Accounting for Discontinuities in Cadastral Data Accuracy: Toward a Patch-Based Approach

Arie CROITORU and Yerach DOYTSHER, Israel

Key words: Positional accuracy, Cadastral data, Random Field, LISA.

SUMMARY

Reliable cadastral data plays an essential role in a variety of activities, such as taxation, and property evaluation and registration. Within these activities positional accuracy plays a fundamental role in the manner by which the data is perceived and used. Users are no longer interested only in the data itself or its derivatives, but also in its reliability and accuracy. Thus, detailed reliable positional accuracy information becomes indispensable.

Consequently, several stochastic models have been recently suggested for modeling positional accuracy in vector cadastral data. One such stochastic modeling tool is the *random field model*. Random fields are well suited for modeling and simulating errors for continuous data, yet Cadastral vector data is not continuous and can not be treated stochastically as such.

This paper proposes an extension of the current random field model, a *patch-wise* random field, that will enable to account for discontinuities in the data. This extension includes a novel mechanism for identifying homogeneous regions within the data. The paper describes the proposed extension and evaluates evaluate its performance using real-world cadastral data.

Accounting for Discontinuities in Cadastral Data Accuracy: Toward a Patch-Based Approach

Arie CROITORU and Yerach DOYTSHER, Israel

1. INTRODUCTION

In many countries currently available cadastral data results from extensive data collection and compilation that was carried out throughout decades or centuries. Along this process the collected data was highly influenced by various factors, such as the available surveying tools and techniques, the processing capacity, or the quality assurance practice that was implemented. Cadastral maps, which often serve as a key data source for cadastral data, has also changed considerably in terms of the production techniques used to generate the maps, the materials used for map production, the scale used, or the conditions in which the maps are stored. Furthermore, the geodetic infrastructure on which the cadastral system is based has also changed considerably in terms of its availability and quality.

An example to the evolution of various factors that affect the characteristics of cadastral data can be found in the history of cadastral data in Israel (Stienberg, 2001). Land registration activities began in the late nineteen century, but its establishment using the registration of titles began only in the late nineteen twenties. During this period cadastral parcels were mapped using chain distance measurements, based on local densification traverses that were laid out on the available geodetic infrastructure. As electro-optical distance measurements became available in the late seventies, higher accuracy was achieved in cadastral surveys, resulting in regions ("patches") of high data accuracy. The overall accuracy of field data was also highly influenced by various regulations that controlled the characteristics of the local traverses: until 1987 no restriction on the hierarchy of the densification traverses was imposed and liberal surveying regulations (in terms of the required accuracy) were used. These regulations were modified in 1987, resulting in stricter surveying regulations and the implementation of a traverse hierarchy for densification purposes. In addition to these, a transition to the new Israeli Geodetic grid was recently implemented. This transition is likely to invoke an additional phase in the evolution of the characteristics of cadastral data in Israel.

As a consequence of these various factors currently available cadastral data is likely to be inhomogeneous in terms of its characteristics (Morgenstern et al., 1989). When examining a single parcel this lack of homogeneity is likely to be of little effect, yet when attempting to create a continuous cadastral database over more extensive areas, as in the case of the generation of a digital cadastral layer, this may result in inconsistent results unacceptable for cadastral purposes (Hebblethwaith, 1989). Consequently various operations are required in order to assure the consistency of the cadastral layer.

In order to overcome this inconsistency problem various methods to incorporate cadastral data, mainly from cadastral map sheets, into a homogeneous cadastral layer were suggested. The incorporation problem is commonly resolved by employing a variety of geometric transformations. These transformations are realized by mathematical models with various

degrees of freedom, ranging from a rigid-body transformation with three degrees of freedom up to an Affine transformation with six degrees of freedom, or a projective transformation with eight degrees of freedom (Fagan and Soehngen, 1987). Although polynomial based transformations with higher degrees of freedom may also be considered for this purpose, in practice they are not recommended due to their potentially erratic behavior.

The transformation process begins with the measurement of homologous points in both data sets. If redundant points are identified the transformation parameters may then be estimated using the Least Squares adjustment technique, during which weights may be assigned to each measurement (Greenfeld, 1997[a]), (Greenfeld, 1997[b]). In the case of control points, weights may be assigned by the rank of each point in the control network hierarchy (Greenfeld, 1997[b]). For non uniform homologous point distribution a modified least squares scheme is required in order to eliminate the effect of leverage points (Kampmann, 1996). Various constraints may also be incorporated in order to maintain the consistency of the existing data and possibly upgrade it (Fradkin and Doytsher, 2002).

Although a geometric transformation and constraints may bring both data sets into the same datum (thus eliminating the systematic effect), discrepancies between the overlapping area of the two data sets are still likely to occur. This difficulty is commonly encountered during the vectorization process of scanned cadastral map sheets (Doytsher and Gelbman, 1995), where each map sheet is treated separately by vectorizing the required data in the map, followed by a transformation (usually an Affine transformation) of the resulting vector data using the state-plane coordinate grid that was overlaid on the map sheet.

When several map sheets with overlapping boundaries are merged, discrepancies in overlapping areas still exist. This is caused by the inability of the Affine transformation to account for the random part of the discrepancies between the two data sets. Although proper averaging of the overlapping vector data may eliminate the discrepancies, it may also introduce distortions in the vector data, and by that cause a violation of the relationships between data elements (such as fixed length or angle, parallelism, perpendicularity, etc.) (Doytsher and Gelbman, 1995). In order to account for the resulting random distortions, a rubber-sheeting process can be employed. During this process the distortions are spread linearly toward the center of the map sheet, where linearity is assumed along the boundary of the map as well as perpendicular to it. A rubber-sheeting algorithm for non-rectangular map regions was also suggested by Doytsher (2000).

