Article

A Data-Driven Approach for Urban Heat Island Predictions: Rethinking the Evaluation Metrics and Data Preprocessing

Berk Kıvılcım [ID] and Patrick Erik Bradley *[ID]

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Englerstr. 7, 76131 Karlsruhe, Germany; bkivilcim21@gmail.com
* Correspondence: bradley@kit.edu Tel.: +49-721-608-47304

Abstract: A 2D raster data representing building volumes of each grids are derived from 3D vector-format urban data for use in machine learning applications. Since the task is to explore patterns, i.e., urban heat islands, Gaussian blurring is implemented on these generated 2D raster data before the training process. This strengthens the visual capturing of spatial relationships, and as a result the correlation rate between air temperature and building volume data is also increased. After the model training, the prediction results are not simply evaluated with most widely used shallow metrics like the Mean Square Error (MSE), but thanks to the raster format of input and output results, some image similarity metrics such as Structural Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) that are able to detect and consider spatial relations are used during the evaluation and interpretation process, because of their higher usefulness in mimicking human visual judgements. The trained models with Random Forest and XGBoost methods which are capable of predicting the spatial distribution of air temperature by using building volume information are compared. By doing so, this research aims to assist urban planners in incorporating environmental parameters into their planning strategies, thereby facilitating more sustainable and inhabitable urban environments.

Keywords: building volume; air temperature; urban planning; urban morphology; machine learning for environment

1. Introduction

Environmental parameters such as air temperature are critical determinants of human quality of life and energy efficiency management. Urban areas are densely populated and also highly correlated with some of these natural phenomena through urban morphology and landscape spatial patterns. Buildings are the most active areas of human activity and have a significant impact on the urban thermal environments by altering the heat exchange [1–5]. Consequently, predicting the effects of urban plans on environmental parameters is essential for proper decision making and planning to enhance the living conditions of cities. Moreover, a significant majority of the world's population is predicted to live in urban environments in the future [6]. Additionally, temperature increases contribute to health issues with potential for heat-related illnesses and disrupt ecosystems and adversely affect biodiversity [7]. Therefore, it is essential to research the urban thermal environment containing buildings [8]. Previous studies have actually highlighted the strong correlation between urban morphology and air temperature, underscoring the importance of employing three-dimensional data in those analyses. Although it is recognised that climate issues still have limited impact on urban planning processes [9,10], it is partly

because of a gap between urban planners and climatologists [9]. To help urban planners and decision makers better understand and use the research findings, linking the climate issues to planning parameters can be more helpful than geographic or morphological parameters [2]. Therefore, in this article, a data-driven machine learning model is trained to predict urban near-surface air temperature by using building volumes to rethink about how to increase the correlations between the indicator parameters and perform better assessments for model accuracy evaluations.

The power of machine learning algorithms allows to evaluate environmental indicators on a large scale and to map urban air temperature [11–15], since those algorithms have the advantage of solving complex non-linear problems with fewer computing sources and less time [16]. Thanks to machine learning methods, we can now better model the patterns of these urban forms to refine those of future cities to meet the needs of rapid urbanization [13]. Although the results obtained when using the machine learning models might differ from the actual measurements while exhibiting a similar trend as the measurements, this makes it still reasonable and acceptable [17]. This issue also occurred in the trained models of the present work. In the end, with the aid of trained models in this study, urban planners can manipulate the building volume values as they desire for the input of the machine learning model in order to observe how their plans will impact environmental factors. It is expected that changing the spatial arrangement of urban components may affect the land surface energy distribution [18]. This could allow them to balance and control their plans based on these impacts, potentially reducing the occurrence of flawed urban planning.

Many of the previous studies focused on investigating the close relationship between temperature and buildings [8,19–25], while many others implemented machine learning models to predict land surface temperature (LST) or air temperature by considering urban morphology [6,16,17,26–28]. According to these studies, a higher building volume contributes to a warmer environment within the city centre. Therefore, optimisation of the building volume should be seriously considered, especially during the urban planning decision-making [29]. However, most of these studies utilise the average building height of regions while focusing on large non-uniform selected regional blocks or local climate zones for predictions. On the other hand, our methodology incorporates building heights and footprints directly without averaging, and also associates volumes with two-dimensional air temperature raster data, enabling predictions not just for specific selected region blocks, but on an adjustable per-pixel basis. This approach will allow for very high-resolution predictions to be made swiftly as higher-resolution meteorological data become available in the future.

In addition, by performing pixel-level predictions, it is possible to express the results in raster format and test the model's accuracy by comparing them with ground truth data in the same format. This approach allows us to not just evaluate the models with shallow metrics like Mean Square Error (MSE), but also enables the capture of spatial patterns and the ability to perform quantitative assessments similar to human visual perception using metrics such as SSIM (Structural Similarity Index) [30], and LPIPS (Learned Perceptual Image Patch Similarity) [31] which are widely adopted similarity metrics [32]. These two perceptual similarity metrics are also used in various different tools such as Nerfstudio [33] to evaluate trained models accuracy which makes them standardized metrics. Moreover, our model's process involved data from ten different cities, with seven used in the training phase and the remaining three for testing (validation). For this reason, the present model still provides a generalised and robust prediction capability.

Besides, land surface temperature (LST) draws significant attention, as it modulates the air temperature of the lower layer of the urban atmosphere [34]. However, since air temperature data over 2 m above the surface is already provided by the German Weather Service as an publicly available data. LST also being sensitive to surface emissivity and reflectivity which might fool the trained model using many different cities, the air temperature data is used instead of LST data for the present methodology.

One of the greatest challenges of urban planning today is to produce urban forms that meet the challenge of today's cities [13]. It has been mentioned in past studies that the landscape pattern of two-dimensional space alone is inadequate for explaining the complex thermal phenomena occurring in urban areas [18]. Also, the correlation between thermal phenomena and the 3D-Volume-Index was higher than the 2D-Area-Index [8]. However, the role of urban morphology, such as building height is often overlooked in many cases [27]. Since height-related indicators have been typically chosen as the major parameters to characterise the three-dimensional landscape morphology [35,36], one of the major limitations has been the difficulty of obtaining high-resolution 3D information about the scale of entire metropolitan areas [18] and accurately estimating the height of buildings on a large scale to obtain the 3D structure of buildings. This is a tough challenge [37]. In addition to that, many analyses from previous studies rely on official urban datasets provided by governments or profit-making organizations, which include building information [17]. Therefore, available and accessible 3D urban morphology data have become essential for extensive academic research on the built environment and urban climate, and a rapid methodology for extracting urban morphology information is urgently needed [38].

Voxel data is deemed suitable for volumetric calculations in order to generate a 2D raster dataset that contains building volume information for each grid. This choice was made because the capability for volumetric calculations is a key advantage of voxel models, which is absent in other model types [39]. However, since city-scale voxel data were not available for our study areas, we generated voxel models by applying a set of extremely simple steps onto publicly accessible CityGML data. CityGML is a widely used open data model based on the extensible markup language (XML), capable of describing model elements in five levels of detail [40]. This approach allowed us to rapidly obtain voxel representations, though with a lower precision. The LoD1 and LoD2 data are available for the corresponding study region. LoD1 data is selected since we are interested with overall building volumes instead of fine details of the buildings. In addition, previous studies indicate that LoD1 models provide a relatively high information content and usability compared to their geometric detail [41–43]. Unlike previous complex and comprehensive studies that have extensively focused on the voxelization issues of CityGML or CityJSON data, the method used in this study was intentionally simplified to accelerate our experimental processes, as we were not concerned with fine building details while focusing on the city-scale since the voxel models were only used to generate 2D raster data.

Some of the comprehensive CityGML-to-voxel conversion algorithms introduced in previous studies [44–47] implement only simulations on the scale of individual buildings instead of a whole city scale [39,48,49]. Cf. also [50–52] for extraction methods of watertight volumetric models from wireframe data like CityGML, and their robustness. Those methodologies are based on geometric intersection procedures in 3D spaces, since we only used intersections in 2D spaces to swiftly create city-scale data for multiple urban regions. The other conventional methodologies for obtaining the voxel data makes the need for a large amount of laser scanning, and other sensor data are necessary [53].

Furthermore, the indicator parameters are not just sufficient for a reliable model training since temperature is also affected by the surroundings on a larger spatial scale [54]. Some of the previous studies shows that implementing some convolutional kernels especially the Gaussian Kernel reduced the noise in the raster values [55,56]. In this study, before the training process a Gaussian blurring algorithm is implemented on 2D raster building volume data which increases the correlation rate with air temperature. A higher correlation tends to yield a more accurate prediction [57].

Some of the contributions of this study to the literature includes: (1) The implementation of Gaussian blurring on building volume data increases the correlation between air temperature and building volume across all study regions, (2) It is observed that shallow metrics, such as the Mean Square Error (MSE), which do not account for spatial relationships, may be misleading when evaluating models for urban heat island predictions. Instead, metrics such as the Learned Perceptual Image Patch Similarity (LPIPS) and the Structural Similarity Index Measure (SSIM) provide more valuable insights by incorporating spatial dependencies.

Consequently, this study might help future studies for forecasting various other natural phenomena in the future by enhancing the insights about data processing steps and result evaluations. The findings of this study are intended to provide a foundational framework for future research, in particular the ongoing research project Distributed Simulation of Processes in Buildings and City Models, funded by the German Research Foundation (DFG), where they can provide a basis for testing mathematical simulation models.

