A Spatial Loss Function for Gradient Boosted Trees

Margot Geerts[1,*], Seppe vanden Broucke[2,1] and Jochen De Weerdt[1]

[1]Research Centre for Information Systems Engineering, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
[2]Department of Business Informatics and Operations Management, Ghent University, Tweekerkenstraat 2, 9000 Gent, Belgium

Abstract

While Gradient Boosted Trees (GBT) are generally preferred in predictive tasks with tabular data, they are unable to fully capture spatial effects in geographically referenced data. Recent advancements prove that significant levels of spatial autocorrelation remain present in the errors of these models. Residual spatial autocorrelation suggests the potential to gain predictive power by pushing the model to better capture spatial relationships in the data. To address this, we propose MI-GBT, a novel framework that accounts for spatial autocorrelation in GBT. Using the local variant of Moran's I, the framework integrates spatial autocorrelation into the objective function of GBT models so as to learn the spatial relationships of the data. An extensive evaluation of our proposed method on four real-life spatial regression datasets demonstrates the predictive gains with respect to regular GBT models as well as competitive spatial and deep learning baselines. Moreover, the residual spatial autocorrelation on a global as well as a local level is analyzed in-depth. With this comprehensive assessment, we can conclude that the MI-GBT framework gains predictive power while reducing the amount of residual spatial autocorrelation in a range of spatial regression tasks.

Keywords

Geospatial data, Gradient Boosted Trees, Spatial autocorrelation

1. Introduction

Geographically referenced data, crucial in fields like earth science, urban informatics, and public health, is challenging to model accurately due to its geospatial nature [1]. Conventional statistical and machine learning methods struggle to capture the spatial autocorrelation, leading to spatial bias [2]. Numerous spatial statistical methods, including spatial lag models and Gaussian Process models, can address this bias to some degree. However, with the rise of big data, there is an increasing need for scalable machine learning solutions, as spatial statistical models are typically computationally intensive [1].

In this study, tabular geospatial data is considered for geospatial regression tasks, where a continuous target is modeled based on a set of variables and the location (X-Y coordinates). Given the tabular nature of the data, related literature predominantly focuses on tree-based ensembles such as Random Forest (RF) and Gradient Boosted Trees (GBT) [3]. However, even though tree-based models such as RF, enhanced with spatial features, are considerably reducing prediction errors, they continue to have difficulties in capturing spatial bias [4].

More recently, methods have been proposed to account for spatial autocorrelation in the loss function of deep learning models [5, 6, 7]. This allows for a more direct approach to learn spatial dependencies in the data and simultaneously increase prediction accuracy [6]. Given that tree-based models outperform deep learning on tabular data, as has been shown in literature and as our experiments will prove, we extend this idea to Gradient Boosted Trees. Leveraging a tailored loss function, we propose a framework that accounts for spatial autocorrelation in GBT using Moran's I: MI-GBT[1]. Evaluating the proposed framework on four real-life geospatial regression datasets, its performance is quantified in terms of prediction accuracy but also residual spatial autocorrelation. Our contributions can be summarized as follows:

- A novel GBT-based framework, MI-GBT, is proposed that tailors the learning procedure to spatial dependencies.
- We extend the traditional approach of evaluating ML models with respect to their prediction error on the test set towards a spatial evaluation in terms of residual spatial autocorrelation.
- We demonstrate that the MI-GBT framework is able to considerably improve predictive performance while reducing the amount of spatial autocorrelation in the residuals.
- We show that GBT models outperform common spatial methods such as Gaussian Process Regression, and state-of-the-art deep learning methods for spatial data.

2. Related work

As defined in [1], there are three approaches for addressing spatial dependencies in prediction methods: spatial feature generation, spatial dependency in model structure, and spatial regularization in the objective function. In this study, we primarily focus on machine learning models in the context of spatial regression. As for spatial feature generation, distances to Points of Interest (POIs) are often used but also buffer distances to other observations have been shown to mimic kriging with Random Forests [8]. Spatial lags and eigenvectors also improve performance of ML methods for spatial prediction tasks, but accuracy gains can already be realized by simply incorporating the X- and Y-coordinates as features [4, 9].

