A Step Toward Accuracy – Based Vector-Data Prioritizing

Arie CROITORU and Prof. Yerach DOYTSHER, Israel

Key words: random process, geostatistics, spectral analysis, least-squares collocation.

ABSTRACT

The large variety of available spatial data sets poses new challenges and potential pitfalls for end-users as well as for data providers. End-users must now prioritize data sets according to their needs and resources. One such prioritizing scheme is based on assessing the uncertainty in a desired end product caused by the usage of a given data set, and comparing it to a predefined uncertainty (risk) threshold. Such a prioritizing scheme requires detailed knowledge of errors in the input spatial data, as well as errors introduced by the analysis process. Unfortunately, in many cases such information regarding the input spatial data is either missing or ambiguous. Consequently, end-users must verify and estimate the applicability of the data. This contribution describes various aspects of such estimation techniques, and examines their applicability for spatial data.

CONTACT

Mr. Arie Croitoru
Technion – Israel institute of technology
Faculty of Civil Engineering, Division of Geodetic Engineering
Technion City, Haifa 32000
ISRAEL
Tel. + 972 4 829 2663
Fax + 972 4 823 4757
E-mail: ariec@tx.technion.ac.il
Web site: www.technion.ac.il/~ariec

Prof. Yerach Doytsher
Technion – Israel institute of technology
Faculty of Civil Engineering, Division of Geodetic Engineering
Technion City, Haifa 32000
ISRAEL
Tel. + 972 4 829 3183
Fax + 972 4 823 4757
E-mail: doytsher@geodesy.technion.ac.il

A Step Toward Accuracy-Based Vector-Data Prioritizing

Arie CROITORU and Prof. Yerach DOYTSHER, Israel

1. INTRODUCTION

At the turn of the millennium, users seeking spatial data can now benefit from a large number of available data sets. This information wealth originated from the rapid technology developments in data capture, data processing, and information networking, and led to the construction of clearinghouses, from which users can extract ready-to-use data directly into their desktop GIS systems. An example of the large variety of information sources available can be illustrated through a simple search conducted in the FGDC geospatial data clearinghouse website. Looking for information on "Elevation and Derived Products" for the state of Israel, using the FGDC NSDI search wizard, (FGDC, 2002) revealed 19 possible data sets, while a similar search for the state of Washington, USA, indicated 34 possible data sets. Though some of these data sources may not be appropriate for certain applications, it is still likely that several of these data sets will be usable, leaving the user to prioritize them.

Although this diversity results in numerous advantages from the user's perspective, it also encompasses a responsibility. Unlike the situation where data is collected by an organization according to its particular specifications and for its exclusive usage, data is no longer "tailor-made" and direct communication between the data provider and the end user ("user") is no longer assured. This may result in data misuse, forcing the user to check the applicability of the data. Data providers may also be affected by this new situation, as there is no guaranty of proper usage of the data, and as the proficiency of the user is unknown. This may expose data providers to undesirable liability and other legal difficulties.

A key point in resolving these difficulties is the ability to provide additional descriptive information regarding the data set at hand. Such additional information may be delivered to the user by adding metadata to the data set, in which various aspects of the data (such as data quality information, spatial reference information, or temporal information) are detailed. Standards for metadata are already available (such as the FGDC-STD-001-1998 standard), and their incorporation into existing data sets is gaining acknowledgment. Yet for many applications the available metadata is not detailed enough, and its implications for a specific usage are unclear (for example, DTM based applications). Consequently, users may still face uncertainty regarding the applicability of the data, even where metadata are available.

An assessment of the "fitness for use" of data may assist in coping with prioritizing (Agumya and Hunter, 1999). Agumya and Hunter suggest that this assessment should be based on an estimation of the uncertainty in the end product, caused by errors and uncertainties in a given data set. This end-product uncertainty can then be used for assessing the total risk facing the user. Assuming an acceptable risk threshold is established, a decision regarding a given data set may be reached.

Although a risk-based analysis may provide an appropriate prioritizing framework, its

implementation is not straightforward. Assessing the end-product uncertainty requires a proper mathematical error propagation model, as well as a detailed description of the error behavior in the source data set. Without such information the risk of using a particular data set can not be estimated, data prioritizing cannot be carried out, and a decision on the fitness for use can not be made. Thus the error characteristics and the error propagation model are interdependent.

