Spatial Interpolation in Mystic Vale

Magdalen Thot

2023-11-04

1 Introduction

1.1 Context

Understanding temperature variation in the study area, particularly due to varying elevations in Mystic Vale, provides crucial insights into the potential impacts of climate change at a larger scale. Spatial interpolation techniques, emphasized in academic literature, are instrumental for practical applications in this context. For instance, Chen, Shen, Jin, and Chen (2011) demonstrated how spatial interpolation methods can create accurate gridded climatological temperature datasets for studying climate change and natural disasters in northeast China. Similarly, Dhamodaran (2021) highlighted the significance of spatial interpolation, especially ordinary Kriging and inverse distance weighted (IDW), for predicting data and generating high-resolution climate maps. This approach allows for the assessment of topographical features and climate change impacts, including disasters like floods and landslides.

Additionally, Irmak (2010) used spatial interpolation to analyze rainfall patterns in Nebraska over a 30-year period. The study compared three common interpolation methods (inverse distance weighted, spline, and kriging) and found no significant differences in predicting spatial variations for 30-year climate normals. Long-term data averaging appeared to homogenize spatial variability, resulting in similar errors across interpolation techniques. It's noteworthy that kriging produced smoother maps compared to spline and inverse distance weighted.

Furthermore, Li et al. (2011) investigated temperature and precipitation data from 1961 to 2005 across 65 meteorological stations in Xinjiang, China. They used statistical methods like the Mann-Kendall test, inverse distance weighted (IDW) interpolation, and Remote sensing analysis to uncover climate change trends. The key findings indicate that both temperature and precipitation have increased over the past 45 years, with temperature exhibiting a more significant increase. The rate of temperature rise is influenced by latitude and elevation, with higher latitudes and elevations experiencing faster increases. In conclusion, these studies collectively underscore the important role of spatial interpolation techniques in understanding and assessing climate variations.

1.2 Research Topic

Spatial interpolation plays a pivotal role in climate change studies by estimating values at unobserved locations, particularly when assessing the influence of elevation on temperature variation. It fills data gaps, creating detailed climatic maps. Granville et al. (2022) emphasized the importance of selecting the right interpolation method for collaborative interdisciplinary research, as demonstrated in their Fire Weather Index case study in Ontario. Hussain, Faisal, Shad, Hussain, and Ahmed (2015) underscored the significance of interpolating elevation data for predicting environmental parameters in Pakistan, with model-based Bayesian kriging proving the most effective approach. Hoffman, Ziegler, Tinkham, Hiers, and Hudak (2023) compared four spatial interpolation methods in fire-prone ecosystems, highlighting regression kriging's superiority. Palmer et al. (2009) evaluated spatial interpolation techniques in New Zealand's Pinus radiata forests, revealing the complexities and performance variations among methods. Despite the valuable research contributions in spatial interpolation, there is still a gap in the literature related to evaluating these methods and their adaptability in areas with varying elevation characteristics.

1.3 Research Objective

In spatial analysis, a gap in knowledge warranting further exploration is the use of spatial interpolation methods in regions characterized by varying elevation profiles. This gap in the literature necessitates an in-depth investigation to shed light on the most effective spatial interpolation techniques for these regions. This study seeks to address this critical question by rigorously assessing the performance of various interpolation methods, including kriging, inverse distance weighting, and Thiessen polygons.

2 Methods

2.1 Study Site & Data

The study site for the report is Mystic Vale, a forested ravine located southeast of the University of Victoria campus, outside the ring road. Its canopy contains various tree species, including Douglas-fir, grand fir, western red cedar, conifers, bigleaf maple, black cottonwood, willow trees, and more. The study site covers 11.6 acres and is regularly used by students and members of surrounding communities for recreational activities. It comprises a portion of a parking lot, a sizable forested area, and a section of a grassy field. The average annual temperature in Mystic Vale is 18 degrees Celsius in September, and the region has a moderate oceanic climate. Conducting this study in Mystic Vale is significant due to the region's highly varied terrain. Elevation varies greatly across different parts of the area, which can have a significant impact on biodiversity and is crucial for predicting weather changes and studying climate trends.

The data for this study was collected in Mystic Vale using spatial sampling methods. Specifically, a modified stratified proportional sampling approach was employed, wherein the study area was subdivided into distinct sections, and data points were systematically collected within each section. For instance, the largest section received the highest number of sample points. Data collection took place on September 14th, 2023, starting at 2:30 PM. At each selected sample point, temperature measurements were recorded using a temperature gun. Additionally, the collected data is stored in a CSV file. To prepare the data for analysis, it underwent a series of steps. First, it was organized and converted into a shapefile format. The data was then projected using the EPSG:32610 projection. An image raster of the study area, available in TIFF format, was also incorporated into the dataset. Furthermore, a corresponding shapefile of the study area was included, which was also projected using EPSG:32610.

The data has a right-skewed distribution with a higher mean (22.24) than median (19.17), which indicates that there might be some higher temperature outliers. The skewness value of approximately 1.58 further supports the presence of positive skewness, which means the distribution is not symmetrical.

