Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2012.749194

51 (2013) 2609–2616
March

Taylor & Francis
Taylor & Francis Group

Analysis of seasonal variations in the quality of water in a reservoir using GIS techniques

Concepción González-García[a], Sara G. Condado[b], Maria Jesús García-García[b,*], Daniel Werenitzky[c]

[a]*Deparment Economics and Forestry Management, Escuela Técnica Superior de Ingenieros de Montes, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain*
[b]*Department of Projects and Rural Planning, Escuela Técnica Superior de Ingenieros de Montes, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain*
Tel. +34 913367110; email: mariajesus.garcia.garcia@upm.es
[c]*Facultad de Agronomia, Universidad Nacional Santiago del Estero, Santiago del Estero, Argentina*

Received 12 January 2012; Accepted 9 September 2012

ABSTRACT

This paper analyzes the seasonal behavior of water quality in the reservoir of Rio Hondo (Sali-Dulce river basin, Argentina) using Geographic Information Systems (GIS) interpolation techniques. Due to its geographical location, the water of this lake is influenced by human activities from the upper basin, such as agricultural development, urban and industrial wastewater discharges which vary in intensity depending on the season. In this study the techniques of interpolation IDW (Inverse Distance Weight) and Kriging were applied to values of a water quality index named ICA2. The integrated tool Geostatistics Analyst in ArcGIS 9.2 software was used to check if it is possible to get a proper ICA2 spatial interpolation with the employed methods and which of them would be the best technique to estimate the quality of water. With the observed and interpolated values of ICA2 index, a series of maps were obtained that allow us to conclude that both methods are valid for estimating trends in water quality. However, slightly better results have been obtained with the Kriging method.

CE Database subject headings: Argentina; Spatial analysis; Water quality; Reservoirs Geographic Information Systems

Keywords: Interpolation; Water quality index; Seasonal analysis; Environmental impact

1. Introduction

Basin of the Rio—Dulce, Argentina, covers part of the territories of the provinces of Salta, Tucuman, Catamarca, Santiago del Estero and Cordoba (Fig. 1).

This basin is characterized in four sectors: upstream or input, middle basin or area of use, zone downstream, and seasonal floods and drainage area of the basin (lake of Mar Chiquita), of which analyses were performed only in the watershed upstream, since it is where we found the activities (anthropogenic pollution: urban, industrial and agricultural waste) that influence the variation in the quality of reservoir water [1].

The Salí river basin covers 72% of the territory of the Tucumán province. It consists of a dense network

*Corresponding author.

