bd_{ij} \sqrt{cpri_i / cpri_j}$$

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 68 $$\begin{aligned} \sum_i PerCap_i \; cpri_i &= \frac{\sum_i G_i cpri_i}{G}\left(y - F\right) + \sum_i d_i cpri_i \\ &= \frac{G}{G}(y-F) + F = y \end{aligned},$$

---

[34] 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.

[35] 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.

and homogeneity is guaranteed by the functional forms as well. The expenditure function follows from rearranging Equation 63:

Equation 69 $$y = e(U, cpri) = F - \frac{G}{U}$$

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 70 $$hcom_{i,r} = pop_r\ perCap_{i,r}$$

4.4.5 Behavioural equations for the processing industry

Processing demand for oilseeds is modelled by using behavioural functions derived from a normalised quadratic profit function under the assumption of a fixed I/O relation between seeds, cakes and oils. Consequently, the processing demand *proc* depends on processing margins *procMarg* which are differences between the value of the outputs (oil and cake) per unit of oilseed processed and the value of the oilseed inputted:

Equation 71 $$proc_{i,r} = ac_{i,r} + \sum_j bc_{i,j,r} \frac{procMarg_{j,r}}{p_{index.r}}$$

where the processing margin is defined from the producer prices *ppri* and crushing coefficients derived from observed supply quantities as:

Equation 72 $$\begin{aligned} procMarg_{seed,r} = & -ppri_{seed,r} \\ & + ppri_{seed\rightarrow cak,r} \frac{supply^{bas}_{seed\rightarrow cak,r}}{supply^{bas}_{seed,r}} \\ & + ppri_{seed\rightarrow oil,r} \frac{supply^{bas}_{seed\rightarrow oil,r}}{supply^{bas}_{seed,r}} \end{aligned}$$

Finally, output of oils and cakes *supply* depends on the processed quantities *proc* of the oilseeds and the crushing coefficients:

Equation 73 $$\begin{aligned} supply_{cake,r} &= proc_{seed,r} \frac{supply^{bas}_{seed\rightarrow cak,r}}{supply^{bas}_{seed,r}} \\ supply_{oil,r} &= proc_{seed,r} \frac{supply^{bas}_{oil\rightarrow cak,r}}{supply^{bas}_{seed,r}} \end{aligned}$$

Special attention is given to the processing stage of **dairy products** for the EU Member states. 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 74 $$supply_{"milk",r}\ cont_{"milk",fp} = \sum_{mlk} supply_{mlk,r}\ cont_{mlk,fp}$$

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 75 $$\begin{aligned} &\mathrm{supplky}_{\mathrm{mlk,r}} = \mathrm{am}_{\mathrm{mlk,r}} \\ &+\sum_{\mathrm{j}} \mathrm{bm}_{\mathrm{mlk,j,r}} \left(\mathrm{ppri}_{\mathrm{j}} - \mathrm{cont}_{\mathrm{j,"fat"}}\mathrm{ppri}_{\mathrm{fat,r}} - \mathrm{cont}_{\mathrm{j,"prot"}}\mathrm{ppri}_{\mathrm{prot,r}}\right) \big/ \mathrm{p}_{\mathrm{index,r}} \end{aligned}$$

And lastly, prices of raw milk are derived from its fat and protein content valued with fat and protein prices and a processing margin.

4.4.5.1 Update note

In 2007 processing yields of oilseeds have been derived from a CET specification with low transformation elasticities (0.1), which turned out less prone to infeasibilities that the earlier Leontief form with perfectly fixed relationships. This does not affect the equations in Section 4.4.5 but renders the processing coefficients price responsive.

Equally, the market balances in the current version comprise exogenously given bio-fuel demand.

4.4.6 Trade flows and the Armington assumption

The *Armington*[36] *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 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 76 $$U_{i,r} = \alpha_{i,r}\left[\sum_{r1} \delta_{i,r,r1} M_{i,r,r1}{}^{-\rho_{r,i}}\right]^{-1/\rho_{r,i}}$$

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*:

---

[36] Armington, Paul S. (1969), A Theory of Demand for Products Distinguished by Place of Production, *IMF Staff Papers* 16, pp. 159-178

Equation 77 $$\frac{M_{i,r,r1}}{M_{i,r,r2}} = \left[\frac{\delta_{i,r,r1}}{\delta_{i,r,r2}} \frac{P_{i,r,r2}}{P_{i,r,r1}} \right]^{1/(1+\rho_{r,i})}$$

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 internal market price and the average import price. The lower stage determines the import shares from different origins. 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.

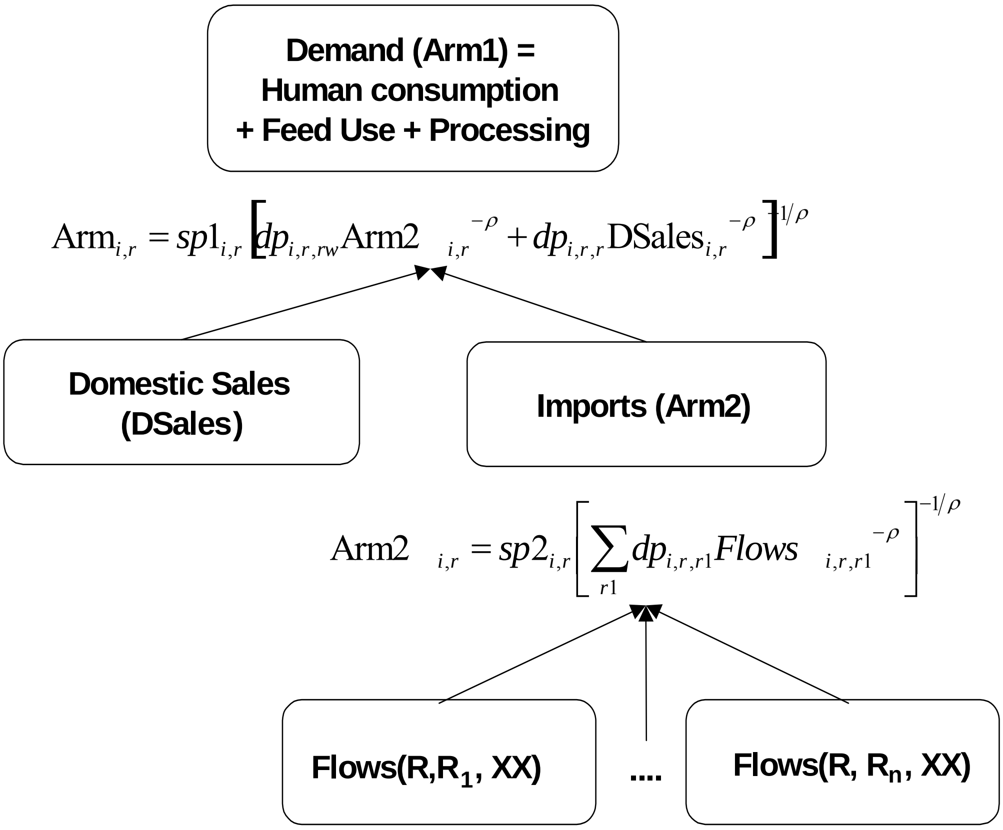
The following table shows the substitution elasticities used for the different product groups. Compared to most other studies, we opted for a rather elastic substitution between products from different origins, as agricultural products are generally more uniform then aggregated product groups, as they can be found e.g. in CGE models.

Substitution elasticities for the Armington CES utility aggregators[37]

Product (group)	Substitution elasticity between domestic sales and imports	Substitution elasticity between import flows
Butter & Cream, Meat	4	8
Cheese, fresh milk products	2	4
All other products	10	25

Source: Own calculations

Figure 6. 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.

Secondly, the Armington aggregator defines an utility a ggregate and not a physical quantity. That second problem is healed by re-correcting in the result listing 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.

---

[37] A sensitivity analysis on those elasticities is given in section 4.7

4.4.7 Market clearing conditions

All quantities in the model are measured in 1000 metric tons. The **quantity balances** first state that production must be equal to domestic sales plus export flows plus changes in intervention stocks:

Equation 78 $$supply_{i,r} = dsales_{i,r} + \sum_{r1 \neq r} flows_{i,r1,r} + isch_{i,r}$$

Further on, **imports and exports** are defined from bilateral trade flows as:

Equation 79 $$imports_{i,r} = \sum_{r1 \neq r} flows_{i,r,r1}$$

Equation 80 $$exports_{i,r} = \sum_{r1 \neq r} flows_{i,r1,r}$$

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 81 $$arm1_{i,r} = feed_{i,r} + hcon_{i,r} + proc_{i,r}$$

4.4.8 Price linkages

All prices in model are expressed as € per metric ton. **Import prices** $impp_{i,r,r1}$ from region *r1* into region *r* of product *i* are determined from market prices *pmrk* taking into account bilateral ad valorem (*tariffa*) and specific (*tariffs*) tariffs minus export subsidies *expsub*:

Equation 82 $$impp_{i,r,r1} = pmrk_{i,r1}\left(1 + tariffa_{i,r,r1}/100\right) + tariffs_{i,r,r1} - expsub_{i,r1}$$

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 using direct and indirect PSEs price wedges, 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.

Equation 83 $$ppri_{i,r} = pmrk_{i,r} + PSEd_{i,r} + PSEi_{i,r}$$

The **average prices of imports** derived from the Armington second stage aggregate are labelled *arm2p* and defined as total import value divided by the Armington second stage utility aggregate *arm2*:

Equation 84 $$arm2p_{i,r} = \frac{\sum_{r1 \neq r} flows_{i,r,r1}\ impp_{i,r,r1}}{arm2_{i,r}}$$

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 85 $$arm1p_{i,r} = \frac{arm2_{i,r}\ arm2p_{i,r} + dsales_{i,r}\ pmrk_{i,r}}{arm1_{i,r}}$$

Consumer prices *cpri* are derived from the composite good price index *arm1p* taken into account policy introduced price wedges as direct and indirect consumer subsidy equivalents plus a fix margin covering transport, processing and all other marketing costs:

Equation 86 $$cpri_{i,r} = arm1p_{i,r} - CSEd_{i,r} - CSEi_{i,r} + cmrg_{i,.r}$$

Unit value exports net of border protection are defined as average market prices in the export destination minus tariffs as:

Equation 87 $$uvae_{i,r} = \frac{\sum_{r1\neq r}\left(pmrk_{i,r1} - tariffs_{r1,r,i}\right)/\left(1 - tariffa_{i,r,r1}\right) flows_{r1,r,i}}{exports_{r,i}}$$

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 88 $$expsub_{i,r} = \frac{exps_{i,r}}{exports_{i,r}}\left(pmrk_{r,i} - uvae_{r,i} + c\exp s_{r,i}\right)$$

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 89 $$\frac{arm2_{i.r}}{dsales_{i,r}} = \left(\frac{dp_{i,rw,r}}{dp_{i,r,r}}\frac{pmrk_{i,r}}{arm2p_{i,r}}\right)^{\frac{1}{1+\varphi_1}}$$

Similarly, relations between import shares are determined by:

Equation 90 $$\frac{flows_{i.r,r1}}{flows_{i,r,r2}} = \left(\frac{dp_{i,r,r1}}{dp_{i,r,r2}}\frac{impp_{i,r,r2}}{impp_{i,r,r1}}\right)^{\frac{1}{1+\varphi_2}}$$

4.4.9 Endogenous policy instruments in the market model

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 91 $$Sigmoid(x) = \exp\left(\ \min(x,0)\ /\ \left(\ 1 + \exp(-abs(x))\ \right)\ \right)$$

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.

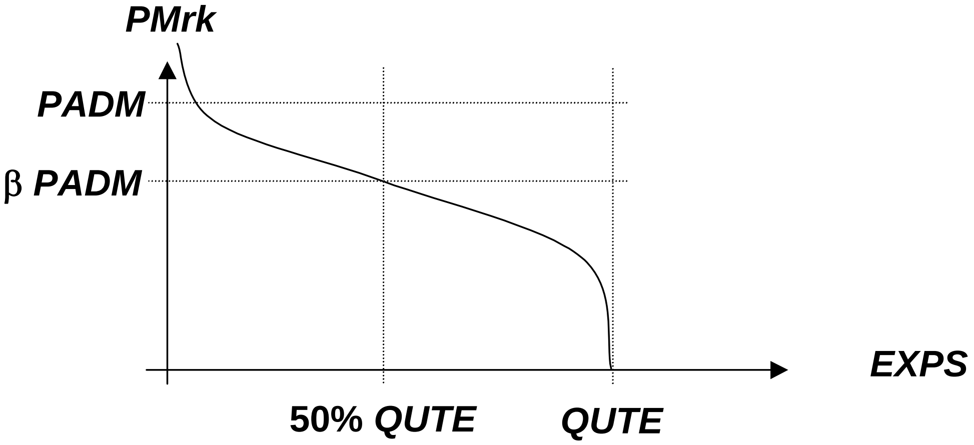
Equation 92 $$exps1_{i,r} = QutE_{i,r}\left[1 - sigmoid\left(\frac{\alpha_{i,r}}{\beta^{E}_{i,r}PADM_{ir}}\left(pmrk_{i,r} - \beta^{E}_{i,r}PADM_{i}\right)\right)\right]$$

The parameters β are determined based on observed price and quantities of subsidised exports. In order to ensure that subsidised exports do not exceed actual exports, the following smooth approximation is used:

Equation 93 $$exps_{i,r} = \tfrac{1}{2}\left(exps1_{i,r} + exports_{i,r} - \sqrt{\left(exps1_{i,r} - exports_{i,r}\right)^2 + \tfrac{1}{20}^2} + \tfrac{1}{20}\right)$$

The relation is shown in the figure below.

Figure 7. Modelling of subsidised exports by a logistic function



Purchases to intervention stocks *intp* depend on the probability of the current market price *pmrk* 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 94 $$intp_{i,r} = IntM_{i,r} errf\left(\left(padm_{i,r} - pmrk_{i,r} + \gamma^p_{i,r}\right)/stddev_{i,r}\right)$$

A decrease of the administrative price or an increase of the market price will hence decrease purchases to intervention stocks.

Releases from intervention stocks *intd* depend on the probability of market prices *pmrk* 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 95 $$intd_{i,r} = \left(intk_{i,r} + intp_{i,r}\right) errf\left(\left(uvae_{i,r} - pmrk_{i,r} + \gamma^d_{i,r}\right)/stddev_{i,r}\right)$$

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 96 $$ints_{i,r} = intp_{i,r} - intd_{i,r}$$

4.4.9.1 Update note

Modelling of subsidised exports has been changed from a quantity to a value base. Hence the logistic function determines the outlays on subsidised exports and it is the value ceiling from the WTO which is controlled in general rather than the quantity ceiling. Furthermore, some modifications have been introduced to acknowledge the double zero agreements involving the EU.

4.4.10 Endogenous tariffs under 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. 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.
- *Quota exactly filled*: the in-quota tariff is applied. The willingness to pay of consumers and thus the price paid is somewhere between the border plus the in-quota tariff and the border price plus the MFN tariff. The difference between the price in the market and the border price plus the in-quota tariff establishes a quota rent. Depending on property rights on the quota and the allocation mechanism, the quota rent is shared in different portions by the producers, importing agencies, the domestic marketing chain or the administration. Typically, the quota rent can neither be observed nor is their knowledge about distribution of the rent.
- *Quota overfill*: the higher MFN-tariff is applied. The quota rent is equal to the difference between the MFN and the in-quota tariff. Again, how the quota rent is distributed to agents is typically not known.

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 7) is applied in the CAPRI market part.

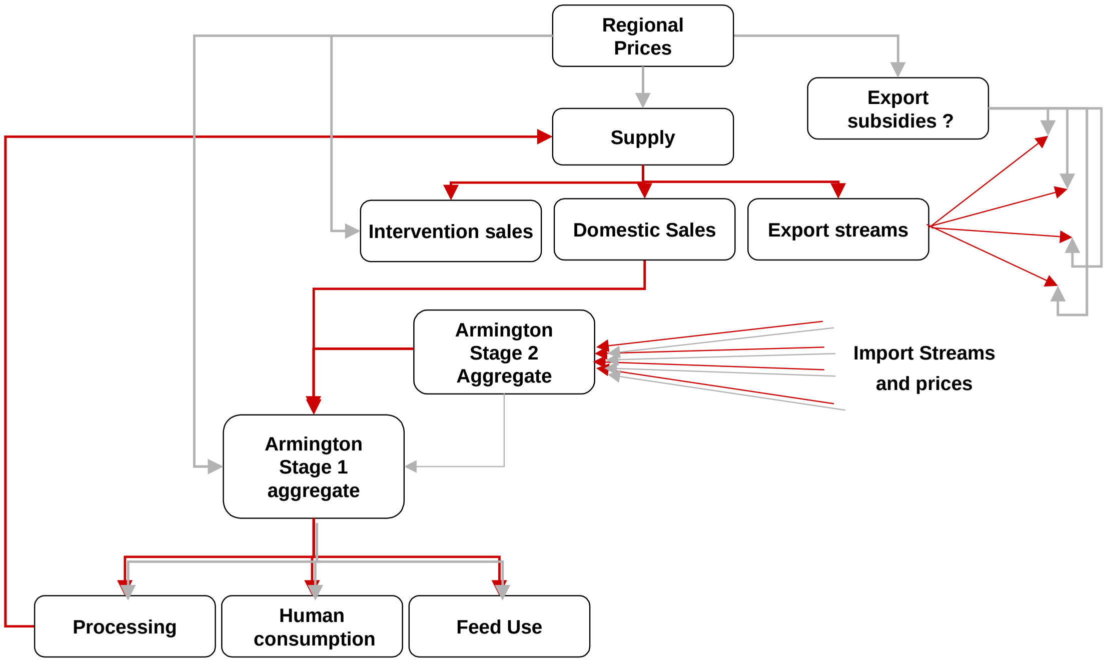
4.4.10.1 Update note

Some time ago two further details have been added to the representation of tariff regimes, evidently in view of their relevance to the EU: The model allows for the definition of variable import levies that reduce tariffs if the import price (CIF plus tariffs) is higher than a specified minimum border price. 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.