This emphasizes the need for the development of suitable error propagation models, as well as techniques for quantifying and describing the behavior of errors in the source data. This contribution addresses the problem of quantifying and describing the behavior of errors in a data set. It should be noted that in this context the term "error" refers to the positional accuracy of the data. According to Kyriakidis et al. (1999), it is assumed that two data types are available to the user:

- *Hard data* $\{X^H, Y^H\}$ – a data set of high accuracy and low volume (data points are sparse and spread throughout the interest area). It is assumed that this data can not be used for application purposes.
- *Soft data* $\{X_i^S, Y_i^S, i=1...n\}$ – a group of n data sets ($n \geq 1$) available to the user (via a clearinghouse, for example). These data sets are of unknown quality, but have adequate volume to be useful for application purposes.

Using these two data sets, a set of residuals may be computed by:

$$dx_i = X_i^S - X^H$$
$$dy_i = Y_i^S - Y^H \qquad (1)$$

It should be noted that it is assumed in Equation (1) that hard data and soft data share a considerable amount of common data features. Based on Equation (1), it is now necessary to extract the characteristics of each of the n soft data sets, which will then serve as a criterion in the prioritizing process. This requires the ability to model and quantify the errors, as will be detailed below.

2. SPATIAL UNCERTAINTY CALCULATIONS

Each of the soft data sets is a product of a variety of acquisition and compilation steps. These steps may include raw data capturing, computational data processing procedures, and various maintenance activities (for example: coordinate transformations, continuous updating, or data integration by conflation). Consequently, the accuracy characteristics of such a data set are no longer dominated by the primary acquisition method that was used, but by a combination of all processes that were applied to the data during its life cycle. Although these aspects of the data may be documented in the metadata (lineage), it is most likely that the effect of these various steps can not be estimated on the resulting accuracy of the data, and classical error propagation techniques can not be implemented. It is therefore necessary to devise other methods for describing the magnitude, relations, and the variability of errors ("*spatial uncertainty calculations*", (Bardossy and Fodor, 2001)).

2.1 Summary statistics

One of the well-known techniques of describing the population of the quantities obtained from Equation (1) is by using scalar quantities such as the mean, $(\overline{dx}, \overline{dy})$, and the variance (s_{dx}, s_{dy}):

$$\overline{dx_i} = \frac{1}{m_i}\sum_{j=1}^{m_i} dx_i(j) \qquad \overline{dy_i} = \frac{1}{m_i}\sum_{j=1}^{m_i} dy_i(j)$$

$$s_{dx_i} = \sqrt{\frac{\sum_{j=1}^{m_i}\left(dx_i(j) - \overline{dx_i}\right)^2}{m_i - 1}} \qquad s_{dy_i} = \sqrt{\frac{\sum_{j=1}^{m_i}\left(dy_i(j) - \overline{dy_i}\right)^2}{m_i - 1}} \tag{2}$$

The computation of these quantities, also known as *summary statistics*, may be accompanied by various statistical tests for significance assurance (for example, Barbato, 2000). Although the quantities described in Equation (2) are indeed a description of the population of Equation (1), they fall short of taking into account the spatial relations within the population. Furthermore, the computation of Equation (2) is statistically justified only when the values of Equation (1) are random and uncorrelated. Thus it can not be expected that summary statistics will be able to account for any correlations in the data, nor describe it (Ehlschlaeger and Goodchild (1994); (Kyriakidis et al., 1999). Consequently, a different framework for describing the population of Equation (1) should be employed.

2.2 Random process theory

The shortcomings of summary statistics necessitate an extension of the statistical framework for analyzing and description of errors. Such a framework can be formed by treating the errors as a *random process*. A random process X is an extension of the random variable space, implemented by introducing a function that depends on a given space t (for example, time or location) over the random variable space s (Peebles, 2001):

$$X = \{x(t,s), t_1 \le t \le t_2\} \tag{3}$$

This indicates that although the values of s are random, the relationships between these values are described by the function over t. Hence, the spatial relations between data elements are accounted for. As a result of this advantage, the random process framework was suggested by several authors for describing errors in spatial data (Goodchild at al. (1992); Ehlschlaeger and Goodchild (1994); Hunter and Goodchild (1996); and Church et al. (1998)). Since the space domain of spatial data is location, the term *random field* is frequently used to describe a random process over this space.