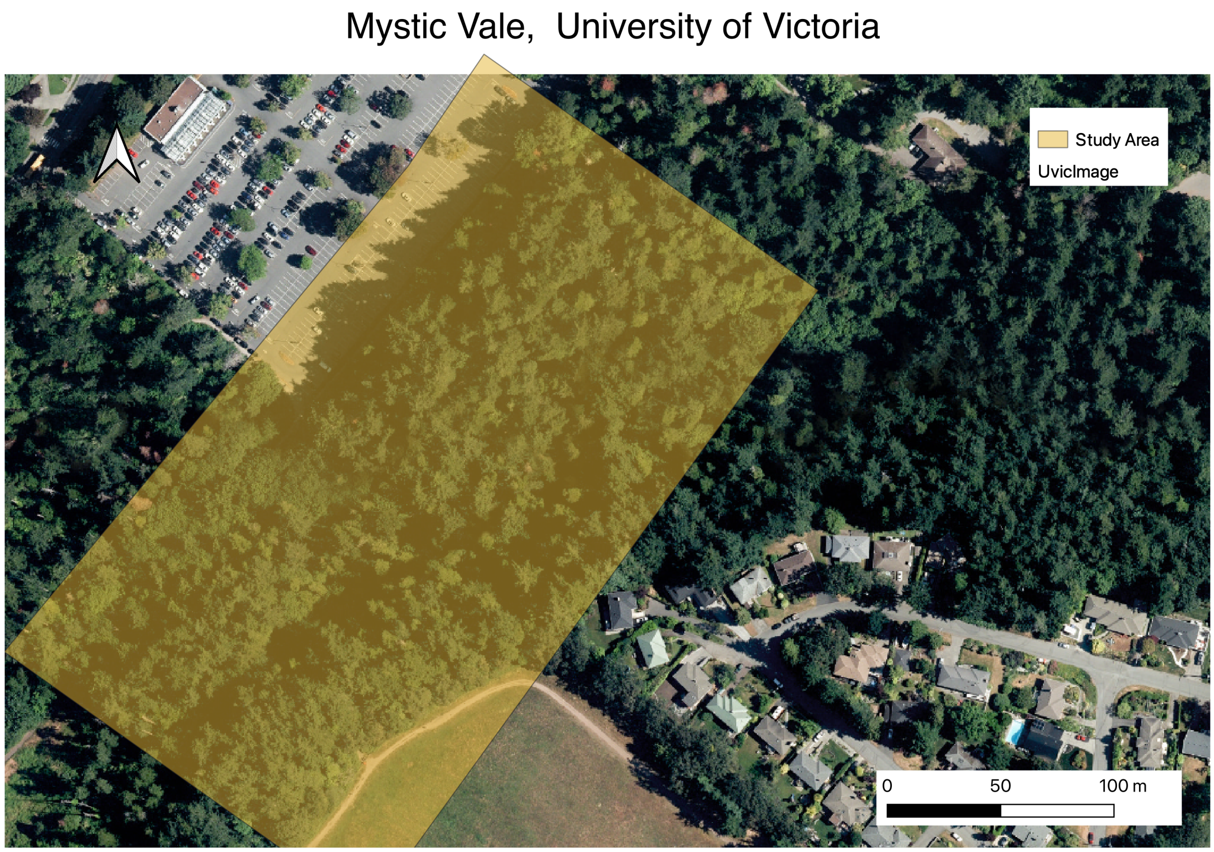

Figure 1: A map of the study area, Mystic Vale

2.2 Description of Methods

2.2.1 Data Preperation

In our study, we assessed several spatial interpolation methods, namely Thiessen polygons, inverse distance weighting (IDW), and ordinary kriging. Before applying these methods, we undertook a structured data subsetting process. Initially, a seed value was chosen to ensure the reproducibility of random row selection. Subsequently, random row numbers were generated and assigned appropriate identifiers for sampling. The selected rows were then segregated into a training dataset. Following this, the remaining rows were subdivided once more, and another set of random row numbers was generated to create a distinct sample. These newly selected rows were again allocated to a training dataset. This refined subset of the training data formed the foundation for our subsequent spatial interpolation methods.

2.2.2 Thiessen Polygons

Thiessen polygons is a method for the partitioning a geographical area into distinct regions or polygons based on the proximity of a set of points. Each Thiessen polygon corresponds to one of these points and encompasses all locations that are closer to that specific point than to any other point in the dataset. Our approach began by duplicating the training dataset and appending the x and y coordinates into new columns. Subsequently, we created an observation window, defining the boundary for our analysis. Following that, we formed a class representing a two-dimensional point pattern, known as a PPP object, which includes essential information about the observation window. We then employed a Dirichlet function to compute the Dirichlet tessellation of the point pattern. This tessellation effectively divided the study area into polygons based on the PPP object as input, ensuring consistent projection information. To enrich the analysis, attribute information such as mean temperatures was linked to their corresponding polygons within the tessellated surface. The resulting output was clipped to the defined study area, and the outcomes were visually presented for further analysis and interpretation.

The we evaluate the models performance. First, we perform leave-one-out cross-validation to assess the method's predictive performance. Next, we create an empty vector, to store cross-validation results and systematically iterate through the training dataset, excluding one data point during each loop. We define an observation window, generate a cross-validated dataset for each iteration, compute Thiessen polygons, and predict temperature values at test points. Then a scatter plot is used to visualize differences between observed and predicted values, and RMSE provides an overall accuracy measure. Next, we compare the interpolated surface created using the training data with validation points. After defining the validation dataset, we compute differences between predicted temperature values from the interpolated surface and observed validation points. A scatter plot illustrates these differences, along with a linear regression fit and RMSE for performance evaluation.

2.2.3 Inverse Distance Weighting (IDW)

The analysis proceeds with Inverse Distance Weighting (IDW), a method for estimating values at unsampled locations based on their proximity to known data points. This adheres to Tobler's First Law of Geography, which states that closer points exert more influence. IDW is based on the formula below

$$Z_i = \frac{\sum(\frac{Z_j}{d_{ij}^p})}{\sum(\frac{1}{d_{ij}^p})}$$

The choice of the power parameter determines the strength of this influence, with higher values emphasizing nearby points and lower values producing smoother surfaces. The IDW process involves creating a grid aligned with the study area's coordinate reference system, performing IDW interpolation using training data, and clipping the results.

To determine the optimal power value (IDP), a leave-one-out validation approach is applied, aiming to minimize the Root Mean Square Error (RMSE). Subsequently, IDW interpolation is performed using the optimal power value. Next, the interpolated surface is compared with validation data using a scatter plot, including a red dashed line for linear regression and RMSE for assessing performance. Finally, a jackknife technique is utilized to assess accuracy by iteratively excluding data points, performing IDW interpolation, and recording predicted values. This comprehensive approach ensures both accurate interpolation and reliable validation.

2.2.4 Ordinary Kriging

The last interpolation method we applied was Ordinary Kriging, a spatial statistical technique used to estimate attribute values at unsampled locations within a study area by considering spatial correlations between known data points.

Our analysis began with the creation of a semi variogram model to assess spatial autocorrelation. More specifically, a crucial decision point in this phase of the analysis was to select the most appropriate variogram model that best fits the data. This decision was made after conducting multiple tests and fine-tuning parameters, and the results of these tests can be seen in Table 1, the model with the loest RMSE is chosen.

This process involved initiating an initial variogram model by specifying parameters for the sill, range, and nugget. The choice of these parameters is essential for accurately capturing the spatial variability and structure within the dataset. It is important to note that the variogram model must reflect the actual spatial characteristics of the data (Figure. Adjusting these parameters and trying different variogram models helped us identify the model that best described the spatial correlation patterns in the dataset. This iterative approach ensures that the final model provides the most accurate estimation of attribute values across the study area. For this analysis a liner model was selected. Next, after selecting the model, we create a spatial grid, apply Ordinary Kriging to estimate mean temperature values across the grid, and visualizes the interpolated surface. This processs uses a the pre-defined variogram model for interpolation accuracy, and the resulting kriging surface is displayed with training data points in a plot.

To assess the model's accuracy, a leave-one-out cross-validation procedure was employed, involving a for loop to predict values for omitted data points using the Kriging model. The differences between predicted and observed values were visualized in dedicated plots, and the Root Mean Square Error (RMSE) was calculated for prediction accuracy. Subsequently, the Kriging-based surface was used to generate variance and confidence interval maps, offering insights into data variability. These maps were compared to validation data points using for loops, resulting in various plots to comprehensively evaluate the model's performance.

