Spatially explicit farm and environmental indicators at a scale of 1 km x 1 km

Deliverable No. D4.6

SUSFANS DELIVERABLES

Adrian Leip (JRC), Markus Kempen (UBO), Xavier Rotllan-Puig (JRC), Maria Bielza (JRC)

Land use diversity and soil erosion are amongst the aggregated variables required for describing environmental sustainability in the domains 'biodiversity' and 'natural resources'. Both aggregated variables need to be quantified at high spatial resolution. The CAPRI model is able to do this, but the calculation procedure required improvements. This report describes basic features of the methodology, scrutinizes deficiencies in the current implementation and identifies possibilities to update and improve the method.

Version	Release date	Changed	Status	Distribution
V1	30/03/2017	--	Final	--

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 633692

SUSFANS DELIVERABLE DOCUMENT INFORMATION

Project name	SUSFANS
Project title:	Metrics, Models and Foresight for European SUStainable Food And Nutrition Security
Project no	633692
Start date:	April 2015
Report:	D4.6 Spatially explicit farm and environmental indicators at a scale of 1 km x 1 km
Work package	WP 4
WP title (acronym):	Drivers and data: primary agricultural and fisheries production
WP leader:	UBO (Andrea Zimmermann), JRC (Adrian Leip)
Period, year:	2, 2017
Responsible Authors:	Adrian Leip, Markus Kempen, Xavier Rotllan-Puig, Maria Bielza
Participant acronyms:	JRC, UBO
Dissemination level:	**Public**
Version	1
Release Date	30/03/2017
Planned delivery date:	30/03/2017
Status	Final
Distribution	

Dissemination level of this report

Public

ACKNOWLEDGEMENT & DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 633692. Neither the European Commission nor any person acting on behalf of the Commission is responsible for how the following information is used. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

TABLE OF CONTENT

DELIVERABLE SHORT SUMMARY FOR USE IN MEDIA

For the quantification of environmental sustainability, at least two aggregate variables require assessment at farm level or high spatial resolution: agricultural land use diversity, and soil erosion. Both aggregate variables are important metrics describing dimensions of the performance metrics for biodiversity conservation (land use diversity) and preservation of natural resources (soil erosion, quantifying the aggregate indicator 'maintenance of soil fertility').

The estimation of these two aggregate variables at regional or national scale is not possible, as it would 'average out' and providing meaningless information.

The CAPRI model has a long-term experience in calculating environmental indicators at high spatial resolution as detailed as 1x1 km (Britz and Leip, 2009; Leip et al., 2008, 2015, 2011a); however a need for updating the procedures and improving some of the algorithms was identified.

This report describes how these updates and improvements are possible, focusing on three stages of the overall simulation framework:

- Quantification of a priori crop shares
- Disaggregation of crop yield and farm inputs
- Quantification of potential loss of soil through water erosion

Therefore, the quality of SUSFANS estimates of environmental sustainability, calculated with CAPRI, will be substantially higher than estimates previously available.

TEASER FOR SOCIAL MEDIA

This deliverable is only relevant for readers who are really interested in technical details how statistical data available at regional scale in the CAPRI model are used to generate high quality agri-environmental indicators at high spatial resolution. Keep your interest and curiosity for SUSFANS results on spatial data of CAPRI simulations!

The report describes basic features of the methodology, scrutinized deficiencies in the current implementation and identifies possibilities to update and improve the method.

High quality, high resolution indicators allow to monitor the environmental integrity of agri-environmental policies at local conditions.

Loss of #biodiversity and soil #erosion are amongst the biggest environmental challenges; their calculation in the CAPRI is improved.

ABSTRACT

For the quantification of environmental sustainability, at least two aggregate variables require assessment at farm level or high spatial resolution: agricultural land use diversity, and soil erosion. We performed a thorough review of the procedures used to calculate environmental indicators at high spatial resolution with the CAPRI model. The need to update and improve was found for different stages of the procedure:

- Quantification of a priori crop shares
- Disaggregation of crop yield and farm inputs
- Quantification of potential loss of soil through water erosion

This report presents the results of feasibility studies on the possibility of improvements in those three stages.

For the quantification of a priori crop shares, an update of the LUD model (Lamboni et al., 2016) is proposing several improvements on model performance and model quality, which overall were shown to generate better prediction both of 'frequent' and 'less frequent' crops; the latter mainly through the introduction of environmental suitability ranges.

The disaggregation of crop yields can be improved through an update of the prior crop yield estimates obtained from a crop model, but in particular also through using results from the SUSFANS yield gap analysis. These results allow that the fertilization rates beyond crop needs can be 'disconnected' from crop yield and linked to the yield gap estimate at high spatial resolution, so that over-fertilization is estimated at those units where nutrient input does not limit crop growth.

Potential soil losses by water erosion was found to be over-estimated for fallow land in non-arid regions. An update of the calculation procedure is proposed, accompanied with a comprehensive literature review not only on soil erosion in Europe for validating CAPRI estimates, but also on soil erosion thresholds that can be used to differentiate sustainable erosion rates from medium and severe erosion rates.

1 INTRODUCTION

Agri-environmental indicators are suitable instruments to identify agri-environmental problems, such as hot-spots, and to monitor changes in environmental quality and in the efficiency of agri-environmental policies (EC, 2006). Quantifying agri-environmental indicators is a challenging task, in particular for large and heterogeneous regions such as the area covered by EU-countries and if high spatial resolution is required. Indeed, many agri-environmental threats ask for indicators at high spatial resolution or at scales that are different from administrative entities. Reliable agri-environmental indicators hinge on the availability of high-quality input data, which in those cases where national-scale data are insufficient often poses considerable problems.

During the last decade, the partial equilibrium model for agriculture CAPRI (Britz and Witzke, 2014) was continuously extended towards the integration of agri-environmental indicators in the modelling framework. In particular, this implied the development of a spatial layer into which regional data are disaggregated. Many agri-environmental indicators are depending on local environmental conditions and require the availability of data at high resolution. Also, it allows the quantification of indicators through linkage with biophysical models (Leip et al., 2008) or using detailed meta-models (Britz and Leip, 2009).

Currently, CAPRI covers a full activity- and product-based GHG accounting system, calculates nitrogen balances and differentiated emissions of reactive nitrogen ($N2_O$, NH_3, NO_x, NO3). The CAPRI-RD research project (2009-2013) included additional, such as the risk of soil erosion, biodiversity friendly farming practices, farmland bird index, agricultural landscape structure, and an indicator related to environmental compensation zones which were recently used for an assessment on the 'greening of the CAP' (European Commission, 2011).

In the SUSFANS project, CAPRI is one of the economic models used in the SUSFANS toolbox (Rutten et al., 2016a, 2016b) and will provide both economic as well as environmental indicators (Götz et al., 2017). The CAPRI model is described in detail in previous SUSFANS deliverables (Götz et al., 2017; Rutten et al., 2016b); here the focus is on the *spatial layer* of CAPRI which was not yet addressed in the documents mentioned.

Zurek et al. (2016 and Deliverable 1.3 forthcoming) define the environmental aspects (performance metrics) that are aimed at assessing the societal goal of 'reducing environmental impacts'. This includes 'climate stabilization', 'clean air and water', 'biodiversity conservation', and 'preservation of natural resources'.

The objective of SUSFANS is the assessment of the agri-food system to contribute to the societal goals; most of the aggregated indicators required to perform this assessment can be quantified with the CAPRI model at the regional (NUTS2) level. However, two of the aggregate indicators require more detailed analysis:

- Agricultural land use diversity where evaluation at regional level introduced an aggregation bias and cannot be used as a proxy for 'agricultural patchiness' (Weissteiner et al., 2016) which on its own is used as a proxy for land use diversity (see Zurek et al., 2016 and Deliverable 1.3 forthcoming).
- Losses of soil organic carbon via water erosion largely depends on the local soil conditions (such as slope) in combination with crop cultivation and farm practices which require evaluation at high spatial resolution. Again, assessment at the aggregated level is not suitable for the quantification of soil erosion (Leip et al., 2015).

The calculation of disaggregated indicators in CAPRI follows a sequential procedure as illustrated in Figure 1. First, crop shares are distributed in to the spatial units within an administrative region, using *a priori* land use shares and associated uncertainties, and available statistical data at NUTS2 and NUTS3 level. Yield and irrigation levels are then subsequently estimated for each crop-spatial unit combination on the basis of irrigation shares obtained from statistical surveys (EC, 2008, 2003a) and scientific assessments (Siebert et al., 2007). These are combined with crop yields which were simulated for six crops (barley, grain and fodder maize, potatoes, pulses, sugar beet, sunflowers, and soft wheat) under irrigated and rain-fed conditions (Orlandini and van der Goot, 2003). Livestock densities depend from fodder production, environmental and economic factors. Densities of land-based animals (ruminants) and land-free animals (monogastric animals) are regressed using animal numbers from the Farm Structure Survey (EC, 2003a). In a last step, farm management in terms of (mineral) N input is estimated on the basis of crop N requirements and N availability from atmospheric deposition, biological N fixation, crop residues, manure (as a function of livestock density, assuming transport of manure not for distances larger than 10 km).

The resulting 'data base' is consistent with the parent CAPRI data at the regional scale (for any scenario that was used) and provides all the individual variables required for the calculation of terrestrial agro-environmental indicators. However, each of these steps obviously introduces also an additional layer of uncertainty. It is therefore preferable to use indicators at highest aggregation level which is 'fit for purpose'. As indicated above, the use of the CAPRI spatial layer is essential for the two variables 'soil erosion' and 'land use diversity'. For other indicators linked to 'emissions' the spatial layer might be used to evaluate the probability distribution of the magnitude for an individual variable.

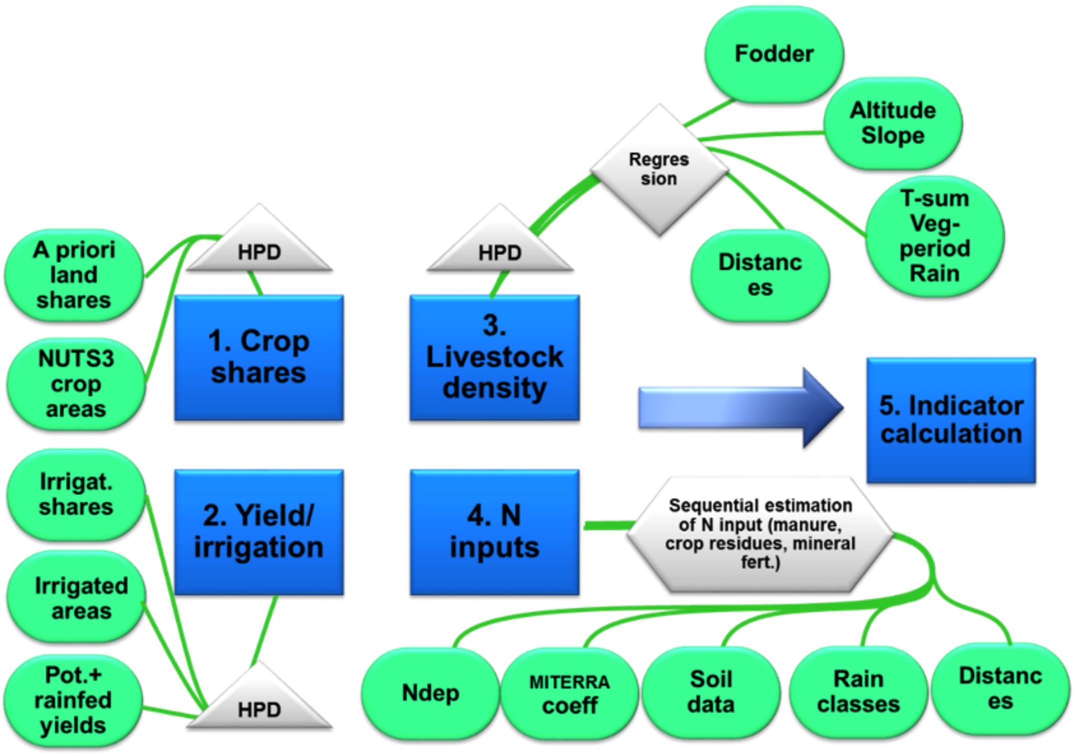

Figure 1. Schematic representation of the disaggregation procedure in CAPRI.

Even though the CAPRI spatial layer is operational (Kempen, 2013; Kempen et al., 2005; Leip et al., 2008, 2015, 2011a), Leip (2011b) identified the need for an update and scientific improvements:

a) Update of the spatial layer, facilitating the link between agro-economic models, better alignment with INSPIRE guidelines, and climate models and extending the geographic scope to continental Europe.
b) Updating and improving the *a priori* land use shares.
c) Improving the distribution of farm inputs within an administrative region.
d) Improving the quantification of agri-environmental indicators.

The process of implementing the necessary changes and developing new methodologies is an ongoing process and had been supported by a number of previous projects, including CC-TAME[1], C-SCAPE[2], GHG-EUROPE[3], and imap8[4].

Within these projects,

- The update of the spatial layer (so-called Homogeneous Spatial Units, HSU) has been finalized. The HSU data set will be made available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 license. For more explanation regarding the license[5]. The HSU data set has been already uploaded to PANGAEA[6] data repository and the publication process is ongoing. In parallel, a publication with a detailed description of the HSU data set (mapping methodology and underlying data sets) is in preparation. PANGAEA ensures a permanent link to the data set and ensures that the data can be referred to by a unique DOI (Digital Object Identifier) and be cited similar to a publication in a journal.
- A new Land Use Disaggregation (LUD) model was developed (Lamboni et al., 2016). However, this model still had some quality deficiencies in predicting non-frequent land uses due to the presence of 'outliers' in the predictions (Lamboni et al., 2016). The **first objective** of this SUSFANS deliverable is to advance with the LUD model and improve the *a priori* land use shares to be used in the CAPRI disaggregation model (see section 2)
- Input of nitrogen to the crops depends currently on crops nutrient requirements (which is linked to crop yields) but is not further differentiated; thus within a region, the level of 'over-fertilization' is independent from the 'yield gap'. In practice, one would expect a large yield gap might be correlated with lower (over) fertilization rates than a situation with a lower yield gap where nutrient limitations are largely eliminated. The **second objective** of this SUSFANS deliverable is therefore to use results from the SUSFANS yield gap analysis (Zimmermann and Latka, 2017) (see section 3).

[1] http://www.cctame.eu/index.html
[2] Contract JRC Ref n° 31699-2010-09 TPW ISP, Project "Accounting for Carbon and GHG Emissions" (C-SCAPE) with the Norwegian Agricultural Economics Research Institute (NILF)
[3] FP7 EU contract No. 244122. http://ghg-europe.eu/index.php
[4] Administrative Agreement N°AGRI-2015-0213-JRC N°33919-2015-06 between DG Agriculture and Rural Development and the Joint Research Centre
[5] see e.g. http://creativecommons.org/licenses/by-nc-sa/4.0/
[6] https://pangaea.de/

- Furthermore, the indicator describing the potential for soil erosion by water has been checked and found to be overestimating soil erosion on fallow land in non-arid regions. The **third** objective of this SUSANS deliverable was thus to do an update of this indicator (see section 4).

2 IMPROVING A PRIORY CROP SHARE ESTIMATES

Authors: Adrian Leip, Xavier Rotllan-Puig

2.1 Introduction

The first step in the spatial disaggregation of CAPRI regional data and the calculation of agri-environmental indicators at high spatial resolution is the provision of good *a priori* crop shares for each spatial unit. CAPRI so far uses an agricultural land use map developed in the CAPRI-DynaSpat project (Kempen, 2013; Kempen et al., 2005; Leip et al., 2008) which is based on statistical information collected around the year 2000. The JRC has started working on an updated and improved land use share map on the basis of more recent data (2008-2010) (Lamboni et al., 2016).

The results were compared with land use observation at high resolution for France, using data from the LPIS data base (Cantelaube and Carles, 2015). This gave confidence in the performance of the model for the frequent crops, while the model still had some quality deficiencies in predicting non-frequent land uses due to the presence of 'outliers' in the predictions (Lamboni et al., 2016).

For the use of the CAPRI spatial layer in the SUSFANS project, it is therefore crucial to continue this work and improve the model as proposed by Lamboni et al. (2016) by better constraining un-frequent crops.

2.2 Improvements of the LUDM

2.2.1 The Land Use Disaggregation model

The Land Use Disaggregation model (LUD) model developed by Lamboni et al. (2016) spatially disaggregates agricultural land use data from a coarse-scale (e.g. administrative region, NUTS2) into the finer-scale Homogenous Spatial Units (HSU), a grid cell of 1 km x 1 km, or a collection of these grid cells having similar properties (Leip et al., 2011b). In other words, LUD predicts the agricultural land-use areas within each HSU for the EU-28 countries.

LUDM combines available statistical data on land-use at NUTS2 and NUTS3 level and point-based observations of land-use from the LUCAS survey (EC, 2003b), which allows to link land use choices with environmental (climate, soil, and land cover classes) and topographical information (relief).

The model uses a three-step approach as shown in Figure 4 of the paper.

- Step 1: develops a multinomial logit model estimating the land use under given environmental conditions and farm gate prices for certain agricultural products, using LUCAS point land use observations. Results are prior estimates of model parameters. Lamboni et al. (2016) derived one vector of model parameters for each NUTS2 region.
- Step 2: improves the model parameters using a Bayesian approach and on the basis of the land use distribution within the NUTS2 over the NUTS3 regions. In this step, crops that have not been included in the first step are included in the model. This steps results also in predicted land use shares for all spatial units.
- Step 3: takes up the predictions from Step 2 and constrains them in order to match them with statistical data at NUTS3 level. Also known shares of area which is of non-agricultural use are fixed; the authors assume that forest area (Kempeneers et al., 2013; Pekkarinen et al., 2009) is accurate and therefore not available for agricultural land use. Step 3 results in final land use shares which can be used by the CAPRI disaggregation model as *a priori* estimates (Britz et al., 2011).

The model was tested on LPIS data available for France. As can be seen in Figure 5 of the paper below, the model performed very well for frequent crops such as rapeseed where Figure 9 of the paper shows the spatial distribution of rapeseed in France in the LPSI data and in the disaggregation results.

