

# DEM构建的多面函数加权抗差算法

陈传法, 李 伟, 李明飞, 戴洪磊

(山东科技大学测绘科学与工程学院, 青岛 266590)

**摘要:** 为了抑制采样点中粗差对DEM构建精度影响, 本文以较高精度的多面函数(MQ)为基函数, 发展了一种MQ迭代加权抗差算法(MQ-R)。MQ-R以传统MQ计算结果为初始值, 以MQ函数模拟值与对应采样点的差值确定采样点权重, 以加权MQ优化初始值, 重复迭代直至收敛。以数值模拟曲面为研究对象, 本文比较并分析了采样误差服从正态分布、被污染的正态分布, 以及Cauchy分布时MQ-R与MQ模拟结果精度。数值分析表明, 当采样误差服从正态分布时, MQ-R计算精度和传统MQ相当; 随着污染率的提高, MQ计算精度急剧降低, 而MQ-R计算结果受粗差影响较小; 当采样误差来源于 $C(0,1)$ 分布时, MQ计算结果完全失真, 而MQ-R可在一定程度上抑制粗差影响。总之, 相比于传统MQ算法, MQ-R不仅具有较高的计算效率, 而且具有较高的抗差性。

**关键词:** 抗差; 薄板样条; 精度; DEM

**DOI:** 10.3724/SP.J.1047.2013.00840

## 1 引言

DEM是对地球表面地形地貌的离散化表达<sup>[1-2]</sup>, 其精度主要取决于采样点质量<sup>[3-4]</sup>。数据采集集中, 受仪器噪声、恶劣天气, 以及人为操作不当等影响, 采样数据中不可避免含有异常值<sup>[5-6]</sup>, 进而影响DEM构建的精度。为了克服采样数据中异常值影响, 相关研究人员提出了一系列算法, 以抑制粗差对DEM影响: Hannah<sup>[7]</sup>提出了一种以局部区域坡度及坡度变化的粗差探测方法; Felicísimo<sup>[8]</sup>发展了一种概率统计算法( $t$ 分布)的粗差探测算法; López<sup>[9]</sup>借助主成分分析剔除DEM中的粗差。尽管以上算法具有较高的粗差探测效率, 但其仅适用于栅格DEM粗差探测。为此, Liu, *et al.*<sup>[10]</sup>发展了一种“两步走”DEM构建算法, 首先, 借助局部自适应抗差统计分析算法剔除离散采样点中的粗差, 然后, 用剩余采样点构建DEM。尽管该算法能构建较高精度DEM, 但该算法在粗差探测时需要人为设置7个参数, 故主观性较强; 且当粗差聚集时, 该算法失效。大量试验表明, DEM构建时, 抗差插值算法计算效率要优于“两步走”技术<sup>[11]</sup>。因此, Chen and Li<sup>[12]</sup>提出了最小绝对偏差(LAD)的多面函数(MQ)抗

差算法(MQ-L)。MQ-L首先借助空间填充算法选择部分采样点作为MQ节点, 然后借助LAD解算MQ误差方程。该算法与传统算法相比, 在一定程度上抑制了DEM构建时采样数据中的粗差影响, 但MQ节点数选择需要充分利用先验知识, 致使该算法主观性较强, 进而限制其推广使用。

为了克服上述抗差算法的缺陷, 本文以MQ为基函数, 发展了一种MQ加权抗差算法(MQ-R)。相比于MQ-L, MQ-R以所有采样点作为其基函数节点, 进而避免了人为选择节点数的主观性<sup>[13]</sup>。MQ-R以传统MQ计算结果为初始值, 以MQ对各个采样点的模拟残差确定采样点权重, 以加权MQ优化初始值, 重复迭代直至收敛。MQ-R无需设置节点数, 且能自动搜索光滑参数最优值, 故具有较好实用性。本文首先借助岭回归重新推导了MQ计算公式, 然后, 以迭代加权为抗差基础, 提出了MQ加权抗差算法, 最后, 以数值试验验证该算法的高效性。

## 2 模型与算法

### 2.1 基于岭回归的MQ公式推导

曲面模拟旨在借助采样点  $z(x_i, y_i)$  ( $i = 1, 2, \dots, n$ )

收稿日期: 2013-11-08; 修回日期: 2013-12-01.

