Application of Geostatistical Conflation Techniques to Improve the Accuracy of Digital Elevation Models

Carlos A. Felgueiras, Jussara O. Ortiz[1], Eduardo C. G. Camargo[1]

[1]Divisão de Processamento de Imagens - Instituto Nacional de Pesquisas Espaciais (INPE) -
Caixa Postal 515 – 12201-970 – São José dos Campos – SP – Brazil

{carlos,jussara,eduardo}@dpi.inpe.br

Abstract. *This short paper describes and analyzes the results of a methodology that allows to conflate existing Digital Elevation Models with a sample set of elevation points in order to obtain more accurate results on modeling elevation information. The set of elevation points has higher vertical accuracy than the DEM and it is used the geostatistical procedure, known as kriging with an external drift, to perform the conflation. An initial case study is presented integrating a Shuttle Radar Topography Mission - SRTM - data and a sample set of elevation points obtained from a region of Campinas, city of São Paulo in Brazil. Others procedures of estimations and simulations will be considered in the future to explore the potential of the conflation techniques using geostatistics.*

1. Introduction

Digital Elevation Models - DEMs -, and their derivative products, are very important information used as input for spatial models performed in Geographical Information Systems - GIS - environment [Burrough 1987]. From a DEM it is possible to derive slope and aspect maps, drainage networks, contour lines, profile and volume calculations, etc. Nowadays it is possible to find DEM information for free, without financial costs, of almost any region of the earth surface. Unfortunately, the vertical (the heights) accuracy of these free DEMs are not appropriated for some spatial models. On the other hand elevation information of the earth surface can also be obtained in a set of spatial locations, 3D points, sampled in a geographical region of interest. These samples can be collected with very high vertical accuracy using Global Positioning System - GPS - equipments, for example.

Geostatistical tools has been used successfully to analyze and to model environmental attributes represented as a set of sample points of geophysical and geochemical indices, concentrations of soil elements, elevations, temperatures, etc. [Isaaks and Srivastava 1989, Goovaerts 1997]. Therefore, geostatistics can be applied to a sample set of elevation points, hereinafter referred to as "sample points", in order to create DEMs using estimations and simulations procedures. Also the geostatistical procedures allow performing conflations by integrating different sources of environmental attributes [Hengl et al 2008, Karkee et al 2008]. In the case of elevations there are kriging and simulation procedures that can be used to conflate existing DEMs with sample points of the same geographical region. The objective of this conflation is to get a more accurate final DEM compared with the original, or the input, DEM.

In this context, the objective of this short paper is to describe and analyze a methodology to conflate DEMs with a sample set of elevation points in order to obtain more accurate results on modeling elevation information. The set of sample points has higher vertical accuracy than the original DEM and it is used the geostatistical procedure, known as *kriging with an external drift* - KED -, to perform the conflations. A case study is presented over a region of Campinas, city of São Paulo State in Brazil, to illustrate the application of the methodology to a real information of the earth surface.

2. Concepts and Methodology

2.1. Main Concepts

Geostatistical approaches for estimations and simulations are based on previous analysis of the spatial correlation of a set of sample points to represent the spatial variability, spatial dependence, of the attribute in a geographical region. This variability in function of the spatial distance is represented by variogram models. Empirical , or experimental, variogram models, $2\gamma^*(\mathbf{h})$, can be estimated, directly from a set of sample points, according to Equation 1 below:

$$2\gamma^*(\mathbf{h}) = \frac{1}{\mathrm{N}(\mathbf{h})} \sum_{(i,j)/\mathbf{h}_{ij} \approx \mathbf{h}} \left(z(\mathbf{u}_i) - z(\mathbf{u}_j)\right) * \left(z'(\mathbf{u}_i) - z'(\mathbf{u}_j)\right) \qquad (1)$$

where $z(\mathbf{u}_i)$ and $z(\mathbf{u}_j)$ are attribute values observed at spatial locations $\mathbf{u}_i$ and $\mathbf{u}_j$ separated by the distance **h.** N(**h**) is the number of samples found inside a circumference with radius distance approximately equal to **h**.