The characteristics of a random process are described by its stationarity and ergodicity. The interested reader may find further information in the extensive literature available on the subject of random processes (for example, Peebles (2001); Mikhail (1976)).

2.3 Describing the spatial behavior of a random field

The random process theory serves as a statistical framework, which must be accompanied by proper indices. These indices can no longer be single scalars as they must account for the spatial dimension of the random filed. Such indices are widely used in the field of Geostatistics.

Two primary indices, namely the *variogram* and the *correlelogram*, are used to characterize a random field. The (experimental) variogram (Equation (4a)) describes the variation of the variance between elements in the field, while the correlelogram (Equation (4b)) describes the correlation between data elements (Cressie, 1993):

$$\gamma(t) = \frac{1}{2N}\sum_{\Delta t \approx h}\left[x(t) - x(t+\Delta t)\right]^2 \quad \text{(a)}$$
$$R_x(t) = \frac{1}{N}\sum_{\Delta t \approx h}\left[\left(x(t) - \mu_x\right)\left(x(t+\Delta t) - \mu_x\right)\right] \quad \text{(b)} \qquad (4)$$

where μ_x is the variance of the data set, and N is the number of data pairs. For the discrete case, both indices are computed by dividing the space, (t), into equally spaced *lags*, (h), where for each lag an average value is taken. It should be noted that these indices assume an isotropic random scalar filed, while the field described by Equation (1) is a random vector field. An attempt to address this issue by modifying Equation (4) to account for direction may be found in Wingle and Poeter (1998) and Funk et al. (1999).

As with summary statistics, the significance of these indices should be reviewed. In the case of the variogram, the variance of each point is estimated by (Cressie, 1985):

$$\mathrm{var}\left[\gamma(t)\right] \approx \frac{2\left(\gamma(t)\right)^2}{N(h)} \qquad (5)$$

The variance of the corrlelogram can be estimated by (Jenkins and Watts, 1968); (Bennett, 1979); (Hearn and Metcalfe, 1995):

$$\mathrm{var}\left[R_x(t)\right] \approx \frac{1}{N} \qquad (6)$$

It should be noted that the variance estimate is constant throughout the correlelogram range. Equation (6) could be used for testing the significance of a given data set when its value is interpreted as the expected boundary limits of the correlations in white-noise. Consequently any correlation value beyond these boundaries indicates that the data set at hand differs from white-noise.

2.4 Describing the spectral behavior of a random field

A further description of the behavior of a random process could be attained by applying a *spectral analysis*, where a transformation of the data into the frequency domain is performed by a Fourier transform. In the case of time series sampled in constant intervals or in the case of spatial data series sampled as a regular grid pattern, the Fourier transform could be applied directly to the series (Bendat and Piersol, 1980). In other cases, where spatial data is sampled

randomly in space domain, it might be more convenient to use the Wiener-Khintchine relations, describing the relationship between the power spectrum of a process and its correlation function (Peebles, 2001):

$$F_x(\omega) = \int_{-\infty}^{+\infty} R_x(t) e^{-j\omega t} dt \quad \overrightarrow{\leftarrow} \quad R_x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_x(\omega) e^{j\omega t} d\omega \tag{7}$$

Where F_x is the Fourier transform of the correlation function that was derived from Equation (4b). Thus the power spectrum and the correlation function of a given random data set are a Fourier transform pair.

