# Salvage the treasure of geographic information in Farm census data

**Alexander Gocht and Norbert Röder**

Johann Heinrich von Thünen-Institute; Institute of Rural Studies
Bundesallee 50, 38116 Braunschweig, Germany

Alexander.Gocht@vti.bund.de

**Paper prepared for presentation at the EAAE 2011 Congress**
**Change and Uncertainty**
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

*Abstract*

*In Germany, since several decades the RAUMIS modelling system is applied for policy impact assessments to measure the impact of agriculture on the environment. A disaggregation at the municipality level with more than 9.600 administrative units, instead of currently used 316 counties, would tremendously improve the environmental impact analysis. Two sets of data are used for this purpose. The first are geo-referenced data, that are, however, incomplete with respect its coverage of production activities in agriculture. The second set is the micro census statistic itself, that has a full coverage, but data protection rules (DPR) prohibit its straightforward use. The paper show how this bottleneck can be passed to obtain a reliable modelling data set at municipality level with a complete coverage of the agricultural sector in Germany. We successfully applied a Bayesian estimator, that uses prior information derived a cluster analysis based on the micro census and GIS information. Our test statistics of the estimation, calculated by the statistical office, comparing our estimates and the real protected data, reveals that the proposed approach adequately estimates most activities and can be used to fed the municipality layer in the RAUMIS modelling system for an extended policy analysis.*

*Keywords: Highest Posterior Density estimator (HPD), RAUMIS, Down scaling*

*JEL classification: C11, C61, C81, Q15.*

## 1. INTRODUCTION

Frequently, the impact of agricultural activities on the environment can only be properly assessed if the underlying distribution is well-covered. For instance, the likely impact of new pests such as the western corn rootworm (*Diabrotica virgifera ssp. virgifera* LeConte), which is relevant to the debate on bT-maize, depends on the share of maize in the crop rotation. Namely, if the share of maize exceeds 50%, western corn rootworm may have a serious impact (CARRASCO et al., 2009). If we analyse the cultivated area in 2007 at the county level which are 316 regions in Germany, the results indicate that the cultivation of maize in Germany should barely be affected by the rootworm (FDZ, 2010). However, if we conduct the same analysis on the municipality level, almost 13% of the maize cultivating areas would be affected by the rootworm. Thus, because agricultural land use and its dynamics are site-dependent, the utilisation of wider regional averages to model specific situations can be misleading (e.g., OSTERBURG *et al.*, 2009, p. 40 ff.).

The agricultural and environmental modelling and information system RAUMIS (HENRICHSMEYER *et al.*, 1996) is a mathematical programming, modelling and information platform used to cover Germany's agricultural sector. RAUMIS is used to analyse agricultural and agri-environmental policy instruments and currently operates at the county level. Similar to economic models such as CAPRI (BRITZ and WITZKE, 2008), the RAUMIS model simulates an aggregate over all farms in a particular region. To overcome problems related to data aggregation, the underlying heterogeneity of farming patterns must be represented. Thus, several different approaches have been applied to disaggregate regional models. For example, a specifically tailored component in the CAPRI model has been used to disaggregate crop shares, stocking densities and fertilizer application rates from about 250 administrative regions across Europe into clusters of 1x1 km grid cells (LEIP et al., 2008) that are based on homogeneous spatial mapping units (KEMPEN *et al.*, 2005). Other downscaling approaches of agricultural statistical data with the help of geographical and/or remote sensing data are presented by DENDONCKER et al., (2006), VERBURG et al. (2006), YOU and WOOD, (2006). However, the resulting resolution with respect to animal and crop categories is very limited and therefore less useful in modelling agricultural decision process. Also if the results are spatially disaggregated into clusters of grid cells, the borders of the clusters do not necessarily coincide with administrative boundaries. Alternatively, a disaggregation of regional production levels into farming groups such as done by GOCHT and BRITZ (2010) is an option. However, this approach also has serious disadvantages because of the missing territorial representation which in turn does not allow spatially geo-referenced data to be linked, an important feature for regional models as RAUMIS.

Alternatively and in the focus of this study, county data are disaggregated to the municipality level using Agricultural Census and GIS data. In contrast to gridding that distributes data published by statistical offices according to a rule set we develop an approach that is capable to exploit the geographic information in the Agricultural Census as far as possible. However, the public availability of high-resolution data (both regarding topological and / or geographic aspects) is limited by legal constraints. In particular, many production activities at the municipality level fall under the data protection regulation (DPR) and are not reportable because the number of observations is limited. Currently, the DPR is ensured by censoring data if they are derived from less than three observations or if a one or two observations dominate the result. A result is viewed as being dominated if a single observation contributes more than 80% to the aggregate (EUROSTAT, 2009). Furthermore, additional aggregates are censored to ensure that data censored in step one cannot be retrieved from the published data. As result, the likelihood that the data will be censored increases with increasing resolution.

If we want to overcome this and disaggregate the county data for the RAUMIS model to the municipality level using Agricultural Census data we need a method to extract additional information from official statistical offices without violating DPR. In contrast to GOCHT and ROEDER (2010) who apply a method based on locally weighted averages and restricted their analysis to a specific region in Germany, we propose an algorithm that recovers local information with the help of the activities' median at the municipality level German wide. These medians are calculated for clusters of similar municipalities. The aim of the present study is to develop an algorithm that is capable to depict the distribution of agricultural land use with the spatial resolution of municipalities. We evaluate the estimated results with respect to both relative intensities (i.e. shares in the crop rotation and stocking levels) and absolute values (i.e. ha or livestock units (LU)). To our knowledge no attempt has been made so far at this coverage and administrative resolution, which results in a public and not traceable dataset for policy impact assessment.

The remainder of the paper is organized as follows. Section 2 highlights some key characteristics of the data. In Section 3, we describe the applied data manipulation algorithms and introduce the estimation framework. Lastly, Section 4 presents the results, and we conclude in a final section.

