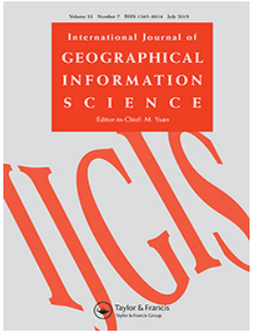

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Assessing the performance of 38 machine learning models: the case of land consumption rates in Bavaria, Germany

Julian Hagenauer, Hichem Omrani & Marco Helbich

To cite this article: Julian Hagenauer, Hichem Omrani & Marco Helbich (2019) Assessing the performance of 38 machine learning models: the case of land consumption rates in Bavaria, Germany, International Journal of Geographical Information Science, 33:7, 1399-1419, DOI: 10.1080/13658816.2019.1579333

To link to this article: https://doi.org/10.1080/13658816.2019.1579333

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2019, VOL. 33, NO. 7, 1399–1419
https://doi.org/10.1080/13658816.2019.1579333

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Assessing the performance of 38 machine learning models: the case of land consumption rates in Bavaria, Germany

Julian Hagenauer[a], Hichem Omrani[b] and Marco Helbich[c]

[a]Stuttgart, Germany; [b]Urban Development and Mobility Department, LISER, Belvaux, Luxembourg; [c]Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, Netherlands

ABSTRACT

Machine learning (ML) is at the forefront of land-use change modeling. Due to numerous available ML approaches, the model choice is complex and usually based on ad hoc decisions, though informed through a few comparative studies that considered a limited number of models. This study contributes a comprehensive comparison of 38 ML models to examine land consumption rates (LCR) (i.e. the transition of landscapes to built-up areas). We modeled LCR for 2009–2015 in Bavaria, Germany, and predicted rates for 2015–2021 at a municipality level. To assess the performance of each approach, we measured the mean absolute error (MAE), the root-mean-square error (RMSE), and the coefficient of determination (R^2) using cross-validation. All algorithms consistently predicted that the land consumption rate for Bavaria will increase. eXtreme gradient boosting decision trees performed best with respect to the RMSE (0.500) and R^2 (0.183), while the support vector machine with polynomial kernel has the lowest MAE (0.288). The generalized additive model and the random forest models also performed well. We recommend these ML approaches for future land consumption and land-use change studies. A poor performance was found for recursive partitioning by decision trees, self-organizing maps, and partitioning using deletion, substitution, and addition moves.

ARTICLE HISTORY
Received 17 June 2018
Accepted 3 February 2019

KEYWORDS
Land consumption; land-use; machine learning; model comparison; Germany

1. Introduction

Land-use mapping and modeling of environmental data using machine learning (ML) (Witten *et al*. 2016) have gained increasing interest within the geospatial community (Lary *et al*. 2016). It has become a vital methodology to monitor (Rogan *et al*. 2008, Qian *et al*. 2014, Heung *et al*. 2016, Omrani *et al*. 2019) and forecast land-use change (Samardžić-Petrović *et al*. 2016, Shafizadeh-Moghadam *et al*. 2017, Du *et al*. 2018, Hagenauer and Helbich 2018). ML comprises a set of inductive models that recognize patterns and/or minimize the prediction error of complex regression functions, by means of a repeated learning strategy from training data, linking an output such as land-use change to several underlying drivers. Once learned, the model can then be used to estimate previously unseen cases and predict future land-use change (Brown *et al*. 2013, Tayyebi and Pijanowski 2014, Lary *et al*. 2016, Maxwell *et al*. 2018).

CONTACT Marco Helbich m.helbich@uu.nl
Supplemental data for this article can be accessed here.

Statistical land-use models such as ordinary least squares regression or logistic binary regression are vital to understand land-use change processes and have thus gained popularity (Veldkamp and Lambin 2001, Lambin and Geist 2008, Jokar Arsanjani *et al*. 2013, Shafizadeh-Moghadam and Helbich 2015, Mustafa *et al*. 2018). While these basic statistical models have a long tradition and have contributed to our understanding of how land-use change takes place across space and over time due to anthropogenic influences (Brown *et al*. 2004, Seto *et al*. 2012), the results are only reliable when model assumptions are fulfilled. These assumptions include that the drivers underlying land-use change are not correlated and that residuals (i.e. the difference between observed and predicted values) follow a specific distribution while being homoscedastic (i.e. having similar variance) (Fahrmeir *et al*. 2013). Of equal importance, these models are frequently restricted to linear associations (Dubovyk *et al*. 2011, Jokar Arsanjani *et al*. 2013), which seems a rather implausible assumption, as shown by a few studies (Tayyebi and Pijanowski 2014). Although non-linearity can be incorporated through polynomial terms or splines (Fahrmeir *et al*. 2013), some prior knowledge about the underlying correlations is still required, though it is rarely available. Similarly, it seems of ultimate importance to consider not only the drivers' main effects but also the interactions between them (Pijanowski *et al*. 2002). Both issues can be addressed through ML.

Data-driven models from the field of ML have recently emerged as powerful alternatives to parametric land-use models, and they hold great promise (Hagenauer and Helbich 2012, Tayyebi and Pijanowski 2014, Rienow and Goetzke 2015, Samardžić-Petrović *et al*. 2016, Omrani *et al*. 2017, 2019, Shafizadeh-Moghadam *et al*. 2017, Du *et al*. 2018). ML models have manifold benefits. For example, they are capable of dealing with massive amounts of data and a large number of variables, and they have the ability to model complex non-linear relationships as well as interactions between drivers, while not being grounded in restrictive distributional assumptions of the input data that are hard to achieve in practice (Witten *et al*. 2016).

The repertoire of ML is large (Lim *et al*. 2000, Fernández-Delgado *et al*. 2014), which makes the selection of a well-performing model a challenging task. However, frequently only a single approach is applied (Azari *et al*., 2016, Huang *et al*. 2010, Hagenauer and Helbich 2012, Linard *et al*. 2013, Samardžić-Petrović *et al*. 2016, Omrani *et al*. 2019), often either artificial neural networks (Haykin 2009), random forests (Breiman 2001), and support vector machines (Scholkopf and Smola 2001). To support an evidence-based algorithmic selection, a small number of land-use studies have compared multiple ML algorithms (Rogan *et al*. 2008, Tayyebi and Pijanowski 2014, Samardžić-Petrović *et al*. 2017, Shafizadeh-Moghadam *et al*. 2017). While supportive, these studies assessed the performance of only a small number of algorithms. For example, Shafizadeh-Moghadam *et al*. (2017) compared six models to generate land-use transition probability maps. Multivariate adaptive regression splines and artificial neural networks were found to predict most accurately. Another study (Samardžić-Petrović *et al*. 2017) concluded that support vector machines outperform their competitors, including neural networks and logistic regression (Huang *et al*. 2010). While consensus is building that ML outperforms parametric models (e.g. regression), findings are less uniform and partly contradictory when ML models are contrasted. However, since many other ML models exist (Fernández-Delgado *et al*. 2014), more powerful and accurate approaches than the established ones might remain unacknowledged.

Previous studies formulate land-use change as a classification problem based on binary raster cells (Tayyebi and Pijanowski 2014, Samardžić-Petrović *et al*. 2017, Shafizadeh-Moghadam *et al*. 2017), however, research dealing with land-use change form a regression-based perspective is virtually not existent. A special case of land-use change, which can analytically be explored on an area level, is the proportion of land consumption per municipality. Land consumption refers to the conversion of agricultural and forestry landscapes into built-up areas (Nuissl *et al*. 2009). As land consumption proceeds and causes an irreversible anthropogenic process (e.g. a loss of biodiversity, atmospheric pollution) (Seto *et al*. 2012, Bren D' Amour *et al*. 2016), it is of importance for policymakers to have the most accurate forecasts available in order to formulate sustainable planning policies.

To address these research gaps, a comprehensive comparison of numerous models is needed to evaluate the suitability of individual regression-based models and to identify their individual strengths and weaknesses. The present study rigorously cross-compared 38 ML models based on data on land consumption data in Bavaria, Germany. Our selected algorithms included tree-based models, artificial neural networks, support vector machines, linear regression models, nearest neighbor algorithms, and rule-based learners. The selection was guided by their proliferation in other domains. Yet, to the best of our knowledge, some of them had never been tested, and such an extensive model comparison had neither been conducted in the context of modeling land-use change nor modeling land consumption. The research question was as follows: Which algorithm predicts land consumption rates most accurately, while still being computationally moderately intense?

This article is structured as follows: Section 2 outlines the study area and the data; Section 3 introduces the methods; Section 4 summarizes the results; Section 5 discusses the findings in the context of the existing literature; and Section 6 draws conclusions.

