ON THE USE OF THE GEOSTATISTICAL ESTIMATION TECHNIQUES TO THE GENERATION, DEBUGGING AND ANALYSIS OF DIGITAL SURFACE MODELS

Jorge DELGADO[(1)], Javier CARDENAL[(1)], Alfonso GÓMEZ[(2)]
[(1)] Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría
Escuela Politécnica Superior. Universidad de Jaén. 23071-Jaén (Spain)
[(2)]Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría. Expresión Gráfica
E.T.S. de Ingenieros Agrónomos. Universidad Politécnica de Madrid. 28040-Madrid (Spain)
jdelgado@ujaen.es, jcardena@ujaen.es, "a_gomez" info@stereocarto.com

IC-15

KEY WORDS: Digital Surface Models, Data Processing, Mathematical Models, Modelling, Spatial Data

ABSTRACT

To utilise DSMs effectively it is necessary to ensure that they are sufficiently accurate to meet the requirements of our specific application. In order to assess the accuracy of a model two basic aspects must be considered: the measured data accuracy and the modelling quality. In this work we show a geostatistical approach to the surface modelling. The geostatistical methods are well suited to assessing accuracy and they allow treat the entire modelization process. These methods allow the study of phenomena that fluctuate in space. The process begins with a data capture using a dense grid in a representative zone of the study area. Those data are used for the spatial variability estimation, obtaining information about the influence zones and presence of anisotropies. The modeled spatial variability functions will be used for the final grid design considering the maximum admissible estimation error, that depends only of the spatial variability functions and the data positions –not data values-. Once the grid to be measured is defined, the final capture is made and the data are debugged using the cross validation method. Using the debugged data and the information derived from the structural analysis the modelization is made. The estimation will be optimum (non-biased and minimum estimation error). The estimation error will be known for the entire model and it will be a very important information for the evaluation of the model quality. Finally, we present an example of analysis using the Factorial Kriging method that allows the filtering using a structural space depending filter schema.

1 INTRODUCTION

Digital Surface Models (DSM) are used widely for many applications. DSMs may be used, for example, in cartographic works such as building extraction, true orthophotos generation or modelling environmental process. The use of DSM needs to ensure the accuracy of the generated models. In this work we present a geostatistical approach to the digital surface modelling. The geostatistical methods are well suited to assessing accuracy and they allow treat the entire process of the modelization (data debugging, spatial variability characterization, estimation of the models –and errors- and models analysis –filtering-). The geostatistical methods are based in the Regionalized Variables Theory (Matheron, 1965) and they are a group of methods for the spatial data treatment. These methods are widely used in several branches of the Earth Sciences (Mining, Environment, Hidrology, Petroleum, etc.) providing excellent results.

The proposed methodology begins with a data capture using a dense grid in a representative zone of the study area (using for example, a DPW). Those data are used for the spatial variability estimation (structural analysis), that consists in the semivariogram estimation in several directions obtaining the influence zones and variability for each direction (anisotropy detection). The modeled semivariogram functions will be used for the final design of the measurement grid considering the maximum admissible estimation error of the model, that depends only of the spatial variability functions and the data positions –not depend of the data values-. Once the final grid is defined, the data capture is made and the obtained data are debugged using the cross validation geostatistical method. The modelization is made using the debugged data and the information derived from the structural analysis. The estimation will be optimum in order the assure the non-bias of the model and the minimization of the estimation error. The estimation error will be known for the entire model and it will be a very important information for the evaluation of the model quality. Finally, using the Factorial Kriging method it is possible the model analysis in order to, for example, detect building or another obstacles using a structural depending filter schema (Figure 1).

