A new GIS toolbox for integrating massive heterogeneous GIS data for land use change analysis

by Birgit Laggner and Natascha Orthen

Thünen-Institute of Rural Studies, Braunschweig, Germany. `birgit.laggner@ti.bund.de`

Abstract

Agricultural land use in Germany and related impacts on the environment and the use of natural resources are key research topics at the Thünen-Institute of Rural Studies. As spatial context is essential for the analysis of causal connections, GIS data regarding all necessary information was gathered during different research projects and prepared for processing in a database. In particular, the Integrated Administration and Control System, which was available for certain project purposes for several Federal Laender and years, serves as a very detailed data source for agricultural land use. We use different Open Source GIS software like PostgreSQL/PostGIS, GRASS and QuantumGIS for geoprocessing, supplemented with the proprietary ESRI product ArcGIS. After introducing the used input data and the general processing approach, this paper presents a selection of geoprocessing routines for which Open Source GIS software was used. As an exemplary 'use case' for the conclusions from the consecutive statistical analysis, we summarize impacts of increased biogas production on agricultural land use change highlighting the trend in biogas maize cultivation and the conversion of permanent grassland to agricultural cropland.

Keywords: PostGIS, GRASS, geoprocessing, multiple data, Open Source, pre-processing of data, environmental impacts, biogas maize, permanent grassland, Germany.

1. Introduction

Agricultural land use and land use changes are related to environmental impacts like groundwater pollution, greenhouse gas emissions or loss of biodiversity (Schramek et al. 2012). These spark political efforts which result in legal acts (e.g. Nitrates Directive, FFH Directive). Agricultural subsidies or legal restrictions add to the driving forces influencing the usage of agricultural land, while land use must suit the natural conditions in order to optimize farm income.

The Thünen-Institute of Rural Studies analyzes land use change in agricultural areas of Germany focusing on environmental impacts and the effects of legal regulation with research projects like "Evaluation of the Common Agricultural Policy from a nature conservation point of view" (2008-2009) (see Nitsch et al. 2012) and "Analysis of land use change and development of methods for identification and quantification of measures for greenhouse gas abatement in the agricultural sector (parts I and II)" (2009-2013). A broad variety of GIS data is gathered in order to build a database for statistical analysis. The data contains official national and federal digital maps regarding land use, natural conditions and protected areas.

The central objective described in this article relates to the development of a GIS toolbox enabling the processing and analysis of massive and heterogeneous geodata useful for the statistical analysis of land use change patterns. The GIS toolbox subsumes all necessary processing steps for data preparation and for the integration of all input datasets into a combined dataset preserving relevant information on every site.

We base the development of the GIS toolbox on several conceptual considerations, whose applicabilities are to be proven as one result of the development process. Decisions with respect to the conceptual approach of the GIS toolbox include the choice of the main data format and the software environments as well as the definition of the necessary processing steps.

A conceptual precondition for the development of the GIS toolbox is the decision to chose a vector polygon approach as the main data format. This decision is motivated by the aim to preserve as much information as possible and necessary from the input data. Mostly all following conceptual decisions depend on this precondition. The related research challenges concern the feasibility of this approach within the given circumstances as well as the evaluation of its expandability and appropriateness. We are aware

of the uncertainties caused by the heterogeneous resolutions of the input datasets, which have to be considered at the discussion of the results of the statistical analyses.

Another conceptual challenge relates to the applicability of a modular design for complex processing steps against the background of the data heterogeneity. While a modular structure would reduce the effort of recurring tasks (addition of data updates, processing of similar datasets for different Federal Laender), even data presumed to be very similar differ in structure and content between different dates or origins.

The necessity to find a balance between rivaling requirements when it comes to software selection represents the next conceptual challenge. First, the chosen software has to fit into the existent infrastructure in order to allow easy usage and to benefit from existing abilities. Second, the efficiency of different software solutions at handling specific tasks on massive geodata has to be considered. Third, the number of interfaces has to be minimized. In this context, it was important to evaluate the use of Open Source software as an alternative to the proprietary software ArcGIS for the intended GIS toolbox.

Challenges with respect to data heterogeneity include handling of different data formats (vector and raster data, not georeferenced data) and spatial references as well as the extraction of relevant data from differently structured sources. All datasets contain a variety of attribute information while the quality of their documentation is varying. Additionally, some input datasets consist of several files in a database-like structure. Concerning the geoprocessing aspect, the attribute-related heterogeneity mostly affects the elimination of overlapping polygons. Further geoprocessing issues derive from sliver and invalid polygons.

The analysis of the expansion of biogas production based on renewable resources from agriculture, an example for land use change induced by political actions, illustrates the use of the GIS toolbox. Superficially, the substitution of fossil energy resources with renewable bioenergy seems like a good way to reduce greenhouse gas emissions. Though, the cultivation of bioenergy crops competes with food production for the same agricultural areas and may increase environmental impacts with respect to, e.g., biodiversity and water quality.

2. Material and Methods

2.1. Input Data

For the analysis of agricultural land use, the IACS (Integrated Administration and Control System) is a very detailed data source. It is available for project purposes for several years for seven Federal Laender of Germany (see Table 1). Since 2005, the IACS combines GIS data on the field scale with land use and land use related information. Because of its application for the administration of agricultural subsidies, the IACS contains only areas for which subsidies have been claimed. Outside this setting, information about land use is complemented by the German Digital Basic Landscape Model (Basis-DLM) which provides nationwide high resolution data with an up-to-dateness of 5 years (BKG 2011). The years 2000, 2005 and 2010 of the Basis-DLM have been selected as input datasets.

