Improved Fast Adaptive IDW Interpolation Algorithm based on the Borehole Data Sample Characteristic and Its Application

Liming Sun*, Yingqi Wei¹, Hong Cai¹, Jun Yan¹ and Jianzhang Xiao¹

¹ China Institute of Water Resources and Hydropower Research, Beijing, China.

*E-mail: sunlm@iwhr.com

Abstract. The inverse distance weighted algorithm (IDW) is one of the most widely used algorithms in geoscience calculation based on its university and simple theory. However, in the practical application based on borehole sampling data, the distance between boreholes is further than the sampling distance along the bedding direction of one borehole, and it needs to search all the near points through a large number of distance calculation, and the process is very slow, so this lacks practicality in the spatial interpolation of the large amount and non-uniform distribution data. Based on the special features of the borehole sampling data, this paper presents a fast adaptive and improved inverse distance weighted interpolation method. And firstly the sample point position could be indexed by the adaptive cube grid raster, then determine the index grid in the horizontal and vertical directions of the interpolation point, and combine these grids for the finally IDW calculation, which significantly improves the search speed of the adjacent interpolation calculation points. It can improve the engineering applicability of the IDW method, especially the three-dimensional spatial interpolation and computing has a good application prospect.

1. Introduction

Spatial interpolation is an important approach to solve the shortage of data sampling in the 3D geoscience related issues. The borehole data is one of the important data sources[1-3], and the 2D and 3D data interpolation method based on borehole data is widely used because of the difficulty in subsurface sampling. For example in the 3D geological modeling using borehole sample data, the 2D smooth interpolation of the geological interface can obtain the smooth geological surface and the 3D spatial interpolation of the geological property field can get the heterogeneous distribution of the geological property [4,5]; in the groundwater analysis application, the spatial interpolation method can be used to calculate the 3D distribution field of the chemical elements in groundwater according to the borehole data; property interpolation and grade calculation of metallic minerals are completed in 3D stratigraphic soil modeling, etc.[6], so the 3D spatial interpolation theory based on borehole sample data is an important foundation work for numerical calculation and analysis. The suitable spatial interpolation method can calculate continuous space field data more quickly and accurately, and it could provide basic data for subsurface geological modeling and numerical analysis application. Furthermore, studying a suitable spatial interpolation method based on the borehole data characteristics is the core issues in the geological geometric modeling and calculation analysis [7].

At present, the most commonly used interpolation methods are two categories: one is geometric smooth interpolation based on mathematical and geometric methods, including B-spline[8], NURBS
surface[9], polynomial and fuzzy mathematics, etc.; and the other is the geoscience interpolation method commonly used in the geology field designed based on the data distribution characteristics of geoscience space, including kriging[10], the inverse distance weighted algorithm (IDW) [11], discrete smooth interpolation, neural network and other interpolation methods, and this kind of interpolation method is more widely used in geospatial interpolation, because the distribution characteristics of geospatial data are usually considered. The IDW interpolation method is suitable for the case of heterogeneous distribution and continuous change, the calculation theory is comprehensive and effective, but it is difficult to deal with the case of uneven and irregular distribution of raw data, and the calculation speed of search neighbor points is slow[12]. The kriging interpolation method reflects not only the spatial distance relationship between the original sample points and the interpolation point, but also the directivity and aggregation characteristics of the raw points distribution, which are obtained by creating a semi-variogram function, these characteristic guarantees higher accuracy and better indicators in most cases[13]. If it is assumed that the equations don't converge or have too few raw points, and the kriging calculation results will be worse. The borehole data are often sampled along the borehole line densely, and they do not follow the spherical, cube or exponential distribution model obviously, so the existing general geological interpolation methods (such as IDW and kriging interpolation algorithm) cannot fully reflect the borehole data characteristics. In addition, due to the dense sampling and linear distribution of borehole data, the traditional interpolation method used in the interpolation calculation could not meet the actual needs of the accuracy and calculation time [14].

