Adjustment computation of HASM: a high-accuracy and high-speed method

T. X. Yue\textsuperscript{*}, S. H. Wang\textsuperscript{*}

\textsuperscript{*} State Key Laboratory of Resources and Environment Information System, Institute of Geographical Sciences and Natural Resources Research, Beijing, China

First published on: 30 September 2010
Adjustment computation of HASM: a high-accuracy and high-speed method

T.X. Yue* and S.H. Wang

State Key Laboratory of Resources and Environment Information System, Institute of Geographical Sciences and Natural Resources Research, Beijing, China

(Received 7 May 2009; final version received 17 December 2009)

We developed a method for high-accuracy surface modeling (HASM) in terms of the fundamental theorem of surfaces, which has theoretically found a solution for error problems. However, the computing speed of HASM was too slow to be widely applied in practice. Thus, adjustment computation of HASM (HASM-AC) is developed in this article. For comparatively testing HASM-AC accuracy, a mathematical surface is first selected so that the true value is able to be predetermined to avoid uncontrollable data errors. The numerical test indicates that HASM-AC has the highest accuracy and its accuracy is 20.67, 15.67, and 14.67 times higher than the inverse distance weighting (IDW), kriging, and spline, respectively. Then, a 0.4 km × 0.5 km rectangular area is used to test the effects of different spatial resolutions and sampling intervals on accuracy. This real-world test demonstrates that HASM-AC accuracy increases at a much better and stable pace as the spatial resolution becomes finer and sampling intervals get shorter, compared to the classic methods. Finally, the computing speed is tested in an area with 6000 × 6000 grid cells where Qinghai, Gansu, and Sichuan provinces meet. The computing speed of HASM-AC is 11, 8, and 563 times faster than IDW, spline, and kriging, respectively, which makes HASM-AC able to process data in a huge size and make real-time visualization realizable. In short, HASM-AC performs best in both the numerical and real-world tests.

Keywords: HASM; adjustment computation; accuracy; computing speed; real-time visualization

1. Introduction

Surface modeling refers to the process of simulating a surface through a scattered point-form data set, a line-form data set, and/or an area-form data set. Various methods for surface modeling have been developed since the early 1950s. Krige (1951) proposed the basic idea of kriging interpolation, which belongs to the family of estimation algorithms of linear least squares and interpolates the value in a random field at an unobserved location from the observations of its value at nearby locations. Bengtsson and Nordbeck (1964) suggested a method for linear interpolation in a triangulated irregular network (TIN) to fit irregularly distributed points into a surface in each triangle. Zienkiewicz (1967) developed the finite element method (FEM) and Akima (1978a, 1978b) programmed the FEM algorithm for surface fitting. Shepard (1968) proposed a surface simulation formula, that is, the inverse distance weighting (IDW) method, which is very dependent on an appropriate weight...
function (Franke 1982). Hardy (1971) formulated a multiquadric method (MQM) under the assumption that any smooth mathematical surface and smooth irregular surface might be approximated to any desired degree of exactness by the summation of mathematically defined surfaces. Harder and Desmarais (1972) discovered a method of surface splines, which is based on the small deflection equation of an infinite plate. Maude (1973) inspired a rectangle-based blending method that generalized the idea of deficient quintic splines to several variables. Triangle-based blending methods were developed for surface estimation by improving IDW (Gold et al. 1977, McLain 1976). Talmi and Gilat (1977) constructed a method for smooth approximation of data on the basis of spline idea, which had a great advantage compared to polynomial interpolation. A method of delta iteration was proposed for producing a surface approximation (Foley and Nielson 1980), which is based on Bernstein polynomials and bicubic spline.

The accuracy of classic methods for surface modeling has been comparatively analyzed by many scholars. For instances, Yakowitz and Szidarovsky (1985) presented that nonparametric regression was more robust. Laslett and McBratney (1990) compared IDW, kriging, and spline by using a carefully and specially designed survey of soil pH; they believed that kriging was the best method. Weber and Englund (1992) concluded that IDW is superior to kriging. Wood and Fisher (1993) assessed the simulation accuracy of IDW, contour flood filling, simultaneous over-relaxation, and spline; they concluded that each of the techniques emphasized a particular aspect of the elevation model and no one technique was sufficient. Hutchinson and Gessler (1994) showed that spline was more accurate than kriging. Myers (1994) concluded that interpolation from spatial data could be performed in various ways by invoking different assumptions and models. Brus et al. (1996) used kriging, IDW, and spline to estimate soil properties at points and found that kriging was more reliable in the sense that it estimated all soil properties well. Desmet (1997) applied IDW, kriging, and spline to constructing a digital elevation model from irregularly spaced sample points and found that spline yielded the best results with regard to accuracy. Borga and Vizzaccaro (1997) compared kriging with MQM; they indicated that kriging performed better at a lower gauge density, whereas the accuracies of both estimators were similar at a higher gauge density. Carrara et al. (1997) found that each method exhibited its own advantages and pitfalls; their respective applicabilities are largely dependent on the way that source data are collected and stored. Caruso and Uarta (1998) compared IDW, kriging, MQM, and tension finite difference methods; they found that each method had its own advantages and drawbacks and depended strongly on the characteristics of the point data set. Mitas and Mitasova (1999) found that different methods could produce quite different spatial representations and an in-depth knowledge of the phenomenon was needed to evaluate which one was the closest to reality. Laslett and McBratney (1990) calculated the accuracies of kriging and spline in terms of several real data sets; they reported that kriging never performed worse than spline (Laslett 1994). Zimmerman et al. (1999) demonstrated that kriging was substantially superior to IDW over all levels of surface type, sampling pattern, noise, and correlation. Kravchenko (2003) found that kriging with known variogram parameters performed significantly better than IDW for most studied cases, while it was much less accurate than IDW when a reliable sample variogram could not be obtained because of either an insufficient number of data points or too large a distance between the data points. Aguilar et al. (2005) comparatively analyzed the IDW and radial basis function (RBF) group, which included MQM, inverse multiquadric method (IMQM), multilog interpolation method (MIM), natural cubic splines (NCS), and thin plate splines (TPS); results showed that MQM in the RBF group was the best method in terms of accuracy and numerical stability. Chaplot et al. (2006) evaluated the performance of IDW, ordinary kriging (OK), universal kriging (UK), MQM,
and regularized spline with tension (RST); their results indicated that few differences existed between the methods when the sampling density was high: kriging had a higher accuracy if the spatial structure of altitude was strong; IDW and RST performed better when the spatial structure of height was weak; MQM performed well in the mountains. Huang and Chen (2007) proved that kriging performed better than spline in some situations, but was outperformed by spline in others.

