Int J Appl Earth Obs Geoinformation 65 (2018) 46–56

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Research Paper

Local modelling of land consumption in Germany with RegioClust

Julian Hagenauer[a], Marco Helbich[b,*]

[a] Leibniz Institute of Ecological Urban and Regional Development, Weberplatz 1, 01217 Dresden, Germany

[b] Department of Human Geography and Spatial Planning, Faculty of Geosciences, Utrecht University, Heidelberglaan 2, 3584CS Utrecht, Netherlands

ARTICLE INFO

Keywords:
Land use
Drivers
Spatial clustering
Spatial heterogeneity
Regression
Germany

ABSTRACT

Germany is experiencing extensive land consumption. This necessitates local models to understand actual and future land consumption patterns. This research examined land consumption rates on a municipality level in Germany for the period 2000–10 and predicted rates for 2010–20. For this purpose, RegioClust, an algorithm that combines hierarchical clustering and regression analysis to identify regions with similar relationships between land consumption and its drivers, was developed. The performance of RegioClust was compared against geographically weighted regression (GWR). Distinct spatially varying relationships across regions emerged, whereas population density is suggested as the central driver. Although both RegioClust and GWR predicted an increase in land consumption rates for east Germany for 2010–20, only RegioClust forecasts a decline for west Germany. In conclusion, both models predict for 2010–20 a rate of land consumption that suggests that the policy objective of reducing land consumption to 30 ha per day in 2020 will not be achieved. Policymakers are advised to take action and revise existing planning strategies to counteract this development.

1. Introduction

Urbanization is occurring at an unprecedented rate (United Nations, 2015), whereby natural, agricultural, and forestry landscapes are converted into built-up areas. This irreversible anthropogenic process, commonly termed land consumption (Nuissl and Schroeter-Schlaack, 2013), has severe and long-lasting consequences for natural habitats, causing a loss of biodiversity, atmospheric pollution, etc. (D'Amour et al., 2016; Foley et al., 2005; Seto et al., 2011).

Land consumption is of particular relevance in Germany, which has one of the highest rates within the European Union (Kroll and Haase, 2010; Siedentop and Kausch, 2004). Even in areas with a declining population, the expansion of built-up areas continues across Germany (Haase et al., 2013). To prevent a further increase, the federal government implemented policies to limit the land consumption rate (LCR) to 30 ha per day up to the year 2020 (Die Bundesregierung, 2016). Although these policies are implemented at the national level, local authorities at the municipal level are still granted spatial planning autonomy by the government. As a consequence, restricting land consumption is a continuous process of reconciling interests across different administrative hierarchies (Jakubowski and Zarth, 2003; Malburg-Graf et al., 2007). Both economic incentives for stakeholders to promote the reuse of formerly built-up land (Borchard, 2011; Schultz and Dosch, 2005) and evidence-based local policymaking are prerequisites to counteract uncoordinated and excessive land consumption (Shafizadeh-Moghadam and Helbich, 2015). Long-term policies need to be founded on precise and data-driven land consumption models (Mas et al., 2014; van Vliet et al., 2016; Veldkamp and Lambin, 2001; Verburg et al., 2004).

While numerous studies deal with urban growth of specific metropolitan areas (Basse et al., 2016; Zeng et al., 2015), to the best of our knowledge, this is the first study addressing LCRs on a nationwide level using local regression-based modeling. A wide spectrum of approaches to modeling land use change has been proposed (Brown et al., 2004; Shafizadeh-Moghadam et al., 2017b; Triantakonstantis and Mountrakis, 2012; Verburg et al., 2004). Markov-cellular automata is a frequently applied model (Aburas et al., 2017; Arsanjani et al., 2013; Guan and Rowe, 2016; Li et al., 2017; de Noronha Vaz et al., 2012). However, the calibration and validation of cellular automata together with the development of transition rules (e.g., neighborhood definitions) is challenging and relies mostly on ad hoc definitions and heuristics (Li et al., 2017; Shafizadeh-Moghadam et al., 2017a). Further, Markov-cellular automata does not consider the underlying drivers. To circumvent these limitations, artificial neural networks (Shafizadeh-Moghadam et al., 2017b), random forests (Haas and Ban, 2014), and support vector machines (Samardzic-Petrovic et al., 2016), or a combination of these approaches (Arsanjani et al., 2013; Omrani et al., 2017), have been suggested to determine the effects of environmental and socioeconomic drivers on land consumption.

Although these machine learning approaches have methodical

* Corresponding author.
E-mail address: m.helbich@uu.nl (M. Helbich).

http://dx.doi.org/10.1016/j.jag.2017.10.003
Received 17 August 2017; Received in revised form 13 October 2017; Accepted 13 October 2017
Available online 19 October 2017
0303-2434/

advantages over statistical methods like regression (e.g., being free of assumptions concerning the input data) (Haykin, 2009), there is no consensus on how to integrate spatial autocorrelation and spatial heterogeneity in these models. Spatial autocorrelation means that areas that are close to each other are subject to similar land consumption processes (Anselin, 2010), while spatial heterogeneity (Fotheringham et al., 2003) means that the associations between land consumption and its drivers vary spatially (Shafizadeh-Moghadam and Helbich, 2015). As stressed by several studies (Anselin, 2010; Brunsdon et al., 1996), ignoring either issue can seriously bias results and might lead to false conclusions and inappropriate policies. Nevertheless, non-spatial regression models are still often applied to assess drivers of land use change (Achmad et al., 2015; Arsanjani et al., 2013; Hu and Lo, 2007; Van Dessel et al., 2011). Whereas spatial autocorrelation has received some attention in the literature (Ay et al., 2017; Dendoncker et al., 2007; Ku, 2016), far less has been devoted to spatial heterogeneity (exceptions are Maimaitijiang et al., 2015; Bagan and Yamagata 2015; Shafizadeh-Moghadam and Helbich 2015; Luo and Wei 2009). Explicitly modeling spatial heterogeneity is particularly important when conducting nationwide studies, where relationships can be *a priori* expected to vary across space due to different levels of regional economic wealth, environmental differences, and local planning policies.

Only a few models exist to explore spatially heterogeneity. A widespread approach is geographically weighted regression (GWR) (Brunsdon et al., 1996; Fotheringham et al., 2003). GWR uses the spatial distance of neighboring observations in order to estimate local coefficients. Both Luo and Wei (2009) as well as Shafizadeh-Moghadam and Helbich (2015) estimated GWR-based urban growth models and confirmed that GWR not only produces more accurate results compared to a global (i.e., study area-wide) regression model, but also reduces residual spatial autocorrelation. No less important, there is statistical evidence that the underlying drivers vary significantly across space (Hennig et al., 2015), which is paramount for place-based planning – a fact that global models had not uncovered (Achmad et al., 2015). Despite these appealing advantages, GWR is subject to methodological debate. In addition to the high volatility of the resulting coefficient surfaces, multicollinearity amongst the estimated GWR coefficients is reported (Griffith, 2008; Wheeler and Tiefelsdorf, 2005). While mild correlations obfuscate coefficient interpretation, strong correlations make it hardly possible to make a reliable separation of individual variable effects (Helbich and Griffith, 2016; Wheeler and Tiefelsdorf, 2005). Others note that GWR itself artificially introduces correlations among coefficients, though the input variables are uncorrelated, potentially artificially causing sign reversals (Páez et al., 2011). Fotheringham and Oshan (2016) refuted this critique through simulations demonstrating that GWR is rather robust against coefficient multicollinearity. However, GWR is not recommended as an inferential tool (Páez et al., 2011) because model calibration (e.g., bandwidths selection) as well as interpretation of the model output (e.g., continuous parameter surfaces) remain challenging.

To circumvent some of these limitations, an alternative approach, termed RegioClust, was developed. RegioClust identifies regions with similar associations between the dependent and the independent variables and calculates local parameter estimates for each region. Such a region-based approach is useful because it facilitates the definition of place-based policies, ensures that local policies have a homogeneous impact, and supports scenario development (de Noronha Vaz et al., 2012; Fischer, 1980). In addition, this study addressed the local drivers of LCRs in Germany at the level of municipalities for the period 2000–10, something that had not been done before, and the predicted rates for 2010–20. The research questions were as follows:

- To what extent did the relationships between LCRs and the drivers vary across Germany in 2000–10?
- Does RegioClust predict actual and future LCRs more accurately than GWR?
- Will the predicted LCRs in 2010–20 be below the targeted 30 ha per day?