基金项目: 国家自然科学基金项目(41101433、41371367); 山东省优秀中青年科学家科研奖励基金项目(BS2012HZ010); 青岛市科技计划基础研究项目(13-1-4-239-jch)。

作者简介: 陈传法(1982-), 男, 山东沂源人, 博士, 副教授, 研究方向为地学曲面建模。E-mail: chencf@lreis.ac.cn

获得曲面函数  $f(x, y)$  准确表达式, 进而估计未知点函数值<sup>[14]</sup>。曲面模拟的一般模型可表达为<sup>[15]</sup>:

$$z(x_i, y_i) = f(x_i, y_i) + e(x_i, y_i), \quad i = 1, 2, \dots, n \quad (1)$$

式(1)中,  $z(x_i, y_i)$  表示第  $i$  个采样点值;  $f(x_i, y_i)$  表示函数模拟值;  $e(x_i, y_i)$  表示采样误差;  $n$  为采样点数。

对 MQ 而言, 其函数表达式为<sup>[12]</sup>:

$$f(x, y) = \sum_{i=1}^m \alpha_i p_i(x, y) + \sum_{i=1}^n \beta_i q(r_i) \quad (2)$$

式(2)中,  $p(x, y)$  和  $\alpha$  分别为 MQ 多项式基函数和多项式系数;  $r_i$  表示第  $i$  个采样点与待求点距离;  $q(r)$  和  $\beta$  分别为 MQ 径向基函数及其权重。吕言<sup>[16]</sup>、龚健雅<sup>[17]</sup>等分别从数值试验和理论分析表明, 三次曲面作为 MQ 的核函数精度高于 Hardy 采用的双曲面方法, 故本文取  $q(r) = 1 + r^3$ 。

式(2)的矩阵表达式可表示为:

$$f = P\alpha + Q\beta \quad (3)$$

式(3)中,  $f = [f_1 \ \dots \ f_n]^T$ ;  $\alpha = [\alpha_1 \ \dots \ \alpha_m]^T$ ;

$$\beta = [\beta_1 \ \dots \ \beta_n]^T; P = \begin{bmatrix} p_{11} & \dots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{n1} & \dots & p_{nm} \end{bmatrix}; p_{ij} = p_j(x_i, y_i);$$

$$Q = \begin{bmatrix} q_{11} & \dots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \dots & q_{nn} \end{bmatrix}$$

因此, 要想得到 MQ 模拟曲面, 必须先计算系数  $\alpha$  和权重  $\beta$ 。

由岭回归基本原理<sup>[18-19]</sup>得 MQ 的目标函数为:

$$\min_{\alpha, \beta} (f - P\alpha - Q\beta)^T (f - P\alpha - Q\beta) + \rho \beta^T Q\beta \quad (4)$$

式(4)中,  $\rho$  为光滑参数, 其值可以借助交叉验证<sup>[20]</sup>求得。

设  $F = (f - P\alpha - Q\beta)^T (f - P\alpha - Q\beta) + \rho \beta^T Q\beta$ , 对  $F$  求关于  $\beta$  的导数得:

$$\frac{\partial F}{\partial \beta} = -(f - P\alpha - Q\beta)^T Q + \rho \beta^T Q = 0 \quad (5)$$

即

$$(Q + \rho I)\beta + P\alpha = f \quad (6)$$

对  $F$  求关于  $\alpha$  的导数得:

$$\frac{\partial F}{\partial \alpha} = (f - P\alpha - Q\beta)^T P = 0 \quad (7)$$

将式(6)代入式(7)得:

$$P^T \beta = 0 \quad (8)$$

由式(6)和式(8)得:

$$\begin{bmatrix} Q + \rho I & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \quad (9)$$

