European Journal of Geography

*The publication of the European Journal of Geography (EJG) is based on the European Association of Geographers' goal to make European Geography a worldwide reference and standard. Thus, the scope of the EJG is to publish original and innovative papers that will substantially improve, in a theoretical, conceptual, or empirical way the quality of research, learning, teaching, and applying geography, as well as in promoting the significance of geography as a discipline. Submissions are encouraged to have a European dimension. The European Journal of Geography is a peer-reviewed open access journal and is published quarterly.*

**Received:** 29/07/2024

**Revised:** 10/10/2024

**Accepted:** 16/10/2024

**Published:** 17/10/2024

**Academic Editor:**
Dr. Alexandros Bartzokas-Tsiompras

*Research Article*

# Introducing Spatial Heterogeneity via Regionalization Methods in Machine Learning Models for Geographical Prediction: A Spatially Conscious Paradigm

**Lukas Boegl** [1] **&** **Ourania Kounadi** [1]✉

[1] Department of Geography and Regional Research, University of Vienna, Universitätsstraße 7, 1010 Vienna, Austria

✉ Correspondence: ourania.kounadi@univie.ac.at

**Abstract:** This study addresses the challenge of incorporating spatial heterogeneity in predictive modeling by introducing regionalization methods in the preprocessing step of the modeling workflow. Spatial heterogeneity, where the mean of attribute values varies across spatial units, poses difficulties for traditional models. To tackle this, we propose a novel approach called Regionalization Random Forest (RegRF), which combines Random Forest with regionalization techniques to enhance predictive performance. Regionalization combines multiple spatial objects into homogeneous regions, which are incorporated into predictive models, allowing models to capture local variations. This research investigates three key questions: (1) How does the predictive performance of RegRF vary when constructed using different regionalization techniques? (2) How does RegRF compare to benchmark methods, including both spatial statistical approaches and spatially conscious machine learning models like Geographically Weighted Random Forest (GW-RF)? Five regionalization methods—WARD, AZP, Kmeans, SKATER, and Max-p—are tested on datasets of varying sizes. Results show that RegRF significantly improves performance over "non-spatial" Random Forest models with minimal additional computation time. While RegRF performs competitively with Geographically Weighted Regression, it requires much less computational effort. GW-RF was not outperformed on smaller datasets but failed to complete for larger datasets. These findings suggest that RegRF can enhance machine learning models by accounting for spatial phenomena, with potential for further optimization.

**Keywords:** spatial heterogeneity; regionalization; spatial clustering; geographical modelling; machine learning

**Highlights:**

- RegRF significantly increases the performance of the predictive models in comparison to "non-spatial" Random Forest models, while only taking a few seconds longer to compute.
- It competes the well-established Geographically Weighted Regression, while only requiring a fraction of the computational effort.
- It can be used for larger datasets while the Geographically Weighted Random Forest may not be able to finish computation.

**DOI:** 10.48088/ejg.l.boe.15.4.244.255

**ISSN:** 1792-1341



## 1. Introduction

This paper introduces and evaluates a novel method for incorporating spatial heterogeneity in geographical predictive modeling. Spatial heterogeneity describes spatial phenomena where attribute values vary across different spatial locations. This variation poses challenges for traditional models, as they often assume uniformity across a study area.

Tobler's first law of geography, stating that "everything is related to everything else, but near things are more related than distant things" (Tobler 1970), underscores the foundational challenge of spatial analysis, which is the non-random distribution and interdependence of spatial data points (Rogerson 2021). Traditional statistical methods often fail to account for these dependencies, leading to biased or inaccurate results (Grekousis 2020; Haining 2010). Various techniques are being used to capture the properties of spatial data such as Moran's I and the Geary Ratio, which quantify spatial autocorrelation (Moran 1950; Geary 1954) as well as geographically weighted regression (GWR) that allows and reveals local variations in parameter estimates (Brunsdon et al. 1996). GWR as well as its successor Multiscale Geographically Weighted Regression have consistently proven their relevance since their introduction, underscored by their continued use in spatial analysis and specifically urban planning (Roy and Chowdhury 2024; Majumder et al. 2023).

More recently, machine learning techniques such as geographically weighted random forest (GW-RF), have been proposed to incorporate spatial consciousness into predictive models (Georganos and Kalogirou 2022). GW-RF is based on the idea that spatial phenomena might be modeled better with an ensemble of local models instead of one global model. Thus, the GW-RF algorithm creates a model for each data point based

on data close to that point, instead of one model for the entire study area. GW-RF has been initially tested for the field of population modeling, and later in other domains such as health (Quiñones et al. 2021; Lotfata et al. 2022), geomorphology (Quevedo et al. 2021) and ecology (Santos et al. 2019).

Other approaches to the use of Machine Learning in spatial modeling include the work from Hengl et al. (2018), where buffer distances from locations are used to incorporate the effects of geographical proximity into the model. Talebi et al. (2022) proposed an approach that is able to incorporate heterogeneity, spatial dependencies and complex spatial patterns through the use of vectorized patterns from spectral information. Spectral information hints at the practical application of this method, which was done in the field of remote sensing. More specifically, the researchers tried to automate geology mapping in order to carry out geochemical prediction. Also, Mueller et al. (2018) incorporated cluster analysis results into models that estimate health insurance status, where the focus was put on localizing the model procedure in order to increase performance and maximize the insights that can be gained by the process. Stojanova et al. (2011) proposed ~~the~~ "Predictive Clustering Trees", a combination of clustering and prediction. This method identifies clusters of similar observations, before designing a predictive model for each of the clusters. The studies by Mueller et al. (2018) and Stojanova et al. (2011) can be seen as the most closely related works, with the important difference being the use of clustering instead of regionalization which will be used in this study. On the topic of spatial heterogeneity, more specifically in the realm of urban living, research is carried out by Roy et al. (2022) which shows how exploring spatial patterns can help in the detection of people's needs.

More recently, efforts to automate spatial machine learning have been discussed in the literature, such as in the work of Liang et al. (2024), who emphasize the need for broader frameworks that encompass both Machine Learning and spatial statistical methods. This paper addresses this by integrating regionalization, a key spatial statistical method, with machine learning through the Regionalization Random Forest (RegRF) model. By doing so, we aim to close the identified gap by evaluating new spatial prediction methods, extending beyond purely machine learning-focused approaches. The application of spatial machine learning methods in exploring variability in health outcomes, such as stunting in Rwanda (Nduwayezu et al. 2024), highlights the value of incorporating spatial heterogeneity in predictive modeling. By revealing local patterns that global models fail to detect, these methods underscore the importance of spatial models like RegRF, which is designed to capture such local variations across spatial datasets. Similarly, the application of Geographically Weighted Neural Networks to traffic crash injury severity (Zhang et al. 2024) showcases the importance of spatial heterogeneity in transportation studies. Methods like RegRF, which combine regionalization with Random Forest, provide a computationally efficient way to address these spatial dependencies, ensuring that local patterns are accurately captured in predictive modeling.

The discussed research projects collectively illustrate the recent efforts to address spatial heterogeneity in traditional machine learning modelling workflows. This paper builds on these efforts by proposing the Regionalization Random Forest (RegRF) method, which combines regionalization techniques with the Random Forest machine learning algorithm. Regionalization describes the procedure of grouping a large number of spatial objects into a smaller number of subsets of objects that are internally homogeneous and occupy continuous regions in space (Assunção et al. 2006). Unlike clustering, regionalization ensures that the grouped areas are spatially contiguous, making it particularly useful for geographical data. By combining regionalization with Random Forest, RegRF aims to improve predictive accuracy while maintaining computational efficiency. This paper contributes to the ongoing discourse on integrating spatial analysis into machine learning algorithms by offering a novel approach to capturing spatial heterogeneity in predictive modeling.