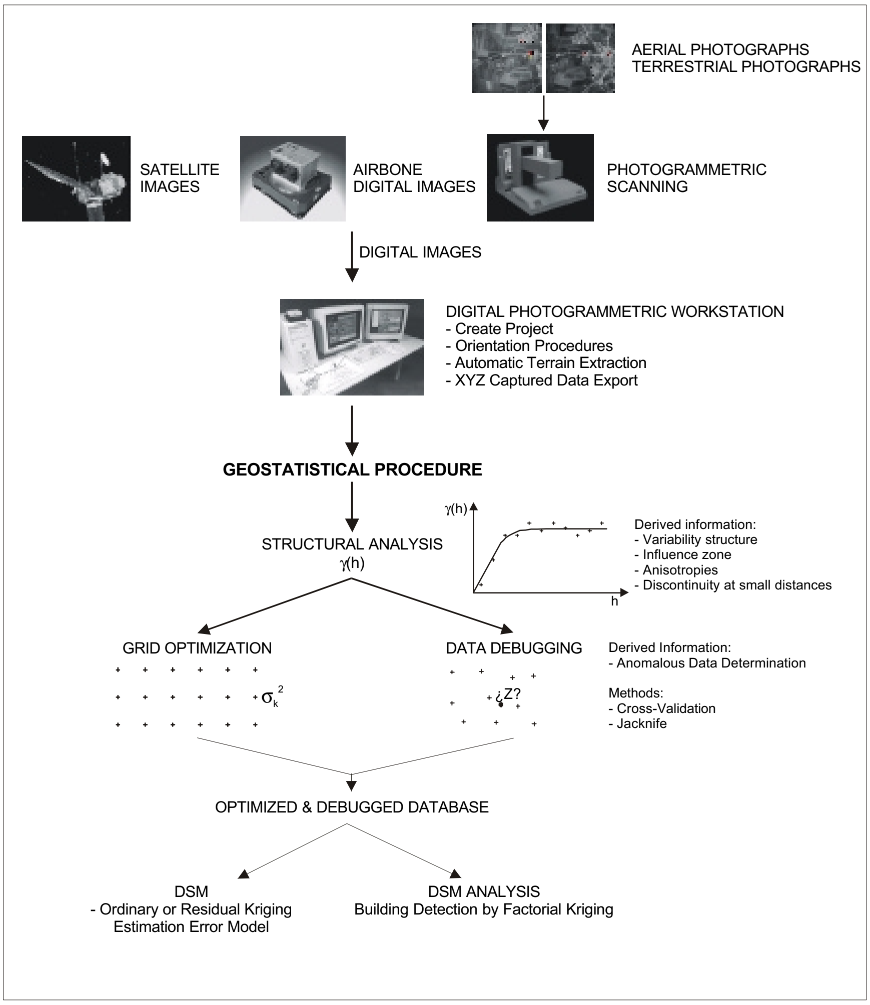

Figure 1. Workflow of the DSM geostatistical generation and analysis

2 APPLICATION

For the experimental investigations we pick up a well-known block “Echallens”. This is one of the sets of training and demonstration materials from Leica Aarau Ltd.. The characteristics of the block and the basic photogrammetric procedures in order to obtain the data for the DSM was shown in the Table 1 and Figures 2 and 3.

3 STRUCTURAL ANALYSIS

After the initial data capture, the geostatistical procedure can begin. The first phase is the preliminary estimation of the spatial variability behaviour or structural analysis. The structural analysis consists in the estimation and modelization of the spatial variability functions (semivariograms) in several directions of the space in order to establish differences (variability or scale anisotropies). A semivariogram is defined as the measure of one half the mean square error produced by assigning the value z(x+h) to the value z(x). In a probabilistic sense, the semivariogram, γ(h) is one half of the variogram of order 2 (Olea, 1991):

$$\gamma(h) = \tfrac{1}{2} Var\{Z(x) - Z(x+h)\} = \tfrac{1}{2} E\left[\{Z(x) - Z(x+h)\}^2\right] \quad (1)$$

expression that is estimated using the following equation:

$$\gamma^*(h) = \frac{1}{2NP(h)} \sum_{i=1}^{NP(h)} \left[Z(x+h) - Z(x)\right]^2 \quad (2)$$

where NP(h) is the number of pairs h apart and Z(x+h) and Z(x) are the experimental values in the x and x+h positions.

ECHALLENS BLOCK

Flight and camera data
Region: Echallens, Canton de Vaud, Switzerland, Date: 25.3.1980
Camera: Wild Aviophot RC10, f= 153.18mm
Photo scale: Approx.1/4300; Flying height: Approx. 1280m/MSL; Mean terrain altitude: 620m/MSL
1 strip (flight line 214), 3 images (4679, 4682, 4685) with approx. 70% forward lap
Terrain type: Varied (from flat –county- to urban)