This type of spectral analysis can serve as an analysis tool for of the behavior of the correlations in a series as it reveals which frequencies contribute the most to the correlation function (Jenkins and Watts, 1968). The dominant frequencies can be detected using a *periodogram*, which shows the power of a given correlation frequency, and is computed by (Jenkins and Watts, 1968):

$$I(\omega_k) = \frac{N}{2} \left| F_x(\omega_k) \right|^2, \quad k = 1 \ldots n \tag{8}$$

The spectrum of a given data set should be tested for significance by a white-noise spectrum, using a normalized cumulative spectrum (Fuller, 1976):

$$C_k = \frac{\sum_{l=1}^{k} I_l(\omega_l)}{\sum_{m=1}^{n} I_m(\omega_m)} \tag{9}$$

where for white-noise data, Equation (9) should produce a diagonal straight line. The test is carried out using the Kolmogorov-Smirnov statistic, $S_{K\text{-}S}$, (Jenkins and Watts, 1968); (Fuller, 1976):

$$S_{K-S} = \frac{\lambda}{\sqrt{\frac{N}{2} - 1}} \tag{10}$$

where λ is 1.63, 1.36, and 1.02 for significance levels of 99%, 95%, and 75% respectively. Using this statistic, a confidence band, which represents the cumulative spectrum bounds of a white-noise, could be constructed. Deviations from this band indicate that the data at hand differs significantly from white-noise.

3. TESTING THE ERRORS FOR RANDOMNESS

Prior to the implementation of any signal-based technique, the significance of the errors must be reviewed against the null hypothesis of white-noise. In practical terms, summary statistics should be preferred for its simplicity, and the need to use of other techniques should be statistically justified. For this purpose, the correlation and the spectral characteristics of a data set can serve as complementary information regarding the nature of the errors. While the correlelogram reveals the similarity between adjacent data points, the periodogram checks for the existence and power of any periodic elements in the errors. Consequently, the periodogram can provide a better indication of the correlation length when the correlelogram and the periodogram are both significantly different from white-noise. In other cases, the

periodogram can provide information on the type of filtering that should be carried out prior to any correlelogram estimation, as periodic effects may create fluctuations in the corrlelogram (Čačko et al., 1988).

A more intricate situation arises when one of the indices does not prove to be significant. For example, a situation where the correlelogram is significant up to a certain correlation length, though the spectrum does not significantly differ from white-noise. Such situations may arise as local correlations could occur even in white-noise errors. Jenkins and Watts (1968) indicate that this may occur up to a certain extent "on the basis of chance" or due to the sampling technique used, and describe the correlelogram as an intermediate step that should be followed by a spectral analysis with an appropriate significance test.

An example of this situation is depicted in Figure 1 and 2. Figure 1 shows the analysis of a set of 400 grid points that were assigned errors in the form of white-noise (marked by the arrows in Figure 1a). From this grid a random set of 200 points were selected at random (Figure 1a), and their correlative and spectral characteristics were computed. Figure 1b shows a histogram of the number of pairs that participated in each lag distance. As can be seen from Figures 1c and 1d, which depict the correlelogram in x and y directions, that while most of the correlelogram points are not significant, about 15% of the correlelogram points are found to be significant. A spectral test of the errors (Figures 1e and 1f) reveals that the spectrum of the errors does not differ significantly from a white-noise spectrum.

In Figure 2 a similar analysis was carried out using real data. This data was collected from a photogrammetric model and was considered to be the soft data set, while a hard data set was collected in the field using a GPS RTK technique. Both data sets share approximately 260 points, as can be seen in Figure 2a. The errors were estimated using Equation (1) (marked by arrows in Figure 2a). A correlation analysis of this data (Figures 2c and 2d) reveals that more than 40% of the correlelogram points are significant, yet a spectral analysis of the data (Figures 2e and 2f) reveals that only the x direction spectrum is significant.

4. ESTIMATING THE SIGNAL

After establishing a signal and obtaining its spatial and spectral characteristics, estimation of the signal should be attained. The quantities obtained by (1) can not be regarded as estimation of the signal, since they are usually a combination of white-noise and non white-noise signals. This calls for a computation scheme that will enable the estimation of the signal and the noise components based on the correlation that was derived, as described above.