4.4.11 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 8. Graphical presentation for one region of a spatial market system



Source: CAPRI modelling system

4.4.12 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.

4.5 Parameter calibration and sources for the behavioural equations

4.5.1 Calibration of the system of supply functions

The supply equation was already introduced in Equation 61. 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:

Equation 97
$$\varepsilon_{i,j} = bs_{i,j} \frac{\overline{ppri}_{j,r} \Big/ \overline{p}_{index.r}}{\overline{supply}_i}$$

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.

4.5.2 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[38] 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:

Equation 98

$$\begin{aligned}
&\frac{\partial PerCap_{r,i}}{\partial y} = \frac{G_i}{G_r} \\
&\frac{\partial PerCap_{r,i}}{\partial cpri_{r,j}} = \left(\frac{G_{ij}}{G} - \frac{G_i G_j}{{G_r}^2} \right) (y - F) \wedge i \neq j \\
&where: \\
&G_{ij} = \frac{\partial G_i}{\partial cpri_j} = \tfrac{1}{2} bd_{ij} \Big/ \sqrt{cpri_i cpri_j} \quad \wedge i \neq j
\end{aligned}$$

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 is tested which is likely to become 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. The current calibration uses Hicksian price elasticites rather than the Marshallian price effects from Equation 98, but both types of marginal effects are closely related.

4.5.3 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

---

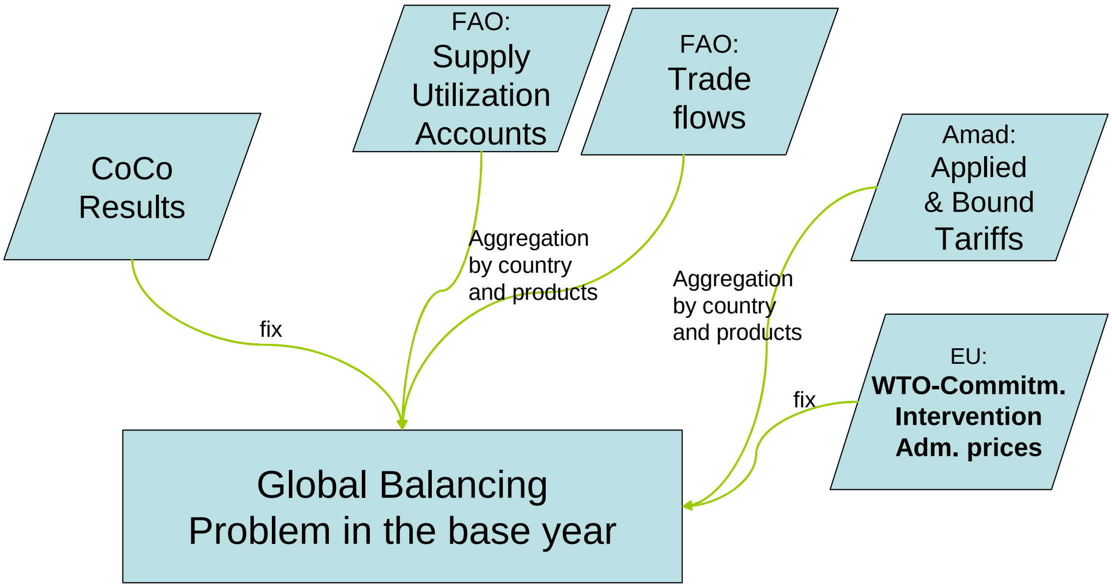
[38] As an alternative, a normalized quadratic expenditure system was tested. According to the family of indirect utility functions discussed above, the function G is then replaced by a from quadratic in normalized prices. However, a Cholesky decomposition is then necessary to ensure correct curvature 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.

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.

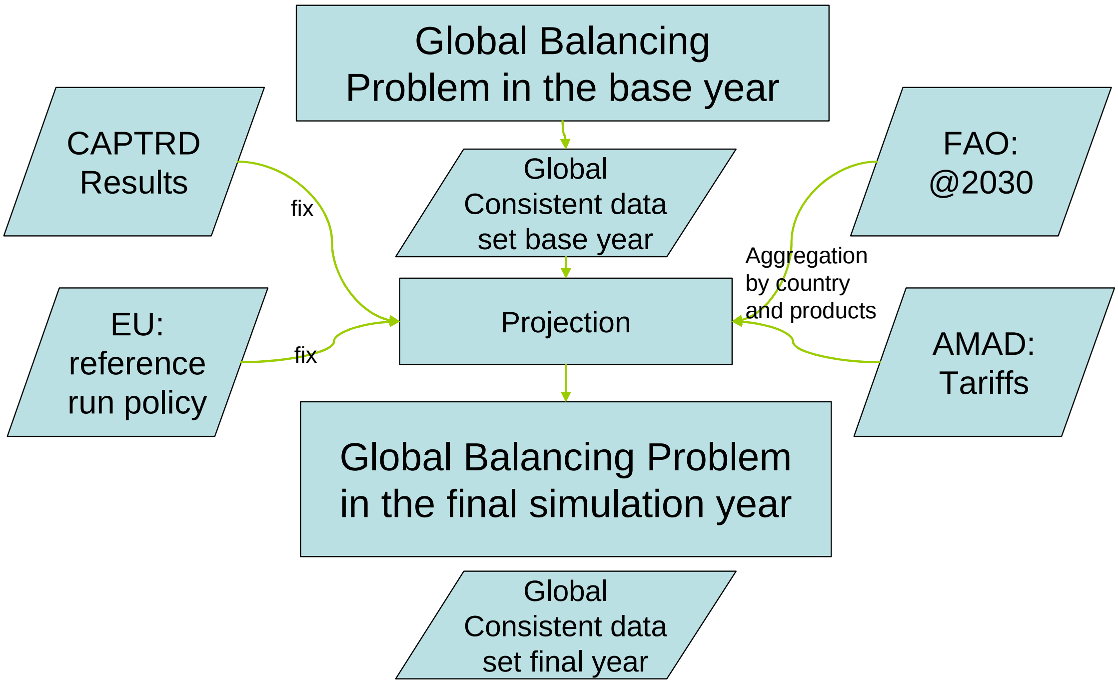
Figure 9. Balancing the market model for the base year



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 on the FAO publication "Agriculture at 2030"and publicly available FAPRI medium term projection. Indeed, there is ready to use GAMS code to convert the FAPRI baseline into a GAMS data table in CAPRI definitions.

Figure 10. Balancing the market model in the final simulation year.



4.6 Linking the different modules – the price mechanism

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 Sensitivity of the CAPRI model to the Armington substitution elasticities

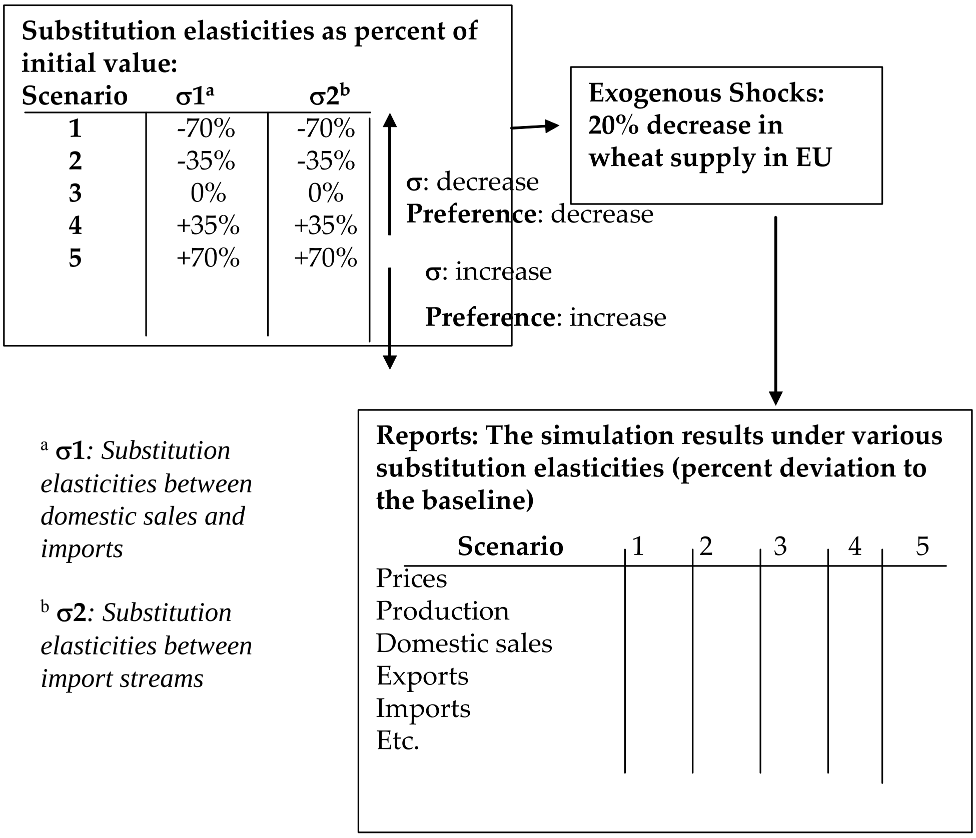
A conventional sensitivity analysis consists to run the model using initial Armington elasticities to obtain the baseline, then to rerun it under various elasticity values, all other things held constant, and finally to compare the reference and simulation results. In our sensitivity study, the implementation 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 Figure 11.

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.

Figure 11. Illustration of Sensitivity Analysis on the CAPRI Market Module



Substitution elasticities as percent of initial value:

Scenario	σ1[a]	σ2[b]
1	-70%	-70%
2	-35%	-35%
3	0%	0%
4	+35%	+35%
5	+70%	+70%

Reports: The simulation results under various substitution elasticities (percent deviation to the baseline)

Scenario	1	2	3	4	5
Prices					
Production					
Domestic sales					
Exports					
Imports					
Etc.					

[a] **σ1**: *Substitution elasticities between domestic sales and imports*

[b] **σ2**: *Substitution elasticities between import streams*

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 (0), 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%.

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 0 and 0, 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.

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 0, 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.

Deviation of the simulation results to the baseline under high and low substitution elasticities with an increase in tariff rate quotas[a]

	Elasticity of substitutionb	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 0, 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 (0).

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

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.

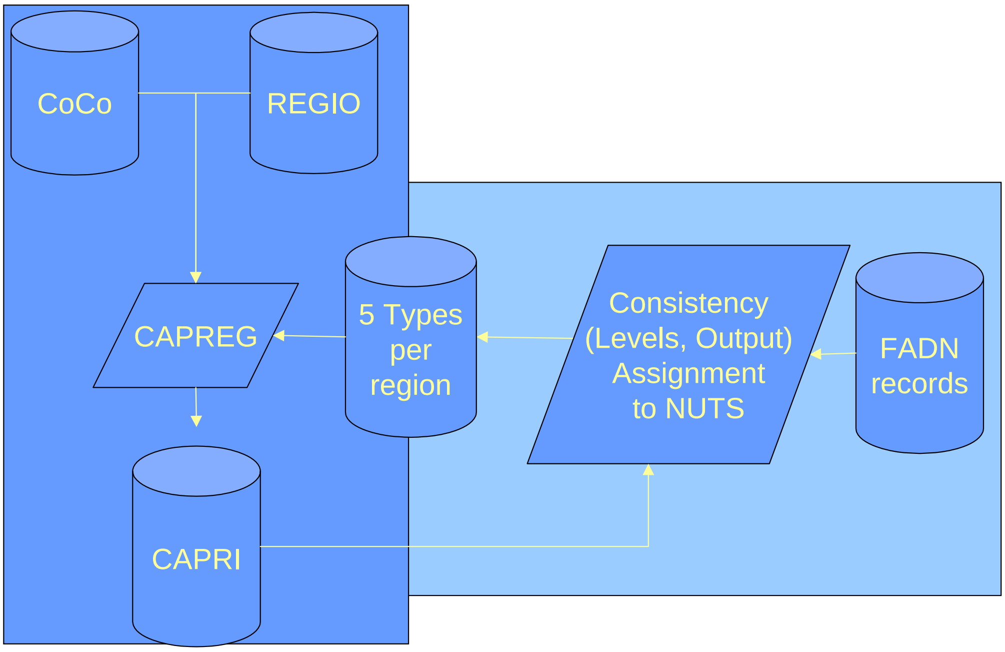
But even with the farm typology according to European standards applied, a number of issues need to be addressed for its application in CAPRI:

(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 0).

The following diagram shows the relation between the FADN data base and the elements of the CAPRI data processor.

Figure 12. 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 0) 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.

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

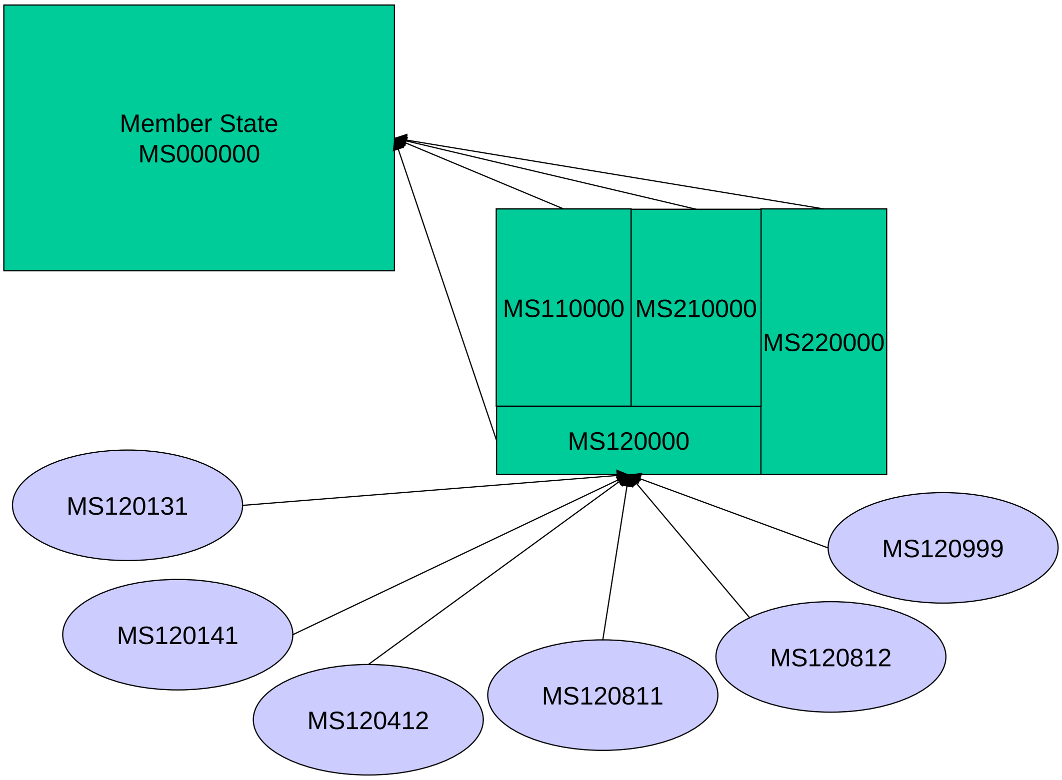
Source: FADN (http://europa.eu.int/comm/agriculture/rica/index_en.cfm).

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 13 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 13. 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 0.

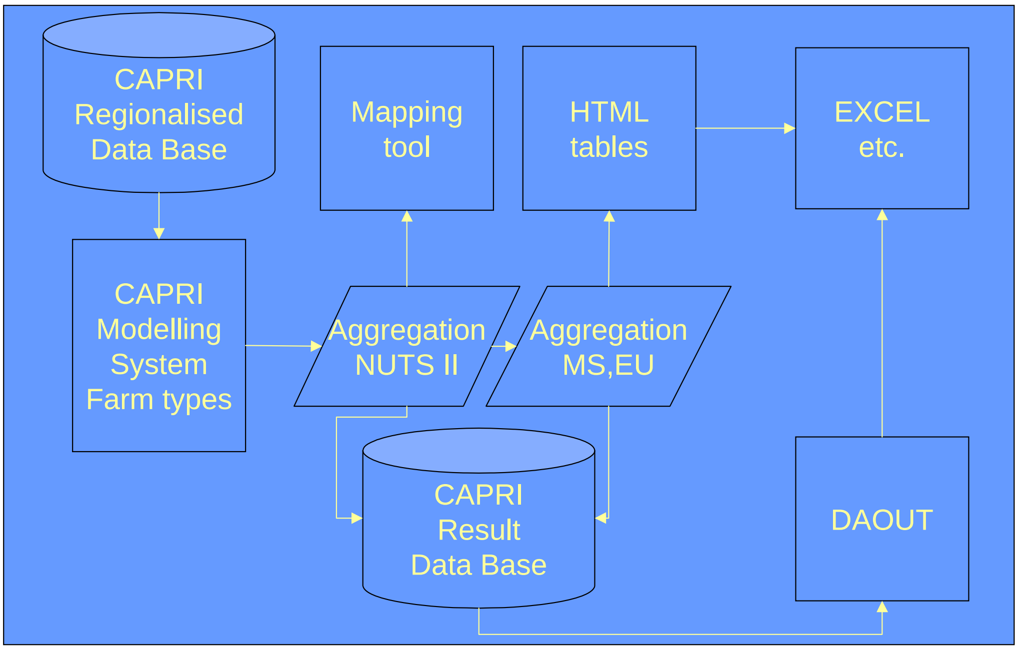
Aggregated farm types used for impact assessment

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 14 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.

