seamless

System for Environmental and Agricultural Modelling; Linking European Science and Society

Quantitative models of SEAMLESS-IF and procedures for up-and downscaling

G. Flichman, M. Donatelli, K. Louhichi, E. Romstad, T. Heckelei, D. Auclair, E. Garvey, M. van Ittersum, S. Janssen, B. Elbersen.

Partners involved: IAMM, CRA, UMB, UBONN, INRA, NUI, WU, ALTERRA

Report no.: 17
November 2006
Ref: PD1.3.2
ISBN no.: 90-8585-044-4

Institut Agronomique Méditerranéen de Montpellier

CRA

UMB Norwegian University of Life Sciences

Logo's main partners involved in this publication

Sixth Framework Programme

SEAMLESS integrated project aims at developing an integrated framework that allows ex-ante assessment of agricultural and environmental policies and technological innovations. The framework will have multi-scale capabilities ranging from field and farm to the EU25 and globe; it will be generic, modular and open and using state-of-the art software. The project is carried out by a consortium of 30 partners, led by Wageningen University (NL).

Email: seamless.office@wur.nl
Internet: www.seamless-ip.org

Authors of this report and contact details

Name: G. Flichman, M.K. Louhichi Partner acronym: IAMM
Address: Mediterranean Agronomic Institute of Montpellier, 3191 route de Mende, 34093 Montpellier Cedex 5, France
e-mail: gflichman@iamm.fr, louhichi@iamm.fr

Name: M. Donatelli Partner acronym: CRA-ISCI
Address: Agricultural Research Council, Via di Corticella, 133, 40128 Bologna, Italy
e-mail: m.donatelli@isci.it

Name: E. Romstad Partner acronym: UMB
Address: Agricultural University of Norway, Dept. of Economics & Resource Management, PO Box 5033, NO-1432 Ås, Norway
e-mail: eirik.romstad@umb.no

Name: T. Heckelei Partner acronym: UBONN
Address: Bonn University, Nussallee 21, 53115 Bonn, Germany
e-mail: thomas.heckelei@ilr.uni-bonn.de

Name: D. Auclair Partner acronym: INRA
Address: Institut National de la Recherche Agronomique, 34398 cedex 5, Montpellier, France
e-mail: auclair@cirad.fr

Name: E. Garvey Partner acronym: NUI Galway
Address: National University of Ireland, University Road, Galway, Ireland
e-mail: eoghan.garvey@nuigalway.ie

Name: M.K. van Ittersum, S. Janssen Partner acronym: WU
Address: Wageningen University, Plant Production Systems, Haarweg 333, 6709 RZ Wageningen, The Netherlands
e-mail: martin.vanittersum@wur.nl , sander.janssen@wur.nl

Name: B. Elbersen Partner acronym: ALTERRA
Address: Alterra research institute, Droevendaalsesteeg 3, PO Box 47, 6700AA Wageningen, The Netherlands
e-mail: berien.elbersen@wur.nl

Disclaimer 1:

"This publication has been funded under the SEAMLESS integrated project, EU 6th Framework Programme for Research, Technological Development and Demonstration, Priority 1.1.6.3. Global Change and Ecosystems (European Commission, DG Research,

Disclaimer 2:

Within the SEAMLESS project many reports are published. Some of these reports are intended for public use, others are confidential and intended for use within the SEAMLESS consortium only. As a consequence references in the public reports may refer to internal project deliverables that cannot be made public outside the consortium.

When citing this SEAMLESS report, please do so as:

Flichman, G., Donatelli, M., Louhichi, K., Romstad, E., Heckelei, T. et al. 2006. Quantitative models of SEAMLESS-IF and procedures for up-and downscaling, SEAMLESS Report No.17, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 112 pp, ISBN no. 90-8585-044-4.

Table of contents

General information

Task(s) and Activity code(s):	Task 1.3, Activity 1.3.2
Input from (Task and Activity codes):	Task 3.2, Task 3.3, Task 3.5, Task 3.6, Task 3.7, Task 3.8, Task 3.9
Output to (Task and Activity codes):	Task 1.4
Related milestones:	M1.3.2

Executive summary

This project deliverable presents the quantitative models of SEAMLESS-IF and the procedures for up- and downscaling.

SEAMLESS has a large number of different models and processing tools. These models are different concerning the methods that are used and the scales. This is the reason why there are specific models and processing tools just for permitting the linkage between some of the basic models.

The Quantitative Models in SEAMLESS are:

1. APES (biophysical model, creates information used in FSSIM);
2. FSSIM (farm bio-economic model, creates information used in CAPRI and other models);
3. CAPRI (agricultural sector model, creates information used in FSSIM - feedbacks - and other models);
4. GTAP (global general equilibrium model, creates information used in other models);
5. ECONOMETRIC LABOUR MODEL (econometric model to estimate labour impact of policies);
6. EXPAMOD (econometric model to interpolate results obtained in a sample of Regions and Farms to the whole EU, links FSSIM results to CAPRI);
7. TERRITORIAL MODELS (model used to analyse the impact of policies in regions defined out of their territorial attributes);
8. STRUCTURAL CHANGE MODEL (used to forecast farm structural changes in size and orientation);
9. DEVELOPING COUNTRIES MODELS (agricultural sector models for developing countries).

In some cases, there are feedbacks between the models, in others; the results obtained using some models are used as input by others without feedback.

1 Introduction

In this document there is a description of the different models that will be used in SEAMLESS-IF.

One of the more difficult challenges in SEAMLESS-IF is related to the different disciplines, methods and scales that are used. In this document we provide the definitions of:

- Models that are used, their principal features, scientific disciplines involved;
- Scales that correspond to each type of model;
- Flows of information between the models;
- Processing Tools used to facilitate and adapt the information flows between models and data-bases and models;
- Methods for linking the different scales;
- Methods for providing a spatial dimension to results;
- Necessary data to make use of the models;
- Indicators able to be calculated by the models;
- Indicators that will be used by models.

The different scales and methods require specific linkage procedures. In many cases, it is not possible to make direct "aggregation" for moving from one scale to another one, because the models that are used in the first level and those used in the second one apply different methods, have different specifications of variables. This is the reason for creating intermediate models that allow those linkages.

2 The models [1]

2.1 APES[2]

The Agricultural Production and Externalities Simulator is a modular simulation model targeted at estimating the biophysical behavior of agricultural production systems in response to the interaction of weather and different options of agro-technical management. Although a specific set of components will be available in the first release, the system will be built to incorporate, at a later time, other modules which might be needed to simulate processes not accounted in the first version.

Using mostly modelling approaches already made available by research and previously tested in other simulation tools, APES will run at a daily time-step in the communication among components. This means that the rate variables estimated by components will use as time unit day-1; however, some components may use a different time-step internally (e.g. soil water). APES is meant to work at **field scale**, simulating 1-D fluxes (a second version may use 2-D fluxes to account for multiple cropping (e.g. vineyards and grasses).

The processes are simulated in APES with deterministic approaches which are mostly based on mechanistic representations of biophysical processes. The criteria to select modelling approaches will be based on the need of: 1) accounting for specific processes to simulate soil-land use interactions, 2) input data to run simulations, which may be a constraint at EU scale, 3) simulation of agricultural production activities of interest (e.g. crops, grasses, orchards, agroforestry), and 4) simulation of management implementation and its impact on the system.

A key aspect of APES is the simulation of management, which requires both models (rules) for management application, and models to simulate the impact on the system of management events. Management is an input for APES, meaning that the management strategy (production technology) is decided a priory and it is converted into atomic operations which are simulated to occur in the production enterprise under evaluation, under specific combinations soil-weather.

The deterministic models of APES will be run in a stochastic fashion by using long series of weather data to account for climatic variability. Outputs will be then available, for each production activity, as average responses and as variability associated to the responses.

The expected use of APES output is in terms of production activity/yearly values, whereas daily outputs will be of interest mostly for APES calibration and evaluation. Production activities and yearly summaries will be used to estimate technical coefficients in the TCG (Technical Coefficent Generator, WP3, task 3.3). It should be pointed out that APES outputs should be evaluated primarily in terms of comparisons among systems rather than as absolute values. A more complete description of APES, inclusive of the references relevant to the release version 0.3, is available in the deliverable PD3.2.19

The following paragraphs contain a summary description of the current version of APES model components and of Modcom (figure 1), which is the simulation engine used to link components. First, a UML (Unified Modelling Language) component diagram of APES is

[1] Part of the models' descriptions are excerpts from the DOW or from previous PDs

[2] For more detailed information concerning APES, see PD3.2.19.

presented. A more complete description of components, inclusive of links to public documentation when released, is available in the deliverable PD3.2.18

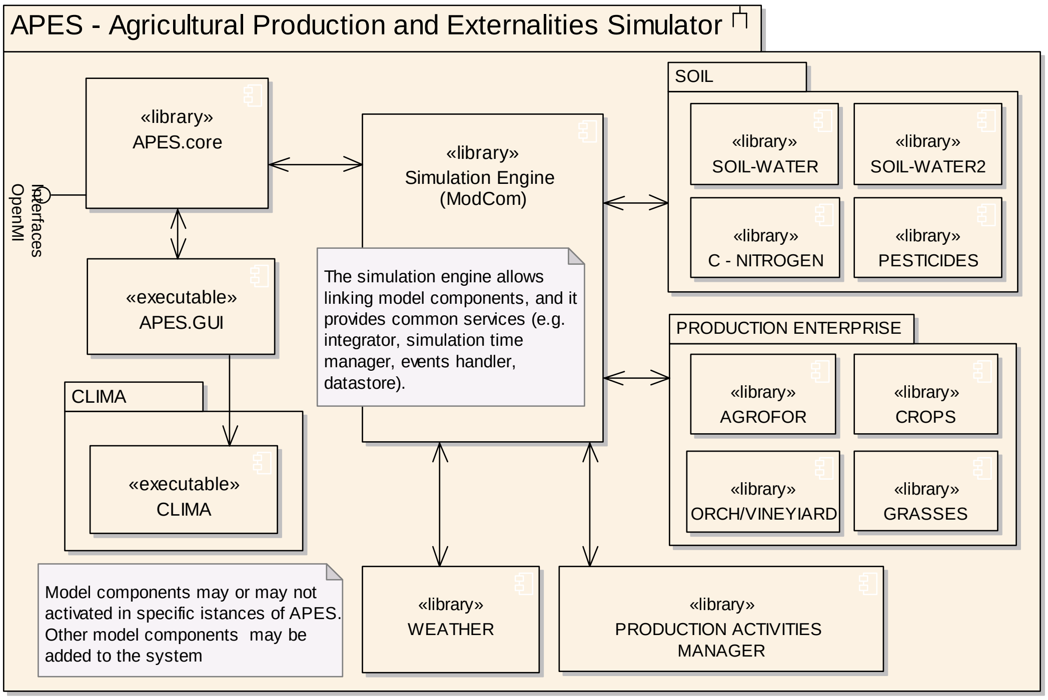


Figure 1 APES component diagram. APES is composed of three main groups of software units: the graphical user interface and the core services component to run Modcom; the simulation engine Modcom, and the model components. Model components can be grouped as soil components, production enterprise components, weather and agricultural management. During implementation, some of the components of the figure have been further splitted, as in tha case of WEATHER and SOIL-WATER, as detailed in the text.

2.1.1 The Soil water and Soil erosion components

2.1.1.1 Models

The SoilWater component describes the infiltration and redistribution of water among soil layers, the changes of water content, fluxes among layers, the effective plant transpiration and soil evaporation, and the drainage if pipe drains are present. Two algorithms have been selected to simulate the water dynamics, a cascading algorithm and a cascading with travel time among layers. The cascading method simulates the soil as a sequence of tanks that have a maximum and a minimum level of water, fixed respectively at the field capacity (FC) and wilting point (WP). Water in excess to the water content at FC for a given layer is routed into the lower layer, and if all the profile has reached the FC, the water in excess is removed from the soil as percolation. The main advantage of this approach is the simplicity and the calculation speed. The main difficulties are that the model has not a strong physical background, because the concept of field capacity is arbitrary and represents a simplification of soil water holding features, and because the time needed to water to move between layers is not considered. Other relevant difficulties of this approach is the impossibility to have soil water contents greater than FC and lower than WP (the latter with exception of the evaporative layer), and the possibility to have allowed movement of water only downwards. This approach is not suitable in presence of layers of different texture and/or water table, even if it is possible to use some approximation to simulate the capillary raise. The cascading method with travel time is an extension of the simple cascading, taking into account the time needed to percolate the layer. Tillage simulation is done following the approach of the models Wepp and SWAT, where each equipment used on the soil have specific parameters and coefficient for the intensity of tillage (mixing among layers), for surface roughness after tillage, ridge high and distance This allows for the simulation of the evolution of bulk density in time, because also a simple model of soil settling after tillage was developed.At the actual stage, all the variables are simulated with daily time step, but algorithms and software structure are ready to work with an hourly or shorter time step.

The SoilErosionRunoff component simulates dynamically the water runoff and the soil erosion. In detail, it represents the runoff volume, the amount of soil eroded, the interception by vegetation, and the water available for infiltration. This component has been structured in a hierarchical way with the above-described Water component, but has its own data-type and related interfaces. As for the Water component, all the variables are simulated using a daily time step, but the algorithms and software structure are already designed to work with an hourly or shorter time step.

2.1.1.2 Current implementation

Soil water and erosion components are developed for the .NET framework, using the C# language. Components are developed according to "Developing Biophysical Models as Components in .NET" work paper. Components are organized in sub-components (water dynamic simulation, runoff simulation, tillage, soil temperature simulation) and within each component there are "strategies" to compute output variables. For several processes we have developed or are under development different strategies. The choice of different strategies does not imply the recompilation of DLL, as the development of new strategy from the client side. A version that works in the ModCom environment, under continuous updating, already linked to other components (clima and crop) in the APES context, linked is available for the WP 3.1 members in the CVS used for APES development.

2.1.2 The Soil water2 component

2.1.2.1 Models

All the physical properties of a soil come from the interaction of the soil structure, its hierarchical units, with the water which is moving between and within them. The soil structure is an organization of the solid phase, nested into several scales levels, which will be considered here as a "container", spatially organized and referenced. Its principal variable is its structural specific volume Vms = V/Ms (inverse of the "apparent" density: Ms= mass of solids in the internally structured volume V of the soil medium).

This container contains water and air that are distributed through the soil structure, and interact with its elements and circulate between them. The module Kamel simulates dynamics of both soil structure and soil-water, interacting together.

The maximum soil depth simulated is fixed to 1.2 meter. It is composed by a surface layer and 4 top-bottom superposed horizons named A1, A2, B1, and B2. The zone bellow horizon B2 can be considered either as a crust (flux =0) either infinite (same parameters as B2). The surface layer objective is to reproduce the impact of technical practices as tillage or effect of a crust on water infiltration and evaporation where surface hydraulic conductivity, layer thickness and maximum surface storage are the three principal modified factors. Each horizon is a homogeneous zone, in term of structure and organisation, called pedostructure. Sixteen physical parameters grouped by functionality are used for its description: organizational parameters that are provided by the soil characteristic shrinkage curve, and the functional parameters of : i) the water potential curves for both micro and macro pore systems, ii) the conductivity curve for the inter-pedal pore space (macro-porosity) and iii) the swelling curve (volume change with time) corresponding to the absorption of water by swelling aggregates immersed in water.

The soil is discretized by 10 layers. To preserve a modelling logic between layers and horizons, the height of each layer is determined by the model using the height of the horizons provided by the user (HorizonDepthA1, HorizonDepthA2, HorizonDepthB1, and HorizonDepthB2). The equation used allows the uniformity of the layer's height in each horizon and differences between horizons. The depth (or height) of each layer is furnished in the output variable LayerDepth. The release provided at the end of January 2006 considers that each layer sate is homogeneous and there is no source of heterogeneity generated using drip irrigation for example. The initial water content of each horizon is furnished by the user using hydrostructural state .

Parameters estimation
All parameter inputs are the physically based parameters of the four pedostructure characteristic curves, namely, the shrinkage curve, the tensiometric curve, the swelling curve (soil specific volume function of time in immersed condition) and the conductivity curve. These curves are measured in laboratory but can also be estimated using pedotransfer functions [5, 6]. Examples of hydro-structural parameter data set for different soil types (USDA classification) are provided in the appendix. Interface software for estimating the

2.1.2.2 Current implementation

The component is being implemented using C#.

2.1.3 The Soil Carbon and Nitrogen component

2.1.3.1 Models

The nitrogen and carbon dynamics are described in the routines of the Soil Carbon-Nitrogen component, for which the SUNDIAL model is used as a baseline. The SUNDIAL model simulates all of the major processes of C and N turnover in the soil/plant system using only simple input data. This feature makes this model an ideal choice to be implemented as base for the C and N modelling in the current framework. In SUNDIAL, the microbial processes of carbon and nitrogen turnover are described together with mineralization and immobilisation occurring during decomposition of soil organic matter. Furthermore, the bypass flow following addition of fertiliser, the nitrification of ammonium to nitrate, and the nitrogen losses by denitrification are also represented in details. In synthesis, therefore, the model should:

- simulate microbial and physical processes influencing the C and N content of the soil, greenhouse gas emissions and leaching losses from the soil;
- allow addition of C and N to the soil as crop residues, organic manures, fertilisers and atmospheric deposition using information supplied by other components;
- use input information about the soil water and temperature provided by other components to simulate the microbial and physical processes of C and N turnover and loss;
- output the distribution of mineral N down the soil profile, so that other components can determine the availability of N to a plant root at a given depth;
- output the nature of losses of C and N from the system so that pollution events can be investigated.

2.1.3.2 Current implementation

The SUNDIAL soil C and N routines have been modularised so that they are separated from crop, water and cultivation routines. The initialisation, addition, microbial and physical processes are distinct in the new code. A C# version of the code has been completed as a Modcom class and included in the release 0.3 of APES.

2.1.4 The Pesticides component

2.1.4.1 Models

The AgroChemicalsFate component predicts the fate of agrochemicals in the environment. The model considers 5 compartments where pesticide is stored: canopy surface, plant, available fraction of the soil, aged fraction of the soil and bound fraction of the soil, even though it is possible by strategies to exclude the bound and aged fractions. The available fraction is partitioned in 3 phases: gas, liquid, and solid.

Models are implemented in four composite strategies:

- Air
- Crop
- Canopy
- Soil

The air strategy considers the processes that occurr before the pesticide reaches the soil, and it simulates the processes of drift and plant interception. The applied pesticide may be deposited on soil surface, lost in drift, or intercepted by the crop canopy. The pesticide on a crop canopy can be volatilized or degraded, penetrate into the leaves, or washed off to the ground by rainfall or irrigation. From the surface, the chemical may enter the soil system transported by infiltrating water and is partitioned among the gas, liquid and solid phases of the soil. The soil compartment is divided in two parts, the first represents the process over the soil surface, the second describes the soil profile. Chemicals are degraded in the soil profile by chemical, photochemical and microbial processes and might be taken up by plant roots.

The component has to be linked to other components to run and to describe the behaviour of pesticides in of the modelled system. It is well known that the main determinant of pesticide flow along soil profile is advection. It is necessary, therefore, that the component reads information about water content and water fluxes from the soil water component. Soil has to provide also temperature because several processes are affected by it. The crop strategy requires information about the crop, in particular about ground cover to estimate crop interception of pesticides during application.

2.1.4.2 Current implementation

The approach used in developing the Pesticide Component rely on a three tier structure: a macro-level that describes the macro-structure of the component and which is made of several sub-components; each sub-component represents a particular environmental compartment and it contains the description of several pesticide interactions with each compartment; finally, each interaction can be quantitatively calculated using different approaches.

Users can configure the Pesticide Component to follow a specific pattern in order to obtain the requested variable(s); this can be done selecting the desired strategy for each interaction between environmental compartments, building the so-called "computation chain".

The pesticides component is written in C# language version 1.1 and is avaialble both as a stand alone component and as a Modcom component implemented in APES.

2.1.5 The Crop component

2.1.5.1 Models

The LINTUL model has been implemented in the current framework to simulate the biomass production as a function of intercepted radiation and its conversion efficiency. The crop growth is limited by two factors, the water stress and the nitrogen limitation. Water stress is modelled via the ratio between actual and potential transpiration; when a water stress event occurs, the simulated crop allocates more biomass to the roots and less to the shoot in order to increase the potential access the soil water. The simulation of nitrogen stress follows the growth dilution concept as implemented in the crop model CropSyst. Radiation use efficiency is reduced by a fraction when the available percentage of nitrogen is between the minimum nitrogen requirement and the critical nitrogen requirement.

The crop model is linked to the nitrogen turnover assuming that roots uptake the required nitrogen over the whole soil profile implying that only one dynamic soil layer need to be considered. The nitrogen model, however, divides soil horizons into a number of discrete fixed model layers. Therefore, given a certain depth of the roots, the nitrogen model should provide an average nitrogen concentration to the crop model.

The model reacts also to the irrigation and fertilization regime, including soil nitrogen mineralization, which depends on soil temperature. Since the susceptibility of crops to water and nitrogen availability depends on crop development stage, the impact of different management strategies could be investigated by the model.

The current model assumes that pests, diseases, weeds and pollutants are under total control so that the crop does not suffer any impact. Phenology depends on temperature, the crop will reach full maturity and ready to be harvested at a certain temperature sum, but the harvest itself will usually take place somewhat later. Possible losses between these delays are not accounted for in the current model. At harvest either the whole of the crop or only crop compartments may be taken from the field. The parts of the crop that remain on the field after a harvest will be used as an input to the soil organic matter module.

Future development of the crop component will aim at decoupling the different model processes to meet the object oriented design paradigm and to allow the user to create, combine and assess different modelling approaches.

2.1.5.2 Current implementation

The model in the APES crop component is yet identical to the LINTUL model. LINTUL is written in FST, Fortran Simulation Language. The Fortran code containing the rate equations, which is generated by FST, is encapsulated in a Fortran dynamic link library (dll). The dll interfaces with a 'Fortran wrapper class'. The 'Fortran wrapper class' implements the modcom interfaces IOdeProvider and ISimObj. These interfaces ensure the proper handling of the rate equations in the modcom simulation environment. Moreover, the 'Fortran wrapper class' converts the semantic datatypes of the crop component into original LINTUL variable names. The semantic datatypes are given in the annex of this document.

The linkage to another component e.g. a soil water component that could model a part of the crop system implies that definition of the crop system changes and the crop model must be changed as such. Because the present crop model is encapsulated in a fortran dll, the fortran dll should be adapted and recompiled meaning that this particular adapted dll is only valid for the particular link that one wants to make. The creation of these high dependencies, or couplings, is neither desired in object oriented design nor is it a characteristic of a good software system. Future development of the crop component will aim to decouple the different model processes so it meets the object oriented design paradigm better.

2.1.6 The Grasses component

2.1.6.1 Models

The grassland model should simulate biomass accumulation for a wide range of grasses species and react dynamically to management practices, such as defoliation and fertilization. Thus, we chose for basis the biophysical sub-model of SEPATOU developed by Cros et al., simulating herbage growth under different management strategies.

This model was extended to a large range of grass species by including the concept of plant functional type, based on a typology developed within INRA, Toulouse. These plant functional types are defined according to grassland utilisation (grazing, cutting) and sward nutrient status (defining through fertilization and plant available nitrogen, given by the soil component). Therefore, such definition of criteria allows (1) predicting herbage accumulation rate under different management practices and (2) evaluating the impact of these practices on biomass production. Plant functional type permit to group species according to their common responses to the environment (response trait) and/or common effects on ecosystem processes (effect trait). Therefore, inclusion of this concept into the grasses model by defining specific parameters applicable to multi-species grassland made the model generic and therefore applicable to the European level.

The grasses model target of simulation is mainly permanent grasslands. It can be extended to temporary grasslands, considering them as PFT A or B, depending on their attributes, especially for phenology. However, it does not consider (1) extensive rangelands, (2) summer pasturing (in mountainous regions) and (3) fallows. Furthermore, the model was developed in the perspective to simulate grassland production from North to South of Europe with a good sensitivity to management practices and climatic differences within a specific zone.

To determine thermal time within the model and consequently phenological variables such as leaf life span, average daily temperature out of the range from 0 to 18°C were set to these limit values. As climatic conditions become more and more contrasting in Northern or Southern part of Europe, there may be a need of some recalibration of our model for Baltic or Mediterranean regions. So such threshold values may need to be revaluated for more extreme conditions, usually leading to the presence of other graminea or dicotyledons that the one considered within the typology from Cruz et al.

Finally, the primary goal of the implemented model within the grassland component was to establish impact of management on grassland production for specific regions. Therefore, upscaling of the model to the European level may lead to some discrepancies in taking into account weather variability (as mentioned previously) but should still be effective to consider impact of management practices.

2.1.6.2 Current implementation

For now, grassland model is implemented as a "one-model per class" (one strategy) and directly inherits methods from ModCom. The component implements the interface ISimObj and IOdeProvider. Further development will be needed to make it as an independent reusable, replaceable and extensible component. A wrapper class will be developed to interface the future component with ModCom.

INRA.Grassland project (a ModCom component) is available through the CVS APES-GUI and Modcom components. It's linked with input data from CLIMA.

2.1.7 The Vineyards / Orchards component

2.1.7.1 Models

The vineyard component is being developed on purpose to match the objectives of APES. It is based on general concepts commonly admitted for modelling potential crop production at field scale. Most of theses concepts have been retrieved from the literature and have been validated. However, even if tests are being performed on different parts of the model, the whole component has not been validated yet with field data.

For the time being, the model is parameterized for grapevine; this choice was driven by the data base at our disposal. Yet the adopted formalisms are generic for perennial crops (fruit trees and wood trees in agroforestry systems). In its present version (month 15), the model is able to simulate:

- yield, average sugar and water content of the product, and the time-course of biomass production in leaves, shoots and fruits;
- the harvest and winter cane pruning (only stand-alone version);
- the biomass of senesced leaves and pruned stems (outputs for soil components);
- potential transpiration, potential soil evaporation and root length distribution throughout the profile.

Only climate data are required to compute the potential production of the annual aboveground biomass. A computation of a water and nitrogen stress index is in progress to allow the linkage with the soil components.

To reach the first objective, namely to provide a prototype version of the fruit tree component running under the Modcom environment, several assumptions/simplifications were made:

- only mature trees (i.e. with a standard architecture) are simulated;
- the soil surface is considered as bare and only one species is growing on the plot;
- perennial woody crops such as grapevine and fruit tree can be simulated the same way;
- only the annual aboveground organ production is taken into account; that is to say leaves, shoots and fruits;
- the biomass is allocated to the different organs of the crop using look-up tables;
- the inter-annual impact of carbon storage is neglected;
- the product quality is described by the water content and the sugar content of fruits;
- root length growth is driven by soil temperature and is disconnected from the biomass production.

For the second prototype, once the software structure will be satisfactory enough, more efforts will be put in testing and improving the concepts to reach the objective of modelling the growth of two species (grapevine or fruit tree, and intercrop) concurrently on a single plot.

The model computes the annual growth of aboveground organs (fruits, leaves, stems) for grapevine; some quality variables such as fruit sugar content and fresh weight are also estimated. To allow the future linking with soil components, the root length growth and its distribution throughout the soil profile is also calculated as well as the potential transpiration and evaporation.

Even if for many points orchards and vineyards can be simulated the same way, discrepancies between them exist due to the specificity of orchard management or to physiological behaviors of fruit trees closer to forest trees. Modeling orchards may require some predictable adjustments. At present our ongoing activity consists mainly in assessing the relevance of adding such adjustments into the present version of the component. Apple tree has been chosen as the species simulated in the APES vineyard/orchard component.

At present, the model does not cope with an environment with a limiting supply of water and nutrients. Impacts of water and nitrogen shortage on growth will be integrated in the forthcoming version. Most recent developments dealt with the development of a wrapper and a domain classes to provide the component with a set of OOP features in phase with the modularity aspect of APES. Such modifications aimed at improving reuse, interchangeability and extensibility of the software unit.

2.1.7.2 Current implementation

The component has been developed in C# to facilitate its integration into the Modcom environment. In parallel a stand-alone model has been written in FST (Fortran Simulation Translator) to test different algorithms.

2.1.8 The Agroforestry component

2.1.8.1 Models

The agroforestry component should be able to predict both the productivity of agroforestry systems, and some of their environmental impacts. However, agroforestry systems are very diverse as they combine numerous tree species with most major crops of Europe. The simultaneous presence of trees and crops represents the major challenge in simulation agroforestry systems, given also the 1D simplification of other APES components.

Modelling agroforestry implies to model competition between trees (usually individual trees) and crop components (usually crop population of plants). Competition occurs for all the resources needed by plants : light, space, water, nitrogen, mineral nutrients. Availability of a below-ground water table plays a key role in such competition.

The tree growth module in APES will dynamically model the tree growth over decades. This module will be generic and could be used for any perennial crop with a canopy (vineyards, orchards, large trees). This dynamic tree component will interact with the crop component. The tree component will be described by an average tree (tree to tree variability will not be described by the model). The tree will have access to a surface that depends from the tree density in the stand. Modelling perennial plants implies to take into account carbohydrates and nitrogen reserve pools, which make the growth model more tricky. These pools are essential to model correctly the rapid leaf area setting at budburst, the fruit production, or the reaction of the plant after pruning.

The APES tree module will also include a fruit pool, but the prediction of the fruit yield is considered not attainable with the simple structure of APES. It is therefore suggested to introduce the number of fruits as a forcing variable in the APES tree module. The number of fruits that will be forced should also take into account any farmer action of fruit number reduction (mechanically or chemically). The tree module will then predict the fate of this pool of fruits, taking into consideration the competition between the various tree sinks for carbon.

C allocation will be governed by two types of rules

- Teleonomic (or goal driven) allocation rules based on allometric equations defining the relative sizes of aboveground sub-compartments and below ground sub-compartments.

- An optimal allocation assumption ('functional equilibrium') between above ground and below ground mediated through stress indices

Six structural tree parts are considered

- Stem
- Branches (distinction between stem and branches is necessary because of alteration of the branch / stem allometry following pruning)
- Foliage
- Coarse (structural) roots
- Fine roots (feeder roots)
- Fruits

Light interception by spaced trees (or rows of vineyards) is a matter of geometry. However, our will is to maintain a 1D model in APES. A possibility for modelling the light interception by the tree is to take into account the structure of the tree stand (spacing of the trees, shape of the canopies) and calculate the true amount of direct and diffuse radiation that reach the crop. This means that some aspects of 2D or 3D modelling are introduced in the model, but that these effects are incorporated in parameters of a 1D model. A geometric description of the tree canopies must be done via an appropriate algorithm. This could be the module of the Hi-sAFe model.