2. Borehole sample data

The main feature of the borehole data is that it is composed of discrete points, each point has \((x, y)\) coordinates and elevation \(z\) coordinates. In addition, the borehole data is spatially, the vertical stratigraphic point are orderly distributed along the borehole line. The same stratigraphic points of the different boreholes are located at the same stratigraphic interface, which is the basis for 3D stratigraphic modeling with the borehole data. Since the borehole exploration data is a linear structure perpendicular to surface, each geological borehole could be abstracted as a point set element with 3D structure [15, 16], as shown in Figure 1.

The definition of borehole data model is as follows:

1. Borehole: borehole number, exploration type, hole coordinates \((x, y, z)\), depth \(h\), number of number of strata: \(n\).

2. Stratum: borehole number, stratum sequence number \(s\), top surface point \((x_1, y_1, z_1)\) and bottom surface point \((x_2, y_2, z_2)\) of the stratum, the corresponding property sample value \(p\), other description information of stratum.

3. Borehole incline: depth sounding \(h_0\), dip Angle \(\alpha\) and azimuth \(\beta\), etc.

4. Property sample: borehole sample data, which is usually the distributed along the borehole line, sample depth \(h_0\), sample number, sample time, and sample property value.

![Figure 1. Borehole sample data.](image-url)
3. Improved fast adaptive IDW interpolation algorithm

IDW interpolation algorithm is a kind of deterministic algorithm widely used in the geoscience field, for its theoretical principle is simple[17]. The IDW principle is to measure distance influence between the interpolation point and the adjacent point. Obviously, the closer the distance is, the greater influence on the interpolation point; and the larger weight on the corresponding point; otherwise, the smaller influence on, so it is called the IDW method. In addition, the interpolation mathematical expression (power exponent), the radius range of adjacent points (number of adjacent points) or the number of adjacent points in different quadrants could be adjusted to improve the interpolation accuracy. Also, IDW is a multidimensional nonlinear data interpolation method and an important tool in scientific visualization of 3d discrete data fields.

Figure 2. 2D IDW method, the different colour points represent the different weight value, (1) the points inside the interpolation radius range, (2) the distance between the original sample point and the interpolation calculation point.

As shown in Figure 2, the general mathematical expression of IDW is as formula (1):

\[ z_j = k_j \sum \frac{1}{d_{ij}^\alpha}z_i \]

(1)

Among them, \( z_j \) is the value of the i point, \( k_j \) is the adjustment value that guarantees that the final weight sum is 1, \( d_{ij} \) is the distance from the i point to the current point, and for N-D interpolation, the distance calculation formula between two points is as formula (2):

\[ d = \sqrt{(x_i - x_j)^2 + (x_2 - x_k)^2 + \ldots + (x_n - x_0)} \]

(2)

And \( \alpha \) is power exponent.

IDW mainly uses power parameters to control the influence of known points on interpolation points. The power exponent is a positive number, when the power exponent is larger, and the effect of distance on the result can be further increased, which varies due to the specific calculation. Borehole sample data is usually sparse in horizontal direction, dense in vertical direction, and the two value is greatly difference. Therefore, the traditional adjacent points searching from spherical surround box are non-uniform distribution along and vertical the borehole line direction, in addition, the available points are not enough, and there are too many similar points, so the accuracy is limited by the traditional sphere bounding box for the search calculation[18], as shown in Figure 3.

