Technical Paper

Computation of a European Agricultural Land Use Map
–
Statistical Approach and Validation

Markus Kempen, Thomas Heckelei, Wolfgang Britz
Adrian Leip, Renate Koeble, Giulio Marchi

Preliminary manuscript which is not (yet) published in a professional journal. Comments and criticisms are welcome and should be sent to the author(s) directly. All citations need to be cleared with the corresponding author or the editor.

E-mail: Markus.Kempen@ilr.uni-bonn.de
Computation of a European Spatial Land Use Map

The underlying Statistical Procedures

Markus Kempen, Thomas Heckelei, Wolfgang Britz
Adrian Leip Renate Koeble, Giulio Marchi

Abstract
In this paper we describe a data-consistent procedure for estimating agricultural land use choices in about 100,000 homogeneous spatial units all over the EU territory, using 100,000 sampling points and aggregate data from administrative regions. The disaggregation procedure requires two steps. In the first step, we regress the share of a specific crop on natural conditions (soil, relief, climate) using the information from the sampling points. The estimated coefficients are then used to predict land use choices in each homogeneous spatial unit. Consistency with the administrative statistics is achieved by maximising the posterior density of estimated a priori information. Chapter 4 contains results for some European countries.

Keywords: Land Use Map, Spatial Disaggregation, Binary Choice, Bayes

1 Introduction

Not at least due to the so-called multi-functional model of European agriculture, there is growing interest in modelling environmental effects of the agricultural sector in the EU. In many cases, results beyond rather crude passive indicators can only be obtained. The general distinction of different land cover classes is based on the CORINE land cover map (European Topic Centre on Terrestrial Environment, 2000) describing land cover (and partly land use) according to a nomenclature of 44 classes, based on the visual interpretation of satellite images and ancillary data (aerial photographs, topographic maps etc.).

Within these administrative boundaries the natural conditions of soil, relief and climate usually differ in such a manner, that the assumption of identical cropping pattern, yields or input use cannot be maintained. Simulations with bio-physical models thus require breaking down results from the economic models into a smaller regional scale. We aim to develop a procedure computing a land use map for 30 cropping activities and one aggregated non-agricultural land use class at a spatial resolution of 1x1km covering the entire European Union.
Most of these approaches developed in recent years were applied at a scale comparable to a 1x1km grid. The accuracy of the employed land use classes varies depending on the aimed results and the complexity of the allocation procedure. The issue of spatial allocation of land use has been addressed from different perspectives during the last years which could be assigned to three major categories:

- Modeling approaches
- Remote sensing
- Land suitability

Howitt (2003), Nielson (2004) and Delden (2005) proposed methods where models based on analysis of available small scale data or expert rules predict values for small spatial entities. Additional information on larger regional units can be used in a procedure ensuring consistency of different scales. Expected values serve as prior information in a constrained optimization problem ensuring consistency of regional data and corresponding spatial units.

Interpretation of satellite images technically allows differentiation of single crops at a scale desired in our study. Although these techniques have been successfully applied in small scale applications like river basins their application at a European level appears too resource demanding at the desired land use classification. However the distinction of about 10 different land cover classes related to agricultural use can be based on the CORINE land cover map for Europe.

Soil suitability is the degree to which soil physical and/or chemical characteristics are in agreement with characteristics that are required for a certain land use. Soil maps give a number of permanent characteristics of the soil which in combination with other characteristics can be used to determine the suitability of locations for certain crops.

The procedure developed to establish a European land use map combines characteristics of various previously mentioned approaches. The CORINE land cover map serves as a starting point which has to be refined to meet the aimed land use classes. This subdivision is based on a model regressing point observations of cropping decisions all over the European territory on soil, relief and climate parameters describing the land suitability. Spatial statistical techniques are used to allow for spatial heterogeneity of the coefficients using a locally weighted logit model. Since statistical data of agricultural production is available at administrative regions a consistency ensuring step will finalize the dissaggregation procedure. The applied Bayesian method fully and transparently accounts for the available prior information – prior distributions derived from the binary choice model – when searching for consistency between the scales.