As measures of the natural conditions influencing the agricultural land use, general soil maps on different scales, a digital elevation model (DEM, 25 m raster data) and 1 km raster climate data (precipitation, temperature, sunshine duration, transpiration) are available. Within certain protected areas, e.g. nature protection areas, drinking water protection areas and flood areas, land use is legally restricted. In order to identify areas of special sensibility and high nature value, we obtained datasets on organic soils as well as habitat and biotope types.

Most input datasets are vector data and are available as shapefiles. The climate raster has ASCII format and the DEM raster has GRID format.

The available official datasets derive from many different sources due to the federal system and the distinct topics related to specialized governmental organizations. Hence, datasets differ with regard to accuracy, data format and significance of the contained attribute information. As a result, every dataset has to be treated individually.

The growing number of biogas plants represents an economic incentive for the cultivation of biogas maize (Osterburg 2010). According to the Renewable Energy Law (EEG 2012), the Federal Network Agency annually provides a list of all biogas plants, including information on capacity, produced energy and location. For the described case, we applied the 2009 list. Furthermore, we included a dataset with georeferenced nationwide address data in order to integrate the address data of the biogas plants into the GIS database.

Table 1: Input Geodatasets.

Dataset, year of publication, availability	Data source	Data volume (number of geometries)

Digital Basic Landscape Model (avail. for all Laender)
    2000
    2005
    2010
	BKG	up to ~2 Mio.
		per year and Federal Land

IACS
    2003 (RP)
    2005 (BB, BY, NI, NW, RP)
    2006 (BB, BY, SH)
    2007 (BB, BY, NI, NW, RP)
    2008 (BB, BY, MV, NI)
    2009 (NI, SH)
    2010 (SH)
	State Ministries of Agriculture	up to ~2 Mio.
		per year and Federal Land

Soil maps Soil map of Germany 1:1 100 000 (avail. for all Laender) Soil map 1:300 000 (BB) Soil map 1:200 000 (incomplete available for all Laender, complete for SH) General geological map of Germany 1:200 000 (available for all Laender) Soil map 1:50 000 (NW, NI) Medium-scale agricultural mapping 1:100 000 (BB, MV) Maps of geomorphology, soil function and soil potential (MV) Soil inventory map (SH)	BGR LBGR BGR BGR NW: NRW Geological Service, NI: LBEG BB: LBGR, MV: LUNG LUNG LLUR	5 681 (nationwide) 11 534 11 352 (SH) 270 288 (nationwide) 138 247 (NW) 90 316 (NI) 22 076 (BB) 19 338 (MV) up to ~12 000 364 367
Digital elevation model (available for all Laender)	BKG	~888 Mio. raster tiles
Raster climate data (available for all Laender)	German Weather Service (DWD)	566 364
Protected areas *Water protection areas* Drinking water protection areas Mineral spring protection areas *Flood areas* *Nature protection areas* Nationwide dataset of several conservation areas (sites of Flora-Fauna-Habitat Directive, special protection areas for birds, nature protection areas, landscape conservation areas, natural reserves, national parks, biosphere reserves) Biotope and biotope type mapping, habitat types of Flora-Fauna-Habitat Directive (BB, BY, NI, NW, RP, SH) Peatland maps (BB, BY, MV) Areas related to European Water Framework Directive (BB, NI)	Various federal State Offices Various federal State Offices Federal Agency for Nature Conservation (BfN) Various federal State Offices Various federal State Offices Various federal State Offices	up to ~10 000 per Federal Land up to ~48 000 per Federal Land ~1 Mio. up to ~1 Mio. per Federal Land up to 33 550 5341
Administrative units (2008)	BKG	12 533
List of biogas plants (nationwide, 2010)	Federal Network Agency	920 878 lines
Georeferenced Address Database (2010)	BKG	22 Mio.

Abbreviations:

