Proceeding of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences
Florianopolis-SC, Brazil, July 10-13, 2012.

Using Spatial Uncertainties to Create Probability Maps for Continuous Attributes

Carlos Alberto Felgueiras[1], Eduardo Celso Gerbi Camargo[2], Jussara de Oliveira Ortiz[3] and Sérgio Rosim[4]

[1,2,3] DPI-INPE, Av. Dos Astronautas, 1758, Jardim da Granja, São José dos Campos, SP
carlos@dpi.inpe.br [1], eduardo@dpi.inpe.br [2], jussara@dpi.inpe.br [3], sergio@dpi.inpe.br [4]

Abstract

*This work presents a methodology to create probability maps for spatial continuous attributes based on indicator geostatistical approaches. The indicator kriging and the indication simulation approaches can be used to infer approximations of conditional **c**umulative **d**istribution **f**unctions (**cdf**) for continuous attributes at different spatial locations of interest. The cdfs are conditioned to a set of spatial points containing continuous attribute values and sampled in a geographic region of interest. The conditional cdfs are then used to infer probability maps of exceeding, or being smaller than, a given threshold, or a predefined attribute, value. In this work it was used an elevation data set sampled in Florianópolis Island, the capital of Brazilian state Santa Catarina, as a case study to illustrate the methodology to create such probability maps.*

Keywords: uncertainty modeling, indicator geostatistics, kriging, simulation, probability maps.

1. Introduction

The modeling of continuous variables has been a necessary task for studies related to measures and estimates that can assist in the understanding of many phenomena that occur in nature, such as mapping studies of soil properties (Burgess and Webster, 1980), studies involving atmospheric (Lajaunie, 1984), hydrologic (Kitanidis and Vomvoris, 1983), geological (Deutsch, 2002) and many other variables. In these cases, it is important to employ appropriate modeling procedures in order to capture the behavior of the investigated variables within the region of study. To accomplish this task the literature points to several methods and modeling procedures. The simplest are the deterministic methods which have limitations because they do not consider reliable intrinsic spatial correlation structures and anisotropy behaviors of the data and uncertainties of the estimates. To overcome these weaknesses more complex models with probabilistic approaches can be used, for example, geostatistics. The geostatistics offers a set of methods and procedures which allow describing the spatial continuity of the variables involved in a more realistic modeling, obtaining more accurate estimates, for example, and allow more accurate investigations of the process, along with its uncertainties, which can later be displayed on maps (Isaaks and Srivastava, 1989).

For geostatistical approaches continuous attributes of spatial data are considered Random Variables (RVs) at each spatial location in a region of interest. A continuous RV has its uncertainty model represented by a local or spatial conditional cumulative distribution function (cdf). The uncertainty model, the cdf, of a continuous RV $Z(\mathbf{u})$, at a specific spatial location $\mathbf{u}$, conditioned to a (n) sample points can be denoted by $F(\mathbf{u};(z)|(n)) = Prob\{Z(u)<=z|(n)\}$ (Deutsch and Journel, 1998).

From the uncertainty model $F(\mathbf{u};(z)|(n))$ one can derive different optimal estimates for unsampled values $z(\mathbf{u})$ in addition to the conditional cdf mean, which is the least–squares error estimate (Deutsch and Journel, 1998). So, the conditional cdf of continuous attributes is mostly used to evaluate mean, median or any other quantil, values in order to create predictions maps. Also the cdf allows evaluating confidence intervals that are used for representing error or uncertainty maps (Felgueiras, 1999). The uncertainty maps can be propagated to the results of spatial models using four different methods including Taylor expansions and Monte Carlo simulations (Heuvelink, 1998). Goovaerts, 1997, presents different ways to account for conditional cdf models in the decision making process: exceeding a probability threshold, exceeding a physical threshold and minimization of the expected loss.

This work explores indicator geostatistical tools, kriging and simulation, that allows estimate an approximation of the cdf for a continuous spatial attribute at any location $\mathbf{u}$ inside a chosen spatial region. So, two different maps can be created: a map of *cdf probability values* that contains, for each spatial location $\mathbf{u}$, the probability values of being smaller or equal to the predefined attribute value and; a map of *1-cdf probability values* representing the probability values of being larger than, or exceeding, the predefined attribute value. Therefore, these two types of map are used to represent probabilities related to a predefined value and can be considered as a preprocessing for classifications based on being smaller than, or exceeding, a probability threshold value.

