Systematic Data Mining into Land Consumption in Germany

Alfred Ultsch, Odette Kretschmer and Martin Behnisch

Abstract

This paper presents a systematic approach for discovering comprehensible, valid, potentially innovative and useful structures in multivariate municipality data. Techniques from statistics, machine learning and data mining are applied in logical consecutive steps. This allows the validation after each step and the generation of important results during the investigation. In particular, the approach does not end with a clustering of the data. If a structure has been identified, then the question is posed: what does the cluster mean? Symbolic machine learning methods are used to produce an understandable description of the clusters in form of classification rules. The approach is demonstrated on a data set of nine variables concerning land consumption of all municipalities in Germany. Selected results demonstrate the capacity of the method.

A. Ultsch
Philipps-Universität Marburg, Datenbionik FB 12, Hans-Meerwein-Straße, 35032 Marburg, Germany
Email: ultsch@informatik.uni-marburg.de

O. Kretschmer (Corresponding author) • M. Behnisch
Leibniz Institute of Ecological Urban and Regional Development, Weberplatz 1, 01217 Dresden, Germany
Email: o.kretschmer@ioer.de

M. Behnisch
Email: m.behnisch@ioer.de

1. Introduction

The rededication of open space into settlement and transportation area is a repeated subject of debate with international scope. In the case of Germany, one result of this debate was the formulation of a national program of measures created within the context of the national sustainability strategy in the year 2002. An essential contribution of scientific work within the political discourse is the provision of a better data basis for decision-making processes regarding smallest administrative levels. With respect to this a considerable demand exists in the precise quantitative description of land used for settlement and transportation purposes at a certain point in time as well as its development over years (land consumption) on small scales. Corresponding analyses mainly relate to selected regions of investigation or to spatial units on higher administrative levels (e.g. districts, federal states). Furthermore, spatially differentiating patterns of land consumption have not hitherto been brought out in detail by the use of multiple characteristics concerning land consumption structures. In addition, most of the recent approaches within this context end with the clustering partition and presentation of a narrative description of selected properties or spatial relationships

On account of the growing availability of georeferenced data as well as factual data, the opportunities to quantitatively explore characteristics of land consumption have clearly improved during recent years and allow multivariate analyses in high spatial resolutions. Modern methods from statistics and computer science subsumed under the labels “data mining” and “knowledge discovery” are needed to discover non trivial, novel, valid and useful structures in such multidimensional data. The contribution at hand seizes on these research needs and proposes a knowledge discovery approach for smallest administrative units using German municipalities as an example. A methodical sequence of steps that serve the detection and explanation of such patterns within multidimensional data is presented. The approach is demonstrated on a data set describing land consumption properties in all 11441 German municipalities. The obtained results comprise a spatial presentation of the identified municipality types as well as a description of their corresponding specific characteristics. New aspects connected to land consumption are regarded through the inclusion of variables such as daytime population density and trade taxes.

2. Land Consumption Data

The data used here (LandConsumptionData) consists of values for all 11441 German municipalities (data valid as of: 31.12.2010). The properties of each municipality are specified by 9 variables measured in the year 2000 and the year 2010, respectively (see table 1). Land-related data was provided by the official land use statistics in Germany. Data uncertainties of this set are discussed, for example, in Deggau 2008, Destatis 2013 and Krüger et al. 2013. Not land-related data include statistics on employees at the place of residence as well as place of work, on population numbers, on housing stock and on trade tax. This data was provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development.

Table 1. List of variables concerning land consumption (LandConsumptionData)

Variable	Description	Unit
[1] land consumption 2000	Settlement and transportation area in proportion to the municipal area	%
[2] land consumption 2010	Settlement and transportation area in proportion to the municipal area	%
[3] daytime population density 2000	Daytime population number in proportion to the area for buildings and associated open spaces	Persons /ha
[4] daytime population density 2010	Daytime population number in proportion to the area for buildings and associated open spaces	Persons /ha
[5] trade tax 2000	Municipal taxation revenue per resident	Euro / person
[6] trade tax 2010	Municipal taxation revenue per resident	Euro / person
[7] inhabitants 2000	Amount of inhabitants	persons
[8] inhabitants 2010	Amount of inhabitants	persons
[9] Municipal Area 2010	Municipal area	ha
Data sources:	Federal Statistical Office (variables [1] to [4] and [9]) and Federal Institute for Research on Building, Urban Affairs and Spatial Development (variables [5] to [8])	

3. Methodological Steps towards Knowledge Discovery

This chapter presents a structured approach to data mining, aiming at the identification of patterns in the data concerning land consumption on smallest administrative units in Germany. The proposed approach follows a sequence of distinctive steps. To ensure the correctness of each step a validation should be performed before proceeding to the next step. The analysis starts at the investigation of single variables and proceeds with the calculation of dynamics, i.e. the change of the values over one decade, towards the classification and clustering of the data. The following step of our data mining is the generation of possible explanations for the identified types of changes. At each of these steps new knowledge can eventually be detected. The approach is described by using the variable 'land consumption' as an example.