2. Materials

2.1. *Study area*

Germany faces one of the highest rates of land consumption of all EU member states (Siedentop and Kausch 2004, Kroll and Haase 2010). The federal government of Germany aims to limit the amount of land consumption to 30 ha per day up to the year 2020 (Statistisches Bundesamt 2012). This corresponds to a land consumption rate (LCR) of 0.43% of Germany's total area in six years. For this model comparison, we selected the federal state of Bavaria, Germany, as the study area. Its approximately 12.977 million inhabitants make it the second most populated federal state in Germany, and its 70,550 km^2 make it the largest one. The area of Bavaria accounts for 19.47% of the total area of Germany.

2.2. *Data*

We analyzed LCR at a municipality level for 2009–2015. After removing unincorporated municipalities (e.g. Chiemsee and Veldensteiner Forst), the dataset included 2,056 municipalities.[1] For each municipality, the LCR for both time periods were determined by dividing the difference between the consumed land (e.g. built-up areas, transportation infrastructure) in the beginning and end of the period by the total area of the

municipality and multiplying it by 100 to obtain percentages. Data were extracted from the IÖR Monitor,[2] which is based on the digital landscape model ATKIS®–Basis-DLM (Meinel and Schumacher 2010).

A set of 10 demographic, socioeconomic, and environmental covariates were used as explanatory variables that were assumed to influence land-use change (Dubovyk *et al*. 2011, Kretschmer *et al*. 2015, Hagenauer and Helbich 2018). Unless stated otherwise, the data were obtained from Germany's regional database (Regionaldatenbank Deutschland),[3] whose tables are based on the Regional Statistical Data Catalog and the Regional Statistical special program of the Federal Statistical Office and the statistical offices of the federal states. The demographic and socioeconomic variables were collected for 2009 and 2015.

In line with urban theory (Duranton and Turner 2012), the spatial distance (in km) from each municipality to the nearest regional center (Oberzentrum) was considered to reflect that accessibility stimulates urban growth. To incorporate the wealth per municipality which was found to be associated with urban growth (Bloom *et al*. 2008), employment rate (in %) served as a proxy variable. We used the in- and out-commuter rates (in %) as indicators to adjust for urbanization pressure through the population in- and outflows. As urban and rural municipalities may be affected differently, we controlled for the degree of urbanity by means of the logged population density (in 1,000 people per km^2) (Hagenauer and Helbich 2018). Data on the trade tax (in €1,000 per capita) were included to represent the municipalities' economic prosperity. Since the potential for further urban change decreases with the amount of existing built-up land, we considered the proportion of built-up area per municipality (in %). The variable was log transformed to make it Gaussian-like distributed. We included the mean of the terrain ruggedness index (Riley *et al*. 1999) to account for different building potentials due to topologic variation. Elevation data with a resolution of 90 m obtained from the Shuttle Radar Topology Mission served as the basis. Finally, the longitude and the latitude of the center of each municipality were included in order to account for the (unmeasured) locational characteristics (Jokar Arsanjani *et al*. 2013). Both longitude and latitude were scaled to zero and one, maintaining the aspect ratio. Table 1 provides descriptive statistics for all variables.

3. Methods

This section briefly describes each of the 38 ML models and their tunable parameters. All models have either already been used in land-use research or have shown promising results in other model comparisons (Fernández-Delgado *et al*. 2014, Xu *et al*. 2014).

All analysis steps were performed in the *R* programming environment (R Core Team 2018) using the caret package (Kuhn 2008). The latter is particularly useful for model comparisons as it harmonizes and streamlines the workflow for predictive models, parameter tuning, and validation, while providing a unified interface for each algorithm. A list of the tested models together with a short description is presented in Table 2. Note that the tested parameter values are given in the notation A:B:C, referring to the integer values A to B with a step size C. We also extended the default values of the tunable parameter values of caret to provide more comprehensive parameter tuning. The *R* code is provided as supplementary materials.

Before the ML models were built, not available values (NAs) were imputed using a bagging ensemble of trees (Feelders 1999). The input variables were then scaled to

Table 1. Descriptive statistics for the response and the covariates for 2009 and 2015.

Response variable	Min.	1st quartile	Median	Mean	3rd quartile	Max.	NAs
Land consumption rate for 2009–2015	−5.200	0.300	0.500	0.611	0.800	5.500	0
Covariates 2009							
Distance to regional center	0.000	8.780	15.600	16.800	23.800	54.400	0
In-commuter rate	0.238	0.611	0.685	0.682	0.759	0.952	27
Latitude	0.000	0.272	0.480	0.499	0.741	1.000	0
Longitude	0.000	0.298	0.447	0.464	0.627	0.980	0
Out-commuter rate	0.255	0.781	0.873	0.828	0.917	0.988	0
Population density (log)	0.766	1.830	2.000	2.080	2.260	3.630	0
Proportion of built-up area (log)	0.000	0.778	0.875	0.920	1.030	1.920	0
Terrain ruggedness (log)	2.938	11.931	15.896	18.525	21.567	112.918	0
Trade tax	−1.320	0.082	0.149	0.254	0.280	19.500	0
Unemployment rate	0.003	0.014	0.018	0.019	0.023	0.048	0
Covariates 2015							
Distance to regional center	0.000	8.780	15.600	16.800	23.800	54.400	0
In-commuter rate	0.228	0.634	0.702	0.701	0.774	0.964	22
Latitude	0.000	0.272	0.480	0.499	0.741	1.000	0
Longitude	0.000	0.298	0.447	0.464	0.627	0.980	0
Out-commuter rate	0.230	0.790	0.875	0.834	0.918	0.993	0
Population density (log)	0.817	1.830	2.010	2.080	2.270	3.670	0
Proportion of built-up area (log)	0.079	0.806	0.908	0.950	1.060	1.910	0
Terrain ruggedness (log)	2.938	11.931	15.896	18.525	21.567	112.918	0
Trade tax	−0.026	0.147	0.246	0.380	0.413	16.500	0
Unemployment rate	0.002	0.011	0.014	0.015	0.017	0.040	0

have zero mean and unit variance in order to make them comparable. The out-of-sample performance of the models (i.e. MAE, RMSE, and R^2) with selected parameters was evaluated through 10-fold cross-validation (Kuhn 2008). This procedure randomly partitions the data into 10 disjoint subsets. One subset at a time is then used for testing the model, while the remaining sets are used to build the model. This reduces the bias in performance estimation since the testing and training data sets are independent of each other (Kohavi 1995). Cross-validation was repeated four times to lower the variance of the estimated performance and the results were averaged. We used the covariates of 2009 to perform cross-validation and determine the best parameters. Then, using the best parameters, the final model was built using the same data. Finally, the covariates of 2015 were applied to the model to predict LCRs for 2015–2021. Figure 1 summarizes the workflow.

To evaluate the different classifiers, the distribution of the performance statistics needed to be taken into account (Hothorn *et al.* 2005). We evaluated the statistical significance of the models' differences in performance as follows: Kruskal–Wallis tests with a 5% significance level were employed to test the null hypothesis that the performance estimates of the models are not systematically different from each other. Two-sided Wilcoxon rank-sum tests were applied to determine the statistical significance of systematic pairwise differences between models. To control for a false discovery rate at the 5% level, the *p*-values were adjusted by means of the Benjamini–Hochberg procedure (Benjamini and Hochberg 1995).

4. Results

Figure 2 summarizes the performance of the classifiers. The lowest mean MAE was achieved by *svmPoly* (0.288), while *xgbTree* had the highest mean R^2 (0.183) and the lowest mean RMSE (0.500). However, the differences between *xgbTree* and *svmPoly* in

Table 2. Summary of the tested models.