2.3 Description of Evaluation

2.3.1 Thiessen Polygons

The evaluation process for Thiessen Polygons employs leave-one-out cross-validation, where each data point is excluded iteratively. The results are stored in a vector, and a comparison is made between observed and predicted values through scatter plots. To measure overall accuracy, the Root Mean Square Error (RMSE) is calculated. In the second segment, the interpolated surface is assessed by comparing it to validation points using scatter plots and RMSE calculations.

2.3.2 Inverse Distance Weighting (IDW)

IDW evaluation revolves around selecting the optimal power value (IDP) to minimize RMSE via leave-one-out validation. The chosen IDP is then used for IDW interpolation, and the results are compared to validation data using scatter plots, linear regression fits, and RMSE calculations. Additionally, a jackknife technique is employed to assess accuracy by systematically excluding data points.

2.3.3 Ordinary Kriging

In the case of Ordinary Kriging, accuracy is assessed through leave-one-out cross-validation, accompanied by RMSE calculations. The methodology also includes the generation of variance and confidence interval maps to gain insight into data variability. These maps are compared to validation points using various plots. The comprehensive evaluation methods used in this study are described in a manner that allows for reproducibility and are not specific to any particular software, ensuring clarity and transparency in the assessment process.

3 Results

3.1 Figures

3.1.1 Thiessen Polygons Figures

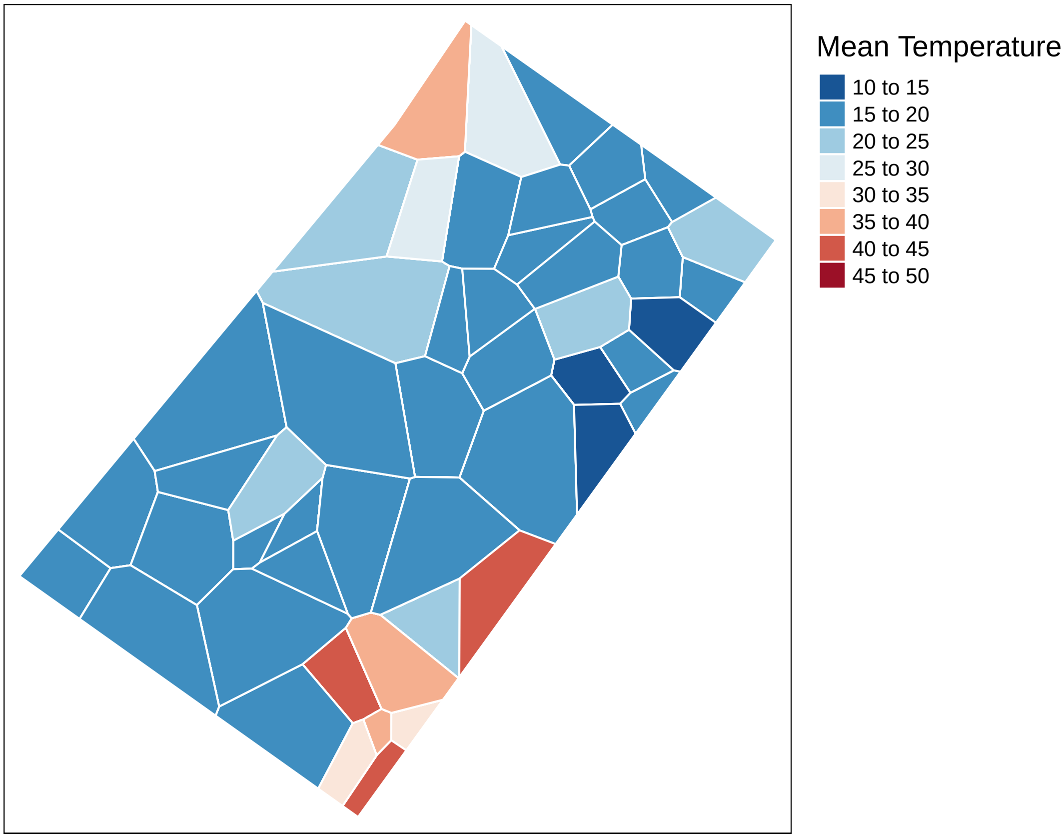

Figure 2: This map shows the mean temperatures in degrees celsius at each of the sampled locations in the study area as Thessien Polyons.

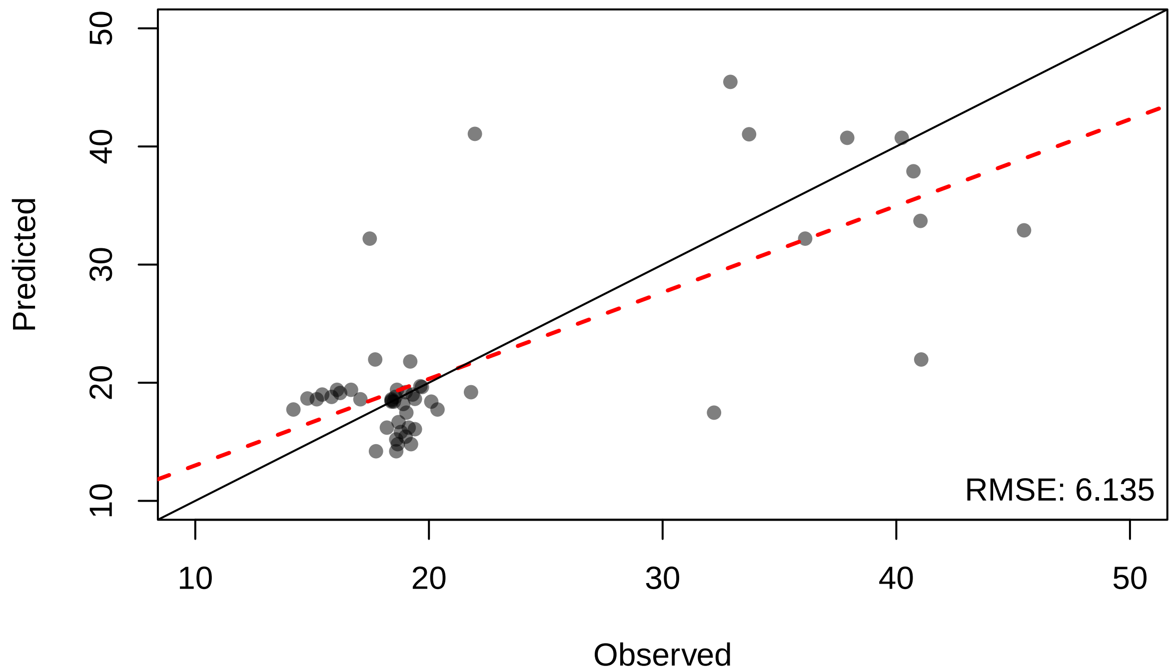

Figure 3: This plot shows observed vs. predicted values in the Thiessen polygon analysis and includes the RMSE (Root Mean Square Error) value as text on the plot for model evaluation