2. Methodology

This study exclusively utilized open-access data and open-source software tools. The employed datasets encompass CityGML data pertaining to the Thuringia state in Germany [58], coupled with hourly air temperature measurements provided by the German Weather Service. The temperature datasets present air temperature at a height of 2 m above ground level and feature a 1 km spatial resolution [59]. The air temperature data can be accessed in [60]. Data processing procedures were mainly conducted using Python 3.10. Additionally, Paraview [61] was used for some visuzalitazion tasks, while QGIS [62] and several of its plugins such as CityJSON plugin [63] and GDAL rasterize tool [64], is needed for data preparation steps to create 2D building volume raster data. In addition, another open-source tool named 'citygml-tools' [65] was used for converting CityGML data into the JSON format.

2.1. Study Area

In this study, ten cities from the state of Thuringia in Germany, varying in size and population density, were selected for the analysis. Seven were used for training and three for testing. The dimensions of the selected areas for these cities are as follows: Erfurt (10 km × 8 km), Jena (8 km × 10 km), Weimar (10 km × 8 km), Suhl (8 km × 10 km), Altenburg (10 km × 8 km), Sondershausen (12 km × 6 km), Gotha (10 km × 8 km), Sonneberg (8 km × 10 km), Schmalkalden (10 km × 8 km), and Gera (8 km × 10 km). When associating building volume data with air temperature measurements, the air temperature datasets used were those recorded in the same year that the CityGML datasets were created for each respective city. This approach was implemented to minimize inconsistencies arising from temporal resolution discrepancies. To present the results more clearly and to better highlight the investigations of this study, instead of using multi-temporal datasets, we utilised air temperature data by averaging the values obtained specifically at 01:00 a.m. during the month of July. The rationale for selecting this particular time is based on previous studies, which demonstrated that the correlation between air temperature and

urban morphology reaches its maximum at 01:00 a.m. during the summer [2]. This high correlation has made it possible to present and interpret the results in a clearer and more comprehensible manner.

The coordinate system of the CityGML data used in this research is EPSG:25832, while the air temperature data possesses latitude and longitude coordinates under the EPSG:4326 system, in addition to X and Y coordinates under the EPSG:3034 system.

The air temperature data, covering all of Germany, was cropped using the upper right and lower left coordinates of the CityGML data to align two datasets accurately. The consistency of data alignment following the cropping process was assessed by converting the air temperature data into raster format from netCDF format and subsequently loading it into QGIS. The alignment was checked through qualitative comparisons between CityGML region and cropped air temperature region within a shared coordinate system in QGIS to confirm consistency. The visual representation of these comparisons is given in Figure 1.
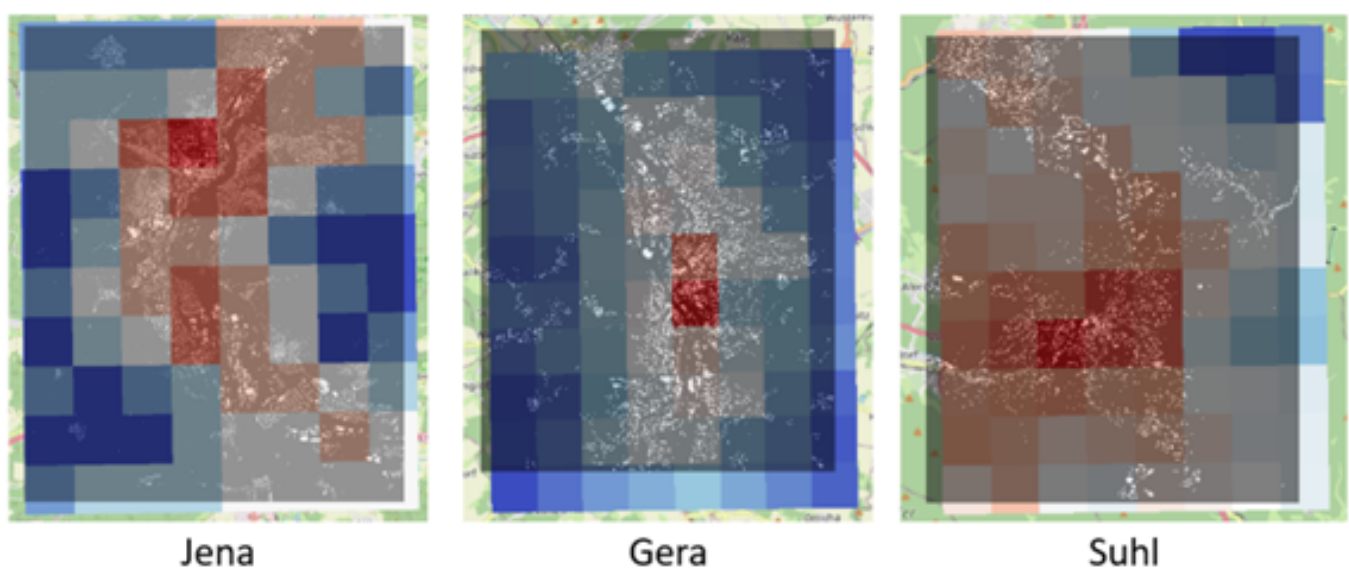

Figure 1. The overlapped visualization of air temperature raster region and rasterized CityGML region for the cities of Jena, Gera, and Suhl. In the air temperature data, red regions indicate areas with high heat stress, while blue regions indicate areas with low heat stress. The white regions in the rasterized CityGML data represent building footprints. Areas with a high concentration of buildings tend to exhibit higher levels of heat stress.

2.2. Creating 2D Building Volume Data in Raster

Traditional techniques for converting CityGML data into voxels operate by calculating intersections between CityGML and potentially billions of grid points for high-resolution and extensive areas. Although existing methods can model the many details of buildings and produce complex building voxels, it requires substantial computational power and time. In this study, these complex methods are avoided in order to be able to conduct our experiments swiftly and to have a simplified method that focuses solely on regions with buildings within a two-dimensional plane, thereby assigning a single height value to each building, thus enabling the rapid construction of less detailed buildings. Consequently, the aim of this process is to derive a 2D raster building footprint data. A brief workflow illustration is given in Figure 2.
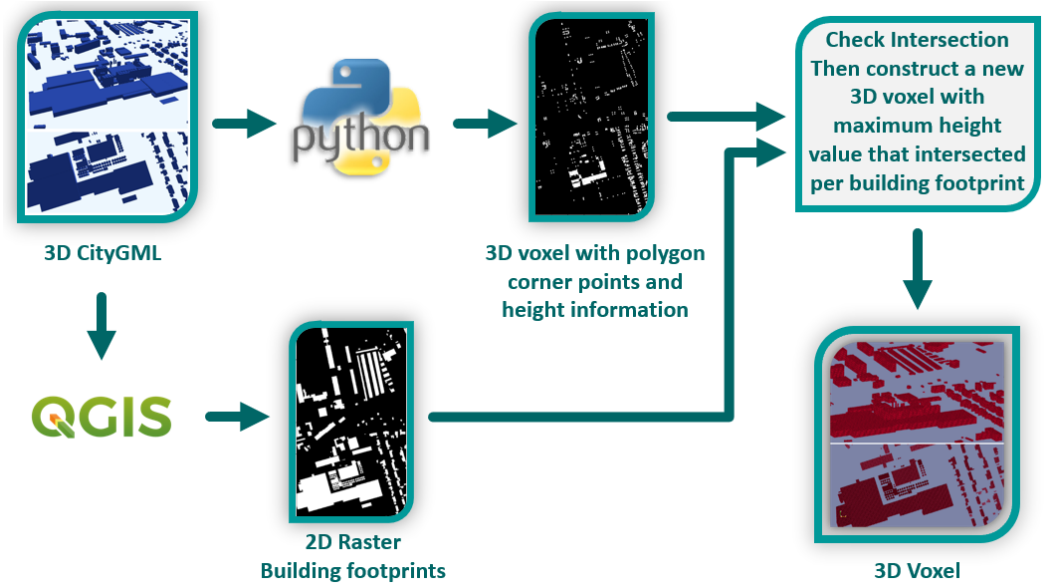
Figure 2. Workflow of the voxelisation process.

2.2.1. Retrieving 2D Building Footprint Areas

Initially, the CityGML data, downloaded via [58], covered an area of 2 km × 2 km. Therefore, these data were merged to create a single comprehensive CityGML file for each city. This resulting CityGML file was then converted into the CityJSON format and imported into QGIS using the "CityJSON Loader" plugin. Subsequently, GDAL's rasterisation tool was employed to produce raster data for any desired region at any specified resolution. In this study, the raster resolution was set at 1 m, and the necessary boundary regions for the raster image were extracted from the CityJSON data. The reason behind setting the voxel resolution to 1 m is to maintain the highest possible resolution. Since the procedures of data processing implemented in this study were deliberately simplified, the primary constraint was not computational speed but rather memory consumption. As the resolution increases, the storage size of voxel-based data grows exponentially. For example, with a 1 m resolution, the voxel representation of the city of Jena alone resulted in a data size of approximately 14 GB. Resolutions higher than 1 m began to exceed the available Random Access Memory (RAM) capacity of our computing environment. Therefore, a resolution of 1 m was selected as the highest feasible value that balances spatial precision with hardware limitations. Additional parameters selected during the use of the tool included: "A fixed value to burn: 1", "Assign a specified no data value to output bands: −999", "Output data type: Int16", "Pre-initialize the output image with value: 0". The output of this process is extracted in the GeoTiff format in the coordinate system of EPSG:25832.

