A Model to Estimate Horizontal Errors within Existing Manually Digitized Maps

Georgios A. Achilleos

Technical University of Crete Sternes Akrotiriou, Chania 73100, GREECE
E-mail: georgiosachilleos@yahoo.com

Abstract
A vast majority of the spatial linear data, which is converted to a digital format by means a manual digitizing process, is used regularly, without properly verifying the reliability of such converted data. It is essential to identify this level of reliability as such data is employed in numerous big projects, which continue to be designed and implemented without knowing the uncertainty and/or risk associated with the outcome. A posteriori knowledge of the level of uncertainty might be enabled using the data per se. In this paper, an effort is made to predict the horizontal accuracy of digitized contours by means of their given digital geometry. This forecast is made by developing a model which makes use of contour geometry indices and provides an estimate of their horizontal accuracy. Subsequently, this knowledge can be utilized to interpolate the surface represented by these digitized contours.
Keywords
contours, horizontal errors, digitizing, model, simulation

I. INTRODUCTION

The availability of reliable digital spatial data is a significant factor influencing the efficiency of spatial information systems (Gong, 1992; Gong, et al., 1995; Kraus, et al.,2007; Leyk, et al., 2005). Analog maps compiled by laborious processes are a notable source of such information. These maps were introduced in these systems following extensive digitization. Since mid-80s and almost throughout the 90s, this digitizing was performed manually by means of a typical digitizer (digitizing table), by integrating all errors included in a map, plus the errors during the digitizing process. Efforts for estimating the final horizontal error of the analog map resulted in values ranging from 0.43 mm—0.77mm (Thapa, Bossler, 1992). Efforts to estimate the digitizing errors led to a more qualitative approach of the problems rather than an estimation of the size of these errors (Amrhein, Griffith, 1991; Aronoff, 1993; Brudson, Openshaw, 1992; Dunn, et al., 1990; Dutton, 1992; Goodchild, et al., 1991; Gong, 1992; Gong, et al., 1995, 2000; Keefer, et al., 1988; Maffini, et al., 1992; Ore, 2001; Otawa, 1987; Warner, Carson, 1991; Wenzhong, Tempfli, 1994). Whilst some were methodical, several others resulted in simple conclusions.

In the case of contours on a map, it is necessary to know the accuracy of their position. Typically, the only detail available is the geometry of these contours and the scale of the map from which they were derived. The elevation accuracy within an elevation surface can be calculated by knowing the positional accuracy of points constituting the contour lines as a measure of the variance of the coordinates of these points (σ_x and $\acute{o}_y = f(\sigma_r)$) (Achilleos, 2006, 2008).

This paper aims to outline an approach for predicting the horizontal accuracy of these contours by means of their digital geometry available. This prediction is made by developing a model that employs the geometric indices of contours and estimating their horizontal position accuracy. The result is an estimation of the variance of coordinates of every point consisting the contour line, in the form of σ_x, σ_y (function of σ_r) (Achilleos, 2006, 2008), as mentioned earlier.

II. ERRORS WITHIN THE SURFACE ELEVATION DATA

Constructing elevation surfaces using contours from an analog map as elevation data, presumes that such data are of good and, preferably, known quality. This way of surface construction is quite rare since other ways faster and more accurate means are available to accomplish this. Nevertheless, huge databases containing elevation surfaces constructed from contours continue to exist . From the earlier passages, it should be evident that these databases need to be examined and verified (Brudson, et al., 1993; Dunn, et al., 1990; Dutton, 1992; Maffini, et al., 1992).

A. The surface construction procedure

The digitized contours embody horizontal errors which are propagated through the procedure of surface construction to the final result, the elevations. The contour horizontal errors are propagated via the geometric transformation during the coordinates stage (Achilleos, 2006) This stage prepares data for the interpolation phase. The errors are then propagated to the calculated elevations of the surface (Achilleos, 2008). These errors could be quite serious when the interpolated surface is subjected to further use (Achilleos, 2006, 2008),. The digitizing errors accompanying the contour lines were unknown in this work, and hence, these errors were approached by estimation. Incorporating a model for estimating these digitizing errors could definitely provide the capability to estimate the elevation errors of the surface in a detailed and

1082-4006/08/14(01)-63$5.00

accurate manner.

B. Digitizing

The digitizing process aims to convert analog information into digital form (Burrough, 1986; Hunka, 1978; Robinson, et al., 1984). This process encompasses all errors that are primarily resulting from human factors, equipments, and the media used for printing the analog map =(Chrisman, 1990; Goodchild, et al., 1991; Otawa, 1987; Thapa, Bossler, 1992).

The influence of certain characteristics on the analog map information, such as line width, line complexity, analog data density, etc. has been proved categorically . (i.e. geometry elements) affect the digitizing result (Dutton, 1992; Maffini, et al., 1992; Otawa, 1987; Thapa, Bossler, 1991).

The MANTISSA project carried out in 1990 established that the minimum error under the best possible conditions is in the order of 0.19mm (Melling, 1991). The conditions that were applied are nearly prohibitive for everyday work. These approaches are usually based on a series of assumptions. Investigating the digitizing error is a questionable subject, since the traditional manual digitizing is being replaced by modern, faster and more flexible methods (Waters, et al., 1988).

The volume of information available which is abundant, structured, organized, and used widely in applications, is difficult to be replaced readily. However, this digital information includes errors from the digitizing process, and remain unknown. Therefore, there is a need to investigate this existing information and evaluate its accuracy systematically.

III. DESIGNING THE INVESTIGATION OF THE DIGITIZING ERRORS

A. Background

Given that the metadata of digital information is usually unknown, one can only estimate or predict its level of quality by utilizing the digital information itself and any other possible derivative. In other words, one should try to predict the image of the error on the basis of the information which includes this error.

In the case of contours, digital data may provide elements for the geometry of these contours. Therefore, a model that attempts to predict the digitizing errors should be based, in the absence of other data, on the geometry of the contour as described by digital data.

The design of the variables for the construction of the model is based on the rational of the three consecutive points (Figure 1). These three points determine approximately the geometry on the peak of a contour where the digitizing error should be predicted. Based on this geometry, all the input variables of the model are measured. The direction of the tangent and, therefore, its vertical may be defined by these three points in a distinct form.

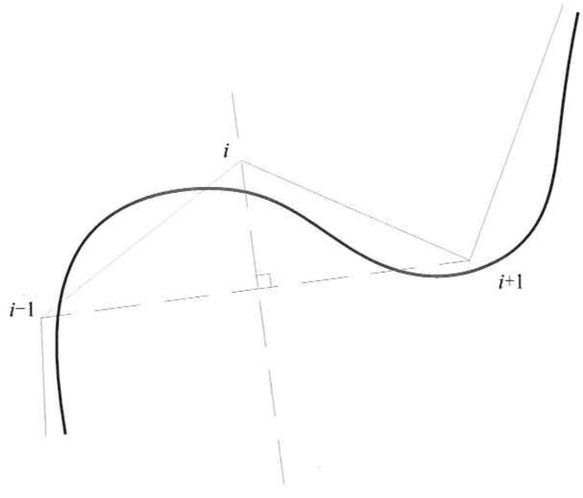

Figure 1. The three consecutive points

B. Defining the digitizing error

The digitizing error is defined as the deviation of the digitized contour lines from their real position, which is defined as the real position of the contour on the analog map.

This deviation is defined as the vector connecting the point on the real contour and the point which is finally effected and digitized. Usually this second point is located outside the contour line (Figure 2).