## 2. METHODS

The section starts with explaining the preparatory steps necessary to overcome inconsistent data definitions between the statistical data bases and the RAUMIS model definition, before we describe the estimation framework and we finalize introducing the test statistic used to evaluate our estimates.

Figure 1 presenting the consecutive processing steps in order to facilitate the understanding of the data processing and handling. It distinguishes between two data processing environments. Processing at the *Research data centre* (FDZ) is done via sending data processing algorithm of standard statistical packages to the FDZ and because a researcher has never direct access to the micro data, one is forced to construct the processing algorithm virtually blind, knowing only the data structure and definition of the data. These conditions are rather uncomfortable because a validation whether a result is an observed trend or just a phenomena resulting from mapping or definition errors is difficult. Also the situation that economic simulation models

are rarely realized in a standard statistical package makes the direct processing in the FDZ environment very cumbersome, and often impossible for economic policy evaluation. However, the big advantage is to have the opportunity to use the high resolution micro data shown in Figure 1 with the AFiD-Panel Agriculture database, to derive indicators. The AFiD Panel Agriculture is derived from the Farm Structure Survey (FSS) and provides extensive information on the agricultural activities in a four year interval for **all** German farms.
All routines to be processed at the FDZ will be checked and results leave the FDZ only when they are in compliance with the DPR, presented in Figure 1 as the dotted rectangle between the two processing environments. Figure 1 also shows the processing at office environment, which is the researcher's office. Here we can use the outcome of the FDZ, which is anonymous not traceable and in compliance with the DPR for further analysis and applications. In Figure 1 step 3 illustrates the setup of an estimation framework, in which we use GIS data together with the FDZ information to obtain a consistent municipality data set.
We now explain step 1 until 3 in more detail: The data preparation in ***Step 1*** comprise the usual preparatory data work, mainly harmonizing definitions. As we need for RAUMIS a consistent data set at municipality level for several years from 1999 onwards we had to adjust and map regional definitions. As example, municipalities merged, split or exchanged and hence significant amounts of land. After harmonizing we remained with 9,679 time consistent municipality units. We had to aggregate some statistical codes to be in line with our 36 RAUMIS agricultural production activities. A complete list of the production activities can be found in GOCHT and RÖDER (2010).

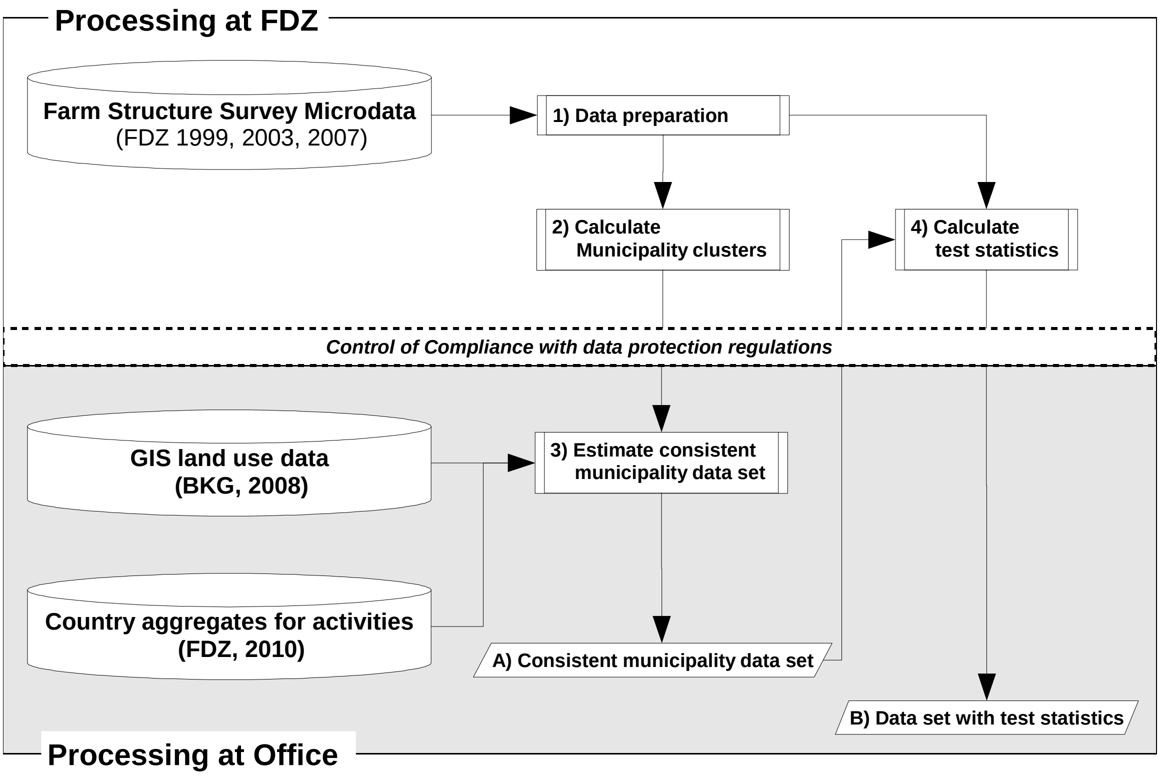


Figure 1: Information flow in the estimation procedure
Source: Own elaboration

As the DPR prevent a direct retrieval of RAUMIS production activities at municipality level, we developed in ***Step 2*** a processing algorithm that complies with the DPR. We clustered the 9,679 regional units into 180 clusters based on several indicators for general land use, arable land use and animal density given in Table 1. For the three groups we independently applied the kMeans-algorithm (WITTEN and FRANK, 2005). The algorithm was sent to the FDZ and applied to the micro data.

