Global Journal of Researches in Engineering: E
Civil And Structural Engineering
Volume 17 Issue 3 Version 1.0 Year 2017
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Prediction of Digital Elevation Model Height by Multivariate Adaptive Regression Splines (Mars) Interpolation Approach

By Zeena Adil Najeeb
Al-Nahrain University

Abstract- The objective of this paper is to assess the applicability and performance of multivariate adoptive regression spline analysis (MARS) for prediction elevation height in digital elevation model. MARS is an adaptive, nonparametric regression approach. Three dimensional coordinates (X, Y, and Z) in Equal-Sized grid Cell observed and recognized vie Differential Global Positioning System (DGPS) at AL-Nahrain university site. Mathematical prediction models with their errors and analysis are established in this paper. As the same time the independent variables X,Y and the dependent predicted variable Z the height which be used in. The data were dividedrandomly into training and testing70% of the entire data set is utilized for training and the remaining30% for testing. MARS depends on two steps for computation logarithm forward and backward to get better performance MARS adopts Generalized Cross-Validation (GCV) with different statistical parameters of standard deviation, root mean square error and residuals.

Keywords: digital elevation model, MARS, Height prediction, DGPS.

GJRE-E Classification: FOR Code: 090599

PREDICTIONOFDIGITALELEVATIONMODELHEIGHTBYMULTIVARIATEADAPTIVEREGRESSIONSPLINESMARSINTERPOLATIONAPPROACH

Strictly as per the compliance and regulations of:

Prediction of Digital Elevation Model Height by Multivariate Adaptive Regression Splines (MARS) Interpolation Approach

Zeena Adil Najeeb

Abstract- The objective of this paper is to assess the applicability and performance of multivariate adoptive regression spline analysis (MARS) for prediction elevation height in digital elevation model. MARS is an adaptive, non-parametric regression approach. Three - dimensional co-ordinates (X, Y, and Z) in an Equal-Sized grid Cell were using Differential Global Positioning System (DGPS) at AL-Nahrain University site. The Mathematical prediction models with their errors and analysis are established in this paper; as same time the independent variables X,Y and the dependent predicted variable Z which be consider the elevation . The data were dividedrandomly into training and testing70% of the entire data set is utilized for training and the remaining30% for testing. MARS depends on two steps for computation which are logarithmic forward and backward solution to get better performance MARS adopts Generalized Cross-Validation (GCV) with different statistical parameters of standard deviation, root mean square error and residuals.

Keywords: digital elevation model, mars, elevation prediction, dgps

I. Introduction

A digital elevation model (DEM) is a numerical representation of topography usually made up of equal-sized grid cells, each with an elevation value [1] DEMs have been widely used in many applications, such as urban planning, civil engineering, landscape building, and mining engineering [2].The accuracy of DEM depends on several criteria such as points distribution, type of instrument and DEM interpolation model. [1] DEMs can be obtained from contour lines, topographic maps, field surveys, photogrammetric techniques, radar interferometry, and laser altimetry [2]. In a gridding method, the corners of regular rectangles or squares are calculated from the scattered control points [3]. Interpolation in digital terrain modeling is used to determine the height value of a point by using the known height of neighboring points [4]. There are many interpolation methods that can be used to generate digital elevation models which include multiple linear regression, Nearest Neighbor, MARS, ANN, Polynomial regression etc. Multivariate Adaptive Regression Splines model is a new method for predicted the DEM height technique [5]. systematic error ,residual error will be determine in regression analysis [6] ANN interpolation is an approximate interpolator, which means that its accuracy is certainly less for known sample points. A regression analysis is used to determine if errors are attributed to any spatial attribute such as the degree of slope, aspect, or distance and direction from the set of nearest neighboring sample points [8].in MARS models data using linear predictor function and estimates unknown model parameters from this data were be used.[9]. The results demonstrate the effectiveness of the new method (MARS) model in prediction elevation height. Error and standard error was then conducted to evaluate the performance of this Interpolation models and MARS approach. Effective to measure the elevation of all points on the terrain surface. Therefore, point densification may be better executed using mathematical models. These models are particular forms of mathematical surface that deals with numerical Representation of the surface of the earth. Digital Elevation Models (DEM). Digital Height Model (DHM), Digital Ground Model (DGM), and Digital Terrain Elevation Model (DTEM) are all common terms.