3.1 Individual Variables

The first step in the investigation of complex patterns in data is the analysis of each variable on its own. This starts with an assumption of the basic type of distribution for each single variable, usually given by the nature of the data (De Gruijter et al. 2006). Probably the most common assumption is that the variable is Gauss (Normal) distributed. This is a consequence from the „central limit theorem" (e.g. Rice 1995). In order to analyze the distribution of single variables the majority of studies in the research field of land consumption uses histograms provided as default method in Geographical Information Systems (e.g. BBR 2003, Job and Pütz 2006, Fina and Siedentop 2008). However, it is crucial how many bins have been used and where the range of the bins starts (Scott 1979). More systematic approaches to investigate the type of the distribution of a variable are to use two visualization tools: the quantile/quantile plot (QQplot) and a Pareto Density Estimation (PDE) of the probability density function. A QQplot compares two distributions by plotting their quantiles against each other. If the two distributions are of the same type the QQplot is a straight line. If the distribution is not a Gaussian and right skewed it is useful to apply a set of standard transformations given by Tukey's „ladder of power" (Tukey 1977).

The Pareto Density Estimation (PDE) is a second visualization tool that is particularly designed to detect different components of a distribution (Ultsch 2005). Fig.1 shows the PDEplot of the $\log_{10}$ of LandConsumption in the year 2010 (LogLandConsumption2010). By fitting a Gaussian (red in Fig. 1) to the distribution, it can be concluded that in first approximation LogLandConsumption2010 is LogNormal distributed (Limpert et al. 2001). The deviation of the PDE (blue in Fig. 1) at the sides of this central Gaussian, however, suggests additional components. Such components can be estimated, for example, as a sum of component distributions. The components of such a mixture model are investigated in more detail when classifying the changes of all variables (section 3.3). It could be determined that all the 9 variables in the LandConsumptionData (see Table 1) are in first approximation LogNormal distributed.

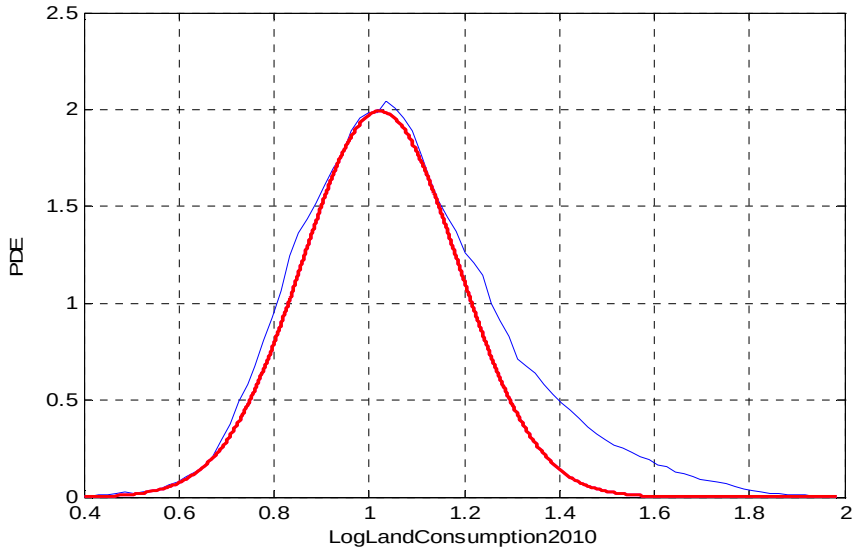

Fig. 1. PDEplot of LogLandConsumption2010 (blue) with a central Gaussian (red) (own computations)

3.2 Measuring Changes

One major focus of our approach is the comparison of the year 2010 to the year 2000, i.e. the characterization of the dynamics of the variables over a time period. In particular for the investigation of land consumption, relative percent change (RPC) is often used in Germany (Stadler 1979, Siedentop and Kausch 2004, Betzhold 2006, Dech and Kausch 2009, BBSR 2012,

Bieling et al. 2013) and also in international projects (Fulton et al. 2001, Ma and Xu 2010, Shrestra et al. 2012). The usage of RPC is, however, problematic in several aspects: the range is not symmetric (i), RPC ranges from -100% to +infinity, the temporal combination of RPC is unexpected (ii) and RPC has some other numerical problems (iii) (Ultsch 2009). In this study the measure "relative difference" (RelDiff) is used. It relates the change of a variable v (Δv) to the average of v_0 (at the time t_0) and v_1 (at a later time t_1). RelDiff(v) = 100*Δv /mean(v), where mean(v) = 0.5*(v_0 + v_1). RelDiff(v) has a symmetric range from -200 to 200. For the range of -25% to 25% RelDiffs are practically (up to a negligible error) equivalent to RPCs (Ultsch 2009).