Chapter 2 describes the data base used while chapter 3 explains and justifies the statistical procedure in detail. Chapter 4 validates the results with out of sample data.
2 Database and Definition of Homogeneous Spatial Mapping Units (HSMUs)

The description of the database is divided in two main parts: (1) sources and definitions of the natural location factors, (2) the construction of HSMU.

2.1 Maps of Natural Location Factors

The relative competitiveness of an agricultural crop at a certain location is determined by natural factors, technology, and market conditions. While market conditions and the generally available technology are assumed to be rather invariant within an administrative region, differences in natural conditions will lead to heterogeneity regarding the optimal crop mix between different locations inside the Nuts II region. Therefore, this study concentrates on natural location factors.

Table 1. Relevant Maps of Natural Conditions

<table>
<thead>
<tr>
<th>Data sources</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Soil Database V2.0 (European Commission, 2004) / Soil Topological Units (STU)</td>
<td>Set of Soil Code (World Reference Base)</td>
</tr>
<tr>
<td></td>
<td>Drainage / water management</td>
</tr>
<tr>
<td></td>
<td>Stones</td>
</tr>
<tr>
<td>Digital Elevation model (CCM DEM 250, 2004)</td>
<td>Slope</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
</tr>
<tr>
<td>JRC-MARS meteodata</td>
<td>Annual Rainfall</td>
</tr>
<tr>
<td></td>
<td>Cumulative temperature sum</td>
</tr>
</tbody>
</table>

According to plant production literature, yield potentials of agricultural crops are mostly affected by soil quality, relief and climate conditions. Small scale informa-

---


2 CCM DEM 250. 2004. EuroLandscape/Agri-Environment Catchment Characterisation and Modelling Activity, Land Management Unit, Institute for Environment and Sustainability, EC-Joint Research Centre. 250 Meter DEM, compiled on the basis of data acquired from data providers and national mapping agencies over Europe for internal use.
tion on location factors stems from different sources and was prepared with the help of geographical information systems (GIS).

The European territory is divided in spatial referenced Soil Mapping Units (SMU). Each SMU consist of up to ten Soil Topological Units (STU). These STU are not spatial referenced and only the percentage of each STU within the corresponding SMU is known. The parameters of interest are given on STU level, mostly as qualitative information like “well drained” or “poorly drained”. Afterwards the percentage of “well” or “poorly” drained areas can be calculated for the SMU or to simplify the attribute of the dominant STU is assigned to the entire SMU. In the following analysis the percentage values are used.

According to the World Reference Base each STU is given a Soil Code like “Albic Luvisol”. These Soil Codes already indicate the suitability for farming. In the relevant SMU 95 WRB Soil Code are present. A first clustering was made based on Driessen et al (2001) who rearranged the 30 WRB soil groups into 10 so-called Sets, based on the dominant soil forming factors that determined the soil conditions. The list of 95 soil codes could be aggregated into 9 of those Sets.

Next, as some Sets were heterogeneous in its characteristics relevant for land use we have subdivided those Sets further into more homogeneous Sets, while other Sets were combined into one as the soils were rather similar from the land use perspective. Finally, from the resulting Sets some soil units have been placed in other Sets because of some common prominent feature. The distinction of new Sets was based on similarity of STUs in terms of soil characteristics which are judged relevant for land use, notably rooting depth, organic matter, texture, drainage class, Available Water Holding Capacity, presence of stones and slope. The grouping was based on judgement. (Diepen)

2.2 **CORINE Land Cover Map (CLC)**

The general distinction of different land cover classes is based on the CORINE land cover map (European Topic Centre on Terrestrial Environment, 2000) describing land cover (and partly land use) according to a nomenclature of 44 classes, based on the visual interpretation of satellite images and ancillary data (aerial photographs, topographic maps etc.).