## 2. Research Scope

Despite the described advancements, there remains a significant gap in the limited exploration of how regionalization techniques can be integrated into machine learning models to improve predictive accuracy. Furthermore, the ubiquitous uncertainty surrounding spatial data, described by terms like spatial heterogeneity and spatial non-stationarity, presents a persistent challenge. The goal of spatial data analysis is to build models that can account for these phenomena. Box (1979) famously noted that "all models are wrong, but some are useful," emphasizing the continual need for developing and testing new models to find "more useful" ones. This research aligns with this perspective by proposing and evaluating a novel approach aimed at enhancing the utility of spatial models in practical applications.

The primary objective of this paper is to introduce the Regionalization Random Forest (RegRF) method and evaluate its performance against benchmark methods. The method proposed here incorporates regionalization techniques into machine learning models. The goal is to evaluate how good predictions made with a combination of Random Forest (RF) and regionalization are compared to benchmark methods like Random Forest, statistical methods like Geographically Weighted Regression (Brunsdon et al. 1996), and novel spatially conscious approaches such as Geographically Weighted Random Forest (GWR) (Georganos and Kalogirou 2022). Our research is guided by the following three research questions: 1. How does the predictive performance of RegRF vary by different regionalization techniques?, 2. How does RegRF perform compared to benchmark methods, either these being spatial statistical methods or machine learning algorithms?, and 3. How does RegRF perform compared to existing spatially conscious machine learning approaches such as geographically weighted random forest (GW-RF)?

The remainder of this paper is structured as follows: The following section (3) elaborates some important concepts in Regionalization and Machine Learning. Performance metrics which are used to assess the results are also discussed there. In the following section (4), a detailed description of the research design, data sources, and the implementation of the RegRF method are given. This is followed by section 5, where the results will be presented and discussed by comparing the performance of RegRF to benchmark methods, as well as discuss limitations of this research. Section 6 provides answers to the research questions before the final section 7 summarizes the key findings, lists implications for spatial analysis, and gives suggestions for future research efforts.

## 3. Methodological bases

### *3.1 Regionalization Methods*

Regionalization describes the procedure of grouping "a large number of spatial objects into a smaller number of subsets of objects, which are internally homogeneous and occupy continuous regions in space." (Assunção et al. 2006) The outcome of this procedure are newfound areas, referred to as regions. Influential work on the subject - especially in a socio-economic context - was pioneered by Openshaw (1977), where the process was mainly termed zone design. The Regionalization methods in use here are: WARD, AZP, SKATER, spatially constrained Kmeans, and

Max-p. The choice to use these specific methods was made based on a comprehensive literature review, ensuring that they are not only widely recognized but also the most relevant to the methodological framework of our experiment. These approaches align with the experimental design, allowing for accurate and reliable analysis of the variables in question.

**WARD** is a form of spatially constrained hierarchical clustering (SCHC) which uses Ward's linkage. An agglomerative hierarchical clustering approach is applied using Ward's linkage and an additional spatial connectivity constraint. The number of regions, as well as which attribute to use for the task needs to be defined by the user. (Feng et al. 2022). WARD is able to compute very efficiently while maintaining good performance, which is essential given the structure of our experiments.

The **automatic zoning procedure (AZP)** was originally described by Stan Openshaw (1977). It tries to group spatial units into a number of regions which also needs to be defined by the user. The criterion for forming the regions, the objective function, is the within region error sum of squares. The algorithm starts with an initial solution, with n units grouped into k regions. There are different approaches to constructing this initial solution, one example is to randomly select k units and add neighboring features until arriving at an end result. By doing this, each n is grouped exactly into one region k and contiguity is guaranteed. After this, one random k is chosen and all n neighboring this region are considered for a move to the chosen region k. The premise for a possible move is the preservation of contiguity. The decision whether the unit n is moved to the other region is made based on the objective function. This is repeated with all the neighbors until there are no more neighbors left to consider for possible moves. Then, another region is chosen at random and the process is repeated with its neighbors, until all regions have been evaluated in this fashion. After that, the whole process is repeated with the updated regions, until there is no more improvement in the objective function. In this research, AZP takes an important role, as its computation time is manageable for various dataset sizes which – like WARD – fits into the methodological framework of the experiment. The reason for AZP's efficiency lies in its design, as it already starts with an initial solution on which it tries to improve.

**SKATER** was developed by Assunção et al. (2006) and stands for "Spatial 'K'luster Analysis by Tree Edge Removal." This algorithm uses a connectivity graph to define the relationships between objects. Each object is represented by a vertex, and these vertices are linked to neighbors with edges. The edges are assigned with a cost, thus integrating the (dis-)similarity of the two objects the edge connects. This graph is then cut at appropriate places, resulting in connected clusters of objects, the regions. This transforms the regionalization task into a task of graph partitioning. The goal is to lower the graph's complexity by cutting it at edges that have a high cost, which means cutting it at edges that connect two dissimilar vertices. Removing edges splits the graph into multiple so-called subgraphs, which represent the regions. SKATER takes a considerable amount of time for larger dataset sizes, which arises from the way the algorithm is architected. Nonetheless SKATER should not be underestimated in its significance as the performance metrics obtained when using it are well balanced.

**Kmeans** is the most used clustering algorithm. Starting from a number k and n data points, the algorithm tries to find k centers which minimize the sum of squared distances from each data point to the closest center. In the beginning, k centers are chosen randomly and each point is allocated to the nearest center. After that, the center is calculated again as the center of all points allocated to it. These steps, allocation and center re-computation, are then repeated until the center point does not change anymore. Before running it, three parameters should be defined: the choice of where to place the initial cluster centers, the number of clusters and which variables to incorporate. Spatial constrained Kmeans adds spatial contiguity constraints to the classic Kmeans algorithm (Feng et al. 2022). The importance of Kmeans to this work is limited for the same reasons as SKATER, as it is not suitable for datasets with larger sample sizes. Reasons for this can be found when looking at the design of the algorithm with its extensive re-computation.

**Max-p** distinguishes itself from the other methods by not needing the number of regions as an input parameter. In fact, it makes this part of the solution process. Instead, the algorithm requires another parameter: a minimum size. This can either be a number of features a region is made of, or a minimum attribute value that a region needs to have. In the first step, the largest possible p value (max-p, p is the number of clusters) which can fulfill the minimum size and the contiguity constraints is determined. For this, different solutions are considered, which are found by growing the regions from random starting points and adding neighbors until the minimum size criterion is met. During this, some units will not be assigned to a region, due to them breaking at least one of the two constraints, which will lead to them being stored in enclaves. These need to be assigned to one region before the optimization process can be started. When this has been done, all of the possible solutions for a value p are considered and processed in the same way as for AZP (Wei et al. 2021; Duque et al. 2012). This leads to the fact that Max-p requires considerable computation time, even for small sample size datasets, which makes it difficult to integrate into the workflow of our experiments. Nonetheless, when results can be obtained, Max-p shows promising results.

