seamless

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

Protocols for spatial analysis to be implemented in the domain editor by WP5

Allocation of farm types spatially including the new Member states

Elbersen, B.S., Kempen, M., Andersen, E., Staritsky, I.G.

Partners involved: Alterra, Ubonn, UoC

Report no.: 51
February 2010
Ref: PD4.7.2-4.7.3
ISBN no.: 978-90-8585-594-1

ALTERRA WAGENINGEN UR
universität bonn
Skov & Landskab

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: Berien Elbersen Partner acronym: Alterra
E-mail: Berien.Elbersen@wur.nl

Name: Markus Kempen Partner acronym: UBonn
E-mail: markus.kempen@ilr.uni-bonn.de

Name: Erling Andersen Partner acronym: UoC
Address: Rolighedsvej 23, 1958 Frederiksberg C, Denmark
E-mail: eran@life.ku.dk

Name: Igor Staritsky Partner acronym: Alterra
E-mail: Igor.staritsky@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, contract no. 010036-2). Its content does not represent the official position of the European Commission and is entirely under the responsibility of the authors."

When citing this SEAMLESS report, please do so as:

Elbersen, B. et al., 2010. Protocols for spatial analysis to be implemented in the domain editor by WP5 - Allocation of farm types spatially including the new Member states, Report No.51, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 32 pp, ISBN no. 978-90-8585-594-1

Table of contents

Objective within the project

The objective of this report is to provide documentation on the final farm type allocation approach and results for EU-15. All allocation results described in this report are included in the final version of the database accessible at www.seamless.slnet.dk.

General Information

Task(s) and Activity code(s): 4.7

Input from (Task and Activity codes): 4.7

Output to (Task and Activity codes): WP5

Related milestones:

Executive summary

In this PD the approach to spatially allocating farm typology information to a specific environmental context is presented. At this moment the farm type information is only available at the administrative level of the FADN regions (HARM 1 regions). The spatial allocation approach adds a spatial dimension to all farm types making it possible to aggregate the types both to natural and to administrative regions. The spatially allocated farm types therefore facilitate the model linking, as they relate different scales to each other, just as different dimensions/domains (administrative, environmental, social). This spatial flexibility provides input data to the bio-economic/physical models in SEAMLESS (FSSIM and APES) in which a link has already been established between the socio-economic and farm management characteristics of farms and their environmental endowment (climat and soil attributes). Such input data also enable the linking of bio-economic/physical models (FSSIM), in which the farm in its bio-physical environment is central, to the market model (CAPRI), in which the market share of a specific farm type in a region is a crucial model input and output.

The spatial allocation of FADN farm information is a complicated process which involves several steps to allocate the FADN farm information and present the allocation results in a form that is in line with the disclosure rules for FADN data and that is useful as the basic input for the environmental and economic modelling in SEAMLESS.

The result of the allocation approach is a methodology that enables us to add a locational dimension to every individual farm contained in the FADN data base. This locational dimension is a reference to either a Homogenous Spatial Mapping Unit (HSMU) or a Farm Mapping Unit (FMU) (a cluster of HSMUs). Since HSMUs can be clustered to administrative or bio-physical entities the farms can also be grouped to these different spatial entities. For the first presentation of the allocation results we have chosen to group the farms to Agri-environmental zones (AEnZ).

The disaggregation approach for FADN farm information in SEAMLESS delivers good results in terms of validation. However, the usefulness of the allocated farm information as input for modelling in FSSIM (and APES) is not reported here.

PD4.7.2 presents farm results for the EU-15 regions and PD 4.7.3 presents the results for the new Member State (EU-10). The presented results in this combined PD are the final ones as validation results of former steps have been used to further improve the presented allocation procedure and results.

1 Introduction

The main objective of SEAMLESS is to develop an integrated framework (SEAMLESS-IF) to support ex-ante analysis of policies that enables analysis at the full range of scales. It therefore requires an integration of both spatial and statistical data and model inputs and outputs. In order to make this spatial integration of different data and modelling inputs and outputs possible WP4 develops procedures to combine data at different spatial levels. In this PD the approach to spatially allocating farm typology information to a specific environmental context is presented.

Because of the regional variation in climate, natural resources (soils, vegetation etc) and social structures, the increasing move towards de-centralisation of policy implementation and reorientation of the Common Agricultural Policy (CAP) to deliver more environmental benefits, there is an increasing need to appraise the multifunctional effects of agriculture at a range of scales. This integrated and multi-scale approach requires the use of farm information that is as spatially explicit as possible as this enables to relate market response behaviour to environmental performance of farms. In this context a methodology was developed to spatially allocate farm information to a specific environmental endowment (see also Seamless PD4.4.2 and PD 4.7.1). At this moment the EU farm information is only available at the administrative level of the regions (NUTS 1 or NUTS 2 regions). The spatial allocation approach adds a spatial dimension to all individual farms contained in the Farm Accountancy Data Network (FADN) making it possible to aggregate these both to natural and to administrative regions.

This spatial dimension is a reference to biophysical units with relatively homogenous conditions for farming, either a Homogenous Spatial Mapping Unit (HSMU) or a Farm Mapping Unit (FMU) (a cluster of HSMUs). Since HSMUs can be clustered to any administrative or bio-physical entity the farms can also be grouped to these different spatial entities. For the presentation of the allocation results we have chosen to group the farms to Agri-environmental zones (AEnZ).

In this report we are presenting the final methodology and the results for EU-15. The same allocation approach was also followed for the individual FADN farms in the new Member States. The slight difference in approach to the allocation and the results are presented in PD4.7.3.

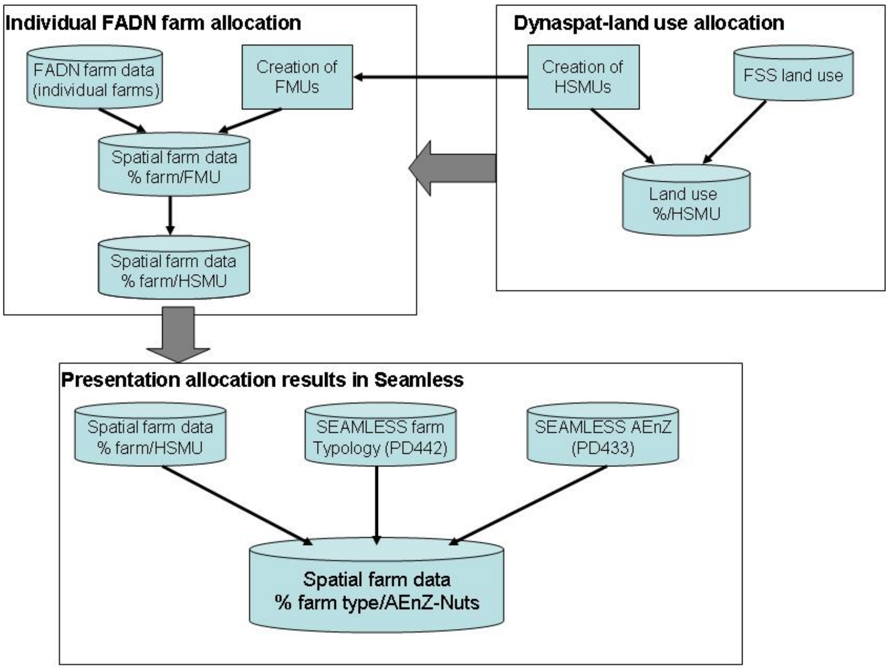
The final methodology of spatial allocation of FADN farm information is first discussed followed by a presentation of the results. All results of the allocation are included in the final Seamless database but in this report results are presented for a selection of regions. The report finishes with a conclusion and further work in this field beyong the scope of SEAMLESS.

