

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Modeling and Visualization of Scattered Positive Data

¹Ghulam Mustafa and ²Abad Ali Shah

¹Department of Computer Science and Engineering, UET Lahore, Pakistan

²R and D Center of Computer Science, UET, Lahore, Pakistan

Abstract: This study discusses the problem of modeling inherently positive data sets using Modified Quadratic Shepard interpolation method. There are number of scientific and business domains where data is inherently positive. For example numbers of people, mass, volume and percentage mass concentration are meaningless when negative. This interpolant generates negative values while modeling such data that may cause ambiguity. We present a scaling technique that constrains the interpolant to produce non-negative graph through scattered positive data sets.

Key words: Visualization, positivity, interpolation and scattered data

INTRODUCTION

Data visualization is an important tool used for analysis and understanding of observed phenomena. Scattered data samples are encountered in many areas of business, science and engineering. Since common visualization methods require an underlying grid, it is required to approximate the reality at the same grid. To approximate the unknown reality at the required grid, we need to construct the representation of the sampled data. Many interpolation methods are available to construct an empirical model from the scattered data sets. Characteristics of the data to be visualized are important consideration while selecting the interpolant for construction of the empirical model. We know number of characteristics of the data to be explored like spatial distribution (1D, 2D, 3D), type of data (scalar, vector etc), data distribution (scattered or ordered data), coordinate system, scale, continuity etc.

In addition to these features our visualization must not contradict the known characteristics about data otherwise the visualization will not be trustable. For example one would not trust in the reality discovered using a visualization that produces physically erroneous results. The problem of visualization based on inherent characteristics of data exists in many application domains. Examples of few such features are positivity, monotonicity, convexity and bounds that are encountered in various scientific and engineering domains. These are usually known facts about the datasets to be explored. When we are presenting data in visual format the visualization must not contradict such known realities inherent to the data.

Study of many researchers has been reported in literature that preserve the above mentioned features of data on regular grid, but relatively less attention has been given to scattered data. Few researchers have investigated the visualization of the scattered data based on the inherent features of the data. We refer to the work of (Sarfranz, 2000; Fritsh and Carlson, 1980; Schmidt and Hess, 1988) to preserve monotonicity of data on regular grid and (Han and Schumaker, 1979) for scattered data. Only a few researchers have worked to preserve the convexity of data. We refer to the work of (Schmidt, 1990) for convex, monotone and positive data on regular grid.

Positivity of data is encountered in many situations in science, engineering and business. The significance of positivity lies in the fact that sometimes it does not make sense to talk of some quantity to be negative. For example, percentage mass concentrations in a chemical reaction, volume, absolute temperature and pressure and mass of something are meaningless when negative. Many researchers have also considered the problem of positivity. For related literature and background we refer to the study of (Nadler, 1992; Sarfranz *et al.*, 2000; Schumaker, 1982; Schmidt and Hess, 1988; Brodlied and Butt, 1993).

For scattered data, the inverse distance weighted methods are considered better due to their efficiency, extendibility to higher dimensions, global in nature and ease in implementation. Modified Quadratic Shepard (MQS) method, being smooth with C^1 continuity, is a commonly used method in the field of science and engineering such as geophysics, astronomy and meteorology. However, it has problem when it is used for visualization of positive data as it interpolates negative

values for inherently positive data. In this study we will present a scheme namely fixed point scaling scheme to preserve positivity while modeling scattered datasets that are inherently positive.