### *3.2 Random Forest*

The predictive models of RegRF were created using the prominent machine learning algorithm **Random Forest (RF)** in its regression form. It is made out of an ensemble of decision trees, with the multitude of trees constructing the "forest". The thought underlying random forest is averaging a multitude of decision trees in scope of creating a robust model with good generalization performance and insusceptibility to overfitting (Raschka and Mirjalili 2017). The algorithm can be described in three steps:

1. Generate $n_{tree}$ samples through bootstrapping from the data (bootstrapping = sampling with replacement).
2. For every sample, create a regression tree that is split at its nodes based on a random sample of the predictor variables, called $m_{try}$.
3. Aggregate all predictions made by the $n_{tree}$ decision trees into the final prediction

Reasons for why the RF algorithm was used in this study are manifold. They are very user-friendly, as they can be used for both classification and regression tasks. Additionally, they only need two input parameters: $m_{try}$ being the number of variables in each random subset, and $n_{tree}$ describing how many trees the forest should consist of (Liaw and Wiener 2002, 18). Adding to that, it has shown to be suitable for geographical modeling, for example by Hengl et al. (2018), Nussbaum et al. (2018), and Prasad et al. (2006). Moreover, tree-based methods have shown to be the most suitable of all statistical learning methods when carrying out spatial data analysis. This is due to their aforementioned simplicity, their ability to express nonlinear relations, their strengths in handling big amounts and dimensions of data, as well as missing values (Kuhn and Johnson 2013).

To use spatial data in Machine Learning, spatial information needs to be added to the learning process, which in this case is done through Regionalization. When integrating spatial properties into the observation matrix, standard Machine Learning methods and algorithms can be used without adapting them. This of course postulates that the observation matrix is as good as possible, which requires consideration of various important aspects, these being proper sampling techniques, the inclusion of spatial features, dimensionality reduction and the handling of missing data (Nikparvar and Thill 2021, 8).

### *3.3 Performance Metrics*

The results of the predictive models are evaluated using two performance metrics. One of those is the **R-squared value ($R^2$),** which is a statistical measure introduced by Wright (1921) that evaluates the direct contribution of a given factor in explaining the variability within a system. In this context, $R^2$ quantifies the proportion of variance in the dependent variable that is predictable from the independent variables. An $R^2$ value of 1 indicates a perfect fit, where the model's predictions exactly match the observed data. Additionally, the spatial autocorrelation in the residuals is used. This refers to the degree to which residuals from a regression model are associated across space. This was calculated using the **Local Indicators of Spatial Association (LISA)** (Anselin 1995), which identifies spatial groupings. The percentage of points that form insignificant LISA clusters indicates how well the model has accounted for spatial autocorrelation (Liu et al. 2022). High percentages of insignificant LISA clusters suggest that the residuals are randomly distributed in space, implying that the model effectively captures the spatial processes influencing the dependent variable. Conversely, lower percentages indicate significant spatial autocorrelation in the residuals, suggesting that the model may have missed some spatial dependencies.

## 4. Materials and Methods

### *4.1 Dataset*

The dataset used in this research is the California Housing dataset, originally introduced by Pace and Barry (1997). This dataset comprises 20,640 observations with 9 variables each that are listed and described in Table 1, which can be seen below. The dataset is highly regarded in spatial autocorrelation analysis and machine learning tasks due to its detailed and extensive nature (Klemmer et al. 2019; Liu et al. 2022). As some of the modelling processes that are described in the next sections were not able to run with the full dataset, we also created a smaller subset of it. For this, all overlapping points were averaged into one point. Afterwards, a new column was created and populated with random numbers, before randomly selecting one of these numbers and only keeping points with the corresponding number in the generated column. This led to a decrease in sample size from 20,640 to 2051. As we can see in the two maps underneath the table (Figure 1), the two datasets exhibit a similar pattern of the locations of the samples as well as the attribute values.

**Table 1.** Variables of the California Housing dataset and their description.

| Variable | Description |
|---|---|
| median house value | Median House Price |
| median income | Median Income (per 10,000 population) |
| housing median age | Median House Age |
| total rooms | Total number of rooms |
| total bedrooms | Total number of bedrooms |
| population | Total number of inhabitants |
| households | Total number of households |
| latitude | Coordinate (WGS84) |
| longitude | Coordinate (WGS84) |

### *4.2. Data Pre-processing*

Data pre-processing involved several steps to ensure the dataset was ready for analysis:

1. Transformation of Variables: The "total rooms" and "total bedrooms" variables were converted to per household metrics to standardize the data.
2. Geometric Conversion: Latitude and longitude coordinates were converted to point geometries with GeoPandas, facilitating spatial analysis.
3. Displacement of overlapping points: Overlapping points were minimally displaced to ensure spatial accuracy. This was necessary because of a function that is part of feature engineering, which creates Voronoi polygons for each point.
4. Reprojection of Dataset: The dataset was reprojected to a projected coordinate reference system (NAD83 / California Albers (EPSG: 3310)) to maintain consistency in spatial analysis.
5. Nested-cross-validation: Nested CV was used to generate more reliable estimates on the predictive performance. By splitting the data in 3 different ways (folds) and thus carrying the experiment out 3 times, the impact of one specific train-test-split was decreased (Raschka and Mirjalili 2017).
6. Train-Test subsets: Train-test sample subsets were created and used identically to ensure fair comparison across different modelling methods.

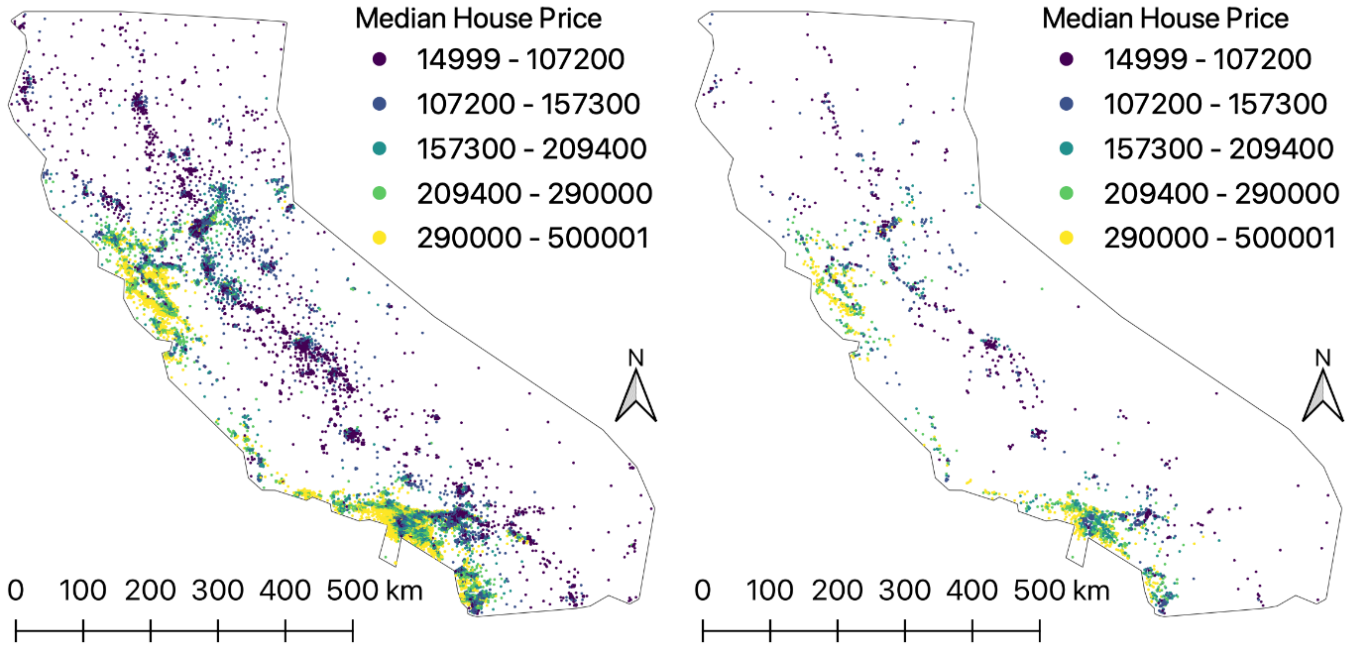


**Figure 1.** Median house price: full dataset, 20,640 points (left) and small dataset, 2051 points (right).

### *4.3. Regionalization Random Forest (RegRF)*

The workflow for RegRF consists of 3 steps:

1. Carry out the regionalization procedure on the training data. The clusters (regions) are then inserted as a spatial feature into the observation matrix.
2. A spatial join is performed between the test data and the training data to identify the cluster region of each test point. Thus, the spatial feature of the test data is engineered in this step.
3. Carry out the RegRF prediction. After fitting the model, predictions on the test set are carried out and reported.