The rest of the article is structured as follows. Section 2 outlines the materials and methods; Section 3 summarizes the results; Section 4 discusses the results in the context of the existing literature, and Section 5 presents the conclusions.

2. Materials and methods

2.1. Study area

Germany is the most populous country in Europe: In 2010, the country's 357,375 km^2 of land was home to about 81 million people. The present study was longitudinal and based on the administrative units of municipalities. Municipalities are an appropriate analyses level, as they are small in size and represent the lowest planning level in Germany. Non-contiguous regions, such as islands or exclaves (e.g., Sylt), and unincorporated regions without populations (e.g., Sachsenwald) were removed. This resulted in a total of 11,357 municipalities.

2.2. Data

The central variable was LCR per territorial unit. The built-up areas comprised settlement and transportation infrastructure for the years 2000 and 2010. Data were extracted from the IÖR Monitor (Meinel and Schumacher, 2010), which is based on the official German digital landscape model ATKIS®–Basis–DLM (Bundesamt für Kartographie und Geodäsie, 2016). The LCR, subsequently serving as continuous response variable, was computed by dividing the difference between the consumed land (e.g., built-up areas, transportation infrastructure) in 2010 and 2000 by the total area of the municipality.

Selecting the explanatory variables was guided by data availability and a literature review (Dubovyk et al., 2011; Kretschmer et al., 2015). Six area-level covariates were collected for 2000 and 2010. As strong evidence exists that accessibility is one of the major drivers of urban growth (Duranton and Turner, 2012; Iacono et al., 2008), the spatial distance (in km) from the center of each municipality to the nearest major highway was calculated. To approximate a Gaussian distribution, the square root (srDistHwy) was taken. The highway data were retrieved from the ATKIS®–Basis–DLM. Employment rate (EmplRate) served as proxy variable for wealth, which is highly correlated with urbanization (Bloom et al., 2008). In order to differentiate between municipalities with different housing characteristics, the proportion of family houses (FamHouse) was included. To model urbanization pressure through population in-flow, the net migration rate (NetMig) was incorporated. To control for the degree of urbanity, population density in 1,000 people per km^2 was included. To approximate a Gaussian-like distribution, the variable was log transformed (logPopDens). Data on the average tax revenue in €1,000 per capita (TaxRev) was collected to represent social deprivation. All these data were obtained from the Federal Institute for Research on Building, Urban Affairs and Spatial Development. Fig. 1 depicts the spatial distribution of each covariate for the year 2000.

2.3. Methods

2.3.1. RegioClust

RegioClust consist of two steps (Fig. 2): While the first step of RegioClust determines spatial clusters, the second step determines regions with similar relationships between the dependent variable and independent variables whereas a separate local regression model is estimated for each region. Thus, RegioClust combines spatial clustering with local modeling.

In detail, in the first step, each observation is considered as a single

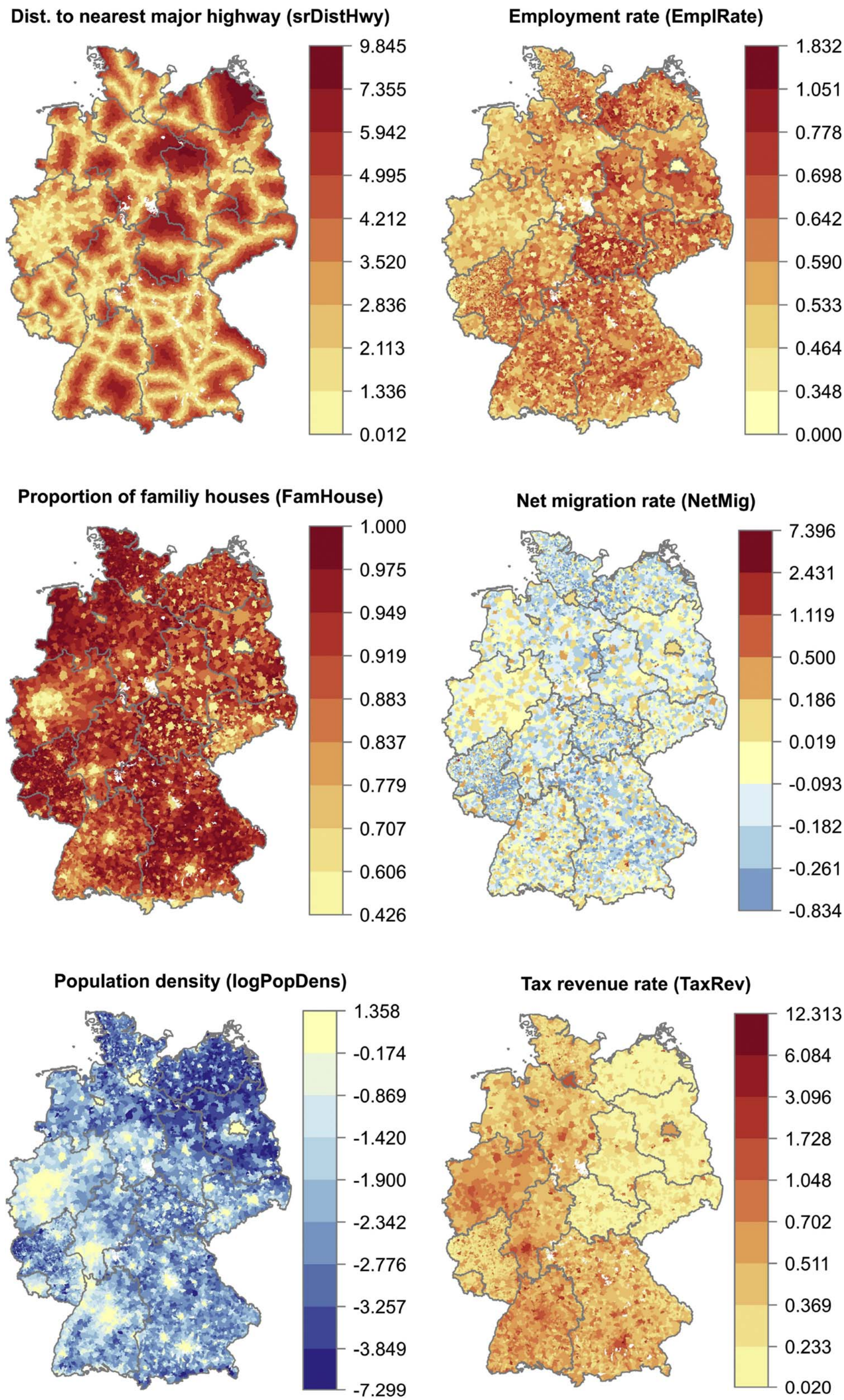

Fig. 1. Drivers of land consumption for the year 2000 on the municipality level.

cluster. Then, the clusters are repeatedly merged. Which clusters are merged is determined by Ward's cluster criterion (Ward, 1963). This criterion determines the clusters that minimize the increase in the total within-cluster variance after merging. The variance is measured by evaluating the spatial distances between the observations (i.e., Euclidean distances between the municipality-based centroids). As a consequence, observations within the same cluster tend to be spatially nearby and thus tend to have similar relationships between the dependent variable and its drivers. As spatially contiguous clusters are essential for many applications (Helbich et al., 2013; Spielman and Folch, 2015), the merging criterion is modified to only consider spatially adjacent clusters (Guo, 2009; Murtagh, 1985; Ruß and Kruse,

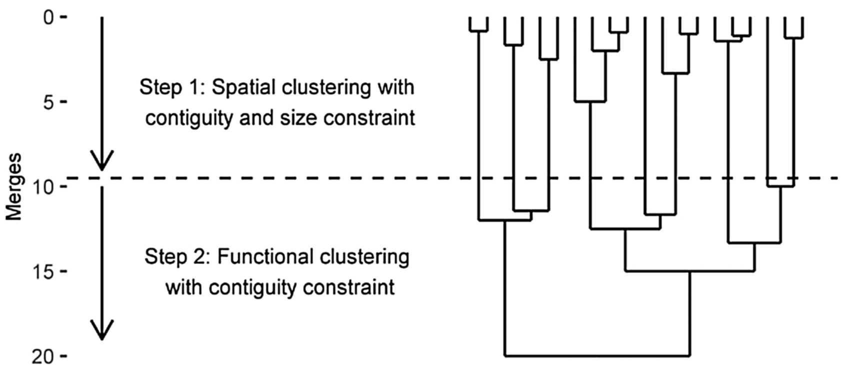

Fig. 2. The hierarchical clustering steps of RegioClust.

