PANAMJAS
Pan-American Journal of Aquatic Sciences

Comparing the performance of two spatial interpolation methods for creating a digital bathymetric model of the Yucatan submerged platform

JAVIER BELLO-PINEDA[1] & JOSE LUIS HERNÁNDEZ-STEFANONI[2]

[1]*Centro de ecología y pesquerías de la Universidad Veracruzana, Hidalgo 617 Col. Rio Jamapa, Boca del Rio, Veracruz Mexico. (229) 956 70 70. (E-mail:bellopj@yahoo.com).*
[2]*Centro de Investigación Científica de Yucatán A.C., Unidad de Recursos Naturales. Calle 43 # 130, Colonia Chuburná de Hidalgo, Mérida Yucatán. Tel (999) 981 3923 ext 372. (E-mail: jl_stefanoni@cicy.mx). Author for correspondence.*

Abstract. Bathymetry is one of the most conspicuous variables to consider in any study involving marine environments. In developing countries, accurate and up to date bathymetric charts are rare or in most cases limited to local areas, most of them available only in hardcopy format. This is the case of the Yucatan submerged platform. In this study, we compare and discuss the performance of two spatial interpolation techniques for creating a digital bathymetric model (DBM): Inverse distance functions and Kriging. The DBM for Yucatan Peninsula submerged platform was generated using 2650 depth point-data values digitized from navigational charts previously published by the Mexican “Secretaria de Marina” at a 1:850,000 scale. The exponential Kriging model produced the most accurate estimates, reducing the error in 18.2% compared with the inverse distance functions. Our results might become helpful to other researchers trying to decide the type of interpolation technique and selection of the model when elaborating digital elevation models for regional mapping purposes.

Key words: Inverse distance functions, kriging, geostatictics, bathymetry, Yucatan Platform.

Resumo. Comparación del desempeño de dos métodos de interpolación espacial para crear un modelo batimétrico digital de la plataforma sumergida de Yucatán. La batimetría es una de las variables más sobresaliente a considerar en estudios de medio ambientes marinos. En los países en desarrollo son raras las cartas batimétricas precisas y actualizadas o en la mayoría de los casos estas se limitan a áreas muy particulares. Muchas de estas cartas están disponibles solamente en formato impreso. Este es el caso particular de la plataforma de Yucatán. En este estudio comparamos y discutimos el desempeño de dos técnicas de interpolación espacial usadas para crear un modelo batimétrico digital (MBD): funciones de distancia inversa y kriging. El MBD fue generado usando 2650 valores puntuales de profundidad, los cuales fueron digitalizados de cartas de navegación publicadas por la Secretaria de Marina, a una escala de 1:850,000. El modelo de kriging exponencial produjo las estimaciones más precisas, reduciendo el error de estimación en al menos un 18.2% comparado con las funciones de distancia inversa. Nuestros resultados podrían ser útiles para otros investigadores tratando de decidir el tipo de técnica de interpolación y la selección del modelo a emplear cuando se elaboran modelos de elevación digital con propósitos de mapeo regional.

Palavras-chave: Funciones de distancia inversa, kriging, geoestadística, batimetría, plataforma de Yucatán.

Introduction

Bathymetry is the process of measuring seafloor water depths and producing realization of underwater topography. The development of bathymetric models is of great importance for the study of underwater environments, and they frequently are the only type of data available for inferring the geology over much of the ocean floor (Bowles *et al.* 1998).

The importance of having accurate bathymetry estimations is evident for a variety of scientific fields related with monitoring, evaluation and assessment of marine environments. The study of geology and structure of ocean floor (Wright *et al.* 2000, Ramírez-Herrera & Urrutia-Fucugauchi 1999), the study of physical oceanographic processes such as currents, tides, water mix, and nutrient transport (Klenke & Schenke 2002, Merino 1997) as well as the study of biological processes such as larval transport, algal blooms and species abundance and distribution, relies on the availability of accurate estimations of bathymetry (Carlon 2002, Epifanio & Garvine 2001). However, in most developing countries the availability of accurate and up to date bathymetric charts is rare or in most cases is limited to a few charts for local areas, most of them available only in hardcopy format.

Geostatistical techniques are useful in providing estimates of sample attributes at locations with sparse information (Burrough & McDonnell 1998). These methods work by defining the spatial structure of the phenomena (i.e. by autocorrelation or auto-covariance functions such as semi-variograms), then estimating values between measured points based on the degree of spatial autocorrelation or covariance found in the data (Robertson 1987, Isaak & Srivastava 1989). Kriging procedures and their required variography are not, however, without critics. It is argued that the structural analysis (variography) may be a rather involved and even a somewhat subjective process. Consequently, simpler alternatives to kriging, such as the inverse distance weighting have been used as interpolation methods (Merwade *et al.* 2006, Kravchenko & Bullock 1999). This technique is easier to implement due to the fact that the estimation of values does not require any measure of either spatial autocorrelation or spatial auto-covariance.