Conventional algorithms based on volumetric soil water content or water potential are not able to simulate correctly water competition between different species. This is another case in which the 1D simplification requires strong assumptions. An algorithm that meets the required criteria, and is based on the matrix flux potential can be used simplifying the algorithm in 1D. Water uptake by mono-specific stands at seasonal scale tends to be dominated by the net supply to the soil (rainfall minus soil evaporation) and evaporative demand (determined by the energy balance), rather than by details of root distribution. This is no longer true in mixed stands.

2.1.8.2 Current implementation

The implementation of the first prototypes of the agroforestry component is on going.

2.1.9 The Weather component

2.1.9.1 Models

Weather components implement several models, from peer reviewed sources, to estimate variables subdivided in five domains. Emphasis is placed in sharing and making available for operational use modelling knowledge produced by research. Weather components can be considered as a realization of a part of "Numerical recipes in agro-ecology", implemented using an updated technology. The reason for the subdivision in components is mostly placed in easier, specialized reuse and maintenance. The reference to the peer reviewed sources of the models is available in the documentation.

AirTemperature
The generation of daily maximum (Tmax, °C) and minimum (Tmin, °C) air temperatures is considered to be a continuous stochastic process with daily means and standard deviations, possibly conditioned by the precipitation status of the day (wet or dry). Three alternative methods are implemented for generating daily values of Tmax and Tmin, all based on the assumption that air temperature generation is a weakly stationary process. The multi-stage

generation system is conditioned on the precipitation status with two approaches. Residuals for Tmax and Tmin are computed first, than daily values are generated - independently (Richardson-type) or with dependence of Tmax on Tmin (Danuso-type). A third stage, that adds an annual trend calculated from the Fourier series, is included in Danuso-type generation. Another approach even accounts for air temperature-global solar radiation correlation. A third approach generates Tmax and Tmin independently in two stages (daily mean air temperature generation first, Tmax and Tmin next), making use of an auto-regressive process from mean air temperatures and solar radiation parameters. Daily values of Tmax and Tmin are used to generate hourly air temperature values, according to alternative methods. Sinusoidal functions are largely used to represent the daily pattern of air temperature. Six approaches, are used to generate hourly values from daily maximum and minimum temperatures. A further approach derives hourly air temperatures from the daily solar radiation profile. Mean daily values of dew point air temperature are estimated via empirical relationships with Tmax and Tmin and other variables. A diurnal pattern (hourly time step) of dew point air temperature is also modelled via two alternative methods.

Evapotranspiration
Evapotranspiration for a reference crop (ET0) is calculated from alternative sets of inputs and for different canopies, conditions and time steps, using one-dimensional equations based on aerodynamic theory and energy balance. A standardized form of the Penman-Monteith equation is used to estimate daily or hourly ET0 for two reference surfaces. According to FAO Irrigation and Drainage Paper n. 56, the reference surface is a 0.12-m height (short crop), cool-season extensive grass such as perennial fescue or ryegrass. A second reference surface, recommended by the American Society of Civil Engineers, is given by a crop with an approximate height of 0.50 m (tall crop), similar to alfalfa. The Priestley-Taylor equation is useful for the calculation of daily ET0 for conditions where weather inputs for the aerodynamic term (relative humidity, wind speed) are unavailable. The aerodynamic term of Penman-Monteith equation is replaced by a dimensionless empirical multiplier. As an alternative when solar radiation data are missing, daily ET0 can be estimated using the Hargreaves equation. An adjusted version of this equation, according to Allen et al. is given. Stanghellini revised the Penman-Monteith model to represent conditions in greenhouse, where air velocities are typically low (<1.0 m s-1). A multi-layer canopy is considered to estimate hourly ET0, using a well-developed tomato crop, grown in a single glass, Venlo-type greenhouse with hot-water pipe heating. The Stanghellini model includes calculations of the solar radiation heat flux derived from the empirical characteristics of short wave and long wave radiation absorption in a multi-layer canopy. A leaf area index is used to account for energy exchange from multiple layers of leaves on greenhouse plants. The constituent equations of the Stanghellini model are in accordance with the standards of the American Society of Agricultural Engineers.

Rain
The occurrence of wet or dry days is a stochastic process, represented by a first-order Markov chain as described by Nicks et al. The transition from one state (dry or wet) to the other (dry or wet) is governed by transition probabilities, as characterized monthly by analyzing historic long-term daily precipitation data for the site. According to multi-transition model from Srikanthan and Chiew, the daily precipitation amounts are divided into a number of up to seven states (dry or wet) from 1 (lowest level of rainfall) to 6 (highest level of rainfall). On days when precipitation is determined to occur, the precipitation amount is generated by sampling from alternative probability distribution functions. The most approaches are based on the two-state transition for dry/wet days. The Gamma distribution is even used to model precipitation amounts for the last state (highest level of rainfall) in the multi-state transition probability matrix by Srikanthan and Chiew, while linear distribution is applied for the other states. The pattern of Gamma plus linear distribution across various occurrence states exhibits a combined J shaped function.

Short-time rainfall data are generated by disaggregating daily rainfall into a number of events, then deriving the characteristics (amount, duration and starting time) for each event. Four approaches have been implemented to disaggregate daily amounts into 6-hour or shorter resolution (as small as 10 minutes) amounts. The method described by Arnold and Williams targets at 0.5-hour time resolution and assumes that daily rainfall falls in only one event. The peak location is generated first according to a broken linear distribution. The other 0.5-hourly amounts are generated from an exponential distribution and relocated on both sides of the peak. The other methods are more flexible and able to capture bursts of storm occurring discontinuously over the day. In the approach by Meteoset an autoregressive process and a Gaussian daily profile model are combined to simulate the possibility of precipitation at any hour. Two options are available to generate sub-daily precipitation events at varying time steps. The cascade-based disaggregation from Olsson (1998) brings about the break-up of a time interval into two equally sized sub-intervals. The total amount is redistributed into two quantities according to two multiplicative weights from a uniform distribution: 24-hour rain into two 12-hour amounts, 12-hour amount into 6-hour amounts, and so on until 1.5-hour resolution is achieved. The approach by Connolly et al. allows disaggregation of daily rainfall into multiple events on a day, and the simulation of time-varying intensity within each event: (1) distinct storms are assumed independent random variables from a Poisson distribution, (2) the storm origins arrive according to a beta distribution, (3) storms terminate after a time that is simulated by a simplified gamma distribution, (4) each storm intensity is a random value exponentially distributed, (5) time from the beginning of the event to peak intensity is given by an exponential function, (6) as well, peak storm intensity for each event is determined from an exponential function, (7) internal storm intensities are represented with a double exponential function.

SolarRadiation
Solar radiation outside earth’s atmosphere is calculated at any hour based on routines derived from the solar geometry. Daily values are an integration of hourly values from sunrise to sunset. The upper bound for the transmission of global radiation through the earth’s atmosphere (i.e., under conditions of cloudless sky), can be set to a site-specific constant or estimated daily by diverse methods.

Broadband global solar radiation (about 0.3-3.0 μm wave-band) striking daily horizontal earth’s surfaces is estimated from alternative sets of weather inputs according to strategies based on either physical relationships or stochastic procedures. A sine-curve assumption is used to prescribe the hourly distribution of solar radiation from its daily value, assuming changes with solar elevation angle. The most simplified models relate diurnal temperature range to solar energy transmission through the earth’s atmosphere. As one of the most important phenomena limiting solar radiation at the earth’s surface are clouds, a cloud cover measure is incorporated in the model from Supit and van Kappel to estimate transmissivity. The radiation model from Winslow et al. uses saturation vapour pressures at minimum and maximum air temperature as a measure of the atmospheric transmission of incident solar radiation. The model from Ångström and Prescott is the most common choice to estimate global solar radiation when sunshine measurements are available. As an alternative, an implementation of the model Johnson et al. and Woodward et al. is given. Stochastic generation is based on the dependence structure of daily maximum and minimum temperature, and solar radiation. Such variables are reduced to time series of normally distributed residual elements with mean zero and variance of one. An autoregressive, weakly stationary multivariate process is used to generate the residuals series. Daily values of global solar radiation are generated for dry and wet days as daily deviations above and below the monthly average value. An implementation by Garcia and Garcia & Hoogenboom is given as well. The flux density at the earth's surface, on a horizontal plane, is comprised of a fraction of direct beam, coming directly from the direction of the sun, and a diffuse radiation coming from many directions simultaneously. The irradiance on a tilted surface even includes ground reflected fraction. The current implementation derives from the general approach from Liu and Jordan. The estimation of diffuse radiation on a horizontal surface depends on the extra-terrestrial irradiance and a transmission function. Hourly transmission relies on the anisotropic assumption for estimation on inclined surfaces and is further divided into the isotropic, circumsolar and horizontal ribbon sub-fractions. These sub-fractions are calculated separately and then summed to provide the diffuse irradiance. Ground reflected irradiance is estimated from a slope-dependent factor. Direct fraction of solar radiation is the complement to global solar radiation. The visible band (0.38-0.71 μm wavelength) is estimated daily by the diffuse/direct radiation ratio, and hourly by the solar elevation course. PAR amount can be also disaggregated into direct and diffuse fractions. Slope is the angle the surface makes with the horizontal plane, and aspect is the clockwise orientation to south. One or both are required to compute geometric factors that convert radiation estimates from horizontal to non-horizontal surfaces. An ESRI-based approach is implemented to derive slope and aspect from digital elevation data grids.

Wind
Daily mean values of wind speed, are generated by sampling from alternative probability distribution functions. Following generation of daily mean wind speed, alternative approaches are available to estimate the maximum and minimum wind speeds for the day. Like most climatic variables, wind speed tends to be both random and cyclic as time varies. Probability distribution functions are used to randomly distribute daily mean wind speed within day. Alternatively, wave functions are used to describe average diurnal wind speed variations using reference values of both maximum and minimum wind speeds for the day as inputs.

ClimReader
Crop, cropping systems, hydrological models at field level, often require meteorological data at daily or hourly time resolution. Such data may include different set of variables (e.g. maximum and minimum daily air temperature, daily rainfall, daily evapotranspiration). Also, such data may include missing data. Meteorological data also require site (location) data (e.g. latitude, clear sky transmissivity). Even if not related to meteorological data, other site data of interest are soil data which need to be loaded at start run. CRA.ClimReader.dll is a component which allows loading location and soil data, and provide met data at run-time. The component allows loading data in different formats (txt, XML, and from MS Access), and different sets of data, allowing for flexibility of data sources. The component allows estimating some meteorological variables if missing in the input file: reference evapotranspiration, vapour pressure deficit, day length, global solar radiation. Reference evapotranspiration can be estimated using the Hargreaves, Priestley-Taylor, and Penman-Monteith method according to data availability. The component uses ET, AirT, Wind, Rain, and GSRad components.

2.1.9.2 Current implementation

CLIMA components are implemented as fully independent components and are available with documentation and sample applications.

A ClimReader component is currently made available as Modcom application on the APES CVS server.

2.1.10 The Management component

2.1.10.1 Models

The AgroManagement component is designed to implement production management actions within the system. A production activity is defined, in this context, as a production enterprise (e.g. a crop rotation, an orchard) associated to a production technique (e.g. irrigated, high nitrogen fertilization, minimum tillage). Such an integrated system must be implemented in a way to allow imitating as closely as possible farmers' behaviour. Limiting the drivers of the decision making process to the biophysical system implies that each action must be triggered at run time via a set of rules, which can be based on the state of the system, on constraints of resources availability, and on the physical characteristics of the system. Besides, implementing the simulation of management in a component based system poses challenges in defining a framework which must be reusable and able to account for a variety of agricultural management technologies applied to different enterprises. Finally, the implementation of management must allow using different approaches to model its impact on different model components.

The AgroManagement component formalizes the decision making process via models called rules, and it formalizes the drivers of the implementation of the impact on the biophysical system via set of parameters encapsulated in data-types called impacts (figure 2). Each operation must have a rule to be applied at run time; when the rule is satisfied, a set of parameters are made available to model components for the implementation of the impact. The component is easily extendable for both rules (which have the structure of strategies) and impacts, so that the implementation of management allows using different modelling approaches. Furthermore, the information on the biophysical system is passed via a data-type called states, which can also be extended. The output drivers (management actions published to model components) can be fully customized by the user as well.

The rule-based model is characterized by 3 main sections:

- Inputs: states and time
- Parameters (values are compared to rules via the rule model)
- A model which returns a true/false output

Rules can be based on relative date or based on a set of state variables and is implemented as a class encapsulating its parameters declaration and test of pre-conditions (this also allows to validate management configuration files via pre-conditions tests).

Parameters are needed by model components to implement the impact of management. There are few parameters which are common to a generic management event and to a specific management event (e.g. irrigation and tillage). Other parameters are needed by specific management approaches and generally at least partially differ even within specific management event types. Using a set of rules and management impacts parameters (which can be extended easily to account for new rules and new approaches to model management impact).

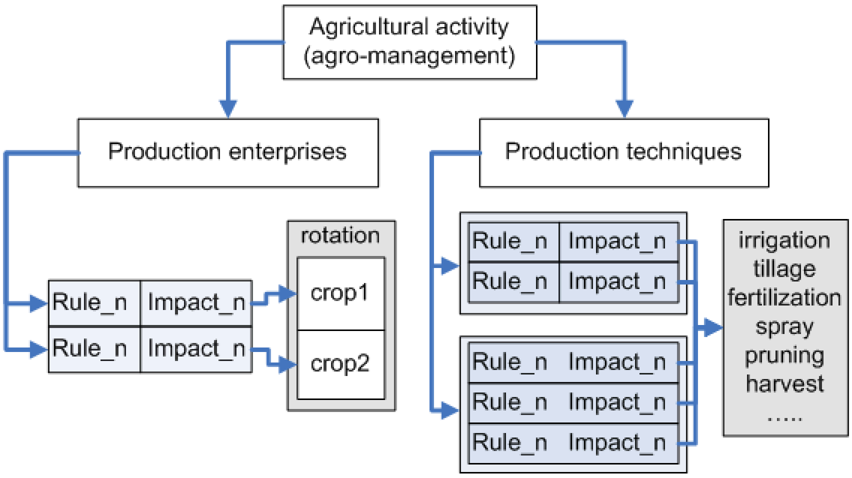


Figure 2 Example of management configuration for a two years rotation. Rules are used to trigger agro-management action which may regard production enterprises (e.g. plant or not a given crop) or production techniques (e.g. apply tillage, spray, harvest a crop)

2.1.10.2 Current implementation

The management component is implemented in two ways: as a stand alone C# component, and as a Modcom component which uses the first one (the Modcom is an application of the generic management component). An application to build agro-management files is made available.

2.1.11 Modcom

Agro-ecological simulations typically present numerous requirements which are common across systems, including (1) standardized public interfaces defining object access and action initiation, (2) high-level communications capabilities for components of the system to communicate with other components in a non-specific manner, (3) standardized methodologies for collecting and transferring information between components of the system, (4) standardized methods for data import, representation, analysis, visualization and export, and (5) mechanisms for synchronizing the sequencing of flow execution among system components. The framework paradigm provides potentially useful capabilities in all of these areas. Modcom is a realization of a framework to meet the needs above.

The objective of the Modcom project is to develop a framework that facilitates the assembly of simulation models from previously and independently developed and tested component models. ModCom provides a set of interface specifications that describe components in a simulation and it provides implementations of the core simulation services. Main core services are integration capabilities using different methods, events handling, and data communication. Modcom allows running simulation setting initial and final time, time step, type of integrator. Once components are registered with Modcom, simulation configurations (components and link among components, parameters values, type of outputs) can be controlled via an XML file. Modcom is a software library implemented for the .NET framework and written in C#

2.1.12 Linking components

The modular approach via components to build simulation systems has advantages but it also demands for specific solutions due to the flexibility of the system. At each time step, model components are called receiving the information made available by other components to the simulation, and sending their own output.

Matching of input-output needs, during whole model design, is built on defining an info-flow matrix, in which each component states needs from other components. Such a matrix is currently at version 7 and constantly refined.

A more difficult problem to handle is given by mass balance and resources arbitration. In a "monolithic" model, all quantities are know and available within each time step, making easy the computation of intermediate variables. In a component based model, variables must be communicated even within time-step across components to estimate intermediate variables. Design features as computation ordering and intermediate variables handling becomes an issue to deal with to ensure resources arbitration and correct mass balance. As an example, a crop component may estimate at each time step the rate plantWaterUptakeAttainable (demand), but that demand must be matched with soil water availability, and it may compete with soil evaporation or even weeds water demand. This implies that the component managing the resource water (SoilWater in this example) makes available a variable which sets the plantWaterUptakeActual, to be used by the crop component to continue in the computations. Also, other components, still in the same example, as the SoilCarbonNitrogen component, need that value of plantWaterUptakeActual to estimate the plantNitrogenUptakeActual. Modcom was modified to account for the need of sharing intermediate variables within the time step. All components implement (or will implement) mass balance to be made available as output.

Problem related to the modular structure of the simulation system are being analyzed to provide a flexible and yet easily usable platform. A first solution and the relevant implementation has been made available with the current release of APES, both in SeamFrame and as APES stand alone application.

2.2 FSSIM[3]

2.2.1 Introduction

The Farm System Simulator (FSSIM) is an integrated modelling system developed to assess the economic and ecological impacts of agricultural and environmental policies and technological innovations. Based on the link of biophysical and micro-economic models, FSSIM seeks to describe the technical aspects at the farm level given specific biophysical conditions, using different sets of constraints to derive a set of feasible technological alternatives for each farm type. The principal characteristic of this type of models is the application of engineering production and environmental functions derived from biophysical models (APES) and other sources (experiments, expert knowledge, surveys, etc.). Built in the philosophy of a *generic* model, FSSIM has to be valid for every type of arable farming system and every geographic location in the EU25 but also for multiple research issues applicable to the micro level. FSSIM consists of an agricultural management module (FSSIM-AM) and a mathematical programming model (FSSIM-MP). FSSIM-AM aims to identify current and alternative activities and to quantify their input output coefficients (both yields and environmental effects). FSSIM-AM includes: 1. a procedure to derive current activities using a survey and a complementary procedure using FADN data, 2. Production Enterprise Generator, 3. Production Technique Generator and 4. Technical Coefficient Generator. Both current and alternative activities can be simulated through the use of the biophysical modelling systems APES. FSSIM-MP seeks to represent the actual farmers' behaviour using the knowledge of technical and socio-economic constraints, the relation between production factors, the amount of output obtained and the costs of each production activity and future market prices. FSSIM-MP includes: 1. the objective function describing the farmers' behaviour and goal, based on the maximization of expected income minus some measure of its variability, according to different states of nature and market; 2. the set of explicit constraints related to technical (land, water, equipment) and socio-economic resources (labour, finances, cash flow) as well as those related to policy and environmental measures (price and market support, quota and set-aside obligations, cross-compliance policies, agri-environmental measures, etc).

FSSIM is calibrated using the risk approach complemented by an extension of the Positive Mathematical Programming (PMP) technique, inspired from Röhm and Dabbert (2003). The base year information for which the model was calibrated stems from a three-year average around 2001. In term of policy representation, FSSIM includes the major policy instruments related to arable crops such as price and market support and set-aside schema as well as certain cross-compliance and specific agro-environmental measures. The implementation of these instruments depends on the analysed policy in different scenario assumptions which are the Agenda 2000 for the base year and the recent CAP reform of June 2003 in Luxembourg, as it would be implemented in 2012, for the baseline scenario. For the Prototype 1, only the policy scenarios defined under Test Case 1 are implemented. Most of these policies are handled in the market models, so no "shocks" was entered at farm level. However, impact assessments of these policy scenarios are performed at different levels (market, region and farm levels).

FSSIM is aimed to be applied to a set of farms representing the farm types across the EU-25. This issue has required the development of a farm typology taking into account the

[3] For more detailed information concerning FSSIM, see D3.3.6

heterogeneity in farming and biophysical endowment. For the Prototype 1, only a set of farm types representing the arable farming system in Flevoland (Netherlands) and Midi-Pyrénées (France) was taken into account.

The first version of FSSIM will concern only arable farming sectors and will have a limited number of variants. Starting from conventional Farm Accountancy Data Network (FADN) and working with exogenous prices, FSSIM seeks to represent major arable activities captured by the Economic Accounts for Agriculture (EAA). To reach these objects, the following specifications are retained:

FSSIM-AM

- FSSIM-AM uses a conceptual approach for the specification of agricultural activities in terms of rotations, crop management, yields, costs and labour requirements, differentiating between currently observed activities and potentially promising activities for the future.
- FSSIM-AM is implemented as a flexible, extensible set of components, e.g. Survey on Current Activities, Production Enterprise Generator, Production Technique Generator, and Technical Coefficient Generator that can be adapted for livestock activities and perennial activities.
- FSSIM-AM is designed with the aim to be generic in its structure and can be applied to any region and conditions.
- The knowledge base has been used to formalize the data-types of FSSIM-AM in a structured, ordered and logical way. This helped to achieve a clear description and definition of many of the concepts used by FSSIM-AM by making the properties of each concept explicit and relating concepts to each other.
- The components of FSSIM-AM are linked to APES and FSSIM-MP, and can be linked to other components through OpenMI.

FSSIM-MP

- FSSIM-MP is **a static model** i.e. one period decision, with the possibility of subdividing the period in sub-periods. This means that the model optimizes an objective function for one period over which decisions are taken considering the specificity of each individual time-period and the trade-offs between the time sub-periods. However, for incorporating some temporal effects, agricultural activities will be defined as “crop rotations” and “zoo-technical units” instead of individual crops and animals and the constraints will be included in a specific structure as developed by Pacini (2003). The development of a FSSIM dynamic version is foreseen for the Prototype 2; this is essential to model perennial activities (orchards, vineyards, olive production, etc) or simulating investments to assess non-equilibrium conditions.
- FSSIM-MP is a **risk programming model**; many alternative formulations/approaches could be used to modelling risk from relatively complex specifications using expected utility theory to a simplified one, just taking into account expected profit and standard deviation of it, according to states of nature and market. The most commonly used methods for including risk in mathematical programming farm models are Mean-Variance, Mean-Standard deviation, MOTAD (Hazell et al, 1986, Mc Carl, 1989), Target MOTAD (Tauer, 1983) and Focus Loss (Boussard, 1971). For this version we start with a basic specification relating to the

Mean-Standard deviation method in which expected utility is defined under two arguments: expected income and risk.

- FSSIM-MP is **a positive model** in the sense that its empirical applications exploit the observed behaviour of economic agents. Traditionally, optimization models (OM) such as linear programming are often based on *normative* assumptions aimed at identifying the "best" production combination under the hypothesis that the initial situation is not binding in terms of production choices. This assumption that induces a wide divergence between base period model outcomes and observed production patterns is unacceptable. In this way that *normative* OM has been left behind for the *positive* type model, where the main objective is not to reach the first-order optimality conditions, which may be infeasible, but to, as precisely as possible, reproduce the observed production situation such that the likely behaviour of the farmers is simulated when varying parameters determined by agricultural policy interventions or technological innovations.
- Given the positive character of the model, and the necessity for reproducing a historical behaviour, a calibration procedure was developed. This procedure is based on the risk approach complemented, by an extension of the Positive Mathematical Programming (PMP) technique, inspired from Röhm and Dabbert (2003).
- FSSIM-MP is a **non-linear programming model** (non-linear objective function), calling sometimes for binary variables[4] that can only take values of 0 or 1 in order to implement cross-compliance policies.

According to all these specifications, the general structure of the first version of FSSIM-MP can be summarized as:

Max: $\mathbf{Utility = C'X - \phi\sigma}$

Subject to: $\mathbf{AX \leq B;}$

$\mathbf{X \geq 0}$

Where **Utility** is the objective function to maximise, **C** is the (n x 1) vector of expected income from agricultural activities, **X** is the (n x 1) vector of agricultural activities' level (in a simulated farm plan), ϕ is the risk aversion coefficient according to the Mean-Standard deviation method, σ is the standard deviation of income according to states of nature and market defined under two different sources of instability: yield (due to climatic conditions) and price due to market conditions, **A** is the (m x n) matrix of technical coefficients, and **B** is the (m x 1) vector of available resource levels. Utility, X and σ are endogenous variables**.**

2.2.2 FSSIM-AM: Agricultural Management

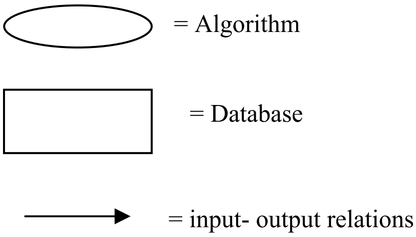
2.2.2.1 Purpose of the FSSIM-AM: current and alternative activities

The purpose of the Agricultural Management Module is to describe, generate and quantify alternative and current activities that can be evaluated by APES (in terms of yields and environmental effects or other quantitative methods). The fully quantified agricultural activities i.e. the complete sets of inputs and outputs are inputs for FSSIM to assess their contribution to the goals considered. Alternative activities are activities that are not currently

[4] A mixed-integer nonlinear programming (MINLP) solver will be used in order to allow an elegant solution for this type of problem.

used, but are technically feasible alternatives for the future, often technological innovations or newly developed cropping or husbandry practices (PD3.3.1), while current activities are activities that are currently practiced and can be derived from observed data.

The procedures for constructing production enterprises and production techniques are quite different for current and alternative activities, while the addition of costs and labour requirements and the processing of APES outputs are largely the same for current and alternative activities. Current activities e.g. current production enterprises and current production techniques will be identified on the basis of expert knowledge. Variable and fixed costs and labour requirements will be derived from statistical sources, which can also be used for alternative activities.



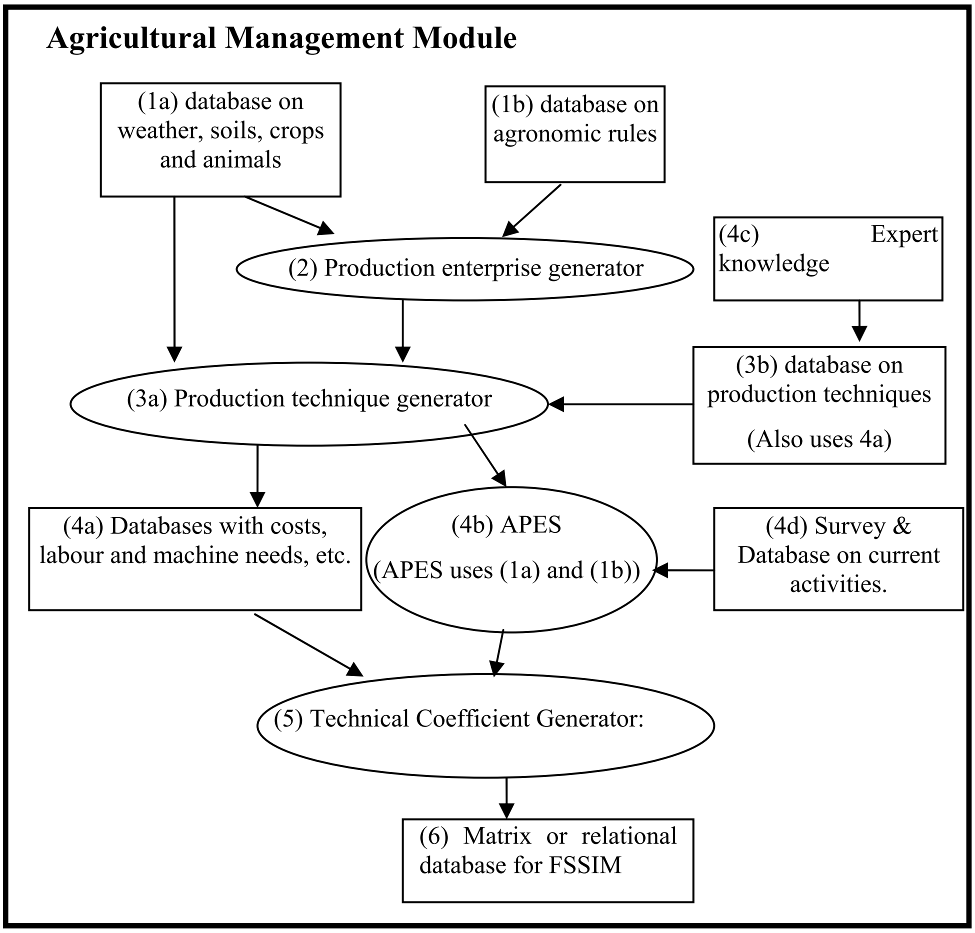


Figure 3 Agricultural Management Module and its components: algorithms, databases and connections

The main calculation components or modules of the Agricultural Management Module in FSSIM are the Production Enterprise Generator (PEG), Production Technique Generator (PTG), Survey on Current Activities, APES and the Technical Coefficient Generator (TCG) (figure 3). Database structures are used to collect and (temporary) store input and output information for the different components. The different components result together in the quantification of agricultural activities in terms of technical coefficients (inputs and outputs) that are offered to the FSSIM model to assess their contribution to goal achievement. The different components are linked through a framework that is part of SEAMLESS-IF. This framework is developed according to the guidelines provided by WP5. Figures 4 to 8 show in more detail how the PEG and the PTG result in the construction of agricultural activities.

Starting point for the Agricultural Management Module is the farm typology developed by WP4. This farm typology provides for each region a set of typical well defined farms in terms of size and specialisation and which are spatially allocated with certain soil and climate characteristics. The soil and climate data will be used both in the PEG and in the PTG, while specialisation and size of the farm determines the availability of equipment.

2.2.2.2 Deriving and assessing current activities

A current activity (CA) is an agricultural activity currently in use on farms. For crops, these activities are represented by data that describe a rotation, measured per hectare.

Current activities may provide some extra data to APES and FSSIM. We may require data about the environmental impact of a particular way of growing a rotation. Such information is not necessarily easily available, but can be assessed by the use of models (APES). To run such models, we need more detailed information about the timing of soil treatment and water availability/use. In general, this requires more detailed data than those used in FSSIM, but the information is important in order to calculate the environmental impacts of changes in prices and policy.

2.2.2.3 Data availability

The FADN data set is an EU-wide data source of annual accounting data from farms. Use of area, number of animals and products produced are also reported. The FADN data set applies the same standards/data groups as far as possible in the various regions. The data represents actual average data regarding yield and costs for farms, as opposed to best practice data.

FADN data offer a lot of information about use of area, crops choice and yields, etc. Crop rotations are not directly observed from these data, neither if we had access to farm level data.