An important consideration with respect to the change of a variable v in land consumption over time is the dependence of RelDiff(v) on the magnitude of v. To address this question a RAplot is used. The RAplot depicts the RelativeDifferences R (RelDiff(v)) vs the logarithm of the average A (log(mean(v)) of the variables. The RAplot is similar to the MAplot used to analyze microarry data in genetics (Dudoit et al. 2002). Fig. 2 shows the RAplot for the dynamics of LandConsumption. The linear interpolation (red in Fig. 2) shows that for the LandConsumption the change is practically independent of the magnitude of LandConsumption. However, the variance of RelDiffLandConsumption decreases with larger LandConsumption.

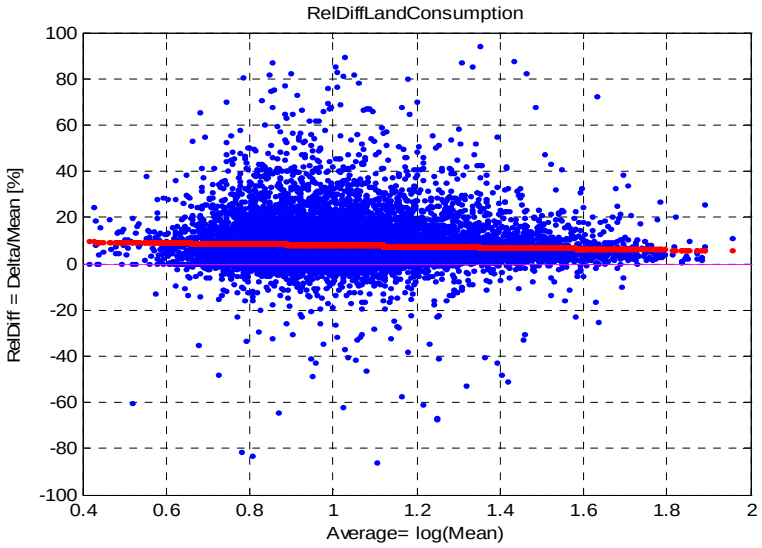

Fig. 2. RAplot of the change in LandConsumption, red is a linear interpolation (own computations)

3.3 Classifying Changes

A first qualitative exploration of the variables used here is derived from a closer inspection of the distributions using again the Pareto Density Estimation (PDE) (see section 3.1). If the variable under consideration is not a simple Gaussian, a sum of components (modes) is fitted to the probability density function. As components for the dynamics of land consumption (RelDiffs) a central Gaussian distributions (N) and two peripheral LogNormal (LN) distributions are used. Fig. 3 shows such a mixture model for the dynamics of the variable RelDiffLandConsumption. It can be seen that the relative differences follow a central Gaussian with a mean of 4.7%. This gives the expected value of land consumption dynamics from the years 2000 to 2010. The tails of the distribution can be modeled appropriately using LogNormal (LN) distributions. The validity of such a model can be assessed using statistical testing, for example, the chi-squared test. A QQplot allows also to judge the suitability of the model (see section 3.1).

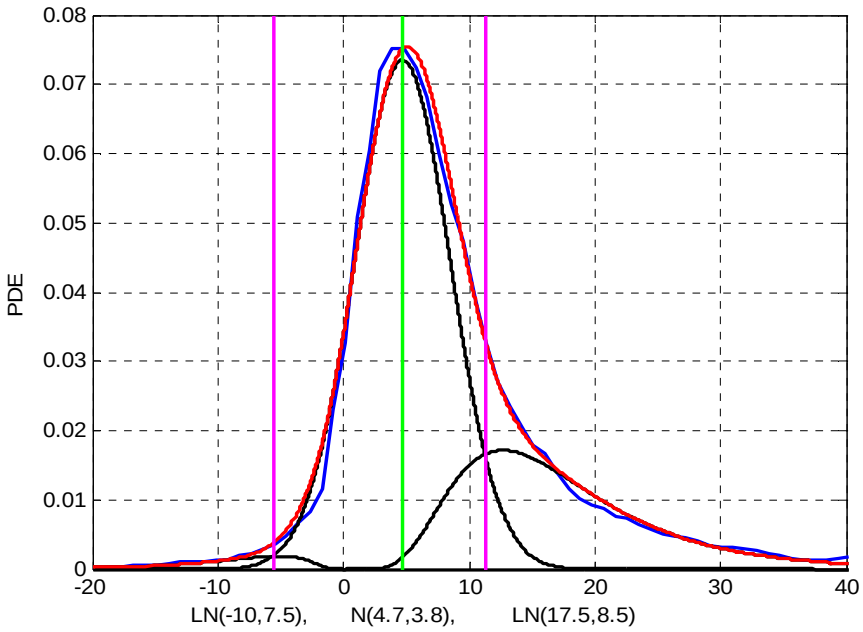

Fig. 3. PDEplot of RelDiff of LandConsumption (blue) with a three component model consisting of one central Gaussian and two LogNormal distributions for the tails (black). The combined model is shown in red. Decision Boundaries are in magenta. The mean of the central Gaussian is marked in green (own computations)

