Mundo Fesc E-ISSN: 2216-0388 P- ISSN: 2216-0353 Artículo Original

Applying the kriging geostatistical technique to model parameters of water quality, TSS, and pH in a section of the Quindío river, Colombia

Aplicando la técnica kriging para modelar parámetros de calidad del agua, sólidos en suspensión TSS y pH en una sección del río Quindío, Colombia

Recibido: 26 de agosto de 2022
Aprobado: 4 de diciembre de 2022

Forma de citar: M. J. Mesa Mazo, C. Cubides, A. M. Pulgarín Galvis, C. A. Acosta Minoli,"applying the kriging geostatistical technique to model parameters of water quality, TSS, and pH in a section of te Quindia river, Colombia", *Mundo Fesc*, vol. 13, no. 25, pp. 289-305, 2023. https://doi.org/10.61799/2216-0388.1284

Mónica Jhoana Mesa-Mazo*
Doctora en Ingeniería
mmesa4@cue.edu.co
Corporación Universitaria Empresarial Alexander von Humbold
Armenia Colombia

Carlos Eduardo Cubides Burbano
Maestría en Ingeniería
carloscubides22@gmail.com
Universidad del Quindío
Popayán Cauca.

Alejandra María Pulgarín-Galvis
Maestría en Biomatemáticas, alejandra.pulgarin@correounivalle.edu.co, Universidad Del Valle, Caicedonia Valle del Cauca, https://orcid.org/0000-0001-6873-9432, Armenia Quindío, Colombia.

Cesar Augusto Acosta-Minoli
Doctor en Matemáticas
cminoli@uniquindio.edu.co
Universidad del Quindío
Quindío, Colombia

Jorge Mario García-Usuga
Doctor en Ingeniería
jmgarcia@uniquindio.edu.co
Universidad del Quindío
Quindío Colombia.

*__Autor para correspondencia:__
mmesa4@cue.edu.co

Aplicando la técnica kriging para modelar parámetros de calidad del agua, sólidos en suspensión TSS y pH en una sección del río Quindío, Colombia

Resumen

Antecedentes: El Río Quindío, principal fuente de agua en el Departamento de Quindío, se ve afectado por la contaminación, especialmente de origen doméstico, agrícola e industrial, siendo la zona de La María una de las más perjudicadas. **Objetivo:** Este estudio se propuso recopilar datos durante la temporada de lluvias en el sector de La María para analizar la calidad del agua en términos de pH y sólidos totales en suspensión (TSS). Se empleó la técnica de Kriging para estimar estos parámetros en puntos no muestreados, considerando la dependencia espacial y temporal de los datos. **Métodos:** Se llevaron a cabo muestreos en diferentes puntos a lo largo del río y en distintos momentos. La técnica de Kriging se aplicó para analizar y estimar los valores de pH y sólidos totales en suspensión en ubicaciones no muestreadas, minimizando errores y permitiendo una validación cruzada. **Resultados:** Los errores más pequeños se obtuvieron cuando las muestras de pH y sólidos totales en suspensión se ajustaron a una distribución normal, destacando la eficacia del Kriging en datos con fuerte autocorrelación espacial positiva. Esto garantizó una fiabilidad ideal en los resultados de estimación para pH y sólidos totales en suspensión en ubicaciones no muestreadas. Conclusión: La aplicación exitosa de la técnica de Kriging para estimar la calidad del agua en puntos no muestreados destaca su utilidad en la gestión de la incertidumbre en parámetros críticos para la calidad del agua, como el pH y los sólidos totales en suspensión, contribuyendo así a estrategias efectivas de preservación y manejo de recursos hídricos.

Palabras clave: Potencial de hidrógeno; sólidos suspendidos totales; método de interpolación Kriging, autocorrelación espacial; ríos; recursos hídricos.

Mónica Jhoana Mesa-Mazo, Carlos Cubides, Alejandra María Pulgarín-Galvis , Cesar Augusto Acosta-Minoli, Jorge Mario García-Usuga.

Applying the kriging geostatistical technique to model parameters of water quality, TSS, and pH in a section of the Quindío river, Colombia

Abstract

Background: The Quindío River, a crucial water source in the Quindío Department, faces pollution issues, particularly from domestic, agricultural, and industrial discharges, with the La María area being notably affected. **Objective:** This study aimed to collect data during the rainy season in the La María sector to analyze water quality in terms of potential of hydrogen (pH) and total suspended solids (TSS). Kriging technique was employed to estimate these parameters at non-sampled points, considering spatial and temporal data dependence. **Methods:** Sampling was conducted at various points along the river and different time intervals. Kriging analysis was applied to estimate potential of hydrogen and total suspended solids values at non-sampled locations, minimizing errors and allowing for cross-validation. **Results:** Smallest errors were obtained when potential of hydrogen and total suspended solids samples adhered to a normal distribution, highlighting Kriging's effectiveness in spatially autocorrelated data. This ensured ideal reliability in estimated results for pH and total suspended solids at non-sampled locations. Conclusions: The successful application of Kriging for estimating water quality in non-sampled points underscores its utility in managing uncertainty in critical water quality parameters like potential of hydrogen and total suspended solids, contributing to effective strategies for water resource preservation and management.