The number of clusters was empirically optimized by inspecting the distribution of R-squared values within the range between 2 and 70 clusters, for the regionalization techniques WARD and AZP, and for 3 folds (414 models). Finally, 23 clusters (regions) were chosen and applied to all methods. Kmeans, Max-P, and SKATER were not included in this optimization task due to time constraints as they require extensive computation time to create the regions. The results are available in the Supplementary Information - Optimization of WARD and AZP.

The workflow above was applied to the small dataset and 23 regions for the five regionalization techniques, for the three benchmark models, namely RF, GWR and GW_RF, and for the three folds, thus resulting in 24 models to be compared. It was not possible (due to computation constrains) to run the models with the full dataset for all regionalization techniques except for WARD and also not possible to run it for GW_RF. Specifically, the GW-RF script for the full dataset required over 100 hours to run on a powerful machine, indicating that it exceeds the computational capabilities of most private devices, thus limiting its practical usability for large datasets.

Therefore, models with the full dataset and 23 regions were created for WARD, Random Forest, and Geographically Weighted Regression, and for the three folds, thus resulting in 9 models to be compared. The predictive results of all models are available in the Supplementary Information - Performance metrics of all tested models.

### *4.4. Resulting regions by technique*

The points of the train dataset were transformed into polygons to enable a “point to polygon” spatial join as it was needed to transfer the regions spatial feature between test and train data. Thus, the regionalization techniques were applied to the train polygon data which were created from Voronoi polygons from the train points. Figure 2 shows the resulting 23 regions (clusters) produced by each method. Although the areas of the regions vary greatly, one can still observe some prominent patterns. First, for all methods, a large number of small clusters is observed in the south west part of the study area that indicates greater variability of the attribute values. Similarly, for all methods except Max-p, a prominent cluster can be observed around the middle part of the coast in the west of the study area. If we look at these regions in Figure 2, we can see that they have a lower number of samples and greater variability of attribute values. Last, for all methods smaller number of clusters that are larger in size are observed on the east part of the study area. This is an effect of fewer sample points there but also because in the east part the attribute values are generally lower and less diverse compared to the west part.

https://doi.org/10.48088/ejg.l.boe.15.4.244.255

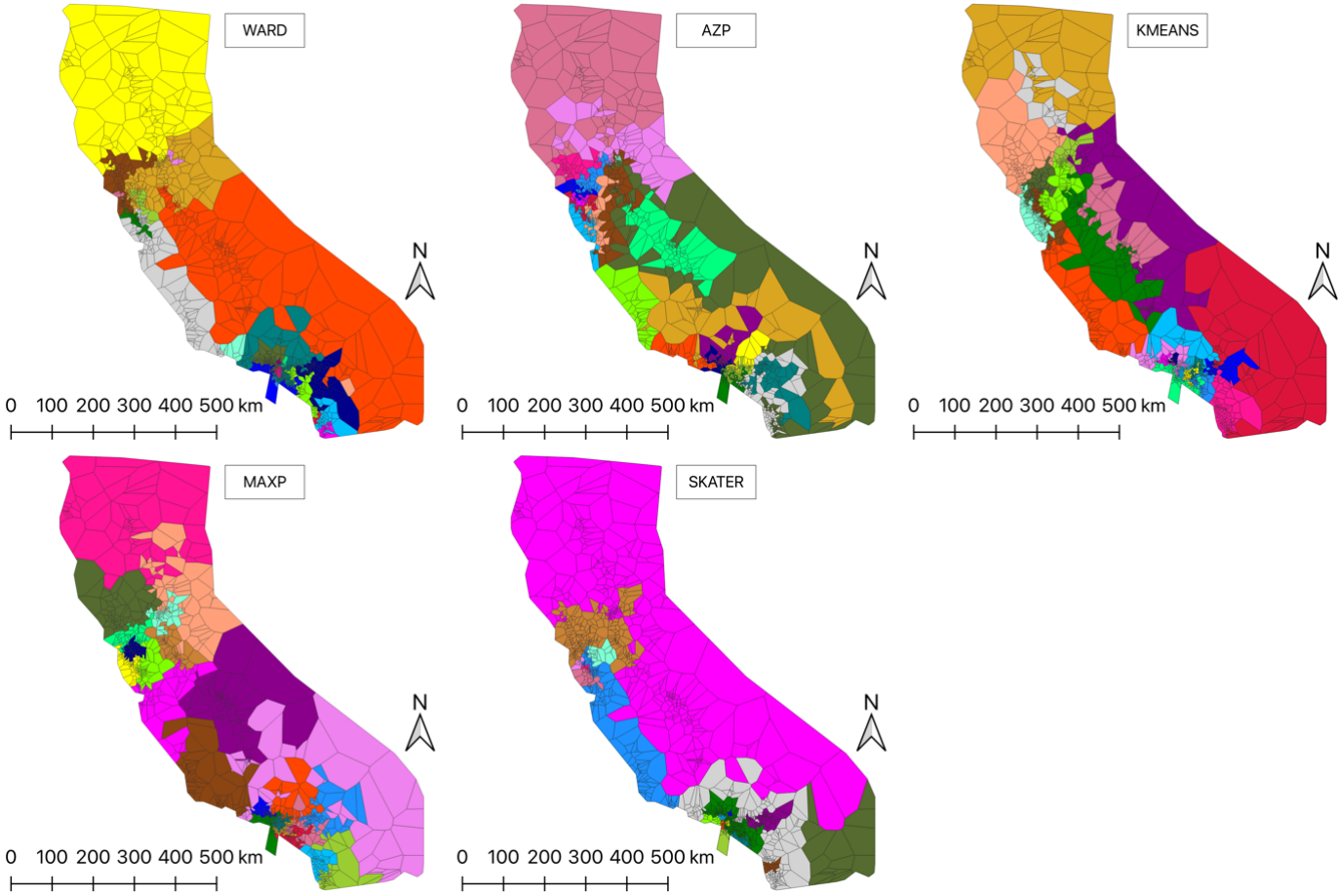


**Figure 2.** The 23 regions for the small dataset and each regionalization method (WARD, AZP, Kmeans, SKATER, Max-p).

### *4.5. Software*

Python and R were used for processing, data analysis and modeling. Python was used for numerical operations, data manipulation, and machine learning tasks. Key Python libraries included are Pandas, Scikit-learn, GeoPandas and spopt. R was employed specifically for implementing Geographically Weighted Random Forest (GW-RF) through the SpatialML package. Additionally, QGIS was used for GIS tasks, such as displacing overlapping points and creating maps. The modelling workflow is publicly available as Jupyter Notebooks and can be found on GitHub – repository "Regionalization Random Forest"[1].

## 5. Results

### *5.1. Comparison of regionalization methods*

Table 2 below lists the R2 values as well as the percentage of residuals that were not spatially clustered for each of the five RegRF models (WARD RF, AZP RF, Kmeans RF, SKATER RF, and Max-p RF) that were built and tested using the small dataset. In each model the spatial feature of the clusters was calculated using a different regionalization technique. Each row in the table presents a different fold, the final row shows the average of the evaluation metrics across folds.