Label	Description	R package	Tested parameters	References	Model class
brnn	Bayesian regularized neural network	brnn	neurons $=2^{(0:5:1)}$	(Dan Foresee and Hagan 1997; MacKay 1992)	NN
cforest	Random forest using conditional inference trees	party	mtry = 1:10:1	(Strobl *et al.* 2008)	Bag, Tree
ctree	Conditional decision tree	party	mincriterion = 0.8:0.99:0.01	(Hothorn *et al.* 2006)	Tree
cubist	Rule-based model that extends Quinlan's M5 model tree	cubist	committees = 1:10:1; neighbors = 0:9:1	(Quinlan 1993a, 1993b, 1992)	Boost, Tree, LM
earth	Multivariate adaptive regression splines	earth	Degree = 1:4:1; nprune $=2^{(0:5:1)}$	(Friedman 1991)	Spline
elm	Extreme learning machine	elmNN	actfun = radbas, sind, purelin, tansig; nhid $=2^{(0:5:1)}$	(Guang-Bin Huang *et al.* 2012)	NN
enet	Elastic Net	elasticnet	fraction = 0:1:0.05; lambda $=10^{(-6:-1:1)}$	(Zou and Hastie 2005)	LM
gam	Generalized additive model	mgcv	select = True,False; method = GCV.Cp	(Hastie and Tibshirani 1986)	Add
gbm	Generalized boosting regression machine	gbm	n.trees = 500; interaction.depth = 4:16:4, shrinkage = 0.01, 0.02, 0.05; n.minobsnode = 4:32:4	(Friedman 2001)	Boost, Tree
gcvEarth	Earth using generalized cross-validation	earth	degree = 1:3:1	(Friedman 1991)	Spline
glm	Generalized linear model	stats		(McCullagh and Nelder 1998)	LM
glmnet	Generalized linear model using penalized maximum likelihood with L1 and/or L2 regularization		alpha = 0:1:0.05; lambda $=10^{(-6:-1:1)}$	(Friedman *et al.* 2010; Simon *et al.* 2011)	LM
kknn	Weighted sum of k nearest neighbors	kknn	kernel = gaussian, optimal, rank, inv; k = 10:90:2; distance = c	(Hechenbichler and Schliep 2004; Samworth 2012)	Avg
knn	Average of *k* nearest neighbors	class	k = 1:100:1		Avg
lars	Least angle regression	lars	fraction = 0:1:0.05	(Efron *et al.* 2004)	LM
lasso	Linear regression with L1 regularization	elasticnet	fraction = 0:1:0.05	(Zou and Hastie 2005)	LM
monmlp	Ensemble of neural networks with monotonic constraints	monmlp	hidden1 $=2^{(0:5:1)}$; n.ensemble = 1:4:1	(Lang 2005; Zhang and Zhang 1999)	Bag, NN
nnet	Neural network	nnet	size $=2^{(0:5:1)}$; decay $=10^{(-6:-1:1)}$	(Haykin 2004)	NN
partDSA	Piecewise constant estimation sieve of candidate estimators based on a comprehensive search over the covariate space	partDSA	cut.off.growth = 1:10:1;MPD = 0.1:1:0.1	(Molinaro *et al.* 2010)	Tree

(*Continued*)

Table 2. (Continued).

Label	Description	R package	Tested parameters	References	Model class
pcaNNet	Neural network with principal component preprocessing	nnet	size $=2^{(0:5:1)}$; decay $=10^{(-6:-1:1)}$	(Eleyan and Demirel 2005)	NN
pcr	Principal component regression	pls	n.comp = 1:9:1	(Jolliffe 1982)	LM
penalized	Generalized linear model with L1 and/or L2 regularization	penalized		(Goeman 2010)	LM
pls	Partial least squares regression	pls	ncomp = 1:4:1, 8	(Wold *et al.* 1984)	LM
ppr	Projection pursuit regression	ppr	nterms = 1:9:1	(Friedman and Stuetzle 1981)	Add
ranger	Random forest	ranger	mtry = 1:10:1; min.obs.size = 5; splitrule = variance, extratrees	(Wright and Ziegler 2017)	Bag, Tree
rf	Random forest	randomForest	mtry = 1:10:1	(Breiman 2001)	Bag, Tree
ridge	Linear regression with L2 regularization	elasticnet	lambda = 0:1:0.05	(Zou and Hastie 2005)	LM
rlm	Robust linear regression that uses Huber's M-estimators	MASS	intercept = True,False; psi = humber, hampel, bisquare	(Huber 1981)	LM
rpart	Decision tree	rpart	complexity = 0.001:0.1:0.001	(Breiman *et al.* 1984)	Tree
rqnc	Non-convex penalized quantile regression	rqPen	penalty = SCAD, MCP; lambda $=10^{(-6:-1:1)}$	(Fan and Li 2001; Tibshirani 1996)	LM
spikeslab	Spike and slab model using a continuous bimodal prior	spikeslab	var = 1:9:1	(Ishwaran *et al.* 2010)	LM
spls	Sparse partial least squares regression	spsl	k = 1:9:1; kappa = 0:05:01; eta = 0:1:0.005	(Chun and Keleş 2010)	LM
svmLinear	Support vector machine with linear kernel	kernlab	C = 0.05:1:0.05	(Cortes and Vapnik 1995; Drucker *et al.* 1997)	SVM
svmPoly	Support vector machine with polynomial kernel	kernlab	degree = 1:3:1; scale = 0.01, 0.05, 0.1, 0.2; C = 0.1:1:0.1	(Cortes and Vapnik 1995; Drucker *et al.* 1997)	SVM
svmRadial	Support vector machine with radial basis kernel	kernlab	C = 0.1:1:0.1; sigma = 0.01, 0.02, 0.05, 0.1, 0.15, 0.2	(Cortes and Vapnik 1995; Drucker *et al.* 1997)	SVM
treebag	Bagging ensemble of decision trees	ipred		(Breiman 1996; Hothorn and Lausen 2005)	Bag, Tree
xgbTree	eXtreme gradient boosting decision trees	xgboost	eta = 0.02; gamma = 0; nrounds = 500; max_depth = 8:24:4, colsample_bytree = 0.3:0.7:0.2; subsample = 08:1.0:0.1, min_child_weight = 0.8:1.0:0.1	(Chen and Guestrin 2016)	Boost, Tree
xyf	Self-organizing map	kohonen	topo = hexagonal; xdim = 1,2,4,6,8,12; ydim = 1,2,4,6,8,12; user. weights = 0.2:0.8:0.2	(Kohonen 2001)	Avg

LM = linear model, NN = neural network, Avg = local averaging, SVM = support vector machine, Tree = decision tree, Bag = bagging, Boost = bootstrap aggregating, Add = additive model, Spline = model using splines.

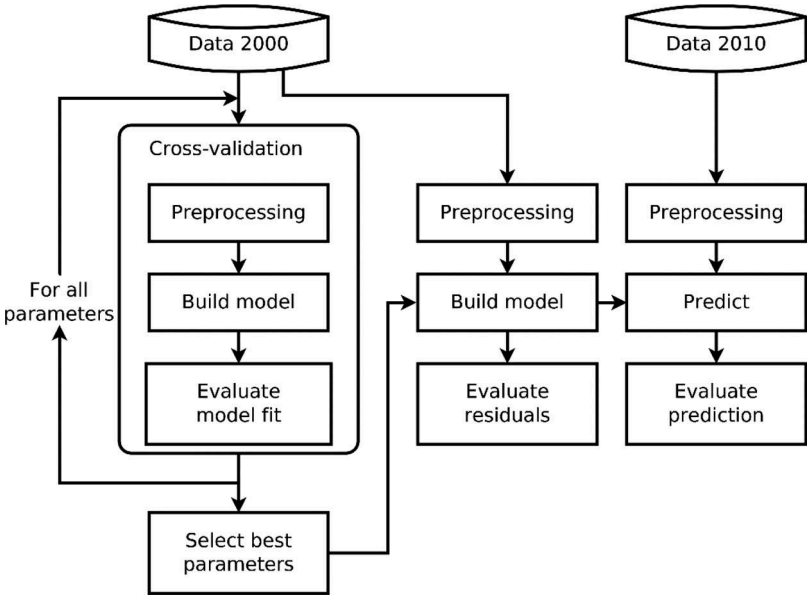

Figure 1. Workflow of parameter selection, model building, and prediction.

mean MAE were minor. The *partDSA* model generally achieved the lowest performance; its mean RMSE (0.535) and mean MAE (0.321) were higher and its mean R^2 (0.075) was lower than any other model.

The null hypothesis of no performance differences between the classifiers was rejected by the Kruskal–Wallis test at 5% significance level for MAE and R^2 but not for RMSE ($p < 0.05$). The null hypothesis of the Wilcoxon rank-sum of equal medians was rejected for many pairs of models for MAE and RMSE ($p < 0.05$); for RMSE, the null hypothesis was not rejected for any pair of models. With respect to RMSE and R^2, the null hypothesis of equal medians was not rejected for any pair of *gam, svmRadial*, and the tree-based ensemble models *rf, xgbTree, ranger, gbm*, and *cforest*. However, the null hypothesis was mostly rejected when comparing these models with linear models (e.g. *glm, lasso*, and *ridge*) or single tree models (e.g. *rpart* and *ctree*), indicating significant performance differences. Detailed results are given in the supplementary materials (Table S1-S3).

As an example, Figure 3 depicts the residuals of the *xgbTree* model (left panel), which is among the best performing models and, for comparison purposes, the residuals of *nnet* (right panel). While in general the residual means were similar across the study area, some minor patterns were observable. For instance, the figure shows *xgbTree* and *nnet* both underestimate the LCR for the city of Straubing and its surroundings (a), while they overestimate the LCR for the city of Augsburg and its surroundings (b). However, despite the overall better performance of *xgbTree* (Figure 2), it underestimates the LCR for region (a) more than *nnet*. Furthermore, *xgbTree* underestimates and *nnet* overestimates the LCR for the municipality of Hof (c).