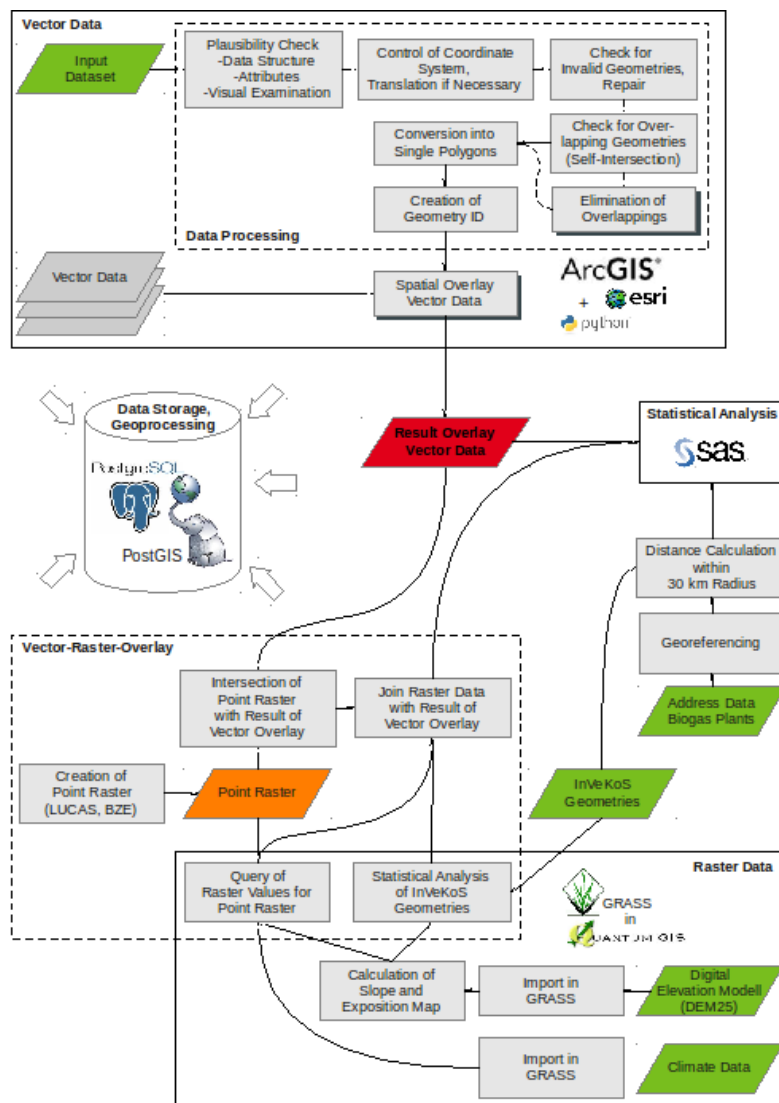|					
BB	*Brandenburg*	*BGR*	*Federal Institute for Geosciences and Natural Resources*	*LUNG*	*State Agency for the Environment, Nature Conservation & Geology Mecklenburg-West Pomerania*
BY	*Bavaria*				
MV	*Mecklenburg-West Pomerania*				
NI	*Lower Saxony*	*LBGR*	*State Office for Mining, Geology and Minerals of Brandenburg*		
NW	*North Rhine Westphalia*			*LLUR*	*State Agency for Agriculture, Environment and Rural Areas Schleswig-Holstein*
RP	*Rhineland-Palatinate*				
SH	*Schleswig-Holstein*	*LBEG*	*State Authority for Mining, Energy and Geology Lower Saxony*		
BKG	*Federal Agency for Cartography and Geodesy*				

Figure 1: Schematic Overview on the Processing Workflow and the Used Software Products.

2.2. Data Processing – Conceptual Approach

The processing of GIS data is intended to integrate all available and useful information into a combined database ready for statistical analysis. Because of the differing data models of vector and raster data, processing divides into a vector part and a raster part at first and requires a combination of both for full integration.

Most input data are vector data. This and the aim to preserve as much accuracy as possible lead to the decision to use a spatial overlay of all vector data as the main database. Area sums calculated for the resulting geometries build the basis for statistical analysis. Due to this decision, the input datasets have to meet a set of basic conditions. At first, all datasets must not contain larger overlappings, because they would generate errors in area calculation. Further, the extent and content of any dataset has to be complete within nation boundaries or those of the analyzed Federal Laender. With respect to processability, it has to be assured that the vector geometries are valid.

Besides the cleaning of overlapping polygons and the repairing of invalid geometries, standard data preparation procedures include transformation of multi polygons into single polygons for identification purposes and harmonization of spatial reference systems. Figure 1 presents an overview scheme of all

processing steps.

Another issue deriving mainly from the heterogeneity of data relates to the appearance of so called sliver polygons – small, elongated polygons at the borders of overlaying polygons with similar but not identical boundaries (see Section 3.2.3).

After the processing of vector data is finished, we have to link the information of the raster datasets to the resulting database (see Section 3.4.2). Additional information such as the situation of biogas plants are integrated into the database by their relation to certain datasets like the IACS data (see Section 3.4.1).

The results chapter describes the implementation of the most important of these steps in more detail with a focus on the implementation in an Open Source environment.

As a part of the statistical analysis, a relation is established between areal values resulting from GIS processing and useful attribute data. In order to identify transition areas where agricultural land use changes have taken place, an explicit assignment and grouping of spatially related IACS land parcels is necessary for the comparison of different years.

2.3. Used Software Products

Not only the number of datasets, but also the size of some of the input data and the resulting database make processing challenging. During consecutive projects, we tested different approaches. Proprietary ESRI product ArcGIS being the standard software in use at our research facility, it was our starting point and our reference for testing. Demands for alternative software resulted from insufficiencies in table management (laborious and very slow at massive data), from increasing failures at overlay operations for massive geodata without the possibility to track the causes and to find a workaround and from slow operation times concerning the combined processing of raster and vector data. The already high license costs of ArcGIS lead to a focus on Open Source software at the search for alternative solutions.

Currently, the data is stored in a PostgreSQL database with a PostGIS extension (PostgreSQL 8.4.5, PostGIS 1.5). This solution guaranties high-performance data access to all project team members while enabling them to execute geoprocessing within the database.

GRASS 6.4, mostly accessed by us in its form as a QuantumGIS extension, is a Open Source software specialized in high-performance raster processing. Therefore, we chose this environment to handle raster data sets like the digital elevation model. In different data manipulation steps it is possible to link the existing raster data with the available vector data.