One such scheme is the Least-Squares Collocation (LSC) scheme. In this technique the relationship between a set of observations $l=\{dx_1,dx_2...dx_N\ ;\ dy_1,\ dy_2,...dy_N\}$ and a parametric model (A) with unknowns (x), is decomposed into a set of *signal* components (s) and a set of *random* components (n) (Moritz, 1972):

$$l = Ax + s + n \qquad (11)$$

A least-squares solution of (11) is obtained by (Moritz, 1972) ; (Cross, 1983):

$$\begin{aligned}\hat{x} &= \left(A^T\left(C_s + C_n\right)^{-1} A\right)^{-1} A^T\left(C_s + C_n\right)^{-1} l \\ \hat{s} &= C_s\left(C_s + C_n\right)^{-1}\left(l - A\hat{x}\right) \\ \hat{n} &= C_n\left(C_s + C_n\right)^{-1}\left(l - A\hat{x}\right)\end{aligned} \qquad (12)$$

where C_s is the variance-covariance matrix of the signal (a full matrix), which can be obtained from Equation (4b), and C_n is the variance-covariance matrix of the noise (a diagonal matrix) that can be estimated using Equation (4a). The estimates of the signal and the noise, as obtained from Equation (12), are also accompanied by a covariance matrix, from which their significance could be estimated (Cross, 1983):

$$\Sigma_{\hat{s}} = C_s^{\ T}\left(C_n + C_s\right)^{-1} C_s - C_s^{\ T}\left(C_n + C_s\right)^{-1} A\left(A^T\left(C_n + C_s\right)^{-1} A\right)^{-1} A^T\left(C_n + C_s\right)^{-1} C_s \qquad (13)$$

$$\Sigma_{\hat{n}} = C_n^{\ T}\left(C_n + C_s\right)^{-1} C_n - C_n^{\ T}\left(C_n + C_s\right)^{-1} A\left(A^T\left(C_n + C_s\right)^{-1} A\right)^{-1} A^T\left(C_n + C_s\right)^{-1} C_n \qquad (14)$$

Using Equation (13) and (14) the significance of the signal and noise could be tested. In addition to its estimation capabilities, the LSC scheme could also serve as a prediction tool, with which the signal and noise could be predicted in locations where no hard data was collected.