Figure 4 shows the predicted mean LCRs for 2015–2021 in percentage (left panel) and the absolute LCR for the same period in ha per day (right panel). The latter refers to an

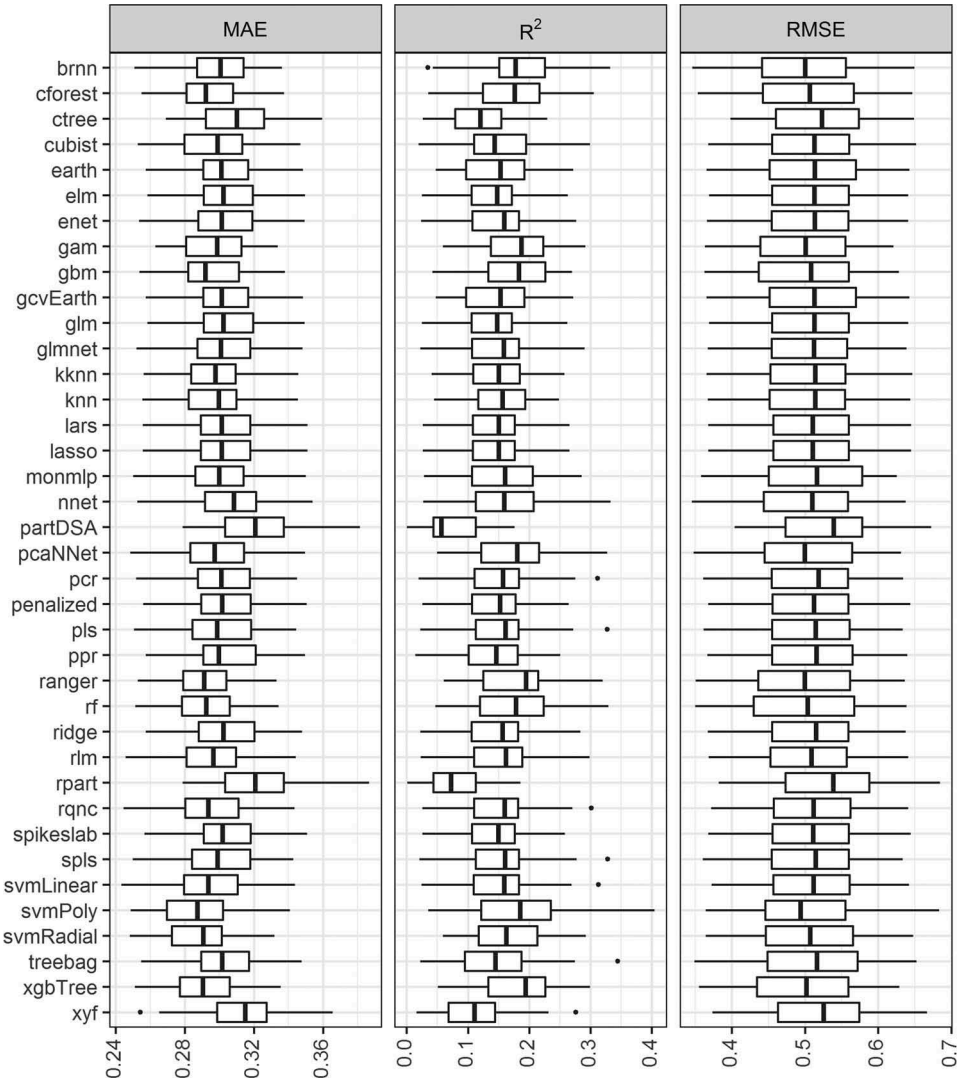

Figure 2. Performance boxplots calculated from the cross-validation results.

absolute measure that takes the municipality's area into account. The lowest mean LCR and absolute LCR were predicted by *cubist* (0.580% and 16.924 ha per day). The highest mean LCR was predicted by *ridge* (0.646%), while the highest absolute LCR was predicted by *rpart* (19.056 ha per day).

For 2009–2015, the observed mean LCR for Bavaria was 0.611%, which corresponds to an absolute LCR of 17.444 ha per day. For 2015–2021, almost all models predicted an increase in mean LCR and absolute LCR compared to 2009–2015. Since Bavaria is the federal state of Germany with the largest area, this result seriously challenges the German policy goal to limit LCR to 30 ha per day until 2020 (Statistisches Bundesamt 2012). The few models that predicted a decrease in both measures were *cubist, rqnc*, and

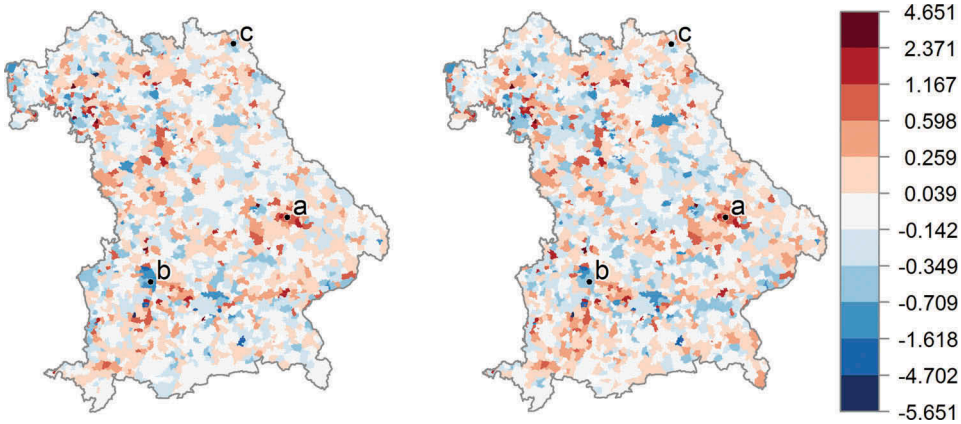

Figure 3. Residuals of the *xgbTree* model (left) and *nnet* model (right).

svmLinear, svmRadial, and *svmPoly*. A decrease in mean LCR and an increase in absolute LCR was predicted only by *rlm.*

Figure 5 compares the observed LCR for 2009–2015 (i.e. the response variable, left panel) and the predicted LCRs for 2015–2021 based on the *xgbTree* model (right panel). The model predicted for most municipalities an increase in LCR. A decrease in LCR was predicted for Straubing (a) and Nuremberg (c) and their vicinities, while a substantial increase was predicted for Augsburg (b) and its vicinity. Moreover, it can be seen that Augsburg's increase in LCR is much higher than that of Munich (d).

To show the influence of the different covariates on LCR, the importance of the variables for the *xgbTree* model is shown in Figure 6. Gain refers to the improvement in performance that is brought by a variable to the branches it is on. Cover measures the relative number of observations related to a variable. Frequency is the proportion of how many times a variable decides on a split in the trees. The figure shows that the present amount of built-up area, population density, and terrain ruggedness are the most important variables, while the least important are employment rate and trade tax. Latitude seems to be more important than longitude.

Figure 7 shows one-way partial dependence plots (Friedman 2001) for the four most relevant variables and a joint partial dependence plot for longitude and latitude (lower panel). Partial dependence plots visualize the change in the average predicted value as one or more variables vary over their marginal distribution (Goldstein *et al*. 2015). While all variables show a non-linear association with LCR, the strength of the association varies. The LCR increases with commuter rates substantially only from values higher than 0.85. Similarly, population density shows a positive association with LCR for values lower 2.0. While the LCR increases substantially with the proportion of built-up area, in particular from values 1.0 onwards, the LCR decreases marginally with terrain ruggedness. The LCR of municipalities close to the east of Upper Palatine (a) is associated with a lower LCR than municipalities that are far apart, in particular in the far south and north-west.