Looking at those values shows that WARD performs the best regarding the R2 values, followed by Max-p with SKATER being very close to it. Kmeans, but especially AZP are lacking behind in this category. One possible reason for this has been mentioned in a previous part of this work, which is that the numbers given here were optimized using the WARD regionalization algorithm. Given this fact, both SKATER and Max-p produce very respectable outcomes. The superior performance of WARD in terms of R2 values indicates that this regionalization method captures the spatial structure of the data more effectively, which could have important implications for the practical application of RegRF in real-world scenarios where spatial structure plays a key role, such as in environmental modeling or socio-economic planning.

When looking at the number of insignificant LISA clusters, the performance of the respective algorithms is very different. For this performance metric, AZP performs the best, where 83% of LISA clusters are insignificant. The second and third highest values for the LISA clusters can be found for Kmeans and WARD. SKATER and Max-p perform the worst in this category with only 75% of clusters being insignificant. The higher number of insignificant LISA clusters for AZP suggests that it is particularly effective in minimizing spatial autocorrelation in the residuals, which is

[1] Regionalization Random Forest – RegRF. GitHub repository: https://github.com/Digital-Geography/Regionalization-Random-Forest

crucial for ensuring the independence of model errors and thereby improving model reliability. This may make AZP a more suitable method in contexts where reducing spatial dependence is a priority.

Another aspect that deserves some attention is the runtime of the algorithms. There are big differences to be found, for example WARD is very fast and takes about 13 seconds for this dataset. AZP is able to finish computation in about 170 seconds, SKATER and Kmeans both need about 130 seconds to deliver a result. By far the longest computation time can be found for Max-p, which takes more than 40 minutes. One reason for this could be the fact that Max-p determines the number of regions on its own. While it offers strong performance in terms of accuracy, its significantly longer runtime makes it impractical for time-sensitive applications. This trade-off between computational efficiency and model performance is a critical factor to consider when applying these models in fast-paced environments such as emergency response.

**Table 2.** Performance of RegRF for each regionalization method and across three folds, small dataset. LISA column shows the percentage of residuals that were not spatially clustered.

| | WARD RF | | AZP RF | | SKATER RF | | KMEANS RF | | MAX–P RF | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Fold** | **$R^2$** | **LISA** | **$R^2$** | **LISA** | **$R^2$** | **LISA** | **$R^2$** | **LISA** | **$R^2$** | **LISA** |
| **1** | 0.749 | 83% | 0.659 | 86% | 0.714 | 77% | 0.703 | 79% | 0.705 | 74% |
| **2** | 0.716 | 78% | 0.727 | 81% | 0.75 | 73% | 0.734 | 80% | 0.734 | 71% |
| **3** | 0.758 | 73% | 0.717 | 82% | 0.707 | 74% | 0.702 | 81% | 0.738 | 79% |
| $\overline{X}$ | 0.741 | 78% | 0.701 | 83% | 0.724 | 75% | 0.713 | 80% | 0.726 | 75% |

### *5.2. Comparison of RegRF and Benchmark Methods*

This section compares RegRF to the chosen benchmark methods (RF, GWR, GW-RF). For the comparison, the WARD results were selected, which was the optimal regionalization method according to the results of the previous section. Furthermore, WARD RegRF was also able to run when using the full dataset, which was also possible for RF and GWR (not for GW-RF). The models using the small dataset are shown in the table below (Table 3), and with a composite map of the LISA clusters of residuals in Figure 3.

**Table 3.** Performance for RegRF and Benchmark Methods (RF, GWR, GW-RF) across three folds for the small dataset. LISA column shows the percentage of residuals that were not spatially clustered.

| | RegRF | | RF | | GWR | | GW-RF | |
|---|---|---|---|---|---|---|---|---|
| **Fold** | **$R^2$** | **LISA** | **$R^2$** | **LISA** | **$R^2$** | **LISA** | **$R^2$** | **LISA** |
| **1** | 0.749 | 83% | 0.636 | 76% | 0.74 | 86% | 0.758 | 80% |
| **2** | 0.716 | 78% | 0.677 | 64% | 0.74 | 86% | 0.774 | 86% |
| **3** | 0.758 | 73% | 0.643 | 68% | 0.755 | 88% | 0.765 | 86% |
| $\overline{X}$ | 0.741 | 78% | 0.652 | 69% | 0.745 | 86% | 0.766 | 84% |

The traditional RF as the only non-spatial method underperforms regarding both the R2 values and the number of insignificant LISA clusters. The best performance regarding the R2 is obtained by GW-RF with a value of 0.766. RegRF is the third best by a close margin (0.004), being slightly worse in this regard than GWR. As these two are very close, looking at the LISA clusters can give additional insights, which shows that GWR performs better regarding this performance metric (78% vs 86%). In fact, GWR also outperforms GW-RF in this category. The comparison between RegRF and traditional methods such as RF and GWR demonstrates that RegRF strikes a valuable balance between computational efficiency and model performance, particularly when applied to large datasets. This reinforces the central hypothesis that regionalization improves the predictive power of spatial models while maintaining computational feasibility, making it a practical choice for large-scale applications.

The spatial distribution of the LISA clusters can provide valuable insights as well. The main things that can be seen there are very pronounced coldspots in the interior of the state for RF, as well as hotspots near the coast. GWR shows very few hot- and coldspots which do not show a distinct pattern in their distribution, apart from one bigger hotspot in the south of California. For GW-RF, a few hotspots can be found near the coast as well as sprinkles of coldspots in the interior of the state. RegRF also shows those coastal hotspots, and also has some coldspots in the interior of the state.

Taking a closer look at the folds can also uncover some interesting patterns: This shows that the variance for R2 across the three folds is the highest for RegRF when looking at the spatial methods. For RegRF, it is 5e-4, while for GWR and GW-RF the variance is 8e-5 and 6e-5, respectively. This means that both GWR and GW-RF do not seem to be that reliant on a "fitting" dataset, which is likely due to their integrated optimization mechanisms. In contrast to that, the optimization for RegRF was done manually which very likely leads to it not being able to account for differences in the data in the same manner. This is also underlined by the following: For GWR and GW-RF, the bandwidth value was estimated separately for each fold (82 for folds 1 and 2, and 77 for fold 3). For RegRF, all parameters were exactly the same in all three splits. So, GWR and GW-RF are better at adapting to the data at hand which is likely to lead to their results having a smaller variance and higher performance across the folds.