Keywords: Potential of hydrogen; total suspended solids; Kriging interpolation method, spatial autocorrelation; rivers; water resources.

Introduction

Water is a renewable, but fine, resource; it covers > 70% of the surface of the Earth. It is an element of great importance for life and a determining factor to comply with the physical, chemical, and biological processes that govern the natural environment, however, with 97.5% being salt water, only close to 2.5% is associated with fresh water. This liquid fulfills a transcendentally important role in the survival of living beings and development of society. It is essential for sustainability and growth of life, constituting a determining factor in biological processes that provide life to entire ecosystems that depend mainly on water [1]. The Quindío River is the principal supply source for the department of Quindío, its main channel rises to the northwest of the department at a height of 3780 masl and ends at a height of 1040 masl, whose channel flows in a southwesterly direction. This hydrographic unit is related directly with the municipalities of Salento, Armenia, Calarcá, and La Tebaida. However, nearly 300,000 people live on the river bank in the sector of La María. Water pollution is produced by factors directly related with the human metabolism and activities of domestic use such as dumping of sewage, food waste, paper, glass, and organic material. This includes factors generated by industrial-type activities; the presence of tanneries and slaughter plants, livestock, pig farming, and – lastly – that associated with mining given by extraction of construction materials, and use of heavy metals, like lithium, used in product refining. These activities seriously compromise the supply of drinking water, causing diseases, like hepatitis, gastroenteritis, and typhoid fever, affecting crops due to the increase of mineral salts from irrigation systems, and leading animal species to high mortality rates [6]. Given the vast importance of water for humans and the diverse species inhabiting the study zone, it becomes necessary to study its quality, where different types of analyses exist – among them, the microbiological analysis that encompasses total coliform detection; the physical-chemical analysis that ranges from measuring pH, nitrites, nitrates, turbidity to absorbance; and, lastly, aesthetic analysis that consists in determining parameters, like temperature, color, odor, flavor, salinity, and total suspended solids.

Quality indicators permit identifying two intrinsic aspects of water; what it contains and in what amount. Many parameters exist to evaluate water quality, but this study specifically collected information on the pH and TSS indices, which are described ahead:

Concentration of hydrogen ions (pH): it is a unit of measure of alkalinity or acidity of a solution, measuring the amount of positive hydrogen ions present in a substance. Analysis of pH is fundamental to characterize water quality, allowing to identify the acid or alkaline conditions of a current, where extreme can seriously affect aquatic flora and fauna [1].

Total suspended solids (TSS): refers to the particulate matter maintained in suspension on surface water currents; its determination quite important for river ecosystems and water quality, given that suspended inorganic solids attenuate light through the dispersion phenomenon, affecting the photosynthesis process of aquatic flora and the capacity of fish to acquire food [2].

This research project used the Kriging interpolator, a method approximate and stochastic, given that it incorporates the effect of spatially uncorrelated residuals [3]; determines the value of the elevation supported by calculating the average of the closest elevations. Local-type interpolators are used due to their capacity to evaluate elevation variation at short distances by linking information from nearest neighbors. These are also used massively due to their low demand with computer machines [4]. This geostatistical method is of greatest use within the type of probabilistic interpolators based on the concept of regionalized random variable, where the value of an attribute tends to be dependent of its spatial location [5], [6], [7] and [8]. For which, the semivariogram is used instead of the standard deviation to analyze spatial variability, it is used to describe the correlation between observations as the distance increases, meaning that nearby observations are more similar than distant ones [9], [10], [11] and [12].

The notion of the spatial autocorrelation of these variables is associated with the idea the values observed in adjacent geographic areas will be more similar than those expected under the assumption of linear independence [5] and [13]. The distribution analysis of a variable must consider the formation of patterns in function of proximity relations, given that these can affect, in non-random manner, other peculiarities of the same place.

A method exists to corroborate the result of the predictions made by spatial interpolation algorithms [14], it is known as Cross Validation and consists in eliminating an elevation value, executing the interpolation algorithm, and estimating the value eliminated.

Materials and Methods

Study area

Data collection took place during the 2018 rainy season in a section of the Quindío River (sector of La María), a place chosen because of the high concentration of residential and industrial structures located on the riverbank, in the department of Quindío, Colombia. The sampling was conducted between October and December for a total of eight weeks where 20 monitoring stations were established, which captured information by following rigorous temporal pattern, which permitted storing 160 pH and TSS samples.

Fig 1 shows the exact geographic location where the 20 monitoring stations were set up. It also shows the stretch of the river involved in the study, together with the municipalities of Armenia (Colombia) and Calarcá, which depend directly on the water tributary and add contamination, and finally identifies the settlement known as "La María", where activities such as material extraction, slaughter plants and tanneries are carried out, another contributor to the decrease in water quality.