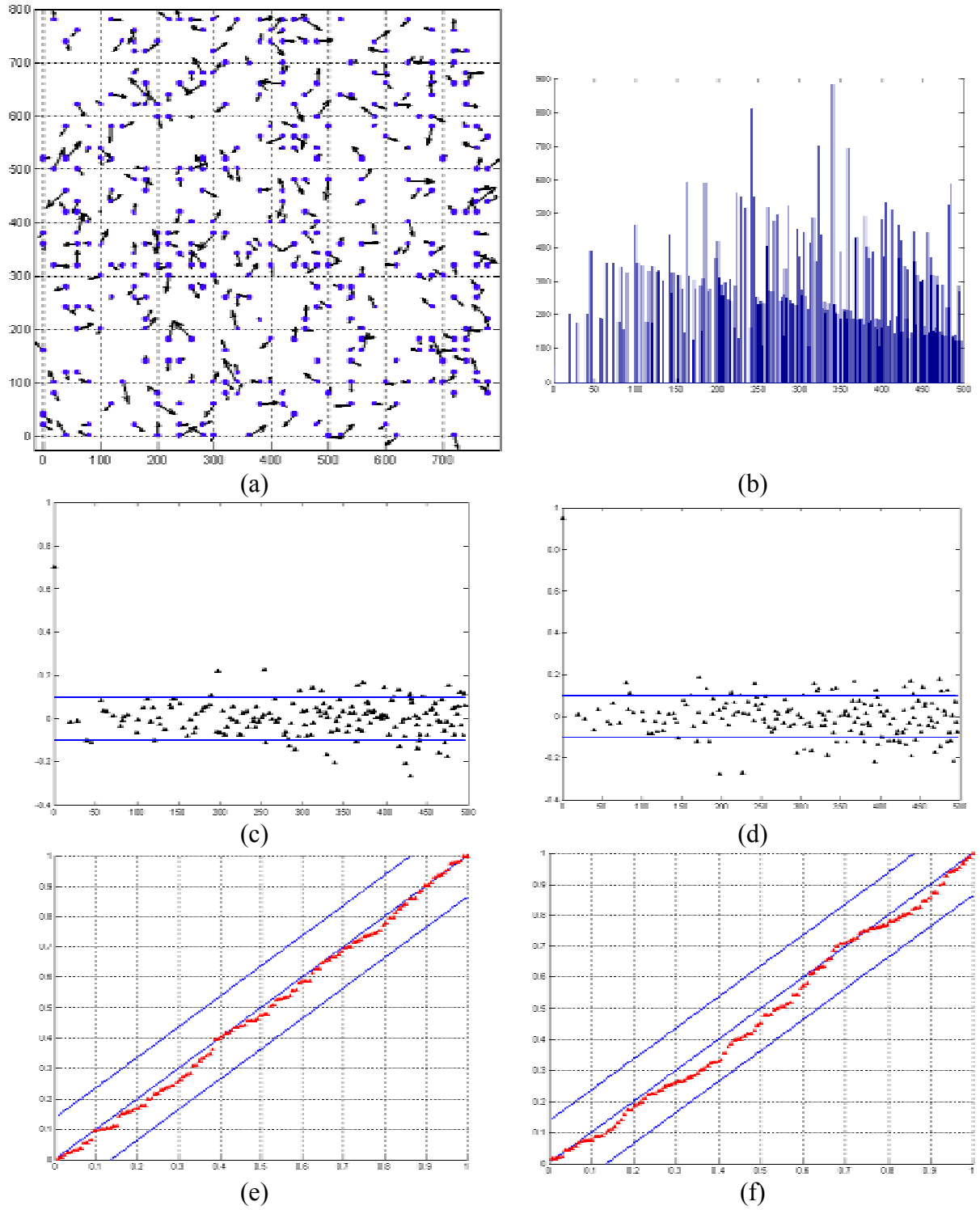

Figure 1: (a) the random grid data points; (b) histogram of the number of pairs in each lag distance; (c) *x* direction correlelogram; (d) *y* direction correlelogram; (e) spectral analysis test for *x* direction; (f) spectral analysis test for *y* direction