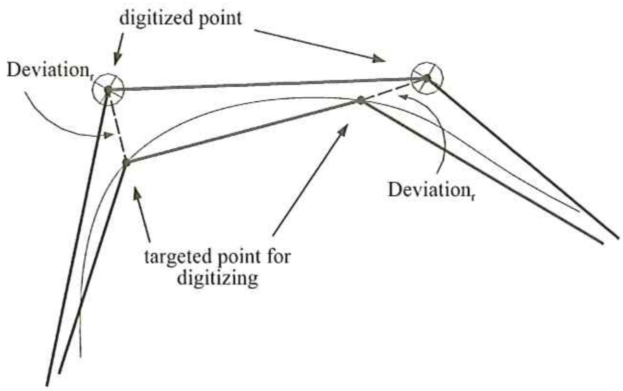

Figure 2. Deviation from the digitizing process

As illustrated in Figure 3, the deviation may be analyzed in two components, vertical (r_k) and tangential (r_e).

This paper deals with the vertical component of the error. The tangential component defines a transposition of this point on the same contour line (Figure 3). This transposition has no effect on the accuracy of the horizontal position of the contour line, provided that it is relatively small and does not significantly affect the shape of the contour.

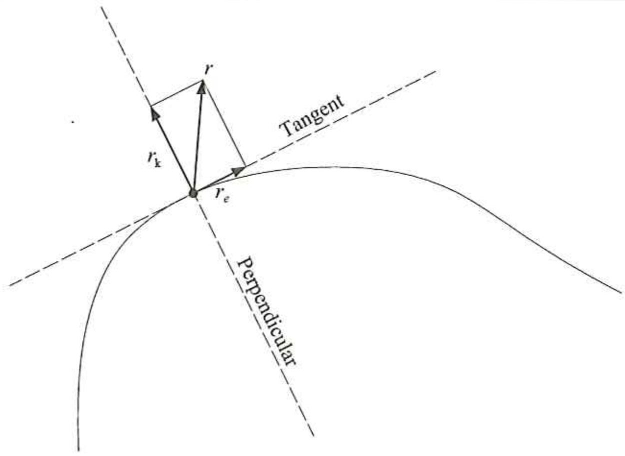

Figure 3. Analysis of the deviation in its two components

The mean digitizing error, derives from the existence of many digitizing errors that are found on the vertical line at each point of the contour, something that results from the participation of a group of people (Figure 4).

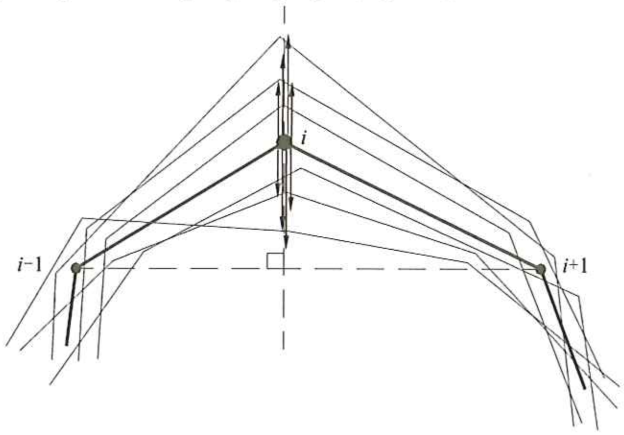

Figure 4. Vectors of the digitizing error at point i of the contour

The vectors of these errors, is a common direction (vertical to the tangent of the contour line), a common definition point (point i of the contour line) but a different course (left or right of the contour line) and a different measure (length) (Figure 4).

Digitizing errors present a random behaviour and for specific digitizing conditions they follow a regular distribution the mean value and their standard deviation as parameters (Brudson, Openshaw, 1993; Gong, 1992; Gong, et al., 1995; Nakos, 1990; Keefer, et al., 1988; Warner, Carson, 1991).

Following this, the digitizing error is analyzed and processed on the basis of two parameters that can describe its distribution and provide its image.

- Mean value of the vectors of the digitizing error
- Mean value of the measure of the digitizing error vectors

When addressing the problem of digitizing errors as a distribution of vectors on a specific vertical direction, the expected mean value would be very close to the definition point of these vectors, namely point i (Figure 5).

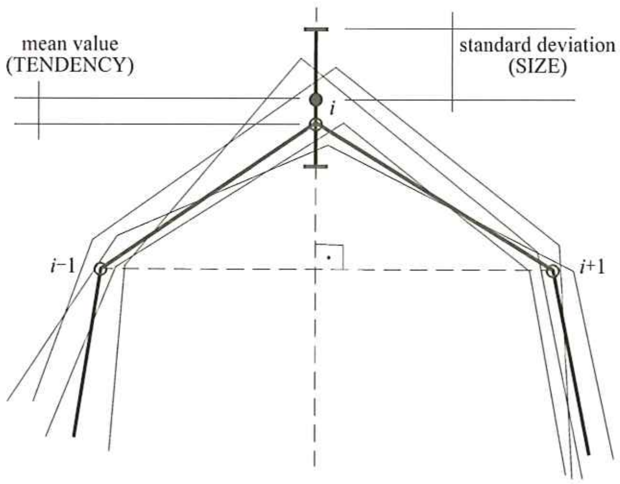

Figure 5. Error distribution

Addressing the problem in this manner defines the potential real position of point i, with a confidence interval (Figure 5). The mean value of these vectors is an indication of the error's tendency to be distributed towards one direction or course (left or right of the contour). In contrast, the size of the digitizing error is determined by the mean measure of the error vectors, irrespective of the course (left or right). The resulting size approaches the deviation of these vectors and describes the size of the digitizing error.

IV. ANALYZING THE DIGITIZING ERROR DATA

A. The research data

The conditions on the basis of which the whole process would be carried out were defined from the beginning of this research.

A topographic map was created, which resulted from the analog topographic map of a 1:10,000 scale. This map was digitized very carefully; an intervention was also made in order to integrate specific forms of contour lines of particular interest (addition and removal of parts of the map). Following this, the contour lines were approached by 3rd degree polynomial, so that their distorted parts are smoothed out and appear more realistic. When the map was completed in a digital form, it was printed on a 1:10,000 scale at an extra high resolution printer. This product was the "analog" map of the research to be digitized (Figure 6).

A group of people with experience in digitizing was selected by sampling among the companies in the field of gathering, organizing and marketing spatial data. This group also included undergraduate and postgraduate students with digitizing background. A remarkable number of people were screened and 34 were finally selected to digitize the map, thus participating in the research.

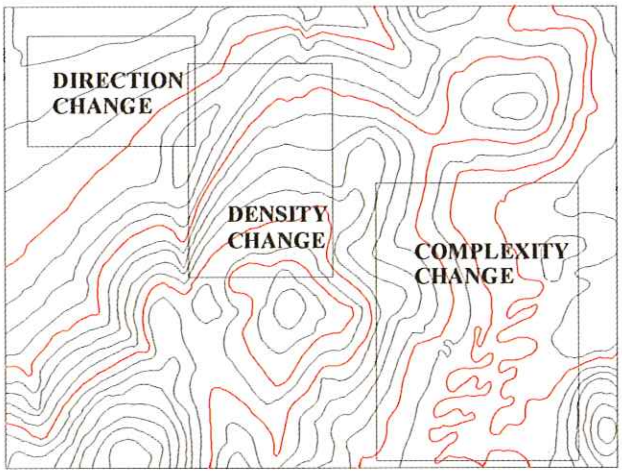

Figure 6. Digitizing map

B. Analyzing data

(i) Statistical data for the digitizing error vector

Having measured the digitizing error at every point of the contour lines (Figure 6, 7) from every digital map of the participants, the mean vector of the digitizing errors and the mean size of these vectors were calculated for every point. The total number of points that are examined on the map is 1032.