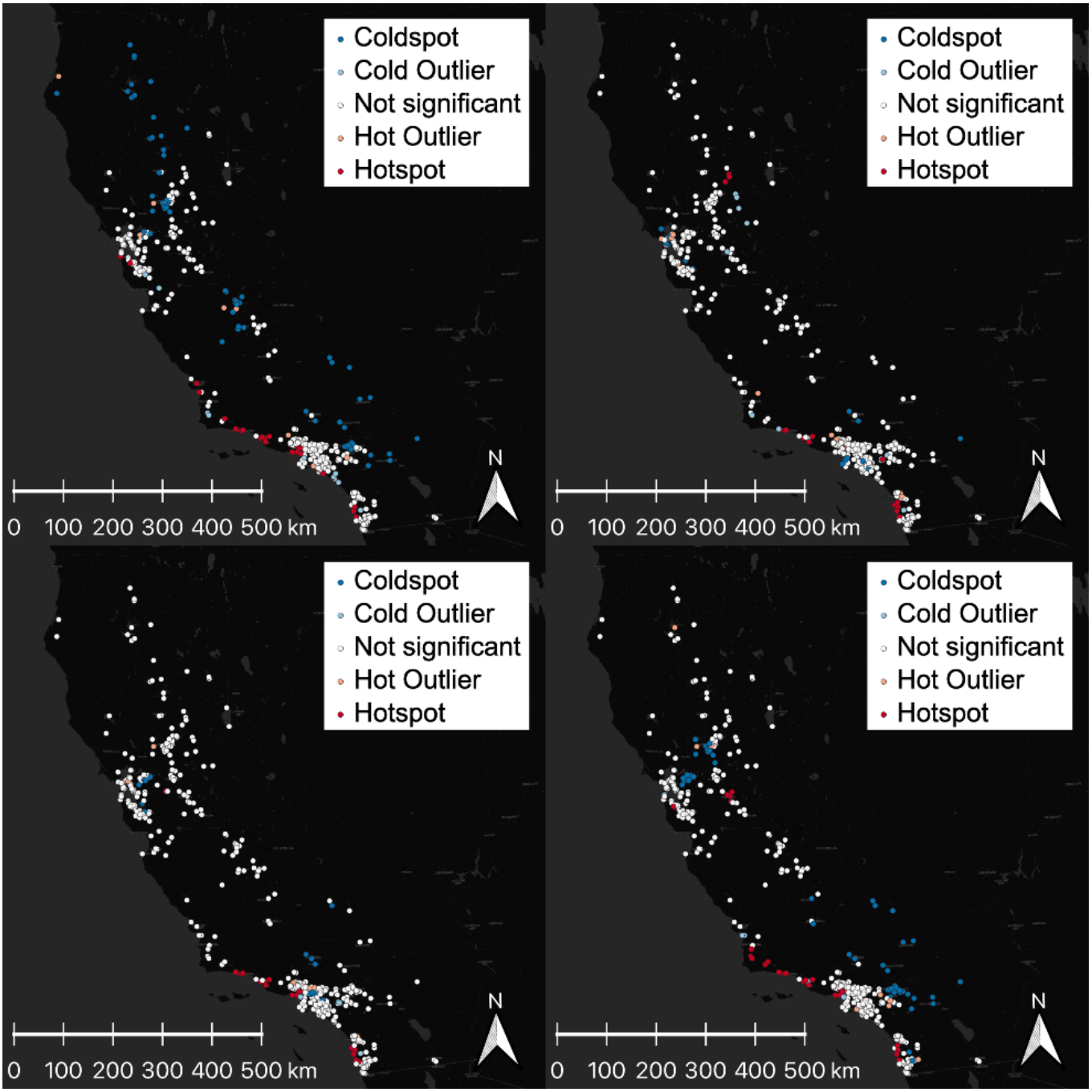


**Figure 3.** LISA clusters of residuals (Fold 3, Small Dataset): Top left: RF; Top right: GWR; Bottom left: GW-RF; Bottom right: RegRF.

For the comparison to the benchmark methods, the time the methods need to deliver results is an important factor. The variation of RegRF under investigation here takes about 13 seconds to finish, while RF has a runtime of 10 seconds. GWR needs 11 seconds to deliver results, and GW-RF finishes computation in 51 seconds. These numbers show that RegRF can significantly increase model performance while adding very little computational effort in comparison to Random Forest. GWR takes a similar amount of time, but only increases the accuracy by a small amount. GW-RF takes by far the longest, a tradeoff that has to be made if the user wants to obtain the best performing model. Additional insights can be gathered by inspecting the Table 4 that shows the results for the full dataset. This demonstrates that the selection of models is not only dependent on accuracy but also on the computational feasibility, particularly in cases where time constraints are critical.

In the case of the bigger dataset, the performance across all methods is higher, which can be attributed to the substantially increased amount of data that is available for training the models. At the same time, the differences between the respective methods are more pronounced. As mentioned before, GW-RF was not able to finish computation for this dataset, so GWR is the best performing method with an average R2 of 0.82. This is again followed by RegRF at 0.791, and RF is way behind at 0.674. This is also reflected by the number of insignificant LISA clusters, which is higher than for the small dataset at 90% for GWR, but a bit lower for RegRF at 77%. Random Forest is again outperformed significantly, where just 67% of LISA clusters are insignificant.

https://doi.org/10.48088/ejg.l.boe.15.4.244.255

The differences between the folds regarding the R2 values tend to be less pronounced with the full dataset, which makes sense due to the bigger size of the dataset, whereby the choice of which samples to use for training and testing becomes less impactful. For RegRF, it is lower than for the small dataset at 1e-4. For RF, the performance is very constant with a variance of 4e-6. The biggest differences can be found for GWR with a variance of 3e-4, which is in fact higher than for the small dataset. Not just for GWR, also for RegRF the first split underperforms in comparison to the other two splits.

The runtimes for some of these methods drastically increase when increasing the sample size. For the full dataset, Random Forest needs 41 seconds to finish, the configuration of RegRF under investigation here follows closely behind this at 45 seconds. GWR also sees a substantial increase in computational effort necessary, as it needs more than 8.5 minutes to arrive at a result. GW-RF was not able to finish computation at all. These aspects show that RegRF can be advantageous in the scenario at hand, as its runtime is only slightly higher than RF, while substantially increasing the performance of the predictive model. GWR is more accurate, but also takes more than 11 times longer to deliver a result.

**Table 4.** Performance for RegRF and Benchmark Methods (RF, GWR) across three folds, full dataset. LISA: LISA column shows the percentage of residuals that were not spatially clustered.

| | RegRF | | RF | | GWR | |
|---|---|---|---|---|---|---|
| **Fold** | **$R^2$** | **LISA** | **$R^2$** | **LISA** | **$R^2$** | **LISA** |
| **1** | 0.78 | 77% | 0.672 | 70% | 0.801 | 89% |
| **2** | 0.797 | 79% | 0.676 | 73% | 0.827 | 91% |
| **3** | 0.797 | 76% | 0.674 | 61% | 0.833 | 90% |
| $\overline{X}$ | 0.791 | 77% | 0.674 | 68% | 0.82 | 90% |

## 6. Discussion

There are significant differences regarding the performance of the predictive models, depending on which regionalization methods is used in the workflow. WARD performed the best regarding the R2 values, followed by Max-p and SKATER. The weakest performance was obtained by Kmeans and AZP. Different conclusions can be drawn when looking at the percentage of insignificant LISA clusters, where AZP performed best, followed by Kmeans and WARD. In this category, SKATER and Max-p had the lowest scores.

It should be noted that there are other factors than just the regionalization algorithms that influence the performance metrics. One of these is the number of regions the dataset is split into. In this work, this was only possible to be explored for a small dataset with the methods WARD and AZP (414 models). We detected small R2 variations across the two methods, the three folds, and 69 combinations of cluster numbers. Although this was a positive sign regarding the optimization requirements of the RegRF approach, more insights can be gained by extending the evaluation to additional regionalization methods.

Another distinguishing factor between the respective regionalization algorithms is the computational effort. There are big differences regarding this aspect, which also greatly influenced the findings of this work. The fastest one was WARD with 13 seconds, while AZP took 170 seconds. SKATER and Kmeans took ten times as long as WARD with 130 seconds for both methods. Max-p required by far the biggest computational effort with more than 40 minutes. Naturally, the dataset size and its influence on the computational demand play a significant role in determining the practical feasibility of algorithms. Higher computational requirements can limit their scalability and real-time applicability, particularly in fields where rapid processing of large datasets is crucial for timely decision-making, which limits the importance of some of the evaluated methods, namely SKATER, Kmeans, and Max-p.