Table 1: Indicators obtained from each cluster

| Indicator group | Unit | Indicators |
|---|---|---|
| General land use | % of utilized agricultural area (UAA) | Arable land, cereals, root crops, vegetables, main forage area, fruits, grassland, rough pastures |
| Arable land use | % of arable land | winter wheat, summer barley, rye, other winter cereals, other cereals, grain maize, rape seed, potatoes, sugar beet, green maize, other forage crops on arable land, other crops, set aside |
| Livestock husbandry | Livestock units (LU) per ha of UAA | Suckler cows, dairy cows, heifers, bulls, calves, sheep, horses, poultry, pig fattening, pig breeding |

Source: Own elaboration

From the processing at FDZ we obtained for each cluster, and hence the municipalities belonging to it, a median and standard deviation of the respective indicators given in Table 1. In Step 3 we setup an estimation framework with the aim to estimate the municipality production structure of the 36 RAUMIS production activities. We setup the model per county. Hence aiming for a complete German wide coverage we had to solve 316 models. With each model we estimate the maximum 36 possible production activities for all municipalities. The number of municipalities per county range from 6 to 159 with a median of 25. In addition, the estimation algorithm uses GIS information on the extent of five land use types (utilized agricultural area (UAA), arable land, grassland, wine yards and orchards) and the agricultural production statistic at the country level, which is publicly available.

The cluster median for each indicators is interpreted as *a priori* information in the Bayesian sense, whereas the data information consists of the given county production values, sum of production activities over the municipalities is equal to the county level, and the constraint that the estimated activity levels add up to observed land use type, observed in GIS data (see Gocht and Roeder, 2010).

Our Bayesian Highest Posterior Density estimator (HPD) maximizes the log of the joint posterior density (see Heckelei et al., 2008), i.e. it searches for the most probable deviations from the cluster median fitting our data information on country activity level and the land type GIS information. Without knowledge about the exact distribution of the error terms in the clustered data, normally distributed errors with a co-variance of zero between the different medians and the obtained variance from FDZ are assumed.

The constraints alone do not allow a unique solution to be identified as there are too many unknown vectors of estimated cropping hectares and livestock herd sizes, exceeding the number of data constraints from GIS and county level statistic. Therefore, prior information must be included in combination with a penalty function. Generalised maximum entropy (Golan et al., 1996) has frequently been applied to this end. However, we used the HPD estimation, which allows a direct and transparent formulation of prior information and reduces the computational complexity of the model (Heckelei et al., 2008). Subject to the constraints, the objective function, assuming a normal distributed error (Heckelei et al., 2008), is a loss function, which minimize the sum of the standardized proportional deviations between our prior expectation and the estimates:

$$(1)\quad \min \sum_{M} \sum_{A} \left( \frac{\left(s^{e} - s^{p}\right)^{2}}{\sigma} * \left(X^{p}\right) \right),$$

where $M$ are the municipalities in a county and $A$ either the GIS land use types (UAA, arable land, ..) or the RAUMIS activities, $s^e$ the estimated share, $s^p$ the respective prior information on median and $\sigma$ obtained by the cluster algorithm (RAUMIS) or the GIS analysis (extent of land use types). $X^p$ is a weight expressed as the expected level of the production activity in a municipality. We standardize the difference between $s^e$ and $s^p$ by σ to account for differences in the confidence we have in $s^p$. Hence, the objective function minimizes the deviations between estimated and observed cropping shares, livestock densities, and composition of the municipalities stock.

After we applied the estimation we obtained absolute and relative shares for all RAUMIS activities. In Step 4, we calculate test-statistics to verify our findings by comparing the estimates with the micro census data. This is possible using the real micro census data. Hence we had to use the virtually blind approach, sending the estimates together with the test statistic routine to the FDZ and could validate our results. We evaluated the distribution of the differences between estimated and observed cropping shares and livestock densities weighted with the respective local production level to assess the overall quality of the results.

The following software was used for the analysis at the FDZ: SAS 9.1 for regression and cluster analysis and the Conopt3-solver in GAMS 23.5 for the Bayesian minimisation problem.

## 3. RESULTS

In section 3.1 we present the general fit of the prior data & constraints compared to our estimates for the 316 models. In section 3.2 we analyse the estimates compared to the real observations. This evaluation is possible because we could compare our estimates with the real data population at FDZ and calculate certain test statistics. We finalize with an analysis of the distribution and development over time of land use of maize in Germany to illustrate the potential of the obtained high resolution data at municipality level and to come back to our illustrative example from the introduction.

### *3.1. Regional variation in the consistency of the prior information*

We start by investigating how consistent the different prior information (clusters based on the FSS and GIS) are in comparison to our obtained results. As aggregated indicator overall production activities and municipalities in a county we present the deviations according to formula (1) in Figure 2. The deviations are in relative terms low in Southern Germany, medium in the North and reach high values in the East. An explanation for these regional differences is the relation of farm size to municipality size. The FSS attributes the farm's activities according to the *situs principle* to the municipality of the farm's headquarter (farmstead). In contrast, the GIS data are attributed according to the location of the plot. This implies that the larger a farm is in relation to the municipality it is located in, the higher is the likelihood that some of the farm land or livestock herd, is located in reality, compared to the statistical data at FDZ, outside of the municipality. Therefore, we get a biased estimate from the cluster analysis. Figure 2 shows that this is particular the case in Eastern Germany.