With an expert survey, we may identify rotations, and we may make sure that data are collected in a uniform standard across regions. The detailed management information that we need to classify the observed CA's cannot be extracted from aggregated databases such as FADN. However, with a survey it may be possible to identify the rotations, not only normal production methods of particular crops. Thus, the survey will be designed to provide data for the particular "FADN-farms" that are identified for the region.

2.2.2.4 Generating alternative activities

Alternative activities are activities that are not currently used, but are technically feasible alternatives for the future, often technological innovations or newly developed cropping or husbandry practices (PD3.3.1). To generate alternative activities a theoretical approach is used to link production enterprises (rotations) to production techniques (crop management). The Agricultural Management Module consists of three components: (i) Production Enterprise Generator generating production enterprises, (ii) Production Technique Generator generating and quantifying the management practices of the production enterprises, and (iii) Technical Coefficient Generator quantifying, collecting and formatting the technical coefficients.

2.2.2.5 Production orientation

A Production orientation is a set of value driven aims and restrictions of the agricultural activity that direct the input and output levels (Van Ittersum and Rabbinge, 1997). Production orientations could, for instance, be labelled as 'integrated', 'organic', 'conventional' or 'low labour input.' In our approach, the production orientation defines the maximum or minimum value of certain variables like minimum and maximum rotation length, the set of possible crop protection methods, the set of possible nutrient sources (organic and inorganic) and the set of possible tillage methods.

In the framework of the Agricultural Management Module, there is the possibility and flexibility to define many different production orientations, allowing to address the wide diversity of possible farming systems and policy questions within SEAMLESS. Production orientations are thus user- and context-specific, e.g. dependent on the (policy) question the user wants to address.

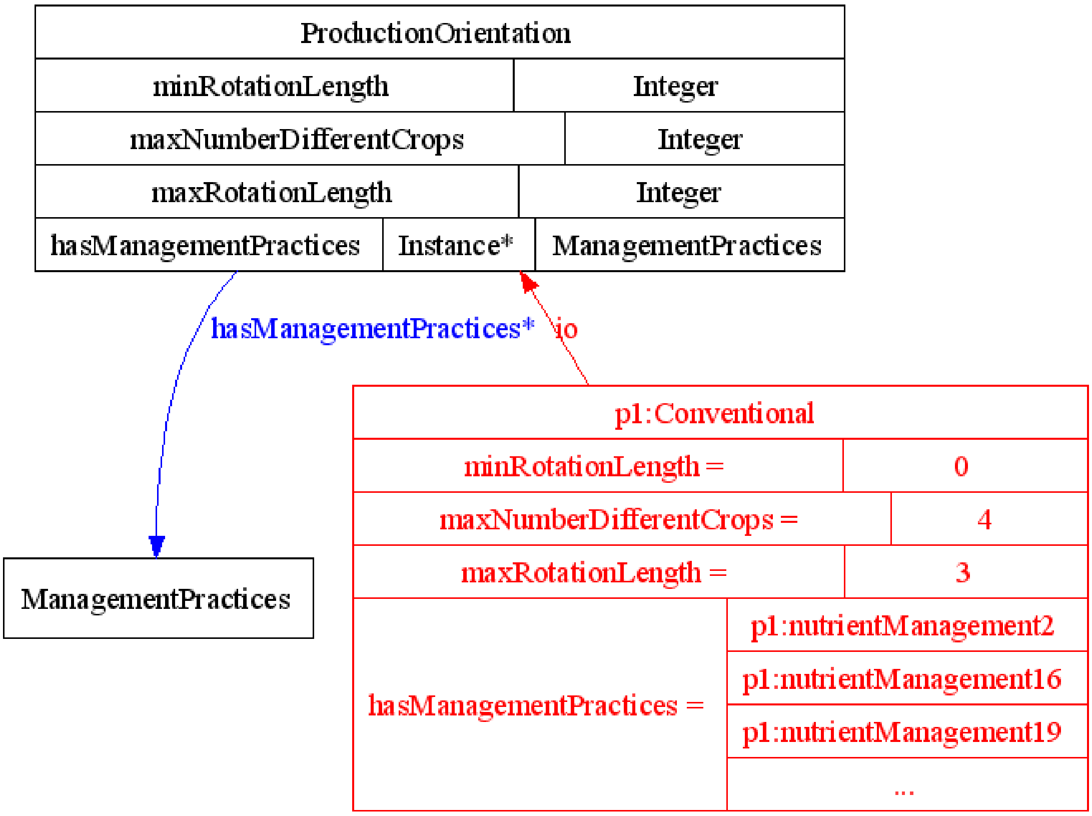


Figure 4 Production orientation class with its properties and an example of an instance 'conventional' of a production orientation

2.2.2.6 Production Enterprise Generator

Production Enterprise Generator (PEG) is a tool to generate a feasible set of production enterprises (crop rotations) of the farm based on crop suitability filters, such as soil and climate characteristics and for annual arable crops rotation suitability filters (or for animal husbandry systems herd composition constraints). The PEG aims to design production enterprises in a coherent transparent and reproducible way. In principle, all crops that may be grown in a given environment, can be combined into different cropping sequences. However, not all of these combinations are agronomically feasible or desirable. The PEG contains a number of 'crop and rotation filters' that limit in an early stage the number of crop rotations for which production techniques need to be defined and that limit the number of simulations to be carried out by APES.

The PEG creates rotations by a three step procedure:

1. Creation of a set of suitable crops from the list of possible crops using suitability filters. The list of possible crops is an input to the PEG and as such user-dependent.
2. Generate one possible combination of suitable crops in a rotation and checked by a set of rotation suitability filters. If a rotation is suitable, it is added to a set of suitable rotations, otherwise it is discarded. One by one all possible rotations are generated and checked, till finally all rotations have been checked. The rotation suitability filters use properties of crops and of crop groups that are related to the requirements of rotations (figure 4).
3. Attaching information on farm type and production orientation to the feasible set of rotations. Each set of rotations is thus linked to a farm type and production orientation. One rotation might be linked to more than one farm type and production orientation.

Based on crop requirements as mentioned by Reinds et al. (1992), 7 crop suitability filters with respect to soil were developed. With respect to climate 3 crop suitability filters were developed as based on Alterra & INRA (2005), Russel (1990), and Wolf et al. (2004), which account for low temperatures over a prolonged period, risks of climatic hazards (e.g. excess of water) and lack of suitable land in such areas. The crop suitability filters are:

- Rainfall Surplus during Harvest;
- Temperature Sum during growing season;
- Slope;
- Altitude;
- Clay Content;
- Rooting Depth;
- Roughness;
- Salinity;
- Alkalinity;
- Drainage.

The crop suitability filters will be further evaluated on the quality of their results after applying the filters to 2-4 sample regions. Also, it could be considered to link crop suitability filters to regions or crops, so that not in all regions or for all crops all crop suitability filters are applied.

The rotation generation tool of the PEG is based upon ROTAT (Dogliotti et al., 2003), which combines crops (maximum 30) from a predefined list to generate all possible rotations (maximum 250.000). An adapted and extended version of ROTAT has been made, which uses part of the 'rotation filters' of the original ROTAT.

2.2.2.7 Production Technique Generator

The Production Technique Generator is a tool to describe production techniques of agricultural activities for the feasible set of production enterprises. A production technique is a complete set of agronomic inputs characterized by type, level, timing and application technique (Van Ittersum and Rabbinge, 1997). The PTG combines different management practices to production techniques and takes out infeasible production techniques according to production orientations. Each management practice is characterized by a set of management aspects. There is a danger of too much detail and over-specification. Thus, it is important to determine the major inputs affecting outputs. The complete set of inputs consists of the following management practices (figure 5):

- General Management: General management refers to operations that always have to take place in order for a successful harvest: sowing, harvesting, clipping, pruning and field inspection. Sowing is defined by a timing and the amount of seed needed, while harvesting is defined by a reduction in biomass or Leaf Area Index of the crop, just as clipping and pruning with the difference that harvesting terminates the crop.
- Water management: Offered to APES are whether or not irrigation is used, the method of application (sprinkler, furrow, etc.) and the timing rule used to initiate an irrigation event combined with a plant available water threshold for irrigation. APES can then calculate the achieved yield level and the amount of water needed.
- Nutrient management: To APES the level of application, type of nutrient, method of application and dose/timing of application can be offered. APES can then calculate the achieved yield levels. The input parameters for the nutrient application events in APES are calculated on the basis of expected target yield, expected recovery, surplus of nutrients supplied, distribution between first and second nutrient application and distribution between organic and inorganic fertilizers.

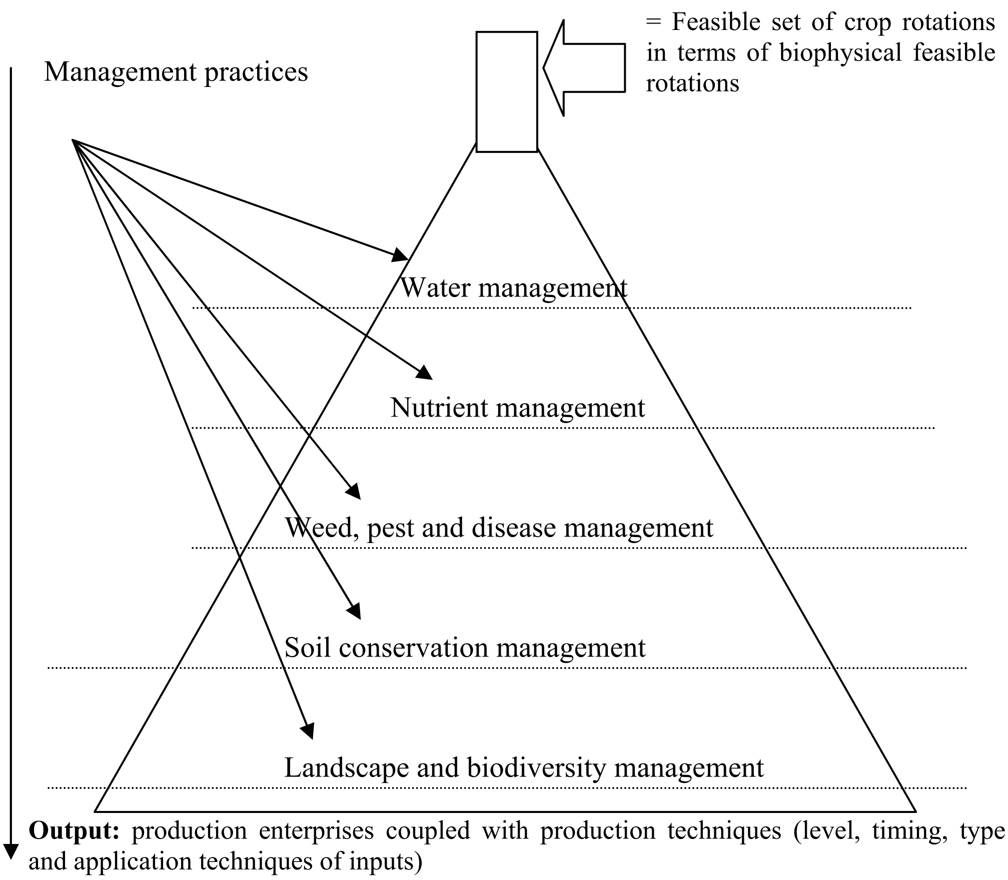


Figure 5 Procedure within Production Technique Generator

- Weed pest and disease management: For weed, pest and disease management control packages will be defined that describe all measures required to achieve a certain control level of weed, pests and diseases for a certain crop. These packages are based on expert knowledge. APES cannot quantify the effect on yield of a certain level of weed, pest and disease pressure. Therefore, weed management control packages are based on an average 95% control of weeds. This control level does not affect the simulated yield level. For pests and diseases such a control level is probably impossible for all control packages (e.g. organic) and yield reductions are inevitable.

- Conservation Management (includes Soil conservation management and Landscape and biodiversity management): conservation management is not primarily aimed at improving environmental conditions to facilitate the production of food or fibre products, but serves other objectives (which might have an effect on production of food or fibre products), like soil conservation and landscape and biodiversity management.

Each of these management practices consists of several aspects, for example, the method of application and the level of application (full replacement of water, e.g. 100% or allowing some water shortage, i.e. 50% of unlimited water supply), the timing of the application (both time window and timing rule used, for example irrigate every 14 days starting from the 1st of July) and input type (for example, manure or fertilizers). Each production technique encompasses for each crop a set of different management practices, each of which can be characterized by such management aspects.

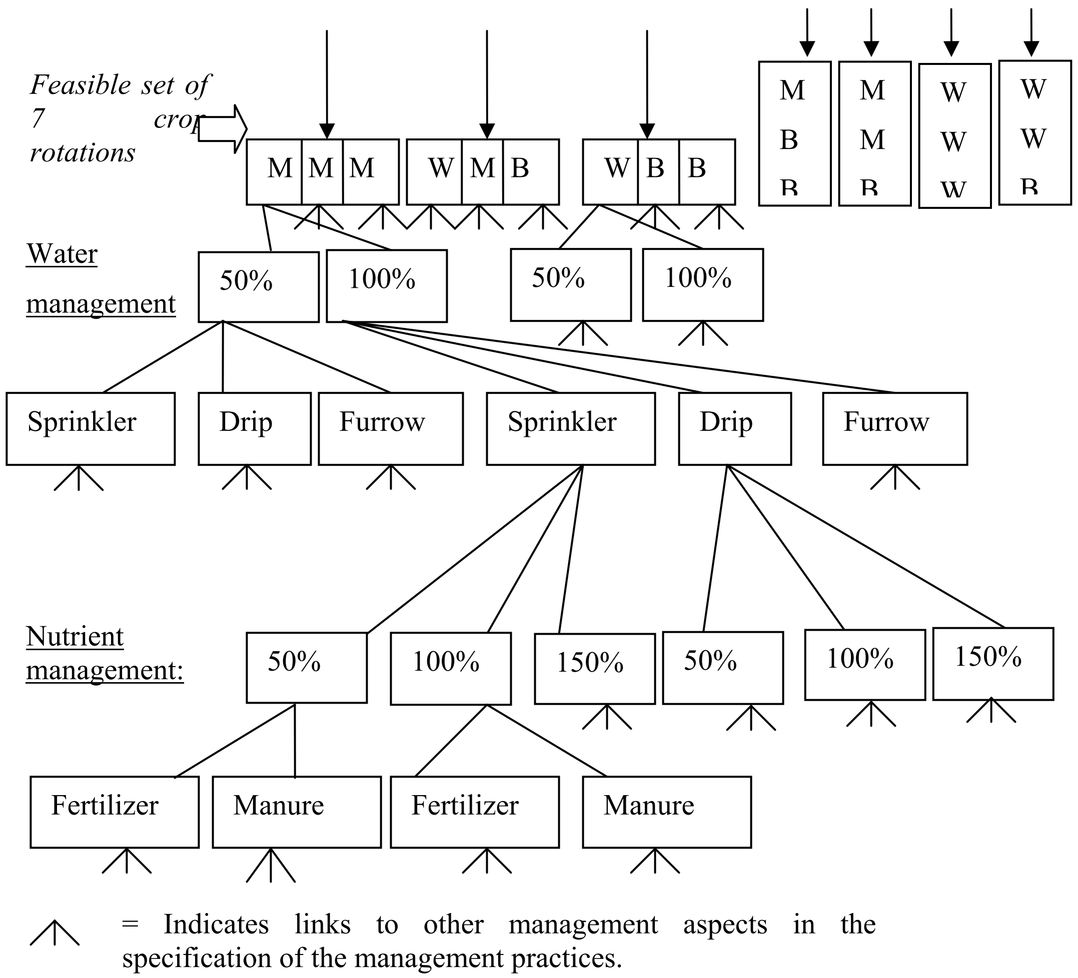


Figure 6 Simplified description of two management practices (water and nutrients) as part of the production technique of an activity

Figure 6 shows how on the basis of 7 rotations combinations of two management practices lead to production techniques. The number of production techniques increases rapidly with increasing management practices. In Fig. 6 the total number of production techniques for each crop with two management practices is 36 (=2 water input levels × 3 water application techniques × 3 nutrient input levels × 2 nutrient input sources). Still three other type of management practices (weed, pest and disease management, conservation management and general management) need to be added, indicating that the final number of production techniques per crop will be far larger than 36 as in this example. The number of production techniques increases rapidly by identifying different management aspects, as each alternative leads to a doubling of the production techniques and a tripling or quadrupling of the agricultural activities (dependent on the rotation length). This is called a combinatorial explosion. It is important to be explicit about the relevancy of management practices to be incorporated to avoid a combinatorial explosion, which leads to extremely large numbers of agricultural activities, increased computing time and difficulties in processing and analysing the results.

To quantify inputs and outputs of a certain production technique, the following procedure will be used, which is an output-oriented approach. In an output-oriented approach the production target (output) is set dependent on the most limiting growth factor and on the objectives of the agricultural activity and then the most efficient set of inputs to realize this target is defined (Van Ittersum & Rabbinge, 1997; Hengsdijk & van Ittersum, 2002). At the start of the growing season, the management will depend on a target yield level, for example 8 t/ha for wheat or 80% of the potential yield level. Consequently, the amount of nutrients that need to be applied is calculated. On the basis of these inputs APES is run for the rest of the growing season, with information on threshold values for water management (like soil water content never drops below 75% of maximum soil water content), threshold values for pest and disease management (e.g. based on temperature sum and humidity) and a management scenario leading to 95% control of all other weeds, pests and diseases resulting in no yield reduction due to weeds, pests and diseases. APES calculates the achieved actual yield level and the amount of water needed. In summary, the proposed procedure is: target yield =>inputs for nutrient, weed, pest and disease and conservation management=> simulation with APES => simulated yield and input levels for water management and some pests and diseases.

2.2.2.8 Assessing agricultural activities through APES

The Agricultural Production and Externalities Simulator (APES) comprises a dynamic crop growth simulation model. While FSSIM as a farm level model operates on a yearly or growing season basis, APES simulates the crop growth and soil processes on a daily basis. APES can evaluate the detailed crop management specifications supplemented by the Agricultural Management Module of FSSIM on yields and environmental effects. This implies that the Agricultural Management Module should provide a detailed description of all crop management events taking place during the growing season with their timings, and on the basis of these specifications APES is able to calculate the effects in terms of yields and environmental effects. Here a distinction needs to be made between current and alternative activities, as current activities contain a yield value for different crop products, which can be used by APES to compare its own yield estimation with the observed yield. Alternative activities hold no information on the outputs as it is frequently unknown what the yield is for alternative production techniques.

An event is one operation that takes place during the growing season of the crop, for example sowing, fertilization, irrigation, harvesting, field inspection, etc. The five management generators define the events based on the implements and inputs associated with the event, the specification of the management practices as part of the production orientation and the rotations that are an output of the PEG. The output of the PTG are agricultural activities specified in terms of biophysical conditions and management events including associated inputs (figure 7), which are used to parameterize the dynamic crop growth simulation model APES. This model simulates yields and environmental effects for each crop with associated management in a rotation.

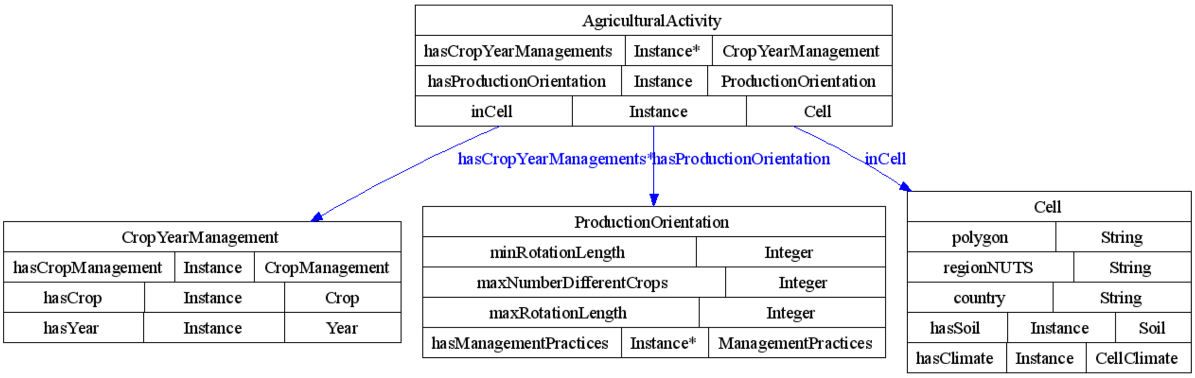


Figure 7 Agricultural activity with its relationships with other classes

The link between APES and the PTG (figure 8) is thus an agricultural activity, which refers to a rotation, several crops in the rotation and a crop management for each of the crops in the rotation. The crop management of a crop in rotation exists out of several events, for example two nutrient events, one irrigation event, a sowing event and a harvesting event. Each of these events has certain properties that are required by APES, for example, a mean tillage depth, the irrigation implement used for irrigation, or the amount of organic N. The properties of the different events are given in figure 9.

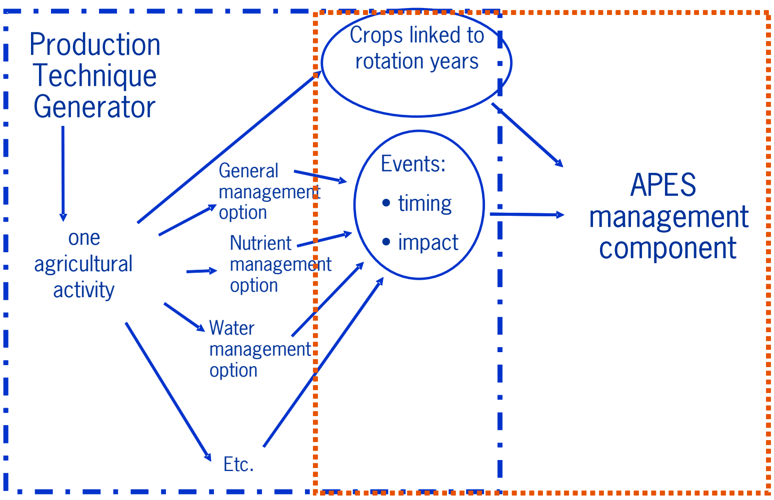


Figure 8 Link through crops in rotation with associated events between the PTG and APES

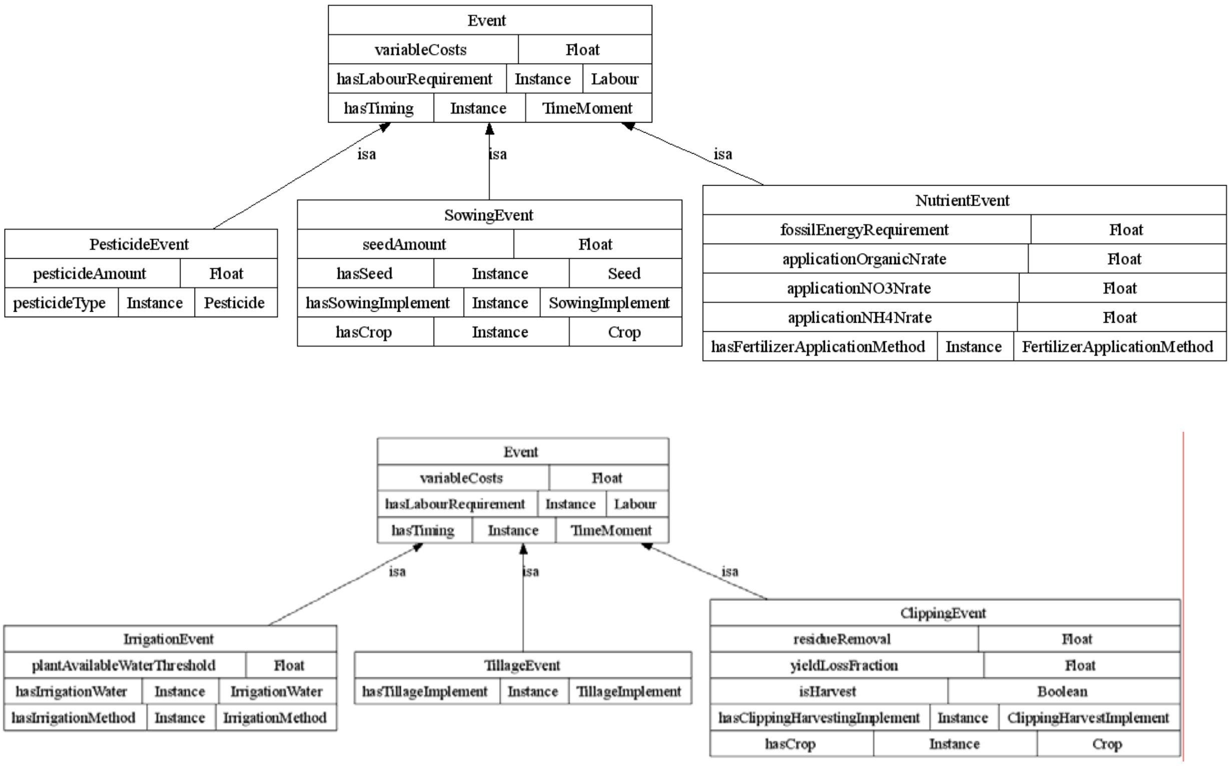


Figure 9 Different types of events and their properties

2.2.2.9 <u>Technical Coefficient Generator</u>

The Technical Coefficient Generator (TCG) links the agronomic input and output coefficients generated by PEG, PTG and APES to socio-economic inputs and outputs by simple calculations. The TCG also describes and adds the standard management operations that do not lead to alternatives in production techniques. In addition, the TCG quantifies other or remaining inputs of each crop in each agricultural activity, i.e. the inputs not simulated through APES. These inputs, for example, refer to all inputs associated with management operations not considered critical for the performance of crop activities (e.g. harvesting operations), and labour and machinery requirements associated with management operations simulated with APES. In addition, the TCG may be used to convert physical inputs into monetary values and to calculate compound environmental performance indicators, for example, incorporating human toxicities and environmental damage related to biocide use. An important task of the TCG is to summarize output generated by APES into information and formats suitable for FSSIM-MP. For example, N leaching losses are calculated on daily basis in APES, while only seasonal or annual losses will be of interest for FSSIM.

The result of the TCG is a fully quantified set of agricultural activities (Technical Coefficient Matrix) that can be offered to FSSIM. This Technical Coefficient Matrix contains sets of technical coefficients, which are organised by Production Coefficients. A Production Coefficient can be understood as a row in the input-output matrix, which describes for a crop in a rotation with a certain management what the technical coefficients are. A Production Coefficient is an ordered collection of technical coefficients as requested by FSSIM-MP. The TCG can produce input-output matrices on different temporal scales: on a daily, yearly or seasonal basis, or per crop or rotation, depending on the needs of FSSIM-MP.For example, a rotation including a potato crop has been linked to a certain production technique which has been evaluated by APES: 150 kg N is used, together with 3 herbicide sprayings in which 4 kg of active ingredients are used. On the basis of this info, the TCG calculates labour and machinery requirements, specific costs of each input item, and possibly other characteristics/indicators of agricultural systems (e.g. biocide index).

Variable costs are calculated by adding input costs for fertilizers, seeds, biocides and the costs of implements associated with each event. Each implement is characterized its (hourly) rental price and by its capacity (hours needed per hectare). The rental price includes all costs related to the implement, with the exception of labour costs, as labour costs are calculated in FSSIM. Typical rental prices include costs related to the depreciation and maintenance of the implement, and the tractor and fuel costs. Labour requirements are calculated by summing the labour needed in all events. Optionally a distinction can be made between skilled and unskilled labour although this is currently not required in FSSIM-MP.

2.2.2.10 Data requirements and data availability to identify the Farm Type

FSSIM is aimed to be applied to a set of farms representing the farm types across the EU-25, since it is impossible to apply the detailed farm models to the entire territory of the EU. This issue required the selection of 20-30 sample regions (4 of them were taken for the Prototype 1) and then the development of a farm typology taking into account the heterogeneity in farming and biophysical endowment. Based on Farm Accountancy Data Network (FADN) and Farm Structural Survey (FSS), this farm typology, developed by WP4, has provided for each sample region (NUTS2 level) a set of typical well defined farms in terms of size, intensity, land use and specialisation. The developed typology was then linked to bio-physical characteristics for spatially allocating the farm types to sub-regional areas, with more homogenous bio-physical endowments. The aim was to divide each sample region into 3-6 types of areas, depending on the bio-physical characteristics (climate, soil, altitude…) and then to spatially allocate the farm types in these bio-physical areas. The specification of each biophysical region was based on one series of weather data; one or several soil types and other unique geographical attributes.

The consequence of this in terms of modelling is to generate, with FSSIM-AM, a specific set of agricultural activities for each biophysical region (i.e. input output matrix) and then to link these sets to farm types and feed them into FSSIM-MP. For reaching this aim, two alternatives were envisaged:

- Alternative 1: one FSSIM model per farm type and biophysical region. That is, for each farm type in each region only one input-output matrix has to be specified. The principal advantage of this option is that it is easier in the interpretation of results, as it closer to real farms. The major disadvantages are: (i) the large number of FSSIM models and (ii) the difficulty to calibrate FSSIM, as the observed data are of average farm types in NUTS 2 level across biophysical regions. Even if they are spatially located, farms will be the same in terms of observed data.

- Alternative 2: one model per farm type with differentiated data and constraints in sub-matrices dealing with the different biophysical regions inside the administrative region. This means that for each farm type which can appear in several biophysical regions, an FSSIM model will be specified. The relationship between a farm type and one or more biophysical regions is determined by the allocation procedure. Lets take an example of one farm type 'arable/medium size/intensive' which can appear for 40% of its acreage in biophysical region 'heavy clay and dry', for 25% in biophysical region 'loamy clay and dry' and for 35% in biophysical region 'sand and wet.' So for this farm type 'arable/medium size/intensive' one FSSIM model has to be specified, which has three different biophysical environments and thus three different set of agricultural activities; one for each biophysical environment.

For the Prototype 1, we aim to use the first alternative; i.e. for each farm type in each biophysical region, one FSSIM model will be specified. For example, if there are three farm types and three biophysical regions, we have to build nine FSSIM models.

2.2.2.11 Representation of the Farm Type: the "average" or the "typical" farm

The farm type consists of a group of farms with similar characteristics (e.g. size, specialization, land use). In order to simulate the behavior of a certain farm type with a farm model (like FSSIM-MP), it is important to select/construct a farm that represents adequately the whole group of farms that are classified in the same farm type.

Representation of the farm type could be achieved either by the "average" or the "typical" farm. The average farm could be defined as a virtual (not observed in reality) farm which is derived by averaging historical data from farms that are grouped in the same farm type. A "typical" farm is an existing (observed) farm with representative (for a certain farm type) properties and characteristics. Different approaches could be used when trying to identify a representative typical farm (e.g. selecting the farm that is closer to the "average" farm or the one with the median profit).