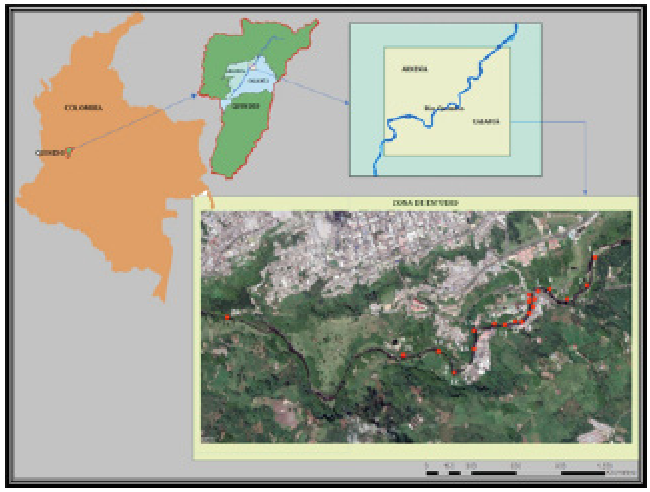

Figure 1. La María Sector in Armenia Quindío Colombia South America. (own source)

Descriptive statistical analysis

The Statgraphics Centurion XVII open software was used to perform a statistical summary for all the pH and TSS sets to discard the presence of atypical data, defining variables. In addition, normality tests were conducted, like Shapiro-Wilk, Kolmogorov-Smirnov, and Chi-squared with 5% significance level, to determine if the data fit a Gaussian distribution. Classification of training groups

Through the subset module, the original data were divided into two subsets, both for the specific samples of pH and TSS for the eight weeks and the 20 stations in which the information was collected. A subset corresponding to 90% of the data (calibration) will be used to generate the interpolation surfaces and the remaining 10% will be used as control points - validation -.

Interpolation with probabilistic algorithms

For each of the training sets (90%), for PH and TSS, ArcGIS 10.7 open software was used – common within the SIG community, through Geostatistical Analyst module. Likewise, this software permitted calculating the Moran and Getis-Ord indices to identify the way data from the pH and TSS variables radiate through the spatial units, - spatial autocorrelation -.

Calculation of the interpolation error

Cross validation was conducted by comparing the values taken as real from the validation sets and those estimated by the training sets. Thereafter, by using MSE and ME, the quality of the interpolation algorithm will be assessed.

Results

Data exploratory analysis

As shown in Table 1.a) no data symmetry exists, given that the mean, median, and mode do not coincide in their values, it is assumed that this distribution has negative asymmetry, although somewhat imperceptible. Furthermore, the pointing coefficient is close to zero (mesokurtic), inherent in a normal distribution. It indicates the possibility of adjusting the pH data in week 5 to a normal distribution. Additionally, Table 1.b) likewise, shows that the mean and median have different values, indicating a slight trend of the distribution toward asymmetry or bias to the right, besides, the pointing coefficient is somewhat distant from zero, with which it is not possible to affirm that TSS data for week 4 adhere to the normal distribution. To corroborate normality, the Chi square, Kolmogorov-Smirnov, and Shapiro-Wilks tests were used.

Table I. Descriptive statistical analysis. a) exploratory summary pH variable week 5. b) exploratory summary TSS variable week 4. (own source)

(a) pH 5 Parameter	Units	value	(b) TSS 4 Parameter	Units	value
Count	und	18	Count	Und	18
Average	m	7.841	Average	M	275.18
Median	m	7.825	Median	M	277.05
Mode	m	7.81	Mode	M	
Variance	m	0.00902	Variance	M	2960.75
standard deviation	m	0.09498	standard deviation	M	54.41
variation coefficient	dimensionless	1.211%	variation coefficient	dimensionless	19.77
standard error	m	0.022	standard error	M	12.83
Minimum	m	7.63	Minimum	M	172
Maximum	m	8.02	Maximum	M	407.7
Range	m	0.39	Range	M	235.7
Bias	m	-0.021	Bias	M	0.396
Kurtosis	dimensionless	0.367	Kurtosis	dimensionless	1.083

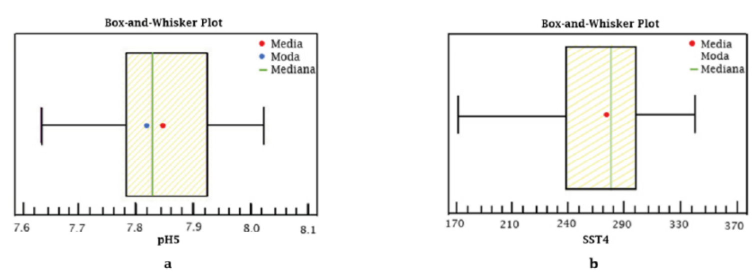

Figure 2. Box and whisker plots. a) Box and whisker plot pH week 5. b) Box and whisker plot TSS variable week 4. (own source)

From Figure 2.a), it may be noted that 50% of pH values between 7.78 and 7.92 are comprised between the 25th and 75th quartiles. The 50th quartile, located exactly on the median value (green vertical line) within the box, shows clearly how large data appear more frequently than the small data, due to negative asymmetry; the extreme values on the whiskers mark the minimum and maximum values for pH and indicate the absence of atypical values that are likely caused due to mistakes - gross errors - .Likewise, Figure 2.b) indicates the absence of atypical data beyond the extreme values on the whiskers for the TSS variable corresponding to week 4. Where one, non-symmetry can be seen due to the absence of the median (50th quartile) in the box's geometric center.