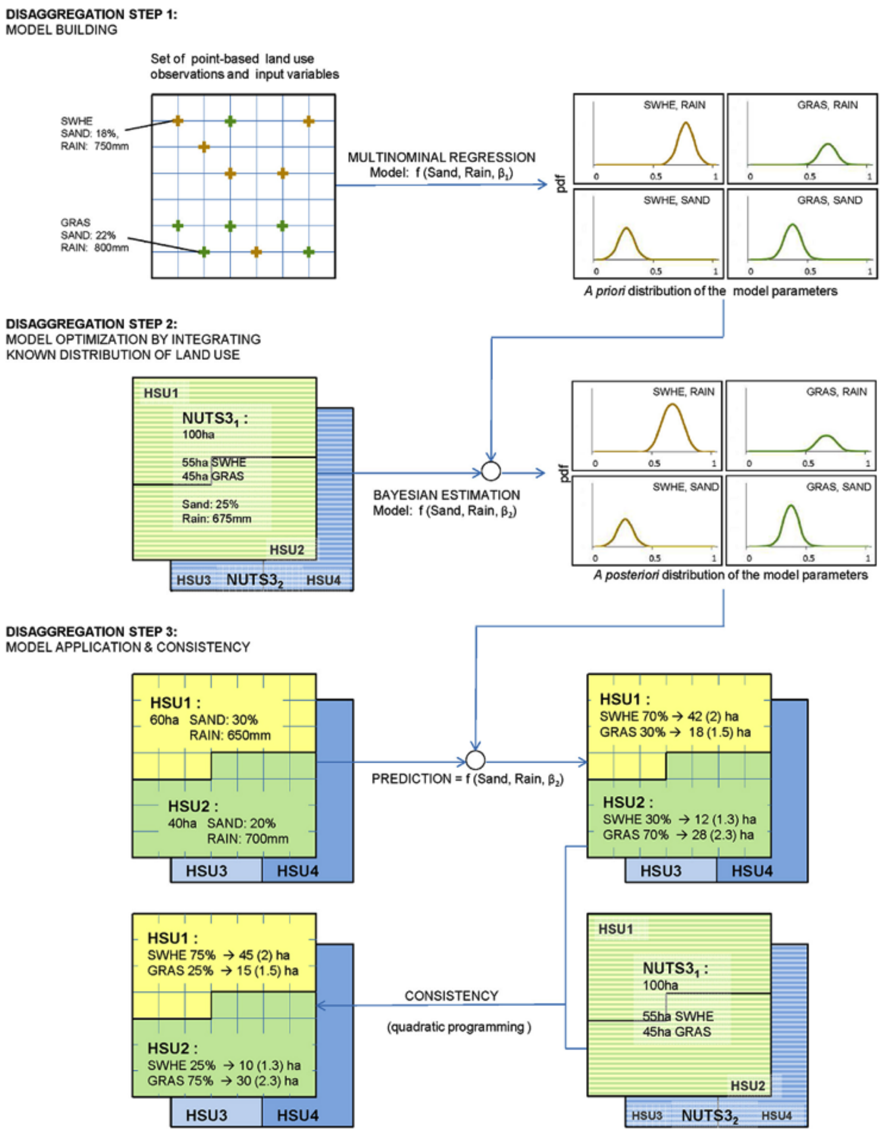

Fig. 4. Global structure of the land-use disaggregation model. Point-based observations and input variables are combined to get the *a priori* distribution of parameters (Step 1). The *a priori* distribution is combined with the land-use statistics to get *a posteriori* distribution (Step 2). The predictions of land-use areas at HSU level (using *a posteriori*) are constrained to match with both the HSU areas and the land-use statistics (Step 3).

Figure 2. Fig4 of Lamboni et al (2016)

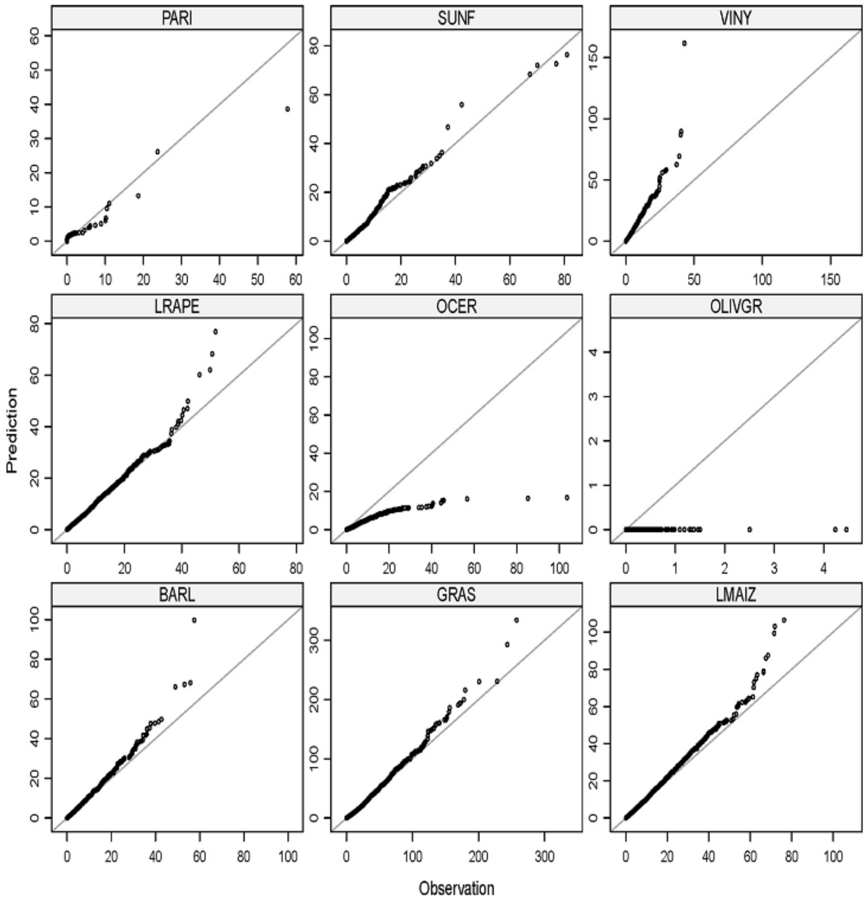

Fig. 5. Q-Q plots of the predicted land-use areas (×100 ha) and the observed land-use areas (×100 ha) at the HSUs level for France. The land-uses BARL, GRAS, LMAIZ, LRAPE, OCER, OLIVGR, PARI, SUNF and VINY stand respectively for barley, grassland, maize, rapeseed, other cereal, olive, rice, sunflowers and vineyard. The peculiar behaviours of the plots VINY, OCER, OLIVGR are mainly due to the uncertainties in the data we used for the comparison and some deviations between predicted and observed land use are evident (see paragraph 6.2).

Figure 3. Fig5 of Lamboni et al (2016)

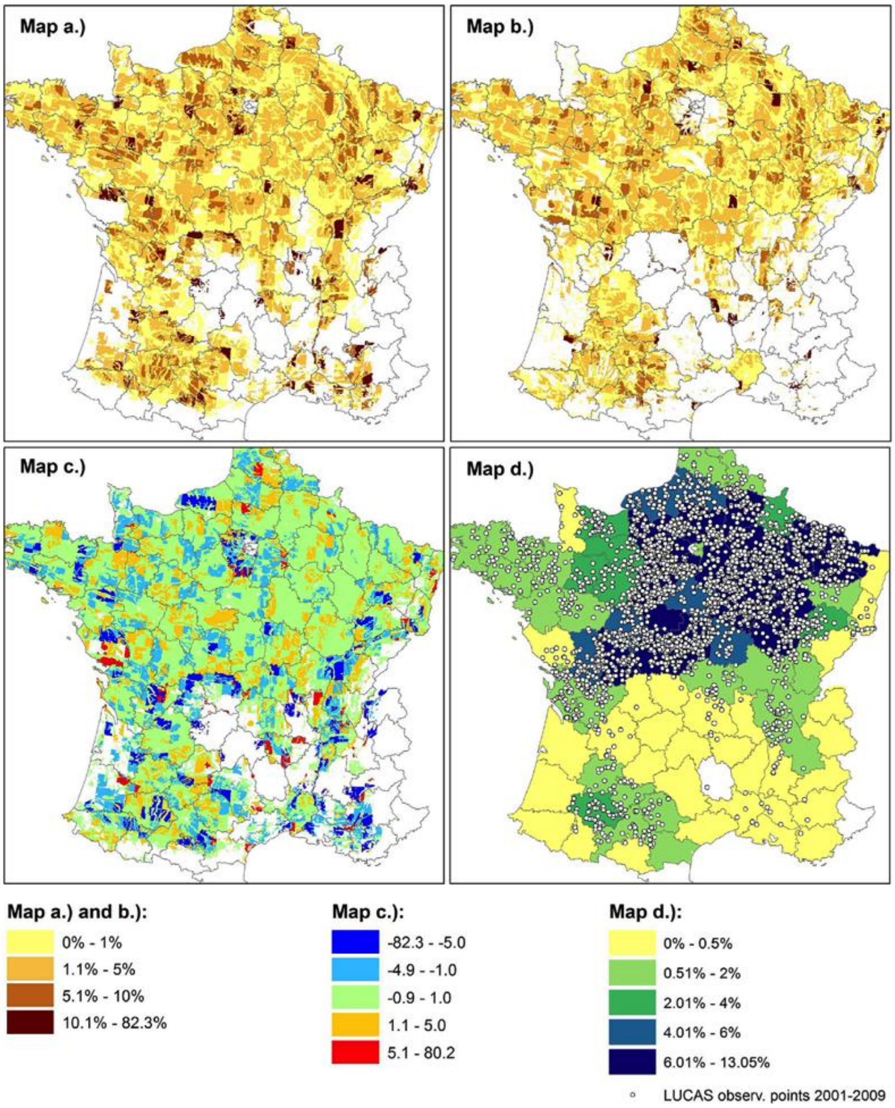

Fig. 9. Comparison of French LPIS data and the constrained disaggregation results for rapeseed for the year 2010. The upper maps show rapeseed area in the HSU as % of tot rapeseed area in the NUTS3 region based on French LPIS data (Map a.) and from the disaggregation (Map b.). Map c. gives the difference between LPIS data and the disaggregati results (Difference between Map b. and Map a.). Map d. presents the original FSS data at NUTS3 level (rapeseed as % of total NUTS3 area) and the LUCAS rapeseed observatio (points) in the years 2001–2009.

Figure 4. Fig9 of Lamboni et al (2016)

As the authors pointed out, LUD model predicted accurately the main land uses of the NUTS2 region. However, caution needed to be taken with predictions of non-frequent crops and/or crops requiring specific cultivation conditions. Such model limitations were probably due to the presence of outliers in the regional (NUTS2) data used to develop it. Potential solutions to improve its accuracy were related to the inclusion of agronomic constraints in the explanatory variables or giving capacity to the model to make predictions from sub-regional data (e.g. NUTS3; Lamboni et al., 2016).

This is discussed as follows:

"The process of the prediction of land-use areas involves a combination of different input variables such as altitude, slope, rain and land cover (CORINE) classes. While the rain can be more important for growing cereals for instance, it can be less important in the case of rice (as the rice field is often irrigated) and in the case of some permanent crops like olive. Our model predicts some small area of cereals like maize or wheat also at altitudes which are not suitable for growing cereals in Europe. Equally, the model predicts small area of rice when the slope is relative high. These results seem to be unrealistic as the rice field requires flat terrain for irrigation purpose. Although we have distinguished the individual CORINE classes for rice and olive as explanatory variables, the model still faces some difficulties to better predict these non-frequent crops. These miss-predictions are likely due to:

- *i) using only one set of model parameters at NUTS2 level to predict all the results in different locations (HSUs) of the NUTS2 region. Indeed, rice, olive and wheat can be seen in one sub-region (NUTS3 for instance) and not at all in others sub-regions of the same NUTS2 region;*
- *ii) not including, in this paper, agronomic constraints for the suitability or limits of growing certain crops under certain environmental conditions;*
- *iii) using a land cover classes (corine) of year 2006;*
- *iv) using a non-informative prior for the crops which are not found in the first step.*

The objective of the current study is to overcome these deficiencies. In addition, efforts are made to improve the performance of the model.

Note that part of the work described below had already been started under the imap8 project mentioned above. However, most of the work on the model's performance (2.2.2) and in particular the – most important - work on the environmental limits (2.2.3.3) was entirely done under the SUSFANS project.

2.2.2 Improvements on performance (required CPU time)

Given the high resolution and big extension of the study area (EU-28), it is essential to minimise the process time (CPU time). With this purpose, the main improvements included to the LUD model were changing some R packages and instructions used, which are less time-consuming especially for such big datasets. The two main R packages included in the process were *data.table* (Dowle and Srinivasa, 2016) and *mnlogit* (Hasan et al., 2016). Also, the first step of the model (see Figure 4 of Lamboni et al. 2016 above) was found to be the most computer resource intensive, as the logit model has to be generated several times until the optimum bandwidth is found, with some bottlenecks on which code improvements could focus.

On the one hand, with *data.table* all operations made on large tables (e.g. read, join, add/modify/delete columns by group) are drastically faster.

On the other hand, *mnlogit*, a package for estimation of multinomial logit models using maximum likelihood, was used instead of *nnet* (Venables and Ripley, 2002) for two reasons. The first was because it is more time and memory efficient (see Hasan et al. 2016 for packages comparisons, also with *nnet*). The second reason is because *mnlogit* deals internally with correlated explanatory variables, while in *nnet* it needs to be done separately. It calculates their correlation and removes the ones more correlated, with a level of tolerance defined by the user.

Besides that, the source code was improved at several places. One of the most important was the re-coding of the quantification of the weights given to the LUCAS point within the maximum bandwidth (loweight.r in the original source code) replacing a loop with direct formulae. Figure 5 shows time requirement (s) for different bandwidths. The time required for the new piece of code was far below 1 s.

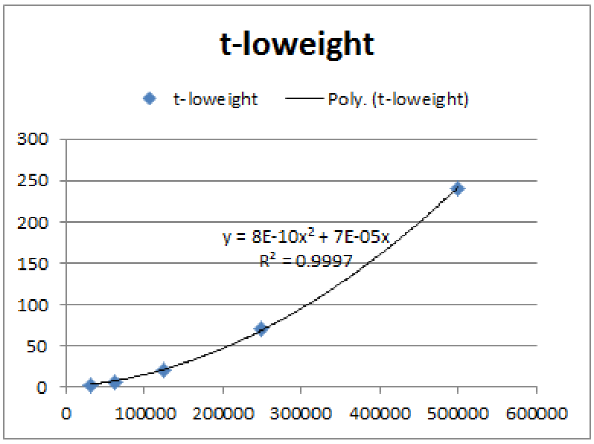

Figure 5. Time requirement (s) for different bandwidths

2.2.3 Improvement on quality

2.2.3.1 Optimum bandwidth – Selection criterium

The 'optimum bandwidth' is an important parameter in the LUD model as it determines how many observations (= LUCAS points) are used in the first step to parameterize the model. In principle, the larger the bandwidth the better the accuracy (as more points are taken under consideration) but this has two drawbacks:

(i) As one model is generated for each NUTS2 region, an increase of the bandwidth beyond the borders of the NUTS2 region 'dilutes' regional characteristics not included in the model explanatory variables but potentially influencing farmers choices for crop (e.g. infrastructure, market access for certain crops, local cultural / traditional preferences ...).

(ii) Required computing resources increase dramatically with the bandwidth.

Therefore, a performance criterion is used to determine the 'optimum' bandwidth. As performance indicator, the F-measure is used. The F-measure is a harmonic mean between the precision and recall or sensitivity. The precision (P) is the ratio of observations that have been well predicted to the total number of predictions of a certain land use. The recall (R) for that land use is the number of well-predicted observations to the total number of observations of the land use. The F-measure can have values between 0 and 1:

$$F = 2\,x\,\frac{P \cdot R}{P + R}$$

The F measure is quantified for all land uses.

For the final performance indicator however different possibilities exist:

(a) quantify the final score as average (or weighted average using the share of observations for each land use as weight) for all land uses available in the model (micro F-measure) or

(b) quantify the final score as (weighted) average over by a pre-defined selection of crops, for a example the crops occurring within a maximum bandwidth. This is called macro F-measure.

The micro F-measure optimizes the LUDM for the land uses around the center of each NUTS2 region. The macro F-measure penalizes models that exclude crops which are contained in the pre-selection.

The model as proposed by Lamboni et al. (2016) used the micro F-measure to improve model speed, at the cost of not considering less-frequent crops in the model at the first step. The LUDM was modified so that the macro F-measure was used on the basis of the share of all crops occurring in the NUTS2. This forces the model to 'grow' until all crops with a significant share in the NUTS2 are included and overcomes the problems mentioned above that some crops which are concentrated in some areas distant from the center of the NUTS2 might be excluded so that no model estimate exists after the first step.

In order to not test bandwidths that are unlikely to perform well, a criterion was included to check the number of crops and Corine land use classes. In the original version, the minimum number was set to three for both criteria. A further criterion was the variance for the explanatory variables (regressors), as the model can only be built if for each regressor at least two observations exist in the sample.

The minimum number of crops that need to be in the sample has been linked to all observations in the NUTS2 region, such that all 'major crops' in the region need to be included in the model. Major crops are identified which occupy at least one permille of the land (estimated from the number of observed occurrences in the LUCAS points) or – for NUTS2 regions with less than 2000 observation points, all crops with at least two observations.

2.2.3.2 Optimal bandwidth – Iteration

The original of the LUDM used a minimum and a maximum potential bandwidth and looped through the calculation of the F-measure score with a predefined bandwidth step to identify the optimum bandwidth for each NUTS2 region. Table E3 of the paper shows the min and max potential bandwidths by country as well as the bandwidth step and the range of optimum bandwidths found across all NUTS2 regions in a country.