An important advantage of using the "average" instead of the "typical" farm is related to the major objectives within SEAMLESS which are:

- To provide information on supply in market modeling across EU 25.
- To make detailed analysis of the consequences of policy decisions on specific regions.

To achieve these goals, it is important to aggregate the results of the farm model to the regional and market scale. It seems that the simulated farm plans of an "average" farm are less specialized than the results of a "typical" farm. This is mainly because production activities of farms that belong to the same farm type, are represented in the base year data (i.e. calibration data). Even less common activities that might not be interesting to be included in a farm level analysis but are probably important in higher level analyses, are also represented in the base year data. On the contrary, the base year data of a "typical" farm are restricted to the few production activities that are part of the farm plan of a single farm. Consequently, predictions at aggregated (regional) level will be more realistic and questions related to the above objectives will be more accurately addressed when the behaviour of the "average" farms is simulated.

Another important advantage of the "average farm" is its lower data requirements (from FADN or other data sources). Information at farm type level and the number of farms represented in each farm type is sufficient in order to do calculations for average farms. It is not necessary to have farm specific FADN data which are not always available due to confidentiality restrictions. On the contrary, farm specific data are required in order to select a "typical" farm.

The major disadvantage of simulating the "average farm" is that the "average farm" does not actually exist (it is a virtual/hypothetical farm). There is no evidence that a farmer would ever choose a system with multiple crops and management techniques (e.g. 24 different crops have been derived for the average farm type of Flevoland). It is also possible that the "average farm" due to multiple cropping and management patterns would not be very efficient (due e.g. to economies of scale) and would probably produce low levels of outputs. Consequently, even if it is assumed that such farming system exists, it is not guaranteed that it would be classified in the same farm type from which it was derived.

Moreover, it is expected that the calibration of the farm model with data of an "average" farm is more difficult. More calibration constraints or (and) non-linear terms are required to fit the simulated results with historical data. Models, which are tightly constrained, can only produce the subset of results that the calibration constraints dictate (Howitt, 2005).

Both approaches seem to have advantages and disadvantages and this makes it very difficult to decide which of them should be used to adequately represent the Farm type. The purpose of the study seems to be the main criterion to select the appropriate approach. Due to the above mentioned objectives of SEAMLESS, which focus on the provision of information and the consequences of policies at regional and EU level rather than at individual farm level, it was decided to use data from the "average" farm to represent the Farm type.

2.2.3 FSSIM-MP: Mathematical Programming Model

The Mathematical Programming Model (FSSIM-MP) constitutes the core of the bio-economic farm model (FSSIM). Based on mathematical programming approach, FSSIM-MP seeks to solve a set of equations describing the relationships between decision variables of a given system, using external input data coming from FSSIM-AM. The use of a mathematical programming approach has the advantage to enable the inclusion of compensation payments, set aside obligations, voluntary set aside and sales quotas, as well as important relations between agricultural production activities (Britz et al, 2002).

FSSIM-MP is build on the philosophy of a *generic* model (i.e. the same structure (equations and variables) but different parameters) in order to be valid for every type of arable farming system and every geographic location in the EU25. However, this doesn't imply that all constraints and equations which are included in FSSIM-MP, will be activated in each farm type. The activation of equations/constraints will depend on the liking of the user and the specific application. That is, before running FSSIM-MP for a representing farm, the user has to select the set of real equations/constraints according to the observed situation.

As we said above, the first version of FSSIM-MP will have a limit number of variants depending on the farm types and conditions to be simulated. The challenge for FSSIM-MP will be (1) to reproduce, as closely as possible, the farmer's observed behaviour, (2) to appreciate the cognitive structures and the decision rules which govern family farmer's decision making (positive assumption: calibration phase) and (3) to assess the impact of agricultural and environmental policies, behavioural changes and new production technologies on production, revenue, environmental resources, etc (normative assumption: simulation phase).

The principal components of FSSIM-MP are:

- A set of decision variables that describe the agricultural activities and state of the system;
- An objective function describing the farmers' behaviour and goals in particular concerning risk;
- A set of explicit physical, financial, technical, economic, agronomic … constraints, representing specifications for system operation;
- A set of policy and environmental measures (price and market support, quota and seta-side obligations, cross-compliance restrictions, etc) as included in the Common Market Organisations (CMOs) regulations and some specific Regulations.

The following section describes in detail and in mathematical term all the elements of these components.

2.2.3.1 Objective function

Most studies that have modelled farmer decision-making, have assumed that farmers are rational profit maximizers and their production decisions are influenced mainly by the relative prices of inputs and products (Falconer and Hodge, 2000). They have nevertheless ignored the reality that decisions of farmers are generally influenced by the issue of risk, of responding to uncertain events and maybe minimizing the probability of adverse states.

FSSIM-MP is a Risk Programming Model taking into account the risk and uncertainty. Many alternative formulations/approaches could be used to model risk, from relatively complex specifications using expected utility theory to a simplified one, just taking into account expected profit and standard deviation of it, according to states of nature and market. The most commonly used methods for including risk in mathematical programming farm models are Mean-Variance, Mean-Standard deviation, MOTAD (Hazell et al, 1986, Mc Carl, 1989), Target MOTAD (Tauer, 1983) and Focus Loss (Boussard, 1971).

For the moment we start with a basic specification relating to the Mean-Standard deviation method in which expected utility is defined under two arguments: expected income and risk. The objective function will thus presume that the farmers make their decisions in order to maximise the expected income minus some measure of its variability, according to different states of nature and market defined under two different sources of instability: yield (due to climatic condition) and price.

The risk aversion coefficient (ϕ) is exogenously specified. To estimate this parameter, the more common method is to parameterize the model for different values of risk aversion and then to choose the value of the parameter that gives the best fit between the model's predicted crop pattern and the observed values in the base year (i.e. this parameter is used to partially calibrate the model). Producer prices are also taken as an exogenous set.

Equation (1) $Max\ U = Z - \varphi\sigma$

With:

U : Utility (expressed in Euro)

ϕ: the risk aversion coefficient,

σ: the standard deviation of income according to states of nature and market defined under two different sources of instability: yield (due to climatic condition) and price,

Z : expected income (i.e. the average annual income across the number of years (p) of each activity *i* (crop rotation)) (in Euro)

Equation (2) $$Z = \left[\sum_j P_j S_j + \sum_{i,p} P\mathrm{rme}_{i,p} X_i - \sum_{i,p} Cv_{i,p} X_i \right] / N_i$$

*(Average gross revenue: price*quantities + premium - variable cost)*

$$- \sum WTl - \left[\sum Ct\Delta t \right] / N_i$$

- (Average labour cost) - (Average land cost or land income)

i: agricultural activities (i.e crop rotation)

p: year

j: crop and animal products (grain, fodder, sugar, milk, meat,…)

Decision variables

X_i: level of selected activity i (in ha per crop rotation and head per animal)

Q_j: total production of each product *j* within activity *i*

$$Q_j = \sum_{i,p} \frac{\overline{Y_{i,j,p}}}{N_i} X_i = S_j + U_j$$

S_j: sold production

U_j: on-farm used production

ΔT: (land in - land out)

Parameters (i.e. exogenous variables)

P_j: producer price for product *j*

$\overline{Y_{i,j,p}}$: average yield over various states of nature and market for each product *j* and per year

$Prme_{i,p}$: compensation payment for each activity *i* per ha/year (depending on the Common Market Organisations (CMOs))

$Cv_{i,P}$: variable cost of activity i per ha/year or head/year

Ct: land cost per ha (land is valued at its market price without transaction cost)

W: labour cost per hour

Tl: average number of hours rented labour

N_i: number of years within each activity *i*, e.g. the length of crop rotation (2 years, 3 years, 4 years …)

2.2.3.2 Technical and socio-economic constraints

The idea is to develop as much as possible the real constraints operating in different farms in the EU25, and not to build artificial ones, as a good specification of the constraints will help much in the comprehension of the farming system and facilitating the calibrating process of FSSIM. Anyway, it is necessary to check the results of FSSIM before starting the simulation; it is not appropriate to use the model if some constraints are forgotten or violated. Parametric simulations can be done in order to analyse the sensitivity of some constraints, particularly when we are not sure of their exact level.

These constraints should be related to technical (land, water, equipment…) and socio-economic resources (labour, finances, cash flow, risk…) and will be activated for some farm types and inactivated for others according to the observed situation in the base year. When a constraint is inactivated, the corresponding resource is not limited and requirements are calculated *ex post*.

Arable land requirement

This constraint limits the level of the crop production activities to on-farm availability of land, for both irrigated and rain-fed crops.

Several soil types will be defined in the FSSIM model. For each type of soil the technical coefficients of each crop rotation are specified. Some may be the same but others will differ (irrigation and fertilisation levels, yields, etc.). These features imply that we should introduce a number of constraints concerning the allocation of activities to these soils.

Additionally, since it's not easy to implement land market in a farm model, an upper bound representing the available land around each farm will be estimated and included. The idea is to restrict the land enlargement possibility through the acquisition or the hire of land.

For each type of soil (ST), the cultivated arable land (Tused) should not exceed initial arable land endowment (Ti) plus land in (Tin) minus land out (Tou). The land in (Tin) which depends on land market availability can not exceed the estimated upper bound (LM-Av) for each farm.

Equation (3) $$\sum_i X_{i,ST} = Tused_{ST} \le Ti_{ST} + Tin_{ST} - Tou_{ST}$$

X_i: level of selected activity i per soil type (ST)

Ti: initial arable land endowment per soil type

Tused: cultivated arable land per soil type

Tou: land out per soil type

Tin: land in per soil type

Equation (4) $$Tin_{ST} \le LM\_Av_{ST}$$

Tin: land in per soil type

LM_Av: upper bound of land in per soil type

Permanent activities constraints

The static programming model described above is not well suited to handle, in an elegant way, activities with a different planning horizon than arable crops. For these reasons we will not develop these constraints for the moment.

Grassland activities

The specification of these constraints is more complicated than those concerning arable land, as the quality and the access to grassland are very different from one farm to another. For easier modelling, the initial grass land endowment will be specified according to some criteria such as the pasture access, the slope gradient and gravel density and the land access. The aim of this specification is to separate between "easy grassland" (Le) devoted to temporary meadows (renewable and fit to reap) and "difficult grassland" (Ld) devoted to permanent meadows which can be only grazed.

For modelling grassland constraints, we assume, at least in the current version of FSSIM that the sum of grassland activities levels in each farm cannot exceed the initial grassland endowment. Hence, we assume that there is:

- No land market for grassland neither for selling nor for the hiring;
- No interaction between arable land and grassland; i.e. possibilities of converting arable land to grassland and, inversely, from grassland to arable land are not taken into account for the moment.
- No exchange between grassland and forest; forest areas were not considered in the model, since a primary interest was to find strategies for the existing pasture land, rather than considering the possibility of clearing forest for pasture production or abandoning grassland to return to the forest.

Equation (5) $\sum X_{'Grassland', ST} \leq GLi_{ST}$

X_i: level of grassland activity per soil type (ST)

Gli: initial grassland endowment

Equation (6) $GLi_{ST} = Le_{ST} + Ld_{ST}$

Le: easy grassland devoted to temporary meadows

Ld: difficult grassland devoted to permanent meadows

This constraint will be developed further in the dynamic version of FSSIM, as we will introduce here the possibility of renewing temporary meadows and perhaps transfer of these meadows to arable land.

Perennial activities
All perennials crops (Citrus, Apples, Olives, Tobacco, Table grapes, Table olives, Table wine, Tobacco…) are linked to long-term investment decisions and should be analysed with a dynamic programming model or an econometric framework to forecast these crops.

Since the first version of FSSIM is a static one and concerns only annual crops, the perennial crops will be considered in a very simple way, assuming that their levels are equal to the ones observed in the base year (i.e. no investment possibility was introduced). However, this will be improved when applying the dynamic version of FSSIM.

Equation (7) $X_{'PER-CROP',ST} = X^{0}_{'PER-CROP',ST}$

$X_{PER-CROP,ST}$: level of perennials crops per soil type (ST)

$X^{0}_{PER-CROP,ST}$: the observed level of perennials crops in the base year.

Labour requirement

This constraint should be considered carefully, because normally there is not a strict amount of labour available but firstly, there are different types of labour for doing different things, with different costs and secondly, available working days according to weather are limited.

For the first one which is the simplest, the "gender issue" can be tackled, when there is a clear division of labour across genders. In some cases it is necessary to take into account "supervision labour", when using temporary labour force. This means that for a certain number of temporary workers the farm needs a supervisor that can be a member of the family or somebody with a specific higher wage. To take into account all these specificities, different types of labour (K) will be included in FSSIM and each one will have its specific labour limitations.

Concerning the second point, everybody knows that the other real labour constraint is the available working days according to weather conditions. It is a constraint that limits in many cases the access to fields and prevents farmers from carrying out certain tasks. To take into account this constraint, the elegant way is to incorporate time periods in FSSIM even if this will require a large amount of data.

Formally, these constraints can be specified as follows: for each labour type (k), the sum of labour required for each activity (Lreq), expressed in hour, should be less than the amount of family (LF) and permanent (LP) labour available in the farm, plus the amount of temporary labour (TL) if needed. LF and LP are exogenous parameters. However, since it is difficult to define strict limits to market labour availability, we assume that the available temporary labour is unlimited and if in a certain region this appears to be not true, we introduce an upper bound. For example, we can specify that the temporary labour cannot exceed its observed level in the base year or that it can exceed but with a higher labour price.

Equation (8) $$\sum_i X_i * Lr_{i,k,p} / N_i = Lreq_k \leq LF_k + LP_k + TL_k$$

$$LF_k = JO * NFW_k$$

$$LP_k = JO * NPW_k$$

$$\sum_k TL_k = TL$$

X_i: level of selected activity i

$Lr_{i,k,p}$: labour type required per year (P) for each activity *i* (hour/year)

Lreq : sum of labour required for each activity (in hour/year)

LF: family labour available (expressed in hour/year)

LP: permanent labour available (in hour/year)

TL: temporary labour available (in hour/year)

N_i: number of years within each activity *i*

NFW: number of family workers per labour type (k)

NPW: number of permanent workers per labour type (k)

JO: hours available regarding weather conditions

LF: available family working day regarding weather conditions (per labour type (k))

LP: available permanent working day regarding weather conditions (per labour type (k))

Equipment requirement

For each kind of equipment (K'), the sum of equipment required for each activity (Ereq), expressed in hour per year, should not exceed the available equipment (Ea) plus bought equipment (Eb) if needed. This implies that new equipment can be bought if operations are carried out with own mechanisation. The possibility of investing in new equipment, which is an endogenous variable, will depend on the supply and demand of equipment.

Such as for the labour constraint, the availability of equipment will depend strongly on weather conditions, because even if equipments are available, fields may not be accessible.

The equipment constraint can be expressed as follows:

Equation (9) $$\sum_i X_i * Er_{i,k',p} / N_i = Ereq_{k'} \leq Ea_{k'} + Eb_{k'}$$

$$Ea_{k'} = JOe * NE_{k'}$$

$$\sum_k Eb_{k'}' = Eb$$

X_i: level of selected activity *i*

Ereq: sum of equipment required for each activity (expressed in hour/year)

Ea: available equipment (in hour/year)

Eb: bought equipment (in hour/year)

$Er_{i,k',p}$: equipment type required per year for each activity *i* (hour/year)

N_i: number of years within each activity *i*

NE: number of equipments per type

JOe: hours available for each equipment regarding weather conditions. JOe and JO can be the same or different according to operation types.

Water requirement and irrigation

These constraints are very important for irrigated agricultural regions. Generally, two types of constraints are identified concerning the water use: irrigable land and water availability.

Irrigable land requirement
For each type of soil (ST), the sum of area devoted to different irrigated activities cannot exceed the available irrigable land (ISi) in each soil.

Equation (10) $\sum_i X_{i,ST} \leq ISi_{ST}$

X_i: level of selected activity *i* per soil type (ST)

ISi: available irrigable land per soil type (ST)

The available irrigable land (ISi) can be taken into account under two formulations:

- If investment possibility is neglected, availability of irrigable land (ISi) is specified as an exogenous variable, and its suitable value will be the one observed on the base year.
- If investment possibility is taken into account, the available irrigable land (ISi) will be an endogenous variable which depend on initial irrigable land observed on the base year (ILi) but also on investment in irrigation equipment (Inv_R).

Equation (11) $ISi = Ili + \sum_R Inv_R$

ISi: available irrigable land per soil type (ST)

Ili: initial irrigable land observed on the base year

Inv_R: investment in different irrigation techniques (R)

Water requirement
In each year (P), the sum of water used for each selected activity (Wused) should not exceed the water volume available (Wa).

Equation (12) $\sum_i X_i * Wr_{i,p} / N_i = Wused \leq Wa$

X_i: level of selected activity *i*

$Wr_{i,p}$: irrigation water requirement per year for each activity *i* (m^3/year)

Wused: sum of water used for selected activities (m^3/year)

N_i: number of years within each activity *i*

Wa_p: total available water per year (m3/year)

2.2.3.3 Risk constraints

As written earlier in this report, risk is introduced in the FSSIM according to the Mean-Standard deviation approach, inspired by Freund (1957). This method computes for each combination of states of nature the negative deviation of actual income from its expected value. Each state of nature and market (k) is defined under two different sources of instability: yield (due to climatic condition) and price.

For the moment, we assume that there is no dependence (no covariance) between yield and price variation, as prices in EU25 are generally defined at higher scales (international market or EU25) and not on local scales.

Equation (13) $Z_k + Dev_k = Z$

Dev_k: deviation dependent upon different states of nature and market (k)

Z: expected income

Z_k: income of different states of nature and market

Equation (14) $$Z_n = \left[\sum_j Pn_{k,j} S_{k,j} + \sum_{i,p} Prme_{i,p} X_i - \sum_{i,p} Cv_{i,p} X_i \right] / N_i - \sum WTl - \left[\sum Ct\Delta t \right] / N_i$$

X_i: level of selected activity *i*

$Pn_{k,j}$: price defined according to different states of nature and market (k)

$Q_{k,j}$: total production of each product *j* within activity *i* according to states of nature and market

$$Q_{k,j} = \sum_{i,P} \frac{Y_{i,k,j,P}}{N_i} X_i = S_{k,j} + U_{k,j}$$

$Y_{i,k,j,p}$: yield defined according to different states of nature and market (k)

$S_{k,j}$: sold production according to states of nature and market

$U_{k,j}$: on-farm used production according to states of nature and market

Equation (15) $$\sigma = \left[\frac{\sum_k (\mathrm{DEV}_k)^2}{N} \right]^{1/2}$$

Dev_k: deviation dependent upon different states of nature and market (k)

N: number of states of nature and market

σ: the standard deviation of income according to states of nature and market

2.2.3.4 Policy description

One of the main purposes of the SEAMLESS modelling system is the analysis of agricultural and environmental policy measures, either proposed or actual, and the quantitative assessment of the effects of policy and market changes. For this purpose one should consider all the instruments of these policies and how they are, or will be, implemented in EU25.

The modelling of agricultural and environmental policies in FSSIM-MP will be developed in the course of time and in response to needs for increasingly detailed analyses. For the first version of FSSIM-MP, policy instruments linked to arable crops will be implemented in the model. These are the CAP support regime (price and market support, set-aside schema, quota system…) included in the Common Market Organisation (CMOs) regulations, as well as certain cross-compliance and specific agro-environmental measures included respectively, in Horizontal and Rural Development Regulations. Specific local regulations such as extensification premium as well as additional premiums granted by Members States will also be modelled. FSSIM-MP will provide a large flexibility to be able to include new policy measures and new issues that will certainly appear.

The implementation of these instruments will be different according to the analysed policies and will depend strongly on information availability. Currently, we plan to implement in FSSIM-MP the following policy description:

- **Base year policy:** the Agenda 2000 (since 2000) reform constitutes the base year policy for which the model was calibrated.
- **Baseline scenario:** the recent CAP reform of June 2003 in Luxembourg, as it would be implemented in 2012 in the EU25 as well as all future changes already foreseen in the current legislation (e.g. sugar market reform), will be considered as the principal policy assumption operating in the baseline scenario. Performed in year 2012 taken as simulation year, the baseline scenario will be the reference for the interpretation and analysis of different policy scenarios.
- **Policy scenarios:** several policy scenarios are scheduled in the SEAMLESS framework through the two test cases (TC1 and TC2). Most of the policy scenarios defined under the Test Case 1 are handled in the market models, as the shocks come from the "market level", whereas the implementation of policy scenarios designed under Test Case 2 are handled in FSSIM-MP, as the shocks enter at the "farm level". For the Prototype 1, only the policy scenarios defined under Test Case 1 are implemented. Thus, all the shocks are entered at market level. However, impact assessment of these policy scenarios will be performed at different levels, for example, using price changes and trade effects at the market level, and production quantities, land use and environmental indicators at regional and farm level.

The following Table 1 gives an overview of the different policy instruments linked to arable crops that are taken into account and how they are considered in FSSIM-MP.

Table 1. Policy instruments to implement in FSSIM-MP

Instrument	**Modelling**	**Data Source**
CAP compensation payment	Linked to agricultural activities	CMOs
Sugar beet quota	Constraints in the system (upper bounds on sales)	CMOs
Compulsory set-aside	Constraints in the system, restrict seta-side to minimum 10% of COP (cereals, oilseeds and protein) crops	CMOs
Voluntary set-aside	Constraints in the system, restrict total set-aside to 33% of COP crops	CMOs
Environmental condition/cross-compliance	Restrictions in the system (controlled by binary-variables)	CMOs + specific national and regional implementation
Agri-environmental measures	Restrictions in the system (controlled by binary-variables)	CMOs+ specific national and regional implementation
Modulation of payment	Linked to premium	CMOs

CMOs: Common Market Organisation

Modelling all these instruments is an important challenge for FSSIM-MP, even if some of them are implemented in an identical way everywhere in EU25 (e.g. direct payment), others such as environmental measures have quite different national/regional implementations. In addition, the information on the administrative implementation of these specific measures is usually scarce, and often not systematically monitored, not published or even not open to the public. These issues have putted several limits to handle these measures in an elegant way and have induced, in some cases, a simplification of reality.

2.2.4 FSSIM Calibration

2.2.4.1 Calibration and validation in optimization models

In model evaluation we generally distinguish model calibration and validation. Calibration is a process of adjusting model parameters in order to reproduce the observed reality of a base period. It can be understood as a way to find the best parameters that characterize a given model and as a method to maximize the similarity between the output of the model and the observed data (Santillana, 2005). According to Howitt (2005), the calibration process is one of using a hypothesized function and data on input and output levels in the base year to derive specific model parameter values that “close" the model. By closing the model, we mean that the calibration parameters lead to the objective function being optimized for the base year conditions, such that it matches the observed base year values. The calibration process can be done against a base year or an average over several years.

The calibration process is necessary, but insufficient to ensure the model’s robustness for different applications, that is why we should also validate it. Validation consists of testing calibrated models with respect to additional sets of input data (independent data) (Thomann et al 1982). It consists of checking that the calibrated model applied for an *ex-post* analysis would be able to track, as closely as possible, historical situations. According to Hazell and Norton (1986), validation is a process that leads to (i) numerical report of the model’s fidelity to the historical data set; (ii) confirm that, the calibrated model is applicable over the limited range of conditions defined by the calibration and validation data sets; (iii) improvements of the model as a consequence of imperfect calibration; (iv) a qualitative judgement on how reliable the model is for its stated purposes, and; (v) a conclusion for the kinds of uses that it should not be used for. Hence, it is important that collection of calibration and validation data must cover the range of conditions over which predictions are desired. The data used for calibration should be fully independent from the validation data (Himesh, 2000).

In this document, we deal only with calibration issues. Traditional optimization models using linear programming are often based on *normative* assumptions aimed at identifying the “best” production combination under the hypothesis that the initial situation is not binding in terms of production choices. For policy assessment, this assumption that induces a wide divergence between base period model outcomes and observed production patterns, is unacceptable. This *normative* LP has been replaced by the *positive* type of model, in which the main objective is to reproduce the observed production situation in order to be able to simulate a likely behaviour of the farmers when varying parameters determined by the agricultural policy intervention. However, in order to reproduce the observed situation, an LP model has to be calibrated.

Linear programming rely on data based on observed average conditions (e.g., average production costs, yields, and prices), which are expressed as fixed coefficients. As a result, these models tend to select crops with highest average returns until resources (land, water, and capital) are exhausted. The predicted crop mix is therefore less diverse than observed in reality and overspecialisation of the solution occurs. The most widespread reason for overspecialisation is that the basis matrix of valid empirical constraints has a rank less than the number of observed base year activities. Since the number of nonzero activities in an LP framework is upper bounded by the number of resource constraints, overspecialisation must occur by design (Heckelei, 2002).

To consider the calibration problem described above and reduce the specialization errors in optimizing models, a common approach was used which consist of constraining the crop supply activities by adding linear constraints. However, calibration by adding linear constraints is also unsatisfactory, as models that are tightly constrained can only produce the subset of normative results that the calibration constraints dictate (Howitt, 1995b). To account for this calibration problem, two other broad categories of approaches have been used: approximate and exact calibration methods. Based on some of the theoretical and methodological aspects of mathematical programming, these approaches try to make the model better able to represent the production choices made by a farmer, and to provide greater capacity to analyse the problems of agricultural policy.

2.2.4.2 The Approximate Calibration Approaches

The common point between what we called "the approximate calibration approaches" is that they do not seek to calibrate exactly optimisation models but rather to fit model's predictions to observed data accepting a residual deviation between simulation and reality.

The approximate calibration approach includes all traditional calibration methods, which require more complicated constraint structures to reproduce the observed cropping pattern. Among these methods we have those based on the use of rotational constraints or step functions over multiple activities to curtail the overspecialization (Meister et al., 1978), and those imposing upper and lower bounds on certain production activities as constraints in a recursive procedure (Day, 1961), etc. The principal limitation of these methods is that the used constraints determine the optimal solution not only for the base year, for which they are appropriate, but they also affect policy simulation runs that attempt to predict the outcome of changed prices, costs or resource availability. The solution of the model under policy runs is therefore significantly restricted by the base year solution constraints.

This approach includes also all the methods which are based on the addition of risk and uncertainty to the linear programming model such as MOTAD, Target MOTAD, Mean-Variance, Safety first, etc (Hazell and Norton, 1986). Most of these risk methods presume that a non-correspondence between model results and observed situation is related to one of the following two factors: (i) omission of some important element of the cost structure, such as specialized management skills in growing high-value crops; (ii) inadequate specification of the risk associated with different activities and farmers' risk aversion (Hazell and Norton, 1986). To capture this last factor adequately the objective function should include the risk cost as well as risk-averse behaviour. Several methods have been proposed to take into account the risk. Some of them treat risk in the objective function (MOTAD, Target MOTAD, Mean-standard deviation, Mean-Variance, Safety first …) and other in the constraint set (Discrete stochastic programming, Chance constrained programming…). The wider used risk approach is the mean-standard deviation which assumes that a farmer's preferences among alternatives farm plans [$U = E-\phi\sigma$] are based on expected income [E] and its standard deviation [σ] multiplied by risk aversion parameter [ϕ]. To estimate this parameter the more common method is to parameterize the model for different values of risk aversion (ϕ) and then to choose the value of the parameter that gives the best fit between the model's predicted cropping pattern and the actual values observed in some base period (i.e. this parameter is used for calibrating the model) (Hazell and Norton, 1986). The type of risk and uncertainty, which are generally taken into account in these models, is income risk due to the variability of yield due to climatic conditions, prices and subsidies. The risks in resources supplies, for example, seasonal labour, water for irrigation, and forage supplies for livestock feed, are generally neglected. These methods have been applied to different agricultural systems under various agro-ecological conditions and at different analysis levels (i.e., farm, region, sector, country) (Boussard, 1988; Boussemart et al, 1996).

Another range of methods that can be classified in this broad approach are the multi-criteria approaches such as the weighted goal programming (WGP) and the Min-Max goal programming (MINMAX GL). These two approaches minimize the deviation of each objective from observed value of the same objective. The unwanted deviations (either positive or negative or both) are weighted, indicating the importance attached to each objective. To allow for summation, the deviations are normalised by dividing them by the expressed range of their respective objectives (Charnes et al., 1955; Simon, 1979; Rehman and Romero, 1993; Tamiz et al., 1998; Ballestero and Romero, 1998).

As we said above the common feature of all these approximate approaches of calibration is that they have reduced the calibration problem, but they cannot calibrate a model exactly and substantial calibration problems remain in many cases.

There is no consensus on the statistic to be used in evaluating the fit, but in most case, simple measures such as the mean absolute deviation[5] (MAD), or the percentage absolute deviation (PAD), have been used. The Theil index[6] and Nash coefficient[7] have been suggested as well. However, there does not exist a threshold value, for example of the PAD, which clearly determines acceptance or rejection of the model. For example, for sector models, Hazell and Norton suggest that a percentage of absolute deviation (PAD) for production and acreage below 10% is good, equal to 5% is exceptional and more than 15% indicates that the model may need improvement before it can be used. These thresholds are always subjective.

Hazell and Norton (1986) suggest also six tests to improve model calibration. First, a capacity test checks whether the set of model constraints allows the base year production. Second, a marginal cost test ensures that the marginal costs of production, including the implicit opportunity costs of fixed inputs, are equal to the output price. Third, they suggest a comparison of the Dual value on land with actual rental values. Three additional comparisons of input use, production level, and product price are also advocated (Howitt, 2005).

---

[5] Mean absolute deviation:

$$d = \frac{1}{n}\left|\sum_i (\hat{X}_i - X_i)\right|^2$$

Where $\hat{X}_i$ is the observed value of the variable *i,* X_i is the simulated value (the model prediction) and *n* is the number of samples..

[6] The Theil index:

$$\frac{\left[\frac{1}{n}\sum_i(\hat{X}_i - X_i)\right]^{/2}}{\left(\left[\frac{1}{n}\sum_i(\hat{X}_i)^2\right]^{/2} + \left[\frac{1}{n}\sum_i(X_i)^2\right]^{/2}\right)}$$

Where $\hat{X}_i$ is the observed value of the variable *i,* X_i is the simulated value and *n* is the number of samples.

[7] The Nash coefficient (widely used in hydrology):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Where $\hat{X}_i$ is the observed value of the variable *i*, Xi is the simulated value, $\overline{X}$ is the mean of the measured values and n is the number of samples.