To illustrate the methodology applied to obtain such probability maps a case study was performed with a set of sample points representing the elevation data in Florianópolis Island, capital of the Brazilian state known as Santa Catarina. It shows how to create the probability maps, considering a minimum, or maximum, elevation value, to prevent problems related to floods, caused by heavy rains in residential areas, for example. Also these maps are then classified, considering different probability intervals values, in order to facilitate decision making activities based on such information.

2. Methodology

The entire methodology applied in this work was performed using the functions of the geostatistical module available in the Analysis menu of the SPRING (Câmara et all, 1996) Geographical Information System (GIS).

The study area is the Florianópolis Island, which is the capital of Santa Catarina State, in the Brazilian country. The bounding box coordinates of the Florianópolis region is: w 48^{o} 37’, w 48^{o} 20’, s 27^{o} 51’, s 27^{o} 23. The sample set is composed by 278 sample points of elevations distributed in the Florianopolis area as shown in the Figure 1 below.

Figure 1: The sample set of elevation distributed in the Florianópolis Island.

Indicator geostatistical approaches allow estimate approximations of conditional cumulative distribution functions, cdf, for variables, representing continuous attributes at different spatial locations **u**. The inferred cdfs are used to estimate local or global probabilities related to a predefined attribute value. The estimate probabilities in the locations of a regular rectangular grid allow the creation of probability maps of being smaller than, or exceeding, the predefined value. The methodology to obtain such probability maps follows the steps below:

1. Perform an exploratory analysis to determine statistical properties of the sample set;
2. Create a surface variogram to verify the presence of anisotropy behavior for the continuous variable;
3. Apply a variogram analysis to obtain experimental and theoretical, mathematical, variability models representing the variability of the variable in the spatial region;
4. Apply the indicator geostatistical procedures, kriging or simulation, to get cumulative distribution functions at each location of a rectangular grid in the area of interest;
5. Use the cdfs to calculate the probability values, at each location of the rectangular grid, of being smaller than a given threshold;
6. Classify the probability maps according several probability interval values.

3. Results and analysis

The statistical results of a exploratory data analysis for the sample set of 278 elevation points above are shown in Table 1.

The results of the Table 1 make possible the knowledge of the statistical properties of the data supporting hypotheses a priori, as for example, the probability distribution function of the used data.

Table 1: Statistics of the elevation sample set.

Mean	95.528	Minimum Value	2.000
Standard Deviation	107.433	First Quartil	13.000
Variation Coefficient	1.124	Median	33.000
Asymmetry Coefficient	1.171	Third Quartil	169.000
Kurtosis Coefficient	3.280	Maximum Value	437.000

Then a surface semivariogram analysis was performed in the elevation data set and an anisotropic continuity behavior was detected. As can be seen in Figure 2 the continuity is greater in the direction of 15° and smaller in the direction of 105°.

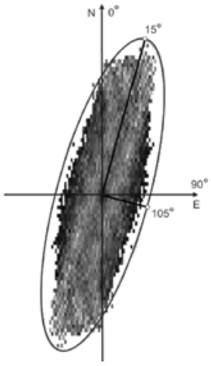

Figure 2: The surface semivariogram analysis of the elevation data set.

In order to model the conditional cdf, using indicator geostatistical approaches, it was used the first, the second (median) and the third quartiles as threshold cut values, 13, 33 and 169. Experimental indicator semivariograms were fitted with spherical models for the chosen cut values and for the two anisotropic directions, 15° and 105°. The semivariogram models are presented in the Figures 3, 4 and 5.

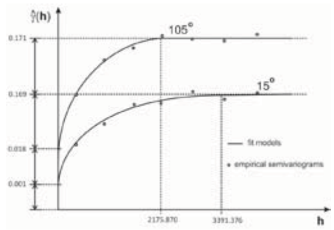

Figure 3: Semivariogram models fitted for the first quartile value.

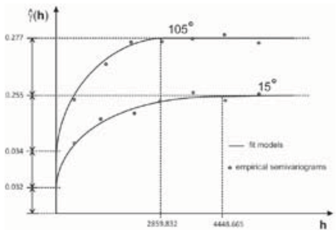

Figure 4: Semivariogram models fitted for the median value.

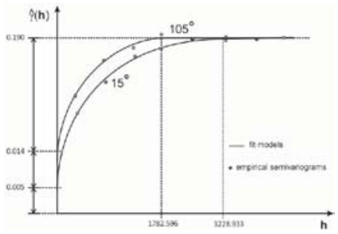

Figure 5: Semivariogram models fitted for the third quantil value.

