Interpolated Surface of Daily Average Wind Speed in the Central Region of Portugal

Beatriz Mendes[1],
[1] NOVA University of Lisbon – NOVA IMS, Tel.: +351 926 074 551; 20230919@novaims.unl.pt

Postgraduate in Geospatial Data Science - Spatial Statistics

Abstract: This work aims to develop an interpolated surface of the daily average wind speed in the Central region of Portugal using the Empirical Bayesian Kriging Regression method. This method is notable for its ability to integrate data on the daily average wind speed measured at meteorological stations and a Digital Terrain Model (DTM) as an explanatory variable. The analysis begins with an exploratory investigation of spatial data, evaluating spatial autocorrelation, and culminating in the selection of the most suitable interpolation method. The goal is to create an accurate interpolated surface to better understand the distribution and variability of wind in the central region of Portugal. This is crucial for the planning and development of wind energy projects, providing insights into potential areas for these projects. The approach used in the study highlights the importance of advanced spatial analysis techniques in meteorology and wind project planning.

Keywords: Interpolated Surface, Spatial Statistics, Wind Speed Analysis and Empirical Bayesian Kriging

1. Introduction

This work, developed within the scope of the Spatial Statistics course, aims to produce an interpolated surface of the daily average wind speed in the central region of Portugal, using a DTM as an explanatory variable. The objective is to address the following questions through the results produced:

- What is the geographical location of the extreme values of wind speed?
- Are there areas that stand out for their high or low wind speeds?
- Is there a global identifiable trend in wind speed values?
- Was global spatial autocorrelation identified in wind speed data? Does this mean that wind speed in a certain area is influenced by values in neighboring areas?
- Analyze the spatial correlation between wind speed (dependent variable) and the DTM (independent variable).

Several methods can be used to develop interpolated surfaces, including Inverse Distance Weighting (IDW), which assumes that closer locations have similar wind measurements, being simple and effective for nearby data but less accurate at longer distances. Ordinary Kriging (OK) is ideal for consistent wind patterns through spatial autocorrelation. The Generalized Additive Model (GAM) models complex nonlinear relationships with smoothing functions, excelling in varied terrains. Other methods include Support Vector Machine (SVM) and Neural Networks (NN), which are robust for predictive analysis and complex relationships but require large datasets. Hybrid models combine methods to balance accuracy and simplicity, such as IDW and OK. Finally, Regression-Kriging (RK) offers a robust approach integrating multivariate linear regression to model deterministic trends and kriging for residual interpolation (Reinhardt & Samimi 2018).

The chosen method for this work is Empirical Bayesian Kriging (EBK) Regression, which combines the advantages of geostatistical methods, incorporating neighboring values to calculate the estimated value, and integrates additional explanatory variables, in this case, the DTM (Reinhardt & Samimi 2018).

2. Study Area and Data

The selected area for developing the interpolated surface of daily average wind speed is the Central region of Portugal (Figure 1).

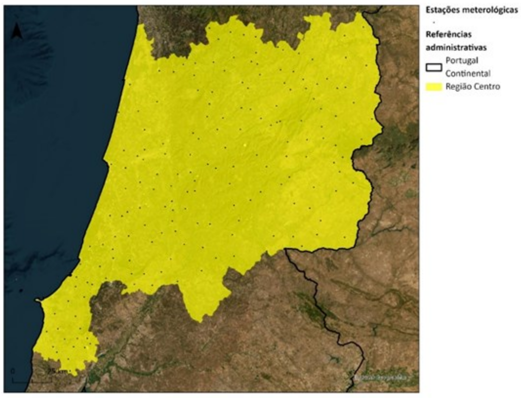


Figure 1. Study area

The Central region is composed of 100 municipalities organized into eight Intermunicipal Communities. According to the 2021 Census, the Central region recorded 2,327,755 inhabitants, reflecting a decrease of almost 4% compared to 2011 (CCDR Centro 2023).

Data on daily average wind speed were collected from 180 meteorological stations in the Central region of Continental Portugal for the period 2001-2023 through the National Information System on Water Resources (SNIRH-APA). Additionally, data on the characteristics of each station, including meteorological station code, altitude, latitude, longitude, coordinates, responsible entity, district, municipality, parish, type of meteorological station, among others, were also collected.

These data were obtained in two .csv files (one with daily average wind speed data and the other with meteorological station data) with a common column for the Meteorological Station Identification Code. To associate the daily average wind speed data with the location of meteorological stations, both tables were imported into ArcGIS Pro, and a join was performed using the common field of the Station Codes. This allowed the spatial distribution of daily average wind speed values to be verified and consequently produced an interpolated surface of daily average wind speed in the Central region of Portugal.

3. Methodology

The development of the work followed the steps identified in Figure 2, with data collected and cleaned (pre-processing), followed by various analyses leading to the final EBK Regression model to develop the interpolated surface of daily average wind speed in the Central region of Portugal.

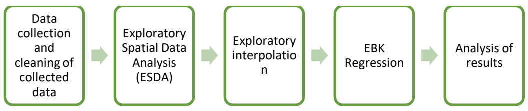
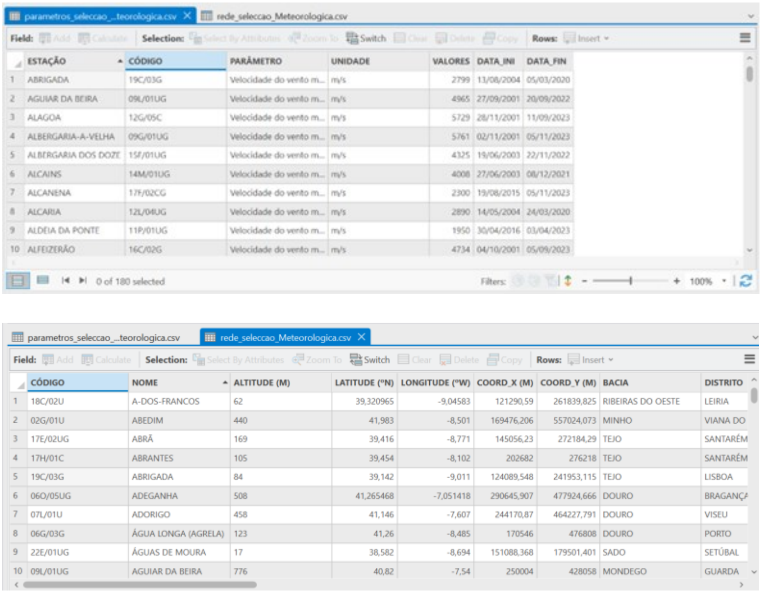


Figure 2. Flowchart with the methodology adopted

The collected data on daily average wind speed from 180 meteorological stations in the Central region of Continental Portugal for the period 2001-2023 were downloaded into two .csv files, as shown in Appendix 1.



	ESTAÇÃO	CÓDIGO	PARÂMETRO	UNIDADE	VALORES	DATA_INI	DATA_FIN
1	ABRIGADA	19C/03G	Velocidade do vento m...	m/s	2799	13/08/2004	05/03/2020
2	AGUIAR DA BEIRA	09L/01UG	Velocidade do vento m...	m/s	4965	27/09/2001	20/09/2022
3	ALAGOA	12G/05C	Velocidade do vento m...	m/s	5729	28/11/2001	11/09/2023
4	ALBERGARIA-A-VELHA	09G/01UG	Velocidade do vento m...	m/s	5761	02/11/2001	05/11/2023
5	ALBERGARIA DOS DOZE	15F/01UG	Velocidade do vento m...	m/s	4325	19/06/2003	22/11/2022
6	ALCAINS	14M/01UG	Velocidade do vento m...	m/s	4008	27/06/2003	08/12/2021
7	ALCANENA	17F/02CG	Velocidade do vento m...	m/s	2300	19/08/2015	05/11/2023
8	ALCARIA	12L/04UG	Velocidade do vento m...	m/s	2890	14/05/2004	24/03/2020
9	ALDEIA DA PONTE	11P/01UG	Velocidade do vento m...	m/s	1950	30/04/2016	03/04/2023
10	ALFEIZERÃO	16C/02G	Velocidade do vento m...	m/s	4734	04/10/2001	05/09/2023

	CÓDIGO	NOME	ALTITUDE (M)	LATITUDE (°N)	LONGITUDE (°W)	COORD_X (M)	COORD_Y (M)	BACIA	DISTRITO
1	18C/02U	A-DOS-FRANCOS	62	39,320965	-9,04583	121290,59	261839,825	RIBEIRAS DO OESTE	LEIRIA
2	02G/01U	ABEDIM	440	41,983	-8,501	169476,206	557024,073	MINHO	VIANA DO
3	17E/02UG	ABRÃ	169	39,416	-8,771	145056,23	272184,29	TEJO	SANTARÉM
4	17H/01C	ABRANTES	105	39,454	-8,102	202682	276218	TEJO	SANTARÉM
5	19C/03G	ABRIGADA	84	39,142	-9,011	124089,548	241953,115	TEJO	LISBOA
6	06O/05UG	ADEGANHA	508	41,265468	-7,051418	290645,907	477924,666	DOURO	BRAGANÇA
7	07L/01U	ADORIGO	458	41,146	-7,607	244170,87	464227,791	DOURO	VISEU
8	06G/03G	ÁGUA LONGA (AGRELA)	123	41,26	-8,485	170546	476808	DOURO	PORTO
9	22E/01UG	ÁGUAS DE MOURA	17	38,582	-8,694	151088,368	179501,401	SADO	SETÚBAL
10	09L/01UG	AGUIAR DA BEIRA	776	40,82	-7,54	250004	428058	MONDEGO	GUARDA

Figure 3. Attributes of the .csv tables used for this work

Both tables share an attribute "ID_Estacao," which allows the tables to be related and thus associate values with points through the join tool in ArcGIS Pro. The original data coordinates were in the "Lisboa Hayford Gauss IGeoE" coordinate system, and for project consistency, this table was converted to the "ETRS 1989 Portugal TM06" coordinate system after transforming it into a shapefile.

The first type of analysis performed on the data was Exploratory Spatial Data Analysis (ESDA), allowing a preliminary analysis of the properties of daily average wind speed data. This included developing:

- A histogram to visualize wind speed distribution and identify patterns or anomalies.
- Analysis of the geographic distribution of meteorological stations to provide insights into spatial coverage and potential sampling areas.
- Identification of outliers using two methods: visualization of the geographic distribution of daily wind speed values recorded at meteorological stations and the quartile method to ensure data quality by eliminating possible measurement errors or outlier phenomena.
- A Voronoi map using the Neighborhood Summary Statistics tool in ArcGIS Pro, visualizing areas influenced by each meteorological station based on spatial proximity, and understanding how observations at each station contribute to the overall model.
- Spatial Autocorrelation using the Spatial Autocorrelation (Global Moran's I) tool in ArcGIS Pro to determine if there are significant spatial patterns in wind speed distribution, influencing decisions about the appropriate interpolation method.

The next phase involved choosing the best method to use among other explained reasons, using the Exploratory Interpolation tool in ArcGIS Pro, which allows obtaining interpolation results from various methods using customizable criteria and cross-validation statistics (ESRI n.d.-b). Figure 11, Appendix 2, shows the chosen options in the tool.