Even if the mentioned Open Source alternatives work really well, we prefer proprietary ESRI product ArcGIS for combining all vector data available for a Federal Land with a spatial overlay. Until now, this is the most effective possibility to handle this manipulation step (see Section 3.3). During project time, an upgrade from ArcGIS 9 to version 10.1 has taken place. For statistical analysis, we use proprietary SAS 9.2.

3. Results

3.1. Preliminary Results regarding the GIS Toolbox Concept

Concerning the general development and architecture of the the GIS Toolbox, we pursue a strategy of handling the heterogeneous massive geodata within an adequate time frame. The individual approach depends on the requirements and restrictions defined by the project objectives as well as the available data and software.

During development, we often tested different approaches for individual processing modules in order to select the most effective solution (see following sections). In order to limit development effort, we preferred solutions existent or realizable within our software infrastructure (ArcGIS, PostgreSQL/-PostGIS). Criteria for the evaluation of the approaches include feasibility, satisfactory results, processing time and adaptability (for additional datasets or updates). At these decisions, the interactions between software features and the different possibilities for the implementation of certain processing steps have to be considered. Based on the subsequent projects, we advanced the concept of the GIS toolbox continuously.

The system design visualized in Figure 1 therefore represents a preliminary result of the current state and will possibly be enhanced in future projects.

3.2. Elements of GIS Toolbox – Data Preparation

3.2.1. Repair of Invalid Geometries

Invalid geometries occur in many original input datasets because of digitalization errors and different definitions of valid geometries in ArcGIS and Post-GIS. Further, invalid geometries can result from geo-

processing routines. However, most geoprocessing routines are very sensible to invalid geometries. This necessitates the application of a repair tool, which was not available as a genuine function until PostGIS 2.0.

In general, approaches like *ST_Buffer*, *ST_SnapToGrid* or the *CleanGeometry* pl/pgsql function created by Horst Düster (2008) were suggested for repairing invalid polygons. Although these approaches work for a variety of invalid polygons, they fail in some cases producing empty or NULL geometries. Therefore, we extended the *CleanGeometry* pl/pgsql function (Düster 2008). The new function combines the described approaches and complements them with additional repair operations like reordering and eventually selecting the nodes for polygon build. Further, we implemented success testing in order to avoid empty geometry entries. Unfortunately, there are still cases, which the function is unable to handle properly. The most frequent invalidity reason is wrong ring order (interior ring is listed before the exterior ring). A plausibility check of the results is therefore necessary to ensure the reliability of the dataset.

3.2.2. Overlapping Polygons

The elimination of overlapping polygons is an important and time-consuming preparation step. We start the procedure with a self-intersection of the dataset, resulting in the identification of all overlappings within it. Only overlappings larger than 1 m^2 are selected for further processing. They are analyzed regarding the reasons for overlapping and the possibilities for their revision. In general, the following types of overlapping polygons occur:

1. duplicates:

    (a) complete duplicates – geometry and attributes, both are similar

    (b) geometrical duplicates – geometries correspond, but attributes are different

2. partly overlapping polygons

    (a) corresponding attributes

    (b) differing attributes

In case of complete duplicates (a.1), we delete one of the polygons. Partly overlapping polygons with corresponding attributes (b.1) are aggregated with PostGIS *ST_Union*. If the attributes of the overlapping polygons differ (a.2, b.2), another approach is used:

First, we analyze the attribute data to figure out if some sort of priority ranking exists enabling us to prefer one geometry. This could concern the zoning in a drinking water protection area, or timestamps in the attribute table making it possible to identify the current or valid entry. Then, with *DELETE* (geometrical duplicates (a.2)) or *ST_Difference* (partial overlappings (b.2)), we cut off geometries of lower priority. However, frequently no information on priorities is provided helping to select which geometries to keep. Then, for partial overlappings (b.2), the only solution is to aggregate them with *ST_Union*. Geometrical duplicates (a.2) are reduced to a single geometry without attributes. We document the relation between the resulting geometries and the original attributes in a translation table. This attribute related approach does implicate the design of tailormade solutions for almost every dataset.

3.2.3. Removing Sliver Polygons in a Timeline Dataset

As mentioned, the IACS datasets are available for several years making it possible to create a timeline of agricultural land use. According to changes in ownership or usage structure and as a result of the correction of the digitalized field borders, the polygons are not static over time. The shifting of field borders between the years leads to numerous sliver polygons after spatial overlay.

Sliver polygons expand a dataset without really adding information. Furthermore, they are very vulnerable for geoprocessing exceptions in later overlay operations. Therefore, we developed a pl/pgsql-function with the aim to identify and eliminate sliver polygons resulting of a spatial overlay. We tested the function for the Federal Land Lower Saxony for which 4 years of IACS data are available (2005, 2007, 2008, 2009).

The identification of sliver polygons within the function is based on the compactness index (cmp) describing the relation of the polygon area (*A*) to the area of a circle with the same perimeter (*U*) as the polygon (Nakos 2001). It relies on the fact that circles represent the geometric shape with maximized ratio of area to perimeter.

$$cmp = \frac{4\pi A}{U^2} \quad (1)$$

(Nakos 2001)

In the pl/pgsql function, polygons are characterized as sliver polygons if their cmp is smaller than 0.13, a value which is derived from the definition of slim, very elongated polygons (Nakos 2001). As additional conditions, the function contains upper and

lower boundaries regarding polygon area. Polygons smaller than 10 m^2 are always classified as a sliver polygon. We chose the upper boundary of 150 m^2 to prevent sliver polygon classification for polygons of slim shape but large area.