SHEPARD METHOD

MQS is an inverse distance weighted method that is based on the approach introduced by Shepard (1968). Let a set of N non-negative data values, f_i at associated scattered sampling locations, $X_i = (x_{1i}, x_{2i}, x_{3i}, \dots)$, where $i = 1, 2, \dots, N$, is given. Shepard interpolant is defined as follow:

$$F(X) = \frac{\sum_{i=1}^N w_i(X) f_i}{\sum_{i=1}^N w_i(X)} \quad (1)$$

where

$$w_i(X) = \frac{1}{d_i(X)}$$

and

$$d_i(X) = [(x_1 - x_{1i})^2 + (x_2 - x_{2i})^2 + \dots]$$

This interpolant provides one of the solutions to the problem of visualization of positive data sets. However it is not a suitable choice for many visualization applications. Few problems have been discussed in (Xiao *et al.*, 1996) and (Xiao and Woodburry, 1999). For example $F(X)$ is bounded between maximum and minimum in the data set. However the sampled values do not always contain all extremes of the quantity. The bound preservation property may also result in loss of information carried by the original sampled data in the subsequent steps of visualization process. This interpolant also has an unnecessary property that slope at each reference point is zero as shown in Fig. 1. Continuity of this interpolant is not acceptable for many applications. Also as it is a global method so it becomes inefficient for large data sets.

Modified quadratic Shepard method: Numbers of modifications have been suggested to overcome the drawbacks of the original Shepard's method. The most interesting modification for data visualization perspective is due to (Frank and Neilson, 1980). They improved continuity of the method by replacing constant basis function, f_i , by quadratic basis function, $Q_i(X)$, with the following characteristics:

$$Q_i(X_i) = f_i$$

$Q_i(X)$ is inverse distance weighted least square fit to the other data points. This made the method a C^1 continuous method.

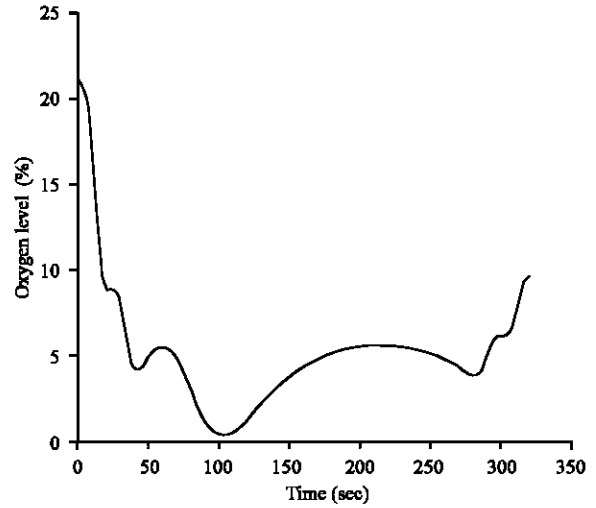


Fig. 1: Plot of data in Table 1 using Shepard method

The resulting interpolating function, $F(X)$, called Modified Quadratic Shepard method, is defined as follows:

$$F(X) = \frac{\sum_{i=1}^N w_i(X) Q_i(X)}{\sum_{i=1}^N w_i(X)} \quad (2)$$

The basis functions $Q_i(X)$ is defined as follow:

$$Q_i(X) = f_i + g_i^T (X - X_i) + (X - X_i)^T A_i (X - X_i) \quad (3)$$

By definition $Q_i(X_i) = f_i$ and is an inverse distance weighted least squares approximation to the other data points. The matrix, A_i is Hessian matrix of the quadratic and g_i^T is gradient vector. The modifications given above not only improve continuity of the interpolant but also eliminate the problem of missing data values with the original Shepard method.

To overcome the inefficiency due to global nature of Shepard method, modified weight functions suggested by (Frank and Neilson, 1980) and (Renika, 1988) are important. Weight function due to Frank and Neilson are defined as follows:

$$W_i(X) = \left[\frac{R - d_i(X)_+}{R d_i(X)} \right]^2$$

$$\text{where } [R - d_i(X)]_+ = \begin{cases} R - d_i(X) & \text{if } R > d_i(X) \\ 0 & \text{otherwise} \end{cases}$$

R is the radius within which the nodes take part for construction of $Q(X)$ and calculation of weight function

$w_i(X)$. Frank and Neilson (1980) suggested two fixed values for R i.e., R_q for construction of Q and R_w for calculation of weight.

$$R_q = \frac{D}{2} \sqrt{\frac{N_q}{N}} \quad \text{and} \quad R_w = \frac{D}{2} \sqrt{\frac{N_w}{N}}$$

where $D = \max_{i,j} |X_i - X_j|_2$

N_q are the number of control points used to calculate $Q(X)$, N_w are the number of control points used to calculate $w(X)$ and N is total number of control points. D is maximum distance between two points in the data set. Suggested values for evenly distributed 2D data sets are $N_w = 9$ and $N_q = 18$. For sparse data or where data sets are small ($N < 25$) considerable increase in N_q and N_w is required with $N_q/N_w = 2$.