The focus of the work is the development of an interpolated surface. However, despite the complexity, the initial aim was to develop an interpolated surface for wind potential in the Central region of Portugal. It was decided to include at least one factor in the analysis, the Digital Terrain Model, as an explanatory variable, given that wind speed is influenced by altitude and terrain slope, and the DTM can help identify wind speed variations affected by terrain elevation.

EBK Regression

EBK Regression is a geospatial interpolation method distinguished by its adaptive statistical approach. Unlike traditional kriging, EBK Regression uses a simulation process to create multiple semivariograms, estimating a model that captures spatial variability of data with greater accuracy. This process is especially useful when data exhibit non-uniform or complex spatial variability, as EBK Regression adapts to these variations, providing more accurate and reliable estimates (Gribov & Krivoruchko 2020).

This method's ability to handle the complexity of spatial data allows for a more accurate representation of reality, essential for the precision of the analyses and conclusions of this work. Additionally, EBK Regression includes the uncertainty associated with variogram parameter estimation, offering a more complete view of prediction accuracy.

For developing the interpolated surface using this method, the K-Bessel Semivariogram Type was chosen for its better flexibility and higher accuracy compared to other options, and for its ability to describe the spatial correlation of processes that may not follow a Gaussian distribution. This choice is justified by the frequent violation of normality assumption in spatial data and the need to model processes with different degrees of spatial smoothness (ESRI 2023).

Regarding Transformation, Log Empirical was chosen to normalize data and ensure predicted values are not negative, considering that the work only deals with positive values.

The fine-tuning process of the EBK Regression model involves adjusting the number of neighbors and sectors to optimize prediction accuracy. These adjustments balance local precision and global generalization of predictions, ensuring model reliability for the specific dataset.

In summary, the EBK Regression method is a spatial interpolation tool that, through some model adjustments, increases model applicability and accuracy.

4. Results

4.1. Exploratory Spatial Data Analysis (ESDA)

The distribution of wind speeds according to the histogram in Figure 4 is approximately normal, indicated by the well-fitted overlay of the normal distribution curve (grey line). The median is close to the mean, suggesting a balanced distribution of speeds above and below the mean. The standard deviation (1.34343599 m/s) indicates moderate variation, suggesting that wind speed generally approaches the mean despite days with significantly calmer or stronger winds in the study area. Most wind speed measurements are concentrated between about 3500 m/s and 5000 m/s, with few occurrences of very low (<2000 m/s) and very high (>6000 m/s) speeds.

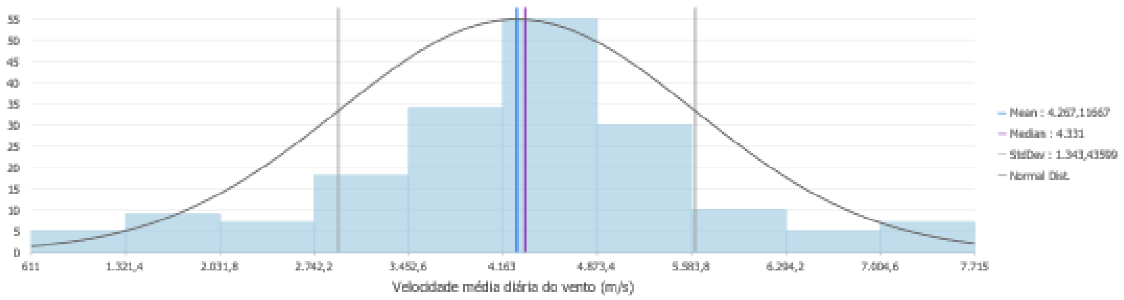


Figure 4. Histogram of average daily wind speed (m/s) in the Central region of Portugal

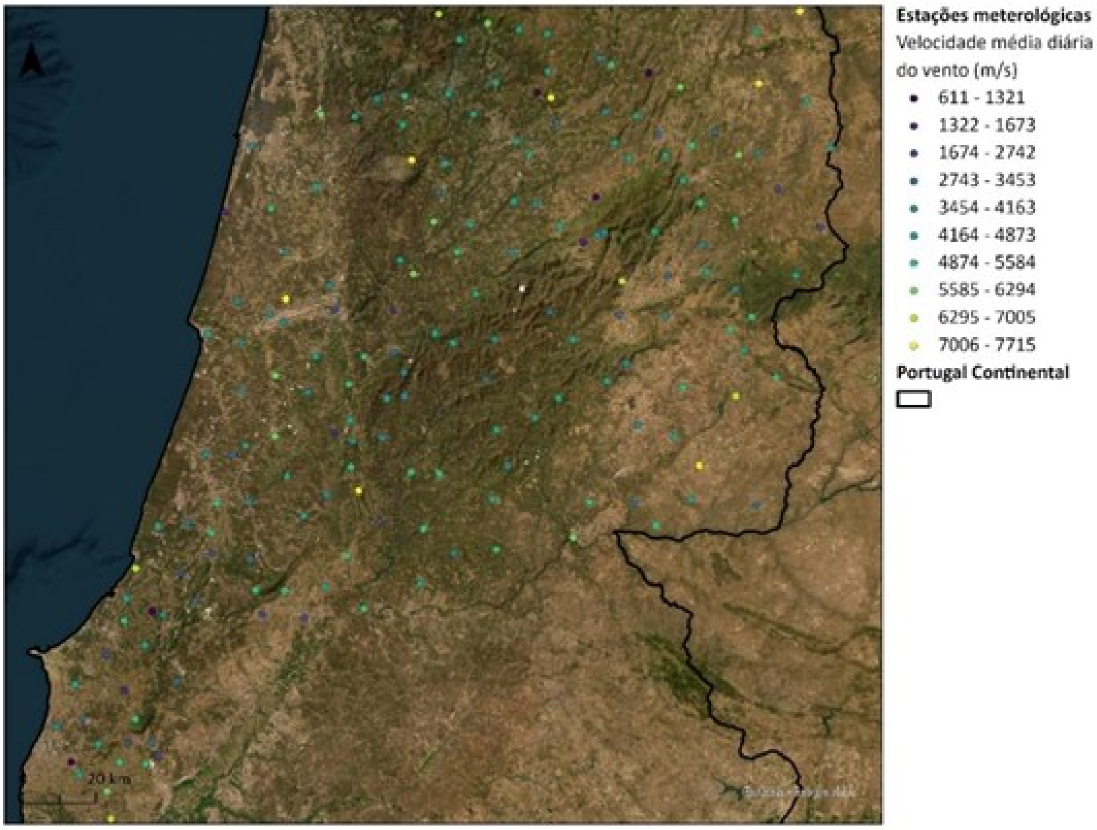


Figure 5. Distribution of meteorological stations and average daily wind speed (m/s)

Table 1. Criteria for Outlier Identification

Criteria	Description	Value
First Quartil (Q1)	The value below which 25% of data lies	3 633
Second Quartil (Q2)	Divides the data into two equal parts	4 325
Third Quartil (Q3)	The value below which 75% of data lies	4977
Interquartile Range (IQR)	Difference between Q3 and Q1 (IQR = Q3 - Q1)	1 344
Lower limit for Outliers	Q1 - 1,5 * IQR	1 617
Upper limit for Outliers	Q3 + 1,5 * IQR	6 993

Figure 5 shows the distribution of meteorological stations and their daily average wind speeds, with the lowest and highest classes represented by different colors for identification.

Outliers were identified using the quartile method, defining criteria for outlier identification (Table 1) and visualized in ArcGIS Pro (Figure 12, Appendix 2).

The Voronoi map (Figure 7) helps analyze sample variability based on its relation to surrounding sample points. Each polygon represents the area closest to a meteorological station, with color indicating the average daily wind speed for that area. Darker polygons indicate higher average wind speeds, potentially highlighting areas of interest for wind energy projects.

Figure 6. Identification of outliers

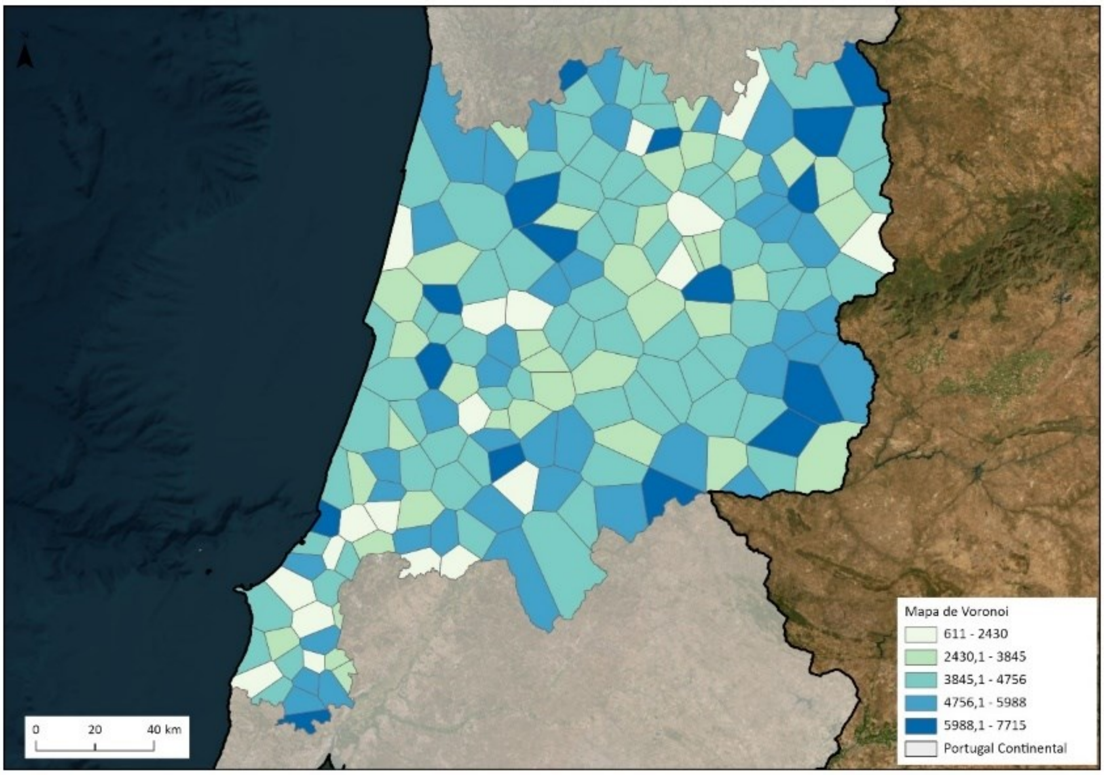


Figure 7. Voronoi map of average daily wind speed (m/s) in the Central region of Portugal

Spatial Autocorrelation (Global Moran's I) (Figure 8) showed no significant spatial autocorrelation in the data, indicated by a Moran's Index of -0.039319 and a p-value of 0.474200. This suggests that, contrary to Tobler's Law, measurements at nearby meteorological stations do not always have similar values, indicating other factors may influence the data more than geographic proximity.

Global Moran's I Summary

Moran's Index	-0,039319
Expected Index	-0,005587
Variance	0,002222
z-score	-0,715662
p-value	0,474200

Distance measured in Meters

Figure 8. Result of Global Moran's I Summary

4.2. Exploratory Interpolation

Using the Exploratory Interpolation tool in Geoestatistical Analyst (ArcGIS Pro), various interpolation methods were evaluated (Table 2).