由式(9)可以获得 MQ 系数, 将系数带入式(3)即可以求算其他点函数值, 进而实现曲面模拟。

## 2.2 MQ 加权抗差算法(MQ-R)

由式(4)知, 当采样点含有粗差时, 其计算结果严重失真, 因此, 必须发展抗差 MQ 算法以抑制粗差对计算结果影响。加权后的 MQ 目标函数可表达为:

$$\min_{\alpha, \beta} (f - P\alpha - Q\beta)^T W (f - P\alpha - Q\beta) + \rho \beta^T Q\beta \quad (10)$$

式(10)中,  $W$  为采样点权重矩阵, 其为对角矩阵;  $W = \text{diag}\left(\frac{1}{w_1} \ \dots \ \frac{1}{w_n}\right)$ ,  $w_i$  表示第  $i$  个采样点权重。

设  $F = (f - P\alpha - Q\beta)^T W (f - P\alpha - Q\beta) + \rho \beta^T Q\beta$ , 对  $F$  求关于  $\beta$  的导数得:

$$\frac{\partial F}{\partial \beta} = -(f - P\alpha - Q\beta)^T W Q + \rho \beta^T Q = 0 \quad (11)$$

$$\text{即} \quad (WQ + \rho I)\beta + WP\alpha = Wf \quad (12)$$

$$\text{即} \quad (Q + \rho/W)\beta + P\alpha = f \quad (13)$$

对  $F$  求关于  $\alpha$  的导数得:

$$\frac{\partial F}{\partial \alpha} = (f - P\alpha - Q\beta)^T WP = 0 \quad (14)$$

将式(13)代入式(14)得:

$$P^T \beta = 0 \quad (15)$$

由式(13)和式(15)得:

$$\begin{bmatrix} Q + \rho/W & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \quad (16)$$

比较式(9)、(16)可知, MQ 和 MQ-R 表达式形式基本一致, 唯一区别是式(16)考虑采样点权重。权重大小主要取决于采样点模拟残差  $e = f - P\alpha - Q\beta$ 。理论上, 权重与残差呈反比, 即, 残差越小, 权重越大。本文使用的权重与残差关系表达为:

$$w_k = \begin{cases} 1, & |e_k/\hat{s}| < c_1 \\ c_2 - \frac{|e_k/\hat{s}|}{c_2 - c_1}, & |e_k/\hat{s}| \in [c_1, c_2] \\ 10^{-4}, & \text{otherwise} \end{cases} \quad (17)$$

式(17)中,  $c_1$  和  $c_2$  为常数, 取  $c_1 = 2.5$ ;  $c_2 = 3$ ;  $\hat{s}$  表示标准差。为了提高标准差的抗差性及计算效率, 本文采用 Rousseeuw and Croux<sup>[21]</sup>提出的标准差

计算公式:

$$\hat{s} = 1.1926 \text{med}_i \left\{ \text{med}_j |e_i - e_j| \right\} \quad (18)$$

综上所述, MQ加权抗差算法是一个不断迭代的过程, 其计算步骤为:

- (1) 用交叉验证获得最优光滑参数  $\rho$ ;
- (2) 利用式(9)计算MQ标准方程系数;
- (3) 计算函数残差  $e = f - P\alpha - Q\beta$ ;
- (4) 利用式(17)计算权重;
- (5) 利用式(16)计算MQ方程系数;
- (6) 重复步骤(3)–(5), 直至系数收敛;
- (7) 利用式(3)进行曲面模拟。

### 2.3 模型验证与精度分析

选择数学合成曲面为研究对象, 分析MQ-R的抗差性, 并与传统MQ算法计算结果进行精度比较。含有误差的数学合成曲面表达式为:

$$f(x, y) = 3 + 2 \sin(2\pi x) \sin(2\pi y) + \exp(15(x-1)^2 - 15(y-1)^2) + \varepsilon \quad (19)$$

式(19)中,  $\varepsilon$  为随机误差。本文试验中的随机

误差主要来源于以下3个分布:

Case1: 标准正态分布  $N(0, 1)$ 。

Case2: 被污染的正态分布  $(1-\theta)N(0, 0.1^2) + \theta N(0, 3^2)$ 。式中,  $\theta$  为污染率, 分别为10%, 20%和30%。