3.1.2 Inverse Distance Weighting Figures

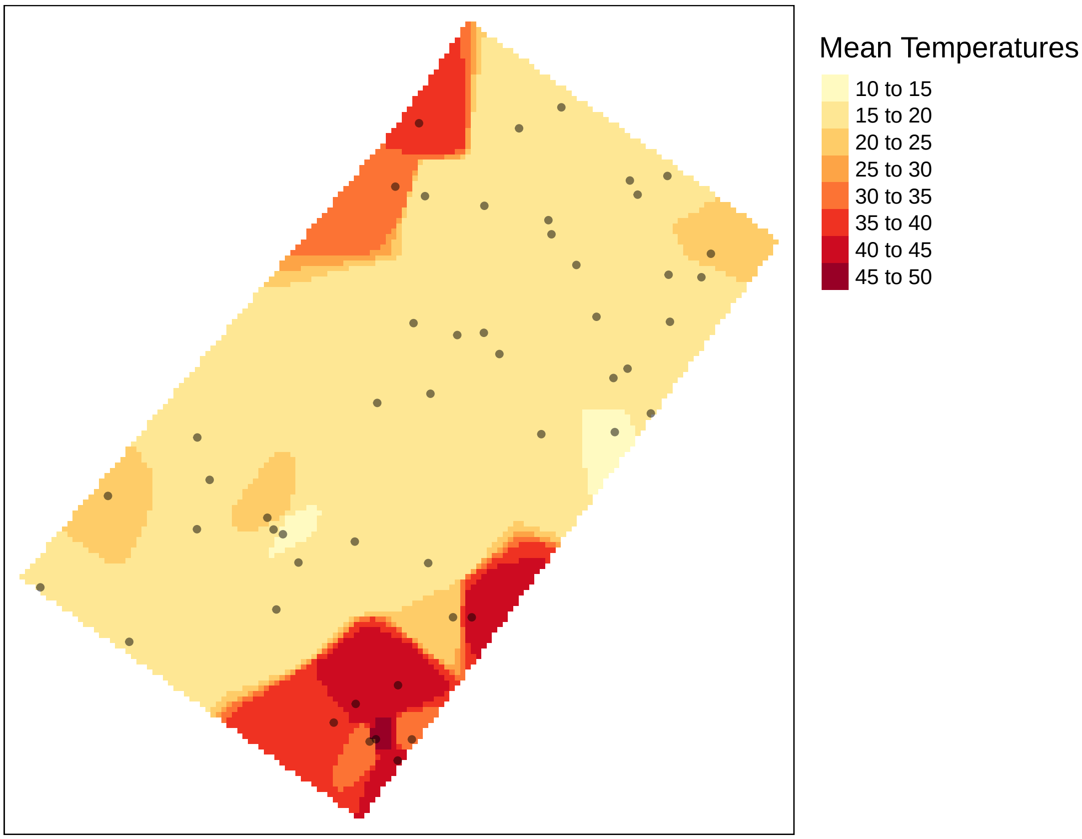

Figure 4: This map shows the IDW surface of mean temperatures in degrees Celsius at each of the sampled locations in the study area using a power function of 20

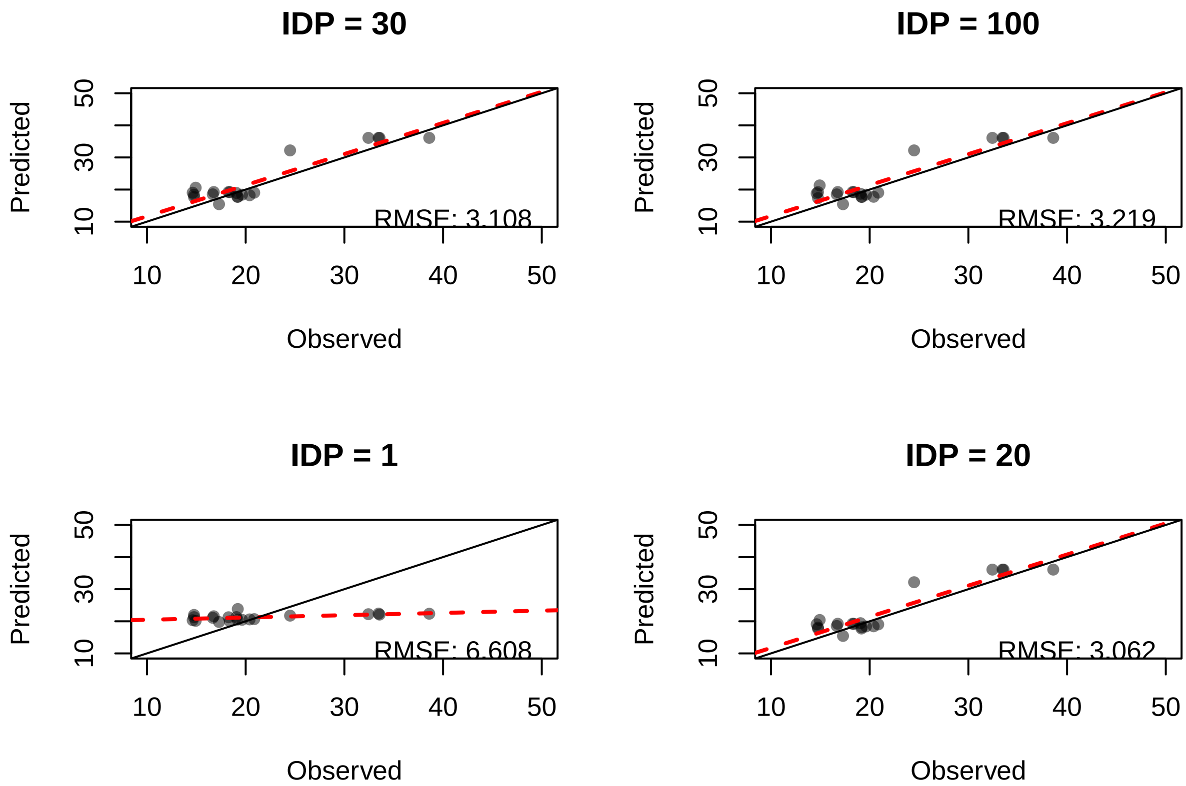

Figure 5: This figure shows how different IDP values affect the RMSE (root mean square error). The higher the error, the less reliable the model is.