Table 2. Ranking of interpolation methods – Result of Exploratory Interpolation (ArcGIS Pro)

Method	Rank	RMSE*	ME*	ME_STD*	RMSE_STD*	ASE*
Simple Kriging _ Default	1	1339,106179	-0,932690487	-0,000663961	1,007262932	1329,390695
Simple Kriging _ Trend	2	1348,780859	-1,209915625	-0,000913745	1,018617907	1324,128361
Simple Kriging _ Optimized	2	1338,582254	-3,138948785	-0,002294915	1,007248266	1329,04475
Ordinary Kriging _ Default	2	1370,223248	-6,453714978	-0,004223239	1,069674585	1278,310782
Simple Kriging _ Trend and transformation	2	1348,804507	8,078050087	0,006123534	1,022455976	1319,181009
Ordinary Kriging _ Optimized	2	1357,967417	-14,70572828	-0,010053734	1,016885762	1334,090442
Empirical Bayesian Kriging - Default	2	1360,316848	-23,83040138	-0,017188941	0,998325704	1361,206211
Empirical Bayesian Kriging - Advanced	2	1327,405902	-24,64606638	-0,017620959	0,989931259	1339,468617
Kernel Interpolation	2	1366,934194	63,6002812	0,04767489	1,007606224	1345,673933
Universal Kriging _ Default	2	1381,672403	68,06985184	0,050531564	1,040269083	1319,152831
Universal Kriging _ Optimized	2	1381,383858	69,03190455	0,050181251	1,025239506	1338,760521
Inverse Distance Weighted - Optimized	12	1400,189003	-36,30472633	-1,7977E+308	-1,7977E+308	-1,7977E+308
Inverse Distance Weighted - Default	12	1476,322591	-27,35550691	-1,7977E+308	-1,7977E+308	-1,7977E+308

*Root-mean-square error (RMSE), Mean Error (ME), Mean Standardized Error (ME_STD), Root Mean Square Standardized Error (RMSE_STD), Average Standard Error (ASE).

The choice of EBK Regression is justified by its ability to incorporate the uncertainty associated with variogram estimation, a crucial aspect given the moderate standard deviation observed in the wind speed histogram. This feature is particularly relevant considering the approximately normal data distribution reflected by the histogram (Figure 4).

EBK Regression is designed to work well with data that follow or can be transformed to follow a normal distribution, offering an advantage over methods that do not consider this transformation. EBK Regression's interpolation adapts to local variations, resulting in a model that can provide a more accurate representation of the complex spatial conditions inherent in wind data.

4.3. EBK Regression

For interpolating daily average wind speed data in the Central region of Portugal, the ArcGIS Pro Geoestatistical Wizard tool was used, choosing the EBK Regression Prediction method, with steps shown in Figure 15, Appendix 2. Various tests were conducted, altering the number of neighbors and sectors while keeping the same specific location. The results are presented in Appendix 3.

The model with 3-5 neighbors and 1 sector shows the highest Standard Error of Prediction, unfavorable for model accuracy. The model with the same range of neighbors but 4 sectors has a lower Standard Error of Prediction than the previous model but still higher than models with 8 sectors. The model with 5-8 neighbors and 8 sectors stands out for its low Standard Error of Prediction, predicting higher model accuracy. Additionally, this model has satisfactory values for Mean Standardized and Root-Mean-Square Standardized, suggesting precise and reliable predictions. Increasing the number of neighbors to 10-15 with 4 sectors shows promising results, but the Standard Error of Prediction is slightly higher compared to the model with 8 sectors and 5-8 neighbors.

The model with 8 neighbors and 8 sectors presents the lowest Standard Error of Prediction among all analyzed models, along with notable performance in terms of Mean Standardized and Root-Mean-Square Standardized, approaching ideal values.

Despite the chosen criteria for defining the selected method, several tests were conducted to ensure that the model with 8 neighbors and 8 sectors was the best fit for daily average wind speed data. The results of the IDW method and EBK Regression with Exponential Transformation resulted in much less smoothed surfaces. Future exploration of these hypotheses and comparisons would be interesting.

Considering all evaluation metrics, the model with 8 neighbors and 8 sectors emerges as the most accurate choice for interpolating daily average wind speed data, optimally balancing precision and reliability according to the provided data. In the following figures and Figure 16, Appendix 2 show the results of the EBK Regression and Standard Error, identifying the area with the highest predicted average wind speed in the Serra da Estrela region.

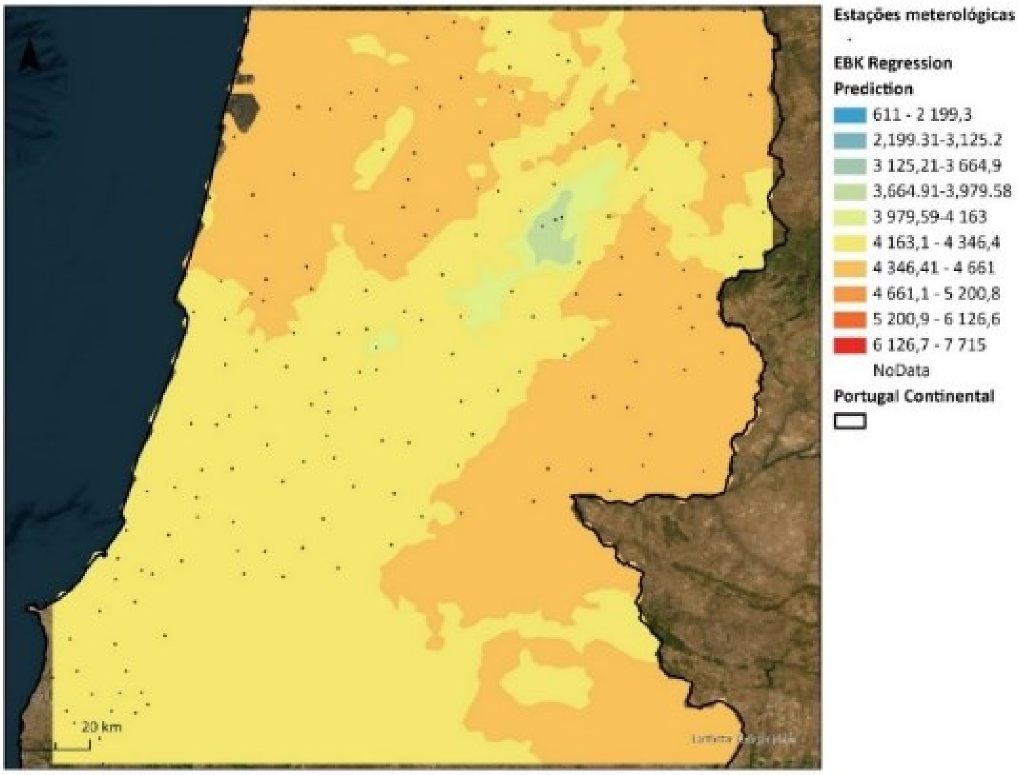


Figure 9. Interpolated surface of daily average wind speed in the central region of Portugal - Result of EBK Regression Prediction

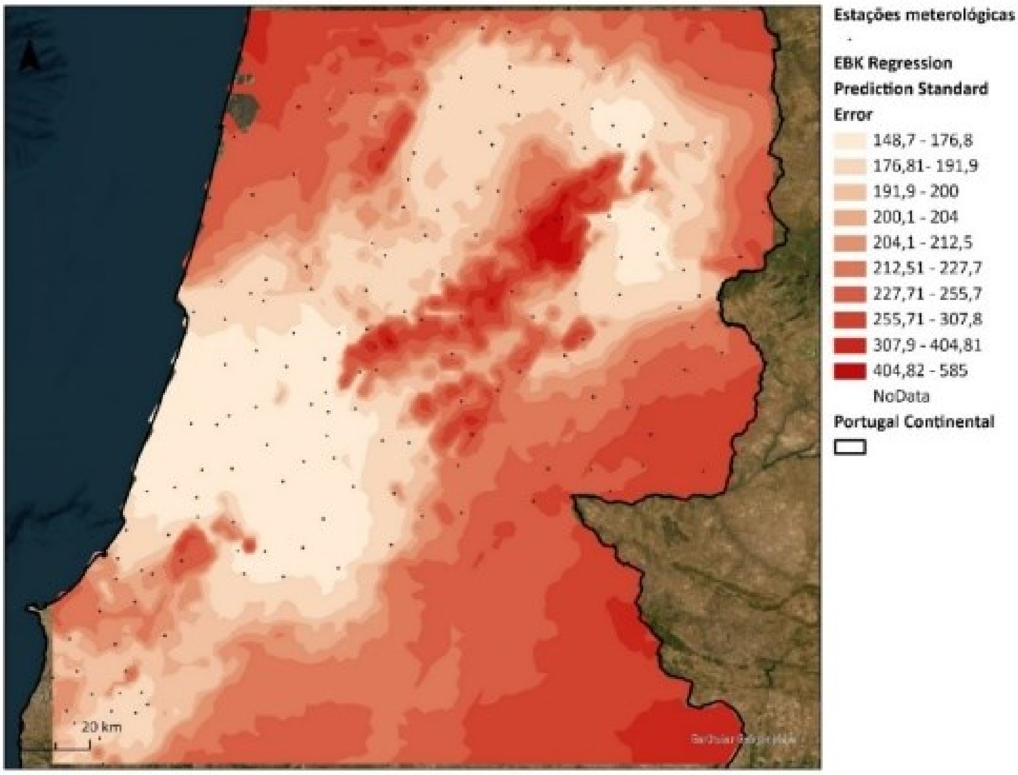


Figure 10. Standard Error EBK Regression Result

Since the EBK Regression method was used, the Semivariogram Sensitivity tool (ESRI n.d.-a) cannot be used, and exact values of nugget, sill, and range cannot be obtained. However, the semivariogram of the used model allows estimating these values (Appendix 3):

- Nugget: 0.98
- Range: 0.377
- Sill: 10.067

The estimated values suggest a low nugget value, indicating little variation in wind speed over short distances, meaning the data is consistent with little random variability or measurement errors at meteorological stations. The low range value suggests wind conditions change over relatively short distances, likely due to local topography or regional climatic variations. The high sill value indicates greater overall variability in wind speed in the studied region.

According to the Normal QQ Plot result (Appendix 3), the interpolation residues are close to a normal distribution, as most points align with the reference line.

5. Conclusions

The detailed analysis of wind speed data in the central region of Portugal using EBK Regression methodology revealed the complexity and variability of wind conditions. No trend was identified in data distribution or outliers. The spatial autocorrelation of the data is insignificant despite being negative, and the values of nugget, range, and sill reflect the high complexity of wind patterns in the central region of Portugal.