It should be noted that the underlying assumption in the rubber sheeting process is that both data sets are identical in terms of their accuracy characteristics. This assumption serves as the basis to averaging of coordinates and to the linear diffusion of the distortions. Yet, in many cases this assumption is not fully justified: averaging is not straightforward as it is not clear whether both data sets share the same accuracy characteristics, while weight assignment can not be carried out without explicit knowledge on the accuracy relations between the data sets. The same difficulty applies in the case of the linear distribution of distortions: this process can be justified by the assumption that the change in the distortions is linear throughout the data set. Yet due to surveying practices, map compilation techniques, and map sheet handling this may not be the case.

These various considerations, in addition to the heterorganic nature of the accuracy of cadastral data, indicate that a more comprehensive approach to the problem of cadastral data incorporation should address the accuracy relations of cadastral data sets and the spatial variation of accuracy throughout each data set. For this purpose a stochastic model should be employed.

2. THE STOCHASTIC MODEL

2.1 Positional accuracy

Prior to the introduction of the stochastic model an explicit definition of the term positional accuracy is required. Given a set of homologous points P in $\Re^n$ (n=2 in the case of a 2D cadastre or 3 for 3D cadastre), let us assume that each point is described by two sets of coordinates: a set $t=\{x_1, x_2,\ldots, x_n\}$ that represents the *estimated* coordinates, and a set $t'=\{x'_1, x'_2,\ldots, x'_n\}$ that represents the *true* coordinates. Based on this, the positional accuracy of each point is given by (Drummond, 1995), (Kyrakidis et al., 1999):

$$dt = t - t' = \{x_1 - x'_1, x_2 - x'_2, \ldots, x_n - x'_n\} \quad (1)$$

The terms "hard data" and "soft data" are usually used for t and t' respectively. As the true coordinates are usually unknown the "best" values available are commonly taken as the hard data set. This is usually carried out by employing a superior surveying technique that provides accuracies which are significantly higher than the accuracies obtained for the soft data. It should be noted that due to the high accuracy requirements the hard data is usually sparsely distributed and is not as dense as the soft data set. Once coordinates are measured, quantifying the positional accuracy for each point can be easily carried out using Equation (1), thus providing a set of positional error values

2.2 Positional accuracy modeling

In order to describe and summarize the characteristics of the positional accuracy of a given soft data set statistical descriptors are required. One of the well-known techniques for describing the behavior of these values is by using summary statistics (Goovaerts, 1997), such as the mean and the variance. Using summary statistics is well-known in mapping and surveying, and was also suggested for describing the positional accuracy of spatial data. An example to this approach may be found in (Barbato, 2000), in which various statistical significance tests were also suggested.

The main drawback of summary statistics is two fold. The use of summary statistics is statistically justified only when the values of Equation (1) are random and uncorrelated, yet for spatial data this may not necessarily be the case. In addition, the values obtained from Equation (1) are treated as simple scalar values without taking into account the spatial distribution of the data. As a result, it can not be expected that summary statistics will be able to account for any correlations in the data, nor describe it (Kyrakidis et al., 1999).

Although the precise definition of the term correlation was not yet introduced, the assumption that correlations are likely to be present in spatial data is based on the nature of the errors present in the various surveying practices used for collecting spatial data. Although the

measurements themselves may not be correlated in these practices (in fact special care is taken in order to avoid dependent or correlated measurements), the errors in these measurements are likely to be correlated since the same surveying equipment and practice is used. Thus the source of correlations is the correlated errors present in surveying operations. An example to these inherent correlations can be found in the compilation of a photogrammetric stereo pair. As lens and film distortions usually exhibit some degree of spatial patterning errors, adjacent points are likely to bare the same lens and film distortions. This similarity introduces correlations into the errors.

The shortcomings of summary statistics necessitate an alternative stochastic framework that would enable accounting for positional accuracy as a spatial phenomenon that has inherent correlations. One such framework, which serves as a fundamental Geostatistical modeling tool, would be treating positional accuracy as a set of random variables that have some spatial dependencies (Goovaerts, 1997). Based on this model, the random variable $z(t)_i$ that represents positional accuracy at point *i* would be:

$$z(t)_i = \{x_1 - x_1', x_2 - x_2', \ldots, x_n - x_n'\}_i = \{s_1, s_2, \ldots, s_n\}_i \quad (2)$$

Where $x_1, x_2, \ldots x_n$ are the Euclidian coordinates of point t_i It should be noted that in this case the random variable is n-dimensional (Yaglom, 1986). If the components of this vector are considered as independent of each other then at each location *t*, a set of random variables can be defined:

$$\begin{aligned} z_1(t) &= x_1 - x_1' \\ z_2(t) &= x_2 - x_2' \\ &\vdots \\ z_n(t) &= x_n - x_n' \end{aligned} \quad (3)$$

A set of dependent random variables constitutes a random function (RF) if for each location *t* the value of the random variable is known. Thus, for a predefined bounded area *D*, the set:

$$\{z(t) : \forall t \in D \subset \mathbb{R}^n\} \quad (4)$$

constitutes a random function. A random function defined over a two or three dimensional domain is commonly termed a random field.

A necessary condition that should be fulfilled for ensuring the statistical stability of the random function, is the existence of the probability distribution functions *P* for each element *t* in the set {Yaglom_72}:

$$\begin{aligned} F(x_1) &= P(z(t_1) < x_1) \\ F(x_1, x_2) &= P(z(t_1) < x_1, z(t_2) < x_2) \\ &\vdots \\ F(x_1, x_2, \ldots x_n) &= P(z(t_1) < x_1, z(t_2) < x_2, \ldots, z(t_n) < x_n) \end{aligned} \quad (5)$$

When applied to real-world data, the question whether the random field at hand is stable and has the same characteristics over the whole domain *D* should be addressed. The stability of the process is of importance here, as in the case of an unstable random field the characteristics should be a function of the location in *D*. This stability is termed stationarity for random functions. In terms of the distribution function, a random process is referred to as

stationary if all the distribution functions of type (5) remain the same when the random variable set used is shifted by h in D (Yaglom, 1962):

$$P\left(z(t_1) < x_1, z(t_2) < x_2, \ldots, z(t_n) < x_n\right) = \\ = P\left(z(t_1 + h) < x_1, z(t_2 + h) < x_2, \ldots, z(t_n + h) < x_n\right) \quad (6)$$

In practice, this implies that the statistical characteristics of the random process do not change throughout D. Stationarity of the type described in Equation (6) is defined as strict stationarity.

