ERJ
Engineering Research Journal
Faculty of Engineering
Minoufiya University

ASSESSING THE ROLE OF POINT SAMPLE NEIGHBORHOODS IN DIGITAL ELEVATION MODELING

Dr. Raaed Mohamed Kamel Hassouna
Department of Civil Engineering, Faculty of Engineering in Shebin El-Kom, Minoufiya University, Shebin El-Kom, Egypt

ABSTRACT:
In this study, mathematical models for the self and relative neighborhoods of spatial point configurations have been proposed. This was intended for data/check elevation point patterns. These models were based on some realized point patterns within an illustrative rectangular geographical window. This window includes the Nile Valley, Eastern Desert and South Sinai in Egypt. The investigated data and check patterns had point elevation values, which were based on the SRTM30 global elevation model. The trends of the neighborhood models agreed with those of the interpolation quality, as expressed in terms of the standard deviation of discrepancies. Moreover, closed expressions were proposed, which relate the terrain interpolation qualities with the relevant neighborhood models. It is recommended to use such criterion for planning the spatial distributions of elevation point patterns in digital terrain modeling.

في هذا البحث تم اقتراح نماذج رياضية لحساب التوزيع المحلي الذاتي والنسبي لأنماط نقط البيانات والتحقيق المستخدمة في نمذجة مناسيب سطح الأرض. وقد تم عمل محاكاة لبعض أنماط توزيع النقط داخل نافذة مستطيلة تتضمن وادي النيل والصحراء الشرقية وجنوب سيناء في مصر. وأخذت مناسيب أنماط البيانات والتحقيق بناءا على النموذج الرقمي لسطح الأرض الطبيعي (SRTM30). وتمت دراسة العلاقة بين نماذج التوزيع النسبي لنقط البيانات والتحقيق، ودقة حساب نماذج سطح الأرض. وقد تطرق البحث الى اقتراح دالة تصف هذه العلاقة. ولذا، يوصى باستخدام هذا المدخل الرياضي في مقارنة أنماط التوزيع الجغرافي للنقط أثناء عمل النماذج الرقمية لمناسيب سطح الأرض.

KEYWORDS: Digital elevation modeling – point samples - neighborhood

1 INTRODUCTION

Digital terrain modeling (*DTM*) encounters discrete cases of both the elevation data and check samples. In particular, the distributions of both the data and check points affect the estimated quality [1], [6]. In this respect, the sparseness of the elevations' point samples plays a significant role [7]. Intuitively, the data and check sample neighborhoods govern such sparseness.

So, the objective of the current study is to propose mathematical models for computing the neighborhoods of spatial point patterns over a rectangular test window ($22° N \le \varphi \le 30° N;\ 30° E \le \lambda \le 36° E$). This geographical region includes the Nile Valley, Eastern Desert and South Sinai in Egypt; and a part of the Red Sea. The modeling strategy encounters both the self neighborhoods of single data or check patterns, and the relative neighborhoods among the data and check patterns. For this purpose, several types of point configurations are realized over the above test region. Such wide spectrum of point sample types may guarantee the reliability and applicability of the obtained results. Nevertheless, the relevant results could be generalized to any existing or planned point samples.

The elevations of the simulated data/check point configurations (with different point densities) are derived from the *30"x30"* global terrain model *SRTM30*. Such resolution corresponds to about 900 meters nominal spacing. The relevant elevations are based on the *EGM96*-derived geoid [11]. In the used version (2.0) of *SRTM30*, the oceans and seas

bathymetry is set to zero, whereas the lakes and rivers have computed elevations [11].

Series of terrain interpolations are performed; and the corresponding accuracies are evaluated at the check patterns. Finally, a closed mathematical expression is proposed, which relates both the self and relative neighborhoods of the data and check patterns to the resulting interpolation accuracy.