II. Materials and Methodology

a) Description of the study area

The research site lies between 44° S Longitude and 33° N Latitude which districts in the middle south of Baghdad city/Iraq in Al-Nahrain university, it was cover from the total area and it is denoted by Baghdad university from west, AL-Jaderiya region from the north east site, easting(E) northing (N) and estimation elevation Height (Z) are represent surface coordinates. About 1780 points as gridding cells were be observed. these points were used in mathematical model, with an average elevation of 30 m above sea level. as illustrated in figure (1) and (2).

Author: *Department of civil engineer, College of Engineering, Al-Nahrain University, Baghdad, Iraq. e-mail: Zinah_adil@yahoo.com*

Figure (1): Iraq UTM Zone

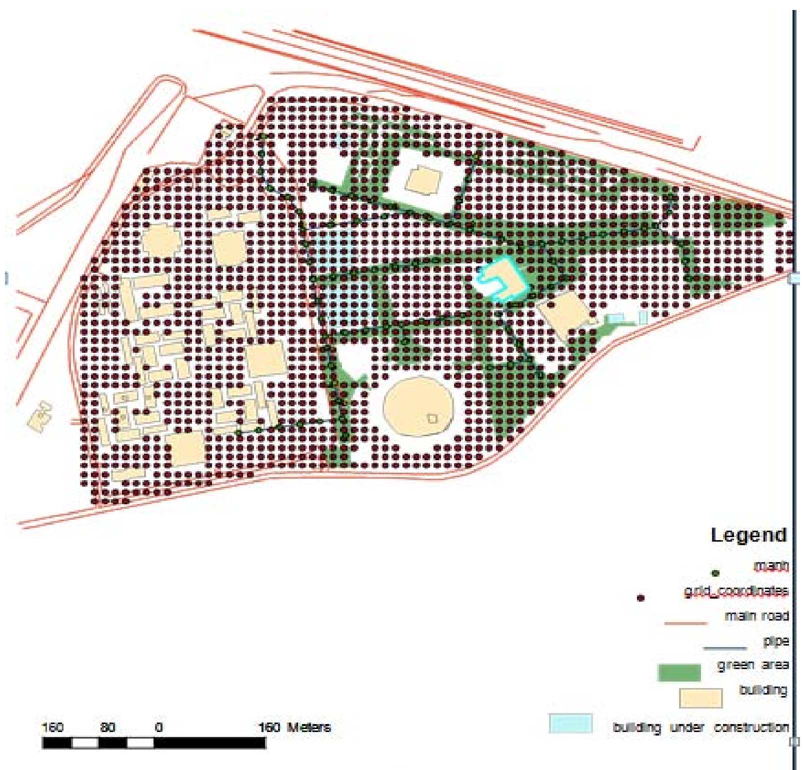

Figure (2): Selected the Study Area Showing the Gridding Cells of DEM *b) Interpolation Method*

A digital elevation model is a mathematical presentation for the surface .It employs one or more mathematical function according to some specific methods based on the set of measured data points [5].Interpolation method is used to determine the height value of a point by using the known heights of neighboring points .Also interpolation techniques can be classified according to different criteria and can be used for different purposes.

i. *Multivariate Adaptive Regression Splines (MARS)*

MARS is a powerful nonparametric model which has been successfully used in many applications of science and technology such as predicting object-oriented software maintain ability [7].MARS divides the whole space of input variable into various sub regions in different mathematical equation for each area and relates to input and output variable using spline basis functions [11].

$$[-(x-t)]_+^q = \begin{cases} (t-x)^q \,, if \; x < t \\ 0 \,, otherwise \end{cases} \quad \cdots\cdots \quad (1)$$

$$[+(x-t)]_+^q = \begin{cases} (t-x)^q \,, if \; x \geq t \\ 0 \,, otherwise \end{cases} \quad \cdots\cdots \quad (2)$$

where q is the power and t is knot

The final MARS model has the following form:

$$\acute{y} = \acute{f}(x) = a_0 + \sum_{M=1}^{M} a_m B_m\,(x) \quad (3)$$

Where

y output variable

a_0 the coefficient constant term

M the number of spline function and

$B_m$and a_m the m_{th} spline function and its coefficient respectively.