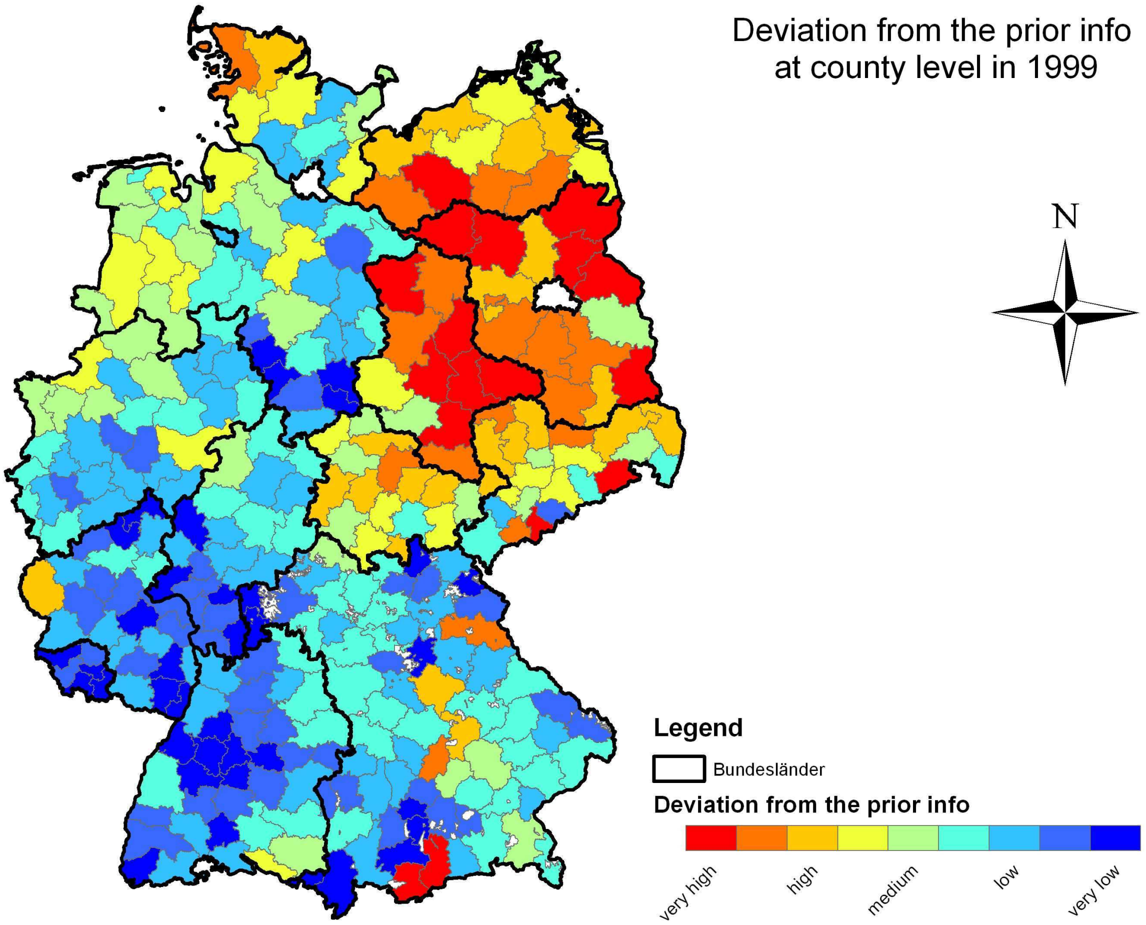


Figure 2: Deviation from the prior value aggregated over all municipalities and activities for 1999

Source: FDZ, and own calculation.

### *3.2. Error Distribution*

The indicator in Figure 2 does not provide us with quality measure for our estimates. To obtain this we need to compare the estimates with the true "observed" production activities at the municipality level. Although the DPR at FDZ prohibit a test statistic for individual data estimates, we can derive, sending our estimates and the test statistic to the FDZ, an aggregated test statistic including the error distribution. In order to avoid a bias by municipalities with no or only a very small stock, we weighted for each municipality the deviation between the observed and estimated stocking density with the respective observed stocking level. This test statistic is presented in Figure 3 for livestock husbandry. It shows that for the livestock activities the estimated livestock densities on municipality level match the observed ones very well. In general, more than 50% (the interval between the 25% and 75% quantile = blue box) of the respective total German stock is attributed with an error regarding the stocking density of well less than ±0.05 LU per ha. For most activities even 90% of the respective stock (Whiskers) are attributed with an error of roughly ±0.1 LU per ha. However, the proposed method is not capable to fully depict the high local intensities characterising pig and poultry production. Here, the interpolation associated with the use of cluster medians implies a large aggregation error. The Box Plot for the plant production activities is depicted in the Annex.

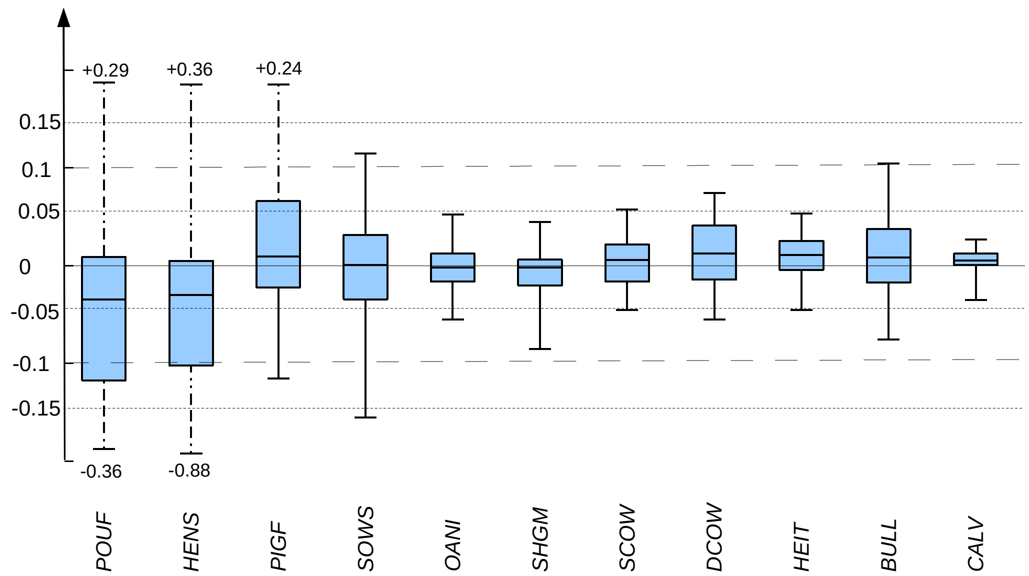


Figure 3: Boxplot of the deviations on municipality level for animal activities in 2007
Description of the activities see Table 2;
Box: 25% and 75% Quantile; Whiskers: 5% and 95% Quantile
Source: FDZ, own calculation.