In this study, two spatial interpolation methods namely inverse distance weighting functions and kriging, were compared in terms of accuracy of the estimates for creating the best DBM for Yucatan submerged platform.

Methods

The Yucatan Peninsula is a large calcareous platform that extends into a submerged area called the "Campeche Bank". It is located between 19° 40' and 21° 37' N and 87° 30' and 90° 26' W. Two bodies of water, the Gulf of Mexico and the Caribbean Sea, border the coast of the Yucatan Peninsula (Figure 1). The Peninsula attained its present shape in the late Pliocene; with important depositions on the coast during the Holocene, and platform reefs which are continually developing to its North and East ends. Variations in water depth have been documented, the maximum interglacial sea level was 30 m higher than today, at the Pliocene; and in the early Holocene sea level was some 100 m lower than today, and present sea level was attained only 5500 years ago. (Schmitter-Soto *et al.* 2002).

Depth data used in this study were obtained from the navigational charts S.M 800, and S.M. 900 published at a 1:850,000 scale by "Secretaria de Marina", Mexico in 1994 and 1995, which updated the information from previous charts from the same source published during 1972 and 1977 respectively. The charts were acquired in hardcopy format and digitized to raster format using a personal scanner. Raster images were imported to the Idrisi GIS-software (Eastman 1999) and geometrically corrected to the Latitude-Longitude coordinates system using degrees and decimal degrees as units, referenced to the ellipsoid GRS 1980 and North America datum 1983. Figure 2 shows the distribution of 2650 punctual depth data values and the coast line, obtained from: http://crusty.er.usgs.gov/coast.getcoast.html.

To produce Digital Bathymetric Models that "best" represent the depth variability for the Yucatan submerged platform, we compare the performance of two spatial interpolation techniques: Inverse distance weighting and Kriging. All geostatistical analyses were made using the GS+ ™ software (Robertson 2000) and then exported to Surfer 8 ™ and Arc View 3.2 ™ for final enhancement and display.

Inverse distance weighting functions is a nearest neighbor interpolation technique that combines both the neighborhood and gradual change notions (Burrough & McDonnell 1998). Estimates of depth values at unvisited points are obtained as a weighted average of his neighbors (sampled points), where the closest points have more weight (importance) than those that are far away. The weighted values are based on an inverse function of the distance to the neighbors. The inverse distance function is expressed with equation a):

$$Z = \sum_{i=1}^{n} \left(\frac{1/d_i^p}{\sum 1/d_i^p} Z_i \right) \qquad \text{a)}$$

Where Z is the estimated depth value, Zi is the depth value calculated at the location i, d is the separation between the estimated point and the sampled location, p is an analysis-defined power parameter and n represents the number of sampling points used for estimation.

The main factor affecting the accuracy of inverse distance interpolator is the value of the power parameter p (Isaak & Srivastava 1989). In this study, we compared estimates of inverse distance interpolator using different integer powers parameters 1, 2, 3, 4, 5 and 6, which are the most commonly used in literature (Kravchenco & Bullock 1999, Burrough & MacDonnell 1998). Since the goal of using inverse distance functions as estimators is giving more weight (importance) to the closest sampled points (Webster & Oliver, 2001), in this study we just considered integer values of p parameter, because the values lower than 1 are closest to a simple average estimation (Isaaks & Srivastava 1989). In addition, the size of the neighborhood and the number of neighbors are also relevant to the accuracy of the results. Here, the closest 16 sampling point in a radius of 1.7 decimal degrees were used to perform the estimations. The choice of neighborhood size was obtained as the maximum separation in which autocorrelation exists between two points located in space (Figure 3).