MARS uses the following two steps:

Forward Algorithm: Basis functions are introduced to define Equation (3). Which Many are added in to get better performance. However the developed MARS can have over fitting problem due to a large number of basis functions.

Backward Algorithm: to prevent over fitting, redundant basis functions are deleted from MARS adopts Generalized Cross-Validation (GCV) method to delete the redundant basis functions [12]. The GCV function is shown in Equation (4).

$$GCV = \frac{\frac{1}{N}\sum_{t=1}^{N}[y_{t-f}\;(x_t\,')]^{\,2}}{[1-\frac{c(B)}{N}]^{\,2}} \quad \cdots\cdots \quad (4)$$

WhereN the number of data and C*(B)* is a complexity penalty that increases with the number of basis function (BF) in the model and it is defined as:

$$C(B) = (B+1) + dB \quad (5)$$

Whered is a penalty for each BF included into the model and B the number of basis functions in Equation (3) [13].

A sensitivity analysis has been done to extract the cause and effect relationship between the inputs and outputs of the MARS model. The basic idea is that each input of the model is offset slightly and the corresponding change in the output is reported. The procedure has been taken from the work of Pijush and Kothari [14]. Accordingly the sensitivity (S) of each input parameter has been calculated by the following formula:

$$S\,(\%) = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\%\,change\;\;in\;output}{\%\,change\;\;in\;input}\right)_j \times 100 \quad (6)$$

Additionaly Comparison made in terms root mean squareerror (RMSE) and mean absolute error (MAE). The values of RMSE and MAE have been determined by using the following relation below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_{ai} - Q_{pi})^{2}}{n}} \quad (7)$$

$$MAE = \frac{\sum_{i=1}^{n}|Q_{ai} - Q_{pi}|}{n} \quad (8)$$

Where Qai and Qpi are the actual and predicted Q values, respectively, and n the number of data.

ii. *Data Mining the Mars Mathmatical Model*

Data mining (DM) tools in Statistica version 12was used to predicted height of observed points by making regression logarithm using flexible model building strategy.

In this study software was used in developing mutative adoptive regression spline mathematical model to predict height Continuous dependent variables (Elevations) and continuous independent variables (Eastings and Northings) for all ground control points were used in the MARS. The data were divided in two sets with about 70% of them used for training and the rest for testing the results. The training data points were used to train the network and compute the weights of the inputs [10].The cross-validation method computes the error in a test set at the same time that the model is being trained. The test data points were used to measure the performance of the selected MARS model.

In order to develop the best possible model, all patterns that are contained in the data need to be included in the training set. Similarly, since the test set is used to determine when to stop training, it needs to be representative of all data, and thus should contain all of the patterns that are present in the available data.

Data mining in Statistica was used since it uses a large-data prediction for variables with MARS model specification having a minimum number of basis functions and showing good result in the prediction.

III. Results for Mars Model and Discussion

Although MARS is widely applied in different fields, it is not widely used in surveying applications.

Thus, we investigated the performance of the MARS digital elevation model interpolation technique to predict the heights from the digital elevation model. DGPS survey data revealed the accuracy of prediction on different points. These results were compared with the testing and training summarization of the GCV using the statistical parameters of standard deviation, root mean square error, and residuals on the both testing and training points, which are shown in Table 1.

Table 1: Summarization Model

Model specifications 30% points	Value	Model specifications 70% points	Value
Independents	2	Independents	2
Dependents	1	Dependents	1
Number of Terms	6	Number of Terms	6
Number of Basis function	5	Number of Basis function	5
GCV	0.114645	GCV	0.183028

Regression statistics were used to compare the validation method, which indicates the best performance with respect to the mean and to the standard deviation as illustrated in Tables 2 and 3.

Table 2: Regression Testing Points 30% Results

Regression statistics	Regression statistics Testing points 30%
	Elevation
Mean (observed)	30.04858
Standard deviation (observed)	0.37972
Mean (predicted)	30.04858
Standard deviation (predicted)	0.17059
Mean (residual)	0.00000
Standard deviation (residual)	0.33925
R-square	0.20182
R-square adjusted	0.19492

Table 3: Regression Training Points 70% Results

Regression statistics	Regression statistics Training points 70%
	Elevation
Mean (observed)	30.60330
Standard deviation (observed)	2.08747
Mean (predicted)	30.60330
Standard deviation (predicted)	2.04379
Mean (residual)	0.00000
Standard deviation (residual)	0.42478
R-square	0.95859
R-square adjusted	0.95842