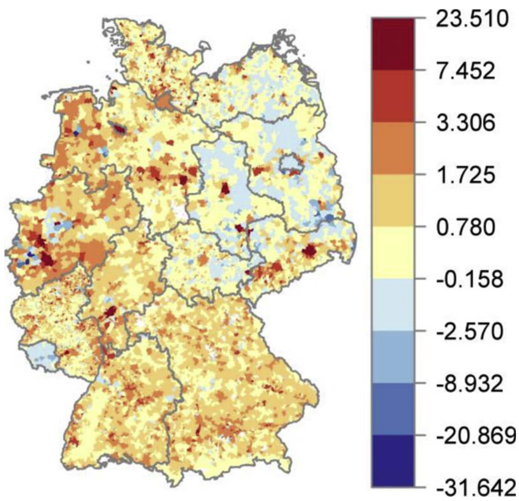

Fig. 3. LCRs in 2000–10 (in % of the municipalities' areas).

2011) whose size is smaller than i observations. The clustering ends when no more clusters can be merged, which is typically when all clusters consist of at least i observations.

In the second step, to obtain spatial clusters with a similar response–covariate relationship, the resulting clusters from the first step are merged employing hierarchical clustering with a contiguity (but not size) constraint. The merging criterion is different from the one in the first step. For each cluster a linear model is estimated, and the clusters with the lowest increase in the residual sum of squares (RSS) are merged. The clustering stops when a given number of clusters j is obtained or no more clusters can be merged. Since each resulting cluster refers to a contiguous geographic region for which a uniquely associated linear model exists, RegioClust refers to a pooled piece-wise linear model.

An open-source software implementation of RegioClust can be downloaded from https://github.com/jhagenauer/regioclust.

2.3.2. Geographically weighted regression

GWR is a locally weighted regression (Brunsdon et al., 1996) that was used as a benchmark for RegioClust to model spatial heterogeneity. Briefly, GWR estimates coefficients for a sub-set of locations by taking the distance of observations into account. The local coefficients are estimated by solving a location-specific weighted least squares model (Fotheringham et al., 2003). The weights are given by a weight matrix that is specified using a local kernel function that models a distance decay between locations. Nearby observations receive higher weights than distant ones. The choice of the kernel function generally has little impact on the results, given that it is smooth. Either a Gaussian or a bisquare kernel function is commonly used. More crucial than the kernel type is the choice of the bandwidth, which is frequently determined using cross validation (Brunsdon et al., 1996; Fotheringham et al., 2003). Moreover, the bandwidth can vary across space depending on the distribution of the data. If the regression points are sparsely distributed across space, a larger bandwidth is selected, and vice versa. Using an adaptive rather than a predefined bandwidth has the advantage that it reduces the number of extreme coefficients (Fotheringham et al., 2003). The GWR models were estimated by means of the 'GWmodel' package (Lu et al., 2014) using the R programming environment (R Core Team, 2017).

3. Results

3.1. Descriptive statistics

The descriptive statistics show that in the year 2000, 9.800% of Germany's total area was covered by built-up areas and transportation infrastructure. In 2010, the proportion had increased to 10.603% (+0.803%). This corresponds to an LCR of 77.455 ha per day. The amount of built-up area differs between east and west Germany. In west Germany, the proportion of covered area increased from 10.555% in 2000–11.651% in 2010 (+1.097%), whereas in east Germany the proportion increased from 8.094% in 2000–8.233% in 2010 (+0.014%). This corresponds to an LCR of 73.330 ha per day in west Germany and 4.126 ha per day in east Germany.

Fig. 3 depicts the LCRs for each municipality. A value of zero indicates that the amount of consumed land did not change between 2000 and 2010, while a value of 50%, for example, indicates that the amount of consumed land increased by 50% of the municipality's total area. Using first-order queen contiguity, the Moran's I statistic confirms that the LCRs are not randomly distributed across Germany and that significant regional differences exist ($I = 0.240$, $p < 0.05$).

3.2. Model fit

Fig. 4 shows the Akaike information criterion (corrected for finite sample sizes) (AICc) (Burnham and Anderson, 2004) of RegioClust for $i = 8$ to 12 (i.e., the minimum number of observations per cluster) and for $j = 1$ to 250 (i.e., the total number of clusters). A low AICc value refers to a good compromise between model fit and model complexity (i.e., number of parameters). The Figure illustrates that for fixed values of j, the AICc mostly increases with i. Also, for each value of i and beginning with $j = 1$, the AICc score decreases with increasing j until its minimum is reached. Beyond this, the AICc increases with j. The value of j for which the minimum is obtained is generally lower the smaller the value of i. Fig. 5 shows the number of outlying coefficients of RegioClust for $i = 8$ to 12 and for $j = 1$ to 250. A coefficient is considered an outlier if its value is three times the interquartile range below the first quartile or above the third quartile of the ranked values. The presence of outlying coefficients is an indicator of local overfitting which is not directly taken into account by the AICc. Generally, the number of outlying coefficients increases with i and j. Also, RegioClust tends to estimate extreme coefficients for population density, in

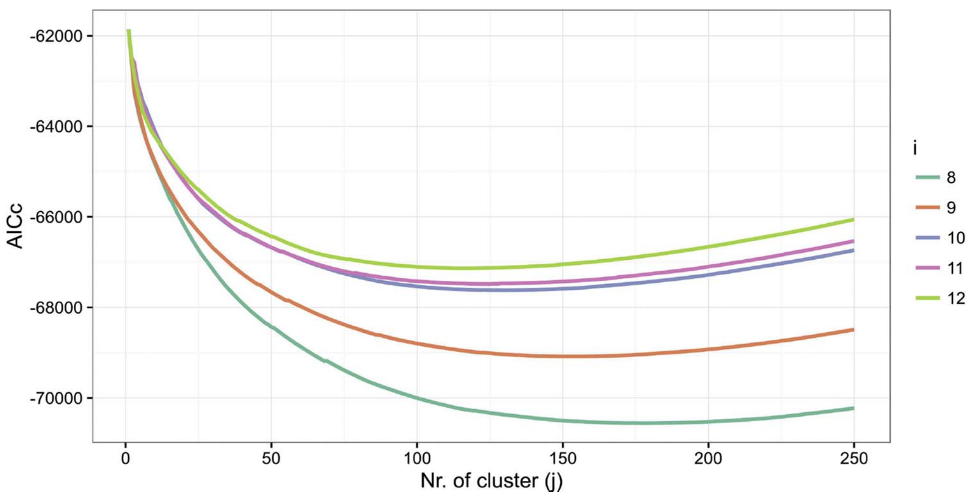

Fig. 4. AICc statistics for RegioClust.