The elimination of sliver polygons in the IACS timeline relies on the assumption that sliver polygons are mostly the result of corrections of the field border digitalization. In this case, the most recent geometry should form the best representation of the actual field border. Therefore, we adjust sliver polygons according to the course of the border line in the most recent dataset.

In a timeline consisting of four years, the function for the identification and elimination of sliver polygons must be executed for each overlay operation. Although, removing sliver polygons has advantages for further overlay operations and statistical analysis, the pl/pgsql function did not yet become part of the general data processing because of its time consuming application.

3.3. Core of GIS Toolbox - Spatial Overlay of Vector Data

After data preparation, we combine all vector data available for a Federal Land with a spatial overlay. For this operation, we use ArcGIS function *UNION* which implies the export of all datasets in .shp format and reimport of the result into the PostgreSQL database again. As ArcGIS 10.1 provides a PostgreSQL database client library enabling direct communication between ArcGIS geodatabases and PostgreSQL, it has to be tested if this time consuming im- and export proceedings are still necessary in the future.

In the meantime, we conducted experiments to implement the functionality of ArcGIS *UNION* in the PostgreSQL environment using pl/pgsql-functions. The result can be found as one of the examples of spatial SQL in the OSGEO User Wiki (Laggner 2011). It consists of a combination of PostGIS functions together with routines for exception handling and data sub-selections in order to speed up processing time. The approach is based on a discussion in the PostGIS users mailing list (postgis-users 2007) and has proven to be most effective in the current software environment (PostgreSQL 8.4, PostGIS 1.5). The overlay dataset is composed of the results of five separate steps:

- intersection (clips the intersecting parts of both input datasets)

- difference A-B (clips all non-intersecting parts of dataset A belonging to intersecting geometries)

- difference B-A (same as above, but vice versa)

- remaining A (selects all geometries of dataset A which are not intersecting with geometries of dataset B at all)

- remaining B (same as above, for dataset B)

Although the PostGIS/PostgreSQL implementation does work, there are still some issues leading to the preference of ArcGIS *UNION*: First of all, only two datasets can be processed at a time which is wearisome if there are at least 10 datasets per Federal Land. This is aggravated by the effort of repairing the invalid polygons after each overlay step and the special treatment of geometries causing geoprocessing exceptions. Also, the resulting geometries contain more sliver polygons and complex polygon boundaries (e.g., unnecessary linestrings) than ArcGIS *UNION* results because of the very high accuracy of the geoprocessing functions in PostGIS.

For the approach using ArcGIS *UNION*, we developed a Python script to control the geoprocessing. The script analyzes the spatial extent of the input data and subsequently calculates the number of parts to which the data should be broken down in order to optimize processing time. Afterwards, the script manages all necessary geoprocessing steps until the spatial overlay operation is finished.

3.4. Special Extensions of GIS Toolbox - Additional Data

3.4.1 Distance calculation biogas plants – fields

Nationwide address data of biogas plants opens the possibility of considering developments of land use changes resulting from the growing number of biogas plants. Besides the postal address, the dataset does not provide any other georeferenced information (coordinates or geometry). In order to georeference the biogas plants, we link their address data to the BKG database with georeferenced nationwide address data.

We use PostGIS for data manipulation. First, we transform the address data of the biogas plant table into a structure conforming with the table containing the georeferenced addresses. Corrections and error checking are a vital part of this step. Because of the heterogeneous enumeration of various address information within one field (person names,

street names, house numbers, plot numbers, etc.), it was impossible to extract precise address information such as street name and house number. Therefore, joining biogas address data with the georeferenced address database is mainly based on postcode and place name. For 35 % of the biogas plants, it was possible to use a substring representing the beginning of the street name as an additional join condition. The joining approach leads to the assignment of clusters of points matching the address join condition (1:n relation). For each biogas plant, we aggregate the matching points with *ST_Collect* and identify the centroid of the resulting multi-point geometry.

Availability of matching street names as well as differences regarding the place names between biogas plant and georeferenced address data are handled by a ranked joining procedure taking information quality into account. In the remaining cases without successful matches, we use the centroid of the postcode areas.

For the distance analysis, we calculate the distances between the agricultural areas enclosed in the IACS of 2009 and the biogas plants within a 30 km and a 15 km radius. The agricultural areas are represented by their centroids in order to obtain the average distance. We create the 30 km and the 15 km radius around IACS centroids with *ST_Buffer* and select the biogas plants located inside with *ST_Intersects*.

3.4.2 Processing of Topographical Raster Data in GRASS

The digital elevation model (DEM) is a 25 m raster dataset providing height information for Germany. In order to get more detailed information about the topography as an important local factor, we link the DEM data with the available vector data. We analyze the topography for three of the original seven Federal Laender (Lower Saxony, Bavaria and North Rhine-Westphalia).

Because of the amount of data, processing is very difficult and we tested various approaches. Currently, we add topographical information to the vector data in two separate processing steps. First, we extract DEM information for the agricultural acreage represented by the most recent IACS geometries on a highly detailed level. We chose a more condensed approach for the areas outside the IACS (point grid described below).

