Nearest Tri-Points Interpolation (NTPI) Technique

Arkan Radhi Ali

Civil Department, Technical Institute of Babylon, AL-Furat AL-Awsat Technical University, Babylon, Iraq

inb.arkn@atu.ed

Submission date:- 3/5/2018	Acceptance date:- 3/6/2018	Publication date:- 2/9/2018

Abstract

Numerous methods have been proposed to interpolate and extrapolate of missing temporal data series such as rainfall, temperature, humidity, runoff storms…etc. Whereas there are many of other method for estimating spatial missing data such as groundwater levels and topography. On the other hand,many methods were developed to evaluate their efficiencies,but the uncertainty of results is rarely calculated. In this study, three interpolation methods have been compared to estimate missing spatial topographical data in the ancient Babylon City. The model domain was discretized into a number of horizontal (19) cells and vertical (23) cells. Five observed elevation remarks were used to estimate the unknown elevations of 257 remarks. The new method namely Nearest Tri-Point Interpolation (NTPI) was compared with the Inverse Distance Weighting (IDW) and surfer techniques. The efficiency of these techniques was calculated by the Average Error (AE) and Standard of Deviation (SD). The (NTPI) technique offers results of less AE and SD as well as more accuracy in ground surface elevations distribution

Keywords: Nearest Tri-Points Interpolation (NTPI), Inverse Distance Weighting (IDW), Average Error (AE), Standard of Deviation (SD).

Introduction

Interpolation may be defined as a mathematical technique to determine missing spatial in a certain location (boundary space) or temporal data (historical time series). The spatial interpolated data should be intended to be the best fit to the reality. Inverse Distance Interpolation (IDW) is based on the estimation that the characteristics value of unknown point is the weighted average of measured point value within the area, and the weights inversely connected to the distances between the prediction point and measured point [1] .The surfer techniques are a software developed golden company (USA), which includes twelve interpolation methods for different needs, one of them is Nearest Neighbor Method, this method assigns the nearest point value to each grid node [2]. The selected method should be referenced corresponding to accuracy assessment [3]. The comparison of the error of estimation obtained at 5 sites, the results show that the optimal interpolation and kriging methods are fitter than other. The IDW and Thiessen polygon shown approximately satisfactory results, while PI did not show good predictions [4].

Interpolation algorithm criteria selection is highly based upon actual and accurate data and the required accuracy. Unfortunately, in engineering practice a gap in historical data usually encountered and needs to be filled with very approximated to the reality [5]. The power of one of IDW was the fit choice, which may due to the low drift of the interpolation of soil properties [6]. The IDW technique was effective to predict 50 % and 65 % of the exact positions of the 20 higher and lower measurements respectively [7]. On the other hand, [1] concluded that inverse distance weighting technique is more important than ordinary kriging and constant parameter technique in many cases. [8] Indicated there are numerous interpolation methods to fill a time series gaps, uncertainty quantification and efficiency criteria namely; interpolation, regression, autoregressive and machine learning and many others to estimate their efficiencies. [9] outlined that good interpolation technique should have adequate known data, estimation missing data, efficient and fast, capable of applicable for all type of data and also be accurate and robust. The NTPI represents a simplified technique basically based on the IDW method, but differs in a methodology of estimation. Spatial missing data in IDW are usually determined depending on all known remarks. But in NTPI the estimated data of missing remark are determined by the nearest tri points.

1. Location of Study Area

The area of Ancient Babylon City is selected to be a sample of the study. The area lies in the Mesopotamia alluvial plain [10]. It is characterized with a rugged and folded nature. It is located between Longitudes (44˚ 24′ 45″-44˚26′ 15″) and latitudes (32˚ 31′ 30″-32˚ 32′ 51″). It is bounded by Al Hillah River from the West, whereas an artificial ditch is located to the east. Fig.(4) represents a location map of the study area. The area undertaken in this mathematical simulation is about 17,562,500m^2 in size. The maximum and minimum ground levels are about 56m and 32m above mean sea levels respectively.