2 THE GENERATED POINT PATTERNS

In general, six types of point patterns were simulated, based on the theory of geometric probability [3], [10]. A main input for the simulation of a specific point pattern, within a given window, is the point density. Figure (1) to (6) show examples of the point pattern types. The plotted samples have a common average density of *3 (points /degree^2).*

The cluster sample is rather irregular. In a random pattern, the point locations are totally independent [10]. Such sample is more regular than a clustered one. The Gibbs sample has a varying point pair interaction [9]. It is relatively evener than a random one and is less regular than a simple inhibition pattern [2]. In a simple inhibition pattern, the points are added randomly to the window one-by-one, subject to the constraint that each new point must lie away from any existing point by at least some specific distance [3]. Both the Gibbs and simple inhibition point samples have considerable random features (see Figures (3) and (4)).

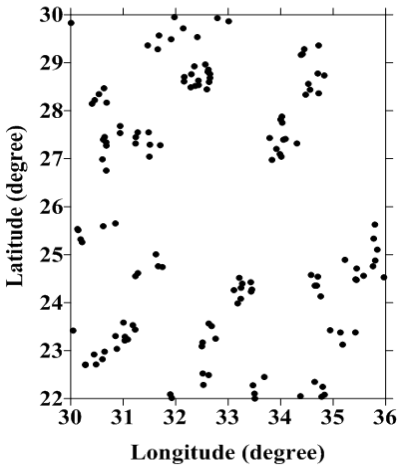

Figure (1): *Cluster sample (CT)*

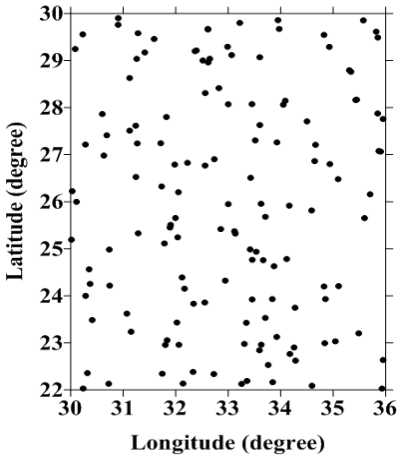

Figure (2): *Random sample (RN)*

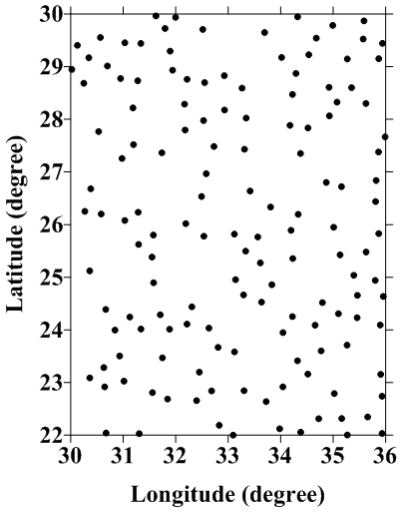

Figure (3): *Gibbs sample (GB)*

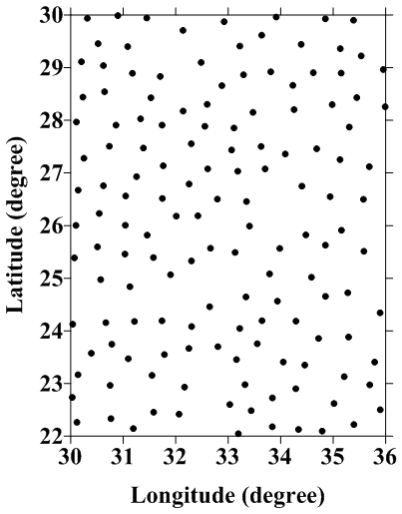

Figure (4): *Simple inhibition sample (IH)*

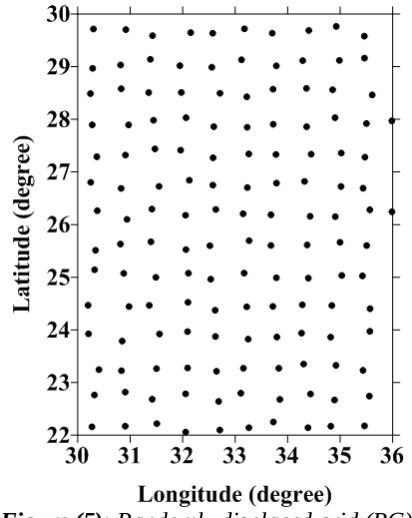

Figure (5): *Randomly displaced grid (RG)*

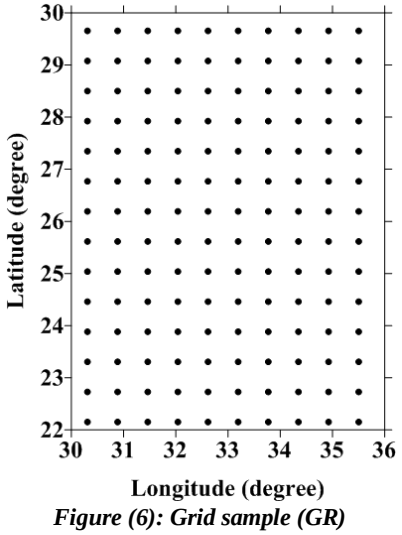

Figure (6): Grid sample (GR)

From Figures (5) and (6), the randomly displaced grid is more uniform than the above four types, whereas it is less regular than a grid pattern. It still exhibits slight randomness. Among all types, the grid is the most regular point pattern.

Such small point density (*3 points /degree^2)* is just illustrative, regarding the relevant easiness of visually comparing the various point configurations in Figure (1) to (6). In this study, however, much higher point densities are used in the assessment of the proposed point neighborhood models and in the digital terrain modeling procedures.

3 SELF AND RELATIVE NEIGHBOR-HOODS: PROPOSED MODELS