Kriging is a spatial interpolation technique based on the spatial structure of sampled points. Using kriging, estimates of depth values at unsampled locations are obtained from the information provided by the structures of spatial variability, as depicted by an autocorrelation function, in this case the semi-variogram of depth values. Such structures help in defining the size and shape of the neighborhood for interpolation (i.e. sampling points that are spatially auto-correlated to the point to be estimated). The semi-variogram is computed by using equation b):

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} \left(Z(x_i) - Z(x_i + h) \right)^2 \qquad \text{b)}$$

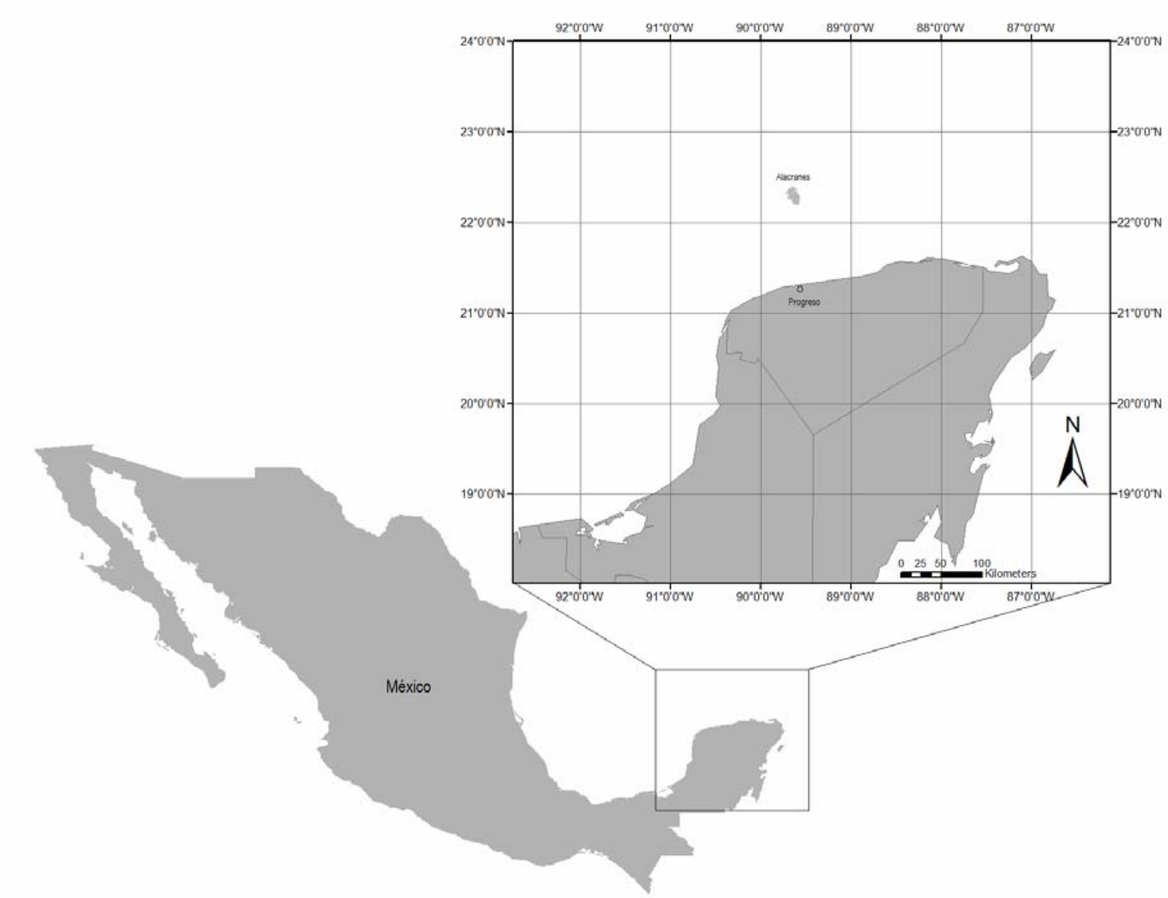

Figure 1. Geographic location of Yucatan peninsula.

Where $Z(x_i)$ is the depth in the location *i*, $Z(x_i + h)$ is the depth value of other points separated from x_i, by a discrete lag h; n represents the number of pairs of observations separated by *h*, and $\gamma(h)$ is the estimated or "experimental" semi-variance value for all pairs at lag *h*.

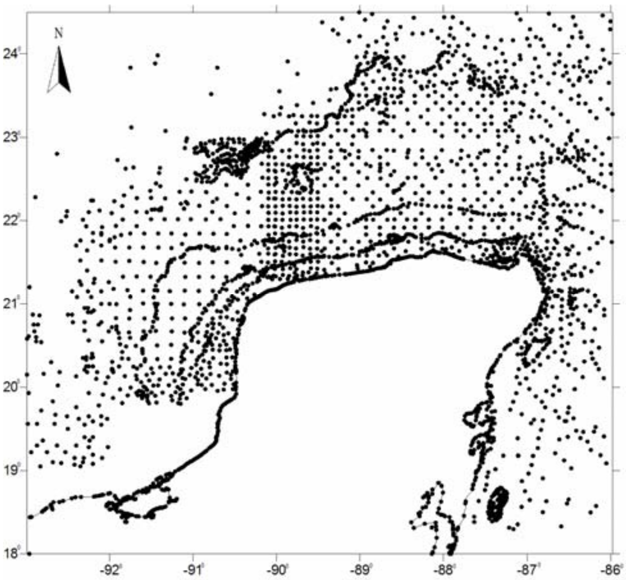

Figure 2. Distribution of punctual depth values as digitized from the navigational charts S.M 800, and S.M. 900 published by "Secretaria de Marina".