Orientation
Interior orientation
4 fiducial marks (affine transformation), Automatic localization supervised by operator (SOCET SET v.4.2.1)
RMS: 0.50 pix, Max. Error X: 0.38 pix, Y: 0.57 pix
Exterior orientation
Automatic Aerial-triangulation (LHS Multisensor Triangulation System v.4.2.1)
34 tie points (obtained using a 3x5 tie point pattern), 19 full control points (data precision: 0.02 m X, Y, Z)
Orientation residuals (Image residuals):
Tie points: RMS: x_p: 0.057 pix, y_p: 0.008 pix; Max: x_p: 0.275 pix, y_p: 0.087 pix
Ground control points: RMS: x_p: 0.054 pix, y_p: 0.008 pix; Max: x_p: 0.572 pix, y_p: 0.172 pix
Orientation residuals (Terrain residuals):
RMS: X: 0.03 m Y: 0.02 m Z: 0.03m Max: X: 0.06 m Y: 0.06m Z: 0.06m (GSD=0.09m)

Terrain Extraction
Automatic Terrain Extraction (LHS TERRAIN v.4.2.1) Adaptative strategy (no smooth nor filters)
% Measured points: 83.61, Good points (Figure of Merit): 897782

Table 1. Characteristics of the Echallens block

Figure 2. Experimental information (photographs and ground control points) and selected areas

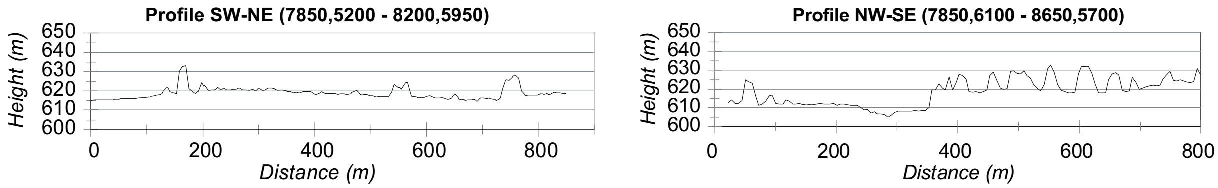

Figure 3. Topographical profiles along the area

The initial data were used for the experimental semivariogram calculation in the four main plane directions with an angular tolerance of 22.5º (Figure 4). This figure shows a clear trend that produces a strongly anisotropic behaviour of the variable (elevation) on the considered areas (except for the HOUSES area).

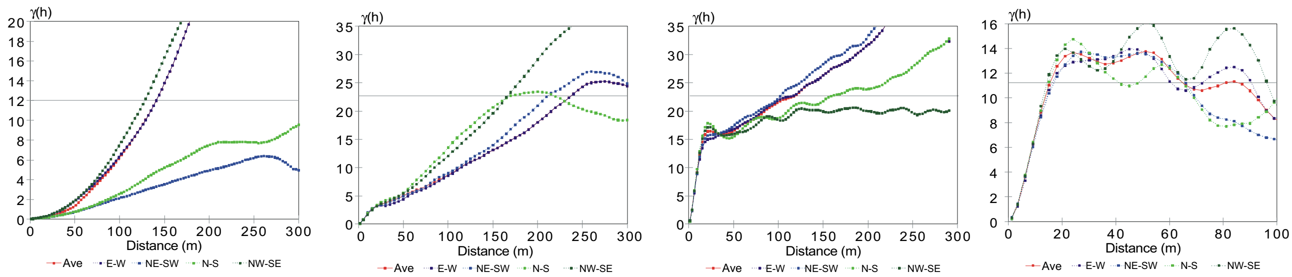

Figure 4. Experimental directional semivariograms for the selected areas (OPEN, MIXED, URBAN, HOUSES) (see the non-stationary behaviour –drift- of the three first obtained semivariograms)

Because the aim of this work is the development of an easy approach for people not trained in Geostatistics, the trend was eliminated in the zones where it was detected and the further calculations were based on residuals. The trend was estimated by least squares adjustment using a quadratic model ($z(x,y)=A+Bx+Cy+Dx^2+Exy+Fy^2$). It is important to point out that the selected model do not affect to the final results. The proposed methodology is based in the drift elimination and work with the residuals ($r(x,y)=z(x-y)-m(x,y)$), and the final results are equal to the estimated residual plus the estimated drift ($z^*(x,y)= m(x,y) + r^*(x,y)$). The experimental semivariograms of residuals and models are shown in Figure 5 and the parameters of the fitted models are summarized in the Table 2 (see Journel and Huijbregts, 1978, for more detailed explication about semivariogram fitting).