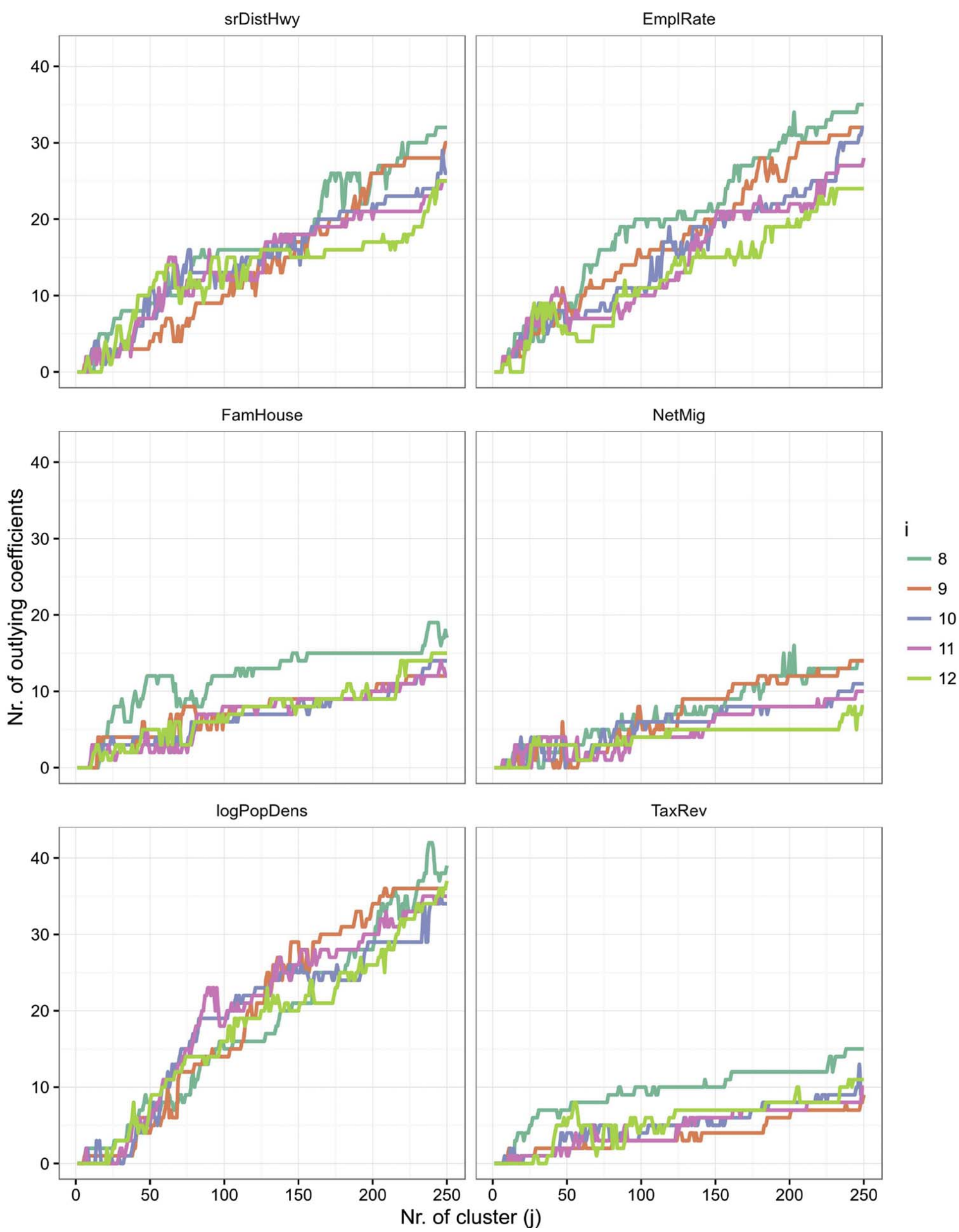

Fig. 5. Number of outlying coefficients for RegioClust.

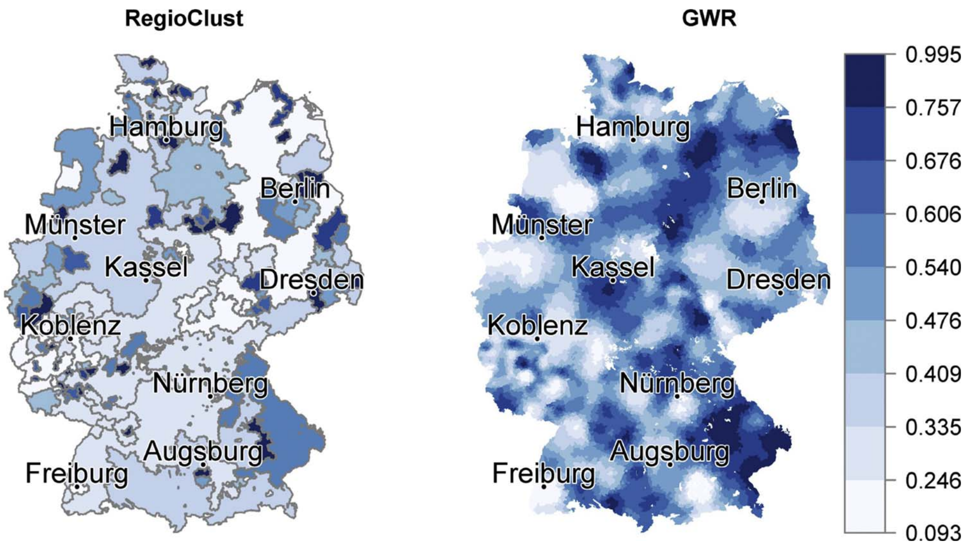

Fig. 6. Model fit (R^2) of RegioClust and GWR.

particular for large values of j. Overall, RegioClust represents for a wide range of i and j a reasonable trade-off between model fit, model complexity, and local overfitting. In the following, the RegioClust model with $i = 10$ and $j = 130$ will be considered. The Moran's I test statistic shows marginal significant spatial dependence of the residuals ($I = 0.030$, $p < 0.05$).

For comparison purposes, a GWR model with Gaussian kernel and adaptive bandwidth using the 20 nearest neighbors was estimated. The number of neighbors was selected by optimizing the AICc. Monte Carlo tests confirmed the significant spatial variability of all coefficients ($p < 0.05$). The AICc is −63,857.010. The Moran's I test statistic refers to marginal spatial dependence of the residuals ($I = 0.033$, $p < 0.05$).

On a Lenovo Thinkpad X230, Intel(R) Core(TM) i5–3320M CPU@ 2.60 GHz with 16 GB RAM, it took 32.510 min to estimate the GWR model and 72.026 min to compute the RegioClust model.

Fig. 6 depicts the spatial regions outlined by the RegioClust model and the local model fits for both RegioClust and GWR. It shows that in the southwest of Germany (i.e., Rhineland–Palatinate and Saarland), the regions are smaller than in the rest. Many of the larger regions consist of rural municipalities as well as medium-sized cities (e.g., Nürnberg). A few small regions consist of only a single large city and its close surroundings (e.g., Hamburg and Dresden). The largest region, with an area of about 45,133 km^2, is located in the west and comprises the cities of Kassel and Münster, and some outskirts of the Rhine–Ruhr metropolitan region.

Both models show a pronounced volatility of the model fit in the southwest (i.e., Rhineland–Palatinate and Saarland) while fitting the data particularly well in eastern Bavaria (i.e., east of Augsburg and Nürnberg). However, some differences are notable. For example, whereas RegioClust provides a better fit for Berlin, Dresden, and Hamburg, GWR shows an improved fit for the south of Kassel and the northwestern surroundings of Berlin. Although the R^2 is spatially randomly distributed, smaller RegioClust regions tend to have higher R^2 than large regions. This was confirmed by analysis of the Pearson's correlation coefficient ($\rho = -0.272$, $p < 0.05$).

3.3. Coefficients of RegioClust and GWR

The significant coefficients (excluding the intercept, $p < 0.05$) are shown for RegioClust in Fig. 7 and for GWR in Fig. 8. Spatial heterogeneity in the associations is evident. For both RegioClust and GWR, population density is a key driver of LCR at the national level. It reaches significance more often than any other coefficient. In contrast, employment rate and distance to nearest major highway are relevant drivers for only a few municipalities. They less often reach significance compared to other coefficients. Overall, with the exception of population density, the number of municipalities showing a significant relationship is higher for the coefficients of RegioClust than for those of GWR.

A detailed inspection of Figs. 7 and 8 reveals some notable similarities across the coefficient surfaces. For instance, whereas for most municipalities population density is positively related to LCR, RegioClust and GWR estimate a significant negative relationship for the federal state of Saarland (g) and for the north of the district of Dennim (a). Analogously, for the federal state of Bremen including its surroundings (b), both models estimate a strong and significant negative association between proportion of family houses and LCR. Other coefficients are not significant for this region. In the south of Augsburg (h), RegioClust and GWR estimate a significant positive effect of employment rate, net migration rate, and population density on LCR, whereas the tax revenue rate has a negative effect. Distance to nearest major highway, however, is only weakly positive and statistically significant for GWR, while proportion of family houses shows a strong positive and significant association only for RegioClust.