Semi-variances were calculated for each possible pair of sampling points, and the mean values of semi-variances were plotted for increasing lag intervals (*h*) to produce the experimental semi-variogram. Spherical, gaussian and exponential models were then fitted to experimental semi-variograms (Robertson 2000). The fitted models provided the following parameters; the total variance -also known as the "sill" variance- which defines the asymptotic value of semi-variance with respect to the lag separation. The sill variance is split in two, the variance due to spatial dependence and the random or "nugget" variance. In turn, the nugget variance, the *y* intercept in the semi-variogram model, reflects both, the spatial variation at shorter lags than the minimum sample spacing and the unexplained variance. The range of influence is the maximum separation at which depth values are still spatially dependent (Isaak & Srivastava 1989, Burrough & Mcdonnell 1998, Webster & Oliver 2001). The coefficient of determination (r^2) resulting from fitting of models to experimental semi-variograms, and cross-validation procedures described later on, were both used as criteria to select the best models in each situation.

Depth value estimates were obtained by using block kriging as expressed with the equation c):

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \qquad \text{c)}$$

Where λ_i are the optimal weights selected to minimize the estimation variance (Webster y Oliver, 2001), $Z(x_i)$ are the observed values of depth and $Z(x_0)$ is the optimal and unbiased estimate of depth values. The bathymetric maps were obtained using at least 16 sampling locations within a maximum radius of 1.7 (Decimal degrees). Neighborhood characteristics were determined by the range of influence of the semi-variogram.

The performance of each interpolation technique, in terms of the accuracy of estimates, was assessed by comparing the deviations of estimates from the measured data through the use of a "jack-kniffing" technique or cross-validation (Isaak & Srivastava, 1989, Webster & Oliver 2001). In such a procedure, sample values are deleted from the data set, one at a time and then the value in turn is interpolated by performing the interpolation algorithm with the remaining sampling values. This yields a list of estimated values of depth data paired to those measured at sampled locations. Therefore, the comparison of performance between interpolation techniques was achieved by using the following statistics: correlation coefficient between measured and estimated depth values, the mean error (*ME*), the mean absolute error (*MAE*) and the root mean square error (*RMSE*) (Zar 1999).
The *ME* is used for determining the degree of bias in the estimates and it is calculated with equation d):

$$ME = \frac{1}{n}\sum_{i=1}^{n} \hat{Z}(x_i) - Z(x_i) \qquad \text{d)}$$

The MAE provides an absolute measure of the size of the error. MAE is calculated with the equation e):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| Z(x_i) - Z(x_i) \right| \qquad \text{e)}$$

The RMSE provides a measure of the error size that it is sensitive to outliers. RMSE values can be calculated with equation f):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{Z}(x_i) - Z(x_i)\right)^2} \qquad \text{f)}$$

Finally, the relative improvement of the best method compared with the other procedures is calculated with equation g):

$$RI = \frac{100\left(RMSE_{Best} - RMSE_{Currentl}\right)}{RMSE_{Best}} \qquad g)$$

Where $RMSE_{Best}$ are the minimum value of *RMSE* and $RMSE_{Current}$ represent the *RMSE* of the current model.

Results

The spatial variation depicted by the semi-variogram models are shown on Table I. Spherical, Gaussian and Exponential models were found to fit well the experimental semi-variograms, and to explain the spatial autocorrelation present in the depth variable (Figure 3), yielding an r^2 ranging from 0.95 to 0.97 (Table I). The structural variance, which determines the variance explained by the model and calculated as (total variance–nugget variance)/total variance*100, ranged from 85.0% to 99.9%. This not only suggests that most of the variability of depth values is explained by the models, but also that a small fraction of variability is attributable to the nugget variance, which was ranged from 0.1 to 15.0 %. The range of influence showed values between 1.4 and 3.0 decimal degrees, indicating that one would reasonably expect that the depth values in places separated as far as in between 1.4 and 3.0 decimal degrees are still somewhat related.