Theoretically, in order to further describe a stationary random field it is necessary to know all the distribution functions of the type (5) (Yaglom, 1962). In practice these are never known and their empirical retrieval trough experiments are not practical. Consequently moments are used for describing a random field. The two basic moments can be defined, namely the mean m of the whole set:

$$m = E\left[z(t)\right] \quad (7)$$

and the correlation C between two locations t_i and t_j:

$$C\left(t_i, t_j\right) = E\left[z(t_i)\, z(t_j)\right] \quad (8)$$

It should be emphasized that in this definition the correlation function may be dependent on the separation (distance) between the two locations t_i and t_j, as well as on the direction between these two locations. A random field, for which the mean is constant and the correlation function depends only on the separation, is termed a second order stationary field (Chiles and Delfiner, 1999) (The term wide sense stationary is also used, for example (Yaglom, 1962). Furthermore, a field, for which the correlation function is rotation invariant, is termed an isotropic random field.

The ability of the random field model to handle positional accuracy as a spatial random phenomenon, while incorporating the influence of correlations, makes the random field the preferable stochastic model for spatial data. Due to its ability to account for a continuous domain D the random field model was successfully implemented for continuous data sets, such as DEMs and categorical data (for example geological or soil maps). An example to the implementation of this approach can be found in (Goodchild et al., 1992), where an estimation of the uncertainty between two categorical regions was obtained using error simulations based on random fields. An example of the application of random fields for DEM error modeling can be found in the work of (Ehlschlager, 1998), who suggested estimating errors by random error field simulations. It should be noted that in these examples the random field model is used as an error realization mechanism, with which simulations are carried out. These simulations are then used for mapping areas likely to suffer considerable errors.

2.3 Geostatistical descriptors

Two primary geostatistical descriptors, namely the *variogram* and the *covariogram*, are used to characterize a random field. The (experimental) variogram (Equation (10a)) describes the variation of the variance between elements in the field, while the covariogram (Equation (10b)) describes the correlation between data elements (Cressie, 1993):

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} \left[z(t_i) - z(t_j) \right]^2 \tag{10a}$$

$$\hat{C}(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left[\left(z(t_i) - \overline{z} \right) \left(z(t_j) - \overline{z} \right) \right] \tag{10b}$$

where:

$$\overline{z} = \frac{1}{n} \sum_{i=1}^{n} z(t_i) \tag{11}$$

$|N(h)|$ is the number of data pairs (t_i,t_j) that are h units apart:

$$\left\{ (t_i, t_j) : \| t_i - t_j \| = h \; ; \quad i = 1 \ldots n \; ; \quad j = 1 \ldots n \right\}, \tag{12}$$

n is the size of the data set, and $\|.\|$ is the Euclidian distance operator. Both indices are computed by dividing all possible distances within D into equally spaced *lags* h, where for each lag an average value is taken. It should be noted that these indices assume a homogenous and isotropic random scalar filed.

2.4 Detecting inhomogeneous regions

Unlike continuous data, there is little experience in applying the random field model for vector data. Early attempts to use this model for continuous data (such as DEMs and categorical maps) were not directly linked to the random field theory, and instead "error grids" were suggested (Hunter and Goodchild, 1996). In their approach, a displacement grid that represents possible shifts of the vector data is generated and applied to the data for error assessment. Kiivery (1997) suggested using a rubber-sheeting like approach, where a transformation of an erroneous map into a "true" map is defined by a linear combination of a set of basis functions e and f (trigonometric functions, similar to Fourier series):

$$\begin{aligned} T_1(t) &= x_1 + \alpha^T e \quad , \quad \alpha \sim N\left(0, \sigma_\alpha^2 I\right) \\ T_2(t) &= x_2 + \beta^T f \quad , \quad \beta \sim N\left(0, \sigma_\beta^2 I\right) \end{aligned} \tag{13}$$

Where σ^2_α and σ^2_β are the variances of the displacement from the given map to the "true" map. Kiivery (1997) also suggested an estimation method for σ^2_α and σ^2_β based on ground truth data. Correlations can be introduced to this model by using a full matrix instead of I in Equation (9).

In contrast, Church et al. (1998) indicate that the model defined in Equation (9) requires that the error field would be smooth, while the errors in vector spatial data may not necessarily be smooth. Consequently, Funk et al. (1999) suggested classifying smooth sub-regions in the data, using directional information as a classifying criterion. For this purpose a modified variogram that is based on direction was suggested:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} \left(1 - \| r_{ij} \| \right) \tag{13}$$

where:

$$r_{ij} = \left\{ \frac{1}{2}\left(\frac{(s_1)_i}{\|s_i\|} + \frac{(s_1)_j}{\|s_j\|} \right), \frac{1}{2}\left(\frac{(s_2)_i}{\|s_i\|} + \frac{(s_2)_j}{\|s_j\|} \right) \right\} \tag{14}$$

It can be easily seen that r_{ij} is the average cosine and sine components of each point pair. Based on the above variogram, the dependence between each point and its neighborhood could be assessed, using a Ratio of Aerial Dependence (RAD) (Funk et al., 1999):

$$RAD_i = 1 - \frac{V_i}{M_i} \tag{15}$$

where:

$$V_i = \frac{1}{n_r}\left[\sum_{j=1}^{N(r)} \left(1 - \|r_{ij}\|\right) \right] \quad , \quad M_i = \frac{1}{n_r}\left[2\sum_{j=1}^{N(r)} \hat{\gamma}(h) \right] \tag{16}$$

$N(r)$ and n_r are the indices and the number of the points within a predefined range from a point i for which the computation is carried out.