2 Approach to spatial allocation of individual FADN farms

2.1 General approach

The spatial allocation of FADN farm information involves several steps (see Figure 2.1). The challenge is to identify the optimal allocation, but also present the allocation results in a form that is in line with the disclosure rules for FADN data and that is useful as the basic input for the environmental and economic modelling in SEAMLESS.

Figure 2.1 Workflow and inputs required to spatially allocate FADN farm information



In Figure 2.1 it becomes clear that first a land use or cropping zones database was developed by the University of Bonn and the Joint Research Centre (JRC, Climate Change Unit, Ispra) within the Dynaspat project. This database is an important basis for the spatial allocation of the farm information. Within the Dynaspat project the Homogeneous Spatial Mapping Units (HSMUs) have been created and land use information has been assigned to these units in a statistical allocation procedure.

The allocation of FADN farms then followed a similar statistical and econometric procedure as the land use allocation and the results were then aggregated into dimensions of a farm typology and linked to Agri-environmental zones before they became available in Seamless as model input.

2.2 The allocation of land use in Dynaspat as basic input to farm allocation approach

In order to use the farm type information as input data for the bio-economic models additional information on the location of the farm types was added to all farm types making it possible to aggregate the farm types both to natural and to administrative regions. This locational dimension is a reference to either a Homogenous Spatial Mapping Unit (HSMU) or a Farm Mapping Unit (FMU) (a cluster of HSMUs).The spatially allocated farm types facilitate the model linking of bio-economic/physical models (FSSIM), in which the farm in its bio-physical environment is central, to the market model (CAPRI), in which the market share of a specific farm type in a region is a crucial model input and output. Since HSMUs can be clustered to administrative or bio-physical entities the farms can also be grouped to these different spatial entities. For the presentation of the farm type information in the database we have chosen to group the farms to Agri-environmental zones (AEnZ). The spatial allocation of FADN farms is done using 2 inputs:

- The allocation of crops to the so-called homogenous spatial mapping units (HSMUs) elaborated in the Dynaspat project.[1]
- The allocation of farms to altitude zones and less favoured areas based on the information included in the FADN data.

The allocation builds on the methodology developed in the CAPRI-Dynaspat project and also uses the detailed land use maps resulting from this project as a main input. In the CAPRI-Dynaspat project a statistical approach for spatial allocation of crop production in the EU was developed. The result of this allocation, a detailed land use map, available for EU25, is the basic input for the allocation of the farm type information in SEAMLESS. The Dynaspat approach disaggregates the FSS crop information from the Nuts 1/2 regions to the much smaller Homogeneous Spatial Mapping Units (HSMUs) by developing allocation algorithms in a statistical procedure. This procedure combines a logit model with a Bayesian highest posterior density estimator. The HSMUs are defined by homogeneous production conditions rather than administrative boundaries.

For the spatial allocation of the FADN farm information the land use information and other attributes assigned to the HSMUs in the Dynaspat project are taken as the main input basis. The methodology for the farm allocation is very similar to that used for producing the land use allocation in Dynaspat. The main difference is however, that instead of using the HSMUs as the basic spatial entities to which farms are allocated a clustering of HSMUs, so-called Farm Mapping Units, are used. In addition, additional information on altitude and less favoured areas are also used to find the best locational match between the farms and the FMUs.

Input data used for the Dynaspat land use allocation

The input data used for the allocation of land use consists of Farm Structural Survey (FSS) data on crop shares per NUTS 2/3 region which was distributed to HSMUs. The allocation of this information was done with a whole range of other data sources which help to predict the

[1] See http://www.ilr1.uni-bonn.de/agpo/rsrch/dynaspat/dynaspat_e.htm

presence of certain crops in an HSMU. For this prediction, both attribute information at the level of an HSMU on climate, soil and topography is used and LUCAS point information on the occurance of crops (see also Table 2.1).

Table 2.1 Distribution data and additional data sources used for the allocation of land use

Distribution data	**Indicators used**	**Source**
I. Land use information at NUTS 1/2 level	30 different crops	Farm Structural Survey (FSS), EUROSTAT 2000

1

Attribute information to predict crop shares in a HSMU	**Indicators used**	**Source**
I. Soil attributes	Soil types: Set of soil codes (World Reference Base)	*Soil data:* *European Soil Database* V2.0 (European Commission, 2004). Set of soil codes: FAO, 1998 and Driessen, et al. 2001.
	Drainage/water management	
	Presence of stones	
II. Relief	Slope	Digital Elevation Model (CCM DEM, 250 meters). EC, JRC-IES, 2004).
	Elevation	
III. Climate	Annual rainfall	JRC-MARS meteodata. Interpolated metereological data, contained in the JRC-MARS-database (Boogaard et al. (2002) and Micale and Genovese (eds.) (2004)): all climate information in this database is provided for every 0.5 long-lat grid based on a 20 year average.
	Cumulative temperature sum	
IV. Land Cover information	11 agricultural classes were selected from the 44 land cover classes	CORINE land Cover (CLC) 2000, ETC-TE, 2000)
Point information	**Indicators used**	**Source**
V. Land Use information from the area frame survey LUCAS	Observed crop information in about 40,000 sampling points in EU15 on 38 crop classes	European Commission (2003), Lucas Survey

The selection of the soil, climate and topographic factors influencing presence of crops, builds on the suitability rules already developed in the MARS project (Boogaard et al. (2002) and Micale and Genovese (eds.) (2004)). Within MARS yield potentials for specific crops have been calculated using expert information and linking this to soil and climate data.

The soil map of Europe shows *soil mapping units* (SMU) which are soil landscapes formed by a characteristic pattern of distinct soil units. These soil units are the *soil typological units* (STU) to which the parameters of interest for the land use allocation are connected. These parameters include e.g. the genetic soil unit (soil name in the soil map legend), texture, slope, soil phase, parent material etc. A SMU may be composed of several STUs, the number of STUs in a SMU ranges from 1 to 6. The exact location of a STU within a SMU is not known, there is only an indication of the percentage distribution of STUs within a SMU. The same STU may occur in several SMUs, in different combinations with other STUs. To determine the specific soil attributes per SMU percentage estimates first need to be made of the agricultural land within a SMU and than of the dominant soil attributes present. To simplify the information and make it useful in the allocation procedure the attribute information of the dominant STU is assigned to the entire SMU, or, if no dominant STU can be identified (no >50% coverage), a percentage value is used.

One of the main soil indicators used in the allocation is the soil type. The soil types are defined in the World Reference Base (WRB) as part of the Soil Code. Each STU in the *European Soil Database* has been given a Soil Code defined and characterised in the WRB. In the relevant SMUs a total of 95 WRB Soil Codes are present. In order to use this soil type information in the allocation procedure, the number of WRB soil codes had to be reduced and clustered according to their suitability for certain crops. A clustering was therefore made in two steps. In the first step the soil grouping of Driessen et al (2001) was used who rearranged the 30 WRB soil groups into 10 so-called Sets, based on the dominant soil forming factors that determined the soil conditions. Since these resulting Sets were still very heterogeneous in terms of suitability for crops and yielding capacities they had to be subdivided again in a second step. In this second step new Sets werer created which were more homogeneous in terms of agronomic capacity notably rooting depth, organic matter, texture, drainage class, Available WaterHolding Capacity, presence of stones and slope. This further subdivision of the Driessen set was based on expert judgement. In order to maintain the logic of the distinction of the soil units on the soil map, we preferred that the Sets were defined by the highest hierarchical level in the WRB, *the Soil Reference Group* and then by the second level of the *Soil Units.*

The Land Cover information used comes from the CLC 2000 (European Topic Centre on Terrestrial Environment, 2000) and divides the land cover into 44 classes. It is produced by combining information from visual interpretation of satellite images and ancillary data (e.g. aerial photographs and topographic maps). For the allocation of the crop information to the HSMUs only the 11 agricultural Corine land cover classes are used (see Table 2.2). Since a 25 hectare area is the minimal mapping unit of CLC and several CLC classes are composite land cover classes, the CLC information can only be used as an additional indicator for predicting crop shares. CLC only gives information on the dominant land use within the minimal mapping unit (250 meters^2) and if the land use is very diverse within the mapping unit, CLC assigns heterogeneous land cover classes to the square (e.g. 'land principally occupied by agriculture with significant areas of natural vegetation'). The CLC therefore only provides additional information for specifying the allocation algorithms as every specific Corine land cover class can only be associated with a limited number of agricultural crops.