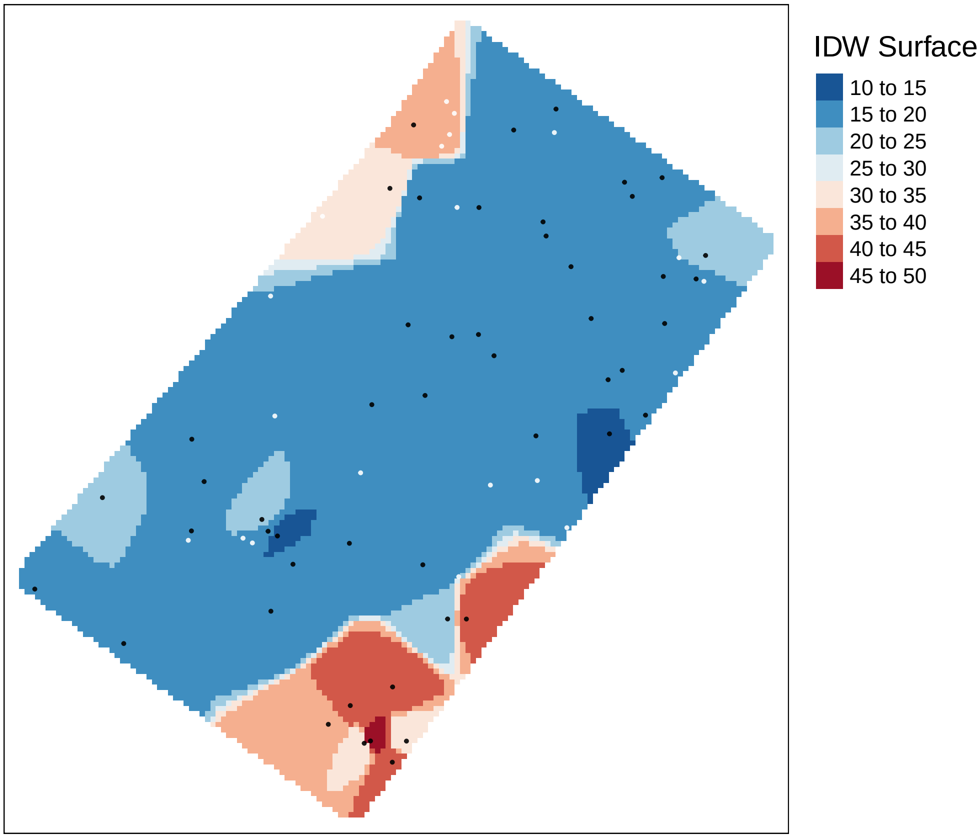

Figure 6: This map shows the final IDW surface of mean temperatures in degrees Celsius.

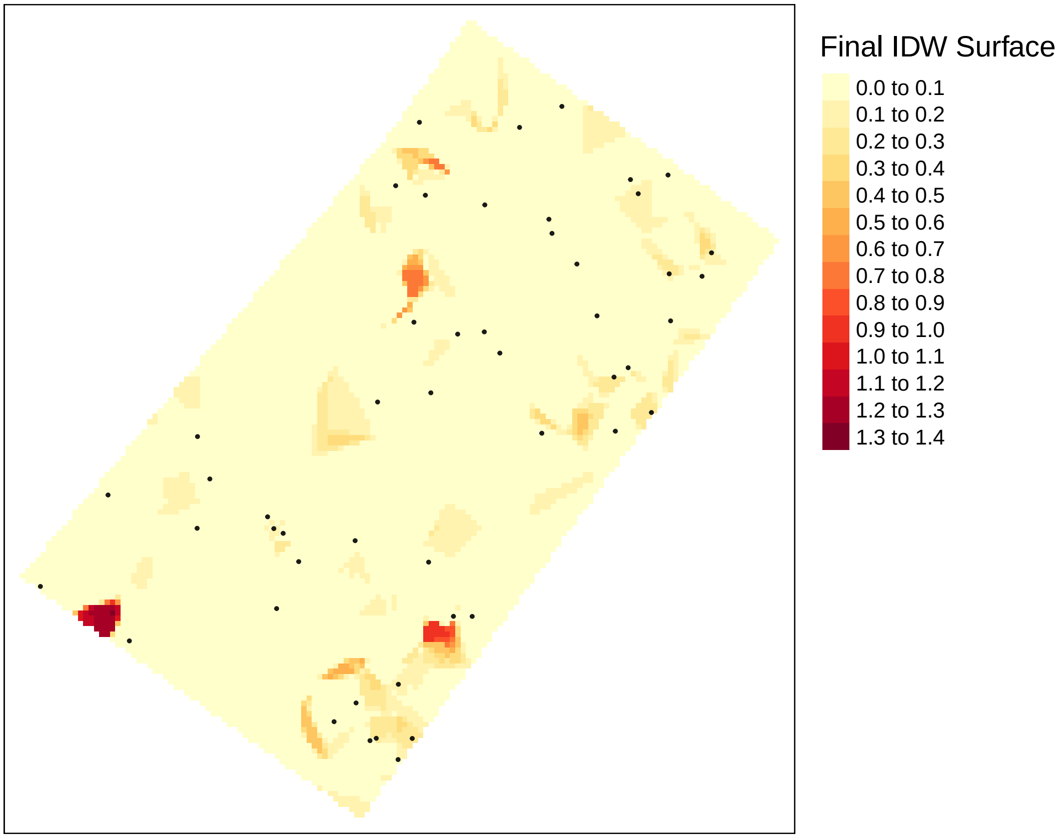

Figure 7: This a map has two main components: a raster layer representing pixel-level confidence intervals using a color palette, and a set of dots representing the locations of data points. This map is intended to visualize the spatial distribution of temperature data along with confidence intervals based on IDW interpolation