Traditionally, the weight correction method is used to correct the points of clustering effect, and assume the correction coefficient \( f \) [0,1], if \( f \) is 0, it will not be corrected, and if \( f \) is 1, it will be the max influence. The \( f \) could be calculated based on the angle between the connecting line between the current point and the point in the neighborhood and the two points in the neighborhood[19]. This method has a good effect on the local clustering effect, but the borehole sample point is along line, the data clustering effect is not regular, because it is difficult to directly use this method. So it is necessary to study an improved IDW method based on the characteristics of the borehole data[20].
3.1 Improved method flow
In this paper, an improved IDW method based on borehole data features is proposed, and the flowchart is shown in Figure 4 below. Firstly, a 3D search grid is built according to the maximum range of the borehole sample points, and the grid parameters are determined by the interpolation accuracy and the borehole distribution. Secondly, all the borehole sample points are iteratively assigned to the corresponding grids by the surround box coordinates, a 3d grid (as surround box) may contain many points, and big grids is used to replace points to speed up the search process in the improved IDW algorithm. Thirdly, different parameters of the bounding box are set according to the horizontal and vertical directions, and the corresponding points of search grid are merged to be used for the final interpolation calculation. Finally, the parameterized interpolation calculation of the entire subsurface will be completed. The entire interpolation process is introduced as follows, as shown in Figure 4.

![Figure 3](https://example.com/figure3.png)  **Figure 3.** The search extent of borehole sample data in the IDW method.

![Figure 4](https://example.com/figure4.png)  **Figure 4.** The method process.

3.2 The determination method of interpolation calculation parameters
The cube search grid is used for fast indexing all the original sample points. The whole subsurface is seamlessly filled with adjacent cube grids with the same size, and all the sample points will be assigned to their own grids by position relationship, the grid search will replace the original spherical search in the improved method. First, the side length parameter $m$ of the cube grid needs to be determined. According to the characteristics of the borehole data, two search bounding boxes in the horizontal and vertical directions of the borehole will be introduced to determine the interpolation calculation parameter, each bounding box was composed of a different number of grids, and then the size parameters of the two bounding boxes will be determined.

According to the description in section 2, the borehole data model is firstly built, and the borehole sample data is stored structurally to maintain the topological relationship between the point and the borehole for grid query and calculation. And then, set the minimum number of original sample points involved in the IDW calculation is $S_{\text{min}}$, the maximum search distance is $D_{\text{max}}$, and the maximum search distance ensures that the search range will not be too large when the original sample points are extremely sparse. Generally, the more borehole sample points are used in the calculation, the higher the calculation accuracy will be, but the search speed and calculation speed will slow. Assuming the vertical sampling distance of borehole is $d$, the distance of the adjacent borehole is $l$, and $d < l$, the number of adjacent borehole is $B_{\text{num}}$, $B_{\text{num}} > S_{\text{min}}$ must be guaranteed, and the size of horizontal search bounding box $B_{\text{horizontal}}$ and vertical search bounding box $B_{\text{vertical}}$, it is ensure that both $S_{\text{min}}$ and $D_{\text{max}}$ requirements are met. The calculation rules must be followed:
(1) The cube grid: points are not too much and uniform distribution, assuming that the search radius is $a$, and the radius $a$ is depended on the distribution of the points, $a>d$.

(2) For the horizontal direction box, the bounding box in vertical borehole line direction, its side length is $B_{\text{horizontal}} \times 1^4$, and the number of boreholes is guaranteed that $B_{num} > S_{\text{min}}$.

(3) For the vertical direction, the bounding box in the direction of the borehole line, which its length is related to the length of the borehole, the length in the horizontal direction is greater than the vertical direction $B_{\text{vertical}} > d^2 12$.

3.3 Creating search cube grid
According to the section 3.2, the grid length size $m$ will be calculated, and the search grid could be built according to the grid size. The size of the search grid is usually large, which is a collection of a points with similar location in a single grid. Firstly, the maximum bounding box of the entire subsurface is calculated according to all the borehole data points, and then the space is divided according to the search grid size $m$. The storage data structure of single grid only needs to store the maximum and minimum XYZ value of the maximum bounding box, and it needs to calculate the number of corresponding grids in the XYZ three directions. In this case, only row and column numbers need to be stored, and all the grid vertices can be calculated according to the maximum and minimum values of row and column numbers and the bounding box. In the end, all sample points need to be iterated once and the point number is recorded in the grid object. Then, when using the IDW interpolation, the corresponding search grids within the specified distance can be directly obtained according to the XY coordinate of the interpolation point, and the interpolation distance is not required to be calculated for many times. In this way, the search speed of the interpolation points involved in the calculation can be greatly accelerated. Since the two squared operation of all sample points is changed to a simple linear sum, so the search time is greatly reduced, which is the core idea of this paper.