Table 2.2 The 11 Agricultural Corine Land Cover Classes used for the allocation of crops

Level 1	Level 2	Level 3
2. Agricultural areas	2.1 Arable land	2.1.1 non-irrigated arable land
		2.1.2 Permanently irrigated land
		2.1.3 rice fields
	2.2 permanent crops	2.2.1 Vineyards
		2.2.2 fruit trees and berry plantation
		2.2.3 olive groves
	2.3 Pastures	2.3.1 Pastures
	2.4 heterogeneous agricultural areas	2.4.1 annual cops associated with permanent crops
		2.4.2 complex cultivation patterns
		2.4.3 land principally occupied by agriculture with significant natural vegetation
		2.4.4 agro-forestry areas

The final and most important source of information on which the allocation is based comes from the LUCAS survey point information (EUROSTAT, 2000). The LUCAS survey was done in 2000 and 2003 in the EU15. In 2006 the sampling design will be modified and the survey is extended to most of the new Member States (MS). When this information becomes available is not clear yet. However, for this first allocation the LUCAS 2000 and 2003 data were used. These LUCAS survey data only cover the territory of the EU15. It is based on a 2-stage sampling method:

- Within a regular grid sample units, so-called Primary Sampling Units (PSU), with a size of 18*18 km are defined. For these whole grids information on land cover/land use is collected using up-to-date aerial photographs. This results in a selection of around 10,000 PSUs in the EU15
- Within every PSU 10 points, so-called Secondary Sample Units (SSU), are selected regularly distributed (in a rectangular of 1500*600 meters side length). For these points real field-observation-based crop information is collected on 38 agricultural crop classes.

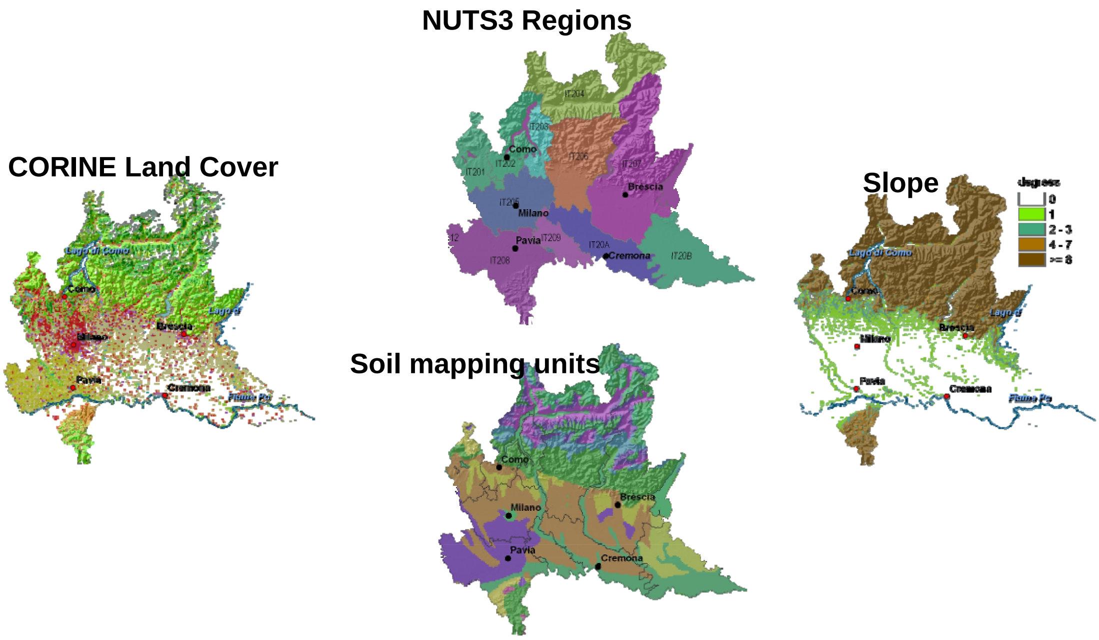
It is the point information in the SSUs that is the main basis for predicting crop shares in the HSMUs. However, since both the LUCAS point information and the CLC information are used as input for predicting crop shares, the influence of measurement errors which most strongly occur in CLC2000 need to be diminished. This is done by making a sub-selection of the SSUs and taking only those that were located more than 100 meters away from the borders of the CORINE classes (Gallego, 2002). This results in a selection of approximately 40,000 LUCAS sample points regularly spread over the whole EU 15 which are the basis for the allocation procedure. For the development of the allocation algorithms it is assumed that the chance for a similar crop pattern to occur in a place outside a LUCAS point depends on the distance to that LUCAS point, the larger the distance the lower the chance, but also on the level of similarity in soil, climate, relief and land cover characteristics.

The 38 agricultural classes found in LUCAS (36 crop land, 2 permanent grasslaclasses) were re-grouped according to the crops found in FSS. All other classes (artificial areas, woodland, etc.) are aggregated in a rest group.

Creation of HSMUs

Since most administrative regions (e.g. NUTS regions) are very diverse from an agro-physical perspective there is a need to split these regions up in small entities, so-called HSMUs with homogeneous bio-physical environment in which similar crop patterns can be assumed. HSMUs are an intersection of land cover (Corine LC 2000), relief (slope in 5 classes), Soil Mapping Units (so-called soil landscapes from the European soil map) and the Nuts 2/3 boundaries (depending on the size of the NUTS regions) (see Figure 2.2). Each HSMU has identical values for land cover class, slope class and Soil SET, other parameters (such as annual rainfall) may differ inside the HSMU. These HSMUs can be multiple polygons (open) which implies that one HSMU can be spread over different locations within a NUTS area. Attributes belonging to every HSMU are calculated (characteristics in terms of soil, climate, land cover, yielding capacity). These attributes are used to allocate the land uses to the HSMUs, but also the farms.

Figure 2.2 An HSMU is an intersection of land cover, slope, soil mapping units and Nuts boundaries



Allocation of land use in Dynaspat

A two step approach is then followed to predict the crop shares in every HSMU. These two steps were applied a couple of times in an iterative process in which the outcome of the following validation provides ideas for calibration of the allocation approach.

<u>Step 1:</u> The first step regresses cropping decisions in each HSMU on bio-physical factors (soil characteristics, climate, slope class and land cover), using results of the

LUCAS survey point information. This is done through the application of a spatial statistical techniques, a *Locally Weighted Logit model,* which results in normally distributed predictions of crop shares per HSMU. This approach results in the expression of expected shares of agricultural crops as probability density functions (pdf), i.e. in each HSMU mean and variance of the shares of 30 agricultural crops and one aggregated non-agricultural land use are estimated.