With such a model Bayes' theorem (e.g. Lee 2012) can be used to calculate posterior probabilities and also to calculate decision boundaries for a classification of the variable. This allows the classification of the variable RelDiffLandConsumption into three meaningful classes. The classification boundaries (-5.6 and 11.3) are at the intersections of the probability density functions of the model (vertical lines in Fig. 3). This model has a straight forward interpretation: the majority of 72% (i.e. the prior probability of the central component) of the change in land consumption is "as expected". "As expected" means an increase in land consumption of m=4.7% with a standard deviation of s =3.8%. This follows from the central Gaussian component N (4.7, 3.8). A subset of 26% of the changes in land consumption is, however, "higher than expected", which means they are larger than 11.3% and follow a LogNormal distribution LN (17.5, 8.5). A subset of 2% of the changes in land consumption are "lower than expected", i.e. less than -5.6, and follow a negative LogNormal distribution LN (- 10, 7.5).

The Bayes posteriors can be used for a qualitative scaling of the variables: the difference (DP) between the posterior of the right component minus the left component scales the values to the interval -100 to 100, such that DP is -100(100), when the variable clearly belongs to the left (right) component. DP is 0 when the variable clearly belongs to the central component of the model. Fig. 4 right shows this qualitative scale for changes in land consumption; class memberships are indicated by colors.

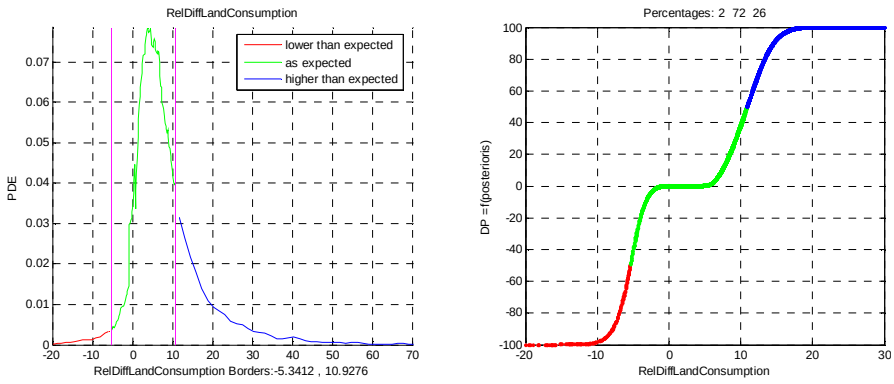

Fig. 4. Qualitative scaling for changes in land consumption (right graph) calculated as the difference in posteriors. Vertical lines as in Fig. 3. The colors red, green and blue indicate the class memberships to the components of the model (left and right graph) (own computations)

3.4 Exploring Structures in More Dimensions

The essential prerequisite for any analysis of high dimensional data is the definition of a valid distance (similarity, dissimilarity) between each of the cases. Typically Euclidean Distances are used as similarity function on the data in land consumption as well as settlement development studies, for example, in Frenkel 2004, Dech and Klein 2009 or Kroll and Haase 2010. However, several problems are posed by the unreflected usage of Euclidean Distances as a dissimilarity measurement. At first, it must be validated that the dimensions are not correlated. Non-correlations could be verified for all variables used in this study. A second concern when using Euclidean Distances is the scaling of the data. The rescaling to a qualitative scale (DP) as described in the last chapter has a large advantage: differences in the data that are clearly within one of the three classes are zero, since for such values the DP values are either -100, 0 or 100 (see Fig. 4 right panel). Therefore the Euclidean Distance on all the qualitative scaling of the variables using Bayes' posterioris measured in percent (QualitativeDistance), reflects perfectly the inner vs inter similarities of the different classes ("less than expected", "as expected" and "higher than expected").

3.5 Classifying and Clustering

The component modeling of the relative differences enabled to assign three classes in each of the four dimensions of the DynamicData (land consumption, daytime population density, trade tax, inhabitants). This allows a total of $3^4 = 81$ different classes. In the DynamicData only 75 classes are present. Of these the largest 6 classes contain more than 300 municipalities in Germany which contain in summary 8346 (73%) of all municipalities. For the remaining 69 classes containing 3095 municipalities a clustering using the QualitativeDistance described above has been performed. A Ward hierarchical clustering (Ward 1963) produced a dendrogram that suggests the existence of 5 clusters of patterns in the dynamics of land consumption. Together with the six classes it amounts to a total of 11 different types of patterns in the dynamics of land consumption in Germany from 2000 to 2010.

3.6 Explaining Land Consumption Changes

The answer to the question, what the types mean, is easy for the types that correspond to a single pattern of classes in the four dimensions of the Dynamic Data. Type 1, for example, represents those municipalities that follow

the average trend in all respects. The types of municipalities that have been generated by the clustering process (see section 3.5) contain a mixture of patterns of land consumption dynamics. To also understand these types a symbolic classifier can be used. Symbolic classifiers make use of explicit rules to classify the data. In our example we used CART (Breiman et al. 1984) on the clustering and extracted the rules from the decision tree of CART.