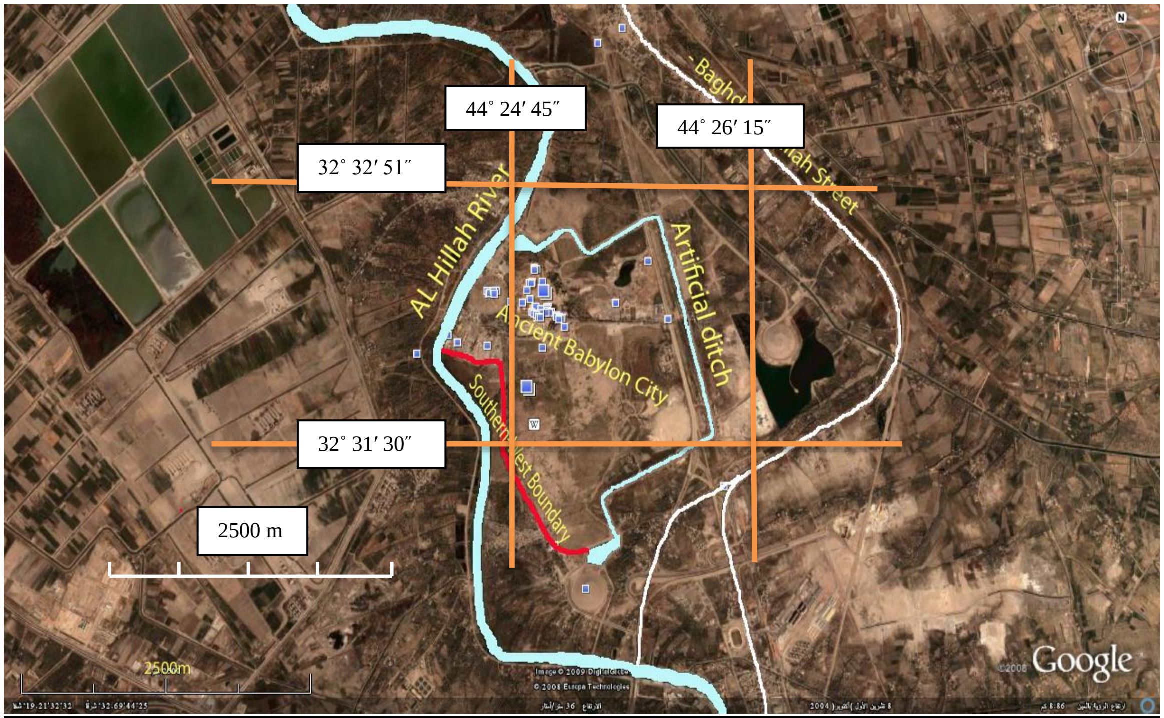

Fig. (1) Location Map

2. Material and Methods

2.1 Mathematical Background

The general formulas of Inverse Distance Weighting method as outlined by [11]is:

$$z_{ij} = \sum_{i=1}^{N} \lambda_i z_{0ij} \qquad \text{.........(1)}$$

$$\lambda_{ij} = \frac{d_{0i}^{-p}}{\sum_{i=1}^{N} d_{0i}^{-p}} \qquad \text{........(2)}$$

$$d_{0i} = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \qquad \text{........(3)}$$

Where:

z_{ij} : predicted elevation for node (i,j).

z_{0ij}:observed value for node (i, j).

N: Number of measured nodes within the modeled domain.

λ_{ij} : Assigned weight for each measured node.

d_{0i} : distance from the unknown elevation cell to the known elevation cell.

$(x,y)_0$,$(x,y)_i$: coordinates of interpolation point and dispersion point.

The power parameter p extremely affects the weighting of the predicted elevation on the bases of the location's value, that is, as the distance between the measured location and the predictive one increases, the weight will exponentially decrease, [12].

The validity of the fitted models was checked on the basis of effective tests. In this method, called jack-knifing procedure, interpolation is performed at all the data locations, ignoring, in turn, each one of them one by one. The differences between estimated and observed values are summarized using cross-validation statistics [13].

2.2 Statistical Evaluation

The differences between estimated and observed values are summarized by using the cross-validation statistics: Average prediction Error (AE), Standard of Deviation (SD). The summary statistics should meet the following criteria [14], [15], [16]: -

$$\boldsymbol{AE} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1j=1}}^{\boldsymbol{N}} (\boldsymbol{Z_{ij}} - \hat{Z}_{ij}) \qquad \text{………… (4)}$$

$$\boldsymbol{SD} = \sum_{\boldsymbol{i=1j=1}}^{\boldsymbol{N}} \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1j=1}}^{\boldsymbol{N}} (\boldsymbol{Z_{ij}} - \hat{Z}_{ij})^{2}} \qquad \text{………… (5)}$$

Where $\hat{Z}_{ij}$ is a measured elevations (*observed*).

Note: measured elevations were obtained by field checking after NTPI map drawing.

Definition: $\boldsymbol{z_{0ij}}$ ***and*** $\hat{Z}_{ij}$ *are both measured elevations but there is a difference in between. The first elevations were used for predetermined the missing data elsewhere. Whereas the second elevations were used for statistical evaluation of the considered methodologies.*

2.3 Discretization of the Area

The considered area, in case of good levels evaluation should initially be surrounded by a hypothetical boundary. To achieve this, the domain was discretized into a number of horizontal and vertical cells. Briefly, the number of cells in x direction was selected to be 19 and in y direction is 23 as shown in Fig. (2).