Algorithmic improvements to solve the spatial dependency problem consist in extending the Generalized Least Square estimation for correlated data in linear models towards RF [10] and GBT [11], ensembles of local RF [12] and GBT [13] models, and a tailored tree learning algorithm with flexible geospatial splits for X- and Y-coordinates [14].

The final category, comprising our study, pertains to adjusting the objective function of the learning algorithm for spatial intricacies. In [15], a clustering tree model is

STRL'24: Third International Workshop on Spatio-Temporal Reasoning and Learning, 5 August 2024, Jeju, South Korea
*Corresponding author.
margot.geerts@kuleuven.be (M. Geerts)
0000-0002-9592-1843 (M. Geerts)
© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1]The Python implementation of the MI-GBT framework is available on GitHub: https://github.com/margotgeerts/MI-GBT.

CEUR-WS.org/Vol-3827/paper1.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

proposed that evaluates candidate splits based on a linear combination of variance reduction and spatial autocorrelation, measured by Geary's C. In [6], the authors propose a deep learning-based method for COVID-19 forecasting with a mean squared error (MSE) loss function adjusted with Moran's I. In the implementation of this loss function, the global Moran's I is calculated across batches and the MSE of the Moran's I is combined with the conventional MSE of the predictions. Instead of an MSE-based loss function, an entropy-based loss adjusted for spatial autocorrelation is proposed in [5] for deep learning energy forecasting models. As the entropy-based criteria measure dependence between observations within a time window, spatial dependency is included by weighting the contribution of observations by their distance. Finally, in [7], the local Moran's I is directly included in the learning process of graph neural networks in an auxiliary learning setting. The final loss function therefore consists of a combination of the loss of the main task and the loss of the auxiliary task, i.e. the predicted local Moran's I. In this study, we extend this approach to GBT models.

3. Methodology

3.1. MI-GBT

Gradient Boosted Trees (GBT) is an ensemble method consisting of a sequence of 'weak' decision tree models that try to improve the predictions of the previous model in the sequence [16]. A Gradient Boosted Tree model with M trees is essentially the sum of decision trees, expressed as $f_M(x_i) = \sum_{m=1}^{M} \nu \cdot T(x_i; \theta_m)$, with θ the tree structure and ν the learning rate. The training procedure consists of M iterations, where a new tree is added in each iteration with respect to a loss function $L(y, f_M(x))$. Given that the model $f_M(x)$ consists of a sequence of trees $T(x; \theta)$, the loss function is minimized to find the optimal tree structure θ_m in each iteration as follows:

$$\begin{aligned}\theta_m &= \arg\min_{\theta} \sum_{i=1}^{N} L(y_i, f_m(x_i)) \\ &= \arg\min_{\theta} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + T(x_i; \theta_m)).\end{aligned}$$

In this way, GBT models considerably decrease prediction bias. Similarly, this provides a way to account for spatial bias. We propose to extend the loss function to not only include the error with respect to the target but also the loss with respect to the spatial autocorrelation of the target. The total loss then becomes:

$$L_1(y, \hat{y}) + \lambda \cdot L_2(S(y), S(\hat{y})),$$

with L_1 and L_2 any regression loss such as mean squared error (MSE), y the target variable, $\hat{y}$ the predicted variable, λ the weighting scheme and S a function of spatial autocorrelation. The spatial autocorrelation is defined in this context by the local indicator of spatial association (LISA), as suggested in [17]. This is a local variant of the Moran's I statistic which quantifies the contribution of each observation to the global spatial autocorrelation. The local Moran's I is able to indicate spatial clusters, where the value of an observation is highly similar to the values of its geographic neighbors, and spatial outliers, where the value of an observation is highly dissimilar to the values of its geographic neighbors. The local Moran's I, LISA, S of a variable x_i is defined as