The two algorithms often estimate different coefficients for some regions. For example, GWR estimates substantially strong and significant coefficients for most variables in the north of the district of Emsland (c), the district of Recklinghausen (e), and the north of the districts of Düren and Rhein–Erft Kreis (f). RegioClust, by contrast, does not estimate any significant coefficient for regions (c) and (e), and for region (f) only the variable tax revenue rate is significant. In the east of Germany, RegioClust identifies region (d), which comprises the municipalities between Frankfurt/Oder and Kolkwitz, close to Cottbus. For this region, RegioClust estimates a significant positive effect for employment rate, proportion of family houses, population density, and net migration rate, while it estimates a significant negative effect for tax revenue rate. For most municipalities in this region, GWR estimates positive coefficients (i.e., proportion of family houses and population density), whereas most drivers are not significant. Moreover, the distinct outline of the region identified by RegioClust is not identified by GWR.

3.4. Predictions of RegioClust and GWR

For the period 2010–20, RegioClust predicts that in Germany 71.011 ha of land will be consumed per day. This corresponds to a decrease of 6.444 ha per day compared to 2000–10. The predicted LCRs differ substantially between east and west Germany. For west Germany, RegioClust forecasts an LCR of 63.385 ha per day, whereas for east Germany the prediction is 7.626 ha per day. Thus, RegioClust predicts a

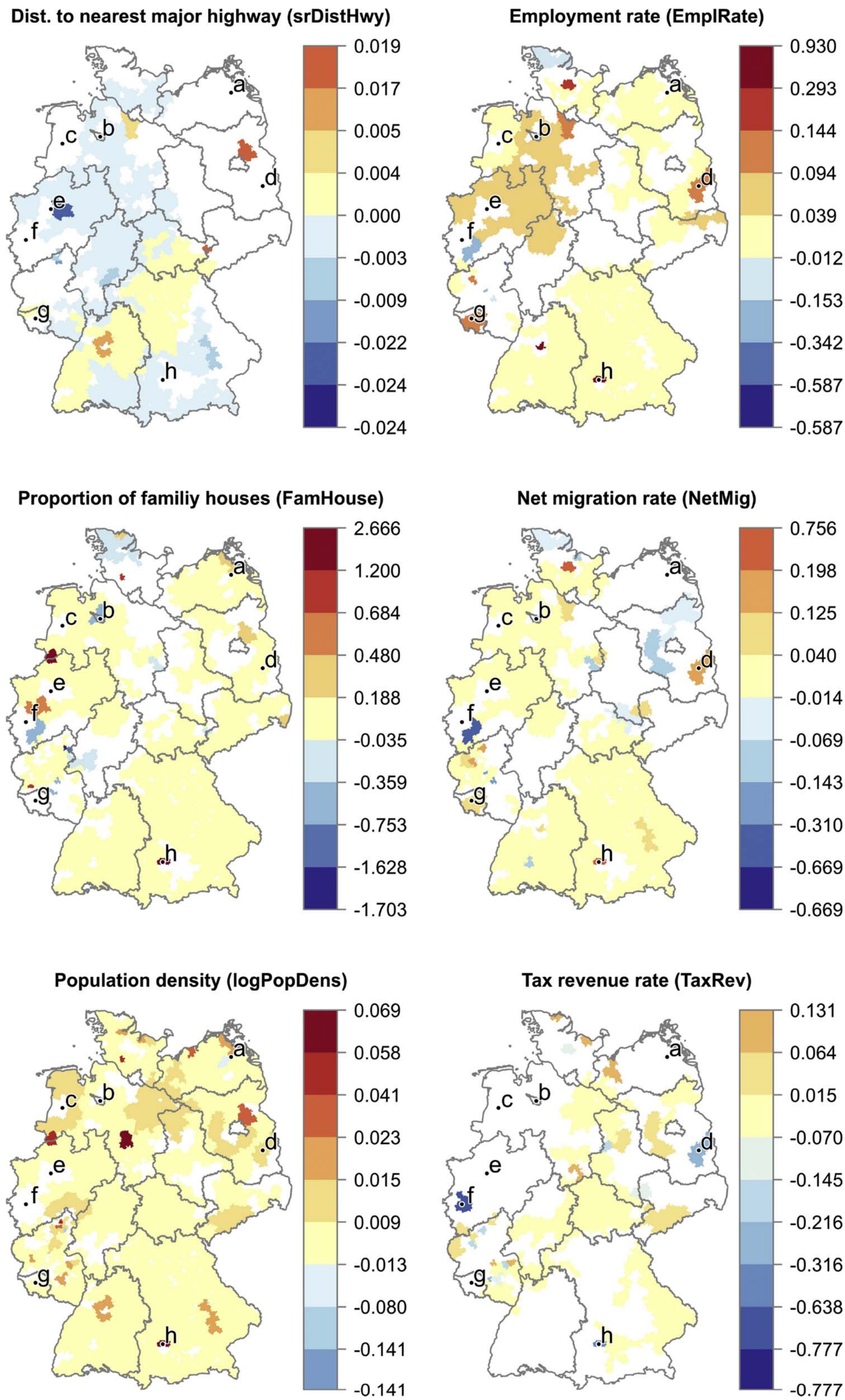

Fig. 7. Estimated local coefficients of RegioClust (areas colored in white refer to insignificant associations at the 0.05 level).

decrease of 9.945 ha per day for west Germany and an increase of 2.318 ha per day for east Germany. The predictions differ when GWR is applied. GWR forecasts that 78.529 ha of land will be consumed per day for the entire country for 2010–20. This corresponds to an increase of 1.074 ha per day in comparison to 2000–10. Again, the LCRs differ substantially between west (73.516 ha per day) and east Germany (5.013 ha per day). This corresponds to an increase of 0.186 ha per day in west Germany and 0.887 ha per day in east Germany.

Fig. 9 compares the predicted LCRs of RegioClust and GWR. It becomes apparent that both models predict a smaller increase in LCR in east than in west Germany. Saxony, consisting of the major cities of Dresden (i) and Leipzig, has the highest predicted increase in east Germany. In west Germany, the increase in land consumption is particularly low for the federal state of Saarland (g). In addition, the