Step 2: The creation of an optimal distribution of the agricultural crops over the HSMUs according to total crop areas at Nuts 1/2 level provided by FSS. This optimalisation is based on a *Bayesian Highest Poterior Density* method and maximizes the posterior density of crop shares within the totals for the Nuts regions. It aims at creating an optimal consistency between scales, i.e. between the totals at Nuts 2 and HSMU levels.

More information on the allocation of land use and the results can be found on the Seamless public portal in report no. 19: http://www.seamless-ip.org/Reports/Report_19_PD4.7.1.pdf (Chapter 2 and especially Box 1).

2.3 The allocation of farm types to agri-environmental zones

The methodology for the farm allocation is very similar to that used for producing the land use allocation in Dynaspat. The result of the allocation is a locational dimension to every individual farm contained in the FADN data base. This locational dimension will exist, like for land use, of a reference to a Farm Mapping Unit (FMU) and a Homogenous Spatial Mapping Unit (HSMU) in which the farm is most likely to be located. The individual FADN farm can than be aggregated to any cluster of farms (SEAMLESS farm types) per cluster of HSMUs. This aggregated information can than be presented provided the FADN disclosure rules, which prescribe a minimal representation of at least 15 FADN sample farms, are not violated. However, the information on the share of the agricultural land managed by the different farm types can always be presented as this is not linked to the FADN variables as such, but are merely a calculated probability.

The individual FADN farm allocation procedure for farms in the EU-15 and in the EU10 was identical. However, there are two important differences:

1. the quality of the input data used was different. The data for the new MS is clearly of lower quality which is merely related to the most recent nature of data collection in these countries. Specific data quality problems for these new MS are therefore explained in a separate sub-section 2.5.
2. Contrary to the EU-15 allocation procedure,, no verification of allocation results against more detailed national data have been done for the new MS.. Further improvements that could come from such a verification have therefore not been available to imrove the allocation procedure in these countries.

Input data used for the farm allocation

The input data used for the allocation of farms within administrative regions are shown in Table 2.3. The individual FADN farms are distributed to Farm Mapping Units (FMUs), which are an aggregation of HSMUs. The use of FMUs is only necessary to simplify the allocation procedure and decreasing the computer calculation time. After the farms have been allocated to the FMUs they are also linked to HSMUs as for every FMU the link to the HSMU is maintained. The distribution uses predictions of the presence of a certain farm in the specific FMU and is based on a range of variables characterizing the farms available in the FADN database, which can be matched with the mapped attribute information for all FMUs. The whole range of other data sources used for the prediction the presence of certain farms in a FMU are given in Table 2.3. The main information source for making the prediction of the presence of a farm is the Dynaspat land use information (see former Section) but in addition to this other attribute information on location in altitude zone and Less Favoured Areas (LFAs) and yielding capacity are also used. LUCAS point information is not used directly, but this information is indirectly incorporated through the use of the Dynaspat land use information for predicting farm shares.

Table 2.3 Distribution data and additional data sources used for the allocation of farms

Distribution data	**Indicators used**	**Source**
Individual farm information at FADN region level	Per farm: - Cropping pattern (total area, area per crop) - Location in altitude zone - Location in LFA - Yield level of main crops	FADN, 2003, EC-DG-Agri
Attribute information to predict farm type shares in a FMU/HSMU	**Indicators used**	**Source**
Dynaspat land use	Shares of 30 different crops per HSMU/FMU	*CAPRI-DynaSpat project (EC-no. 501981)*
Relief	Elevation	Digital Elevation Model (CCM DEM, 250 meters). EC, JRC-IES, 2004).
Soil yielding capacity	Potential yields for main agricultural crops	*JRC-MARS-Yield Forecast System (MARS-CGMS). See:* *http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/crop_yield_forecasting_system.htm*
Less Favoured Areas	EU-LFA boundaries EU15	EC-JRC, LFA boundaries map

For the allocation approach the FADN farms will form the main input. This also implies that the farm information (or parts of it) needs to be disaggregated within the rather large FADN regions which are usually equivalent to NUTS1/2 regions.

For more information on the input data used for allocating the FADN farms See report no. 19: http://www.seamless-ip.org/Reports/Report_19_PD4.7.1.pdf (Chapter 3).

Allocation of FADN farm types

The methodology for the farm allocation is very similar to that used for producing the land use allocation in Dynaspat (see former Section). The main difference is however, that instead of using the HSMUs as the basic spatial entities to which farms are allocated a clustering of HSMUs, so-called Farm Mapping Units, are used. This clustering is necessary to reduce the complexity of the allocation procedure. The final allocated results are still linked back to the original HSMUs of which the FMUs are composed. For the presentation of the results farm allocation results have therefore first be linked to HSMUs and than aggregated to farm types in Agri-environmental zones.

The allocation of FADN farm information is done in steps:

1. Aggregation of HSMUs into FMUs
2. Create fixed distribution of FADN farms over dominant altitude and LFA and non-LFA zones
3. Identify optimal match between farm cropping patterns and potential yield levels and land use patterns in (a regional cluster of) FMUs by applying a *Bayesian Highest Poterior Density* method

Step 1: Definition of FMUs

In order to reduce the complexity of the allocation procedure a clustering of HSMUs was necessary. Whereas the HSMUs were designed to be homogenous regarding land use, the FMU should create continuous regions in which a fit with the UAA of a farm can be made and a link can be established between farms yield levels and soil conditions determining potential yields. Therefore the soil mapping units were chosen as the main attribute according to which the HSMUs were clustered into FMUs. Since location in altitude zones and LFA are the other robust statistical information on which the location predictions are made the dominant altitude class and presence of LFA was the second clustering layer used. Finally it was also ensured that clusters into FMUs could only be created within a Nuts2/3 boundary.

Step 2: Consistent allocation of farms in altitude zones and LFAs

The variable of interest is the probability of finding a certain farm in a specific FMU $p_{f,fmu}$. As a single farm in the FADN sample represents many similar farms this probability can also be understood as the share of these farms being allocated in a specific FMU. From the FADN statistics it can be exactly derived which farms are located in a certain altitude zone and whether located in a LFA. This information is taken as fixed and given, i.e. if the FADN farm and the FMU do not belong to the same combination of LFA and altitude zone the probability $p_{f,fmu}$ of finding this farm in this FMU is fixed to zero. An obvious constraint in the allocation procedure is that the probabilities for each farm must add up to unity:

$$\sum_{fmu} p_{f,fmu} = 1$$

Another natural constraint refers to the agricultural area of farms and FMUs

$$\sum_{f} p_{f,fmu} A_f = A_{fmu}$$

Where A_f is the UAA represented by a FADN farm and A_{fmu} the agricultural area in a FMU respectively. If the area derived from different sources is not fully consistent an adjustment factor is calculated to enforce consistency. This consistency ensures that the number of farms and their related farm area in different altitude zones and LFA and non-LFA areas fits exactly to the available agricultural areas in altitude and LFA zones located in the cluster of FMUs making up the administrative region for which the FADN data are given.

$$\sum_{f} p_{f,fmu} A_f = adj\ A_{fmu}$$

Step 3: Identify optimal match between cropping patterns and yield levels on farms and of FMUs