$$S(x_i) = (n-1)\frac{(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \sum_{j=1, j\neq i}^{n} w_{i,j}(x_j - \bar{x}),$$

with $w_{i,j}$ an element in the spatial weights matrix $\boldsymbol{w}$ which determines the neighboring observations. In this study, a k-nearest neighbors (kNN) matrix is employed with k as a parameter of the model. The kNN matrix can be either unweighted or weighted, resulting in a binary matrix in the former case and a numerical matrix with inverse Haversine distances in the latter case. In summary, the proposed GBT with the Moran-adjusted loss function has four parameters in addition to GBT's parameters as outlined in Table 1. We have abstracted from the specification of the base GBT model, as the proposed methodology can be applied to any GBT specification in theory. In what follows, the LightGBM (LGBM) [18] and XGBoost [19] implementations will be considered for the MI-GBT framework.

Table 1
Additional parameters for MI-GBT.

Parameter	Meaning
λ	Weight of local Moran loss
k	Number of nearest neighbors in spatial weights matrix
$weighted$	Distance-based spatial weights matrix if true binary spatial weights matrix if false
$loss$	Type of local Moran loss (MSE or Huber loss)

3.2. Experimental setup

3.2.1. Datasets

The proposed MI-GBT framework is evaluated on four real-life datasets containing geographically referenced data. As this study focuses on the spatial regression task, all datasets contain a continuous target variable, location variables (X- and Y-coordinates) and possibly other explanatory variables. The following datasets are considered, with N indicating the number of observations and m the number of explanatory variables:

- **Election**: The election dataset contains the election outcomes of 2016 in the different counties of the US (N=3107) [20]. The election outcome can be explained by socio-demographic variables such as median household income, net migration rate, and unemployment rate (m=9).
- **Air temperature**: The air temperature dataset comprises the mean temperature measured by 3076 weather stations positioned across the world [21]. In addition to temperature and X- and Y-coordinates, this dataset also contains measurements of precipitation (m=3).
- **3d road**: The 3d road dataset is a three-dimensional locational dataset containing X-, Y- and Z-coordinates [22]. The elevation Z of roads in North Jutland, Denmark is predicted based on the X- and Y-coordinates (N=397 978, m=2).

- **Housing**: The housing dataset consists of house prices of more than 250 000 homes in Flanders, Belgium that were sold between 2015 and 2023. The house prices can be explained by the housing characteristics in addition to the location (m=15). This dataset is proprietary.

The first three datasets are randomly split into 70% training, 15% validation, and 15% test data, while the housing data is split out-of-time with approximately 60% training data, 20% validation and 20% test data. As explained below, the validation set is used for hyperparameter tuning and the test set for reporting the results.

3.2.2. Methods

We compare the MI-GBT framework with LGBM and XGBoost as base models. Additionally, Gaussian Process Regression, a common spatial prediction method, and PE-GNN, a competitive graph neural network that accounts for spatial autocorrelation [7], are included in the benchmark. For GBT methods, a Bayesian search is performed using the validation set to find the optimal set of parameters. For LGBM and XGBoost, the number of estimators, learning rate, column sample size, maximum depth and subsample size are tuned. The MI-GBT variants are tuned with respect to the same parameters in addition to the λ, k, $weighted$, and $loss$ parameters (see Table 1). The PE-GNN model is specified with GCN layers, is trained for 1000 epochs with learning rate 0.001 and parameters lambda and embedding size are tuned with a grid search. The final model is selected based on the best validation MSE across epochs. For the Gaussian Process model, an approximate specification is trained using GPytorch [23] for 250 epochs with learning rate 0.01 and the same procedure for selecting the final model.