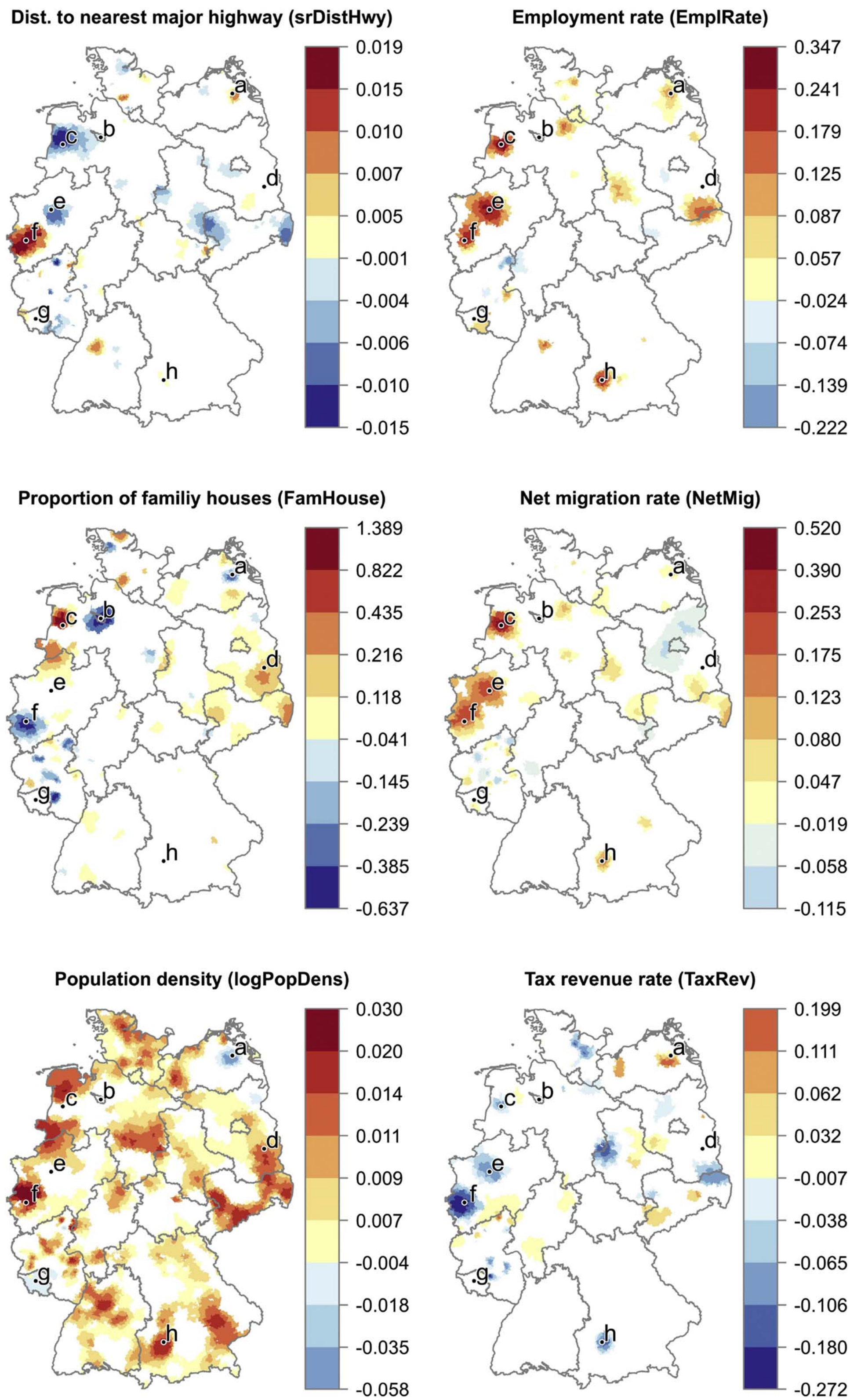

Fig. 8. Estimated local coefficients of GWR (areas colored in white refer to insignificant associations at the 0.05 level).

predicted increase tends to be higher for urban districts, such as Dresden (i), Braunschweig (k), and Bremen (b), than for rural ones. Some differences are also noticeable between RegioClust and GWR. RegioClust tends to predict more pronounced LCRs compared to GWR. This is observable for the Dresden region (i), the district of Düren (f), and the district of Dennim (a). Because unlike GWR RegioClust does not embody a distance-based smoothing, the predictions are often substantially different for adjacent regions. This is in particular noticeable for the south of Lübbenau (d), the district of Dennim (a), and the east of Rhein–Erft–Kreis, next to the district of Düren (f). For the municipalities between Recklinghausen and Dortmund (e) and south of Augsburg (h), RegioClust predicts negative LCRs whereas GWR predicts positive rates. A reverse effect appears, for example, for municipalities south of Lübbenau (d). In contrast to RegioClust, GWR is prone to

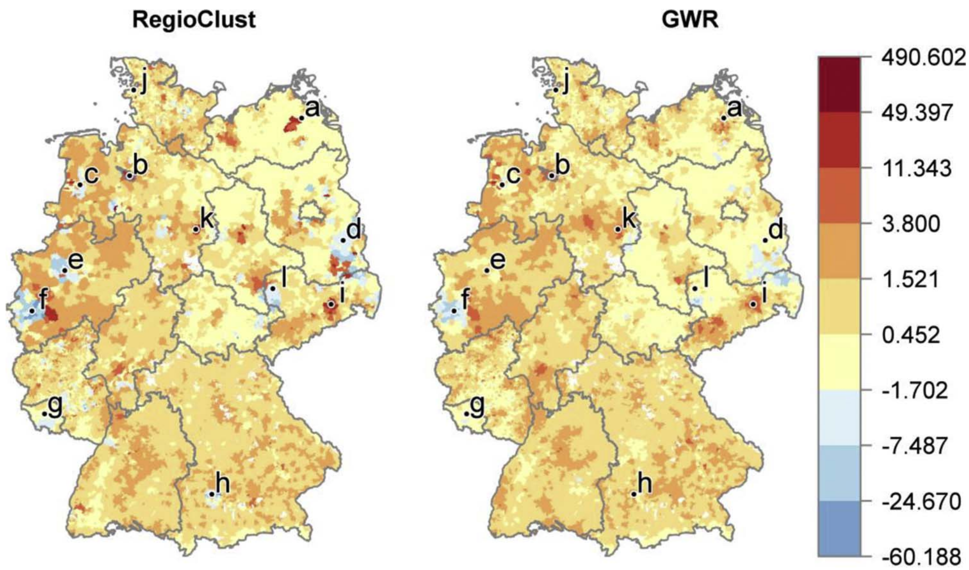

Fig. 9. Predicted LCRs in 2010–20 (in % of the municipalities' areas).

making extreme predictions. For example, GWR predicts for Norderfriedrichskoog (j) an excessive increase in the LCR (+490.6%), while RegioClust predicts a moderate decrease (−12.7%).

4. Discussion

Numerous studies emphasize the socioeconomic and demographic differences between east and west Germany that translate into diverse patterns of land consumption (Kroll and Haase, 2010; Nuissl et al., 2009; Schmidt, 2011). The present findings are congruent with these studies and provide statistical evidence that the LCRs between east and west Germany differed for the period 2000–10. The predictions of RegioClust and GWR indicate that different LCRs can also be expected for the period 2010–20. In particular, both models refer to a further increase in the LCRs for east Germany. For west Germany, however, RegioClust predicts a substantial decrease, whereas GWR predicts a marginal increase in the LCRs. Both models agree that the predicted LCRs for the period 2010–20 exceeds the goal of the German federal government to reduce the LCR to 30 ha per day until 2020 (Die Bundesregierung, 2016), thus bringing into question the effectiveness of currently implemented planning policies to reduce the LCR (Kretschmer et al., 2015).

Due to the diverse pattern of LCRs across Germany, decision-makers are advised to formulate spatially tailored spatial planning strategies to counteract the predicted future increase in LCRs at both the municipal and the federal state level. Complementing a quantitative restriction of land consumption, Siedentop and Kausch (2004) propose qualitative regulations concerning the determination of locations for urban expansion. Such a strategy seems to support the prevention of the further expansion of built-up areas, which increases the transportation infrastructure demand in rural areas. Besides, brown field recycling and urban renewal seem to be feasible approaches to mitigating land consumption (Borchard, 2011; Schultz and Dosch, 2005).

The comparison of the performance of RegioClust and GWR showed that the former provides a better model fit with respect to the AICc. Even more important, the residuals of RegioClust exhibit less spatial dependence, which is essential for model estimation and inference (Ay et al., 2017; Brady and Irwin, 2011), compared to GWR. Heuristic optimization of the RegioClust's clusters has the potential to further improve model performance (Guo, 2009). However, it must be noted that the performance of both models depends on the selected parameters. RegioClust is affected by the parameters i and j. Generally speaking, low values of i and high values of j result in small but spatially homogeneous clusters with respect to the within-cluster sum of spatial distances between locations. This permits the modeling of small-scale spatial variation. Smaller clusters, however, also increase the risk of estimating linear models that only represent small-scale noise or are based on a low number of unrepresentative data points (local overfitting). High values of i and low values of j typically result in larger but spatially inhomogeneous clusters. Linear model estimated for larger clusters, however, may fail to represent regularities that exist only on a small scale (underfitting). To circumvent a subjective parameter selection, the parameters i and j of RegioClust are chosen by considering the AICc score, which is a well-established measure for this purpose (Symonds and Moussalli, 2011) and the number of outlying coefficients.