The kriging procedure allows to infer a mean value of the attribute, in any spatial location **u**, from a number $n(\mathbf{u})$ of neighbor samples $z(\mathbf{u}_\alpha)$, α=1,...,$n(\mathbf{u})$. The general formulation for the kriging estimator is:

$$z^*(\mathbf{u}) - \mu(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \cdot [z(\mathbf{u}_\alpha) - \mu(\mathbf{u}_\alpha)] \qquad (2)$$

where $\mu(\mathbf{u}) = m(\mathbf{u})$ is the tendency, or the mean value, of the attribute in the spatial location **u**, $\mu(\mathbf{u}_\alpha)$ is the mean value in each sampled location $\mathbf{u}_\alpha$. The weights $\lambda_\alpha(\mathbf{u})$ are estimated considering the correlation structure defined by the modeled variogram for the set of sample points considered [Isaaks and Srivastava 1989].

Also, geostatistical approaches, for estimations and simulations, allow incorporating secondary information, along with primary one, in order to obtain more accurate and reliable models of the attribute. Variations of the kriging, using secondary information, can be: cokriging, universal kriging, regression kriging, kriging with varying local means, kriging with a trend model, kriging with an external drift, etc. [Ortiz et all 2007, Wackernagel 1998].

The KED approach is an extension of the kriging with a trend model, where the trend is obtained from a secondary (external) variable related to the primary one. The trends are used as the mean values, in the Equation 2, and the residual covariance, rather than the covariance, of the primary variable must be used to solve this kriging approach. [Deutsch and Journel 1998] point out two conditions that have to be met before applying the external drift algorithm: (1) The external variable must vary smoothly in space and (2) The external variable must to be known at all locations $\mathbf{u}_\alpha$ of the primary data and at all locations $\mathbf{u}$ to be estimated. The advantage in this case is that it is not necessary to know the cross covariance between the primary and secondary variables. In this work the input DEM is considered the external drift yelling elevation data at all locations of the condition (2) above.

2.2. Methodology

The methodology of this work is based on the *kriging with an external drift* procedure presented in [Deutsch and Journel 1998] and follows the steps below:

1. Import a SRTM file from the region of study
2. Import the high vertical accurate sample set of elevation points
3. Create a residual file of the sample points taken the trend off the sample set
4. Create empirical and theoretical variograms for the set of sample points
5. Generate a DEM using the set of sample points and its theoretical variogram
6. Use the SRTM to create an ASCII data file with collocated elevation values and the residuals
7. Generate a conflated DEM using the SRTM data as gridded external drift and the collocated SRTM and the residuals information.

3. Results and Analysis

As a case study it was chosen a small region of Campinas, city of São Paulo State in Brazil. The geographical location coordinates of the bounding box of this region are:
(w 47° 08', s 23° 00') e (o 46° 57', s 22° 50''"). The SRTM DEM information of the Campinas region is shown in the Figure 1.