Renika (1988) obtained improvement in accuracy using different criteria and values of N_q and N_w . These are the number of control points that impose significant effect for the reference point. For 2D data sets suggested values for good results are $N_w = 19$ and $N_q = 13$.

The modification given above also improves extrapolation capability of the method. The resulting interpolant called Modified Quadratic Shepard method is a suitable choice for many scientific visualization domains.

Loss of positivity using MQS method: The Modified Quadratic Shepard Method does not guarantee to preserve positivity. In Table 1 samples of oxygen percentage in flue gases of a boiler with respect to time are given. Mass concentration is always positive. A negative mass concentration value does not make sense. So we expect interpolated values to be positive. Using Modified Quadratic Shepard method a curve has been constructed in Fig. 2 for this data set. The graph goes negative for the inherently positive data. The plot shows negative mass value that is ambiguous. A number of solutions to the problem of positivity preservation using this method have been suggested. We will discuss and compare some of the solutions in the following subsection.

Earlier work for preservation of positivity using MQS method: There are number of ways to preserve positivity using Modified Quadratic Shepard method. Basis functions are common target to achieve the goal of positivity. As the weight functions, $w(X)$, of the interpolant are always positive so the function, $F(X)$, will never go negative if all basis functions are positive. Most of the methods concentrate on basis function for preservation of positivity of the interpolant.

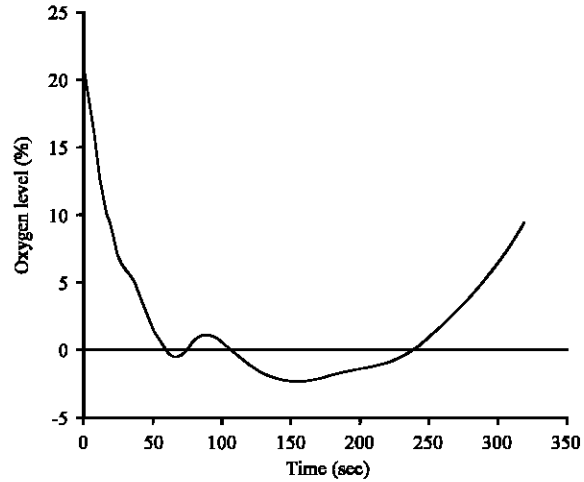


Fig. 2: Plot of data in Table 1 using MQS method

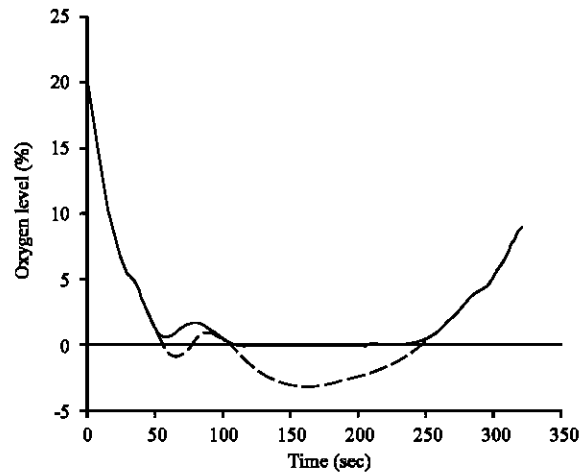


Fig. 3: Plot of data in Table 1 using MQS (Dashed) and truncated basis functions (Solid)