In what follows, the notion self neighborhood (*SNB*) will refer to the neighborhood of a single point sample. Alternatively, the relative neighborhood (*RNB*) will express the neighborhood of a point pattern with respect to another one.

Two models for assessing both the *SNB* and *RNB* are proposed. In particular, the *SNB* of a data or a check sample could be defined as the mean nearest neighbor distance over all points [8], [9]. Thus,

$$SNB = [\ \sum_{i=1}^{n} NND\ ]\ /\ n, \qquad (1)$$

where *NND* is the distance to the nearest neighbor and *n* is the number of the pattern points.

On the other hand, the *RNB* of a data pattern with respect to a check sample, $RNB_{D\text{-}K}$, could be defined as the mean of the nearest data neighbor distances to the check points, as follows

$$RNB_{D\text{-}K} = [\ \sum_{i=1}^{nk} NND_{D\text{-}K}\ ]\ /\ nk, \qquad (2a)$$

where $NND_{D\text{-}K}$ is the distance from a check point to the nearest data point, and *nk* is the number of check points. And the *RNB* of a check pattern with respect to a data one, $RNB_{K\text{-}D}$, is defined as the mean nearest check neighbor distances to the data points,

$$RNB_{K\text{-}D} = [\ \sum_{i=1}^{nd} NND_{K\text{-}D}\ ]\ /\ nd, \qquad (2b)$$

where $NND_{K\text{-}D}$ is the distance from a data point to the nearest check point, and *nd* is the number of data points. Computer software was prepared for assessing

both the *SNB* and *RNB* (in arc-degree) for all treated point samples, according to Eqs. (1) and (2).

4 ASSESSMENTS OF *SNB* AND *RNB*

Figures (7a) and (7b) show the *SNB* trends for the six types of point samples, having densities of *9* and *36* (points/degree^2), respectively. The similarity among the two trends is obvious, in spite of the relevant different *SNB* magnitudes.

Figure (8) illustrates the $RNB_{D\text{-}K}$ relevant to a common point density of *9*. Similar trends are resulting in Figure (9a), which corresponds to data/check point density of *36*. Again, the trends, but not the values, of the *RNB* seem to be independent on the point density. Figure (9b) shows the same features of Figure (9a), but in regard to the type of check patterns. It should be mentioned that the same *SNB* and *RNB* trends were observed for samples having point density of *3*.