Figure 3.a) and Figure 3.b) indicate that collected observations stacked in bars indicate the value and number of times with which data on the pH variable on week 5 and TSS on week 4.

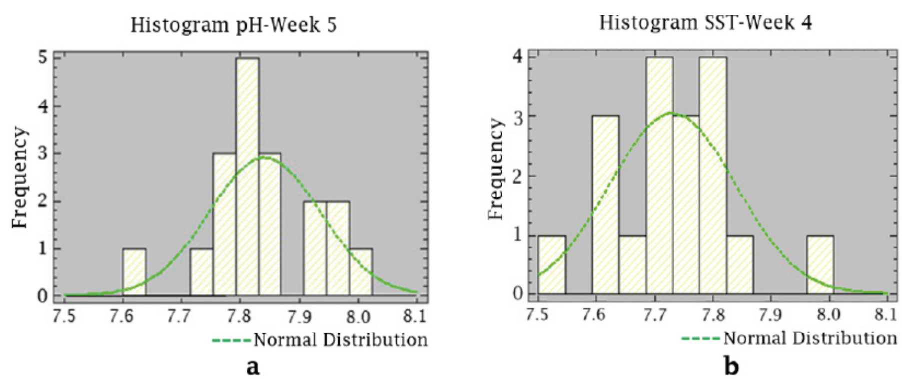

Figure 3. Histogram + normal function. a) Frequency diagram pH week 5. b) Frequency diagram TSS week 4. (own source)

Figure 3.a) shows how the adjustment function (dotted green line) underestimates the pH values from approximately 7.6 to 7.65, to then overestimate from (7.7, 7.76), then underestimates from (7,76, 7.87), thereafter overestimates from 7.9 to 7.93 to finally underestimate from 7.93 to 8.04. Likewise, the diagram in Figure 3.b) showed that the adjustment function underestimates in the interval (7.5, 7.74), as well as for interval (7.76, 7.82), and overestimates in the interval (7.82, 8.02).

The data exploratory analysis suggests a possible trend of the data for pH week 5 and TSS week 4 to adjust to the normal distribution. To validate this affirmation, three tests were performed: Chi-squared, Kolmogorov-Smirnov, and Shapiro-Wilk to corroborate the results shown in the statistical summary.

Table 2. Normality test for pH and TSS. Table a1, b1 and c1 for pH week 5. Table a2, b2 and c2 for TSS week 4. (own source)

Normality Test pH					Normality Test TSS	
Chi-Square	1.3333	a1)		a2)	Chi-Square	1.3333
P-Valor	0.7212				P-Valor	0.7212
Chi-Square	1.3333	b1)		b2)	Chi-Square	1.3333
P-Valor	0.7212				P-Valor	0.7212
Estadístico	0.9704	c1)		c2)	Estadístico	0.9704
P-Valor	0.7876				P-Valor	0.7876

Tables 2.a1), 2.b1) and 2.c1), corresponding to the pH variable on week 5, for 5% significance level, permitted seeing clearly how the p value for the three tests is > 0.05; referring through hypothesis tests to the non-rejection of the null hypothesis, Ho, - the data come from a normal distribution - with 95% confidence interval. Similarly, for Tables 2.a2), 2.b2), and 2.c2), related with the TSS variable week 4, the p value for the three tests is > 0.05, supporting statistically the fit of the data to a normal distribution.

It is notable that Table 2 only indicates statistical results corresponding to two data sets for one week of information (pH and TSS), of the eight weeks collected, given the large amount of information obtained after two months of study. Hence, a brief statistical description was made of the behavior of data from the seven weeks remaining for both variables.

Data for pH3, pH5, and pH6, corresponding to weeks 3, 5, and 6, respectively, fit a normal distribution according to tests performed, and for pH1, pH2, pH4, pH7, and pH8, corresponding to weeks 1, 2, 4, 7, and 8, respectively, did not fit a Gaussian distribution. Moreover, SST1, SST4, SST5; SST6, and SST7, associated to weeks 1, 4, 5, 6, and 7, respectively, showed adjustment to the normal distribution, while SST2, SST3, and SST8 did not fit said Gaussian theoretical distribution.

Interpolation, ME and MSE

Table III. indicates the linear regression equations adjusted to the data observed and which govern the linear interpolation models and additionally highlights one of the advantages of the Kriging method, which permits obtaining information of the estimation variance (prediction error).

Kriging models	pH week 5	TSS week 4
Prediction	pHpredicted=0,5785*pHobserved+3,3107	SSTpredicted=0,4574*SSTobserved+146,1193
Error	ErrorpHpredicted=-0,4215*pHobserved+3,3107	ErrorSSTpredicted=-0,5426*SSTobserved+146,1193

In Figure 4.a) and 4.c), the pH and TSS graphics, respectively, show how colors represent the ranges between which the variables move. Figure 5.a shows that the pH value does not move uniformly as the information collection advances from week to week. In addition, Figure 4.c. indicates how the TSS values vary homogeneously from week 1 to week 8. Figures 4.b and 4.d show the sites where the estimation is more precise and how the variance tends to increase when it is farther from the sampling points.