As a result of local overfitting, outlying coefficients indicate unstable local estimates, which hamper a meaningful interpretation and make predictions less reliable. However, outlying coefficients can also be an issue for GWR. This becomes apparent for the municipality of Norderfriedrichskoog (j). Because this municipality was a tax haven until 2004 for foreign and domestic enterprises (von Schwerin and Buettner, 2016), the tax revenue for the year 2000 was high in absolute terms compared to the adjacent rural municipalities. The high tax revenue did not substantially affect the coefficient estimation of RegioClust, because the municipality is part of a large cluster/region in which each municipality is considered similarly. In contrast, the optimized kernel width of GWR is too small to compensate for the high tax rate, which results in a biased estimate of the coefficient. As a consequence, when using data of 2010, where the tax revenue rate of the Norderfriedrichskoog was already increased due to the introduction of higher taxes in 2004, GWR substantially overestimates the LCRs, whereas the LCR estimates of RegioClust are still within a reasonable range.

This study was innovative in developing a new approach to modeling spatially varying coefficients. RegioClust has the strength to identify regions with a clearly defined boundary and non-volatile coefficients, which supports visual analysis and model interpretation. For instance, the outlined regions of RegioClust show that several urban municipalities are clearly separated from surrounding rural areas while others are not. This means that for some urban municipalities, the relationship between LCR and its drivers is similar to that of its neighboring rural municipalities. This finding is remarkable, because land consumption itself is generally considered to be different for urban and rural areas (Siedentop and Kausch, 2004). As all previous studies focused on selected regions (Bieling et al., 2013; Rienow and Goetzke, 2015), this was the first regional analysis dealing with land consumption and its drivers at the national level, not only retrospectively but also prospectively.

However, this study also had some limitations. Because aggregated data at the municipal level were used, bias due to the size and shape of the municipalities might have affected the results (Openshaw and Taylor, 1979). Although this problem is hardly avoidable, it is important to be aware of it when interpreting the results. Also, even though this study went beyond the usage of distance-based drivers only (Achmad et al., 2015; Arsanjani et al., 2013; Hu and Lo, 2007; Shafizadeh-Moghadam et al., 2017b) and considered socio-demographic, economic, and environmental variables as well, the whole spectrum of possible drivers is still to be explored (Kretschmer et al., 2015). In particular, the effect of other road types besides major highways or means of transportation on LCRs calls for further research. Finally, LCRs were considered only on a single spatial and temporal scale. Land consumption, however, is closely related to urbanization, which is known to occur on many spatial and temporal scales (Wu, 2007). A multi-scale analysis has the potential for more accurate modeling of LCRs and thus a better understanding of land consumption processes (Grant et al., 2015).

5. Conclusions

Germany is faced with exceptionally high LCRs, which are a challenge to sustainability. This study addressed this issue by examining the drivers of LCRs at the municipal level for the period 2000–10 and predicting rates for 2010–20. For this purpose, a new method for modeling spatial varying relationships, termed RegioClust, was developed. Empirical comparison indicated that RegioClust provides better model fits (i.e., AICc scores) than GWR, but tends toward minor local overfitting if parameters are not chosen appropriately. Both models provided clear evidence that LCR drivers vary substantially across Germany and that population density is of the utmost importance. For 2010–20, RegioClust and GWR predicted substantially different LCRs for east and west Germany. Most important, the forecasts provide evidence that the policy target of reducing the LCR to 30 ha per day in 2020 will not be achieved. In order to counteract this development, it is advised to revise local planning policies while intensifying brown field recycling and urban renewal, particularly in those municipalities that show an excess of LCRs for 2010–20.

Acknowledgements

Julian Hagenauer thanks Martin Behnisch from the Leibniz Institute of Ecological Urban and Regional Development for providing the data for this study.

References

Aburas, M.M., Ho, Y.M., Ramli, M.F., Ash'aari, Z.H., 2017. Improving the capability of an integrated CA-Markov model to simulate spatio-temporal urban growth trends using an analytical hierarchy process and frequency ratio. Int. J. Appl. Earth Obs. Geoinf. 59, 65–78.
Achmad, A., Hasyim, S., Dahlan, B., Aulia, D.N., 2015. Modeling of urban growth in tsunami-prone city using logistic regression: analysis of Banda Aceh Indonesia. Appl. Geogr. 62, 237–246.
Anselin, L., 2010. Thirty years of spatial econometrics. Pap. Reg. Sci. 89, 3–25.
Arsanjani, J.J., Helbich, M., Kainz, W., Boloorani, A.D., 2013. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. Int. J. Appl. Earth Obs. Geoinf. 21, 265–275.
Ay, J.-S., Chakir, R., Gallo, J. Le, 2017. Aggregated versus individual land-use models: modeling spatial autocorrelation to increase predictive accuracy. Environ. Model. Assess. 22, 129–145.
Bagan, H., Yamagata, Y., 2015. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. GIScience Remote Sens. 52, 765–780.
Basse, R.M., Charif, O., Bódis, K., 2016. Spatial and temporal dimensions of land use change in cross border region of Luxembourg. Development of a hybrid approach integrating GIS, cellular automata and decision learning tree models. Appl. Geogr. 67, 94–108.
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, 192–203.
Bloom, D.E., Canning, D., Fink, G., 2008. Urbanization and the wealth of nations. Science 319, 772–775.
Borchard, K., 2011. Grundriss der Raumordnung und Raumentwicklung. Verlag der ARL.
Brady, M., Irwin, E., 2011. Accounting for spatial effects in economic models of land use: recent developments and challenges ahead. Environ. Resour. Econ. 48, 487–509.
Brown, D.G., Walker, R., Manson, S., Seto, K., 2004. Modeling land use and land cover change. In: Gutman, G., Janetos, A.C., Justice, C.O., Moran, E.F., Mustard, J.F., Rindfuss, R.R., Skole, D., Turner, B.L., Cochrane, M.A. (Eds.), Land Change Science: Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface. Springer Netherlands, Dordrecht, pp. 395–409.
Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28, 281–298.
Digitales Basis-Landschaftsmodell. Bundesamt für Kartographie und Geodäsie.
Burnham, K.P., Anderson, D.R., 2004. Multimodel inference. Sociol. Methods Res. 33, 261–304.
D'Amour, C.B., Reitsma, F., Baiocchi, G., Barthel, S., Güneralp, B., Erb, K.-H., Haberl, H., Creutzig, F., Seto, K.C., 2017. Future urban land expansion and implications for global croplands. Proc. Natl. Acad. Sci. U. S. A. 34, 8939–8944. http://dx.doi.org/10.1073/pnas.1606036114. 201606036.
de Noronha Vaz, E., Nijkamp, P., Painho, M., Caetano, M., 2012. A multi-scenario forecast of urban change: a study on urban growth in the Algarve. Landsc. Urban Plan 104, 201–211.
Dendoncker, N., Rounsevell, M., Bogaert, P., 2007. Spatial analysis and modelling of land use distributions in Belgium. Comput. Environ. Urban Syst. 31, 188–205.
Die Bundesregierung, 2016. Deutsche Nachhaltigkeitsstrategie.
Dubovyk, O., Sliuzas, R., Flacke, J., 2011. Spatio-temporal modelling of informal settlement development in Sancaktepe district Istanbul Turkey. ISPRS J. Photogramm. Remote Sens. 66, 235–246.
Duranton, G., Turner, M.A., 2012. Urban growth and transportation. Rev. Econ. Stud. 79, 1407–1440.
Fischer, M.M., 1980. Regional taxonomy: a comparison of some hierarchic and non-hierarchic strategies. Reg. Sci. Urban Econ. 10, 503–537.
Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., et al., 2005. Global consequences of land use. Science 309, 570–574.
Fotheringham, A.S., Oshan, T.M., 2016. Geographically weighted regression and multicollinearity: dispelling the myth. J. Geogr. Syst. 18, 303–329.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships. John Wiley & Sons.
Grant, L.P., Gennings, C., Wheeler, D.C., 2015. Selecting spatial scale of covariates in regression models of environmental exposures. Cancer Inform. 14, 81.
Griffith, D.A., 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). Environ. Plan. A 40, 2751–2769.
Guan, C., Rowe, P.G., 2016. Should big cities grow? Scenario-based cellular automata urban growth modeling and policy applications. J. Urban Manag. 5, 65–78.
Guo, D., 2009. Greedy optimization for contiguity-constrained hierarchical clustering. IEEE International Conference on Data Mining Workshops 591–596.
Haas, J., Ban, Y., 2014. Urban growth and environmental impacts in Jing-Jin-Ji the Yangtze river delta and the Pearl river delta. Int. J. Appl. Earth Obs. Geoinf. 30, 42–55.
Haase, D., Kabisch, N., Haase, A., 2013. Endless urban growth? On the mismatch of population, household and urban land area growth and its effects on the urban debate. PLoS One 8, e66531.
Haykin, S.S., 2009. Neural Networks and Learning Machines. Prentice Hall.
Helbich, M., Griffith, D.A., 2016. Spatially varying coefficient models in real estate: eigenvector spatial filtering and alternative approaches. Comput. Environ. Urban Syst. 57, 1–11.
Helbich, M., Brunauer, W., Hagenauer, J., Leitner, M., 2013. Data-driven regionalization of housing markets. Ann. Assoc. Am. Geogr. 103, 871–889.
Hennig, E.I., Schwick, C., Soukup, T., Orlitová, E., Kienast, F., Jaeger, J.A.G., 2015. Multi-scale analysis of urban sprawl in Europe: towards a European de-sprawling strategy. Land Use Policy 49, 483–498.
Hu, Z., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst. 31, 667–688.
Iacono, M., Levinson, D., El-Geneidy, A., 2008. Models of transportation and land use change: a guide to the territory. Journal of Planning Literature 22, 323–340.
Jakubowski, P., Zarth, M., 2003. Nur noch 30 Hektar Flaechenverbrauch pro Tag. Raumforsch. Raumordn. 61, 185–197.
Kretschmer, O., Ultsch, A., Behnisch, M., 2015. Towards an understanding of land consumption in Germany – outline of influential factors as a basis for multidimensional analyses. Erdkunde 69, 267–279.
Kroll, F., Haase, D., 2010. Does demographic change affect land use patterns? A case study from Germany. Land Use Policy 27, 726–737.
Ku, C.-A., 2016. Incorporating spatial regression model into cellular automata for simulating land use change. Appl. Geogr. 69, 1–9.
Li, X., Chen, Y., Liu, X., Xu, X., Chen, G., 2017. Experiences and issues of using cellular automata for assisting urban and regional planning in China. Int. J. Geogr. Inf. Sci. 1–24.
Lu, B., Harris, P., Charlton, M., Brunsdon, C., 2014. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. Geo-Spatial Inf. Sci. 17, 85–101.
Luo, J., Wei, Y.H.D., 2009. Modeling spatial variations of urban growth patterns in Chinese cities: the case of Nanjing. Landsc. Urban Plan. 91, 51–64.
Maimaitijiang, M., Ghulam, A., Sandoval, J.S.O., Maimaitiyiming, M., 2015. Drivers of land cover and land use changes in St. Louis metropolitan area over the past 40 years characterized by remote sensing and census population data. Int. J. Appl. Earth Obs.