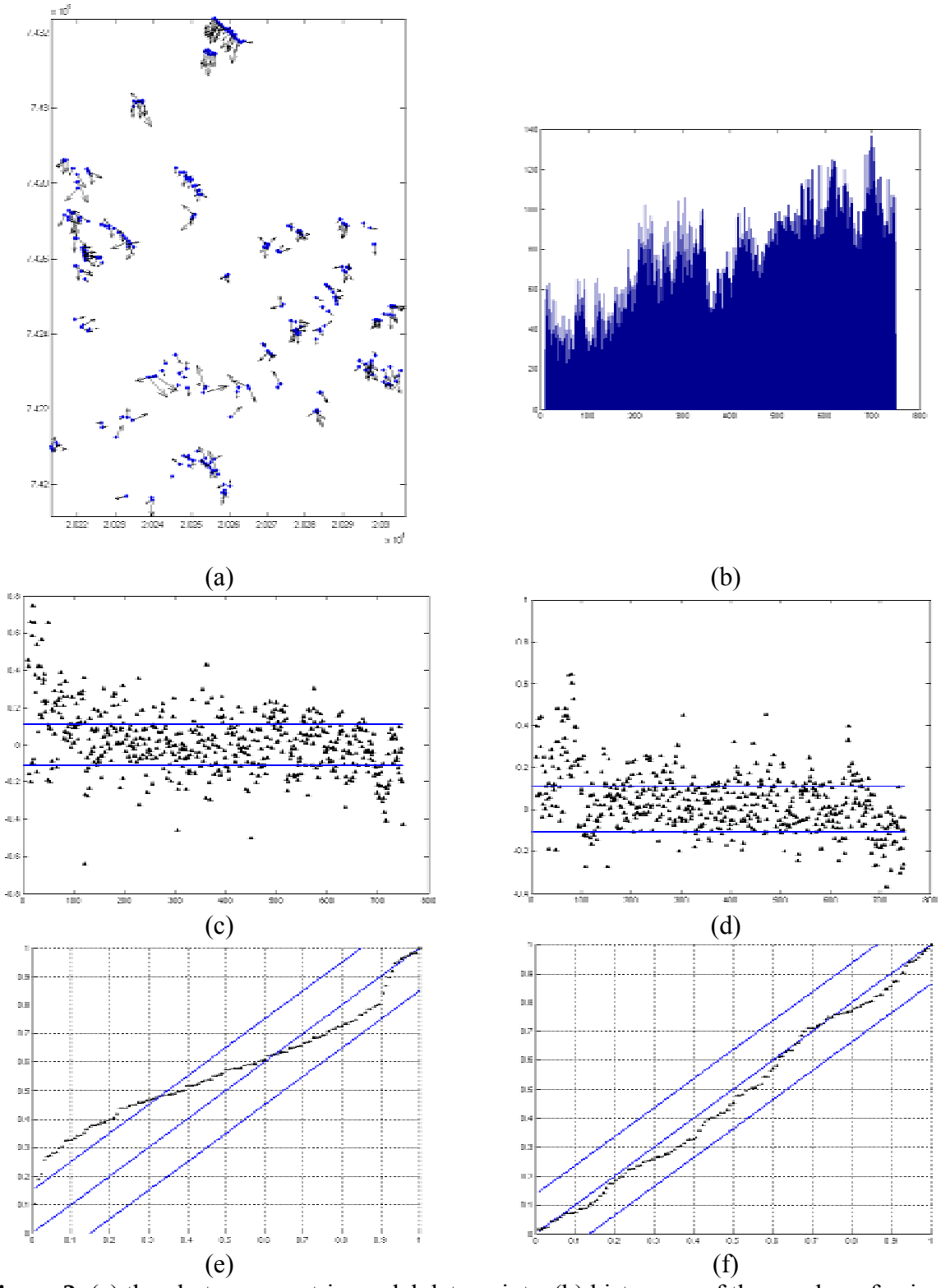

Figure 2: (a) the photogrammetric model data points; (b) histogram of the number of pairs in each lag distance; (c) *x* direction correlelogram; (d) *y* direction correlelogram; (e) spectral analysis test for *x* direction; (f) spectral analysis test for *y* direction

5. CONCLUSION AND FUTURE WORK

Geostatistical and spectral analysis provide a promising framework for analyzing errors in spatial data. Although geostatistical techniques became widely used for this purpose, less attention was given to the advantages of spectral analysis and the complimentary information it provides. Furthermore, a combination of these tools may assist in constructing a statistical significance test mechanism, which will justify any further signal-based analysis, such as the least-squares collocation scheme.

The implementation of these powerful analysis tools in the case of spatial data is not yet fully understood. Although geostatistics originated from spatial data, it is primarily designed to deal with single scalar values in space while errors in spatial data are usually vectors. Spectral techniques may be easily implemented on regularly sampled one or two-dimensional data (for example, a time series or an image), yet its applicability to random sets of spatial vector data is not straightforward.

Another issue that should be addressed in the context of error analysis is the sampling strategy. In the case of a continuous data set the sampling can be carried out in a regular pattern, but in the case of spatial data only random samples can be obtained. In order to assure an optimal description of the data and of the significance of any analysis made, a planning of the sampling should be carried out.

These challenges serve as the basis of ongoing research aimed at developing analysis tools for positional accuracy assessment of a given data set, based on a sparse higher quality data set.

REFERENCES

Agumaya A. and Hunter G.J., 1999. ”Assessing ‘Fitness for Use’ of Geographic Information: What Risk Are We Prepared to Accept in Our Decisions?”. ”Spatial Accuracy Assessment – Land Information Uncertainty in Natural Resources”, Lowell K. and Jaton A. (Editors), Ann Arbor Press, Chelsea Michigan. Pp. 35-43.

Barbato F.D., 2000. ”Accuracy Parameters Determination for GIS Base Map”. In the proceedings of ”Accuracy 2000”, the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences”, Amsterdam, July 2000. pp. 35-38.

Bardossy G. and Fodor J., 2001. ”Traditional and New Ways to Handle Uncertainty in Geology”. Natural Resources Research, Vol. 10, No. 3. pp. 179-187.

Bendat J.S. and Piersol A.G., 1980. “Engineering Applications of Correlation and Spectral Analysis”. John Wiley & Sons, A Wiley-Interscience Publication, New-York. 302 pages.

Bennett R.J., 1979. “Spatial Time Series, Analysis-Forcasting-Control”. Pion Limited, London. 674 pages.