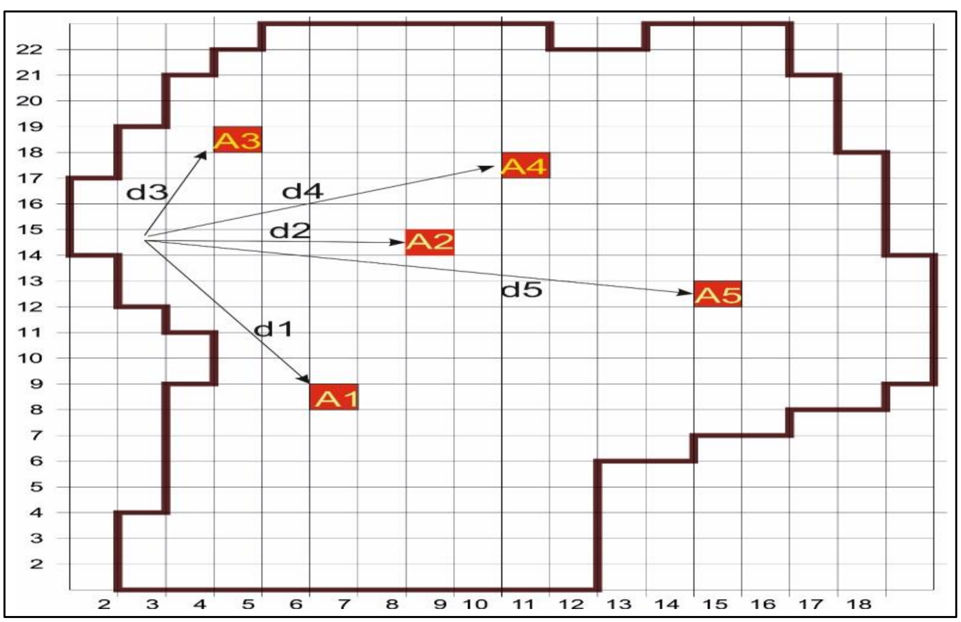

Fig. (2) Mesh Design, Known Cell Elevations and Distances of Unknown Cell Elevation

2.4 Base map Implementation

In order to draw the surrounding boundary, the x and y ordinates should be assigned to all cells that traced the boundary. The coordinated values may be stored in a specified file under extension of (.bln).

2.5 Method of calculation by (IDW)

The methodology of elevation levels estimation corresponding to the IDW methods of (Equations.1, 2 and 3) requires to estimate the distances $(d_1\ d_2\ d_3\ d_4\ and\ d_5)$ for each node of unknown level within the model domain. The previous distances and the observed elevations on the basis of five nodes of known elevations namely; $(A_1\ A_2\ A_3\ A_4\ and\ A_5)$ as shown in Fig. (2) are presented in Table (1).

Table (1) Known Elevations (Observed)

Cell	X	Y	Ẑ
A_1	7	9	48
A_2	9	14	38
A_3	5	19	56
A_4	11	18	33
A_5	15	13	34

2.6 Nearest Tri-Points Interpolation (NTPI) Concept

The NTPI is slightly differed from IDW that the last method uses all known cell elevations to estimate the missing data inside the required rang. Theoretically this concept is true but not exactly real since the missing data is only relevant to surrounding unknown cells data for all types of spatial models such as pressure, temperature, topography, GW levels…etc. The suggested technique in NTPI is that the missing data should be depended on the surrounding nearest three known points, Table (2). Figure (3) presents the procedure of data prediction for unknown cells.

1- Step1: the elevation of cell (7,18) was estimated by using (Eqs. 1, 2, 3) depending upon the nearest tri points of known cells elevations (A2, A3, A4) and distances ($d_{1,}\ d_{1,}\ d_3$).

2- Step2: The elevation of the cell (6, 16) is estimated similarly depending upon the known nearest tri points elevations of (A2, A4, A3) and distances ($d_{4,}\ d_{5,}\ d_6$).

3- Step3: the nearest tri points known cells elevations of (A1, A2, and A3) and distances (d7, d8, d9) obtain the elevation of the cell (6, 13).

4- Step4: The elevation of Cell (11, 11) similarly estimated by known nearest tri points (A5, A2, A1) and distances (d_{10}, d_{11}, d_{12}).

5- The procedure may be circulated for other cells of missing data within the domain of Fig. (3).