A two-dimensional scatter plot of the predictions, observations, and residuals for each elevation are illustrated in Figures 3 and 4. The scatter plots refer to the optimum results from the random data in which testing and training points were selected. The final strategy for the value used in the GCV are goodness of fit, RMSE, and residuals, which leads to the optimum MARS model S obtained from the testing and training

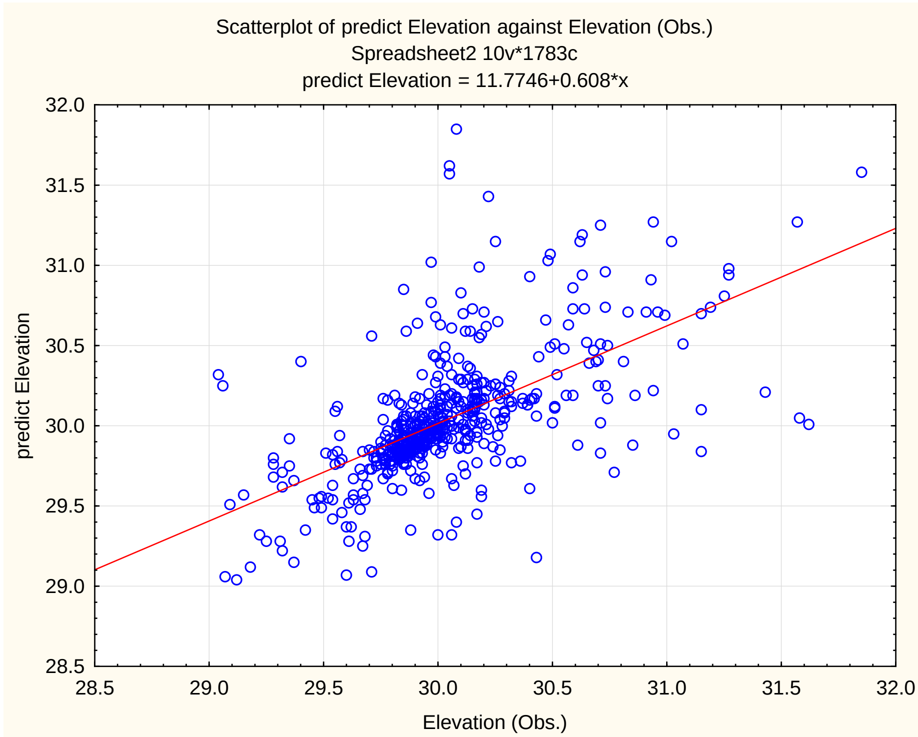

Figure 3: Two-dimensional scatter plot of the predicted elevations against the observed elevations from the testing points.