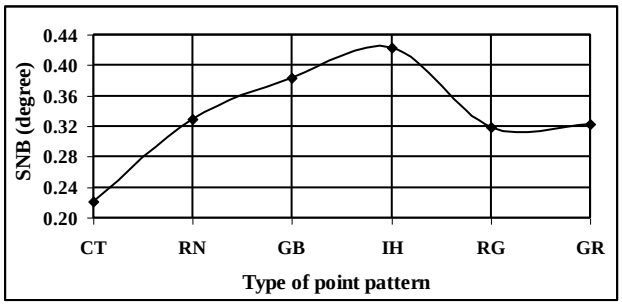

Figure (7a)

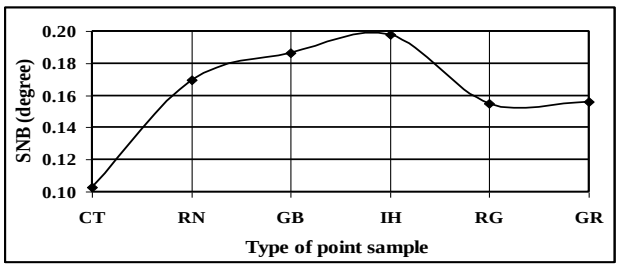

Figure (7b)

Figure (7): *SNB* of different sample types having densities of *9 and 36 (points/degree^2)*

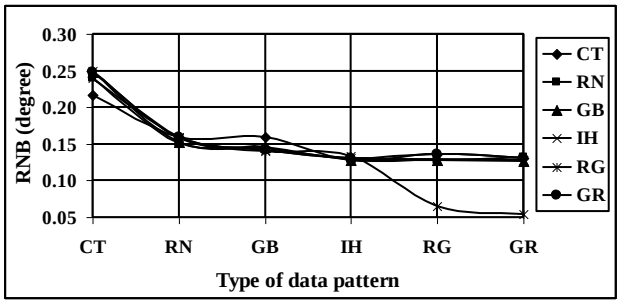

Figure (8): *RNB of data patterns with respect to check* patterns
(Density = *9 points/degree^2*)

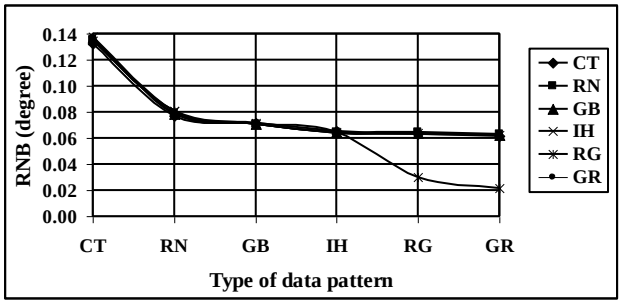

Figure (9a)

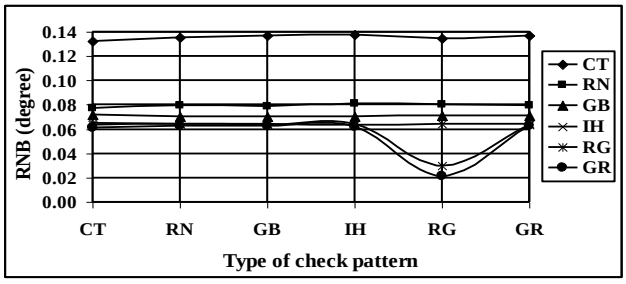

Figure (9b)

Figure (9): *RNB* of data patterns with respect to check patterns
(density = *36 points/degree^2*)

5 NEIGHBORHOOD AND *DTM* QUALITY

The elevations relevant to the above data and check patterns, having densities of 36, were derived from the *30"x30"* global terrain model *SRTM30* [11]. Such elevations were computed via the nearest neighbour interpolation [5]. In this version of *SRTM30*, the Red Sea elevations are set to zero. The elevations in the investigated region exhibit extreme variations, which serves as a general rough terrain case. Using such point density, the nominal data/check point resolution corresponds to *10* arc-minutes.

Series of terrain interpolations were performed. The interpolations were firstly conducted from the data patterns to the nodes of a *5'x5'* grid, and then back to the check patterns. In this respect, the local Kriging, the local Inverse Distance (to power *2*) and the Minimum Curvature interpolation methods were used. The local techniques utilized the nearest twelve data points [4]. The relevant accuracies, in terms of the standard deviations of discrepancies, were evaluated at the check patterns. Figures (10), (11) and (12a) show the resulting accuracy trends, corresponding to the three methods, respectively. It is obvious that the trends of the three methods resemble to a great extent those in Figure (9a). Moreover, Figure (12b) shows the standard deviation trends, relevant to the Minimum Curvature method, as related to the type of check patterns. Also, such trends are correlated with those in Figure (9b).