Case3: Cauchy分布  $C(0, 1)$ 。

当误差来源于标准正态分布时(Case1), 采样数据中不含有粗差; 当误差来源被污染正态分布时(Case2), 约100 $\theta$ %数据来源于  $N(0, 3^2)$ , 故大部分数据被认为是粗差; 考虑到  $C(0, 1)$  有很长的尾部, 且其均值和方差均不存在, 因此误差来源于Cauchy分布时(Case3), 采样数据中含有大量粗差。

将模拟区域归一化为  $[0, 1] \times [0, 1]$ , 区域网格数为  $101 \times 101$ , 随机选择  $51 \times 51$  个采样点用于曲面模拟。除了MQ-R外, 传统MQ方法也用于曲面模拟。模拟结果精度指标分别为中误差(RMSE)、最大误差(Max), 以及最小误差(Min)。不同采样误差条件下, 两种方法模拟结果如表1所示。

由表1可知, 当采样误差服从正态分布时, 两种

表1 不同误差分布条件下, 两种方法模拟结果比较

Tab.1 Accuracy comparison between MQ-R and MQ under different error distributions

误差分布	MQ-R				MQ			
	$\rho$	Min	Max	RMSE	$\rho$	Min	Max	RMSE
$N(0, 1)$	369.05	-0.22	0.27	0.02	310.50	-0.23	0.28	0.02
$(1-\theta)N(0, 0.1^2) + \theta N(0, 3^2) \theta=0.1$	557.11	-0.18	0.17	0.02	17.05	-0.50	0.61	0.13
$(1-\theta)N(0, 0.1^2) + \theta N(0, 3^2) \theta=0.2$	604.63	-0.15	0.12	0.03	8.43	-0.71	0.90	0.17
$(1-\theta)N(0, 0.1^2) + \theta N(0, 3^2) \theta=0.3$	444.34	-0.23	0.25	0.03	1.05	-1.08	1.35	0.48
$C(0, 1)$	11.10	-1.24	0.81	0.21	0.02	-3.09	2.89	1.30

方法的模拟结果精度基本一致, 表明MQ-R具有较高的计算效率。当采样误差来源于被污染的正态分布时, 随着污染率的提高, 两种方法的计算精度

均有降低, 但相比MQ算法, MQ-R受污染率影响较小。例如, 当污染率从10%提高到30%时, MQ-R的RMSE从0.02升高到0.03, 提高率仅为50%, 而MQ

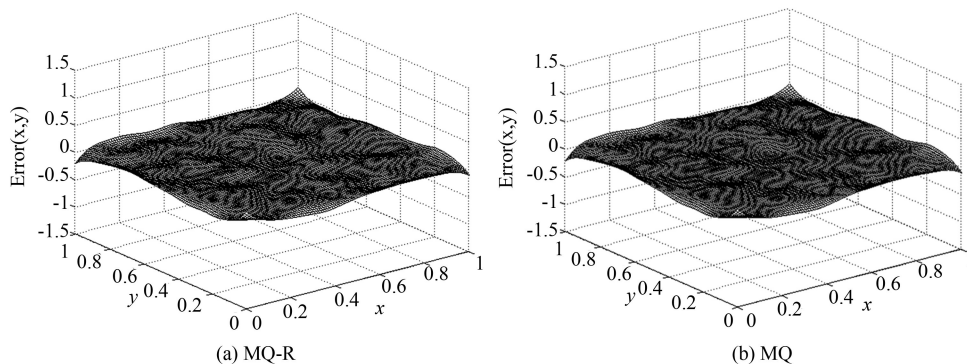


图1 采样误差服从正态分布时两种方法误差曲面

Fig.1 Error surfaces of MQ-R and MQ when sampling errors are from the normal distribution