The absolute levels of deviations between the observed and estimated levels are shown in Table 2 for 2007 for different quintiles. The error mean (50% quantile) locates near zero for all production activities. Our estimation hence fits the underlying population. Further, the table tells us that for example in ~4,200 of ~8,400 municipalities (between the 25% and 75% quantile) the stock of dairy cows (row four) is over(under)estimates by at most ~59(-67) LU. For the majority of cases (municipalities and activities) the error regarding the absolute level of the local stock lies between ± 20 LU. However, larger errors are not unlikely in particular for pigs, bulls, heifers and dairy cattle. At least for the cattle activities these larger errors occur predominantly in municipality with large stocks, therefore limiting the proportional error regarding the attributed stock.

Table 2: Distribution of the absolute differences between the estimated and observed livestock at municipality level in 2007 (in LU)

| RAUMIS | Description | n° of munici-palities | Avg. herd size per municipality | Quantile of the error distribution | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 5% | 25% | 50% | 75% | 95% |
| CALV | Calves | 9074 | 66 | -60 | -9 | 0 | 10 | 48 |
| BULL | Male cattle > 6 month; stock bulls | 8972 | 138 | -134 | -15 | 3 | 28 | 136 |
| HEIT | Heifers | 9191 | 273 | -198 | -37 | 0 | 34 | 156 |
| DCOW | Dairy cows | 8382 | 486 | -363 | -67 | 2 | 59 | 263 |
| SCOW | Suckler and fattening cows | 8826 | 84 | -138 | -23 | 0 | 20 | 107 |
| SHGM | Sheep | 8476 | 24 | -90 | -7 | 1 | 9 | 47 |
| OANI | Other livestock (horses) | 8796 | 59 | -101 | -21 | -2 | 15 | 78 |
| SOWS | Sows for piglet production | 7622 | 117 | -166 | -16 | 1 | 20 | 140 |
| PIGF | Pig fattening | 8614 | 250 | -203 | -18 | 1 | 20 | 154 |
| HENS | Laying hens | 8854 | 25 | -41 | -1 | 0 | 3 | 36 |
| POUF | Poultry fattening (broiler, turkeys, etc.) | 8480 | 34 | -64 | 0 | 0 | 2 | 44 |

Source: FDZ, own calculation.

To finalize the analysis we compare for maize the estimation results at municipal level with an approach in which we assume that county aggregated shares, available from RAUMIS are a good estimate of our municipal shares. Figure 4 shows that although, in many areas in

Germany the county averages are a reasonable estimate for the municipality shares (e.g. Rhineland-Palatine, Hesse, Thuringia, and Saxony) the county averages underestimate drastically the relevance of maize in the Geest of Schlewig-Holstein and Lower Saxony, and in the foothills of the Alps, the Bavarian Forst and the Odenwald. Also the relevance of maize is overestimated for large parts of the Black forest, the marsh land of Lower Saxony and the north eastern part of Schleswig-Holstein.