We conduct the data manipulation with GRASS 6.4. Besides altitude, slope and aspect are relevant for agricultural use. We derive raster maps

for the two last variables from the DEM with the r.slope.aspect module of GRASS. For the linkage between the topographical raster maps and the IACS polygons, we use the GRASS module *v.rast.stats2*. This module calculates univariate statistics like average, standard deviation, minimum, maximum, median or percentiles for all tiles of a raster map intersecting with overlaid vector polygons. The results are attached to the attribute table of the vector map.

In order to complement topographical information for the areas outside the IACS, we chose a different procedure. We generate a Germany-wide regular 100 m vector point grid. The points are based on the 1 km LUCAS grid (Eurostat 2013) and were densified to a 100 m grid. This approach intents to generate a sufficient number of samples for statistical analysis. Furthermore, we add additional points in order to meet the known sampling points of the German agricultural soil inventory (TI-AK 2013). This will allow to easily link information of the soil inventory in future.

We intersect the point grid with the polygons of the vector overlay in order to take a spatial sample of this pattern of land use and local conditions for the three Federal Laender which are topographically analyzed. In the next step, we attach the topographical information of the raster maps to each point of the selected grid matrix. We use GRASS (*r.mapcalc*) to combine vector and raster data.

3.5. Use Case: Biogas-induced Land Use Change Between 2005 and 2007 as an application for GIS approach

A recent development in Germany is the change in land use because of the expansion of biogas production based on renewable resources from agriculture. In this respect, an increase of the cultivation of maize as the dominant fermentation substrate was observed which is attended by a conversion of permanent grassland to agricultural cropland. This causes concerns regarding the impacts on nature and water protection goals which were basic objectives in the WAgriCo2 project (follow-up of EU-LIFE project „Water Resources Management in Cooperation with Agriculture"). The following findings are excerpts of the final project report (Osterburg 2010).

The changes in land use regarding the trends of maize acreage and the conversion of permanent grassland are analyzed for Lower Saxony based on IACS data for the years 2005 and 2007. The IACS data is set in relation to local characteristics such as soil types, topography, the establishment of nature

and water protection areas as well as the distance of the land parcels to biogas plants.

For the statistical analysis, land use changes are quantified as net changes because otherwise normal crop rotation would increase the total area evolved without reflecting real change.

Between 2005 and 2007, the area of maize increased by approx. 60 000 ha in Lower Saxony. The IACS data enables further analysis of these changes classified according to farm or land parcel properties.

From the distance analysis, an agglomeration effect for biogas maize cultivation in the vicinity of biogas plants is observable. 74 % of all biogas maize is cultivated within a radius of 5 km from a biogas plant, (95 % within 8 km radius).

Analyses of correlations with moisture classes and potential soil erosion caused by precipitation show expansion trends of biogas maize cultivation into sensitive areas. The expansions into erosive areas as well as into wet areas have to be considered as a potential concern regarding the future water and soil protection. The intensification of pressure on the wet areas is of special relevance because the often organic soils contribute to greenhouse gas emissions if the organic carbon is mineralized under intensive agricultural use (Nitsch et al. 2012).

Within the same period, a net shift of approx. 19 000 ha from permanent grassland to agricultural cropland has taken place in Lower Saxony. Maize has, with 59 %, the highest percentage of the cultures following the ploughing of grassland. Still, biogas maize is responsible for only 8 % of the conversion areas while other maize comprises the remaining 51 %.

When analyzing the protective effect of reserved areas for nature and drinking water protection aims, significant effects can be found on areas where grassland protection regulation is implemented and where participation in agri-environment measures takes place more frequently. It could be assumed, that some protection effects are caused by already less suitable agricultural conditions on the protected areas.

4. Discussion

The selection and development of adequate processing methods for relevant GIS data is defined as an important part of the research interests within all mentioned projects. The first part of the discussion therefore addresses the appropriateness of the chosen approach including the contribution of the Open

Source software and examines possible future options. In the second part, the discussion focuses on the findings of the 'use case' amongst other things pointing to some data uncertainties.

4.1. Evaluation of GIS Approach

We can state that the current geoprocessing approach leads to the aimed result of comprehensive GIS datasets ready for statistical analysis. However, experiences to date reveal the enormous effort needed for data preparation and processing. With a growing number of datasets through further updates with additional years, the overlay operation becomes more time consuming and results in a larger dataset. This means, the amount of information which can be analyzed together is limited. Different approaches for future handling are possible:

First, the dataset could be more or less fixed for a certain date with long update cycles. During an update, old contents are replaced by new ones in order to avoid excessive database expansion. Apart from the reduced usability because of possibly out-dated information, with this approach the complete, time consuming overlay operation might become necessary on a regular basis.

Second, all available datasets are prepared but overlay operations only take place for specific analyses with the needed subset of data. Advantages would derive from the highest achievable timeliness of data and the minimization of data size per analysis. Hence, this approach is only reasonable, if small data subsets are sufficient for most research questions. If too many datasets are to be included, the subsequent effort could in total even exceed that of the current approach.