Čačko J., Bílý M., and Bukoveczky J., 1988. “Random Processes: Measurement, Analysis and Simulation”. Elsevier, Fundamental Studies in Engineering, Vol. 8. 234 pages.

Church R., Curtin K., Fohl P., Funk C., and Goodchild M. F., 1998. “Positional Distortions in

Geographic Data Sets as a Barrier to Interoperation". ACSM Annual Convention, pp. 377-387.
Cressie N.A., 1985. "Fitting Variogram Models by Weighted Least Squares", Mathematical Geology, Vol. 17, No. 5. pp. 563-86.
Cressie N.A., 1993. "Statistics for Spatial Data (revised edition)". Wiley Series in Probability and Statistics, New-York. 900 pages.
Cross P.A., 1983. "Advanced Least-Squares Applied to Position Fixing". Working paper No. 6, Northeast London Polytechnic, Department of Land Surveying, 205 pages.
Ehlschlaeger C.R. and Goodchild M.F., 1994. "Uncertainty in Spatial Data: Defining, Visualizing, and Managing Data Errors". Proceedings of LIS/GIS '94 Annual Conference, pp. 246-253.
FGDC, 2002. http://130.11.52.184/FGDCgateway.html
Fuller W.A., 1976. "Introduction to Statistical Time Series". John-Wiley and Sons, New-York. 470 pages.
Funk C., Curtin M., Goodchild M.D. and Noronha V., 1999. "Formulation and Test of a Model of Positional Distortion Fields". "Spatial Accuracy Assessment – Land Information Uncertainty in Natural Resources", Ann Arbor Press. Pp. 131-144.
Goodchild M.F., Guoqing S. and Shiren Y., 1992. "Development and Test of an Error Model for Categorical Data". Int. J. Geographic Information Systems, Vol. 6 (2), pp. 87-104.
Hearn G.E. and Metcalfe A.V., 1995. " Spectral Analysis in Engineering – Concepts and Cases". Arnold, London, Great Britain. 307 pages.
Hunter G.J. and Goodchild M.F., 1996. "A New Model for Handling Vector Data Uncertainty in Geographic Information Systems". URISA Journal, Vol. 8 (1), pp. 51-57.
Jenkins G.M. and Watts D.G., 1998. " Spectral Analysis and its Applications". Holden-Day Inc., Oakland California. 525 pages.
Kyrakidis P.C., Shortridg A.M. and Goodchild M.F., 1999. " Geostatistics for Conflation and Accuracy Assessment of Digital Elevation Models". Int. J. of Geographical Information Science, Vol. 13, No. 7. pp. 677-707.
Mikhail E.M., 1976. "Observations and Least-Squares (with contributions by F. Ackermann)". IEP-Dun Donnelley, New-York, 497 pages.
Peebles P.Z., 2001. "Probability, Random Variables and Random Signal Processing, 4th edition". McGraw-Hill Series in Electrical Engineering and Computer Engineering. 462 pages.
Wingle W.L. and Poeter E., 1998. "Directional Semivariogram: Kriging Anisotropy without Anisotropy Factors". AAPG'98 Annual Convention "Advances in Geostatistics", Salt Lake City, Utah, May 17-20. pp. 1-4.

BIOGRAPHICAL NOTES

Arie Croitoru Received his B.Sc. (cum laude) from the Technion – Israel Institute of Technology, Division of Geodetic Engineering in 1992. In 1997 he received his M.Sc. form the Technion. Until 1999 he was engaged largely in photogrammetric and mapping projects. Since 1998 he is a Ph.D. candidate at the Technion, the Division of Geodetic Engineering under supervision of Prof. Yerach Doytsher. He is expected to graduate by the end of 2002.

Prof. Yerach Doytsher graduated from the Technion - Israel Institute of Technology in Civil Engineering in 1967. He received an M.Sc. (1972) and D.Sc. (1979) in Geodetic Engineering also from the Technion. Until 1995 he was involved in geodetic and mapping projects and consultation within the private and public sectors in Israel. Since 1996 he is a faculty staff member in Civil Engineering at the Technion, and has been head of the Geodetic Engineering Division within the Department, as well as head of the Geodesy and Mapping Research Center at the Technion.