One of the most important aspects of our research is how the proposed approach performs in comparison to benchmark methods. For this work, GWR, GWR-RF, and RF have been chosen. In comparison to GWR, RegRF performs worse by a close margin. Given this, the significantly lower computation time of RegRF for the full dataset gives the proposed approach an advantage for future usage and exploration. Furthermore, RF was outperformed by RegRF in every scenario and by a large margin. This means that better performing models can be created without adding a lot of computation time.

RegRF was compared to one additional method, which is the recently proposed approach Geographically Weighted Random Forest (Georganos and Kalogirou 2022). The question at hand can only be answered for the small dataset, as GW-RF was not able to finish computation in a reasonable amount of time for the full dataset, which naturally is not a desirable trait for potential users of the algorithm. Three attempts were made to run it using the full dataset, but the computation was not able to finish in seven days, despite the computer being a powerful machine learning station. So, for the full dataset, RegRF demonstrated a significant advantage. In contrast to this stand the results of the application of GW-RF with the small dataset. Here, it outperformed all the other methods regarding the R2 values by quite a big margin. Nevertheless, the computational effort needed was also the highest.

RegRF's performance is heavily influenced by the presence of spatial patterns in the dataset. In scenarios when spatial dependence is weak or absent, the benefits of regionalization are diminished, and traditional methods may perform equally or better. Furthermore, as RegRF is built upon Random Forest, it requires a sufficient amount of data to effectively learn from. In cases where datasets are very small, the model's performance may suffer, particularly in comparison to methods that incorporate adaptive mechanisms like GWR or GW-RF. Real-world applications, such as urban planning or environmental monitoring, often deal with spatially structured data, which is where RegRF demonstrates its advantages. However, for applications with sparse or less structured data, the performance gains might not be as pronounced.

The term "Spatial-Conscious Machine Learning Model" was introduced by Kiley and Bastian (2020) who showed that such models outperform non-spatial models. Their application domain was real estate, similar to our study as well as the studies by Liu et al. (2022) and Soltani et al. (2022). Other application domains include topsoil silt and clay content (Behrens et al. 2018), robberies (Deng et al. 2023), crime (Yao et al. 2020),

and height-for-age (Nduwayezu et al. 2024). All the above-mentioned studies yielded superior predictive results thus forming a promising landscape for the future of spatial prediction. Additional domains where machine learning is used for spatial prediction, such as traffic and noise pollution (Sofianopoulos et al. 2024) or the classification of urban tree species (Molnar et al., 2020), could also be benefited from the exploration and testing of recent spatially-conscious machine learning approaches.

## 7. Conclusions

The most important conclusion from this study is the vast improvement achieved by the use of RegRF over the traditional Random Forest algorithm. Significant improvements in prediction accuracy can be made with practically no increase in computation time, depending on the regionalization algorithm in use. Also, RegRF was even able to compete with the established spatial statistical method Geographically Weighted Regression in certain scenarios. The third approach RegRF was compared against, Geographically Weighted Random Forest, was not able to be matched, but it should be mentioned that the practical application of this algorithm was not possible for the large-scale dataset, for which RegRF was also able to compute in a very convenient amount of time. Additionally, as RegRF is based on Random Forest, it is possible to use it for both classification and regression tasks, which cannot be said for GWR. Also, GWR can only deal with a linear relationship between the target and the input features, which is not a restriction for RegRF and GW-RF. In scenarios where, non-linearity between the dependent and independent variables exists, RegRF provides a flexible and robust solution, making it highly applicable for real-world challenges, where data relationships are often complex. Due to this, for certain scenarios it can be argued that RegRF is a suitable and optimal choice, which ultimately depends on factors like the size of the dataset, the (non-) linearity between dependent and independent variables, the type of supervised task, as well as the levels of measurement of the input features. Possible practical applications of RegRF are manifold, as it can be used in all fields that work with spatial data. This includes but is not limited to urban planning, public health, agriculture, disaster management, transportation, real estate, or criminology. As per usual for Machine Learning tasks, the data in use will play a crucial role in the success of RegRF's application.

To make RegRF more robust and possibly improve its performance, it is useful to look at future perspectives that should be addressed. In particular, future research could focus on developing an integrated optimization mechanism similar to those employed by GWR and GW-RF, where bandwidth parameters are adjusted based on the specific fold. RegRF lacks such an adaptive mechanism, which could improve its performance further. Key parameters to optimize include the choice of regionalization method, the number of regions, the minimum number of objects per region, and which variables to consider for the regionalization process. Addressing these elements could allow RegRF to adapt more effectively to various data structures and improve its predictive accuracy. It is possible that this would significantly increase the computational effort necessary, which so far can be seen as an advantage over GWR and GW-RF, but to arrive at a reliable conclusion regarding this, a practical experiment would need to be carried out.

Another aspect that should be addressed by future research is the implementation of additional regionalization algorithms that are able to deal with large-scale datasets. In this work, only two of the five tested methods were able to produce results for the full dataset. All of them were able to run for a smaller dataset, but even there some of them took considerable amounts of time. Given the trend in the last years as the processing and analysis of large-scale datasets has become more and more important, this is an opportunity that researchers in the field of spatial analytics should not overlook. The proposed approach has shown its potential which could be boosted further by more modern regionalization algorithms. This is also underscored by the fact that the newest algorithm under evaluation here stems from the year 2012. In the light of the current advancements in the fields of artificial intelligence, data processing and applied informatics, this seems to lack behind in actuality. This also leads to the final issue that needs to be discussed, which is the importance of including spatial aspects in Machine Learning wherever possible. Traditional (non-spatial) Machine Learning algorithms have become very popular over the last years, but the spatial component still seems to not be on the mind of many users of these methods. The experiments carried out in this work have shown the vast improvements that can be achieved by including spatial aspects, even without extensive knowledge about the area of interest.

**Funding:** This research received no external funding.

**Data Availability Statement:** The datasets analyzed during the current study are available in the California Housing repository, [https://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing]

**Conflicts of Interest:** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Assunção, R. M., Neves, M. C., Câmara, G., & Da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7), 797–811. https://doi.org/10.1080/13658810600665111

Behrens, T., Schmidt, K., Viscarra Rossel, R. A., Gries, P., Scholten, T., & MacMillan, R. A. (2018). Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science*, 69(5), 757–770. https://doi.org/10.1111/ejss.12687

Box, G. E. P. (1979). Robustness in the Strategy of Scientific Model Building. In Robustness in Statistics (pp. 201–236). Elsevier. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Deng, Y., He, R., & Liu, Y. (2023). Crime risk prediction incorporating geographical spatiotemporal dependency into machine learning models. *Information Sciences*, 646, 119414. https://doi.org/10.1016/j.ins.2023.119414

Duque, J. C., Anselin, L., & Rey, S. J. (2012). The MAX-P - Regions Problem. *Journal of Regional Science*, 52(3), 397–419. https://doi.org/10.1111/j.1467-9787.2011.00743.x

Feng, X., Barcelos, G., Gaboardi, J. D., Knaap, E., Wei, R., Wolf, L. J., Zhao, Q., & Rey, S. J. (2022). spopt: A python package for solving spatialoptimization problems in PySAL. *Journal of Open Source Software*, 7(74), 3330. https://doi.org/10.21105/joss.03330