Geoinf. 35, 161–174.
Malburg-Graf, B., Jany, A., Lilienthal, M., Ulmer, F., 2007. Strategies and instruments to limit excessive land use in Germany: A proposal to the German Council for Sustainable Development. In: Proceedings of the 2nd International Conference on Managing Urban Land. Dresden, Germany.
Mas, J.-F., Kolb, M., Paegelow, M., Olmedo, M.T.C., Houet, T., 2014. Inductive pattern-based land use/cover change models: a comparison of four software packages. Environ. Model. Softw. 51, 94–111.
Meinel, G., Schumacher, U., 2010. Konzept, funktionalitaet und erste exemplarische Ergebnisse des Monitors der Siedlungs- und Freiraumentwicklung (IOER-Monitor). Flächennutzungsmonitoring II. Rhombos-Verlag.
Murtagh, F., 1985. A survey of algorithms for contiguity-constrained clustering and related problems. Comput. J. 28, 82–88.
Nuissl, H., Schroeter-Schlaack, C., 2013. On the economic approach to the containment of land consumption. PLoS One 12, 270–280.
Nuissl, H., Haase, D., Lanzendorf, M., Wittmer, H., 2009. Environmental impact assessment of urban land use transitions – a context-sensitive approach. Land Use Policy 26, 414–424.
Omrani, H., Tayyebi, A., Pijanowski, B., 2017. Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: an integrated ML-CA-LTM modeling framework. GIScience & Remote Sensing 54 (3), 283–304 Published online: 05 Jan 2017, https://doi.org/10.1080/15481603.2016.1265706.
Openshaw, S., Taylor, P.J., 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. Stat. Appl. Spat. Sci. 21, 127–144.
Páez, A., Farber, S., Wheeler, D., 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environ. Plan. A 43, 2992–3010.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rienow, A., Goetzke, R., 2015. Supporting SLEUTH? Enhancing a cellular automaton with support vector machines for urban growth modeling. Comput. Environ. Urban Syst . 49, 66–81.
Ruß, G., Kruse, R., 2011. Exploratory hierarchical clustering for management zone delineation in precision agriculture. Industrial Conference on Data Mining 161–173.
Samardzic-Petrovic, M., Dragicevic, S., Kovacevic, M., Bajat, B., 2016. Modeling urban land use changes using support vector machines. Trans. GIS 20, 718–734.
Schmidt, S., 2011. Sprawl without growth in eastern Germany. Urban Geogr. 32, 105–128.
Schultz, B., Dosch, F., 2005. Trends der Siedlungsflächenentwicklung und ihre Steuerung in der Schweiz und Deutschland. disP – Plan. Rev. 41, 5–15.
Seto, K.C., Güneralp, B., Hutyra, L.R., 2011. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proc. Natl. Acad. Sci. U. S. A. 109, 16083–16088.
Shafizadeh-Moghadam, H., Helbich, M., 2015. Spatiotemporal variability of urban growth factors: a global and local perspective on the megacity of Mumbai. Int. J. Appl. Earth Obs. Geoinf. 35, 187–198.
Shafizadeh-Moghadam, H., Asghari, A., Taleai, M., Helbich, M., Tayyebi, A., 2017a. Sensitivity analysis and accuracy assessment of the land transformation model using cellular automata. GIScience & Remote Sens. 54 (5), 639–656.
Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., Taleai, M., 2017b. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. Comput. Environ. Urban Syst. 64, 297–308.
Siedentop, S., Kausch, S., 2004. Die raeumliche Struktur des Flaechenverbrauchs in Deutschland. Raumforsch. Raumordn. 62, 36–49.
Spielman, S.E., Folch, D.C., 2015. Reducing uncertainty in the American community survey through data-driven regionalization. PLoS One 10.
Symonds, M.R.E., Moussalli, A., 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav. Ecol. Sociobiol. 65, 13–21.
Triantakonstantis, D., Mountrakis, G., 2012. Urban growth prediction: a review of computational models and human perceptions. J. Geogr. Inf. Syst. 4, 555–587.
United Nations, 2015. World Urbanization Prospects: The 2014 Revision.
Van Dessel, W., Van Rompaey, A., Szilassi, P., 2011. Sensitivity analysis of logistic regression parameterization for land use and land cover probability estimation. Int. J. Geogr. Inf. Sci. 25, 489–508.
Veldkamp, A., Lambin, E.F., 2001. Predicting land-use change. Agric. Ecosyst. Environ. 85, 1–6.
Verburg, P.H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004. Land use change modelling: current practice and research priorities. GeoJournal 61, 309–324.
Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.
Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7, 161–187.
Wu, J., 2007. Scale and scaling: a cross-disciplinary perspective. In: Wu, J., Hobbs, R.J. (Eds.), Key Topics in Landscape Ecology. Cambridge University Press, Cambridge, pp. 115–142.
Zeng, C., Liu, Y., Stein, A., Jiao, L., 2015. Characterization and spatial modeling of urban sprawl in the Wuhan metropolitan area China. Int. J. Appl. Earth Obs. Geoinf. 34, 10–24.
van Vliet, J., Bregt, A.K., Brown, D.G., van Delden, H., Heckbert, S., Verburg, P.H., 2016. A review of current calibration and validation practices in land-change modeling. Environ. Model. Softw. 82, 174–182.
von Schwerin, A., Buettner, T., 2016. Constrained Tax Competition – Empirical Effects of the Minimum Tax Rate on the Tax Rate Distribution, in: Beiträge zur Jahrestagung des Vereins für Sozialpolitik 2016: Demographischer Wandel. Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg.