Acta Sci. Pol., Administratio Locorum 24(3) 2025, 449–468.
https://czasopisma.uwm.edu.pl/index.php/aspal pISSN 1644-0749 eISSN 2450-0771 DOI: 10.31648/aspal.11055

ORIGINAL PAPER
Received: 26.01.2025 Accepted: 13.05.2025

GEOSTATISTICAL METHODS OF MAP PREPARATION ON THE EXAMPLE OF TEMPERATURE DISTRIBUTION IN AFRICA

Patrycja Tarnowska[1]✉, Marek Ogryzek[2]✉, Eufemia Tarantino[3]✉

[1] ORCID: 0009-0002-8352-0281
[2] ORCID: 0000-0002-7493-5488
[3] ORCID: 0000-0002-2468-0771
[1,2] University of Warmia and Mazury in Olsztyn, Faculty of Geoengineering, Institute of Spatial Management and Geography, Department of Socio-Economic Geography
Prawochenskiego Street, 15, 10-720 Olsztyn, **Poland**
[3] Polytechnic University of Bari, Department of Civil, Environmental, Land, Building Engineering and Chemistry
70-125 Bari, **Italy**

ABSTRACT

The main aim of the study was to compare the different geostatistical methods used to map the spatial distribution of temperature and to select the optimal method using the example of maximum temperatures in Africa. An exploratory data analysis was carried out, including statistical tests, distribution verification, and identification of outliers. Maps were produced in ArcGIS Pro using deterministic (IDW, LPI) and stochastic (Kriging Ordinary, Kriging Simple, Kriging Universal, EBK) methods with different parameters. Validation was then carried out using the GIS Data Modelling Validator software, analysing the interpolation error parameters. The validation made it possible to compare the different methods and select the most effective one. Prediction, probability and error maps were developed for the selected method. The analysis carried out showed that the appropriate selection of the interpolation method is crucial for the accuracy and usability of the resulting maps. Several key research questions were posed to identify the optimal geostatistical method for spatial mapping of temperature distribution, using data from Africa as a case study. The effectiveness of various methods in this process was investigated, along with the parameters during the data modelling phase that have the greatest impact on the accuracy of temperature predictions. Additionally, an attempt was made to determine to what extent the presence of outliers influences interpolation results. A critical issue was also to establish which method provides the highest predictability and concordance with actual temperature measurements.

Keywords: geostatistics, validation, kriging, GIS, Walidator, MPQE

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

INTRODUCTION

Geographic Information System (GIS) is a fundamental component in geostatistics, facilitating spatial data analysis and providing critical support for selecting optimal interpolation methods. GIS enables the efficient processing of geospatial datasets, such as temperature measurements from diverse locations, which can then be rigorously analysed and visualised. In the realm of interpolation method selection, GIS offers advanced capabilities for evaluating data distribution, detecting spatial trends, and analysing spatial dependencies between measurement points. This comprehensive understanding of spatial data characteristics allows for the precise alignment of interpolation techniques with the underlying spatial patterns. Moreover, GIS affords robust visualisation tools, enabling the generation of interpolation maps that enhance the interpretation and communication of analytical results (Urbański, 2008). An integral feature of GIS is its standardised approach to spatial data identification, which ensures seamless integration with datasets from various systems (Gaździcki, 1990; Izdebski 2015; Maguire, 1991).

Geostatistics is a branch of statistics dedicated to the analysis and modelling of spatial data, accounting for its geographical distribution and the dependencies between values at different points in space (Matheron, 1962; 1965). Over the following decades, geostatistics found widespread application both in theory and in empirical research across disciplines such as geology, hydrology, and meteorology. Despite its extensive practical applications, geostatistics continues to evolve and expand its theoretical framework (Royle, 1979). Geostatistics is distinguished primarily by its consideration of spatial dependencies between data points, in contrast to classical statistics, which typically assumes the independence of observations. The innovation of geostatistics lies in treating the analysed parameter as a "regionalised variable", whose values are dependent on the spatial coordinates of the measurement points. The First Law of Geography, articulated by Tobler (1970), asserts that "everything is related to everything else, but near things are more related than distant things", also referred to as the First Law of Spatial Analysis (Ogryzek, 2018). Geostatistics adopts this principle, assuming that the values of a phenomenon at nearby locations are more similar than those at distant points. This property, known as spatial autocorrelation, is essential for understanding and modelling many natural and social phenomena (Marmol, 2002). Geostatistical methods constitute a fundamental toolkit for spatial data analysis, enabling precise modelling and interpretation of the spatial variability of geographic data. These methods are primarily applied to spatially correlated data (Cellmer, 2014).

Geostatistics can be used to study the distribution of temperatures (Vicente-Serrano et al., 2003) and maximum temperatures (Davison & Gholamrezaee, 2012). Geostatistical modelling methods are particularly useful for modelling temperature in regions with complex topography and limited availability of meteorological data (Ahmed et al., 2014), as well as in mountainous regions (Raquel et al., 2007). Research has confirmed the achievement of more detailed local forecasts of climatic variables using geostatistical methods (Jha et al., 2013; Jha et al., 2015) and also enables the identification of climate types (Alvares et al., 2013). The Kriging method has also proven effective in predicting temperature distribution in greenhouse environments (Bojacá et al., 2009). The challenge still lies in modelling over time (Pickens et al., 2021). The land surface temperature is an important variable in global climate change because it can prevent the deterioration of air quality in environments using preventive measures (Taoufik et al., 2021). Temperature changes affect the decision to abandon their fields, and consequently, thousands of hectares of land that were previously green areas have turned into deserts (Morsy & Ahmed, 2023).

The goal of this study was to determine the optimal geostatistical interpolation technique analysing mean maximum temperature data across Africa. For this purpose, two hypotheses were formulated, defined as H_1 and H_2. H_1: Stochastic methods generate smaller interpolation errors than deterministic methods; and H_2: Exploratory Data Analysis (EDA)

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

is an indispensable stage in the geostatistical modelling process, as it enables the identification of spatial patterns, data distribution characteristics, and potential anomalies that may influence model accuracy.

Several key research questions were also posed:
- Which methods are the most effective in this process?
- What parameters during the data modelling phase have the greatest impact on the accuracy of temperature predictions?
- To what extent does EDA facilitate the modelling process?
- What methods, techniques, or tools can be applied to identify areas with the highest predictability and agreement with actual temperature measurements?

MATERIALS AND METHODS

The primary objective of this research was to conduct a detailed comparison of various methods for developing maps of mean maximum temperature in the African region using advanced geostatistical techniques. To achieve this, the focus was placed on utilising the geostatistical methods available within the Geostatistical Wizard, which is a component of the ArcGIS Pro software. As part of the study, cross-validation and subset validation were performed to obtain comparative parameters for assessing the model's fit to actual measurements, using the GIS Modelled Data Validator tool. The GIS modelling validator analyses and extracts parameters for comparison through an algorithm (Ogryzek et al., 2020), allowing for the identification of the technique that can be considered optimal, as its results demonstrate the highest accuracy in representing the actual temperature distribution in the African region. The research procedure followed the workflow illustrated in Fig. 1.

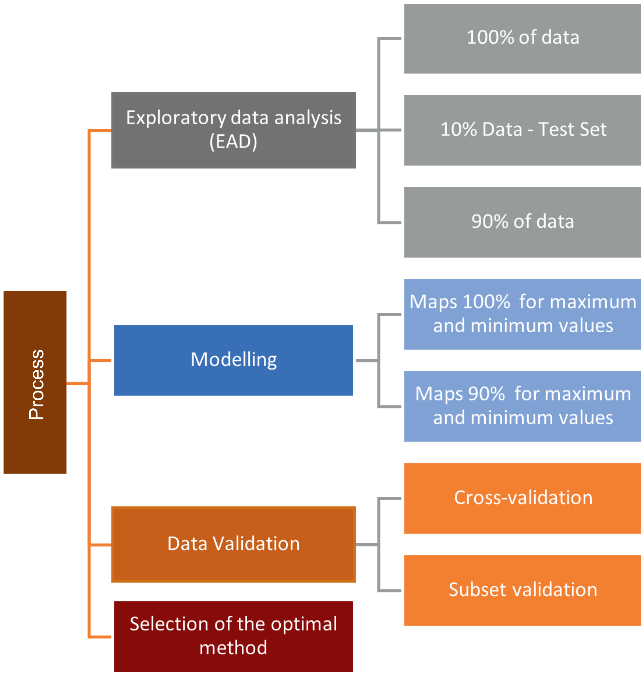


Fig. 1. Research procedure
Source: own elaboration based on Ogryzek (2018).

Geostatistics employs both mathematical and statistical approaches to data analysis, considering both local and global variability. Two main categories of geostatistical methods are distinguished: deterministic methods and stochastic methods. Deterministic methods are used to calculate values at raster cell centres through specific mathematical algorithms that ensure the continuity of the modelled surface. Interpolation using these methods focuses on achieving a smooth transition of values between cells and estimating the most probable (average) values. The result is surface models that are "smoothed" and may not fully capture local data variability. Although there are many deterministic methods, the focus has been on those most recommended for this type of analysis, namely, the Inverse Distance Weighting (IDW) method and Local Polynomial Interpolation (LPI) (Ogryzek et al., 2020).

Stochastic methods in geostatistics are advanced interpolation tools that, unlike deterministic methods based on strict mathematical relationships, account for the random nature of spatial variability. By utilizing statistical models and the concept of spatial autocorrelation (the relationship between variable values at different locations), these methods enable not only the estimation of the most probable value at a given location but also the quantification of the uncertainty associated with that estimation (Krige,

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

1951; 1952). Kriging employs statistical models for spatial data interpolation, based on the assumption that variable values at different points in space are correlated, with the strength of this correlation diminishing as the distance between points increases. This approach enables kriging to provide precise estimates of variable values at unknown locations by leveraging the measured values at neighbouring points, while simultaneously accounting for the spatial structure of variability. One of the key advantages of–kriging, distinguishing it from other geostatistical methods, is its ability to estimate interpolation error. This capability allows for the assessment of uncertainty in the results, which is crucial in many fields such as environmental studies, geology, and hydrology, where the accuracy of estimates is essential (Chiles & Delfiner, 1999). As noted by Marmol (2002), an advanced technique like kriging requires a detailed analysis of the input data. Lack of control over the individual stages of the geostatistical analysis can lead to obtaining misleading or erroneous results.

Kriging methods include:

– Ordinary Kriging (OK): the basic form of kriging, which assumes data stationarity and treats the mean as an unknown value.
– Simple Kriging (SK): assumes that the mean is known and constant across the entire area.
– Universal Kriging (UK): a method that accounts for a global trend in data.
– Empirical Bayesian Kriging (EBK): an advanced technique that automatically adjusts model parameters based on the data, allowing for more accurate predictions.

The Parametric Method for Estimation Quality Assessment (MPQE) is an approach used in geostatistical analysis to compare the quality of different spatial interpolation methods. Within MPQE, estimation quality parameters are determined for each interpolation method in the study area and time frame. These parameters can include various quality indicators such as mean error, standard error, standardised error, and others. Then, using linear programming, the optimal interpolation method for a specific case can be identified. In this context, "optimal" refers

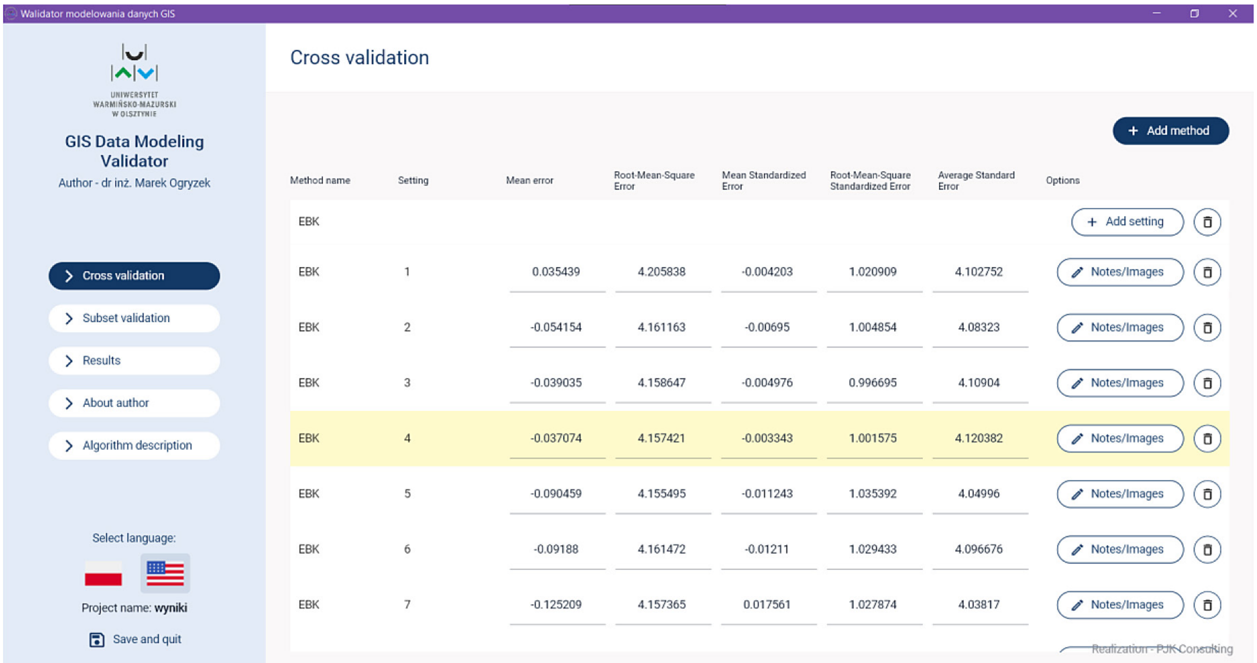


Fig. 2. Interface view of the GIS Data Modelling Validator program
Source: own elaboration based on GIS Data Modelling Validator program.

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

to the method whose Cn parameter (which defines the quality of the estimation) is closest to or equal to zero. This indicates that MPQE identifies the method that provides the most realistic estimation results. The GIS Data Modelling Validator program, developed by Ogryzek (2025), was designed (Fig. 2) to streamline the process of selecting the optimal geostatistical method for spatial data modelling. This tool enables objective evaluation of the quality of different interpolation, approximation, or extrapolation models, thereby supporting decision-making in spatial analysis.

The program utilises Parametric Estimation Quality Assessment (MPQE), which is based on comparing error parameters between actual and predicted values at measurement points (Ogryzek & Kurowska, 2016). It employs both cross-validation and subset validation and then applies an optimization algorithm to identify the model with the lowest weighted prediction error. The program analyses a range of error parameters, such as:

– Mean Error (ME),
– Root Mean Square Error (RMSE),
– Average Standard Error (ASE),
– Mean Standardized Error (MSE),
– Root Mean Square Standardized Error (RMSSE).

Different interpolation methods generate a varied number of comparative parameters. When modelling empirical semivariograms, it is crucial to align the model characteristics with the phenomenon being studied, taking into account the analysis of previously discussed parameters (Dębowska & Zawadzki, 2005; Isaaks & Srivastava, 1988). However, an ideal fit between the model and the data does not always guarantee the lowest errors, which is why result validation is essential before selecting the optimal model. In practice, choosing the best method can be challenging due to the multitude of estimation quality statistics and the lack of clear criteria for selection. The Parametric Method for Estimation Quality Assessment (MPQE) addresses this issue by employing an optimisation algorithm based on estimation parameters from cross-validation (CV) and subset validation (SV).

When selecting the optimal method, a weighting scheme is applied (Formula 1), based on Ogryzek (2018): parameters from cross-validation (CV) + parameters from subset validation (SV) for the optimal model, which has the lowest value.

$$\begin{cases} \Delta_1 a_1 + \Delta_2 b_1 + \Delta_3 c_1 + \Delta_4 d_1 + \cdots + \Delta_n n_n = C_1 \\ \Delta_{21} a_2 + \Delta_{22} b_2 + \Delta_{23} c_2 + \Delta_{24} d_2 + \cdots + \Delta_{2n} n_n = C_2 \\ \cdots \\ \Delta_{n1} a_n + \Delta_{n2} b_n + \cdots + \Delta_m n_m = C_n \end{cases} \tag{1}$$

where:

$C_1, C_2, \ldots, C_n$ – quality assessment of a geostatistical estimation method,
$\Delta_1, \Delta_2, \ldots, \Delta_n$ – weight of a parameter,
$a_1, a_2, \ldots, a_n$ – value of parameter a in the data group at 100%,
$b_1, b_2, \ldots, b_n$ – value of parameter b in the data group at 100%,
$c_1, c_2, \ldots, c_n$ – value of parameter c in the data group at 90%,
$d_1, d_2, \ldots, d_n$ – value of parameter d in the data group at 90%,
$n_1, n_2, \ldots, n_n$ – number of parameters.

The geostatistical method for which the C_n parameter is closest to or equal to 0 is considered the optimal method, with $C_n \leq 0$ = max. The method with the lowest weighted prediction error in MPQE is deemed the optimal method. Based on the weights used in C_{10}, it is recommended to apply MPQE for each model as an algorithm expressed as RMSE (KW) + 5% ME (KW) and for RMSE (WP) + 5% ME (WP).

The universal nature of the method allows it to be applied to various types of data. Once the data is entered, the program generates a PDF report containing the analysis results, summary tables, and an overview of the modelling process. The GIS Data Modelling Validator is a tool that streamlines the process of selecting the optimal geostatistical method by providing an objective assessment of model quality and facilitating decision-making in spatial analysis. Its universal nature makes it applicable to different types of data, which makes it a valuable tool for GIS professionals.

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

By using MPQE, researchers can make more informed decisions regarding the choice of interpolation methods, depending on the specifics of the area and the objective of the analysis. This method is particularly useful when multiple interpolation methods are available, and the best solution for a particular case needs to be identified.

RESULTS AND DISCUSSION

Africa is the second-largest continent on the planet, surrounded by several major bodies of water: the Mediterranean Sea to the north, the Arabian Sea and Indian Ocean to the east, and the Atlantic Ocean to the south and west. The land is largely shaped by elevated plateaus that gradually slope down toward the coastal regions, creating depressions within the interior, the largest being the Congo Basin. Mountainous areas are present in the north (Atlas Mountains), the south (Drakensberg Mountains), and the east (Ethiopian Highlands). In terms of climate, most of Africa falls within hot and very hot zones, as defined by Köppen's classification system. The highest temperatures are recorded in the Sahara, where extreme highs exceed 44°C. In contrast, the lowest maximum temperatures are found in Algeria, along the Mediterranean coast, hovering around 7°C (Makowski, 2018; Martyn, 2000).

The first phase of analysis focuses on identifying trends, irregularities, correlations and key features within the dataset under review. This stage is inherently iterative and involves the use of multiple methods – such as visual exploration, summary statistics and multivariate techniques – to better understand the organisation of the data and identify potential problems. The dataset to be analysed consists of a point layer, showed on Fig. 3, containing maximum temperature values recorded at various locations across Africa.

The Average Nearest Neighbour analysis the spatial distribution of point objects by calculating the average distance between each object and its nearest neighbour. The average nearest neighbour ratio is calculated as the observed average distance divided by the expected average distance (with expected average distance being based on a hypothetical random distribution with the same number of features covering the same total area) (Jackowski, 2004; Ogryzek & Jaskulski, 2025; Śliwicki, 2014). The result of the analysis is the Nearest Neighbour Index (NNI), which helps determine the nature of point distribution:

– NNI < 1: Indicates clustering of points, meaning a tendency to form groups or clusters.
– NNI = 1: Suggests a random distribution of points, with no apparent clustering or dispersion patterns.
– NNI > 1: Signifies point dispersion, indicating a tendency for points to be more evenly distributed across space.

The tool also outputs z-scores and p-values, which help assess whether the results are statistically significant. A p-value lower than a chosen threshold (typically 0.05) signifies that the observed pattern significantly deviates from randomness. As the tool executes, relevant statistical results are shown at the bottom of the Geoprocessing panel and can be monitored live or accessed later via the Geoprocessing History. Additionally, the tool can produce an HTML report summarizing the findings visually, which can be opened using the file path mentioned in the output messages.

To assess the spatial pattern of the dataset, the Nearest Neighbour Index (NNI) was applied. This metric is calculated as the ratio between the observed average distance among points and the expected distance assuming a random spatial distribution. In this analysis (Fig. 4), the test was performed under the assumption of a null hypothesis that the points are randomly distributed. The aim was to determine whether the data display a non-random spatial structure, such as clustering or regular spacing. An NNI value below 1 indicates a clustering tendency, while a value above 1 points to a more dispersed or competitive arrangement.

In the analysed dataset, the NNI was found to be less than 1, pointing to a clear clustering pattern. This result supports the rejection of the null hypothesis and indicates that the data are not randomly distributed, but instead form statistically significant clusters.

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

Fig. 3. Map of the distribution of measurement points processed in ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

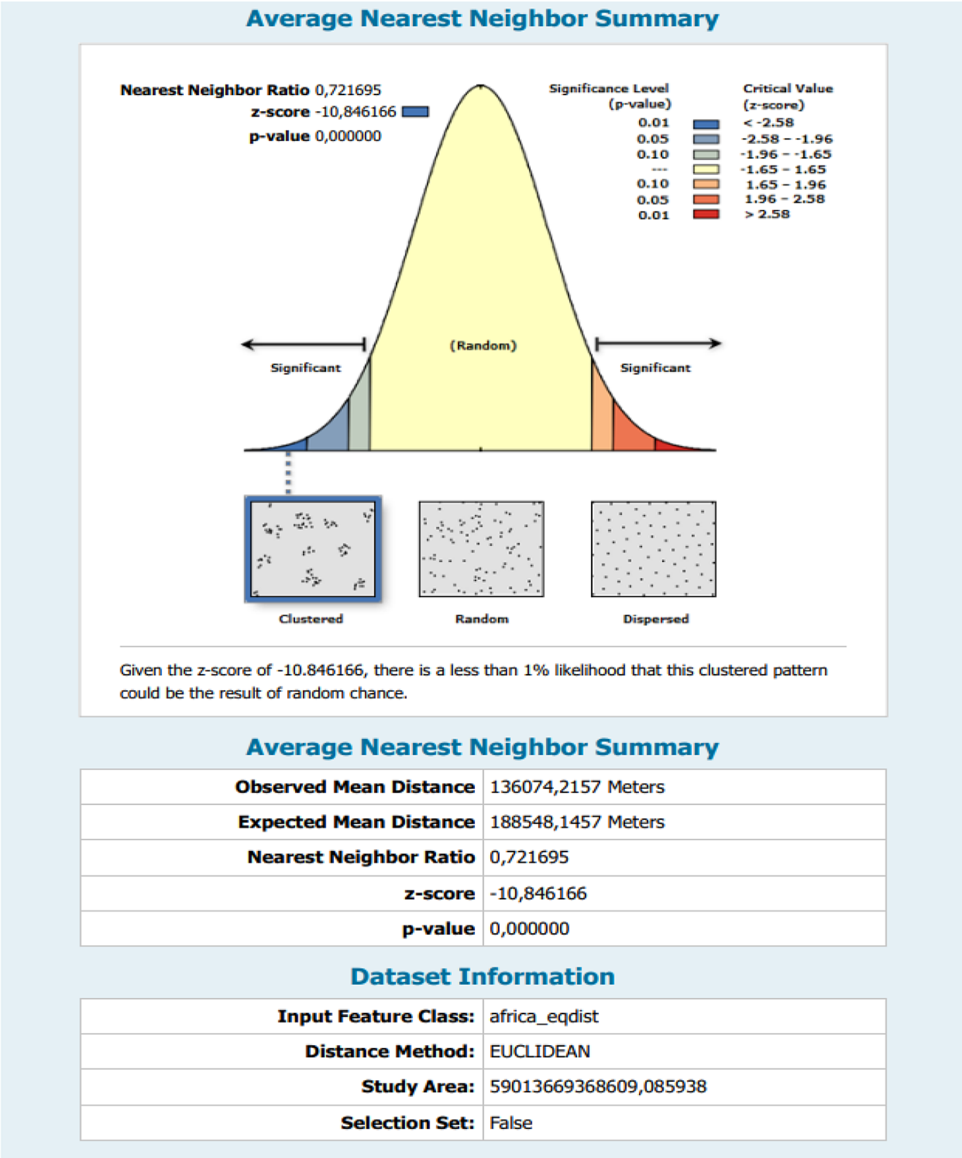


Fig. 4. Report from the Average Nearest Neighbour Test in ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

The Spatial Autocorrelation (Global Moran's I) tool evaluates the overall spatial autocorrelation in a dataset by simultaneously considering both the locations of features and their associated attribute values. The analysis generates the Moran's I statistics, which quantifies the degree of spatial correlation, ranging from -1 (indicating negative spatial autocorrelation) to 1 (indicating positive spatial autocorrelation). Interpretation of results:

– Moran's I < 0: The attribute values of neighbouring objects differ from each other more than expected under a random distribution (dispersion).
– Moran's I = 0: The attribute values of neighbouring objects are independent of each other (random distribution).
– Moran's I > 0: The attribute values of neighbouring objects are more similar to each other than expected

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

under a random distribution (clustering) (Anggani et al., 2023; Ogryzek & Jaskulski, 2025).

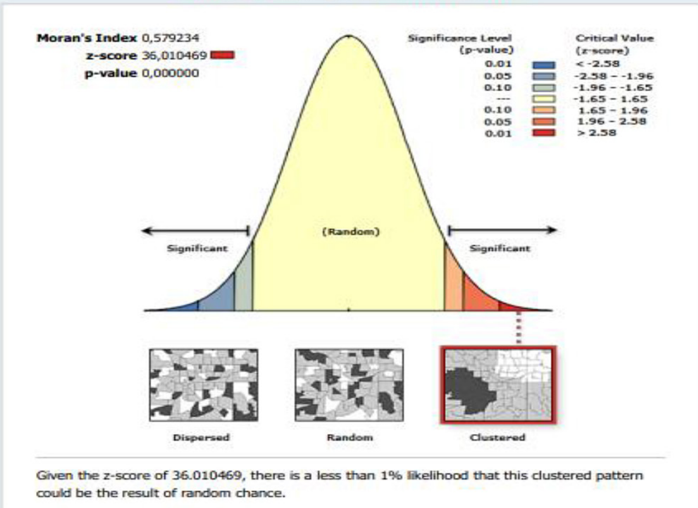
To assess the statistical significance of the result, the tool also provides a z-score and a p-value. The p-value represents the probability that the observed pattern could have occurred by chance, serving as an approximation based on the test statistic and a known probability distribution.

As with the Average Nearest Neighbour tool, this analysis provides real-time messages showing the computed values and offers the option to generate an HTML report that includes a visual summary of the results.

Both tools are useful for spatial analysis but differ in what they measure:
– Average Nearest Neighbour: Assesses the spatial distribution of objects (points) regardless of their attributes.
– Spatial Autocorrelation (Global Moran's I): Assesses the spatial distribution of object attributes, considering their mutual relationships.

In the next stage of the analysis, spatial autocorrelation was assessed using the Global Moran's I statistic (Fig. 5). This method assesses whether the spatial arrangement of the attribute values associated

Spatial Autocorrelation Report



Global Moran's I Summary

Moran's Index	0,579234
Expected Index	-0,002415
Variance	0,000261
z-score	36,010469
p-value	0,000000

Dataset Information

Input Feature Class:	africa_eqdist
Input Field:	MAX_TEMP
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	True
Distance Threshold:	1259034,7735 Meters
Weights Matrix File:	None
Selection Set:	False

Fig. 5. Report from the Moran's I Test using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

with the features exhibits a clustered, dispersed or random pattern. The Moran's I test is useful for detecting spatial dependencies between data points. The z-score and p-value are key indicators of statistical significance when interpreting the results. A significantly positive Moran's I value indicates clustering, while a significantly negative value indicates a dispersed distribution. In this case, the Moran's I value was close to 1, indicating a strong clustering pattern. These results justify rejecting the null hypothesis of a random spatial distribution.

The statistical significance shown by both the z-score and the Moran's I index confirms the presence of spatial relationships in the dataset – indicating that the observations are spatially structured rather than randomly distributed.

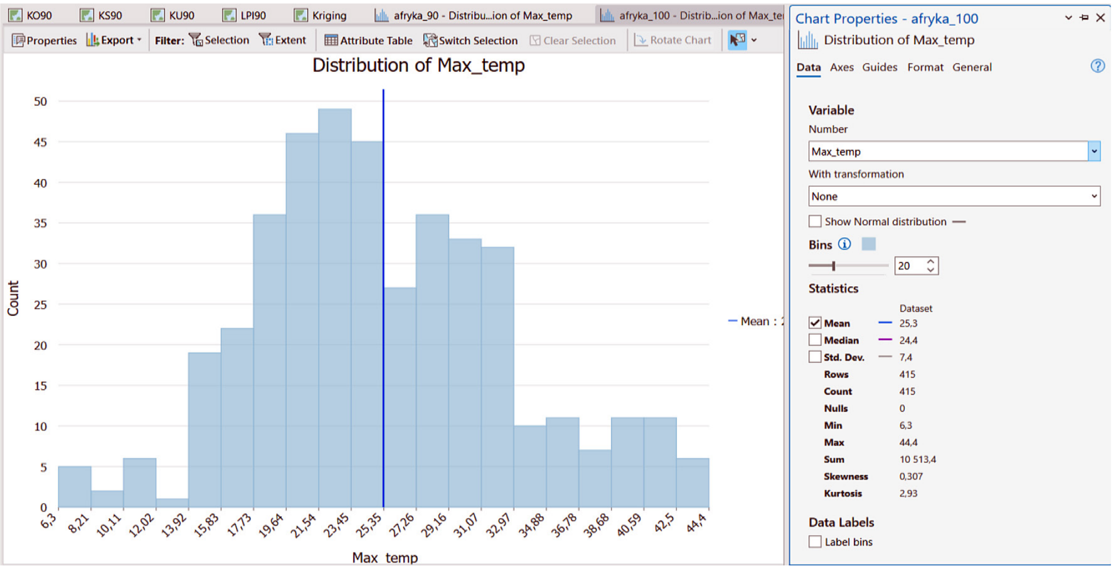


Fig. 6. Histogram of the data distribution for the full dataset using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

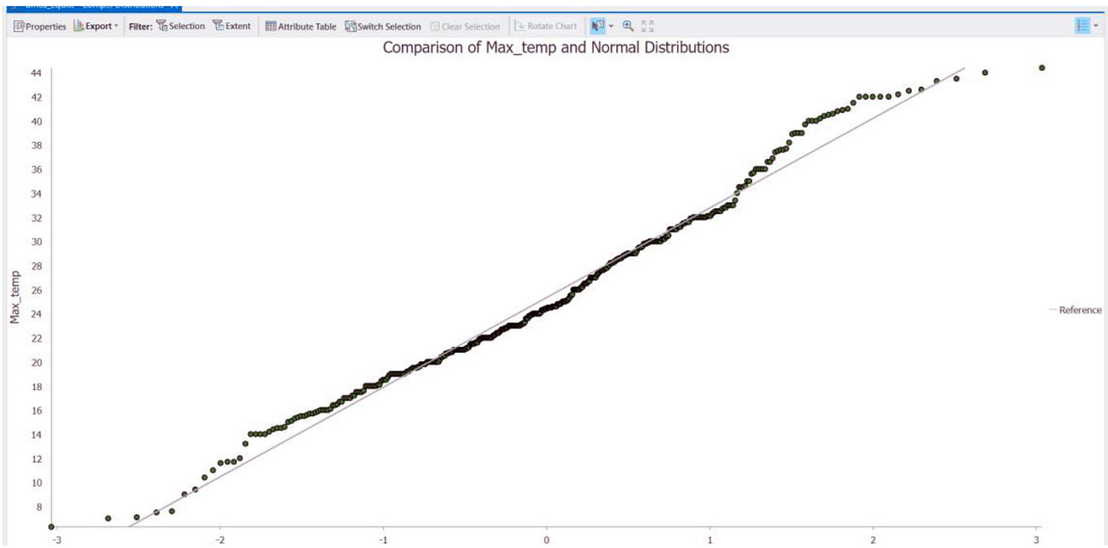


Fig. 7. QQ Plot of the data distribution for the full dataset using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

Checking the distribution of data is an important aspect of geostatistical analysis, particularly as many methods, such as kriging, are based on the assumption of normality. If this assumption is not met, it may be necessary to transform the data or use alternative interpolation techniques. For a data set to be considered normally distributed, the skewness should ideally be close to 0, indicating symmetry around the mean, while the kurtosis should be around 3, indicating a similar peak sharpness to that of a normal distribution. In this analysis, a histogram (Fig. 6) and a QQ plot (Fig. 7) were used to assess the nature of the data distribution. The full data set showed a slight positive skewness of 0.307 and a kurtosis of 2.93. Although no transformation was applied, the visual and statistical evaluations suggested that the data were approximately normally distributed. Several transformations – logarithmic, square root, inverse and Box-Cox – were considered and the results are summarised in Table 1. The Box-Cox transformation performed similarly to the untransformed data. While such a method can improve the normality of a distribution, its use may be of limited benefit in cases where the data are already close to normality.

Table 1. List of parameters

Transformation Method	Skewness	Kurtosis
No Transformation	0.307	2.93
Logarithmic	-0.82	4.8
Square Root	-0.177	3.3
Inverse	2.86	16.0
Box-Cox	0.307	2.93

Source: own elaboration based on Tarnowska (2024).

As a result, the original data values were retained as the exploratory analysis indicated no need for transformation and the final results were unaffected.

Next, the full dataset was split into a test set (10%) and a training set (90%). The Subset Feature tool was used to extract 10% of the data for the test set. Then, by inverting the selection, the remaining 90% of the data was extracted as the training set.

Similar to the complete dataset, both a histogram (Fig. 8) and a QQ plot (Fig. 9) were generated for the training subset (comprising 90% of the data) to assess its distribution. The training data showed a slight positive skewness of 0.26 and a kurtosis of 2.9. Various transformation techniques – including logarithmic, square root, inverse and Box-Cox – were evaluated, with the results summarised in Table 2. The Box-Cox transformation produced results almost identical to those from the untransformed data, suggesting that it may improve the approximation to normality. However, as the dataset was already close to a normal distribution, it was considered unnecessary to apply this transformation. The decision to retain the original values was supported by the exploratory analysis, which revealed no compelling reason for transformation and confirmed that this decision did not compromise the results of the subsequent analysis.

Table 2. List of parameters

Transformation Method	Skewness	Kurtosis
No Transformation	0.26	2.9
Logarithmic	-0.88	4.9
Square Root	-0.226	3.3
Inverse	2.9	16.2
Box-Cox	0.26	2.9

Source: own elaboration based on Tarnowska (2024).

A t-test was performed to evaluate whether there were statistically significant differences between the full dataset (100%) and the training subset (90%). The findings, summarised in Table 3, show that the differences in skewness and kurtosis between the two datasets are not statistically significant, with p-values of 0.85 and 0.99, respectively. These results indicate that the training set is a reliable representation of the complete dataset, thereby validating its use in subsequent analyses and for constructing geostatistical models.

Table 3. Parameters skewness and kurtosis

Parameters	100%	90%	p-value
Skewness	0.31	0.26	0.85
Kurtosis	2.93	2.9	0.99

Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

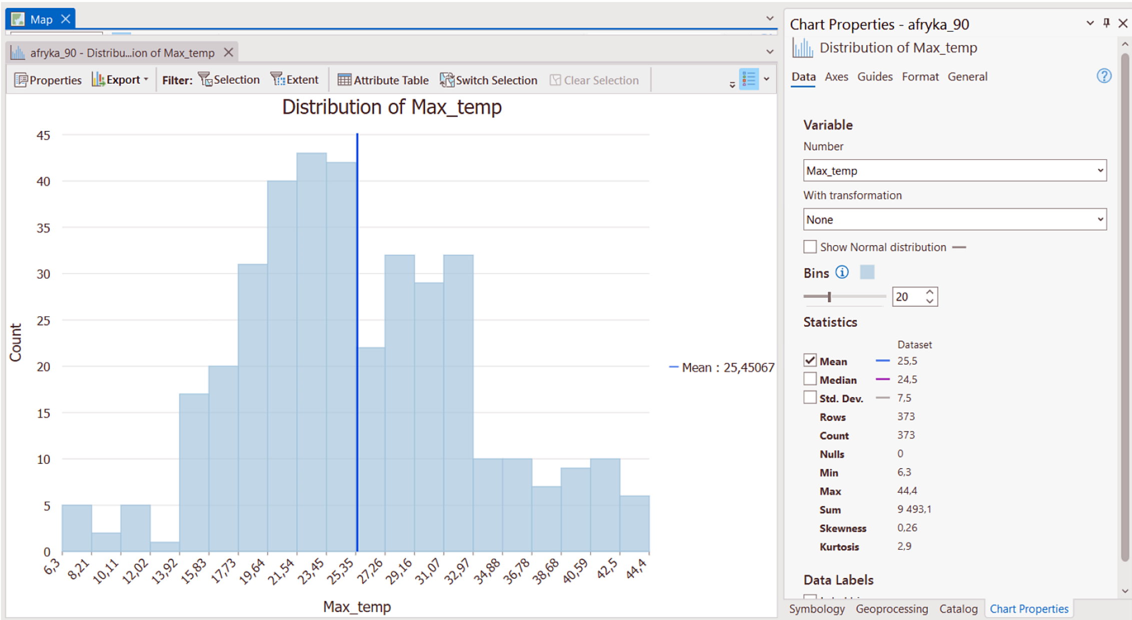


Fig. 8. Histogram of the data distribution for the training set (90% of the data) using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

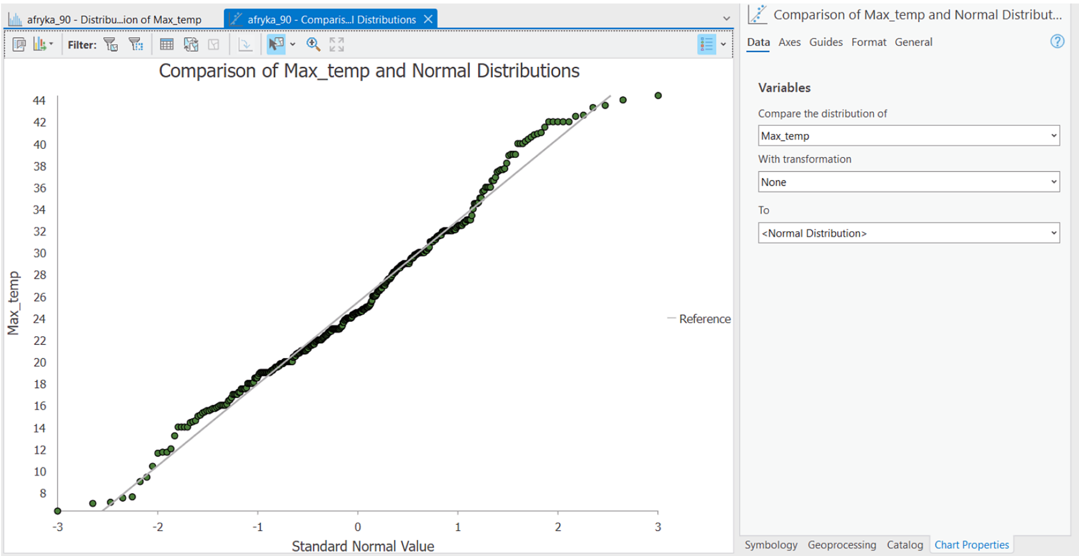


Fig. 9. QQ Plot of the data distribution for the training set (90% of the data) using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

The next stage of the analysis was to use the Create Thiessen Polygons tool. This method is widely used in spatial analysis to delineate zones of influence around individual points. Each resulting polygon – referred to as a Thiessen or Voronoi polygon – encompasses an area in which all locations are closer to the associated input point than to any other point in the dataset. The edges of these polygons are formed by the perpendicular bisectors of the lines connecting adjacent input points.

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

Thiessen polygons facilitate the interpretation and evaluation of spatial point distributions by delineating individual zones of influence. Larger polygons can suggest areas of lower point density or highlight points of potentially greater spatial significance. Analysis of these polygons can provide important insights into the spatial configuration of the dataset and support informed decision making in a variety of areas. In the next stage of the analysis, a Voronoi diagram (Fig. 10) was created using the Create Thiessen Polygons tool available in ArcGIS Pro.

The analysis revealed no significant outliers or anomalies in the dataset, indicating a uniform spatial distribution of values. This supports the reliability and consistency of the data, which is essential for the geostatistical procedures that follow.

Interpolated maps of maximum temperature across Africa were then produced using a range of geostatistical and deterministic techniques. The mapping process was carried out for both the complete dataset (100%) and a 90% training subset. For each method used, multiple maps were produced using different parameter configurations available in the Geostatistical Wizard, such as varying the number of sectors or adjusting the model types. A total of 84 maps were generated (42 for each dataset), providing a broad basis for evaluating the impact of different interpolation settings. The methods used in the analysis included:
– Empirical Bayesian Kriging (EBK),
– Local Polynomial Interpolation (LPI),
– Inverse Distance Weighting (IDW),
– Kriging Simple (KS),
– Kriging Oridinary (KO),
– Kriging Universal (KU).

For each of the maps, several indicators were obtained and considered when comparing the results of spatial analyses related to property values. These indicators are:
– Mean Error (ME),
– Root Mean Square Error (RMSE),
– Average Standard Error (ASE),
– Mean Standardized Error (MSE),
– Root Mean Square Standardized Error (RMSSE).

To perform a parametric evaluation of estimation quality, both cross-validation (CV) and subset validation (SV). These techniques allowed for the assessment of estimation quality using both validation methods. Indicators such as RMSE (Root Mean Square Error) and ME (Mean Error) represent the statistical error characteristics of the interpolation for each examined interpolation model. The MPOJE value was calculated based on the algorithm provided in the GIS Data Modelling Validator, as detailed in the chapter dedicated to this program. This procedure takes into account various aspects of estimation error and enables a more comprehensive assessment of interpolation methods' quality. The summary results are presented below in Fig. 11.

The summary results of the analysis highlight the effectiveness of different interpolation methods depending on the validation technique used and the characteristics of the dataset. In the case of cross-validation, Empirical Bayesian Kriging (EBK) was found to be the most effective method. The optimal configuration for EBK included no data transformation and division into four sectors at 45 degree intervals. For subset validation, Simple Kriging with a Gaussian model proved to be the most accurate and maintained its superiority throughout the analysis. This method provided the most satisfactory results when applied with the same sector configuration and Gaussian model. In addition, prediction, probability and error maps were generated for this method, providing further insight into the spatial distribution and quality of the estimates.

The results of the analysis support hypothesis H_1, which states that stochastic interpolation methods yield lower estimation errors compared to deterministic methods. Both EBK and Simple Kriging with a Gaussian model significantly outperformed deterministic techniques such as Inverse Distance Weighting (IDW) and Local Polynomial Interpolation (LPI) in terms of error reduction. Simple Kriging produced the best results in subset validation, while EBK was optimal in cross-validation, confirming the greater accuracy and precision of stochastic methods.

It is important to emphasize that the selection of an optimal interpolation method depends not only

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

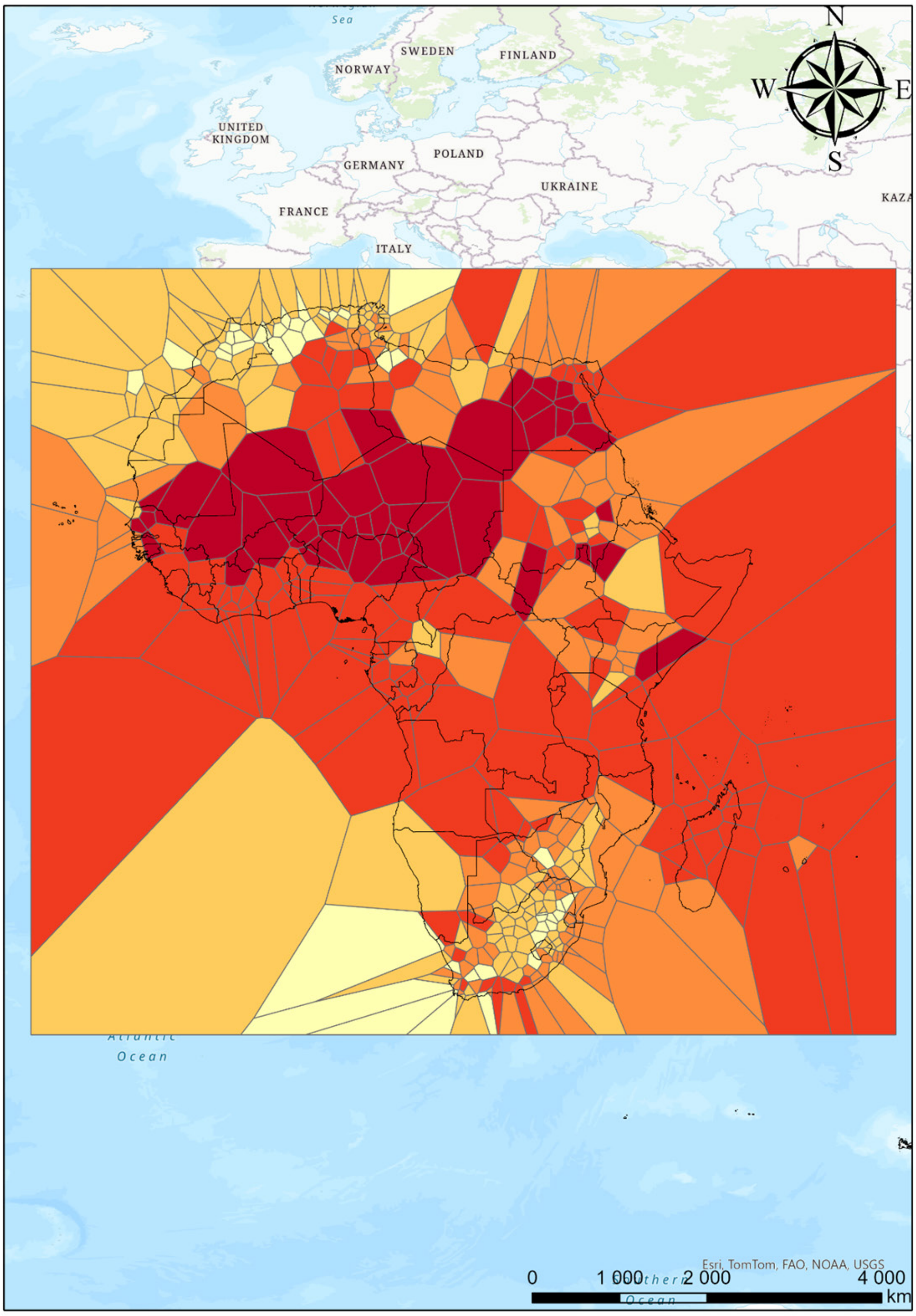


Fig. 10. Map of the division into Thiessen polygons using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

Method name	Cross validation			Subset validation			MPQE
	ME	RMSE	MPQE	ME	RMSE	MPQE	
EBK	-0.037074	4.157421	2.911024	0.596648	4.260149	3.893799	3.402412
IDW	-0.223174	4.326309	3.915996	0.376879	4.463631	4.054956	3.985476
KO	0.024852	4.225503	2.956493	0.646132	4.060992	3.719506	3.337999
KS	0.046040	4.220315	2.952767	0.719749	3.923480	3.603107	3.277937
KU	0.024852	4.225503	2.956492	0.634431	4.062885	3.720039	3.338266
LPI	-0.150326	4.529484	3.195611	0.197271	4.223036	3.820459	3.508035

Fig. 11. Summary result using the GIS Data Modelling Validator

on the technique itself but also on the parameters applied during interpolation and the specific characteristics of the spatial dataset. Therefore, the final selection should result from a holistic evaluation encompassing exploratory analysis, model calibration, and validation. Based on this comprehensive approach, Simple Kriging with a Gaussian model was identified as the most suitable geostatistical interpolation method. The outcomes of this analysis served as the basis for creating three final maps illustrating the spatial distribution of maximum temperature across Africa.

The prediction map (Fig. 12) shows the spatial distribution of predicted maximum temperatures

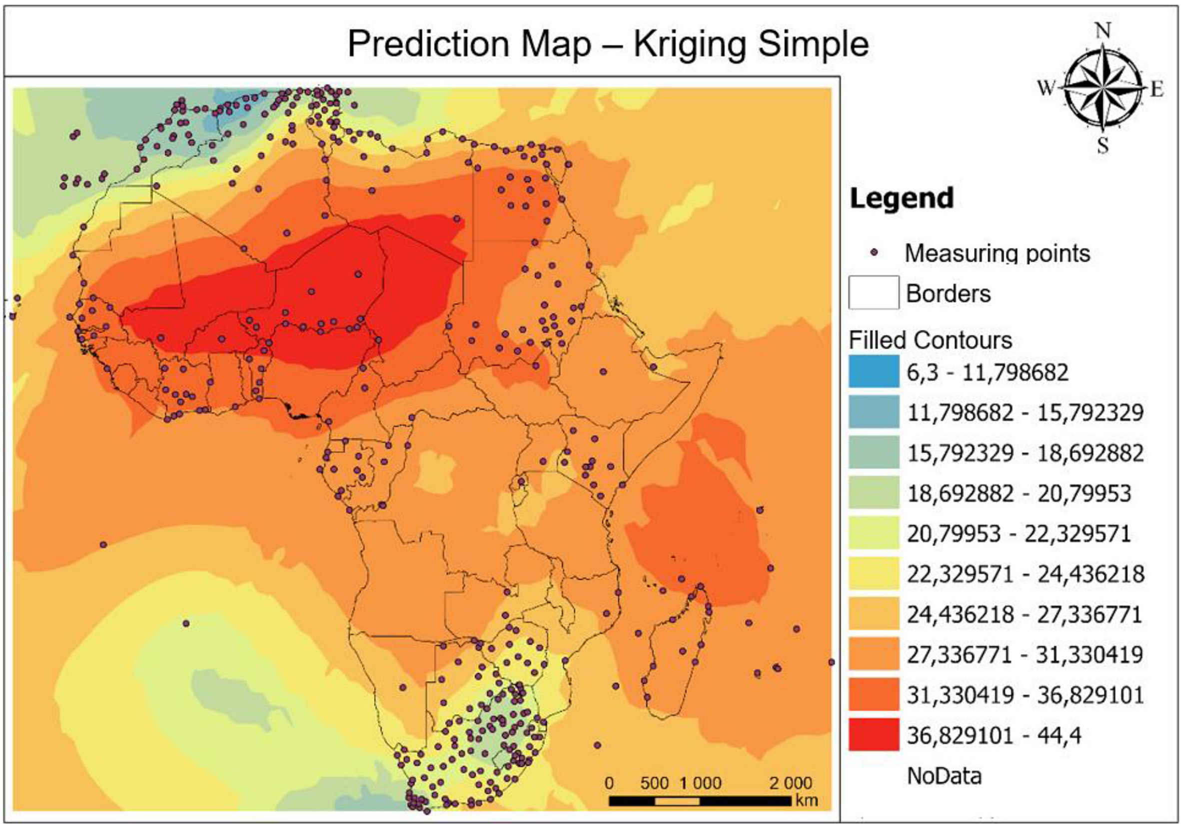


Fig. 12. Prediction map for the optimal geostatistical method using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

across Africa, generated using the Simple Kriging geostatistical interpolation method. This map is based on the available measured data and provides a comprehensive visualisation of temperature variations, supporting the identification of spatial patterns and climatic trends.

The resulting map shows considerable spatial variability in maximum temperatures. The highest temperatures, above 36.8°C, are mainly concentrated in the Sahara, reflecting the extreme aridity that characterises this region. Elevated temperatures are also observed in other parts of the continent, but their intensity is generally lower than in the Sahara. In contrast, the lowest maximum temperatures, around 7°C, are recorded in the northernmost regions of Africa, particularly along the Mediterranean coast of Tunisia. Cooler temperature zones are also found in the southern part of the continent, particularly in the territory of South Africa.

The probability map (Fig. 13) illustrates the probability that the actual maximum temperature at a given location falls within the range estimated by the Simple Kriging method. Higher probability values correspond to greater confidence in the predicted results, indicating regions where the interpolation model has greater reliability and accuracy.

Regions marked in red with probability values close to 1 indicate a high possibility that the interpolation method was applied accurately. Large areas in Central Africa and Sub-Saharan Africa have high probability values, indicating that these regions had more complete and reliable measurement data, which positively influenced the performance of the interpolation model. In particular, the area

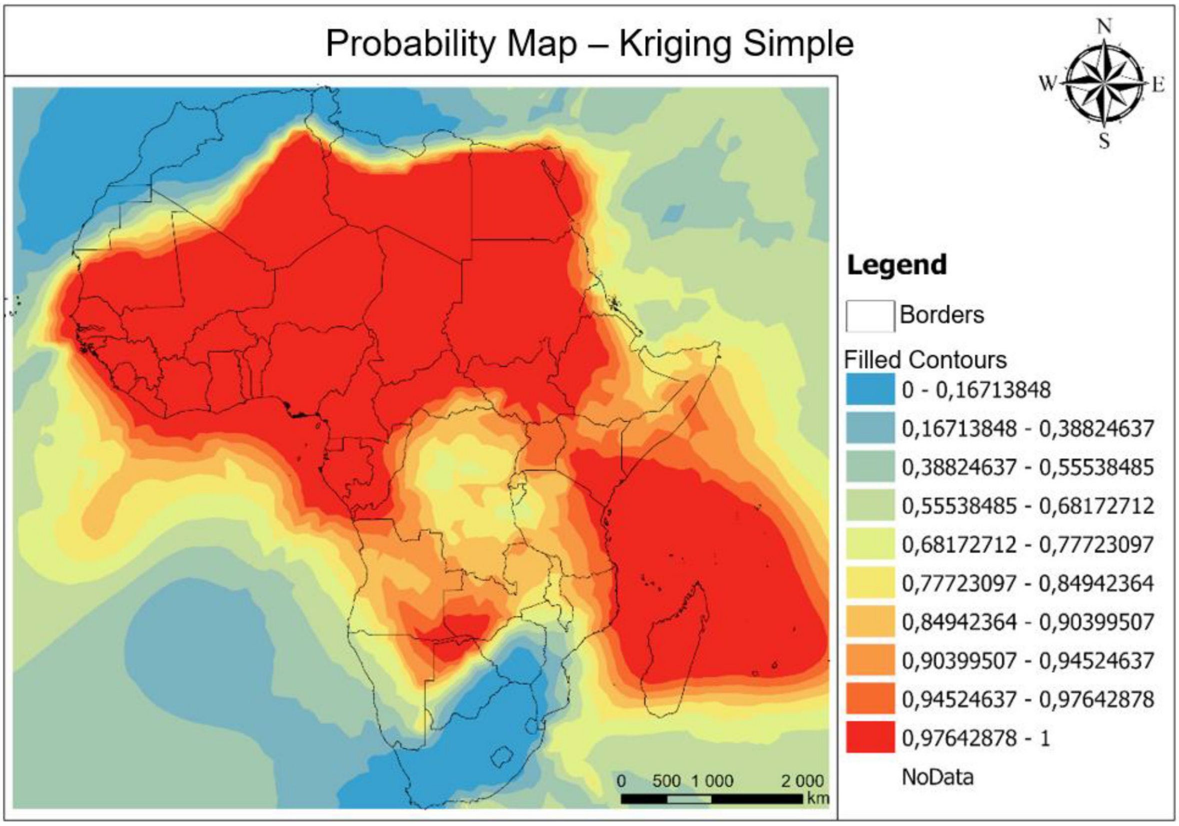


Fig. 13. Probability map for the optimal geostatistical method using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

patrycja.tarnowska11@gmail.com, marek.ogryzek@uwm.edu.pl, eufemia.tarantino@poliba.it

within the Democratic Republic of Congo has a medium probability zone (shown in yellow) where the probability values are between 70–80%, despite the near absence of direct measurement points. This pattern may be due to the ability of the model to incorporate spatial information from neighbouring areas where data availability is higher.

Conversely, regions marked in blue with probability values close to 0 are located at the extremes of the continent – notably in the north near the Mediterranean Sea and in southern Africa. In these areas, despite the presence of some monitoring stations, the data may be less representative or contain inaccuracies. Possible contributing factors include limited coverage by meteorological stations, logistical challenges in data collection or local climatic variability.

The error map (Fig. 14) shows the spatial distribution of prediction errors, highlighting the discrepancies between predicted and actual maximum temperatures across Africa. The prediction error serves as a key metric in assessing the accuracy of the interpolation model. Ideally, the error should be minimised, as lower values reflect greater model precision and a greater ability to accurately reproduce observed temperature.

The regions with the lowest prediction error (represented by the lightest colours) are concentrated in the central part of the continent, as well as in selected areas in northern and southern Africa. These low error values reflect the high precision of the Simple Kriging model, indicating close agreement between predicted and observed temperature values. Areas of medium prediction error, generally surrounding

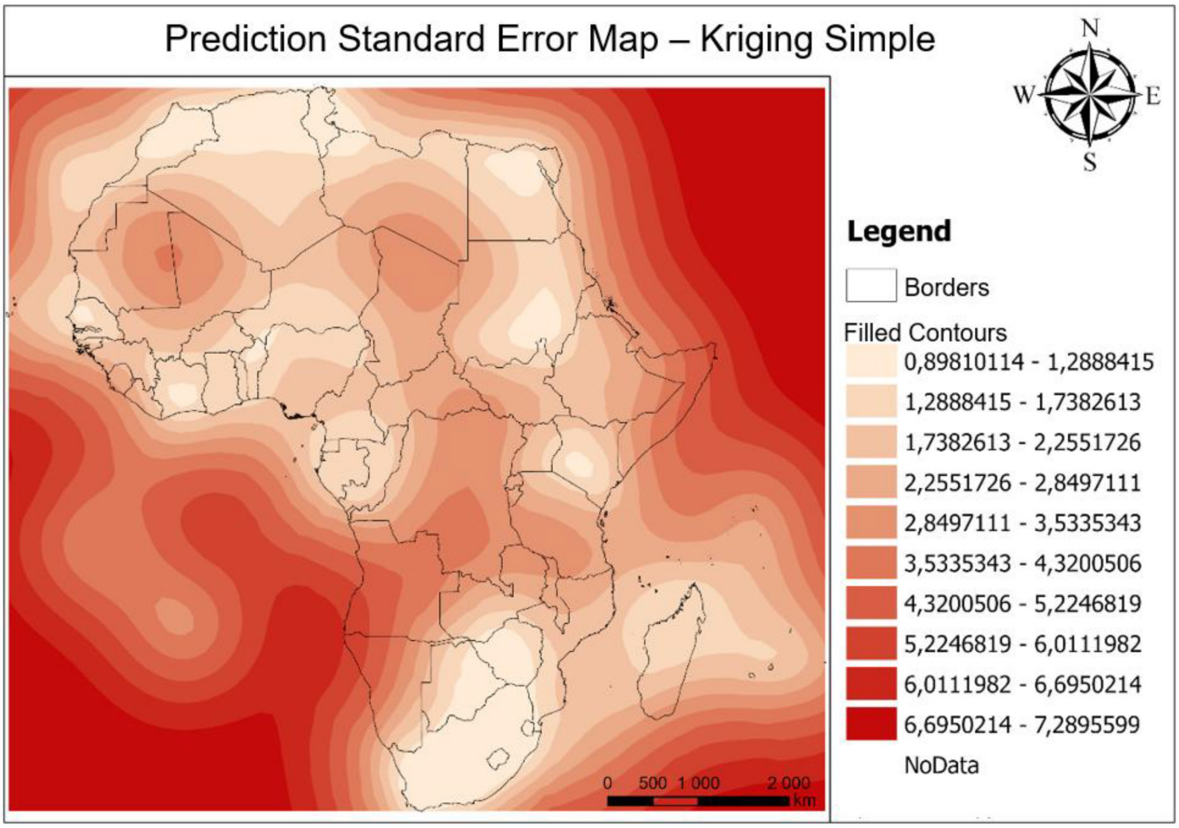


Fig. 14. Error map for the optimal geostatistical method using ArcGIS Pro software
Source: own elaboration based on Tarnowska (2024).

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

the low error zones, show moderate model precision – suggesting that the predictions, although less accurate, are still reliable. In contrast, the highest prediction errors, exceeding 6, are observed along the periphery of the continent. These elevated errors indicate low model precision and a significant deviation between predictions and observed measurements in these regions.

CONCLUSIONS

This study focused on an in-depth analysis of maximum temperature data in Africa, aiming to identify the optimal geostatistical interpolation method. The research involved comprehensive data exploration, including statistical tests, distribution analysis, and outlier identification. Subsequently, using ArcGIS Pro software, maps were generated using various interpolation methods, both deterministic (IDW, LPI) and stochastic (Ordinary Kriging, Simple Kriging, Universal Kriging, EBK).

Exploratory Data Analysis (EDA) proved to be an extremely important stage in the geostatistical modelling process, confirming Hypothesis H_2. Through EDA, it was possible to conduct a thorough analysis of the spatial distribution of the data, identify outliers, and detect spatial dependencies between measurement points. Techniques such as the Moran's I test and Average Nearest Neighbour analysis enabled a deep understanding of the data structure, contributing to a better fit of the interpolation models. These findings demonstrate that EDA was an essential step that allowed for the optimization of modelling parameters and the selection of the most accurate interpolation method.

An equally important step was model validation using the GIS Data Modelling Validator software, which allowed for a comparison of the results of different interpolation methods based on error parameter analysis. This process led to the selection of the optimal method that most accurately reflects the actual distribution of temperatures in Africa with the highest probability.

The analysis results showed that Empirical Bayesian Kriging (EBK) gave the best performance in the cross-validation procedure, while Simple Kriging with a Gaussian model proved to be the optimal method in the subset validation. Furthermore, based on the MPOJE values, Simple Kriging with a Gaussian model was identified as the most efficient interpolation technique among all those evaluated in this study. Prediction, probability and error maps were generated for this method, providing valuable insight into both the spatial distribution of maximum temperatures and the overall accuracy of the interpolation model.

The results of this research highlight the importance of selecting an appropriate interpolation method to ensure the accuracy and applicability of spatial maps. A detailed assessment of model quality was carried out in order to identify areas with the smallest prediction errors and the highest probability of agreement with actual measurements. This assessment allowed the identification of regions where the interpolation methods were the most accurate, thereby supporting more reliable estimates of maximum temperature patterns. The results of this analysis make an important contribution to understanding the spatial variability of estimation errors and provide a solid basis for evaluating the performance of geostatistical models.