***Comment1**: The calculation technique initially and simultaneously requires to delineate the single missing data cell and radially move to locate the nearest three known points which constantly essential for NTPI estimation. Then the previous steps of estimation (1 to 5) may be followed.*

***Comment2**: The IDW procedure mostly correct for interior interpolation but it is slightly untrue for exterior interpolation. This drawback is usually encountered in engineering practice. Whereas by the NTPI estimation, one can interpolate and extrapolate all missing data inside and outside the considered range.*

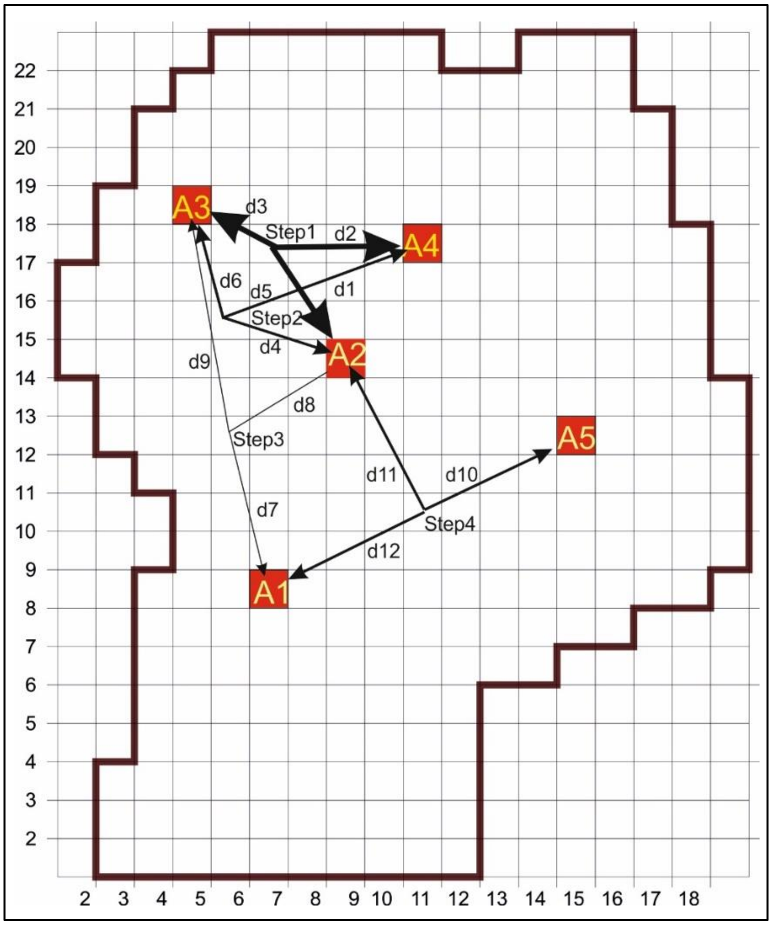

Fig. (3) Estimation Procedure via NTPI Method

Table (2) Sample of NTPI Cells Elevations Calculations

Cell NO.	Coordinates		Distances			Weights			Nearest Known Elevations			NTPI	IDW	surfer
1	x	y	d1	d2	d3	λ1	λ 2	λ 3	Z1	Z2	Z3			
2	10	15	1.5	3.3	5	0.77124	0.15934	0.06941	38	33	34	36.9256	38.0183	36.39
3	10	14	1	1	4	0.48484	0.48484	0.03030	38	36.92	34	37.3578	38.1740	36.77
4	11	14	1	1.5	3.9	0.66216	0.29429	0.04353	37.35	36.92	34	37.0776	38.0274	35.68
5	9	13	1	6	4.5	0.92836	0.02578	0.04584	38	34	48	38.3553	38.4857	39.429
6	10	13	1	5	4.7	0.92143	0.03685	0.04171	38.35	34	48	38.5921	38.5931	37.82
7	8	14	1	5	5.7	0.93389	0.03735	0.02874	38	48	56	38.8909	38.5745	41.12
8	8	13	4	1.5	6.8	0.11824	0.84084	0.04091	48	38	56	39.9188	39.1923	41.77
9	9	15	3.5	1	5.6	0.07331	0.89805	0.02863	33	38	56	38.1489	38.3096	38.42
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
253	17	9	9.4	9.9	4.2	0.14470	0.13045	0.72483	38	48	34	36.4052	40.8029	44.32
254	18	13	9	8.5	3.2	0.09968	0.11176	0.78855	38	33	34	34.2869	40.1454	44.07
255	18	12	9.1	9.4	3.2	0.09975	0.09349	0.80674	38	33	34	34.3055	40.4062	44.37
256	18	11	9.4	9.8	3.5	0.10949	0.10073	0.78977	38	33	34	34.3372	42.8252	45.12
257	18	10	9.6	10.6	4.2	0.14195	0.11643	0.74161	38	33	34	34.4513	40.3433	41.89

2.7 Earthly settings

40 points were selected in the plotted area to measure their real elevations (*measured elevations and then becomes an observed elevations*) by levelling device. These elevations were used to calculate the differences between them and the elevations of the three methods, Table (3).