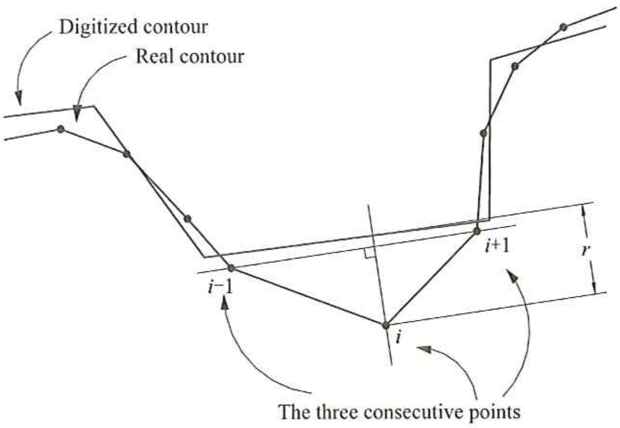

Figure 7. The logic of the algorithm for the calculation of the digitizing error

Table 1 presents the elements of the mean vector of the digitizing error r. It is observed that the mean digitizing error r is 0.00 mm from all points examined, a fact indicating that there are negative and positive values of errors at the same percentage.

Figure 8 below presents an extract of the map, on which the mean vectors of the digitizing error r are indicated together with the corresponding circles. It can be observed that the error tends to switch locally (changes sign from negative to positive and vice versa, at a local level). The same is observed in Figure 9 which presents an extract of the open curve with

Table 1. Statistical data of the mean vector of the digitizing error *r*

Digitizing error measure	
Statistical data	
mean	0.000mm
sd	0.190mm
min	−0.640mm
max	0.536mm

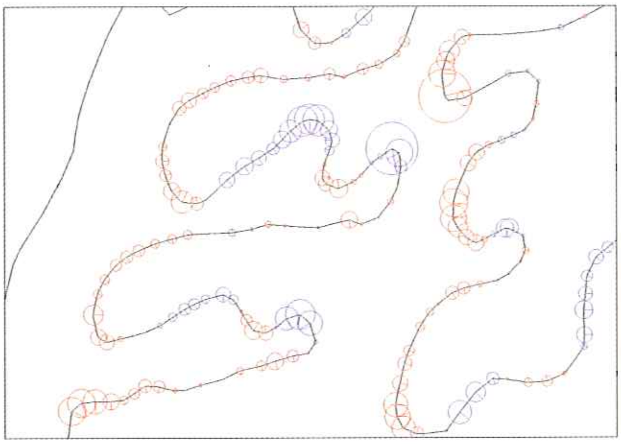

Figure 8. Circles of the mean vector of error *r*

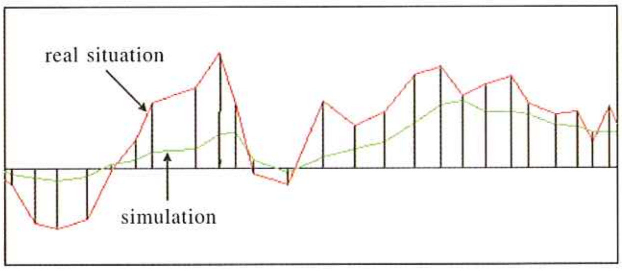

Figure 9. Presentation in length of the mean vector of error r

the error diagram in the Oy axis and the curve length in the Ox axis. The switch is more apparent is this figure (Figure 9), where there is an image of time line.

In Figure 8 2-3 circles seem to have extremely large size and indicate cases where the algorithm that measures these errors failed. These points were excluded from the analysis and the statistical processing.

(ii) Statistical data for the mean vector of the digitizing error

Table 2 presents statistical data of the mean size of the vectors of the digitizing error ||*r*||. These sizes are measured in mm.

It is observed that the mean value of ||*r*|| is:

0.284mm

that is slightly higher than the distinctive capacity of the human eye (0.25mm), which is broadly accepted in cartography.

Table 2. Statistical data of the mean size of the digitizing error $\|r\|$

Digitizing error measure	
Statistical data	
mean	0.284mm
sd	0.084mm
min	0.061mm
max	1.098mm

The maximum value of $\|r\|$ that was observed (except in cases of the measurement algorithm failure) is:

1.098mm

On the whole, the results are found to be satisfactory. When observing the value range within which 99.9% of the errors is found, the maximum limit was found to occur at 0.54mm, very close to 0.50mm, a value that is the feasible error of obtaining data from an analog map. Moreover, the mean value of errors is 0.28mm, which is usually considered to be the distinctive capacity of human vision (0.25mm).

Figure 10) below presents an extract of the map with the circles of error $\|r\|$ at each point of the digital curves. These circles confirm the aspect that the mean digitizing error is bigger at the curved parts of the contours rather than at the straight ones. It is also observed that this fluctuation of the mean measure of the digitizing error is gradual and not abrupt (Figure 10). This observation also indicates the serial dependence of the digitizing error.

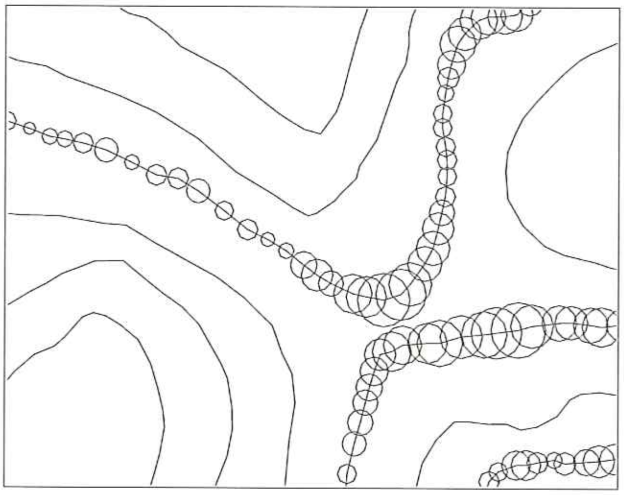

Figure 10. Extract of the map with the digitizing error in the form of error circle

(iii) Auto-correlation of the digitizing error

Figure 11 presents the auto-correlation of the mean vector of the digitizing error r of each point. It is observed that the auto-correlation at each point with the previous 4—6 points is higher in number than the auto-correlation presented by the mean measure of vector $\|r\|$(Figure 12). For one position lag, auto-correlation lies within 0.70—0.75, while with regard to size r, auto-correlation appears to reach even the 16th position (16th lag) and beyond.