This approach has been found to be inefficient as tests needed to be performed beyond the optimum bandwidth. The approach has been replaced by an iterative method. The concept of the iteration is simple: the F measure score is quantified for three equidistant bandwidths (b1, b2, b3) using a relatively large step width (db1). If a bandwidth at the edge scores best, then a next bandwidth outside the range is added, otherwise new intermediate bandwidths are added. This continues until the (predefined) minimum step width has been reached and thus the optimum bandwidth found.

There is however not a unique relationship between a bandwidth and the F measure determined for a specific NUTS2 region. For each bandwidth, the sample of observation is split into 2 groups; while 80% are lumped and used for the calibration of the multilogit model, the remaining 20% is used to validate the model and to calculate the F measure score. This is repeated until a stable F measure (mean) is obtained (i.e. an F measure with a requested precision which is calculated as a function of the current step width (delta bandwidth) and a pre-defined minimum requested variance (0.01) and a maximum allowed variance (0.015). As seen in Figure 6 depending on the sample the F measure (here Fnuts2 as defined above) scatters considerably and a stable F measure is obtained in this example only after several repetitions of the validation procedure.

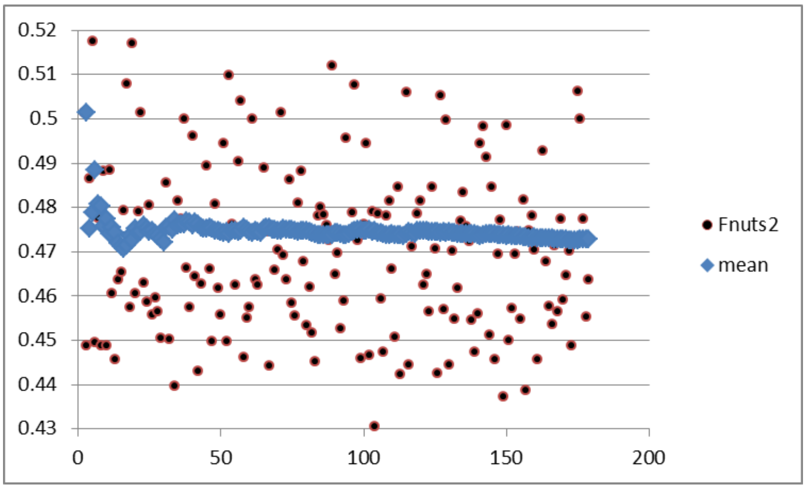

Figure 6. Evolution of the mean of an F measure until stabilizes

Table 1. Table E.3 of Lamboni et al (2016)

Table E.3
Bandwidths for EU-28 countries. The potential bandwidths are a set of values regularly chosen between the minimum (Min) and the maximum (Max) with the step (Step). The optimal bandwidths are selected by the model among the potential ones. As we can have different optimal bandwidths for different regions in the same country, we provide the minimum (Min) and the maximum (Max) of the optimal values. All the values are in km.

Potential bandwidths			Optimal bandwidths		
Countries	Min	Max	Step	Min	Max
Austria	20	150	5	20	40
Belgium	10	150	5	20	60
Bulgaria	300	480	10	300	340
Croatia	50	150	5	135	135
Cyprus	600	700	10	660	660
Czech Republic	20	150	5	20	45
Denmark	50	150	5	50	50
Estonia	180	350	10	280	280
Finland	50	150	5	50	85
France	50	150	5	50	75
Germany	25	150	5	25	60
Greece	50	150	5	50	130
Hungary	50	150	5	50	50
Ireland	50	250	10	60	200
Italy	50	150	5	50	65
Latvia	50	150	5	75	75
Lithuania	50	150	5	60	60
Luxembourg	50	150	5	50	50
Malta	25	150	5	100	100
Netherlands	35	150	5	35	70
Poland	50	150	5	50	50
Portugal	50	150	5	50	85
Romania	200	450	10	200	450
Slovakia	50	150	5	50	60
Slovenia	50	150	5	55	55
Spain	50	150	5	50	80
Sweden	50	150	5	50	80
United Kingdom	15	150	5	15	60

2.2.3.3 Introducing of environmental limits for crop allocation

One of the main limitations of the LUD model already described by their authors was the necessity of including agronomic/environmental constraints to the predictions. Or in other words, LUD sometimes predicted some crops/land-uses in places where the environmental conditions were not suitable for those crops to grow. This happened mainly for non-frequent crops such as rice, vineyards and olive groves, and also sometimes for sunflowers, but not only. Thus, the objective of this part was to calculate the ranges of each explanatory variable included into the model for each crop or land-use. Such ranges represented the maximum and minimum environmental conditions needed for each crop or land-use to grow.

The ranges for each crop/land-use were derived from the range observed in the LUCAS points and calculated from the whole EU. Some tests were run to check if the results were better with ranges calculated for each NUTS2 region. No significant differences were observed. In addition, it makes more ecological or agronomical sense to calculate them for wider regions such as the entire Europe, or at least for eco-regions.

Once these "environmental constraints" were calculated, they were used to remove the out-of-range predictions from the set of predictions obtained at the end of the Step 2 of LUDM. The reason for implementing it in this point, and not at the end of the modelling process, was because doing so the contribution of the Step 3 was kept, as the purpose of the Step 3 is to refine the final results so that they match another known data in a different scale (statistics at NUTS3 level in our case).

2.2.3.4 Other secondary improvements

Another secondary development of the model was the scaling of the explanatory variables. As advised in *nnet* documentation and in the literature, the explanatory variables should be roughly scaled to [0,1], otherwise the model fitting would be slow or might not even converge at all.

2.3 Results and discussion

Table 2 shows the results of a time comparison between *nnet, mnlogit* and the original LUD model. This comparison was done for the Step 1, the most time consuming, and for one NUTS2 region of France (FR10). Even with all the quality improvements included in the model explained above, running LUD model with *mnlogit* is faster on average.

Table 2. Time comparison beetwen nnet (nnet_x(1)), mnlogit (mnlogit_x(1)) and the original LUD model (mati(1)) for one NUTS2 region of France (FR10). neval is the number of times that the expression was evaluated. Time units in seconds

	expr	min	lq	mean	median	uq	max	neval
1	mnlogit_x(1)	640.7908	738.4014	1112.846	954.0632	1528.374	1702.602	5
2	nnet_x(1)	296.7388	492.9698	1248.989	1710.1651	1809.940	1935.130	5
3	mati(1)	1217.0835	1226.1828	1258.643	1238.4757	1277.640	1333.831	5

To assess the quality of the model predictions, we used independent data from the Land Parcel Identification System (LPIS) for France (Cantelaube and Carles, 2015). This is the best high-resolution data available so far. The LPIS data were re-mapped from parcel level into the HSU.

As the land-uses for LPIS data and LUD predictions were not exactly corresponding, we re-classified some of them in order to be comparable. Finally we obtained nine different land-uses as seen in Table 3.

Table 3. Link between land-uses of LPIS and LUD used to assess the model quality

Land-use codes	LPIS land-uses	LUD land-uses
PARI	Rice	Rice
SUNF	Sunflowers	Sunflowers
VINY	Vineyards	Vineyards
LRAPE	Rape and turnip seeds	Rape and turnip seeds
OLIVGR	Olive groves	Olive groves
OTHC	Common wheat, other cereals	Wheat, triticale, oats, rye, other cereals
BARL	Barley	Barley
GRAS	Mountain pasture / heathland, permanent grassland	Grassland with sparse tree/shrub cover, grassland without tree/shrub cover, spontaneously vegetated surfaces
LMAIZ	Maize	Maize

To be consistent with the methodology used in Lamboni et al. (2016), we used the weighted predictor error (Chakir, 2009) to check the quality of the improvements included in LUD compared with the original model. The errors terms were calculated as follows:

$$E_{l,h} = \sqrt{\left(\widehat{a_{h,l}} - a_{h,l}\right)^2 \times \frac{a_h}{\sum_{h=1}^{H_n} a_h}},$$

where $\hat{a}_{h,l}$ is the predicted area of a given land-use within the HSU, $a_{h,l}$ the observed land-use area, a_h the area of a HSU and H_n the number of HSUs in a given region (NUTS3).

To avoid the effect of the outliers, which affect both the mean and the maximum values, during this section we will focus on the median of these errors for the analysis. In addition, it has to be noted that the smallest HSU has an area of 100 ha (or 1 km^2). Thus, for instance, median errors on the order of 2 hectares would mean that more than 50% of the errors are under this value, which could be considered good predictions.

Firstly, we assessed the quality of the model after including two levels of developments. The first level of development was the new method to select the optimal bandwidth explained in sections 2.2.3.1 and 2.2.3.2. The second level was including also the function *mnlogit* instead of *nnet/multinom* for the estimation of multinomial logit models. We run the tests for the entire France. See the highest median errors in Table 4.

For grasslands, rapeseed and other cereals, the results with both new levels of developments were considerable better compared with the original model, yet minimizing the errors with *mnlogit*. Barley also gave better results using *mnlogit* than the original. On the other hand, olive graves, rice and vineyards gave worse results with the new developments, although keeping them around 1 hectare. Finally, maize and sunflowers using *mnlogit* gave errors about 5 and 2 hectares, which are still acceptable if we bear in mind that the minimum area of the HSUs is 100 ha.

Therefore, looking globally at these results could be said that the new developments improved the predictions of the more frequent land-uses and kept the non-frequent ones at low levels.

Table 4. Highest median error terms for the original LUD model and using the new method to select the optimal bandwidth with nnet and mnlogit (Highest Median LUD-orig, Highest Median nnet and Highest Median mnlogit, respectively) for France. Results in ha.

	Crop	Highest Median LUD−orig	Highest Median nnet	Highest Median mnlogit
1	BARL	3.150	6.403	1.985
2	GRAS	100.000	11.585	9.959
3	LMAIZ	3.634	3.103	5.078
4	LRAPE	100.000	6.062	4.647
5	OLIVGR	0.000	0.028	0.937
6	OTHC	100.000	14.413	5.628
7	PARI	0.000	0.000	0.937
8	SUNF	0.027	2.355	2.129
9	VINY	0.028	2.554	1.381

Secondly, to assess the effect of including environmental constraints for the land-use predictions of the model, we focussed on the region of France that shows more land-use predictions out of their environmental range (FR71). The model that includes constraints performed better for most of the crops, including the less frequent sunflower and vineyard (see Table 5).

Table 5. Comparison of median errors for model with and without including environmental constraints to the crop predictions for the region FR71. Results in ha.

	Crop	Highest Median No Constraint	Highest Median With Constraint
1	BARL	0.205779146	0.13171947
2	GRAS	4.185099234	3.68112016
3	LMAIZ	0.491291088	0.43337197
4	LRAPE	0.009458955	0.02577970
5	OLIVGR	0.000000000	0.00000000
6	OTHC	0.656160956	0.52134567
7	PARI	0.000000000	0.00000000
8	SUNF	0.068053302	0.03166883
9	VINY	0.035583687	0.01191906

Figure 7 shows the spatial distribution of the shares of sunflower observations (LPIS; Fig. 7A) and predictions, both before (Fig. 7B) and after (Fig. 7C) the inclusion of environmental constraints to the modelling process, in FR71. Both prediction maps show substantial similarities with observations (Figures 7 D and E) indicating good performance of the disaggregation models. However, the map after including the environmental constraints shows a pattern of fewer differences with observations (more green surface; Fig 7E) than the map without including such constraints (Fig 7D).

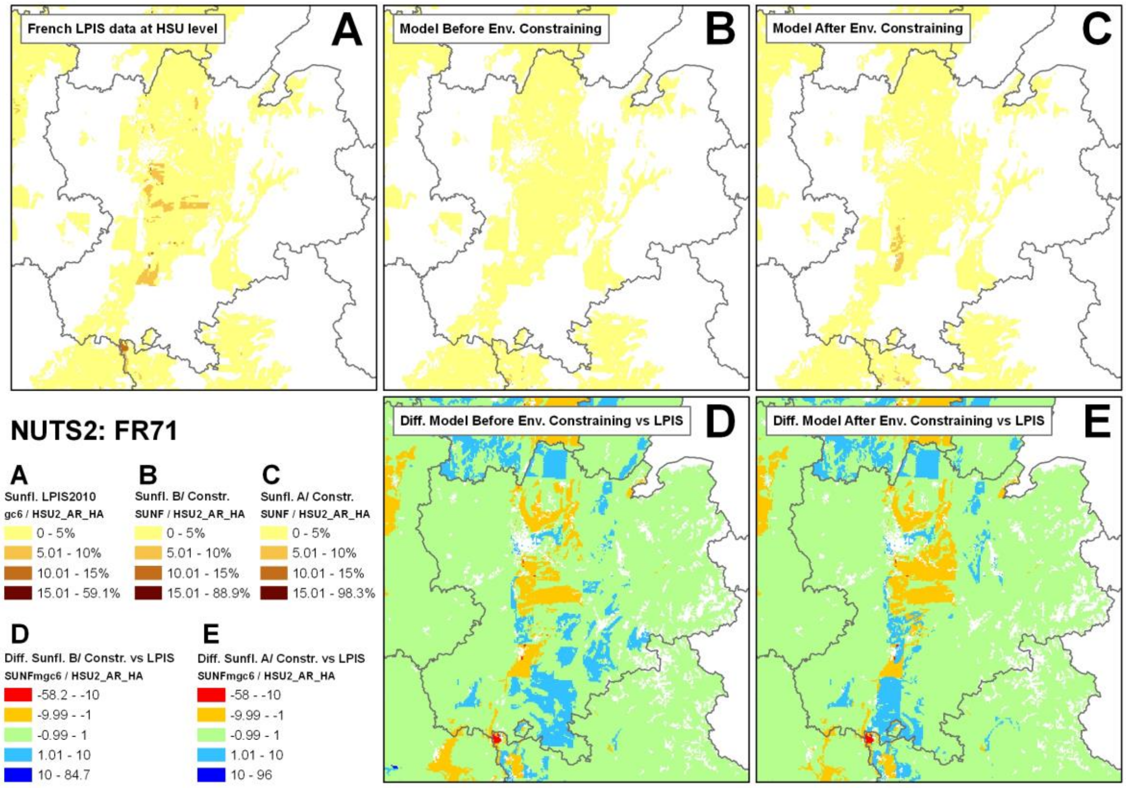

Figure 7. Distribution of shares of sunflowers in FR71. (A) Observation data (LPIS) and predictions of the model, (B) without and (C) with the inclusion of environmental constraints to the estimates. Maps D and E show the difference between observations and predictions

As in this region we did not have presences of olive groves and very few of rice, we checked the result for these non-frequent land-uses in FR81. The median for these two was zero, so we calculated percentiles 90 in order to be able to compare errors with and without including constraints. Table 6 shows that the error terms for olive groves were virtually the same and for rice it was slightly better for the model including environmental constraints.

Table 6. Comparison of percentile 90 errors for model with and without including environmental constraints to the crop predictions for the region FR81. Results in ha.

Crop	Percentile 90 No Constraint	Percentile 90 With Constraint
OLIVGR	0.7006	0.7006
PARI	0.3273	0.3179

Finally, Figure 8 provides the summary (median, 3rd quartile and max) of the error terms for the NUTS2 regions of France after including environmental constraints and using *mnlogit*. It can be seen that, while median and 3rd quartile for all the regions and all the crops were close to zero, there still were several outliers.

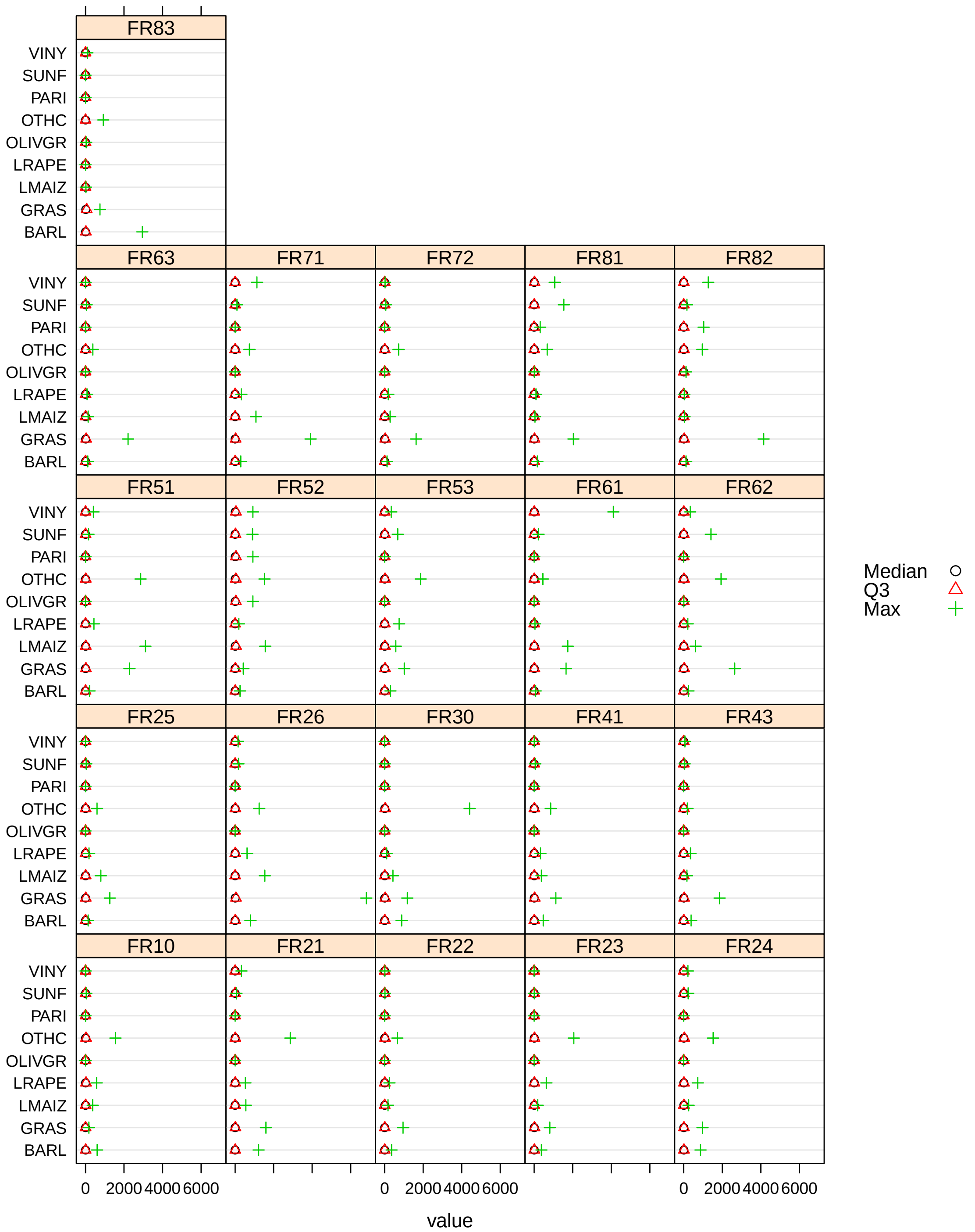

Figure 8. Summaries (median, 3rd quartile and maximum) of the error terms for the NUTS2 regions of France. Units in ha

However, the presence of such extreme predictions was improved for some crops in FR71 after restricting them with environmental constraints (Figure 9A and Figure 9B, without including environmental constraints and including them, respectively). This pattern is clear for all the land-uses except for vineyards and grasslands, although the former corrected some of them.