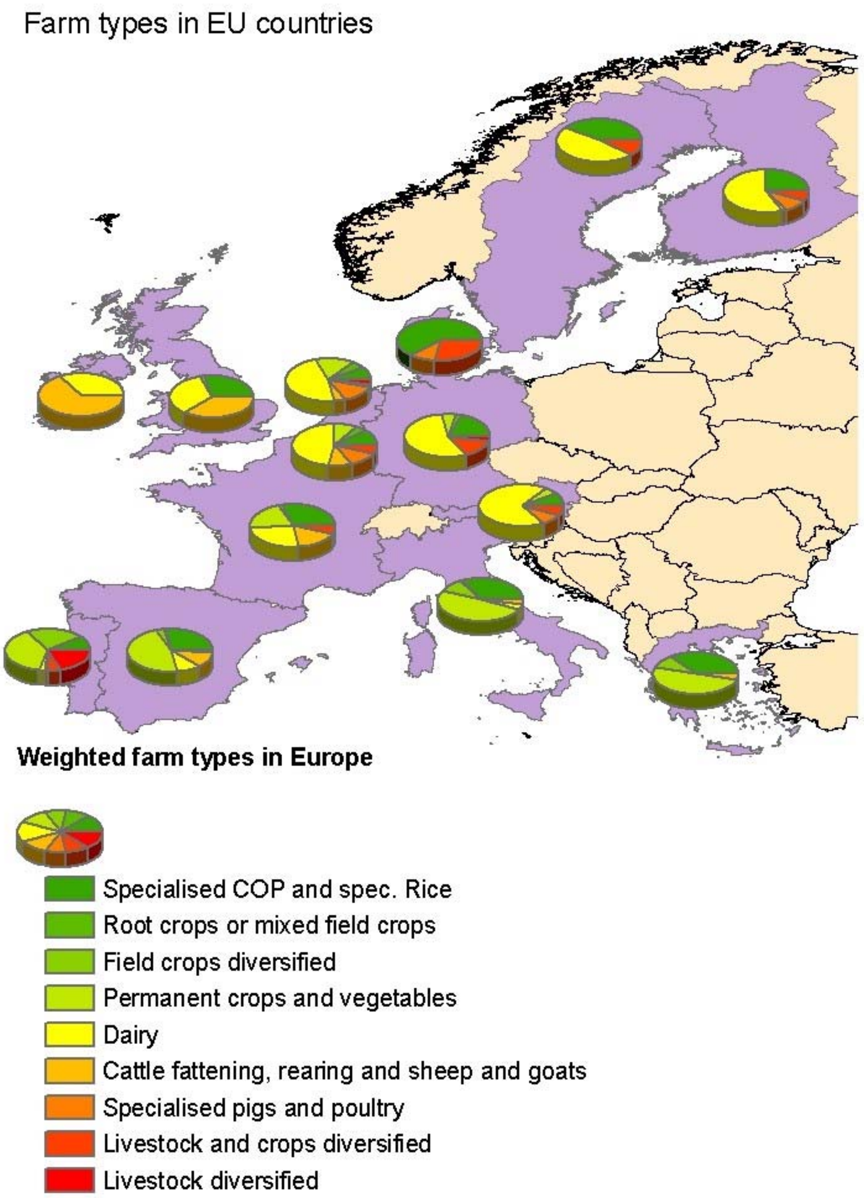
Figure 14. Integration of farm types in the CAPRI modelling system



Source: Own calculations

Figure 15 shows the dominant farm types per country. For reasons of survey research the farm types mentioned in 0 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 15. 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.

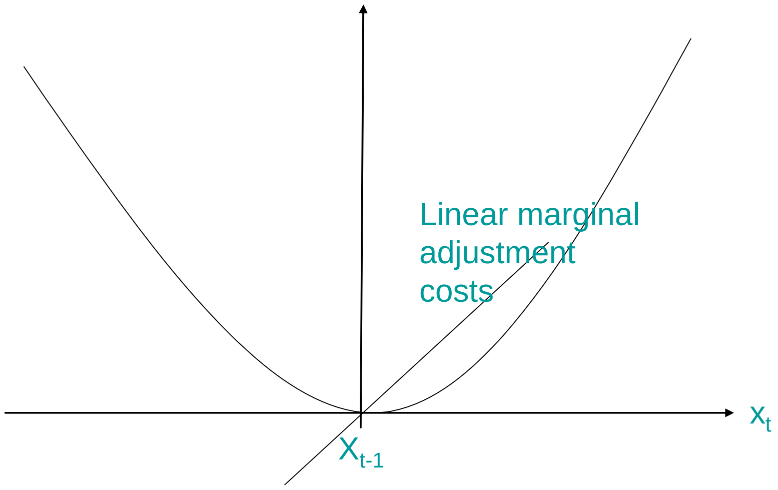
Due to this risk, we propose to pursue both approaches 3 and 4, having the third approach as a backup.

6 A feasibility study for a recursive-dynamic version

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 16. 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 99

$$\frac{C_t(x)}{p_I} = \underbrace{a + \sum_i b_i x_i^t + \tfrac{1}{2} \sum_{ij} q_{ij} x_i^t x_j^t}_{\text{current approach}} + \underbrace{\tfrac{1}{2} \sum_i ac_i \left(x_i^t - x_i^{t-1}\right)^2}_{\text{adjustment cost}}$$

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 100
$$\frac{\partial\left(C_t/P_I\right)}{\partial x_i} = b_i + \sum_j q_{ij} x_j^t + ac_i\left(x_i^t - x_i^{t-1}\right)$$

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 101

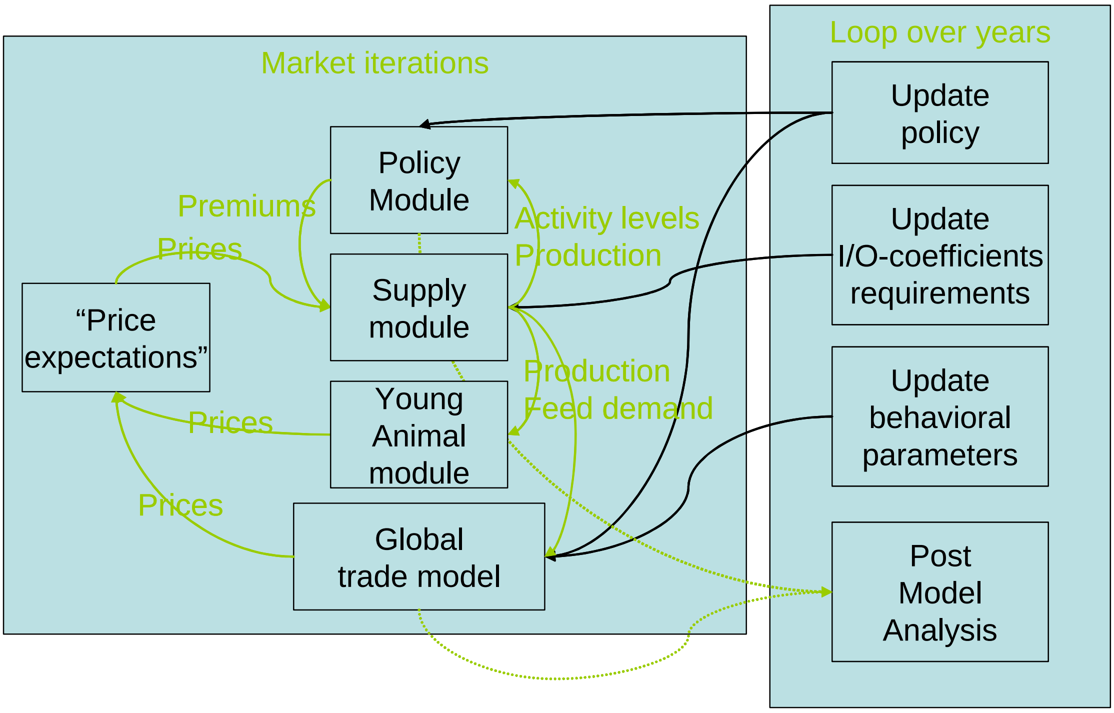
$$\frac{\Pi}{p_I} = \left.\begin{aligned} &\sum_i (1.-PartAdj_i)\, a_i \tilde{p}_i \\ &+ \tfrac{1}{2}\sum_i \sum_j (1.-PartAdj_i)\, B_{ij}\, \tilde{p}_i\, \tilde{p}_j \end{aligned}\right\} \text{Standard profit function}$$

$$\left. + \sum_i PartAdj_i\; X_i^{t-1} \tilde{p}_i \right\} \begin{array}{l}\text{Adjustment part:}\\ \text{A certain part of profits}\\ \text{depends on last year's}\\ \text{production}\end{array}$$

$$\Rightarrow \quad X_i = \frac{\partial \tilde{\Pi}}{\partial \tilde{p}_i} = (1.-PartAdj_i)\left[a_i + \sum_j \tilde{p}_j B_{ij}\right] + PartAdj_i\; X_i^{t-1}$$

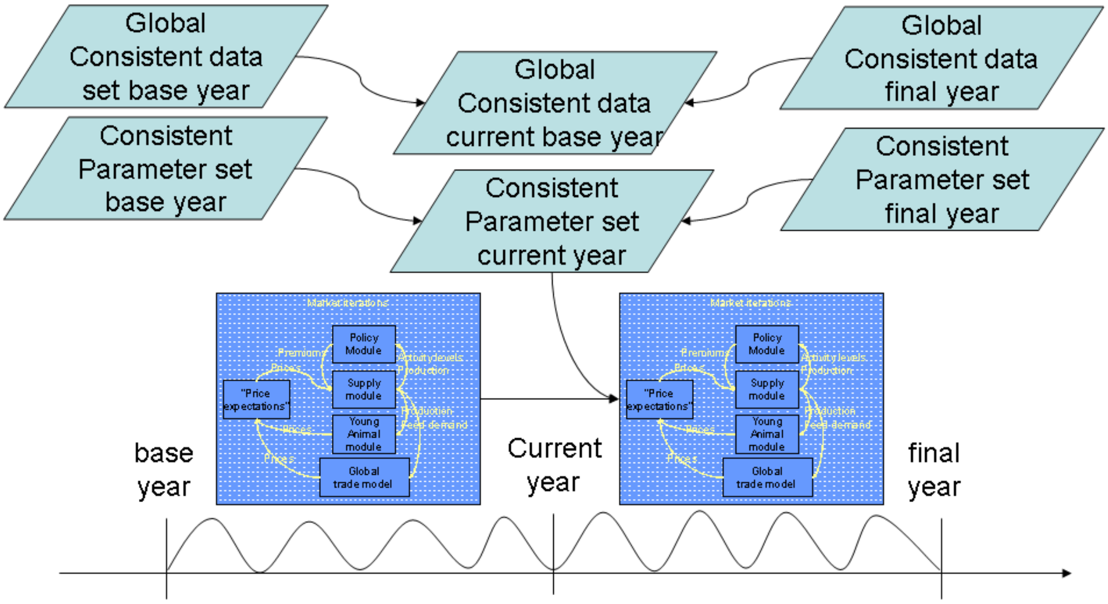
Trials to estimate simultaneously the adjustment cost parameters and the parameters of the long term cost function from a combined time-series cross-sectional estimation using regional time series on crop hectares, yield and the allocated costs from CAPREG as data set did not yield stable regression 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 17. 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 18. 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 A spatial land use map (Markus Kempen and Renate 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.

7.1.1 Spatial calculation unit (HSMU)

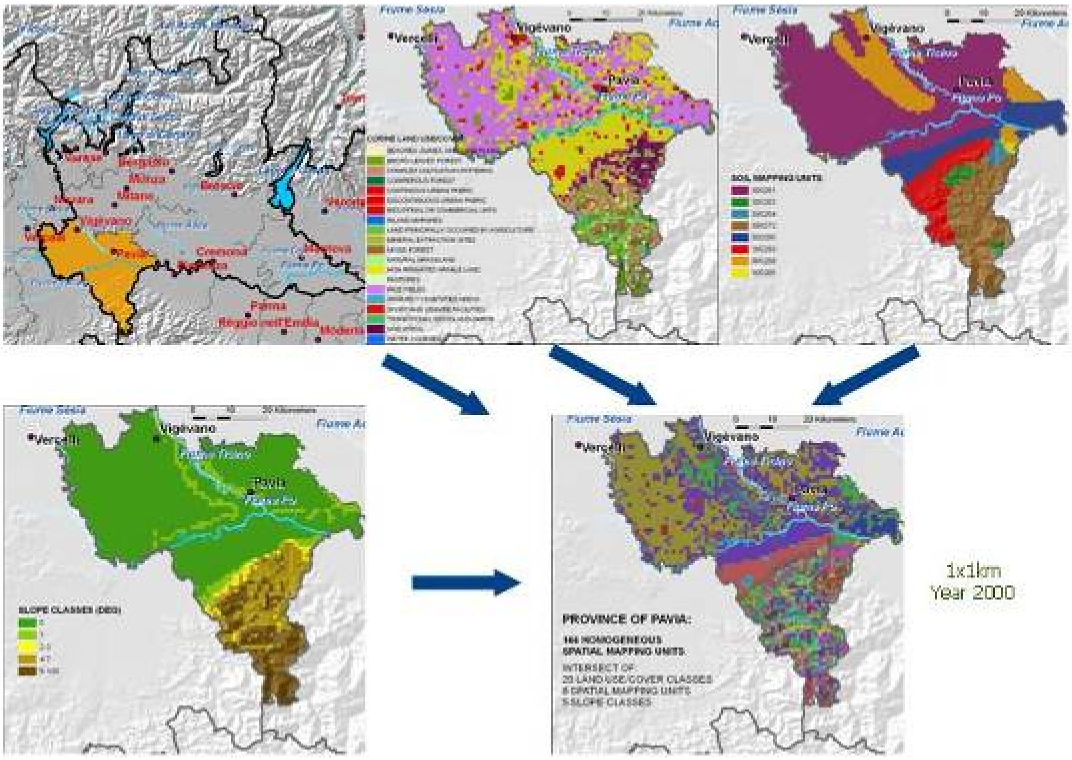
The aim of building HMSU is the definition of areas inside an administrative region where approximate homogeneity according location factors may be assumed. The HMSU serve then as simulation units for the bio-physical models and are constructed by overlaying different maps (land cover, soil map, climatic 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 19. 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.

7.1.2 Consistent disaggregation of land use

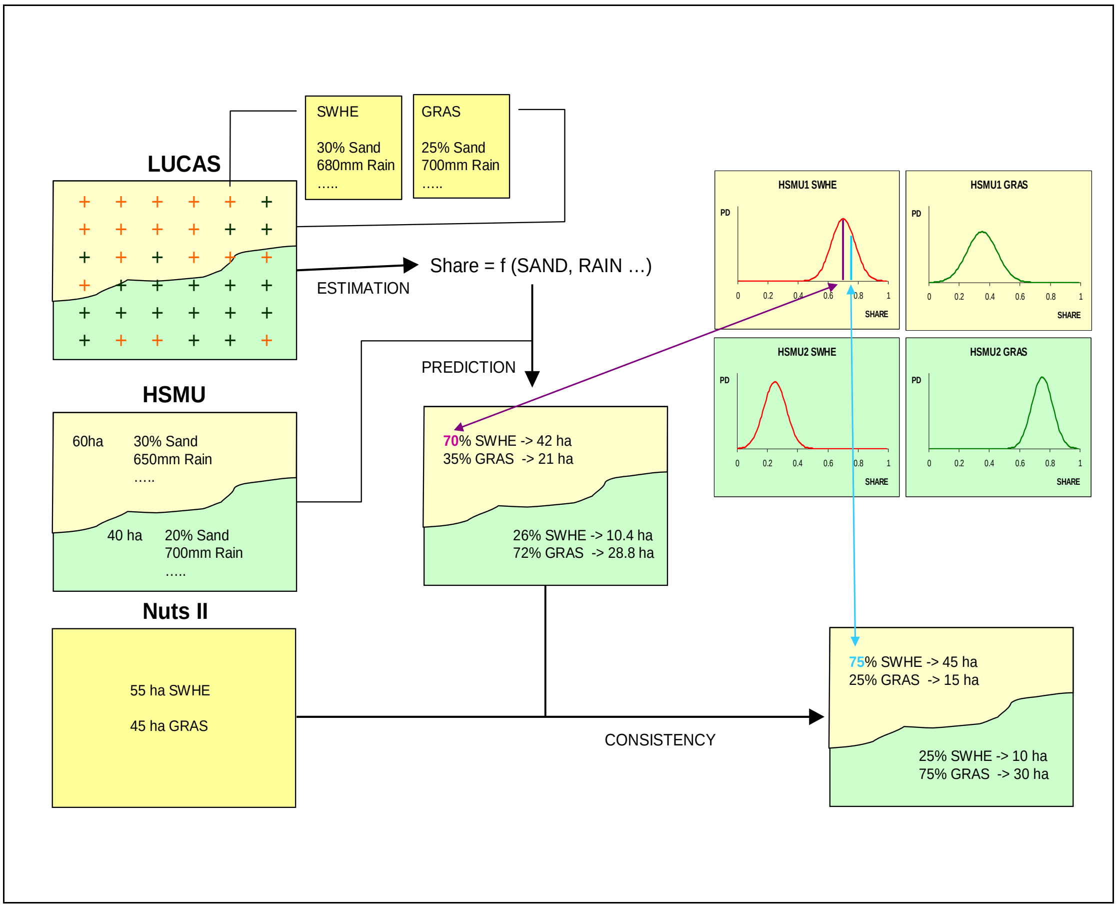
The calculation of a spatial explicit land use map makes use of the LUCAS survey. In opposite to mapping approaches, area frame surveys based on a common statistical sampling method gather land cover and land use data at specific sample points, only, and extrapolate from these to the entire area under 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 20. 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 20. 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:

$$\Lambda(\boldsymbol{\beta}'\mathbf{x}_\mathbf{i}) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_\mathbf{i}}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_\mathbf{i}}}$$

$$\log L = \sum_{i=1}^{n}[y_i \log\Lambda(\boldsymbol{\beta}'\mathbf{x}_\mathbf{i}) + (1-y_i)(1-\Lambda(\boldsymbol{\beta}'\mathbf{x}_\mathbf{i}))]$$

where ***Y*** is a dummy vector indicating whether a certain crop was observed at a location *i* (y_i=1), $\mathbf{x}_\mathbf{i}$ is the design matrix containing data on natural conditions and $\Lambda(\boldsymbol{\beta}'\mathbf{x}_\mathbf{i})$ is the probability that a specific crop is grown at location *i*.