Several scholars have paid attention to computational speed in their comparative studies. Franke (1982) evaluated IDW, spline, TIN, FEM, and MQM and concluded that IDW performed well because it required moderate storage and computation time. Mitas and Mitasova (1999) pointed out that reliable methods of surface modeling should satisfy several important demands such as accuracy and high computational speed. Katzil and Doytsher (2000) compared computational speeds of different methods and indicated that improved cubic spline was 85 times as fast as kriging when they were applied to large area grid. Related comparative studies demonstrated that the high computational cost remains a significant problem of surface modeling (Lee et al. 1997, Aguilar et al. 2005). Cornford (2005) reviewed comparison studies on different methods for surface modeling; he thought that a good prior knowledge of what you are trying to simulate was most important and a good method should be chosen in terms of computational speed, cost of implementation, scaling with data size, and the ability to make probabilistic predictions. Pringle et al. (2009) concluded that the primary disadvantage of kriging was its relatively slow computing speed.

To find a solution for the error problems that had long troubled surface modeling, we developed a method of high-accuracy surface modeling (HASM) in terms of the fundamental theorem of surfaces. Numerical and real-world tests demonstrated that HASM is much more accurate than the classic methods (Yue et al. 2007, 2009, Yue and Song 2008). However, HASM had a slow computing speed because an iterative method was used to solve its partial differential equation set, which limited its capacity of processing huge data. Thus, in this article, adjustment computation of HASM (HASM-AC) is developed to speed up the computation. Performances of HASM-AC are compared with the ones of IDW, kriging, and spline, which have been the most widely used in geographical information systems.

2. Adjustment computation of HASM

A primary purpose of an adjustment computation is to ensure that all observations are used to find the most probable values for the unknowns in the model. The combination of stochastic and functional models results in a mathematical model for the adjustment. The determination of variances, and subsequently the weights of the observations, is known as the stochastic model. A functional model in adjustment computations is an equation or set of equations that represents or defines an adjustment condition (Ghilani and Wolf 2006).

The adjustment computation is rigorously based on probability theory. The sum of the squares of the errors times their respective weights is minimized in the adjustment computation. By enforcing this condition in any adjustment, the set of errors that is computed has the highest probability of occurrence. The adjustment computation permits all observations, regardless of their number or type, to be entered into the adjustment and used simultaneously in the computations by means of least squares. Least squares has the advantage that after an adjustment has been finished, a complete statistical analysis can be made of the results. Various tests can be conducted to determine if a survey meets acceptable tolerances or whether the observations must be repeated.
It is usually convenient to think the sampled values as arising from an underlying function, \( f(x, y) \), which is not necessarily known (Franke 1982). The sampled values are not assumed to satisfy any particular conditions as to spacing or density, that is, scattered. Suppose that the coordinates of grid cells are \((m, n)\) where the sampled points are located, corresponding values are \( f_{m,n} \), and grid-cell size is \( h \). Then, the finite difference equation set of HASM could be formulated as follows (Yue et al. 2007, 2009):

\[
\begin{align*}
\frac{f_{i+1,j} - 2f_{i,j} + f_{i-1,j}}{h^2} &= (\Gamma^1_{11})_{i,j} f_{i+1,j} - f_{i-1,j} + (\Gamma^2_{11})_{i,j} f_{i,j+1} - f_{i,j-1} + \frac{L_{i,j}}{\sqrt{E_{i,j}} + G_{i,j} - 1} \\
\frac{f_{i,j+1} - 2f_{i,j} + f_{i,j-1}}{h^2} &= (\Gamma^1_{22})_{i,j} f_{i+1,j} - f_{i-1,j} + (\Gamma^2_{22})_{i,j} f_{i,j+1} - f_{i,j-1} + \frac{N_{i,j}}{\sqrt{E_{i,j}} + G_{i,j} - 1} \\
\frac{f_{i+1,j+1} - f_{i-1,j+1} - f_{i+1,j-1} - f_{i-1,j-1}}{4h^2} &= (\Gamma^1_{12})_{i,j} f_{i+1,j} - f_{i-1,j} + (\Gamma^2_{12})_{i,j} f_{i,j+1} - f_{i,j-1} + \frac{M_{i,j}}{\sqrt{E_{i,j}} + G_{i,j} - 1} \\
f_{m,n} &= f_{m,n}
\end{align*}
\]