3. AN ALTERNATIVE CLUSTERING SCHEME

3.1 The clustering criteria

In contrast to the clustering approach described above, the approach suggested here is based on the well known classical variogram definition, as it was introduced in Equation (10a). The basis to this approach is an attempt to point out areas which exhibit similar correlation characteristics, and in contrast areas which exhibit irregular correlation characteristics. The basis to this approach is the global indication to spatial association, namely Moran's I indicator (Cliff and Ord, 1973):

$$I = \frac{n}{\sum_{i,j} w_{ij}} \frac{\sum_i \sum_j w_{ij} z(t_i) z(t_j)}{\sum_i z(t_i)^2} \tag{17}$$

Based on Moran's I (Anselin, 1995) suggested using a Local Indication of Spatial Association (LISA) measure I_i:

$$I_i = \frac{n}{\sum_{j=1}^{n} z(t_j)^2} z(t_i) \sum_{j=1}^{n} w_{ij} z(t_j) \tag{18}$$

where w_{ij} is a weight reflecting the relation between location i and location j.

The main advantage of this approach is its ability to evaluate the local association I_i between each data point and its immediate neighborhood. High LISA values indicate pockets of nonstationarity in the data, or extensive differences in the accuracy characteristics of neighboring points, while low LISA values indicate that stationarity can be assumed for the region or that neighboring points exhibit similar accuracy characteristics. It is therefore possible to assess whether an given sub region in the data set could be considered as stationary, or whether it should be isolated from the data set, and regarded as a separate *patch*

due to its statistical dissimilarity. The classification criterion is therefore based on the statistical characteristics of the data. It should be noted that the local nonstationary pockets identified by LISA can result not only from the inhomogeneous nature of the data. Various other factors, such as surveying errors or data processing errors may also inflict local spatial association instabilities. Thus, data analysis with LISA could also serve as a gross error detection scheme in situations where the local instabilities do not exhibit any spatial pattern. For this purpose local Moran's *I* scatter plots could be used (Anselin, 1997).

As LISA indicates the relations between a data point and its "neighborhood", a clearer definition of the neighborhood and of the way by which the weights are assigned should be given. In some cases where geographical boundaries are addressed (as in the case presented by (Anselin, 1995)) the neighborhood is clearly identified. Yet in the case of vector data (as in the case of cadastral data) data points are considered, therefore no definite neighborhood boundaries are available a priori.

In order to accommodate this requirement for a neighborhood and weights definition it is possible to use the covariogram (Equation (10a)) as a global approximation of spatial association. It should be emphasized that as the computation of the covariogram is carried out prior to the detection of nonstationary patches and as the computation of the covariogram is based on the stationarity assumption, it is likely not to be a reliable estimation of the global correlation. Therefore, the covariogram is used only as an initial approximation. Once nonstationary patches are detected, the covariogram could be re-estimated for each of the stationary patches and could be considered as reliable.

3.2 Covariogram estimation

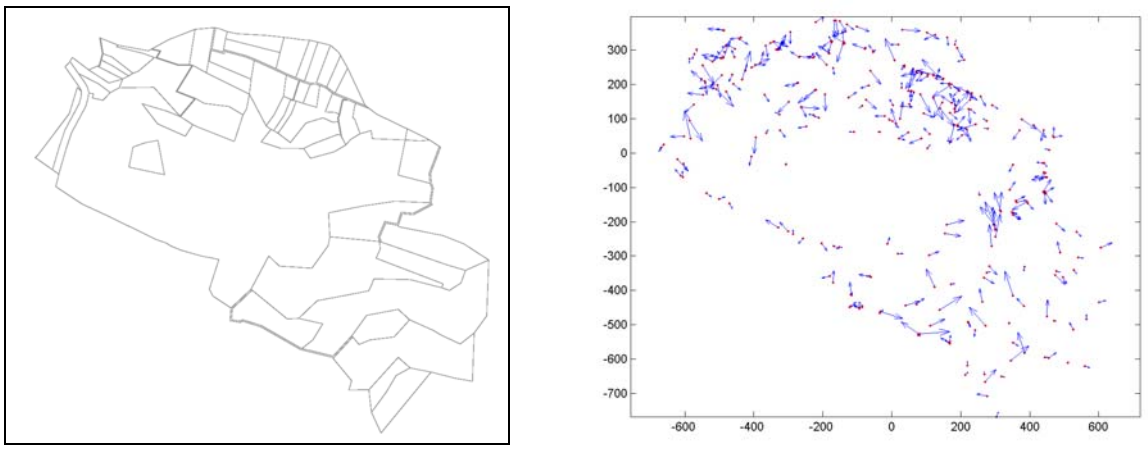
The estimation of the covariogram is based on a non parametric approach that was recently suggested by Croitoru and Doytsher (2003). A non parametric covariogram generation provides the ability to avoid several ambiguities that are commonly associated with covariogram generation, such as the ambiguity in the covariogram cloud generation, ambiguity in the model fitting process, and ambiguity in the selection of the fitting process itself.