Table (3) differences between 40 measured elevations and the elevations of three applied Methods at same cells (diff *=differences)

NTPI	IDW	surfer	Measured elevations	NTPI diff *.	IDW diff *.	Surfer diff*.
36.92561667	38.01833793	36.39	36.8	0.125 616674	1.218337931	-0.41
37.35786667	38.17408919	36.77	37.4	-0.042133333	0.774089189	-0.63
37.07760993	38.02740435	35.68	37	0.077609926	1.027404347	-1.32
38.35530086	38.48576655	39.429	38.4	-0.04469914	0.08576655	1.029
38.59219725	38.59314306	37.82	38.57	0.02219725	0.023143055	-0.75
38.89095592	38.57459078	41.12	38.8	0.090955918	-0.22540922	2.32
39.91889803	39.19232886	41.77	40	-0.081101968	-0.807671137	1.77
38.1489118	38.30968195	38.42	38.2	-0.051088202	0.109681946	0.22
39.40563991	38.68460936	41.28	39.3	0.105639913	-0.615390642	1.98
43.07156192	42.45850655	45	43	0.071561917	-0.541493446	2
41.6264401	40.80294312	44.32	41.5	0.1264401	-0.697056885	2.82
41.25913862	40.14547592	44.07	41.2	0.059138617	-1.054524083	2.87
40.28895987	40.40621576	44.37	41.5	-1.211040128	-1.093784235	2.87
42.18855962	42.82524695	45.12	42.05	0.13855962	0.775246948	3.07
40.05229553	40.34336706	41.89	40	0.052295531	0.343367056	1.89
38.1304969	38.5617905	38.81	38.2	-0.069503098	0.361790497	0.61
35.38975946	37.15154531	36.25	35.28	0.109759457	1.87154531	0.97
34.54722149	36.22557011	34.43	34.45	0.097221485	1.775570108	-0.02
35.78337602	37.500008	34.98	36	-0.216623985	1.500008005	-1.02
36.66666667	38.02740435	35.68	36.6	0.066666667	1.427404347	-0.92
36.59365527	38.36668099	36.66	36.5	0.093655271	1.866680992	0.16
40.91412857	41.18593576	42.82	41	-0.085871426	0.185935763	1.82
39.86086957	40.17719632	40.79	40	-0.139130435	0.177196322	0.79
39.42072538	39.73289063	39.12	39.5	-0.079274615	0.232890629	-0.38
38.76505208	39.0367819	37.82	38.7	0.06505208	0.336781898	-0.88
46.3928804	45.1	47.58	46.4	-0.007119597	-1.3	1.18
44.69743239	43.15438241	47.01	44.5	0.197432393	-1.345617588	2.51
44.24015328	42.277	46.69	44.2	0.04015328	-1.923	2.49
43.88641961	42.077	46.73	43.5	0.386419607	-1.423	3.23
45.12805797	42.418	47.21	45	0.128057974	-2.582	2.21
46.49783341	45.39	48.14	46.3	0.197833406	-0.91	1.84
48.54772244	50.077	49.44	48	0.547722435	2.077	1.44
44.87180323	46.155	46.17	44.6	0.271803233	1.555	1.57
40.55341503	41.488	42.67	40.2	0.353415026	1.288	2.47
37.62268368	37.913	39.25	37.5	0.122683683	0.413	1.75
35.17090025	35.257	36.14	35	0.170900246	0.257	1.14
33.37727743	34.073	33.81	33.33	0.047277427	0.743	0.48
33.45263455	34.55	32.87	33.3	0.15263455	1.25	-0.43
34.29968273	35.548	33.4	34	0.299682734	1.548	-0.6
35.10210146	36.805	34.04	35	0.102101464	1.805	-0.96

3. Results and Discussion

IDW is usually used to extract missing levels in GW, which does not suffer a significant change in its levels [14], while the NTPI can be used to find the missing levels of GW as well as the ground surface and irregular surfaces because it depends on the nearest three points.

The interpolated elevations of (257 cells) on the basis of IDW method and the known cells elevations (A_1 A_2 A_3 A_4 and A_5) are represented graphically in Fig. (4). whereas the obtained interpolated elevations by surfer on the bases of the same cells are shown in Fig. (5). The interpolated cells elevations of Table (2) and Fig. (6) were obtained by the NTPI method.

