CONF-9609185--1

SAND96-1830C

Knowledge Base Interpolation of Path-Dependent Data Using Irregularly Spaced Natural Neighbors

Jim Hipp, Ralph Keyser, Chris Young, Ellen Shepard-Dombroski, Eric Chael

Sandia National Laboratories

work completed under DOE ST485D

sponsored by DOE

ABSTRACT

This paper summarizes the requirements for the interpolation scheme needed for the CTBT Knowledge Base and discusses interpolation issues relative to the requirements. Based on these requirements, a methodology for providing an accurate and robust interpolation scheme for the CTBT Knowledge Base is proposed. The method utilizes a Delaunay triangle tessellation to mesh the Earth's surface and employs the natural-neighbor interpolation technique to provide accurate evaluation of geophysical data that is important for CTBT verification. The natural-neighbor interpolation method is a local weighted average technique capable of modeling sparse irregular data sets as is commonly found in the geophysical sciences. This is particularly true of the data to be contained in the CTBT Knowledge Base. Futhermore, natural neighbor interpolation is first order continuous everywhere except at the data points. The non-linear form of the natural-neighbor interpolation method can provide continuous first and second order derivatives throughout the entire data domain. Since one of the primary support functions of the Knowledge Base is to provide event location capabilities, and the seismic event location algorithms typically require first and second order continuity, this is a prime requirement of any interpolation methodology chosen for use by the CTBT Knowledge Base.

Keywords: delaunay tessellations, voronoi diagrams, natural neighbors, interpolation

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000.

OBJECTIVE

The major objectives of this research are to provide a complete and accurate interpolation scheme capable of evaluating geophysical data important to CTBT verification. This research includes investigating methods of gridding or tessellating data and interpolation methods that provide accurate results given the constraints of sparse data and smoothly varying and continuously differentiable surfaces.

RESEARCH ACCOMPLISHED

Introduction

In this paper we will describe the parameterization techniques and methods of interpolation necessary to satisfy requirements for the proposed CTBT Knowledge Base. As part of its capabilities, the Knowledge Base will require an accurate, robust, and efficient means for interpolating various geophysical parameters required by many related or associated tools needed for automatic data processing, event analysis, and event interpretation.

The Knowledge Base interpolation requirements include the ability to model sparse and scattered data with high accuracy over a smooth surface that possesses both 1st and 2nd order continuity. Of the many interpolation methods which we considered, only natural-neighbor interpolation was found to possess all of the requirements specified for the Knowledge Base. This interpolation method requires the generation of a unique set of triangles to partition the Earth's surface. This partitioning, known as Delaunay tessellation (Delaunay, 1934) constructs a well behaved nodal connectivity over arbitrary data which can then be smoothly interpolated using natural neighbor interpolation techniques.

In the following sections we describe and define the parameterization and interpolation techniques necessary to support the Knowledge Base. First, we define the Knowledge Base requirements relative to data parameterization and interpolation. Next, we define interpolation sufficiency conditions and discuss how these conditions along with the knowledge base requirements led us to choose natural-neighbor interpolation as our preferred interpolatory method. Then we give a brief overview of the tessellation, data searching, and natural-neighbor techniques followed by some current results showing some interpolatory meshes developed from seismicity travel-time data. Finally, we briefly describe future work.

Knowledge Base Requirements

The proposed knowledge base is a vital part of CTBT verification analysis. As a central storage area, it facilitates the integration of and access to information needed to verify compliance with the CTBT. Major tasks in which this information will be used are phase detection and identification, association of detected arrivals with events, event location, event identification and special event analysis (Shepherd-Dombroski, 1996 [draft]). The initial focus of this effort is on path-dependent seismic processing and in particular on event location. Ultimately, hydroacoustic, infrasound and radionuclide data sources will also be modeled.

The Knowledge Base will store much of the information needed for automatic and manual data processing by the USNDC. Most of this information is linked to specific locations on the Earth but the data is generally poorly or irregularly sampled making interpolation into other areas an important issue. Providing an irregular tessellation and an accurate and robust interpolation method that can handle the highly irregular data point distribution would be extremely useful. This would enable the use of a dense mesh in areas where much data is available, or where seismic activity or test ban treaty surveillance interest is high. On the other hand, the tessellation could be coarsened in regions of little data, activity, or surveillance interest.

The selected interpolation method must allow new data to be added easily and should be based on a local data interpolation method to minimize data error broadcast and diffusion throughout the entire tessellation. Also, the data point insertion should be of two varieties: fixed and variable. Fixed points are true data point specifications that are known to be tied to precise locations. Variable data points, on the other hand, can be smoothed into the mesh and represent a known set of values that apply over an entire region of the Earth. Smoothing the data can help to minimize the mesh distortion and thereby help to decrease the local error in the interpolation.