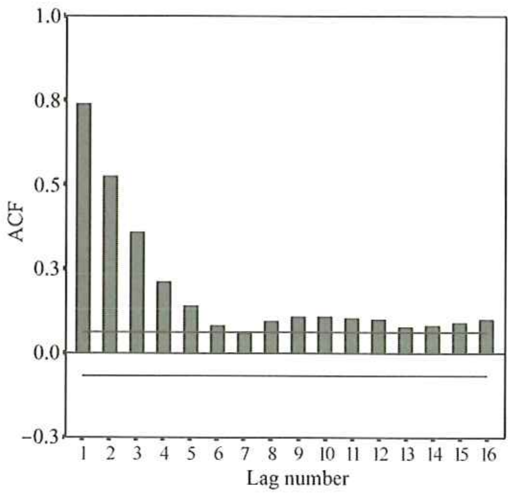

Figure 11. Diagram of the auto-correlation of size r

$\|r\|$

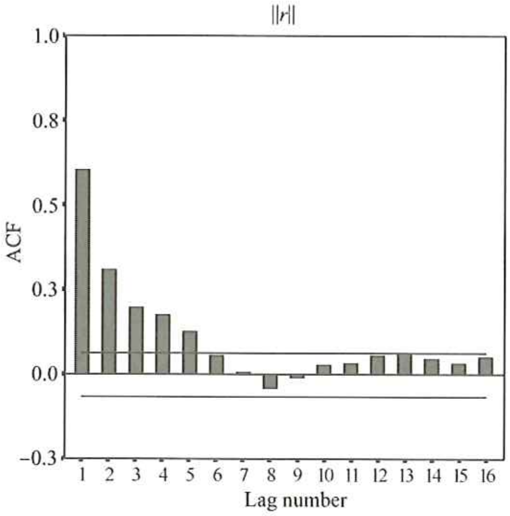

Figure 12. Diagram of the auto-correlation of size $\|r\|$

Figure 12 shows that there is auto-correlation and dependence of the mean measure of the vector of the digitizing error $\|r\|$ at each point with the 4...6 points of the previous positions at the curve (4...6 lags). The auto-correlation for one position lag (1st lag) is 0.60—0.70. (Bora-Senta, Moisiadis, 1992; Neter, et al., 1990).

V. DESIGNING THE GEOMETRY INDICES

Modelling the digitizing error requires the design of certain geometry indices, which would be the independent variables of the model under consideration.

Many indices are designed (35 in particular), which in their majority present a high correlation among them. In the modelling process, some of these indices are finally used; the ones that describe the phenomenon in a better way. These are further presented below.

The design of the geometry indices was implemented in three groups; the point-based measured indices, the locally measured indices, and the indices that describe the prior geometric situation from the point under examination.

A. Point-based indices

The point-based indices are indices that calculate geometrical data of the digital contour at each point and in the very close area around it (three consecutive points). These indices include the following, which are also used.

(i) Central coordinates (lx, ly)

The indices of the central coordinates (lx, ly) are defined as the distance at the X and Y axes of the point examined from the centre of gravity of the map (Figure 13). After the coordinates of the position of the map's centre of gravity (x_c, y_c) are calculated from all the digitized points, the central coordinates are measured from equations 1 (1).

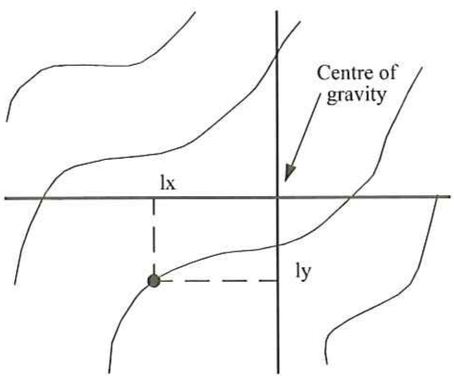

Figure 13. Central coordinates index (lx, ly)

$$lx = x_i - x_c \quad x_c = \sum_{i=1}^{n} x_i / n$$
$$ly = y_i - y_c \quad y_c = \sum_{i=1}^{n} y_i / n \qquad (1)$$

where
(x_i, y_i):the coordinates of the point under examination
(x_c, y_c):the coordinates of the centre of gravity
n: sum of digitized points of contours of the map

The central coordinates determine the relative position of the digitizing point with respect to the entire map (position in relation to the centre of gravity).

(ii) Curvature (cur)

The *curvature* at each point of the contours is defined as the quotient of the change of direction at the point examined to the local length (Figure 14). The change of direction is measured in radians (rad). The higher this measure is, the bigger the curvature of the analog curve that was digitized.

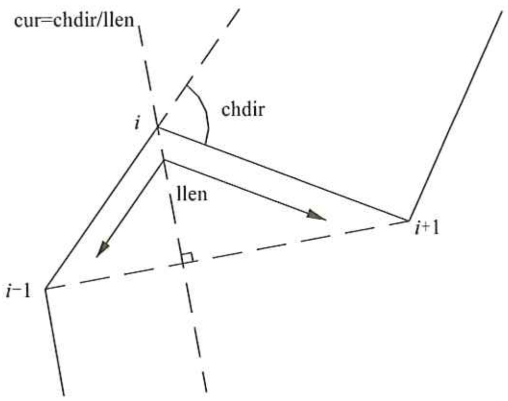

Figure 14. Distinctive curvature index

The curvature is calculated from equation 2 (2):

$$cur=chdir/llen \qquad (2)$$

where
chdir: change of direction [azimuth (i+1,i) - azimuth (i-1,i)] (Brundson C., Openshaw, 1993)
llen: local length [d(i-1),i + d(i, i+1)]

(iii) Diversion(ekt)

The *diversion* index (ekt) is the vertical distance from the last digitizing point towards the straight line defined by the previous two digitized points (Figure 15).

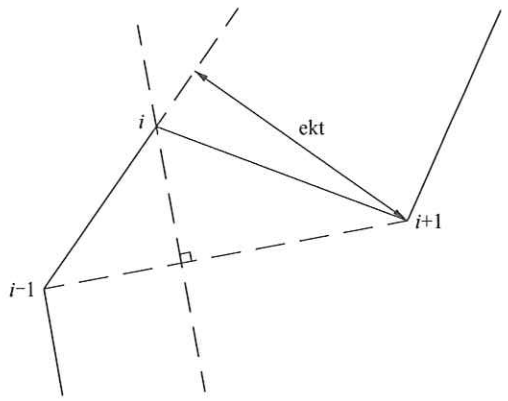

Figure 15. Diversion index

The calculation of diversion is based on the calculation of the distance of a point from a straight line (3).

$$ekt = \left| \frac{A \cdot x_{i+1} + B \cdot y_{i+1} + C}{A^2 + B^2} \right| \qquad (3)$$

where
A,B,C: the coefficients of the equation of the straight line that is defined by points i-1 and i.

(Ax+By+C=0)

(iv) Direction of tangent(dir)

The index of the direction of tangent (Figure 16) is calculated from equation 4 (4) and the unit used is the radian (Warner W, et al., 1991).

$$dir = Azim(i-1, i+1) \qquad (4)$$

This index describes the movement of the digitizer's crosshair

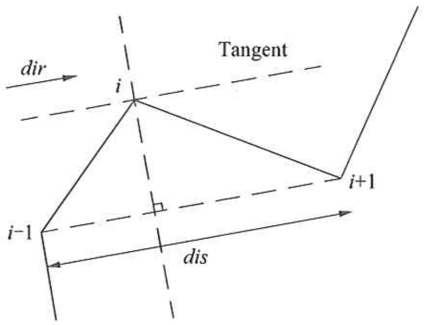

Figure 16. Indices of the tangent's direction

during the digitizing phase.

B. Locally measured indices

The locally measured indices are an expansion of certain indices measured at the point at a broader area around the point under examination. In other words, it regards a generalization process. The three consecutive points are now defined as $(i-1)'$, i, $(i+1)'$ (Figure 17).