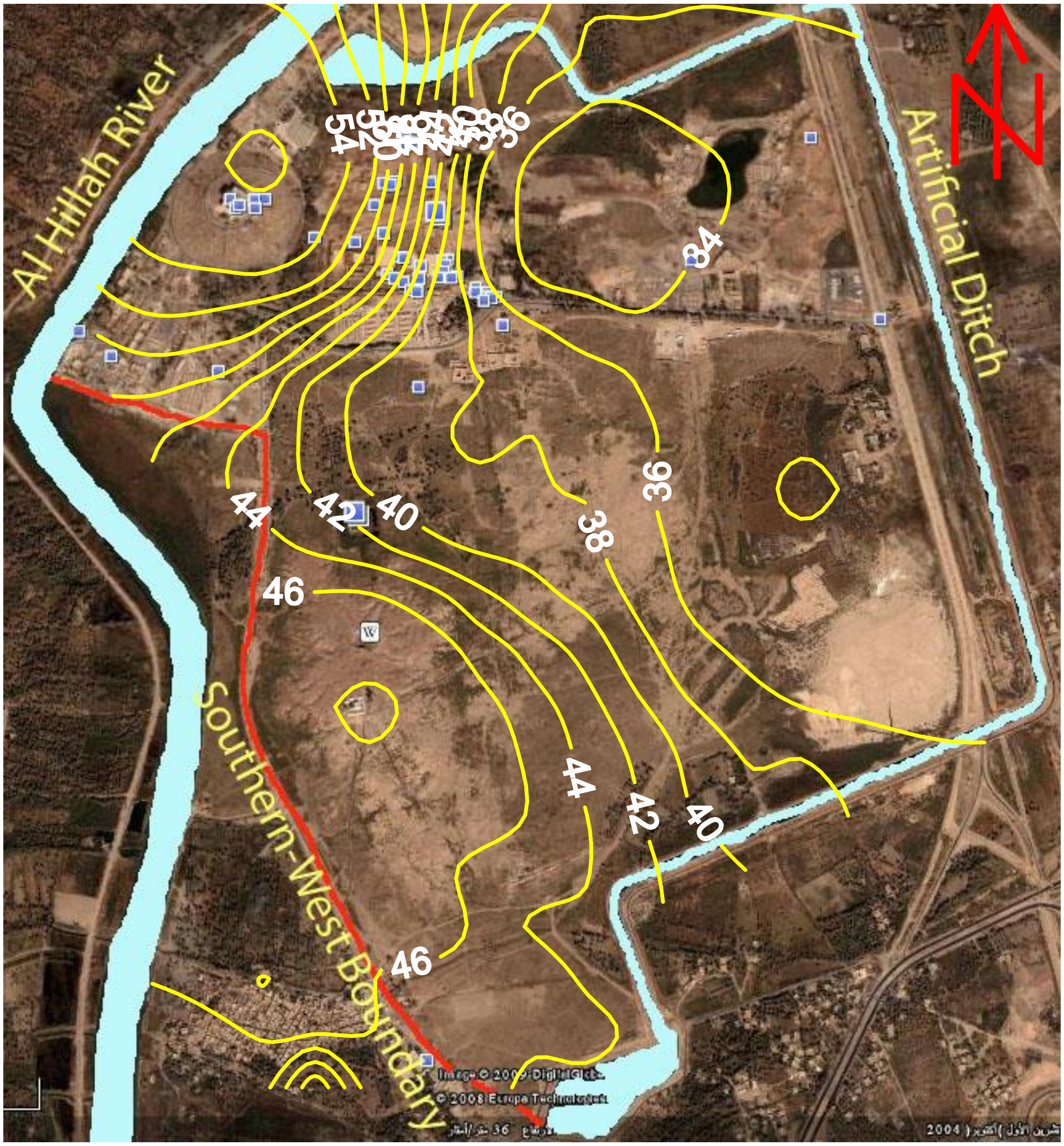

Fig. (4) Topographic Map based on IDW elevations Results

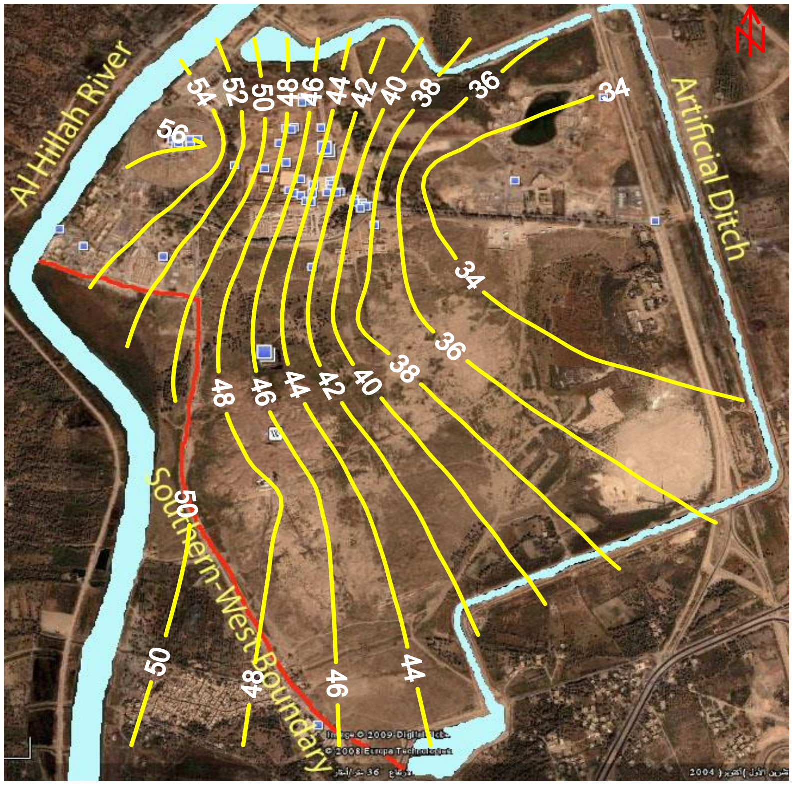

Fig. (5) Topographic Map based on Surfer Values

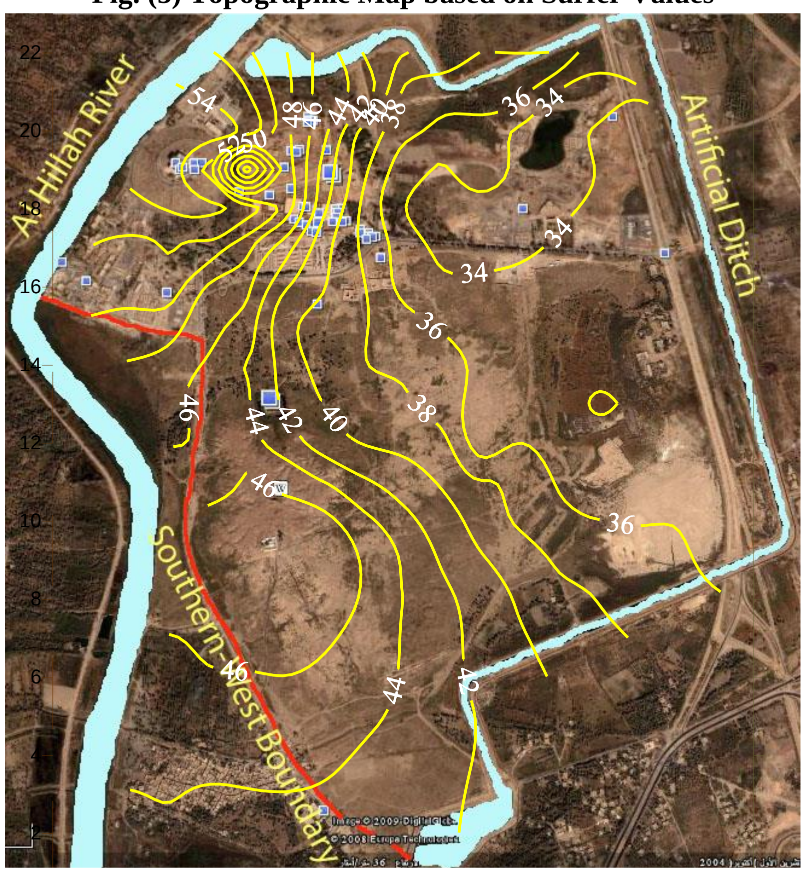

Fig. (6) Topographic Map based on NTPI Results

4. Statistical Evaluation

The Average Error (AE) and Standard Deviation (SD) of Eqs. (4 and 5) were used to evaluate the three methods. The AE and SD were estimated by using 40 measured points within the area, Table (3). It is found that NTPI offered the minimum values of AE and SD as indicated in Table (4).

Table (4) AE and SD results

Method	AE	SD
NTPI	0.057323	0.255239
IDW	0.262747	1.182729
Surfer	1.029475	1.697135

5. Conclusions

It is concluded that the NTPI method offers: -

1- More topographic details and accurate results relevant to reality.
2- Less AE and SD values to anonymous its priority.
3- The best fit for both interpolation and extrapolation techniques.
4- More accuracy in distribution of ground surface elevations.

References

[1] George Y.L., David W.W., " An adaptive inverse-distance weighting spatial interpolation technique", Journal of Computers & Geosciences, Vol.34, Issue 9, September 2008, pp.1044-1055, 2008.