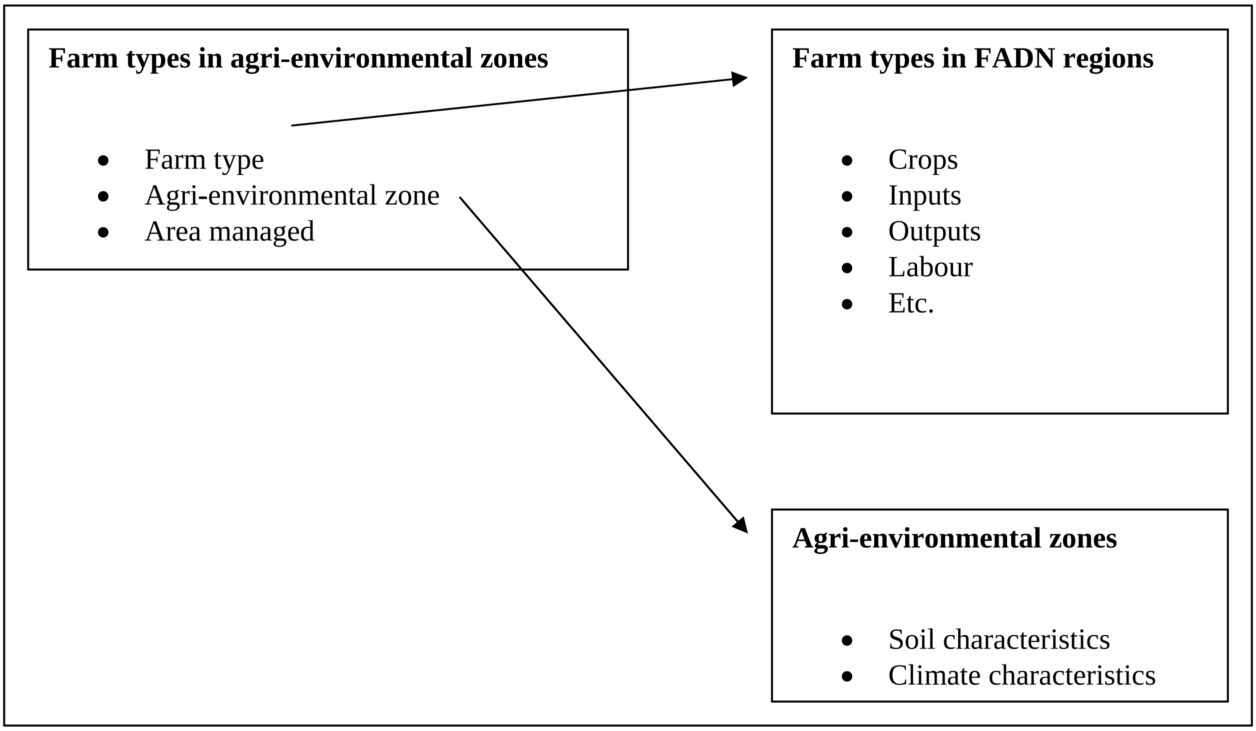
The allocation of farms in this step is based on yields and land use decisions. Whereas in the case of yield the findings on a single farm should be similar to those in a FMU, in the case of land use information this could be interpreted in different ways. On the one hand it could be assumed that farms in a FMU look alike and therefore the predicted land use in a region should be similar to that of farm level. On the other hand, a region could also be managed by different specialized farms. In this case the aggregated land use of all farms allocated in a region should be close to the predictions on this region. This formulation is in line with the predicted mean and variance used in the Dynaspat project for land use allocation and will be used in the following procedure to allocate the farm to the FMUs.

For more detailed information on the input data used for allocating the FADN farms See report no. 19: http://www.seamless-ip.org/Reports/Report_19_PD4.7.1.pdf (Chapter 3). In this report it is also explained what validation/calibration tests were done and how these were used the further improve the allocation procedure in an iterative process.

2.4 The presentation of allocation results: linkages between farm type data and biophysical data

The relations and descriptions of farm types and agri-environmental zones in the SEAMLESS database are summarised in Figure 2.3. The information for farm types in agri-environmental zones includes only one variable: The area managed. However, for the farm types present in FADN regions a whole range of variables is available in the database. The area within one agri-environmental zone is managed by several farm types and one farm type in most cases will manage land in different agri-environmental zones. This information on the distribution of farm types within agri-environmental zones is linked to one agri-environmental zone with a specific description of soil and climate characteristics. The relation to the more detailed descriptions of the farm types is more complicated. This information comes from the FADN data that have been processed to the SEAMLESS farm typology and are included at the level of the FADN regions, but of course only for farm types based on more than 15 sample farms. One description of a farm type in the FADN regions represents this specific farm type wherever it occurs in an agri-environmental zone within this FADN region. Presently, these links between farm types in the agri-environmental zones and at the FADN region level are only included in the database for farm types with more the 15 sample farms at FADN region level.

Figure 2.3 Illustration of the links and descriptive variables available for farm types and agri-environmental zones in the SEAMLESS database.



The relationships between the different data and the specific variables included can be studied in the final Seamless database - especially in the diagrams ‘farm type information’ and ‘Biophysical information’. The key tables to look at to explore the relations directly in the database or in the html files mentioned above are:

- data_agrienviromentalzone
- data_representativefarm
- data_representativefarminagroenvregion

2.5 Data availabilty in EU10

The core procedure applied to allocate FADN farms in space is unique for the entire EU25. Main differences comparing EU15 and EU10 stem from the availability of raw data for those regions. Generally data gaps can be filled by extrapolation from the existing observations. Some indicators used in the overall approach suffered from missing data in some or all EU10 countries. Specific data problems and their handling are the following:

- *Land Use information from the area frame survey LUCAS*

LUCAS point observations on crops are available for the entire EU15 in the years 2000 and 2003 whereas only some EU10 countries report data for 2003. This effects quality of the Dynaspat land use predictions at HSMU level in two different ways. First the higher number of observations done twice in time should enhance the robustness and accuracy of the locally weighted logit estimation since more observations are available within a relatively short distance from the region of interest. Second, the spatial econometrics underlying the locally weighted logit estimation requires to some extending continuous observations in space. As LUCAS observations are missing in some EU10 countries the locally weighted logit had to be replaced by (ordinary) logit estimation. It is hardly impossible to quantify to which extend these limitations of data reduce the quality of the overall land use map. Preliminary validation however indicates that the average mis-classification of land in EU10 countries is not significantly worse than in EU15.

- *Soil yielding capacity*

The JRC-MARS yield forecasts were only available for EU15 countries. Based on those combinations of yield forecasts and soil characteristics a Meta model was estimated to extrapolate yield forecasts to the EU10 countries. These forecasts are used to create spatial heterogeneity of yields whereas more aggregate regional yields from FSS and the CAPRI data base are used in a reconciliation step.

- *FADN data*

FADN records are available in EU15 from 1990 to 2004. In EU10 only 2004 records are available. This difference however is of minor importance since only records of specific years are used in the allocation procedure so far.

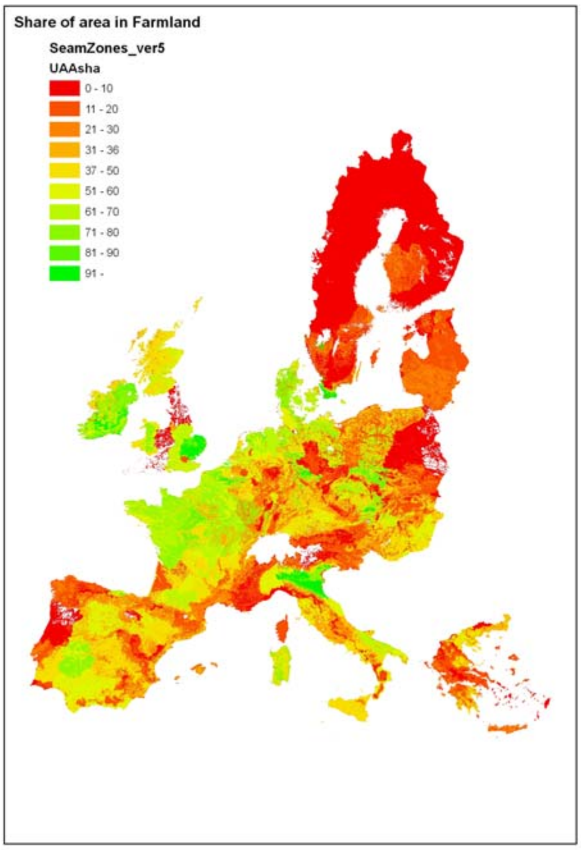
3 Allocation results

In the following sub-sections selected results of the allocation of the farm types are presented. The results are presented for the total allocated farm area and for the allocation of land to different farm types according to the different dimension of the farm typology used in SEAMLESS. These dimensions are: Size, intensity and specialisation (Andersen et al., 2006).