Finally, Figure 8 compares the required time to build the final models. All computations were performed on a standard laptop PC equipped with 32 GB of RAM and an

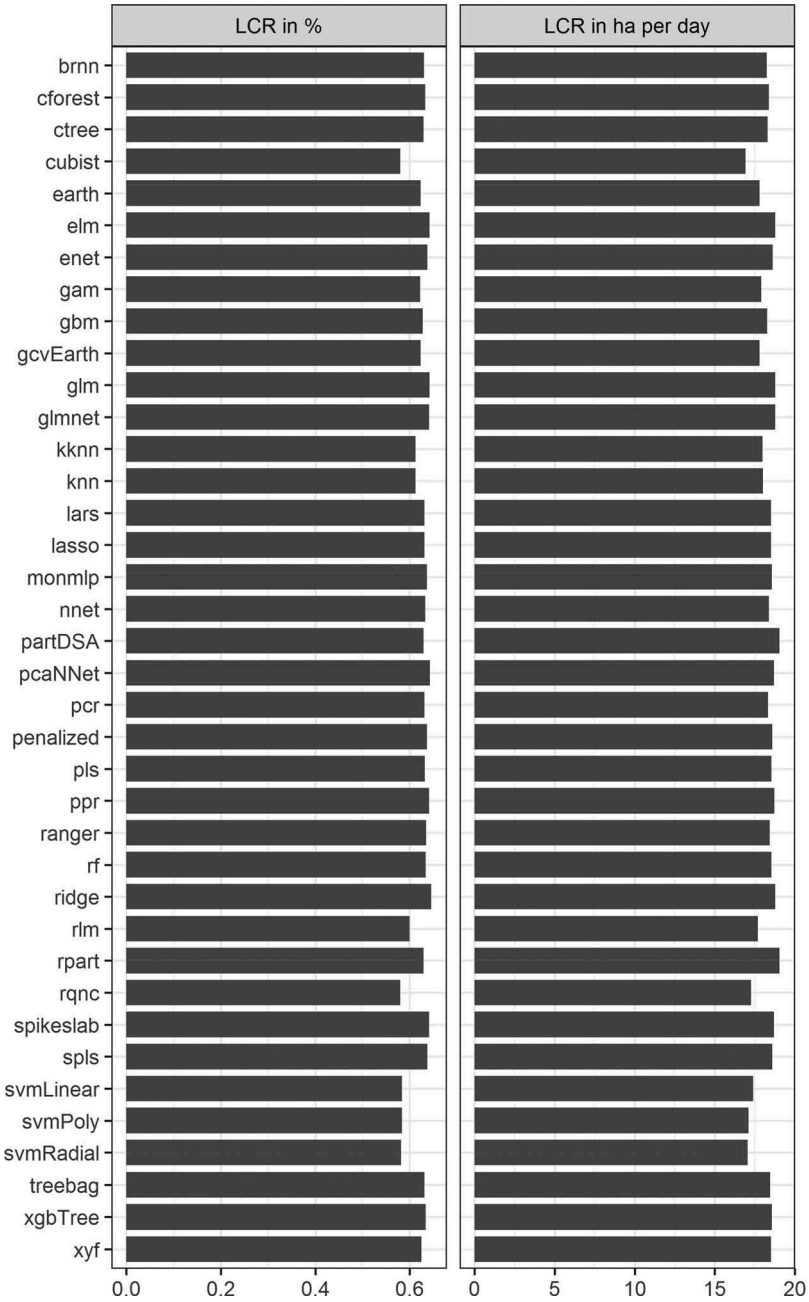

Figure 4. Predicted mean LCRs for 2015–2021 in percentage (left) and absolute LCR for 2015–2021 in ha per day (right).

Intel Core i7-6820HQ CPU @ 2.7 GHz. The most computational intensive was *cforest*, while the *rqnc* model was the fastest. As expected, ensemble models (e.g. *rf, treeBag, xgbTree, monmlp*) and support vector machine models generally took longer to build, while models based on linear regression were far less computationally intensive.

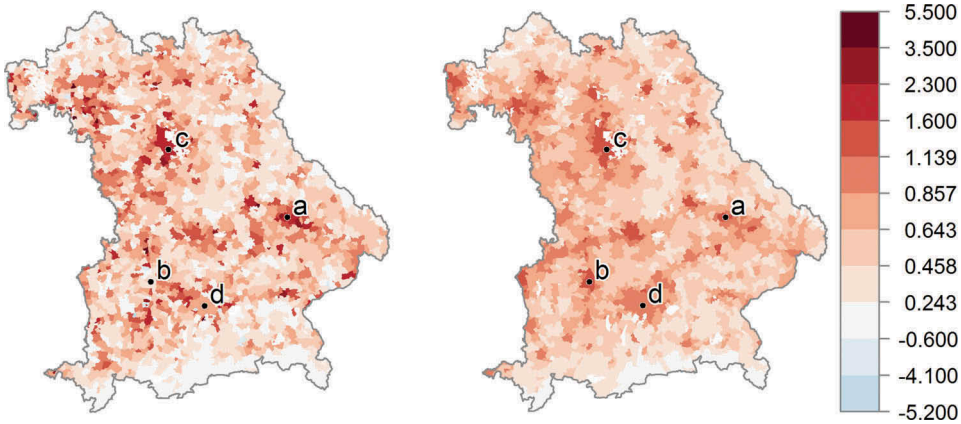

Figure 5. Observed LCRs for 2009–2015 (left) and predicted LCRs for 2015–2021 using *xgbTree* (right).

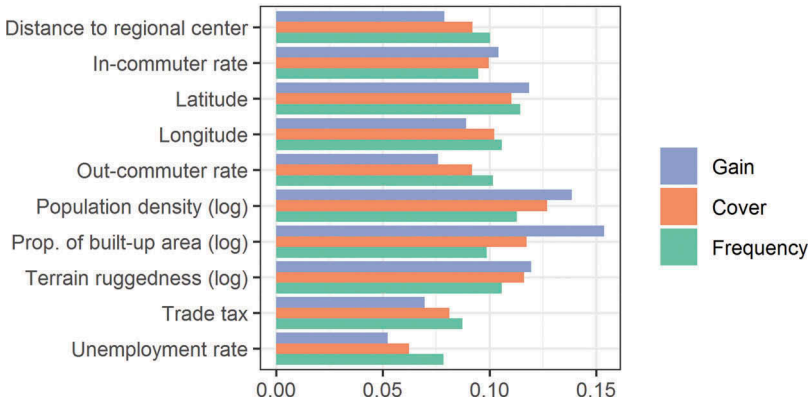

Figure 6. Variable importance for *xgbTree*.

Among the boosting models, running *cubist* took less time than *gbm* or *xgbTree*. In addition, *ranger* was much faster than other random forest models, such as *rf* and *cforest*.

5. Discussion

Rapid progress in ML has resulted in numerous models, necessitating comparative studies to guide application-oriented research (Fernández-Delgado *et al*. 2014). This is particularly true for land-use change science. The present study has contributed, to the best of our knowledge, the most comprehensive comparison of regression-based ML algorithms for a continuous outcome variable, namely LCR.

Generally speaking, ML has performed well in several situations, including land-use modeling where the input data is usually multidimensional (Shafizadeh-Moghadam *et al*. 2017). Our findings confirm this conclusion by indicating that *xgbTree, gbm*, and *gam* achieved the highest predictive accuracy compared to the other 35 models. Although