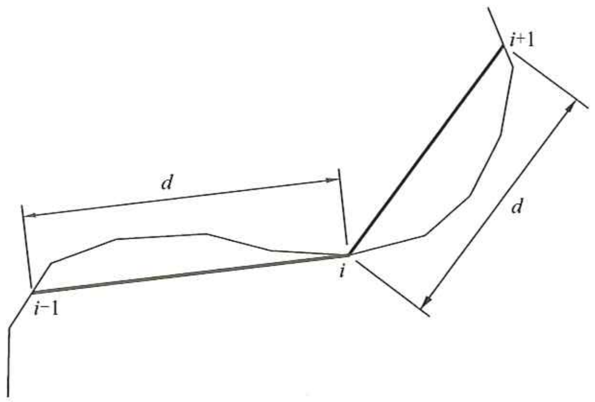

Figure 17. Defining the three consecutive points at local level

In this group, the indices are similar to the indices of the first group (point-based indices), with the difference that they refer to the points $(i-1)'$, i, $(i+1)'$. The length that determines these points is a result of the investigation in conjunction with the study of the auto-correlation presented by the geometrical indices.

Nevertheless, finally, none of the indices from this group was used in the modeling, in this project.

C. Indices of prior geometrical situation

These indices aim to describe the phenomenon of the serial dependence that exists in the data/products of the digitizing process, as observed in research studies.
An index of this group, which is used in the modeling is:

- Distance from centre of gravity(dcom)

The index of the *distance from the centre of gravity* of the contour that is examined is a size that aims to describe how close or far the complex part of the digitized contour is found prior to the point under examination. A complex part of the contour means that there are more dense points-peaks in order that the centre of gravity is transferred near this complex part (Figure 18). High index values indicate that the most complex part of the previous part of the contour is far enough from the point under examination, while low values indicate the opposite (Figure 18).

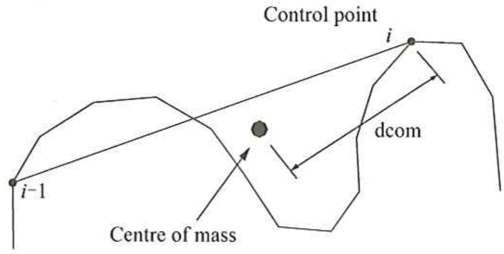

Figure 18. The index of distance from the centre of gravity

The calculation of this index is based on equations similar to equations 1 (1).

VI. MODELING THE MEAN BEHAVIOR OF THE DIGITIZING ERRORS

Two approaches are employed while investigating and modeling the behavior of the digitizing errors:

- Investigation with the use of the geometrical characteristics of the digitized contours;
- Investigation with the use of the geometrical characteristics of the digitized contours and serial dependence (auto-correlation) of the errors

The first approach is based on the assumption that digitizing errors follow a pattern of independent behavior (the error of a point does not depend on the errors of the previous points).

The second approach introduces the concept of the dependent behavior (the error of a point is directly dependent on the errors of the previous points and the following points, in turn, will depend on it).

A. A model for the mean vector of the digitizing error

At a first attempt to model the digitizing error, regression was applied with the stepwise method (gradually selecting the variables). The independent variables that were selected in the linear model, among the geometrical indices, are the following in order of contribution:

- ekt
- lx
- sin(dir)

The degree of adjustment of this model in the data is:
Multiple R 0.5792

R Square	0.2641
Adjusted R Square	0.2597
Standard Error	0.1412
Durbin-Watson Test	0.6571

The independent variables present the highest possible contribution, but they cannot describe the phenomenon of the digitizing error satisfactorily (R Square 0.2641). Efforts were made with other forms of models (polynomial or generic non linear models), but they did not provide better results than the linear model (it is preferred for its simplicity).

On the basis of the value of the Durbin-Watson test, the model presents an intense serial dependence, which should be either eliminated so as to have only the effect of the geometrical indices (Neter, et al. [24]), or used in order to improve the phenomenon's explanation (Bora-Senta and Moisiadis [5]; Chartfield [8]; Xenakis [32]).

The problem of the serial dependence is indicated in the Diagram of the model's residuals (Figure 19).

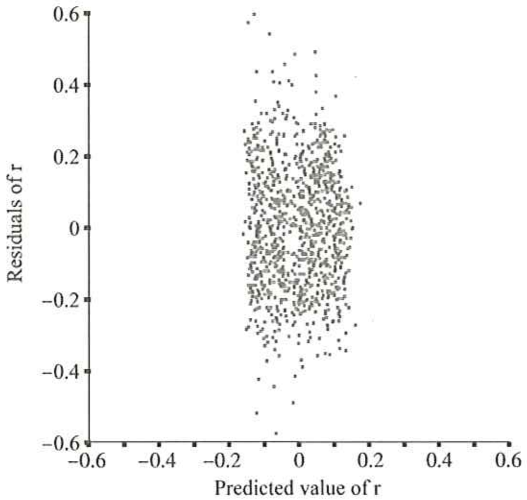

Figure 19. Diagram of the residuals of model *r*/sindir, ekt, lx prior to the elimination of the serial dependence

The integration of auto-correlation in the modeling process of the phenomenon is effected for one position lag. This results from studying the diagrams of auto-correlation of r and ||*r*|| in combination to the auto-correlation diagrams of the geometrical indices.

The elimination of the serial dependence with the use of the autoregression process, leads to a model as depicted in Table 3.

The diagram of the residuals of this model (Figure 20) is clearly corrected and does not present indications of error regularity or heteroscedasticity, however its adjustment in the data remains feeble.

The integration of the serial dependence in the modeling leads to the model of Table 4.

Table 3. Model *r*/sindir, ekt, lx after the elimination of the serial dependence

Multiple R	0.4831
R Square	0.2217
Adjusted R Square	0.2184
Standard Error	0.1391
Durbin-Watson Test	1.9167
EKT	+0.01982
LX	+0.00721
SINDIR	−0.09822
CONSTANT	−0.03876

Table 4. Model *r*/sindir, ekt, lx, *r*_1 with the use of regression

Multiple R	0.7962
R Square	0.6830
Adjusted R Square	0.6781
Standard Error	0.1174
Durbin-Watson Test	1.8752
EKT	+0.01274
LX	+0.00019
SINDIR	−0.02414
R_1	+0.68231
CONSTANT	−0.02391

This model (Table 4) presents a better degree of adjustment in the data than the previous one (Table 3). The serial dependence is integrated in the form of an independent variable in the model (*r*_1). This variable is the value of the dependent variable r at the previous position of the point under examination (position *i*-1).

The existence of the term r_1 on Table 4 produces limitations in the application of the model of the error's tendency (Bora-Senta, Moisiadis, 1992; Neter, et al., 1990). The model offers a satisfactory simulation degree of the digitizing error (Keefer, et al., 1988). This matter is examined below.

With regard to the model's residuals (Figure 20), they present a distribution as the one in Figure 20 and do not indicate

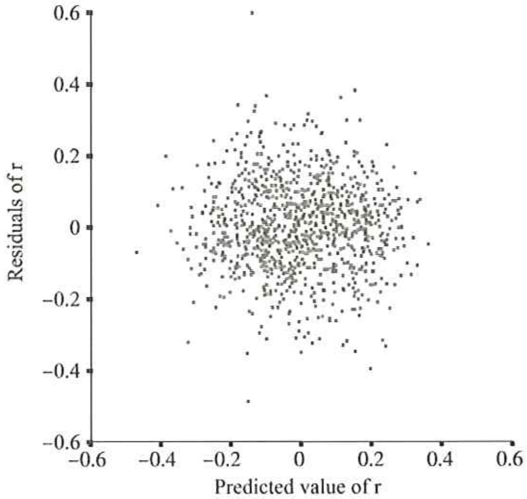

Figure 20. Diagram of the residuals of model r / sindir, ekt, lx after the elimination of the serial dependence

regularity or heteroscedasticity. This observation is documented by applying the F-test control.