3.4 Horizontal and vertical search grid computing
According to the search grid obtained in section 3.3, all the original sample points have been stored in the search grid. The distance between the original sample point and the interpolation point can be obtained using the existing grid directly, and the distance formula needs not to recalculate. On the basis of accelerating the search time and further combining the characteristics of the borehole data, the difference between the horizontal direction and the vertical direction of the search bounding box will be considered, the accuracy is improved on the basis of fast search. In this way, the algorithm optimization based on the borehole sample data can improve the calculation efficiency and accuracy of the IDW algorithm. Since the horizontal and vertical directions need to be calculated separately when searching the grid, it needs to be completed in three steps:
Figure 5. Two search grid surround boxes in horizontal and vertical directions.

(1) Interpolation bounding box of horizontal direction, after determining the position of the interpolation points, the search grids of the interpolation points need to be determined firstly. And then, according to the bounding box side in the horizontal direction calculated in section 3.2, the search grid collection of the interpolation points in the XYZ three direction are obtained directly. Finally, all sample points of the horizontal bounding box can be directly obtained according to the point index in the grid, as shown in Figure 5.

(2) Interpolation bounding box of vertical direction, it is similar to the horizontal interpolation bounding box, the points in the vertical direction can be directly obtained according to the calculated side length in section 3.2. After deduplication with the horizontal grid, the points in both grids can be merged to obtain all the original sample points involved in the calculation.

(3) Put the bounding box data into the original calculation formula of the IDW interpolation method for calculation, and the interpolation result of any point will be obtained, as shown in Figure 5.

3.5 Application

According to the optimized IDW interpolation method based on the borehole data characteristic in this paper, which will be applied to the grade calculation and analysis of metallic minerals to verify the feasibility of the method. Taking the borehole sample data of a gold mine as an example, there are 421 boreholes with a sample interval of 2m in the direction of the borehole. The boreholes are evenly distributed, most of which are inclined boreholes, and there are infill boreholes in the position with high metal content. Finally, according to the borehole data shown in Figure 6, there are a total of 43,000 sample points, with a east-west span of 1500m and a height span of 1000m, as shown in Figure 6. The single search speed is not acceptable, so it is not practical to directly use the inverse distance interpolation method. By using the IDW method proposed in this paper, the full size of 125m grid was calculated in 0.5 seconds, and the grade interpolation of the whole geological space was completed. The accuracy of cross validation method meets the 5% error requirement, which has advantages over the traditional method in terms of time and accuracy. The result is shown as in Figure 7.
4. Summary
The method proposed in this paper is an improved method of IDW interpolation based on borehole sample data, compared with the traditional IDW interpolation, the spherical search radius in the original IDW interpolation is transformed into the horizontal bounding box and the vertical bounding box. The radius length of both bounding boxes is determined by the borehole data characteristics, which not only considers the borehole data characteristics, the clustering effect of original data, and the distance search efficiency of interpolation points, but also improves the final calculation accuracy by adding reasonable original points. These are important improvements to the IDW algorithm. To sum up, this paper puts forward the following two improvements:

(1) Compared with the traditional IDW method, the distance between each interpolation point and the original sample points will be obtained by iterative calculation, the method indexes original sample points by moving search grid, which significantly accelerates the search speed of adjacent interpolation points. Method could quickly index the distance between two points without calculation.

(2) The method of combining horizontal and vertical bounding boxes is proposed by taking into account the borehole sample data characteristics. The different data distribution density in the two directions is considered, which greatly improves the accuracy of the IDW method. The corresponding parameter quantification criterion is proposed, which provides the feasibility for the practical engineering application of the IDW method, and an example is given to verify the algorithm.

Acknowledgements
The research presented is funded by the National Key Research and Development Program of China (No.2018YFC1508602, 2017YFC0804607).