The results, in terms of the accuracy of estimates (estimated errors), obtained from the cross-validation procedures are presented in Table II. The mean error (*ME*) is generally lower for kriging methods as interpolators. The depth values when predicted by kriging resulted in average underestimations of 3.6 and 3.8, being this the lowest values compared with those of inverse distance procedures, which gave a mean underestimation higher than 12.2. The other two measures of error, i.e. *MAE* and *RMSE*, showed similar behavior for all methods. The highest values of these measures of errors were obtained with inverse distance methods. In the same way estimated depth values are more correlated with measured depth data when kriging is utilized (Figure 4). Therefore, there is evidence that the accuracy of depth values estimations is improved when kriging procedures are applied. The relative improvement of the best technique compared with the others is also showed in Table II. Kriging procedure allowed at least a reduction of 18.2% in the error compared with the inverse distance functions.

The results strongly suggest that the accuracy of estimates and therefore the accuracy of mapping depth values were improved by using kriging (Table II). Furthermore, it must be taken into account that the kriging technique has an intrinsic additional advantage over the other interpolation method since its estimates are unbiased and with minimum variance. Thus, they are accompanied by a measure of the error in each predicted value: the estimation variance (Webster & Oliver 2001). This measure of the estimation error is provided by most of the geostatistical software programs, including GS+™. The final DBM produced by using the exponential model with kriging, was imported to the software Surfer ™ for final enhancement and display. Figure 5 shows a 2.5-D representation of the final DBM.

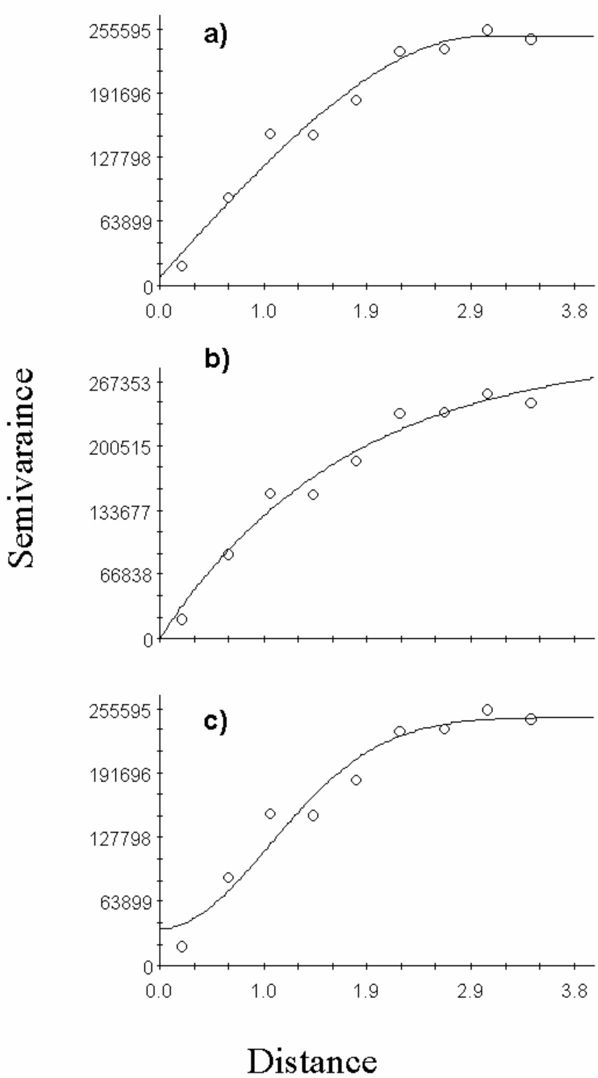

Figure 3. Experimental and model semi-variograms of depth values: a) spherical model, b) exponential model and c) gaussian model.

Table I. Parameters and statistics of semi-variogram models describing the spatial variability of depth values.

Model	Nugget Variance	Total Variance	Range	Relative Structural (%) Variance	r^2
Spherical	8500.00	248500.00	3.005	96.6	0.972
Exponential	100.00	300500.00	1.724	99.9	0.973
Gaussian	36900.00	246800.00	1.417	85.0	0.951

Table II. Results of mean error, mean absolut error, root mean square error, correlation coefficients between measured and estimated depth values.