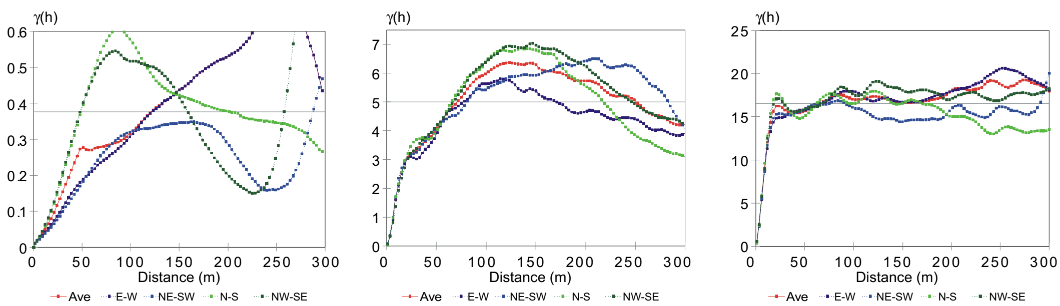

Figure 5. Experimental directional semivariogram of the residual for the non-stationary areas (see the stationary behaviour of the obtained semivariograms)

Zones	Structures and parameters
OPEN	Struct.1: Type: Spherical Range: 132 m Sill: 0.39m² Angle: 0º Anis. Ratio: 0.28 AIC: -433.857 (goodness of fit parameter)
MIXED	Struct.1: Type: Spherical Range: 150 m Sill: 5.90m² Angle: 90º Anis. Ratio: 0.30 AIC: -567.253 (goodness of fit parameter)
URBAN	Struct.1: Type: Exponential Range: 22 m Sill: 16.78m² Isotropic AIC: -643.364 (goodness of fit parameter)
HOUSES	Nugget effect: 0.1 m² Struct.1: Type: Gaussian Range: 13 m Sill: 12.00m² Isotropic Struct.2: Type: Spherical Range: 23 m Sill: 1.00m² Angle: 135º Anis Ratio: 0.3

Table 3. Semivariogram modelling (anisotropy parameters in GSLIB format, Deustch and Journel, 1997)

The main conclusions from this analysis are:

- OPEN ZONE: Slighty anisotropic behaviour with larger continuity in N-S NE-SE directions (125m vs. 75m). This anisotropy was produced by an ondulated terrain (Figure 3).
- MIXED ZONE / URBAN ZONE: Isotropic behaviour with a larger continuity in the MIXED zone (lower variability) with a total range of 100m vs 30m of the URBAN zone.
- HOUSES ZONE: Only this zone does not present trend (stationary). Low continuity produced by the irregular relief due to the presence of houses.
- Logically, the zone of influence defined by the range is smaller when the discontinuities of the zone (variability) are larger.

4 NETWORK DESIGN

One of the most interesting contributions of the geostatistical methodology to the model generation is the optimization of the measurements grid. This optimization is based in the estimation error analysis. The estimation error only depends of the spatial variability structure and the data localization; the specific values do not have any influence. Using this conceptual basis, the measurements grid of the HOUSES zone has been optimized using the ordinary kriging geostatistical estimation method. The results have been summarized in the Figure 6. This figure shows the relation between data density (expressed in No.points/m^2) and estimation error -kriging standard deviation- (expressed in meters) for a square regular grid. The curve presents an asintotic behaviour, and it marks a limit beyond it the error is approximately constant

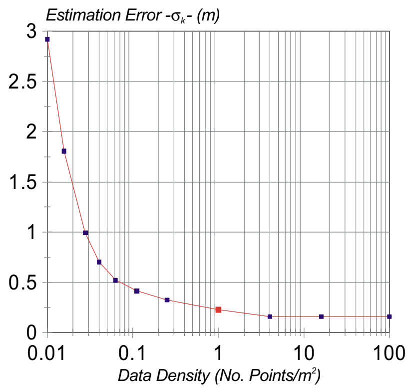

Figure 6. Relation between data density of the regular data grid and the maximum estimation error of the model (in red colour the selected grid 1m x 1m with estimation error equals to 0.23m)

5 DATA DEBUGGING

A very important aspect of the automatic terrain extraction systems is to provide adequate tools for the blunder detection. The quality of these tools is basic in order to assess the quality of the final models and to reduce to the edition time consumption. The geostatistical cross validation method (XVALID) consists in the validation of the data in which the observations are dropped one at a time from a sample of size n, and n estimates are computed using at most the remaining (n-1) measurements (Olea, 1991). The comparison between the real values and the estimated values calculated on the censored subsets is the basis for evaluating the similitude between a data and its neighbourhood. This is a classical criterion used in the data debugging that it is improved using the spatial variability techniques (anisotropies, variability and influence zone).