The CORINE classification system distinguishes 11 agricultural classes (Non-irrigated arable land, permanently irrigated land, rice fields, vine yards, fruit and berry plantations, olive groves, annual crops associated with permanent crops, complex cultivation, pasture, marginal areas and forestry). Some of the classes as “Rice fields”, “Olive groves”, “Vineyard”, “Pasture” or “Arable Land” clearly indicate a special agricultural use. A minimum of 25 ha of homogeneous land cover is defined to build one CORINE mapping unit. That definition of the minimum mapping unit leads to two effects. Firstly, “pure” classes such as “Arable land” may in reality comprise small parcels of other land cover classes as well if these are smaller then 25 ha. Secondly, so-called heterogeneous agricultural areas as e.g. “Land principally occupied by agriculture with significant areas of natural vegetation (marginal area)” comprise no pre-dominant land use >25 ha and give only limited information about the type of agricultural use. The 25 ha limit results from the
mapping conventions and the interpretative limits set by the spatial and spectral resolution of the satellite images.
In this study we assume that only the agricultural classes are suitable for farming. The reader is reminded that agricultural classes may comprise small parcels of non-agricultural uses, as agricultural use may be found in non-agricultural classes.

2.3 Motivation and Construction of Homogeneous Spatial Mapping Units (HMSU)

The aim of building HMSU is the definition of areas inside an administrative region where approximate homogeneity according location factors may be assumed. The HMSU serve then as simulation units for the bio-physical models and are constructed by overlaying different maps (land cover, soil map, climatic factors etc.). In order to allow for a manageable number of HSMUs, the most important factors must be selected, and continuous parameters must be grouped in classes. The CORINE land cover map was used here in combination with two further main factors relating to soil (Soil Mapping Units) and relief (slope in 5 classes). Each HSMU has identical values for these three items, other parameters (such as annual rainfall) may differ inside the HSMU. Weighted averages are defined for the parameters shown in the table above for each HSMU using GIS techniques. The HMSU approach was deemed superior to a grid layout, especially as factors determining optimal cropping patterns may be identical across very large regions (say Northern Finland) so that grid units would be “wasted”, whereas in other regions especially such which high relief changes, the grid units may already comprise huge differences in natural conditions. Further on, the units can be defined so that they do not cross administrative borders, and grid data may be redefined based on the HSMUs.

2.4 Land Use / Land Cover Frame Statistical Survey (LUCAS)

In opposite to mapping approaches, area frame surveys based on a common statistical sampling method gather land cover and land use data (EUROSTAT, 2000) at specific sample points, only, and extrapolate from these to the entire area under investigation (European Commission, 2003a). LUCAS covers the territory of all EU Member States and all kinds of land uses, and is based on a two-stage sampling design: at the first level, so-called Primary Sampling Units (PSUs) are defined as cells of a regular grid with a size of 18 × 18 km, while the Secondary Sampling Units (SSUs) are 10 points regularly distributed (in a rectangular of 1500 × 600 m side length) around the centre of each PSU (Fehler! Verweisquelle konnte nicht gefunden werden.) resulting in approximately 10,000 PSUs for the whole EU (European Commission, 2003).

Due to possible measurement errors regarding the geo-references in the CORINE maps (Gallego 2002), about 30% of the LUCAS points closer then 100 m to the border of a CORINE class were not considered in here. The 38 agricultural classes found in LUCAS (36 crop land, 2 permanent grassland classes) were re-grouped according to the crops found in CAPRI. All other classes (artificial areas, woodland, water, etc.) are aggregated in a residual classed termed “OTHER”.
3 The Disaggregation Procedure