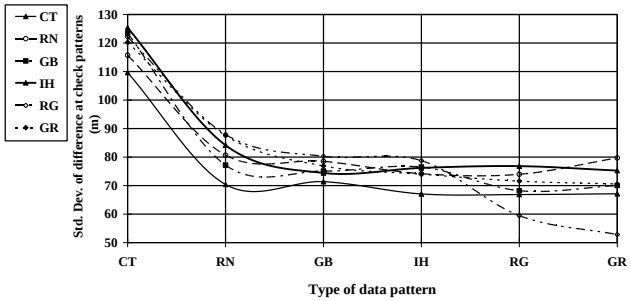

Figure (10): *Interpolation accuracies (Local Kriging)*

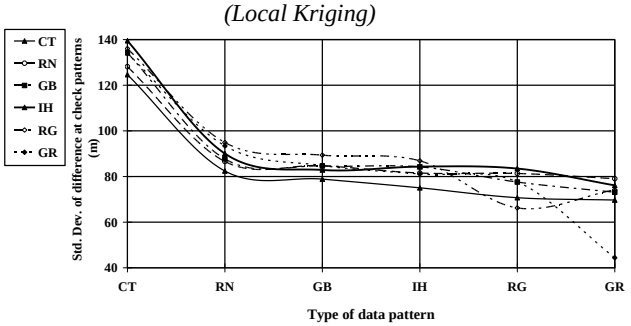

Figure (11): *Interpolation accuracies (Local Inverse Distance to power 2)*

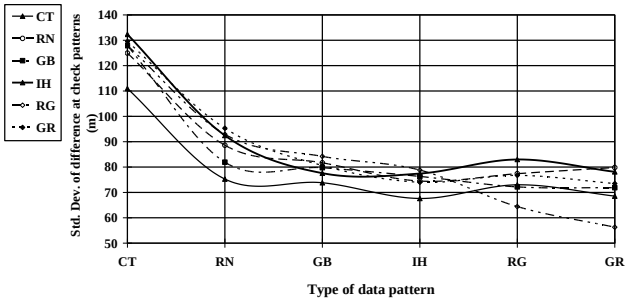

Figure (12a)

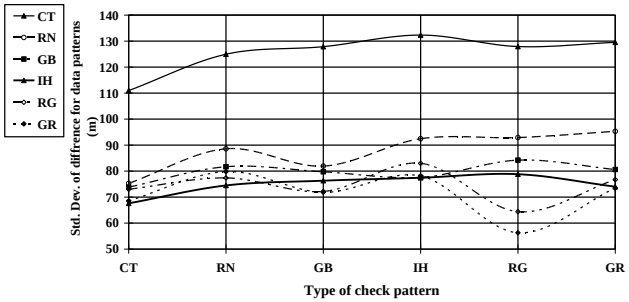

Figure (12b)

Figure (12): *Interpolation accuracies (Minimum Curvature)*

6 PROPOSED EXPRESSION

Based on Sections (4) and (5), it could be noticed that the *DTM* accuracy is greatly dependent on the neighborhoods of both the data and check patterns. So, a mathematical expression will be proposed, which relates such features to the *DTM* quality. A linear quality function *y* is introduced, which accounts for the *SNB* of both the data and check patterns; the RNB_{D-K} and the RNB_{K-D}, as follows

$$y = A \,.\, SNB_D + B \,.\, SNB_K + C \,.\, RNB_{D-K} + D \,.\, RNB_{K-D}\,, \qquad (3)$$

where SNB_D and SNB_K are the self neighborhoods of the data and check patterns, respectively. And the coefficients *A* to *D* are four unknown parameters. The above mathematical function was fitted to the standard deviation series resulting from the three methods. For each technique, the linear formula in Eq. (3) was applied to the relevant *36* values of the standard deviations. For each case, there exist *36* equations in four unknowns,

$$\boldsymbol{Y}_{36,1} = \boldsymbol{E}_{36,4} \;.\; \boldsymbol{X}_{4,1}, \qquad (4)$$

where $\boldsymbol{X}$ is the vector of coefficients and $\boldsymbol{Y}$ is the vector of standard deviations. The coefficient matrix $\boldsymbol{E}$ carries the *SNB* and *RNB* values, after being converted to kilometer units, using a mean Earth's radius of *6370 km*. A least-squares algorithm, with equal weights, was applied to solve for the vector of coefficients,

$$\hat{\boldsymbol{X}}_{4,1} = (\boldsymbol{E}^T \boldsymbol{E})^{-1}_{4,4} \,.\, (\boldsymbol{E}^T \boldsymbol{Y})_{4,1}. \qquad (5)$$