3.1 Allocated farm area

The results of the allocation of farm types to agri-environmental zones in terms of the share of the agri-environmental zones managed as farmland are shown in Map 3.1. Not surprisingly mountainous and upland regions and regions with a high share of forest have low shares of the area of the agri-environmental zones in farmland. The results for most regions show an internal differentiation. This means that the approach for the spatial allocation of the farm types, and thus the farmland, provides results that do show variation in the allocated farmland in different biophysical endowments.

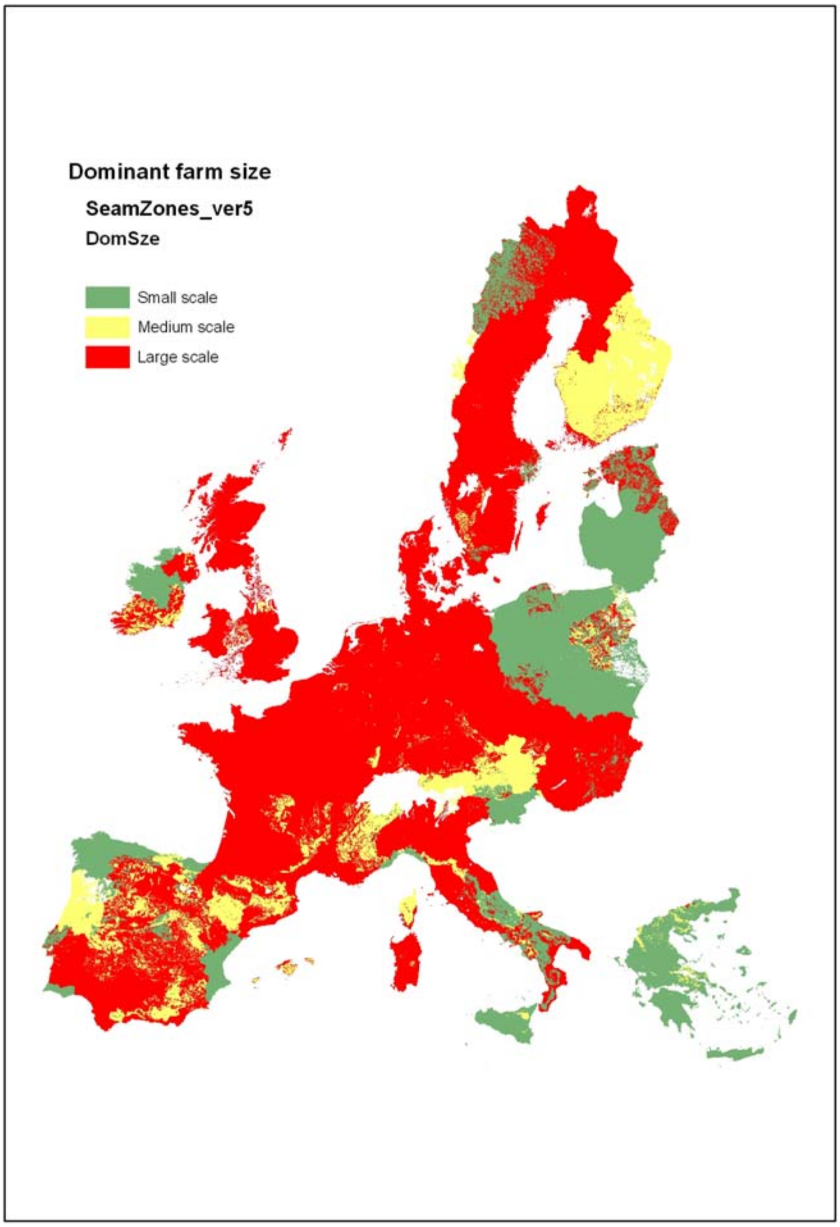
Map 3.1 The share of the total area of the agri-environmentalzones in farmland.



3.2 Allocation of farm types according to farm size

In Map 3.2 the results of the allocation of farm types to agri-environmental zones are shown in relation to the size dimension of the allocated farm types. More specifically the dominating farm size in terms of area managed is shown for each agri-environmental zone.

Map 3.2 *Dominant farm size of the agri-environmentalzones. Measures as the share of the agricultural area managed by the different types of farms according to farm size*



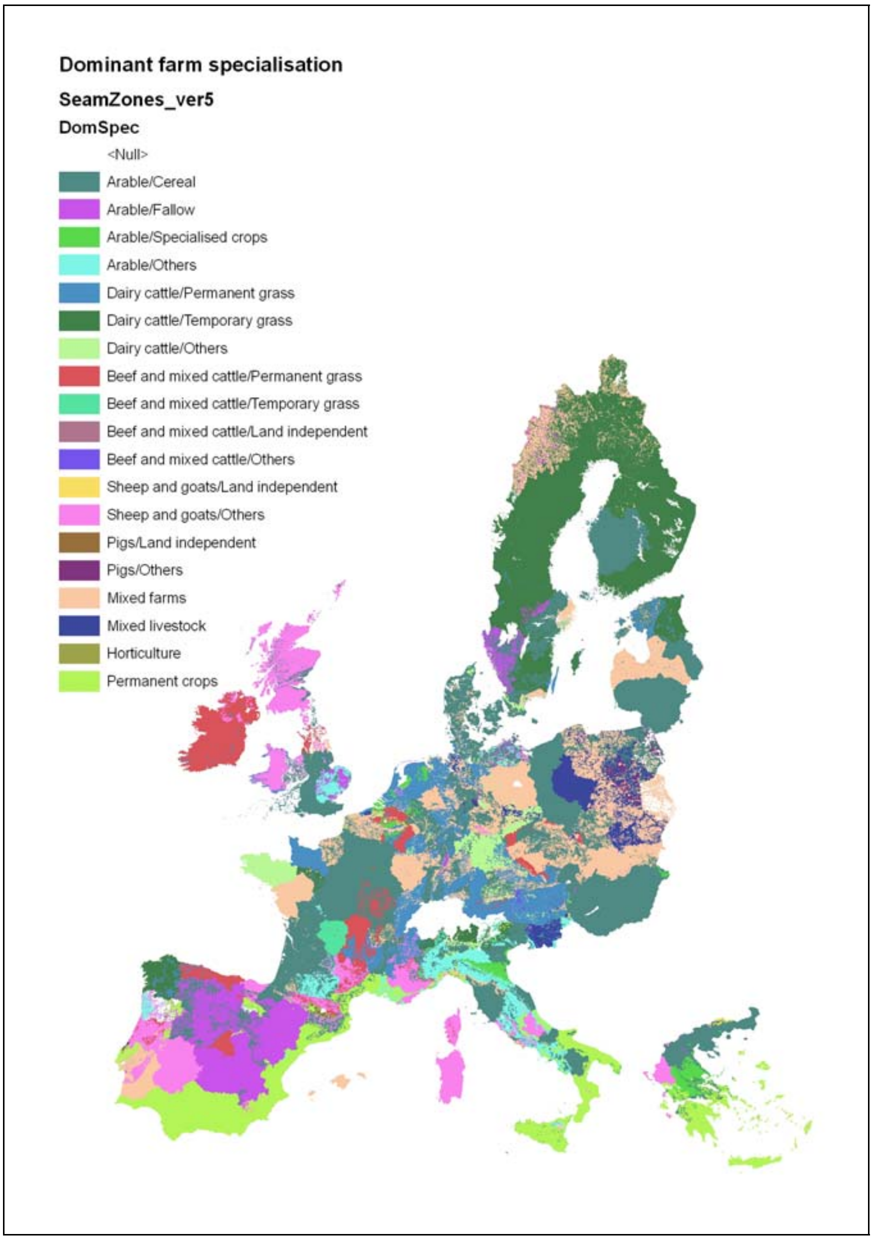
As can be seen in Map 3.2 large scale farm dominate in the North-western part of the European Union, except in Ireland where the Northern part are dominated by small scale farming and the Southern part by a mix of large and medium scaled farm types according to agri-environmental zones. The Southern and Mediterranean part of the Union shows is greater diversity in dominating farm types. In most Member States in this part of the Union all three size classes can be found as dominating. The results for EI-10 also show a diverse picture, where both small, medium and large scale farm types can be found as dominating. The results in general show less differentiation within the administrative regions than for the allocation of farmland in Section 3.1. This indicates that the farm size according to the spatial allocation used here are less depending on the bio-physical endowment than on regional differences. However, the first results shown here are very rough and regional analyses are needed to verify this. It is also worth noting that in most of the administrative regions where more than one farm size is dominating only two farm size are found to be dominating. The two most frequent combinations here are large/medium scaled and large/small scaled – only a few regions can be found where the combination medium/small scale farm types have been found to dominate.