2.2.2. Real-World Coordinate System to Voxel Coordinate System

This step involves calculating which location indices in our voxel system correspond to each building's polygon vertices that possess EPSG:25832 coordinate data. In addition to the horizontal plane coordinates of these corner points, the data also includes building height information. Consequently, this allows us to utilise the heights of buildings to combine them with 2D building footprints from the raster image.

For instance, when for the region encompassing the city of Erfurt with an area of 10 km × 8 km, the voxel resolution is selected as 1 m, the number of voxels in the horizontal plane should be 10,000 × 8000. Consequently, the indices of the voxel array range from 0 to 9999 for width and from 0 to 7999 for height in the Python indexing. Considering all these factors, a normalisation method was employed to transform data from the EPSG:25832 coordinate system to the local voxel coordinate system. The formulas used for the X and

Y axes are provided below, cf. Equations (1) and (2). The reason for employing different formulas for the X and Y axes is the orientation of arrays in Python, where the origin (0,0) index is at the top-left corner, whereas the real-world coordinate system of the study area places the origin at the bottom-left. This discrepancy causes a flip along the Y-axis, leading to inconsistencies. The formula used for the Y-axis adjusts this issue. Furthermore, after the normalisation process in the formula, the resulting values between 0 and 1 are multiplied by the width or height values using *width* $-$ 1 or *height* $-$ 1. This minus 1 subtraction adjustment is made because Python indexing starts at 0. In the end, as voxels constitute discrete grids, the new coordinates derived from the formula must be integers. Therefore, a rounding operation is applied to the computed values to ensure they conform to this requirement.

$$X_{\text{voxel}} = \text{round}\left(\frac{(X_i - X_{\min})}{(X_{\max} - X_{\min})} \cdot (width - 1)\right) \tag{1}$$

$$Y_{\text{voxel}} = \text{round}\left(\frac{(Y_{\max} - Y_i)}{(Y_{\max} - Y_{\min})} \cdot (height - 1)\right) \tag{2}$$

2.2.3. Assigning the Height Information to Building Footprints

In the final stage of the voxelisation process, a voxel is generated based on the normalised building polygon vertex coordinates positioned within the voxel grid and 2D building footprint raster image. If these corner points on the voxels align with a building depicted in a two-dimensional raster image, then the height of the building is derived from the highest height value among the matching vertex coordinates. This method produces buildings in the voxel space that are highly detailed and accurate in terms of their footprint, yet adopt a simplified approach for height representation by assigning a single height value per building. A result of this voxelisation process is demonstrated in Figure 3.
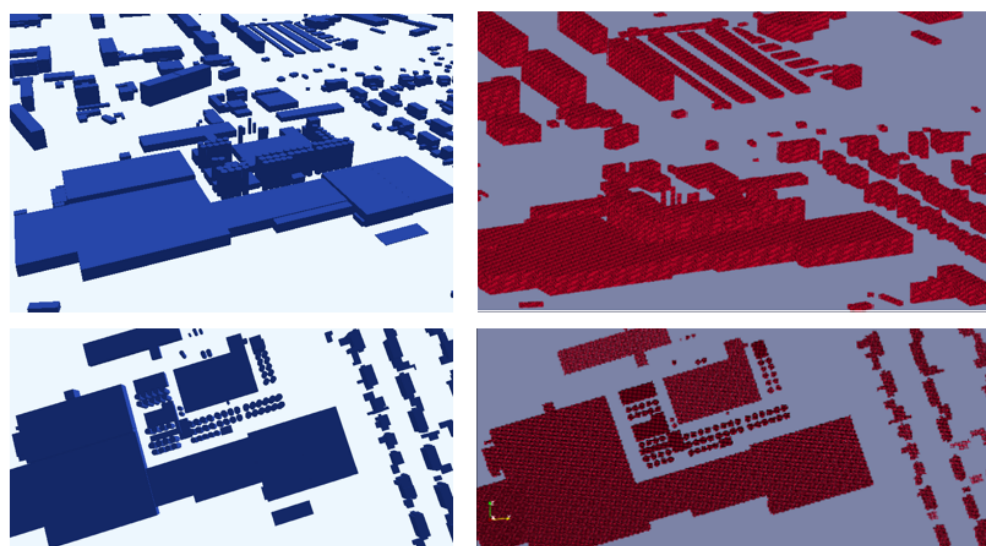

Figure 3. The (**left column**) represents the CityGML visualization of Gotha from different viewing angles, the (**right column**) represents the voxel visualisation of the same region of Gotha.

2.3. Machine Learning Training

In the machine learning model, the phenomenon targeted for prediction is air temperature. Therefore, air temperature data with a resolution of 1 km $\times$ 1 km has been employed as ground truth data in the training phase. Given that the resolution of these air temperatures is 1 km $\times$ 1 km, the area covered by the voxels is divided into a grid commensurable with the voxel numbers. For example, the region with size of 10 km $\times$ 8 km is divided into

a 10 × 8 grid. Afterwards, new two-dimensional building volume data, containing the total building volume for each grid are generated. Consequently, the adjustable resolution of the voxels allows to associate them with higher resolution air temperature data, if available. Thereby our methodology also ensures that the voxelised methods can adapt to improved or different meteorological data resolutions. After the rasterization steps, the other implemented steps before the machine learning training is presented in Figure 4.
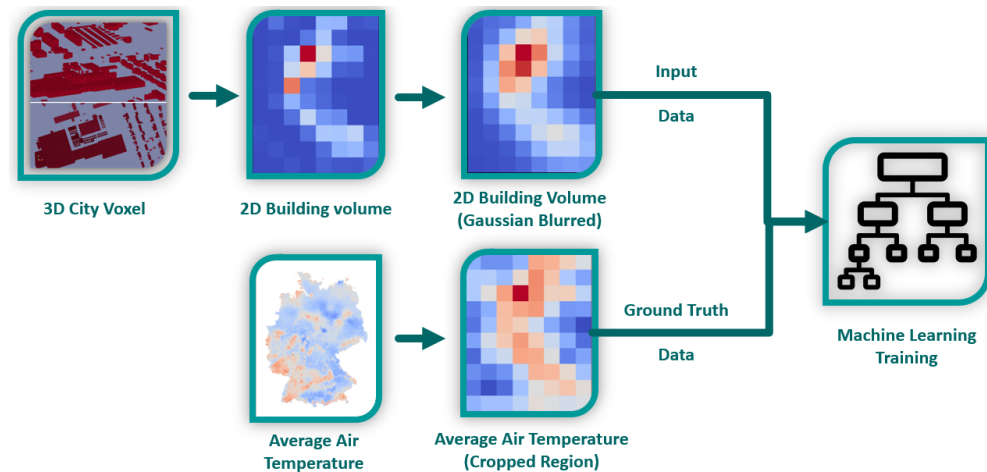

Figure 4. A brief workflow demonstration of data pre-processing steps for the machine learning training. In the air temperature data, red regions indicate areas with high heat stress, while blue regions indicate areas with low heat stress. In the building volume data, red regions indicates high amount of building volume, while blue regions indicate low amount of volume.

Incorporating spatial neighbourhood characteristics during the training phase is crucial, especially when considering urban environments with varying spatial patterns since even if a region has a high building volume but is isolated without many surrounding buildings, it may exhibit a lower urban heat island effect than expected. Similarly, if a region has relatively low building volume but with a high building volume in surrounding regions, the heat might be trapped, resulting in air temperature values higher than expected. To account for this, a Gaussian blurring method has been applied to the 2D building volume input data. This approach allows the representation of a region's building volume to incorporate contributions from its neighboring areas. The Gaussian blur acts as a kernel-based smoothing technique, where the central value of the kernel receives the highest weight, thus preserving data characteristics while simultaneously integrating the influence of adjacent regions. This approach has demonstrated an increase in the correlation between the building volume in each city and their air temperature. The selected Gaussian kernel parameter is; sigma value = 0.85. To determine the optimal sigma parameter for Gaussian blurring, initial experiments were conducted using sigma values of 0.5, 1.0, 1.5, and 2.0. In each of these experiments, the application of Gaussian blurring showed a positive contribution to correlation improvement. However, for sigma = 2.0, correlation enhancement was observed in only 8 out of the 10 cities, unlike the other values which yielded improvements in all cities. Among the tested values, the best performance was achieved with sigma values of 1.0 and 1.5. Furthermore, a comparative analysis of the results obtained with sigma values of 1.0 and 1.5 revealed negligible differences in correlation performance. In order to achieve the most effective outcome with minimal manipulation of the original data, a sigma value of 0.85, which is in between 0.5 and 1.0, was subsequently tested. This value also resulted in high correlation scores, comparable to those obtained with sigma values of 1.0 and 1.5. Based on these findings, sigma value = 0.85 was selected as the optimal value for further analyses. Furthermore, this

correlation tends to rise in cities with higher population densities, indicating a significant interplay between urban morphology and thermal behavior. The amount of correlation of original data and Gaussian blurred data with air temperature is given in Table 1, while the visualisations of those datasets are presented in Table 2, which clearly shows the effect of the Gaussian blur.