References

CCDR Centro. (n.d.). CCDRC - DataCentro. Datacentro.ccdrc.pt. Retrieved November 27, 2023, from http://datacentro.ccdrc.pt/Viewer.aspx

CCDR Centro. (2023). Sobre a Região Centro. CCDRC. https://www.ccdrc.pt/pt/regiao-centro/sobre-a-regiao-centro/

ESRI. (n.d.-a). 045028: The geostatistical layer or XML file does not represent a supported model.—ArcGIS Pro | Documentation. Pro.arcgis.com. Retrieved December 4, 2023, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/tool-errors-and-warnings/040001-050000/tool-errors-and-warnings-45026-45050-045028.htm

ESRI. (n.d.-b). Exploratory Interpolation (Geostatistical Analyst)—ArcGIS Pro | Documentation. Pro.arcgis.com. Retrieved December 7, 2023, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/geostatistical-analyst/exploratory-interpolation.htm

ESRI. (2023). What is EBK Regression Prediction?—ArcGIS Pro | Documentation. Pro.arcgis.com. https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/what-is-ebk-regression-prediction-.htm

Gribov, A., & Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. Science of the Total Environment, 722, 137290. https://doi.org/10.1016/j.scitotenv.2020.137290

Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. Computers & Geosciences, 33(10), 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001

Johnston, K., Hoef, J. M. V., Krivoruchko, K., & Lucas, N. (2004). Exploratory Spatial Data Analysis. In Using ArcGIS Geostatistical Analyst GIS by ESRI (pp. 81–90, 96–105). http://downloads2.esri.com/support/documentation/ao_/Using_ArcGIS_Geostatistical_Analyst.pdf

Moore, D. S., McCabe, G. P., & Craig, B. A. (2014). Introduction to the Practice of Statistics. In Google Books. Macmillan Higher Education. https://books.google.pt/books/about/Introduction_to_the_Practice_of_Statisti.html?id=pX1_AwAAQBAJ&redir_esc=y

Reinhardt, K., & Samimi, C. (2018). Comparison of different wind data interpolation methods for a region with complex terrain in Central Asia. Climate Dynamics, 51(9-10), 3635–3652. https://doi.org/10.1007/s00382-018-4101-y

Appendix 1 – Data collected

Table 3. Contents of the .csv file "parametros_seleccao_Meteorologica"

STATION	CODE	VALUES	START_DATE	END_DATE
ABRIGADA	19C/03G	2799	13/08/2004	05/03/2020
AGUIAR DA BEIRA	09L/01UG	4965	27/09/2001	20/09/2022
ALAGOA	12G/05C	5729	28/11/2001	11/09/2023
ALBERGARIA DOS DOZE	15F/01UG	4325	19/06/2003	22/11/2022
ALBERGARIA-A-VELHA	09G/01UG	5761	02/11/2001	05/11/2023
ALCAINS	14M/01UG	4008	27/06/2003	08/12/2021
ALCANENA	17F/02CG	2300	19/08/2015	05/11/2023
ALCARIA	12L/04UG	2890	14/05/2004	24/03/2020
ALDEIA DA PONTE	11P/01UG	1950	30/04/2016	03/04/2023
ALFEIZERÃO	16C/02G	4734	04/10/2001	05/09/2023
ALJUBARROTA	16D/01UG	1853	05/08/2004	10/09/2009
ALMAÇA	11H/01UG	4346	28/09/2001	02/11/2023
ALMEIDA	09P/02UG	4359	11/10/2001	05/05/2020
ALMEIDINHA	10O/02UG	6247	02/11/2003	05/11/2023
ALTO DA FOZ DO GIRALDO	13K/05UG	4423	25/10/2001	27/10/2022
ALVAIÁZERE	15G/01UG	5526	14/10/2001	05/11/2023
ALVORNINHA	17C/06G	4629	04/10/2001	21/12/2021
ANSIÃO	14G/01C	1389	02/12/2001	21/09/2005
ARANHAS	13O/02UG	4832	25/11/2001	26/09/2022
ARRANHÓ	20C/03G	6367	13/08/2004	16/10/2023
ARRIMAL	17D/03UG	3174	15/11/2001	12/06/2017
ASSEICEIRA	18D/01C	3363	30/11/2001	27/12/2022
BARRAGEM DE MEIMOA	12O/04C	4101	29/11/2001	06/12/2021
BARRAGEM DE ÓBIDOS (DGADR)	17C/08C	2148	02/03/2005	26/10/2011
BATALHA	16E/06C	5254	02/12/2001	13/09/2023
BENDADA	11N/02UG	3984	01/11/2001	19/08/2019
BOLETA (CARAPINHEIRA)	12F/03UG	4132	07/01/2005	07/09/2022
BOUÇÃ (PESSEGUEIRO DO VOUGA)	09G/03UG	4863	13/01/2005	05/11/2023
BRUFE (BARREIROS)	09K/03UG	618	26/09/2001	19/08/2003
CADAFAZ	13I/02UG	4401	29/09/2001	16/12/2021
CALDAS DE FELGUEIRAS	11J/01UG	4296	28/09/2001	24/10/2022
CALDE	09J/03UG	4816	26/09/2001	24/10/2023
CAMPELO	13H/07UG	4423	22/08/2002	16/12/2021
CAMPELOS	18B/03UG	3051	04/10/2001	25/08/2015
CAMPIA	09H/01UG	4348	19/06/2003	05/11/2023
CANTANHEDE	11F/01UG	3504	01/11/2001	17/10/2019
CAPINHA	12M/02UG	4045	12/10/2001	24/06/2021
CARANGUEJEIRA	15E/03G	4033	15/11/2001	22/07/2020
CARAPINHAL	13H/09UG	4893	02/02/2002	05/07/2021
CARIA	12M/01UG	3306	24/08/2002	06/12/2021
CARRAZEDE	16G/02UG	4609	26/10/2001	02/12/2022
CARREGAL DO SAL	11I/03UG	4284	28/09/2001	22/08/2022
CARVOEIRO	16J/02UG	4940	25/10/2001	25/01/2023
CASAL DO RATO	13D/04UG	4416	30/09/2001	06/09/2022
CASTANHEIRA DE PÊRA	13H/05UG	3653	24/03/2004	29/07/2021

STATION	CODE	VALUES	START_DATE	END_DATE
CASTELEIRO	12N/01UG	4340	11/10/2001	21/06/2021
CASTELO NOVO	13M/03G	3997	22/09/2001	17/04/2018
CASTRO DAIRE LAMELAS)	08J/06G	5775	01/03/2002	29/01/2023
CAXARIAS	15F/02C	4317	02/12/2001	04/11/2022
CELA	16C/01C	6821	01/12/2001	23/10/2023
CELORICO DA BEIRA	10M/01G	3061	01/11/2001	05/11/2023
CERNACHE DE BONJARDIM	15H/01C	5153	14/12/2001	17/08/2021
CHÃO DE CODES	16I/02UG	3937	25/10/2001	18/08/2020
CHÃO DE COUCE	14G/05UG	3760	11/11/2001	29/04/2020
CODECEIRO	10N/01UG	3450	01/11/2003	12/06/2017
COENTRAL GRANDE	13H/08UG	3535	29/09/2001	15/10/2019
COIMBRA	12G/02UG	2430	17/10/2001	30/07/2008
CONDEIXA	13G/02UG	4595	27/10/2001	15/03/2022
CONSTÂNCIA	17G/04UG	5221	26/10/2001	28/12/2022
COVILHÃ	12L/03G	6543	01/11/2001	05/11/2023
CRESPOS	16E/01UG	2931	05/10/2001	16/08/2016
CUMIEIRA	14G/04UG	4292	14/10/2001	22/11/2022
CÔJA	12J/01UG	4642	24/10/2001	15/12/2021
DEGRACIAS	13F/02UG	3748	18/06/2003	22/07/2020
ERMIDA (TONDELA)	10I/01UG	3357	19/06/2003	23/11/2022
ESCALHÃO	08P/02G	7479	11/10/2001	05/11/2023
ESTRADA	11F/02UG	5326	01/11/2001	13/10/2023
ESTREITO	14K/04UG	4337	02/03/2002	23/06/2020
FAJÃO	13J/01UG	4076	02/11/2001	26/08/2019
FERREIRA DO ZÊZERE	15H/02UG	2248	26/10/2001	10/01/2008
FERREIRA-A-NOVA	12E/02UG	3845	21/08/2002	07/09/2022
FIGUEIRÓ DOS VINHOS	14H/01UG	3454	14/10/2001	01/08/2017
FORNINHOS	09L/02UG	3733	29/12/2001	31/03/2021
FORNOS DE ALGODRES	10L/01UG	4707	06/01/2002	04/10/2022
FRAGOSELA DE BAIXO	10J/03UG	4169	29/12/2001	02/10/2022
FREIXIANDA	15G/03UG	4162	05/10/2002	22/11/2022
GAFANHA DA NAZARÉ	10E/03UG	4120	02/02/2002	23/06/2020
GOUVEIA	11L/01UG	921	14/07/2004	05/02/2007
GÓIS	13I/01G	4240	10/01/2002	28/07/2021
IDANHA-A-VELHA	14O/01UG	6431	12/10/2001	26/06/2023
ISNA	14J/02UG	3633	15/05/2004	27/05/2020
LADOEIRO	14N/02UG	7561	13/10/2001	05/11/2023
LAGOA COMPRIDA	11L/07CG	1562	15/10/2005	05/11/2023
LEIRIA	15E/01UG	5122	05/10/2001	17/08/2023
LENTISCAIS	15M/02UG	3113	28/06/2003	23/04/2020
LOBAGUEIRA BODIOSA	09J/04UG	4296	27/12/2001	05/11/2023
LOURIÇAL	13E/02UG	4091	21/08/2002	05/09/2022
LOUSÃ	13H/03UG	3414	02/02/2002	04/10/2021
MACEIRA (LIS)	15D/03UG	3351	08/09/2001	14/06/2017
MALPICA DO TEJO	15M/03UG	5215	13/10/2001	27/12/2022
MANGUALDE	10K/01UG	4417	31/10/2001	03/10/2022
MATA	14M/02UG	4575	15/02/2002	21/09/2022
MATA DA BIDOEIRA	14E/02UG	3369	06/10/2001	19/06/2017