In relation to the hypotheses, the research results confirm both H_1 and H_2:

- H_1: The hypothesis stating that stochastic methods generate smaller interpolation errors than deterministic methods was positively verified. The analysis shows that stochastic methods, such as Simple Kriging and Empirical Bayesian Kriging (EBK), indeed proved to be more precise and produced smaller prediction errors compared to deterministic methods like IDW and LPI. Simple Kriging with the Gaussian model achieved the best results, confirming the superiority of stochastic methods in more accurately representing the spatial distribution of temperatures.
- H_2: The hypothesis that Exploratory Data Analysis (EDA) is an essential stage in the geostatistical modelling process was also confirmed. Through

 ✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it

preliminary data analysis, including spatial distribution analysis, outlier identification, and trend assessment, proved to be crucial for the accuracy of the model results. EDA enabled precise adjustment of modelling parameters and the selection of optimal interpolation methods, which directly impacted the quality of the final predictions.

In conclusion, both hypotheses were confirmed during the study, highlighting the importance of stochastic methods in the interpolation process and the necessity of preliminary data analysis to achieve reliable geostatistical modelling results.

Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

Ahmed, K., Shahid, S., & Harun, S. (2014). Spatial interpolation of climatic variables in a predominantly arid region with complex topography. *Environment Systems and Decisions*, *34*, 555–563. https://doi.org/10.1007/s10669-014-9519-0

Alvares, C. A., Stape, J. L., Sentelhas, P. C., De Moraes Gonçalves, J. L., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* [*Meteorological Journal*], *22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Anggani, N. L., Amrullah, H. M., & Gemilang, D. S. A. (2023). Moran I autocorrelation study for level spatial pattern analysis. *Jurnal Indonesia Sosial Teknologi* [*Indonesian Journal of Social Technology*], *4*, 1285–1291. https://doi.org/10.59141/jist.v4i9.686

Bojacá, C., Gil, R., Gómez, S., Cooman, A., & Schrevens, E. (2009). Analysis of greenhouse air temperature distribution using geostatistical methods. *Transactions of the ASABE*, *52*, 957–968. https://doi.org/10.13031/2013.27393

Cellmer, R. (2014). *Modelowanie przestrzenne w procesie opracowywania map wartości gruntów* [*Spatial modelling in the process of developing land value maps*]. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego.

Chiles, J.-P., & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley. https://doi.org/10.1002/9780470316993

Dębowska, U., & Zawadzki, J. (2005). Analiza statystyczna i geostatystyczna zróżnicowania przestrzennego parametrów położenia warstw w NW części Gór Świętokrzyskich [Statistical and geostatistical analysis of the spatial diversity of strata positions in the NW part of Holy Cross Mts, central Poland]. *Przegląd Geologiczny*, *53*(4), 306–310.

Davison, A., & Gholamrezaee, M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *468*, 581–608. https://doi.org/10.1098/rspa.2011.0412

Gaździcki, J. (1990). *Systemy Informacji Przestrzennej* [*Spatial Information Systems*]. Polskie Przedsiębiorstwo Wydawnictwa Kartograficznego.

Isaaks, E. H., & Srivastava, R. M. (1988). Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical Geology*, *20*, 313–341.

Izdebski, W. (2015). *Wykłady z Systemów Informacji o Terenie* [*Lectures on Land Information Systems*]. Wydawnictwo Politechniki Warszawskiej.

Jackowski, A. (2004). *Encyklopedia. Geografia* [*Encyclopedia. Geography*]. Zielona Sowa.

Jha, S., Mariéthoz, G., Evans, J., & Mccabe, M. (2013). Demonstration of a geostatistical approach to physically consistent downscaling of climate modeling simulations. *Water Resources Research*, *49*, 245–259. https://doi.org/10.1029/2012WR012602

Jha, S., Mariéthoz, G., Evans, J., Mccabe, M., & Sharma, A. (2015). A space and time scale-dependent nonlinear geostatistical approach for downscaling daily precipitation and temperature. *Water Resources Research, 51*, 6244–6261. https://doi.org/10.1002/2014WR016729

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, *52*(6), 119–139.

Krige, D. G. (1952). A statistical analysis of some of the borehole values in the Orange Free State goldfield. *Journal of the Southern African Institute of Mining and Metallurgy, 53*(3), 47–64.

Maguire, D. J. (1991). An overview and definition of GIS. *Geographical information systems: Principles and applications, 1*(1), 9–20.

Makowski, J. (2018). *Geografia fizyczna świata* [*Physical geography of the world*]. Wydawnictwo Naukowe PWN.

Marmol, U. (2002). Modelowanie reprezentacji powierzchni topograficznej z wykorzystaniem metody geostatystycznej [Modeling of topographic surface

patrycja.tarnowska11@gmail.com, marek.ogryzek@uwm.edu.pl, eufemia.tarantino@poliba.it

representation using geostatistical methods]. *Geodezja* [*Geodesy*], *8*(2), 259–270.

Martyn, D. (2000). *Klimaty kuli ziemskiej* [*Climates of the Earth*]. Państwowe Wydawnictwo Naukowe PWN.

Matheron, G. (1962). *Traité de géostatistique appliquée* [*Treatise on applied geostatistics*]. Technip.

Matheron, G. (1963). *Principles of geostatistics*. Technip.

Matheron, G. (1965). *Les variables régionalisées et leur estimation: une application de la théorie de fonctions aléatoires aux sciences de la nature* [*Regionalized variables and their estimation: an application of random function theory to the natural sciences*]. Masson.

Morsy, S., & Ahmed, S. (2023). Monitoring of Land Surface Temperature from Landsat Imagery: A Case Study of Al-Anbar Governorate in Iraq. *Geomatics and Environmental Engineering, 17*(3), 61–81. https://doi.org/10.7494/geom.2023.17.3.61

Ogryzek, M., & Kurowska, K. (2016). Geostatystyczne metody opracowywania map średnich cen transakcyjnych gruntów rolnych niezabudowanych [Geostatistical Methods Preparing of Map Average Transaction Prices of Agricultural Land Undeveloped]. *Studia i Prace WNEiZ US, 45*(1), 397–408. https://doi.org/10.18276/sip.2016.45/1-31

Ogryzek, M. (2018). Parametryczna ocena jakości estymacji map opracowanych metodami geostatystycznymi [Parametric evaluation of the quality of estimation of maps developed by geostatistical methods]. *Studia i Prace WNEiZ US, 54*(3), 319–330. https://doi.org/10.18276/sip.2018.54/3-23

Ogryzek, M., Krypiak-Gregorczyk, A., & Wielgosz, P. (2020). Optimal geostatistical methods for interpolation of the Ionosphere: A Case Study on the St Patrick's Day storm of 2015. *Sensors, 20*(10), 2840. https://doi.org/10.3390/s20102840

Ogryzek, M., & Jaskulski, M. (2025). Applying Methods of Exploratory Data Analysis and Methods of Modeling the Unemployment Rate in Spatial Terms in Poland. *Applied Sciences-Basel*, *15*(8). https://doi.org/10.3390/app15084136

Pickens, B. A., Carroll, R., Schirripa, M. J., Forrestal, F., Friedland, K. D., & Taylor, J. C. (2021). A systematic review of spatial habitat associations and modeling of marine fish distribution: A guide to predictors, methods, and knowledge gaps. *PLoS ONE*, *16(*5), e0251818. https://doi.org/10.1371/JOURNAL.PONE.0251818

Raquel, B., Montes, F., Rubio, A., & Osoro, K. (2007). Geostatistical Modelling of Air Temperature in a Mountainous Region of Northern Spain. *Agricultural and Forest Meteorology, 146*(3–4), 173–88.

Royle, A. G. (1979). *A practical introduction to geostatistics*. Department of Mining and Mineral Sciences, University of Leeds.

Śliwicki, D. (2014). Application of nuclear estimators for estimating the effectiveness of active labour market programs. *Acta Universitatis Nicolai Copernici (AUNC), Ekonomia, 45*, 27–40. https://doi.org/10.12775/AUNC_ECON.2014.002

Taoufik, M., Laghlimi, M., & Fekri, A. (2021). Comparison of Land Surface Temperature Before, During and After the Covid-19 Lockdown Using Landsat Imagery: A Case Study of Casablanca City, Morocco. *Geomatics and Environmental Engineering*, *15*(2), 105–120. https://doi.org/10.7494/geom.2021.15.2.105

Tarnowska P. (2024). Geostatyczne metody opracowania map na przykładzie temperatur w Afryce [Geostatical methods of map preparation on the example of temperature distribution in Africa]. (Master's thesis). University of Warmia and Mazury in Olsztyn.

Tobler, W. (1970). A computer model simulating urban growth in Detroit region. *Economic Geography, 46*(1), 234–240.

Urbański, J. (2008). *GIS w badaniach przyrodniczych* [*GIS in environmental research*]. Wydawnictwo Uniwersytetu Gdańskiego.

Vicente-Serrano, S., Saz-Sánchez, M., & Cuadrat, J. (2003). Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research, 24*, 161–180. https://doi.org/10.3354/cr024161

✉patrycja.tarnowska11@gmail.com, ✉marek.ogryzek@uwm.edu.pl, ✉eufemia.tarantino@poliba.it