A related disadvantage of all these calibration approximate approaches is the *ad hoc* nature of the calibration process. The calibration procedure is generally manual and if a model is manually calibrated, the objective may be assessed qualitatively: fits may be obtained by eye and intuition then play a part in choosing appropriate calibrated parameter sets. For automated calibration, an "objective function" or functions giving a mathematical definition of how good a solution is must be formally specified (Jackson et al, 2005).

As an alternative to the approximate calibration approaches, another range of approaches has been proposed since a few years in order to exactly calibrate optimization models exploiting the observed behaviours of economic agents. A brief overview of these approaches and their extensions will be developed in the following section.

2.2.4.3 The Exact Calibration Approaches

An alternative solution for the calibration problems was proposed in the late 1980's by Richard Howitt through a new methodological approach called "Positive Mathematical Programming" (PMP). The standard approach has been described by Howitt (1995a), but the technique has been employed by a series of pragmatic, policy oriented modelling exercises long before this period - following Howitt's non published proposal - (for example: House, 1987; Kasnakoglou and Bauer, 1988; Horner et al, 1992). The term "positive" implies that, like in econometrics, the parameters of the non-linear objective function are derived from an economic behaviour assumed to be rational given all the observed and non-observed conditions that generate the observed activity levels (De Frahan, 2005). This approach stipulates that, a divergence between model's prediction and the observed reality of a base period means that both technical constraints and cost (or yield) specification were not completely taken into account, and so they had to be included in the objective function via a nonlinear cost (or/and production) function (Gohin et Chantreuil, 1999). The principal advantages of this approach -compared to *ad hoc* calibration procedure- include: automatically and exact calibration of optimization models exploiting the observed behaviour of economics agents, lower data requirements, and continuous changes in exogenous variables (Röhn and Dabbert, 2003).

According to Howitt (2005), two conditions should be fulfilled in order to exactly calibrate optimization models: the nonlinear and the dimension calibration conditions.

1. *The nonlinear calibration condition:* if the number of non-zero crop activity levels observed (k) exceeds the number of binding constraints (m), the profit function is nonlinear in land for most crop activities, and the observed crop allocations are a result of a mix of unconstrained and constrained optima (Howitt, 2005). The equilibrium conditions for this case are satisfied if some, or all, of the cropping activities have decreasing returns to land as the crop acreage is increased. The most common reasons for decreasing returns per acre are declining yields due to heterogeneous land quality, risk aversion, or increasing costs due to restricted management or machinery capacity. This condition is necessary but not sufficient, because many models have some nonlinear terms in the objective function reflecting endogenous price formation or risk specifications, but they cannot calibrate precisely the model.
2. *The calibration dimension condition:* The number of calibration terms in the objective function must be equal to or greater than the number of independent variables to be calibrated. The ability to adjust some nonlinear parameters in the objective function, typically the risk aversion coefficient, can improve model calibration. However, if there are insufficient independent nonlinear terms the model is unable to calibrate precisely.

PMP approach involves three phases: calibration, estimation and simulation.

1. *The calibration phase:* consists of writing an LP model as usual but adding to the set of limiting resource constraints a set of calibration constraints that bind the activities to the observed levels of the base year period. The sole purpose of this phase is to obtain an accurate and consistent measure of the vector of dual values associated with the calibration constraints, but as pointed out by Heckelei and Wolff (2003) this phase can be integrated in the estimation phase by means of Lagrangean multipliers (Howitt, 1995a).

   Paris and Howitt (1998) interpret this vector as capturing any type of model mis-specification, data errors, risk behaviour and prices expectations.

   $$\max Z=\mathbf{p'x}-\mathbf{c'x} \qquad \text{Subject to} \quad \mathbf{Ax}\le\mathbf{b}[\lambda], \quad \mathbf{x}\le\mathbf{x}_o+\boldsymbol{\varepsilon}\,[\rho], \quad \mathbf{x}\ge\mathbf{0} \qquad (1)$$

   Where **Z** is the objective function value, **p, x** and **c** are (n x 1) vectors of product prices, non-negative activity levels, and accounting costs per unit of activity. **A** represents an (m x n) matrix of coefficients in resource constraints, **b** and **λ** are (m x 1) vectors of resource availability and their corresponding shadow prices. The (n x 1) $\mathbf{x_0}$ non-negative vector of observed activity levels, **ε** is an (n x 1) vector of small positive numbers for preventing linear dependency between the structural and the calibration constraints, and **ρ** is an (n x 1) vector of duals associated with the calibration constraints.

2. *The estimation phase:* consists of employing the dual values **ρ** delivered by the first phase to specify additional non-linear terms in the objective function which allows reproducing the observed activity levels without calibration constraints. These terms mostly refer to increasing marginal cost (Arfini and Paris, 2000), or/and a decreasing marginal yields (Howitt, 1995a; Barkaoui et Butault, 1998), or a *neutral* form[8] (Röhm and Dabbert, 2003). A frequent case considers calibrating the parameters of a variable cost function $\mathbf{C^v(x_0)}$, such that the 'variable marginal' cost $\mathbf{MC^v}$ of the activities is equal to the sum of the known cost **c** and the 'non-specified' marginal cost **ρ.** In case of a quadratic function form[9], the following condition for calibration is implied:

$$\mathbf{MC^v} = \frac{\partial\, \mathbf{C^v(x)}}{\partial\, \mathbf{x}} = \mathbf{d} + \mathbf{Qx_o} = \mathbf{c} + \boldsymbol{\rho} \qquad (2)$$

   Where d is an (n x 1) vector of parameters of the cost function and Q is an (n x n) symmetric, positive (semi-) matrix.

   To solve this system of n equations for [N+(N+1)/2] parameters, the literature suggests many solutions, which include simple *ad hoc* procedures with some parameters set a priori (Howitt, 1995a), the use of supply elasticities (Helming et al, 2001), the direct derivation of the unknown parameters from the Kuhn-Tucker

8 For neutral form: $=\rho_i * X_i * (1-\frac{X_i}{\hat{X}_i})$ where ρ_i is the dual value associated with the calibration constraint of activity i, $\hat{X}_i$ is the observed level of activity *i,* and X_i is the simulated activity level.

9 Other functional forms are possible. The generalized Leontief and the weighted-entropy variable cost function (Paris and Howitt, 1998) and the constant elasticity of substitution (CES) production function (Howitt, 1995b) in addition to the constant elasticity of transformation production function (Graindorge *et al.*, 2001) have also been used. A von Neumann-Morgenstern expected utility approach has been used to account for a constant absolute risk aversion to price volatility (Paris, 1997).

conditions (Judez et al, 2001), and the employment of maximum entropy criterion (Paris and Howitt, 1998).

3. *The simulation phase*: consists of adding the estimated non-linear terms (cost (production) function) to the PL objective function in order to simulate the farm's behaviour when some condition changes, such as prices, yields, policy, etc.

During the last decade, PMP has become a popular method for farm, regional and sectors models. It established itself as a widely used approach for the specification of programming models designed for analysing agricultural and environmental policies. It has generated numerous applications and extensions. Among the works developed using PMP we should mention the models of the University of Bonn (Heckelei and Britz, 2000), INRA-Nancy (Barkaoui et Butault, 1998), University of Madrid (Judez et al, 2001), the FAL model (Kleinhanss, 2002) and the CAPSET model (Paris et al, 2002). Some other applications are shown in Howitt and Gardner (1986), House (1987), Kasnakoglu and Bauer (1988), Arfini and Paris (1995) and Helming et al (2001).

During the last decade, several expanded frameworks of the PMP methodology have been developed in order to overcome some of the critique as to the original version. Among this we have (i) the new approaches developed in order to estimate the parameters of the non-linear functions (Helming et al, 2001; Paris and Howitt, 1998; Judez et al, 2001); (ii) those used to solve the problem of excluding crops that are not present in the base year (self-selection problem) (Paris and Arfini, 2000); (iii) those to deal with the problems of zero-marginal product (cost) for one of the calibrating constraints (Gohin and Chantreuil, 1999; Paris and Howitt, 2001; Röhm and Dabbert, 2003), (iv) those for fixed coefficient technology problem as well as for the use of many observed data (Paris and Howitt, 2001); etc. To deal with the two last problems, Paris and Howitt (2001) suggested a new version of PMP takes on the structure of a Symmetric Positive Equilibrium Problem (SPEP) that overcomes residual criticisms raised against the original version. This new version will appear substantially different from the original specification but it follows the same inspiration and goal.

In spite of the attractiveness of this technique and its popularity, it has several important problems:

(i) Any version of a PMP method will always exactly calibrate the model to the observed situation, but under different scenarios, model behaviour may be different and not in line with theoretical expectations (Röhm and Dabbert, 2003);

(ii) With only one observed set of data it is possible to construct an infinite number of non-linear curves that 'calibrate' the model and hence any economic interpretation placed on these functions becomes non-justified (Yates and Rehman, 2002);

(iii) This approach cannot take into account, in an explicit way, technology and the same crop grown under two technologies as if they were two separate crops, which may lead to unsatisfying results (Röhm and Dabbert, 2003);

(iv) The PMP approach cannot resolve the self-selection problem if it is applied at farm level (e.g. alternatives activities which are not observed in the base year cannot be easily selected by the PMP approach in a policy simulation).

2.2.4.4 Criteria to design calibration methodology for FSSIM

This part presents certain criteria to take into account in the selection of the calibration methodology for FSSIM. These criteria will help in the choice of the calibration methodology but also in the definition of the assumptions to take into account in the calibration process for FSSIM.

The main specifications that the calibration methodology for FSSIM requires are:

- The calibration process should be easily applicable for all farm types in EU, be transparent and calibrate the model automatically, as in the final version of SEAMLESS we will have about 300 bio-economic farm models, so it is not realistic to calibrate manually all theses models.
- The calibration procedure can be either exact or approximate even if exact calibration is desirable, because the project priority is not to develop models which are able to reproduce exactly the observed data and to reach a consistency with the data base, but rather to describe well the system and to simulate trends as a result of changes in policies and agro-technologies. What is important is that the model describes the system in detail and produces, in the calibration phase, results which are not very different to observed data. However, the challenge in this case (i.e. if we apply an approximate calibration approach) is to propose a list of calibrated parameters, to select a statistical method in order to evaluate the fit, and to choose criteria for acceptance or rejection of the model.
- The calibration procedure should be extremely transparent[10] and should not add parameters, terms or constraints binding the model which have no clearly defined economic or technological background. Each additional parameter, term or constraint able to calibrate the model but potentially restricting the model behaviour under policy runs (simulation phase) should be avoided.
- The calibration procedure should satisfy technical requirements and take into account easily the substitution between production activities, as in FSSIM each crop (animal) will be grown under a large number of production techniques and thus the number of generated production activities (current and alternative) will be important.
- This procedure should calibrate the model not only for one variable such as an activity level, but for a few calibration variables such as activity levels (i.e. rotation and crop levels), production level, and income. The precise list of calibration variables for calibration needs to be defined.

2.2.4.5 Procedure for calibrating FSSIM

As we have explained above, in the literature we cannot find a calibration approach able to calibrate automatically and exactly optimization models without adding doubtful terms, parameters or constraints which could restrict the model behaviour under policy runs (simulation phase). On the other hand, approximate calibration approaches, which are usually more transparent, are generally manual and very subjective.

In order to take into account all these problems, we propose to apply a new procedure which combines approximate and exact approaches. This procedure is based on a risk approach complemented by an extension of the Positive Mathematical Programming (PMP) technique, inspired from Röhm and Dabbert (2003). The idea is to use the risk approach to improve the model calibration until reaching an acceptable threshold and then to apply the PMP approach to exactly calibrate the model. That is, we apply the PMP approach only when we obtain good results with a certain value of a risk aversion coefficient, as PMP method will always exactly calibrate the model to the observed situation even if the used data are erroneous.

However, before applying this procedure we should check whether we do not miss some important constraints or activities which can improve model performance.

[10] When we write "transparent" it is opposed to "black box", it means avoiding terms that are just residuals or accepting residuals if they are small.

The proposed procedure for FSSIM calibration involves four steps (figure 10):

- The first step consist of selecting calibration variables (activity level, income, production, etc) and calibration criteria (MAD, STD, Theil index, etc) to calculate the fit between simulated and observed results and a threshold value to determine acceptance or rejection of the model.
- The second step consists of applying the risk approach in order to calibrate as positive as possible the model. For the Prototype 1, this step consist to run FSSIM under different risk aversion coefficients and then to choose the coefficient that gives the best fit between simulated and observed results. These coefficients are estimated within a range of parameter justified from literature.

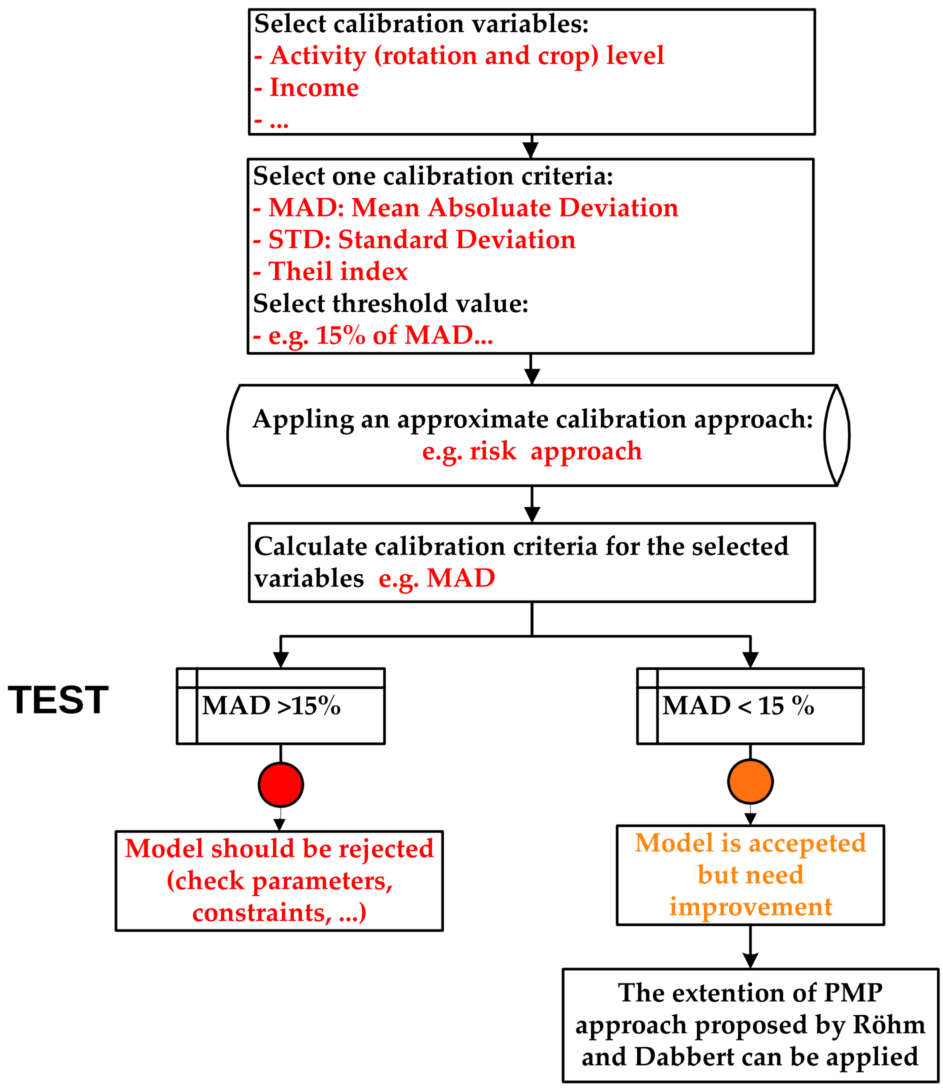


Figure 10 Calibration procedure for FSSIM

- The third step consists of testing model credibility and determines acceptance or rejection of the model. To do this test, some assumptions[11] proposed by Hazell and Norton (1986) will be taken into account: if the percentage of Mean absolute deviation (MAD) for income and activity level is below 15%, the model is acceptable but needs improvement before it can be used; if the MAD exceeds 15%, the model is bad and should be rejected (user should check model parameters, constraints, etc.).
- The forth step concerns only models which have passed the previous step (MAD is below 15%) but that need improvement. It consists of applying an extension of the PMP approach inspired from Röhm and Dabbert (2003). This approach is the only one which is able to take into account explicitly technology requirements of farm models. The idea of this approach is to consider the same crop under different technologies as similar crops and not as different crops as suggested by the standard approach of PMP. This assumption induces that the substitution between similar crops should be stronger than between different crops. To take into account this assumption, Röhm and Dabbert propose to divide the slope of the marginal gross margin function of each crop in two parts. One part depends on the activity level of the crop specified by technology (production technique) and the other on the activity level of total crop.

$$\max Z = \mathbf{p'x} - \mathbf{c'x} \qquad (5)$$

$$\text{Subject to:} \quad \mathbf{Ax} \le \mathbf{b}\,[\boldsymbol{\lambda}], \qquad (6)$$

$$x_{C,T} \le \mathbf{x}^0_{C,T} + \boldsymbol{\varepsilon}_1\ [\rho_1], \qquad (7)$$

$$\mathbf{x}_C \le \mathbf{x}^0_C + \boldsymbol{\varepsilon}_2\ [\rho_2], \qquad (8)$$

$$\mathbf{x} \ge \mathbf{0}$$

Constraints (6) are called resource constraints while constraints (7) and (8) are called calibration constraints. The first calibration constraint (i.e. constraint 7) is related to crop specified by technology and the second one for all production techniques of a specific crop in total.

As in the PMP standard approach the dual values ρ_1 and ρ_2 delivered by the first phase are used to specify additional non-linear terms in the objective function which allows reproducing the observed activities levels without calibration constraints (Röhm and Dabbert, 2003).

However, in order to be applied for FSSIM, the Röhm and Dabbert approach should be adapted and some specificity should be taken into account, as FSSIM contains much complexity such as the large number of crop rotations and production techniques, which can prevent its calibration even with this approach. This adapted approach will be developed in a next paper if all involved partners agree with this procedure.

2.2.5 Conclusion and future development

This section gives a detailed description of the first version of FSSIM, especially its structure, components and model linking. This description will be continuously revised, in order to

[11] This is just an example and other assumption concerning the calibration criteria as well as threshold value can be taken into account.

include all the targeted description of the FSSIM decided in SEAMLESS frame. The principal actions that we plan to develop in the future regarding FSSIM are:

- Completion the development of tools of FSSIM-AM and FSSIM-MP for arable farming activities;
- Including others model specifications (e.g. other possible constraints, more specification of policy, etc.);
- Improving the calibration procedure especially the parameterisation of risk aversion coefficient;
- Extending the tools/models of FSSIM-AM and FSSIM-MP for the other farming sectors (livestock systems, orchards, vineyards, etc.);
- Appling the tools to the others sample regions of Prototype 1 and Prototype 2 and their farm types;
- Appling the tools to test case regions;
- Developing a dynamic version a FSSIM-MP for specific questions at regional level;
- Adding more advanced calculations on variable costs, for example by comparing to the disaggregated FADN variable costs or by making the variable costs depend on the farm size to account for economies of scale;
- Adding more environmental effect and indicator calculations apart from the limited set that is currently available;
- Improving the integration between model components.

2.3 CAPRI[12]

2.3.1 Introduction

CAPRI stands for 'Common Agricultural Policy Regionalised Impact analysis' and is both the acronym for an EU-wide quantitative agricultural sector modelling system and of the first project centred around it13. The name hints at the main objective of the system: assessing the effect of CAP policy instruments not only at the EU or Member State level but at sub-national level as well.

The scope of the project has widened over time: the first phase (FAIR3-CT96-1849: CAPRI 1997-1999) provided the concept of the data base and the regional supply models, but linked these to a simple market model distinguishing the EU and rest-of-the-world. In parallel, a team at the FAL in Braunschweig applied CAPRI to asses 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 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).

During the years, the system was applied to a wide range of different scenarios. The very first application in 1999 analysed the so-called 'Agenda 2000' reform package of the CAP. Shortly afterwards, a team at SLI, Lund, Sweden applied CAPRI to analyse CAP reform option for milk and dairy. FAL, Braunschweig looked into the effects of an increase of biological production systems. WTO scenarios were run by the team in Bonn in 2002 and 2005. Moreover, CAPRI was applied to 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

[12] For more specific and detailed information concerning CAPRI, see PD 3.3.5.1

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

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.

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 2004 at the end of the CAP-STRAT project. It is planned to update the documentation on a regular basis if the need arises.

2.3.2 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)14. 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 from farm type to global scale including input and output coefficients.

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

The economic model is split into two major modules (figure 11). 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.

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

The market 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 40 countries or country blocks in 18 trading blocks. 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.

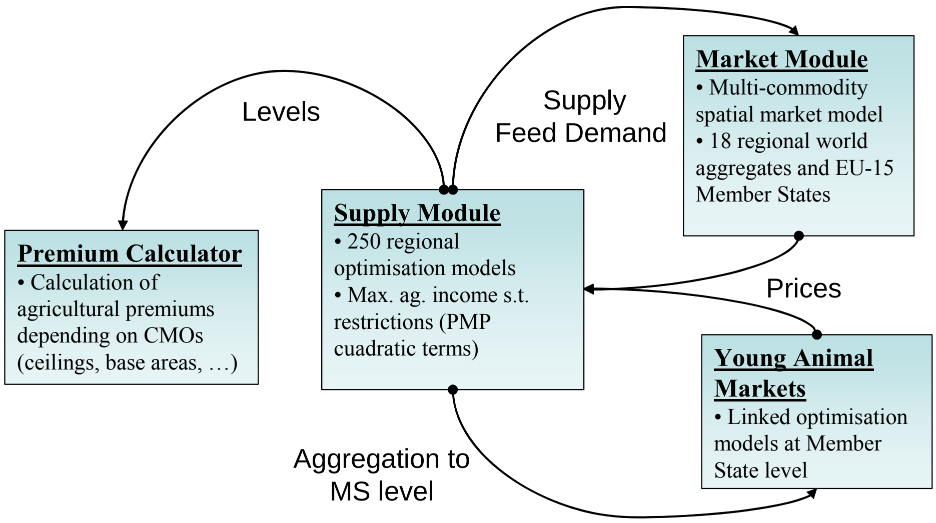


Figure 11 Link of modules in CAPRI (iterative approach)

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.

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 currently realised in C, which interacts with FORTRAN code and libraries which are inter-alias dealing with data base management. Typically, these applications generate run-specific parts of the GAMS code. Exploitation tools apply additionally Java applets for interactive maps and XLM/XSLT to generate interactive HTML tables.

Methodological development, updating, maintenance and application of CAPRI are based on a network approach with is currently centred in Bonn. The team in Bonn acts as a ‘clearing house’: any changes introduced in CAPRI are reviewed by it and, when accepted, become part of the master version. The master version, covering data bases, software and documentation is distributed to all participants of the network usually in the context of training sessions which bring the network together at least once per year. 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.

This network approach has been now expanded through the implementation of CAPRI in several institutions (JRC-IPTS, DG-Agri, LEI and FAL) and the participation in the SEAMLESS Consortium. While in the future most of the mentioned principles should remain (the Bonn team acting as a ‘clearing house’ for model updates, no fees attached to its use and controlled entry to the ‘club’), others might change or be added in the future. Although the Bonn team intends to continue its responsible role for the CAPRI model and to accommodate the needs of the arrangements made, the intensity of the different activities are certainly subject to future funding opportunities.

2.3.3 The CAPRI Data Base

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

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

Wherever possible link to harmonised, well documented, official and generally available data sources to ensure wide-spread acceptance of the data and their sustainability.

Completeness over time and space. As far as official data sources comprise gaps, suitable 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 -currently EU27 plus Norway- need to fit to data at regional level -administrative units at the so-called NUTS 2 level, about 250 regions for EU25- and data at global level, currently 16 non-EU regions broken down to 27 countries or country blocks. 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 crop rotations and herd sizes and a herd flow model for young animals.

Building up the data base at regional or NUTS 2 level, which takes the national data 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.

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).

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.

2.4 Econometric model for estimating Labour allocation to activities

The general objective of this model is to provide activity specific labour inputs at the national and regional level. This forms the basis for a CAPRI post model analysis capturing employment effects of scenario driven changes in the regional production structure. In order to arrive at activity specific labour inputs, FADN data on farm labour input and activity levels are used to estimate labour input per activity level.

2.4.1 General approach

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 per activity level (€/ha or €/animal) or physical terms (hours/ha). The CAPRI data base generally uses physical terms and, where aggregation across different physical measures is required, input coefficients are measured as volume measures using constant prices.

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 analyzing 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.

Input coefficients for different inputs are constructed in different ways. Apart from the estimation of input coefficients for young animals, fertilizer and feed, the remaining inputs in CAPRI are estimated from a Farm Accounting Data Network (FADN) sample and then these estimation results are combined with aggregate national input demand reported in the EAA and standard gross margin estimations, using a Highest Posterior Density (HPD) estimation framework.

As a first step of estimating labour coefficients (family labour and paid labour, both in hours, as well as wage regressions for paid labour) in the SEAMLESS context, standard econometric techniques are employed using 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.

A smaller subset of the resulting estimates were unrealistic. Since a full accounting of activities and labour in the sectoral approach does not allow to simply discard these estimates, a second stage estimation/calibration was set up building a consistency framework around the unrestricted first step estimates.

2.4.2 Econometric Estimation

The general idea of the approach is to explain aggregate input quantities (dependent variable) with the activity levels of all production activities (independent variable) in a pooled cross sectional, time series approach. The estimated regression coefficients represent directly or allow to calculation of labour input coefficients.

One issue to be considered is the "unobserved heterogeneity" or individual effect on labour input coefficients which cannot be explained by activity levels. It is assumed that this varies only across farms and not over time. As a result, it follows that these behavioral differences between individual farms independent of the activity levels are captured by the intercept. Examples of the underlying heterogeneity of farms could be the farms productivity of land, the managerial quality of household members running the farm or other unobserved factors.

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*tons/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.

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 Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) were tested with varying degrees of success.

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 500,000 observations, while most countries have about 15,000-50,000). On average each particular farm appears 5 times in the 13 year panel.

The starting sample sizes and the regional level were, as follows (for the years 1989-2001) unless otherwise stated:

Sample size	Region
•AT - Austria - 2570 farms--→ price data from'95-2001	National
•BL – Belgium/Luxembourg, 2643 farms	National
•DE - Germany, 16745 farms	NUTS I
•DK - Denmark, 7299 farms	National/NUTSII
•EL - Greece, 7152 farms	National
•FI - Finland, 1413 farms--→ price data from'95-2001	National
•IR - Ireland, 3733 farms	NUTS II
•IT - Italy, 52264 farms	NUTS II
•PT - Portugal, 6912 farms	National
•SE - Sweden, 1471 farms--→ price data from'95-2001	National
•UK - United Kingdom, 8102 farms	"Nuts half"
•ES - Spain, 25427 farms	NUTS II
•NL – Netherlands, 4347 farms	National
•FR – France, 15262 farms	NUTS II

Before the econometric regressions the following data cleaning procedures and other data transformations were employed:

- The regressors with less than or equal to 50 observations for both activity levels and yield/herd size were excluded
- All non-zero values were counted
- The data were truncated at zero in order to eliminate reported negative level and yield values and also reported negative real input costs
- Price indices were taken from the COCO database in order to calculate wage input costs in real terms
- The maximum yield/herd size is calculated by economic unit (farm) and it replaced yield and herd size respectively
- The interaction term level*yield/herd size is created
- Variables were weighted by total output
- Indexed data by farm number
- Year dummies were generated
- Data were declared to be a time series

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, but are contracted in the context of another project) : 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.

The main focus for SEAMLESS, obviously, is total family labour and total paid labour. Further development and improvement of the estimations and particularly that of the FADN employment models is anticipated through the incorporation, if possible, of 'engineering' labour requirements for various activities. The work in incorporating both the engineering results and the econometric results into CAPRI should take advantage of the previous general work on inputs as a whole. This will be done in alliance with the Bonn team through the late summer of 2006. The whole set of results is currently being analysed for plausibility.

Finally, it is regarded that the cohort analysis of farm holders, part of the FP6 CAPRI DYNA-SPAT project, will provide information which once linked with the econometric estimation output, is likely to produce more reliable results than what becomes available from the econometric analysis alone regarding predicted labour use coefficients that may be used in scenario building. These results were presented in Brixen, and link regional unemployment levels with numbers of farm holders. They may also be combined with the Markov chain results of the Bonn team (also part of WP3).

2.5 EXPAMOD (from sample to EU)

2.5.1 Introduction

This model will allow for interpolation and aggregation of supply behaviour results obtained by FSSIM models from a sample of regions to all EU regions. The key issue in this regard is that a stratified sample of regions used for the FSSIM simulations allows expansion (transfers) of FSSIM results to regions not in the sample. The idea behind the underlying selection of regions (sampling) is to capture most of the heterogeneity of biophysical and farm variables that exists in the EU thereby allowing for an interpolation of indicators to regions without FSSIM simulations.

The methodology to map the supply behaviour of farm models (FSSIM) to the market model (CAPRI) comprises the following sequence of steps:

1) Simulation of FSSIM supply response to price variations through a set of price experiments to obtain price-quantity data set. The price variation in the experiments shall be realistic with respect to the expected price variations in the application of the system. Possibly a simplified pre-test application of the market model is necessary to obtain the range of price variation in future versions of the system.

2) Estimation of an econometric Meta-model explaining supply response based on prices and other explanatory variables, which determine FSSIM supply response, but are also available for the rest of the EU regions.

3) Use of this Meta-model to extrapolate supply response to farm types in other regions.

4) Aggregation of supply response to the level of CAPRI regions (NUTS2 administrative units) using representativity weights of FSSIM farm types. A mapping from FADN regions to CAPRI regions needs to be included as well as a mapping from FSSIM to CAPRI products at this point.

5) Calibration of regional supply modules in CAPRI to aggregated supply response. In Prototype 1 calibration to elasticities (percentage changes) around base year values is realized. In subsequent versions of the system, a fit to price-quantity data is envisaged.

These steps described above apply to the use of EXPAMOD for the extrapolation and aggregation of the supply response from FSSIM models to the market model. For subsequent prototypes, similar approaches are envisaged for the extrapolation of other indicators from the farm type models to other regions.

2.5.2 Model context

Spatial scale

On the input side, EXPAMOD deals with price-quantity combinations for a particular FSSIM farm type being representative for the respective NUTS2 region. The output is the supply response for a NUTS2 region at the aggregate level. Consequently, we have different spatial scales between input and output of EXPAMOD.