The Random Forest (RF) and Extreme Gradient Boosting (XGBoost) techniques are selected with hyper-parameter optimisation conducted via a trial and error method for machine learning training. The reason for using the more primitive trial-and-error method instead of systematic hyper-parameter optimization approaches such as Bayesian optimization or grid search was to prioritize not only achieving the best results in quantitative metrics but also to qualitatively assess the models. The goal was to identify models that provided the most generic outcomes based on human visual judgment. Moreover, as evidenced in the conclusion of this study, we demonstrated that when evaluating the trained models using image similarity metrics designed to mimic human visual judgment, the metric results do not always align with those obtained from traditional evaluation metrics such as the Mean Square Error, which also supports the idea of using trial-and-error in this study. A previous study demonstrated the effectiveness of XGBoost compared to other techniques for predicting urban heat island effects [66]. For the RF model, the hyper-parameters were established as follows: number of trees = 100,000, maximum depth of trees = 3, minimum number of samples required to split a node = 4, minimum number of samples per leaf = 2, and max_features set to 'sqrt'. For the XGBoost model, the parameters were set to: number of trees = 300,000, maximum depth of trees = 3, and learning rate = 0.000003. Additionally, to mitigate overfitting, augmented data were utilized during training, with parameters specified as number of samples = 100 and noise level = 0.01.

Table 1. The table represents the correlation values of the original 2D building volume data and the Gaussian blurred data with the air temperature

Datasets	Correlation with Building Volume and Air Temperature	Correlation with Gaussian Blurred Building Volume and Air Temperature
Altenburg	0.78	0.93
Erfurt	0.85	0.90
Gera	0.76	0.87
Gotha	0.71	0.83
Jena	0.73	0.82
Schmalkalden	0.48	0.65
Sondershausen	0.62	0.71
Sonneberg	0.53	0.74
Suhl	0.52	0.66
Weimar	0.67	0.77

Table 2. This figure illustrates the spatial distribution of building volumes (expressed in units of m^3) and a monthly average air temperature of July at 01:00 a.m. (expressed in units of °C) in selected urban areas to show the effect of Gaussian blur, and the relation between urban morphology and air temperature.
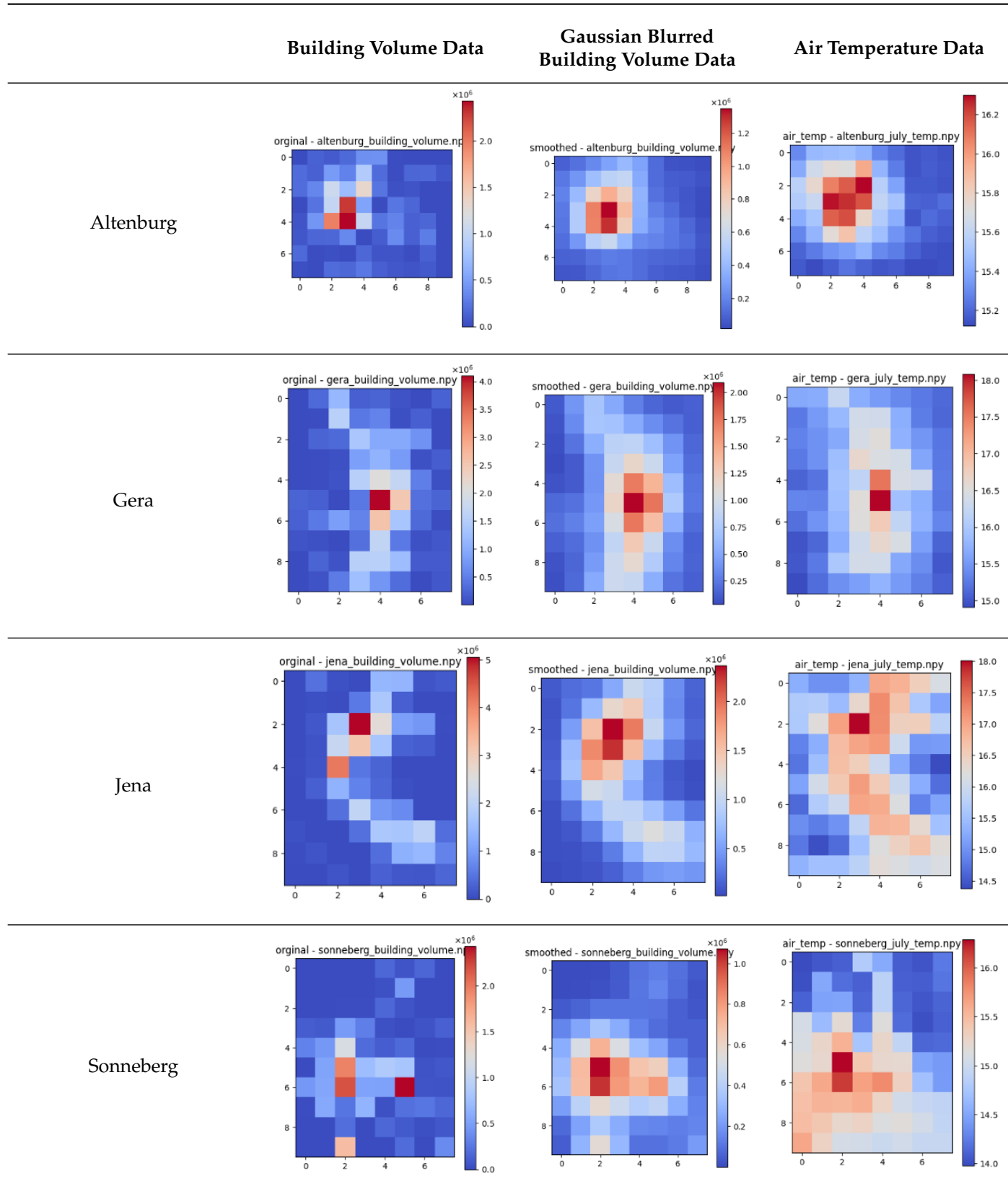
3. Results

Since the accuracy of the trained model and the corresponding parameter selections were determined through a trial-and-error approach, multiple training processes were repeated. The resulting training outcomes were analysed both qualitatively and quantitatively. However, in this section, the model that qualitatively provided the most accurate results and best represented the spatial patterns of the urban heat island effect is presented.

For this reason, rather than focusing solely on achieving a lower MSE (Mean Square Error), visual results were used as the primary basis for accuracy. For instance, in evaluations using deeper trees with both the XGBoost and Random Forest methods, MSE values as low as $0.20\,^{\circ}C^2$ were achieved for XGBoost, while values around $0.45\,^{\circ}C^2$ were observed for Random Forest during our experiments. However, upon reviewing the qualitative results, it became evident that these seemingly satisfactory quantitative outcomes were the result of overfitting. This was especially apparent in the visual outcomes, where the urban heat island patterns of cities were not predicted in a consistent and semantically correct manner, failing to capture the expected spatial patterns. On the other hand, some experiments provide appropriate and consistent visual patterns, despite having higher quantitative error values such as $0.92\,^{\circ}C^2$ MSE for the Random Forest and $0.84\,^{\circ}C^2$ for the XGBoost techniques.

In addition to the visual analysis of the predictions, the differences between the predicted results and ground truth data were also examined. Upon investigation, it was observed that the Random Forest model's error behavior did not follow any discernible spatial patterns. Instead, the error appeared to occur randomly across the spatial domain, suggesting that the model's inaccuracies were distributed without any systematic bias or spatial structure. On the other hand, when examining the error distribution of the model trained with XGBoost, it was observed that the highest error levels were concentrated in regions corresponding to urban areas. This phenomena can be seen in the following figures of this section.

The Random Forest gives better results in SSIM and LPIPS metrics than in metrics like MSE, as shown in Table 3, cf. also the next paragraph. The comparison between the prediction results and ground truth for the test dataset (Erfurt, Suhl, Sonneberg) obtained using the Random Forest model is presented in Table 4. Although the spatial patterns are well predicted and presented in that Table 4, some deviations in air temperature predictions were observed especially for the cities of Schmalkalden (in training dataset) and Suhl (in test dataset). For instance, the air temperature predictions for Schmalkalden ranged between $14.37\,^{\circ}C$ and $15.64\,^{\circ}C$, whereas the ground truth data showed a range of $15.14\,^{\circ}C$ to $16.84\,^{\circ}C$. Similarly, for Suhl, the predictions fell between $14.36\,^{\circ}C$ and $16.02\,^{\circ}C$, while the actual measurements ranged from $11.37\,^{\circ}C$ to $14.84\,^{\circ}C$. These discrepancies directly contribute to higher MSE values, because MSE values do not consider the spatial relations of data but directly focus on the prediction values and their differences with ground truth values. If the prediction interval differs a lot from the ground truth interval, this gives reason for high MSE values. However, for other cities, significant differences in air temperature predictions were not observed. For example, the prediction range for Altenburg was $14.72\,^{\circ}C$ to $16.02\,^{\circ}C$, compared to the ground truth range of $15.12\,^{\circ}C$ to $16.3\,^{\circ}C$. For Erfurt, the prediction range was $14.33\,^{\circ}C$ to $17.61\,^{\circ}C$, while the ground truth range was $14.01\,^{\circ}C$ to $17.05\,^{\circ}C$. Similarly, Gera's prediction interval was $14.79\,^{\circ}C$ to $17.59\,^{\circ}C$, with a ground truth interval of $14.90\,^{\circ}C$ to $18.08\,^{\circ}C$; Gotha's prediction interval was $14.33\,^{\circ}C$ to $16.03\,^{\circ}C$, compared to the ground truth of $13.19\,^{\circ}C$ to $14.88\,^{\circ}C$; Jena's prediction interval was $14.72\,^{\circ}C$ to $17.61\,^{\circ}C$, with the ground truth range at $14.37\,^{\circ}C$ to $18.01\,^{\circ}C$; Sondershausen's prediction interval was $14.33\,^{\circ}C$ to $15.64\,^{\circ}C$, while the ground truth ranged from $13.94\,^{\circ}C$ to $16.42\,^{\circ}C$; Sonneberg's prediction interval was $14.33\,^{\circ}C$ to $16.02\,^{\circ}C$, compared to the ground truth

interval of 13.98 °C to 16.29 °C; and finally, Weimar's prediction interval was 14.58 °C to 16.49 °C, while the ground truth ranged from 12.94 °C to 16.29 °C.