(1)

where \((\Gamma^1_{11})_{i,j}, (\Gamma^1_{22})_{i,j}, (\Gamma^2_{11})_{i,j}, (\Gamma^2_{12})_{i,j}, (\Gamma^2_{22})_{i,j}\) are definite differences of the Christoffel symbols of the second kind; \(E_{i,j}, F_{i,j}, G_{i,j}\) and \(G_{i,j}\) are the first fundamental coefficients; \(L_{i,j}, M_{i,j},\) and \(N_{i,j}\) are the second fundamental coefficients. They can be formulated as follows:

\[
E_{i,j} = 1 + \left(\frac{f_{i+1,j} - f_{i-1,j}}{2h}\right)^2
\]

(2)

\[
G_{i,j} = 1 + \left(\frac{f_{i,j+1} - f_{i,j-1}}{2h}\right)^2
\]

(3)

\[
F_{i,j} = \left(\frac{f_{i+1,j} - f_{i-1,j}}{2h}\right) \left(\frac{f_{i,j+1} - f_{i,j-1}}{2h}\right)
\]

(4)

\[
L_{i,j} = \frac{f_{i+1,j} - 2f_{i,j} + f_{i-1,j}}{h^2} \sqrt{1 + \left(\frac{f_{i+1,j} - f_{i-1,j}}{2h}\right)^2 + \left(\frac{f_{i,j+1} - f_{i,j-1}}{2h}\right)^2}
\]

(5)

\[
N_{i,j} = \frac{f_{i,j+1} - 2f_{i,j} + f_{i,j-1}}{h^2} \sqrt{1 + \left(\frac{f_{i+1,j} - f_{i-1,j}}{2h}\right)^2 + \left(\frac{f_{i,j+1} - f_{i,j-1}}{2h}\right)^2}
\]

(6)

\[
M_{i,j} = \frac{f_{i+1,j+1} - f_{i-1,j+1} - f_{i+1,j-1} - f_{i-1,j-1}}{4h^2} \sqrt{1 + \left(\frac{f_{i+1,j} - f_{i-1,j}}{2h}\right)^2 + \left(\frac{f_{i,j+1} - f_{i,j-1}}{2h}\right)^2}
\]

(7)
\[
\begin{align*}
(G_1^1)_{i,j} & = \frac{G_{i,j} \cdot (E_{i+1,j} - E_{i-1,j}) - 2F_{i,j} \cdot (F_{i+1,j} - F_{i-1,j}) + F_{i,j} \cdot (E_{i,j+1} - E_{i,j-1})}{4 \left( E_{i,j} \cdot G_{i,j} - (F_{i,j})^2 \right) h} \\
(G_1^2)_{i,j} & = \frac{2E_{i,j} \cdot (F_{i+1,j} - F_{i-1,j}) - E_{i,j} \cdot (E_{i,j+1} - E_{i,j-1}) - F_{i,j} \cdot (E_{i+1,j} - E_{i-1,j})}{4 \left( E_{i,j} \cdot G_{i,j} - (F_{i,j})^2 \right) h} \\
(G_2^1)_{i,j} & = \frac{2G_{i,j} \cdot (F_{i,j+1} - F_{i,j-1}) - G_{i,j} \cdot (F_{i+1,j} - F_{i-1,j}) - F_{i,j} \cdot (G_{i,j+1} - G_{i,j-1})}{4 \left( E_{i,j} \cdot G_{i,j} - (F_{i,j})^2 \right) h} \\
(G_2^2)_{i,j} & = \frac{E_{i,j} \cdot (G_{i,j+1} - G_{i,j-1}) - 2F_{i,j} \cdot (F_{i,j+1} - F_{i,j-1}) + F_{i,j} \cdot (G_{i+1,j} - G_{i-1,j})}{4 \left( E_{i,j} \cdot G_{i,j} - (F_{i,j})^2 \right) h} \\
\end{align*}
\]

The finite difference equation set means that 5 \times 5 grid cells are necessary for a solution and there are maximum 3 + \( u \) constraint conditions. \( u \) represents the number of sampling points in the 5 \times 5 grid cells and 0 \( \leq u \leq 9 \). The 9 inner grid cells among the 5 \times 5 grid cells are used for the solving process and the 12 grid cells on the boundary are only used to calculate the first and second fundamental coefficients (Figure 1).

The correction vector, \( V \), can be expressed as follows:

\[
V_{9 \times 1} = \begin{bmatrix} v_{i-1,j-1} \\ v_{i-1,j} \\ v_{i-1,j+1} \\ v_{i,j-1} \\ v_{i,j} \\ v_{i,j+1} \\ v_{i+1,j-1} \\ v_{i+1,j} \end{bmatrix} \rightarrow \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{bmatrix}
\]

The coefficient matrix can be rewritten as
\[ D = \begin{bmatrix} \frac{2 + h(G_{11}^2)_{i,j}}{2h^2} & 0 & -2 & \frac{2 - h(G_{11}^1)_{i,j}}{2h^2} & 0 \\ 0 & \frac{2 + h(G_{22}^2)_{i,j}}{2h^2} & 0 & \frac{2 - h(G_{22}^1)_{i,j}}{2h^2} & 0 \\ -h^2 & 0 & \frac{2 - h(G_{11}^2)_{i,j}}{2h^2} & 0 & -\frac{(G_{11}^1)_{i,j}}{2h} \\ \frac{1}{4h^2} & -\frac{1}{4h^2} & \frac{(G_{12}^2)_{i,j}}{2h} & -\frac{(G_{12}^1)_{i,j}}{2h} & \frac{1}{4h^2} \end{bmatrix} \]

where \( D \) is determined by Gauss equation set and \( R \) is determined by sampling points.