Temporal scale

In Prototype 1, the regression within EXPAMOD is based on data from the static version of FSSIM. Consequently, the time horizon (temporal scale) is implicit in the flexibility of input adjustment modelled in FSSIM. Since the current FSSIM version assumes fixed labour, capital and land resources, a short term horizon is implied (2-3 years). It is to be investigated later how changes to a dynamic version of the FSSIM in combination with the structural change model will influence the temporal scale of the supply response.

2.5.3 Model inputs

For prototype 1 of EXPAMOD it is envisaged to use FSSIM data from 2 sample regions and will extrapolate the results to the other 2 sample regions. In further prototypes and in the final version, it will use data from the whole set of sample regions (between 20 and 30 NUTS2 regions) and extrapolate to the remaining regions of the EU.

Model inputs are

- Price-quantity combinations obtained from FSSIM simulations (per farm type in each sample region);
- Farm type resources such as land, water, etc. (per farm type for all EU regions);
- Biophysical soil and climate data (per farm type for all EU regions);
- Aggregation weights (per farm type for all EU regions).

2.5.4 Model outputs

The prototype 1 version of EXPAMOD produces price elasticities for marketable products in NUTS2 regions. Price elasticity is a percentage change in quantity if price increases by 1%. Plausible range of values is (-5.0; +5.0), but extremes are possible for crops at a very low or zero level in the initial situation. In subsequent versions of EXPAMOD, the output will be a price-quantity data set at NUTS 2 level which will be used to calibrate the market model to this response surface.

2.5.5 Data availability

For prototype 1, data beyond the 4 sample regions are generally not available. Ultimately, the following data are expected to be available:

- Price-quantity combinations obtained from FSSIM simulations (per farm type in NUTS2-region) – available from FSSIM results after performing parametric runs for a relevant set of price variations.
- Farm type resources such as land, water, etc. (per farm type in all EU regions) – available from farm typology.
- Biophysical soil and climate data (per farm type in all EU regions)
- Aggregation weights (per farm type in all EU regions) – at a later stage projected into the future based on the structural change module.

2.5.6 Model structure

FSSIM-MP will be run for a set of exogenously set prices to obtain a supply curve for each farm type in the sample region (for each policy scenario and for the baseline scenario(s) in case that the policy changes imply changes of supply response at the farm level). The left hand side of figure 12 illustrates these runs for farm types in the FSSIM regions.

These results, obtained only for a sample of regions and farm types, will be "expanded" (by EXPAMOD) to the other regions and farms that are out of the sample. From these aggregated results in terms of supply response, it will be possible to estimate the supply function parameters that will allow CAPRI calibration to FSSIM supply responses.

Currently the following sequence of steps is technically implemented in the EXPAMOD module:

1. (OLS) estimation of linear regressions for FSSIM-MP models in 2 out of 4 sample regions (2 out of 4 sample regions);
2. Calculation of statistics for the evaluation of regression results;
3. Extrapolation to "non-FSSIM.MP" regions (the other 2 sample regions);
4. Aggregation of supply response over farm types weighted by farm type representativity factors to NUTS2 regions;
5. Calculation of elasticities for:
    - FSSIM regions (2 out of 4 sample regions);
    - non-FSSIM regions (the other 2 sample regions);
6. Store elasticities for CAPRI.

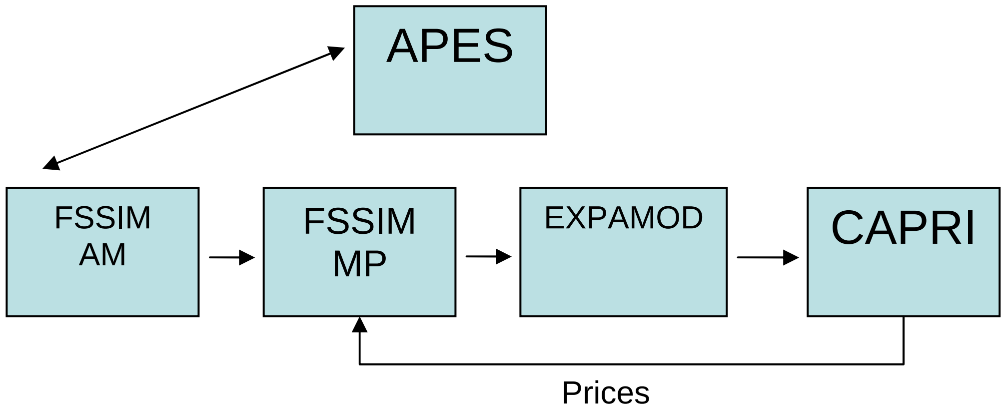


Figure 12 Links between FSSIM-EXPAMOD-CAPRI

2.5.7 Equations

Supply response function to be estimated for *FSSIM-regions*:

$$Quantity_{if} = \beta_0 + \beta_1 \cdot Price_i + \beta_2 \cdot Explanatory\ factors_{f(r)} \qquad (1)$$

Here *i* stands for the product, *f* stands for a farm type and *r* for region. Explanatory factors are *farm* resources (land, water) and *regional* biophysical variables (e.g. soil type, environmental zone).

In Prototype 1, a linear econometric model (parametric) is estimated as a more sophisticated approach cannot be implemented based on the data from 2 sample regions. Extrapolation and aggregation are performed for the other 2 sample regions, which also allows for a first test of the models capabilities in comparison to the available FSSIM results. Further elaborations (after Prototype 1) will implement non-parametric estimation techniques which implies that regression models for a certain farm type are estimated based on observations similar to this farm type and the corresponding biophysical zone; Further testing and potential flexibility in the choice of explanatory variables depending on goodness of fit will be investigated as well.

Extrapolation of supply responses to other (non-FSSIM) FADN regions is done according to the formulae:

$$\hat{Q}uantity_{if} = \beta_0 + \beta_1 \cdot \Pr ice_i + \beta_2 \cdot Explanatory\ \ factors_{f(r)} \qquad (2)$$

Here the estimates $\beta_0 - \beta_2$ are obtained from estimation, prices and explanatory factors are those for non-FSSIM regions.

Aggregation from FSSIM product list to CAPRI product list:

Quantity Wheat(CAPRI) = Quantity Wheat1(FSSIM) + Quantity Wheat2(FSSIM)+ Quantity Wheat3(FSSIM) (3)

The assumption here is that within one crop family the FSSIM optimal solution will select only crops which are close price substitutes. For more detailed information, see PD3.6.2, part 4: Procedure for aggregating products defined in the FSSIM list and disaggregating endogenous prices produced by CAPRI

Aggregation to NUTS2 regions is done using the weights (farm type representativity shares):

$$Quantity_{ir} = \sum\nolimits_f w_{fr} \cdot Quantity_{if}\ , \qquad (4)$$

where w_{fr} is the percentage of farm type *f* in region *r*

Elasticity of product *i* (in NUTS2 region) is calculated according to formulae:

$$E_j^{ir} = \beta_1 \cdot \frac{\Pr ice_j}{Quantity_{ir}}$$

Adjusting new product prices from CAPRI to FSSIM.
This is done with the percentage price change that is obtained in CAPRI keeping the original relative price differences that existed in the FSSIM application before the policy impact simulation:

New Price Wheat1(FSSIM) = Old Price Wheat1 (FSSIM) * (New Price Wheat(CAPRI)/Old Price Wheat(CAPRI))

New Price Wheat2(FSSIM) = Old Price Wheat2 (FSSIM) * (New Price Wheat(CAPRI)/OldPrice Wheat(CAPRI))

New Price Wheat3(FSSIM) = Old Price Wheat2 (FSSIM) * (New Price Wheat(CAPRI)/OldPrice Wheat(CAPRI))

2.6 Territorial models [15]

2.6.1 The basis – spatial allocation of farm types

Territorial models are at the end of the modelling chain of Seamless. Their inputs can be data coming from the SEAMLESS Data Base, or if necessary from other data sources, and outputs from other models (APES, CAPRI, and mainly FSSIM). Some of the results of these territorial models are quantitative, while others are purely qualitative. The quantitative results can be either used directly as indicators or be integrated in the "indicator calculator" for further treatment and analysis, but they do not serve as inputs to other models. The qualitative outputs are to be used in the post-treatment analysis, and/or in the negotiation phases. Being at the end of the modelling chain, the territorial models are implemented in the second phase, after delivery of the first prototype.

FSSIM will provide estimates for changes in acreage for various productions, as well as information on the choice of agronomic practices (fertilization, pesticide use, water use, etc.). By adding extra agricultural-based land uses such as forest, fallow, etc. in FSSIM, acreage shares are likely to be more comprehensive. As such, they will provide a first proxy for changes at the landscape level, although not spatially.

As the basic simulation unit for analysing policy impacts is the farm, it is necessary to give a spatial location to each farm that is simulated, in order to be able to localise the impacts. In principle, this information is not directly available. A procedure, using the available information concerning territorial attributes in space and the land use of each farm type will make possible to locate the results of FSSIM simulations in the space. This will allow performing two types of aggregation from the results at farm level: aggregate into administrative regions and aggregate into territorial regions.

By combining existing mapping data for soil, climate, and access to water with CAPRI DynaSpat, it appears that we may be able to probabilistically spatially place the farm types and hence the changes in land use coming from FSSIM results. If successful, this provides a highly promising avenue for spatial modelling of landscape impacts.

This statistical procedure will allow mapping into space the results obtained by FSSIM, as well as aggregating the type of information that corresponds to specific territorial units, as it is the case for most environmental issues. The environmental results obtained through APES-FSSIM simulations will allow estimating indicators at a territorial level, that is the one making sense for environmental and natural resources analysis.

Since the landscape configuration differs tremendously throughout Europe, a European-wide analysis is unrealistic and will probably not result in useful results. Therefore it is proposed in the frame of the SEAMLESS project, to do an analysis on the basis of selected case studies (from the Test Case regions), and perhaps for areas which we know already from other studies. Representative existing Landscape types and 'Ecoprofiles' will be selected and the results extrapolated to similar landscapes.

Within SEAMLESS the Territorial models are eventually to be applied at the European level. In a first step they will be applied for one specific region, as a proof-of-concept, and in a second stage they will be developed for the Test Case regions, and ultimately (within the Project) in order to be applicable for the biophysical regions included in the Sample of Administrative Regions. The future work, which goes beyond the scope of the present task, will concern all the biophysical regions defined in the EU. However, it will be necessary to

[15] The term "landscape models" has also been used for these models.

evaluate the level of error of this procedure (a large amount of information will be estimated), and the indicators that will be developed at the European level will be less precise (or will only express risks rather than real impact estimates) than the ones for the Sample Regions. It is quite possible that some of the indicators will not be calculated at the European level due the imperfection of the available information. In particular, it may be impossible to visualise the landscapes of some European regions (or sub-regions), due to a lack of available spatially explicit data. What is important is that the models and tools will be ready to be applied in any European situation, given the availability of appropriate information.

The spatial dimension concerning the models and the results obtained using the models are applied in two different steps:

FSSIM farm models are applied to virtual farms obtained through the farm typology.

These farms have a specific spatial location, as well in terms of administrative Regions and agri-environmental regions.

The results obtained out of the simulations done with FSSIM, can be located in space, not only in the territorial Regions, but in any kind of regional aggregation made out of the use of the Homogenous Spatial Mapping Units (HSMU). It is possible to aggregate results of FSSIM and it is also possible to split the results obtained by FSSIM simulations per land use, distributing them in the grid. This procedure may lead to results not necessarily consistent: with other sources of information a methodology has to be developed to solve this problem.

On one side, the spatial allocation of farms has to be done using relatively large territorial Regions (taking into account the limitations of the use of FADN data for statistical secrecy).

On the other side, the allocation per land use in a very detailed grid, after aggregation on small administrative regions may not be identical to the statistical information on land use available per administrative region.

Once the simulation results are spatially allocated, the detailed analyses related to visual attributes, and biological diversity / landscape heterogeneity can be performed using the tools developed for this task. This resulting "landscape" is virtual, not real, and exists only in a probabilistic sense. Note that the environmental impacts like nutrient leaching, erosion and pesticide residues will be assessed in APES-FSSIM, they just have to be spatially allocated.

2.6.2 Modelling visual landscape attributes

One important aspect of cultural landscapes concerns the perception by the public of landscape elements and changes in the landscape. This will be addressed here by developing a model component specifically targeted at visualising changes in the landscape.

Three-dimensional visualisation of the landscape provides means that are better understood than maps, especially for the general public. With such methods, visual changes of the landscape can be shown very impressively, which can allow for an intuitive assessment of the visual landscape quality. Static, web-based landscape visualisation tools have made considerable progress in recent years, such as for example Google Earth (http://earth.google.com/), covering the entire planet in 3D, or the French geographic service (http://www.geoportail.fr/) which is at present in 2D, but the 3D version is planned for autumn 2006. The visualisations are based on aerial (satellite) imagery, at a specific date, but are not dynamic. The challenge in the Seamless project is to view *future changes* in land use, according to scenarios. In Seamless, the landscape visualisation component should be launched at the end of a scenario simulation to allow for exploration of landscape changes. Visualisation could have a significant implication for the choice of effective land-use policy, and could be used as a basis for discussion and negotiation within the community. The pressures causing changes in landscape will come from the FSSIM model which will then be translated into changes in the spatial configuration of the landscape. The mapped results (environmental data such as land cover and land use) will be specific to some particular regions and should be available from the Seamless GIS. They will be used here to compute and visualise a 3D scene.

2.6.2.1 Model

For integration within Seamless-IF a new open-source module (named SLE –Seamless Landscape Explorer–) will be developed, based on the LGPL graphic engine Ogre (Object-oriented Graphics Rendering Engine).

The input data for such visualisation concerns terrain data (DEM), land cover (present land cover, and land cover resulting from scenario outputs), and a library of detailed textures and vegetation models. Various options will be explored, ranging from simple vegetation forms (using basic primitives such as cones, cylinders or spheres) to detailed botanically correct models (produced through the Bionatics commercial partner).

In a first stage, for building Prototype 1, we focused our work on the development of the 3D landscape visualisation tool best adapted for Seamless. The first developments of the visualisation component assume that the appropriate data is available. Further work will focus more in depth on the way "inappropriate data", such as can be expected for direct outputs from FSSIM, can be processed to serve as input to this visualisation tool. The perception of visual characteristics is part of the qualitative "indicators", and this is not an output of the model. The results are to be used in the post-treatment analysis, and/or in the negotiation phases with end-users and/or other stakeholders (WP2, WP6 WP7).

2.6.2.2 Specifications

The objective of the model component for assessing visual attributes is to compute a virtual landscape from accurate spatial data and to deliver a reasonable realistic representation of an existing landscape. Firstly, we need to build a real-time "*renderer*" that is powerful enough to handle large datasets of landscape terrain. Secondly, a matching virtual representation of an existing area has to be constructed, including vegetation and man-made structures.

To develop the "*renderer*", we have chosen to use a free (LGPL) graphic engine called Ogre (ObjectOriented Graphics Rendering Engine). It features the typical functionality of visualisation packages, such as scene graph and LOD/CLOD[16] functionality. It also supports a terrain renderer plugin. But the main advantages of this package are design quality, flexibility and clear documentation.

To develop a realistic virtual landscape visualisation platform, one does not only need the appropriate VR software, but also the content to build and populate the virtual landscape (e.g. appropriate tree models, terrain imagery, etc). The system we are building will be able to automatically generate suitable 3D content models from spatial 2D data. Seamless users should have the option to load a scenario and explore the results into a 2D map on the SeamlessIF graphic user interface and select an area of interest. After the selection is made, the mapping data should be launched to SLE which converts it to a suitable 3D model. The user will then be able to view the model in the custom built 3D landscape visualisation component.

[16] **LOD:** Level of detail; **CLOD:** Continuous level of detail

The 3D data conversion tool could work as an external module of the visualisation one and convert raster layers and shape files into a format the renderer can read. The Seamless GIS should include at least accurate data such as:

• LCC (land cover classification) map (e.g. Corine Landcover 2000),

• Landscape classification map (e.g. LANMAP)

• Digital elevation model (DEM) (e.g. GTOPO).

The availability of such data within the Seamless environment, and more importantly the degree of detail of these data, will be discussed with WP4.

The data will be used to generate terrain models including elevation data, roads and rivers. In case no mapping information exists (e.g. tree locations), algorithms will need to be implemented to generate the necessary additional data.

2.6.2.3 Use of the landscape visualisation module

Although some studies have been undertaken to quantitatively evaluate the scenic beauty of landscapes (see for example de Vries et al., 2006), the present work is not intended to produce indicators, but to produce images which can then be used as a negotiation tool. The system developed here is designed to import scenario-based landscapes that are generated with the help of the farm and environmental process models in Seamless (*i.e.* FSSIM. The results can then be used either individually by the policy-maker as a decision tool, or as a support for discussion and negotiation. For example, end-users could load a scenario that is based on a hypothetical development ten years into the future where the issue of soil salinity has been neglected, which could manifest itself in soil bleaching and dying vegetation. The users can use this platform to explore the landscapes, find environmental issues that are relevant, and discuss their opinions to form strategies for overcoming them. These aspects will be investigated in conjunction with WP2 (Indicators), WP6 (testing) and WP7 (participatory development).

2.6.2.4 Current Implementation

The challenge of our software SLE is, without proprietary tools or database preparation, to pull specific raw data (elevation, imagery and land cover) from a GIS, fuse this data in a procedural manner to enhance its apparent quality, and add vegetation objects to the scene. In a first stage, we make the assumption that land use and land cover classification data exist to describe the placement of vegetation in general terms. Where data are missing or incomplete, it will be necessary to develop a way to extrapolate missing land cover data. Our approach enhances the apparent quality of the given set of terrain elevation data and surface imagery, adds vegetation objects, and generates a plausible synthetic terrain environment where data is not accurate enough. The terrain will be created by SLE within the simulation application at run-time without the need of a precompiled terrain database.

Our goals are that the generation of synthetic terrains can be quick, automatic, based on raw source data. Realistic landscapes (*i.e.* representative of the actual and future environment) can be created automatically from the GIS and easy to edit by novice users. In order to design, construct, and test this concept, we need to:

1. Identify suitable algorithms for landscape visualisation. Compare and contrast these algorithms to determine suitability with respect to implementing an object placement algorithm and ability to increase the apparent resolution of source data using a noise function.
2. Implement the selected continuous level of detail algorithm to create a terrain mesh.

3. Drape imagery or splat texture over the mesh with noise-generated details added to improve the apparent quality of the source imagery if necessary
4. Identify which land cover classification datasets will be available in Seamless
5. Develop a vegetation object placement algorithm based on the processing of raw and derived land cover/landscape classification and elevation data. Placement must be automatic, repeatable, and guaranteed to put the same size and type of object in the same location and orientation given the same input data and configuration parameters (*e.g.* random number seeds).
6. Examine level of detail techniques to realistically represent vegetation objects with reduced geometric complexity. Use a level of detail (LOD) technique to construct vegetation objects and spatially organize these objects within our system to improve rendering performance
7. Develop a GUI to explore the virtual landscape.

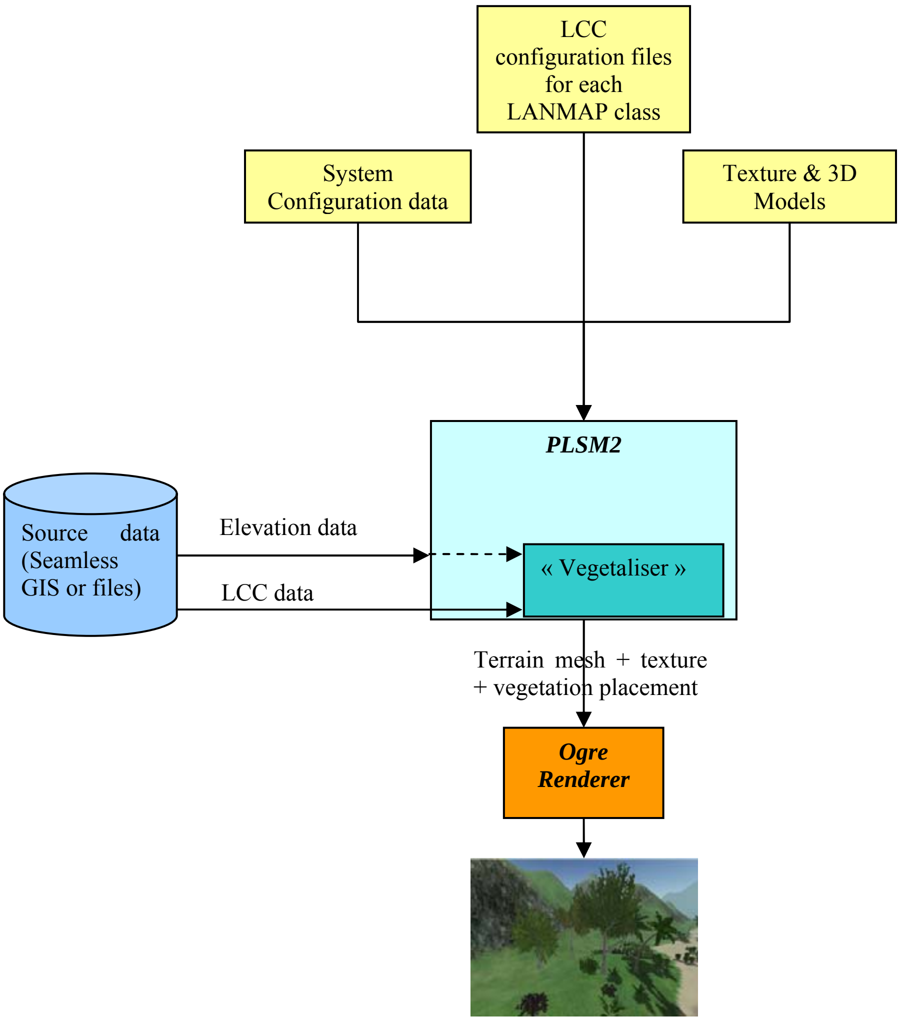


Figure 13 The Seamless Landscape Explorer diagram

It should be noted that large-scale vegetation placement is one of the most significant missing elements within today's simulation systems. It is an area where source code and academic papers are either not publicly available or simply do not exist. Fortunately, vegetation generation can be examined somewhat independently of other terrain visualisation components. While elevation data is needed to determine topological features and their influence on ecotope properties, this data (and its resulting mesh) need not be optimized with a continuous level of detail scheme nor noise enhanced. Similarly, surface textures, lighting and shadows, and other terrain visualisation elements can be studied and implemented independently from the vegetation placement algorithm.

SLE vegetation placement routines will reside within the paging landscape scene manager 2 (PLSM2) plug-in of the open source graphic engine, Ogre. In figure 13, we describe the flow of data through SLE. Source data (either stored locally or accessed from a central GIS), user-defined configuration data, and visual object models serve as inputs into the system. Output to the renderer (Ogre) within our application consists of a scene graph containing the terrain mesh and associated textures and the vegetation objects and their placement within the scene.

In figure 13, we depict the flow of data through SLE. Source data (either stored locally or accessed from a central GIS), userdefined configuration data, and visual object models serve as inputs into the system. Output to the renderer (Ogre) within our application consists of a scene graph containing the terrain mesh and associated textures and the vegetation objects and their placement within the scene. PLSM2 generates the terrain mesh, ground surface base texture, and detail textures. Elevation data, accessed from PLSM2 heightfield data, is used by the "vegetalizer" to create topographic image maps. From the source LCC image, pixels are extracted corresponding to each selected LCC type and saved as a separate "picked points" image which is smoothed to allow for overlapping LCC types. Finally, a probability map for each LCC type is generated based on its smoothed image map and modified by the results of a comparison between the topographic image maps and the LCC type's regime characteristics as specified in the LCC configuration file.

We also use the Geospatial Data Abstraction Layer (GDAL). GDAL is an open source translator library for handling (e.g. reading, writing, manipulating) a wide variety of raster and vector geospatial data formats. GDAL allows SLE to integrate many disparate datasets.

This work should provide the 3D visualisation tool needed for the environmental process model outputs intended for landscape exploration and for the visual characteristics and perceptions assessment. Regarding its integration into Seamless, there still remain some questions to clarify such as:

- Will we be able to have enough accurate data for elevation and land cover?
- How exactly will the changes in land cover be spatialized and be Corinelike classified

2.6.3 Modelling biological diversity

The modelling approach in SEAMLESS will be based on the results of the 5th framework project GREENVEINS. In this project relatively simple, independent relationships were found between Landscape structure (LS), Land Use Intensity (LUI) and the species richness of the major species groups herbaceous plants, birds and arthropods. The statistical analysis (Billeter et al. in prep.) showed for all of the mentioned major groups a specific relationship between species richness, the percentage of surface occupied by Green Veining (GV) and a dominant LUI factor: for herbaceous plants this was the percentage of surface area that was intensely fertilised, for birds the input of Nitrogen in kg/ha and for arthropods the crop diversity (for more information on the LUI measures used in GREENVEINS see Herzog 2006)). These relationships are basically suitable to predict the change in the number of species resulting from relative changes in farmland practices. In GREENVEINS a first attempt was made to develop a simple predictive model form the statistical models. However, a statistical model only predicts for the specific parameters it contains. When predictions from other intensity or diversity parameters are wanted, a conversion is necessary. This conversion should be based on existing correlations between intensity and diversity parameters in the normal agricultural landscape types. Information for this conversion can be obtained from literature and from projects like for instance MIRABEL, but will still need to be made within the scope of SEAMLESS.

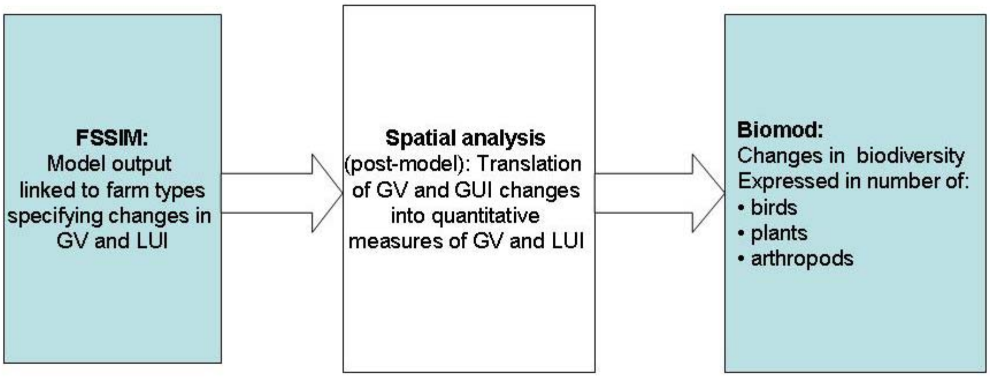


Figure 14 Data flow between FSSIM and Biomod

In SEAMLESS, we will use the GREENVEINS results to develop a simple, generally applicable regression model, Biomod. This model will predict the change in species numbers of birds, arthropods and herbaceous plants that will be caused by predicted changes in farmland practices from the FSSIM model. The changes in farming practice will need to be translated in intensity and diversity parameters. These parameters need to be translated into the quantitative GV and LUI parameters used as input in the Biomod, Biological diversity model (see figure 14). For this quantification the FSSIM output will need to be linked to spatial data sources such as topographic maps (and/or aerial photographs if topo maps are not available or too course) to establish the present and future GV measure. For the translation of the FSSIM output in a quantified LUI measure a link will also need to be established with space (e.g. present and future share of area intensively fertilised). The correct translation of the FSSIM output into spatial GV and LUI measures will be critical and will be a post model activity. The execution of case studies in order to develop the Biomod prototype will help to establish the right procedure and will provide an understanding of whether this can be simplified and standardized and incorporated in a component of SEAMLESS_IF.

For the development of the first prototype of Biomod we will develop a more general predictive model from the statistical models resulting from the GREENVEINS project and test and calibrate this model in Flevoland. More specifically we will select two landscape plots of 5*5 km in the Noordoostpolder, the oldest part of Flevoland.

2.7 Structural change model [17]

Economic incentives simultaneously lead to a redistribution of resources between farms (e.g. through the land market and/or by exits of farmers or switch from full to part-time farming) as well as to changes of the production pattern of the farm, a process termed structural change.

From an aggregate perspective, structural change may be understood and observed by the number of farms found in certain typologies at certain points in time. Firstly, the individual FADN records will be aggregated according to the chosen typology, so that afterwards the number of farms in each ex-post year and each cell is known along with moments for economic indicators such as farm profits. Seen as a time series, the cells describe a Markov process where changes in the number of farms reflect both a re-distribution of resources as land (shift between size classes and specialization) as well as changes in the number of farms (reflecting mainly the fact that the number of farmers is reducing over time).

The transition probabilities of the Markov process will be estimated depending on time, macro-economic variables (e.g. unemployment rate) and farm profits as explanatory variables. The farm profits link the estimation both to the market and the farm type layer: prices lead simultaneously to changes in the production program and in the profit of the individual farm presented in FSSIM and the changes in profits trigger an update of the aggregation weights to aggregate farm type results to the regional level. The update of the scale linkage weights will embed constraints which ensure that the available regional land base is not exceeded. That approach will lead to an endogenous module for “structural change”.

[17] For more specific and detailed information concerning this model, see the DOW.

2.8 Developing country models and impacts on countries outside the European Union[18]

There are two major reasons for assessing third country impacts of EU agricultural policies:

- EU agricultural policies may affect the international competitiveness of European agricultural producers, which simultaneously changes international trade flows and EU agricultural production;
- EU agricultural policies, through international trade, affect livelihoods and natural resource use in developing countries, which may conflict with, or support EU's development and environmental policies.

To analyze the international competitiveness of the European agricultural sector we will link two models: CAPRI (Common Agricultural Policy Regional Impact Analysis) and GTAP (Global trade Analysis Project). CAPRI and GTAP are both global models, but with different regional aggregation and sectoral coverage. The strong point of CAPRI is the detailed modelling of agricultural policies for the European agricultural sector and the differentiated representation of international *agricultural* trade and trade policies. The strong point of GTAP is its *economy-wide coverage*, accounting for all sectors. The objective of linking CAPRI and GTAP is therefore to arrive at a model structure which improves

1. the extent to which CAPRI accounts for feedback with other sectors and countries,

2. the extent to which GTAP accounts for EU agricultural policy and production.

Such a linking improves the performance of both models, contributing to the analysis of the impact of EU agricultural policies on the international competitiveness of European agriculture, international trade and production in the rest of the world. The way in which the two models will be linked is the first major challenge faced in this task.