Interpolation Procedure	ME	MAE	RMSE	Corr	RI (%)
Kriging (spherical)	-3.83	40.98	155.41	0.956	1.42
Kriging (exponential)	-3.55	36.68	153.21	0.957	0.00
Kriging (Gaussian)	-6.06	45.97	162.14	0.939	5.51
Inverse distance-1	-26.77	56.39	221.62	0.917	30.87
Inverse distance-2	-21.77	48.48	201.41	0.929	23.93
Inverse distance-3	-17.64	44.62	191.08	0.934	19.82
Inverse distance-4	-14.90	43.03	187.60	0.934	18.33
Inverse distance-5	-13.28	42.63	187.25	0.934	18.18
Inverse distance-6	-12.33	42.56	187.92	0.933	18.47

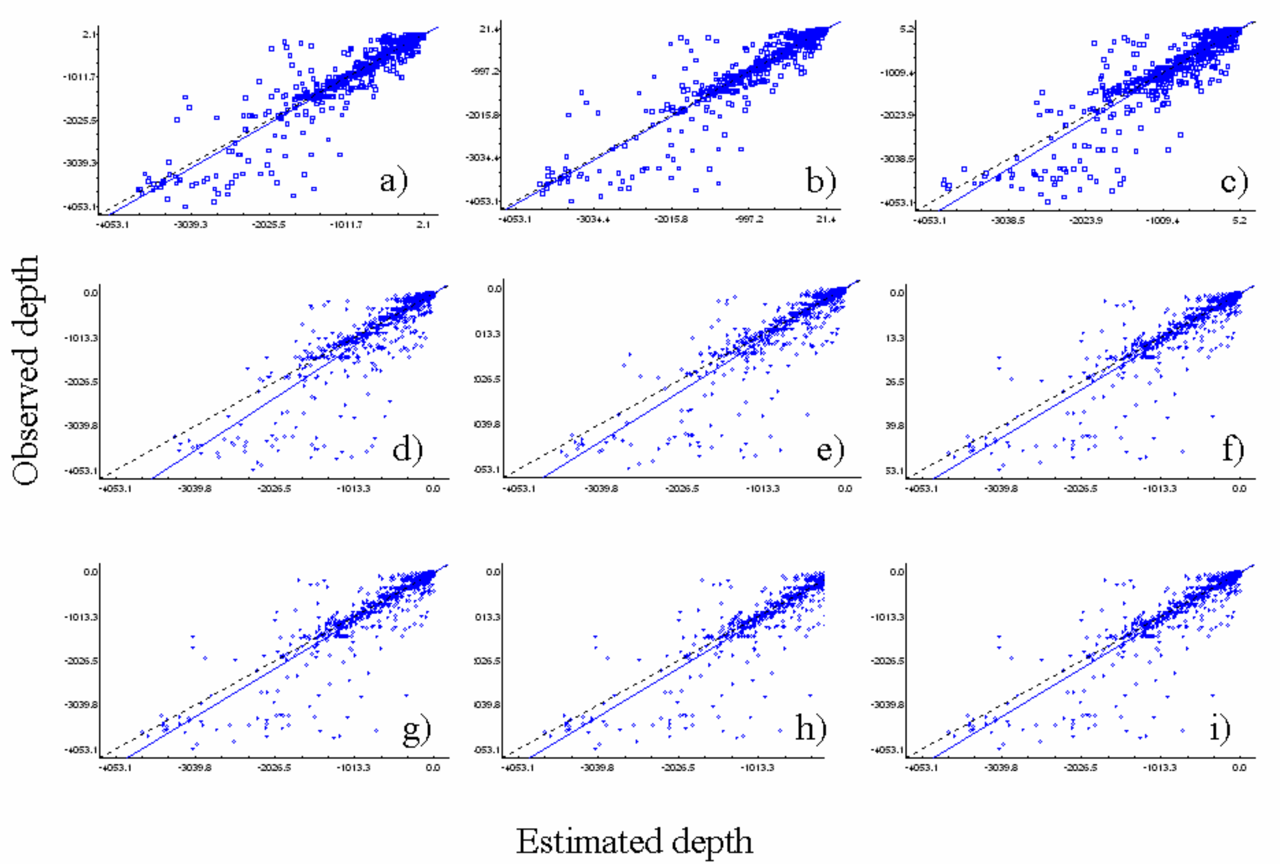

Figure 4. Results of cross validation analysis used to compare the interpolation methods: a) spherical kriging, b) exponential kriging, c) gaussian kriging, d) inverse distance-1, e) inverse distance-2, f) inverse distance-3, g) inverse distance-4, h) inverse distance-5, i) inverse distance-d (Dotted lines represent a perfect fit 1:1).