Geary, R. C. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5(3), 115. https://doi.org/10.2307/2986645
Georganos, S., & Kalogirou, S. (2022). A Forest of Forests: A Spatially Weighted and Computationally Efficient Formulation of Geographical Random Forests. *ISPRS International Journal of Geo-Information*, 11(9), 471. https://doi.org/10.3390/ijgi11090471
Grekousis, G. (2020). Spatial Analysis Methods and Practice: Describe – Explore – Explain through GIS (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108614528
Haining, R. P. (2010). The Nature of Georeferenced Data. In M. M. Fischer & A. Getis (Eds.), Handbook of Applied Spatial Analysis (pp. 197–217). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03647-7_12
Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518. https://doi.org/10.7717/peerj.5518
Kiely, T. J., & Bastian, N. D. (2019). The Spatially-Conscious Machine Learning Model (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1902.00562
Klemmer, K., Koshiyama, A., & Flennerhag, S. (2019). Augmenting correlation structures in spatial data using deep generative models (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.1905.09796
Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3
Liang, P., Qin, C.-Z., & Zhu, A.-X. (2024). Using Automated Machine Learning for Spatial Prediction—The Heshan Soil Subgroups Case Study. *Land*, 13(4), 551. https://doi.org/10.3390/land13040551
Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. 2, 5.
Liu, X., Kounadi, O., & Zurita-Milla, R. (2022). Incorporating Spatial Autocorrelation in Machine Learning Models Using Spatial Lag and Eigenvector Spatial Filtering Features. *ISPRS International Journal of Geo-Information*, 11(4), 242. https://doi.org/10.3390/ijgi11040242
Lotfata, A., Georganos, S., Kalogirou, S., & Helbich, M. (2022). Ecological Associations between Obesity Prevalence and Neighborhood Determinants Using Spatial Machine Learning in Chicago, Illinois, USA. *ISPRS International Journal of Geo-Information*, 11(11), 550. https://doi.org/10.3390/ijgi11110550
Majumder, S., Roy, S., Bose, A., & Chowdhury, I. R. (2023). Multiscale GIS based-model to assess urban social vulnerability and associated risk: Evidence from 146 urban centers of Eastern India. *Sustainable Cities and Society*, 96, 104692. https://doi.org/10.1016/j.scs.2023.104692
Molnár, V. É., Simon, E., University of Debrecen, Hungary, & Szabó, S. (2020). Species-level classification of urban trees from worldview-2 imagery in Debrecen, Hungary: An effective tool for planning a comprehensive green network to reduce dust pollution. *European Journal of Geography*, 11(2), 33–46. https://doi.org/10.48088/ejg.v.mol.11.1.33.46
Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17. https://doi.org/10.2307/2332142
Mueller, E., Sandoval, J. S. O., Mudigonda, S., & Elliott, M. (2018). A Cluster-Based Machine Learning Ensemble Approach for Geospatial Data: Estimation of Health Insurance Status in Missouri. *ISPRS International Journal of Geo-Information*, 8(1), 13. https://doi.org/10.3390/ijgi8010013
Nduwayezu, G., Kagoyire, C., Zhao, P., Eklund, L., Pilesjo, P., Bizimana, J. P., & Mansourian, A. (2024). Spatial Machine Learning for Exploring the Variability in Low Height-For-Age From Socioeconomic, Agroecological, and Climate Features in the Northern Province of Rwanda. *GeoHealth*, 8(9), e2024GH001027. https://doi.org/10.1029/2024GH001027
Nikparvar, B., & Thill, J.-C. (2021). Machine Learning of Spatial Data. *ISPRS International Journal of Geo-Information*, 10(9), 600. https://doi.org/10.3390/ijgi10090600
Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M. E., & Papritz, A. (2018). Evaluation of digital soil mapping approaches with large sets of environmental covariates. *SOIL*, 4(1), 1–22. https://doi.org/10.5194/soil-4-1-2018
Openshaw, S. (1977). A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling. *Transactions of the Institute of British Geographers*, 2(4), 459. https://doi.org/10.2307/622300
Pace, R. K., & Barry, R. (1997). Sparse spatial autoregressions. *Statistics & Probability Letters*, 33(3), 291–297. https://doi.org/10.1016/S0167-7152(96)00140-X
Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, 9(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1
Quevedo, R. P., Maciel, D. A., Uehara, T. D. T., Vojtek, M., Rennó, C. D., Pradhan, B., Vojteková, J., & Pham, Q. B. (2022). Consideration of spatial heterogeneity in landslide susceptibility mapping using geographical random forest model. *Geocarto International*, 37(25), 8190–8213. https://doi.org/10.1080/10106049.2021.1996637
Quiñones, S., Goyal, A., & Ahmed, Z. U. (2021). Geographically weighted machine learning model for untangling spatial heterogeneity of type 2 diabetes mellitus (T2D) prevalence in the USA. *Scientific Reports*, 11(1), 6955. https://doi.org/10.1038/s41598-021-85381-5
Raschka, S., & Mirjalili, V. (04). Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow (Second edition, fourth release,[fully revised and updated]). Packt Publishing.
Rogerson, P. (2021). Spatial statistical methods for geography. Sage publications.
Roy, S., Bose, A., Majumder, S., Roy Chowdhury, I., Abdo, H. G., Almohamad, H., & Abdullah Al Dughairi, A. (2023). Evaluating urban environment quality (UEQ) for Class-I Indian city: An integrated RS-GIS based exploratory spatial analysis. *Geocarto International*, 38(1), 2153932. https://doi.org/10.1080/10106049.2022.2153932
Roy, S., & Chowdhury, I. R. (2024). Intoxication in the city: Investigating spatial patterns and determinants of drugs and alcohol-related illegal activities in India's geostrategic corridor. *Applied Geography*, 171, 103386. https://doi.org/10.1016/j.apgeog.2024.103386
Santos, F., Graw, V., & Bonilla, S. (2019). A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon. *PLOS ONE*, 14(12), e0226224. https://doi.org/10.1371/journal.pone.0226224
Sofianopoulos, S., Stigas, S., Stratakos, E., Tserpes, K., Faka, A., & Chalkias, C. (2024). Citizens as Environmental Sensors: Noise Mapping and Assessment on Lemnos Island, Greece, Using VGI and Web Technologies. *European Journal of Geography*, 15(2), 106–119. https://doi.org/10.48088/ejg.s.sof.15.2.106.119
Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. J. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities*, 131, 103941. https://doi.org/10.1016/j.cities.2022.103941
Stojanova, D., Ceci, M., Appice, A., Malerba, D., & Džeroski, S. (2011). Global and Local Spatial Autocorrelation in Predictive Clustering Trees. In T. Elomaa, J. Hollmén, & H. Mannila (Eds.), *Discovery Science* (Vol. 6926, pp. 307–322). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24477-3_25

Talebi, H., Peeters, L. J. M., Otto, A., & Tolosana-Delgado, R. (2022). A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling. *Mathematical Geosciences*, 54(1), 1–22. https://doi.org/10.1007/s11004-021-09946-w

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234. https://doi.org/10.2307/143141

Wei, R., Rey, S., & Knaap, E. (2021). Efficient regionalization for spatially explicit neighborhood delineation. *International Journal of Geographical Information Science*, 35(1), 135–151. https://doi.org/10.1080/13658816.2020.1759806

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7), 557–585.

Yao, S., Wei, M., Yan, L., Wang, C., Dong, X., Liu, F., & Xiong, Y. (2020). Prediction of Crime Hotspots based on Spatial Factors of Random Forest. 2020 15th International Conference on Computer Science & Education (ICCSE), 811–815. https://doi.org/10.1109/ICCSE49874.2020.9201899

Zhang, Z., Xu, N., Liu, J., & Jones, S. (2024). Exploring spatial heterogeneity in factors associated with injury severity in speeding-related crashes: An integrated machine learning and spatial modeling approach. *Accident Analysis & Prevention*, 206, 107697. https://doi.org/10.1016/j.aap.2024.107697



European Journal of Geography