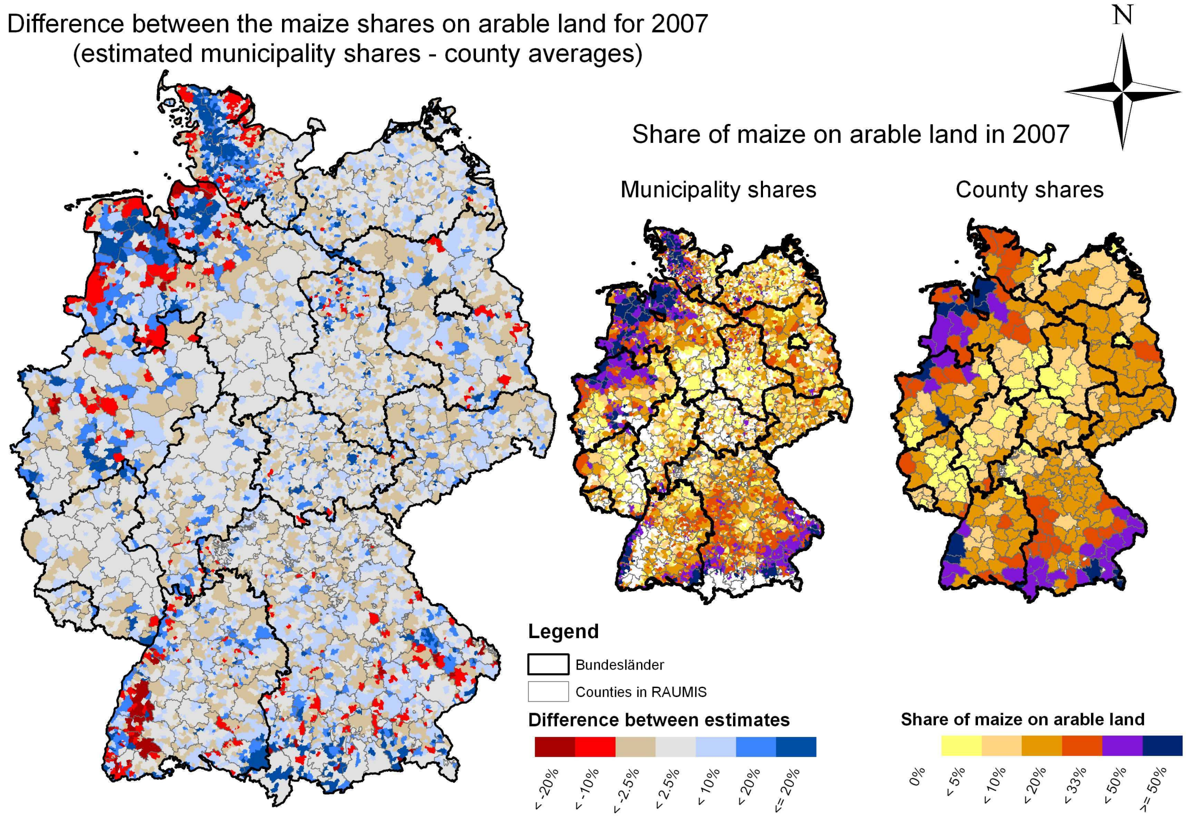


Figure 4: Difference between the estimated shares of maize on arable land for 2007 (estimated municipality shares – county averages)
Source: Own estimation

### *3.3. Development and cultivation of maize in Germany*

After we evaluated the quality of the estimates compared to the real population and for maize compared to a naive approach using equal municipality shares from the county, we will use the obtained results to analyse the distribution and development of maize shares in Germany at municipality levels, to gain more insight into possible phytosanitary problems. To our knowledge, such an exercise is done for Germany for the first time with such a resolution.
Figure 5 depicts the estimated distribution on municipality level of maize (grain and green) in Germany for 2007. Despite the fact that maize was grown only on 16% of Germany's arable land, maize covers more than 33% of the respective arable land in a couple of areas. One centre lies in north-western Germany between the Ruhrgebiet and Rhine in the south-west and the Elbe in the north-east. A second large hot spot is located in south-eastern Bavaria east of the Inn and between the Alps and the Bavarian Forest. Smaller areas with high shares of maize (beyond 33%) can be found in the Geest (Schleswig Holstein), the Upper Rhine valley (Baden-Württemberg), the foothills of the Allgäu (Baden-Württemberg and Bavaria) and the Sauerland (Northrhine-Westphalia). Maize reaches, hence, in several areas quite critical levels regarding phytosanitary issues when the distribution is analysed at municipality level.

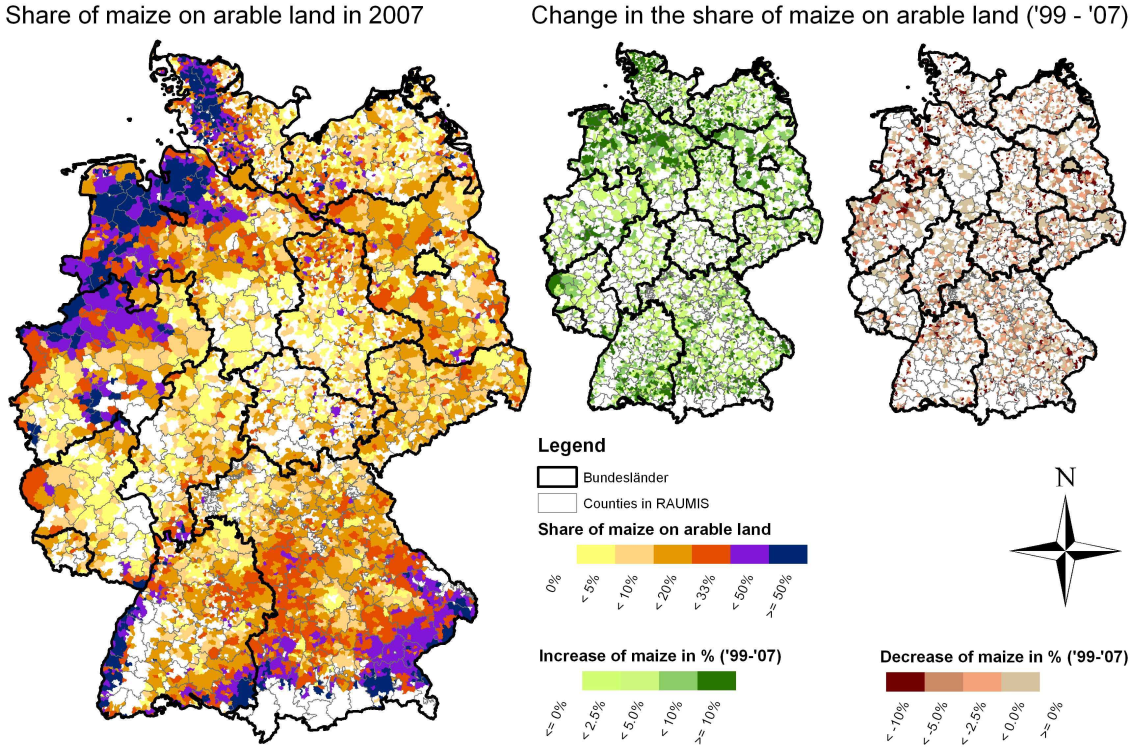


Figure 4: Dynamic of estimated maize shares on arable land 2007 compared to 1999
Source: Own estimation

The area cultivated with maize expanded by 300,000 ha between 1999 and 2007 resulting in a moderate increase of maize's share on total arable land from 13.3% to 15.9%. However, these aggregate figures cover a quite significant dynamic on the local level that we now are able to analyze with the outcome of the estimation. In large parts of North-Western Germany, in the Geest, and in the vicinity of mountain ranges (e.g. Eifel, Sauerland, and Alps) maize's share on arable land increased by more than 10% points. The cultivation of maize declined in the north-western part of Northrhine Westphalia, the eastern part of Bavaria and the northern part of Baden-Württemberg. Till 2002 the cultivation of maize was strongly linked to arable forage cropping in particular dairy farming and bull fattening. This explains the high shares of maize in areas with high cattle densities (e.g. along the North Sea and in the foothill of the Alps). Grain maize including corn-cob mix was important in the Upper Rhine Valley, along the border between Northrhine-Westphalia and Lower Saxony and in south east of Bavaria. While the area of grain maize remained nearly constant over the last decade the area of green maize declined parallel to the declining cattle stock till 2002. From 2002 till 2007 the maize area expanded by more than 360,000 ha due to the promotion of biogas production based on silage maize (BMELV, various years). The described development is critical for two reasons. First, maize cultivation is expanded in areas where maize is already the dominant crop, increasing phytosanitary risks. Second, the cultivation of maize in mountain ranges induces a high risk of erosion, as in these areas the precipitation is high, the terrain is fairly undulated and maize is developing a protective vegetation cover late in the year.