The similar comparison between the predictions obtained using the XGBoost model and ground truth data for the test cities (Sondershausen, Schmalkalden, Erfurt) is presented in Table 5. The air temperature prediction ranges for each city using the model trained with the XGBoost method are as follows: Altenburg: 14.58 °C to 15.43 °C, Erfurt: 14.44 °C to 16.08 °C, Gera: 14.72 °C to 16.08 °C, Gotha: 14.44 °C to 15.60 °C, Jena: 14.68 °C to 16.08 °C, Schmalkalden: 14.01 °C to 15.10 °C, Sondershausen: 14.46 °C to 15.10 °C, Sonneberg: 14.01 °C to 15.43 °C, Suhl: 14.01 °C to 15.43 °C, and Weimar: 14.58 °C to 16.08 °C. The ground truth air temperature intervals for comparison have been provided in the previous paragraph. Unlike the range distribution observed in the Random Forest method, these prediction ranges differ less from the ground truth data. This is expected to result in lower MSE values compared to those obtained with the Random Forest method.

Considering the experiments conducted and the quantitative and qualitative results obtained, it has been demonstrated that metrics such as MSE do not fully reflect the accuracy of the model. It was also observed that hyper-parameter selection plays a critical and direct role in model performance. Additionally, the Random Forest technique was shown to accurately predict overall spatial patterns even on unseen data (test data) during training.

In addition to the MSE comparisons, the similarities between the predicted and ground truth images were analyzed using SSIM (Structural Similarity Index), and LPIPS (Learned Perceptual Image Patch Similarity) metrics. The SSIM value of 1 indicates two images are exactly the same while 0 means no similarity. LPIPS is specifically designed to evaluate the similarities like human visual perception through deep learning techniques to overcome the shallowness of SSIM [31]. If a LPIPS value is closer to 0, that means two images are very similar to each other. The kernel size for the SSIM calculation was set to 5 × 5. For the LPIPS computation, all prediction and ground truth data were resized to 64 × 64, and the AlexNet architecture [67] was used as the multi-layer perceptron model within the deep neural network. That means both of these methods uses kernels to detect the spatial relationships during the evaluations. The results of these metrics are given in the Table 3. While the model trained using the XGBoost method outperformed the Random Forest model in terms of the MSE metric, both qualitative assessments and other quantitative metrics like SSIM and LPIPS indicated that the Random Forest-based method outperformed the XGBoost model across all cities.

Table 3. The table represents SSIM and LPIPS values for Random Forest and XGBoost models across different cities.

Dataset	Random Forest		XGBoost	
	SSIM	LPIPS	SSIM	LPIPS
Altenburg	0.89	0.0000059	0.74	0.0000183
Erfurt	0.82	0.0000504	0.71	0.0002
Gera	0.88	0.0000377	0.72	0.0001
Gotha	0.80	0.0000479	0.77	0.0000430
Jena	0.77	0.00014	0.61	0.0003
Schmalkalden	0.70	0.0000479	0.47	0.0001
Sondershausen	0.65	0.0001	0.45	0.0002
Sonneberg	0.65	0.0001	0.46	0.0001
Suhl	0.62	0.0001	0.46	0.0002
Weimar	0.80	0.0000924	0.71	0.0001

Table 4. Comparison of the air temperature prediction results obtained from the model trained with Random Forest technique and ground truth data for the test dataset which is not included during the training process. The difference map between prediction and ground truth images is given in the bottom row.

	Ground Truth Air Temperature Map	Predicted Air Temperature Map	Difference Between Ground Truth and Predictions
Sonneberg			
Erfurt			
Suhl			