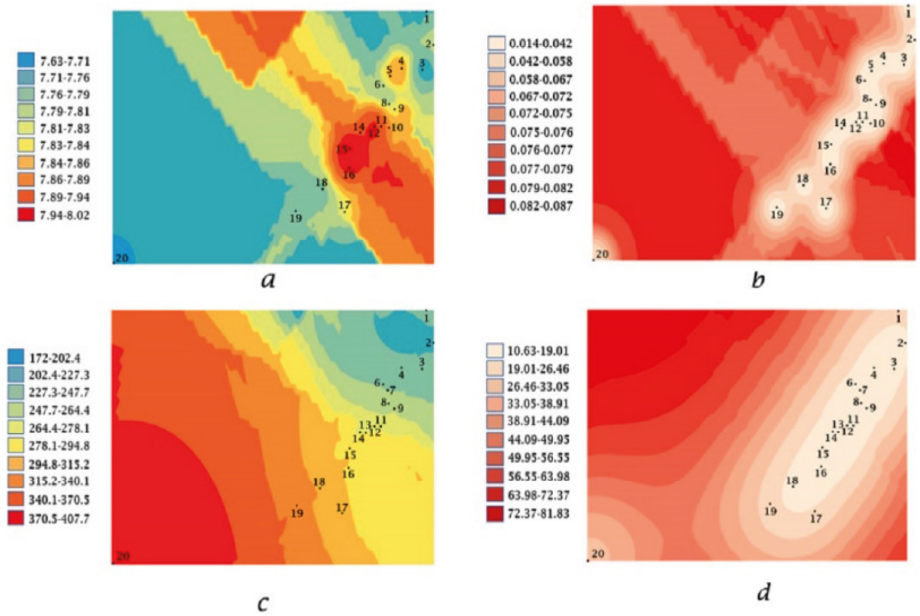

Figure 4. pH and TSS prediction and error surfaces, using Ordinary Kriging. Figures 4.a and 4.b show prediction and error surfaces for pH week 5. Figures 4.c and 4.d show prediction and error surfaces for TSS week 4. (own source)

Discussion

The ME average residues, as a result of interpolation with the ordinary Kriging method, for each training set corresponding to the eight weeks during which information was collected, enable complete statistical evaluation and more detailed analysis of the error behavior [15]. This lies specifically in the investigation of the introduction of systematic errors by processes to which the data were subjected (interpolation), this is why these residues indicate the presence or not of bias in the predictions (under- and over-estimation). Highlighting the high or little fit of the data to the normal curve, additionally, the standard deviation of the residues focuses on identifying the validity of the theory proposed by the authors [16] and [17], which states that: "*a mean of residuals close to zero leads to affirming that the standard deviation of the sample will have a value approximately equal to that given by the mean square error (MSE)*", and – consequently – predictions free of systematic errors within the materialization of the interpolation process.

Table IV. Comparison of average residues and standard deviation for pH and TSS. (own source)

Weeks	Average pH residues	Average TSS residues	pH standard deviation	TSS standard deviation
1	-0.03	-1.50	0.18	1.08
2	-0.07	12.86	0.09	23.37
3	-0.01	11.77	0.06	0.37
4	-0.02	4.72	0.14	8.98
5	-0.01	-0.73	0.03	1.25
6	0.01	2.19	0.05	0.79
7	0.13	-9.49	0.06	0.03
8	0.14	-17.27	0.00	17.52

Table IV shows the average residues for pH and TSS, where the pH variable indicates for weeks 3, 5, and 6 mean residues close to zero, inferring an estimation free from bias (lack of systematic errors), as expected from sets that fit the normal distribution. For residues from the TSS variable, it may be stated that data from each of the eight weeks had mean residues not close to zero, which indicates the presence of bias in the estimations, however, those that fit the normal distribution had the lowest bias, as was the case of SST1, SST4, SST5, SST6, and SST7.

Validation of the quality of ordinary Kriging interpolation was conducted by using the statistical MSE or RMSE, managing to establish and quantify the most significant disparities found between applying the algorithms used to model surfaces and one used as reference - control points – [15], [18] and [19].

Figure 5.a and 5.b permitted easily visualizing how the training sets corresponding to week 3, 5, and 6 and 1, 4, 5, 6 and 7 for pH and TSS, respectively, had the lowest amount of error in the estimation models and ratified how – when fitting the normal distribution – these had the lowest MSE.

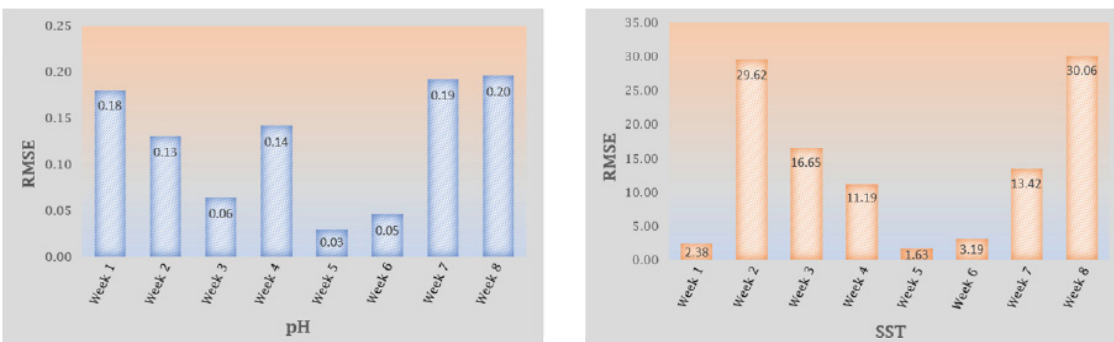

Figure 5. Graphic description of the MSE (RMSE) for pH and TSS. Figure 5.a RMSE for pH week 5 and Figure 5.b RMSE for TSS week 4. (own source)