Table 1: Oxygen levels in flue gases from a boiler

Time (sec)	0	20	40	100	280	300	320
Oxygen (%)	20.8	8.8	4.2	0.5	3.9	6.2	9.6

Basis function truncation: As the weight functions of MQS interpolant are always positive, so the function, $F(X)$, will never go negative if all basis functions, $Q(X)$, are positive. One way to ensure positivity of $Q(X)$ is to truncate its negative part by putting all negative values to zero. This is simple solution but it has drawback that continuity of the graph is reduced to $^{\circ}C$ as shown in Fig. 3. A similar approach is suggested by (Xiao and Woodbury, 1999) however the truncation of the negative value after evaluation. The resulting continuity is usually not acceptable for visualization applications.

Constraining radius of participation: In this scheme we constrain the radius of participation of basis function for

construction of graph, R_w , such that the basis function is positive in this radius. This scheme is simple and low cost as small additional computations are involved at preprocessing stage only. However drawbacks are not tolerable. The most significant drawback is that the number of basis functions participating in construction of the graph may be too small to keep the continuity of the graph.

Simple positive modified quadratic Shepard method: This method, suggested by (Asim, 2000), scales all the basis functions that go negative. Positivity of $Q(X)$ is achieved by scaling and shifting of basis function. The original negative basis functions are replaced by the modified positive basis functions $\hat{Q}(X)$. The modified basis functions, $\hat{Q}(X)$, have following characteristics:

- It interpolate f_i i.e.,

$$\hat{Q}(X_i) = f_i$$
- For 2D data sets it is a circular paraboloid.

$$\hat{Q}(x_1, x_2) = z_k + a_i[(x_1 - x_{1i})^2 + (x_2 - x_{2i})^2]$$

The paraboloid may be a convex or concave depending upon the shape of original basis function. The symmetrical axis of the paraboloid passes through the stationary point, z_s , of the original basis function. Scaling is performed by shifting, z_s to z_k upward or downward, based on the type of original basis function and satisfying the above given conditions. Asim (2000) discussed details of the scheme and implementation results.

This scheme also has draw backs. This scheme unnecessarily perturbs basis functions and the interpolant may sometimes lose continuity with respect to data. The resultant graph may not preserve the structure of original graph.

Scaling and shifting algorithm: This is another scheme given by (Asim, 2000). Implementation of the scheme is discussed in (Asim *et al.*, 2004). This scheme scales Hessian matrix, A_i , of each negative basis function by a positive scaling factor α_i and shift the basis function to interpolate the reference point f_i . More specifically:

- It interpolates f_i i.e.,

$$\hat{Q}(X_i) = f_i$$
- Otherwise. $\hat{Q}(X) = z_k + \alpha_i[(X - X_s)^T A_i (X - X_s)]$

where A is known as Hessian Matrix and $(X - X_s)$ is a vector defining coordinate position and $(X - X_s)^T$ is

transpose of this vector. X_s is the coordinates of the stationary points of the basis function. For 2D data sets the shape of paraboloid after scaling, is similar to the original. The paraboloid may be a convex, concave or hyperbolic depending upon the shape of original basis function. The symmetrical axis of the paraboloid passes through the stationary point, z_s , of the original basis function.

Scaling is performed by shifting, z_s to z_k upward or downward, based on the type of original basis function and satisfying the above given conditions. Details of the scheme and implementation results are discussed in (Asim, 2004).

Extension of the scheme to higher dimensions is difficult as many special cases emerge due to increase of eigenvalues combinations. In addition this technique faces divided by zero problem

This study presents fixed point scaling scheme that will remove the drawbacks of the previous schemes.

Fixed point scaling: In this scaling scheme we scale only those basis functions that go negative in the region of their participation. We constrain each negative basis function to be positive by using fixed point scaling scheme as given below.