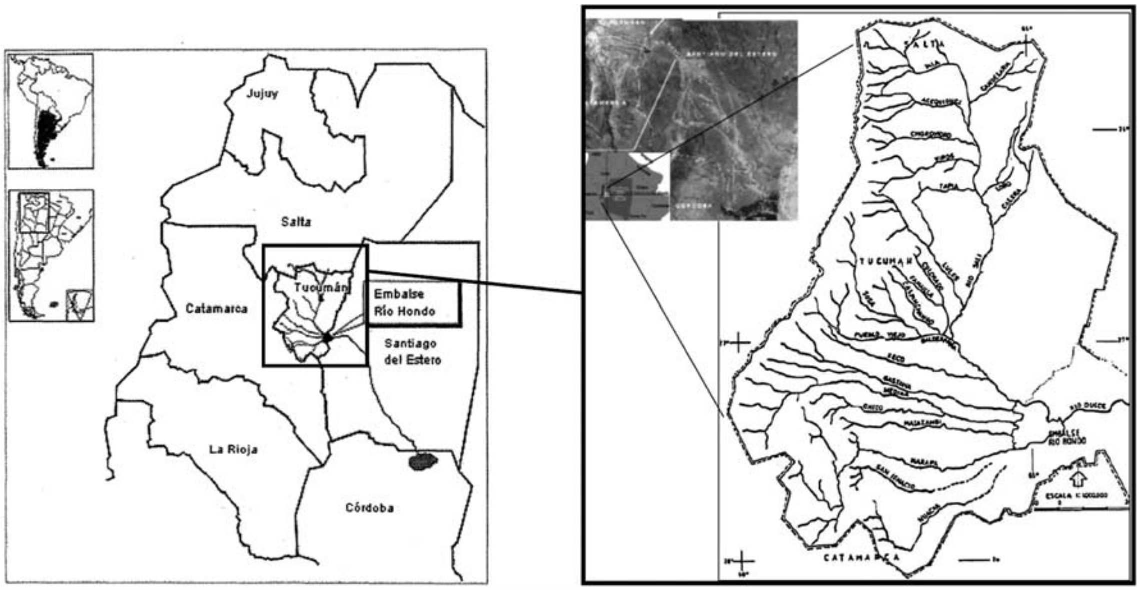

Fig. 1. Location map of the study area.

of drainage which collects the rainfall on the eastern slopes of the mountain chain in the western region of the province. The rivers from these mountains are the main tributaries of the Sali river. The economic activities are concentrated in this basin, and so the provincial water resources represent the main source of contribution to the provincial production. The hydrological network has several important uses: water supply, hydroelectric power production, agricultural–livestock, and industrial uses.

The climatic type of the reservoir area, according to the index Köpen, is semi-arid steppe—dry in winter, hot in summer [1].

In this region, the rains are usually in summer, contributing about 70% of the total annual rainfall [1] and winter is the dry season.

The flow of the rivers in the basin depends mainly on the rainfall. The maximum flow rates are recorded during the months November–April (summer season in the Southern Hemisphere so, late spring–summer–early in the autumn) and are lowest from May to October (late autumn–winter–early in the spring). The lower flow period coincides with the period of greatest industrial activity and therefore, with the higher levels of discharge of industrial effluents. The main industrial activities in the province are those related to sugar and citrus production, beginning approximately in May and ending in November.

At the begining of the raining season, the flow of the Salí river sweeps the discharges of organic origin to the inside of the reservoir.

This paper has been based on the data of physico-chemical and biological parameters of water samples at different points of the water mass in the Rio Hondo reservoir. In the studies by Werenitzky, [1,2] these data have been analyzed statistically and have been used to obtain water quality indexes. Comparative analysis of the performance of each index was done to identify which of them best described the condition of water quality. The water quality index ICA2 was selected as the one that best describes the state of water in each season for which data are available [3].

In the study reported here, ICA2 data in each season have been used to estimate values of the index in locations where data is not available, employing several interpolation techniques with the help of a GIS (Geographic Information System).

There are studies related to the effectiveness of the interpolators applied to the same data-set [4–8], studies related with the interpolation techniques applied to physical characteristics of land [9] or to digital elevation models of the terrain (MDET) [10]. Besides, in recent studies GIS is applied to the spatial distribution of water quality, both in rural areas [11] and in urban water use [12]. However, we believe that there is not much research about performance of different interpolation methods in surface water [13].

The interpolation techniques used in this study are IDW (Inverse Distance Weighting) and Kriging. They were chosen as they are the most commonly used in analysis of spatial variation [14]. From a theoretical point of view, kriging method is highly recommended for inclusion in GIS [15]; but some authors note that, in practice, its effectiveness is comparable to the other methods—that are more simple and with less computational requirements [7,9]—like the IDW interpolation method.

In this case, the comparison of IDW and kriging methods applied to ICA2 values has led to identification of the method that provided a seasonal sequence of maps which best depict the water quality distribution in Rio Hondo.

The interpolation methods have been applied with ArcGIS and Geostatistical Analyst extension (GA), which allow exploratory data analysis and the application of deterministic and geostatistical methods for obtaining interpolation surfaces. To verify the goodness of fit of each method and therefore to compare both, GA provides the prediction errors of each method [16]. To this aim, the prediction errors were also analyzed statistically to verify compliance of linear hypotheses.

The objectives were: (1) to determine the spatial distribution of ICA2 obtaining water quality surfaces by interpolation using two methods, IDW and Kriging that allow ArcGIS tool; (2) to compare the two methods to determine the method that best describes the water quality in Rio Hondo; and (3) finally, to obtain the sequence of more accurate maps of quality for each season of the year for which data are available.