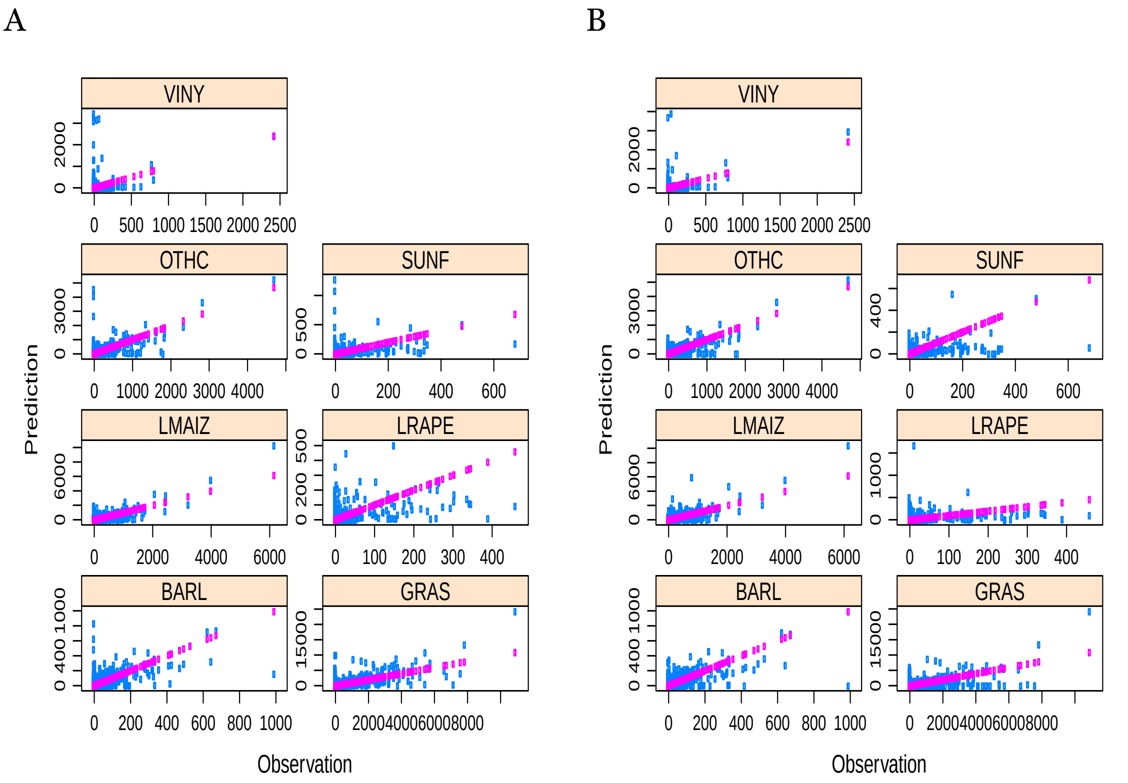

Figure 9. Scatter plots of the predicted and the observed data for FR71 (A) without including environmental constraints and (B) including them. Blue points are the scatter plots of observations versus predictions and pink points are the plot of observations versus observations. Units in ha

2.4 Conclusions

The main objective of this study was to improve the LUD model, both regarding its performance and its quality of the predictions, to contribute to the general objective of having an updated and improved land use share map to be implemented in CAPRI.

Regarding the improvement of its performance, the new developments included in the Step 1 of the model reduced considerably the process time. Although the Step 1 is the more resource consuming, some improvements could still be implemented specially in the Step 3, adapting it to work with *data.tables* and finding alternative R functions/packages more resource-efficient.

With regard to the improvement of the quality of the results, the *mnlogit* function for the estimation of multinomial logit models, together with the new procedure to select optimal bandwidths, give as good results for frequent crops as the original LUD model, or even a bit better for some crops. On the other hand, also the less-frequent crops show some improvements thanks to the inclusion of the environmental constraints to the crops/land-use predictions. In addition, the presence of outliers in the predictions has been reduced as well, although some of them can still be found.

After including some of the possible improvements pointed out by Lamboni et al. (2016) and others, the main objective of the study has been achieved. However, there are still misspredictions of the model that might be due to different reasons. On the one hand, exists uncertainty on both the data used to fit the model (from FSS), especially for non-frequent crops, and on the data to validate its predictions (LPIS). The use of other data to validate the model could give the possibility to check the results for other crops that now are not available. On the other hand, fitting one model for each smaller sub-regions instead of one for each NUTS2 might also significantly increase the accuracy of the model. This could be done, however, only if data at a finer scale than at NUTS3 level becomes available.

3 IMPROVING DISAGGREGATION OF YIELD AND FARM INPUTS IN CAPRI

Authors: Markus Kempen and Adrian Leip

3.1 Introduction

Organic and mineral fertilizer application rates are a highly relevant factor for environmental impacts of agricultural production as they drive realized crop yields and nutrient surpluses, and consequently the whole nutrient and carbon cycle in agriculture.

The CAPRI disaggregation procedure of inputs and outputs from (statistical) data at Nuts2 level to a spatial layer (HSMU or HSU) generally combines two steps:

(1) prior expectation on inputs and outputs coming from various sources (e.g. crop simulation models, statistical analysis) and

(2) a reconciliation procedure minimizing differences from the prior expectations while achieving consistency with the more aggregate Nuts2 data.

The crop yield estimation combines different types of a priori information in a Bayesian estimation framework (HDP, Highest Posterior Density, Heckelei et al., 2008) to derive simultaneously spatially explicit yield estimates and irrigation shares per crop (Kempen et al., 2005). Regression models based on FAO and FSS data on irrigation are used to forecast the irrigated share per crop and HSMU. Prior information on yields is taken from crop model results on potential and water limited yield (MARS-CGMS model).

After the consistent disaggregation of crops yields, the application rates of organic and mineral fertilizer rates are determined. Crop uptakes are derived from yields assuming (so far) a linear dependency of yield and fertilizer application. Organic fertilizer rates are derived from estimated, spatially explicit animal numbers. Following, expected mineral fertilizer application rates are calculated to "fill up" the expected total needs. At sub-regional level, the organic and inorganic application rates per crop are defined as to recover in average the ones at regional level. Hence over (or under) fertilization of crops is taken into account in the final result.

The findings in the following chapters often suggest an update of the prior information, while the design of the reconciliation procedure can be maintained.

3.2 Crop yield

3.2.1 Analysis of current methodology

The current CAPRI methodology uses yield potential and water limited yield potential from a crop simulation model (applied by MARS unit around 2005) as prior information for a data consistent disaggregation at Nuts2 level. Following current literature (e.g. Grassini et al., 2015), yield potential (Yp) is defined as the yield of an adapted crop cultivar as determined by

- solar radiation,
- temperature,
- carbon dioxide,
- and genetic traits that govern length of growing period, light interception by the crop canopy and its conversion to biomass, and partition of biomass to the harvestable organs water (van Ittersum and Rabbinge, 1997).

Water-limited yield potential (Yw) is determined by these previous factors and also by water supply amount and distribution during the crop growth period and field and soil properties that affect soil water availability such as slope, plant-available soil water holding capacity, and depth of the root zone water (Lobell et al., 2009; van Ittersum and Rabbinge, 1997; van Ittersum et al., 2013).

For a specific location and year, the crop yield gap (Yg) is defined as the difference between Yp (irrigated systems) or Yw (rainfed) and average actual farm yield (Ya).

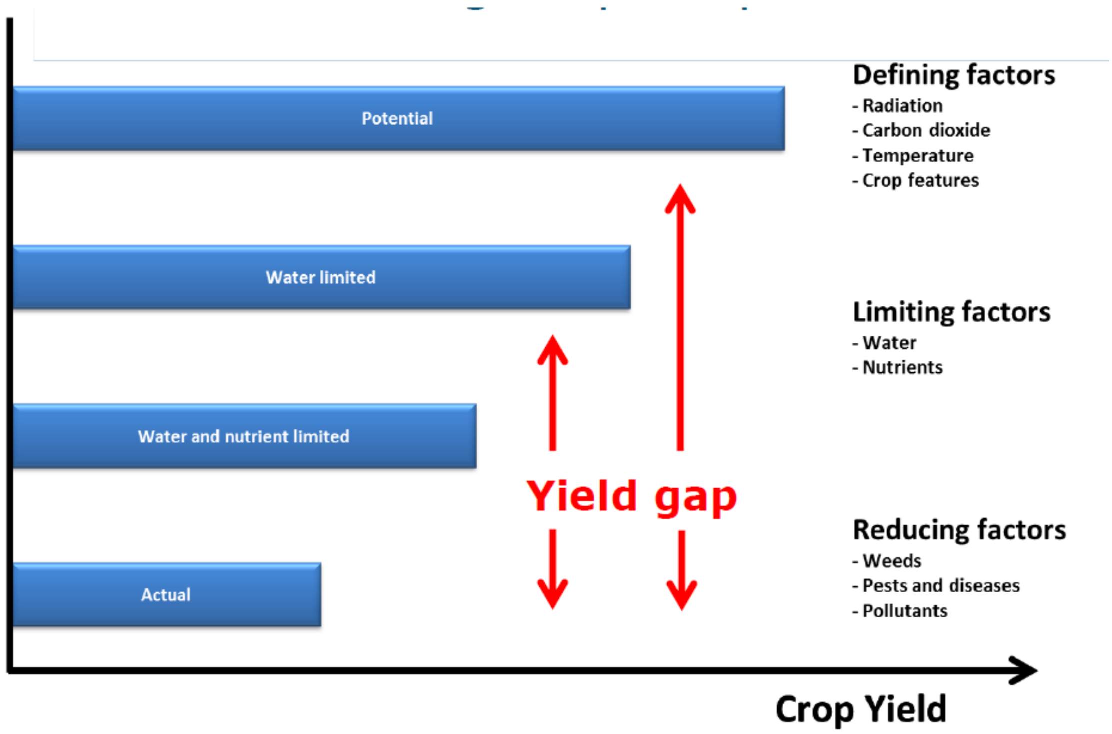

Figure 10. Relation of yield potential, water limited yield potential, actual yield and yield gap. Adapted from van Ittersum et al. *(2013)*.

Several methodologies have been proposed and applied to estimate Yp and Yw and subsequently Yg. van Ittersum et al. *(2013)* compared several methodologies and concluded that the application of crop growth models allows for the most robust estimation of Yp and Yw. Consequently, the CAPRI disaggregation procedure, using yield potential and water limited yield potential from a crop growth model as prior information, can be seen as “state of the art”. However, validation of CAPRI results for Germany reveal some inconsistencies of actual results and the “yield gap” concept.

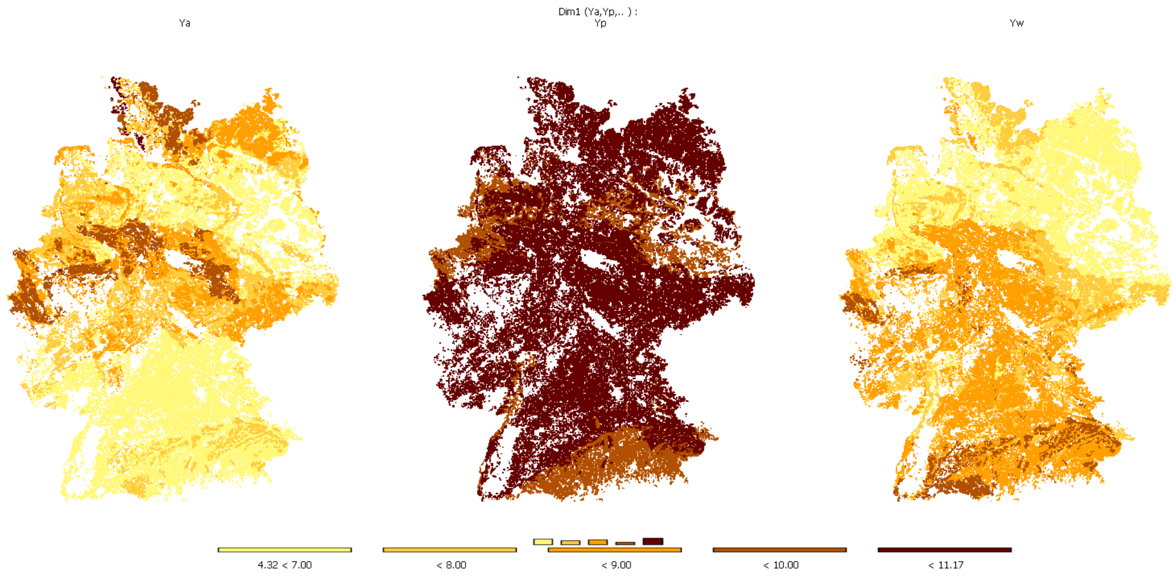

Figure 11. Actual yield, yield potential and water limited yield potential in the CAPRI base year data 2008 for soft wheat.

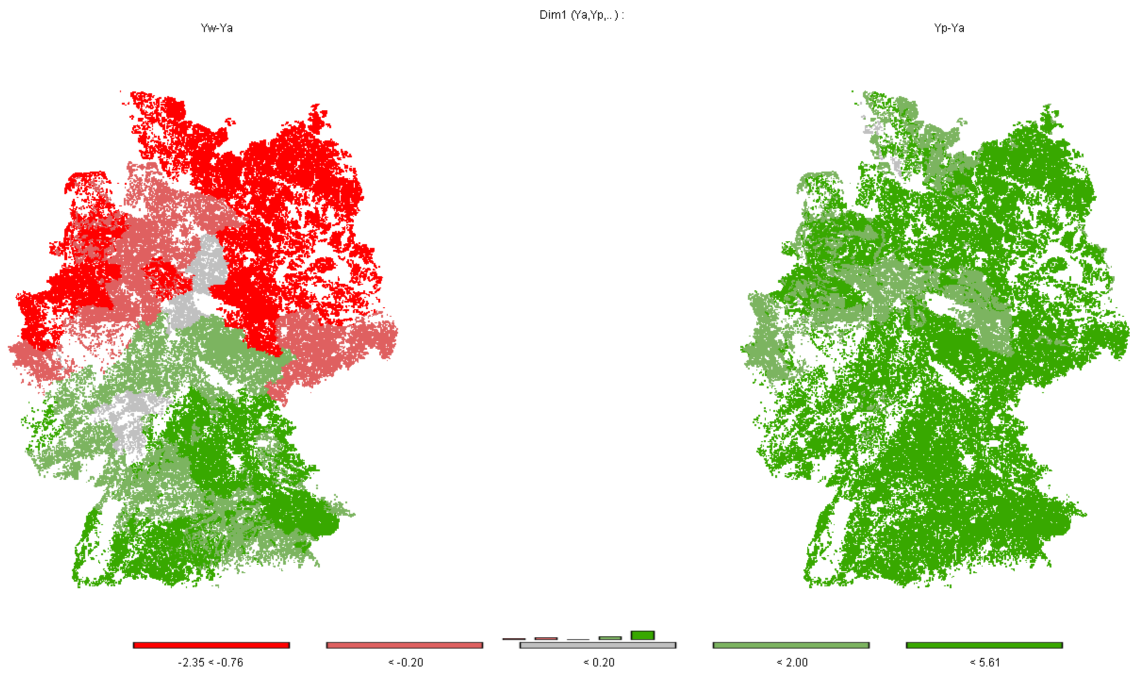

Figure 12. Yield gap compared to water limited yield potential and yield potential in the CAPRI base year data 2008 for soft wheat.

The actual yield Ya stems from a disaggregation procedure using simulated yield (MARS) at HSMU level and actual yield at Nuts2 coming from statistical data. In Germany cereals are typically cultivated as rain fed crops, so comparison can focus on Ya and Yw. The spatial pattern of Ya and Yw differ significantly at country level. While the actual yield tends to be higher in the northern part of

Germany, the water limited yield potential is supposed to be higher in the southern part. Within Nuts2, the pattern of MARS is recovered, but "scaled" to match the Nuts2 totals. This can result sometime in abrupt changes in HSMU yields along administrative boarders (see e.g. the northers boarder of Bavaria).

The contradictious data for Ya and Yw result in negative yield gaps in larger parts of northern Germany. As the statistical data on administrative regions influences the final result of Ya in the HSMUs significantly the result of Ya might be reliable for the analysis currently done with the spatial CAPRI data. As the statistical data on Nuts2 can be assumed to be reliable, the discrepancies might stem from the (outdated) simulations run of the crop growth model.

Figure 13 compares the actual CAPRI results with more recent MARS-CGMS yields as used during the FRAGARIA contract. The shortcomings described before are not visible in this comparison.