- We take reference point of the basis function as fixed point.
- We scale the basis function given in Eq. 3 by a positive scaling factor, such that the following conditions for the modified positive basis function $\hat{Q}(X_i)$ hold:

$$\hat{Q}(X_i) = f_i \quad \text{and} \quad \hat{Q}(X_m) = 0 \quad (4)$$

where X_m is the coordinates of minima of the basis function in the region R_w . The scaled basis function takes the following form:

$$\hat{Q}_i(X) = f_i + \alpha_i [g_i^T (X - X_i) + \frac{1}{2} (X - X_i)^T A_i (X - X_i)] \quad (5)$$

To find scaling factor, α_i , we apply the boundary conditions as stated above and get:

$$\hat{Q}_i(X_m) = 0 \quad \text{i.e.} \quad f_i + \alpha_i [g_i^T (X_m - X_i) + \frac{1}{2} (X_m - X_i)^T A_i (X_m - X_i)] = 0 \quad (6)$$

The minimum of the basis function is given by:

$$Q_i(X_m) = f_i + g_i^T (X_m - X_i) + \frac{1}{2} (X_m - X_i)^T A_i (X_m - X_i) \quad (7)$$

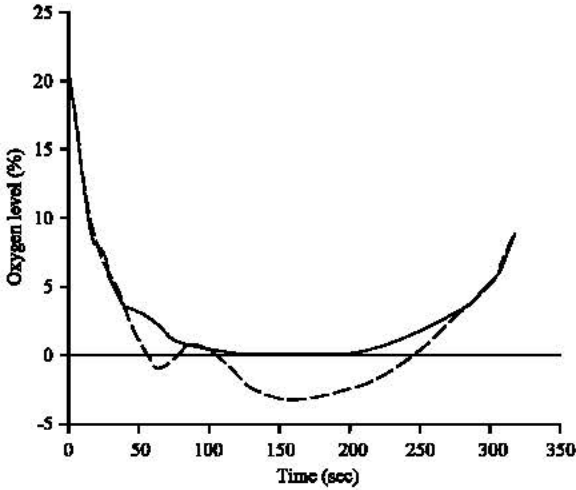


Fig. 4: Plot of data in Table 1 using MQS (Dashed) and fixed point scaling schemes (Solid)

where X_m is the point of minima
Solving Eq. 6 and 7 simultaneously we get:

$$\alpha_i = \frac{f_i}{f_i - Q_i(X_m)} \quad (8)$$

It is worth noting that α_i varies between 0 and 1.

If $f_i = 0$ then $\alpha_i = 0$. This means that the basis function becomes $Q(X) = 0$ for all cases where f_i is zero. So we treat zero reference values as a special case. For semi-positive value of $Q(X_m)$ there is no need of scaling. Scaling is required only if $Q(X_m)$ is negative. In this case scaling factor is defined by the Eq. 8.

The beauty of this scaling scheme is that it scales basis function without changing signs of eigenvalues. So the basis functions do not lose their shape during the constraining process. This preserves the shape (convex, concave and hyperbolic) of the basis function.

The implementation of this study is straight forward. A separate module for scaling is required without disturbing the existing modules. Module for scaling takes the coefficient of quadratic $Q(x)$ and returns the modified coefficients of $\hat{Q}(x)$ after multiplying with the scaling factor. We have implemented and tested the results of the research in 1D and 2D data. Plot of 1D data in Table 1 is shown in Fig. 4 and plot of 2D data is shown in Fig. 6. The results are promising as compared to the earlier variants. We have got the improvements on two fronts:

First: instead of treating the convex, concave and hyperbolic cases separately our fixed point scaling technique treat these cases uniformly. It reduces not only the programming effort but also the computational cost.