The Equation set (1) can be reformulated as follows:

\[
\begin{align*}
(f_{i,j} + v) - 2(f_{i,j} + v) + (f_{i-1,j} + v) \\
= (G_{11}^1)_{i,j} + \frac{h^2}{2h} - (f_{i-1,j} + v) + (G_{11}^2)_{i,j} + \frac{h^2}{2h} - (f_{i-1,j} + v) + N_{i,j} \\
= (G_{22}^1)_{i,j} + \frac{h^2}{2h} - (f_{i-1,j} + v) + (G_{22}^2)_{i,j} + \frac{h^2}{2h} - (f_{i-1,j} + v) + M_{i,j} \\
= (G_{12}^1)_{i,j} + \frac{h^2}{2h} - (f_{i-1,j} + v) + (G_{12}^2)_{i,j} + \frac{h^2}{2h} - (f_{i-1,j} + v) + N_{i,j} \\
\end{align*}
\]

Then, we have

\[
D = \begin{bmatrix} \frac{2 + h(G_{11}^2)_{i,j}}{2h^2} & 0 & -2 & \frac{2 - h(G_{11}^1)_{i,j}}{2h^2} & 0 \\ 0 & \frac{2 + h(G_{22}^2)_{i,j}}{2h^2} & 0 & \frac{2 - h(G_{22}^1)_{i,j}}{2h^2} & 0 \\ -h^2 & 0 & \frac{2 - h(G_{11}^2)_{i,j}}{2h^2} & 0 & -\frac{(G_{11}^1)_{i,j}}{2h} \\ \frac{1}{4h^2} & -\frac{1}{4h^2} & \frac{(G_{12}^2)_{i,j}}{2h} & -\frac{(G_{12}^1)_{i,j}}{2h} & \frac{1}{4h^2} \end{bmatrix} \]

(15)

(16)
\[ W = \frac{W}{(3+u) \times 1} \]

\[
\begin{pmatrix}
\frac{I_1}{2h} \left( f_{i+1,j} - f_{i-1,j} \right) + \frac{I_2}{2h} \left( f_{j+1} - f_{j-1} \right) + \frac{L}{E_{i,j} + G_{i,j} - 1} \left( f_{i+1,j} - 2f_{i,j} + f_{i-1,j} \right) \\
\frac{I_1}{2h} \left( f_{i+1,j} - f_{i-1,j} \right) + \frac{I_2}{2h} \left( f_{j+1} - f_{j-1} \right) + \frac{N}{E_{i,j} + G_{i,j} - 1} \left( f_{i+1,j} - 2f_{i,j} + f_{i-1,j} \right) \\
\frac{I_1}{2h} \left( f_{i+1,j} - f_{i-1,j} \right) + \frac{I_2}{2h} \left( f_{j+1} - f_{j-1} \right) + \frac{M}{E_{i,j} + G_{i,j} - 1} \left( f_{i+1,j} - 2f_{j,j} + f_{j-1,j} \right) \\
0 \\
0 \\
0
\end{pmatrix}
\]

\[(18)\]

The number of rows of \( R \) is determined by the number of sampling points of \( u (u \leq 9) \). The element of every row vector of \( R \) takes its value of 1 when it corresponds to an element of \( V = [v_0, v_1, \ldots, v_8] \) or else it takes 0.

When the solution procedure is completed, spatial distribution of simulation error \( \sigma \) can be expressed as follows:

\[ \sigma^2 = \sigma_0^2 \cdot Q \]

where \( \sigma_0^2 = \frac{V^T \cdot P \cdot V}{(3+u)} = \frac{-W^T \cdot K}{(3+u)} \); \( Q = \frac{P^{-1} \cdot A^T \cdot S^{-1} \cdot A \cdot P^{-1} \cdot V}{(3+u) \times 1} \); \( V \) are correction values; \( P \) is weight matrix; \( K \) are connection numbers; \( W = \frac{-A \cdot V}{(3+u)} \), \( A \) is coefficient matrix of conditional equations; and \( S = \frac{A \cdot P^{-1} \cdot A^T}{(3+u) \times 1} \).

If \( n \) and \( m \) are, respectively, the number of computational units consisting of 5 \( \times 5 \) grid cells in latitudinal and longitudinal directions, the number of computational units in a computational domain would be \( 15(n-1) \cdot (m-1) + n \cdot m \). It should be noted that the computational unit is simply termed ‘unit’ later. The number of equations to be solved is \( 15(n-1) \cdot (m-1) + n \cdot m + u \), in which \( u \) is the number of sampling points. If computational domain would be very big, the computation would become too huge to be conducted. For solving this problem, the huge equation set is grouped into many small equation sets and every small equation set is successively solved by means of the method for successive independent conditional adjustment. The solution vector of each small equation set, which is the initial values plus their correction values, is taken as virtual observed values for solving the next small equation set. This procedure is continued until the final solution is obtained. The solution of one holistic computation equals the solution of grouped successive computations.