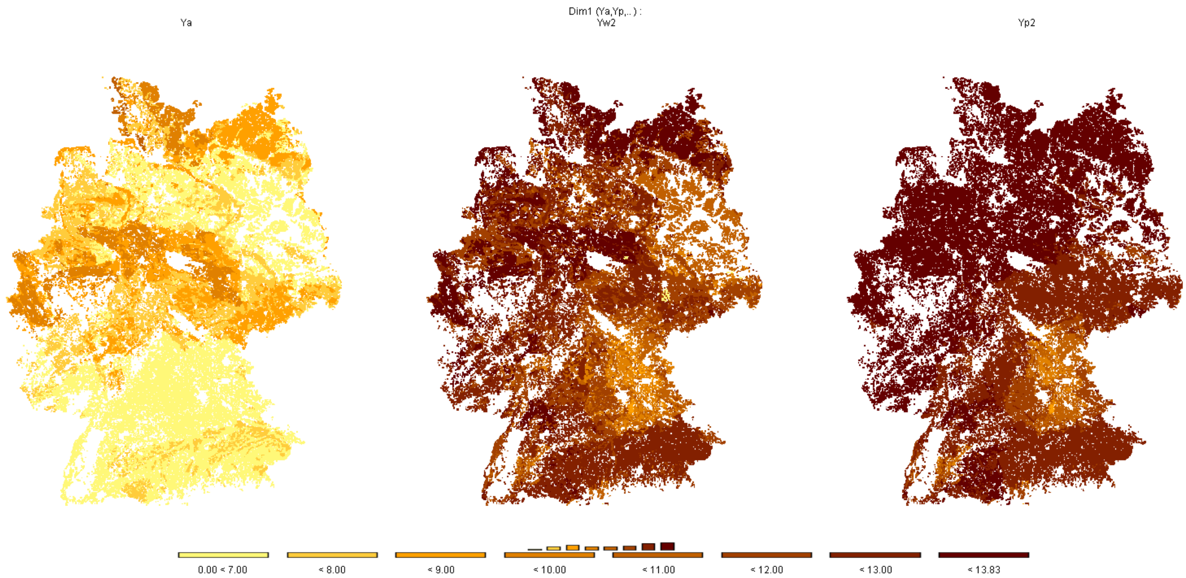

Figure 13. Actual yield, yield potential and water limited yield potential in the CAPRI base year data 2008 and MARS-CGMS yields used in the FRAGARIA service contract.

3.2.2 Updates

The prior information coming from crop simulation models is updated using the MARS-CGMS simulated yields used during the FRAGARIA contract (Pérez-Soba et al., 2015).

The general approach using simulated crops yield in in line with recommendations found in literature. The quality of simulated crop yields should be regularly updated. Maybe it could be helpful to have time series of crops yields. The historic data in the CAPRI data base could then be linked to the actual whether is those years, while for future projections a “average” whether might be preferable.

3.2.3 Results

The calculation of prior expectations of crop yield in spatial units is based now on crop model simulation results as used in the Fragaria service contract and regional yield gap averages coming from Zimmermann and Latka (2017). Technically the CAPRI model can switch between the updated and the standard calculation of priors.

The prior expectations differ significantly between the two specifications since Zimmermann and Latka (2017) predict that realized is in a range of 60-90% of the potential/water limited yield. As the high posterior density (HPD) downscaling procedure consolidates the priors for spatial units with the same average realized yield for administrative regions the difference in the final, data consistent spatial yield is nonetheless relatively small.

In the updated version the final result is close to the prior expectation, which is transparently calculated based on improved plant growth modelling and empirical studies. When the final result differs significantly from the prior expectation (as it did before), specification of objective function and constraints in the HPD framework significantly influence the final result. The decisions on the model specification were partially arbitrary and the calculations of the solver are to some extend a “black box”. Replacing a “by chance” good working consolidation procedure by a transparent, scientifically based prior calculation is the core improvement of the spatial yield data base in CAPRI

3.3 Linking fertilizer and other inputs to the yield gap

3.3.1 Analysis of current methodology

According to literature the yield gap depends on natural conditions. Van Bussel (2015) deals with upscaling yields from sample regions to national level. More observations of yield gap in different natural conditions reduce error in predicting yield gap at national level. These findings might be useful “vice versa” when downscaling from Nuts2 to HSMU. Currently the disaggregation procedure tends to adjust the HSMU priors by the same percentage value to meet the observed Nuts2 totals.

In SUSFANS it is aimed to regress actual yield (or yield gap) of FADN farms on natural conditions. FADN data is used in combination with climate and location data to determine a production function for actual yield depending on potential yield and inputs (see Figure 14) (Zimmermann and Latka, 2017)

Concept – Frontier analysis

Estimation of production functions → Stochastic Frontier Analysis (SFA)

$$q_i = \underbrace{\exp(\beta_0 + \beta_1 \ln x_i)}_{\text{Deterministic component}} \times \underbrace{\exp(v_i)}_{\text{Noise}} \times \underbrace{\exp(-u_i)}_{\text{Inefficiency}}$$

$$\ln q_i = \beta_0 + \beta_1 \ln x_i + v_i - u_i$$

$$\text{where} \quad u_i = \delta z_i$$

Figure 14. Concept of the frontier analysis in SUSFANS (Zimmermann and Latka, 2017)

3.3.2 Updates

Estimated coefficients were applied to the natural conditions in the HSMU resulting in estimates of the yield gap. Following, the prior for each HSMU is calculated as (water limited) yield potential minus expected yield gap. The natural conditions (temperature, radiation, precipitation) used by Zimmermann (2017) show a significant statistical impact on the yield gap at European scale. When disaggregating yields from Nuts2 to spatial layers, the variation of these natural conditions (and following the expected yield gap) among spatial units is rather small.

In the CAPRI disaggregation procedure the required fertilizer and other inputs (e.g. plant protection) are generally assumed to be proportional to the crop yield. In case of N fertilizers, a more sophisticated model accounts for effects of soil types and share of manure.

In case of fertilizer application, the aggregate statistical data often reports inputs that are higher than the actual requirements of the crops. In this case the disaggregation procedure would result (more or less) in a uniform percentage increase of fertilizer input per crop and ha. According to the literature on "yield gap", a yield gap is a consequence of limitation of nutrition, pests and diseases.

Following, a high yield gap would indicate that the inputs by the farmer (fertilizer, pesticides) are low. Consequently, no oversupply of fertilizer is expected. When the actual yield comes close to the (water limited) yield potential, there will be no limitation of nutrients and oversupply is possible and more likely. However, no evidence for this can be found in literature so far.

Fertilizer and plant protection inputs are explanatory variables (xi). In case of significant estimates for the corresponding coefficients, this production function is then be used to calculate the expected input use depending on actual yield and yield potential in the spatial unit.

Another issue concerning the disaggregation of inputs is the transport of manure. The transport of manure between Nuts2 regions is actually not considered in the CAPRI data. Partially data on manure transport is available for specific regions (see Table 7). The manure flow from Netherlands to North Rhine Westphalia is probably the most important flow between Nuts regions in Europe. It can be seen that some communes (Kleve, Viersen, Heinsberg) have significant imports of manure. However, it has to be discussed whether these local transport of manure plays a role at European scale.

The environmental legislation in several countries limits the application of manure per ha (e.g. at 170kg N from manure in Nitrate vulnerable zones). This is so far only partially taken into account when transport of manure is allowed to neighboring HSMU. A fixed upper limit was not included before, but is likely that the control- and sanction systems of “cross compliance” enforces that these rules are observed.

Table 7. Manure import from Netherlands to North Rhine Westphalia. (taken from https://www.umwelt.nrw.de/fileadmin/redaktion/4_guelleimporte_niederlande.pdf)

Kreis		Wirtschaftsdünger		Stickstoff		Phosphat	
		t	%	kg	%	kg	%
111	Düsseldorf	2.051	0,14	19.837	0,16	19.219	0,23
112	Duisburg	5.373	0,38	36.489	0,29	28.136	0,33
114	Krefeld	3.387	0,24	24.760	0,20	14.236	0,17
116	Mönchengladbach	50.432	3,54	392.152	3,13	252.110	3,00
117	Mülheim	240	0,02	2.343	0,02	1.273	0,02
122	Solingen	25	0,002	173	0,00	103	0,001
124	Wuppertal	211	0,01	1.472	0,01	997	0,01
154	Kleve	239.085	16,80	1.957.799	15,64	1.274.869	15,16
158	Mettmann	4.607	0,32	70.758	0,57	66.266	0,79
162	Rhein-Kreis Neuss	107.173	7,53	822.323	6,57	583.784	6,94
166	Viersen	197.316	13,86	1.377.644	11,01	886.536	10,54
170	Wesel	89.367	6,28	675.589	5,40	434.790	5,17
	Reg.-Bez. Düsseldorf	699.268	49,1	5.381.339	43,0	3.562.319	42,3
315	Köln	16.170	1,14	157.662	1,26	100.202	1,19
316	Leverkusen	1.290	0,09	9.695	0,08	7.317	0,09
334	Aachen	32.350	2,27	361.730	2,89	265.454	3,16
358	Düren	124.965	8,78	1.071.846	8,56	695.414	8,27
362	Rhein-Erft-Kreis	133.775	9,40	1.073.983	8,58	640.411	7,61
366	Euskirchen	57.024	4,01	1.082.427	8,65	908.160	10,80
370	Heinsberg	297.571	20,91	2.215.523	17,70	1.483.844	17,64
374	Oberbergischer Kreis	515	0,04	3.239	0,03	2.145	0,03
378	Rheinisch-Bergischer Kreis	1.444	0,10	10.741	0,09	6.816	0,08
382	Rhein-Sieg-Kreis	10.198	0,72	141.009	1,13	89.978	1,07
	Reg.-Bez. Köln	675.302	47,4	6.127.855	49,0	4.199.741	49,9
512	Bottrop	460	0,03	3.173	0,03	1.887	0,02
554	Borken	292	0,02	8.811	0,07	7.922	0,09
558	Coesfeld	84	0,01	579	0,00	345	0,00
562	Recklinghausen	650	0,05	12.129	0,10	10.992	0,13
566	Steinfurt	21	0,001	145	0,00	86	0,001
570	Warendorf	1.221	0,09	41.649	0,33	28.156	0,33
	Reg.-Bez. Münster	2.728	0,2	66.486	0,5	49.388	0,6
711	Bielefeld	983	0,07	21.480	0,17	18.434	0,22
758	Herford	2.928	0,21	99.962	0,80	45.948	0,55
762	Höxter	1.701	0,12	39.800	0,32	44.246	0,53
766	Lippe	4.821	0,34	143.624	1,15	98.637	1,17
770	Minden-Lübbecke	48	0,003	331	0,00	197	0,00
774	Paderborn	2.167	0,15	41.487	0,33	36.559	0,43
	Reg.-Bez. Detmold	12.648	0,9	346.684	2,8	244.021	2,9
954	Ennepe-Ruhr-Kreis	4.263	0,30	37.859	0,30	35.527	0,42
958	Hochsauerlandkreis	500	0,04	14.726	0,12	12.531	0,15
962	Märkischer Kreis	15.502	1,09	106.754	0,85	71.360	0,85
966	Olpe	582	0,04	20.091	0,16	14.892	0,18
974	Soest	12.139	0,85	414.029	3,31	220.913	2,63
978	Unna	300	0,02	2.071	0,02	1.231	0,01
	Reg.-Bez. Arnsberg	33.285	2,3	595.530	4,8	356.454	4,2
	NRW	1.423.231	100	12.517.894	100	8.411.923	100

The disaggregation procedure of fertilizer inputs is extended to consider various aspects when determining the expected input per HSMU and crop:

(1) relation of yield gap and fertilizer input,
(2) transport of manure and
(3) environmental legislation.

This makes the calculation of the prior expectation more complex than is most other cases, where the prior is usually a crude value coming from one other source. Hence, a “pre model” is developed to account for all these aspect when setting the prior expectation. Other inputs will still be distributed proportional to the actual yield since no better (quantifiable) methodology was found so far. The “final” disaggregation step can then be unchanged.

3.3.3 Results

Zimmermann and Latka (2017) used singe farm data from the FADN to estimate effects of climate and management data on the yield of soft wheat (SWHE), barley (BARL), maize (MAIZ), potato (POTA) and as EU-wide (where available) case study soybean (SOYA). The explanatory variables considered for determining the frontier yields are a trend (YEAR), precipitation (PREC), radiation (RAD) and temperature (TEMP). The management variables considered are economic farm size (ESU), fertilizer expenditure per ha (FERT) and plant protection expenditure per ha (PROTEC). Due to limited data availability, just the fertilizer and plant protection expenditure per total Utilized Agricultural Area (UAA) per farm could be used.

$$ln\left(y_{jtr}\right) = \beta_0 + \beta_1 ln\left(YEAR_{jtr}\right) + \beta_2 ln\left(PREC_{jtr}\right) + \beta_3 ln\left(RAD_{jtr}\right)$$
$$+\beta_4 ln\left(TEMP_{jtr}\right) + v_{jtr} - i_{jtr}$$

with ,

$$i_{jtr} = \delta_1\left(ESU_{jtr}\right) + \delta_2\left(FERT_{jtr}\right) + \delta_3\left(PROTEC_{jtr}\right)$$

Given the formulation of the estimation model the coefficients of the variables can be interpreted as follows:

1. Frontier variables (log-log): Coefficients give percent change of yield for a percent change in frontier variables (YEAR, PREC, RAD, TEMP).
2. Management variables (log-linear): Coefficients give percent change of yield for a unit change of management variables (ESU, FERT, PROTEC); in the formulation as used here, management variables are subtracted, i.e. the sign of the actual impact is exactly the opposite.

The estimates of the coefficients of the management variables δ2 and δ3 can be used to predict inputs of fertilizer and plant protection in spatial units. In the following results are shown in Table 8 for the example of δ2 (FERT) and barley (BARL).

Table 8. Calculation of marginal efficiency of additional fertilizer input

		from estimation model			effects of increasing FERT by 1€		
region	crop	Yield (% of max.)	δ2	FERT (mean)	FERT (% diff)	Yield (% diff.)	required FERT as % of additional FERT
50	BARL	87,1%	-0,0022	96,3	1,0%	0,2%	21,4%
60	BARL	87,1%	-0,0062	85,3	1,2%	0,6%	53,1%
90	BARL	88,9%	-0,0081	90,9	1,1%	0,8%	73,2%
113	BARL	79,4%	-0,0036	114,8	0,9%	0,4%	41,6%
115	BARL	76,9%	-0,0017	79,7	1,3%	0,2%	13,6%
116	BARL	83,3%	-0,0089	73,9	1,4%	0,9%	65,7%
136	BARL	79,4%	-0,0036	97,2	1,0%	0,4%	35,2%
153	BARL	77,6%	-0,0033	80,6	1,2%	0,3%	26,6%
163	BARL	80,8%	-0,0025	113,5	0,9%	0,2%	28,0%
182	BARL	70,6%	-0,0033	111,0	0,9%	0,3%	37,2%
183	BARL	74,7%	-0,0075	91,6	1,1%	0,8%	68,9%
184	BARL	74,2%	-0,0058	67,1	1,5%	0,6%	38,6%
192	BARL	73,6%	-0,0077	99,6	1,0%	0,8%	76,3%
193	BARL	75,5%	-0,0089	72,0	1,4%	0,9%	64,3%
201	BARL	66,1%	-0,0121	48,0	2,1%	1,2%	58,0%
222	BARL	74,3%	-0,0034	97,5	1,0%	0,3%	32,9%
244	BARL	75,7%	-0,0026	149,2	0,7%	0,3%	39,1%
250	BARL	86,7%	-0,0202	32,2	3,1%	2,0%	65,2%
260	BARL	74,5%	-0,0064	79,1	1,3%	0,6%	51,0%
270	BARL	71,0%	-0,0099	55,0	1,8%	1,0%	54,5%
281	BARL	80,1%	-0,0046	96,0	1,0%	0,5%	44,3%
282	BARL	75,7%	-0,0095	65,2	1,5%	0,9%	61,7%
291	BARL	80,3%	-0,0051	51,3	1,9%	0,5%	26,3%
292	BARL	80,9%	-0,0020	87,6	1,1%	0,2%	17,4%
301	BARL	83,2%	-0,0041	42,3	2,4%	0,4%	17,5%
302	BARL	81,7%	-0,0022	72,1	1,4%	0,2%	15,7%
303	BARL	76,3%	-0,0081	70,6	1,4%	0,8%	57,1%
340	BARL	78,8%	-0,0031	124,3	0,8%	0,3%	38,0%
360	BARL	79,5%	-0,0019	120,4	0,8%	0,2%	23,3%
380	BARL	82,1%	-0,0031	126,2	0,8%	0,3%	39,7%
411	BARL	84,3%	-0,0085	101,2	1,0%	0,8%	85,7%
413	BARL	81,9%	-0,0050	94,5	1,1%	0,5%	47,5%
421	BARL	81,5%	-0,0022	99,7	1,0%	0,2%	21,8%
431	BARL	84,2%	-0,0021	96,8	1,0%	0,2%	20,1%
515	BARL	83,2%	-0,0050	141,2	0,7%	0,5%	71,2%
530	BARL	66,6%	-0,0154	59,2	1,7%	1,5%	91,0%
540	BARL	60,7%	-0,0074	51,3	1,9%	0,7%	37,9%
545	BARL	72,5%	-0,0029	82,4	1,2%	0,3%	23,9%
550	BARL	89,4%	-0,0135	39,1	2,6%	1,3%	52,6%
555	BARL	74,7%	-0,0109	46,9	2,1%	1,1%	51,2%
560	BARL	70,9%	-0,0072	52,0	1,9%	0,7%	37,3%

Source: Zimmermann and Latka (2017), own calculation for selectet regions

The estimated coefficient δ2, estimated yield and observed fertilizer expenditures are taken from Zimmermann and Latka (2017). Then the percentage change in expenditures for fertilizer and yield are calculated for an increase of fertilizer expenditures by 1 Euro. The percentage change in expenditures for fertilizers should be equivalent to the percentage change in applied quantity. As the requirements for nutrients are proportional to the yield, the percentage changes in yield and fertilizer expenditures should be identical under ideal conditions. However, in practice probably not all fertilizer applied by the farmer is finally taken up by the plant. Hence we define the ratio of percentage change in expenditures and percentage change of yield as the "efficiency" of additional applied fertilizer.

Conceptually this efficiency should not be below 0% or above 100%. As the calculation is based on a regression analysis, a few outliers are not in this range and were removed. The average over all regions the marginal efficiency of additional fertilizer is about 40% at the currently observed realized yield. It should be noted, that this is not the overall efficiency of fertilizer applications.