Applying the estimated $\widehat{\boldsymbol{\beta}}$ 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:

$$\hat{\mathbf{Y}} = \Lambda(\boldsymbol{\beta}'\mathbf{x}_h) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_h}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_h}}$$

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:

$$\varpi_{ij} = \left[1-\left(\frac{\delta_{ij}}{d_j}\right)^3\right]^3 I(\delta_{ij} < d_j)$$

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:

$$CV = \sum_{i=1}^{n}\left(y_i - \hat{y}_{i\neq i}\right)^2$$

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):

$$\log L = \sum_{i=1}^{n} \omega_{ij} [y_i \log \Lambda(\boldsymbol{\beta}_{\mathbf{j}}^{'}\mathbf{x}_{\mathbf{i}}) + (1-y_i)(1-\Lambda(\boldsymbol{\beta}_{\mathbf{j}}^{'}\mathbf{x}_{\mathbf{i}}))]$$

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:

$$V_\beta = Cov\left\lfloor \hat{\boldsymbol{\beta}} \right\rfloor = \hat{\mathbf{H}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{H}}^{-1}$$

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):

$$\mathbf{H} = \frac{\partial^2 LogL}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{'}} = -\sum_i \omega_i \Lambda_i (1-\Lambda_i)\mathbf{x}_{\mathbf{i}}\mathbf{x}_{\mathbf{i}}^{'}$$

$$\mathbf{B} = \sum_i \omega_i (y_i - \Lambda_i)^2 \mathbf{x}_{\mathbf{i}}\mathbf{x}_{\mathbf{i}}^{'}$$

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 $\mathbf{V}_{\boldsymbol{\beta}}$.

$$\mathbf{V}_{\mathbf{Y}} = Var\left\lfloor \hat{\mathbf{Y}} \right\rfloor = \Lambda_i (1-\Lambda_i)\mathbf{x}'\mathbf{V}_{\boldsymbol{\beta}}\mathbf{x}$$

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:

$$V_Y = VAR(\hat{Y}) = \frac{1}{B-1}\sum_{b=1}^{n}\left(\hat{Y}_b - \hat{Y}\right)^2$$

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

$$\sum_{h\in N2} Y_{c,h}^{con} A_h = CA_{c,N2}$$

and the adding up to unity

$$\sum_{C} Y_{c,h}^{con} = 1$$

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:

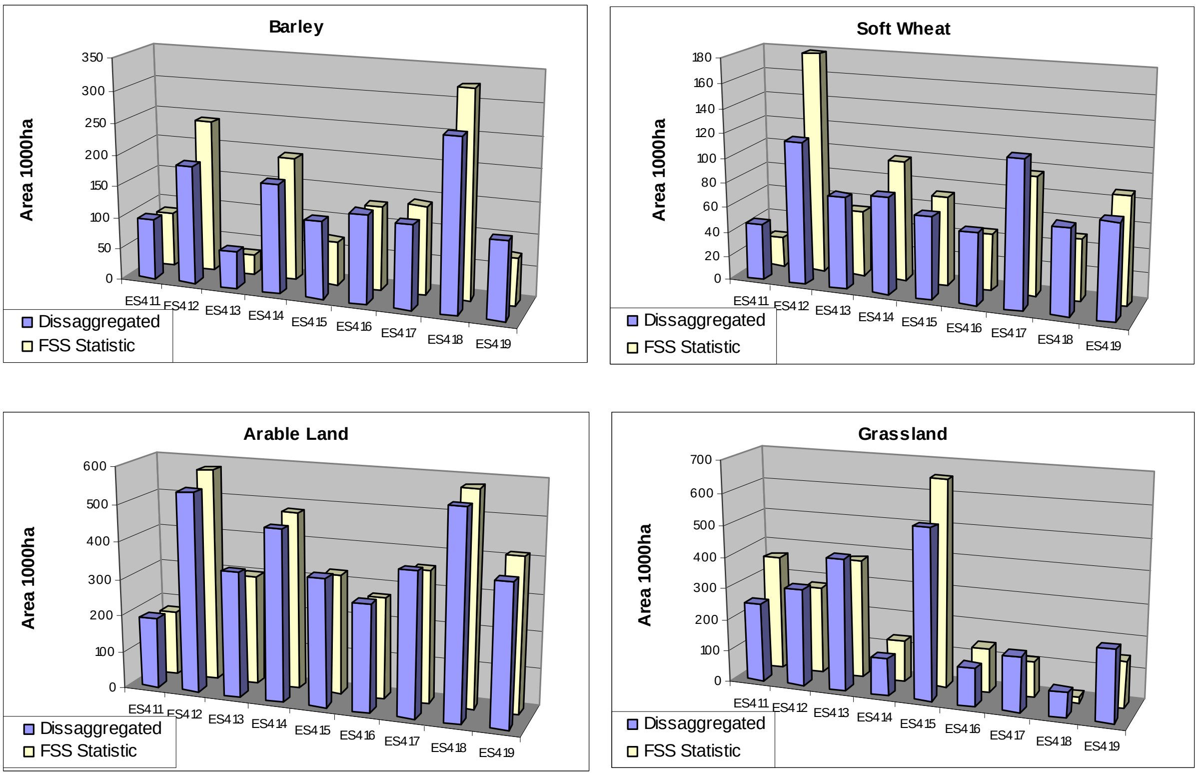
$$-\sum_{c}\sum_{h}\left[\log\left(\sqrt{2\pi}\mathbf{V}_{\mathbf{Y_{c,h}}}\right) + \frac{\left(\mathbf{Y}_{c,h}^{con} - \hat{\mathbf{Y}}_{c,h}\right)^2}{2\mathbf{V}_{\mathbf{Y_{c,h}}}{}^2}\right]$$

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[39]. Adding up over the corresponding HSMU yields crop areas at NUTS III level that can be compared to the observed data. Figure 21 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 21. Comparison of estimated and observed shares in NUTS III region for different crops (Castilla Leon ES410)



In order to measure the overall fit with a few meaningful numbers we calculate the percentage of misclassified area within each Nuts II. For each crop the total of the absolute difference between predicted and actual area in each Nuts III is divided by the area in the Nuts II region. Summation over all crops yields the percentage of misclassified area in the region at the accuracy of a single crop. Alternatively errors can be calculated for groups of similar crops. Often the errors level out within such an aggregate.

---

[39] Usually this disaggregation procedure is applied to the complete and consistent Nuts II database of the CAPRI modelling system, but any other statistic can be used as well. In order to allow a consistent analysis based on FSS Nuts III data the corresponding Nuts II information was used here.

Precentage of misclassified area for different crops (Castilla Leon ES410)

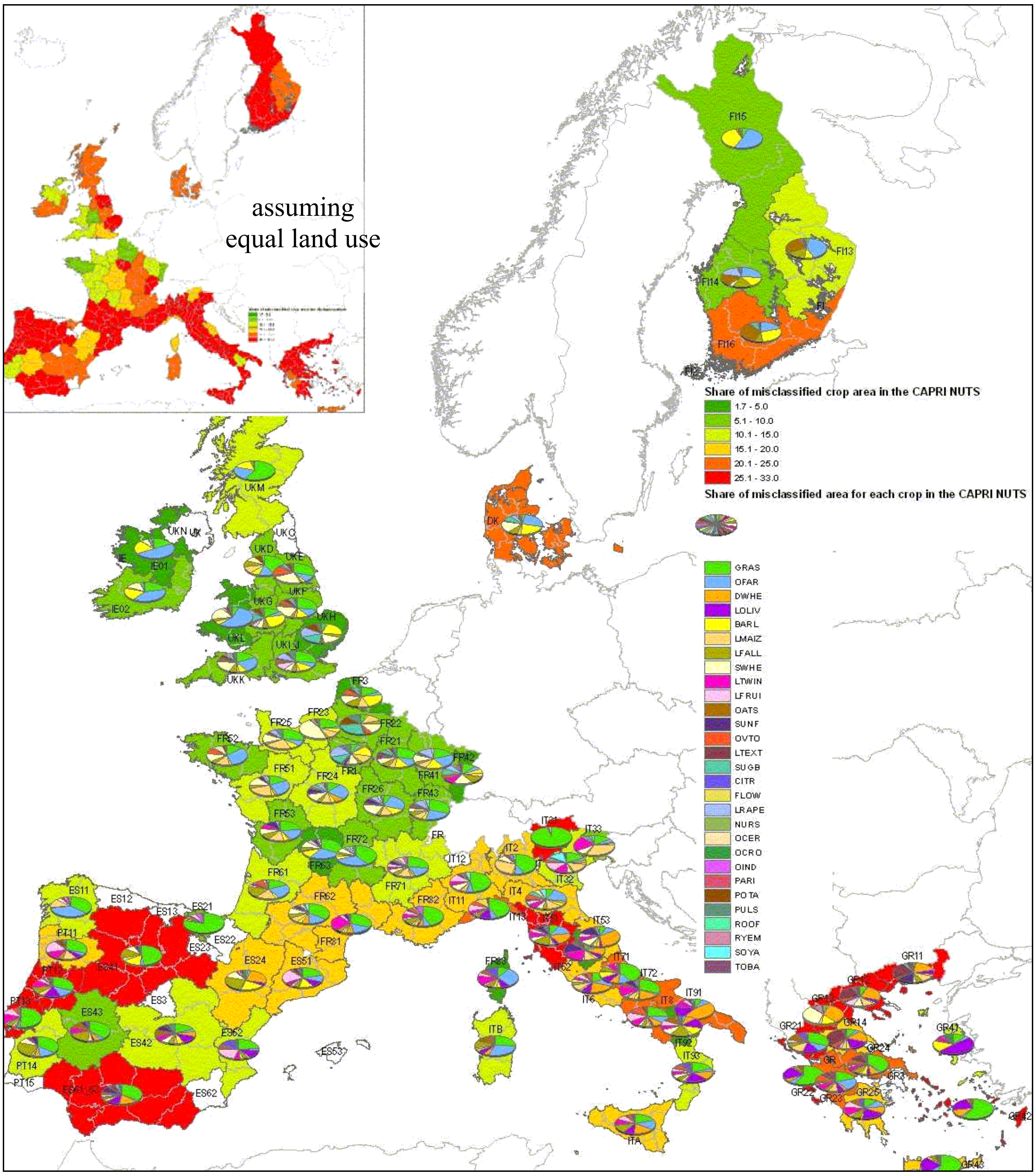
Crop	Missclassified Area in Nuts III (% of UAA in Nuts II)			
	Single Crop	Groups		
Soft Wheat	3.39	6.44	Arable Land 3.96	UAA 8.43
Durum Wheat	0.32			
Barley	5.02			
Rye	0.87			
Oats	1.46			
Maize	1.24			
Other Cereal	0.06			
Fallow Land	2.95	2.95		
Rice	0.00	1.09		
Sunflower	1.19			
Soya	0.00			
Texture 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 22 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[40]. Nonetheless the dissaggregation is a significant improvement compared to the assumption of identical cropping pattern within each Nuts II region.

---

[40] 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 22. Precentage of misclassified areas in validated Nuts II Regions after disaggregation



7.2 Yields, irrigation shares and stocking densities (Wolfgang 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.

7.3 Linkage to process-based modelling (DNDC) (Adrian Leip and Gerry Mulligan)

7.3.1 Introduction

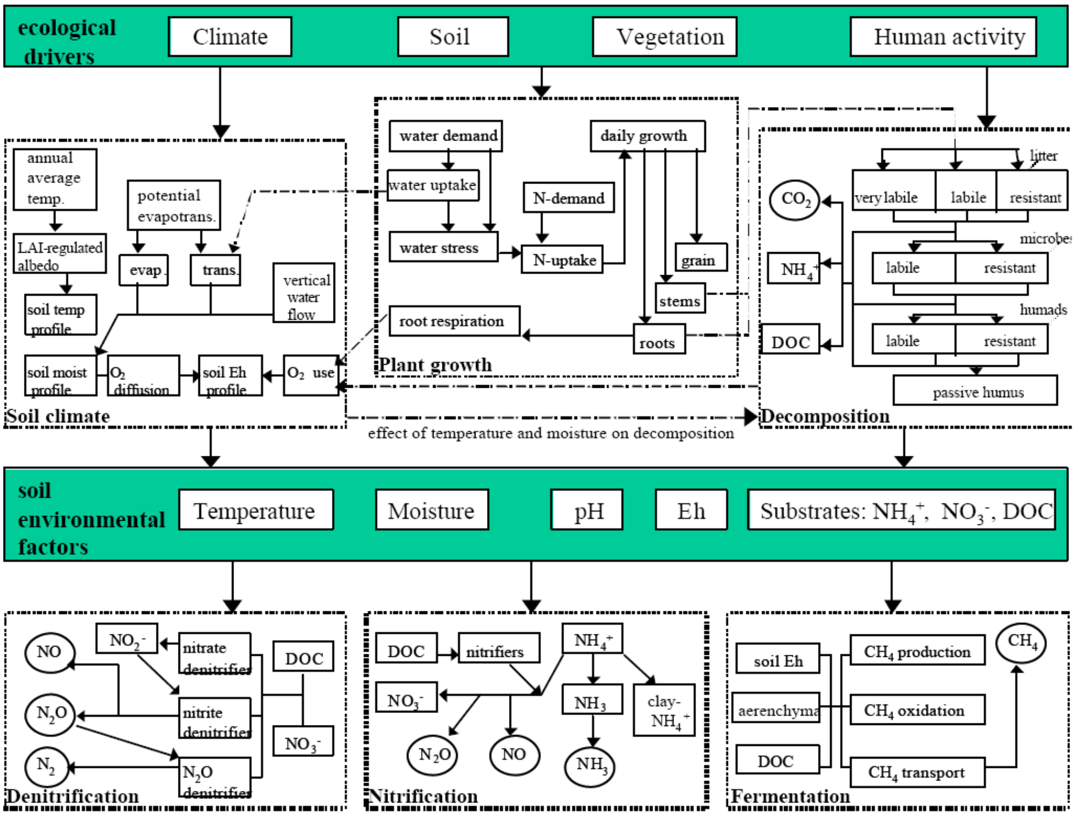
Process-based models are adequate to analyze the impact of changing farming practices, as they are able to cope with the complex interplay of environment and anthropogenic activities. But the accuracy of simulated fluxes with process-based models such as DNDC (Denitrification Decomposition) Model (Li 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 23). 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.

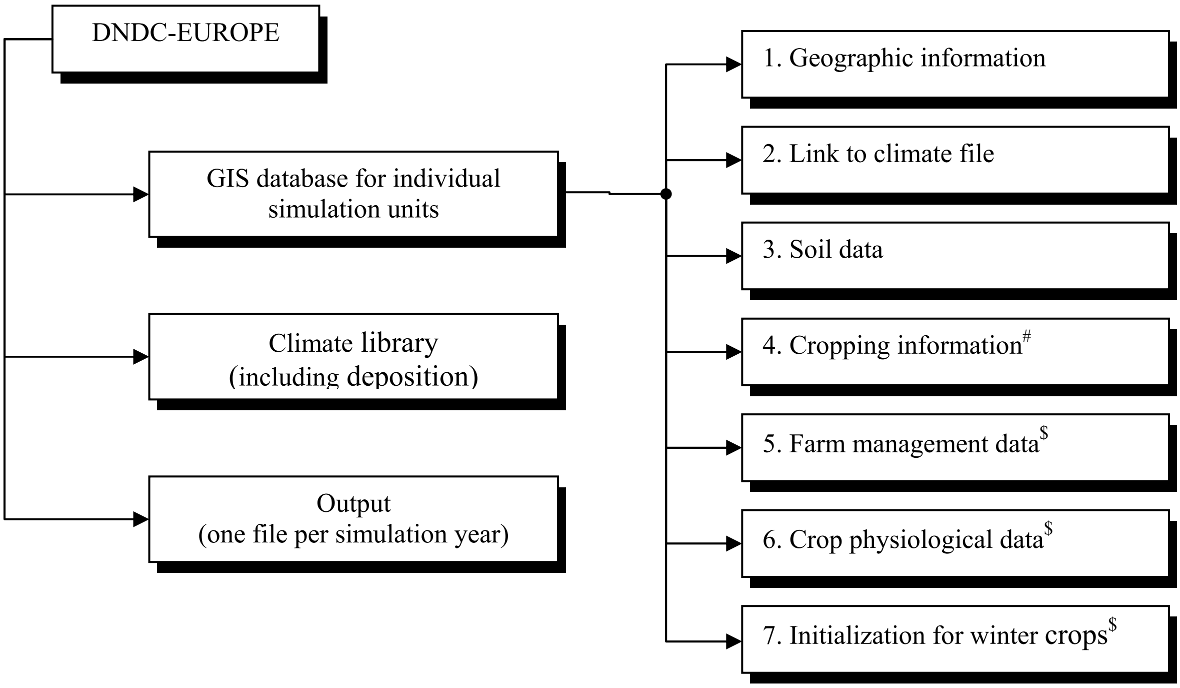
Figure 23. Structure of the DNDC model



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 24. Database structure of DNDC-EUROPE. $^{\#}$Modified GIS file; $^{\$}$Additional GIS file



7.3.2 *Input data*

DNDC data requirements include:

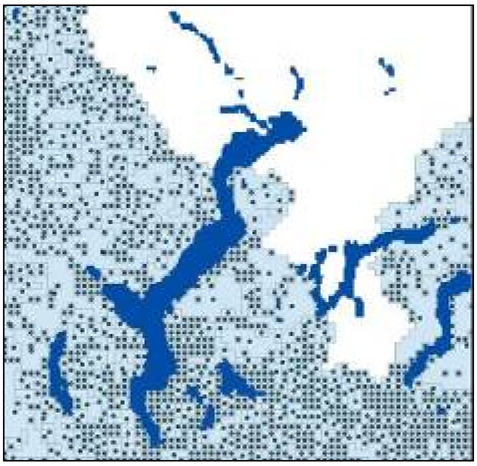
- Geographic location (e.g. region or point)
- Climate data (precipitation, temperature)
- Soil organic carbon, clay content, pH and bulk density
- Nitrogen fertiliser and manure, atmospheric deposition
- Crop area
- Management operations (including sowing, harvesting, tillage, fertiliser application)

They are briefly described in the following sections.