Non parametric covariogram estimation was first exploited by (Hall et al., 1994), who suggested a non-parametric approach to the covariance estimation problem. Their approach consisted of three processing steps: in the first step a non parametric covariance sequence, *C'*, is estimated from the raw covariogrm bin set, *C*, using a kernel function (for example, a smoothing kernel with a bandwidth *h*). In the second processing step a non negative spectrum of *C*' is derived. This is done by computing the Fourier transform of *C'*, followed by an elimination of negative spectral components. In the final step, the non-parametric covariance function is derived by an inverse Fourier transform of the non-negative spectrum. A similar approach was suggested later by Yao and Journel (1998), who applied a moving average smoothing kernel, and Bjornstad and Falck (2001), who utilized a b-spline kernel function.
The scheme proposed by Croitoru and Doytsher (2003) is an extension of the non-parametric approach suggested by (Yao and Journel, 1998). As the covariance function is a real function, complex components should not be present when the inverse Fourier transform is applied.

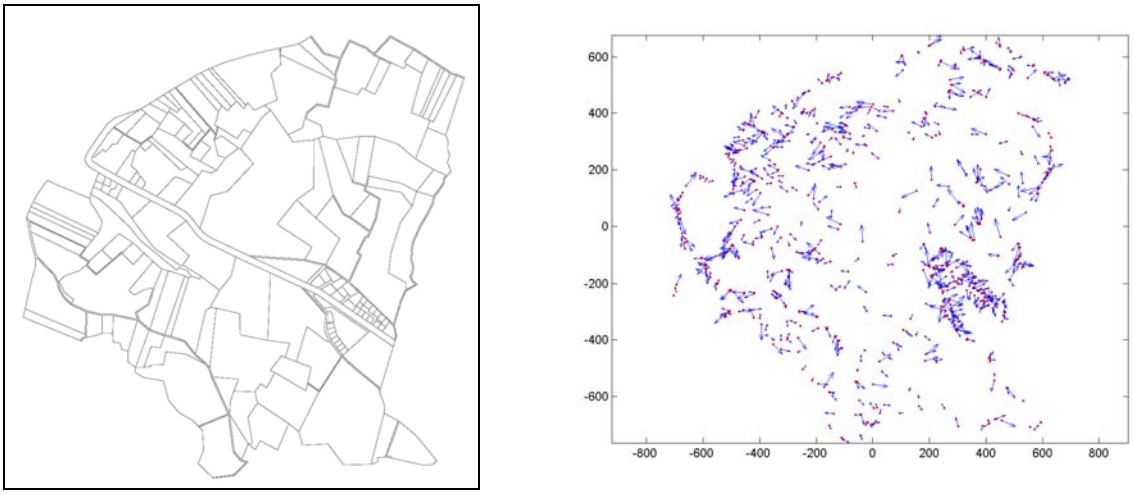
(Yao and Journel, 1998) avoided this by constraining the spectrum to symmetry around the zero frequency. In order to avoid this type of constrains and to assure that a real and even covariance function is obtained, a Discrete Cosine Transform (DCT) is used instead of the Fourier transform. This facilitates an easier application of the non parametric approach for covariogram generation.

4. PRELIMINARY RESULTS

In order to perform an initial assessment of the proposed clustering scheme two real-world cadastral blocks ("block 2" and "block 9") were used. Each block consisted of a set of parcels with various shapes and sizes. Both data sets included a set of digitized parcel border coordinates (soft data) and the corresponding set of coordinates that were computed using the original data that was recorded during the cadastral survey. Each of these blocks was also adjusted using various constraints, such as known lengths or topologic relations. The resulting adjusted coordinates were then used as the hard data set. In order to find correspondence between the soft data and the hard data in each block a point matching algorithm was applied (Valdes et al., 1995). This resulted in 306 matched points out of the 340 available points in "block 2" and 689 matched points out of the available 853 points in "block 9". Consequently, each block contained a list of corresponding points in the hard and the soft data sets. This correspondence was used to compute the positional accuracy of the digitized map sheets. The results obtained from this preliminary processing are depicted In **Fig. 1.**



(a) (b)

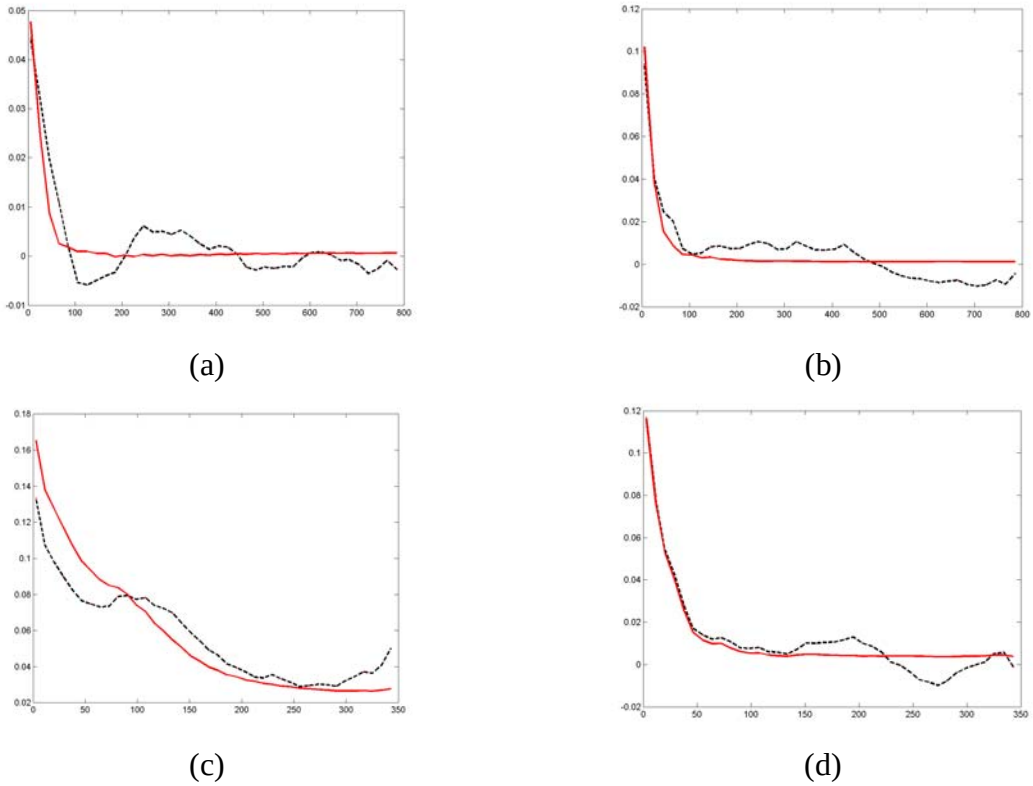


(c) (d)

Fig. 1: The two blocks processed: (a) “block 2”; (b) the derived positional accuracy (depicted as blue arrows) for the matched points in “block 2”; (c) “block 9”; (d) the derived positional accuracy (depicted as blue arrows) for the matched points in “block 9”.