The cross validation method provides for each data position the estimated value and estimation error. With this information we can obtain the zscore coefficient (estimated minus real value divided by the estimation error -kriging standard deviation; σ_k-. The zscore probability distribution aproximately fits to a gaussian probability distribution and allows the probability error estimation. This method has been applied to the entire area data using the xvokb2d program from the GSLIB library (Deutsch and Journel, 1997, Deutsch, 2000). The computed errors only reveal anomalous values with respect to the neighbourhood. The diferences can be produced by a correlation error or, simply by the proper terrain morphology. The most important advantage of this method, comparing to the classical methods (gradient, local parallax), is the use of the variogram information and the error estimation values. The Figure 7 shows an example of the bad correlated point error detection. This point has not been considered by suspicious by the SOCET SET system –figure of merit- (correlation coefficient equals to 84).

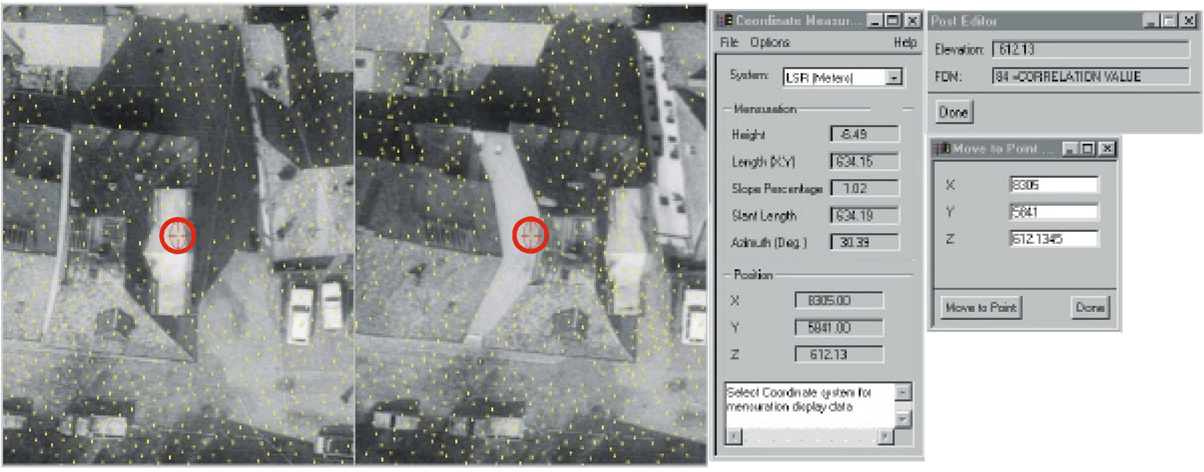

Figure 7. Example of bad correlated point error detected by XVALID (error=11.7m, zscore= 3.215, p=99.87%). Circles mark the position of the point in both photographs.

6 DSM GENERATION

Once the semivariograms have been calculated and fitted to a theoretical model, we know the variability structure of the phenomena. The model will be defined by: the nested structures needed for the correct fit of the model (the different structures indicate the different variability scales of the terrain elevation); the range value (maximum distance separing pairs of data that has any significant statistical dependence); the sill value (variability between pairs of data that has not any significant statistical dependence); and the angle and ratio of anisotropy (that allows the orientation of the directional variability in the plane). The information from the structural analysis will be used combined with the position of the measured data to form the kriging equation systems (eq.3). The solution of these systems (one system for each estimated point) will provide the weight for each point. These weights will be used to obtain the estimated value and the estimation error (kriging variance).

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta \gamma(x_\alpha, x_\beta) - \mu = \gamma(x_o, x_\alpha) \\ \sum_{\beta=1}^{n} \lambda_\beta = 1 \end{cases} \tag{3}$$

$$Z^*(x) = \sum_{\beta=1}^{n} \lambda_\beta Z_\beta \qquad \sigma^2_{OK} = \sum_{\beta=1}^{n} \lambda_\beta \gamma(x_\alpha, x_\beta) + \mu$$

where n are the number of β points that participate in the estimation process (defined by the search ellipse and by the maximum and minimum estimation data parameter); λ is kriging weight (solution of system); γ(i,j) is the semivariogram value for the distance between the points i and j; x_o is the point to be estimated; μ is the Lagrange parameter; $Z^*(x)$ is the estimated value; Z_β are the experimental data and σ^2_{OK} is the error estimation –kriging variance-