Third, a complete change of the approach could be conceivable from the hitherto spatially inclusive, polygonal approach to a statistical sample based on a regular point grid. This approach would offer many advantages for the geoprocessing. It would dramatically reduce the effort of the overlay operation because no complicated polygons have to be processed. Topology exceptions due to invalid input or output polygons are avoided. Adding to convenience, the elimination of overlappings is limited to those affecting the point grid, and reference tables suffice as processing routines regarding the geometries. Furthermore, integration of raster data is easily manageable. For more detailed research aspects, especially with respect to IACS data, the point grid should be complemented with polygonal data in a way corresponding to the second approach.

Which approach will be pursued later on has to be decided within the context of future research projects also keeping software improvements in mind.

With respect to the Open Source software which was used for geoprocessing, it has been proven that it is suited for the processing of massive geodata. The processing steps executed with Open Source software were always finished within hours or days, which is acceptable seen the bulk of data. As a special strength, the PostgreSQL database facilitates storage and manipulation of many datasets, enabling data access for all project team members. PostGIS, the spatial extension for PostgreSQL, provides many geoprocessing functions. Compared to ArcGIS, the numerous possibilities to develop workarounds whenever the preferred solution fails, are an important advantage of PostGIS.

One major drawback of PostGIS is its sensitivity for topology exceptions. The instances of failing operations increase with growing numbers of geometries to be processed. In some cases, this requires a massive use of error trapping routines expanding the geoprocessing operations considerably. The sensitivity for topology exceptions is associated with the high accurateness of some geoprocessing functions. Thus, the addition of a scalable tolerance enabling the user to consider the resolution of a dataset would represent a desirable improvement.

A combination of different software is necessary at the given stage of used software products. The data transfer between them causes additional effort and parallel data storage. A better data interchange with the PostgreSQL database is established with the current version of ArcGIS 10.1. This enables the user to directly access the PostgreSQL database as a data source and storage location for inputs and results of geoprocessing routines in ArcGIS. In addition, new features in PostGIS 2.0, like the implementation of raster data processing and topological models for vector data, open prospects for possible future method adaptations.

A further point for discussion is the informative value of the input datasets. Having IACS data for seven Federal Laender at our disposal, a comparison of the results and the composition of a nationwide database would be desirable. Due to differences in availability between the Federal Laender, the overlay results are unique considering the integrated IACS years as well as specific input data. The latter is not only related to thematically specific datasets (e.g., peatland mappings) but also to the spatial resolution of datasets issuing the same subject. A good example is the variety of available soil maps (see Table 1). The only soil map with nationwide extent is the Soil Map of Germany 1 : 1 000 000. The Soil Map of Germany 1 : 200 000 is in preparation, but is not available in full coverage. Therefore, it is appropriate only for some of the Federal Laender. For some soil related issues, the Geological Map of Germany 1 : 200 000 is the best data source at national scale. Soil maps with finer resolution are only available from federal institutions and the acquired datasets differ in resolution (e.g., 1 : 300 000, 1 : 100 000, 1 : 50 000) and included attribute information.

Resolution differences arise between most of the input datasets. This is inevitable, if information regarding all relevant topics has to be assembled. In consequence, inaccuracies are to be expected from the result of the overlay operation, accumulating especially near polygon borders.

As another issue, input data represent information for a specific period in time. This means that they are limited concerning their comparability for different Federal Laender and the timeliness of their information.

The inaccuracies deriving from georeferencing have to be considered when working with the dataset of the biogas plants. Besides the described difficulties in matching the address data with the georeferenced dataset, the postal address of a biogas plant may not be consistent with its actual location. The generated point geometry can only be an approximation of the actual address.

4.2. Discussion with Respect to 'Use Case' Results

The analysis of land use changes induced by increased biogas production as presented in Chapter 4 bases substantially on the interpretation of IACS data. The main contribution of analyzing the data within a GIS consists in the spatial link of datasets with different subjects or time-references, allowing the identification of causal connections and temporal changes.

IACS provides the most detailed data available with respect to land use in agricultural areas. It contains information regarding the applying farm, the agricultural use and the applied subsidies. The land parcels are mapped with high accuracy because their extension has an effect on the granted subsidies (EC 2004). Nevertheless, the accuracy of IACS data regarding the identification of land use changes is limited.

Many IACS land parcels are aggregates of more than one field with different crops and owners. This leads to a certain fuzziness on the spatial allocation of land use changes. The identification of changes within a land parcel is only possible if an areal difference is reported regarding the agrarian use for this land parcel between 2005 and 2007 (Osterburg 2010). Land use shifts of a corresponding extent within a land parcel remain unnoticed. The described land use changes therefore represent minimum changes. However, the error induced by this inaccuracy is estimated to be negligible because of the small size of land parcels (3 ha on average) in Lower Saxony (Osterburg 2010).

Similar to the localization of land use changes within a land parcel, it is also difficult to track farm-related management changes on land parcels where more than one farm operates.

Furthermore, the differentiation between biogas maize and other maize can only be based on the information that energy crop aid was applied for a land parcel. In comparison with additional statistical data, 20 % of the expected biogas maize acreage could not be identified in the IACS data (Osterburg 2010).

As already discussed in Section 4.1, the georeferencing of the biogas plants is afflicted with uncertainties which are transmitted to the results of the distance calculation. Since the analysis shows the highest effect in the immediate vicinity of biogas plants, it should be kept in mind that in some cases the georeferencing error might be higher than the calculated distance.