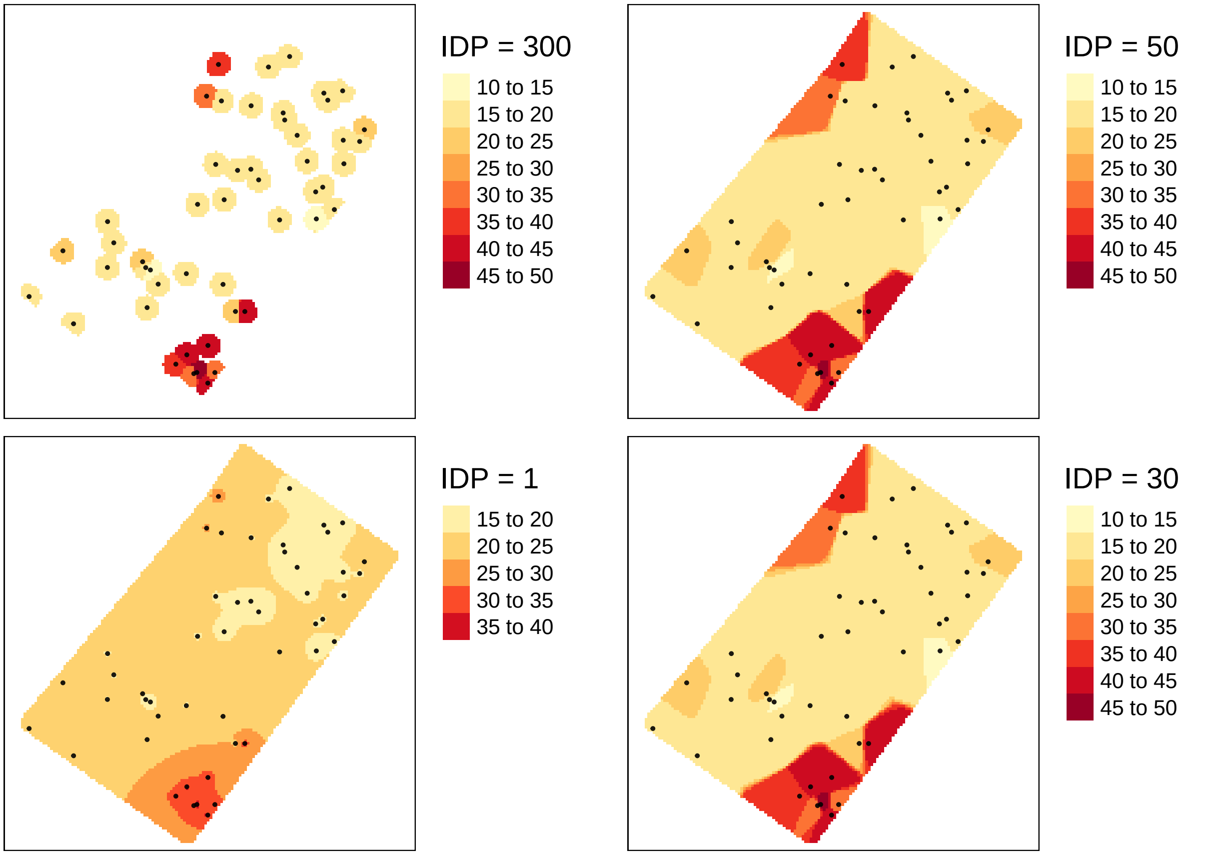

Figure 8: This figure shows how different power functions can change the surface of an IDW map. Higher power functions result in more emphasis on the nearest points, while lower ones result in less emphasis on the nearest points

3.2 Ordinary Kriging

This table displays the different tested models and their values. 'Sph' stands for a Spherical model, 'Exp' stands for an Exponential model, and 'Per' represents a Periodic model. (Table 1).

Table 1: Tested Models

	Nugget	Sill	Range	RMSE
Sph	10	50	100	5.08
Exp	10	30	200	5.07
Per	10	30	100	8.45

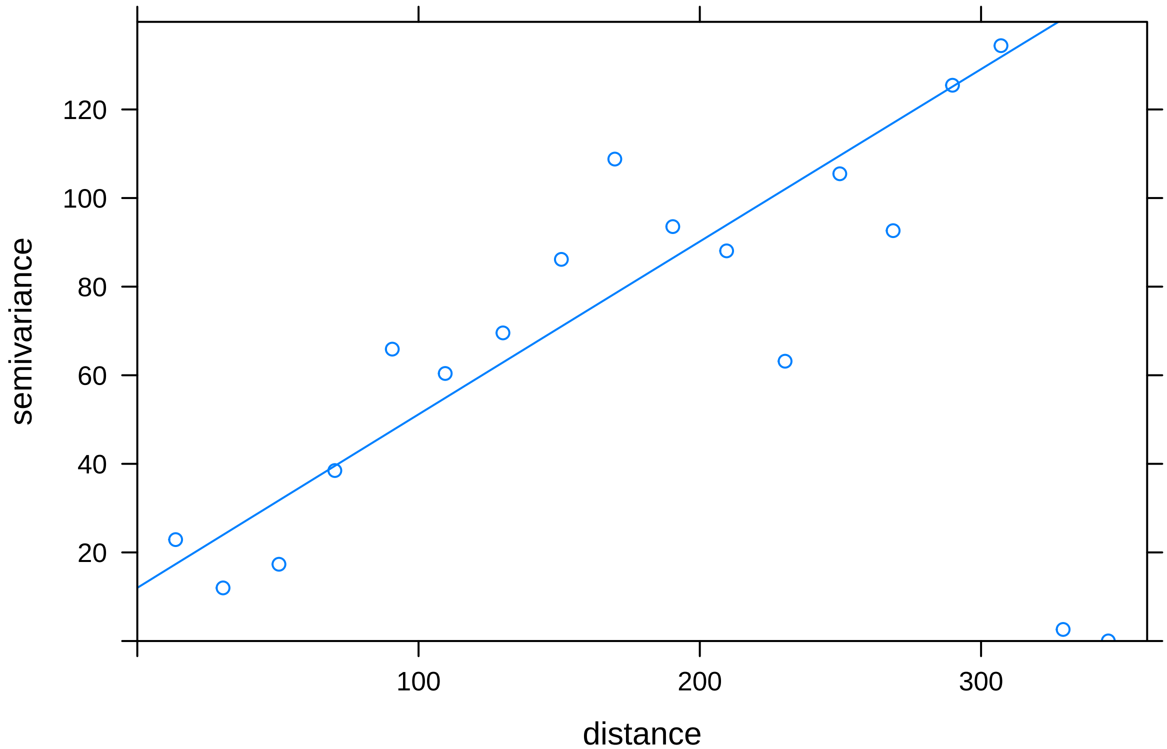

Figure 9: This figure shows the fitted exponential model over the semivariogram points. It is an exponential model with a sill of 30, a range of 200, and a nugget of 10.