The solution vector in unit \( \Psi \) is independent of the one in its neighbor unit \( \Theta \) of which the distance from the center of unit \( \Psi \) is 4 grid cells. Thus, all these independent units can form one group. One computational domain can be classified into 16 such groups at maximum (Figure 2). In each group, distance between centers of the neighbor units is equal to 4 grid cells, so that the units in the group can be dealt with according to the location of the starting unit center and its shift.
In every group, the normal equation set can be expressed as

\[ S \cdot K + W = 0 \]  \hspace{1cm} (20)

where

\[ S = \begin{bmatrix}
q_1 \\
\vdots \\
q_3 \\
\vdots \\
q_i \\
\vdots \\
q_k
\end{bmatrix}
\]

and \([q_1], [q_2], [q_3], \ldots, [q_k]\) are the matrices corresponding to the independent computational units.

According to the principle of matrix partitioning, \( S^{-1} = \begin{bmatrix}
[q_1]^{-1} \\
\vdots \\
[q_3]^{-1} \\
\vdots \\
[q_i]^{-1} \\
\vdots \\
[q_k]^{-1}
\end{bmatrix}\).

Thus, the whole solving process is transformed into successively solving the 16 groups and the solving process of each group is transformed into solving the equation set in independent units successively. According to this principle, the whole solving process is a repetition of solving on the unit of 5 \(\times\) 5 grid cells and the maximum number of equations to be solved is 12. Thus, the computation becomes very small and a rather small storage space is utilized. The computational time of HASM-AC can be calculated by \( t = \sum_{i=1}^{q} t_i \) where \( q = 15(n-1) \cdot (m-1) + n \cdot m \), and \( t_i \) is the computation time of the \( i \)th unit.
3. Tests of accuracy and computing speed

The accuracies and speeds of IDW, kriging, spline, and HASM-AC are comparatively calculated, in which higher accuracy means lower error. The classical methods are performed by the 3D Analyst module in ArcGIS 9.2. Default parameters are employed by the software. For IDW, the power is 2, the search radius is variable, and the maximum number of the searched points is 12. For kriging, the ordinary method is selected, the model of semivariogram is spherical, the search radius is variable, and the maximum number of the search points is 12. For spline, the regularized option is used, the weight is 0.1, and the number of the searched points is 12.

HASM-AC is run through six steps: (1) to compute initial values of all grid cells in the computational domain in terms of the sampled data; (2) to flag the sampled points as control points and to transform coordinate of the control points to grid coordinate if a control point might not locate at a grid-cell center (a lattice center), it would be extended to its closest lattice center by means of the Taylor expansion method; (3) to set a weight threshold for the control points; (4) to create weight matrix in terms of the weight threshold; (5) to start simulation on the basis of the initial values and the weight matrix; and (6) to output the simulated results including simulated surface, histogram of absolute error (AE), error surface, root mean-square error (RMSE), and computational time.

3.1. Errors

An error is defined as the difference between the sampled value and the simulated value. In the tests, AE and RMSE are calculated. They are, respectively, formulated as

\[
AE = f_{i,j} - Sf_{i,j}
\]

and

\[
RMSE = \left( \frac{1}{I \times J} \sum_{i=1}^{I} \sum_{j=1}^{J} (f_{i,j} - Sf_{i,j})^2 \right)^{-1/2}
\]

where \(Sf_{i,j}\) is the simulated value of \(f(x,y)\) at lattice \((i,j)\); \(f_{i,j}\) is the true value or the sampled value of \(f(x,y)\) at the lattice \((i,j)\); and \(i = 1, 2, ..., I, j = 1, 2, ..., J\).

3.1.1 A numerical test

A mathematical surface (Figure 3) is selected to test the accuracies of HASM-AC and the classical methods so that the true value is able to be predetermined to avoid the uncertainty caused by uncontrollable data errors. The mathematical surface has the following formulation:

\[
z(x,y) = 3 + 2 \sin(2\pi x) \sin(2\pi y) + \exp(-15(x - 1)^2 - 15(y - 1)^2) \\
+ \exp(-10x^2 - 15(y - 1)^2)
\]
Computational domain of the test surface is normalized as $[0, 1] \times [0, 1]$ and is orthogonally divided into $100 \times 100$ grid cells, that is, grid spacing was 0.01. Sampling interval was defined as 0.02, which means that the sampling proportion is 25%.

The error histograms indicate the general patterns of AEs from different methods (Figure 4). The vertical coordinate represents frequency, that is, the count of grids at which there is an equal AE. Abscissa axis represents the AE. The AE ranges of IDW, kriging, and spline are, respectively, $[-0.330 0.332]$, $[-0.168 0.168]$, and $[-0.065 0.065]$, while that for HASM-AC is $[-0.008 0.008]$. For IDW, kriging, and spline, the counts of grids where AE is bigger than 0.008 are 4591, 4599, and 4602, respectively; the ones where AE is less than −0.008 are 4597, 4596, and 4606. In other words, only 8.21%, 8.05%, and 7.92% grid cells simulated by IDW, kriging, and spline, respectively, are as accurate as the ones by HASM-AC.

RMSEs of IDW, kriging, and spline are, respectively, 0.062, 0.047, and 0.044, whereas that of HASM-AC is 0.003. In other words, accuracy of HASM-AC is 20.67, 15.67, and 14.67 times higher than IDW, kriging, and spline, respectively.

The error surfaces indicate that IDW, kriging, and spline have similar error magnitudes and distribution patterns (Figure 5a–c). The biggest AE appears in the peak and hollow areas. They evidently have peak-cutting and hollow-filling problems. Systematic errors are created in the simulation processes of IDW, kriging, and spline. But the error of HASM-AC (Figure 5d) has quite smaller magnitudes and an even distribution, which is greatly different from the ones of IDW, spline, and kriging.