In the next section, we evaluate the methods on the four datasets using the prediction error on the test set and the spatial autocorrelation in the residuals. We measure the spatial autocorrelation in the residuals by the Global Moran's I and the LISA based on the optimal spatial weights matrix for respectively the MI-LGBM model and the MI-XGBoost model. Based on these spatial weights matrices, the spatial autocorrelation in the ground truth data is compared to the remaining level of spatial autocorrelation in the absolute errors of the models. The global spatial autocorrelation is quantified by the standardized Moran's I which can be compared across models. In addition, the percentage of observations with a statistically significant LISA value is investigated. The p-values are calculated based on 999 permutations and a significance level of 0.05 is employed. Given the randomness of permutations, the average value over 10 repeated calculations is presented for both the Moran's I and LISA.

4. Results and Discussion

The predictive performance of the models is evaluated using the RMSE, MAE and R^2 scores as presented in Table 2. The MI-GBT models are compared against their GBT counterpart as well as PE-GNN and GPR. From the election dataset, it is apparent that MI-LGBM is superior to the base LGBM model in terms of predictive power. On the contrary, the XGBoost variant shows superiority over the MI-XGBoost model in two out of four datasets. Surprisingly, the PE-GNN model has great difficulty to capture the patterns in the election dataset accurately, while GPR performs on par with the GBT models and presents slight superiority in two metrics. A different storyline unfolds for the temperature dataset, where all metrics indicate the advantage of the MI-GBT framework over the base models with MI-LGBM having the upper hand according to two out of three metrics. In contrast to the election dataset, the PE-GNN model shows similar performance to the GPR model on this dataset, but here both models underperform with respect to GBT models. For the 3d road data, neither GBT model differentiates between the standard loss function and the adjusted loss function in terms of predictive power. Moreover, the XGBoost models perform considerably better than LGBM for this dataset. Further, it is apparent that for the 3d road dataset, the GPR model performs much worse than all other models, including PE-GNN. This indicates that for pure spatial interpolation tasks, where only location and no other explanatory variables are available, ML methods are preferable. Finally, in the housing case, the MI-LGBM model consistently presents higher predictive power than the respective base model, while the reverse is true for MI-XGBoost. Nevertheless, PE-GNN and GPR are considerably outperformed by the GBT models and the MI-LGBM model presents the best performance overall for the housing dataset. In conclusion, all four datasets show a considerable advantage in the MI-GBT framework across GBT implementations, while only in two datasets the base XGBoost model performs better without the adjusted loss function. Moreover, GBT models generally outperform GPR and PE-GNN for geospatial regression tasks.

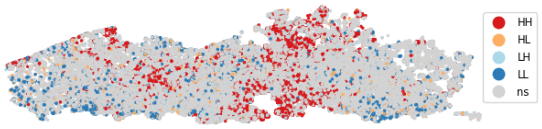
In Table 3, the global and local spatial autocorrelation in the ground truth data and in the absolute errors of the models is reported. While it is clear that all GBT models drastically reduce the level of spatial autocorrelation compared to the original data, though still statistically significant, interesting findings can be drawn from a comparison among GBT and baseline models. As for LGBM, though the indicators of spatial autocorrelation generally differ only slightly between the MI-LGBM and base models across datasets, the indicators are generally in favor of MI-LGBM. For the election dataset, both the global and local indicator appoints the MI-LGBM model to best capture spatial dependencies. For the temperature and 3d road datasets, the same pattern occurs in the Moran's I, but PE-GNN and LGBM respectively presents slightly less significant LISA values. The housing dataset shows the ability to capture spatial effects of MI-LGBM more convincingly, even presenting the least amount of residual spatial autocorrelation overall. In XGBoost, the adjusted loss function does not seem to push the model to better capture spatial relationships in the housing dataset. In contrast, the global and local Moran's I indicates lower spatial autocorrelation in the MI-XGBoost residuals for the temperature and 3d road datasets, even more so than LGBM-based models. While the 3d road dataset did not show any distinction in predictive power between the MI-GBT models and their respective base models, the advantage of the MI-GBT framework is underscored with less residual spatial autocorrelation. In the election case, the Moran's I does not differentiate between the base XGBoost and MI-XGBoost model, but the local indicator prefers the base model. While in general the MI-GBT framework shows improvement in the level of residual spatial autocorrelation, the base models are close contenders in stark contrast to GPR and PE-GNN. With the exception of the election dataset where XGBoost is slightly in favor, the MI-GBT framework demonstrates the least amount of residual spatial autocorrelation overall

Table 2
Prediction error metrics on the test set (RMSE, MAE and R^2 scores) of the proposed methods and baselines, for four different datasets. The best value according to each metric is underlined per group of models and highlighted in bold per column.