## 4. Conclusions and Outlook

The proposed method of disaggregation, which combined the highest posterior density (HPD) and a cluster analysis improved land use estimates at the municipality level and complied with the data protection rules (DPR) at the FDZ.

The correlation between the observed and predicted values was analysed for the entire data set in German, and the results indicated that the proposed approach can adequately depict the spatial and density distribution of most RAUMIS activities while complying with the DPR.
Not surprisingly the described procedure greatly improves the mapping quality for activities whose distribution shows are clear spatial pattern that does not coincidence with the county borders e.g. the distribution of rough pastures or the distribution of maize in Schleswig-Holstein and Baden-Württemberg. If an activity is widespread and dominant the advantage of the estimated results versus a naive downscaling of the county shares is less clear.
On the local level the described procedure generally reaches a high level of accuracy regarding relative indicators as stocking densities and cropping shares. However, the absolute reported values on this level must be interpreted with some caution.
For most activities the described procedure generally covers well the intensity gradient present in Germany's agriculture. There seem to be two main reasons why our estimated results deviate from the census data. First, we are deriving prior information and constraints from two databases (FSS and GIS) which are not consistent in its recording rules. The cluster prior information is derived from the sum of all farmsteads in a municipal (FSS) independently where the fields or herd sizes are located in reality. This is known as *situs principle*. In contrast, the GIS data are attributed according to the location of the plot. The treatment of this error is difficult, because it is part of the definition how to record the statistic. This error could be reduced by aggregating neighbouring municipalities based on their similarity as long as certain thresholds regarding minimum farm numbers and UAA are reached. The delimitation of appropriates rules has to be left for a further study. The second reason for deviation comes from the clustering algorithm and the moments derived for each production activity as prior. Due to the execution times of the estimation problem of several days on a grid cluster server it is not possible to extensively test different assumption as the normal error distribution for the prior information or the weighting of the error term.
Statistical offices in Germany and the EU record each year a lot of data highly relevant for land use policy assessment. Strict data protection rules limit the use and the research community is often forced to smooth data which results in a reduced accuracy (increases the aggregation bias) and often complicates the analysis. We have shown that clustering together with Bayesian estimation applied to different data sources yield a robust estimate of the statistical data at municipality level for land use. Nevertheless it is weird to know that all the invested time and resources could have been saved if the data would be public.

## REFERENCES


BKG (Bundesamt für Kartographie und Geodäsie) (2008): Basis-DLM (Digitales Basis-Landschaftsmodell) 1:25 000. Frankfurt / Main.

BMELV (Bundesministerium für Ernährung,. Landwirtschaft und Verbraucherschutz) (Various years): Statistisches Jahrbuch über Ernährung, Landwirtschaft und Forsten der Bundesrepublik Deutschland.

Britz W. and Witzke, P. (2008): CAPRI model documentation (2008): Available at http://www. capri-model.org/docs/capri_documentation.pdf, pp. 181.

Carrasco L. R., T. D. Harwood, S. Toepfer, A. MacLeod, N. Levay, J. Kiss, R. H. A. Baker, J. D. Mumford and Knight, J. D. (2009): Dispersal kernels of the invasive alien western corn rootworm and the effectiveness of buffer zones in eradication programmes in Europe. Annals of Applied Biology 156 (1): 63-77.

Dendoncker N., P. Bogaert, and Rounsevell, M. (2006): A statistical method to downscale aggregated land use data and scenarios. Journal of Land Use Science 1 (2): 63-82.

EUROSTAT (2009): Statistical disclosure control. Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/methodology/statistical_disclosure_control. Last update: 25.04.2009.

FDZ (Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Länder) (2010), AFID-panel agriculture, 1999, 2003 and 2007.

Gocht A, and Britz, W. (2010), EU-wide farm types supply in CAPRI - How to consistently disaggregate sector models into farm type models, Journal of Policy Modeling (33), 146-167.

Gocht, A. and N. Röder (2010): Recovering localized information on agricultural structure underlying data confidentiality regulations - potentials of different data aggregation and segregation techniques. Paper presented at 50th annual conference of the GEWISOLA "Möglichkeiten und Grenzen der wissenschaftlichen Politikanalyse", Braunschweig, 29.09. - 01.10.2010. URL: http://purl.umn.edu/93975.

Golan A., G. Judge and Miller, D. (1996): Maximum entropy econometrics, Robust Estimation with Limited Data. John Wiley, New York.

Heckelei T., T. Jansson, and Mittelhammer, R. (2008): A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models Discussion Paper 2008:2, University of Bonn, Available at http://www.ilr1.uni-bonn.de/agpo/publ/dispap/ download/dispap08_02.pdf

Henrichsmeyer, W., Cypris, C., Löhe, W., Meudt, M., Sander, R., von Sothen, F., Isermeyer, F., Schefski, A., Schleef, K.-H., Neander, E., Fasterding, F., Helmcke, B., Neumann, M., Nieberg, H., Manegold, D., and Meier, T. (1996): Entwicklung eines gsamtdeutschen Agrarsektormodells RAUMIS96. Endbericht zum Kooperationsprojekt. Forschungsbericht für das BML (94 HS 021), 1996, vervielfältigtes Manuskript Bonn/Braunschweig.