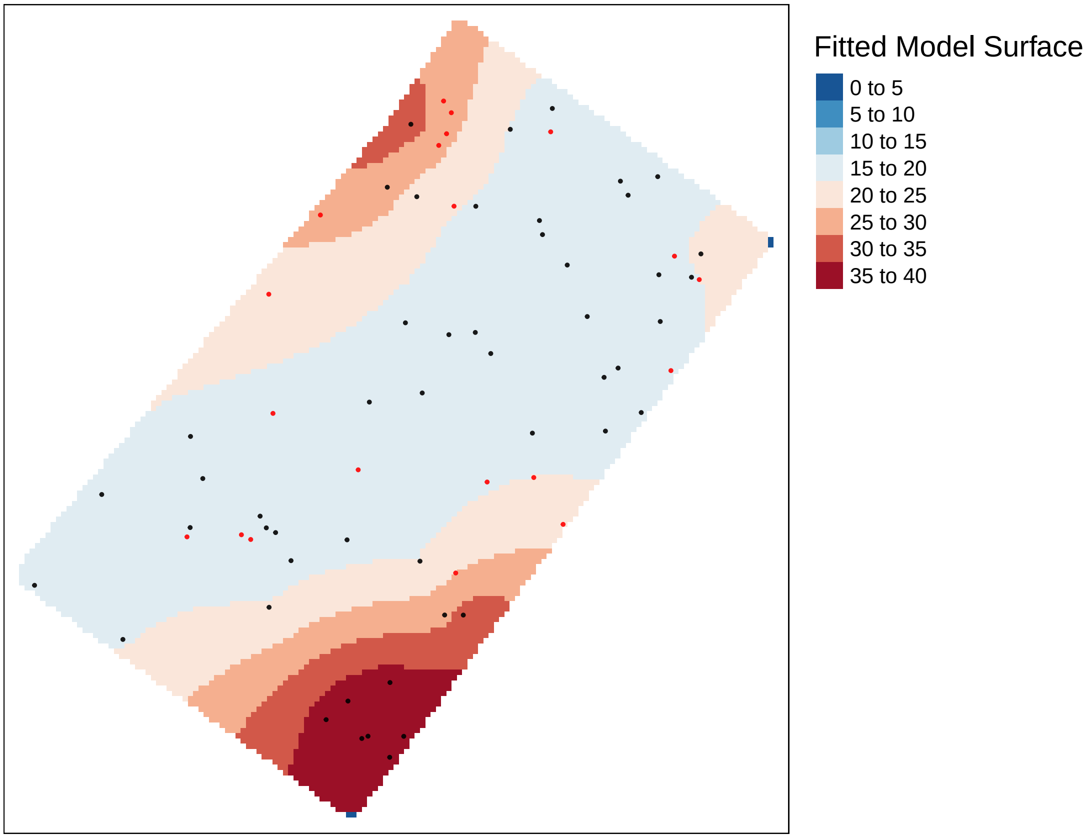

Figure 10: This figure displays the results of a Kriging interpolation model along with training (red) and validation data points (black). The map includes a raster representation of the fitted model surface, using a red-blue color palette, with dark blue representing the lowest and dark red representing the highest temperatures in degrees Celsius

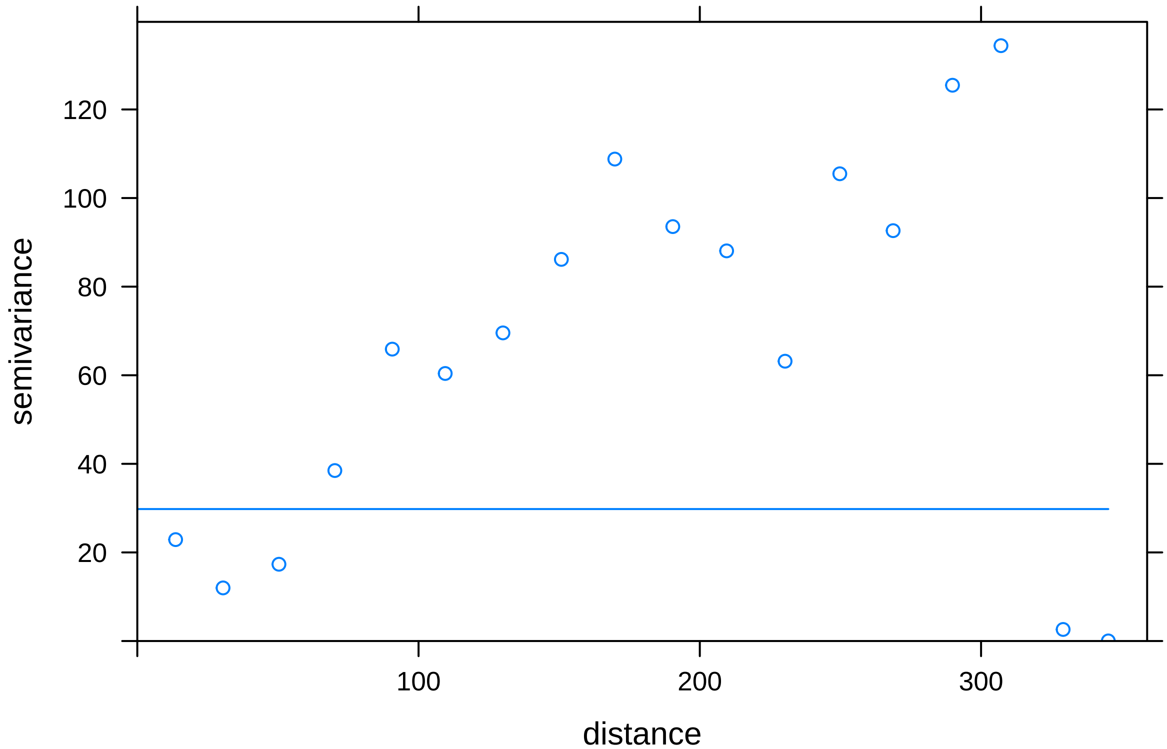

Figure 11: An example of an incorrectly fitted variogram yields a very high RMSE result of 8.45.

3.3 Description of Results

3.3.1 Thiessen polygons

The Thiessen polygon method effectively divided the study area into unique regions, highlighting areas with extreme mean temperatures. It performed well in outlining peripheral zones, such as parking lots and grassy fields.However, its performance was sub optimal for this dataset due to the method's sensitivity to data point distribution. In cases of dense clustering or sparse distribution, Thiessen polygons could not accurately represent the underlying spatial patterns. This sensitivity is evident in Figure 2, where strong clustering makes Thiessen polygons less suitable for this dataset.

3.3.2 Inverse Distance Weighting

The IDW interpolation method relies on inverse distance for estimating values at unsampled locations. Findings demonstrate its sensitivity to the choice of the power parameter (IDP), significantly affecting the results (Figure 8). Optimal performance is achieved with an IDP value of 20, emphasizing the need for parameter optimization. Variations in IDP show a substantial impact on the root mean squared error (RMSE) value and model accuracy, as depicted in Figure 5.

3.3.3 Kriging

In contrast, Kriging excels in capturing spatial dependence and autocorrelation. The variogram modeling process is essential in the Kriging method and facilitates the identification of spatial correlation patterns. The findings emphasize the significance of variogram model fitting, where parameters such as model type, sill, range, and nugget play a crucial role in achieving an accurate representation of the spatial structure.

The Kriging method demonstrates enhanced accuracy with a well-fitted variogram model. In this case, the optimal model is an exponential one with a sill of 30, a range of 200, and a nugget of 10. As these parameters change, the output of the semivariogram and RMSE also changes significantly. Figure 11 illustrates the impact of selecting an inappropriate model on RMSE.

4 Disscussion

4.1 Description of General Findings

This study aimed to identify the most effective method for accurately predicting temperature variations in regions with complex elevation profiles. It compared three interpolation methods: Thiessen polygons, IDW, and Kriging. The findings highlighted that Thiessen polygons are sensitive to data distribution, IDW requires careful power parameter optimization, and Kriging with a well-fitted variogram model performs well.