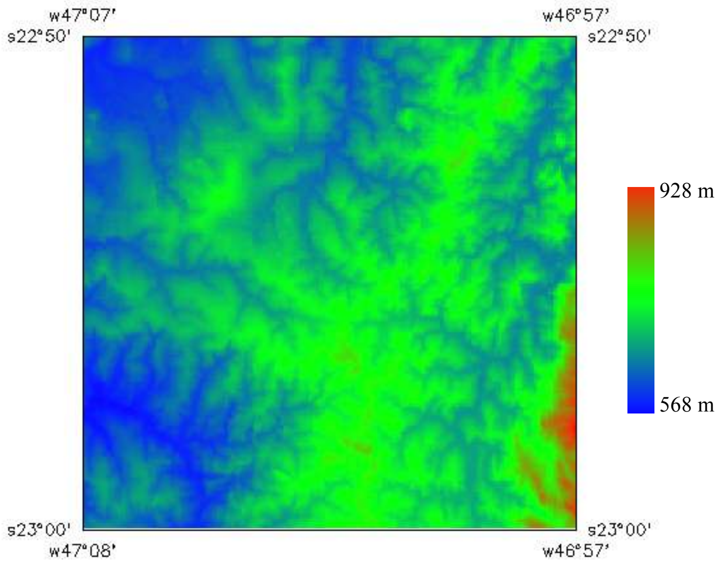

Figure 1. The SRTM DEM information of the Campinas region

The height resolution of this information, inside this region, is approximately 9.3m as assessed by analysis considering an accurate test set of 4549 elevation points.

In this experiment, it was considered a sample set of 505 elevation points sparsely distributed in the Campinas region. The height accuracy of the samples is 2.5m.

It was used geostatistical tools of the GIS known as SPRING, [Camara et al 1996] to perform most of the analysis and the estimations presented in this work. These tools were implemented in SPRING, [Camargo 1998], based on the original functions of the GSLib software developed by [Deutsch and Journel 1997]. Also, the GSLib was used to run the KED estimations to accomplish the conflation results.

Initially, the spatial locations of the sample points were used to obtain collocated elevation values from the SRTM information. This was performed by an external C program, developed for this purpose, that calculates also the residuals subtracting the elevation of the samples from the collocated SRTM elevation values.

Spatial dependence analysis of the sample points and of the residuals was performed to obtain empirical semivariograms. Theoretical semivariogram models (red curves) were fitted over these empirical semivariograms, as shown in Figure 2, representing the spatial variability of the elevations of the sample points (left) and of the residuals (right).

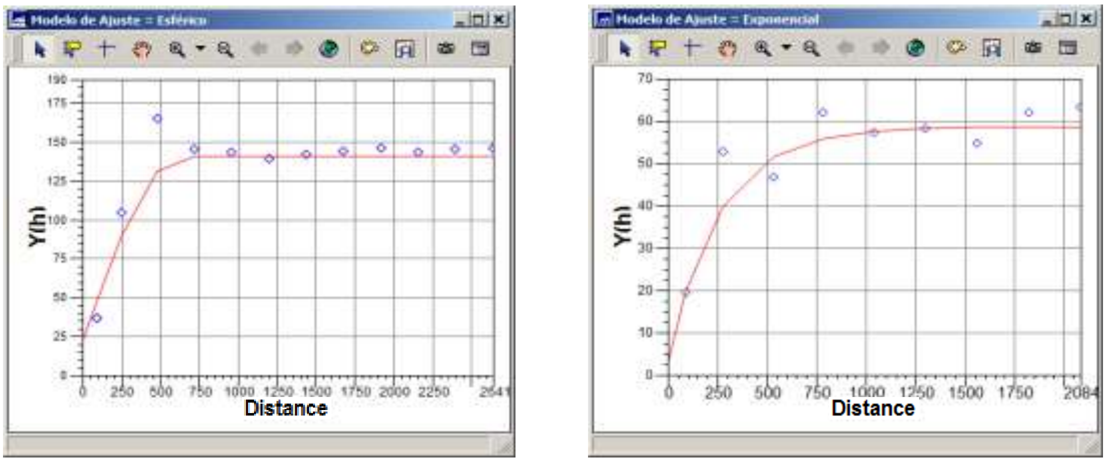

Figure 2. Theoretical semivariograms (red curve) fitted over empirical semivariograms (blue points) of the sample points (left) and of the residuals (right)

The theoretical variogram of the sample points is a spherical model with 21.48 m of nugget effect, 118.86 m of contribution and 645.46 m of range. The theoretical variogram of the residuals is an exponential model with 2.62 m of nugget effect, 55.82 m of contribution and 776.18 m of range.

The ordinary kriging procedure of the SPRING software was applied over the sample set of elevation points using the left theoretical variogram above. The resulting sample DEM, an elevation grid of size 200 columns by 200 rows, is shown in Figure 3. The spatial resolution of this grid is 90 meters in both x and y directions.