7.3.2.1 Geographical data

The longitude and latitude coordinates derived from the centroids of each HSMU (see Figure 25) are used within the DNDC model to drive the day length function of the crop growth model.

Figure 25. 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 0).

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*

7.3.2.2 Weather data

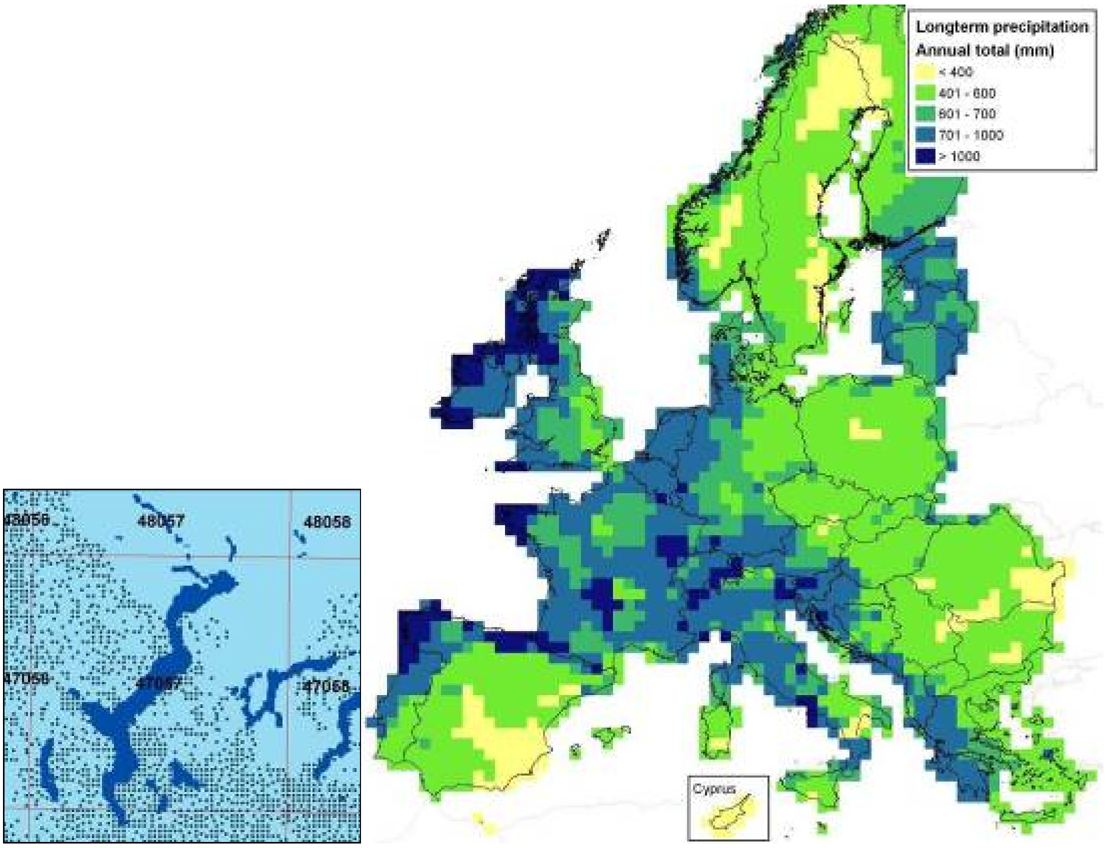
The climate file in the DNDC GIS database contains the atmospheric nitrogen deposition data and provides the link between the modelled (HSMU) units and the meteorological text files containing the daily meteorological data (0)

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*

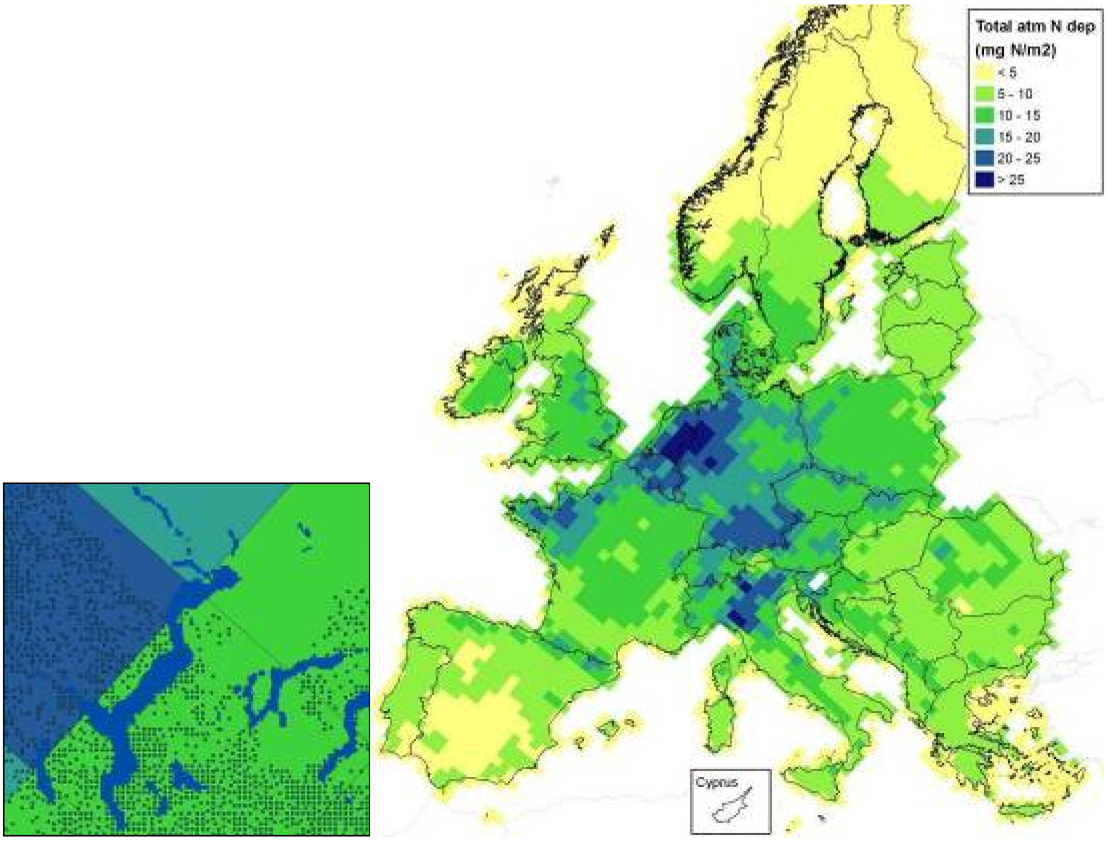
Daily meteorological data for 2000 has been extracted from the Monitoring Agriculture and Regional Information Systems (MARS) database containing data interpolated on a 50 km x 50 km grid (Figure 26) (Orlandi and Van der Goot, 2003). The DNDC model requires minimum and maximum temperature (°C) and precipitation data (cm).

Figure 26. 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 27. Link between HSMU and N deposition grid (EMEP 50 km x 50 km)



7.3.2.3 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)[41].

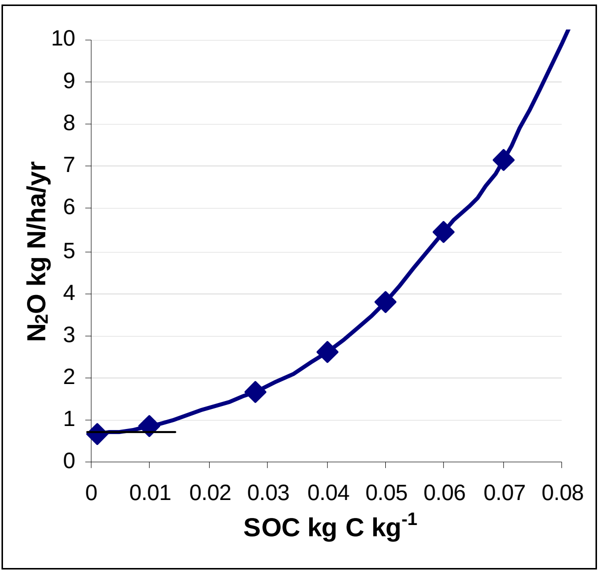
The ESBN have created a series of 1 km x 1 km soil rasters including topsoil organic carbon content that have been calculated using a refined pedo-transfer rule derived from the European Soil Database, an extended CORINE land cover dataset), a digital elevation model (DEM) and mean annual temperature 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).

---

[41] Distribution version 2.0, http://eusoils.jrc.it/ESDB_Archive/ESDBv2/fr_intro.htm

Sensitivity analysis has shown that the DNDC model is very sensitive to SOC (Figure 28) (Mulligan, 2006)

Figure 28. 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)

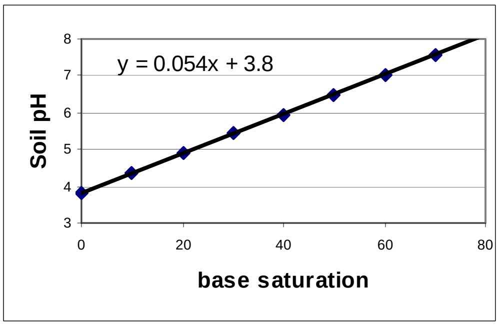
For this project bulk density data (g/cm^3) were derived from packing density (*pers comm*. Jones, 2003) using the pedo-transfer function:

$$\rho_b = P_D - (C_l \times 0.009)$$

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 29) between base saturation and soil pH was estimated based on expert knowledge from R.A. Jones (*pers comm*. 2003).

Figure 29. Linear relationship between soil pH and base saturation

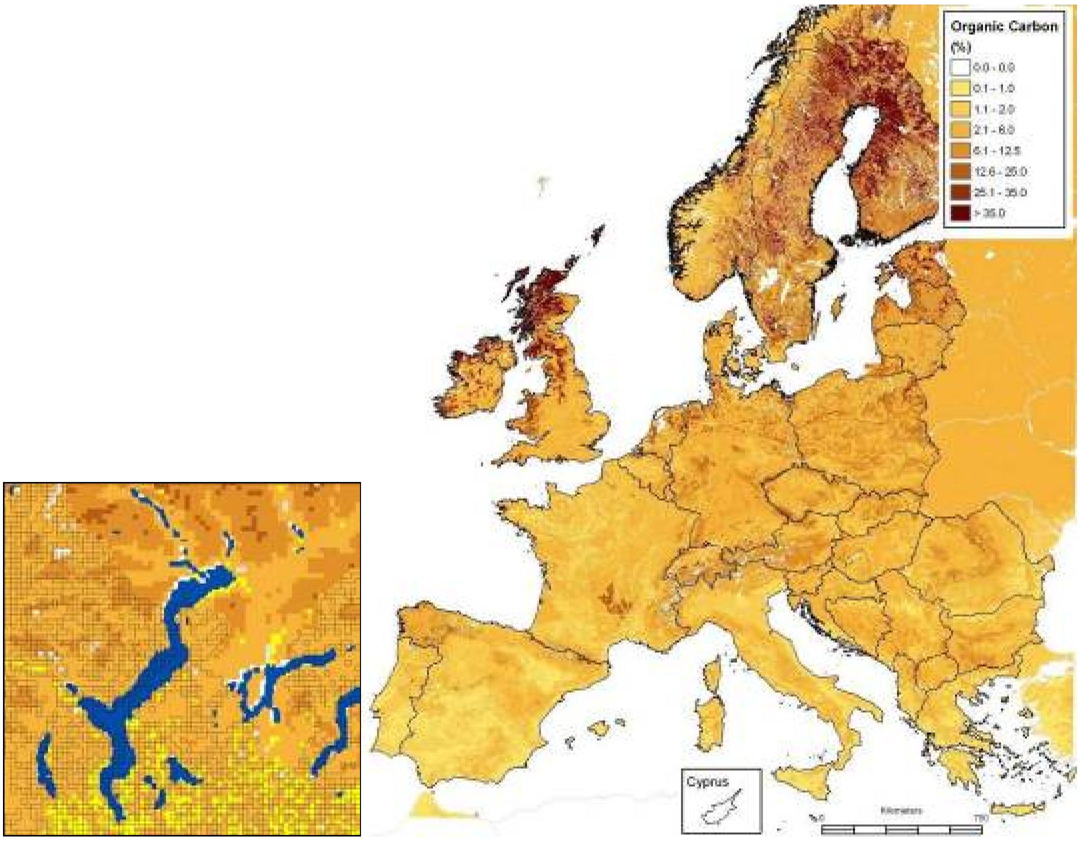


The soil parameters shown in 0 were calculated in ARC GIS zonal statistics using the HSMU shape to define the regions and soil rasters for the values to be summarised (see Figure 30)

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 30. HSMU shapefile used to summarise soil raster values.



7.3.2.4 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).

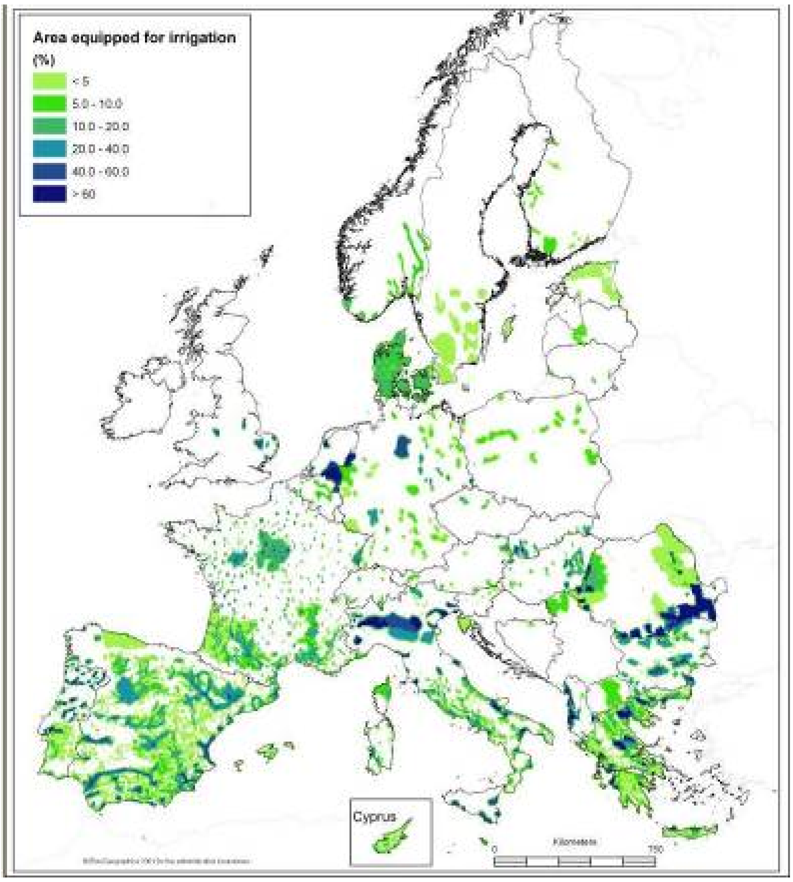
The percentage crop residue requirement for DNDC is defined by (Li, 2002) as the fraction of above-ground crop residue (leaves and stems) left as stubble or litter in the field. Crop residue incorporation rates are available from IPPC guidelines for estimating N_2O emissions due to Nitrogen content in 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 31).

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 31. Area equipped for irrigation (Siebert et al., 2005)



7.3.3 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 0. 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.