3.3 Allocation of farm types according to farm specialisation

The results of the allocation of farm types to agri-environmental zones in terms of dominating specialisation of the allocated farm types are show in Map 3.3. The overall picture is a huge variation in dominating specialisation across EU25. Of 21 possible farm type specialisations 19 are found to be dominating in one or more regions. The two specialisation types that are not found as dominating in any region are ‘Poultry and mixed pigs/poultry’ and ‘Dairy cattle/land independent’. Arable specialisation farm types are the most dominating specialisation across EU25 from Western part of Finland to Scotland, Central Spain, Sicily and Greece. Of the arable specialisation types cereals is the most frequent one. Fallow specialisation can be found in Central Spain, Sweden, Mecklenburg-Vorpommern and Eastern England, specialised crops are found to dominate in parts of Belgium, the Netherlands, Greece and Emilia Romagna, whereas other arable specialisation are dominating in many parts of Italy and in parts of the regions Mecklenburg-Vorpommern, Midi-Pyrenee and Eastern England. Dairy cattle specialisations dominate in many areas, mainly North of the Alps or the Pyrenees. Dairy cattle specialisation based on permanent grassland and the most widespread type dominating in areas of the Netherlands, Belgium, Germany, Austria, Estonia, Galicia, Normandy and several areas in the Alp regions. Temporary based dairy cattle systems dominate in many parts of Sweden, Finland and Estonia and in Pays de Loire, Galicia and parts of the Alp region of Italy. Other dairy cattle systems dominate in many parts of Germany and Bretagne and in parts of Southern Sweden and Denmark. Of the beef cattle specialisations the one based on permanent grassland is the most widespread found in Ireland, Northern Ireland, Central France, North-western Spain, Northern Portugal, Czech Republic and Yorkshire. The temporary grass based beef specialisations are found in Limousine, Piemonte and other Alp regions of Italy. Sheep and Goat specialisation, almost entirely land based systems, are found to dominate large areas of Scotland, Corse, Sardegna, Extremadura, Central and Northern Portugal, the Pyrenee region, Central Italy and North-western Greece. Pig farms dominate in few parts of EU25: Land independent pig systems are found to dominate parts of the area of Limburg, Cataluna and Pays de Vasco, other pig systems are found to dominate in parts of Mazoweckie, Podlaskie, Steirmark, Braunschweig and Lisboa area. Mixed farms are dominating widespread mainly in the Northern part of Continental

Europe: Germany, Poland, Denmark and Czech and Slovak republics. Also many areas in Belgium, Latvija and Estonia, France, Italy and Alentejo are dominated by the mixed farms. Horticulture specialisations dominate in Liguria and in parts of the Netherlands and North-western Greece.

Map 3.3 *Dominant farm specialisation of the agri-environmental zones. Measures as the share of the agricultural area managed by the different types of farms according to specialisation.*

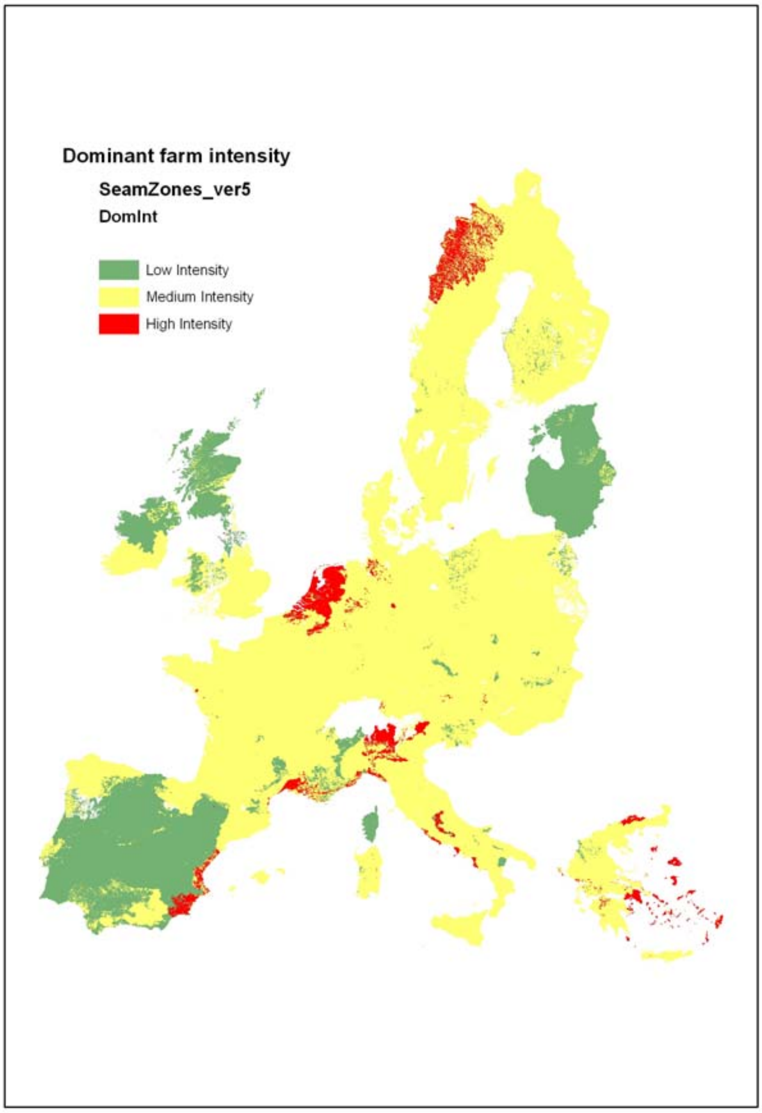


Finally, Permanent crop specialisations are dominating in most of Greece, in Southern Italy, along the Mediterranean coast from Liguria to Algarve and other parts of Portugal. In the overall picture more than half of the administrative region show differences between the agri-environmental zones. This might be expected to the case for more regions as land use is a major discriminator in the spatial allocation of farm types. However, more detailed regional analyses are needed to verify this.