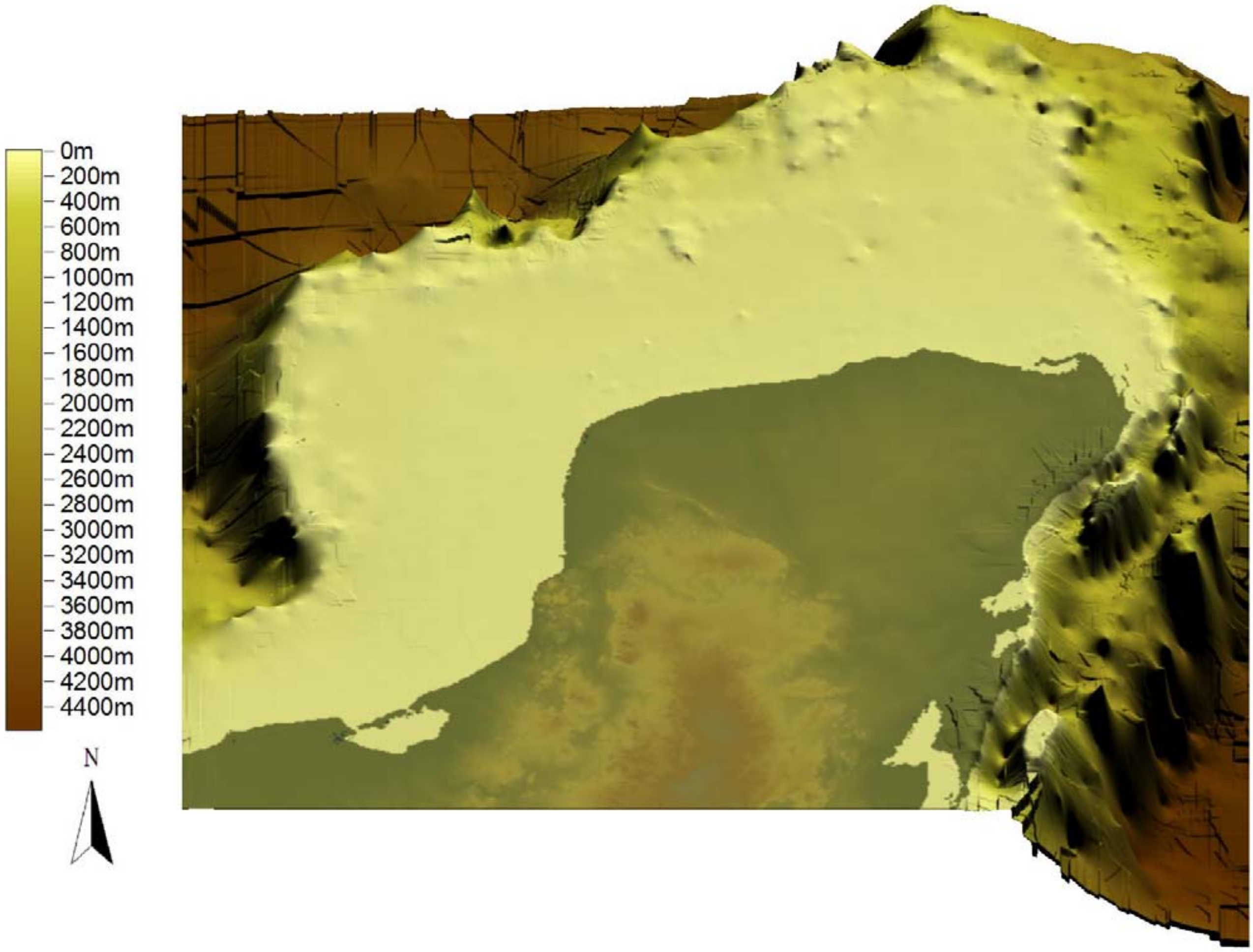

Figure 5. A 2.5-D illustration of the final digital bathymetric model (DBM) for the Yucatan submerged platform produced by using the spherical model with kriging.

Discussion

The results obtained from the comparison of the two interpolation methods analyzed in this study indicated that kriging was the most suitable method for mapping the spatial distribution of depth values at regional scale. The results also revealed that although the inverse distance method has the advantage of relative simplicity and ease of processing, this method is the least accurate, resulting with at least an increase of 18.2% in the error compared with kriging procedures. Other studies have reported similar results (Hernandez-Stefanoni & Ponce-Hernandez 2006, Nalder & Wein 1998, Voltz & Webster 1990), revealing that the estimation is improved when kriging is applied. However this improvement is given only if number of points is large enough to apply this technique and if a careful selection of the models of semi-variograms is undertaken (Kravchenko & Bullock 1999).