For each of these blocks the proposed clustering approach was implemented, where the discrepancies between the soft and the hard data sets were processed in the east and north direction separately. This process began with the extraction of the raw covariogram for each block, from which the non parametric covariogram was estimated in east and north directions (Fig. 2(a) - 2(b) and 2(c) - 2(d) for “block 2” and “block 9” respectively). Based on these results, the LISA indicator was computed for each point in these blocks. For this purpose the estimated non-parametric covariograms that were earlier computed were used for defining the immediate neighborhood of each data point and for weight assignment. The results obtained for “block 2” and “block 9” are depicted in Fig. 3 and Fig. 4 respectively.

As can be seen from these results, “block 2” exhibits stationarity throughout its area (low LISA values) except for several small patches. A closer inspection of these patches reveals that they are inflicted by a single point, which is likely to be an outlier. In contrast, “Block 9” exhibits several well distinguished nonstationary patches, especially in the East direction. An additional inspection of these patches indicate that they are not inflicted by a single point, but by groups of points, and are therefore not likely to be generated by outliers. Consequently, further processing of this block should be carried out only after the isolation of these patches.



(a) (b)

(c) (d)

Fig. 2: The extracted non parametric covariogram - the raw covariogram (black) and the estimated non-parametric covariogram (red) for "block 2" and "block 9": (a) processing results in east direction for "block 2"; (b) processing results in north direction for "block 2"; (c) processing results in east direction for "block 9"; (d) processing results in north direction for "block 9".

5. CONCLUSION AND FUTURE WORK

This paper addressed the problem of detecting non stationary patches in cadastral data. Although stationarity is usually assumed when processing vector data, this assumption is not fully justified in many cases. As the cadastral infrastructure evolves with time its accuracy characteristics may change considerably. Consequently, a given cadastral block may contain data from various periods with a varying accuracy. When processing such data, assumptions, such as linearity and homogeneity, should be verified, and inhomogeneous regions should be isolated. It is therefore required to provide adequate tools that will enable the detection and extraction of inhomogeneous patches. For this purpose the LISA indicator was explored as a classifying criteria. Combined, with a non-parametric covariogram estimation scheme, which is used as an initial neighborhood estimator, the LISA indicator was applied to cadastral data. Using this scheme, patches of nonstationarity were successfully detected.