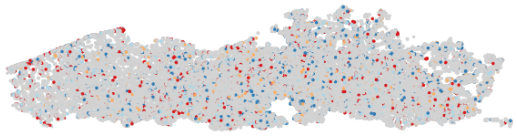
Model	Election			Temperature			3d Road			Housing		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
LGBM	0.0383	0.0125	0.7631	0.0392	0.0255	0.9560	0.0297	0.0195	0.9530	0.0942	0.0684	0.4924
MI-LGBM	0.0381	0.0123	0.7662	**0.0390**	0.0254	**0.9565**	0.0297	0.0195	0.9529	**0.0941**	**0.0683**	**0.4939**
XGBoost	0.0321	**0.0115**	0.8343	0.0419	0.0242	0.9497	**0.0263**	**0.0155**	**0.9631**	0.0943	0.0685	0.4911
MI-XGBoost	0.0334	0.0133	0.8180	0.0391	**0.0241**	0.9562	**0.0263**	**0.0155**	**0.9631**	0.0945	0.0686	0.4894
PE-GNN	0.0630	0.0324	0.3620	0.0685	0.0468	0.8658	0.0509	0.0362	0.8616	0.1112	0.0830	0.2925
GPR	**0.0317**	0.0158	**0.8389**	0.0591	0.0446	0.8998	0.0781	0.0560	0.6746	0.1073	0.0801	0.3414

Table 3
Standardized Global Moran's I and percentage of significant LISA values ($p < 0.05$) from the ground truth data and residuals of the proposed methods and baselines, for four different datasets. Two groups are defined by different spatial weights matrices resulting from the hyperparameter optimization of the two respective MI-GBT models. The best value is underlined per group of models and highlighted in bold per column.

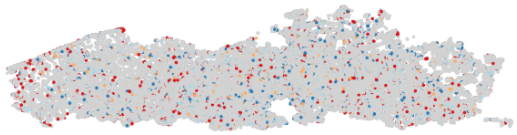
Model	Election		Temp.		3d Road		Housing	
	MI	LISA	MI	LISA	MI	LISA	MI	LISA
Ground truth	61	53.5%	214	82.5%	1612	69.3%	288	23.1%
LGBM	35	34.0%	62	65.8%	1049	52.0%	39	7.2%
MI-LGBM	33	29.6%	59	63.7%	1024	52.4%	**38**	**6.9%**
PE-GNN	39	48.7%	62	62.5%	1167	59.1%	63	8.4%
GPR	55	53.1%	80	67.4%	1520	64.5%	87	9.3%
Ground truth	52	42.0%	79	39.0%	1058	48.5%	622	59.2%
XGBoost	**24**	**15.7%**	**11**	7.4%	787	12.6%	65	20.3%
MI-XGBoost	**24**	21.5%	**11**	**7.2%**	**778**	**12.5%**	68	20.6%
PE-GNN	31	31.8%	36	18.1%	806	32.1%	117	26.3%
GPR	43	39.2%	47	24.8%	1024	41.9%	177	31.6%



(a) Ground truth



(b) MI-LGBM



(c) LGBM

Figure 1: LISA plots from housing data in the ground truth data (a), MI-LGBM residuals (b), and LGBM residuals (c). Points with significant local spatial autocorrelation ($p < 0.05$) are colored red, orange or blue, points with insignificant LISA values are colored grey.

across spatial regression datasets.