The sampling proportion is increased to 34% by adding sample points at the grid cells whose both row number and column number are exactly divisible by 3. Then, the numerical test indicates that the AE ranges of IDW, kriging, spline, and HASM-AC become $[-0.078 0.079]$, $[-0.011 0.011]$, $[-0.00015 0.00014]$, and $[-0.00011 0.00009]$, respectively. RMSEs decrease to 0.012, 0.002, 0.0002, and 0.0001. In other words, the increase in sample points makes the accuracies of all methods become much higher (Figure 6).
A real-world test of accuracy

A rectangular area in Tongzhou District of Beijing is selected as the real-world test area (Figure 7). It has its length of 0.5 km from east to west and its width of 0.4 km from north to south. The North Canal of Beijing flows through the rectangular area from north to south. This test area is relatively flat, which can effectively reveal accuracy difference among the methods of surface modeling. Its elevation is greater than 14 m and less than 22 m. Total station instruments are used to randomly sample height points in the test area. Three hundred and ninety-five height points are sampled, that is, there is about one sample in each 600 m² or approximately a sampling interval of 22 m on an average. Three hundred and forty-seven sampled points were used to simulate the elevation surface and the remaining 48 height points were employed to cross-validate the simulated surface. Cross-validation consists of three steps: (1) 48 height points are removed from the 395 sampled points randomly; (2) the

Figure 4. Histograms of absolute errors when the sampling proportion is 25%: (a) from IDW, (b) from spline, (c) from kriging, and (d) from HASM-AC.

3.1.2 A real-world test of accuracy

A rectangular area in Tongzhou District of Beijing is selected as the real-world test area (Figure 7). It has its length of 0.5 km from east to west and its width of 0.4 km from north to south. The North Canal of Beijing flows through the rectangular area from north to south. This test area is relatively flat, which can effectively reveal accuracy difference among the methods of surface modeling. Its elevation is greater than 14 m and less than 22 m. Total station instruments are used to randomly sample height points in the test area. Three hundred and ninety-five height points are sampled, that is, there is about one sample in each 600 m² or approximately a sampling interval of 22 m on an average. Three hundred and forty-seven sampled points were used to simulate the elevation surface and the remaining 48 height points were employed to cross-validate the simulated surface. Cross-validation consists of three steps: (1) 48 height points are removed from the 395 sampled points randomly; (2) the
We first test the effect of different spatial resolutions, 45 m × 45 m, 35 m × 35 m, 25 m × 25 m, 15 m × 15 m, and 5 m × 5 m, on accuracy of the simulated surface by every method of surface modeling. When spatial resolution is changing from 45 m × 45 m to 5 m × 5 m, the RMSE of HASM-AC is monotonously decreasing from 0.40 to 0.11 m and reaches its minimum AE amplitude in spatial resolution of 5 m × 5 m (Table 1). However, spline, kringing, and IDW have their minimum RMSEs at the spatial resolution of 15 m × 15 m. Spline presents its minimum AE amplitude at the spatial resolution of 5 m × 5 m, kringing at 35 m × 35 m, and IDW at 25 m × 25 m. HASM-AC shows the highest accuracy at almost all spatial resolutions compared to the classic methods, except that IDW has the best performance on spatial resolutions of 25 m × 25 m and 35 m × 35 m. Accuracy of HASM-AC has its stable change trends, but the ones of the classical methods are unstable.

Then, we test the effect of sampling intervals on the simulation accuracies of all methods at the spatial resolution of 15 m × 15 m. The test results indicate that HASM-AC accuracy becomes lower monotonously when sampling points become sparser, whereas accuracies of the classic methods have an oscillating change when sampling points become sparser (Table 2). All the classic methods are unstable, whereas HASM-AC is very stable. When the sampling intervals on an average are shorter than 26 m, HASM-AC has the highest accuracy. IDW has a higher accuracy than HASM-AC when the sampling intervals are longer than 26 m. However, the accuracy of HASM-AC is always higher than the ones of spline and kringing for all sampling intervals. When sampling points are reduced from 347 to
... in the simulation process, AE amplitudes of all methods reach their smallest values, which might mean that some 48 sampling points among the sampled data have a greater impact on the whole data structure. Then, AE amplitudes of HASM-AC and kriging become bigger monotonously when the sampling intervals increase, whereas the one of spline has a considerably oscillating change. In other words, HASM-AC is the most accurate and the most stable method, whereas spline is the least accurate and the most unstable one.

The simulated surfaces (Figure 8) indicated that HASM-AC result fits the real terrain in terms of our investigation on the spot. Spline gives better terrain than IDW and kriging although it has a great oscillatory surge at the right corner, which makes its error much higher than the ones of IDW and kriging. The North Canal is not reflected clearly by IDW and kriging simulations. Especially, a branch on the left side of the North Canal does not appear in the result simulated by kriging.
Figure 7. Tongzhou test area for accuracy, in which black dots represent the sampled points.