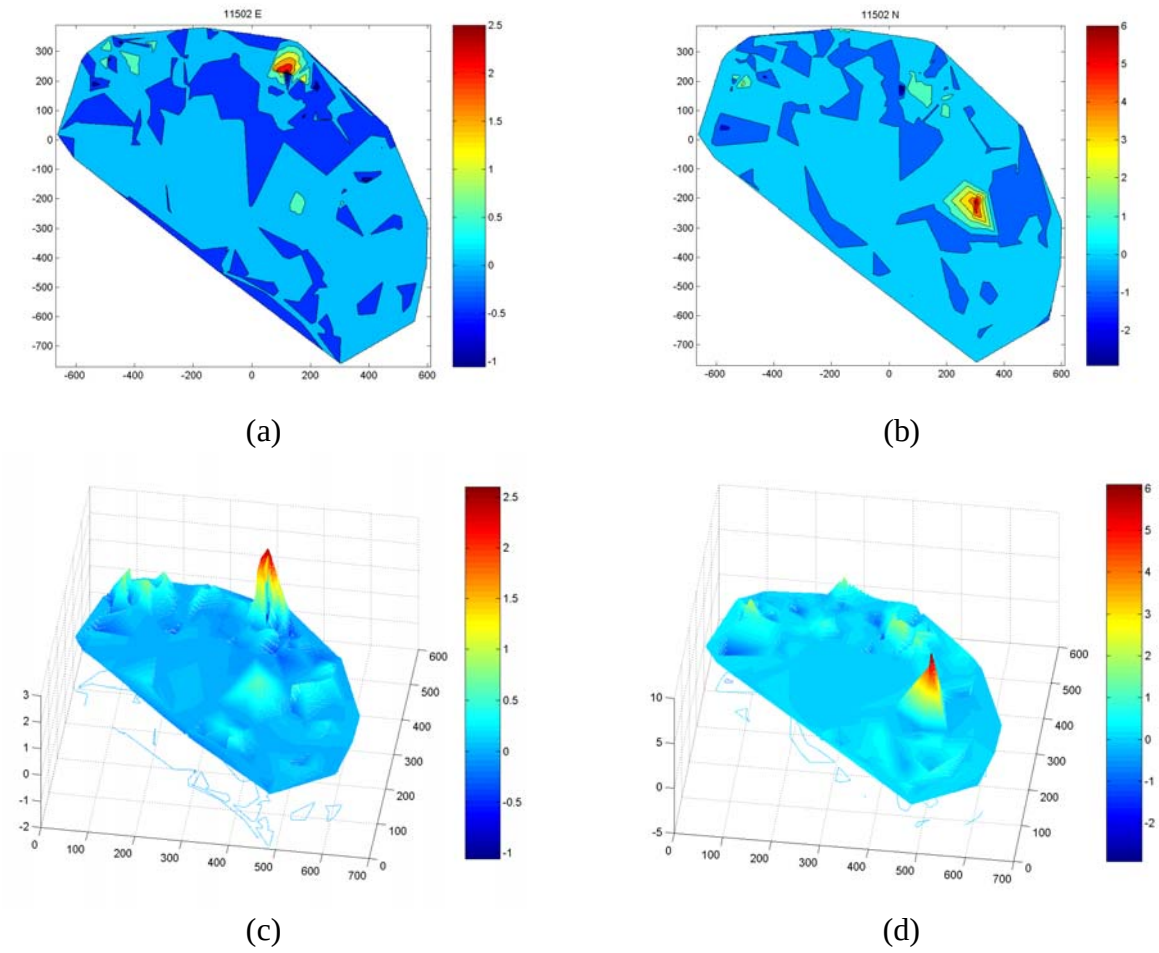


Fig. 3: Detection of nonstationary patches in "block 2": (a) detection results in east direction; (b) detection results in north direction; (c) a 3D view of the detection results in east direction; (d) a 3D view of the detection results in north direction.

In light of the results obtained, further research is still required. The ability to analyze the accuracy of spatial data and identify non stationary regions without the requirement for a decomposition of the data to two orthogonal directions (East/North) should be explored. It is therefore necessary to explore the application of *vector* random fields for this purpose. Furthermore, tools for clustering such vector fields are also required. It should be noted that although the application of the random field model is gaining recognition as a stochastic framework for vector data accuracy, the problem of detecting nonstationary patches in vector data sets was not fully addressed. As the proposed approach can be applied to a variety of vector data sources, its performance on various typed of vector data should be further explored.

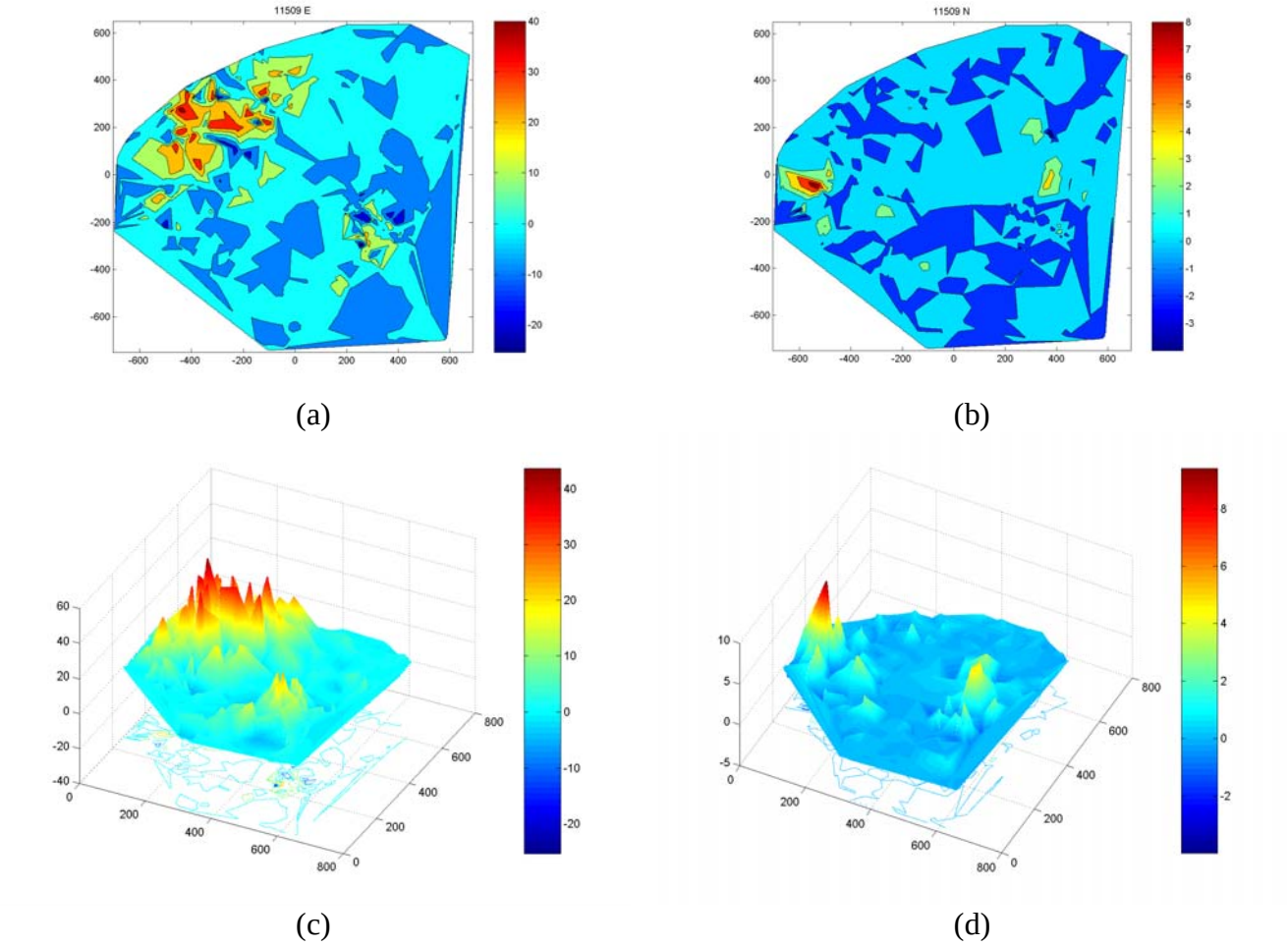


Fig. 4: Detection of nonstationary patches in "block 9": (a) detection results in east direction; (b) detection results in north direction; (c) a 3D view of the detection results in east direction; (d) a 3D view of the detection results in north direction.