Finally, we investigate the LISA values for the housing data from the LGBM models visually in Figure 1. This figure contains the observations colored by the LISA values in the ground truth data (Figure 1a), in the residuals of the MI-LGBM model (Figure 1b), and in the residuals of the base LGBM model (Figure 1c). The colors represent the categorization of LISA values: spatial clusters of high values (red), spatial clusters of low values (blue), spatial outliers with high values surrounded by low-value neighbors (orange), spatial outliers with low values surrounded by high-value neighbors (light blue), and non-significant (grey). From Figure 1a, it is clear that the spatial clusters dominate, while spatial outliers are less prevalent. High-value spatial clusters are more common in urban areas, e.g. Brussels, Antwerp and Ghent. Low-value clusters are scattered across the region, and especially in rural areas in the west and east. These patterns seem to correspond with the typical patterns of house prices (high in urban areas, low in rural areas). More striking is the difference with Figures 1b-1c, where significant LISA values are much more sporadic. The LISA plots suggest that both models are able to capture a large part of the spatial effects indicated by the spatial clusters. For both the MI-LGBM model and base model, the low-value spatial clusters are the most common among the LISA categories. This indicates clusters of observations where the models predict closely to the true value. The MI-LGBM model presents approximately the same amount of low-value spatial clusters than the LGBM model, whereas it has less high-value spatial clusters. These findings suggest that while the MI-LGBM model has as many clusters of small errors, it has less clusters of large errors, boosting predictive power of the proposed method. As for spatial outliers, high-value spatial outliers are slightly more common in MI-LGBM than the base model, while low-value spatial outliers are somewhat less common. However, it could be argued that spatial outliers likely remain difficult to capture for any predictive model. A visual investigation of the geographic location of the significant LISA values in the residuals does not indicate stark differences between the models.

5. Conclusion

As previous research has shown that ML models generally present remaining spatial autocorrelation in the residuals, indicating that although they outperform other models in terms of predictive performance, spatial effects remain uncaptured. In deep learning methods, loss functions adjusted for spatial dependencies have been demonstrated to improve predictive performance. As we focus on the common task of geospatial regression with tabular data, we propose a framework that adjusts the loss function of Gradient Boosted Trees (GBT) for spatial autocorrelation: MI-GBT. Specifi-

cally, the adjusted loss function is a linear combination of the predicted target loss and the loss with respect to the local Moran's I of the target. MI-GBT provides the flexibility to learn the optimal kNN-based spatial weights matrix, type of local Moran loss and the importance of the latter with respect to the conventional loss. The proposed framework is evaluated on four real-life spatial regression datasets using two common GBT implementations, XGBoost and LightGBM. A thorough experimental evaluation showed that the MI-GBT models outperform the base models in all datasets but the housing and election datasets where the advantage is established in the MI-LGBM model while the MI-XGBoost presents on par results. Strikingly, GBT models in general performed considerably better than Gaussian Process Regression, a common spatial prediction method, and PE-GNN, a state-of-the-art deep learning method for geospatial data. Secondly, we evaluated the residual spatial autocorrelation in the models both globally and locally. While Gaussian Process Regression and PE-GNN established consistently higher levels of residual spatial autocorrelation, the MI-GBT framework proved more efficient in capturing spatial relationships than the respective base models evidenced by lower Moran's I and less significant LISA values.

References

[1] Z. Jiang, A survey on spatial prediction methods, IEEE Transactions on Knowledge and Data Engineering 31 (2019) 1645–1664. doi:10.1109/TKDE.2018.2866809.

[2] I. Song, D. Kim, Three common machine learning algorithms neither enhance prediction accuracy nor reduce spatial autocorrelation in residuals: An analysis of twenty-five socioeconomic data sets, Geographical Analysis 55 (2023) 585–620. doi:10.1111/gean.12351.

[3] L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on typical tabular data?, Advances in neural information processing systems 35 (2022) 507–520.

[4] X. Liu, O. Kounadi, R. Zurita-Milla, Incorporating spatial autocorrelation in machine learning models using spatial lag and eigenvector spatial filtering features, ISPRS International Journal of Geo-Information 11 (2022). doi:10.3390/ijgi.