[2] Chin S.Y., Szu P. K., Fen B. L., Pen S. H."Twelve Different Interpolation Methods: A Case Study of Surfer 8.0", Proceedings of the XXth ISPRS Congress, Vol.35, pp. 778-785, 2004.

[3] Azpurua M., Dos R. K."A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude". Progress in Electromagnetic Research M, 2010, 14, pp.135–145, 2010.

[4] Tabios G. Q. and Salas J. D."A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation",Journal of the American Water Resources Association, Vol.21, Issue 3, pp. 365-380,1985.

[5] Beveridge, S."Least squares estimation of missing values in time series". Commun. Stat. Theory Methods, 21, pp.3479–3496, 1992. [CrossRef].

[6] Robinson T.P., Metternicht G."Testing the performance of spatial interpolation techniques for mapping soil properties", Computers and Electronics in Agriculture, Vol. 50, Issue 2, pp.97-108, Feb.2006.

[7] Tomczak M." Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach", Journal of Geographic Information and Decision Analysis, Vol. 2, No. 2, pp. 18-30, 1998.

[8] Mathieu L., Jean-Baptiste A., and François H.L."Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment" Water, 9, 796; doi: 10.3390/w9100796, 2017.

[9] Brubacher, S. R.; Tunnicliffe W."Interpolating time series with application to the estimation of holiday effects on electricity demand", J. R. Stat. Soc. Ser. C 1976, 25, pp.107–116, 1976. [CrossRef].

[10] Arkan Radhi Ali," Vertical Hydraulic Conductivity of Unsaturated Zone by Infiltrometer Analysis of Shallow Groundwater Regime (KUISG)", Journal of University of Babylon for Engineering Sciences, Vol.26, No.4, pp.185-194, Feb.2018.

[11] Agnieszka K., Antoni G."Comparison of Deterministic Interpolation Methods for the of Groundwater". Journal of Ecological Engineering**",** Vol.15, No.4, pp. 55–60 DOI: 10.12911/22998993.1125458, Oct. 2014.

[12] Johnston K., Ver H.J.M., Krivoruchko K. "ArcGIS 9. Using ArcGIS Geostatistical Analyst ESRI", 2003. (User book).

[13] Kumar V., Remadei."Kriging of groundwater levels – a case study", Journal of Spatial Hydrology, Vol.6, No.1, pp.81–92, 2006.

[14] Krivoruchko K."Introduction to Spatial Data Analysis in GIS", ESRI Press, 2006.

[15] Luo W., Taylor M. C. and Parker S. R." A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales", International Journal of Climatology, Int. J. Climatol. 28: 947–959 (2008), Published online 2 August 2007 in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/joc.1583,2007.

[16] Li J, Heap A.D."A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors", Ecological Informatics 6: 228-241, 2011.

تقنية الاستكمال لأقرب ثلاث نقاط

اركان راضي علي

الخلاصة

تم أقتراح العديد من الطرق للاستكمال الداخلي والخارجي لسلسلة البيانات الزمنية المفقودة مثل هطول الأمطار ودرجة الحرارة والرطوبة والجريان السطحي ...الخ. في حين هنالك العديد من الطرق الاخرى لتقدير البيانات المكانية المفقودة مثل مستويات المياه الجوفية والتضاريس. من ناحية أخرى، تم تطوير العديد من الطرق لتقييم كفاءاتها ولكن نادراً ما يتم حساب عدم التيقن من النتائج. في هذة الدراسة تمت مقارنة ثلاث طرق للإستكمال الداخلي لتقدير البيانات الطبوغرافية المكانية المفقودة في مدينة بابل القديمة. تم تحديد مجال النموذج في عدد من الخلايا الأفقية (19) والخلايا العمودية (23). استخدمت خمس نقاط معلومة الارتفاع لتقدير الارتفاع غير المعروف في 257 نقطة. وتمت مقارنة الطريقة الجديدة (NTPI)مع طريقة ترجيح المسافة المعكوس (IDW) وطريقة السيرفر. تم حساب كفاءة هذة الطرق عن طريق معدل الأخطاء والأنحراف المعياري. تقنية الاستكمال لأقرب ثلاث نقاط اعطت نتائج اقل لمعدل الأخطاء ولمعدل الانحراف المعياري وكذلك دقة أكثر في توزيع إرتفاعات سطح الأرض.

الكلمات المفتاحية: الاستكمال الداخلي لأقرب ثلاث نقاط(NTPI)، ترجيح المسافة المعكوس (IDW) معدل الخطأ (AE)، الانحراف المعياري (SD).