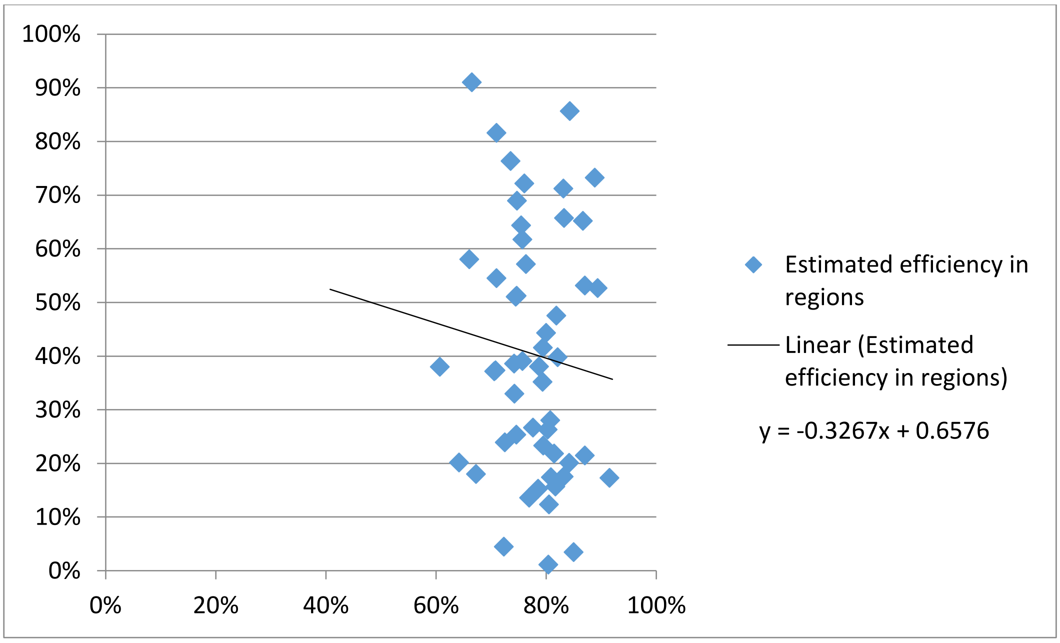

Figure 15. Efficiency of additional fertilizer (y axis) compared to realized yield (x axis) – own calculation

Figure 15 shows the efficiency of additional fertilizer compared to the realized yield in selected European regions. A linear trend suggests that the efficiency of additional fertilizer decreases with increasing yield. The calculations based on empirical analysis are in line with our theoretical considerations on the relation between yield gap and fertilizer application (see section 3.3.2).

The regression results are used to calculate the prior expectation of fertilizer use in spatial mapping units. This calculation of prior expectations is so far realized as an alternative option in the CAPRI downscaling procedure.

3.4 Conclusions

The concept of downscaling yield and fertilizer inputs, using a high posterior density estimator to line up spatially explicit priors with administrative data, can be maintained. The methodology of deriving prior information can be improved using new findings in research. Crop growth models calculating potential and water limited yield have been improved and a new data set is included now. Results of yield gap estimations at regional scale can be used to derive expected realized yields at grid level. The distribution of fertilizer can also be changed based on yield gap estimates. In the previous methodology surplus of fertilizer was allocated proportional to the nutrient needs of the crops. In the updated allocation procedure fertilizer surplus is more likely to occur for crops showing a low yield gap. In case of high yield gaps the fertilizer gifts should be close to the crop needs. The rules for estimating manure applications are revised and account now for the limits given by environmental legislations. Transport of manure across mapping units is allowed. In administrative regions with very high animal numbers transport of manure to neighboring regions is made possible.

4 IMPROVING ESTIMATES OF SOIL LOSS CALCULATED WITH CAPRI

Authors: Maria Bielza (JRC)

4.1 Introduction

Soil erosion, as characterised by the widely used RUSLE equation, is determined by two groups of factors. The first group can be identified as biophysical factors and is represented as RKLS, which is the product of the factors for rainfall-runoff (R), soil erodibility (K) and length and steepness of slope (LS)). This product had been computed at 1Km cell EU27 grid by JRC/SOIL_ACTION. It had been averaged at HSMU level and included in CAPRI in 2012.

The second group consists of the cover, management and support practices factors (C and P factors). The C-factor can be subdivided in Ccrop/cover x Cmanagement:

$$Cfactor = C_{crop/cover} \cdot C_{management}$$

These factors had been improved during the imap-8 project. This approach, however was overestimating soil erosion in fallow land in non-arid regions, where GAEC does not require a green cover but there is a natural green cover during summer due to climatic conditions.

4.2 Methods

The $C_{Msummercover}^{MS,fall}$ factor with a value of 0.5 should be extended to all non-arid regions. An additional improvement was therefore necessary. This was done based on the aridity index (P/PET). The aridity index is calculated as the coefficient between the annual precipitation and the annual ETP. A region is considered arid when the index is lower or equal to 0.5.

For this reason, the arid regions, on which $C_{Msummercover}^{MS,fall}$ = 0.94 will be only a few Mediterranean regions (shown on Table 9). Where half of the region area is arid, an intermediate value of 0.72 has been applied. For the rest of the regions, the Csummercoverfactor takes the value of 0.5.

Table 9 Values for CMsummercover

Code	Region name	CMsummercover
CY00	CY00 - Kypros	0.94
EL11	EL11 - Anatoliki Makedonia, Thraki	0.72
EL12	EL12 - Kentriki Makedonia	0.94
EL14	EL14 - Thessalia	0.94
EL24	EL24 - Sterea Ellada	0.72
EL30	EL30 - Attiki	0.94
EL41	EL41 - Voreio Aigaio	0.94
EL42	EL42 - Notio Aigaio	0.94
EL43	EL43 - Kriti	0.94
ES22	ES22 – Com. Foral de Navarra	0.72
ES23	ES23 - La Rioja	0.94
ES24	ES24 - Aragón	0.94
ES30	ES30 - Comunidad de Madrid	0.94
ES41	ES41 - Castilla y León	0.94
ES42	ES42 - Castilla-la Mancha	0.94
ES43	ES43 - Extremadura	0.94
ES51	ES51 - Cataluña	0.94
ES52	ES52 - Comunidad Valenciana	0.94
ES53	ES53 - Illes Balears	0.94
ES61	ES61 - Andalucía	0.94
ES62	ES62 - Región de Murcia	0.94
FR83	FR83 - Corse	0.94
ITF4 (IT91)	ITF4 - Puglia	0.72
ITG1 (ITA0)	ITG1 - Sicilia	0.94
ITG2 (ITB0)	ITG2 - Sardegna	0.94
MT00	MT00 - Malta	0.94
PT15	PT15 - Algarve	0.94
PT18	PT18 - Alentejo	0.94
RO22 (RO02)	RO22 - Sud-Est	0.94

4.3 Validation of the CAPRI average soil erosion indicator

The values obtained with CAPRI for potential soil erosion are shown on Table 10. The CAPRI results show a UAA weighted average value for EU-28 close to 5.3 t/ha. For comparison, other average values for soil erosion found in the literature are:

- 17 t/ha/yr for arable soils in Europe, based on plot data (Pimentel et al., 1995)
- 11.1 t/ha/yr of average value of soil erosion by water for Europe (Yang et al., 2003)
- 8.8 t/ha/yr (for bare soils ca. 32 t/ha/yr for the Mediterranean zone and ca. 17 t/ha/yr for the rest of Europe) (Cerdan et al., 2006)
- Mean rates in Europe are ca. 3-40 t/ha/yr for actual soil erosion (Verheijen et al., 2009)
- The main publication for soil erosion in Europe (Panagos et al., 2015c) found a mean erosion rate in EU-28 of 2.46 t/hayr for the potentially erosion-prone land cover in EU-28[7]. However, it needs to be taken into account that CAPRI values refer only to agricultural land. We have calculated the average value for agricultural land uses from the Panagos et al. (2015c) results at 100 m resolution, using as mask Corine Land Cover-2XX, and obtained an average value of 3.13 t/ha (see Table 11).

Table 10 Values of average soil erosion from CAPRI (t/ha/yr) for EU-28 minus HR, CY and MT.

Region level	Simulation scenario	WEIGHTED AVERAGE	MAX	MIN	MEDIAN	AVERAGE
NUTS2	Base year 2008	5.3	112.6	0.09	2.3	7.1
	CAPRI baseline 2025	5.3	119.1	0.09	2.3	7.3
HSMU	Base year 2008	5.3	1792.2	0.00	2.4	11.1
	CAPRI baseline 2025	5.3	1115.0	0.00	2.3	10.2

Table 11 Values of average soil erosion from Panagos et al. (2015) in t/ha/yr, for EU-28, masked with CLC-2XX.

Region level	Reference year	WEIGHTED AVERAGE	MAX	MIN	MEDIAN	AVERAGE
NUTS2	2010	3.13	25.7	0.14	2.2	3.4
100 m2	2010	3.13	325	0.00		

According to the results by Panagos et al. (2015c) the CAPRI mean is too high and should be revised. The problem with the maximum erosion values is still more important. While according to Panagos et al. (2015c) the maximum at

[7] The average rate of soil loss falls to 2.22 t/hayr if the non-erosion prone areas are included in the statistical analysis. For arable land, Panagos et al. (2015) found an average value of 2.62 t/hayr.

NUTS2 region is 25.7 (for IT93-Calabria), the CAPRI 2008 maximum is 112.6 (IT13-Liguria), followed by 76.1 in (AT21-Carinthia). At high geographical resolution (HSMU), oscillations are even bigger, the maximum values almost reaching 1800 t/ha in a number of HSMUs in Slovenia in the base year. However, we have observed that these correspond to areas mainly of forest and shrub land, with no or very little agricultural area (those HSMU do not appear in 2025 as they do not have UAA). The maximum in 2025 corresponds to 1,115.0 t/ha for a small HSMU in Liguria with vineyards and citrus fruits, followed by four HSMUs above 800 t/ha: three in Slovenia, with fodder maize and grain maize and one in Tuscany with sunflower). These values are not possible, given that according to Maetens et al. (2012), maximum erosion values found on European experimental plots are 325 t/ha annually. Moreover, Panagos et al. (2015c) have imposed this value as the maximum soil loss rate to avoid model outliers (they had found less than 0.001% of pixels above this value).

An explanation for the CAPRI high values is that the soil erosion indicator has been calculated for small patches of crop fields on high slope areas where there is a majority of non-agricultural land uses (e.g. the highest values correspond to HSUs in Slovenia mainly covered with forest, shrub land and other uses, and around 1% of the land has potatoes or maize). In the mountain regions in Austria, Liguria, Slovenia, only small plots have low-soil protecting crops and practices (e.g. vineyards), and these usually not in the most steep slopes or with the protection of terraces. For this reason, we suggest to follow the example of Panagos et al. (2015c) and cap the indicator at a maximum of 325 t/hayr.

With this change the SiSLOS_perHa indicator value in Slovenia in the base year has changed from 44.67 t/hayr to 42.90 t/hayr (the value by Panagos et al. (2015c) is 14.61).

4.4 Soil erosion thresholds

4.4.1 The CAPRI shares indicator

Apart from the use of average soil loss values, it is necessary to define which values can be considered acceptable and which can pose problems for the environment. Besides, using the average value at NUTS2 region might not show if there are areas with high risk erosion. For this reason, a set of three subdicators were designed in CAPRI, to show the area referred to relevant thresholds. The area was classified according to the following soil erosion classes:

1) <0.5 tons/ha

2) <5.0 tons/ha

3) >5.0 tons/ha

Given the lack of information for interpretation of these thresholds, a small literature review has been performed to understand if they are relevant or should be revised.

4.4.2 Literature review: soil formation, soil tolerance and soil erosion thresholds

There has been much discussion in the literature about thresholds above which soil erosion should be regarded as a serious problem. This has given rise to the concept of 'tolerable' rates of soil erosion that should be based on reliable estimates of natural rates of soil formation. A good illustration of the problem of soil formation and soil tolerance is offered by Hurni (1983). Hurni states that "the destabilizing effects of soil erosion resulting from agricultural activities cannot be evaluated by soil erosion process studies alone. The rate of soil formation must be known so that maximum tolerable soil loss for cultivated slopes can be assessed. Two case studies, one in the Ethiopian high mountains and one in the mountains of Northern Thailand, are used to demonstrate the role of the various factors that influence stability. Despite moderate erosion rates, the Simen ecosystems in Ethiopia have suffered the greatest ecological damage through soil degradation processes. This threatens the food security of the present-day inhabitants. Reasons for this can be found in the inaccurate perception of soil erosion as a problem, and in the low soil loss tolerances of this high mountain environment. In contrast, high soil erosion rates on cultivated swiddens[8] in Huai Thung Choa, Northern Thailand, have been clearly recognized by the local people; this is reflected in their shifting to swidden cultivation practices. Recent trends of reduced fallow periods, however, have resulted in accelerated soil degradation. Soil formation rates, nevertheless, are high enough for recovery once fields are abandoned. Use of a soil degradation ratio, defined as the soil loss divided by the soil loss tolerance[9], was found to be

[8] Agricultural practice consisting on planting one (or more) year of crop after burning the natural vegetation, followed by several years of fallow, in which the regenaration of natural vegetation and soil takes place.

[9] Soil tolerance (T) is defined here as the mean annual soil loss that a given cultivated slope can tolerate. It can be calculated as T=max(F1, min(F2, F3), being (F1) the formation rate of A horizon by accumulation on the soil surface; (F2)

a practical method of evaluating the destabilizing potential of soil erosion in agricultural ecosystems". In Ethiopian mountains, soil formation rates during fallow periods take values of 0.1-0.4 t/ha for elevations above 3500m, 2 t/ha for elevations between 3000-3500, 4t/ha for elevations between 2500-3000 m and 6 t/ha for elevations between 2000-2500 m (Hurni, 1983). The fact that high elevations are associated to lower tolerance levels leads to expect low tolerance in altitude areas, which are precisely areas where soil erosion can be more intense. However, soil formation and soil tolerance are also dependent on climatic conditions (rainfall, temperatures). Always according to Hurni (1983), in Thailand mountains, potential soil losses estimated from the RUSLE equation register 120 t/ha/yr on 40% slope. However, the climate allows soil formation rates of 10-12 t/ha/yr. This rate, in combination with the swiddens system consisting of 1 year crop + 10 years of fallow, allows the soil to recover.

Values of soil loss tolerance in the United States are: 2 to 11 t/ha/yr (McCormack and Young, 1981); 5 to 12 (Schertz, 1983); 2.5-12.3 t/ha/yr (NCRS-USDA, 2002), which is equivalent to ≈0.2–1 mm/yr of erosion (assuming a soil bulk density of 1,200 kg/m3). The suggested tolerance level for most soils in Ontario is 6.7 t/ha/yr or less (Stone and Hilborn, 2012). Researchers have expressed concern that T values themselves are set substantially higher than soil production rates, because of political and economic considerations (Larson, 1981).

In order to apply these concepts, it would be necessary to be able to estimate soil tolerance (soil formation rates) at Pan-European level, from climatic and soil type information. Verheijen et al. (2009) have tried to find tolerable erosion rates in Europe. According to them: "Ideally, soil formation models (e.g. Hoosbeek and Bryant, 1992 and Minasny and McBratney, 2001) would have been developed and validated to such an extent that for any soil type, under any land use, soil management practice, in any region, accurate estimates of soil formation rates could be derived. However, fundamental scientific knowledge on soil formation processes is still insufficient at present to support the use of mechanistic soil formation models for establishing tolerable rates of soil erosion in the context of environmental protection. Therefore, the most useful contribution that science can make to the policy process would be to arrive at a consensus on mean rates of soil formation and soil erosion. Considering soil formation rates by both weathering and dust deposition, it is estimated that for the majority of soil forming factors in most European situations, soil formation

the formation rate of A horizon from B horizon (from biological activity); (F3) the formation rate of B horizon by weathering of parent material into smaller particles.

rates probably range from ca. 0.3 to 1.4 t/ha/yr. Although the current agreement on these values seems relatively strong, how the variation within the range is spatially distributed across Europe and how this may be affected by climate, land use and land management change in the future remains largely unexplored. Future differentiation of soil formation rates for soil–land use–climate combinations is needed, and quantitative pedogenesis modelling (e.g. Hoosbeek and Bryant, 1992 and Minasny and McBratney, 2001) may provide an appropriate methodology." (Verheijen et al., 2009).

Verheijen et al., 2009 report some of the factors affecting soil formation. For example, chemical weathering can be expected to increase where precipitation increases, particularly where the parent material is well draining. Soils formed in limestone or granitic lithology are reported to have formation rates towards the smaller part of the range, although the body of evidence is relatively small and more experimental research is urgently needed into soil formation rates for these lithologies, since they cover a substantial area in Europe. Soil formation by sedimentation in water is only significant in the floodplains of large river systems. Dust deposition is more intense in the Mediterranean areas close to the Sahel region. Goudie and Middleton (2001) report very low values in the Alps (0.002-0.004 t/ha/yr), low in Central France (0.01) and North-Eastern Spain (0.05), high in other Mediterranean regions (0.1 in Corsica and Southern Sardinia, 0.1-0.2 in Crete) and still higher in Pyrinees and some areas of the Aegean See (>0.3).

Egli et al. (2013) have studied soil formation on Alpine soils. They have found that soil production and related tolerable erosion rates (i.e. 50–90 % of the soil production rate) are a strong function of time. Average soil production rate in alpine areas for relatively old soils (>10–18 kyr) is between 0.54 (±014) and 1.13 (±0.30) t/ha/yr, for young soils (>1–10 kyr) between 1.19 (±0.44) and 2.48 (±0.91) t/ha/yr and for very young soils (≤1 kyr) between 4.15 (±2.42) and 8.81 (±5.20) t/ha/yr.[10] Due to the fast glacier retreat after the Late Glacial Maximum (LGM), the surface age of most areas (having meadows) in the subalpine to the alpine range of the European Alps is near 10 to 18 kyr (Ivy-Ochs et al., 2004). European soils have in most cases an age of >10 kyr (Alewell et al., 2015).