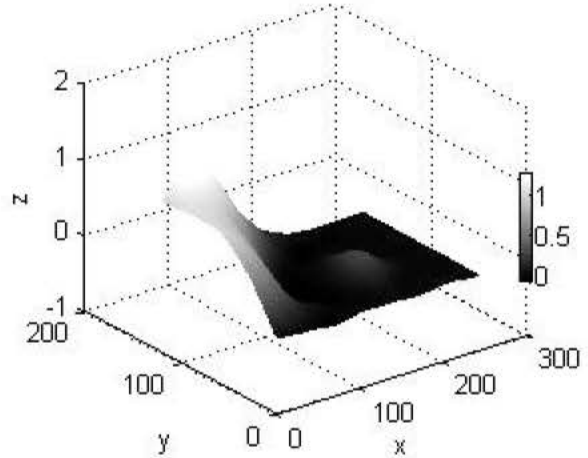


Fig. 5: Plot of test data using Modified Quadratic Shepard method

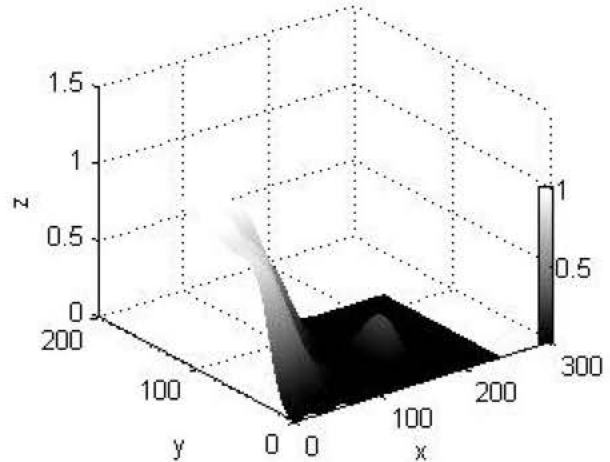


Fig. 6: Plot of test data using fixed point scaling algorithm

Moreover the earlier scaling and shifting technique was facing divided by zero problem. This problem has been eliminated in this scheme.

Second: During implementation we have considered the fact that the cases where data value is zero, the scaling factor too become zero. So we have treated it as special case that further reduces the computational effort.

For construction of examples in 2D and comparison purposes we used following test functions for generation of test data.

$$F(x,y) = \begin{cases} 1 & \text{if } (y-x) > \frac{1}{2} \\ 2(y-x) & \text{if } 0 \leq (y-x) \leq \frac{1}{2} \\ \frac{1}{2} \cos(4\pi P^{\frac{1}{2}}) + \frac{1}{2} & \text{if } P \leq \frac{1}{16} \\ 0 & \text{Otherwise} \end{cases} \quad (9)$$

where $P = (x - \frac{3}{2})^2 + (y - \frac{1}{2})^2$

Data generated at 30 random locations using this function has been used for construction of examples. Plot of a data having many zero values, using MQS, is shown Fig. 5. The graph goes negative where as the test function is always non-negative. Plot of the same data using fixed point scaling is shown in Fig. 6. The positivity is preserved as depicted from the range of graph.

RESULTS AND DISCUSSION

The criteria for comparison are based on the key features of MQS that make it suitable for certain applications. The key features of this interpolant are continuity, efficiency and extendibility to higher dimensions.

Original MQS method is C^1 . This level of continuity is often required for visualization applications. So C^1 continuity is also required for positivity preserving algorithm. It is trivial to prove that the continuity of the positivity preserving algorithm is also C^1 however gradient at reference point is also scaled by the scaling factor. For zero valued reference points the gradients also become zero.

Visualizations applications often involve higher dimensional data sets and MQS is easily extendible to higher dimensions. So extendibility of the positivity preserving algorithms to higher dimensions is also important criteria for comparison. We formulated the fixed point scaling scheme without its dimensional consideration so implementation and extension of the scheme to higher dimensional applications is straightforward.

The results are of comparative assessment of these features is summarized in Table 2. Computational cost relative to MQS method is given for 2D data plotted on a 200X200 grid for comparison purpose. Computational cost increase is small because additional processing required for positivity preservation is done at preprocessing stage. Treating zero reference values as a special case further reduces the computational cost.