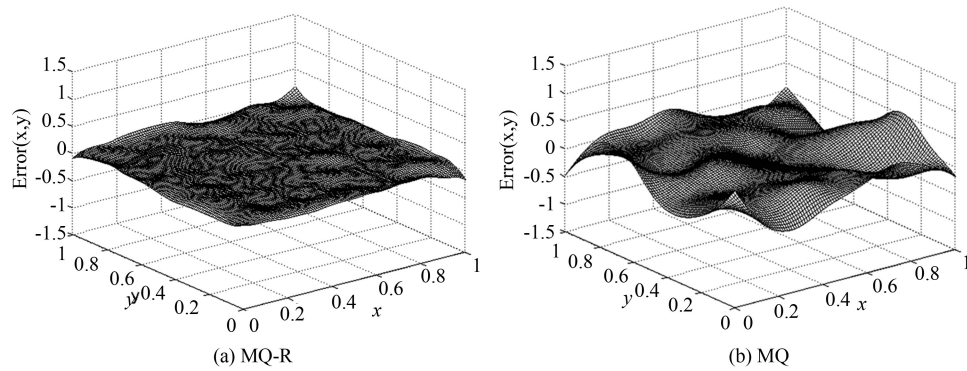


图2 污染率为10%时两种方法误差曲面

Fig.2 Error surfaces of MQ-R and MQ when contaminating proportion is 0.1

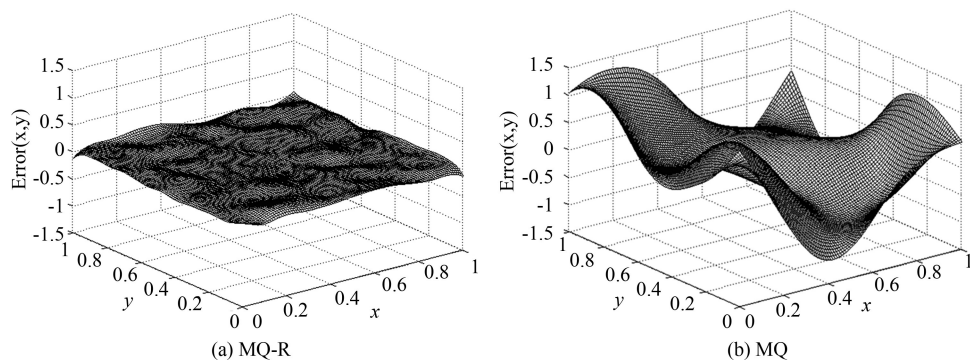


图3 污染率为20%时两种方法误差曲面

Fig.3 Error surfaces of MQ-R and MQ when contaminating proportion is 0.2

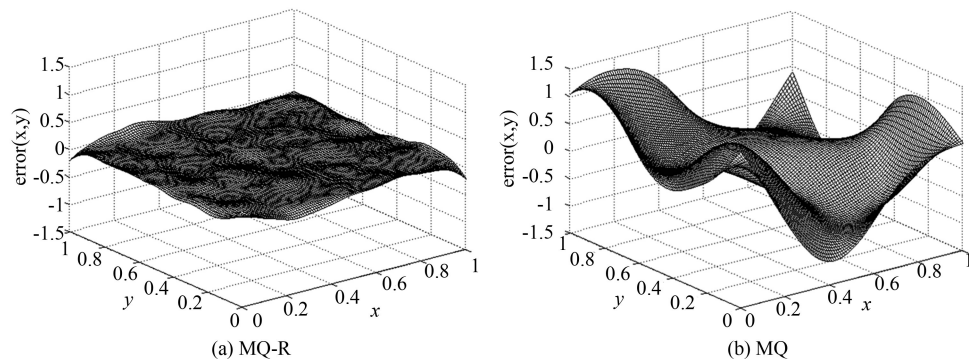


图4 污染率为30%时两种方法误差曲面

Fig.4 Error surfaces of MQ-R and MQ when contaminating proportion is 0.3

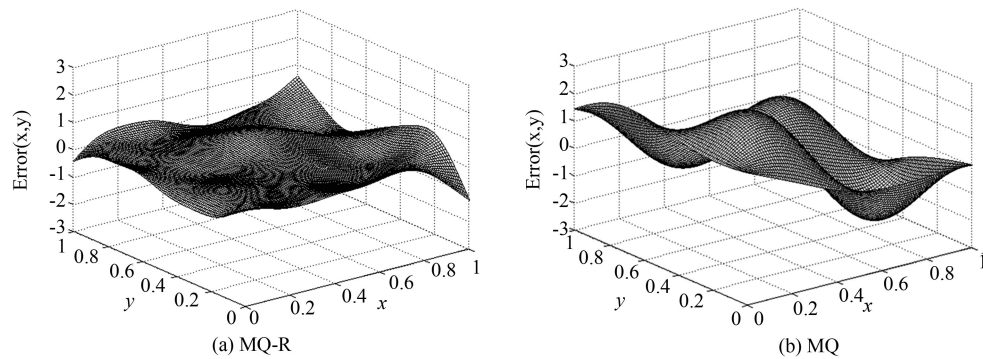
图5 采样误差服从 $C(0,1)$ 时两种方法误差曲面

Fig.5 Error surfaces of MQ-R and MQ when sampling errors are from Cauchy distribution

的RMSE从0.13升高到0.48,提高率为269.23%,且MQ的精度远低于MQ-R精度。当采样误差来源于 $C(0,1)$ 时,相比其他误差来源数据,两种方法的精度均有所降低,但MQ-R精度明显高于MQ。总之,相比于传统MQ算法,MQ-R不仅具有较高计算效率,而且具有较好的抗差性。

当采样误差来源于不同误差分布时,两种方法的误差曲面如图1-5所示。由图1可知,当采样误差来源于正态分布时,两种方法的误差曲面非常平坦,且曲面变化基本一致,表明MQ-R具有较高的计算效率。随着污染率的提高,传统MQ的误差曲面凹凸不平,且复杂性急剧提高,而MQ-R的误差曲面相对比较平坦,即MQ-R具有较高的抗差性。当采样误差服从 $C(0,1)$ 时,MQ计算结果完全被破坏,而MQ-R仍然能较好地抑制粗差影响,使其误差曲面变化相对缓和。

### 3 结论

受数据采集条件的限制,采样数据中不可避免的含有异常值<sup>[22]</sup>,进而严重影响DEM构建精度。为此,本文以较高精度的多面函数(MQ)为基函数,发展了MQ加权抗差算法(MQ-R)。数值试验表明,相比于传统MQ算法,MQ-R在保持较高计算效率的同时,可以在一定程度上抑制粗差对曲面模拟影响,进而提高DEM构建精度。

实际数据采集中,采样点精度除了受高程误差影响外,还受平面位置误差影响。因此,抗差插值时,应考虑杠杆点<sup>[23-24]</sup>影响。为此,后续研究我们将发展能抑制杠杆点影响的MQ抗差算法。

#### 参考文献:

- [1] Chen C F, Yue T X, Li Y Y. A high speed method of SMTS [J]. Computers & Geosciences, 2012(41):64-71.
- [2] Chen C F, Fan Z M, Yue T X, *et al.* A robust estimator for the accuracy assessment of remote-sensing-derived DEMs [J]. International Journal of Remote Sensing, 2012,33(8): 2482-2497.
- [3] Gong J, Li Z, Zhu Q, *et al.* Effects of various factors on the accuracy of DEMs: an intensive experimental investigation [J]. Photogrammetric Engineering and Remote Sensing, 2000,66(9):1113-1117.
- [4] Aguilar F J, Aguera F, Aguilar M A, *et al.* Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy[J]. Photogrammetric Engineering and Remote Sensing, 2005,71(7):805-816.
- [5] Gonga-Saholiariliva N, Gunnell Y, Petit C, *et al.* Techniques for quantifying the accuracy of gridded elevation models and for mapping uncertainty in digital terrain analysis[J]. Progress in Physical Geography, 2011,35(6): 739-764.