Table 5. Comparison of the air temperature prediction results obtained from the model trained with XGBoost technique and ground truth data for the test dataset which is not included during the training process. The difference map between prediction and ground truth images is given in the bottom row.
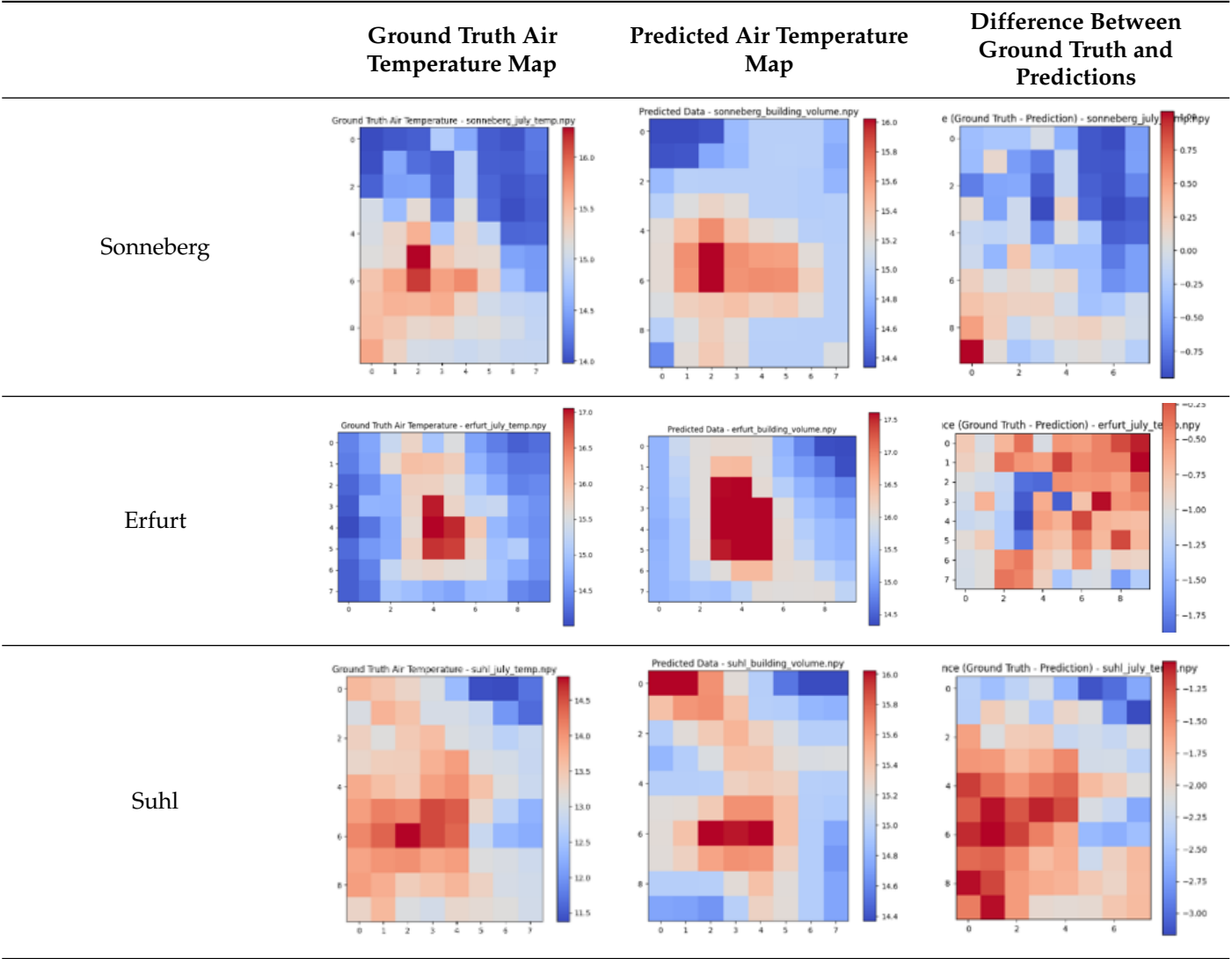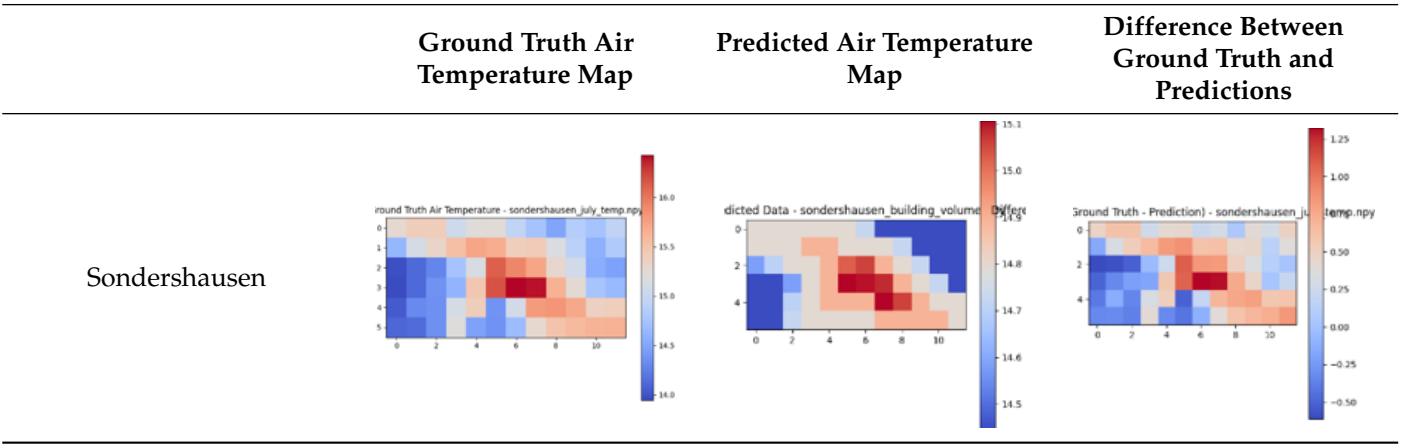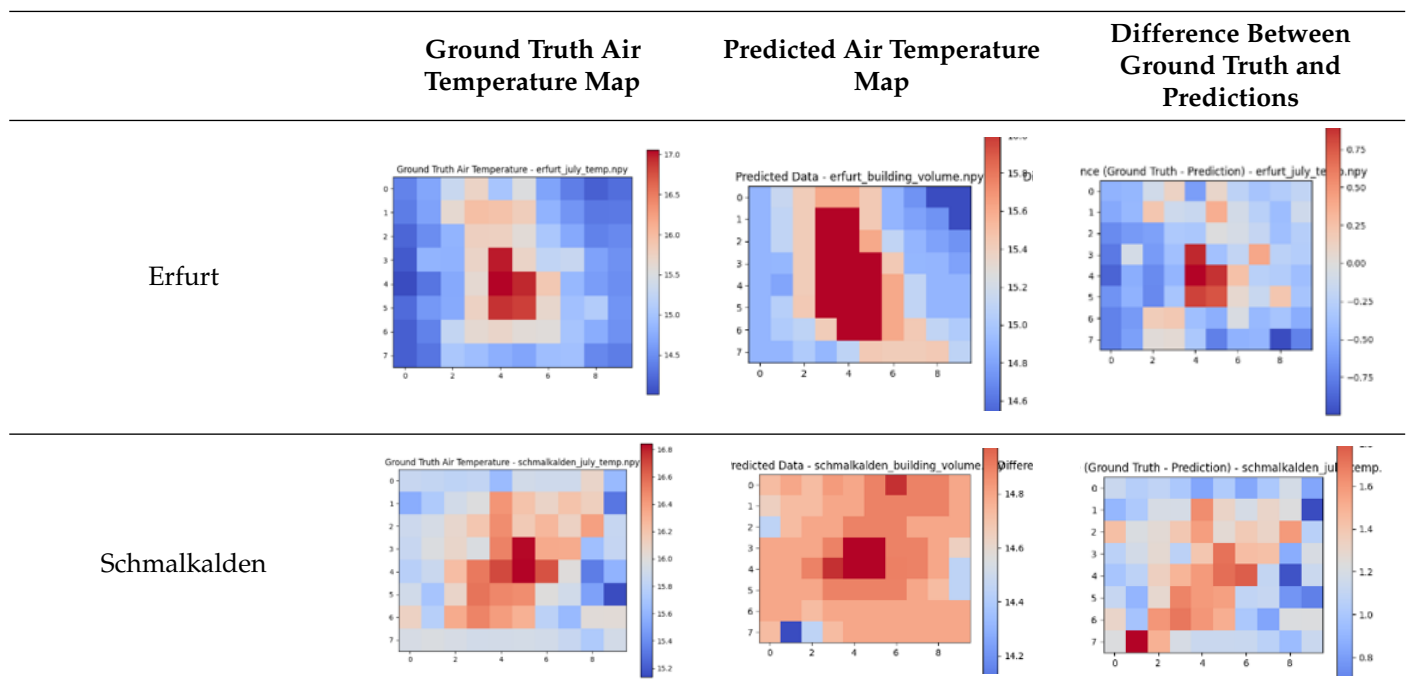
	Ground Truth Air Temperature Map	Predicted Air Temperature Map	Difference Between Ground Truth and Predictions
Sondershausen			

Table 5. *Cont.*

	Ground Truth Air Temperature Map	Predicted Air Temperature Map	Difference Between Ground Truth and Predictions
Erfurt			
Schmalkalden			

4. Discussion

Although the voxelization methodology presented in this study is designed to yield quick results, it is previously mentioned that the results compromise the level of detail of buildings. Since only a single height value is assigned per building, this approach can present challenges with more complex building structures. In such cases, the voxelization process may lack sufficient precision, leading to some degree of compromise in accurately capturing the architectural complexity of certain buildings.

While generating voxels, the intersection between 2D footprints of the buildings and the building polygon corner points in the voxel coordinate system is used. To increase the possibility of intersection, instead of treating these building polygon corner points as single pixels, 3×3 patches are used. It is important to note that the choice of the patch size is related to the resolution being used. For instance, at lower voxel resolutions, the use of patches may not be necessary, whereas at higher resolutions, larger patches might be required to ensure intersection. On the other hand, using larger patches reduces the precision in cases where buildings are densely situated and have varying heights. However, since the voxel resolution is selected as 1 m in this study, the use of 3×3 meter patches does not significantly cause the issues in this context. Moreover, buildings that are closely positioned in many urban areas often share similar height values which minimize the potential impact on precision. Nevertheless, even working with 1 m resolution voxels and using 3×3 patches, some buildings were not generated due to the failure in achieving intersections. For example, 24,990 out of 26,642 buildings were successfully matched and generated in Erfurt, 19,154 out of 20,180 in Jena, 17,313 out of 18,353 in Weimar, 17,551 out of 18,626 in Suhl, 11,526 out of 13,001 in Altenburg, 7803 out of 8454 in Sondershausen, 12,782 out of 13,620 in Gotha, 12,002 out of 12,814 in Sonneberg, 9847 out of 10,670 in Schmalkalden, 19,693 out of 22,255 in Gera.

To minimize regional biases, cities with varying population and population densities from the state of Thuringia were deliberately selected. For instance, Altenburg represents a city with a relatively low population (approximately 32.000) but a moderate population density of 708.6 inhabitants per square kilometer. In contrast, Sondershausen has the

lowest population density in the region (108.2 inhabitants per square kilometer), while Jena exhibits the highest population density of 968.1 inhabitants per km^2). Additionally, Erfurt was included as the most populous city in Thuringia with approximately 212.000 inhabitants. A range of other cities with differing levels of population and population density were also selected to ensure representative sample and generalized predictions of the region. These population and population density values pertain to the year 2017.

Since the correlation between air temperature and building volume data is not perfect, it is not expected to achieve entirely accurate prediction results. However, given that cities with high population densities tend to exhibit stronger correlations with air temperature data, focusing the training process only by using cities such as Tokyo, New York, Istanbul or other cities with a high population density may lead to more consistent model training. This approach could improve the performance of models designed for future analyses of metropolitan urban environments. Also, it is possible to obtain significantly different model outcomes even when using the same hyper-parameters. Therefore, it remains feasible to train models that outperform or underperform the ones presented in this study by utilizing the same hyper-parameters through trial-and-error processes across multiple trainings.

Besides, the methodology presented in this study not just solely focuses on the quantitative accuracy of predictions, but also observed the spatial patterns. This revealed that low MSE values alone are not sufficient to understand model accuracy. While metrics such as the Mean Square Error (MSE) directly focus on the differences between predicted and ground truth data, it does not have the ability to assess the patterns and relational structures within the data. A model can achieve a low MSE value while still suffering from overfitting. In fact, during this study, even though MSE values as low as 0.20 were obtained, it was observed that such values could result from overfitting rather than true generalization. Therefore, only models that provided the most generalized outcomes were considered and presented in this study. This highlights that low MSE values alone are not sufficient to evaluate the true accuracy and robustness of a model. Therefore, the use of additional metrics (SSIM, LPIPS) for comparing prediction and ground truth images after post-training are useful. Unlike MSE, these metrics incorporate kernel-based approaches that assess the relationship of each data point with its surrounding context. This enables an evaluation of whether the model correctly captures and reproduces general patterns within the data. Hence, these metrics not only provide insights into the numerical closeness to the ground truth, but also offer a deeper understanding of the model's consistency and generalization ability. Among the other metrics, LPIPS stands out as one of the most state-of-the-art but there are still some drawbacks of this metric. A previous study indicates that LPIPS is susceptible to such imperceptible adversarial perturbations where the LPIPS values are significantly affected by adding some noise or manipulating just a single pixel [32].