Table 1. Effect of grid spacing on the errors.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Spatial resolution (m)</th>
<th>Minimum AE (m)</th>
<th>Maximum AE (m)</th>
<th>AE amplitude (m)</th>
<th>RMSE (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline</td>
<td>45 × 45</td>
<td>-2.47</td>
<td>1.64</td>
<td>4.11</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>35 × 35</td>
<td>-2.35</td>
<td>1.32</td>
<td>3.67</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>25 × 25</td>
<td>-1.65</td>
<td>1.70</td>
<td>3.35</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>15 × 15</td>
<td>-1.56</td>
<td>1.02</td>
<td>2.58</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>5 × 5</td>
<td>-0.50</td>
<td>1.38</td>
<td>1.88</td>
<td>0.35</td>
</tr>
<tr>
<td>Kriging</td>
<td>45 × 45</td>
<td>-1.35</td>
<td>1.05</td>
<td>2.4</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>35 × 35</td>
<td>-0.89</td>
<td>0.91</td>
<td>1.8</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>25 × 25</td>
<td>-0.96</td>
<td>1.34</td>
<td>2.3</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>15 × 15</td>
<td>-1.09</td>
<td>0.77</td>
<td>1.86</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>5 × 5</td>
<td>-0.74</td>
<td>0.75</td>
<td>1.49</td>
<td>0.36</td>
</tr>
<tr>
<td>IDW</td>
<td>45 × 45</td>
<td>-1.78</td>
<td>1.25</td>
<td>3.03</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>35 × 35</td>
<td>-0.80</td>
<td>0.68</td>
<td>1.48</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>25 × 25</td>
<td>-0.44</td>
<td>0.84</td>
<td>1.28</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>15 × 15</td>
<td>-0.39</td>
<td>0.99</td>
<td>1.38</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>5 × 5</td>
<td>-1.01</td>
<td>1.16</td>
<td>2.17</td>
<td>0.24</td>
</tr>
<tr>
<td>HASM-AC</td>
<td>45 × 45</td>
<td>-0.76</td>
<td>1.66</td>
<td>2.42</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>35 × 35</td>
<td>-1.02</td>
<td>0.56</td>
<td>1.58</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>25 × 25</td>
<td>-0.77</td>
<td>0.95</td>
<td>1.72</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>15 × 15</td>
<td>-0.66</td>
<td>1.06</td>
<td>1.72</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>5 × 5</td>
<td>-0.29</td>
<td>0.38</td>
<td>0.67</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Table 2. The effect of sampling intervals on the errors.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Number of sampled height points</th>
<th>Sampling interval on an average (m)</th>
<th>Minimum AE (m)</th>
<th>Maximum AE (m)</th>
<th>AE amplitude (m)</th>
<th>RMSE (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline</td>
<td>347</td>
<td>24</td>
<td>-1.56</td>
<td>1.02</td>
<td>2.58</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>26</td>
<td>-1.04</td>
<td>0.75</td>
<td>1.79</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>251</td>
<td>28</td>
<td>-2.12</td>
<td>1.60</td>
<td>3.72</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td>32</td>
<td>-2.15</td>
<td>1.03</td>
<td>3.18</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>155</td>
<td>36</td>
<td>-2.03</td>
<td>3.01</td>
<td>5.04</td>
<td>0.64</td>
</tr>
<tr>
<td>Kriging</td>
<td>347</td>
<td>24</td>
<td>-1.09</td>
<td>0.77</td>
<td>1.86</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>26</td>
<td>-0.94</td>
<td>0.55</td>
<td>1.49</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>251</td>
<td>28</td>
<td>-1.12</td>
<td>0.72</td>
<td>1.84</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td>32</td>
<td>-1.26</td>
<td>0.60</td>
<td>1.86</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>155</td>
<td>36</td>
<td>-1.88</td>
<td>0.34</td>
<td>2.22</td>
<td>0.48</td>
</tr>
<tr>
<td>IDW</td>
<td>347</td>
<td>24</td>
<td>-0.39</td>
<td>0.99</td>
<td>1.38</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>26</td>
<td>-0.64</td>
<td>0.61</td>
<td>1.25</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>251</td>
<td>28</td>
<td>-1.16</td>
<td>0.17</td>
<td>1.33</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td>32</td>
<td>-1.21</td>
<td>0.24</td>
<td>1.45</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>155</td>
<td>36</td>
<td>-1.67</td>
<td>0.32</td>
<td>1.99</td>
<td>0.36</td>
</tr>
<tr>
<td>HASM-AC</td>
<td>347</td>
<td>24</td>
<td>-0.66</td>
<td>1.06</td>
<td>1.72</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>26</td>
<td>-0.66</td>
<td>0.46</td>
<td>1.12</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>251</td>
<td>28</td>
<td>-0.67</td>
<td>0.54</td>
<td>1.21</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td>32</td>
<td>-1.15</td>
<td>0.73</td>
<td>1.88</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>155</td>
<td>36</td>
<td>-1.61</td>
<td>1.44</td>
<td>3.05</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Figure 8. Simulated surfaces: (a) from IDW, (b) from spline, (c) from kriging, and (d) from HASM-AC.
3.3. A real-world test of computing speed

An area with 6000 × 6000 grid cells is selected to test the computational speeds of the methods because data set sizes of the mathematical surface and the rectangular area in Tongzhou District of Beijing are not big enough to discriminate the differences in computational speeds. The area for testing computational speed is located in 100°E – 105°E and 30°N – 35°N, where Qinghai, Gansu, and Sichuan provinces meet. The data set has a spatial resolution of 3′ × 3′. This area with an elevation range between 333 and 5739 m includes various landforms such as plateau, ridgeline, hill, and plain. To the west of this area is the Tibet Plateau. The middle part has an undulating topography including crisscross valleys and widely distributed basins. The southeast is a transition area from Hengduan Mountains to the plain of western Sichuan Province. The northeast is a transition area from the plateau of southern Gansu Province to loess plateau (Figure 9).