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

7.4 Landscape indicators (Maria Luisa 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.

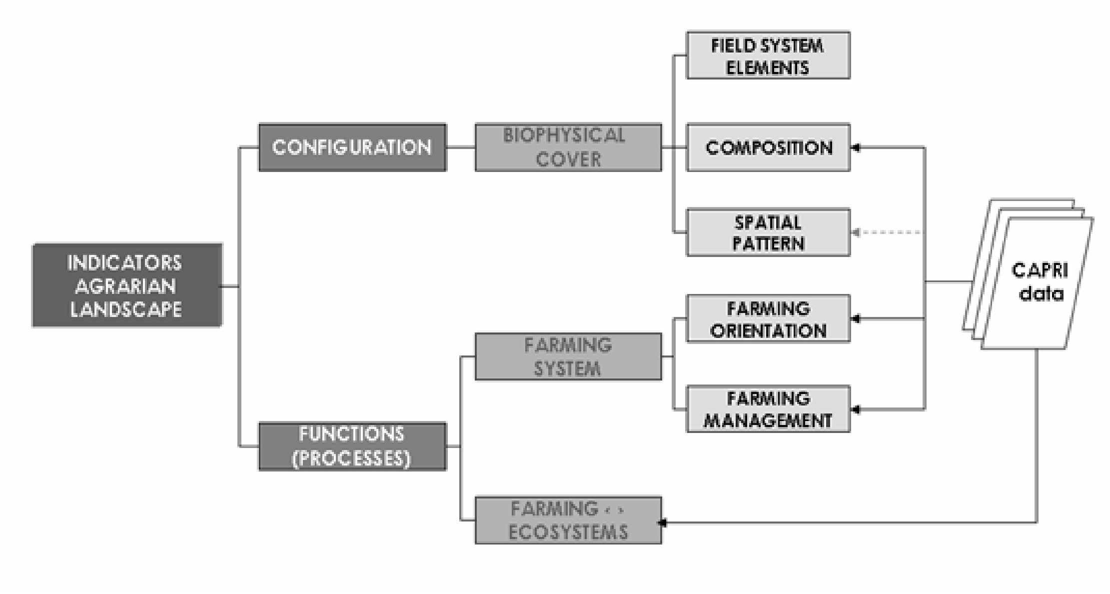
The added value of the new data layers is given by the fact that the information they provide was traditionally available at NUTS2 level; at such scale it is not possible to carry out assessments of the impact of the agricultural policy on the landscape, due to the large size of the administrative regions. 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.

7.4.1 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 32. 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.

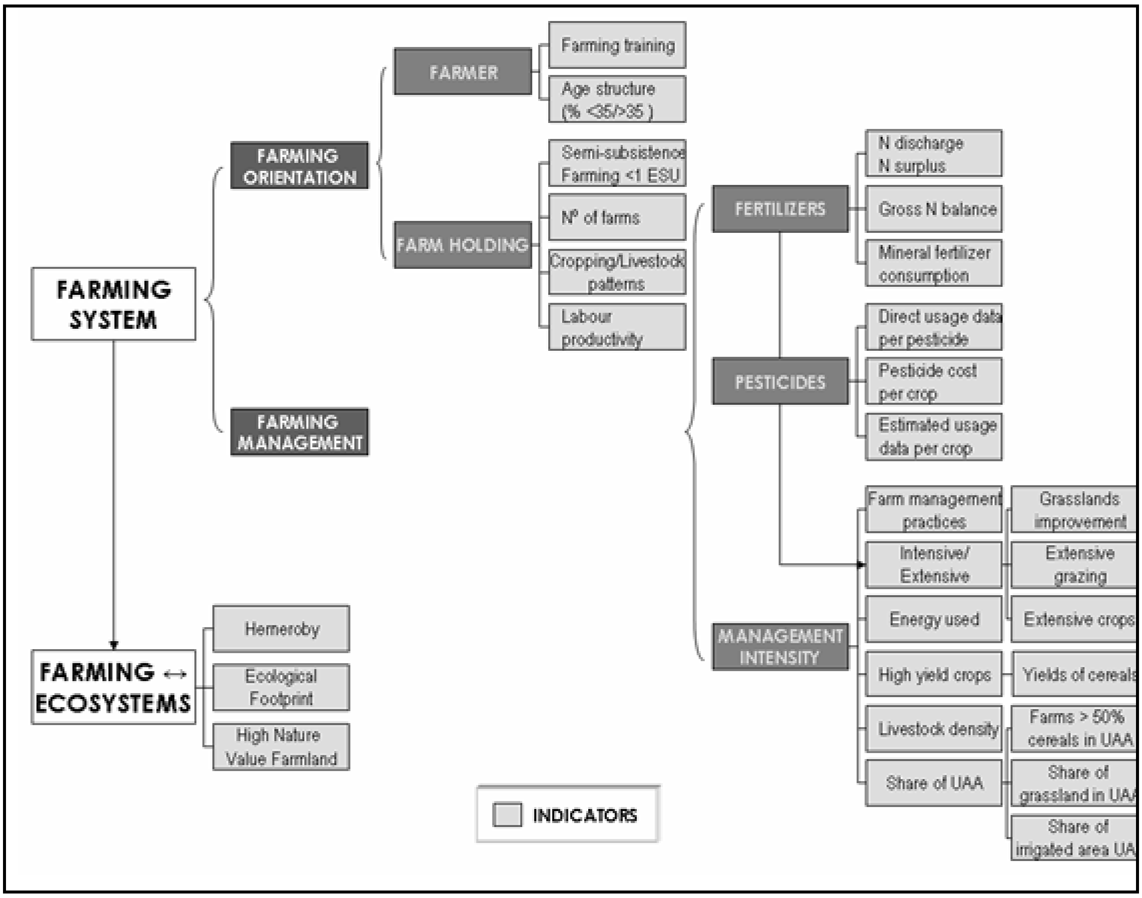
Indicators to be implemented in the CAPRI frame were selected according to their relevance for agrarian landscape, to their frequency in literature, and their potential implementation with CAPRI data.

Figure 32. 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 32). 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 33 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 33. Overview of indicators related to the farming system and to the interaction between farming and ecosystems



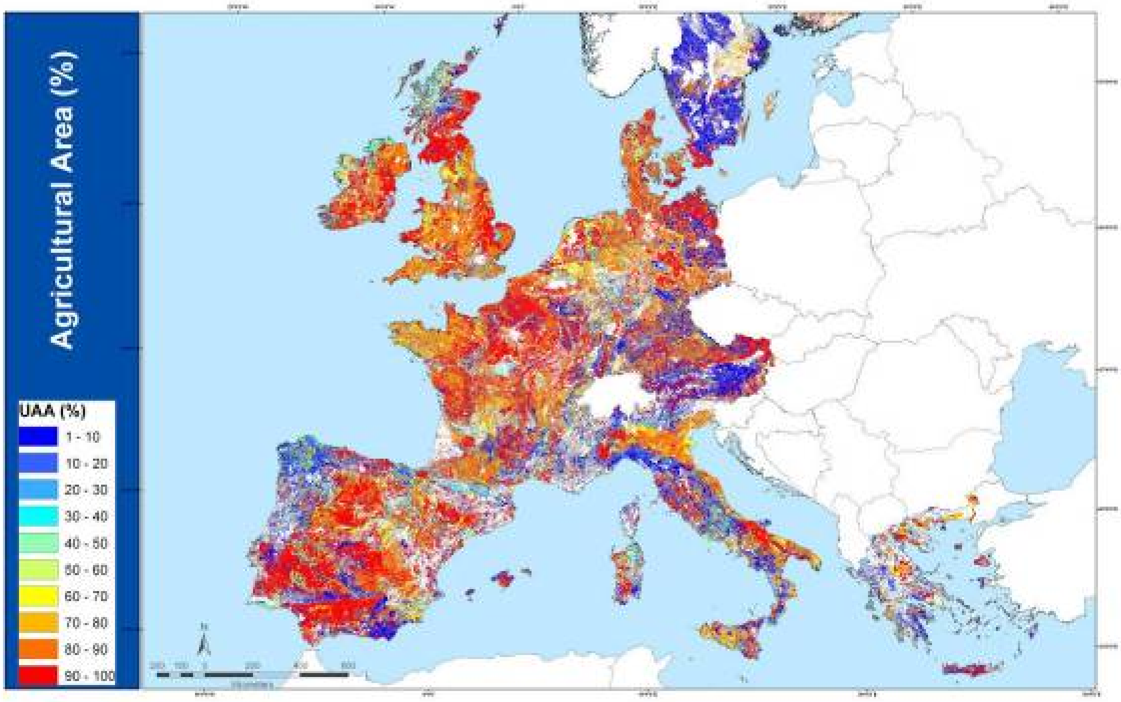
For completeness of information the field system indicators are also reported in fig.1, but the spatial resolution of 1 km and the information on biophysical cover expressed mainly as crop shares makes it not suitable for the retrieval of information regarding linear features or the historical parcel 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.

7.4.2 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 34. 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 35. Prevailing crop with reference to the UAA

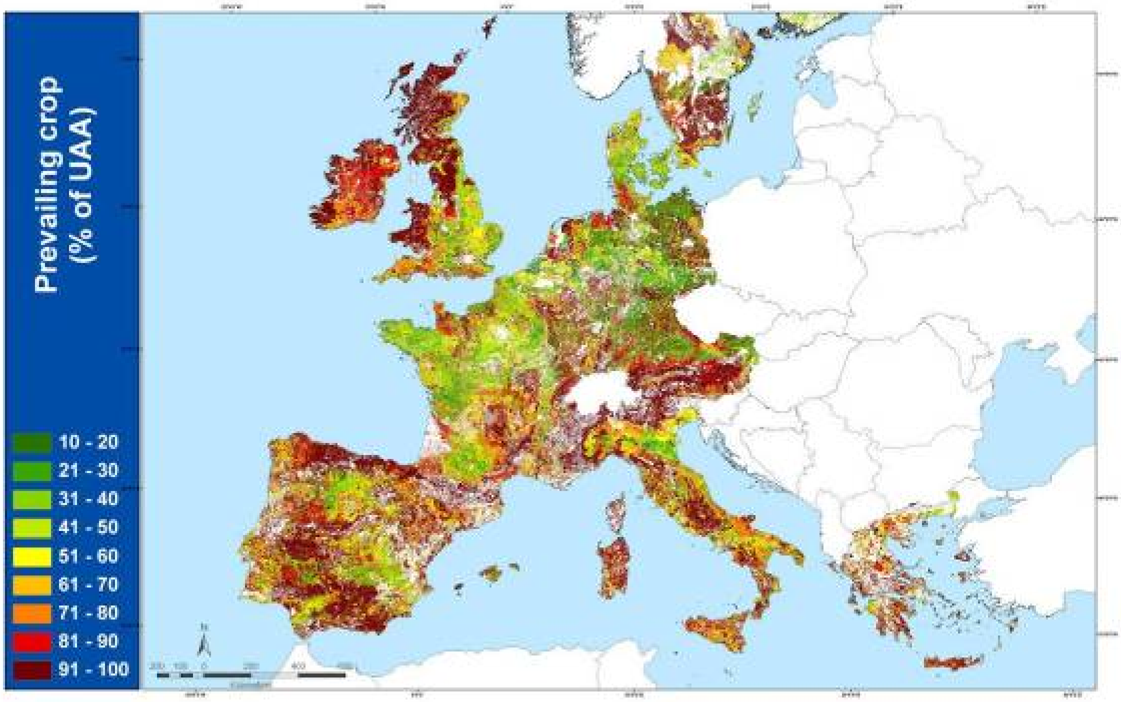
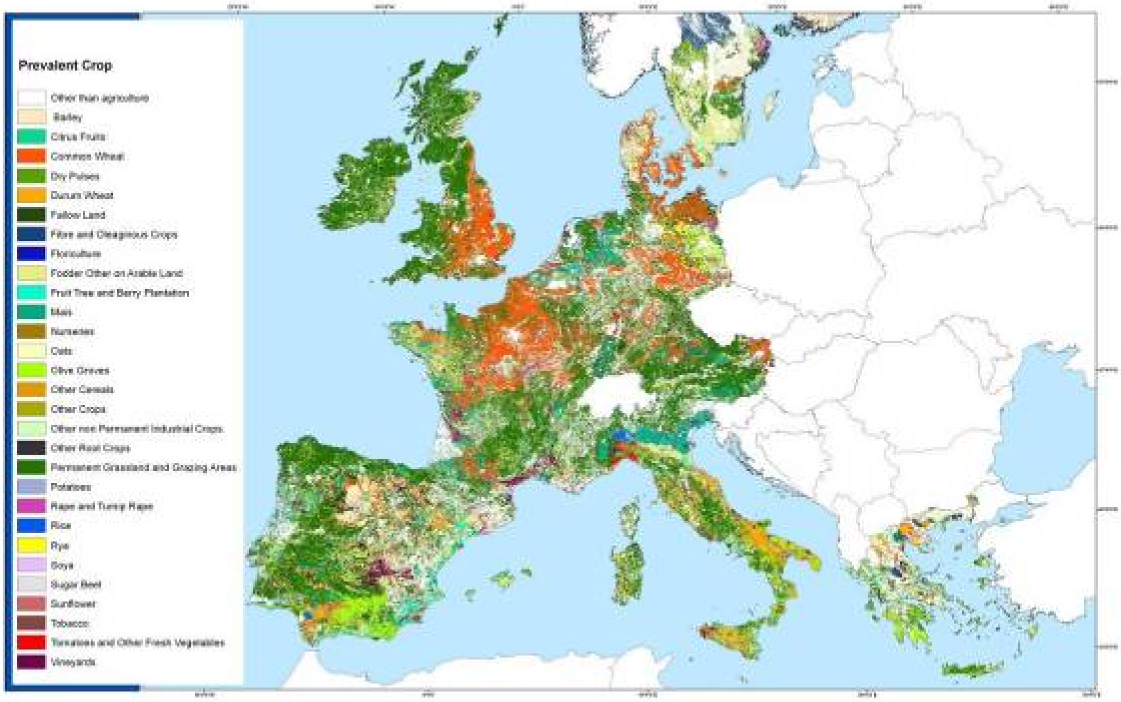


Figure 36. Prevalent crop



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 37. % of cereals in UAA

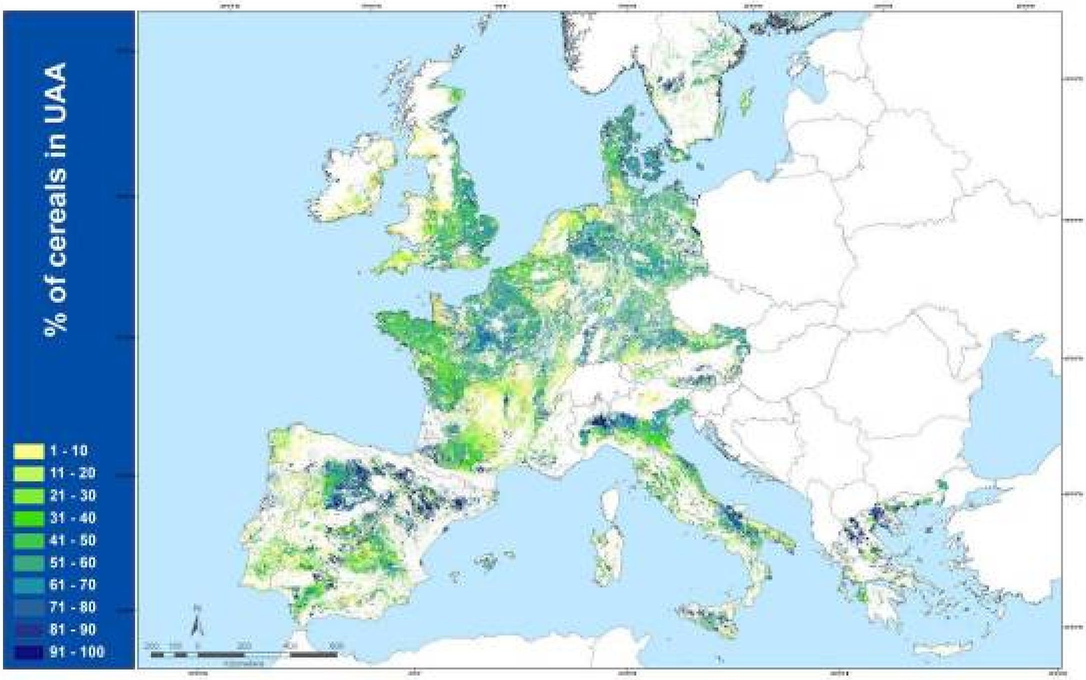
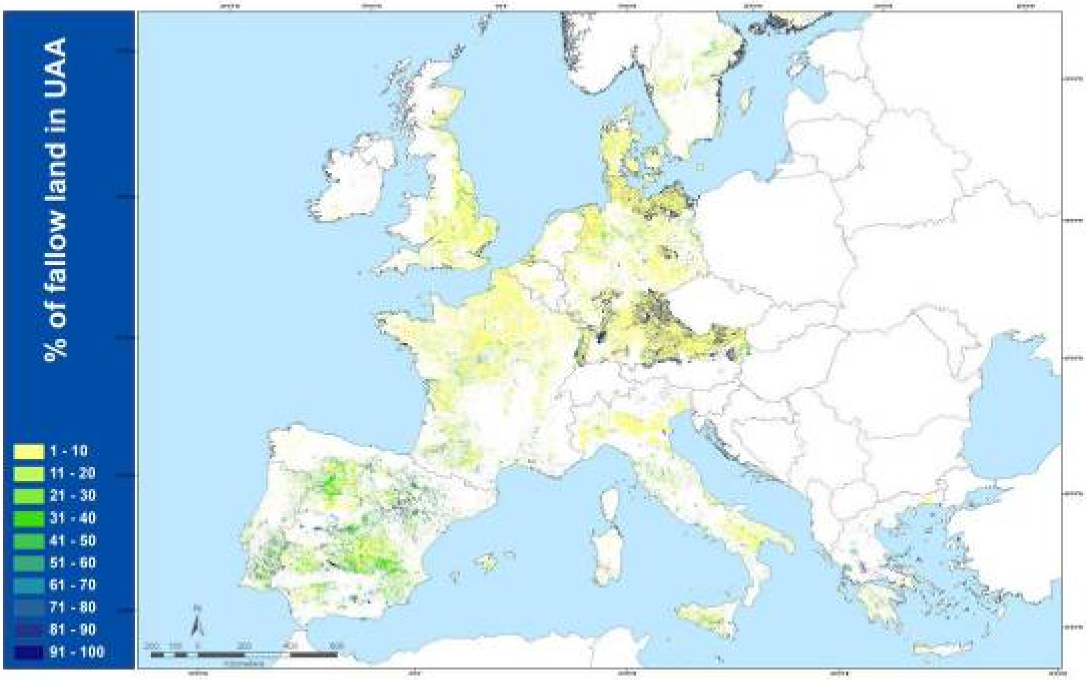