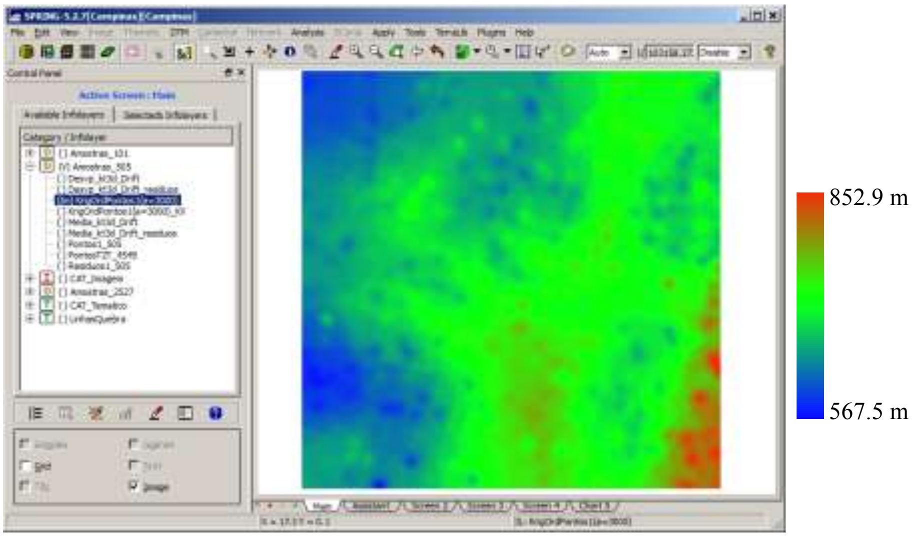

Figure 3. Digital Elevation Model estimated using only the sample points and their theoretical semivariogram

Following this, the kriging procedure of the GSLIB, known as K3td, was applied to the sample points, merged with collocated information of SRTM data. It was used theoretical semivariogram model of the residuals and the option of kriging with an external drift was chosen as parameters for this function. The gridded information of the SRTM was taken as the external drift file required for this KED procedure. The resulting DEM, an elevation grid of size 200 columns by 200 rows, is shown in Figure 4. The spatial resolution of this grid is 90x90 m in x and y directions.

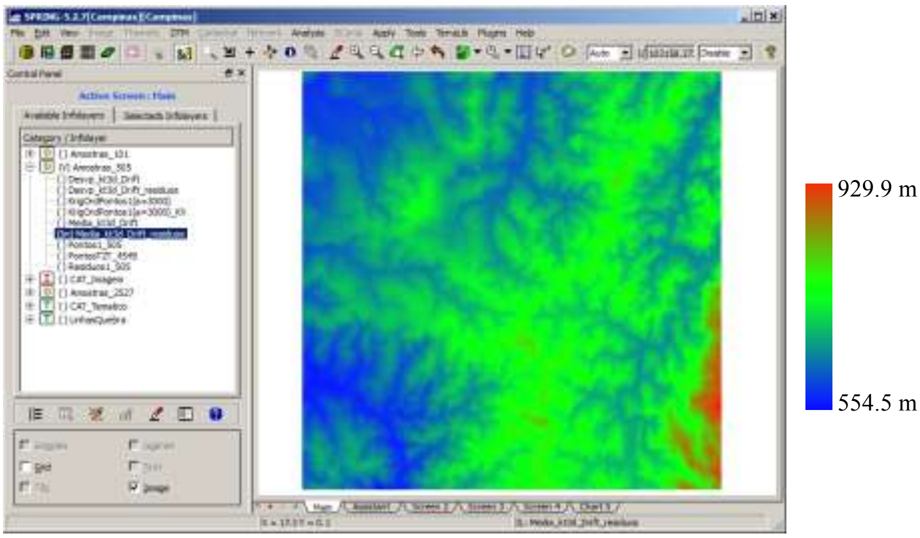

Figure 4. Digital Elevation Model obtained by conflation of the sample points and the SRTM data

A qualitative analysis, by visual inspection, between the maps of the Figures 1 and 3 shows that the SRTM DEM presents more detailed information. The map of Figure 3 is more homogeneous because the number of sample points is low, or it is not enough to represent the heterogeneity presented in the SRTM DEM .