In brief, the tendency of the digitizing error, apart from its high auto-correlation, appears to have a stronger relation to the digitizing direction. This relation is stronger not to the direction as azimuth (dir), but to its sine (sindir) (namely, azimuth 10^g and azimuth 390^g have the same sine and present the same behaviour with regard to the error's tendency).

B. Model for the mean size of the digitizing error

A regression process was applied with the stepwise method and the variables selected from the sum of the geometrical indices for the linear model are the following in order of contribution:

- cur
- dir
- dcom

The degree of adjustment of this model in the data is:

Multiple R	0.3479
R Square	0.1761
Adjusted R Square	0.1659
Standard Error	0.1317
Durbin-Watson Test	0.7782

It is obvious from the statistical Durbin-Watson size that this model presents an intense serial dependence. This is also indicated by the diagram of the residuals (*Y* axis) in connection with the predicted value (*X* axis) (Figure 21). This diagram shows that the residuals tend to be more negative than positive. When resolving the model using auto-regression, the serial dependence can be eliminated and the model as shown on Table 5 below is obtained.

The new model (Table 5), although better than the previous

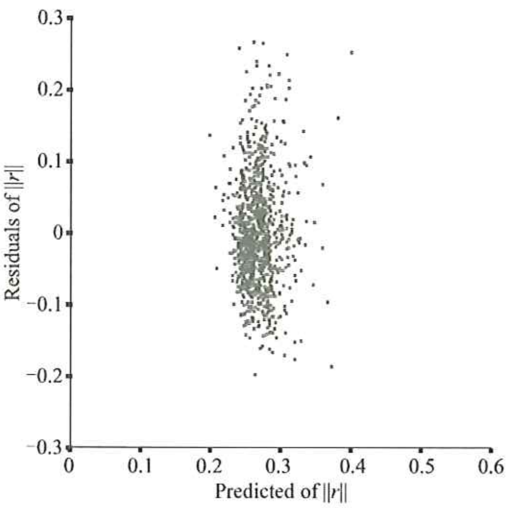

Figure 21. Diagram of residuals of model ||r||/cur, dir, dcom prior to the elimination of the serial dependence

Table 5. Model ||r|| / cur, dir, dcom after the deletion of the serial dependence

Multiple R	0.5128
R Square	0.2960
Adjusted R Square	0.2792
Standard Error	0.1271
Durbin-Watson Test	1.9029
CUR	+0.08730
DIR	+0.01628
DCOM	−0.00571
CONSTANT	−0.25172

one, continues to offer a low degree of adjustment in the data. However, it is a useful tool for applying simulation procedures of ||r|| which is further described.

The diagram of the residuals of this model is improved in terms of serial dependence (Figure 22). It presents a minimum

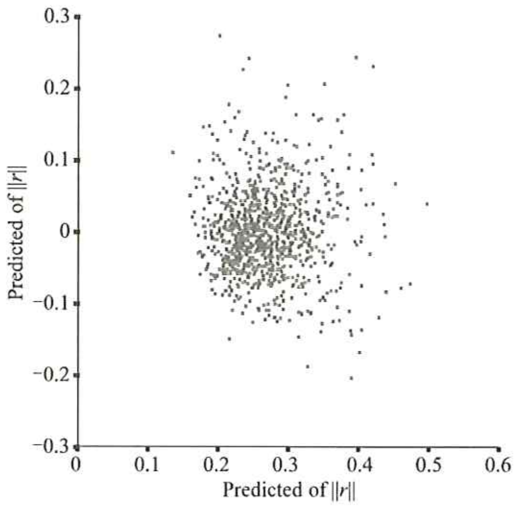

Figure 22. Diagram of residuals of model ||r||/cur, dir, dcom after the elimination of the serial dependence

heteroscedasticity, which is trivial (Neter et al., 1990).

By integrating the serial dependence in the form of an independent variable into a model with the use of simple regression, the following results are obtained (Table 6).

The model of Table 6 is improved, thus offering the possibility to simulate the mean measure of the digitizing error.

Table 6. Model ||r||/cur, dir, dcom, ||r||_1 with the use of regression

Multiple R	0.7233
R Square	0.5711
Adjusted R Square	0.5582
Standard Error	0.1131
Durbin-Watson Test	1.8328
CUR	+0.06821
DIR	+0.00592
DCOM	−0.00672
||R||_1	+0.59261
CONSTANT	−0.07592

The diagram of the residuals of this model is in the form of Diagram 6 (Diagram 6).

In brief, the mean measure of the digitizing error presents a high dependence with the size of curvature (cur) and complexity (expressed by the dcom variable), as opposed to the tendency of the digitizing error, which is highly dependent on the digitizing direction (sindir variable).

This observation is rational, given that the complex form of the contours that are digitized affects the person who digitizes, thus leading to bigger errors in size (error measure). On the contrary, the direction presented by the contours that are digitized, affects the person performing the digitization. This is so, as the one who performs the digitization attempts to preserve the digitizing course close to the analog contour, thus resulting in the digitizing errors continuously changing their sign (right- left of the analog contour / error tendency).

C. Applying the models in simulation procedures

The digitizing error may be approached partially, not with the aim to predict its behavior accurately, but to simulate this procedure for research purposes. The prediction capacity of the model plays an important role in the simulation.

If a model gives the value Y_p as the prediction value, it should be accompanied by the standard deviation of $\acute{o}_{Y_p}$. This standard deviation, for the specific statistical level of confidence, may be translated into a confidence interval of the prediction.

Model:

$$Y_p = [b] \cdot [X] \quad (5)$$

where
[*b*]: table of the model's coefficients
[*X*]: table of vector of the model's independent variables
Y_p: dependent variable - prediction

The variance of the predicted value Y_p derived from (Neter, et al., 1990):

$$S^2_{Y_p} = [X]' \cdot [V_b] \cdot [X] \quad (6)$$

where
[V_b]: The matrix of variance – co-variance of the model's coefficients(symmetric table)

This matrix comes from equation 5 (5) (Neter, et al., 1990):

$$V = \sigma^2 \cdot (X' \cdot X)^{-1} \quad (7)$$

The prediction is considered to be better, i.e. it presents lower variance, at the centre of the experimental area. The quality of the prediction is diminished while moving away from this centre.

Further on, these models are applied in simulation procedures.

(i) Simulation of the mean vector of the digitizing error
The models developed for the mean vector of the digitizing error are as follows:

Model 1:

$$r = 0.0198\,\text{ekt} + 0.0072\,\text{lx} - 0.0982\,\text{sindir} - 0.0387$$

where the serial dependence is deleted; and

Model 2:

$$r = 0.0127\,\text{ekt} + 0.0002\,\text{lx} - 0.0241\,\text{sindir} + 0.6823\,r\_1 - 0.0239$$

in case the serial dependence is integrated in the data.

The variance-covariance matrix for the first model (Model 1) is presented in Table 7, while for the second model (Model 2) the matrix is presented in Table 8.