- [6] Chen C F, Yue T X. A method of DEM construction and related error analysis[J]. Computers & Geosciences, 2010,36 (6):717-725.
- [7] Hannah M J. Error detection and correction in digital terrain models[J]. Photogrammetric Engineering & Remote Sensing,1981,47(1):63-69.
- [8] Felicísimo A M. Parametric statistical method for error detection in digital elevation models[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 1994,49(4):29-33.
- [9] López C. Locating some types of random errors in digital terrain models[J]. International Journal of Geographical Information Science, 1997,11(7):677-698.
- [10] Liu H, Jezek K C, O'Kelly M E. Detecting outliers in irregularly distributed spatial data sets by locally adaptive and robust statistical analysis and GIS[J]. International Journal of Geographical Information Science, 2001,15(8): 721-741.
- [11] Maronna R A, Martin R D, Yohai V J. Robust statistics: Theory and methods[M]. New York: Wiley Blackwell, 2006.
- [12] Chen C F, Li Y Y. A robust multiquadric interpolation for DEM construction[J]. Mathematical Geosciences, 2013, 45(3): 297-319.
- [13] Franke R, Hagen H. Least squares surface approximation using multiquadrics and parametric domain distortion[J]. Computer Aided Geometric Design, 1999,16(3):177-196.
- [14] Chen C F, Li Y Y. A robust method of thin plate spline and its application to DEM construction[J]. Computers & Geosciences, 2012(48):9-16.
- [15] Yue T X, Chen C F, Li B L. An adaptive method of high accuracy surface modeling and its application to simulating elevation surfaces[J]. Transactions in GIS, 2010,14(5): 615-630.
- [16] 吕言. 数字地面模型中多面函数内插法的研究[J]. 武汉测绘学院学报,1981(2):14-28.
- [17] 龚健雅. 关于DTM中多面函数内插法几个问题的研究[J]. 测绘通报,1985(5):12-16.
- [18] Hoerl A E, Kennard R W. Ridge regression: Biased estimation for nonorthogonal problems[J]. Technometrics, 1970,12(1):55-67.
- [19] Hoerl A E, Kannard R W, Baldwin K F. Ridge regression: some simulations[J]. Communications in Statistics, 1975,