According to Eurostat (2016): «Soil formation processes and rates differ substantially throughout Europe. In some cases, rates of soil erosion larger than

[10] According to the same authors, measured recent soil erosion rates in alpine areas at intensively used slopes range from 6 to 30 t/ha/yr, while average catchment values for the Urseren Valley (using the model USLE plus soil loss due to landslides) resulted in an overall loss of 1.8 t/ha/yr (0.18 mm/yr). These values considerably exceed production rates of the soils.

1 t/ha/yr are regarded as tolerable from the wider perspective of society as a whole, for example for economic considerations or the preservation of soil functions. In Switzerland, the threshold tolerated for soil erosion is generally 1 tonne per hectare per year, though this rate can be increased to 2 tonnes per hectare per year for some soil types (Schaub and Prasuhn, 1998). More recently, Verheijen et al. (2009) estimated the average soil formation rate in Europe to 1.4 tonne per ha per year which is much lower than the soil loss rate. In general, losses above 1 tonne per hectare per year are generally considered as irreversible. Nevertheless, there may be a need to propose different thresholds of rates of soil erosion that are tolerable in different parts of Europe. However, this aspect needs further elaboration».

According to Ecologic Institute and SERI (2010), which contains a wide literature review: «In conclusion, an exercise of setting regional or even local or site specific threshold levels of erosion would be valuable and more research is definitely needed in this regard. However the literature review and expert interviews carried out for this study show that a threshold value of 1 t /ha/ year , based on a comparison with mean natural soil formation rates, is generally accepted as a tolerable rate of soil erosion in our current socio-economic context. Site specific exceptions above and below this threshold are of course realistic and necessary in some cases. Indicators of a danger zone or of a zone of tolerable erosion are thus also related to soil formation indicators. (…) Recent research (Verheijen et al., 2009) leads to the conclusion that overall soil formation ranges from ca. 0.3 to 1.4 t /ha/ year. Therefore accepting a threshold of 1 t/ha/y, i.e. above the estimated average rate of soil formation, can critically be interpreted as a rather pragmatic approach to soil protection in order also to maintain current demand for soil productivity in Europe, e.g. food production».

According to Morgan (2009), the tolerable soil loss threshold (T) formulated for Mediterranean environments is 10 T ha-1 yr-1. Borrelli et al. (2016) have calculated potential soil erosion rates in arable land at high resolution, finding that «the predicted soil erosion rates are still about twice the tolerable soil loss threshold (T) formulated for Mediterranean environments (10 T ha−1 year−1, Morgan, 2009)».

4.4.3 Conclusions

From the review performed in the previous section, it can be established that the task of setting individual thresholds or tolerance levels for all EU regions would be too complex at the moment. It would be more reasonable to set some thresholds at EU-level or at large climatic areas (e.g arid or Mediterranean areas against rest of Europe).

According to literature findings the soil formation rates in Europe are on average around 1.4 t/ha annually; in some cases 2t/ha. Thus 76% of the area in Europe is below this sustainable rate. According to a personal communication by Panagos, for scientific purposes, the following classes can be considered:

0-1 t/ha per year: Low rate
1– 2 : Sustainable rates
2-5: Medium erosion rate
5-10: High erosion rate
>10: Severe soil erosion rate

To limit it to 3 classes, mainly oriented for policy makers, the following classes can be used (Panagos et al., 2016):

0-2: Sustainable
2-10: Medium-High
>10 :Severe

5 CONCLUSIONS

For the quantification of environmental sustainability, at least two aggregate variables require assessment at farm level or high spatial resolution: agricultural land use diversity, and soil erosion. We performed a thorough review of the procedures used to calculate environmental indicators at high spatial resolution with the CAPRI model. Need for update and improvement was found for different stages of the procedure:

- Quantification of a priori crop shares
- Disaggregation of crop yield and farm inputs
- Quantification of potential loss of soil through water erosion

This report presents the results of feasibility studies on the possibility of improvements in those three stages.

For the quantification of a priori crop shares, an update of the LUD model (Lamboni et al., 2016) is proposing several improvements on model performance and model quality, which overall were shown to generate better prediction both of 'frequent' and 'less frequent' crops; the latter mainly through the introduction of environmental suitability ranges.

The disaggregation of crop yields can be improved through an update of the prior crop yield estimates obtained from a crop model, but in particular also through using results from the SUSFANS yield gap analysis. These results allow that the fertilization rates beyond crop needs can be 'disconnected' from crop yield and linked to the yield gap estimate at high spatial resolution, so that over-fertilization is estimated at those units where nutrient input does not limit crop growth.

Potential soil losses by water erosion was found to be over-estimated for fallow land in non-arid regions. An update of the calculation procedure is proposed, accompanied with a comprehensive literature review not only on soil erosion in Europe for validating CAPRI estimates, but also on soil erosion thresholds that can be used to differentiate sustainable erosion rates from medium and severe erosion rates.

6 REFERENCES

Alewell, C., Egli, M., Meusburger, K., 2015. An attempt to estimate tolerable soil erosion rates by matching soil formation with denudation in Alpine grasslands. J. Soils Sediments 15, 1383–1399. doi: http://dx.doi.org/10.1007/s11368-014-0920-6

Borrelli, P., Paustian, K., Panagos, P., Jones, A., Schütt, B., Lugato, E., 2016. Effect of Good Agricultural and Environmental Conditions on erosion and soil organic carbon balance: A national case study. Land use policy 50, 408–421. doi: http://dx.doi.org/10.1016/j.landusepol.2015.09.033

Britz, W., Leip, A., 2009. Development of marginal emission factors for N losses from agricultural soils with the DNDC-CAPRI meta-model. Agric. Ecosyst. Environ. 133, 267–279. doi: http://dx.doi.org/10.1016/j.agee.2009.04.026

Britz, W., Verburg, P.H., Leip, A., 2011. Modelling of land cover and agricultural change in Europe: Combining the CLUE and CAPRI-Spat approaches. Agric. Ecosyst. Environ. 142, 40–50. doi: http://dx.doi.org/10.1016/j.agee.2010.03.008

Britz, W., Witzke, P., 2014. CAPRI model documentation. Bonn, Germany.

Cantelaube, P., Carles, M., 2015. Le registre parcellaire graphique: des données géographiques pour décrire la couverture du sol agricole, in: Cahier Des Techniques de l'INRA, Special Issue GéoExpé. pp. 58–64.

Cerdan, O., Poesen, J., Govers, G., Saby, N., Le Bissonnais, Y., Gobin, A., Vacca, A., Quinton, J., Auerswald, K., Klik, A., Kwaad, F.F.P.M., Roxo, M.J., 2006. Sheet and Rill Erosion, in: Soil Erosion in Europe. John Wiley & Sons, Ltd, Chichester, UK, pp. 501–513. doi: http://dx.doi.org/10.1002/0470859202.ch38

Chakir, R., 2009. Spatial Downscaling of Agricultural Land-Use Data: An Econometric Approach Using Cross Entropy. Land Econ. 85, 238–251. doi: http://dx.doi.org/10.3368/le.85.2.238

Dowle, M., Srinivasa, A., 2016. data.table: Extension of `data.frame`. R package version 1.10.0 [WWW Document]. URL https://cran.r-project.org/package=data.table. Available at: https://cran.r-project.org/package=data.table.

EC, 2003a. Farm structure 1999/2000 survey. Office for Official Publication of the European Communities, Luxembourg.

EC, 2003b. The Lucas survey. European statisticians monitor territory, Theme 5:

Agriculture and fisheries. Office for Official Publications of the European Communities, Luxembourg.

EC, 2006. Development of agri-environmental indicators for monitoring the integration of environmental concerns into the common agricultural policy. Communication from the Commission to the Council and the European Parliament. COM(2006) 508 final. Commission of the European Communities, Brussels.

EC, 2008. Regulation (EC) No 1166/2008 of the European Parliament and of the Council of 19 November 2008 on farm structure surveys and the survey on agricultural production methods and repealing Council Regulation (EEC) No 571/88. Off. J. Eur. Union L321, 14–34.

Egli, M., Dahms, D., Norton, K., 2013. Soil production rates on silicate parent material in high-mountains: different approaches-different results?, in: EGU General Assembly Conference Abstracts. p. 8405.

European Commission, 2011. Commission staff working paper. Executive summary of the impact assessment: Common Agricultural Policy towards 2020. European Commission, Brussels.

Eurostat, 2016. Agri-environmental indicator – Soil erosion. Eurostat Statistics Explained [WWW Document]. URL http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_soil_erosion (accessed 3.10.17). Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_soil_erosion.

Götz, C., Zimmermann, A., Leip, A., van Zanten, H., Hornborg, S., Ziegler, F., 2017. Database on farm-level production and sustainability indices for assessing sustainable diets. Deliverable 4.7 of the SUSFANS project H2020 / SFS-19-2014: Sustainable food and nutrition security through evidence based EU agro-food policy, GA no. 633692.

Grassini, P., van Bussel, L.G.J., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., van Ittersum, M.K., Cassman, K.G., 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. F. Crop. Res. 177, 49–63. doi: http://dx.doi.org/10.1016/j.fcr.2015.03.004

Hasan, A., Wang, Z., Mahani, A.S., 2016. Fast Estimation of Multinomial Logit Models: R Package mnlogit. J. Stat. Softw. 75. doi: http://dx.doi.org/10.18637/jss.v075.i03

Heckelei, T., Mittelhammer, R., Jansson, T., 2008. A bayesian alternative to generalized Cross Entropy solutions for underdetermined econometric models. Food Resour. Econ. - Discuss. Pap. 2.

Hurni, H., 1983. Soil Erosion and Soil Formation in Agricultural Ecosystems: Ethiopia and Northern Thailand. Mt. Res. Dev. 3, 131. doi: http://dx.doi.org/10.2307/3672994

Ivy-Ochs, S., Sch�fer, J., Kubik, P.W., Synal, H.-A., Schl�chter, C., 2004. Timing of deglaciation on the northern Alpine foreland (Switzerland). Eclogae Geol. Helv. 97, 47–55. doi: http://dx.doi.org/10.1007/s00015-004-1110-0

Kempen, M., 2013. EU wide analysis of the Common Agricultural Policy using spatially disaggregated data.

Kempen, M., Heckelei, T., Britz, W., 2005. An econometric approach for spatial disaggregation of crop production in the EU. Working paper presented at the EAAE Seminar, Parma, 3-5 February 2005. University of Bonn, Institute for Agricultural policy, market Research and Economic sociology.

Kempeneers, P., McInerney, D., Sedano, F., Gallego, J., Strobl, P., Kay, S., Korhonen, K.T., San-Miguel-Ayanz, J., 2013. Accuracy Assessment of a Remote Sensing-Based, Pan-European Forest Cover Map Using Multi-Country National Forest Inventory Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 54–65. doi: http://dx.doi.org/10.1109/JSTARS.2012.2236079

Lamboni, M., Koeble, R., Leip, A., 2016. Multi-scale land-use disaggregation modelling: Concept and application to EU countries. Environ. Model. Softw. 82, 183–217. doi: http://dx.doi.org/10.1016/j.envsoft.2016.04.028

Larson, W.E., 1981. Protecting the soil resource base. J. Soil Water Conserv. 36, 13–16.

Leip, a., Marchi, G., Koeble, R., Kempen, M., Britz, W., Li, C., 2008. Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe. Biogeosciences 5, 73–94. doi: http://dx.doi.org/10.5194/bg-5-73-2008

Leip, A., Bielza, M., Bulgheroni, C., Ciaian, P., Lamboni, M., Paracchini, M., Terres, J., Weiss, F., Witzke, H., 2015. Spatially Explicit Evaluation of the Agri - environmental Impact of CAP, in: International Conference of Agricultural Economists; ICAE2015.

Leip, A., Busto, M., Winiwarter, W., 2011a. Developing spatially stratified N 2O emission factors for Europe. Environ. Pollut. 159, 3223–3232. doi: http://dx.doi.org/10.1016/j.envpol.2010.11.024

Leip, A., Wattenbach, M., Reuter, H.I., Koeble, R., Balkovic, J., Skalsky, R., Obersteiner, M., 2011b. A new data infrastructure for European terrestrial ecosystem modelling - Development of Unified Spatial Characterisation Identifier for Europe (uscie).

Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. Annu. Rev. Environ. Resour. 34, 179–204. doi: http://dx.doi.org/10.1146/annurev.environ.041008.093740

Maetens, W., Vanmaercke, M., Poesen, J., Jankauskas, B., Jankauskiene, G., Ionita, I., 2012. Effects of land use on annual runoff and soil loss in Europe and the Mediterranean: A meta-analysis of plot data. Prog. Phys. Geogr. 36, 599–653. doi: http://dx.doi.org/10.1177/0309133312451303

McCormack, D.E., Young, K.K., 1981. Technical and societal implications of soil loss tolerance, in: Soil Conservation Problems and Prospects:[proceedings of Conservation 80, the International Conference on Soil Conservation, Held at the National College of Agricultural Engineering, Silsoe, Beford, UK, 21st-25th July, 1980]/edited by RPC Morgan.

Morgan, R.P.C., 2009. Soil erosion and conservation. John Wiley & Sons.

NCRS-USDA, 2002. Technical Guide to RUSLE use in Michigan, NRCS-USDA State Office of Michigan. Available at: http://www.iwr.msu.edu/rusle/tvalue.htm.

Orlandini, S., van der Goot, E., 2003. Technical description of interpolation and processing of meteorological data in CGMS. Available at: http://mars.jrc.ec.europa.eu/mars/Bulletins-Publications/Technical-description-of-interpolation-and-processing-of-meteorological-data-in-CGMS.

Panagos, P., Borrelli, P., Meusburger, K., Alewell, C., Lugato, E., Montanarella, L., 2015a. Estimating the soil erosion cover-management factor at the European scale. Land use policy 48, 38–50. doi: http://dx.doi.org/10.1016/j.landusepol.2015.05.021

Panagos, P., Borrelli, P., Meusburger, K., van der Zanden, E.H., Poesen, J., Alewell, C., 2015b. Modelling the effect of support practices (P-factor) on the reduction of soil erosion by water at European scale. Environ. Sci. Policy 51, 23–34. doi: http://dx.doi.org/10.1016/j.envsci.2015.03.012

Panagos, P., Borrelli, P., Poesen, J., Ballabio, C., Lugato, E., Meusburger, K., Montanarella, L., Alewell, C., 2015c. The new assessment of soil loss by water erosion in Europe. Environ. Sci. Policy 54, 438–447. doi: http://dx.doi.org/10.1016/j.envsci.2015.08.012

Pekkarinen, A., Reithmaier, L., Strobl, P., 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. ISPRS J. Photogramm. Remote Sens. 64, 171–183. doi: http://dx.doi.org/10.1016/j.isprsjprs.2008.09.004

Pérez-Soba, M., Elbersen, B.S., Kempen, M., Braat, L., Staristky, I., Wijngaart, R. van, Kaphengst, T., Andersen, E., Germer, L., Smith, L., Rega, C., 2015. Agricultural biomass as provisioning ecosystem service: quantification of energy flows. Contract No. JRC 97764. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC97764.

Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., others, 1995. Environmental and economic costs of soil erosion and conservation benefits. Science (80-.). 267, 1117.

Rutten, M., Achterbosch, T.J., de Boer, I.J.M., Cuaresma, J.C., Geleijnse, J.M., Havlík, P., Heckelei, T., Ingram, J., Leip, A., Marette, S., van Meijl, H., Soler, L.-G., Swinnen, J., van't Veer, P., Vervoort, J., Zimmermann, A., Zimmermann, K.L., Zurek, M., 2016a. Metrics, models and foresight for European sustainable food and nutrition security: The vision of the SUSFANS project. Agric. Syst. doi: http://dx.doi.org/10.1016/j.agsy.2016.10.014

Rutten, M., Zimmermann, A., Havlik, P., Leip, A., Heckelei, T., Achterbosch, T., 2016b. Modelling Sustainability and Nutrition in Long Run Analyses of the EU Agri-Food system: Work plan for the SUSFANS Toolbox. Deliverable 9.1 of the SUSFANS project H2020 / SFS-19-2014: Sustainable food and nutrition security through evidence based EU agro-f. Available at: https://public.3.basecamp.com/p/ychkkCtZpeo3B74LnNH4nAAB.

Schaub, D., Prasuhn, V., 1998. A map on soil erosion on arable land as a planning tool for sustainable land use in Switzerland. Adv. GeoEcology 31, 161–168.

Schertz, D.L., 1983. The basis for soil loss tolerances. J. Soil Water Conserv. 38, 10–14.

Siebert, S., Hoogeveen, J., Frenken, K., 2007. Irrigation in Africa, Europe and Latin America: Update of the digital global map of irrigation areas to Version 4. Frankfirt Hydrol. Pap. 5. Available at: http://publikationen.ub.uni-frankfurt.de/rewrite/index/id/type/opus3-id/value/3846.

Stone, R.P., Hilborn, D., 2012. Universal Soil Loss Equation Factsheet. Order No. 00-001. Ontario Ministry of Agriculture, Food and Rural Affairs. Available at: http://www.omafra.gov.on.ca/english/engineer/facts/12-051.htm#4.

van Bussel, L.G.J., Grassini, P., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., Saito, K., Cassman, K.G., van Ittersum, M.K., 2015.