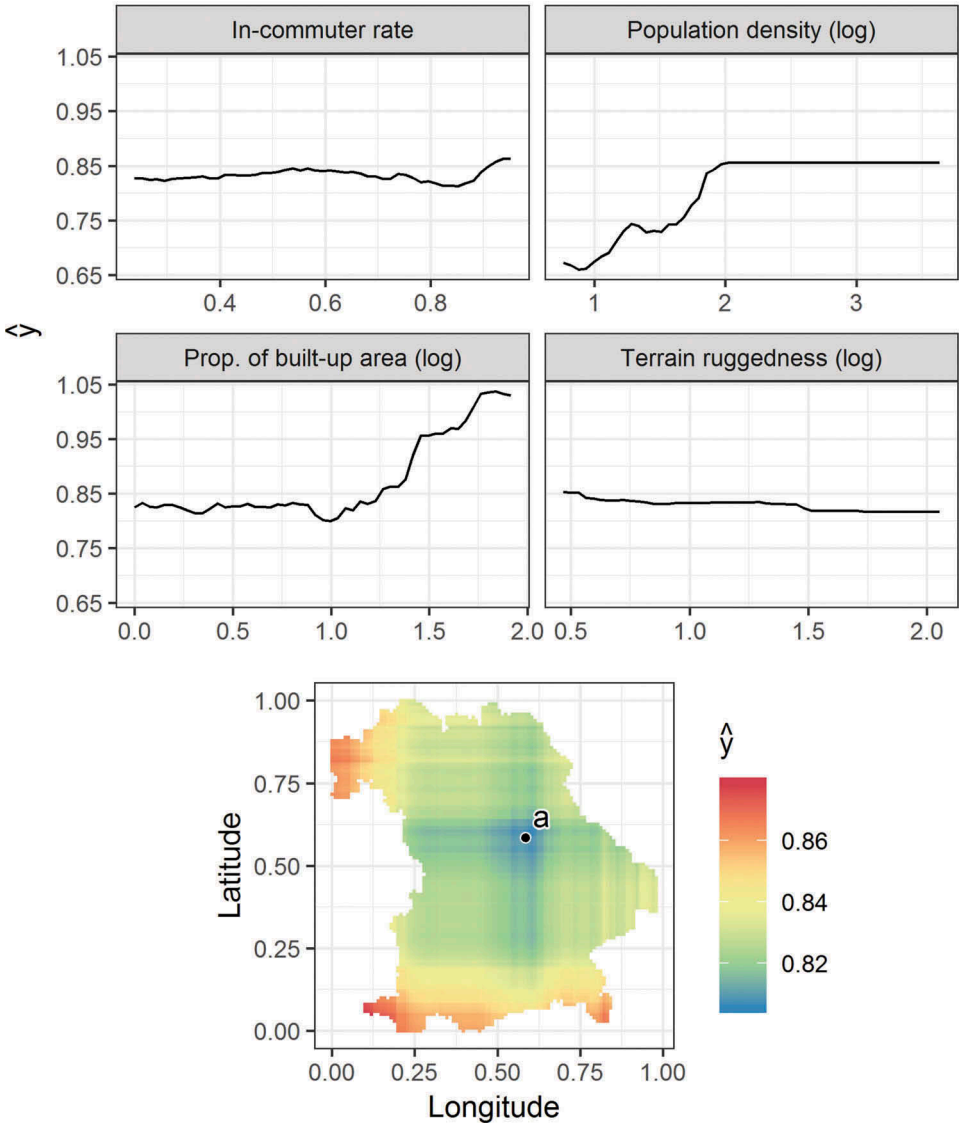

Figure 7. Partial dependence plots for the most relevant variables.

the *treeBag* algorithm is also a tree-based ensemble learner, it did substantially worse. We assume that this can be attributed to the tendency of *treeBag* to create correlated trees, which increases the upper bound for the generalization error (Breiman 2001). An ensemble model performs better if there is significant diversity among the sub-models forming the ensemble (Kuncheva and Whitaker 2003). Further, it is remarkable that *knn*, arguably one of the simplest models considered, performed well: It out-performed both single-tree models (e.g. *ctree* and *rpart*) and linear models (e.g. *lasso* and *glm*). This could imply that covariate interactions were minor, because, in contrast to *rpart, knn* is not able to model interactions among variables. This hypothesis is also supported by the very good performance of *gam*, which also does not model interactions.

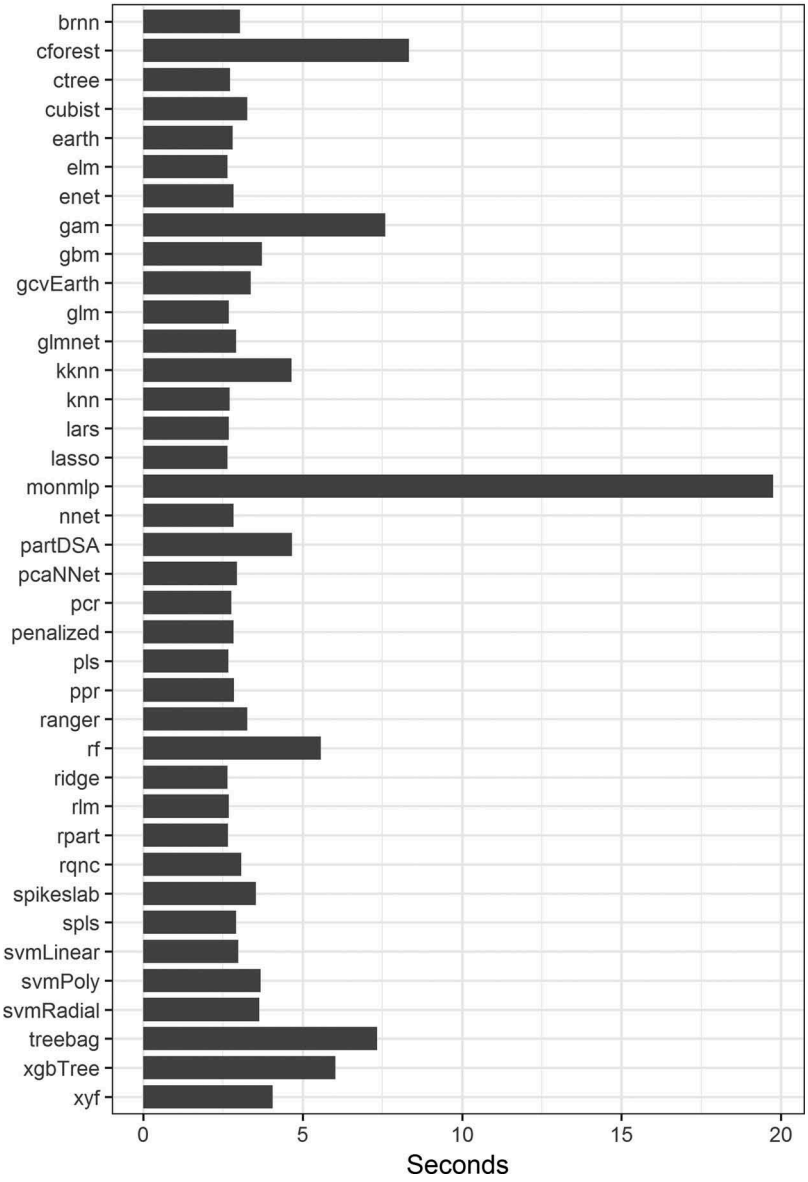

Figure 8. Computational times to build the final models.

As we are not aware of a study similar to ours using LCR as an outcome, we discuss our findings in the general context of land-use modeling (Tayyebi and Pijanowski 2014, Samardžić-Petrović *et al*. 2017, Shafizadeh-Moghadam *et al*. 2017, Du *et al*. 2018). While existing model comparisons consider rather similar models (i.e. artificial neural networks, multivariate adaptive regression splines, classification and regression trees), the literature is inconclusive about the performance of ML algorithms, although it does acknowledge that logit regressions (Jokar Arsanjani *et al*. 2013) are out-performed by ML (Shafizadeh-Moghadam *et al*. 2017). This is in line with our findings that generalized linear models had

a weak predictive performance. The results were less clear when comparing different ML algorithms. For example, others (Tayyebi and Pijanowski 2014) report that an artificial neural network provided higher accuracy, given limited model insights, than tree-based models and multivariate adaptive regression splines. Focusing on tree-based learners, Du *et al*. (2018) found that artificial neural networks are less suited and that extremely randomized trees perform best. In our case, artificial neural networks also had only an average accuracy. Confirming previous studies (Samardžić-Petrović *et al*. 2017), we also found that support vector machines (i.e. *svmRadial*) perform better than decision trees and artificial neural networks, though our study indicates that ensemble and boosting models can achieve better results. In another multi-label classification study (Omrani *et al*. 2015), *knn* was also found to be promising.

In total, no particular model in the reviewed ML studies stands out. Moreover, none of the studies made an effort to consider spatial autocorrelation (i.e. the fact that adjacent units are more similar than distant ones (Anselin 2010)), as is done in econometric-based land-use models (Shafizadeh-Moghadam and Helbich 2015, Ay *et al*. 2016, Hagenauer and Helbich 2018). By considering the longitude and latitude per area, we attempted to adjust for unexplained locational effects. Research should prioritize exploring the impact of autocorrelation on ML models.

5.1. *Strengths and limitations*

Only a limited number of land-use studies have compared ML models, and a large majority of previous studies considered land-use change as a classification problem and used raster data as input (Shafizadeh-Moghadam *et al*. 2017). We have extended this body of knowledge towards regression-based approaches by means of modeling changes in land consumption. Another strength is that the set of tested algorithms (38 models) is significantly larger compared to available land-use change studies. This supports future ML-based studies in selecting a model that will potentially have a high predictive accuracy. Further, the models were challenged through a unified framework using optimized tuning parameters and repeating cross-validation to adjust for sampling fluctuations coupled with several goodness-of-fit metrics. In contrast to others (Du *et al*. 2018), we also compared the algorithms in terms of their computation time. We used not only distance-based variables, which are without doubt essential drivers, but also demographic, socioeconomic, and environmental variables. Finally, we used data available to the public and share the used code to ensure the reproducibility of our findings.