All of estimation methods are affected by smooth effect, but kriging, allows the elimination of this smoothing by using the following expression (Isaaks and Srivastava, 1989):

$$Z_c^*(x) = Z^*(x) - \frac{\sigma_k^2}{\sigma_{k_{max}}^2}\left(Z^*(x) - \frac{Z^*(x) - b}{a}\right) \tag{4}$$

where $Z^*_c(x)$ is the corrected value, $Z^*(x)$ is the estimated value, σ^2_{OK} is the estimation error –kriging variance-and σ^2_{kmax} is the maximum estimation error, a is the slope of the line of regression between experimental and estimated values and b is the intercept of the line (this parameters are obtained using linear regression on the cross validated data).

In the figure 8 are shown the final results of the estimation in the HOUSES zone. The most important results are: the estimated value (in point or block support in a raster o contour level format) and the estimation error (directly related with the availability with experimental points).

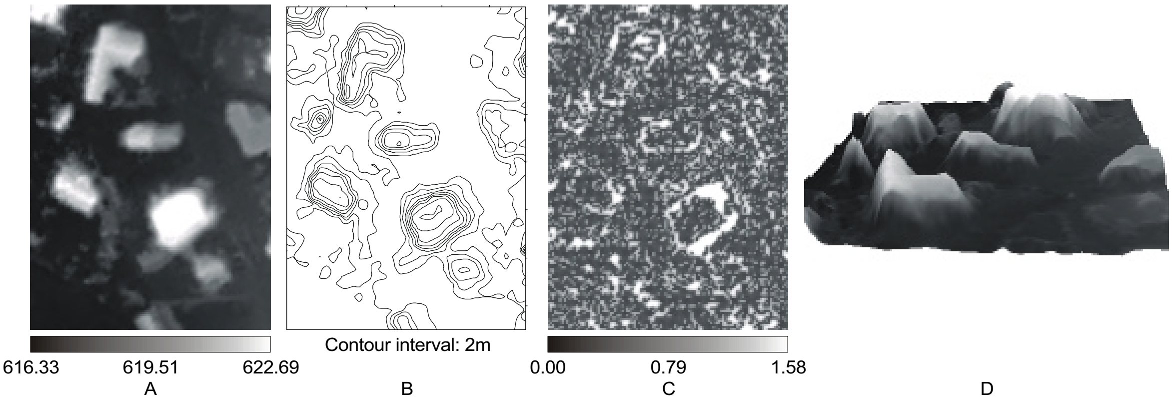

Figure 8. DSM of the HOUSES zone. A) DSM in raster format; B) Contour map derived from DSM; C) Estimation error representation (kriging variance); D) 2.5D representation of DSM.

6.1 Quality Control

The generated model was checked in order to establish its precision. For this quality control we have used the ground control points and tie points from the orientation (aerotriangulation) process. With this procedure we try to eliminate the possible error produced from an incorrect orientation in the zone used for the check control. The number of used check points has been 53 (19 full control points –precision 0.02m-). The admissible point error of the points has been established in 0.09m (ground sample distance). For each point we obtained the estimated height values from the SOCET SET® system and from the Ordinary Kriging model using the jackknife technique (Quenouille, 1956, Tukey, 1958). The jackknife J(θ) is a statistical estimator of an unknown parameter θ of the form

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} J_i(\theta) \quad \text{where} \quad J_i(\theta) = m\Theta - (m-1)\Theta_i \tag{5}$$

where m is the number of subsets in a partition of the data set, Θ is an estimator of θ using the entire data set and Θ_i is an estimation of the same parameter, but computed excluding the ith partition from the data set (Olea, 1991). The statistics of the error (difference between real and estimated value) are summarized in Table 4.

	Jacknife (OK)	SOCET SET®		Jacknife (OK)	SOCET SET®
Mean	-0.03	-0.04	**Minimum**	-1.14	-1.36
Std.Dev.	0.20	0.23	**Lower Quartile**	-0.06	-0.06
Coef.Var.(%)	607.977	540.544	**Median**	-0.01	-0.02
Range	1.47	1.72	**Upper Quartile**	0.03	0.04
RIQ	0.09	0.10	**Maximum**	0.33	0.36

Table 4. Basic Statistics of DSM Quality Control (all values in meters)

In both cases the results are similar but there is some methodological different between them. So the XVALID method provide the estimated value and the estimation error. The estimation error is a basic statistical parameter in order to establish the model quality. This error is function of the model of semivariogram selected and the data localization.