From field to atlas: Upscaling of location-specific yield gap estimates. F. Crop. Res. 177, 98–108. doi: http://dx.doi.org/10.1016/j.fcr.2015.03.005

van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance—A review. F. Crop. Res. 143, 4–17. doi: http://dx.doi.org/10.1016/j.fcr.2012.09.009

van Ittersum, M.K., Rabbinge, R., 1997. Concepts in production ecology for analysis and quantification of agricultural input-output combinations. F. Crop. Res. 52, 197–208. doi: http://dx.doi.org/10.1016/S0378-4290(97)00037-3

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S Fourth edition by. World 53, 86. doi: http://dx.doi.org/10.2307/2685660

Verheijen, F.G.A., Jones, R.J.A., Rickson, R.J., Smith, C.J., 2009. Tolerable versus actual soil erosion rates in Europe. Earth-Science Rev. 94, 23–38. doi: http://dx.doi.org/10.1016/j.earscirev.2009.02.003

Weissteiner, C.J., García-Feced, C., Paracchini, M.L., 2016. A new view on EU agricultural landscapes: Quantifying patchiness to assess farmland heterogeneity. Ecol. Indic. 61, 317–327. doi: http://dx.doi.org/10.1016/j.ecolind.2015.09.032

Yang, D., Kanae, S., Oki, T., Koike, T., Musiake, K., 2003. Global potential soil erosion with reference to land use and climate changes. Hydrol. Process. 17, 2913–2928. doi: http://dx.doi.org/10.1002/hyp.1441

Zimmermann, A., Latka, C., 2017. The drivers of crop production at regional level in the EU: an econometric analysis. Deliverable 4.5 of the SUSFANS project H2020 / SFS-19-2014: Sustainable food and nutrition security through evidence based EU agro-food policy, GA no. 633692.

Zurek, M., Ingram, J., Rutten, M., Zimmermann, A., Leip, A., et al., 2016. Metrics to assess Sustainable Food and Nutrition Security in the EU -a progress report – SUSFANS deliverable D1.2.

7 ANNEXES

7.1 Annex 1 – Spatial layers available for the CAPRI model

The minimum spatial unit of the HSU data set is the 1km2 grid cell. To facilitate the link between additional spatial information required for different purposes (e.g. the mapping of spatial environmental indicators) all data sets are linked at a resolution of 1km2 via a "Unified Spatial Data Characterization Identifier (USCIE). The underlying 1km2 grid builds on the recommendations from the 1st Workshop on European Reference Grids concerning grid alignment and projection. (References: D2.8.I.2 INSPIRE Specification on Geographical Grid Systems – Guidelines http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.1.pdf, Short Proceedings of the Workshop on European Reference Grids, Ispra, 27-29 October 2003[11]).

Currently the following spatial information is linked via the USCIE:

- **MARS meteorological 25km2 grid** (EC JRC AGRI4CAST (2012): Interpolated Meteorological Data. European Commission Joint Research Centre, Institute for Environment and Sustainability, Monitoring Agricultural Resources (MARS) Unit). http://agri4cast.jrc.ec.europa.eu/DataPortal/Index.aspx.
- **LUCAS survey data for the years 2001, 2003, 2006, 2007, 2009 and 2012.** While data until 2007 is unpublished the 2009 and 2012 survey data can be accessed at http://ec.europa.eu/eurostat/web/lucas/data/database)
- **Irrigated area (% of the USCIE grid cell) in the year 2000** based on Siebert, S., Hoogeveen, J., & Frenken, K. (2006). Irrigation in Africa, Europe and Latin America. Update of the Digital Global Map of Irrigation Areas to Version 4. Frankfurt Hydrology Paper 5. Institute of Physical Geography, University of Frankfurt (Main), Germany and Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/nr/water/aquastat/irrigationmap/index50.stm Map Version 4.0.

[11] http://eusoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf

- **Land use/cover data.** Data source: European Topic Centre on Spatial Information and Analysis (2012): CORINE Land Cover 2006 raster data at 250m resolution - version 16 (04/2012), http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-2. European Topic Centre on Land Use and Spatial Information (2011): CORINE Land Cover 2000 raster data at 250m resolution - version 15 (08/2011)) CLC2000 data has been applied only in the case of Greece. http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster-1. European Space Agency (2008): GlobCover Land Cover v2 2008 database. European Space Agency GlobCover Project, led by MEDIAS-France. 2008), http://due.esrin.esa.int/page_globcover.php. GLC2006 data has been applied in the USCIE/HSU2 domain for areas where CLC2000/2006 was not available.
- **HSMU data set.** Leip, A., Marchi, G., Koeble, R., Kempen, M., Britz, W. & Li, C. (2008). Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe. Biogeosciences, 5 (1), 73–94. Copernicus Publications. doi:10.5194/bg-5-73-2008. Data available at ftp://mars.jrc.ec.europa.eu/Afoludata/Public/all_datasets.html Dataset 251.
- **Altitude and slope.** Jarvis, A., Reuter, H.-I., Nelson, A., Guevara, E., 2008. Hole-filed srtm for the globe version 4. Available from the CGIAR-CSI SRTM 90 m database. Available at http://srtm.csi.cgiar.org.
- **Forest cover (% of the 1km2 grid cell) for the years 2000 and 2006.** The forest cover is based on a high-resolution (25m) pan-European forest cover maps for the year 2000 and 2006. Kempeneers, P., Sedano, F., Seebach, L., Strobl, P., San-Miguel-Ayanz, J. (2011) Data Fusion of Different Spatial Resolution Remote Sensing Images Applied to Forest-Type Mapping. IEEE Transactions on Geoscience and Remote Sensing, 49 (12), pp. 4977-4986. http://forest.jrc.ec.europa.eu/download/data/forest-data-download/
- **Global Homogeneous Response Units.** Skalsky, Rastislav; Tarasovicová, Zuzana; Balkovic, Juraj; Schmid, Erwin; Fuchs, Michael; Moltchanova, Elena; Kindermann, Georg; Scholtz, Peter; McCallum, Ian (2012): Global Homogeneous Response Units. doi:10.1594/PANGAEA.775369

- **Annual dry and wet deposition for the years 2006 – 2010 of reduced and oxidized nitrogen as well as oxidized sulphur as mg(N,S) m^{-2}.** The original data at the 50km by 50km EMEP grid resolution was provided by D. Simpson (pers. comm. July 2014). Dry deposition is distinguished for 5 different land cover classes i.e. coniferous forest, deciduous forest, cropland, seminatural vegetation and water surfaces. As the land use/cover information in the EMEP data set is different from the land use/cover data set included in the USCIE data set (s. above), the dry deposition has been re-mapped to the land use/cover of the USCIE data set. However the total dry deposition of a specific compound within the EMEP grid cell remained un-changed.

7.2 Annex 2 – Improvement of soil erosion estimates (imap 8)

7.2.1 Improvement of the Ccrop/cover factor

In the 2012 version of soil erosion indicators, the Cfactor was just based on Ccrop/cover while the effect of management via the Cmanagement factor was ignored. This factor had a single value per crop for all EU. In the current version, these values have been updated according to the literature. For permanent crops and grasslands a differentiation of the value per crop and country has been implemented, while for arable crops still one EU-wide value is used.

The main source used for the update of the C-factor is the data provided by the JRC/soil unit used for the article by Panagos et al. (2015a). For the crops where it was not available, the previous CAPRI value was used, with the following exceptions:

- For some permanent crops (APPL, CITR, OFRU, OLIV, TABO) the values per country by Panagos et al. (2015a) which were quite low (on average 0.22) have been proportionally increased in order to obtain a higher value, more in line with the 0.35 of Bosco and Rigo (2013) and de Vente et al. (2009). Also Rousseva et al. 2003 found a value of 0.42 for fruit trees in Bulgaria, still higher than the 0.39 obtained for Bulgaria. However, these factors should be reviewed.

- Oats: as this cereal is not cold resistant, it cannot be planted in autumn but in spring in northern countries. We assume therefore that oats is a winter cereal in southern countries and a spring crop for colder countries, and used the corresponding factors.
- MAIF (fodder maize): the value by Wall et al. (2002) was used, higher than that of grain maize because of its earlier harvest.
- OFAR (other fodder on arable land, which includes temporary grass, other silage cereals and other fodder plants such as alfalfa): a value of 0.10 was used for most countries, a higher value of 0.15 for Mediterranean countries (CY, ES, EL, IT, MT, PT).
- OCER (other cereals including millet, triticale, buckwheat, sorghum (except fodder sorghum) and summer cereal mix have been assigned a value of 0.28, similar to that of oilseeds, higher than those of winter cereals 0.20-0.22 and lower than that of dry pulses (0.32).
- GSET(Set aside obligatory used as grassland): the same CC factor as for intensive grassland (GRAI). CM=1.
- TSET(Set aside obligatory fast growing trees): CC=0.3 in order to have an intermediate value between GSET (0.09) and FALL and ISET (0.5). In this way, it is close to the value for nurseries (0.296).

7.2.2 Description of the Cmanagement and P factors

Recent research (2015a, 2015b) have estimated the C and P factors for all EU. Based on their results and in other literature, we next explore the possibilities of improvement of the soil erosion indicator by including also the Cmanagement and the P factors.

Following Panagos et al (2015a), the Cmanagement factor is the result of the product of three subfactors:

$$C_{Management} = C_{Mtillage} \cdot C_{Mresidues} \cdot C_{Mcatchcrop}$$

When the exact practices on each crop are not known, but the total area on which these practices are applied is available, the three components of the Cmanagement factor to be applied to arable crops can be calculated as follows (Panagos et al., 2015a):

$$C_{Mtillage} = \%arable_{conventional} \cdot 1.0 + \%arable_{conservation\ or\ ridge} \cdot 0.35 + \%arable_{no\ till} \cdot 0.25$$

$$C_{Mresidues} = (1 - \%arable_{residues}) \cdot 1.0 + \%arablecrops_{residues} \cdot 0.88$$

$$C_{Mcatchcrop} = (1 - \%arable_{catchcrop}) \cdot 1.0 + \%arable_{catchcrop} \cdot 0.80$$

Where $\%arable_x$ is the arable area under a certain management practice X on total arable crops area. Catch crops are crops sown specifically to protect bare soil in winter (and early spring) after the harvesting of summer crops.

The previous equations would apply only to arable crops. Land left fallow and setaside is a special case. Given that by definition there is no standard crop coverage during the summer, the management factor is also affected by the summer cover.

The literature does not distinguish specific management factors for land left fallow. Based on associations with other crops and practices, we propose to estimate it by adding to the standard management factors a factor CMsummercover describing the situation during the summer period. The determination of this factor depends on the share of fallow land being left rough during summer, or covered with crop residues or green cover:

$$C_{Mfallow} = C_{Mtillage} \cdot C_{Mresidues} \cdot C_{Mcatchcrop} \cdot C_{Msummercover}$$

$$C_{Msummercover} = \%fallow_{rough} \cdot 1.0 + \%fallow_{residues} \cdot 0.88 + \%fallow_{greencover} \cdot 0.5$$

In order to justify the values of the selected factors for fallowland, we propose three examples. If a fallow land was planted with rye in the winter which is left for the summer, its final Cfactor would be Cfactor= $C_{crop/cover}$x $C_{Mfallow}$ = $C_{crop/cover}$x $C_{Mtillage}$x $C_{Mresidues}$ x $C_{Mcatchcrop}$ x$C_{summercover}$=0.5 x1 x 1x 0.8 x 0.5 = 0.2, therefore equal to the Cfactor of rye (0.2), a cereal that is usually planted in autumn. If it was planted with grass not tilled in the early autumn and left like that for the whole year: Cfactor= C_C x $C_{Mtillage}$ x $C_{Mresidues}$ x $C_{Mcatchcrop}$ x $C_{Msummercover}$=0.5 x 0.25 x 1 x 0.80 x 0.5=0.5x0.1=0.05. According to the literature review, grasslands have Cfactors between 0.03 and 0.1, therefore this value for fallowland coverd with grass all th year can be considered acceptable. A last example: if the fallow or setaside area was left untouched after the previous year crop, it could be considered no till, with crop residues in winter and in summer, therefore Cfactor = C_C x $C_{Mtillage}$ x $C_{Mresidues}$ x $C_{Mcatchcrop}$ x $C_{Msummercover}$ = 0.5x 0.25x 1 x 0.88 x0.88 = 0.5x0.19 = 0.1.

The CMfactor values we have just described are collected in Table 12.

Table 12 Cmanagement factor values from the literature

	Management practice	C_M Factor
Tillage practice	Conventional tillage	1.0
	Conservation/ridge tillage (Reduced tillage)	0.35
	No till	0.25
Residues	Rough cultivated surface (fall ploughed or disced) (RS)	1
	Winter vegetal cover (including stubble and mulch) and spring plough (WSM)	0.88
Winter green cover (catch crops)	Rough cultivated surface (fall ploughed or disced) (RS)	1
	Winter green cover (WGC) (Ccover)	0.80
Summer cover (for fallow, use with a C_C=0.5)	Rough cultivated surface (fall ploughed or disced) (RS)	1
	Summer vegetal cover (including overwintered stubble and mulch) (SSM)	0.88(1)
	Summer green cover (SGC)	0.5(2)

Source: Panagos et al. (2015a) (1) Estimated by authors, see text. (2) Estimated from crops used for summer green cover (e.g. rye).

Following Panagos et al (2015b), the estimation of the support practices P-factor can be calculated as:

$$P = P_C \cdot P_{SW} \cdot P_{GM}$$

Where P_C is the contouring factor, P_{SW} the stone walls factor and P_{GM} the grass margins factor.

7.2.3 Actual implementation of CM in CAPRI

The objective has been to calculate the C_M subfactors when possible, otherwise to use the average values at region level from the literature. Given that Panagos et al. (2015a) have calculated the average $C_{management}$ factors for arable land at

NUTS2 level, and Panagos et al. (2015b) the P factors, their output values can be used in CAPRI (see Annex 1 with C_M factors for arable land[12] and Annex 2 with P factors). The average regional values have been used in all C_M subfactors except for $C_{Mcatchcrop}$ and $C_{Msummercover}$.

In CAPRI, data on farm practices were not available for endogenous implementation of C_M factor. However, since the CAPRI modelisation of greening measures, the catch crop winter cover has been endogenously included in the model. Therefore, we have endogenised the estimation of $C_{Mcatchcrop}$.

According to personal information by Gocht (2016), catch winter crop (CATC) can be activated in CAPRI when non-winter cover crops are in the crop rotation[13]. However, as we yet do not know which crops have been considered winter crops, we have applied it to arable crops area except fallowland and grassland. Anyway, this should not have an impact on the final value, as the total CATC surface to which $C_{Mcatchcrop}$ applies remains the same. Therefore, $C_{Mgreencover}$ is calculated from the share of catch crops area on total arable crops area (not fallow nor grassland). The values of $C_{Mtillage}$ and $C_{Mresidues}$, exogenous, have been taken from Panagos et al. (2015a) at NUTS2 regions.

Please note that CATC has not been disaggregated to HSMU units, therefore, all the C_M factors are available only at NUTS2 level. Different scenario impacts between HSMU cannot be attributed to the change in management practices but only to the change in crop distribution and thus the linked with the environmental factors in the RUSL equation.

A possible source of information on farming practices, which could be used for the implementation of C_M, is the GAEC features. In the GAEC features, some countries must apply certain soil conservation practices. This would allow implementing $C_{Mresidues}$, $C_{Mcatchcrop}$ and $C_{Msummercover}$ factors for arable crops and fallow. According to their GAEC, some countries require the application of the practices only on fallow land, or only on cultivated land, or only for certain types of farms, zones, or everywhere.

In the case of the application of GAEC, the formula would be:

[12] It is to be pointed out that the CM factors by Panagos et al. (2015) include also what we have called CMsummercover, the additional factor applied to fallow land.

[13] In the policy folder in dat there is a file called wintercover.gms. It defines the maximum shares of wintercover by activity (100 for grass). In the supply model optimization, the potential catch crop area is calculated endogenously based on those shares by farm type (or by NT2 region), which yields the maximum area available for CATC and hence for complying with the cropping obligations in the greening policy. The catch crops per crop would be an approximation, using the available area per crops and the total area of CATC.

$$C_{Mresidues-GAEC} \cdot = \begin{cases} 0.88 \; for \; arable \; land \; in \; the \; countries \; where \; GAEC \; requires \; vegetal \; cover \\ 1 \; for \; the \; rest \; of \; crops \;\; and \; countries \end{cases}$$

$$C_{Mcatchcrop-GAEC} \cdot = \begin{cases} 0.80 \; for \; arable \; land \; in \; the \; countries \; where \; GAEC \; requires \; catch \; crop \\ 1 \; for \; the \; rest \; of \; land \; cover \; and \; countries \end{cases}$$

$$C_{Msummercover-GAEC} \cdot = \begin{cases} 0.5 \; for \; fallow \; land \; in \; the \; countries \; where \; GAEC \; requires \; vegetal \; cover \\ 0.88 \; for \; fallow \; land \; in \; the \; countries \; where \; GAEC \; requires \; stubble \; or \; mulch \; cover \\ 1 \; for \; the \; rest \; of \; crops \; and \; countries \end{cases}$$

This implementation of the factors would be a proxy, given that some farms might be leaving farm residues in those countries where it is not compulsory, so in these countries the indicator would be underestimated. Moreover, in some countries it is compulsory only when the slope is above a certain threshold, so this should be combined with the slope of the HSMU.

In CAPRI, in spite of the lack of precision of this method, we have applied it only for the calculation of CMsummercover, given that there was no specific value for this factor by Panagos et al. (2015a). For fallow-land and set-aside (FALL, ISET and VSET), the management factor has been calculated as:

$$C_M^{fallow} = C_{Mtillage}^{arableN2} \cdot C_{Mresidues}^{arableN2} \cdot C_{Mcatchcrop}^{arableN2} \cdot C_{Msummercover}^{fallow}$$

Where:

$$C_{Msummercover}^{fallow} = \begin{cases} 0.5 \; for \; AT, DK, FI, FR, \; LT, \; NL \; (GAEC \; imposes \; a \; summer \; cover) \\ \\ 1 \; for \; the \; rest \; of \; the \; countries \end{cases}$$

And the other three factors being the average regional value by Panagos et al. (2015a).

Please note that this method can result in an overestimation of the Cmsummercover factor for other countries where we have assumed

CMsummercover=1 but the fallow soil or part of it could have a green cover during the summer.

7.2.4 Implementation of the P factor in CAPRI

The support practices factor P (soil conservation and prevention practices factor), has been subdivided in three subfactors by Panagos et al. (2015b): the contouring, the stone walls and the grass margins factors. Pcontouring is dependent on the slope, so the contour farming factor could potentially be calculated in function of the farm slope (for the countries where this GAEC measure is applied, and also for scenario analysis). The grass margins, and stone walls or terraces factors will be difficult to implement for lack of information.

Given the limited possibilities for the calculation of these factors, it was more advisable to use the values per region by Panagos et al. 2016b.