STATION	CODE	VALUES	START_DATE	END_DATE
MAXIAL	19B/03UG	4361	03/10/2001	19/08/2020
MAÇAINHAS	10N/05UG	4813	10/10/2001	04/10/2022
MECA	19C/08C	5153	30/11/2001	15/06/2023
MERCEANA	19C/07G	4977	04/10/2001	28/12/2022
MESQUITELA	10M/03UG	4296	28/09/2001	23/10/2022
MINDE	16E/02G	5045	05/10/2002	30/11/2022
MIUZELA	10O/03G	2549	01/11/2001	19/01/2010
MOLEDO	18B/01UG	4709	04/10/2001	22/12/2021
MONFORTE DA BEIRA	15N/01UG	4296	27/06/2003	27/12/2022
MONTE CAVALEIRO	12O/03UG	5229	12/10/2001	26/09/2022
MONTE REAL	14D/03C	4008	01/12/2001	07/08/2017
MOSTEIRO DE CABRIL	08I/01UG	6344	31/10/2001	05/11/2023
MOURILHE	10M/02UG	4485	10/10/2001	13/07/2020
OLEIROS	14J/01UG	4200	25/03/2004	20/12/2021
OLIVEIRA DO BAIRRO	10G/01UG	4249	13/01/2005	05/11/2023
OLIVEIRA DO HOSPITAL	11J/02C	3809	27/01/2002	15/10/2017
ORJARIÇA	19B/05C	4187	30/11/2001	31/03/2020
OTA	19D/02UG	2598	01/01/2002	30/01/2015
PAMPILHOSA DA SERRA	13J/03UG	3134	25/03/2004	19/08/2020
PARANHOS DA BEIRA	11K/03UG	3682	13/05/2004	29/06/2021
PATAIAS (GARE)	16D/03UG	4506	05/10/2001	07/10/2021
PEDROGÃO	16F/04C	5479	29/11/2001	31/03/2023
PEGA	11O/01G	5210	11/10/2001	05/11/2023
PENALVA DO CASTELO	10K/02UG	3093	25/08/2002	14/04/2020
PENAMACOR	12O/01UG	5298	12/10/2001	17/12/2022
PENDILHE	08J/05UG	4231	02/05/2002	03/11/2021
PENEDOS DE ALENQUER	19C/04C	1794	24/12/2005	29/04/2020
PENELA	13G/01UG	5347	11/01/2002	23/11/2022
PENHA GARCIA	13O/01UG	5208	12/10/2001	26/09/2022
PENHAS DOURADAS	11L/10G	3574	10/11/2005	05/11/2023
PINHEL	09O/01G	7475	01/11/2001	05/11/2023
POMBAL	14F/01UG	5734	18/06/2003	05/11/2023
PORTO DE MÓS	16E/03UG	1950	05/10/2001	18/06/2007
PRAGANÇA	18C/01G	5410	04/10/2001	02/11/2023
PRAIA DE MIRA	11E/01C	1410	02/09/2006	16/10/2010
PROENÇA-A-NOVA	15J/01UC	4724	24/11/2001	17/08/2021
PÍNZIO	10O/01UG	3550	12/10/2005	05/10/2022
QUEIRIGA	09K/02UG	4648	26/09/2001	22/10/2022
QUINTA DA FUMADINHA	08L/07UG	3648	27/09/2001	20/12/2022
RAMELA	11N/01UG	5082	10/10/2001	02/01/2023
REGO DA MURTA	15G/02G	7132	22/09/2001	02/05/2023
REVELES (ABRUNHEIRA)	13E/01UG	3775	07/01/2005	06/10/2021
RIBEIRADIO	09H/04UG	3649	16/05/2004	24/02/2021
ROSMANINHAL	15O/01UG	3292	13/10/2001	07/10/2020
SALIR DE MATOS	17C/05UG	4934	04/10/2001	22/12/2021
SANTA COMBA DÃO	11I/01G	6142	24/10/2001	30/10/2023
SANTO VARÃO	12F/02C	3901	18/06/2003	17/08/2021
SARDOAL	16I/04UG	4659	26/10/2001	13/09/2021
SARNADAS DE RÓDÃO	15L/02UG	5016	25/10/2001	27/12/2022

STATION	CODE	VALUES	START_DATE	END_DATE
SEIA	11K/01UG	4600	28/09/2001	16/01/2023
SEJÃES (OLIVEIRA DE FRADES)	09H/02UG	5083	02/03/2002	23/01/2023
SERTÃ	15I/01UG	4900	26/10/2001	25/10/2022
SILVARES	13K/02UG	4195	02/11/2001	20/12/2021
SOBRAL DA SERRA	10N/04UG	5229	10/10/2001	03/01/2023
SOBRAL DE MONTE AGRAÇO	19C/01UG	5988	15/09/2001	05/11/2023
SOBRAL DE SÃO MIGUEL	12K/01UG	3252	15/10/2005	20/12/2021
SOBRAL DO PICHORRO	09M/03UG	4380	27/09/2001	03/01/2023
SOURE	13F/01G	6195	29/09/2001	05/11/2023
SÁTÃO	09K/01G	7121	26/09/2001	01/11/2023
SÃO MARTINHO DAS MOITAS	08I/03UG	4633	21/06/2003	05/11/2023
SÃO MIGUEL DE ACHA	13N/01UG	5021	12/10/2001	27/12/2022
SÃO PEDRO DO SUL	09I/01C	4382	28/11/2001	30/10/2023
SÃO VICENTE DA BEIRA	13L/04UG	4517	24/08/2002	19/12/2022
TAMANHOS	09N/01UG	5820	27/09/2001	04/09/2023
TENTÚGAL	12F/01UG	7715	30/09/2001	14/09/2023
TOMAR	16G/01UG	5041	05/10/2002	29/11/2022
TORRES NOVAS	17F/05UG	2288	15/06/2002	18/12/2009
TORRES VEDRAS	19B/01UG	611	04/10/2001	27/09/2006
TOURO	08K/01UG	4756	28/12/2001	20/09/2022
TROUXEMIL	12G/04UG	4135	21/08/2002	25/01/2023
TURQUEL	17D/01UG	3995	05/10/2001	13/06/2017
TÁBUA	11I/04UG	5002	28/09/2001	16/01/2023
VALE AFONSINHO	08O/01UG	4205	01/11/2003	03/10/2022
VALE DE ESPINHO	12P/01UG	4633	11/10/2001	05/10/2022
VALE DO ROSSIM	11L/09G	3718	22/12/2005	30/01/2023
VALE SALGUEIRO	14E/03UG	4396	06/10/2001	05/07/2021
VALHELHAS	11M/01UG	4580	11/10/2001	06/12/2021
VARZIELAS	10H/02G	7375	24/10/2001	05/11/2023
VERMELHA	18C/03UG	1656	12/08/2004	08/04/2009
VILA DE REI	15I/02UG	4923	26/10/2001	21/12/2021
VILA NOVA DE POIARES	12H/02UG	1673	02/02/2002	15/11/2006
VILA NOVINHA	09M/01UG	1210	08/04/2016	20/09/2022
VILA VELHA DE RODÃO	16K/01G	6124	25/10/2001	05/11/2023
VILAR DE BESTEIROS	10I/02UG	4392	24/10/2001	29/06/2021
VILAR FORMOSO	10Q/01UG	4264	11/10/2001	05/05/2020
VIMEIRO (ALCOBAÇA)	17C/02UG	1036	05/10/2001	05/08/2004
VIMEIRO (LOURINHÃ)	18B/04UG	4083	04/10/2001	29/06/2021

Appendix 2 – Tools used in ArcGIS Pro

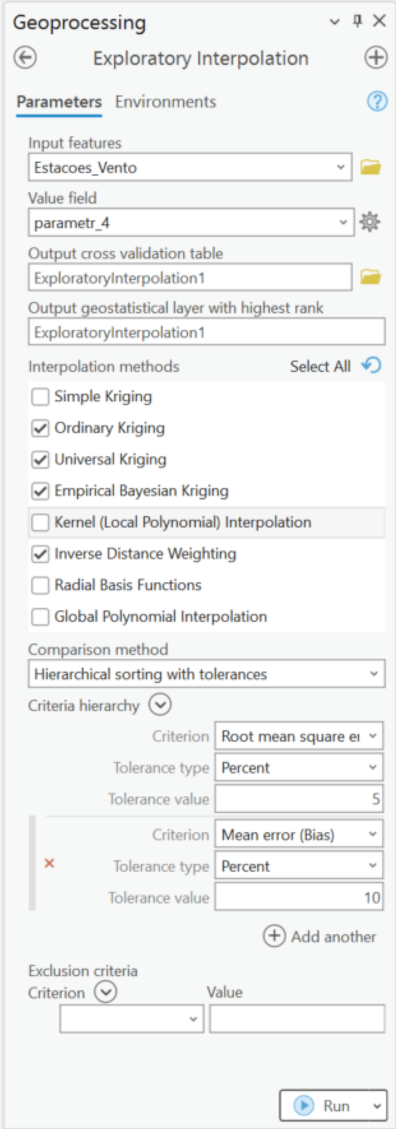


Figure 11. Exploratory Interpolation in ArcGIS Pro

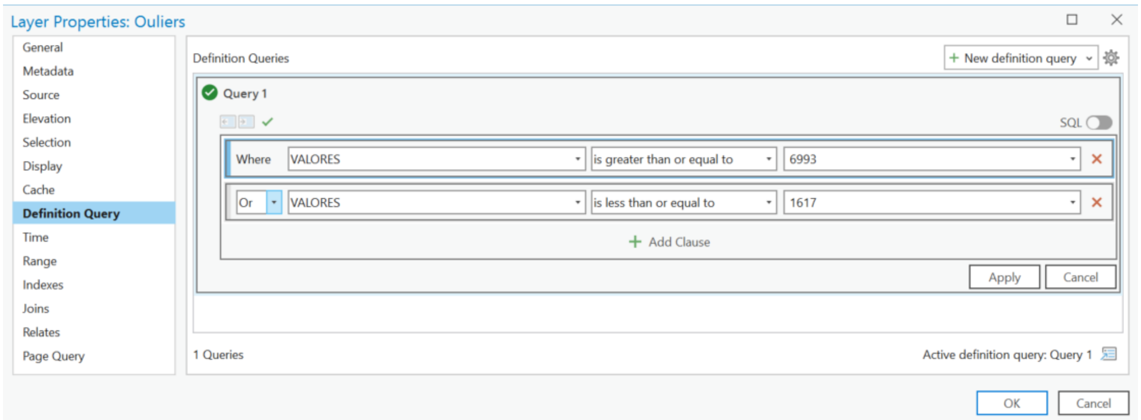


Figure 12. Query of the shapefile with the values of the average daily wind speed, to represent only the outliers

Ouliers

Field: Add Calculate **Selection:** Select By Attributes Zoom To Switch

	FID	Shape *	OBJECTID	ESTAÇÃO	CÓDIGO	VALORES	D
1	160	Point	161	TORRES VEDRAS	19B/01UG	611	0
2	28	Point	29	BRUFE (BARREIROS)	09K/03UG	618	2
3	77	Point	78	GOUVEIA	11L/01UG	921	1
4	178	Point	179	VIMEIRO (ALCOBAÇA)	17C/02UG	1036	0
5	174	Point	175	VILA NOVINHA	09M/01UG	1210	0
6	17	Point	18	ANSIÃO	14G/01C	1389	0
7	127	Point	128	PRAIA DE MIRA	11E/01C	1410	0
8	82	Point	83	LAGOA COMPRIDA	11L/07CG	1562	1
9	151	Point	152	SÁTÃO	09K/01G	7121	2
10	133	Point	134	REGO DA MURTA	15G/02G	7132	2
11	170	Point	171	VARZIELAS	10H/02G	7375	2
12	123	Point	124	PINHEL	09O/01G	7475	0
13	65	Point	66	ESCALHÃO	08P/02G	7479	1
14	81	Point	82	LADOEIRO	14N/02UG	7561	1
15	157	Point	158	TENTÚGAL	12F/01UG	7715	3
	Click to add new row.						