[5] M. Ceci, R. Corizzo, D. Malerba, A. Rashkovska, Spatial autocorrelation and entropy for renewable energy forecasting, Data Mining and Knowledge Discovery 33 (2019) 698–729. doi:10.1007/s10618-018-0605-7.

[6] F. Olsen, C. Schillaci, M. Ibrahim, A. Lipani, Borough-level covid-19 forecasting in london using deep learning techniques and a novel mse-moran's i loss function, Physics 35 (2022) 2211–3797. doi:10.1016/j.rinp.2022.105374.

[7] K. Klemmer, N. S. Safir, D. B. Neill, Positional encoder graph neural networks for geographic data, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2023, pp. 1379–1389.

[8] T. Hengl, M. Nussbaum, M. N. Wright, G. B. Heuvelink, B. Gräler, Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, PeerJ (2018). doi:10.7717/peerj.5518.

[9] L. Hu, Y. Chun, D. A. Griffith, Incorporating spatial autocorrelation into house sale price prediction using random forest model, Transactions in GIS 26 (2022) 2123–2144. doi:10.1111/tgis.12931.

[10] A. Saha, S. Basu, A. Datta, Random forests for spatially dependent data, Journal of the American Statistical Association 118 (2021) 665–683. doi:10.1080/01621459.2021.1950003.

[11] R. Iranzad, X. Liu, W. A. Chaovalitwongse, D. Hippe, S. Wang, J. Han, P. Thammasorn, C. Duan, J. Zeng, S. Bowen, Gradient boosted trees for spatial data and its application to medical imaging data, IISE transactions on healthcare systems engineering 12 (2022) 165–179. doi:10.1080/24725579.2021.1995536.

[12] S. Georganos, S. Kalogirou, A forest of forests: A spatially weighted and computationally efficient formulation of geographical random forests, ISPRS International Journal of Geo-Information 11 (2022) 471. doi:10.3390/ijgi11090471.

[13] L. Deng, M. Adjouadi, N. Rishe, Geographic boosting tree: Modeling non-stationary spatial data, in: 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2020, pp. 1205–1210. doi:10.1109/ICMLA51294.2020.00190.

[14] M. Geerts, S. vanden Broucke, J. De Weerdt, An evolutionary geospatial regression tree, in: M. Sioutis, Z. Long, J. H. Lee, M. Bhatt (Eds.), Proceedings of the 2nd International Workshop on Spatio-Temporal Reasoning and Learning (STRL 2023) co-located with the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023), Macao, S.A.R., August 21, 2023, volume 3475 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023. https://ceur-ws.org/Vol-3475/paper4.pdf.

[15] D. Stojanova, M. Ceci, A. Appice, D. Malerba, S. Džeroski, Dealing with spatial autocorrelation when learning predictive clustering trees, Ecological Informatics 13 (2013) 22–39. doi:10.1016/j.ecoinf.2012.10.006.

[16] T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, volume 2, Springer, 2009.

[17] L. Anselin, Local indicators of spatial association—lisa, Geographical analysis 27 (1995) 93–115.

[18] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems 30 (2017).

[19] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.

[20] J. Jia, A. R. Benson, Residual correlation in graph neural network regression, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 588–598. doi:10.1145/3394486.3403101.

[21] J. Hooker, G. Duveiller, A. Cescatti, A global dataset of air temperature derived from satellite remote sensing and weather stations, Scientific Data 5 (2018) 180246. doi:10.1038/sdata.2018.246.

[22] M. Kaul, B. Yang, C. S. Jensen, Building accurate 3d spatial networks to enable next generation intelligent transportation systems, in: 2013 IEEE 14th International Conference on Mobile Data Management, vol-

ume 1, IEEE, 2013, pp. 137–146. doi:10.1109/MDM.2013.24.

[23] J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, A. G. Wilson, Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 31, Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper_files/paper/2018/file/27e8e17134dd7083b050476733207ea1-Paper.pdf.