In this study, XGBoost and Random Forest (RF) machine learning methods were primarily employed. XGBoost was selected based on prior studies suggesting that it provides one of the most accurate predictions for Urban Heat Island (UHI) modeling. Random Forest, on the other hand, was included as it is one of the most well-established and foundational models in machine learning. Upon applying our experiments to these two models, we observed that XGBoost indeed yielded lower MSE values compared to other machine learning approaches. However, when evaluating the results using both human visual judgment and perceptual similarity metrics such as LPIPS and SSIM, Random Forest was found to produce more visually accurate outcomes. Since the goal of this study is not to compare different techniques for identifying the single most accurate model, but rather to emphasize the importance of perceptual metrics (like LPIPS and SSIM) alongside conventional metrics (like MSE) in urban data analysis, we defer the comparative use of additional machine learning models to future research. The focus in this present work is on

demonstrating the potential of a data-driven approach that incorporates perceptual and structural similarity assessments in the context of urban heat island prediction.

5. Conclusions

In this study, the voxel data allowed for calculating the amount of building volumes and to represent them in a 2D raster format for each air temperature pixel region. Since these building volumes are represented in raster format, it becomes possible to apply image processing techniques such as Gaussian blurring. Gaussian blurring enabled the integration of spatial neighborhood relationships before the training process, as the value of a pixel is influenced by adjacent pixel values. Thereby, the correlation rate is also increased between air temperature data and building volume data after implementing Gaussian blur.

Additionally, the proposed CityGML-to-voxel conversion steps facilitated the rapid generation of city-scale 3D volumetric data, accelerating experiments by allowing a quick acquisition of voxel data from different regions. This rapid conversion might enable urban planners to quickly implement their plans in digital applications and observe the impact on environmental indicators.

In addition, as demonstrated by previous studies, when model accuracy is evaluated using the MSE metric, the XGBoost method produces models with low error rates. However, thanks to the raster format of our rasterized methodology, these errors are observed to be systematic and the spatial distributions are not well captured. On the other hand, the Random Forest method has higher MSE values, it qualitatively demonstrates more consistent spatial patterns and the error distribution appears more randomly. Furthermore, the raster-based format of both predictions and ground truth data allowed for the implementation of image similarity metrics like SSIM or LPIPS that also capture and consider spatial relationships for the evaluation process. Since the Random Forest method qualitatively provided better results than the XGBoost method, these image similarity metrics prove qualitative finding in a quantitative manner. The SSIM and LPIPS scores for each city indicate that the prediction results of the Random Forest method are better than XGBoost. Furthermore, using the testing data confirms that these accurate patterns are not due to overfitting but rather indicate good generalization.

Taking all of these into consideration, it has been demonstrated that previous studies relying solely on MSE to evaluate their models are inadequate and lack depth. The raster-based data driven approach presented here takes into account spatial neighborhood relationships both before and after training, resulting in more accurate outcomes and insightful analyses. The models used in this study are intentionally kept simple by excluding multi-variable and multi-temporal data sources in order to clearly demonstrate the benefits of the data-driven methodology. The findings in this study may serve as a foundation for future research. The potential applications of such research in urban planning lie in the refinement and optimization of urban design strategies before implementing them into the real world, thereby contributing to more data-informed development practices.

Author Contributions: Conceptualization, B.K. and P.E.B.; methodology, B.K.; software, B.K.; validation, B.K.; investigation, B.K.; data curation, B.K.; writing—original draft preparation, B.K.; writing—review and editing, P.E.B.; visualisation, B.K.; supervision, P.E.B.; funding acquisition, P.E.B. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The 2D raster building volume and air temperature datasets used during model training, along with example code and the trained models, are publicly available via the project's GitHub repository under the following link: https://github.com/BerkKivilcim/Urban-

References

1. Han, Y.; Taylor, J.E.; Pisello, A.L. Toward mitigating urban heat island effects: Investigating the thermal-energy impact of bio-inspired retro-reflective building envelopes in dense urban settings. *Energy Build.* **2015**, *102*, 380–389. [CrossRef]
2. Lan, Y.; Zhan, Q. How do urban buildings impact summer air temperature? The effects of building configurations in space and time. *Build. Environ.* **2017**, *125*, 88–98. [CrossRef]
3. Liu, L.; Liu, J.; Jin, L.; Liu, L.; Gao, Y.; Pan, X. Climate-conscious spatial morphology optimization strategy using a method combining local climate zone parameterization concept and urban canopy layer model. *Build. Environ.* **2020**, *185*, 107301. [CrossRef]
4. Ren, Z.; Jiang, B.; Seipel, S. Capturing and characterizing human activities using building locations in America. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 200. [CrossRef]
5. Zhong, C.; Schläpfer, M.; Müller Arisona, S.; Batty, M.; Ratti, C.; Schmitt, G. Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Stud.* **2017**, *54*, 437–455. [CrossRef]
6. Tehrani, A.A.; Veisi, O.; Delavar, Y.; Bahrami, S.; Sobhaninia, S.; Mehan, A.; et al. Predicting urban Heat Island in European cities: A comparative study of GRU, DNN, and ANN models using urban morphological variables. *Urban Clim.* **2024**, *56*, 102061. [CrossRef]
7. Arnfield, A.J. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol. J. R. Meteorol. Soc.* **2003**, *23*, 1–26. [CrossRef]
8. Yang, Z.; Chen, Y.; Zheng, Z.; Huang, Q.; Wu, Z. Application of building geometry indexes to assess the correlation between buildings and air temperature. *Build. Environ.* **2020**, *167*, 106477. [CrossRef]
9. Eliasson, I. The use of climate knowledge in urban planning. *Landsc. Urban Plan.* **2000**, *48*, 31–44. [CrossRef]
10. Ng, E. Towards planning and practical understanding of the need for meteorological and climatic information in the design of high-density cities: A case-based study of Hong Kong. *Int. J. Climatol.* **2012**, *32*, 582–598. [CrossRef]
11. Fathi, S.; Srinivasan, R.; Fenner, A.; Fathi, S. Machine learning applications in urban building energy performance forecasting: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *133*, 110287. [CrossRef]
12. Liu, L.; Silva, E.A.; Wu, C.; Wang, H. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* **2017**, *65*, 113–125. [CrossRef]
13. Tekouabou, S.C.K.; Diop, E.B.; Azmi, R.; Jaligot, R.; Chenal, J. Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: Potential, issues and challenges. *J. King Saud-Univ.-Comput. Inf. Sci.* **2022**, *34*, 5943–5967. [CrossRef]
14. Venter, Z.S.; Brousse, O.; Esau, I.; Meier, F. Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. *Remote Sens. Environ.* **2020**, *242*, 111791. [CrossRef]
15. Yoo, S.J.; Kwon, T.; Lyoo, Y.S. Challenges of influenza A viruses in humans and animals and current animal vaccines as an effective control measure. *Clin. Exp. Vaccine Res.* **2018**, *7*, 1–15. [CrossRef] [PubMed]
16. Fan, C.; Zou, B.; Li, J.; Wang, M.; Liao, Y.; Zhou, X. Exploring the relationship between air temperature and urban morphology factors using machine learning under local climate zones. *Case Stud. Therm. Eng.* **2024**, *55*, 104151. [CrossRef]
17. Lau, T.K.; Lin, T.P. Investigating the relationship between air temperature and the intensity of urban development using on-site measurement, satellite imagery and machine learning. *Sustain. Cities Soc.* **2024**, *100*, 104982. [CrossRef]
18. Zheng, Z.; Zhou, W.; Yan, J.; Qian, Y.; Wang, J.; Li, W. The higher, the cooler? Effects of building height on land surface temperatures in residential areas of Beijing. *Phys. Chem. Earth Parts A/B/C* **2019**, *110*, 149–156. [CrossRef]
19. Hu, Y.; Dai, Z.; Guldmann, J.M. Modeling the impact of 2D/3D urban indicators on the urban heat island over different seasons: A boosted regression tree approach. *J. Environ. Manag.* **2020**, *266*, 110424. [CrossRef]
20. Li, H.; Li, Y.; Wang, T.; Wang, Z.; Gao, M.; Shen, H. Quantifying 3D building form effects on urban land surface temperature and modeling seasonal correlation patterns. *Build. Environ.* **2021**, *204*, 108132. [CrossRef]
21. Oke, T.R. The heat island of the urban boundary layer: Characteristics, causes and effects. In *Wind Climate in Cities*; Springer: Dordrecht, The Netherlands, 1995; pp. 81–107.
22. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [CrossRef]

23. Voogt, J.A.; Oke, T.R. Thermal remote sensing of urban climates. *Remote Sens. Environ.* **2003**, *86*, 370–384. [CrossRef]

24. Wu, J. Urban sustainability: An inevitable goal of landscape research. *Landsc. Ecol.* **2010**, *25*, 1–4. [CrossRef]

25. Zhou, W.; Huang, G.; Cadenasso, M.L. Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landsc. Urban Plan.* **2011**, *102*, 54–63. [CrossRef]

26. Lin, A.; Wu, H.; Luo, W.; Fan, K.; Liu, H. How does urban heat island differ across urban functional zones? Insights from 2D/3D urban morphology using geospatial big data. *Urban Clim.* **2024**, *53*, 101787. [CrossRef]

27. Liu, B.; Guo, X.; Jiang, J. How urban morphology relates to the urban heat island effect: A multi-indicator study. *Sustainability* **2023**, *15*, 10787. [CrossRef]

28. Raaymakers, T. Understanding Urban Temperature Differences through 2D/3D Urban Morphology. In *Context of the Belt & Road Initiative*; Wageningen University and Research: Wageningen, The Netherlands, 2024.