4. Results

This paper concentrates on the methodical approach to data mining and knowledge discovery, however we would like to present some results and thereby concentrate on the spatial distribution and meaning of the identified patterns. With the approach presented above, each of the 11 types shown in Fig. 5 can be given a meaning. Table 2 presents selected results focussing on such types of municipalities identified by the classification process.

Table 2. Description of selected types of changes in land consumption between 2000 and 2010 (L = land consumption, D = daytime population density, T = trade tax, I = inhabitants; * = 'as expected', + = 'more than expected', – = 'less than expected')

Type Nr.	Classified changes				Description
	L	D	T	I	
1	*	*	*	*	Common dynamics
2	+	*	*	*	Progressive land consumption
3	*	*	*	-	Rapidly shrinking municipalities
4	*	*	+	*	Increasing role of trade tax revenues
5	*	*	-	*	Diminishing role of trade tax revenues
6	+	-	*	*	Towards considerable inefficiency of land use

The largest identified type contains those municipalities (n = 4844) with all changes in "as expected" range. The expected changes are characterized by the mean and standard deviation of the central Gaussian: land consumption (m=4.7%, s=3.8%), daytime population density (m= -2.7%, s=10.2%), trade tax (m=47.5%, s=47.1%) and inhabitants (m= -0.8%, s=5.8%). This type is called "Common Dynamics" and is mostly located in the western parts of Germany (type 1). In contrast to some previous studies denoting the core areas of German agglomerations as 'hot spots' of land consumption

(e.g. Siedentop and Kausch 2004), Fig. 5 illustrates that most urban regions and their close vicinities developed 'as expected' between 2000 and 2010.

The second largest type containing 1343 municipalities deviates from the first class only in that respect that changes in land consumption are higher than expected. Such municipalities are therefore labelled as "Progressive land consumption" (type 2). It can be seen, that the problem of land consumption is obvious in the extended surroundings of urban areas or in peripheral regions. Progressive land consumption occurs in all federal states of Germany concentrating outside densely populated core cities and their vicinities. A similar development can be observed for some municipalities in rural areas of Bavaria, Rhineland-Palatinate, North Rhine-Westphalia and Schleswig-Holstein (type 6). These municipalities (n = 342) differ from type 2 only with regard to the development of their daytime population density. For type 6 the daytime population density was lower than expected. Therefore type 10 can be referred to as municipalities "Towards considerable inefficiency of land use"

The fifth largest type of municipalities (n = 706) is characterized by losses in population higher than expected (type 3) and therefore is labelled as "Rapidly shrinking municipalities". Under the assumption that all variables are independent from each other, it can be calculated how many municipalities are expected for each of the 75 classes and whether the observed class population deviates from this. Under the independency assumption type 3 should contain only 85 municipalities. The fact that this type exceeds the expected amount by more than ten times indicates an important trend. Such municipalities are almost exclusively located in former East German federal states. In view of a sustainable and sufficient use of land, particular attention needs to be paid at the adapted proportion of area supply and demand in such municipalities.

Further remarkable results refer to municipalities summarized by types 4 and 5, which are characterized by opposing developments in trade tax revenues. While type 4 (n = 589) contains municipalities with increases in trade tax revenues higher than expected, type 5 (n = 552) is composed of municipalities with increases in trade tax revenues lower than expected. The latter municipalities with a "Diminishing role of trade tax revenues" mainly occur in former West German federal states whereas municipalities with "Increasing role of trade tax revenues" can be found throughout the country. Direct relations between the amount of collected trade taxes and the intensity or productivity of land use cannot be derived. Nevertheless, the recognition of such so far unknown structures in municipal data is useful in order to obtain a deeper understanding of influential factors of land consumption processes through future detailed analyses (see section 5).