Figure 38. % of fallow land in UAA

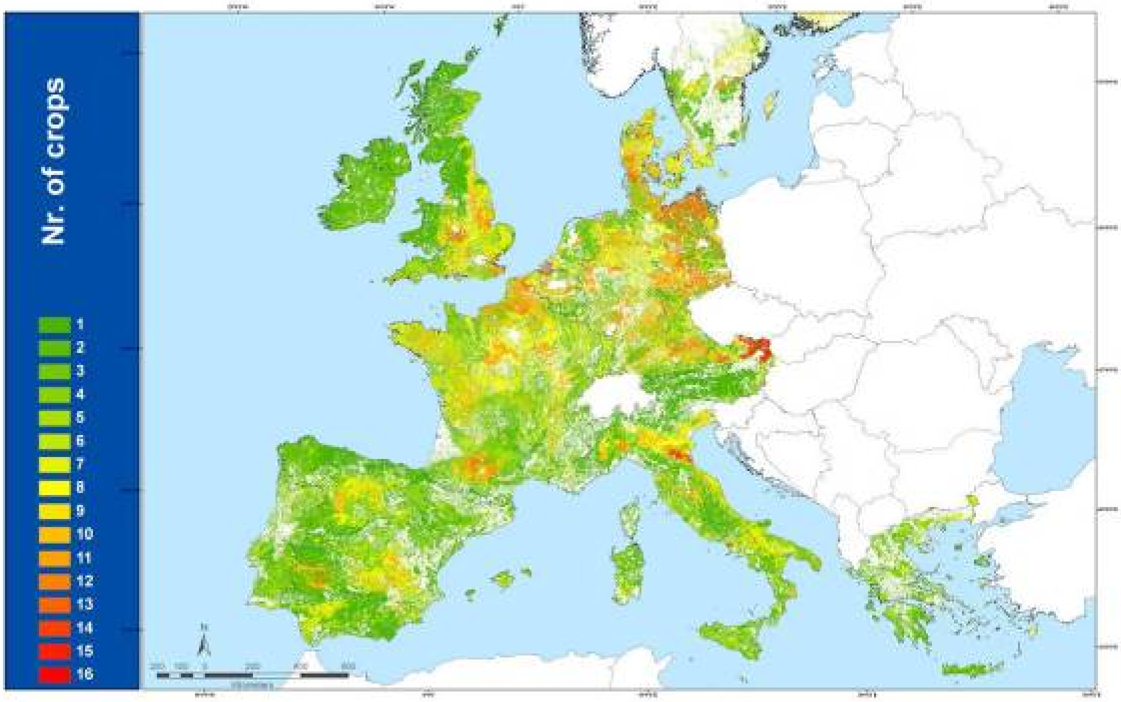


Crop diversity

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 39. Crop diversity (see text for explanation)



A more elaborated index is Simpson's diversity index:

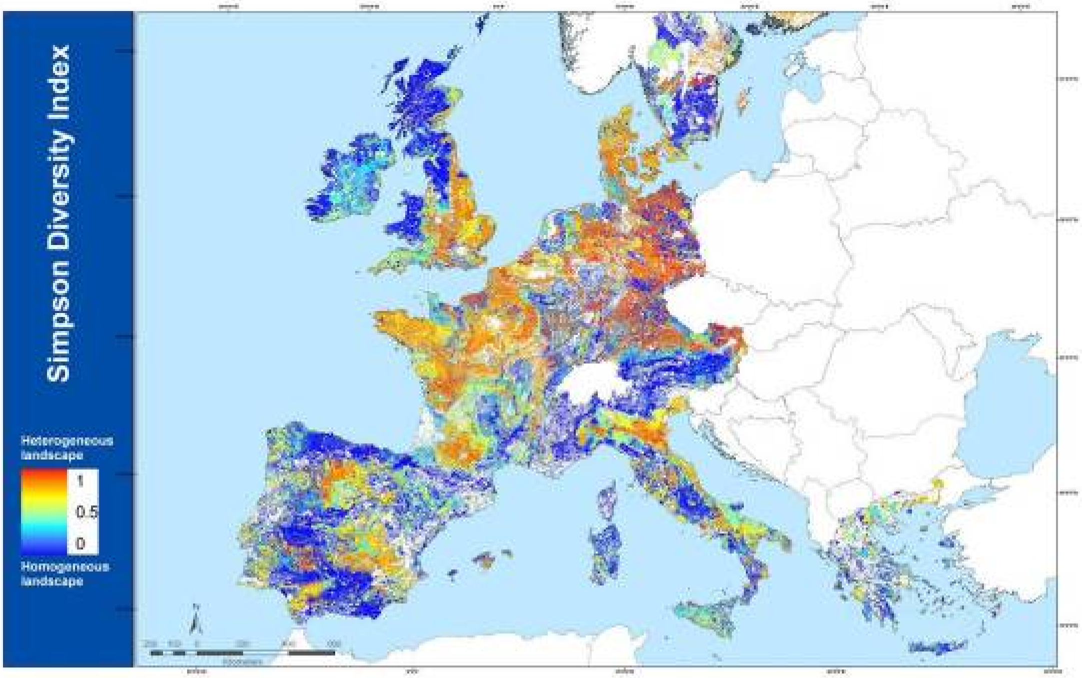
$$\mathrm{SIDI} = 1 - \sum_{i=1}^{m} P_i^2$$

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 40. 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.

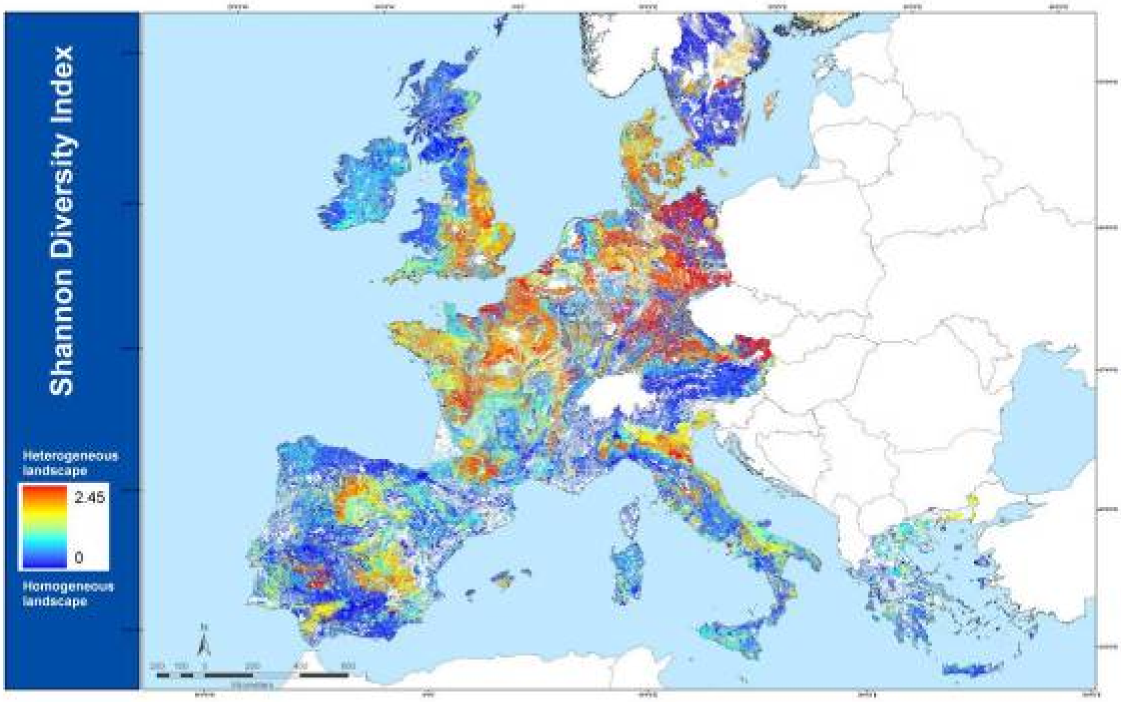
A third, well known but still discussed index is Shannon's Diversity Index

$$\mathrm{SHDI} = -\sum_{i=1}^{m} (P_i * \ln P_i)$$

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.

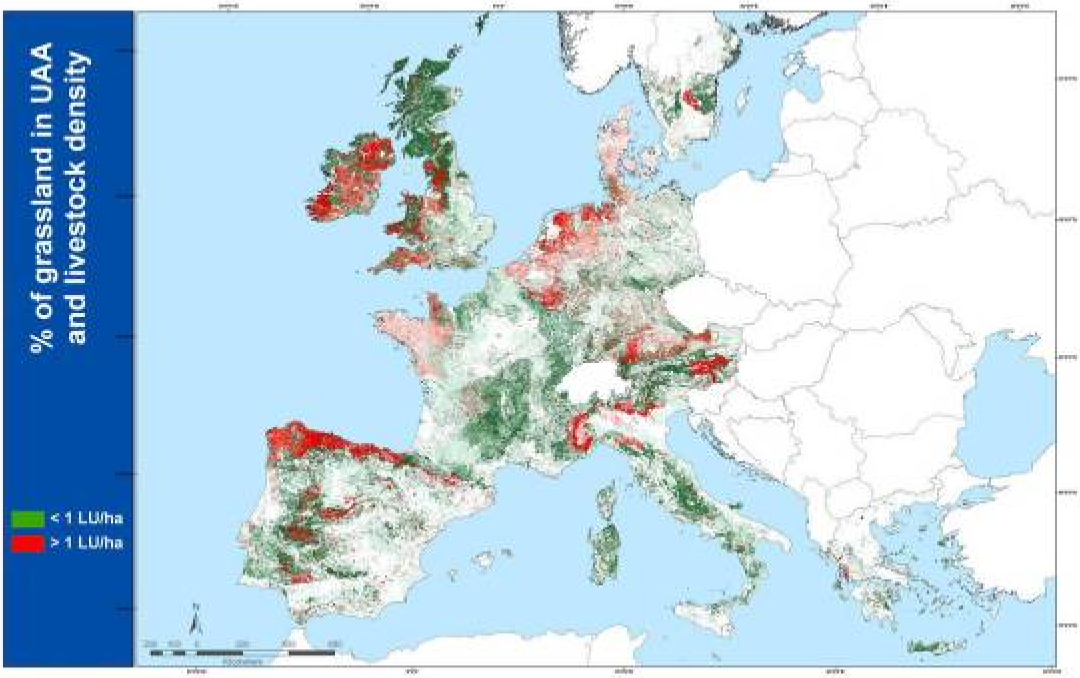
Figure 41. Shannon Diversity Index



7.4.3 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 42. % of grassland in UAA in areas with different livestock pressure



7.4.4 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:

$$\text{Index } 1 = 10 + ((1 - C1/UAA*10)) + (1 - (C2/UAA*10)) + \ldots$$

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 0

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
⬜	no direct correspondence

Figure 43. – Crop diversity and share of grassland index from French FSS statistics (left) and CAPRI data (right)

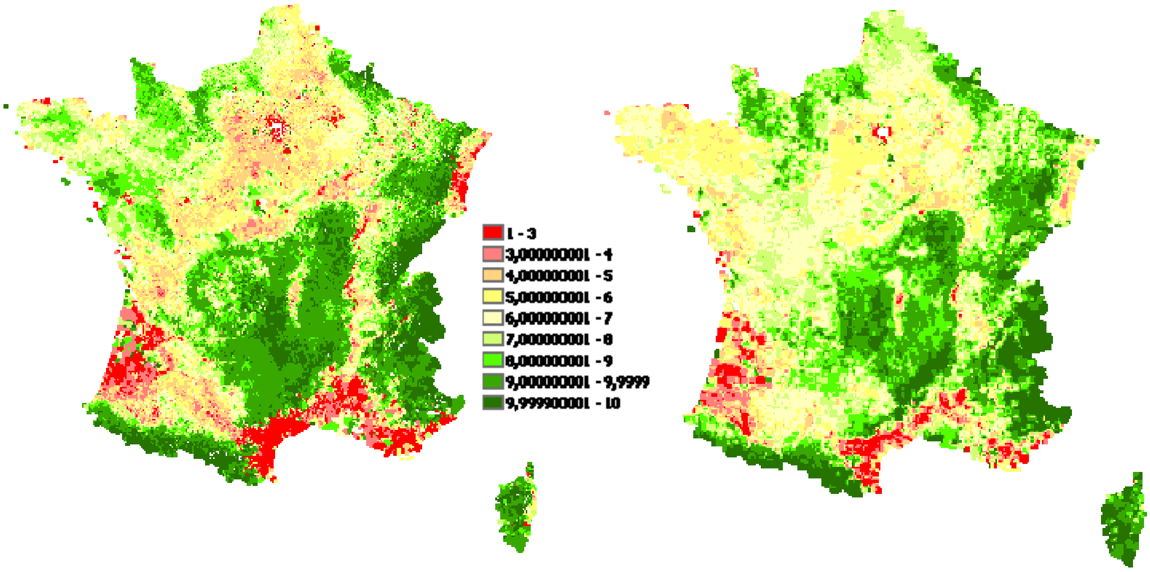


Figure 43 shows the results, obtained by calculating the index at NUTS5 level. The level of approximation is rather high, considering that in 95% of municipalities the difference in the score is equal or less than 3, and that in 87% of cases is equal or less than 2.

Discrepancies may be due to both to the differences in the aggregation of crops 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 0, the indicator in Figure 44.

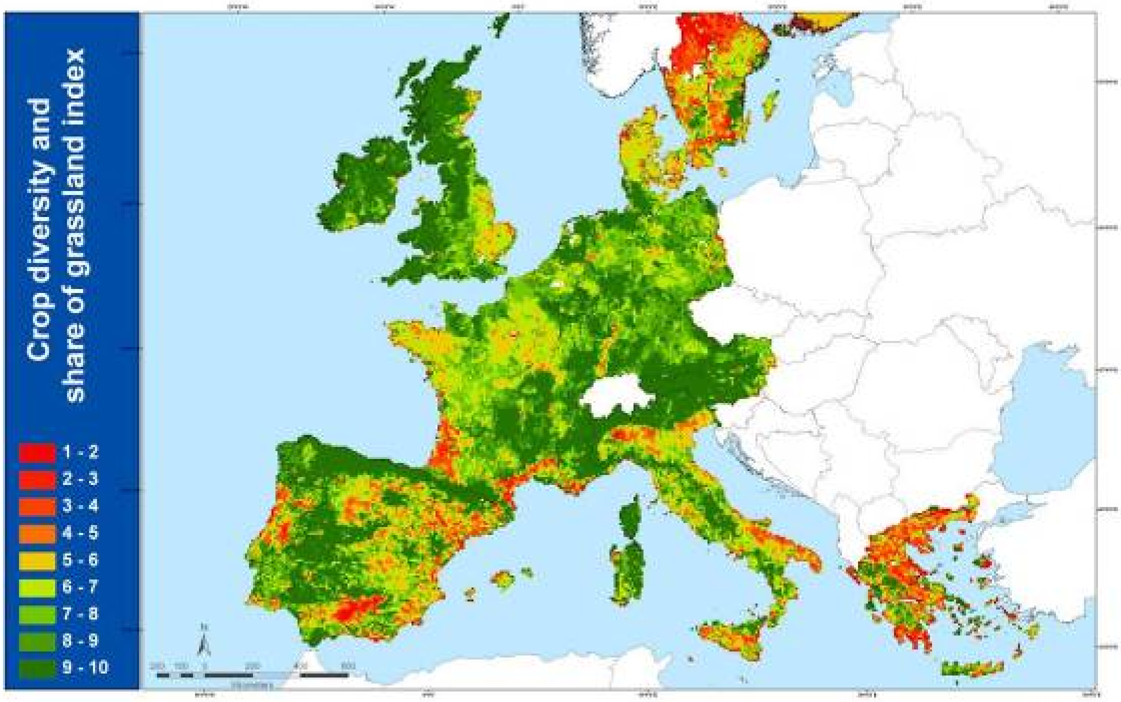
Crop classes used in the calculation of the EU15 indicator

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 44. 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 0.