Before describing the crucial steps in detail the general approach of the disaggregation procedure is illustrated in Figure 1. Suppose there is a Nuts II region divided in only two HSMUs each comprising two crops – grassland (GRAS) and soft wheat (SWHE). Combining the LUCAS survey with digital maps provides us with several observations of crops grown at a defined point with a set of natural conditions. Using an adequate estimation model we can regress the probabilities of finding a crop at a certain location on the natural conditions. As this probability can be interpreted as the share of the crop in a homogeneous region, applying these estimated coefficients to the average natural conditions in a certain HSMU yields normally distributed predictions of crop shares for this HSMU under corresponding assumptions on the stochastic processes governing crop choice. These a priori information on cropping shares are generally not consistent with the “known” cropping area in the Nuts II region. The “best” set of data-consistent shares given the prior information is identified by a Bayesian highest posterior density approach. The concept of the HPD estimator allows the direct inclusion of the uncertainty of the prior mean. The variance can be derived from asymptotic properties or bootstrapping procedures.

Figure 1. Scheme of Disaggregation Procedure
3.1 Locally Weighted Binomial Logit Estimation

Generally, shares for each crop $\hat{Y}_c$ are regressed on the following explanatory variables describing natural conditions:

- Set of soil code
- Drainage
- Presence of stones
- Slope
- Elevation
- Rainfall
- Sum of temperature in vegetation period

The regressions were estimated independently for each crop $c$ in each CORINE class $clc$.

The arguments for using specific coefficients for each CORINE class are as follows. Assume grass land parcels are found in the LUCAS survey in the “non-irrigated land” CORINE class. We would assume that slope has a positive effect on the probability to find grass. In the “pasture” class of CORINE, we would eventually find the opposite effect: with increasing slope, grass land could be replaced by forest. For convenience the indices $c$ and $clc$ are omitted in the following.

The LUCAS survey reports one point in time observations and hence does not deliver cropping shares (or rotations), but requires a binary choice model. Both logit and probit models (see e.g. Green 2000) were originally tested, with the logit approach giving slightly better results. The likelihood function of finding crop $c$ at a specific LUCAS point $i$ for the binomial logit model is defined as:

$$ \Lambda(\beta'x_i) = \frac{e^{\beta'x_i}}{1 + e^{\beta'x_i}} $$

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

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

Applying the estimated $\hat{\beta}$ to the average natural conditions in a HSMU ($x_h$) give us a prior estimate for the share of a specific crop in a certain HSMU:

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

Binomial versus Multinomial Regression

The approach discussed above examines the crops independently from each other and thus neglects the information that crops compete for the available land, with two possible effects. Firstly, the error terms for the different crops are probably correlated, and secondly, the individual estimated shares don’t add up to unity. The
multinomial probit model would be ideal as it allows for an unrestricted variance
covariance structure of the error terms and satisfies the additivity condition, but is
computationally infeasible for 30 crops and 10,000 points. The assumption of an
identity matrix for the variance covariance matrix underlying the multinomial logit
model was deemed as too inflexible (Nelson et al. 2004), albeit it is easier to solve.
The way out might be a nested logit model, a possible expansion in further analysis.
However, both problems were not deemed crucial for the application at hand.
Given the large number of observations, the possible gain of taking correlations
between the error terms across crops into account is most probably small. Furthermore,
the violation of the adding up condition for the shares is explicitly accommodated
in the second step of the disaggregation procedure, where the estimated
shares serve as prior information, only.

**Single crops versus Groups of crops**
The land use shares are estimated for 30 single crops and 10 groups of similar crops
(see also Table 2). The grouping of single crops was done with respect to natural
requirements of specific crops. This can help to allocate crops where sparse informa
tion from the Lucas sample is available, because the overall procedure will tend to allocate those crops in HSMUs where crops requiring similar natural conditions
are assigned.