The GTAP model is a global general economic model covering all sectors of the economy. What makes it unique within the SEAMLESS-IF collection of economic modelling tools is not its global character (that is also true for CAPRI), but its ability to model feedback between the agricultural sector and the general economy. This enables to analyse a wider set of scenarios and increases substantially the analytical power of SEAMLESS-IF. A crucial issue in GTAP integration is the link to the agricultural market model CAPRI which is dealt with first. Then, the link to developing country models is explained and the suggested sectoral and regional aggregation for GTAP motivated and described. Finally, some scenario issues in connection with GTAP are treated.

2.8.1 CAPRI-GTAP link

The specific problem of integrating CAPRI and GTAP into SEAMLESS-IF is the potential overlap in results generated by both models. A parallel application of both models would, for example, produce two sets of results on changes of agricultural output values in the EU. Those results would not coincide perfectly even if scenario assumptions, data, and structural parameters would be made consistent as far as possible, as the general model structure still differs. This approach would not only imply continuous and tedious work on model

---

[18] This section is an excerpt of PD 3.8.1. For more details, see PD 3.8.1. The linkages of these models with the other models in SEAMLESS is described in the last section of this deliverable.

specification for consistency purposes, but also require a sophisticated strategy in interpreting and communicating differences in indicator values generated by both models.

Therefore, we suggest a full link of both models, driven by the general vision to make best use of the comparative advantages of both models in the SEAMLESS context and to avoid generation of conflicting results.

2.8.2 Economic model link

The basic idea of the CAPRI-GTAP model link is to replace the agricultural sector within GTAP by results generated from the CAPRI model. The major motivation is that the CAPRI model has more functionality in representing specific agricultural policies such as premiums or quotas and has a considerably more disaggregated representation of the agricultural sector regarding product and regional differentiation. Consequently, the CAPRI model is generally more suitable to simulate impacts of changes in agricultural policies on the agricultural sector. In turn, the strength of the GTAP model is the incorporation of all other sectors of the economy. This allows for the specification of more general policy scenarios (trade liberalisation versus *agricultural* trade liberalisation), endogenous modelling of prices which are exogenous in CAPRI (e.g. input prices), and feedbacks between the economy and the agricultural sector.

The specific link foreseen is based on an iterative procedure similar to the connection of regional programming models and market module in CAPRI. In each iteration, CAPRI supplies the following entries for the Social Accounting Matrix (SAM) in GTAP:

- Output value of the primary agricultural sector (all products according to the Economic account of agriculture);
- Value of raw milk production (input value to the Dairy processing sector);
- Output value of dairy processing sector;
- Value of sugar beet production (input value to the sugar processing sector);
- Output value of sugar processing sector;
- Value of oilseed production (input value to the oilseed processing sector);
- Output value of oilseed processing sector.

The SAM will be rebalanced based on these values and all other values in other sectors from the last GTAP run (or data base in the first iteration) and then the GTAP model will be calibrated to this new SAM. GTAP then supplies price changes of intermediate inputs to CAPRI in the next step:

- Price change for machinery goods;
- Price change for energy goods;
- Price change for fertilizer.

This allows recalculating input cost for CAPRI production activities for the next simulation run. The iteration is terminated, when no significant output value changes (from CAPRI) and price changes (from GTAP) occur. By this procedure, we replace the supply behaviour of the GTAP sectors mentioned above by the supply behaviour of the CAPRI model for the agricultural sector and at the same time allowed for scenario or feedback driven impacts on input prices to enter the CAPRI model. The results of both models regarding the overlapping sectoral output values and prices are consistent.

2.8.3 Technical model link

Two versions of the GTAP model, one coded in the software package GEMPACK, the other in GAMS, are potentially available for SEAMLESS. The GEMPACK version is the one mainly used by GTAP modellers and would facilitate analysis of model results through decomposition tools like AnalyseGE and introduction of new model developments, but its use requires a license. The GAMS version has the advantage that GAMS is already used in SEAMLESS-IF. Existing wrappers could be used and the exchange between economic modellers within the project would be enhanced. However, both versions would allow for the link described above. Given the current expertise within SEAMLESS developing a link through GAMS is the preferred option which does not preclude translating the link between the two models to the GEMPACK version of GTAP in the future.

Exchange of data between the models would be done via GDX data files. This would suffice for the GAMS version of GTAP, but there also exists a tool able to translate GDX format into the specific GEMPACK format.

Both, GTAP and CAPRI, are suggested to use the same scenario editor to avoid duplication and ensure consistency regarding overlapping exogenous model shifters by design.

2.8.4 GTAP link to developing country models

This is a rather straightforward link based on the assumption, that the developing country (DC) considered is sufficiently small to not generate any relevant feedback for the world economic markets. In this case the information flow between GTAP and the developing country models is one-way. Price changes of relevant inputs would be delivered by the GTAP regional aggregate (see below) relevant for the specific DC (changes in agricultural output prices are delivered by the CAPRI model).

2.8.5 Sectoral aggregation in GTAP

The CAPRI-GTAP above allows for a rather lean sectoral differentiation in GTAP which almost complies with the general disclosure rules of the GTAP consortium, as only 11 sectors are distinguished. The suggested differentiation is the following:

- Agriculture (excluding Forestry and Fishing);
- Sugar processing;
- Dairy processing;
- Oilseed processing;
- Other food processing;
- Energy sector;
- Machinery sector;
- Chemical sector;
- Rest of manufacturing;
- Services;
- Trade services.

2.8.6 Regional aggregation in GTAP

The regional aggregation should reflect the differentiated information requirements to be expected for scenarios which also impact on the non-agricultural sectors. Those include competitiveness issues with major trading partners as well as impacts on transition, threshold, and developing countries. Furthermore, it is considered necessary to distinguish between the EU member states. Consequently, the suggested regional aggregation with 40 countries or country aggregates (Table 2) does not comply with the current GTAP disclosure rules:

Table 2 GTAP regions

1) Belgium/Luxemburg	15) Cyprus	29) NAFTA
2) Denmark	16) Czech Republic	30) Brazil
3) Germany	17) Estonia	31) Rest of America
4) Greece	18) Hungary	32) China
5) Spain	19) Latvia	33) India
6) France	20) Lithuania	34) Japan and East Asian Tigers
7) Ireland	21) Malta	35) Asian LDC
8) Italy	22) Poland	36) Former Sovjet Union
9) Netherlands	23) Slovak Republic	37) Australia and New Zealand
10) Austria	24) Slovenia	38) Rest of Sub-Saharan Africa
11) Portugal	25) Bulgaria	39) African LDC
12) Finland	26) Romania	40) Rest of world
13) Sweden	27) Croatia	
14) United Kingdom	28) Turkey	

2.8.7 Scenario issues

Like CAPRI, GTAP is not a projection tool. Consequently, the scenario definition needs to rely on projections of other institutions. The OECD outlook work provides developments for major economic variables which can be used for the calibration of the reference run (just as CAPRI uses agricultural outlook work of DG-Agri).

2.8.8 Linking GTAP with developing country models

The standard specification of GTAP is based on the economic structure in high-income countries. Changes will be made to the modelling of developing countries to account for the importance of tropical crops, weaker supply response resulting from market imperfections and imperfect labour markets. These changes to GTAP are the second modelling challenge faced in this task.

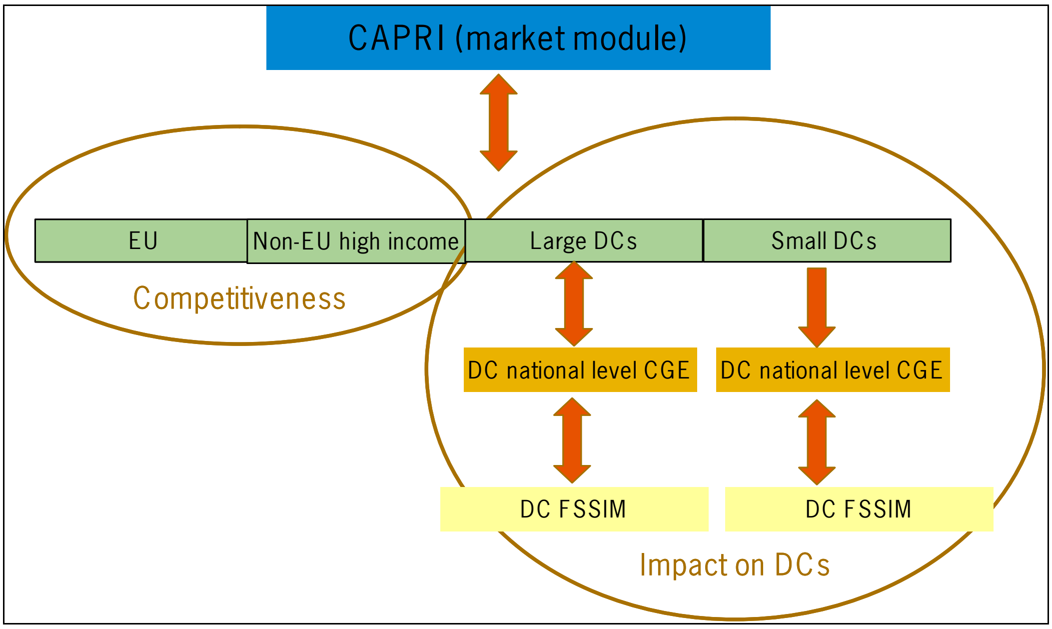


Figure 15 Link between the different models

To analyze the impact of EU agricultural activities and policies on poverty and sustainability in developing countries national level general equilibrium models will be developed for developing countries. Because of the importance of agriculture in developing countries and the need to analyze both rural and urban incomes an agricultural sector model like CAPRI does not suffice. In terms of the overall SEAMLESS project the national models serve the same role for developing countries as CAPRI does for European agriculture: transmitting changes at the international and national level to rural households. Development of a national level CGE model that allows analysis of household incomes, production and consumption decisions is the third modelling challenge faced in this task. The national level CGE models of developing countries will be used to analyze poverty and income distribution. The CGE model will be linked to FSSIM/type farm household models (figure 15) to analyze changes in agricultural production and sustainability indicators. Modifications will be needed to FSSIM models as used for Europe to address interactions between consumption, off-farm incomes and production decisions that result from market imperfections in developing countries. Adjusting the FSSIM framework for use in developing countries is the fourth modelling challenge faced in this task. This includes ‘tropicalizing’ the APES model so as to simulate the necessary production functions.

A successful modelling framework linking EU agricultural production to production in developing countries yields indicators to assess:

- the competitiveness of EU agriculture (border prices versus domestic prices, trade flows, value-added agri-food and non/agricultural goods, regional changes inn agricultural production),
- poverty (poverty headcount, employment and income distribution, caloric intake,
- rural/urban distribution of income and expenditures),
- environmental indicators (natural resource(soil)productivity, nutrient balances, toxicity index of herbicides and pesticides).

While the joint CAPRI-GTAP modelling will provide useful insights at the macroeconomic level and the global level, the picture has to be completed with country specific analysis that is able to zoom in on country- and sector specific issues in selected LDCs. A macro-micro linkage will be used to assess the distributional impacts on poverty in some developing countries. The outcomes of the CAPRI-GTAP modelling provide the macro and sectoral effects and these will be linked with a micro-economic household model for some developing countries (see, Bourguignon et al. 2002). The micro models use a similar methodology as the one used in FSSIM-MP for linkages between the farm and market level. The methods that will be used are the same as for linkages between FSSIM and CAPRI.

3 Model linkages and model's integration

3.1 Introduction

The models in SEAMLESS constitute a whole system, with different type of linkages. However, inside this system, there are sub-systems (models) that can work - assuming certain hypothesis - in an independent way. The advantage of this possibility consists essentially in allowing relatively easy updating and amelioration of the different models, or part of the models without affecting the other models.

This modular structure does not mean that using the models in an integrated manner or separately is indifferent to the results obtained.

Basic models in SEAMLESS:

- APES: can be used in a completely independent manner; it is not affected by other models in the system;
- FSSIM: can be used independently, instead of using APES results as part of its inputs; it could use other sources of information for externalities and for technical coefficients of alternative activities; however, it was designed to be used with APES;
- CAPRI: it existed before SEAMLESS; of course, it can work independently; however, the linkage with the APES-FSSIM sub-system implies a relevant change in CAPRI supply module;
- TERRITORIAL MODELS: they use outputs of APES-FSSIM sub-system, but potentially they can work with other information;
- GTAP: it existed before SEAMLESS; of course, it can work independently; nevertheless, the linkage with the APES-FSSIM-CAPRI sub-system implies changes in the information concerning the EU that is used in GTAP;
- DEVELOPING COUNTRIES MODELS: these models can work independently, but the linkage with the rest of SEAMLESS system provides inputs that allow analysing the effects of European Policies on developing countries.

Auxiliary models in SEAMLESS (not less important for being auxiliary)

- EXPAMOD: expanding results from the Sample of Regions and Farms to the EU
- SPACEMOD: Model that spatially allocates farms in a specific region.

The Spatial allocation model SPACEMOD works before all the others. Once the Sample of regions and farms has been decided on, the farms that will be simulated in the APES-FSSIM sub-system, have a geographical location.

3.2 APES-FSSIM sub-system

The APES-FSSIM sub-system will be described in this section. This sub-system will be adapted to be used for specific assessment at regional level. The principal limitation is that it will use always exogenous prices, because it is lacking a market model. The principal advantage is the quality of the information that will be used, i.e., it is higher as the requirements will be only for a sample of regions and farm types and not for the whole of the European Union. However, this sub-system is integrated with others that will provide this market dimension allowing dealing with the totality of the European Level. These aspects will be developed in the following sections.

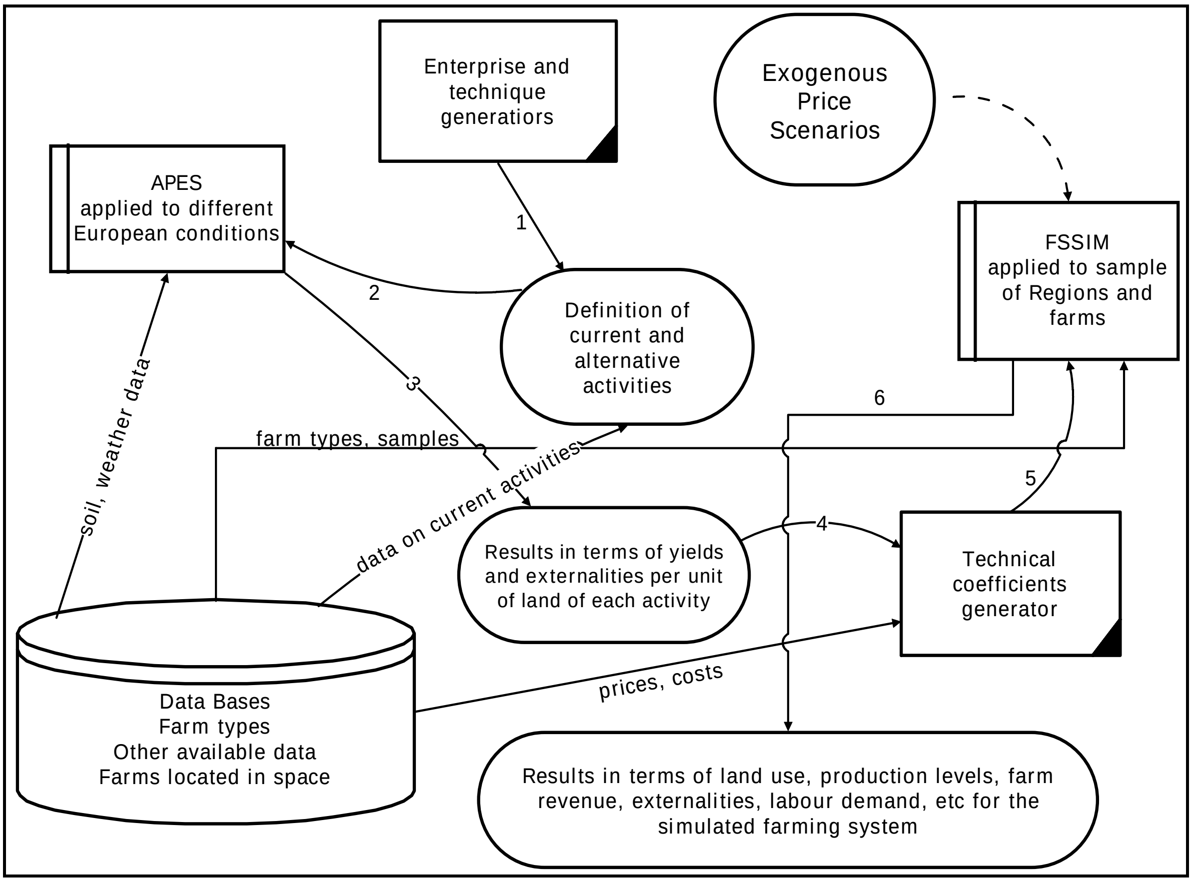


Figure 16 SUB-SYSTEM APES-FSSIM

Figure 16 shows the flow of information between FSSIM-AM, APES and FSSIM-MP. The initial point is in the FSSIM-AM. This module has four components: The survey for current activities, the Production Enterprise Generator (PEG), the Production Technique Generator (PTG) (Dogliotti et al, 2003) and the Technical Coefficient Generator (TCG) (Dogliotti et al, 2004). These three components are processing tools that prepare and adapt data inputs for APES and FSSIM-MP.

The PEG is a processing tool that defines the set of alternative agricultural enterprises that will be proposed to the system. Combined with the PTG, that generates different types of management, it is possible to define the Agricultural Activities. One example can be, in the case of crop production, a wheat-soybean rotation (production enterprise), using high level of irrigation and fertilisation (management technique, specified with the PTG). These two generators work with rules that allow to define a large number of activities, but a number that is strongly reduced compared with all the logical combinations that would appear crossing all the possibilities. This is true particularly in the case of crop production activities. These activities are defined as rotations, because it is the better way to analyse the interactions between production activities, externalities and impacts on natural resources.

In SEAMLESS models, two types of agricultural production activities are defined:

- Current activities are those observed in the base period. The base period is the one used as benchmark for calibrating the models. It will be an average of several recent years;
- Alternative activities are those that can appear in the future, out of changes in the policy environment and technological developments.

The number of these alternative activities could be extremely high, particularly in the case of crop production, if we apply a mechanical calculation, making all the possible combinations of crops and techniques in the rotations. Fortunately, not all combinations are feasible from a technical point of view. The role of these two generators, PEG and TCG is to generate activities, arriving to create a realistic number of activities using "filters". In any case, the number of alternative activities will be quite high. There is a potential problem in the use of this tool. On one side, it is necessary to arrive to a reduced number of activities and, at the same time, the filters should not eliminate activities that even being not the most suitable from an agronomic point of view, could be applied by farmers in certain policy environments. As our purpose is to build models representing as close as possible the real behaviour of farmers, we should avoid adopting a normative point of view while defining activities.

Data flows in the sub-system APES-FSSIM are the following:

- Arrows 1 - 2: The PEG and the TCG define activities to be simulated both in APES and in FSSIM. These activities are defined by their inputs and outputs. The input requirements for defining each agricultural activity need as well exogenous data (from databases, surveys, expert knowledge) or endogenous data (results of other tools or models);
- Arrow 3: Indicates the results obtained with APES in terms of yields and externalities;
- Arrow 4: Results obtained in APES as well as exogenous information from the databases (prices, costs, level of constraints) flow into the TCG;
- Arrow 5: The TCG, after receiving and processing this information prepares the input data in the format needed by FSSIM-MP and send these coefficients to FSSIM-MP;
- Arrow 6: Finally, the results of FSSIM-MP are obtained, as well for economic, social and environmental variables, for a certain type of policy scenario.

3.2.1 Principal outputs of this sub-system

A simulation of policy impacts using this sub-system provides different type of outputs. The principal ones are:

- Economic
    - Farm revenue;
    - Budgetary cost of the policy;

  - Level of supply of different products;
  - Changes on shadow prices of resources (land, water, labour).
- Social
  - Changes in labour demand for the simulated farming systems;
  - Changes on gender employment in the farms.
- Environmental
  - Changes on potential pollution levels (nitrates, chemicals) ;
  - Changes on potential soil erosion;
  - Impacts on other sources of soil degradation;
  - Impacts on water demand.

These outputs allow building appropriate indicators for policy assessment.

It is important to clarify that APES provides outputs concerning environmental issues, per unit of surface (e.g. one ha of a certain rotation) and FSSIM-MP, using this information, will provide the environmental outputs out of the solution of the farm bio-economic models, that give the weight of the different production activities that create the externalities.

3.2.2 Type of assessments by this subsystem

- At the level of this sub-system, the results are limited to the farms simulated in a Sample of regions, and are expressed only at the farm level.
- Principal limitations:
  - Lack of market dimension (exogenous prices, no feed-back from supply changes induced by the simulated policy impacts);
  - Lack of spatial dimension (some spatial issues, albeit very partial and imperfect, could be handled).

3.2.3 Data requirements for models of this sub-system

The principal data requirements, ordered per principal model are as following.

3.2.3.1 <u>APES</u>

- Weather
- Soil
- Management definitions
- Crop parameters

3.2.3.2 <u>FSSIM-MP</u>

- Yields (exogenous for current activities, endogenous for alternative activities);
- Prices (exogenous from past information, endogenous from CAPRI simulation results);
- Premiums (exogenous);

- Policy definitions (exogenous);
- Costs (exogenous);
- Inputs (exogenous);
- Equipment (exogenous);
- Labour requirements (exogenous);
- Land prices, land rent (exogenous);
- …..

3.3 APES-FSSIM-Territorial sub-system

In this section, we include the territorial models. The farm types are spatially allocated to agri-environmental zones within FADN regions. This spatial attributes of FSSIM results will allow to expand environmental impact results from models developed at a local scale and farm scale (APES and FSSIM) to larger scales (meso level – landscapes and small regions).

This sub-system is an extension of the APES-FSSIM one. The territorial models use the results obtained in the sub-system APES-FSSIM, representing and analysing those results at the level of agri-environmental zones. This expansion of results is done at two levels:

- For the biophysical regions included in the Sample of Administrative Regions;
- For all the agri-environmental zones defined in the EU.

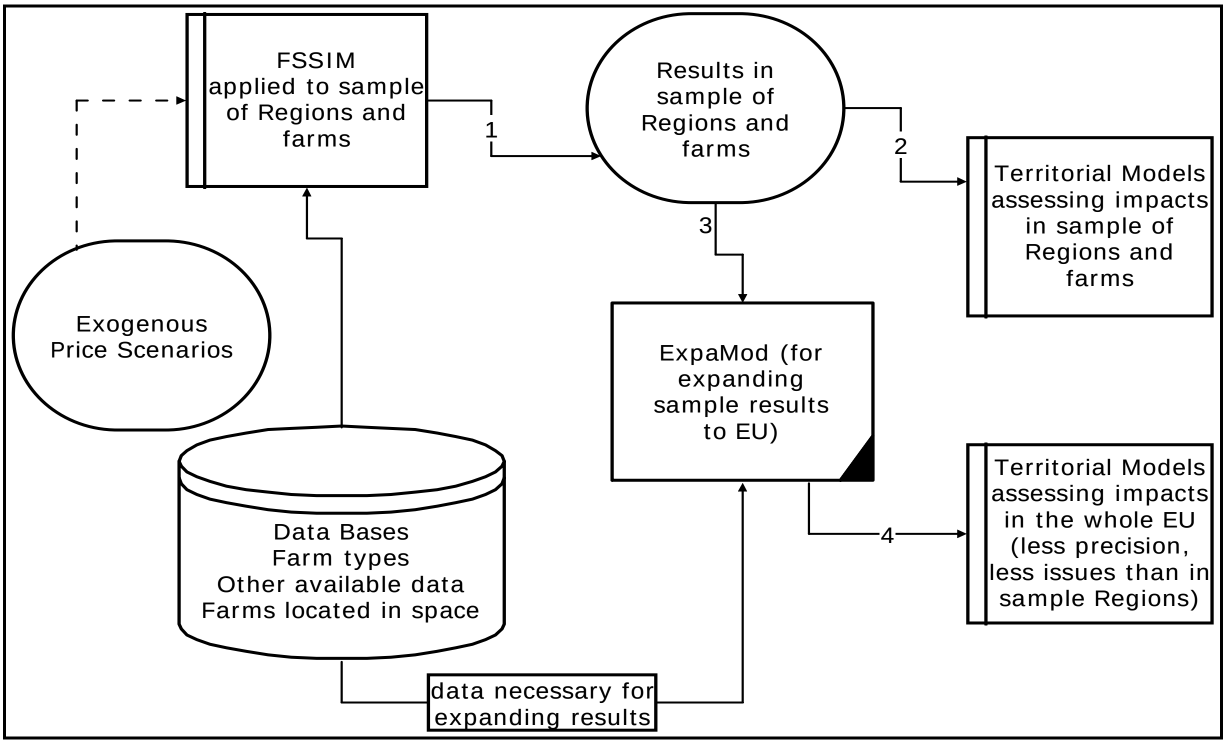


Figure 17 SUB-SYSTEM APES-FSSIM-TERRITORIAL MODELS

Data flows in the sub-system APES-FSSIM-TERRITORIAL (figure 17) are the following:

- Arrows 1 - 2: The results obtained by the APES-FSSIM sub-system are used as inputs by the Territorial Models for the sample (or Test Case) regions;
- Arrow 3: These results, expanded to the whole EU, will allow making an assessment at a European level. This is possible because we will have information on an important set of attributes for all farm types and regions in Europe. With this information that is available both in the Sample regions and in all the regions, it will

be possible to use statistical methods (described in EXPAMOD) for expanding these results.

- Arrow 4: The statistical models will be applied at the European level. However, it will be necessary to evaluate the level of error of this procedure (large part of the information will be estimated), and the indicators that will be developed at the European level will be less precise than the ones for the Sample Regions. It is well possible that a part of the indicators will not be calculated at the European level due the imperfection of the available information. What is important is that the models and tools will be ready to be applied in any European situation, given the availability of appropriate information.

3.3.1 Principal outputs of this sub-system

Part of the outputs of this sub-system are the same as those of the APES-FSSIM one, with the important difference that they are upscaled to regional levels. And this upscaling allows to obtain through the territorial models a different type of results.

A simulation of policy impacts using this sub-system provides different type of results.

3.3.1.1 Outputs similar to those of the APES-FSSIM sub-system, at regional levels

- Economic
    - Farm revenue;
    - Budgetary cost of the policy;
    - Level of supply of different products;
    - Changes on shadow prices of resources (land, water, labour).
- Social
    - Changes in labour demand for the simulated farming systems;
    - Changes on gender employment in the farms.
- Environmental
    - Changes on potential pollution levels (nitrates, chemicals) ;
    - Changes on potential soil erosion;
    - Impacts on other sources of soil degradation;
    - Impacts on water demand.

The difference with the previous sub-system is that these results will be aggregated at regional levels.

3.3.1.2 New outputs, out of the use of territorial models:

- Rough assessment of the spatial pattern or structure of landscapes;
- Relationship between pattern and process in landscapes;

- Relationship of human activity to landscape pattern, process and change; and
- Effect of scale and disturbance on the landscape:
    - o environmental issues at the landscape level;
    - o visual characteristics of landscapes and their perception;
    - o biodiversity and landscape heterogeneity;
    - o policy impacts on biodiversity.

3.3.2 Type of assessment by this subsystem

- At the level of this sub-system, the results will allow to assess policy impacts at farm and at regional level, as it appears observing the outputs of the sub-system.
- Principal limitation:
    - o Lack of market dimension (exogenous prices, no feed-back from supply changes induced by the simulated policy impacts).

3.3.3 Data requirements for models of this sub-system

The principal data requirements, in addition to the ones already defined, concern attributes needed for the territorial models. This is the same information that is also required for SPACEMOD that allows spatially allocating farms. Some of these new specific data are:

- Altitude;
- Weather conditions;
- Population;
- Infrastructure facilities;
- Education;
- …….

3.4 APES/FSSIM-CAPRI Linkages

In this section, we do not define a sub-system[19]. We present here the procedure of obtaining endogenous prices in CAPRI using FSSIM supply response behaviour.

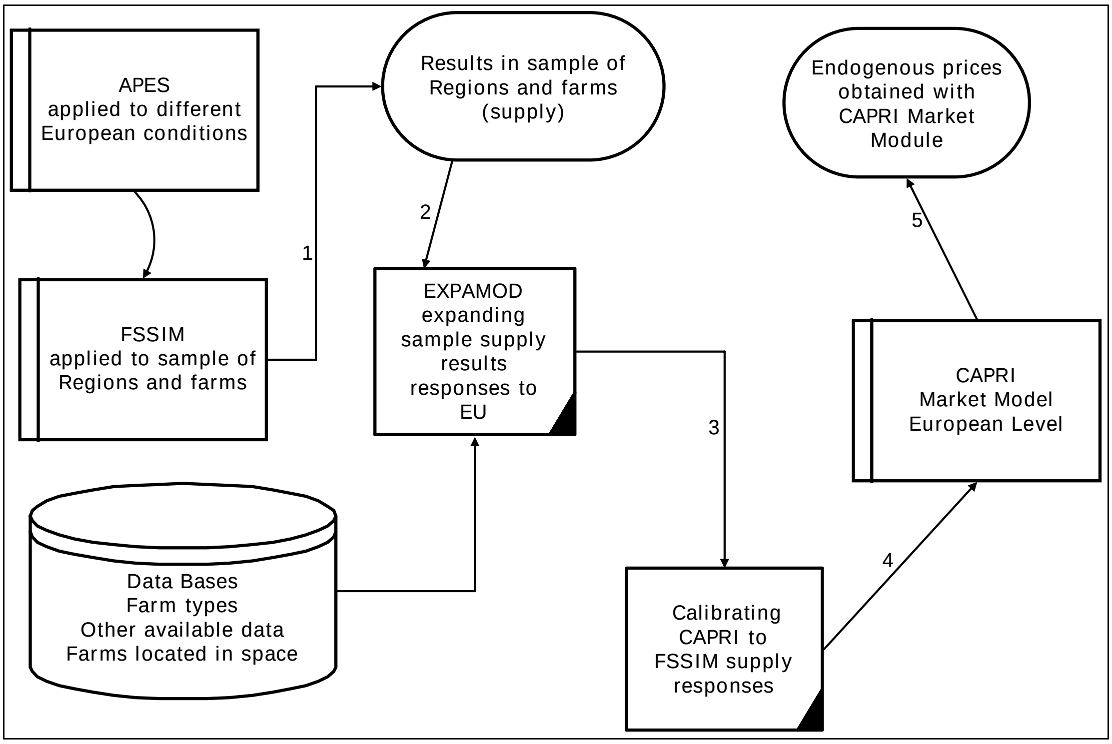


Figure 18 APES/FSSIM-CAPRI LINKAGES

CAPRI itself can work independently of the rest of SEAMLESS-IF. The integration of CAPRI in SEAMLESS is essentially done using the supply responses obtained through the use of FSSIM for calibrating the supply behaviour of CAPRI. This integration implies very important methodological changes concerning the CAPRI supply module. Once this supply response is introduced in CAPRI, it will be confronted - in the CAPRI market module - with the demand and this confrontation generates endogenous prices. These endogenous prices are a feedback to FSSIM-MP models. One iteration is necessary to perform this procedure. At the end, the results obtained by FSSIM-MP on a sample of farms and regions regarding other indicators then supply can be expanded to the entire EU. For this expansion, the EXPAMOD will be used. For obtaining (with CAPRI) the endogenous prices that are one of the consequences of the changes induced by the policy simulation, all the other results (environmental, social, landscape, etc) are not necessary.