A qualitative analysis, also by visual inspection, presents no differences between the DEMs of Figures 1 and 4. On the other hand, a quantitative analysis by a validation procedure, using a test set of 4549 elevation points, has shown that the vertical accuracy, measured by the standard deviation value, for the SRTM DEM is 9.32 meters.

When the same validation approach is applied to the DEM of the Figure 4 the calculated standard deviation value is 8.75 meters. This represents an increasing of 5.85% in accuracy of the final map (9.32-8.75)/9.32=0.0585. Table 1 presents other results obtained with different sets of sample points showing that the vertical accuracy increases with the number of sample points considered.

Table 1. Vertical DEM accuracy considering different number of sample points

Number of Samples (Tests) Points	Ordinary Kriging DEM (m)	SRTM DEM (m)	External Drift Kriging DEM (m)	Accuracy Improvement (%)
101 (4953)	34.53	9.28	9.09	2.20
505 (4549)	20.22	9.32	8.75	5.85
2527 (2527)	9.73	9.33	7.63	17.79

4. Conclusions

This article has shown that it is feasible to apply successfully the conflation technique using the geostatistical procedure known as kriging with an external drift. In this approach the SRTM DEM is considered the tendency of the information that is updated, or improved, with the information of a better accurate sample set of elevation points.

As shown by the case study it was obtained an improvement on the vertical resolution of 5.85% when it was used 505 sample points along with the SRTM considered. The improvement is better when the number of sample points is larger, as shown in Table 1. The results could be improved if a sample set with a better vertical resolution was used in this work.

This short paper presents only a small prototype representing the beginning of the researches related to the conflation techniques to improve existing DEMs.

Other spatial regions and geostatistical options will be considered in the future such as: cokriging and co-simulations, kriging and simulations with regression and with varying local means and the same variations for indicator krigings and simulations.

References

Burrough, P. A. (1987) "Principles of geographical information systems". New York: Oxford University Press. 344p.

Camara, G., Souza, R. C. M., Freitas U. M. and Garrido, J. (1996) "SPRING, Integrating Remote Sensing and GIS by object-oriented data modeling". Computer & Graphics, v. 20, n. 17, p.395-403.

Camargo, E. C. G. (1997) "Desenvolvimento, implementação e teste de procedimentos geoestatísticos (Krigeagem) no sistema de processamento de informações georreferenciadas (SPRING)". 1997-06. 146 p. (INPE-6410-TDI/620). Available at: <http://urlib.net/sid.inpe.br/iris@1912/2005/07.20.08.47.41>.

Deutsch, C. V. and Journel, A. G. (1998) "GSLIB: geostatistical software library and user's guide". New York: Oxford University Press, 369p.

Goovaerts, P. (1997) "Geostatistics for natural resources evaluation". New York: Oxford University Press, 483p.

Hengl, T., Bajat, B., Blagojevic, D., Reuter, H. I. (2008) "Geostatistical modeling of topography using auxiliary maps". ;Computers & Geosciences, 1886-1899.

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

Karkee, M.; Steward, B. L.; Azis, S. A. (2008) "Improving quality of public domain digital elevation models through data fusion". Biosystems Engineering, v.101, p.293-305, 2008.

Ortiz, J.O., Felgueiras, C.A.; Druck, S.; Monteiro, A.M.V. (2007) "Avaliação do procedimento de geoestatística de cokrigeagem para determinação da distribuição espacial de propriedades do solo". Geomática, Embrapa, cap. 7.

Wackernagel, H. (1998) "Multivariate Geostatistics". New York, Springer, 291p.