2. Materials and methods

2.1. Data and description of study area

The initial data correspond to the physico-chemical parameters for monitoring water quality stipulated by the Secretaría de Recursos Naturales de la Nación (Ministry of Natural Resources of the Nation) to the concessionaire for the operation and management of the hydroelectric plant of Río Hondo in the province of Santiago del Estero. The data collection was done during the years 1995, 1996, 1997, 1998, 1999 and part of 2000, in months in which there are problems with water quality, i.e. winter, spring, and summer seasons. So, there was no sampling in autumn.

The situation of the sampling points in Rio Hondo reservoir was georeferenced using ArcGIS 9.2 software and then the values of water quality index ICA2 calculated previously were assigned. Werenitzky [1] analyzes several water quality indices and the best index to describe the variation of water quality in Rio Hondo was ICA2. So, it was selected to this study.

The dam or reservoir of Rio Hondo has an average depth of 5 m and an approximate surface of 33,000 hectares. The ratio volume/surface is 1,740 hm^3/33,000 ha, according to data registered by the Regulatory Body Safety of Dams (ORSEP) under the Ministry of Public Works, Argentina.

Sampling points were in accordance with general rules for sampling with multiple objectives of a reservoir. The literature [17,18] suggests a minimum of three sampling points: one at the top, one in the middle and a third near the dam, i.e. in the tail, body, and head of the reservoir. Also, the points of the mouths of the tributaries were taken into account [19]. In this case, one sample point in the tail (no. T16), four in the body (points no. T1, T8, T10, T17) and six in the head (point no. T2, T9 and points of mouth of tributaries T3, T4, T5, T7) have been considered. The sampling dates were decided based on the time of year during which most of the contributions of anthropogenic waste (industrial and agricultural) take place in each season of the year. The interpolation techniques allow us to represent the approximate state of quality of the water body at each time sampled on a continuous basis.

2.2. General procedure

First, an exploratory data analysis was done to check normality of ICA2 data by a normal QQ-plot. From the data-set at each season, ICA2 interpolation techniques were applied to obtain continuous surfaces by estimating the unobserved values (predictions) of the variable under study. Interpolation techniques employed are IDW and Kriging. An analysis of errors was done to compare both the methods. This analysis was to test the lineal hypothesis: zero mean, constant variance, and independence. The Mean Square error (MSE) of each method also served to verify the goodness of fit of each one. The lowest MSE is the best interpolation technique.

The selected interpolation method was used to obtain water quality maps of each season for which observed data were available (no data in autumn).

2.3. Spatial interpolation methods

Interpolation is the estimation of Z values of a surface at an unsampled point based on the known Z values of surrounding points [20]. In this case, the Z values are the ICA2 data.

2.3.1. Inverse of distance weighted (IDW) method

IDW is a deterministic local method and interpolates values giving more weight to the values of closest stations and less to those that are farthest away. To predict the value of a nonsampling site, the values of the measured points around the place that has to be predicted are used. Hence, it is considered that each weight is inversely proportional to the distance of the point being estimated [21].

The expression of the model is:

$$Z(S_0) = \sum_{i=1}^{N} \lambda * Z(S_\mathrm{i})$$

where $Z\ (S_0)$ is the estimate value for the site S_0; N is the number of sampling points around of unsampled point and whose value is going to estimated; λ is the weight assigned to each sampling point—these weights decrease with distance; and (S_i) is the observed value at site S_i.

The expression to calculate the weights is:

$$\lambda_\mathrm{i} = d_{i0}^{-p} / \sum_{i=1}^{N} d_{i0}^{-p}$$

with d_{i0} being the distance between the unsampled site S_0 (whose value is to be estimated) and each sample site, S_i. As the distance increases, the weight is reduced by an exponent p. That is, while increasing the distance between sampling points and estimated points, the weight λ_i of sample point used to predict will decrease exponentially.

Another important issue of the IDW technique is determination of the number of neighbors to take into account when calculating the predicted value. This will depend on the type of data and the surface that is sought. Therefore, the area calculated using the weighted average IDW will depend on the parameter p and the neighborhood search strategy.