Figure 6 shows three maps representing the probabilities of being smaller than the elevation values 50, 150 and 250, respectively. The dark areas represent low probabilities, closer to 0, and the white areas represent high probabilities, closer to 1 or 100% of probability.

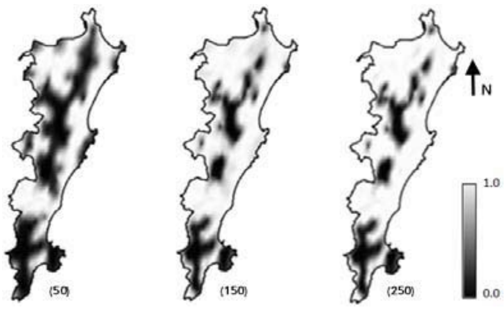

Figure 6: Probability Maps of being smaller than the elevation values 50, 150 and 250.

The figures above show that as the threshold values get larger the probability maps show larger white regions. These white regions can be considered as regions of flood risks and problems related to natural disasters caused by excessive rains in this region. The maps of Figure 7 are classified representations of the maps depicted in Figure 6. Maps ranked by probability intervals can facilitate the determination of subregions which should be considered for mitigation of adverse effects in the study area. The darkest areas of the maps below should be the first candidates to receive attention of the decision makers for flood risks applications

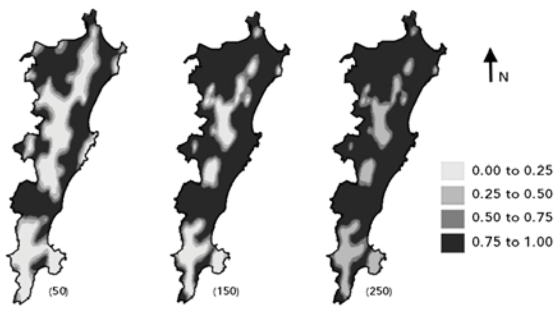

Figure 7: Classified Probability Maps of being smaller than the elevation values 50, 150 and 250.

3. Conclusions

This work presented a methodology to create unclassified, and also classified, probability maps of being smaller than a predefined threshold value. An elevation data set was used in order to illustrate the steps involved in the methodology. These probability maps are useful for decision making process in order to prevent and avoid disasters related to flood areas when heavy rains occurs in a short period of time at one small spatial region, for example. It is important to note that the probability maps were obtained not from estimates but directly from the uncertainties models obtained with geostatistical indicator procedures, kriging or simulation. This methodology can be applied to other environment attributes, as soil and air toxic elements for example, to support decisions making on whether a geographic region must be cleaned or not. This work shows, also, that a GIS containing a geostatistical module can be a powerful tool to manipulate spatial data in order to obtaining important results, as maps or reports, based on uncertainty analysis.

References

Burgess, T. M.; Webster, R. (1980), “Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging”. In: *Journal of Soil Science*, Vol 31:315-31.

Burrough , P. A.; Mcdonnell , R.A. (1998), “*Principles of geographical information system”,* Oxford University Press, New York, U.S.A, 344p.

Camara G., Souza, RCM, Freitas UM and Garrido J. (1996), “SPRING, Integrating Remote Sensing and GIS by object-oriented data modeling”. *In: Computer & Graphics*, Vol. 20:395-403.

Deutsch, C.V.; Journel, A.G. (1998), “*GSLIB: geostatistical software library and user’s guide*”, Oxford University Press, New York, U.S.A, 369p.

Deutsch, C.V. (2002), “*Geostatistical Reservoir Modeling*”, Oxford University Press, New York, USA, 376 p.

Felgueiras, C.A. (1999) “*Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: o paradigma geoestatístico por indicação*”, 165p. PhD Thesis, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brazil.

Goovaerts, P. (1997) “*Geostatistics for natural resources evaluation”*. Oxford University Press, New York, U.S.A, 483p.

Heuvelink, G. B. M. (1998) “*Error Propagation in Environmental Modeling with GIS”*, Taylor and Francis Inc, Bristol, 345p.

Isaaks, E. H; Srivastava, R.M.(1989), “An introduction to applied geostatistics”, New York: Oxford University Press, New York, 561p.

Kitanidis, P.K.; Vomvoris, E. G. (1983). “A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations”. *In: Water Resour. Res.*, 19(3), 677-690.

Lajaunie, G. (1984). "A geostatistical approach to air pollution modeling." In: “*Geostatistics for Natural Resources Characterization*”. G. Verly, M. David, A. G. Journell, and A. Marechal eds, Reidel, Dordrecht, Netherlands, 877–891.