29. Isa, N.A.; Salleh, S.A.; Mohd, W.M.N.W.; Chan, A.; Ooi, M.C.G.; Zakaria, N.H.; Islam, M.A. Building Volume Effects on Ambient Temperature In The Kuala Lumpur City. In *IOP Conference Series: Earth and Environmental Science, Proceedings of the 15th International Conference on Atmospheric Sciences and Applications to Air Quality, Kuala Lumpur, Malaysia, 28–30 October 2019*; IOP Publishing: Bristol, UK, 2020; Volume 489, p. 012011.

30. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

31. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.

32. Ghildyal, A.; Liu, F. Attacking perceptual similarity metrics. *arXiv* **2023**, arXiv:2305.08840.

33. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A.; et al. Nerfstudio: A modular framework for neural radiance field development. In Proceedings of the ACM SIGGRAPH 2023 Conference Proceedings, Los Angeles, CA, USA, 6–10 August 2023; pp. 1–12.

34. Voogt, J.A.; Oke, T. Effects of urban surface geometry on remotely-sensed surface temperature. *Int. J. Remote. Sens.* **1998**, *19*, 895–920. [CrossRef]

35. Cai, Z.; Han, G.; Chen, M. Do water bodies play an important role in the relationship between urban form and land surface temperature? *Sustain. Cities Soc.* **2018**, *39*, 487–498. [CrossRef]

36. Guo, G.; Zhou, X.; Wu, Z.; Xiao, R.; Chen, Y. Characterizing the impact of urban morphology heterogeneity on land surface temperature in Guangzhou, China. *Environ. Model. Softw.* **2016**, *84*, 427–439. [CrossRef]

37. Wu, C.D.; Lung, S.C.C.; Jan, J.F. Development of a 3-D urbanization index using digital terrain models for surface urban heat island effects. *ISPRS J. Photogramm. Remote Sens.* **2013**, *81*, 1–11. [CrossRef]

38. Ren, C.; Cai, M.; Li, X.; Shi, Y.; See, L. Developing a rapid method for 3-dimensional urban morphology extraction using open-source data. *Sustain. Cities Soc.* **2020**, *53*, 101962. [CrossRef]

39. Heeramaglore, M.; Kolbe, T.H. Semantically enriched voxels as a common representation for comparison and evaluation of 3D building models. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *10*, 89–96. [CrossRef]

40. Padsala, R.; Gebetsroither-Geringer, E.; Bao, K.; Coors, V. The Application of CityGML Food Water Energy ADE to Estimate the Biomass Potential for a Land Use Scenario. In *CITIES 20.50–Creating Habitats for the 3rd Millennium: Smart–Sustainable–Climate Neutral, Proceedings of the REAL CORP 2021, 26th International Conference on Urban Development, Regional Planning and Information Society, Vienna, Austria, 7–10 September 2021*; CORP—Competence Center of Urban and Regional Planning: Vienna, Austria, 2021; pp. 851–861,

41. Biljecki, F.; Ledoux, H.; Stoter, J. An improved LOD specification for 3D building models. *Comput. Environ. Urban Syst.* **2016**, *29*, 25–37. [CrossRef]

42. Henn, A.; Römer, C.; Gröger, G.; Plümer, L. Automatic classification of building types in 3D city models: Using SVMs for semantic enrichment of low resolution building data. *GeoInformatica* **2012**, *16*, 281–306. [CrossRef]

43. Hofierka, J.; Zlocha, M. A new 3-D solar radiation model for 3-D city models. *Trans. GIS* **2012**, *16*, 681–690. [CrossRef]

44. Mulder, D. Automatic Repair of 3D City Building Model Using a Voxel–Based Repair Method. Master Thesis, Delft University of Technology, Delft, The Netherlands, 2015.

45. Nourian, P.; Gonçalves, R.; Zlatanova, S.; Ohori, K.A.; Vo, A.V. Voxelization algorithms for geospatial applications: Computational methods for voxelating spatial datasets of 3D city models containing 3D surface, curve, and point data models. *MethodsX* **2016**, *3*, 69–86. [CrossRef]

46. Willenborg, B.; Sindram, M.; Kolbe, T.H. Semantic 3D city models serving as information hub for 3D field based simulations. *Lösungen Für Eine Welt Im Wandel* **2016**, *25*, 54–65.

47. Sindram, M.; Machl, T.; Steuer, H.; Pültz, M.; Kolbe, T.H. Voluminator 2.0–Speeding up the approximation of the volume of defective 3D building models. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 29–36. [CrossRef]

48. Konde, A.; Saran, S. Web enabled spatio-temporal semantic analysis of traffic noise using CityGML. *J. Geomat.* **2017**, *11*, 248–259.

49. Ridzuan, N.; Ujang, U.; Azri, S. 3D vectorization and rasterization of CityGML standard in wind simulation. *Earth Sci. Inform.* **2023**, *16*, 2635–2647. [CrossRef]

50. Jahn, M.; Bradley, P. Computing watertight volumetric models from boundary representations to ensure consistent topological operations. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *VIII-4/W2-2021*, 21–28. [CrossRef]

51. Jahn, M.; Bradley, P. A Robustness Study for the Extraction of Watertight Volumetric Models from Boundary Representation Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 224. [CrossRef]

52. Jahn, M. Distributed & Parallel Data Management to Support Geo-Scientific Simulation Implementations. Ph.D. Thesis, Karlsruhe Institute of Technology, Karlsruhe, Germany, 2022.

53. Pusacker, K.; Coors, V.; Eckhardt, J.D.; Rupf, I. A Concept for 3D Geological and Urban Subsurface Modeling with a Unified Voxel Model Examined by a Case Study for the City Center of Stuttgart (Baden-Württemberg), Germany. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *10*, 193–200. [CrossRef]

54. Konarska, J.; Holmer, B.; Lindberg, F.; Thorsson, S. Influence of vegetation and building geometry on the spatial variations of air temperature and cooling rates in a high-latitude city. *Int. J. Climatol.* **2016**, *36*, 2379–2395. [CrossRef]

55. Kim, H.J.; Shrestha, A.; Sapkota, E.; Pokharel, A.; Pandey, S.; Kim, C.S.; Shrestha, R. A study on the effectiveness of spatial filters on thermal image pre-processing and correlation technique for quantifying defect size. *Sensors* **2022**, *22*, 8965. [CrossRef] [PubMed]

56. Zumwald, M.; Knüsel, B.; Bresch, D.N.; Knutti, R. Mapping urban temperature using crowd-sensing data and machine learning. *Urban Clim.* **2021**, *35*, 100739. [CrossRef]

57. Du, Z.; Wang, H. Is a higher correlation necessary for a more accurate prediction? *Sci. China Phys. Mech. Astron.* **2011**, *54*, 172–175. [CrossRef]

58. Open Source CityGML Data of Thuringia/Germany. Available online: https://geoportal.thueringen.de/gdi-th/download-offene-geodaten/download-3d-gebaeudedaten (accessed on 10 February 2024).

59. Krähenmann, S.; Walter, A.; Brienen, S.; Imbery, F.; Matzarakis, A. High-resolution grids of hourly meteorological variables for Germany. *Theor. Appl. Climatol.* **2018**, *131*, 899–926. [CrossRef]

60. DWD Climate Data Center (CDC): Annual Mean of Station Observations of Daily Air Temperature at 2 m Above Ground in °C for Germany. Available online: https://opendata.dwd.de/climate_environment/CDC/grids_germany/hourly/hostrada/air_temperature_mean/ (accessed on 3 August 2024).

61. Sandia National Labs; Kitware Inc.; Los Alamos National Labs. Paraview: Parallel Visualization Application. Available online: https://www.paraview.org/ (accessed on 5 January 2023).

62. QGIS Development Team. Organization: Open Source Geospatial Foundation. QGIS Geographic Information System (Used QGIS version: 3.18.0 with GRASS 7.8.5). Available online: https://qgis.org/ (accessed on 23 April 2025).

63. Vitalis, S.; Arroyo Ohori, K.; Stoter, J. CityJSON in QGIS: Development of an open-source plugin. *Trans. GIS* **2020**, *24*, 1147–1164. [CrossRef] [PubMed]

64. GDAL/OGR Contributors. *GDAL/OGR Geospatial Data Abstraction Software Library*; Version 3.2.1 Is Used—This Version Is Introduced in 2021; Zenodo: Geneva, Switzerland, 2024. [CrossRef]

65. Ledoux, H.; Arroyo Ohori, K.; Kumar, K.; Dukai, B.; Labetski, A.; Vitalis, S. CityJSON: A compact and easy-to-use encoding of the CityGML data model. *Open Geospat. Data Softw. Stand.* **2019**, *4*, 1–12. [CrossRef]

66. Tanoori, G.; Soltani, A.; Modiri, A. Machine Learning for Urban Heat Island (UHI) Analysis: Predicting Land Surface Temperature (LST) in Urban Environments. *Urban Clim.* **2024**, *55*, 101962. [CrossRef]

67. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. Available online: https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (accessed on 23 April 2025). [CrossRef]