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 45. Simulation of High Nature Farmland areas with CAPRI data, based on crop diversity and share of grasslands, and livestock density

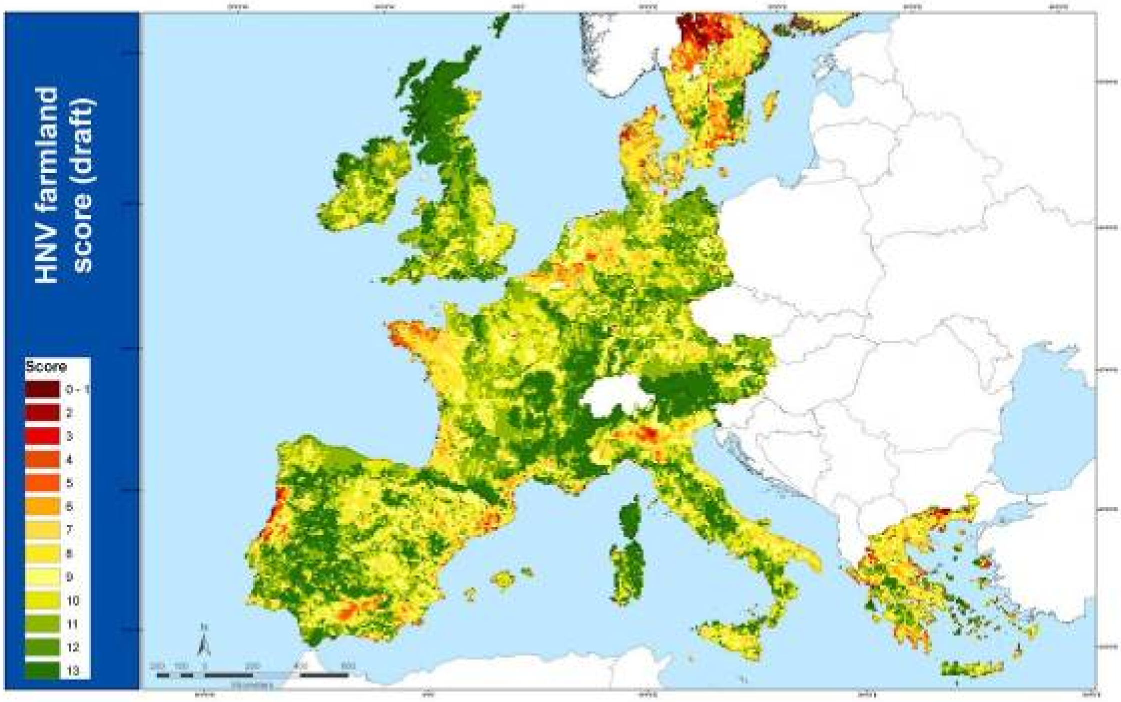
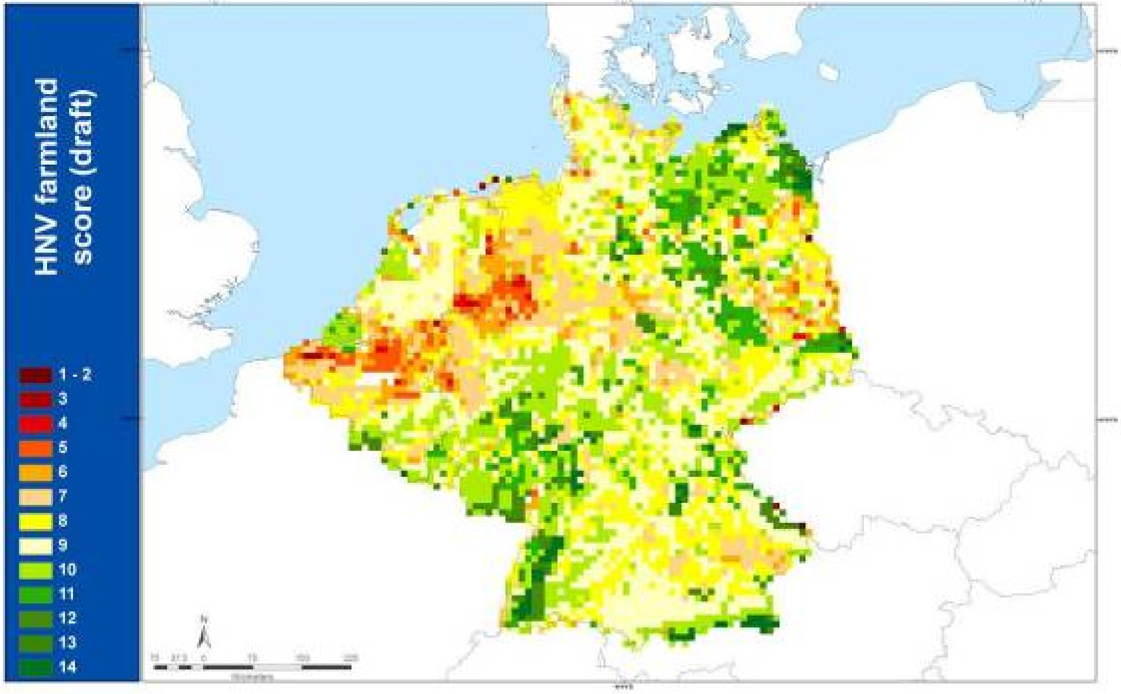


Figure 46. Simulation of High Nature Farmland areas with CAPRI data, based on crop diversity and share of grasslands, and N-surplus

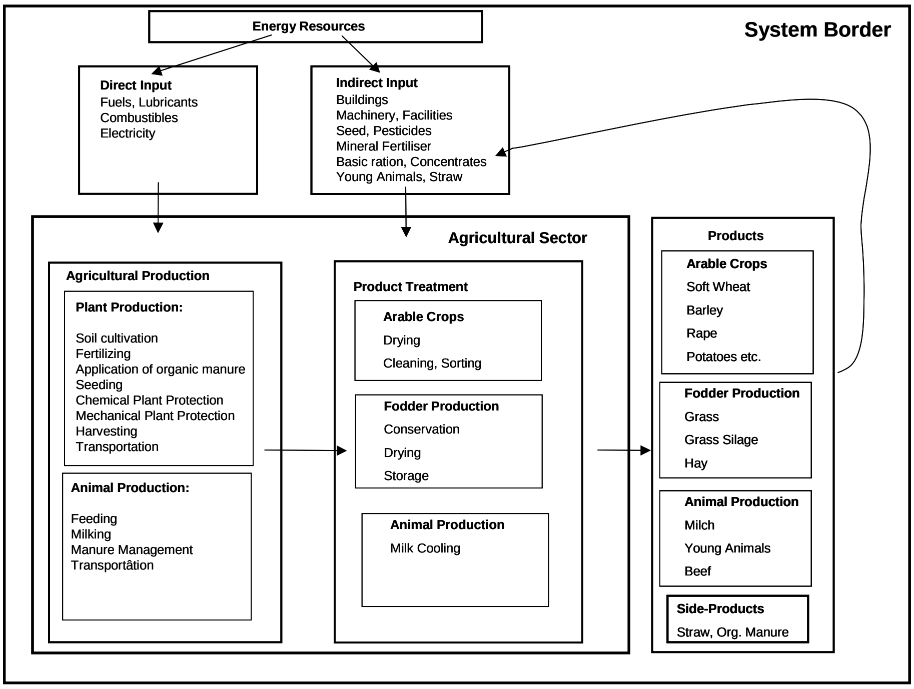


7.5 Energy Use in Agriculture (Tim Kränzlein)

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 47 shows the relevant borders and integrated processes for CAPRI energy module. The term “Agricultural Production” in Figure 47 is the main interface between LCA and CAPRI production activities.

Figure 47. 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). 0 shows the main direct energy sources used in CAPRI.

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 7.3.2.11. 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.

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 0. 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 0. 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 0, 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. Additionally 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 “Machinery”. 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 (0).

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 0. 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 0.

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. 0 gives an overview. Further down, each parameter is shown in more detail.

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 7.3.3) 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 –	MJ/€	The output level of a CAPRI activity is assessed by its energy content (See Chapter 7.3.3) whereas allocation	Aggregated on NUTS-II, NUTS-0 and EU

Type "Finance"		between main product and side-products is done for some activities. The result is divided by the income for the activity.	level
Energy balance	MJ	The output level of all CAPRI activities of a region are assessed by its energy contents (See Chapter 7.3.3) 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 shown on in detail.	Region-specific and activity-specific; weighted 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 shown in detail..	Region-specific and activity-specific; weighted averages on NUTS-0 level
Energy content products	MJ/kg product	On an activity-based level, the energy content for products can be shown; 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

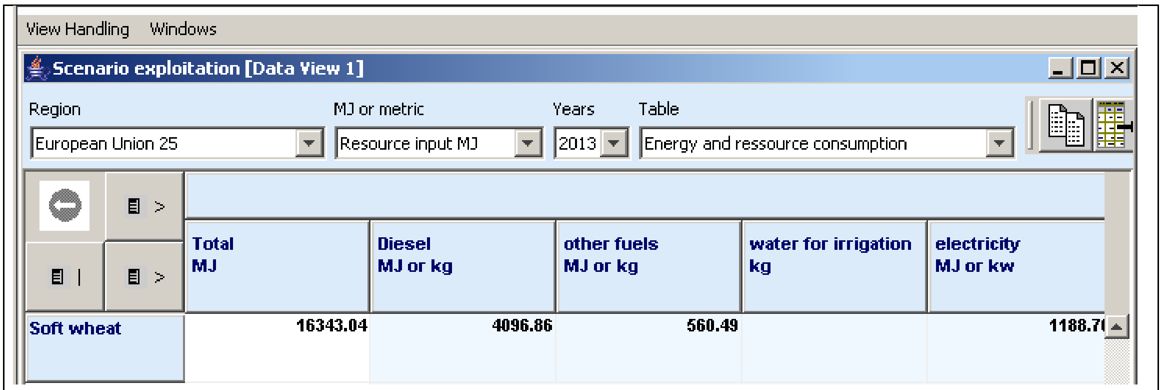
7.5.4.1 Application notice

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 48 shows some examples for display modes of the energy indicator.

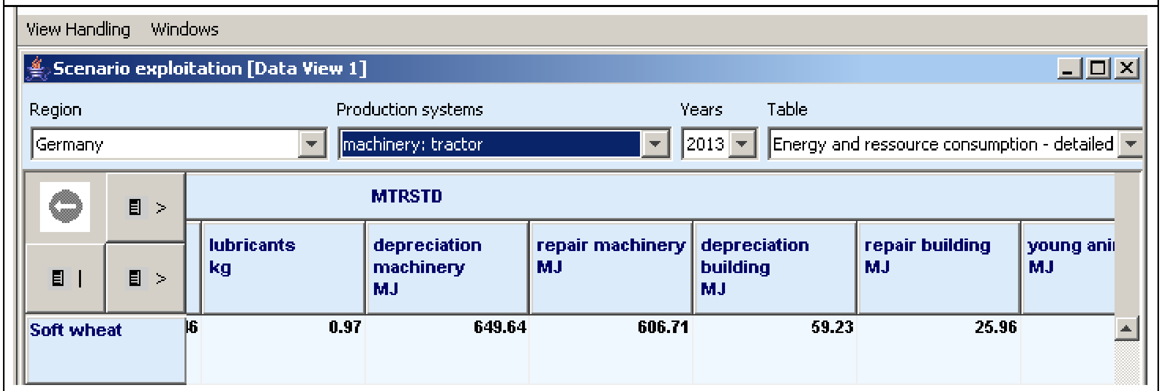
Figure 48. Energy parameters: examples for results displaying



Region: European Union 25
MJ or metric: Resource input MJ
Years: 2013
Table: Energy and ressource consumption

	Total MJ	Diesel MJ or kg	other fuels MJ or kg	water for irrigation kg	electricity MJ or kw
Soft wheat	16343.04	4096.86	560.49		1188.7

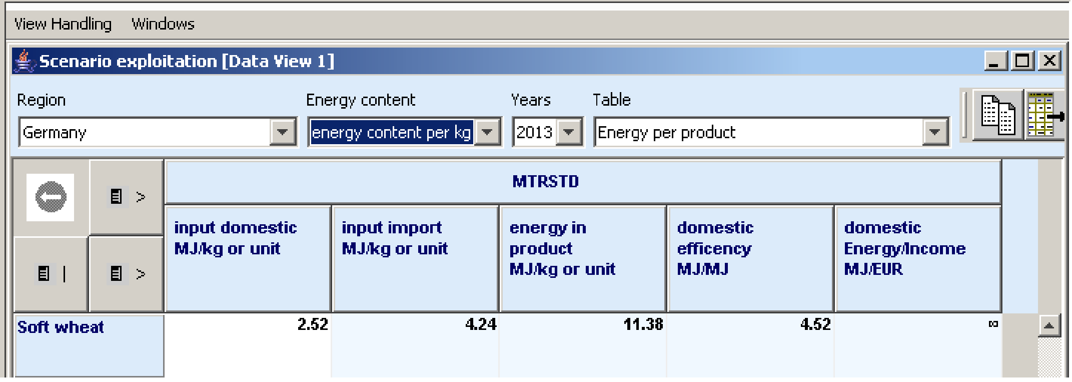
Example 1: Energy consumption - overview



Region: Germany
Production systems: machinery: tractor
Years: 2013
Table: Energy and ressource consumption - detailed

	MTRSTD						
		lubricants kg	depreciation machinery MJ	repair machinery MJ	depreciation building MJ	repair building MJ	young ani MJ
Soft wheat	6	0.97	649.64	606.71	59.23	25.96	

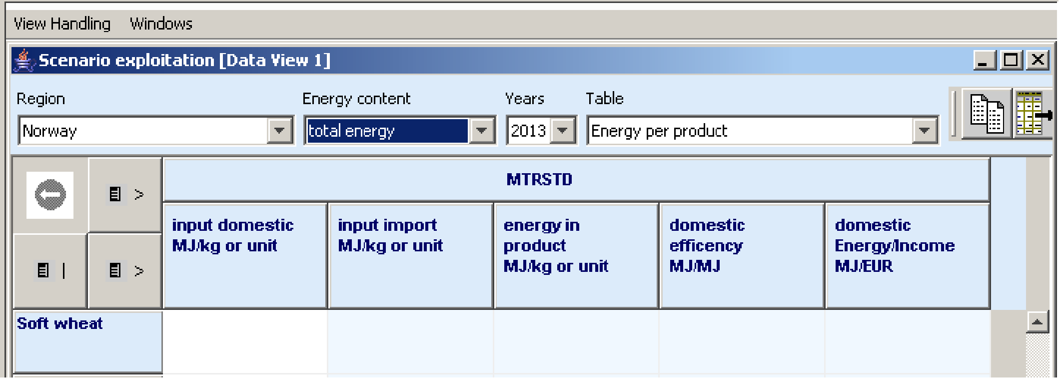
Example 2: Energy consumption - detailed



Region: Germany
Energy content: energy content per kg
Years: 2013
Table: Energy per product

	MTRSTD				
	input domestic MJ/kg or unit	input import MJ/kg or unit	energy in product MJ/kg or unit	domestic efficency MJ/MJ	domestic Energy/Income MJ/EUR
Soft wheat	2.52	4.24	11.38	4.52	∞

Example 3: Energy parameters with reference to the product



Region: Norway
Energy content: total energy
Years: 2013
Table: Energy per product

	MTRSTD				
	input domestic MJ/kg or unit	input import MJ/kg or unit	energy in product MJ/kg or unit	domestic efficency MJ/MJ	domestic Energy/Income MJ/EUR
Soft wheat					

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.

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

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

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

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

List of activities in the supply model

Group	**Activity**	**Code**
Cereals	Soft wheat Durum wheat Rye and Meslin Barley Oats Paddy rice Maize Other cereals	SWHE DWHE RYEM BARL OATS PARI MAIZ OCER
Oilseeds	Rape Sunflower Soya Olives for oil Other oilseeds	RAPE SUNF SOYA OLIV OOIL
Other annual crops	Pulses Potatoes Sugar beet Flax and hemp Tobacco Other industrial crops	PULS POTA SUGB TEXT TOBA OIND
Vegetables **Fruits** **Other perennials**	Tomatoes Other vegetables Apples, pear & peaches Citrus fruits Other fruits Table grapes Table olives Table wine Nurseries Flowers Other marketable crops	TOMA OVEG APPL CITR OFRU TAGR TABO TWIN NURS FLOW OCRO
Fodder production	Fodder maize Fodder root crops Other fodder on arable land Graze and grazing	MAIF ROOF OFAR GRAS
Fallow land and set-aside	Set-aside idling Non food production on set-aside Fallow land	SETA NONF FALL

Group	**Activity**	**Code**
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
	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

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 **Fruits** **Other perennials**	Tomatoes	TOMA
	Other vegetables	OVEG
	Apples, pear & peaches	APPL
	Citrus fruits	CITR
	Other fruits	OFRU
	Table grapes	TAGR
	Table olives	TABO

Group	Item	Code
	Table wine Nurseries Flowers Other marketable crops	TWIN NURS FLOW OCRO
Fodder	Gras Fodder maize Other fodder from arable land Fodder root crops Straw	GRAS MAIF OFAR ROOF STRA
Marketable products from animal product	Milk from cows Beef Pork meat Sheep and goat meat Sheep and goat milk Poultry meat Other marketable animal products	COMI BEEF PORK SGMT SGMI POUM OANI
Intermediate products from animal production	Milk from cows for feeding Milk from sheep and goat cows for feeding Young cows Young bulls Young heifers Young male calves Young female calves Piglets Lambs Chicken Nitrogen from manure Phosphate from manure Potassium from manure	COMF SGMF YCOW YBUL YHEI YCAM YCAF YPIG YLAM YCHI MANN MANP MANK
Other Output from EAA	Renting of milk quota Agricultural services	RQUO SERO
Inputs		
Mineral and organic fertiliser **Seed and plant protection**	Nitrogen fertiliser Phosphate fertiliser Potassium fertiliser Calcium fertiliser Seed Plant protection	NITF PHOF POTF CAOF SEED PLAP
Feedings tuff	Feed cereals Feed rich protein Feed rich energy Feed based on milk products Gras Fodder maize Other Feed from arable land Fodder root crops Feed other	FCER FPRO FENE FMIL FGRA FMAI FOFA FROO FOTH

Group	Item	Code
	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
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

Group	Item	Code
	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
	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

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}