The two parameters (λ_i and p) are chosen optimally according to a minimum root mean square error (RMSE) criterion [22].

Kriging method. Kriging is a geostatistical interpolation technique that utilizes the statistical properties of the measured points. This technique estimates values at nonsampled locations using weights that reflect the correlation between data at two sampled locations or between a sample location and the location to be estimated [23]. This method provides a measure of error of estimates, which is also an indicator of the goodness of fit and predictions.

This procedure provides the best linear unbiased estimator (BLUE) of the variable under study. It is "linear" since the estimated values are weighted linear combinations of the available data. It is "unbiased" because the mean of error is 0. It is "best" since it aims at minimizing the variance of the errors.

$$Z(u_\alpha)\,,\ \alpha = 1, 2, \ldots N$$

where u_α denotes the sampled points with information about studied variable Z and $Z^*(u)$is the estimation of unknown *Z(u)* from $Z(u_\alpha)$.

The basic approach of the estimation by Kriging is considering $Z(u)$ as a linear combination of observations (in this case, values of ICA2)

$$Z^*(u) \ = \ \sum_{\alpha=1}^{N} \lambda_\alpha(u)\, Z(u_\alpha)$$

This method uses variogram to express the spatial variation and it minimizes the error of predicted values which are estimated by spatial distribution of the predicted values. So, the weights are chosen on an approach which considers that this estimate is optimal. The optimal weights, λ_i, should produce unbiased estimates,

$$E[Z^*(u) - Z(u_\alpha)] = 0$$

and have errors with a minimum variance

$$\mathrm{var}[Z(u) - Z^*(u)] = \min$$

The difference between Kriging and other linear estimation methods is its aim of minimizing the error variance. Kriging is distinguished from IDW and other interpolation methods by taking into consideration the variance of estimated parameters.

2.4. Validation of the interpolation methods

The validation of the interpolation methods was performed using analysis of errors. The statistical analysis used for estimation errors of stochastic models such as Kriging, is applied, by extension, to the IDW model to facilitate comparison of both interpolation methods.

The chosen model can be validated by interpolating observed values. So if n observations $Y(x_i)$; $i = 1, \ldots, n$ are available, the validation process proceeds as follows:

For each j, j = 1, …, n discard point $(x_j;\ Y\ (x_j))$; estimate the $Y^*(x_j)$ by solving the IDW or the Kriging system having set $x_0 = x_j$ and using the remaining points x_i, $i \neq j$ for the interpolation; evaluate the estimation error $e_j = Y_j^* - Y_j$.

The model can be considered theoretically valid if the error distribution is approximately gaussian with zero mean and unit variance (N(0; 1)), i.e. satisfies the following:

Zero mean,

$$\frac{1}{n}\sum_{1}^{n} e_{\mathrm{i}} = 0$$

The estimation variance σ_i^2 is coherent with the error standard deviation:

$$\frac{1}{n}\sum_{1}^{n}\left(\frac{Y_{\mathrm{i}}^{*} - Y_{\mathrm{i}}}{\sigma_{\mathrm{i}}^{2}}\right)^{2} \cong 1$$

Besides, under linear hypothesis $\sigma_{\mathrm{i}}^2 = \sigma^2 = cte$.

Also, it is interesting to look at the behavior of the interpolation error at each point, using mean square error (MSE) of the vector e_j:

$$\mathrm{MSE} = \frac{1}{n}\sum_{1}^{n} e_{\mathrm{i}}^{2}$$

and its root (RMSE).

3. Results and discussion

The IDW and Kriging methods were applied to ICA2 values of each season. The exploratory data analysis of ICA2 at each season provides the QQ-plots. It is possible to assume normality of data in all cases. Kriging methods work best if the data are normally distributed.

The maps—obtained by applying the interpolation methods IDW (Fig. 2) and ordinary kriging (Fig. 3) to the distribution of water quality in each season—were interpreted according to climatic characteristics and human activities upstream, in each season.

As specified in the introduction, in this region the rainy season is during summer which contributes 70% of the total annual rainfall [1] and winter is usually dry.