Table 7. Variance-covariance matrix of Model 1 (r/ekt, lx, sindir)

	ekt	lx	sndir	const.
ekt	5.3526E-3			
lx	−3.8728E-3	0.1198E-3		
sindir	1.6041E-3	2.8975E-3	7.7694E-3	
const.	−4.2251E-3	−3.6882E-7	−1.6269E-5	10.1328E-3

Table 8. Variance-covariance matrix of Model 2(r/ekt, lx, sindir, r_1)

	ekt	lx	sndir	r_1	const.
ekt	4.0612E-3				
lx	−1.6494E-8	0.0962E-3			
sindir	−5.8858E-7	1.3694E-7	6.3920E-3		
r_1	−1.3841E-5	−2.257E-7	5.8882E-5	23.693E-3	
const.	−2.3827E-5	−2.127E-7	−6.686E-6	2.2115E-5	7.6064E-3

For Model 2, it is necessary to know the error for the first point of each contour as input data in the prediction of the error of the second point of the contour.

A random number is selected within the range that is determined by the standard deviation of the mean vector of the digitizing error, in combination with a certain statistical confidence level (e.g. level of confidence 95%). In this particular case, the mean vector of the digitizing error presents a value of standard deviation ±0.19. Therefore, for a confidence level of 95%, the value range within which an initial value may be selected for the first point is:

±0.3812

In the simulation under consideration, each contour is selected so that it is approached by thirty (-30-) repetitions and that the mean value is calculated.

The results of the simulation with the application of models 1, 2 compared to the data from the errors that were measured, are shown on Table 9. It is observed that the simulation presents the digitizing error better than it is in reality, given that lower range values appear (min, max) and the standard deviation of the predicted values of the digitizing errors is nearly half the

Table 9. Simulation results based on Models 1 and 2

	OBSERVED	MODEL 1: PREDICTED	MODEL 2: PREDICTED
min r	−0.6403	−0.8891	−0.1754
max r	+0.5360	+0.9123	+0.1726
mean r	−0.0041	+0.0012	+0.0039
sd r	±0.1872	±0.1051	±0.0816

real size.

Furthermore, it is observed that model 2 gives lower predicted digitizing errors than the corresponding ones of model 1.

The two distributions of digitizing errors (observed, predicted) are compared statistically in order to find out whether or not and to what extent they are identified serially. This control is performed using the test-x^2 (Hammond, McCullagh, 1974) and the result is affirmative.

Ç ï: the two distributions are approached statistically at a satisfactory level
$x^2 = 24.0698$
$x^2_{critical} >> 43.77$
n=998−1=997 freedom degrees and confidence level 95%

The two distributions are checked in parts (e.g. per forty points) in order to confirm this behavior at a smaller scale. The result is the same, namely that the two distributions are approached. This check is a good indication that the simulation applied approaches the phenomenon satisfactorily.

This statistical approach of the two distributions can be seen in Figure 23 (Figure 23) where an extract of the two distributions is depicted. Based on this diagram, the level of approach of the two distributions can be compared, which in statistical terms is regarded as reliable (error probability 5%).

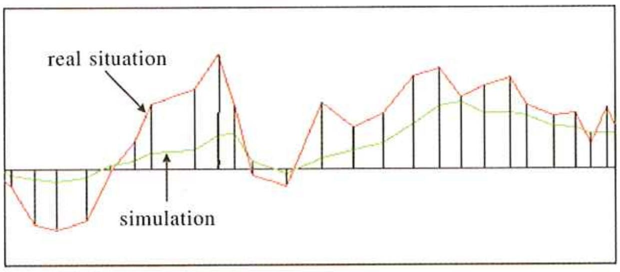

Figure 23. Comparative presentation between the real digitizing error and the error predicted from the simulation

(ii) Simulation of the mean size of the digitizing error
A similar approach as in the mean vector of the digitizing error is followed in the case of the mean size $\|r\|$.

The models that were developed and used in the simulation of this size are:

Model 3:

$$\|r\| = 0.0873\, cur - 0.0163\, dir + 0.0057\, dcom + 0.2517$$

Model 4:

$$\|r\| = 0.0682\, cur - 0.0059\, dir + 0.0067 dcom + 0.5926\, \|r\|\_1 + 0.0759$$

The variance-covariance matrix of these two models are shown on Tables 10 and 11 (Tables 10, 11).

Table 10. Variance-covariance matrix of model 3(r/cur, dir, dcom)

	Cur	abs_dir	dcom	const.
cur	1.0615E-2			
abs_dir	−0.818E-8	2.8395E-3		
dcom	1.3461E-3	−0.1782E-3	2.5612E-3	
const.	−4.335E-3	−1.9016E-3	−5.318E-3	1.3563E-2

Table 11. Variance-covariance matrix of model 4 ($\|r\|$/cur, dir, dcom, $\|r\|$_1)

	Cur	dir	dcom	$\|r\|$_1	const.
cur	8.5782E-3				
dir	−0.281E-3	2.3177E-3			
dcom	1.1094E-3	−0.1803E-3	2.0788E-3		
$\|r\|$_1	−1.1008E-3	1.2297E-3	−0.349E-3	24.822E-3	
const.	−2.8295E-3	−1.8311E-3	−3.8507E-3	−9.349E-3	12.8591E-3

The values range within which the selection of the absolute value of the digitizing error is made for the first points of the contours is:

$$\pm 0.1592$$

The results of the simulation are shown on Table 12 .

Table 12 shows that in both cases of models, the minimum value predicted is negative (very close to zero), something that in fact cannot happen. Once again, it can be noted that the values of this table (Table 12) are the statistical result of thirty (–30–) successive simulation repetitions.

It is observed that the predicted values of the error size that resulted from the two models are almost identical with the mean value of the error size observed. A difference is observed at the maximum value predicted by model 3, as well as at the variations of the errors that are predicted in both models. On the whole, it is observed that model 4 predicts smaller errors.

With a procedure similar to the one for the mean error vector,

Table 12. Simulation results based on Models 3 and 4

	Observed	Model 3: Predicted	Model 4: Predicted
min $\|r\|$	+0.0675	−0.0068	−0.0042
max $\|r\|$	+0.6527	+0.6211	+0.3988
mean $\|r\|$	+0.2685	+0.2692	+0.2637
sd $\|r\|$	±0.0797	±0.0582	±0.0367

one may observe the uncertainties of the predictions of these models (models 3, 4). Moreover, a test-x^2 may be carried out in order to ascertain whether and how the observed error sequence is approached by the predicted sequence.

The test provides the following results:

$$x^2=24.0863$$
$$x^2 \text{critical} >> 43.77$$

which determine that the two distributions are approached statistically for a confidence level 95%.

VII. GENERAL OBSERVATIONS REGARDING THE MODELING AND SIMULATION

Modeling a phenomenon is a complex procedure (usually in connection with the phenomenon's complexity), that requires meticulous attention and deliberation, and also many repetitions in order to succeed.

In this research effort, the nature, complexity, and multi-variability of the digitizing process make the modeling of the phenomenon difficult.

The fact that the data presents an intense serial dependence, affects the adjustment of a simple linear model and imposes the use of a time series model, so that the explanation of the phenomenon is satisfactory at a level of approximately 70%. The possibilities to utilize such a model are limited, mainly to simulation cases and to comparisons of the various geometric forms of parts of contours with the digitizing error.