It is also important to notice that in addition to the better performance of kriging procedures, the semi-variogram analysis required for kriging interpolation provides interpretative values beyond its role in kriging estimation (Rossi *et al.* 1992). Such information is not produced and made available by using the inverse distance functions. For example, semi-variogram models were able to explain the nature, intensity and extent of the spatial distribution patterns of depth values. They also showed that such values are spatially structured from patches separated between 1.4 and 3.0 decimal degrees, which correspond to the "range of influence" parameter on the semi-variogram.

As a final remark and considering that in many developing countries the accuracy of bathymetric data in digital format are rare, our approach might constitute a suitable option not only to researchers in our region but to others trying to decide the type of interpolation technique and the model to choose when elaborating digital elevation models for regional mapping purposes.

References

Bowles, F., Vogt, P. & Jung, W. 1998. Bathymetry (Part I), sedimentary regimes (Part II), and abyssal water disposal potential near the conterminous United States. **Journal of Marine Systems**, 14: 211-239.

Burrough, P. A. & McDonnell, R. A. 1998. **Principles of Geographical Information Systems. Spatial Information Systems and Geostatistics.** Oxford University Press, New York, 333 p.

Carlon, D. B. 2002. Production and supply of larvae as determinants of zonation in a brooding tropical coral. **Journal of Experimental Marine Biology and Ecology,** 268: 33– 46.

Eastman, J. R. 1999. IDRISI 32: user's guide. **Clark labs for cartography, technology and geographic analysis.** Clark University, Massachusetts, USA.

Epifanio, C. E. & Garvine, R. W. 2001. Larval Transport on the Atlantic Continental Shelf of North America: a Review. **Estuarine, Coastal and Shelf Science**, 52: 51–77.

Hernández-Stefanoni, J. L & Ponce-Hernandez, R. 2006. Mapping the spatial variability of plant diversity in a tropical forest: Comparison of spatial interpolation methods. **Environmental Monitoring and Assessment,** 117: 307-334.

Isaaks, E. H. & Srivastava, R. M. 1989. **An introduction to applied geostatistics.** Oxford University press. New York, 561 p.

Klenke, M. & Schenke, H. W. 2002. A new bathymetric model for the central Fram Strait. **Marine Geophysical Researches**, 23: 367–378.

Kravchenco, A. & Bullock, D. G. 1999. A comparative study of interpolation methods for mapping soil properties. **Agronomy Journal**, 91: 393-400.

Merino, M. 1997. Upwelling on the Yucatan Shelf: hydrographic evidence. **Journal of Marine systems,** 13: 101-121.

Merwade, V. M., Maidment, D. R. & Goff, J. A. 2006. Anisotropic considerations while interpolating river channel bathymetry. **Journal of Hydrology**, 331: 731-741.

Nalder, I. A. & Wein, R. W. 1998. Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest. **Agriculture and Forest Meteorology**, 92: 211-225.

Ramírez-Herrera, M. & Urrutia-Fucugauchi, J. 1999. Morphotectonic zones along the coast of the Pacific continentalmargin, southern Mexico. **Geomorphology,** 28: 237–250.

Robertson, G. P. 1987. Geoestatistics in ecology: Interpolating with known variance. **Ecology**, 68(3): 744-748.

Robertson, G. P. 2000. **GS+: geostatistics for environmental science**. Gamma Design Software. Plainwell, Michigan.

Rossi, R. E., Mulla, D. J., Journel, A. G. & Franz E. H. 1992. Geostatiscal tools for modeling and interpreting ecological spatial dependence. **Ecological Monographs,** 62: 277-314.

Schmitter-Soto, J. J., Comun, F. A., Escobar-Briones, E., Herrera-Silveira J., Alcocer J., Suarez-Morales E., Elías-Gutierrez M., Díaz-Arce V., Marín L. E. & Steinich B. 2002. Hydrogeochemical and biological characteristics of cenotes in the Yucatan Peninsula (SE Mexico). **Hydrobiologia**, 467: 215–228.

Secretaria de Marina. 1994. **Navigational charts**. S.M 800.

Secretaria de Marina. 1995.) **Navigational charts**. S.M. 900.

Voltz, M. & Webster, R. 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. **Journal of Soil Science,** 41: 473-490.

Webster, R. & Oliver, M. A. 2001. **Geostatistics for environmental science.** John Wiley and Sons, LTD. Toronto, Canada, 271 p.

Wright, D. J., Bloomer, S. H., MacLeod, C. J., Taylor, B. & Goodlife, A. M. 2000. Bathymetry of the Tonga Trench and Forearc: a map series. **Marine Geophysical Researches,** 21: 489–511.

Zar, J. H. (1999). **Biostatistical Analysis.** Prenctice Hall, New Jersey, 663 p.

Received January 2007
Accepted May 2007
Published online September 2007