The gradient of quality from the mouths of the rivers to the retaining wall of the dam is evident. The index values are increasing, for example, from Sali river to dam wall. In winter, the index shows minimum values at the mouth in the reservoir of Salí, Marapa, and Gastona rivers because they carry the waste from the sugar industry, which began the harvest in Tucumán at the end of autumn. Coinciding with the dry season (early autumn in May to October, beginning of spring), the pollution load of rivers increasesas its volume decreases. This results in an extreme pollution (the green in spring and summer) in the area of the mouths. In late spring and early summer, with the first rains as well as the cessation of discharges by the end of the harvest, the eutrophication process is stopped by a dilution effect on the pollutants and water clarity. This process is spatially better represented in Fig. 2 than in Fig. 1.

3.1. Results of analysis of errors

Table 1 shows the statistical summary of the errors obtained with each method. The values provided by the GIS (means and standard deviations) have been arrived at by applying a statistical software (Statgraphics Plus 5.1).

The statistical summary includes measures of central tendency, variability, and form. Since the measured data are normal, we can see that the values of skewness and kurtosis of residuals are within the range between –2 and + 2, so there is no significant deviation from normality in any of the three cases. The same conclusion is obtained by the chi-square test of goodness of fit. It cannot be rejected by the hypothesis of normality at 95% confidence (p values greater than 0.05).

Independence of the errors is checked using the Box-Pierce test. It is based on the sum of the squares of the first 10 autocorrelation coefficients. The p-value for this test is greater than 0.10. So for a 90% or higher confidence level, it cannot reject the null hypothesis that the data are random.

The hypothesis that the average error in each case is zero, is verified with a t-test which does not reject the null hypothesis (p-value > 0.05) in any of the six cases.

Homoscedasticity requires that the variance of the errors is constant for all values. This condition may be analyzed using graphs of the distribution of errors about the zero mean. If there is a uniform dispersion of 18 data points about the zero line, it would be possible. To avoid any doubt that the observation of such graphics sometimes produce, they have been divided into two subsamples of 18 values of 9 values