**Local versus Global Regressions**
The assumption of European wide invariant relationships between the share of each
crop and a limited number of location factors describing natural conditions may be
problematic if other omitted explanatory factors are not randomly distributed in
space, but “clustered”. Suppose, for example, two HSMUs with identical natural
conditions, the first one close to a sugar refinery, and the second one far way from
the next sugar plant. The share of sugar beets in the first unit will be probably much
higher, an effect not linked to the natural conditions. Clearly, omitted variables as
the effect of sugar refineries could lead to seriously biased parameter estimates.
Adding more explanatory variables would certainly help, but it is simply impossible
to collect information on all probably relevant factors (market points, transport
infrastructure, environmental legislation, etc.). Instead, spatial econometric tech
iques are applied to overcome the problem of omitted variables that are correlated
over space.
The basic idea behind Locally Weighted Regression, which was proposed by
Cleveland and Devlin (1988), is to produce site specific coefficient estimates using
Weighted Least Squares to give nearby observation more influence than those far
away. Further on, the estimation for any specific site is limited to a number of ob
servations within a certain bandwidth around the site. Locally Weighted Regression
are mostly found combined with Least Squares estimators, but application to
Maximum Likelihood Estimation as needed in the case of discrete dependent vari
ables are described as well (Anselin et al. 2004).
The weight given to any observation \( i \) in constructing the estimate for site \( j \) is given by \( \omega_{ij} \). The tri-cube is a commonly used weighting function:

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

Where \( \delta_{ij} \) is the distance between site \( i \) and observation \( j \), \( d_j \) is the bandwidth and \( I(.) \) is an indicator function that equals one when the condition is true. The effect of any one location in space on near points thus falls depending on the distance and becomes zero once the distance exceeds the bandwidth. There are other common weighting schemes like the Gaussian function or several Kernel weighting functions (see: Anselin et al. 2004 or Fotheringham et al. 2002). But it has been shown that opting for a proper bandwidth is more significant than choosing a certain spatial weighting function.

When there is no prior justification for applying a particular bandwidth, an appropriate bandwidth can be found by the minimising either the cross-validation score (CV), the Akaike Information Criterion (AIC) or the Schwartz Criterion (SC). The AIC and the SC are offered by most software packages. The CV is calculated as:

\[
CV = \sum_{i=1}^{n} (y_i - \hat{y}_{i,i})^2
\]

where \( n \) is the number of data points and the prediction for the \( i \)th data point \( \hat{y}_{i,i} \) is obtained with the weight for that observation set to zero. Each of the criteria can be minimised by a golden section search (see Press et al. 1989). In our study all criteria led to similar results. We opted to minimise the Schwartz Criterion, because according to Boots et al. (2002) it seems to have better large sample properties.

In typical applications, sites and observations would be identical. In our context, that would require estimates per crop and CORINE class for each LUCAS point, which is computational impossible. Instead, the NUTS II regions were chosen as sites. When estimating for a particular NUTS II region, all LUCAS point inside that NUTS II region received uniform unity weight, and points in neighbouring NUTS II regions weights equal or smaller unity according to (4). That still leads to a large number of possible estimations: 150 Nuts II regions times 10 agricultural CORINE classes times 30 crops, but fortunately, many of the combinations do not comprise any observations. Weighting each likelihood contribution with \( \omega_{ij} \) gives (Fotheringham et al. 2002):

\[
\log L = \sum_{i=1}^{n} \omega_{ij} [y_i \log \Lambda(\beta_j|x_i) + (1 - y_i)(1 - \Lambda(\beta_j|x_i))]
\]
3.2 Attaining Variance of Land Use Shares

Given the non-linear character of the estimations, the variance-covariance matrices offered by the statistical packages are not analytically calculated, but are instead numerically approximated which proved to be not suitable. Quite small predicted mean values in combination incredibly high variances led to shaky final results. Consequently the estimation of the prior variance attracts our attention. Statistical formulas can be used to derive the variance of a predicted mean. Furthermore results from bootstraps from the finite sample can be taken to calculate the variance.