Table (1) lists the estimated coefficients relevant to each interpolation technique, along with their standard errors. The three methods give the same signs of the estimated coefficients, which tend to agree in their

magnitudes. In particular, an increase in SNB_D, SNB_K and RNB_{D-K} leads to a de-gradation in the resulting *DTM* quality. Conversely, the larger is the RNB_{K-D}, the better is the resulting prediction quality.

Figures (13a) and (13b) show the quality trends relevant to the Minimum Curvature method, as estimated from Eq. (3). Obviously, such trends mirror those in Figure (12a) and (12b), respectively, in terms of the neighborhood features.

Table (1): *The estimated functional parameters for the three interpolation techniques*

Method & Variables	Coefficient	Value (m/km)	Standard error (m/km)
Local Kriging			
SNB_D	*A*	0.968	0.257
SNB_K	*B*	0.731	0.269
RNB_{D-K}	*C*	6.796	0.316
RNB_{K-D}	*D*	-0.584	0.308
Local Inverse Distance			
SNB_D	*A*	1.022	0.463
SNB_K	*B*	0.785	0.485
RNB_{D-K}	*C*	7.531	0.570
RNB_{K-D}	*D*	-0.668	0.557
Minimum Curvature			
SNB_D	*A*	0.924	0.292
SNB_K	*B*	0.854	0.306
RNB_{D-K}	*C*	7.203	0.360
RNB_{K-D}	*D*	-0.691	0.352

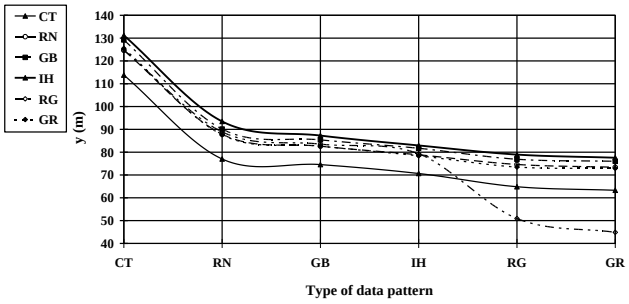

Figure (13a)

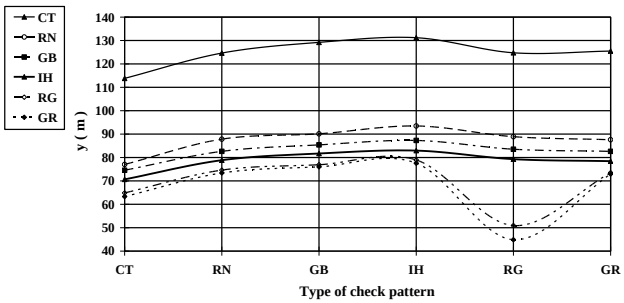

Figure (13b)

Figure (13): *Model computed accuracies (Minimum Curvature)*

7 MODEL'S VERIFICATION

Now, the validity of the proposed mathematical expression will be verified, regarding its ability to express the expected terrain interpolation accuracy in terms of the *SNB* and *RNB* features. In particular, the above three methods was used for interpolating terrain elevations using two pairs of data and check samples, which have arbitrarily different distributions and densities. The elevations of such samples were derived from the *SRTM30* model.

The first data and check sample pair have point densities of *250* and *100*, respectively. The relevant data/check *SNB* and *RNB* features were computed. Such data has a nominal resolution of about *3.8* arc-minutes. The interpolations were firstly performed into a *4'x4' DTM*, and then backward into the check points. Table (2) lists a comparison among the model quality, based on the coefficients in Table (1) and Eq. (3), and the actual standard deviation of differences.