HASM is run by means of successive independent conditional adjustments. The holistic computational process is transformed into computing independent computational units in 16 different groups. HASM-AC’s dynamically reading and storing the data reduces the occupancy of computer resources and shortens the computational time. The tests on different methods are conducted by using the same computer with Intel Core Duo, 2.67 GHz CPU, DDR2 800 2GB memory, P35 chipsets, SATA2 320GB hard disk, operating system of Windows XP, and .NET Framework 2.0. The test results demonstrate that HASM-AC spends 1.6 hours for simulating the test area, whereas spline, IDW, and kriging spend 12.84, 17.06, and 906.27 hours, respectively. It means that the computing speed of HASM-AC is 8, 11, and 563 times higher than spline, IDW, and kriging, respectively.

4. Conclusions and discussion

4.1. Conclusions

The advantages of HASM-AC can be summarized as follows: (1) the finite difference of HASM can be used to establish condition equations and conduct constraint evaluation in a computational domain under the principle of linear unbiased estimates with the minimum error variance; (2) the principle of linear unbiased estimates with the minimum error variance ensures that the simulation approaches the most probable surface; (3) the successive least squares method based on grouping computational units reduces the occupancy of computer resources considerably, which greatly shortens the computing time and makes the HASM-AC able to process a huge data; (4) data error in a computational unit does not only impact the
simulation in its own computational unit but also propagates among all computational units; HASM-AC makes accuracy very high and error impact evenly distributed spatially.

The numerical test indicates that HASM-AC has the highest accuracy. Spline accuracy is always higher than IDW and kriging. Spline accuracy improves much faster than IDW and kriging accuracies when density of the sample points increases. In the real-world test of Tongzhou District of Beijing, HASM-AC is the most accurate on average and the most stable method. Spline is the least accurate and the most unstable method because of its great oscillatory surge at the right corner. The real-world test of computing speed, in the area where Qinghai, Gansu, and Sichuan provinces meet, shows that HASM-AC has the highest computing speed. Computational time of spline is shorter than that of IDW. Kriging computation spends much longer time than all other methods. In short, HASM-AC performs the best in both numerical and real-world tests. Spline, IDW, and kriging have their own advantages and drawbacks.

4.2. Discussion

HASM-AC is theoretically based on the fundamental theorem of surfaces. In terms of the fundamental theorem, a surface is uniquely defined by the first fundamental coefficients and the second fundamental coefficients (Henderson 1998). The first fundamental coefficients of a surface yield information about some geometric properties of the surface, by which we can calculate lengths of curves, angles of tangent vectors, areas of regions, and geodesics on the surface. Those geometric properties and objects that can be determined only in terms of the first fundamental coefficients of a surface are called the intrinsic geometric properties. The collection of these geometric properties and objects forms the subject of intrinsic geometry of a surface, which studies those of its properties that do not depend on the shape of the surface, but depend only on measurements that we can carry out while staying on a surface itself (Toponogov 2006). The second fundamental coefficients reflect the local warping of the surface, namely its deviation from tangent plane at the point under consideration (Liseikin 2004), which can be observed from outside Earth. HASM-AC is theoretically perfect, which insures that it performs best.

The classical methods, such as IDW, kriging, and spline, have various theoretical shortcomings. IDW uses an inverse distance weighting function of sampling values to determine the interpolation value for any given point within the calculated area, but it fails to incorporate the spatial structure and ignores information beyond the neighborhood of the sampling values (Magnussen et al. 2007). Kriging tries to have the mean error equal to zero and aims at minimizing the variance of the errors, but its goals are practically unattainable since the mean error and the variance of errors are always unknown (Isaaks and Srivastava 1989). Spline uses univariate cubic basis-splines to approximately simulate surfaces, but a few types of surfaces fit the formulation of univariate cubic basis-splines (Watt 2000). Therefore, errors of all these classical methods are difficult to be controlled in the processes of surface modeling because these methods are not based on the inherence of surfaces.

HASM-AC conducts the computation in 16 groups and data fusion among the groups must be implemented in the simulation process, which might spend much time. Therefore, HASM-AC needs to be further improved. Theoretically, a complete simulation of a surface needs to conduct, one by one, the 16 grouping computations. The accuracy improvement will become less and less when the grouping computation process continues further. Some of the grouping computations might be unnecessary to be further conducted when accuracy requirement has been satisfied. In further research work, configuration of the groups will be studied to find out the dominant connection groups and do rational planning for each of them, through which HASM-AC would be more applicable and more efficient.
It comes naturally to HASM-AC to be extended to a parallel algorithm. Most of the currently used parallel systems fall into single instruction stream/multiple data streams (SIMD) or multiple instruction streams/multiple data streams (MIMD) categories (Kontoghiorghes 2006). In SIMD, all processors are given the same instruction; each processor operates on different data, and the processors may sit out a sequence of instructions. In MIMD, each processor runs its own instruction sequence; each processor works on a different part of the problem; each processor communicates data to other parts; and processors may have to wait for other processors or for access to data. The parallel algorithm of HASM-AC would be based on SIMD.

Acknowledgments
This work was supported by the China National Science Fund for Distinguished Young Scholars (40825003), by the National High-Tech R&D Program of Ministry of Science and Technology of the People's Republic of China (2006AA122219), by the Free Research Project of State Key Laboratory of Resources and Environment Information System (081105), and by the National Key Technologies R&D Program of the Ministry of Science and Technology of the People's Republic of China (2006BAC08B).

References


