ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Environmental Sciences 7 (2011) 395–400
Procedia Environmental Sciences
1st Conference on Spatial Statistics 2011 – Mapping Global Change

Comparative study of models for estimating heights well not measured by the file of phase III of the drainage system in the Arzobispo subwatershed Bogota city

Natalia Niño[a], Cesar Basurto[b], Fabio Jaramillo[c], Fernando Santa[d]

[a,b,c]*Cadastral Engineering Student Distrital University Francisco José de Caldas, Bogotá, Colombia*
[d]*Cadastral Engineering Professor Distrital University Francisco José de Caldas, Bogotá, Colombia*

Abstract

We compared different models including spatial dependence through geostatistical techniques such as kriging in estimating heights well unmeasured sewerage system of the city of Bogotá. It also incorporates information on the structural characteristics of the network and physical environment. The tests referred to were carried out by using the software R version 2.11.1. The results showed a better fit of the models with spatial effects compared to currently use by the sewerage of the city, which has allowed a model of heights consistent with the logical flow of the hydraulic model.

Selection and peer-review under responsibility of Spatial Statistics 2011

Keywords: interpolation methods, heights, network, sewerage, predictor, cross validation

1. Introduction

The sewerage system describes an operating configuration characterized by its elements' heights, since the flow moves through the system because of gravity. Due to the importance of a right performance in the sewerage system in order to ensure life's quality of every citizen in cities like Bogotá, it is essential the implementation of techniques which provide reliable information, and to make it possible it is completely necessary solve the problems generated by the lack of information in certain network elements such as wells. In the specific case of wells, all expansive development, reconstruction or

a *E-mail address: natalianino@universitarios.com*
b*E-mail address: cesarbasurto7@gmail.com*
c *E-mail address: fajp22@gmail.com*
d*E-mail address: fernando.santa@gmail.com*

Selection and peer-review under responsibility of Spatial Statistics 2011
doi:10.1016/j.proenv.2011.07.068

beautification works advanced by infrastructure labors, prevent wells inspection, locating them “hidden”, either under parkland or under asphalt, making really difficult to access them to make the respective measurements.

Since the system depends on the model of heights or levels, it can be obtained by several interpolation techniques. The spatial interpolation process consists on estimate the possible values that the heights in a set of points located or georeferenced can take, from the variable’s values found in a sample of known points. All interpolation methods are based on the fact that the closer the points are located on the surface the higher spatial correlation they have.

This study aims to analyze the spatial correlation, interpolation and simulation of the sewerage network, based on the fact that the heights values of nearby points in space tend to be similar, which provides us information to study the behavior of the area’s attribute and compare the results for different spatial interpolation methods.

2. Research Methodology

The basis information of this study corresponds to the altimetric dates collected on terrain by the drainage system phase III of the land registry besides dates relating to the structural characteristics of the network setting and of the system inspection wells, information of the Bogota Aqueduct and Drains Company.

Deterministic and stochastic interpolation methods were implemented for the non-measured levels estimation, on the statistical software R.

Through the obtained results comparison with the different interpolators, it is determined, which the most effective and exact model is, through statistic techniques like the crusade ratification, analyzing the root mean square error of each model and the resume statistics in order to establish how good every interpolator estimates the depth prediction of the inspection wells.

3. Literature review

3.1. Spatial interpolation methods

The spatial prediction is based on the dependence that exist between the objects, fact that obeys the geography’s first principle which says that nearby objects tend to be more similar than objects more distant between them, because if the data have no spatial dependence the geographic analysis is meaningless. [1]

From the spatial dependence between objects there are two prediction models groups: deterministic and stochastic. Deterministic models are designed under the assumption that the outcome of an experiment is determined by the conditions under which it performs. This kind of models is divided into global ones, which use all the data to do the prediction, and the local ones which use a subset of the entire sample for estimation.

On the other hand, stochastic models are those where no one knows the expected outcome is not known but the probability is known and, as a consequence, there is uncertainty. In addition, prediction of these models provide an estimation error.

3.2. Deterministic interpolators

Inverse Distance Weighted (IDW): It's an exact interpolation method which estimates the weighted average based on a weighting coefficient corresponding to the inverse of the distance between each point and the point to estimate raised to an exponent. The IDW assigns more weight to points closer to the position to predict that those who are further away.

Triangulate Irregular Network (TIN): Is a form of vectors based on digital geographical data, constructed by triangulation of a set of vertices (points connected to form a network of triangles). The resulting triangulation satisfies the Delaunay triangle approach, which ensures that no vertex is inside any of the circles circumscribed to the triangles of the network.