Robust Covariance Matrix Estimation

The prior variance $\hat{Y}$ is based on the asymptotic covariance matrixes for the coefficients. A robust covariance matrix can be calculated analytically (see White (1982)) as:

$$V_\beta = Cov[\hat{\beta}] = \hat{H}^{-1}\hat{B}\hat{H}^{-1}$$

where for the weighted logit model the elements of Hessian $H$ and the Brendt, Hall, Hall and Hausman matrix $B$ are given by (Green 2000):

$$H = \frac{\partial^2 \text{LogL}}{\partial \beta \partial \beta'} = -\sum \omega_i \Lambda_i (1 - \Lambda_i) x_i x_i'$$

$$B = \sum \omega_i (y_i - \Lambda_i)^2 x_i x_i'$$

As insignificant parameter estimates might influence the efficient calculation of a robust covariance matrix although they do not influence the forecasted value, insignificant variables were removed from the estimations. The variance of $\hat{Y}$ builds upon the calculated covariance matrix $V_\beta$.

$$V_Y = Var[\hat{Y}] = \Lambda_i (1 - \Lambda_i) x' V_\beta x$$

Using specific $x_{HSMU}$ yields variances of the predicted land use share in each HSMU (Green 2000).

3.3 Data-consistent Disaggregation

The second step of the disaggregation procedure identifies crop shares in each HSMU using the prior information on the estimated crop shares from the first estimation step under two data constraints: Firstly, adding up the areas per crop in each HSMU must recover the cropping areas $CA$ for that crop at NUTS II level. Secondly, the posterior shares in each HSMU must add to unity, including all non-agricultural land use from the LUCAS survey aggregated to the category “OTHER”. In opposite to the first step this requires simultaneous accounting for all...
crops $c$ in all relevant HSMUs $h$. The notation is therefore extended, e.g. from $Y$ to $Y_{c,h}$.

The crop areas in each HSMU are defined by multiplying the posterior shares $Y_{c,h}$ with the entire area $A_h$, thus

$$\sum_{h \in N^2} Y_{c,h} A_h = C A_{c,N^2}$$

While shares of single crops must be in line with the corresponding groups of crops:

$$\sum_{c \in g} Y_{c,h} = Y_{g,h}$$

and the adding up to unity must be imposed:

$$\sum_{c} Y_{c,h} = 1$$

As the predicted unrestricted shares will typically violate the constraints, a penalty function is necessary to define the optimal deviations from the predictions. Generalized Maximum Entropy (GME) techniques (Golan, Judge and Miller 1996) have often been used for this type of data balancing exercises in recent times. Here, however, a Bayesian highest posterior density (HPD) estimator is applied allowing for a direct and transparent formulation of prior information and considerably reducing the computational complexity compared to the GME approach (Heckelei et al. 2005). The prior information for single crops $c$ and groups of crops $g$ is expressed as normal densities of predicted shares, with mean vector $\hat{Y}_{c,g,h}$ and variance derived by the methods described before. After taking logs, the prior density function for the consistent shares $Y_{c,h}^{con}$ is:

$$-\sum_{c \in g} \sum_{h} \left[ \log \left( 2\pi \nu_{cg,h} \right) + \frac{(Y_{c,g,h}^{con} - \hat{Y}_{c,g,h})^2}{2\nu_{cg,h}} \right]$$

4 Results

As it is hardly possible to present here the actual outcome of the disaggregation procedure – 30 land use shares in 100,000 HSMU. In order to enable the use of the data for further calculation the outcome is translated to a GIS map. In the following section we want to focus on validation of the results.

For some European regions, land use statistics at a lower administrative level, called Nuts III, are available from the farm structure survey (FSS; EUROSTAT,
2002). This information is used as out-of-sample observation to validate the results of the disaggregation algorithm, which predicts cropping areas for the HSMUs consistent to NUTS II\(^3\). Adding up over the corresponding HSMU yields crop areas at NUTS III level that can be compared to the observed data. Figure 2 exemplary contrasts actual and predicted cropping areas for selected crops in the nine Nuts III regions in Castilla-Leon, Spain. Although the disaggregation reflects the principal pattern quite well there are sometimes large differences.