Spatial autocorrelation

After the average and variance, it is the most important property of any geographic variable [13] and is explicitly associated with its spatial pattern.

To measure the type of correlation the same variable has on different spatial units, the study calculated Moran's index. In Figure 6.a, Moran's index for pH has a value of 0.1884, indicating that the disposition of the data to spatial grouping (positive AE), the standard deviation values (z-score), and probability (p value) support the hypothesis behind this statistical test, which guarantees with 99% certainty the presence of the spatial grouping in the pH data. This means there are sampling data that share similar information on contiguous locations that irradiates through many spatial geographic units. In Figure 6.b, Moran's index for the TSS variable TSS has a low value of 0.0875, which indicates a weak relationship between the sampling data. This points to no correlation, a trend supported by standard deviation values (z-score) and probability (p value). Hence, the values of neighboring spatial units are produced randomly.

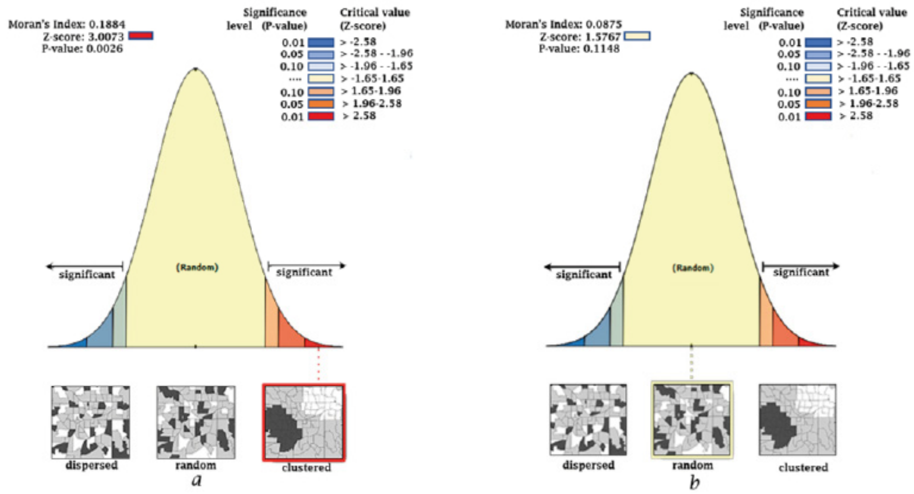

Figure 6. Moran's index source ArcGIS 10.7 for the pH and TSS variables. Figure 6.a Moran's index for pH week 5. Figure 6.b Moran's index for TSS week 4. (own source)

Given that Moran's index is used to identify the type of spatial configuration followed by the data, the Getis-Ord index is eventually a measure of the concentration of the type of spatial configuration present, allowing to identify if a spatial pattern (grouping or dispersion) occurs in low bajo or high degree.

Figure 7.a shows how for the pH variable the degree of spatial grouping of the data for week 5 is high; the p value < 0.05 shows with 95% CI how the null hypothesis is rejected and the alternate hypothesis is accepted – data have a high degree of spatial grouping, in other words, it indicates that besides having a high degree of similarity between pH data for week 5, their locations are quite close among each other, which is supported by Tobler's principle. Figure 7.b demonstrates that when no correlation is found between the data for TSS week 4, the degree of concentration is equally random. This clearly implies that the high and low concentrations given by the Getis-Ord index is only possible for cases with presence of spatial autocorrelation.