3.3. Stochastic interpolators

Kriging: Is an optimal linear predictor, which means, is unbiased and minimum variance. This statistical method in order to estimate uses one spatial correlation model for data collection, setting the weights for each point used in prediction. Kriging is associated to the realization of a stochastic process and is based on the assumption that the parameter which is interpolated can be treated like a regionalized variable that varies continuously from one place to the following points according to the spatial correlation grade, but they are statistically independents. The objective of all Kriging methods is to find the weight,λ, optimal values n, and observations $z(s_i)$, in order to predict the unknown value of the position x_0. The Kriging predictor $z(s_0)$ is expressed as follows: [2]

$$Z(s_0)=\Sigma_{i=1}^{n}\lambda_i\, z(s_i) \qquad (1)$$

When the mean is known and constant for the entire region of interest and the predictor is unbiased, it means, the expected value of the prediction error is zero; weights are obtained of minimizing the error variance. [2]

Cokriging: This spatial prediction method consists on make spatial prediction of one variable based on its information and based on some auxiliary variables information which are spatially correlated with the main variable. The Cokriging predictor expression is:

$$Z_{v1}(s_0)= \Sigma_{i=1}^{n1}\lambda_i\, Z_{v1}(s_i)+ \Sigma_{j=1}^{n2}\lambda_j\, Z_{v2}(s_j)+ \ldots +\Sigma_{k=1}^{n2}\lambda_k\, Z_{vk}(s_k) \qquad (2)$$

The left side of equality in the above equation represents the prediction of the variable of interest in the position s_0 unsampled. $Z_{v1}(s_i)$ with $i=1,2,\ldots,n_1$, represents the primary variable. In the same way, $Z_{v1}(s_j)$ with $j=1,2,\ldots,n_k$, represents the k auxiliary variables. λ_i, $\lambda_j,\ldots,\lambda_k$ with $i=1,2,\ldots,n_1$, $j=1,2,\ldots,n_2,\ldots,$ $k=1,2,\ldots,n_k$ respectively, represent the observations weights of primary and auxiliary variables and are estimated based on the Coregionalization Linear Model (CLM) adjusted to the simple and cross semivariograms. Weightsλ_i, $\lambda_j,\ldots,\lambda_k$ are estimated similarly to the process described for the ordinary kriging method, ie, these are those who minimize the prediction error variance subject to the restriction that the predictor is unbiased. [3]

This method requires that both the co-variable and the target variable count with a spatial structure which can be modeled; in addition, it also requires that both of them have to be jointly second order stationary, which means, that the mean is constant (mean exists and is no position dependent) and the covariance only depends on the vector distance between the points and also a spatial covariance dependent. [4]

3.4. Cross validation: Cross validation removes one data point from a sample and uses the remainder of the population to interpolate a surface. The measured value is compared to the predicted value. This procedure is done for the entire population.

4. Analysis and results

4.1. Inverse Distance Weighted (IDW): using the *gstat* library of R, the prediction method IDW was established. This method assigns a weighted average of the near points values. The weighting coefficient is the inverse of the distance between the points, elevating to an index, in this specific case, the index is 2. The result is situated between the values that intervened in the process, because this interpolator calculates a weighted average

In the prediction map shows a fairly uniform with depths tend to depth values between 2.4 and 1.8 meters.

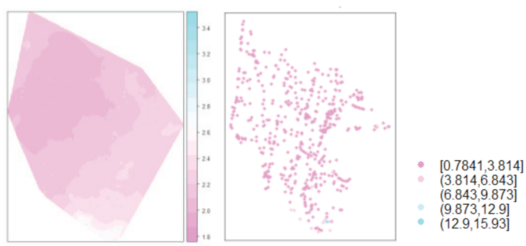

Fig. 1. (a) map depth prediction by IDW; (b) Depth estimation map for wells not measured by IDW

4.2. Triangulate Irregular Network (TIN): The *tripack* library was used in order to generate the triangulate irregular network. Afterwards the Delaunay triangulation was intersected, that maximizes the triangles interior angles with the predictable points, in order to estimate an average of the vertex depth values. The rounding mistakes are minimal, thanks to the Delaunay triangulation.. The prediction of the depth of the wells is mostly towards measures below 4 meters.

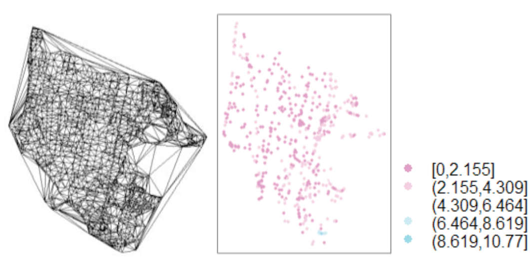

Fig. 2. (a) Delaunay Triangulation; (b) Depth estimation map for wells not measured by TIN

4.3. Kriging: This interpolation method was implemented thanks to the gstat library, it was necessary to apply the box cox transformation in order to guarantee a symmetric depth. The transformed depth doesn't represent an inclination in the space. The semi variance model that measures the space correlation level, was adjusted by minimal squared or weighted, through three different methods, to a gaussian covariance model and the obtained indicators are showed for each case, in the table 1 below.