Data flows in this diagram FSSIM-CAPRI (figure 18) are the following:

[19] Even if the relation APES-FSSIM is not shown here, it is implicit. APES and FSSIM are developed with the purpose of being used jointly, even though FSSIM can work with information from different sources. On the other hand, the Territorial Models, being applied out of the results of other models and without feedbacks, could not be applied if the necessity of their use is not apparent.

- Arrows 1 - 2: The supply results obtained by the APES-FSSIM sub-system applied to the Sample of regions and farms are used as inputs for the EXPAMOD that expands these results to the whole EU;
- Arrow 3: These results, expanded to the whole EU, are used by a model that calibrates CAPRI supply behaviour to FSSIM behaviour;
- Arrow 4: Supply results at the EU level are used in CAPRI Market Model for obtaining new endogenous prices;
- Arrow 5: The endogenous prices will be used for new simulations in FSSIM.

3.4.1 Principal output of this linkage

- endogenous prices, feedback to FSSIM for a second iteration.

3.5 CAPRI- APES/FSSIM-Territorial sub-system

This section presents a sub-system with all the models, except the agricultural employment, GTAP and the Developing Countries Models. The information flow begins with the endogenous prices defined by CAPRI market module.

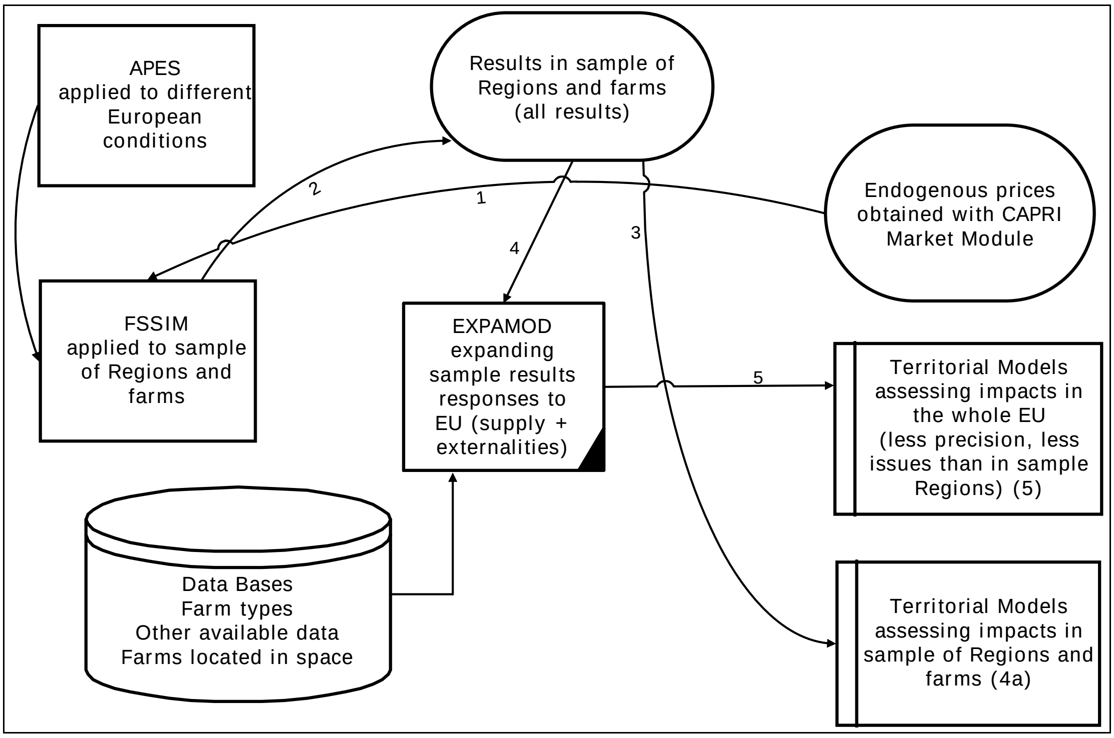


Figure 19 SUB-SYSTEM CAPRI-APES/FSSIM-TERRITORIAL

In figure 19 we can see a sub-system similar to the one represented in figure 17 (sub-system APES-FSSIM-TERRITORIAL), with a very important difference that is the feed-back from CAPRI that allows, after one iteration, to input endogenous prices in FSSIM and to do new simulations.

Data flows in this sub-system (figure 19) are the following:

- Arrow 1: The prices obtained in the CAPRI Market Module are used for new runs with FSSIM;
- Arrow 2: The results obtained by FSSIM runs will be exploited, as well concerning supply of products and externalities (the amounts of externalities was calculated by APES for all agricultural activities; FSSIM output gives the result of these externalities taking into account the amount of each activity in the optimal solutions);
- Arrow 3: The results are transferred to the Territorial models of the Sample Regions; for these regions, they will be exploited with relatively high detail;
- Arrows 4 and 5: The results are transferred to EXPAMOD which allows expanding them to the whole EU; however, it will be necessary to evaluate the level of error of this procedure (big part of the information will be estimated) and the indicators that will be developed at the European level, will be less precise than the ones for the

Sample Regions; it is quite possible that a part of the indicators will not be calculated at the European level due the imperfection of the available information; what is important is that the models and tools will be ready to be applied in any European situation, given the availability of appropriate information.

3.5.1 Principal outputs of this sub-system

The outputs of this sub-system are the same as those of the APES-FSSIM-TERRITORIAL one, with the important difference that the prices used come from the CAPRI Market Module. A simulation of policy impacts using this sub-system provides a different type of results

3.6 CAPRI-APES/FSSIM-Territorial-Structural change -Labour demand sub-system

In this sub-system, we add the model for estimating the structural change in the composition of farms as a consequence of simulated policies. The results obtained with this model may produce a feedback to FSSIM and through it to the whole sub-system. The econometric model estimating impacts on labour and gender is also considered here. In principle, the results obtained by this model do not produce a feedback on other models.

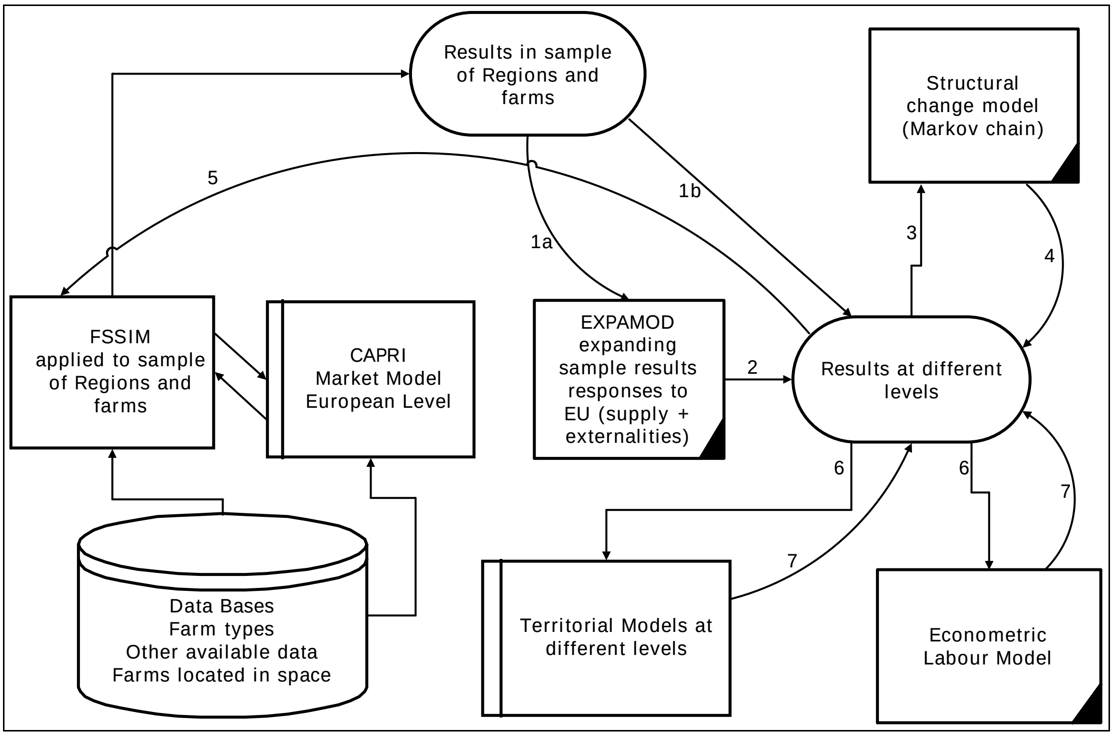


Figure 20 SUB-SYSTEM CAPRI-APES/FSSIM-TERRITORIAL-STRUCTURAL CHANGE-LABOUR DEMAND

Data flows in this sub-system (figure 20) are the following:

- Arrows 1a and 1b: The results obtain in the Sample of regions and farms can be used at different levels, directly and only in the Sample, and going through EXPAMOD for the entire EU, with the limitations already explained;
- Arrow 3: The results in terms of activity levels are inputs to the Structural Change Model that estimates the impacts in the composition and numbers of farm types;
- Arrow 4: The obtained results in terms of structural change, if important, produce a feedback;
- Arrow 5: Represents the feedback to the FSSIM-CAPRI;
- Arrow 6: The results become inputs for the Territorial Models;

- Arrow 7: Territorial Models produce the type of results already developed, with additional information concerning changes in farm structure and in labour demand.

3.6.1 Principal outputs

Outputs are similar to those of the CAPRI- GTAP-APES/FSSIM-TERRITORIAL sub-system with additional information concerning social issues. In the previous sub-system, labour demand could be assessed only from the results of FSSIM. In this sub-system, the results obtained through the structural change model and the econometric model for labour will allow to obtain a different and more complete type of assessment concerning social impacts of policies.

3.7 CAPRI-GTAP-Developing countries sub-system

The impact created by European Policies on the international market conditions will be reflected in outputs of CAPRI and GTAP. They consist essentially of price changes as well as trade regulations. These changes will affect developing countries. The models dealing with these countries have special characteristics, impossible to be generalized completely. However, the structure of these models will keep some of the basic SEAMLESS approaches, essentially with respect to the linkages between biophysical models, farm models and market models. This means that what appears in figure 21 as two boxes "Developing Countries Models" and "Territorial Models" corresponds to what is represented in figure 20 for the EU.

In this phase, we incorporate to the system two different models: GTAP, a general equilibrium global model and the developing countries models. The developing countries models will allow to assess impacts of European Policies on developing countries, but feedbacks are not going to be analyzed. GTAP will be used to analyze the impact of European Policies on a global level, with the possibility of providing assessment to all the countries considered in the GTAP system. At the same time, the currently envisaged link between CAPRI and GTAP allows to obtain consistent results for the agricultural sector and non-agricultural sectors at the European and global level.

This subsystem can work independently. Of course, the rich information provided by the other sub-systems will strongly ameliorate its performance.

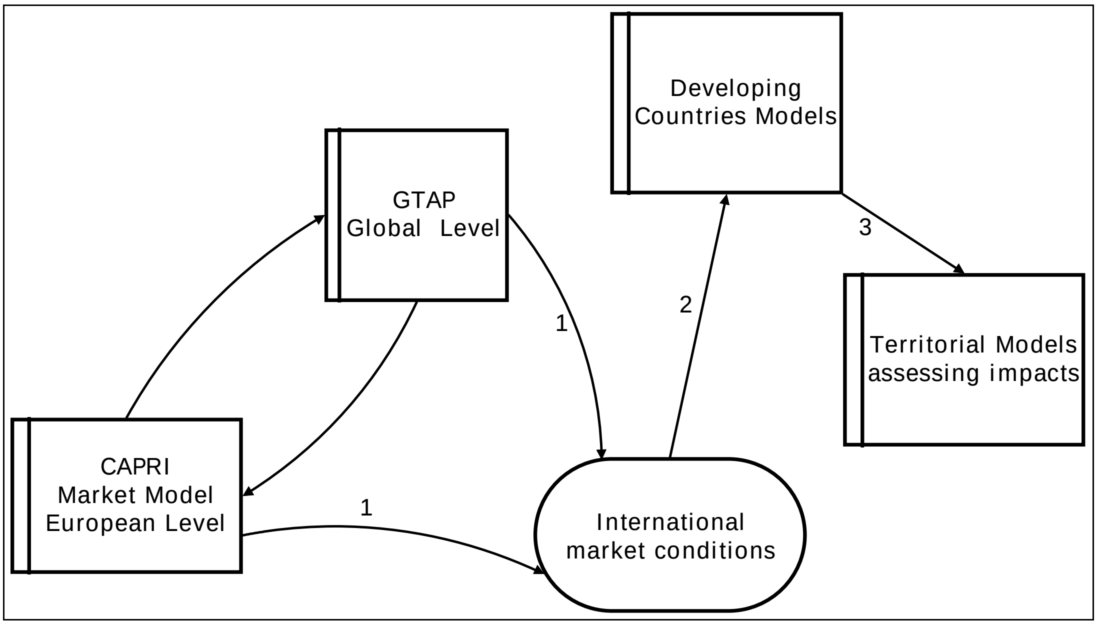


Figure 21 SUB-SYSTEM CAPRI-GTAP-DEVELOPING COUNTRIES sub-system

3.7.1 Principal outputs of this sub-system

Principal outputs are impacts of European policies on developing countries and, more general, on the global system. This sub-system gives the possibility to simulate impacts of policy changes outside the agricultural sector of the EU or in other countries on the European agricultural sector.

3.8 Conclusions about model linkages and integration

Linking models in the context of SEAMLESS has different meanings for each one of the sub-systems that have been described. The different situations can be summarized as follows:

- Outputs of one model are - after adapting that output - inputs for other model. This is the case of the sub-system APES-FSSIM. Results in terms of yields and externalities obtained through APES became technical coefficients of agricultural activities in FSSIM. This operation seems quite straightforward, but it implies complex procedures for adapting the receiving model to "digest" properly the outputs of the provider model. But the information does not change its nature. Yields are yields, kg of nitrate pollution are kg of nitrate pollution, etc.
- Outputs of one model play the role of "observations" that allow the development of an intermediate model (sometimes called a meta-model). Only the output of this meta-model can provide the appropriate input that can be "digested" by the receiving model. This is the case of the linkage FSSIM-CAPRI. The model we called EXPAMOD does this kind of work.

What is important to emphasize, is that model linkages in SEAMLESS have some specific characteristics. The different models may have their independent life, but the linkages and integration give an additional added value that, we hope, will be extremely important. Let us put an example concerning the linkage APES-FSSIM. In economic models, the usual way of representing externalities is to use a matrix of emission coefficients (this is the case for greenhouse gas emissions, or nitrate pollution). This means that a linear relationship between the level of the activity - wheat production - and the level of nitrate pollution is assumed to exist. In the case of industrial pollution, this linear relation can be accepted in many situations. It cannot be accepted, by no means, in agriculture production (Louhichi, 2001; Flichman, 2002, Flichman & Jacquet, 2003). In many cases, the relation between level of production and physical level of the externality may not only be non linear but sometimes non monotonous and non convex. This explains why taxes on Nitrogen Fertilizers cannot be efficient. Hence, the integration of biophysical and economic farm models allows to represent in a substantially ameliorated manner this type of relationships. This results in a better assessment concerning policies that are addressed towards a diminution of negative externalities. For doing this, it is necessary to use a primal approach for representing technology. Most economic models use a dual approach through costs, not having a direct representation of physical quantities of inputs. In our case, the sub-system APES-FSSIM can work in a very efficient way. But this approach would be extremely difficult to apply in a market model, working at the whole European scale. That is the reason why the use of an intermediate meta-model is necessary, allowing to "graft" responses obtained with FSSIM in the market model, CAPRI. And the linkage implies in this case a feedback of endogenous prices obtained in CAPRI, to FSSIM. This is why this linkage implies integration; it is more than just using outputs from A as inputs for B.

Most of the other linkages are simpler, with the outputs of a certain model becoming the inputs of another, but as a whole, the synergy of these models should produce something more than the sum of the results of the independent models.

References

Boussard J.M., 1971. Time horizon, Objective function, and Uncertainty in a multiperiod model of firm growth, American Journal of Agricultural Economics 53 (2), 467-477.

Cogneau D and A. Robilliard (2000). Growth, Distribution and Poverty in Madagascar: Learning from a Microsimulation Model in a General Equilibrium Framework. TMD

Dogliotti S., Rossing W., Van Ittersum, M. K., 2004. Systematic design and evaluation of crop rotations enhancing soil conservation, soil fertility and farm income: a case study for vegetable farms in South Uruguay. Agricultural Systems 80, 277-302.

Dogliotti S., Van Ittersum, M. K., Rossing W., 2003. ROTAT a tool for systematically generating crop rotations. European Journal of Agronomy 19, 239-250.

Falconer K., Hodge I., 2000. Using economic incentives for pesticide usage reductions: responsiveness to input taxation and agricultural systems. Agricultural Systems 63, 175-194.

Flichman, G. "Building Agro-Environmental Indicators by the Integration of Biophysical and Economic Models for Assessing Sustainability of Agricultural Trade Liberalisation". Workshop "Methodological Tools for Assessing the Sustainability Impact of EU's Economic Policies, with applications to Trade Liberalisation Policies, Brussels 7-8 November 2002. CEPII - European Commission, November 2002;

Flichman, G & Jacquet, F, "Le couplage des modèles agronomiques et économiques : intérêt pour l'analyse des politiques". Cahiers d'économie et sociologie rurales, nº 67 2003.

Flichman G., Garrido A., Ortega C. V., 1994. Environmental and land use issus in the Mediterranean basin: an economic perspective. 34th EAAE seminar, Spain: CIHEAM-IAM Saragossa, pp. 1-50.

Gaaff, B. Koole, J. Dekkers (2005). Future land-use change in the Netherlands: an analysis through a chain of models, forthcoming in : Agrarwirtschaft.

Hazell P.B.R., Norton R.D., 1986. Mathematical Programing for Economic Analysis in Agriculture. Macmillan Publishing Co, New York, 400p.

Hengsdijk, H., Van Ittersum, M. K., 2002. A goal-oriented approach to identify and engineer land use systems. Agricultural Systems 71, 231-247.

Junker F., Wieck C., Jansson T., Perez I., 2003. Policy description in the CAPRI model. Working paper, University of Bonn.

Kuhlman, T, F. van Tongeren, J. Helming, A. Tabeau,, D. Verhoog, R. Groenevoeld, A. Discussion Paper 61. IFPRI, Washngton D.C.

Kuiper, M. (2005). Village Modeling: A Chinese Recipe for Blending General Equilibrium and Household Modeling. PhD Thesis, Wageningen University.

Louhichi, K "La relation agriculture environnement, Thèse de Doctorat, Université Montpellier I (2001)

Louhichi K., Alary V., Grimaud P., 2004. A dynamic model to analyse the bio-technical and socio-economic interactions in the dairy farming systems on the Réunion Island. Animal Research 53, 1-19.

Pacini G.C., 2003. An environmental-economic framework to support multi-objective policy-making: a farming systems approach implemented for Tuscany. PhD, Wageningen University PhD thesis, Wageningen, 173p.

Reardon, T., K. Moulis, et al. (1998). Rural non-farm income in developing countries. In: *The state of food and agriculture, 1998*. FAO. Rome.

Schroeer L., 2004. EU-25 Trade Policy Monitoring CAP Reform 2003 – Deconstructing Decoupling 2004. Global Agriculture Information Network.

Tauer L.W., 1983. Target MOTAD. American journal of Agricultural Economics 65, 605 610.

Ten Berge, H.F.M., van Ittersum, M.K., Rossing, W.A.H., van de Ven, G.W.J., Schans, J., 2000. Farming options for The Netherlands explored by multi-objective modelling. European Journal of Agronomy 13, 263-277.

Van Ittersum, M.K., Rabbinge, R., 1997. Concepts in production ecology for analysis and quantification of agricultural input-output combinations. Field Crops Research 52, 197-208.

Van Meijl, Hans and Frank van Tongeren (2002), The Agenda 2000 CAP reform, world prices and URAA GATT-WTO export constraints, *European Review of Agricultural Economics*, Vol. 29 (4) (2002) pp. 445-470.

Van Tongeren, Frank (2005), Brazil and China: the real winners of the Doha round?, Entwicklung und Ländlicher Raum (Agriculture and development), Vol. 39, no. 5, pp 12-14.

ANNEX A Some ideas about relationships between Models and Indicators.

In principle, the general relation between models and indicators is that indicators are built out of model outputs, building indexes from single outputs or from different kinds of output combinations.

However, in some cases, there are situations in which indicators can be used as inputs in certain models. This is the situation when there is the possibility of obtaining indicators for one field used for one agricultural activity. This may be the case for obtaining information in the cases, in which the available models are not able to provide, or as an alternative source to model outputs. An example can be the externalities produced by chemicals, concerning human health or specific environments. An indicator of this type, calculated for a field, can be an input for FSSIM, a technical coefficient. The use of fuzzy logic procedures can be appropriate in these cases. When a solution is obtained for FSSIM, in which several activities are defined with a certain weight for the farm, it is possible to obtain an indicator for the farm, that will be the weighted average of the field indicators that where introduced in the model.

Observing the chain of models in SEAMLESS it is possible to clarify what model is providing what type of indicator.

- APES: does not provide outputs directly used in indicators; its outputs, concerning agricultural externalities, will be inputs for FSSIM;
- FSSIM: provides outputs that can be used for creating indicators; it gives a weight at farm level (from the importance of each agricultural activity in the solution) to the outputs generated by APES;
- Territorial Models: provide outputs that allow creating indicators at a meso-level; these indicators in some cases are more than a simple aggregation of farm indicators; this is the case concerning landscape and biodiversity;
- CAPRI: provides outputs for creating indicators at different levels, from farm to EU; however, for environmental indicators the Territorial Models will be the most appropriate for providing information in order to build these indicators;
- Models in Developing Countries: they will provide the same type of outputs for creating indicators, but adapted to the different characteristics of each type of situation; information related with poverty issues will be given special attention.

ANNEX B About Test Case Regions and sample Regions

In first versions of the project, Sample Regions did not exist. There were only Test Case Regions, where we should "test" the system. The procedures to aggregate from farm level to EU level were not defined. On one side there was FSSIM, an "agronomist's" farm simulator, FARM-INT, an "economist's farm model" and CAPRI. After long reflection and discussion, we realized that for linking farm models to CAPRI it was necessary to define a Sample of Regions, because it was materially impossible to use the farm models in all the Regions and farm types of the EU. Short time afterwards we realized that there was only one farm model and we abandoned FARM-INT.

On the other side, the Test Case Regions were already defined, but it was clear that if the farm model was going to be applied only on the Test Case Regions, it would be impossible to expand the results to the whole EU, because of lack of representativity.

Up to a certain point, the system will be applied entirely in the Sample Regions. A logical question appears then: do we still need Test Regions, what is the difference?

Even if this construction has been done while we were beginning the project and it is possible to find it not completely coherent, we think it is possible to find a rigorous manner to deal with it, defining precisely the role of the Sample of Regions (and farms) on one side, and the Test Case Regions (and farms) on the other.

B-1 For the Sample of Regions and Farms

B-1.1 What is the purpose of a Sample of Regions and Farms?

The basic unit for the integrated biophysical and economic analysis is the farm. The bio economic farm model, FSSIM, is a farm model, using very detailed information. Part of the FSSIM inputs comes from APES, the biophysical model. Even if theoretically it is possible to build models for all representative farms in Europe, practically it is completely impossible. However, we need to build a system able to be applied in the whole EU and especially a system able to assess policy impacts that affect market prices of agricultural products. The only way to assess price impacts is to use a model that simulates endogenous prices out of policy impacts. It was decided that this model, in SEAMLESS, is CAPRI. The dilemma that appears in a clear way is the following one. On one side we have farm models, able to integrate biophysical detailed information, but only applicable in a limited number of cases. On the other hand a market model that has a supply module completely different to FSSIM, particularly in what concerning the details in the definition of management and technology and concerning the treatment of externalities. We need to build an appropriate bridge to make these two different "animals" communicate. The relationship between the two models is described in this document (section 3.4). But for establishing that linkage, we need information coming from FSSIM models developed and run for an adequate Sample of Farms and Regions. The Regions chosen for the Test Cases cannot be considered as an adequate sample.

The first step is to define a Sample of Regions, at NUTS 2 level. The criteria for defining this sample should combine socio economic attributes of the Regions (level of regional income, agricultural income, weight of agriculture in the regional economy, employment in agriculture…) and biophysical attributes (latitude, temperature, dominant soils, precipitation…). Inside these Regions, agri-environmental regions will be defined.

The second step is to define the farms (virtual farms) that are simulated with FSSIM. For doing this, we use the farm typology defined in WP4. As each farm that will be simulated by FSSIM will have already a spatial definition, different aggregates of farm results can be obtained.

Once the results obtained in the Sample of Regions and farms, the results concerning supply of agricultural products are used to adapt CAPRI supply response (see section 2.3).

However, the most important use (not of small importance even if it is only one) of the Sample is to obtain a supply response at the EU level, out of policy impacts of different nature. For example, cross compliance policies related with detailed management, can be represented only in models such as FSSIM, and the supply impacts of these policies may be relevant at the European level and have strong implication on market prices.

Having FSSIM models in the Sample of Regions selected is the only way to make in the best possible rigorous way a linkage between FSSIM and CAPRI.

B-1.2 What are the data requirements for the Sample?

The data required are those necessary to APES and to FSSIM. Part of them can be obtained from European databases, but some of the data are available only from regional databases or from expert knowledge.

The essential information, in addition to what is currently available in EU databases is the following:

1. Resources availability:
   1.1. Available land with soil quality specification consistent with APES simulations;
   1.2. Available labour;
   1.3. Available water for irrigation;
   1.4. Equipment, capital goods available;
2. Policy parameters detailed per region:
   2.1. Premium levels;
   2.2. Specific agro environmental measures;
   2.3. Specific cross-compliance measures;
3. Regional input prices
4. Regional product prices
5. Production costs
6. Yields
7. Very importantly: farm management data

Besides these data, we need specific information for the calibration process. A period of 5 years can be used, at least a period corresponding to the longest crop rotation that will be used:

1. Crop pattern;
2. Production level;
3. Farm revenue.

B-1.3 Data concerning specific policy implementation

For the sample Regions, it is necessary to obtain with the maximum possible detail the information about policy implementation. Many policies have quite different regional implementation. A description of the policies in these Regions demand a specific activity: what is the level of premiums, what are the most important agro-environmental measures in the region, how conditionality is applied, etc.[20]

B-1.4 What are the simulations to perform?

In the Sample Regions, it will be possible to simulate specific regional policies in an imperfect manner, or to make the assumption they are the same. It is impossible to obtain good information on the type of implementation of policies, especially concerning environment, in all European Regions. As the principle of subsidiarity is applied, the way in which specific policies, defined only in their principles at the European level, are implemented at regional and local levels may be quite different. No centralized information on this issue exists. The regional regulations are not translated to a common language. This issue puts severe limits to the possibility of making a global European assessment on the impact of policies, fully applying SEAMLESS-IF. Any exercise of this type will have the limitation of severe simplifications concerning policy scenarios at regional levels.

B-1.5 What are the analyses of results?

Concerning always the Sample Regions, the analysis of results will be limited, taking into account the imperfection of the definition of policy scenarios. For the moment it is not possible to define precisely these limits, they should be fixed according to the quality of the information used in the models.

Considering the analysis of the expanded results to the whole EU, the environmental results have also to be cautiously analysed, with a previous assessment from natural scientists.

B-2 For the Test Case Regions and Farms

B-2.1 What is the purpose of the Test Case Regions and Farms?

It may appear to be not necessary anymore to take account of the existence of the Sample. But for the use of Test Case regions it is necessary and justified to take into account the limitations, we mentioned concerning the Sample Regions.

On one side, the Sample Regions/farms exercise allows the expansion (even limited) of results to the whole EU, on the other side, the level of information introduce the limits we mentioned in the previous pages.

[20] It is clear that these data are not all available in EU databases, but they may be available in regional databases. Data for management in physical units are the same for APES and for FSSIM

The Test Case regions/farms, having much more detailed information, allow a detailed analysis of the policy impacts that cannot be performed in the Sample Regions/farms, nor expanded to the whole EU.

Two objectives should be achieved using the Test Case regions/farms:

- Make a refined "proof of concept" of part of the system (not the market dimension)
- Produce detailed assessment on policy impacts at regional and local level

Sample Regions are going to be defined first at NUTS 2 level, and inside them, territorial regions will be also defined, but Test Case regions are smaller than NUTS 2 regions and do not need to coincide with administrative boundaries.

B-2.2 What are the data requirements for the Test Regions/farms?

The data are the same as those required for the Sample, but as much more effort will be used for these Regions, the quality will be higher. The typology of farms, being consistent with the general SEAMLESS typology, will be much more detailed, using local expertise and specific regional databases. The quality of the information on costs, quality of resources, management, will be superior to the one used in the Sample.

An important difference is also the specification of the policies that in the Test Cases will correspond in detail with the real regional implementation. This will not be possible to achieve at the same level in all the Sample Regions.

B-2.3 What are the simulations to perform?

In this case, it will be possible to simulate specific regional policy scenarios as well as technical change induced by policy changes (bottom-up). This type of simulations cannot be performed in the Sample, and also it cannot be expanded to the EU.

We have, in our opinion, to "sacrifice" the possibility of providing a feedback to the market model of supply impacts of regional policies that will be only simulated in the Test Regions/farms. In terms of what has been developed in this document, we will use a variant of the sub-system APES-FSSIM-TERRITORIAL MODELS (section 3.3).

In any case, when simulations will be performed at EU level, the Test Case regions/farms will be included and a more refined assessment on policy impacts will be done for these Regions, even if the results **in terms of supply** will not be used for the feedback procedure between FSSIM and CAPRI.