Figure 13. Resultado a Query da figura anterior

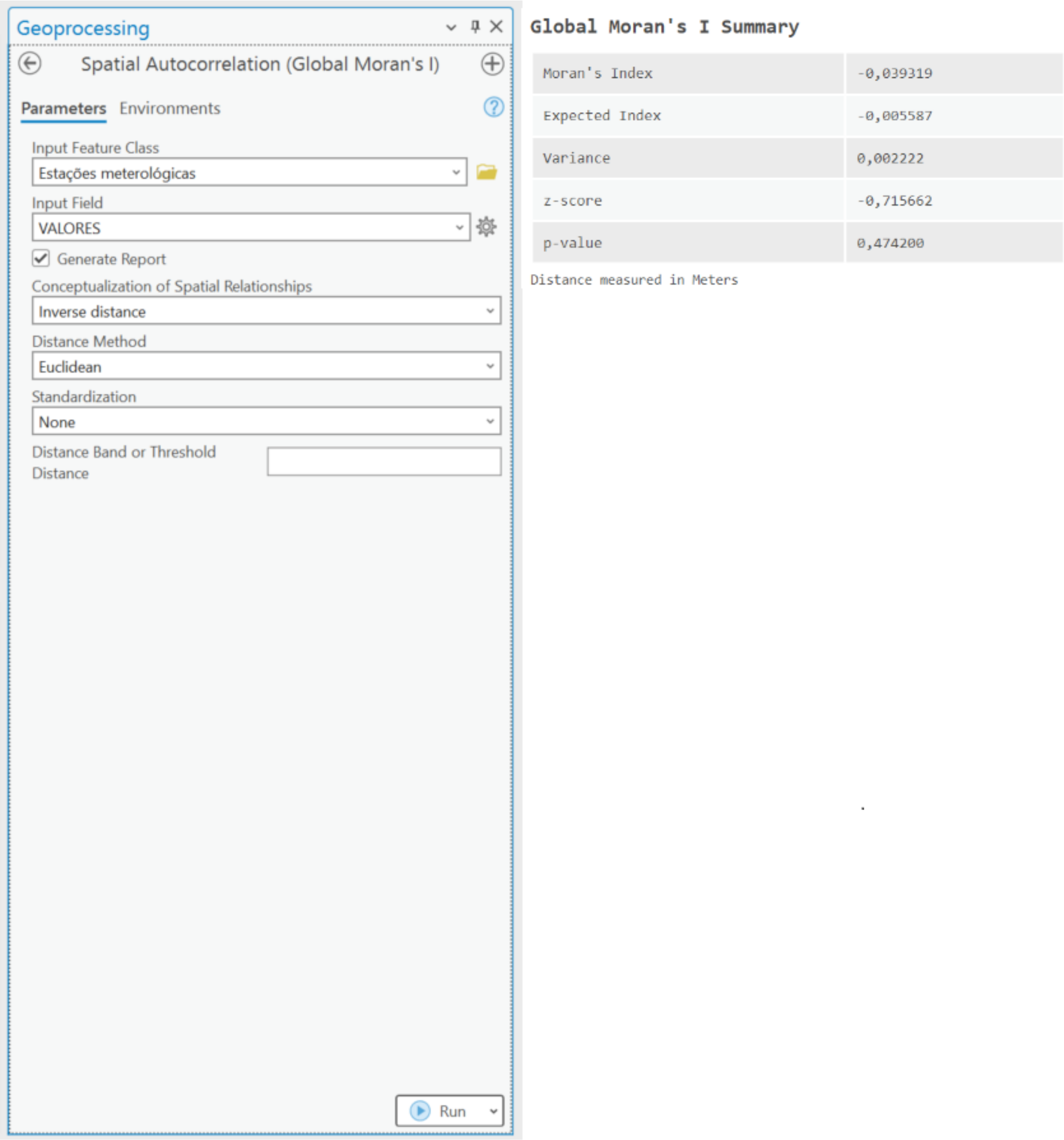


Figure 14. Spatial Autocorrelation (Global Moran's I)

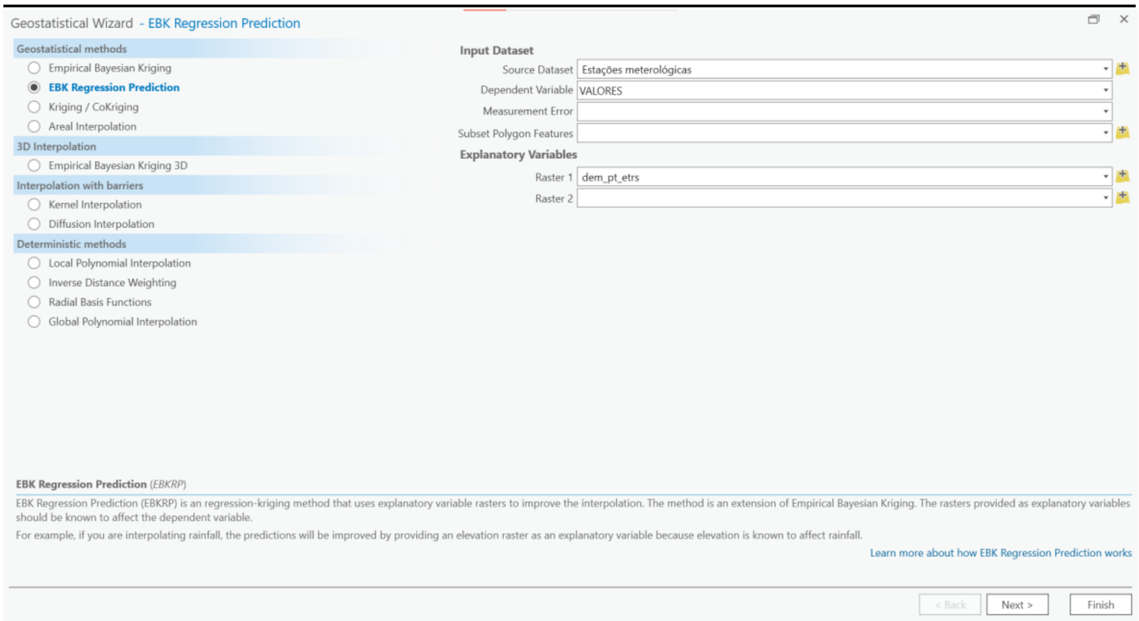
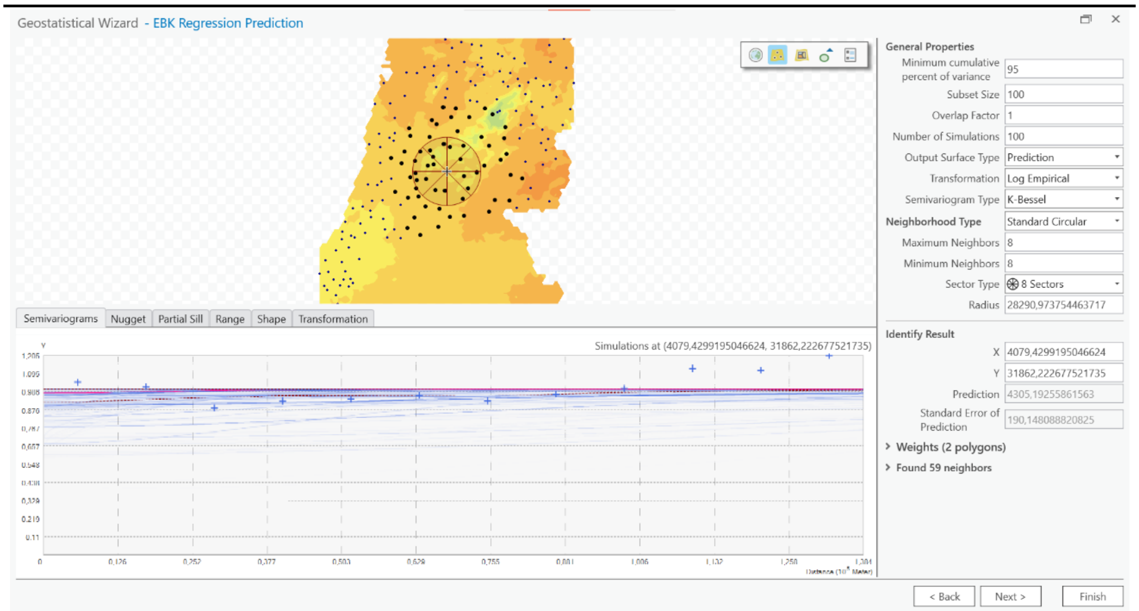
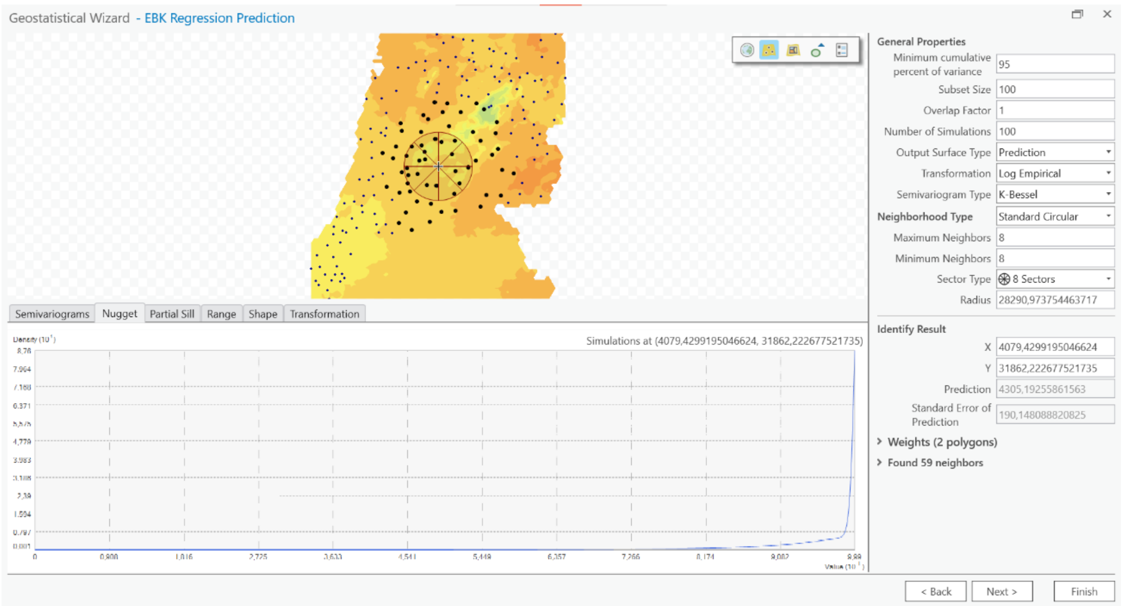
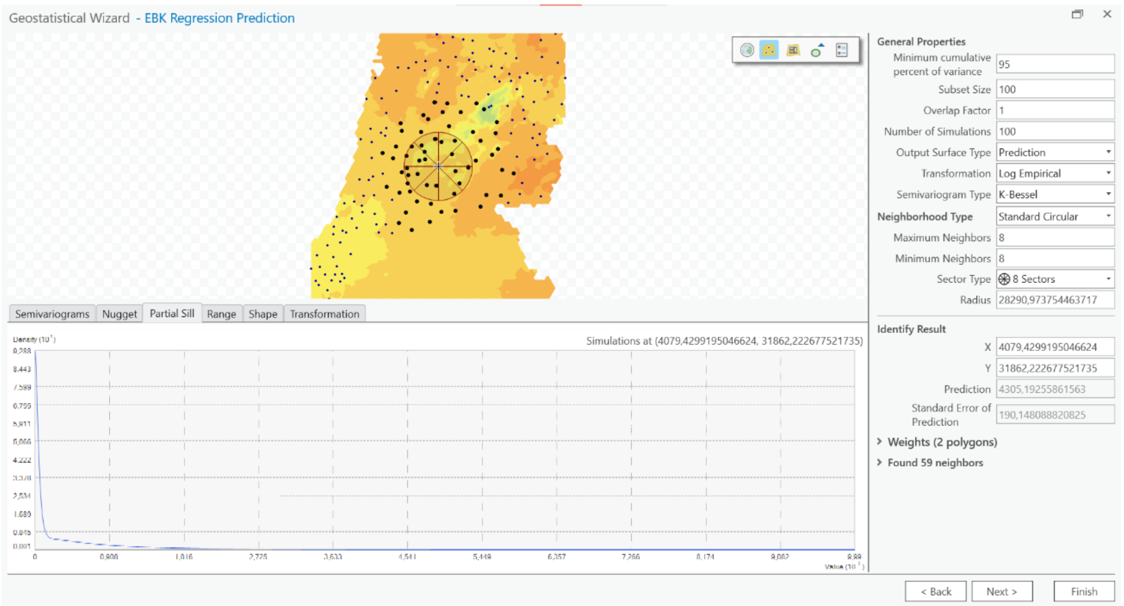
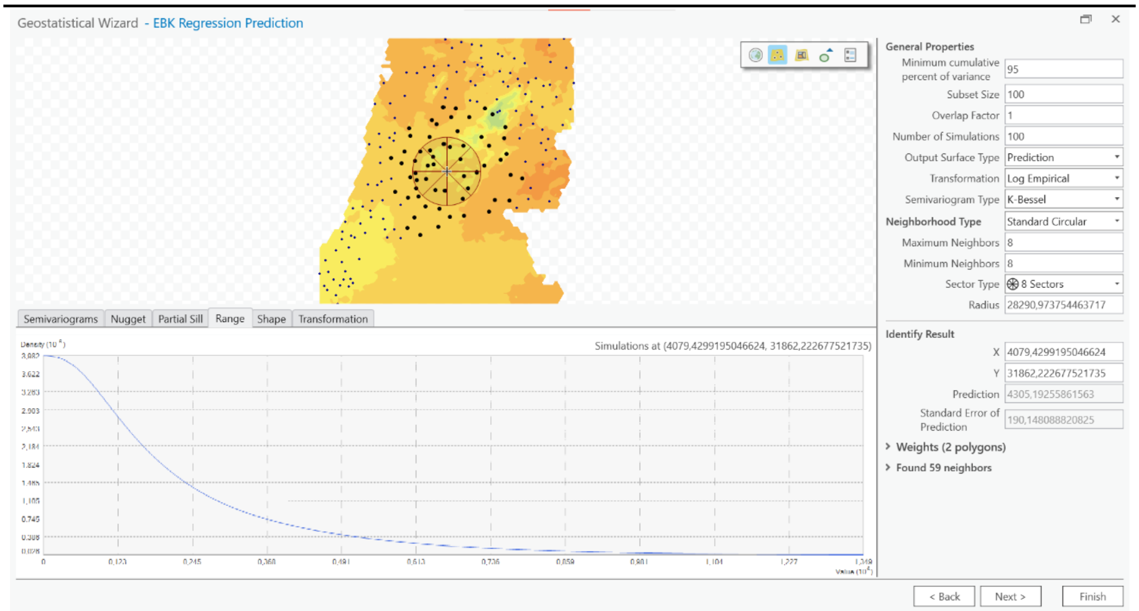
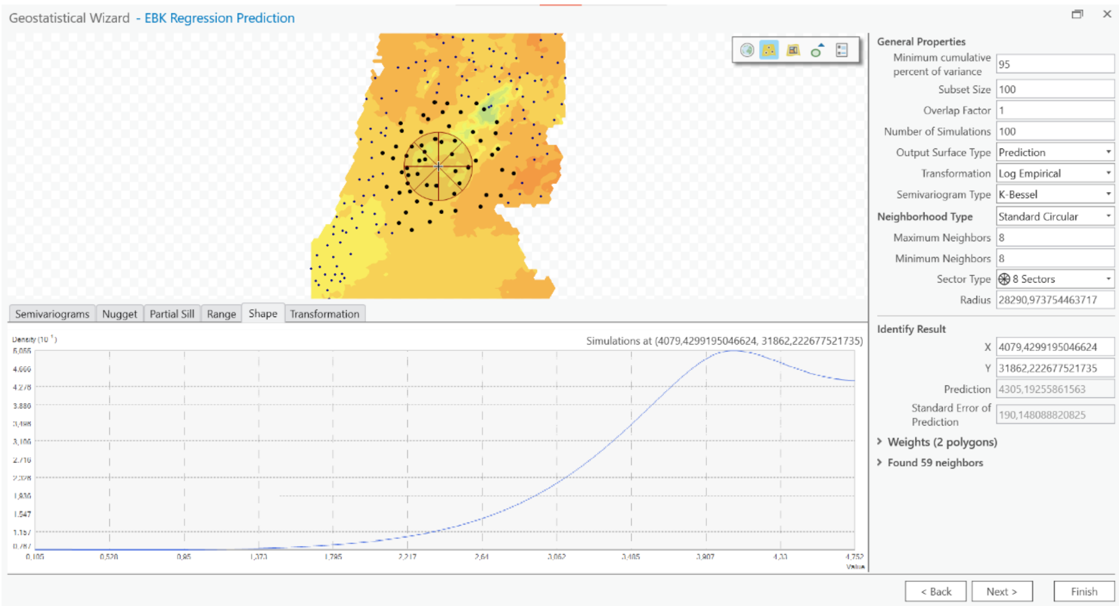
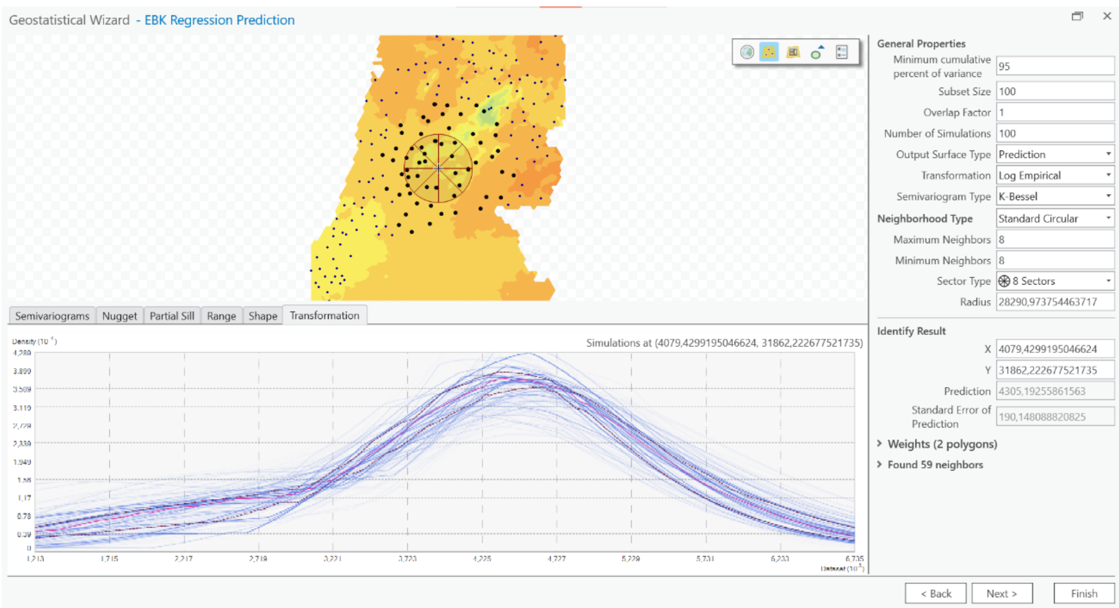