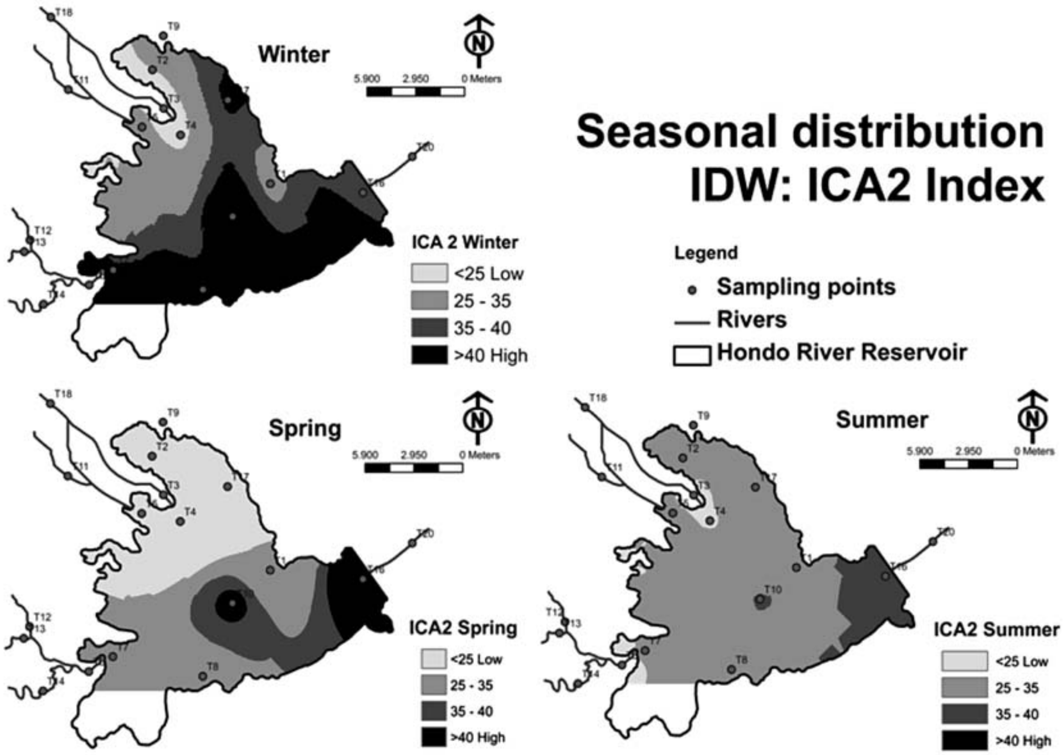

Fig. 2. Maps of distribution of estimated values of ICA2 with IDW interpolation method.

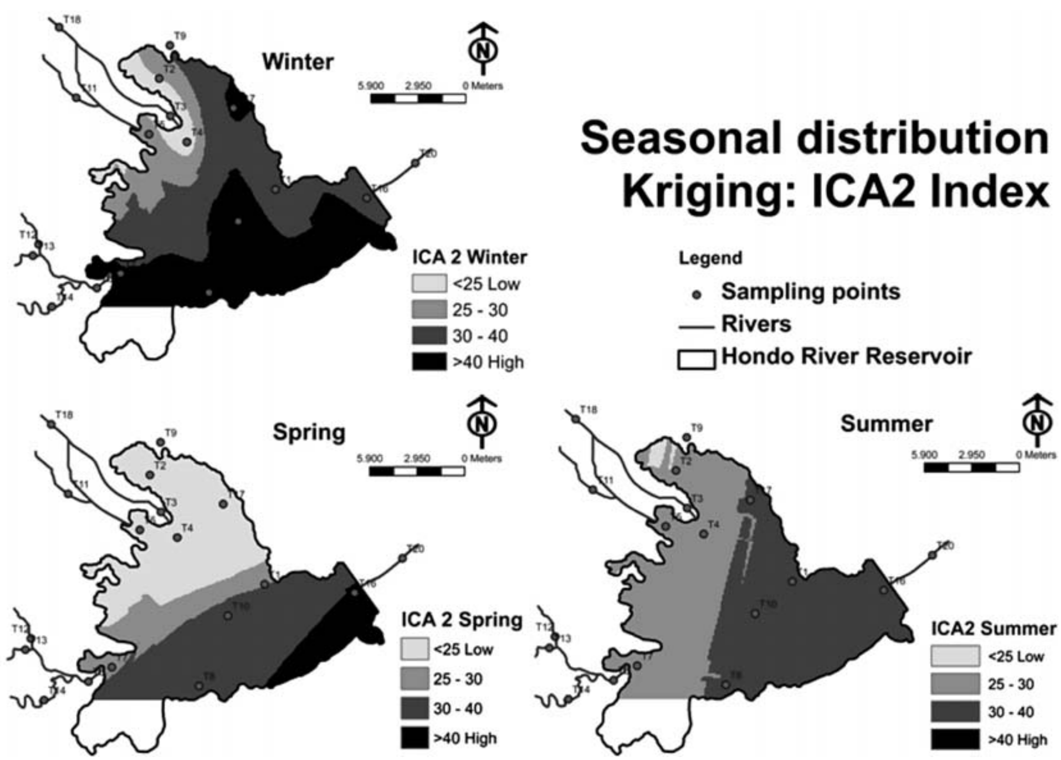

Fig. 3. Maps of distribution of estimated values of ICA2 with kriging interpolation method.

Table 1
Statistical summary of errors from IDW and Kriging method

Statistics	IDW error winter	IDW error spring	IDW error summer	Kriging error winter	Kriging error spring	Kriging error summer
Frequency	18	18	18	18	18	18
Mean	−0.098	−0.657	−0.410	−0.224	−0.194	−0.408
Standard deviation	8.674	9.499	7.490	7.262	8.470	7.098
Minimum	−17.82	−16.06	−11.01	−15.63	−12.44	−9.79
Maximum	13.44	16.18	15.4	14	18.04	16.52
Skewness	−1.012	0.256	1.020	−0.785	0.764	1.592
Kurtosis	−0.16	−0.74	−0.27	0.51	−0.27	0.55