- 4(2):105-123.
- [20] Falivene O, Cabrera L, Tolosana-Delgado R, *et al.* Interpolation algorithm ranking using cross-validation and the role of smoothing effect: A coal zone example[J]. *Computers & Geosciences*, 2010,36(4):512-519.
- [21] Rousseeuw P J, Croux C. Alternatives to the median absolute deviation[J]. *Journal of the American Statistical Association*, 1993,88(424):1273-1283.
- [22] Rousseeuw P J, Debruyne M, Engelen S, *et al.* Robustness and outlier detection in chemometrics[J]. *Critical Reviews in Analytical Chemistry*, 2006,36(3-4):221-242.
- [23] Rousseeuw P J. Least median of squares regression[J]. *Journal of the American Statistical Association*, 1984(79):871-880.
- [24] Rousseeuw P J, Van Driessen K. Computing LTS regression for large data sets[J]. *Data Mining and Knowledge Discovery*, 2006,12(1):29-45.

## A Robust Multiquadratic Method and Its Application to DEM Construction

CHEN Chuanfa\*, LI Wei, LI Mingfei and DAI Honglei

(*Geomatics College, Shandong University of Science and Technology, Qingdao 266590, China*)

**Abstract:** In order to resist the effect of outliers on DEM construction, a robust multiquadric method (MQ-R) has been developed. MQ-R firstly takes the estimation of the classical MQ as the initial values to compute the residuals of all sampling points, and then a weighted function has been constructed to determine the weights of sampling points based on the above residuals. Finally, a iteratively re-weighted MQ is formed to decrease the effect of outliers on DEM construction. At the same time, the smoothing parameter of MQ and MQ-R is determined based on a k-fold cross-validation. A synthetic surface was employed to comparatively analyze the estimation accuracies of MQ-R and the classical MQ, where the sampling points are contaminated by three groups of errors with different distributions. These include the standard normal distribution, contaminated normal distribution with the contaminating proportion of 10%, 20% and 30%, and Cauchy distribution. Numerical tests indicate that when sampling errors are from the standard normal distribution, the accuracy of MQ-R is comparative to that of MQ. As the contaminating proportion increases, the accuracy of MQ becomes lower, whereas MQ-R can resist outliers very well. When sampling errors are from Cauchy distribution, the results of MQ are completely destroyed, but those of MQ-R are still satisfactory. In conclusions, MQ-R with a high efficiency and a high robustness can be used to resist outliers in DEM construction.

**Key words:** robust; thin plate spline; accuracy; DEM

\*Corresponding author: CHEN Chuanfa, E-mail: chencf@lreis.ac.cn