Table 1. Adjusted parameters semivariance model

Variofit by	Nugget	Sill	Range
OLS (Ordinary Least Squares) - equal	0.1153	0.1227	3612.7035
WLS (Weighted Least Squares) - npairs	0.1122	0.0560	2001.0494
Cressie	0.1170	2.6608	19453.7650

The Krigging adjusted by ordinary least squares showed less variability in the distribution of prediction which has less error in the estimation of the deep wells for unmeasured.

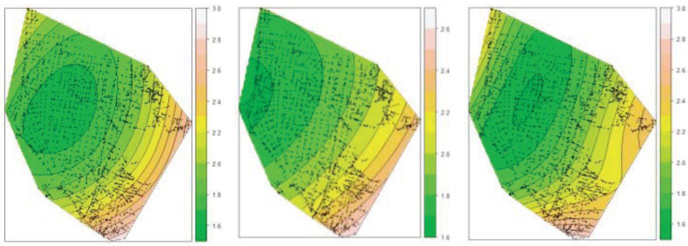

Fig. 3. (a) Kriggin prediction map, semivariogram adjusted by equals; (b) Kriggin prediction map, semivariogram adjusted by Cressie. (c) Kriggin prediction map, semivariogram adjusted by npairs

4.4. Cokriging: At the beginning, three variables or maybe co-variables were taken into account in this study case: soil type, topography and underground water level, which were co-located at the well depth. However, these variables explain very little the goal variable because the spatial correlation with the co-variables was very low, achieving to explain just through the underground water level value, no more than the 20%, that's because that was the unique selected variable to be used for the co-kriging estimation, obtaining the next crossed and direct semi-variograms:

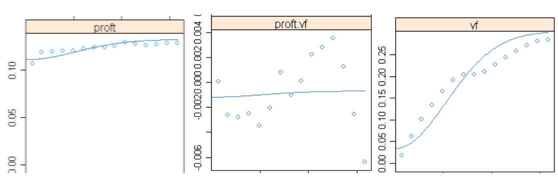

Fig. 5. (a) Semivariogram depth of wells; (b)Cross validation; (c) Semivariogram of the ground water level

5. Model Comparison

When comparing the mean square error of interpolation models implemented in this study, we see that the stochastic interpolators spatial prediction models are more efficient and accurate than deterministic estimates, since the mean square error of the residuals is considerably lower as shown in Table 2.

Table 2. Root mean square error (RMSE).

Spatial interpolation methods	RMSE
IDW	1.03689048
TIN	1.06246108
Krigging by OLS (Ordinary Least Squares) - equal	0.04404195
Krigging by WLS (Weighted Least Squares) - npairs	0.05930033
Krigging by Cressie	0.05773132
Cookrigging	0.34059910

The interpolation model that best explains the behavior of the hidden depths of the wells is the Krigging, since it has better goodness of fit by least squares and has less variability in the prediction error.

6. Conclusions

• The deterministic methods provide a good approximation to the spatial estimation of a phenomenon in space but these models are clearly susceptible to a strong variation in the measured data over relatively short distances.

• The goodness of fit of a spatial model proposed which seeks to explain a spatial phenomenon, is framed in different statistical tools of interest, such as the areas of prediction and cross validation, but a strong parameter held for the election of best prediction model is one that has a minimum value of least square error in making the estimate.

• For the proper development of geostatistical tools, such as cokrigging, it is important from the scoop and to ensure that there is a high spatial correlation between the secondary variables will be used as a function of improving the estimation of the primary variable.

References

[1] Ashish Sen, Muni Srivastava, *Regresion Analysis. Theory, Methods, and Applications.* Springer – Verlag. USA, New York, 1990.

[2] Christopher D. Lloyd, *Local Models for Spatial Analysis*. Queen's University Belfast. Taylor & Francis Group.USA, New York, 2007. p. 159

[3] Giraldo Henao Ramon, *Introduction to Geoestatistics*. Universidad Nacional de Colombia. Colombia, Bogotá. p. 63

[4] Diaz Viera Martin A., *Applied Geostatistics*. Institute of Geophysics, UNAM. Institute of Geophysics and Astronomy, CITMA, Cuba. 2002. p. 10