Kempen M., Britz W. and Heckelei T. (2005): A Statitical Approach for Spatial Disaggregation of Crop Production in the EU, In: Arfini Filippo (ed.). Modelling agricultural policies: state of the art and new challenges; proceedings of the 89th European Seminar of the European Association of Agricultural Economists (EAAE), Parma, Italy, February 3rd-5th, 2005. Parma : Monte Universita Parma Editore, pp. 810-830.

Leip A., Marchi G., Koeble R., Kempen M., Britz W. and 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.

Osterburg B., H. Nitsch, B. Laggner and W. Roggendorf (2009): Auswertung von Daten des Integrierten Verwaltungs- und Kontrollsystems zur Abschätzung von Wirkungen der EU-Agrarreform auf Umwelt und Landschaft. Arbeitsberichte aus der vTI-Agrarökonomie 07/2009. Braunschweig.

Verburg P., C. Schulp, N. Witte and A. Veldkamp (2006): Downscaling of land use change scenarios to assess the dynamics of European landscapes. Agriculture, ecosystems and environment 114: 39–56.

Witten, I.H. and E. Frank (2005): Data Mining: Practical Machine Learning Tools and Techniques. Elsevier. Amsterdam

You L. and S. Wood (2006): An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems 90: 329-347.

Zellner A. (1971): An introduction to Bayesian inference in econometrics. New York: Wiley.

Annex:

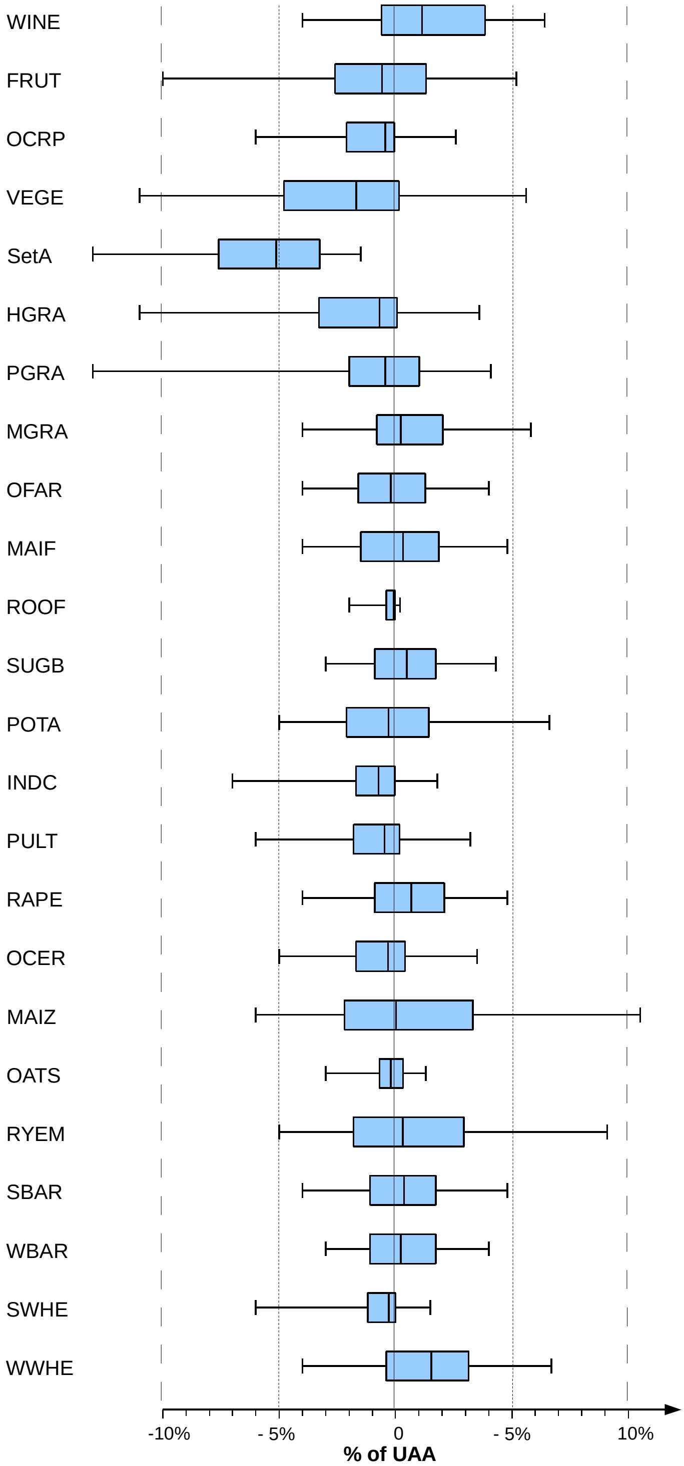


Figure 3: Boxplot of the deviations on municipality level for plant production activities in 2007

WWHE: Winter wheat, spelt; SWHE: Summer wheat, durum wheat; WBAR: Winter barley; SBAR: Summer barley; RYEM: Rye, and winter cereal mixes; OATS: Oats and summer cereal mixes; MAIZ: Grain maize (including CCM); OCER: Other cereals, triticale; RAPE: Rape and turnip rape; PULT: Pulses; INDC: Other oilseeds and industrial crops (hops, tobacco, etc.); POTA: Potatoes; SUGB: Sugar beet; ROOF: Other root crops (fodder beet, etc.); MAIF: Green and silage maize; OFAR: Grass on arable land (including all other fodder on arable land); MGRA: Meadow; PGRA: Pasture; HGRA: Rough pastures; SetA: Set aside; VEGE: Vegetables, strawberries; OCRP: Other plant production (flowers, nurseries, etc.); FRUT: Fruits (without strawberries); WINE: Wine

Source: FDZ, own calculation.