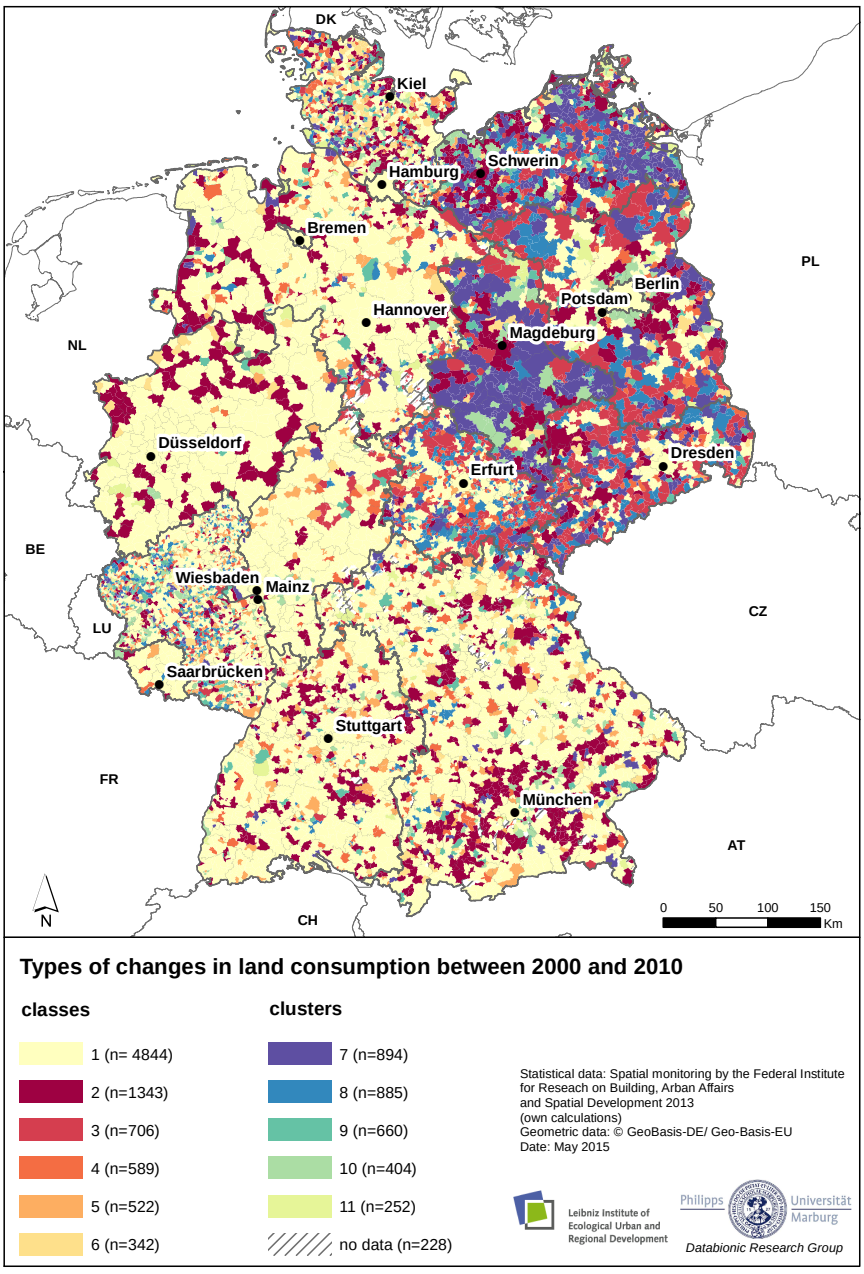

Fig 5. Location of the 11 types of dynamics in land consumption data. Types 1 to 6 are identified by the classification process (colors red to yellow) while the types 7 to 11 are identified by the clustering process (colors blue to yellow)

A method of giving a meaningful interpretation to the remaining types of municipalities identified by the clustering process is elaborated for our approach (see section 3.6). To give an example, the characteristics of a municipality belonging to type 7 are as follows:

- change in Land Consumption is high AND (i)
- change in Daytime Population Density is low AND (ii)
- change in Trade Taxes is high AND (iii)
- change in Inhabitants is low (iv)

Thus type 7 consists of municipalities with a consumption of land and trade tax revenues per capita mostly higher than average. The population related variables (daytime population and inhabitants) show that these municipalities are experiencing an above average loss in the number of people. Such municipalities could be called “More trade taxes, less open space and people”. The municipalities of this type occur in peripheral regions and concentrate in the eastern federal states such as Saxony, Thuringia, Schleswig-Holstein, Brandenburg, Mecklenburg-West Pomerania and Saxony-Anhalt. For a deeper understanding of the other four clusters descriptions should be elaborated in a similar way.

As has been described by Siedentop and Kausch 2004, German federal states having a strong administrative fragmentation as in Rhineland-Palatinate, Schleswig-Holstein and Thuringia, are characterized by a disperse land consumption pattern without noticeable spatial principle of order. The considerable mixture of different types of municipalities in these states emphasizes that the described approach classifies and clusters municipalities independently from the size of their administrative area.

5. Discussion

Land consumption is often observed or discussed as a whole regarding the entire settlement and transportation area (BMVBS and BBSR 2009, German National Strategy of Biological Diversity, Sustainability Strategy of the Federal Government). However it is well-known that the qualitative development of land consumption and urban form is not understood sufficiently in this way (BMVBS and BBR 2007). Furthermore, repeated calls have been made for increased investigation into issues of the multi-scale characteristics of land consumption systems as well as their temporal dynamics and the multi-scale effects of influential factors (Jörissen and Coenen 2007, BMVBS and BBSR 2009, Hersperger and Bürgi 2009). Spatial differentia-

tion and process description might promote the extraction of specific regularities, influential factors and indicators of land consumption. For example, it is relevant to analyze the structure and development of settlement and population (e.g. inhabitants, population density, settlement density, daytime population density, buildings under construction etc.) as well as important economical and socio-spatial parameters related to land consumption development (e.g. place of residence or work, financial strength of the public sector, attractiveness as a commercial location, number of commuters etc.). Against this background the identification and explanation of land consumption types is an important task (ESPON 2012, BMVBS and BBSR 2009).