Table 2
MSE for each method

Interpolation method/ parameters	Prediction error: IDW		Prediction error: Kriging	
	Mean	MSE	Mean	MSE
Winter	−0.84	8.53	−0.23	7.06
Spring	−0.66	9.26	−0.19	8.23
Summer	−0.41	7.29	−0.41	6.91

each, and each station is ranked by matching the variances of each subsample in it, using a test of the F. In all cases, the p-value>0.05 and so the null hypothesis of equal variances of the two sets of values in each sample of errors for each station cannot be rejected.

In Table 2, the average of square error is recorded for each station and methodology. The MSE is a measure of error that is most used to assess the adequacy of the models in the GIS [24].

By analyzing the estimation error of each method, their average error was found to be next to zero, in both the methods. With respect to MSE, the method of higher goodness of fit to the phenomenon studied is the one that has a lower MSE. Therefore, although the difference of the average square values does not discriminate specifically between Kriging and IDW in the three seasons, it is the Kriging method which obtains lower values i.e. a better fit than the IDW model.

In the estimation of water quality, there are a few known studies that compare different interpolation techniques. The de la Mora's study [25] compared the performance of three interpolation methods using a Water Quality Index (ICA) in a water body whose interest was the water supply to a metropolitan area. This study was one of the first attempts to assess the spatial variation of an ICA by ordinary kriging (KO) as a stochastic interpolation technique in addition to the comparison with other deterministic interpolation methods.

In this paper, after analyzing the goodness of fit of both the interpolation methods, the Kriging method has been proved to be the most suitable to describe the seasonal trend in water quality through maps of the Hondo River Reservoir.

4. Conclusions

By comparing the IDW deterministic method and kriging geostatistical method, it can be observed that the values were not very different in terms of prediction error. But we found that the Kriging method presents a best fit and a better depiction of the seasonal variation of water quality in the Río Hondo reservoir. This conclusion is similar to the results of other relevant studies.

Graphic analysis of the sequence of water quality mapsof Río Hondo reservoir shows a low water quality in all its seasons—with higher quality in winter and lower quality in summer. On the other hand, the quality of the reservoir decreases from northeast to southwest, due to the discharge upstream, and this phenomenon is common in every season. The temporal evolution of Río Hondo reservoir quality is accordance with the industrial processes that occur upstream.

In the Hondo River reservoir, the pollution problem posed by the influence of human activity requires analyses of spatio-temporal variation and this study is a contribution to the representation of the problem that can help policy-makers and managers of the area.

References

[1] D. Werenitzky, Evaluación de la contaminación en la cuenca alta del río Salí Dulce, Estudio estadístico de los contaminantes del embalse del río Hondo (Provincias de Tucumán y Santiago del Estero). Tesis doctoral, ETSI de Montes, Universidad Politécnica de Madrid, (2003) https://serviciosgate.upm.es/tesis/tesis/3767 (último acceso 5 agosto 2010).

[2] D. Werenitzky, C. González García, E. Ayuga Téllez, D. Hermosilla Redondo, Análisis Estadístico de la Variación Temporal y Espacial de la Calidad del Agua en la Cuenca del Río Sali-Dulce (Argentina), En CD del III Congreso de Ingeniería Civil, Territorio y Medio Ambiente: 13 pp., Libro de resúmenes, ISBN: 84-380-0335-4, Zaragoza, España, 2006, p. 46.

[3] C. González García, D. Werenitzky, E. Ayuga Téllez, M.J. García García, A.P. Espluga González de la Peña, Evaluación de la contaminación estacional de las aguas del embalse de Río Hondo (Argentina) mediante la aplicación de índices de calidad, Actas del IV Congreso Nacional de Evaluación de Impacto Ambiental (IV CONEIA). Eds. Asociación Española de Evaluación de Impacto Ambiental, ISBN: 978-84-96437-73-9, Madrid, 2007.

[4] F.A.N. Declercq, Interpolation methods for scattered sample data: accuracy, spatial patterns, processing time, Cartography and Geographic Information Systems 23(3) (1996) 128–144.