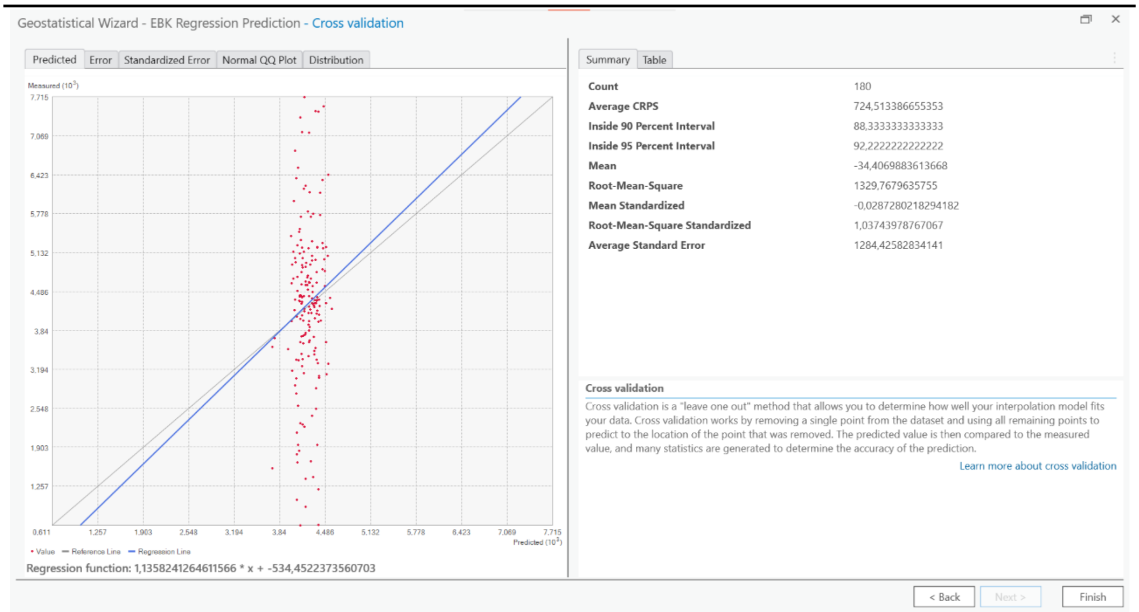
Figure 15. Geostatistical Wizard in ArcGIS Pro

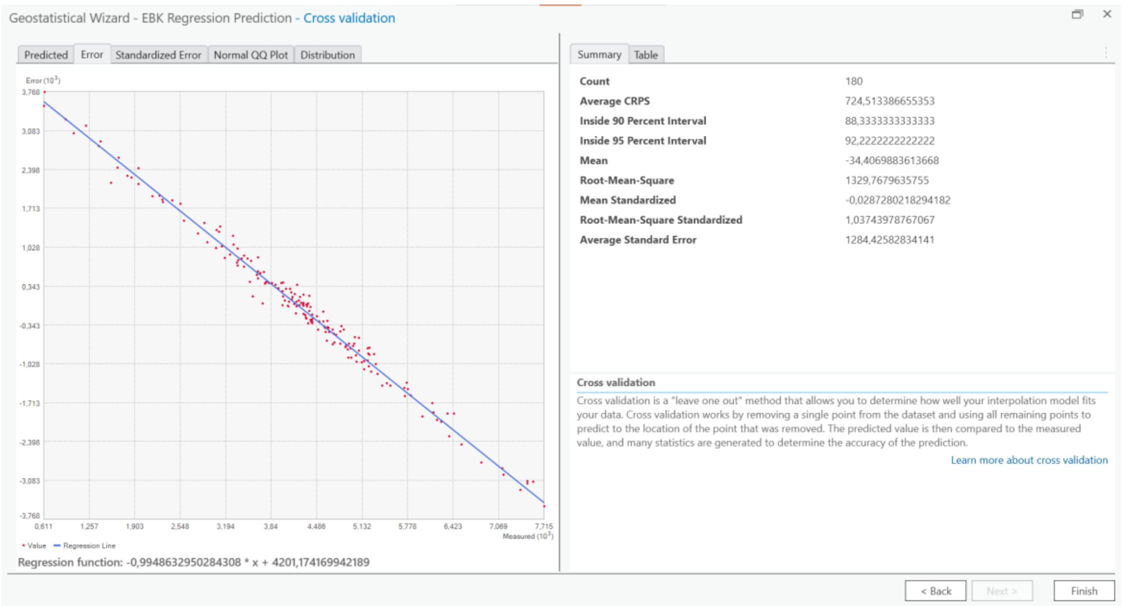

Geostatistical Wizard - EBK Regression Prediction
Semivariograms
Nugget
Partial Sill
Range
Shape
Transformation
Simulations at (4079,4299195046624, 31862,222677521735)
General Properties
Minimum cumulative percent of variance 95
Subset Size 100
Overlap Factor 1
Number of Simulations 100
Output Surface Type Prediction
Transformation Log Empirical
Semivariogram Type K-Bessel
Neighborhood Type Standard Circular
Maximum Neighbors 8
Minimum Neighbors 8
Sector Type 8 Sectors
Radius 28290,973754463717
Identify Result
X 4079,4299195046624
Y 31862,222677521735
Prediction 4305,19255861563
Standard Error of Prediction 190,148088820825
Weights (2 polygons)
Found 59 neighbors
< Back
Next >
Finish