Despite these strengths, the following limitations need to be acknowledged. Most of these ML models use several parameters, which influences their performance. To circumvent subjective parameter selections beyond the default values, we systematically investigated values from a manually specified subspace. A more exhaustive parameter search is computationally expensive. Though done with care, we cannot entirely rule out that the optimal parameters are covered within the considered search space. As our focus relied on land consumption, an infrequently used but highly relevant outcome variable for spatial planning, it remains unclear how generalizable our results are for other land-use change outcomes. Further, data for all German municipalities were not available to us and we cannot rule out that other relevant explanatory variables were missing. Different study areas as well as diverse input data may have an impact on the model performance. We

caution against blindly using a single algorithm; instead, we recommend testing at least a few of the ones that perform well. Despite this restriction, our results provide applied land scientists with essential insights into ML model performance. To address this restriction, we advise developing a benchmark data to support future model assessments. Because we used municipality-level data, we cannot rule out that the aggregation to municipalities affected the results (Openshaw and Taylor 1979). Moreover, although our study grounds on the smallest territorial level for which data were available, the size of the tested data is moderate, as is the complexity through the number of considered covariates. Future studies should consider higher dimensional input data.

6. Conclusion

The selection of ML algorithms is not a straightforward task due to the large number of available alternatives. The present study performed a systematic and comprehensive comparison of 38 regression-based ML algorithms for modeling land consumption, using a variety of performance measures grounded on repeated 10-fold cross-validation.

While the ranking of the models depended to a minor extent on the consulted goodness-of-fit metrics when assessing land consumption data, our results showed that eXtreme gradient boosting decision trees (*xgbTree*) performed substantially better with respect to the RMSE (0.500) and R^2 (0.183), while the support vector machine with radial basis kernel (*svmPoly*) had the lowest MAE (0.288). Of similar importance, other frequently applied ML models in land-use science only performed either moderately (e.g. *earth, nnet*, and *glm*) or poorly (e.g. *rpart* and *xyf*). Rarely used models such as partitioning using deletion, substitution, and addition moves (*partDSA*) and *rpart* did not perform well, independent of the performance measure. Due to their outstanding predictive performance, *xgbTree, gam*, and random forest models like *ranger* and *rf* seem to be a good initial choice when conducting a case study. However, we also suspect that they will not outperform the other models in all situations, and advise evaluating at least a few alternative models. Though not yet used, these well-performing ML approaches should play a major role in future land consumption and land-use studies to explore under which conditions they perform well and what results in a limited performance.

Notes

1. https://opendata.bayern.de/detailansicht/datensatz/verwaltungsgebiete .
2. https://monitor.ioer.de .
3. https://www.regionalstatistik.de .

Acknowledgments

We thank the anonymous reviewers for their constructive comments that greatly improved the article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Julian Hagenauer received his PhD in geography from Heidelberg University, Germany, in 2013. Currently, he is working as a data scientist at a global car manufacturer. His research interests are geocomputation, machine learning, and spatial data mining.

Hichem Omrani is senior researcher at Luxembourg Institute of Socio Economic Sciences (LISER). His research focuses on complex systems modeling, agent based systems, urban mobility and policy evaluation, and information technology and decision making.

Marco Helbich is an Associate Professor in the Department of Human Geography and Spatial Planning, Utrecht University. His research interests focus on geocomputational techniques and spatiotemporal analytics to address human–environment relations in cities.

References

Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science*, 89, 3–25. doi:10.1111/j.1435-5957.2010.00279.x

Ay, J.S., Chakir, R., and Le Gallo, J., 2016. Aggregated versus individual land-use models: modeling spatial autocorrelation to increase predictive accuracy. *Environmental Modeling and Assessment*, 22, 1–17.

Azari, M., et al., 2016. Integrating cellular automata, artificial neural network, and fuzzy set theory to simulate threatened orchards: application to Maragheh, Iran. *GIScience & Remote Sensing*, 53, 183–205. doi:10.1080/15481603.2015.1137111

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B,* 57, 289–300.

Bloom, D.E., Canning, D., and Fink, G., 2008. Urbanization and the Wealth of Nations. *Science*, 319, 772–775. doi:10.1126/science.1153057

Breiman, L., et al., 1984. *Classification and Regression Trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24, 123–140. doi:10.1007/BF00058655

Breiman, L., 2001. Random forests. *Machine Learning*, 45, 5–32. doi:10.1023/A:1010933404324

Bren D' Amour, C., et al., 2016. Future urban land expansion and implications for global croplands. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 8939–8944, 201606036.

Brown, D.G., et al., 2004. Modeling land use and land cover change. *In*: G. Gutman, et al., eds. *Land change science: observing, monitoring and understanding trajectories of change on the earth's surface*. Dordrecht: Springer Netherlands, 395–409.

Brown, D.G., et al., 2013. Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Green and Sustainable Chemistry*, 5, 452–457. doi:10.1016/j.cosust.2013.07.012

Chen, T. and Guestrin, C., 2016. XGBoost : A scalable tree boosting system. *In*: *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 785–794, San Francisco, CA. *arXiv*, 1–6.

Chun, H. and Keleş, S., 2010. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society. Series B (Methodological)*, 72, 3–25. doi:10.1111/j.1467-9868.2009.00723.x

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20, 273–297. doi:10.1007/BF00994018

Drucker, H., et al., 1997. Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155–161.

Du, G., et al., 2018. A comparative approach to modelling multiple urban land use changes using tree-based methods and cellular automata: the case of Greater Tokyo Area. *International Journal of Geographical Information Science*, 32, 757–782. doi:10.1080/13658816.2017.1410550

Dubovyk, O., Sliuzas, R., and Flacke, J., 2011. Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 235–246. doi:10.1016/j.isprsjprs.2010.10.002

Duranton, G. and Turner, M.A., 2012. Urban growth and transportation. *The Review of Economic Studies*, 79, 1407–1440. doi:10.1093/restud/rds010

Efron, B., et al., 2004. Least angle regression. *Annals of Statistics*, 32, 407–499. doi:10.1214/009053604000000067

Eleyan, A. and Demirel, H., 2005. Face recognition system based on PCA and feedforward neural networks. *In*: *Proceedings of the 8th international work-conference on artificial neural networks*. Berlin: Springer, 935–942.

Fahrmeir, L., et al. 2013. *Regression: models, methods and applications*. Berlin: Springer.

Fan, J. and Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348–1360. doi:10.1198/016214501753382273

Feelders, A., 1999. Handling missing data in trees: surrogate splits or statistical imputation? *European Conference on Principles of Data Mining and Knowledge Discovery*, 329–334. doi:10.1007/978-3-540-48247-5_38

Fernández-Delgado, M., et al., 2014. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15, 3133–3181.

Foresee, F.D., and Hagan, M.T., 1997. Gauss-newton approximation to bayesian learning. *In*: *Proceedings of the 1997 international joint conference on neural networks,* Houston, TX, 1930–1935.

Friedman, J., Hastie, T., and Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33. doi:10.18637/jss.v033.i01

Friedman, J.H., 1991. Multivariate adaptive regression splines. *Annals of Statistics*, 19, 1–67. doi:10.1214/aos/1176347963

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29, 1189–1232. doi:10.1214/aos/1013203451

Friedman, J.H. and Stuetzle, W., 1981. Projection pursuit regression. *Journal of the American Statistical Association*, 76, 817–823. doi:10.1080/01621459.1981.10477729

Goeman, J.J., 2010. L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, 52, 70–84.

Goldstein, A., et al., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24, 44–65. doi:10.1080/10618600.2014.907095

Hagenauer, J. and Helbich, M., 2012. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *International Journal of Geographical Information Science*, 26, 963–982. doi:10.1080/13658816.2011.619501

Hagenauer, J. and Helbich, M., 2018. Local modelling of land consumption in Germany with RegioClust. *International Journal of Applied Earth Observation and Geoinformation*, 65, 46–56. doi:10.1016/j.jag.2017.10.003

Hastie, T. and Tibshirani, R., 1986. Generalized additive models. *Statistical Science*, 1, 297–310. doi:10.1214/ss/1177013604

Haykin, S., 2004. A comprehensive foundation. *Neural Networks*, 2, 41.

Haykin, S.S. 2009. *Neural networks and learning machines*. Upper Saddle River, NJ: Pearson.

Hechenbichler, K. and Schliep, K. 2004. *Weighted k-nearest-neighbor techniques and ordinal classification*. Discussion Paper 399, SFB 386, Ludwigs–Maximilians University Munich. Available from: http://epub.ub.uni-muenchen.de/1769

Heung, B., et al., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, 62–77. doi:10.1016/j.geoderma.2015.11.014

Hothorn, T., et al., 2005. The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics*, 14, 675–699. doi:10.1198/106186005X59630

Hothorn, T., Hornik, K., and Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics*, 15, 651–674. doi:10.1198/106186006X133933

Hothorn, T. and Lausen, B., 2005. Bundling classifiers by bagging trees. *Computational Statistics & Data Analysis*, 49, 1068–1078. doi:10.1016/j.csda.2004.06.019

Huang, B., Xie, C., and Tay, R., 2010. Support vector machines for urban growth modeling. *Geoinformatica*, 14, 83. doi:10.1007/s10707-009-0077-4

Huang, G.-B., et al., 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42, 513–529. doi:10.1109/TSMCB.2011.2168604

Huber, P.J. 1981. *Robust statistics, analysis*. Hoboken, NJ: John Wiley & Sons

Ishwaran, H., Kogalur, U.B., and Rao, J.S., 2010. spikeslab: prediction and variable selection using spike and slab regression. *The R Journal*, 2, 68–73.