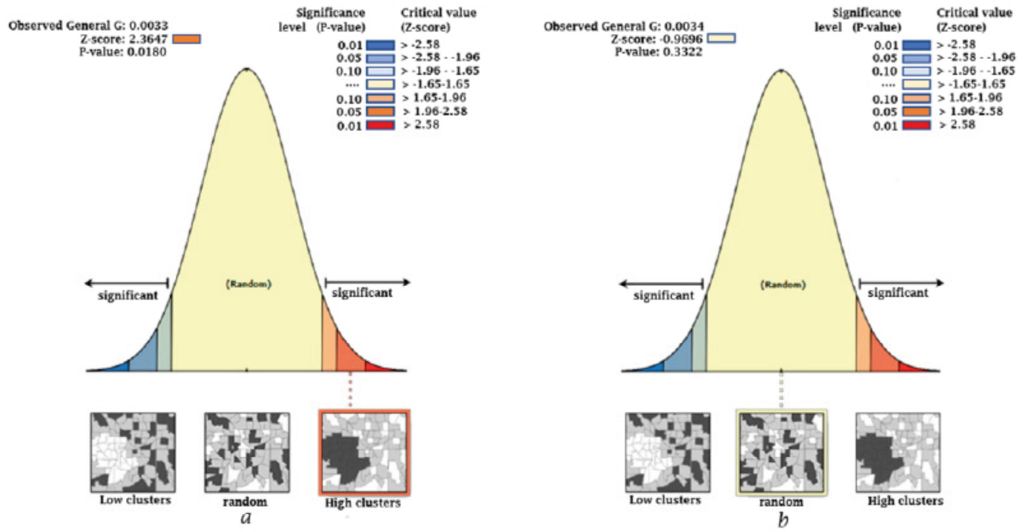

Figure 7. Getis-Ord index for pH and TSS variables. Figure 7.a Getis-Ord index for pH week 5. Figure 7.b Getis-Ord index for TSS week 4. (own source)

These models conducted, as shown in Figures 6 and 7. Of the eight weeks for pH, weeks 3, 5, and 6 had high degree of spatial grouping, this indicates that not only the positions of the stations where the measurements were taken are very close to each other, but additionally their values also are; that is, pH for those 20 stations maintains very similar values. This guarantees a much more effective estimation process, as shown in Table 5, as long as they fit the normal distribution. Likewise, permitting to monitor the variability of the contamination induced by diverse factors, given the high grouping of the pH values in these weeks, associating this behavior to no dumping of any pollutant, which would induce an abrupt change in pH in any of the stations, especially those between stations 10 and 15, sites where the highest deposit of residues takes place onto the tributary, possibly due to low industrial and mining production during said weeks. In addition, TSS had very variable behavior for all the data within each week, not managing to establish any positive autocorrelation for the 20 data within each week; rather, following a mostly random behavior. Mutability possibly attributed to the dumping of human, mining, and industrial waste and helped by progressive increase in heavy rains.

The importance of identifying the type of relationship the pH and TSS variables have with each other is closely linked to the propagation of errors in an interpolation model, where it is clearly expected that a prediction model as its uncertainty (error) be normally distributed to keep the estimation error in a location to radiate spatially toward its neighbors. If the sampling values of the study variable are highly autocorrelated (high clusters), it is likely for the values in the estimation models to also be and, hence, their errors will surely share said property [15], [9]. Now, achieving little bias in the predictions will guarantee a very small propagation of such errors to close spatial units.

Conclusions

The result described converges into an optimal ordinary Kriging behavior for week 5 of the pH variable and week 4 of the TSS variable, as well as for all the sets that perfectly fit the normal distribution within their respective sets. This is undoubtedly an important result to determine water quality parameters with high degree of reliability, given that the relevance implied by having precise pH and TSS measurements, in the development and sustainability of the environment, and avoiding acquisition and propagation of diseases related with contamination. Likewise, it is a useful tool for environmental control and sanitation agencies that establish methodologies that allow to mitigate and record the quality of the water resource periodically at all points of the tributary; thus, minimizing the high index of gastrointestinal diseases caused by its consumption, which have placed negatively the department of Quindío in first place.

Using the RMSE statistic, recognized universally as the parameter in charge of evaluating the quality of the DEM, provided a report on the overall quality of the study surface; however, an additional measure known as the mean error (ME) was used to achieve a more reliable and specific analysis, guaranteeing a much more precise interpretation of the behavior of errors of the estimation models. This is how the mean error identified anomalies associated with systematic errors (Bias) introduced in the data processing.
The exploratory statistical analysis permitted identifying the qualities associated with each study phenomenon to guarantee that discrete data, like pH and TSS can be reproduced through a continuous adjustment function. That is, those training sets that fulfilled the normality test ensured better quality in the estimations.

The spatial autocorrelation is a vitally important property and its study in geostatistical analyses is indispensable; its knowledge permits defining the spatial pattern (distribution) of the sampling points, as well as the tendency of their values, indicating if the values taken by a variable in a specific location depend or not on the values of that variable in other positions. Ideally, it is preferred for data not be correlated, however, the environmental information Will always follow the concept of regionalized variable (Tobler's principle: "in the geographic space, everything is related to everything else, but nearby things are more related than distant things") on which algorithms, like Kriging, focus. Besides, the presence of positive spatial autocorrelation is also highly useful to identify and monitor the variability of water quality, due to dumping of contaminants, given that it studies closely the pattern of similarity between the pH and TSS values within each week, respectively, indicating anomalies reflected by abrupt changes in their values, abducted to spills of polluting materials.

The sets with positive spatial autocorrelation (spatial grouping) and which adjusted to normality, guaranteed low error propagation, that is, that is, as there were no high over- and under-estimation of values (no presence of bias), there was no irradiation of considerable errors through nearby spatial units.

Kriging is an interpolator with many virtues with respect to other prediction methods; optimization of the search radius for the neighbors to be used in a specific location (based on the analysis of the experimental semivariogram), effective determination of the weights assigned to each neighbor, based not only on the distance between the points and the location estimate, but also on the (relationship) distance between observation points, inclusion of an error prediction model; lastly, guaranteeing a linear, unbiased estimation of minimum variance. However, the advantage of Kriging with respect to other methods, attributed to this last virtue, was only achieved when the data were free of bias, that is, when they fit the Gaussian curve or normal distribution.