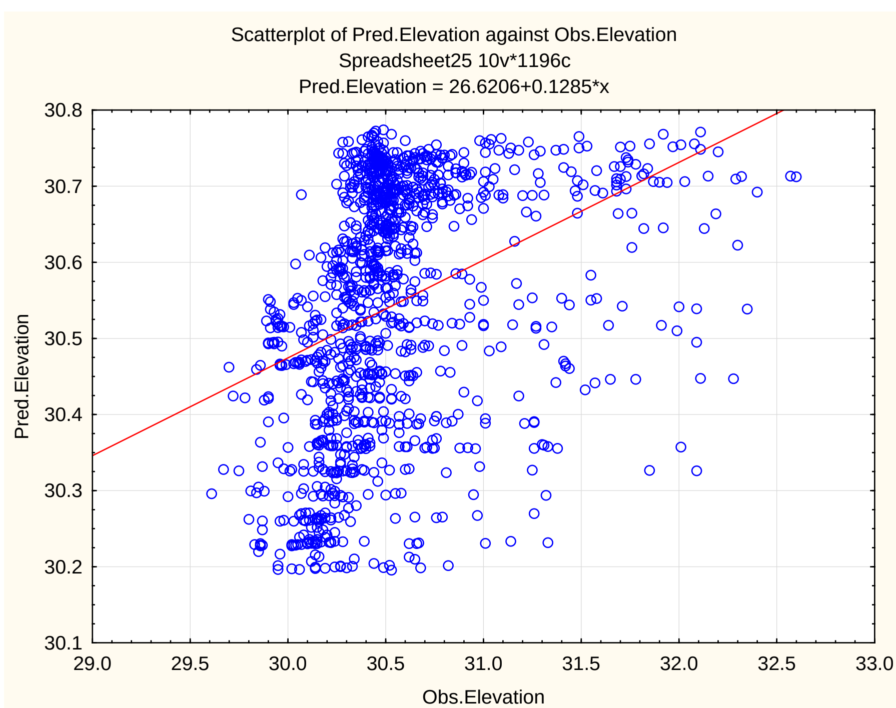

Figure 4: Scatter plot of the predicted elevations against the observed elevation derived from the training points.

IV. Conclusions

This work is part of series of interpolation DEM methods coming soon, the author wishes to express his appreciation to the Al_Nahrain University Survey Team for providing us with the data. It successfully adopted the MARS model to predict elevations showing high accuracy using a flat terrain digital elevation model. In addition, it shows the efficiency of computing residuals from a large set of data. The MARS model resulted in acceptable performance using the developed equations for determining minimum and maximum elevation. This study shows that MARS can be used as a robust tool for solving different problems in Geomatics engineering.

References Références Referencias

1. Vincent Chaplot, Frédéric Darboux, Hocine Bourennane, Sophie Leguédois, Norbert Silvera, Konngkeo Phachomphon" Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density" Geomorphology 77 (2006) 126–141.
2. W. Z. Shi & Y. Tian (2006) A hybrid interpolation method for the refinement of a regular grid digital elevation model, International Journal of Geographical Information Science, 20:1,53-67, DOI: 10.1080/13658810500286943
3. P.V. Arun "A comparative analysis of different DEM interpolation methods" The Egyptian Journal of Remote Sensing and Space Sciences (2013) 16, 133–139.
4. Okwuashi O, Ndehedehe C. Digital terrain model height estimation using support vector machine regression. S Afr J Sci. 2015;111(9/10), Art. 2014-0153, 5 pages. http://dx.doi.org/10.17159/sajs.2015/20140153
5. H.Karaborka, O.K. Baykanb, C. Altuntasa, F.Yildza," Estimation of Unknown Height With Artificial Neural Network on Digital Terrain Model"
6. Zhilin Li, Qing Zhu and Christopher Gold" DIGITAL Terrain Modeling" Principles and Methodology 2005 by CRC Press.
7. David A. Merwin, Robert G. Cromley & Daniel L. Civco (2002) Artificial Neural Networks as a Method of Spatial Interpolation for Digital Elevation Models, Cartography and eographic Information Science, 29:2, 99-110, DOI: 10.1559/152304002782053323
8. Zhou, Y., Leung, H.: Predicting object-oriented software maintainability using multivariate adaptive regression splines, Journal of Systems and Software, 80(8) 1349-1361 (2007)
9. Jianya Gong, Qing Zhu, YaolinLiu, "DEM interpolation based on artificial neural networks" Geospatial Information, Data Mining, and Applications, Proc. of SPIE Vol. 6045, 604528, (2005) • 0277-786X/05/$15 • doi: 10.1117/12.651405
10. Aman.O.C, Wu Beiping Ziggah.Y.Y." Testing Simple Regression Model for Coordinate Tranformation by Comparing Its Predictive Result for two Regions" Academic Research International IISSN-L:2223-9553,ISSN:2223-9944 VOL.4 NO .6 November 2013
11. Bashar Tarawneh Geoscience Frontiers 8 (2017) 199-204 "Predicting standard penetration test N-value from cone penetration test data using artificial neural networks"
12. S. Sekulic and B. R. Kowalski, “MARS: A tutorial,” J. Chemom., vol. 6, pp. 199–216, 1992.
13. P. Craven and G. Wahba, “Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation,” Numer Math, vol. 31, pp. 317-403, 1979.
14. J. H. Friedman, “Multivariate adaptive regression splines,” The Annals of Statistics, vol. 19, pp. 1–14, 1991.
15. PijushSamui , and Dwarkadas Pralhaddas Kothari2 "A Multivariate Adaptive Regression Spline Approach for Prediction of Maximum Shear Modulus (Gmax) and Minimum Damping Ratio (εmin)" ENGINEERING JOURNAL Volume 16 Issue 5 9553, ISSN: 2223-9944