The second data/check pattern pair has densities of *300* and *3*, respectively. Such data has a nominal resolution of about *3.5* arc-minutes. Similarly, interpolations were performed into a *4'x4'* grid, and back into the check points. Table (3) compares the model quality *y* with the actual quality at the check points for this case.

Finally, the three interpolation methods were used to directly interpolate the full *30"x30" SRTM30* elevation grid over the test window. An independent set of *705* scattered check points were used. This corresponds to data and check point densities of *14400* and *15*, respectively. Table (4) shows a comparison among the model and actual accuracy.

Table (2): *Model and actual quality relevant to the data and check samples with densities of 250 and 100, respectively*

Method	*y* (m)	Actual standard deviation (m)
Local Kriging	**52.797**	**50.189**
Local Inverse Distance	**54.672**	**55.864**
Minimum Curvature	**50.975**	**53.397**

Table (3): *Model and actual quality relevant to the data and check samples having densities of 300 and 3, respectively*

Method	*y* (m)	Actual standard deviation (m)
Local Kriging	**51.850**	**52.850**
Local Inverse Distance	**55.006**	**57.487**
Minimum Curvature	**56.989**	**56.279**

Table (4): *Model and actual quality based on the SRTM30 data grid and 705 independent check points*

Method	*y* (m)	Actual standard deviation (m)
Local Kriging	**24.132**	**25.794**
Local Inverse Distance	**25.942**	**26.169**
Minimum Curvature	**28.091**	**26.046**

8 CONCLUDING REMARKS

Based on the current study, it can be concluded that the *DTM* quality are greatly dependent on the *SNB* of both the data and check patterns; and the *RNB* among them. Regarding the three investigated prediction techniques, the proposed closed expression appropriately describes the terrain interpolation quality. So, it is recommended to apply the *SNB* and *RNB* criteria for the planning and pre-analysis of the distributions of elevation point patterns in digital terrain modeling.

ACKNOWLEDGMENTS The two reviewers are acknowledged for the critical review of the manuscript.

9 REFERENCES

[1] Aguilar; F.; Agüera, F.; A. Aguilar, M. and Carvajal, F. (2005): "Effects of Terrain Morphology, Sampling Density, and Interpolation Methods on Grid DEM Accuracy", *Photogrammetric Engineering & Remote Sensing*, Vol. 71, No. 7, July, pp: 805–816.

[2] Allard, D. (1995): "Non Parametric Maximum Likelihood Estimation of Features in Spatial Point Processes using Voronoi Tessellation", *Department of Statistics, University of Washington Seattle*, Washington, USA, August 22.

[3] Baddeley, A. and Turner, R. (2005): "spatstat: An R Package for Analyzing Spatial Point Patterns", *Journal of Statistical Software*, Vol. 12, No. 6, pp: 1 - 42.

[4] Cimmery, V. (2010): "User Guide for SAGA (version 2.0.5), Volumes 1 and 2", August, SAGA System for Automated Geoscientific Analyses, available at: http://www.saga-gis.org (accessed May 2011).

[5] Golden Software, Inc. (2011): "Surface Mapping System, Surfer Version (10.1.561)", *Golden, Colorado*, Mar 1.

[6] Li, Z.; Zhu, Q. and Gold, C. (2005): "Digital Terrain Modeling: Principles and Methodology", *CRC Press*, Florida.

[7] Oksanen, J., (2006): "Digital Elevation Model Error in Terrain Analysis", *Ph.D. Thesis, Faculty of Science, University of Helsinki*.

[8] Schumacher, D. (2005): "Estimation of Distances between Point Process Distributions", *Dissertation (Dr. sc. nat.), Fakultät für Mathematisch-naturwissenschaft, Universität Zürich*.

[9] Stoyan, D. and Penttinen, A. (2000): "Recent Applications of Point Process Methods in Forestry Statistics", *Statistical Science*, Vol. 15, No. 1, pp: 61-78.

[10] Stoyan, D. and Stoyan, H. (1994): "Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics", *John Wiley & Sons Ltd*, Baffins Lane, Chichester, England.

[11] USGS (2006): "SRTM30", available at: http://dds.cr.usgs.gov/srtm/, (Accessed 31 May 2012).