Generally, the Kriging geostatistical method showed its outstanding capacity to adjust its function to the limited amount of sampling points and its irregular distribution, where the statistic used globally reflected the Kriging performance for pH and TSS, where the worst results point-by-point, as expected, were obtained in monitoring station number 20, given its remote location with respect to the other sampling points, complicating inclusion of neighbors for the interpolation process.

Acknowledgments

To the Universidad del Quindío and the Corporación Universitaria Empresarial Alexander von Humboldt.

References

[1] J. F. Ortiz-Giraldo, M. J. Mesa y O. E. Molina, “Regular time series for physical-chemical water parameters: Total suspended solids and hydrogen potential”, *Journal of Physics: Conference Series*, vol. 1448, no. 1, 2020. Doi: 10.1088/1742-6596/1448/1/012021

[2] J. M. García-Usuga, G. Olivar-Tost, M. J. Mesa-Mazo, y C. A. Acosta-Minoli, “Modeling the dynamics of total suspended solids in a mountain basin using network theory”, *River Research and Applications*, pp. 1-1, 2 febrero 2021. DOI: 10.1002/rra.3828

[3] W. R. Tobler, “A computer movie simulating urban growth in the Detroit region”, *University of Michigan-Clark University*, vol. 52, no. 2, pp. 95-102, 1970

[4] E. S. Anderson, J. A. Thompson, y R. E. Austin, “LIDAR density and linear interpolator effects on elevation estimates,” *International Journal of Remote Sensing*, vol. 26, no. 18, pp. 3889-3900, 2005. DOI: 10.1080/01431160500181671

[5] R. Giraldo, Estadística espacial. Bogotá: Universidad Nacional de Colombia, 2011

[6] X. Liu, “Airborne LiDAR for DEM generation: some critical issues”, *Progress in Physical Geography*, vol. 32, no. 1, pp. 31-49, 2008. Doi: 10.1177/0309133308089496

[7] M. Villatoro, C. Henríquez, y F. Sancho, "Comparación de los interpoladores IDW y KRIGING en la variación espacial de pH, Ca, CICE y P del suelo", *Agronomía Costarricense*, vol. 1, no. 32, pp. 95-105, 2008

[8] A. Sajid, R. Rudra, y G. Parkin, "Systematic Evaluation of Kriging and Inverse Distance Weighting Methods for Spatial Analysis of Soil Bulk Density", vol. 55, 1983, pp. 1-13, 2013. Doi: 10.7451/CBE.2013.55.1.1

[9] A. Hartemink y A. McBratney, GEOMORPHOMETRY: Concepts, Software, Applications (Development). Developments in Soil Science, 2009. Doi: 10.1007/978-3-319-28295-4

[10] J. Márquez Pérez, Modelos Digitales de Elevaciones, Métodos de Interpolación y Parámetros Morfométricos del Relieve. Sevilla-España: Departamento de Geografía Física y Análisis Geográfico Regional, 2004

[11] A. Felicísimo, "Modelos digitales del terreno: introducción y aplicaciones a las ciencias ambientales," Universidad de Oviedo, Oviedo, 1994

[12] J. S. Duque Martínez, "Comparación Estadística de Métodos Interpolación Determinísticos y Estocásticos para la generación de Modelos Digitales del Terreno a partir de datos LIDAR, en la parroquia de Tumbabiro, cantón San Miguel de Urcuquí, provincia de Imbabura", Universidad San Francisco de Quito, 2015

[13] J. P. Celemín, "Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación", *Revista Universitaria de Geografía,* vol. 18, no. 1, pp. 11-31, 2009

[14] J. Garzón, "Validación cruzada como método de investigación en la calidad de Modelos Digitales de Elevación," Universidad San Francisco de Quito, 2013

[15] P. F. Fisher y N. J. Tate, "Causes and consequences of error in digital elevation models," *Progress in Physical Geography,* vol. 30, no. 4, pp. 467-489, 2006. Doi: 10.1191/0309133306pp492ra

[16] A. Vega y J. F. Mas, "Error assessment of digital elevation models obtained by interpolation", *Investigaciones Geográficas,* vol. 69, no. 8701, pp. 53-67, 2009

[17] W. E. Salinas, C. U. Paredes, X. Martínez, y F. Guevara, "Evaluación de la exactitud posicional vertical de una nube de puntos topográficos LiDAR usando topografía convencional como referencia," *Investigaciones Geográficas*, vol. 85, pp. 5-17, 2013

[18] Q. Guo, W. Li, H. Yu, y O. Alvarez, "Effects of topographic variability and Lidar sampling density on several DEM interpolation methods", *Photogrammetric Engineering & Remote Sensing*, vol. 76, pp. 701-712, junio 2010. Doi: 10.14358/PERS.76.6.701

[19] X. Liu, Z. Zhang, J. Peterson, y S. Chandra, "The effect of LiDAR data density on DEM accuracy," *en International Congress on Modelling and Simulation (MODSIM07),* pp. 1363-1369, 2007