The objective of this contribution was the presentation of a systematic step-by-step approach to data mining, starting from the raw data and ending with the discovery of some new data structures, which have been subjected to rigorous (statistical) testing. One advantage of this approach is the evaluation of interim results after each step. Any evaluation of results can be realized best through an interdisciplinary collaboration of computer science (data mining) and the spatial sciences.

Selected variables concerning land consumption are analyzed in high spatial resolution on the level of municipalities. The investigation of each variable on its own leads to an understanding of the general type of the distribution. One advantage is the discovery of different types of municipalities on the basis of a close data inspection. For example the distribution of land consumption could be described by a mixture of three components. When measuring changes in land consumption data relative difference (RelDiff) is presented as a superior index to common relative percent change measurements. For example the numerical values of RelDiff have a straight forward interpretation. Relative differences are numerically stable. In contrast to relative percent change the compensation for variance is much easier for RelDiff. When analyzing characteristics of variables in high dimensional space in former studies of land consumption, the selection of suitable distance measurements is often not regarded in its importance for the clustering. It should be emphasized that before using Euclidean Distances the correlation, scale and distributions of the variables need an intensive analysis to obtain valid clustering results. In this paper a qualitative scaling of the variables using Bayes' posterioris (QualitativeDistance) reflects perfectly the inner vs inter similarities of the different classes ("less than expected", "as expected" and "higher than expected"). A combined clustering and classification approach of all German municipalities could be carried out. Nearly three quarters of the German municipalities (n = 8346) could be described by the six largest classes that are formed on the basis of the qualitative modelling of the variables. To obtain a concise view, a clustering has been performed for

the remaining municipalities. This resulted in the identification of five additional types of land consumption changes.

Usually other approaches end with a clustering. This paper goes an important step further. The knowledge discovery approach implies a continuous and ongoing search for appropriate abstractions (e.g. “More trade taxes less open space and people” or “Progressive land consumption”). Such a meaningful description of structures is supported by machine generated explanations. The extraction of knowledge is therefore based on specific data properties and a deeper multidimensional description of land consumption characteristics as well as human interpretations or subsequent validations in mind of the involved spatial expert. The presented approach demonstrates how the applied processes help to understand variables separately and how to understand multivariate structures. The process of generating abstractions is presented in an exemplary way.

6. Conclusion

This paper demonstrates the scope and opportunities of a knowledge discovery approach applied to spatial data. It aims at the identification of structures in data concerning land consumption processes on the level of German municipalities. To find such structures the analysis proceeds step by step from the (1) separate inspection of all variables in the dataset and (2) the calculation and inspection of their changes over one decade to (3) a classification and clustering as well as (4) the generation of possible explanations as well as the assignment of semantics for the identified structures. As a result not only the clustering but also every single methodological step serves as a basis for the extraction of new and useful knowledge. The presented approach produces a map showing 11 types of dynamics in land consumption data. The investigation and mapping of the spatial distributions of variables associated with land consumption is essential in particular for the application of supporting monitoring systems as well as for the investigation of land consumption drivers. The analysis of factors potentially influencing the structures presented in this paper is going to be focused on in another publication soon coming up.

References

BBR (Bundesinstitut für Bauwesen und Raumordnung) (Ed.) (2003). Siedlungsstrukturelle Veränderungen im Umland der Agglomerationsräume. Bonn: *Bundesinstitut für Bauwesen und Raumordnung* (self-publishing).

BBSR (Bundesinstitut für Bau-, Stadt- und Raumforschung) (Ed.) (2012). Raumordnungsbericht 2011. Bonn: *Bundesinstitut für Bau-, Stadt- und Raumforschung* (self-publishing).

Betzhold, T. (2006). Trendwende beim Flächenverbrauch?, *Statistisches Monatsheft Baden-Württemberg*, 3(2006), 3-9.

Bieling, C., Plieninger, T., Schaich, H. (2013). Patterns and causes of land change: Empirical results and conceptual considerations derived from a case study in the Swabian Alb, Germany. *Land Use Policy*, 35(2013), 192-203.

BMVBS (Bundesministerium für Verkehr, Bau und Stadtentwicklung) and BBR (Bundesamt für Bauwesen und Raumordnung) (Eds.) (2007). Nachhaltigkeitsbarometer Fläche. Bonn: *Bundesinstitut für Bau-, Stadt- und Raumforschung* (self-publishing).

BMVBS (Bundesministerium für Verkehr, Bau und Stadtentwicklung) and BBSR (Bundesinstitut für Bau-, Stadt- und Raumforschung) (Eds.) (2009). Einflussfaktoren der Neuinanspruchnahme von Flächen. Bonn: *Bundesinstitut für Bau-, Stadt- und Raumforschung* (self-publishing).

Breiman, L., Friedman, J., Olshen, R.A., Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth.

Dech, S., Klein, R. (2009). *Entwicklung und Evaluierung eines fernerkundungsbasierten Flächenbarometers als Grundlage für ein nachhaltiges Flächenmanagement*. Schlussbericht zum Verbundvorhaben (self-publishing).