7 DATA EXTRACTION

Finally, we analyze the model for detect different size obstacles (filtering). The geostatistical method of filtering is the Factorial Kriging –FK- (Galli et al., 1984). The FK is similar than the Fourier analysis and consists in the descomposition of the spatial variability of the regionalized variable (spectrum descomposition) using the semivariogram information. This method was applied to the filtering of the HOUSES DSM obtained the results shown in the Figure 9. At medium range scales (12 and 20m) we can detect the buildings of the zone (larger density of anomalous data), at a very short range the anomalies detected are produced by small irregularities of the terrain.

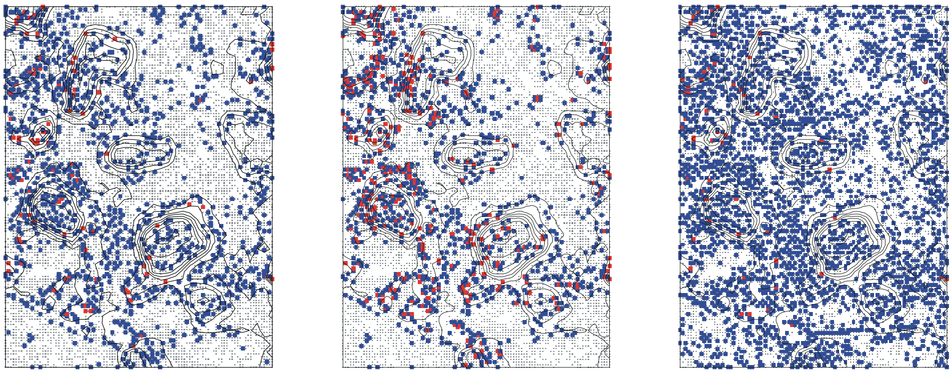

Figure 9. Obstacles detection by FK. Left: 20m scale; Center: 12m scale; Right: 1m scale. In all the figures the blue colour represents positive obstacles and the red colour negative ones.

8 CONCLUSIONS

In this work, we show the capabilities of the geostatistical estimation methodology in the topocartographic data treatment. We present an example of the application to the surface modelling. The principal advantages of this methodology are based in the topoprobabilistic point of view that the Geostatistics has. The structural analysis provide a very interesting information for the modelling (zone of influence, anisotropy presence) and they are very useful for the grid desing and blunder data detect. The estimation of the model uses the structural analysis information and provide the estimation error values –a basic statistical parameter in order to establish the model quality-. Finally, we have applied the FK method for the model filtering and the detection of several scale obstacles.

ACKNOWLEDGEMENTS

The work of this paper was partially supported by the Research Groups “Ingeniería Cartográfica” (Cartographic Engineering -TEP 164-) and “Recursos Hídricos” (Hydric Resources -RNM 126-) of the Andalusian Regional Goverment. We also thank to LH Systems the information provided (photographies and control ground data of project Echallens) for the realization of the practical example. Original colour images are available by request to the authors.

REFERENCES

Deutsch, Ch. (2000). GSLIB software web page, http://www.gslib.com.
Deutsch, Ch.; Journel, A. (1997). GSLIB-Geostatistical Software Library and User’s Manual. Oxford University Press, 2nd edition, New York, 369 p.
Galli, A.; Gerdil-Neuillet, F.; Dadou,C. (1984). Factorial Kriging Analysis: a substitute to spectral analysis of magnetic data. In: Verly, G. et al. (eds). Geostatistical for natural resources characterization. Ed. Reidel Publ.Co., Dordercht, pp.543-547.
Isaaks, E.H. and Srivastava, R.M. (1989). An Introduction to Applied Geostatistics. Oxford University Press. New York. 561 p.
Journel, A. and Huijbregts,Ch. (1978). Mining Geostatistics. Academic Press, New York, 600 p.
Matheron,G. (1965). Les variables regionaliseés et leur estimation. Ed. Masson et Cie, Paris, 212 p
Olea, R.A. (1991). Geostatistical Glossary and Multilingual Dictionary. In: International Association for Mathematical Geology Studies in Mathematical Geology, No. 3, 177 p.