Table 2: Comparison of MQS with positivity preserving algorithm

	MQS	Scaled MQS
Continuity	C^1	C^1
Relative computational cost	1	1.04
Extendibility to higher dimensions	Yes	Yes
Deviations from original		
RMS	0.205	0.199
Max absolute	0.987	0.999
Jack Knifing error		
RMS	0.099	0.130
Maximum	0.293	0.486

The deviation of the graph from original MQS method increases with the number of its basis function undergoing scaling. This deviation is high when the most of the reference values are zero or near zero. The error terms in Table 2 indicate that the modeling capability of the interpolant has reduced a lot due to the scaling for preservation of positivity.

CONCLUSIONS

The fixed point scaling provides an efficient solution to preserve positivity while visualizing positive data using Modified Quadratic Shepard Method. This research can be implemented using few additional software modules without any major disturbance to existing software modules.

The error estimates indicate that accuracy of the interpolant decreases due to this modification. More work is required to improve the accuracy of the interpolant while preserving positivity.

Regarding future work:

- We are working to improved accuracy of the positivity preserving interpolant.
- We are exploring the possibility of achieving goal of positivity by fitting positive least square quadratic to data.

ACKNOWLEDGMENT

The authors acknowledge enabling role of PAEC and Higher Education Commission, Islamabad, Pakistan and appreciates financial supports through Development of S and T Manpower through Indigenous Ph.D (300 Scholars).

REFERENCES

Asim, M.R., 2000. Visualization of data subject to positivity constraint. Ph.D Thesis, school of computer studies, University of Leeds, Leeds, England, pp: 62-75.

Asim, M.R., G. Mustafa and K. Brodliie, 2004. Constrained visualization of 2D positive data using modified quadratic shepard method. Proceedings of WSCG'2004, February 2-6, 2004.

Brodliie, K.W. and S. Butt, 1993. Preserving positivity using piecewise cubic interpolation. *Comput. Graphics*, 17: 55-64.

Frank, R. and G. Neilson, 1980. Smooth Interpolation of large set of scattered data. *Intl. J. Neum. Meth. Eng.*, 15: 1691-1704.

- Fritsh, R.D. and R.E. Carlson, 1980. Monotone piecewise linear interpolation. *SIAM J. Number Anal.*, 17: 238-246.
- Han, L. and L.L. Schumaker, 1997. Fitting monotone surfaces to scattered data using C^1 piecewise cubics, *SIAM J. Number Anal.*, 34: 569-585.
- Nadler, E., 1992. Non-negativity of bivariate quadratic functions on a triangle. *Computer Aided Geometric Design*, 9: 195-205.
- Renika, R.J., 1988. Multivariate interpolation of large set of scattered data. *ACM Transactions on Mathematical Software*, 14: 139-148.
- Sarfraz, M., M.Z. Hussain and S. Butt, 2000. A Rational Spline for Visualizing Positive Data. *The Proceedings of IEEE International Conference on Information Visualization-IV, 2000-UK* IEEE Computer Society Press, USA.
- Sarfraz, M., 2000. A rational cubic spline for the visualization of monotonic data. *Computers and Graphics*, 24: 509-516.
- Schumaker, L.L., 1982. On shape preserving quadratic spline interpolation *SIAM J. Number Anal.*, 20: 854-864.
- Shepard, D., 1988. A two-dimensional interpolation function for irregularly spaced data. In *Proceedings of 23rd National Conference*, New York, ACM., pp: 517-523.
- Schmidt, J.W. and W. Hess, 1988. Positivity of cubic polynomial on intervals positive spline interpolation. In *BIT*, 28: 340-352.
- Schmidt, J.W., 1990. S-convex, monotone and positive interpolation with rational bicubic spline of C^2 continuity. *BIT.*, 33: 496-511.
- Xiao, Y. and C. Woodbury, 1999. Constraining global interpolation methods for sparse data volume visualization. *Intl. J. Comput. Applic.*, 21: 56-64.
- Xiao, Y., J.P. Ziebarth, B. Rundell and J. Zijp, 1996. The challenges of visualizing and modeling environmental data. *Proceedings of the 7th IEEE Visualization Conference (VIS'96)*, pp: 413-416.