Deggau, M. (2008). Die amtliche Flächenstatistik – Grundlage, Methode, Zukunft. In G. Meinel, U. Schumacher (Eds.), *Flächennutzungsmonitoring – Grundlagen, Statistik, Indikatoren, Konzepte*. Aachen: Shaker Verlag.

De Gruijter, J., Brus, D., Bierkens, M., Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. Berlin, Heidelberg: Springer-Verlag.

Destatis (Statistisches Bundesamt) (eds.) (2013). Flächenerhebung nach Art der tatsächlichen Nutzung. https://www.destatis.de/DE/Publikationen/Qualitaetsberichte/LandForstwirtschaft/Flaechenerhebung.pdf?__blob=publicationFile. Accessed 13 Feb 2015.

Dudoit, S., Yang, Y.H., Callow, M.J., Speed, T.P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat. Sin*, 12(1), 111-139.

ESPON (European Spatial Planning Observation Network) (Ed.) (2012). EU-LUPA – European patterns of land use. http://www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/EU-Lupa. Accessed 02 March 2015.

Fina, S., Siedentop, S. (2008). Urban sprawl in Europe - identifying the challenge. In: Schrenk, M., Popovich, V.V., Engelke, D., Elisei, P. (Eds.), *REAL CORP 008: Mobility Nodes as Innovation Hubs* (Proceedings Real Corp: 13th International Conference on Urban Planning, Regional Development and Information Society).

Frenkel, A. (2004). Land-use patterns in the classification of cities: the Israeli case. *Environment and Planning B: Planning and Design*, 2004 (31), 711-730.

Fulton, W., Pendall, R., Nguyen, M., Harrison, A. (2001). Who Sprawls Most? How Growth Patterns Differ Across the U.S.. http://www.brookings.edu/~/media/research/files/reports/2001/7/ metropolitanpolicy-fulton/fulton.pdf. Accessed 24 Feb 2015.

Hersperger, A. M., Bürgi, M. (2009). Going beyond landscape change description: Quantifying the importance of driving forces of landscape change in a Central Europe case study. *Land Use Policy*, 26(3), 640-648.

Job, H., Pütz, M. (2006). Aktuelle Struktur und Entwicklung der Flächennutzung in Bayern. In: Job, H., Pütz, M. (Eds.), *Flächenmanagement – Grundlagen für eine nachhaltige Siedlungsentwicklung mit Fallbeispielen*

aus Bayern (pp. 84-97). Hannover: Akademie für Raumforschung und Landesplanung (self-publishing).

Jörissen, J., Coenen, R. (2007). Sparsame und schonende Flächennutzung, Studien des Büros für Technikfolgen-Abschätzung beim deutschen Bundestag. 20. Berlin: edition sigma.

Kroll, F., Haase, D. (2010). Does demographic change affect land use patterns?: A case study from Germany. *Land Use Policy*, 27(3), 726-737.

Krüger, T., Meinel, G., Schumacher, U. (2013). Land-use monitoring by topographic data analysis. *Cartography and Geographic Information Science*. doI: 10.1080/15230406.2013.809232.

Lee, P.M. (2012). *Bayesian Statistics: An Introduction*. Chichester: Wiley.

Limpert, E., Stahel, W.A., Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, 51(5), 341-352.

Ma, Y., Xu, R. (2010). Remote sensing monitoring and driving force analysis of urban expansion in Guangzhou City, China. *Habitat International*, 34(2), 228-235.

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis* (Second ed.). Belmont: Duxbury Press.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika* 66(3), 605–610.

Shrestha, M.K., York, A.B., Boone, C.G., Zhang, S. (2012). Land fragmentation due to rapid urbanization in the Phoenix Metropolitan Area: Analyzing the spatiotemporal patterns and drivers, *Applied Geography*, 32(2), 522-531.

Siedentop, S., Kausch, S. (2004). Die räumliche Struktur des Flächenverbrauchs in Deutschland. Eine auf Gemeindedaten basierende Analyse für den Zeitraum 1997 bis 2001. *Raumordnung und Raumforschung*, 1(2004), 63-49.

Stadler, R. (1979). Zum Problem des Landschaftsverbrauchs. *Baden-Württemberg in Wort und Zahl*, 4, 102-111.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Ultsch, A. (2005). Pareto Density Estimation: A Density Estimation for Knowledge Discovery, In: Baier, D., Wernecke, K.D. (Eds), *Innovations in Classification, Data Science, and Information Systems* (Proceedings of the 27th Annual Conference of the Gesellschaft für Klassifikation e.V. Brandenburg University of Technology, Cottbus, March 12–14, 2003).

Ultsch, A. (2009). Is log ratio a good value for measuring return in stock investments?, In: Fink, A., Lausen, B., Seidel, W., Ultsch, A. (Eds.), *Advances in Data Analysis, Data Handling and Business Intelligence* (Proceedings 32nd Annual Conference of the German Classification Society, Helmut-Schmidt-University, Hamburg, July 16-18, 2008).

Ward, J. H. (1963). Hierarchical Grouping to optimize an objective function. *Journal of American Statistical Association*, 58(301), 236–244.