3.4 Allocation of farm types according to farm intensity

Map 3.4 shows the results of the allocation of farm types to agri-environmental zones in terms of dominating farm type according to intensity. More specifically it is dominating intensity of farm types in terms of the share of the area of the agri-environmental zones managed. Most regions are identified as dominated by farm types of medium intensity. High intensity farms dominate in the Netherlands and in minor parts of Italy, Greece, Spain, France, Germany and Sweden. Low intensity farm types dominate in the majority of the area of Portugal, Spain, the Baltic States, Scotland, Wales, Northern Ireland and Northern part of Ireland.

Map 3.4 Dominant farm intensity of the agri-environmentalzones. Measures as the share of the agricultural area managed by the different types of farms according to intensity.



The results shown as dominating intensity of farm types shows little variation between the agri-environmental zones within administrative regions. This is a little surprising as yields are one of the variables that are used for the spatial allocation and which could be expected to vary between agri-environmental zones as these per definition represents different bio-physical endowments within the administrative regions. This might indicate that yields not always follow intensity or that the scale of these analyses hides some differences. Regional analyses are needed to determine this.

4 Conclusions and recommendations for further work

The overall conclusion of this report is that both the disaggregation approach for land use in the Dynaspat project and the allocation of SEAMLESS farm types are delivering good results in terms of validation, at least for the EU-15. A further validation in new MS is still necessary. We know however that a validation would confirm a worse match between allocation results and statistical farm distribution then found in the EU-15. The reason is related with the data quality and the relative lower number of observations in the EU-12 data sources used for the allocation.

In spite of this we can confirm that the disaggregated farm type and land use information can be used well to relate the farm type information to a bio-physical context and will therefore enable:

- The differentiation of farm types according to bio-physical environment within regions
- to integrate market response behaviour with environmental performance of farms
- to up-scale environmental performances of farms to farm type groups

It is clear that the usefulness of the allocated farm type information as input for modelling in FSSIM still needs extensive testing. This should be part of follow up of Seamless work. Presently the allocated farm information has been compiled at the level of farm types per Nuts2 regions and environmental zone, including information on the share of the land on different soiltypes. It should be confirmed by the FSSIM modellers (T3.3) whether this is a useful way of organising the data.

Another complicated issue is the disclosure problem of the used FADN information. When preparing the presentation of the results we have assumed that disclosure rules are not violated when the disaggregated results are presented at agri-environmental zone level even though they are not necessarily representing more than 15 farms. This assumption is based on the fact that the disaggregation results are only based on a statistical estimation, they are therefore not real and they only provide information on the area present of a certain farmtype. This means that any attributes derived directly from the FADN data and presented for the allocated farm types will be based on a regional average (per FADN region or HARM1) and will only be disclosed when represented by 15 or more farms.

Finally it should again be stressed that although the FADN database is the only available EU wide farm information source containing individual farm information, it still has major disadvantages. These should be kept in mind when working with the allocated farm data. The major disadvantage is that the FADN sample does not include all the small farms and all the part-time farms. This means that especially the farms in the more marginal farming areas which mostly coincide with the Agri-mask 1 and 2 areas in the AEnZ, are not well represented. This also explains the low farm area allocated to these agri-mask regions. The low representation of farm area in the Agri-mask regions is further aggrevated by the exclusion of common land, seasonal lets and wintering/summering arrangements in FADN. This aspect should be further discussed in the SEAMLESS project when specific attention is paid to the High Nature Value farmland areas as these coincide strongly with the areas in agri-mask 1 and 2 and include common land use categories.

References

Andersen, E., Verhoog, A.D., Elbersen, B.S., Godeschalk, F.E., Koole, B., 2006. A multidimensional farming system typology, SEAMLESS Report No.12, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 30 pp, ISBN no. 90-8585-041-X.

Elbersen B., Kempen, M., van Diepen K., Andersen E., Hazeu G., Verhoog D. 2006. Protocols for spatial allocation of farm types, SEAMLESS Report No.19, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 107 pp, ISBN no. 90-8585-046-

Hazeu, G.W., Elbersen, B.S., van Diepen, C.A., Baruth, B., Metzger, M.J., 2006. Regional typologies of ecological and biophysical context, SEAMLESS Report No.14, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 55 pp, ISBN no. 90-8585-042-8.

Verhoog, D. and Andersen, E. (2006) First prototype of the global data component of the SEAMLESS datasets for Prototype 2 of SEAMLESS-IF. SEAMLESS D4.6.2.

Glossary

Agri-Environmental Zonation A biophysical typology based on environmental zones and soil data.

Agri-Environmental Land Types The elements of the Agri-Environmental Zonation (AEnZ)defined by the combination of Environmental Zones, Organic Carbon content and Agri_mask classes.

Agri-mask A mask indicating which areas in Europe are have no or relatively small constraints, which areas are not suitable and which areas are strongly handicapped for arable agriculture

Allocation This is a methodology that enables to add a (below regional) locational dimension to every individual farm contained in the FADN data base and every land use in the FSS database.

Enironmental Stratification A statistical environmental stratification of Europe consisting of 84 strata based on 20 most important environmental variables.

Environmental Zones An aggregation of the 84 environmental strata into 13 environmental zones.

FADN Farm Accountancy Data Network of the European Union (FADN) has been established since 1965. The aim of the network is to gather accountancy data from farms for the determination of incomes and business analysis of agricultural holdings. Based on sample farms covering information on farms in EU-15.

FADN farm One sample farm in the Farm Accountancy Data Network. FADN is based on a representative sample of all agricultural holdings. The sample covers about 60,000 holdings in EU15

Farm type A classification of farms according to different dimensions. In SEAMLESS a farm typology for the whole EU has been developed. The different dimensions of this typology are:

- Size: Measured as the economic size of farms
- Intensity: Measured as the total output in Euro per ha
- Specialisation: Measured as the standard gross margins from different types of crops and livestock
- Land use: Measured as the proportion of the agricultural area covered by specific types of crops.

To reduce the number of farm types the two last dimensions are combined in one dimension. This is possible because not all combinations of these two dimensions are relevant. In total of 189 farm types are identified. This is the aggregate of 3 size types, 3 intensity types and 21 combined specialisation/land use types.

FMU	Farm Mapping Unit. FMU is a continuous region with similar soil soil conditions determining potential yields and similar altitude and LFA characteristics. FMUs are a cluster of HSMU and were created to reduce the complexity of the allocation procedure of FADN farms.
FSS	Farm Structure Survey data are used to collect information on agricultural holdings in the Member States at different geographic levels (Member States, regions, districts) and over periods (follow up the changes in agricultural sector), thus provide a base for decision making in the Common Agricultural Policy. Responsible Institution at EU level is Eurostat.
HSMU	Homogeneous Spatial Mapping Units are an intersection of land cover (Corine LC 2000), relief (slope in 5 classes), Soil Mapping Units (so-called soil landscapes from the *European soil map*) and the Nuts 2/3 boundaries (depending on the size of the NUTS regions) (see Figure 2.1). Each HSMU has identical values for land cover class, slope class and Soil SET, other parameters (such as annual rainfall) may differ inside the HSMU.
OCTOP	The Organic Carbon content of the TOPsoil (OCTOP) (in %) calculated for every 1km2 in Europe