References
[1] Caumon G, Grosse O, Lepage F. Unstructured Stratigraphic Grids: Construction, Population and Visualization Issues[J]. 2004(1): 94305.
[2] Durlofsky L J. A general gridding, discretization, and coarsening methodology for modeling flow in porous formations with discrete geological features[J]. Advances in Water Resources. 2016.
[3] Yamamoto J K, Kikuda A T, Koike K, et al. Determination of volumetric uncertainty for geological bodies detected by boreholes[J]. Measurement. 2015, 66: 45-53.
[4] Wang W, Sun L, Li Q, et al. Representing the geological body heterogeneous property field using the quadratic generalized tri-prism volume function model (QGTPVF)[J]. Arabian Journal of Geosciences. 2017, 10(5): 115.
[5] Zhu L, Li M, Li C, et al. Coupled modeling between geological structure fields and property parameter fields in 3D engineering geological space[J]. Engineering Geology. 2013, 167: 105-116.

[6] Touch S, Likitlersuang S, Pipatpongsa T. 3D geological modelling and geotechnical characteristics of Phnom Penh subsoils in Cambodia[J]. Engineering Geology. 2014, 178: 58-69.

[7] Hwang Y, Clark M, Rajagopalan B, et al. Spatial interpolation schemes of daily precipitation for hydrologic modeling[J]. Stochastic Environmental Research and Risk Assessment. 2012, 26(2): 295-320.

[8] Fu Q, Wu Z, Wang X, et al. An algorithm for finding intersection between ball B-spline curves[J]. Journal of Computational and Applied Mathematics. 2018, 327: 260-273.

[9] Ruij JL, Caumon G, Viseur S, et al. Modeling Channel Forms and Related Sedimentary Objects using a Boundary Representation Based on Non Uniform Rational B-Splines[J]. In Proc. 34th Gocad Meeting. 2014(Umr 7330): 1-23.

[10] Ma YZ, Royer J J, Wang H, et al. Factorial kriging for multiscale modelling[J]. Journal of the Southern African Institute of Mining and Metallurgy. 2014, 114(8): 651-657.

[11] Setianto A, Triandini T. Comparison of Kriging and Inverse Distance Weighted (IDW) interpolation methods in lineament extraction and analysis[J]. Journal of Southeast Asian Applied Geology. 2013, 5(1): 21-29.

[12] Mei G, Tian H. Impact of data layouts on the efficiency of GPU-accelerated IDW interpolation[J]. SpringerPlus. 2016, 5(1): 104.

[13] Stein M L. Interpolation of spatial data: some theory for kriging[M]. Springer Science & Business Media, 2012.

[14] Setianto A, Triandini T. Comparison of Kriging and Inverse Distance Weighted (IDW) interpolation methods in lineament extraction and analysis[J]. Journal of Southeast Asian Applied Geology. 2013, 5(1): 21-29.

[15] Wang Y, Ge F, Shao Y. The disperse ore bodies rebuilt based on sparse borehole data[J]. 2011 International Conference on Computer Science and Service System, CSSS 2011 - Proceedings. 2011: 2710-2713.

[16] Charifo G, Almeida J A, Ferreira A. Managing borehole samples of unequal lengths to construct a high-resolution mining model of mineral grades zoned by geological units[J]. Journal of Geochemical Exploration. 2013, 132: 209-223.

[17] Woodard R. Interpolation of Spatial Data: Some Theory for Kriging[J]. Technometrics. 2000, 42(4): 436-437.

[18] Cressie N. The origins of kriging[J]. Mathematical Geology. 1990, 22(3): 239-252.

[19] Shepard D. A two-dimensional interpolation function for irregularly-spaced data[J]. 23rd ACM national conference. 1968: 517-524.

[20] Lu G Y, Wong D W. An adaptive inverse-distance weighting spatial interpolation technique[J]. Computers & Geosciences. 2008, 34(9): 1044-1055.