**Figure 2.** Comparison of estimated and observed shares in NUTS III region for different crops (Castilla Leon ES410)

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

\(^3\) Usually this disaggregation procedure is applied to the complete and consistent Nuts II database of the CAPRI modelling system, but any other statistic can be used as well. In order to allow a consistent analysis based on FSS Nuts III data the corresponding Nuts II information was used here.
Table 2. Percentage of misclassified area for different crops (Castilla Leon ES410)

<table>
<thead>
<tr>
<th>Crop</th>
<th>Missclassified Area in Nuts III (% of UAA in Nuts II)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single Crop</td>
</tr>
<tr>
<td>Soft Wheat</td>
<td>3.39</td>
</tr>
<tr>
<td>Durum Wheat</td>
<td>0.32</td>
</tr>
<tr>
<td>Barley</td>
<td>5.02</td>
</tr>
<tr>
<td>Rye</td>
<td>0.87</td>
</tr>
<tr>
<td>Oats</td>
<td>1.46</td>
</tr>
<tr>
<td>Maize</td>
<td>1.24</td>
</tr>
<tr>
<td>Other Cereal</td>
<td>0.06</td>
</tr>
<tr>
<td>Fallow Land</td>
<td>2.95</td>
</tr>
<tr>
<td>Rice</td>
<td>0.00</td>
</tr>
<tr>
<td>Sunflower</td>
<td>1.19</td>
</tr>
<tr>
<td>Soya</td>
<td>0.00</td>
</tr>
<tr>
<td>Texture Crops</td>
<td>0.59</td>
</tr>
<tr>
<td>Pulses</td>
<td>0.34</td>
</tr>
<tr>
<td>Other Crops</td>
<td>0.00</td>
</tr>
<tr>
<td>Potatoes</td>
<td>0.24</td>
</tr>
<tr>
<td>Sugar Beet</td>
<td>0.60</td>
</tr>
<tr>
<td>Root Crops</td>
<td>0.01</td>
</tr>
<tr>
<td>Rape</td>
<td>0.02</td>
</tr>
<tr>
<td>Tobacco</td>
<td>0.01</td>
</tr>
<tr>
<td>Other Industrial</td>
<td>0.04</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>0.00</td>
</tr>
<tr>
<td>Other Vegetable</td>
<td>0.17</td>
</tr>
<tr>
<td>Flowers</td>
<td>0.00</td>
</tr>
<tr>
<td>Other Fodder</td>
<td>1.88</td>
</tr>
<tr>
<td>Grassland</td>
<td>9.27</td>
</tr>
<tr>
<td>Nursery</td>
<td>0.01</td>
</tr>
<tr>
<td>Fruits</td>
<td>0.10</td>
</tr>
<tr>
<td>Citrus</td>
<td>0.00</td>
</tr>
<tr>
<td>Olive</td>
<td>0.11</td>
</tr>
<tr>
<td>Vine</td>
<td>0.40</td>
</tr>
<tr>
<td>Nuts II</td>
<td><strong>30.29</strong></td>
</tr>
</tbody>
</table>

Figure 3 illustrates the misclassified areas in Europe where out of sample data is available from the FSS statistics. In regions with a high percentage of misclassified area often grassland accounts for a significant part of the errors. This is astonishing since grassland has its “own” Corine land cover class and indicates that misclassification might not only be a consequence of a poor disaggregation procedure but also
a result of contradictious data sources. Nonetheless the dissaggregation is a significant improvement compared to the assumption of identical cropping pattern within each Nuts II region.

Figure 3. Percentage of misclassified areas in validated Nuts II Regions after disaggregation

---

4The Corine land cover map reports indeed about 2 Mio ha “Pasture” and “Natural Grassland” in Spain while in the FSS statistic about 9 Mio ha Grassland are declared.
5 References