The study's implications extend to climate change, emphasizing the importance of method selection in climate mapping and parameter optimization for accuracy. However, the study has limitations. It examined only three methods, leaving room for future research to explore more. Future studies could explore more efficient parameter optimization methods.

4.2 Description of Specific Findings

4.2.1 Thiessen Polygons

Thiessen polygons offer simplicity and clear boundaries, making them a suitable choice for straightforward spatial interpolation tasks. They are best used when the dataset is evenly distributed and when distinct boundaries between areas of influence are desired. However, their sensitivity to data distribution makes them less suitable for datasets with clustered or varying points. Their discontinuous nature can result in unrealistic maps when applied to irregularly spaced data. Their lack of smoothness may lead to jagged representations in regions with sparse data, as seen in the Thiessen Polygons surface (Figure 2). Hence, Thiessen polygons are best suited for uniform datasets where clear, non-overlapping divisions are needed and should be avoided in cases with clustering, such as with the data in the study.

4.2.2 Inverse Distance Weighting (IDW)

IDW is a flexible and relatively easy-to-implement method, making it a good choice for general spatial interpolation tasks. It is most effective when you need to emphasize the influence of nearby data points over distant ones and when a simple and computationally efficient approach is required. However, IDW's accuracy heavily depends on the choice of the power parameter (IDP), making it best suited for cases where this parameter can be accurately determined, as evidenced by the impact of different IDP values on the surface and RMSE (Figure 5 and 8). Consequently, it should be avoided when dealing with complex spatial trends or intricate spatial autocorrelations as it tends to oversmooth data, potentially obscuring local variability, as evident in Figure 8 where IDP = 1.

4.2.3 Ordinary Kriging

Kriging excels in capturing spatial autocorrelation and producing continuous, realistic surfaces. It is best used in regions with complex spatial relationships, intricate elevation profiles, or varying point distributions. Kriging's optimal performance is realized when a well-fitted variogram model is used (Figure 9). However, the use of Kriging should be avoided in cases where a variogram model fitting might be challenging or time-consuming, making it less practical for beginners or when computational resources are limited. Additionally, Kriging assumes stationarity in the data (data doesn't change over time), so its use may be less suitable in situations with non-stationary data. In summary, the choice between Thiessen polygons, IDW, and Kriging should be made based on the dataset characteristics and research goals.

4.3 Justification of Most Appropriate Method

Based on the extensive evaluation of spatial interpolation methods in our study, the Ordinary Kriging method was chosen. This decision is strongly rooted in the compelling findings derived from our results. Ordinary Kriging, with its variogram modeling approach, consistently demonstrated superior performance in capturing spatial autocorrelation and providing accurate predictions of temperature variations, particularly in regions with complex elevation profiles. For instance, when inspecting the map created for Kriging, there is a clear image of the elevation profile, where the blue area with consistent low values of temperature indicates a valley at lower elevation as compared to the higher-temperature clustered areas on the periphery, which indicate higher elevations (Figure 10). In addition, our research revealed that the Kriging model, characterized by an exponential variogram model with specific parameter values, had fewer drawbacks for this particular dataset than the other methods. These results, backed by data preparation, model parameter optimization, and comprehensive evaluation, have conclusively established Ordinary Kriging as the most suitable method for our temperature surface dataset estimation.

4.4 Description of Findings in Context to Literature

The results from this research have shown than our findings align with and provide further support to the existing literature on spatial interpolation methods. Notably, the research by Jain and Flannigan (2017) emphasizes the effective performance of geostatistical methods, particularly Kriging, in learning about regions with complex elevation profiles. This is consistent with our decision to favor Ordinary Kriging as the most effective method for accurately predicting temperature variations in the study area. The study conducted by Li and Heap (2013) underlines the challenges in selecting an appropriate interpolation method due to various factors affecting their performance, including data quality and correlation between variables. These insights echo our emphasis on the importance of rigorous parameter optimization and understanding the characteristics of the dataset and study area. Furthermore, the work by Foehn et al. (2018) highlights the relevance of combining different data sources and interpolation techniques, a concept that aligns with our systematic approach to evaluating various methods and their suitability for complex elevation profiles. Additionally, the study by Manies and Mladenoff (2000) demonstrates that interpolation methods are effective in estimating relative landscape composition, which is consistent with our focus on accurately predicting temperature variations in regions with differing elevations.Thus, build upon the literature, our findings suggest that Ordinary Kriging is the optimal method for temperature surface estimation in regions with complex elevation profiles.

5 Refernces

Jain, P., & Flannigan, M. D. (2017). Comparison of methods for spatial interpolation of fire weather in Alberta, Canada. *Journal Name*, *Volume*(Issue), Page numbers. DOI or URL

Li, J., & Heap, A. D. (2013). Spatial interpolation methods applied in the environmental sciences: A review.

Foehn, A., García Hernández, J., Schaefli, B., & De Cesare, G. (2018). Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments.

Manies, K. L., & Mladenoff, D. J. (2000). Testing methods to produce landscape-scale presettlement vegetation maps from the U.S. public land survey records.

Chen, Y. N., Shen, S. L., Jin, L., & Chen, Y. P. (2011). Spatial interpolation techniques: Their applications in regionalizing climate-change series and associated accuracy evaluation in Northeast China. International Journal of Climatology, 31(4), 501-514. https://doi.org/10.1002/joc.2099

Irmak, A., Ranade, P. K., Marx, D., Irmak, S., Hubbard, K. G., Meyer, G. E., & Martin, D. L. (2010). Spatial interpolation of climate variables in Nebraska. Source, 53(6), 1759-1771.

Dhamodaran, S., & Lakshmi, M. (2021). Comparative analysis of spatial interpolation with climatic changes using inverse distance method. Source, 12(6), 6725-6734. https://doi.org/10.1007/s12652-020-02296-1

Li, Q. H., Chen, Y. N., Shen, Y. J., Li, X. G., & Xu, J. H. (2011). Spatial and temporal trends of climate change in Xinjiang, China. Source, 21(6), 1007-1018. https://doi.org/10.1007/s11442-011-0896-8

Granville, K., Woolford, D. G., Dean, C. B., Boychuk, D., & McFayden, C. B. (2022). On the selection of an interpolation method with an application to the Fire Weather Index in Ontario, Canada.

Hussain, I., Faisal, M., Shad, M. Y., Hussain, T., & Ahmed, S. (2015). Assessment of spatial models for interpolation of elevation in Pakistan. International Journal of Geoinformatics, 7(3), 409-422. https://doi.org/10.1504/IJGW.2015.069371

Palmer, D.J., Hock, B.K., Kimberley, M.O., Watt, M.S., Lowe, D.J., & Payn, T.W. (2009). Comparison of spatial prediction techniques for developing Pinus radiata productivity surfaces across New Zealand. *Forest Ecology and Management, 258*(12), 2589-2600.