These models may also be combined, so that the digitizing phenomenon is approached in a better fashion. The use of "mean vector of digitizing error" and "mean size of digitizing error" may realize an uncertainty zone around each digitized contour, within which the contour may be positioned.

The simulation results may be utilized appropriately in cases where the question is the effect of the digitizing error on secondary products, that resulted from the processing of the digitized information. These products, such as terrain slopes, length-sections and cross-sections, digital terrain models, visibilities, calculations of solids (volume, expanded surfaces), reports, etc., are influenced by these errors, given that they make use of the horizontal position of the digitized contours. The possibility to simulate the digitizing error offers the opportunity to study the distribution of the problem in these secondary products and to estimate the level of risk undertaken by using these products.

REFERENCES

[1] Achilleos G., 2006, *Propagation of Uncertainty within the Affine Transformation Application on Contours*, Journal of Geographic Information Sciences, 12(2): 79—91.

[2] Achilleos G., 2008, *Errors within the Inverse Distance Weighted (IDW) interpolation procedure*, Geocarto International, 1—21, iFirst article (http://dx.doi.org/10.1080/10106040801966704).

[3] Amrhein C., 1991, Griffith D., *A Statistical Model for Analyzing Error in Geographic Data in an Information System*, Dept. of Geography, University of Toronto, Discussion Paper No. 38, Canada.

[4] Aronoff S., 1993, *Geographic Information Systems: A Management Perspective*, WDL Publications, Canada, Chapters 4—5, pp. 103—149.

[5] Bora-Senta E, Moisiadis Chr., 1992, *APPLIED STATISTICS: Multiple regression—Dispersion analysis—Time series*. 2nd edition. Ziti publications, Thessaloniki (in Greek).

[6] Brudson C., Openshaw S., 1993, "Simulating the Effects of Error in GIS.", *Geographical Information Handling - Research and Applications, (editor: Mather P. M.)*, John Wiley & Sons, England, Chapter 7, pp. 47—61.

[7] Burrough P. A., 1986, *Principles of Geographical Information Systems for Land Resources Assessment*, Oxford University Press, Great Britain.

[8] Chartfield C., 1980, *The Analysis of Time Series: An Introduction*, Second Edition, Chapman and Hall Ltd, USA.

[9] Chrisman N., 1990, "Efficient Digitizing through the Combination of Appropriate Hardware and Software for Error Detection and Editing.", *Introductory Readings in GIS (editors: Peuquet D., Marble D.)*, Taylor & Fransis, pp. 233—246.

[10] Dunn R., Harrison A. R., White J. C., 1990, "Positional Accuracy and Measurement Error in Digital DataBases of Land Use : an empirical Study.", *International Journal of Geographical Information Systems*, 4(4): 385—398.

[11] Dutton G., 1992, "Handling Positional Uncertainty in Spatial Databases.", *Proceedings of 5th International Symposium on Spatial Data Handling*, USA, pp. 460—469.

[12] Goodchild M., Davis F., Painho M., Stoms D., 1991, *The Use of Vegetation Maps and Geographic Information Systems for Assessing Conifer Lands in California*, NCGIA, Technical Report 91—23, USA.

[13] Gong P., 1992, "Boundary Uncertainties in Digitized Maps I: Some Possible Determination Methods", *Proceedings of GIS / LIS '92*, San Jose, California, pp. 274—281.

[14] Gong P., Zheng X., Chen J., 1995, "Boundary Uncertainties in Digitized Maps: An Experiment on Digitization Errors", *Geographic Information Sciences, Vol. 1, Issue 2*, pp. 65—72.

[15] Gong P., Mu L., 2000, "Error Detection through Consistency Checking", *Geographic Information Sciences, Vol.* 6, Issue 2.

[16] Hammond R. Mccullagh P. S., 1974, *QUANTITATIVE TECHNIQUES IN GEOGRAPHY: An Introduction*, Oxford University Press, William Clowes & Sons Limited, London.

[17] Hunka G.W., 1978, "Aided-Track Cursor for Improved Digitizing Accuracy.", *Photogrammetric Engineering & Remote Sensing*, 44(8): 1061—1066.

[18] Keefer B., Smith J., Gregoire T., 1988, "Simulating Manual Digitizing Error with Statistical Models.", *Proceedings of GIS-LIS '88*, ACSM, ASPRS, AAG and URISA, San Antonio, pp. 475—483.

[19] Kraus Y., Shoshani U., Doytsher Y., 2007, "Precise Coordinate LIS for Improving Planning and Land Registration Processes in Modern Real Estate Markets", *FIG Working Week 2007, Strategic Integration of Surveying Services*, Hong Kong SAR, China, 13—17 May.

[20] Leyk S. T., Weibel R., Boesch R., 2005, "A Conceptual Framework for Uncertainty Investigation in Map-based Land

Cover Change Modelling", *Transactions in GIS,* 9(3): 291—322.

[21] Maffini G., Arno M., Bitterlich W., 1992, "Observations and Comments on the Generation and Treatment of Error in Digital GIS Data.", *Accuracy of Spatial Databases (editors: Goodchild M., Gopal S.)*, NCGIA, Taylor & Fransis, England, pp. 55—67.

[22] Melling P., 1991, "Evolution of Digitizers at TDS-Numonics", *The Cartographic Journal*, Vol. 28, pp. 51—54.

[23] Nakos V., 1990, *Digital Imaging of cartographic phenomena based on the theory of Fractal Geometry. Application to the topographic relief with the use of digital models,* PhD. Dissertation, Dept. of Rural & Surv.Eng., NTUA, Athens (in Greek).

[24] Neter J., Wasserman W., Kutner M., 1990, *Applied Linear Statistical Models*, IRWIN, USA.

[25] Ore R A., 2001, "A Comparison of Digitizing: FastCAD, Macromedia FreeHand and PC Arc/Info", *International User Conference 2001*, Proceedings-Workshops, GIS ArcInfo-GIS ArcView, San Diego, USA.

[26] Otawat T., 1987, "Accuracy of Digitizing: Overlooked Factor in GIS Operations.", *Proceedings of GIS/LIS 87'*, San Fransisco, pp. 295—299.

[27] Robinsona A., Sale R., Morrison J., 1984, *Elements of Cartography*, Fifth Edition, John Wiley & Sons, USA.

[28] Thapa K., Bossler J., 1992, "Accuracy of Spatial Data Used in Geographic Information Systems", *Photogrammetric Engineering & Remote Sensing,* 58(6): 835—841.

[29] Warner W., Carson W., 1991, "Errors associated with a Standard Digitizing Tablet.", *ITC Journal*, 2, pp. 82—85.

[30] Waters R., Meader D., Reinecke G., 1988, "Data Capture for the Ninties: VTRAK.", *Proceedings of Ninth International Symposium on Computer Assisted Cartography,* Baltimore, Maryland, April 2—7, AUTOCARTO 9, pp. 377—383.

[31] Wenzhong S., Tempfli K., 1994, "Modelling Positional Uncertainty of Line Features in GIS.", *ASPRS / ACSM, Annual Convention & Exposition Technical Papers*, Nevada, April 25—28, pp. 696—705.

[32] Xenakis A. S., 1991, *Analysis of time series and predictions: analysis in the field of time*. Dept of Economic Sciences, University of Athens, Athens (in Greek).

[33] Zheng X., Gong P., 1997, "Linear Feature Modelling with Curve Fitting: Parametric Polynomial Techniques", *Geographic Information Sciences,* Vol. 3, Issue 1, pp. 7—19.