Geostatistical Wizard - EBK Regression Prediction
Semivariograms
Nugget
Partial Sill
Range
Shape
Transformation
Simulations at (4079,4299195046624, 31862,222677521735)
General Properties
Minimum cumulative percent of variance 95
Subset Size 100
Overlap Factor 1
Number of Simulations 100
Output Surface Type Prediction
Transformation Log Empirical
Semivariogram Type K-Bessel
Neighborhood Type Standard Circular
Maximum Neighbors 8
Minimum Neighbors 8
Sector Type 8 Sectors
Radius 28290,973754463717
Identify Result
X 4079,4299195046624
Y 31862,222677521735
Prediction 4305,19255861563
Standard Error of Prediction 190,148088820825
Weights (2 polygons)
Found 59 neighbors
< Back
Next >
Finish


Geostatistical Wizard - EBK Regression Prediction
Semivariograms
Nugget
Partial Sill
Range
Shape
Transformation
Simulations at (4079,4299195046624, 31862,222677521735)
General Properties
Minimum cumulative percent of variance 95
Subset Size 100
Overlap Factor 1
Number of Simulations 100
Output Surface Type Prediction
Transformation Log Empirical
Semivariogram Type K-Bessel
Neighborhood Type Standard Circular
Maximum Neighbors 8
Minimum Neighbors 8
Sector Type 8 Sectors
Radius 28290,973754463717
Identify Result
X 4079,4299195046624
Y 31862,222677521735
Prediction 4305,19255861563
Standard Error of Prediction 190,148088820825
Weights (2 polygons)
Found 59 neighbors
< Back
Next >
Finish


Geostatistical Wizard - EBK Regression Prediction
General Properties
Minimum cumulative percent of variance 95
Subset Size 100
Overlap Factor 1
Number of Simulations 100
Output Surface Type Prediction
Transformation Log Empirical
Semivariogram Type K-Bessel
Neighborhood Type Standard Circular
Maximum Neighbors 8
Minimum Neighbors 8
Sector Type 8 Sectors
Radius 28290,973754463717
Semivariograms
Nugget
Partial Sill
Range
Shape
Transformation
Simulations at (4079,4299195046624, 31862,222677521735)
Identify Result
X 4079,4299195046624
Y 31862,222677521735
Prediction 4305,19255861563
Standard Error of Prediction 190,148088820825
Weights (2 polygons)
Found 59 neighbors
< Back
Next >
Finish

Geostatistical Wizard - EBK Regression Prediction
General Properties
Minimum cumulative percent of variance 95
Subset Size 100
Overlap Factor 1
Number of Simulations 100
Output Surface Type Prediction
Transformation Log Empirical
Semivariogram Type K-Bessel
Neighborhood Type Standard Circular
Maximum Neighbors 8
Minimum Neighbors 8
Sector Type 8 Sectors
Radius 28290,973754463717
Semivariograms
Nugget
Partial Sill
Range
Shape
Transformation
Simulations at (4079,4299195046624, 31862,222677521735)
Identify Result
X 4079,4299195046624
Y 31862,222677521735
Prediction 4305,19255861563
Standard Error of Prediction 190,148088820825
Weights (2 polygons)
Found 59 neighbors
< Back
Next >
Finish

Geostatistical Wizard - EBK Regression Prediction
General Properties
Minimum cumulative percent of variance 95
Subset Size 100
Overlap Factor 1
Number of Simulations 100
Output Surface Type Prediction
Transformation Log Empirical
Semivariogram Type K-Bessel
Neighborhood Type Standard Circular
Maximum Neighbors 8
Minimum Neighbors 8
Sector Type 8 Sectors
Radius 28290,973754463717
Semivariograms
Nugget
Partial Sill
Range
Shape
Transformation
Simulations at (4079,4299195046624, 31862,222677521735)
Identify Result
X 4079,4299195046624
Y 31862,222677521735
Prediction 4305,19255861563
Standard Error of Prediction 190,148088820825
Weights (2 polygons)
Found 59 neighbors
< Back
Next >
Finish


Geostatistical Wizard - EBK Regression Prediction - Cross validation
Predicted
Error
Standardized Error
Normal QQ Plot
Distribution
Summary
Table
Count 180
Average CRPS 724,513386655353
Inside 90 Percent Interval 88,3333333333333
Inside 95 Percent Interval 92,2222222222222
Mean -34,4069883613668
Root-Mean-Square 1329,7679635755
Mean Standardized -0,0287280218294182
Root-Mean-Square Standardized 1,03743978767067
Average Standard Error 1284,42582834141
Cross validation
Cross validation is a "leave one out" method that allows you to determine how well your interpolation model fits your data. Cross validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.
Learn more about cross validation
Regression function: 1,1358241264611566 * x + -534,4522373560703
< Back
Next >
Finish


Geostatistical Wizard - EBK Regression Prediction - Cross validation
Predicted
Error
Standardized Error
Normal QQ Plot
Distribution
Summary
Table
Count 180
Average CRPS 724,513386655353
Inside 90 Percent Interval 88,3333333333333
Inside 95 Percent Interval 92,2222222222222
Mean -34,4069883613668
Root-Mean-Square 1329,7679635755
Mean Standardized -0,0287280218294182
Root-Mean-Square Standardized 1,03743978767067
Average Standard Error 1284,42582834141
Cross validation
Cross validation is a "leave one out" method that allows you to determine how well your interpolation model fits your data. Cross validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.
Learn more about cross validation
Regression function: -0,9948632950284308 * x + 4201,174169942189
< Back
Next >
Finish

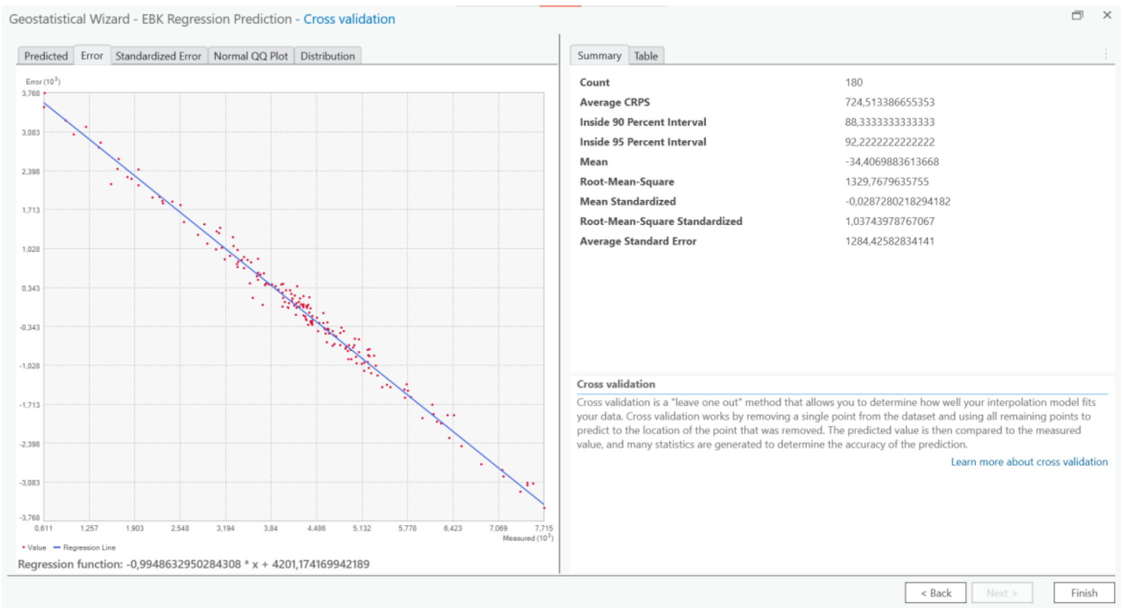

Geostatistical Wizard - EBK Regression Prediction - Cross validation
Predicted
Error
Standardized Error
Normal QQ Plot
Distribution
Summary
Table
Count 180
Average CRPS 724,513386655353
Inside 90 Percent Interval 88,3333333333333
Inside 95 Percent Interval 92,2222222222222
Mean -34,4069883613668
Root-Mean-Square 1329,7679635755
Mean Standardized -0,0287280218294182
Root-Mean-Square Standardized 1,03743978767067
Average Standard Error 1284,42582834141
Cross validation
Cross validation is a "leave one out" method that allows you to determine how well your interpolation model fits your data. Cross validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.
Learn more about cross validation
Regression function: -0,9948632950284308 * x + 4201,174169942189
< Back
Next >
Finish

Figure 16. EBK Regression Prediction results for model with 8 sectors and 8 neighbors

Appendix 3 – Results of adjustments to the EBK Regression model

Table 4. Results of adjustments to the EBK Regression model

N.º Neighbors		**3-5**				**5-8**	**8**	**10-15**		
N.º Sectors		1	4	4 with 45*	8	8	8	1	4	4 with 45º
EBK Regression Prediction	Prediction	4183,451179	4243,602603	4243,602603	4278,660499	4274,755453	4305,192559	4245,72269	4280,059405	4299,278878
	Standard Error of Prediction	305,7575933	202,7792882	202,7792882	197,2953304	190,9200638	190,1480888	208,3628357	189,8208997	193,212867
Cross Validation	Count	180	180	180	180	180	180	180	180	180
	Average CRPS	728,3712407	723,5751656	724,7699534	725,5659529	723,2262051	724,5133867	722,025245	723,0691487	722,7624894
	Inside 90 Percent Interval	88,88888889	88,88888889	88,88888889	88,33333333	88,33333333	88,33333333	88,88888889	88,33333333	88,33333333
	Inside 95 Percent Interval	92,22222222	92,22222222	92,22222222	92,22222222	92,22222222	92,22222222	92,22222222	92,22222222	92,22222222
	Mean	-37,93539115	-37,7449322	-35,39770992	-38,53335659	-35,77038811	-34,40698836	-35,54646349	-38,40241884	-32,21620653
	Root-Mean-Square	1335,812201	1327,916716	1330,164365	1331,586619	1326,992323	1329,767964	1324,295142	1326,970699	1326,406882
	Mean Standardized	-0,030792289	-0,031222292	-0,028817589	-0,031775295	-0,029929192	-0,028728022	-0,029174426	-0,031866936	-0,027190891
	Root-Mean-Square Standartized	1,0315865	1,031331058	1,033440906	1,035987072	1,034343747	1,037439788	1,02912718	1,033824477	1,03412198
	Average Standard Error	1298,776967	1290,249413	1290,11672	1287,757297	1285,552308	1284,425828	1290,011769	1285,912888	1284,829066