Jokar Arsanjani, J., et al., 2013. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observation and Geoinformation*, 21, 265–275. doi:10.1016/j.jag.2011.12.014

Jolliffe, I.T., 1982. A note on the use of principal components in regression. *Journal of Applied Statistics*, 31, 300. doi:10.2307/2348005

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *In*: C. S. Mellish, ed. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95)*, Montreal, Canada. Los Altos, CA: Morgan Kaufmann, 1137–1143.

Kohonen, T. 2001. *The self-organizing map*. Berlin: Springer.

Kretschmer, O., Ultsch, A., and 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. doi:10.3112/erdkunde.2015.03.05

Kroll, F. and Haase, D., 2010. Does demographic change affect land use patterns? A case study from Germany. *Land Use Policy*, 27, 726–737. doi:10.1016/j.landusepol.2009.10.001

Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, 28, 1–26. doi:10.18637/jss.v028.i05

Kuncheva, L.I. and Whitaker, C.J., 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51, 181–207. doi:10.1023/A:1022859003006

Lambin, E.F. and Geist, H.J. 2008. *Land-use and land-cover change: local processes and global impacts*. Berlin: Springer.

Lang, B., 2005. Monotonic multi-layer perceptron networks as universal approximators. In: Duch, W. *et al*, eds. *Artificial neural networks: Formal Models and Their Applications - ICANN 2005*, pp. 31–37. Lecture notes in Computer Science 3697. Berlin: Springer.

Lary, D.J., et al., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7, 3–10. doi:10.1016/j.gsf.2015.07.003

Lim, T.-S., Loh, W.-Y., and Shih, Y.-S., 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40, 203–228. doi:10.1023/A:1007608224229

Linard, C., Tatem, A.J., and Gilbert, M., 2013. Modelling spatial patterns of urban growth in Africa. *Applied Geography*, 44, 23–32. doi:10.1016/j.apgeog.2013.07.009

MacKay, D.J.C., 1992. Bayesian Interpolation. *Neural Computation*, 4, 415–447. doi:10.1162/neco.1992.4.3.415

Maxwell, A.E., Warner, T.A., and Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39, 2784–2817. doi:10.1080/01431161.2018.1433343

McCullagh, P. and Nelder, J.A., 1998. *Generalized linear models*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Meinel, G. and Schumacher, U., 2010. Konzept, Funktionalitaet und erste exemplarische Ergebnisse des Monitors der Siedlungs-und Freiraumentwicklung (IOER-Monitor). *In*: *Flaechennutzungsmonitoring II*. Berlin: Rhombos-Verlag, 183–200.

Molinaro, A.M., Lostritto, K., and van der Laan, M., 2010. partDSA: deletion/substitution/addition algorithm for partitioning the covariate space in prediction. *Bioinformatics*, 26, 1357–1363. doi:10.1093/bioinformatics/btq142

Mustafa, A., et al., 2018. Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computers, Environment and Urban Systems*, 67, 147–156. doi:10.1016/j.compenvurbsys.2017.09.009

Nuissl, H., et al., 2009. Environmental impact assessment of urban land use transitions – a context-sensitive approach. *Land Use Policy*, 26, 414–424. doi:10.1016/j.landusepol.2008.05.006

Omrani, H., et al., 2015. Multi-label class assignment in land-use modelling. *International Journal of Geographical Information Science*, 29, 1023–1041. doi:10.1080/13658816.2015.1008004

Omrani, H., et al., 2019. The land transformation model-cluster framework: applying k-means and the Spark computing environment for large scale land change analytics. *Environmental Modelling and Software*, 111, 182–191. doi:10.1016/j.envsoft.2018.10.004

Omrani, H., Tayyebi, A., and 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, 283–304. doi:10.1080/15481603.2016.1265706

Openshaw, S. and Taylor, P.J., 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical Application in the Spatial Science*, 21, 127–144.

Pijanowski, B.C., et al., 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. *Computers, Environment and Urban Systems*, 26, 553–575. doi:10.1016/S0198-9715(01)00015-1

Qian, Y., et al., 2014. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7, 153–168. doi:10.3390/rs70100153

Quinlan, J.R., 1992. Learning with continuous classes. *Machine Learning*, 92, 343–348.

Quinlan, J.R., 1993a. Combining instance-based and model-based learning. *Machine Learning*, 76, 236–243.

Quinlan, J.R. 1993b. *C4. 5: programs for machine learning, machine learning*. Amsterdam: Elsevier.

R Core Team, 2018. R Development Core Team. *R: A Language and Environment for Statistical Computing*.

Rienow, A. and Goetzke, R., 2015. Supporting SLEUTH – enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, 49, 66–81. doi:10.1016/j.compenvurbsys.2014.05.001

Riley, S.J., DeGloria, S.D., and Elliot, R., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *International Journal of Internet Science*, Available from: https://doi.org/citeulike-article-id:8858430

Rogan, J., et al., 2008. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. *Remote Sensing of Environment*, 112, 2272–2283. doi:10.1016/j.rse.2007.10.004

Samardžić-Petrović, M., et al., 2016. Modeling urban land use changes using support vector machines. *Transactions in GIS*, 20, 718–734. doi:10.1111/tgis.12174

Samardžić-Petrović, M., et al., 2017. Machine learning techniques for modelling short term land-use change. *SPRS International Journal of Geo-Information*, 6, 387. doi:10.3390/ijgi6120387

Samworth, R.J., 2012. Optimal weighted nearest neighbour classifiers. *Annals of Statistics*, 40, 2733–2763. doi:10.1214/12-AOS1049

Scholkopf, B. and Smola, A.J. 2001. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Cambrdige, MA: MIT press.

Seto, K.C., Guneralp, B., and Hutyra, L.R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 16083–16088. doi:10.1073/pnas.1211658109

Shafizadeh-Moghadam, H., et al., 2017. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, 64, 297–308. doi:10.1016/j.compenvurbsys.2017.04.002

Shafizadeh-Moghadam, H. and Helbich, M., 2015. Spatiotemporal variability of urban growth factors: A global and local perspective on the megacity of Mumbai. *International Journal of Applied Earth Observation and Geoinformation*, 35, 187–198. doi:10.1016/j.jag.2014.08.013

Siedentop, S. and Kausch, S., 2004. Die räumliche Struktur des Flächenverbrauchs in Deutschland. *Raumforschung und Raumordnung*, 62, 36–49. doi:10.1007/BF03183466

Simon, N., et al., 2011. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39. doi:10.18637/jss.v039.i05

Statistisches Bundesamt, 2012. *Nachhaltige Entwicklung in Deutschland – Indikatorenbericht 2012*. Wiesbaden: Statistisches Bundesamt.

Strobl, C., et al., 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9. doi:10.1186/1471-2105-9-307

Tayyebi, A. and Pijanowski, B.C., 2014. Modeling multiple land use changes using ANN, CART and MARS: comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *International Journal of Applied Earth Observation and Geoinformation*, 28, 102–116. doi:10.1016/j.jag.2013.11.008

Tibshirani, R., 1996. Regression selection and shrinkage via the lasso. *Journal of the Royal Statistical Society: Series B,* 58, 267–288.

Veldkamp, A. and Lambin, E.F., 2001. Predicting land-use change. *Agriculture, Ecosystems & Environment*, 85, 1–6. doi:10.1016/S0167-8809(01)00199-2

Witten, I.H., et al. 2016. *Data mining: practical machine learning tools and techniques*. Cambridge, MA: Morgan Kaufmann.

Wold, S., et al., 1984. The collinearity problem in linear regression. The Partial Least Squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific Computing*, 5, 735–743. doi:10.1137/0905052

Wright, M.N. and Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77, 1–17. doi:10.18637/jss.v077.i01

Xu, L., Li, J., and Brenning, A., 2014. A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery. *Remote Sensing of Environment*, 141, 14–23. doi:10.1016/j.rse.2013.10.012

Zhang, H. and Zhang, Z., 1999. Feedforward networks with monotone constraints. *International Joint Conference on Neural Networks (IJCNN)*, 3, 1820–1823.

Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Methodological)*, 67, 301–320. doi:10.1111/j.1467-9868.2005.00503.x