Although the presented 'use case' focuses on a specific research aspect, the already developed database and GIS toolbox provide a very useful basis for further land use related research. Corresponding to the research interests of the Thünen-Institute of Rural Studies, other use cases might include the agricultural use of organic soils and the related effects on greenhouse gas emissions or the problem of nutrient contamination in groundwater and surface water.

5. Summary

The described geoprocessing approach is sufficient for handling massive and heterogeneous geodata and provides the targeted results. Nevertheless, the continuation of the approach is limited because of the long processing time and the massive amount of data resulting from data acquisition and processing.

The requirements of future projects will influence the decision on how the approach will change for their purposes.

Open Source software is able to substitute proprietary software like ArcGIS in most processing steps. The main objection concerning the performance of the used PostGIS version relates to the topological sensitivity of its geoprocessing functions.

Statistical analysis based on the processed geodata allows reliable statements regarding relevant phenomena in agricultural land use change. Hence, it has to be taken into account, that the heterogeneity of the available geodata together with the uncertainties inherent to the input data causes inaccuracies in the spatial overlay and the resulting area calculation.

The use case demonstrates how relevant variables can be analyzed together based on the available data and which conclusions can be drawn from the results. There is an increase in biogas maize cultivation probably leading to a regional reorganization. Direct cumulative effects of biogas plants are observed mainly within very short distances. The gain in biogas maize cultivation contributes to the pressure on sensible and valuable areas. A protective effect is strongest if explicit regulations exist regarding agricultural land use.

The use of the database and the GIS toolbox is not limited to a specific research aspect but can be transferred to other land use related issues.

References

Düster, H. (2008), 'cleanGeometry - remove self- and ring-selfintersections from input Polygon geome-

tries'. URL: http://www.kappasys.ch/pgtools/cleangeometry/cleanGeometry.sql (Download: 01.09.2008, Last Checked: 29.01.2013)

EEG (2012), 'Erneuerbare-Energien-Gesetz vom 25. Oktober 2008 (BGBl. I S. 2074), das durch Artikel 5 des Gesetzes vom 20. Dezember 2012 (BGBl. I S. 2730) geändert worden ist'. URL: http://www.gesetze-im-internet.de/bundesrecht/eeg_2009/gesamt.pdf (Last Checked: 06.02.2013)

European Commission (2004), 'Commission Regulation (EC) No 796/2004 of 21 April 2004 laying down detailed rules for the implementation of cross-compliance, modulation and the integrated administration and control system provided for in of Council Regulation (EC) No 1782/2003 establishing common rules for direct support schemes under the common agricultural policy and establishing certain support schemes for farmers'. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32004R0796:en:NOT (Last Checked: 29.01.2013)

Eurostat (Last Checked: 29.01.2013), 'Land use/cover area frame survey (LUCAS) – Introduction'. URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/lucas/introduction

BKG (Federal Agency for Cartography and Geodesy) (2011), *Digital Base Landscape Model – Basis-DLM*, Status of documentation: 20.09.2011. URL: http://www.geodatenzentrum.de/docpdf/basis-dlm_eng.pdf (Download: 29.01.2013)

Laggner, B. (2011), 'Set of functions for spatial overlay of 2 polygon tables'. URL: http://trac.osgeo.org/postgis/wiki/UsersWikiExamplesOverlayTables2 (Last Checked: 29.01.2013)

Laggner, B., Orthen, N. (2013), 'Verwendung räumlicher Massendaten bei der Analyse von Landnutzungsänderungen', *in proceedings of* 'Massendaten-management in der Agrar- und Ernährungswirtschaft, 33. GIL – Jahrestagung, 20. - 21. Februar 2013, Potsdam', pp.183–186.

Nakos, B. (2001), 'On the Assessment of Manual Line Simplification Based on Sliver Polygon Shape-Analysis', *in* '4th Workshop on Progress in Automated Map Generalisation (ICA, Beijing, China, August 2001)'.

Nitsch, H., Osterburg, B., Roggendorf, W., Laggner, B. (2012), 'Cross compliance and the protection of grassland – Illustrative analysis of land use transitions between permanent grassland and arable land in German regions', *Land Use Policy* 29, 440-448.

postgis-users (2007), 'Union of 7 datasets', thread in postgis-users mailing list archive. URL: http://lists.osgeo.org/pipermail/postgis-users/2007-August/016804.html (Last Checked: 06.02.2013)

Osterburg, B. (2010), 'Wirkungen von Biogasanlagen auf die Flächennutzung in Niedersachsen – eine Analyse von Daten des Integrierten Verwaltungs- und Kontrollsystems (InVeKoS)', Arbeitsberichte aus der vTI-Agrarökonomie 05/2010, Braunschweig, September 2010.

Schramek, J., Osterburg, B., Kasperczyk, N., Nitsch, H., Wolff, A., Weis, M., Hülemeyer, K. (2012), 'Vorschläge zur Ausgestaltung von Instrumenten für einen effektiven Schutz von Dauergrünland', *BfN-Scripten* 323, Federal Agency for Nature Conservation (BfN), Bonn.

TI-AK (Thünen-Institute of Climate-smart-agriculture), 'German agricultural soil inventory'. URL: http://www.ti.bund.de/en/startseite/institutes/climate-smart-agriculture/research-projects/german-agricultural-soil-inventory.html (Last Checked: 29.01.2013)