ACKNLOWLEDGEMENTS

The authors would like to thank Mr. Eitan Gelbman for providing the cadastral data sets that were used in this paper. The authors would also like to acknowledge the kind support of the Survey of Israel and of the Ministry of Science.

REFERENCES

Anselin L., 1995. Local Indicators of spatial association – LISA. *Geographical Analysis* 27(2): 93-115.

Anselin L., 1997. The Moran scatter-plot as an ESDA tool to assess local instability in spatial association. In Spatial Analytical perspective on GIS, ed. Manfred Fischer, Henk Scholten, and David Unwin. Taylor and Francis, London.

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.
Bjornstad O. N. Falck W., 2001. Nonparametric spatial covariance functions: estimation and testing. *Environmental and Ecological Statistics*, 8(2001): 53-70.
Chiles J. P., Delfiner P., 1999. Geostatistics – modeling spatial uncertainty. John Wiley and Sons, New-York. 695 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.
Cliff A. D., Ord J. K., 1973. Spatial autocorrelation. Pion Limited, London. 178 pages.
Cressie N.A., 1993. Statistics for spatial data (revised edition). Wiley Series in Probability and Statistics, New-York. 900 pages.
Croitoru A., Doytsher Y., 2003. Improved non parametric covariance function for spatial data error interpolation. Submitted for publication.
Doytsher, Y. 2000. A rubber sheeting algorithm for non-rectangular maps . *Computers & Geosciences*, 26 (2000), pp. 1001-1010.
Doytsher, Y., Gelbman, E., 1995. Rubber sheeting algorithm for cadastral maps. *Journal of Surveying Engineering*, November 1995, pp. 155-162.
Drummond J., 1995. Elements of spatial data quality. Elsvier Science, New-York, pp. 31-58.
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.
Ehlschlager C. R., 1998. The stochastic simulation approach: tools for reporting spatial application uncertainty. Research Thesis, University of California, Santa Barbara, California, USA.
Fagan, G. L., Soehngen, H. F., 1987. Improvement of GBF/DIME file coordinates in a geobased information system by various transformation methods and rubbersheeting based triangulation . Proceedings of Auto-Carto 8, eighth international symposium on computer-assisted cartography, pp. 481-491.
Fradkin K., Doytsher Y., 2002. "Urban Digital Cadastre: Analytical Reconstruction of Cadastral Boundaries". Computers, Environment and Urban Systems, Vol. 26(5): 447-463.
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., Shiren, Y., 1992. Development and test of an error model for categorical data . *Int. J. Geographic Information Systems*, Vol. 6 (2), pp. 87-104.
Goovaerts P., 1997. Geostatistics for natural resources evaluation. Oxford University press, New-York. 483 pages.
Greenfeld, J. 1997[a]. Consistent property line analysis for land surveying and GIS/LIS. *Surveying and Land Information systems*, Vol. 57 (2), pp. 69-78.
Greenfeld, J. 1997[b]. Least squares weighted coordinate transformation formulas and their application . *Journal of Surveying Engineering*, November 1997, pp. 147-161.
Hall P., Fisher N. I., Hoffmann B., 1994. On the Nonparametric Estimation of Covariance Functions. *Annals of Statistics* 22(4): 2115-2134.

Hebblethwaite D. H., 1989. Concepts for coping with a shifting cadastral model. *The Australian Surveyor*, 34(5): 486-493.

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.

Kampmann, G., 1996. New adjustment techniques for the determination of transformation parameters for cadastral and engineering purposes. *Geomatica*, Vol. 50 (1), pp. 27-34.

Kiivery H. T., 1997. Assessing, Representing, and Transmitting Positional Uncertainty in Maps. International Journal of Geographical Information Science, 11(1): 33-52.

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.

Morgenstern D., Prell K. M., Riemer H. G., 1989. Digitization and geometrical improvement of inhomogeneous cadastral maps. Survey review 30 (234): 149-159.

Steinberg G., 2001. Implementation of legal digital cadastre in Israel. Proceedings of FIG working week 2001, 6-11 may, Seoul, Korea, CD-ROM.

Valdes G. F., Campusano L. E., Velasquez J., D., Stetson P. B., 1995. FOCAS Automatic catalog matching algorithm. *Publications of the Astronomical society of the Pacific*: 107: 1119-1128.

Yaglom A. M., 1962. An introduction to the theory of stationary random functions. Prentice-Hall, Englewood Cliffs, New-Jersey.

Yaglom A. M., 1986. Correlation theory of stationary and related random functions, Vol. – 1: Basic results. Springer Verlag, New-York. 526 pages.

Yao T., Journel A. G., 1998. Automatic modeling of (cross) covariance tables using fast Fourier transform. *Mathematical Geology*, 30(6): 589-615.

BIOGRAPHICAL NOTES

Dr. 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, and in 2002 he has completed his Ph.D in Geodetic Engineering. He is currently a post doctoral research fellow at the Geospatial Information and Communication Techmology (GeoICT) laboratory at York University, Toronto, Canada.

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 and Environmental Engineering at the Technion, and is currently the head of the Department of Transportation and Geo-Information Engineering. He is also heading the Geodesy and Mapping Research Center at the Technion.

CONTACTS

Dr. Arie Croitoru
GeoICT Laboratory, Department of Earth and Atmospheric Science
Faculty of Pure and Applied Science, York University
4700 Keele Street
Toronto ON
CANADA M3J 1P3
Tel. +1 416 736-2100 Ext.77771
Email: arie@yorku.ca

Prof. Yerach Doytsher
Technion – Israel institute of technology
Faculty of Civil and Environmental Engineering
Department of Transportation and Geo-Information Engineering
Technion City
Haifa 32000
ISRAEL
Tel. + 972 4 8293183
Fax + 972 4 8234757
Email: doytsher@geodesy.technion.ac.il