[5] R. Franke, Scattered data interpolation: Tests of some methods, Compu. Math. 38(157) (1982) 181–200.

[6] D. Weber, E. Englund, Evaluation and comparison of spatial interpolators, Math. Geol. 24(4) (1992) 381–391.

[7] D. Weber, E. Englund, Evaluation and comparison of spatial interpolators II, Math. Geol. 26(5) (1994) 589–603.

[8] X. Yang, T. Hodler, Visual and statistical comparisons of surface modeling techniques for point-based environmental data, CaGIS 27(2) (2000) 165–175.

[9] G.M. Laslett, A.B. McBratney, P.J. Pahl, M.F. Hutchinson, Comparison of several spatial prediction methods for soil pH, J. Soil Sci. 38 (1987) 325–341.

[10] M.A. Aguilar, F.J. Aguilar, F. Agüera, F. Carvajal, F. Agüera, Evaluación de Diferentes Técnicas de Interpolación Espacial para la Generación de Modelos Digitales del Terreno agrícola, Mapping Interactivo, ISSN: 1.131–9.100, 2001.

[11] M. Elbana et al., Preliminary planning for reclaimed water reuse for agricultural irrigation in the province of Girona, Catalonia Spain, Desalin. Water Treat. 22(1–3) (2010) 47–55.

[12] Lee Ji-Heon, Ha Sung-Ryong, Bae Myong-Soon, Calculation of diffuse pollution loads using geographic information, Desalin. Water Treat. 19(1–3) (2010) 184–190.

[13] R.R. Murphy, F.C. Curriero, W.P. Ball, Comparison of spatial interpolation methods for water quality evaluation in the chesapeake bay, ASCE J. Environ. Eng. 136(2) (2010) 160–171.

[14] M. Villatoro, C. Henríquez, F. Sancho, Comparación de los interpoladores IDW y Kriging en la variación espacial de pH, Ca, CICE y P del suelo, Agronomía Costarricense: revista de ciencias agrícolas 32(1) (2008) 95–105.

[15] M.A. Oliver, R. Webster, Kriging: A method of interpolation for geographical information systems, Int. J. GIS 4(3) (1990) 313–332.

[16] A. Moreno, coord, Sistemas y Análisis de la Información Geográfica, Manual de autoaprendizaje con ArcGIS, 2ª ed. Alfaomega Grupo Editor, México, 2008.

[17] W.W. Walker, Simplified Procedures for Eutrophication Assesment and Prediction: User Manual, Headquarters, U.S. Army Corps of Engineers, 2003.

[18] R.G. Wetzel, Limnology, 2nd ed., Saunders College, Fort Worth, TX, 1983.

[19] Ryding, S. y Rast, W. 1992. "El control de la eutrofización en lagos y pantanos". Ediciones Pirámide, Madrid, España, 375.

[20] B. Nas, A. Berktay, Groundwater quality mapping in urban groundwater using GIS, Environ. Monitoring and Assessment 160(1–4) (2010) 215–227.

[21] A. Moreno, Sistemas y análisis de la información geográfica, Manual de autoaprendizaje con ArcGIS. Ra-Ma, 2005, 782–790.

[22] D. Rojas-Avellaneda, Spatial interpolation techniques for estimating levels of pollutant concentrations in the atmosphere, Revista Mexicana de Física 53(6) (2007) 447–454. http://www.scielo.org.mx/pdf/rmf/v53n6/v53n6a4.pdf.

[23] H. Wackernagel, Multivariate geostatistics, Third edition, Springer-Verlag, Berlin, NY, 2003.

[24] M. Morad, A.I. Chalmers, P.R. O'Regan, The role of root-mean-square error in the geo-transformation of images in GIS, Int. J. of GIS 10 (1996) 347–353.

[25] O. De la Mora, J.G. Fores-Garnica, C. Ruiz, V. García, Modelaje Estocástico de la Variabilidad Espacial de la Calidad de agua en un Ecosistema Lacustre, Rev. Int. Contam. Ambient 20(3) (2004) 99–108.