From the perspective of the knowledge base, successful interpolation will have to satisfy two basic criteria: accuracy and algorithm compatibility. Accuracy refers to the fact that the interpolation must yield values at certain known locations which are within a specified range of tolerance. This can be broken down further into requirements for accurate interpolation of information at known data points (e.g. master events) and for accurate interpolation into areas where there are no known data points but where ancillary information imposes constraints (e.g. tectonic provinces). The algorithm compatibility criteria is simply that interpolated information must satisfy the requirements of current operational algorithms which are deemed mission-critical. As an example, the event location algorithm used at the USNDC requires continuous first and second derivatives from the interpolation to allow proper convergent behavior. Although discontinuities are readily observed in the real Earth, compliance with the derivative continuity requirement is essential to ensure proper performance from the location algorithm. On the other hand, should these real discontinuities require modeling, they can easily be approximated by fixing the gradient along the boundaries edge (this requires data nodes to be placed along the boundary).

In addition to satisfying accuracy and compatibility issues, the data connectivity should be implemented to support fast searching techniques. This will be extremely advantageous, relative to performance, to ensure reasonable response time from the Knowledge Base. In summary, the selected interpolation method must provide the following capabilities:

1) The method should support a sparse unstructured data distribution, as is typically associated with seismic data,

2) The method should allow new data to be added easily and should support both fixed and variable data locations,

3) The method should be based on a local interpolation scheme to minimize error propagation,

4) The method should provide accurate results within a specified tolerance,

5) The method should possess first and second order continuity to support data processing algorithms, and finally

6) The method should provide for rapid searching techniques to maintain reasonable performance margins.

In the next section interpolation sufficiency will be discussed in a general manner and it will be shown that almost all of the requirements specified above correspond to general requirements for an optimum interpolation method and that non-linear natural-neighbor interpolation meets all of these requirements.

Interpolation Sufficiency

Many interpolation methods are available for general use and all vary with how well they fit arbitrary data representations. Generally speaking they can be organized into the two distinct classes of *fitted functions* and *weighted average* methods. Fitted function methods generally require that a set of coefficients to some polynomial be determined by solving a related set of linear equations involving the data and constraints or criteria that control the fit of the function. The fitted function approach summarizes data behavior in a global manner. Weighted average techniques, on the other hand, sum the data influences at an interpolation point from all neighbor points that lie within the interpolation points "influence" region. Weighted average techniques are local in nature and allow local surface trends to be captured which, generally speaking, is not possible with the fitted function approach.

Fitted function techniques are generally much faster than local interpolation schemes in that a single function evaluation is required while weighted average methods require evaluation of the weights at each new interpolation point. On the other hand, fitted function methods are typically poor models for sparse data distributions in that the resulting surface can behave unexpectedly; over- and under-shoot between data points being a typical problem.

Watson (1992) gives an excellent survey of the major types of computational interpolation methods employed today. He also describes five generalizations that define the optimum or ideal interpolation. These generalizations include:

1) The interpolated surface must fit the data to a user specified level of precision.

2) The interpolated surface should be continuous and possess a finite slope throughout the interpolation domain.

3) The interpolated surface should be locally dependent on the data which minimizes datum error propagation throughout the surface and helps to provide surface stability by negating the possibility that a small change in any datum will result in a large or spurious change in the interpolated surface.

4) The tautness of the interpolated surface should be adjustable to provide the user some control in the rate of change of surface to suit the data. Tautness of the surface should be adjusted automatically at each data point as a function of the local surface variability.

5) The interpolation method should behave equally well regardless of data density or pattern. When data density is not homogeneous (e.g. seismic data), methods employing fixed distance or area subsets cannot be used.

Given sufficient data, in both density and accuracy, any interpolation scheme will give good results since the sampled surface is so well defined. At the opposite extreme, however, peaks, valleys, and rapid changes in the data gradients may all be impossible to infer from a sparse data set. As a result sparse data should be interpolated with methods that incorporate the influence of local gradients. Blending gradients with the interpolated values ensures a continuous and smooth surface throughout the interpolation region.

Of the many available interpolation techniques, natural-neighbor interpolation appears to best fit all of the properties outlined above in Watson's five generalizations. More over, interpolation of seismic data, as required by the Knowledge Base, seems to require that all 5 generalizations be satisfied. The first generalization is necessary to maintain high accuracy. The second and fourth generalizations are required by some data processing algorithms that need first and second order derivative continuity. The third generalization is required by the small scale length features in many of the data sets. The fifth generalization is required since the available data is by default scattered and irregular and contains many large regions with little or no data at all.

Data Tessellations

Proper implementation of a natural-neighbor interpolator requires some form of a triangular tessellation of the data. Delaunay (1934) tessellations, which are triangles on 2-D or 3-D surfaces, are of particular interest because they minimize the maximum angles (or inversely maximize the minimum angles) of the triangles created in the tessellation. Fortune (1992) refers to this property as the maximum-minimum angle property. This property tends to form equilateral triangles and leads to configurations where the average node valence (the number of edges connected to a node) is approximately six. Generally speaking, well proportioned triangles fair better with regard to minimizing local data variability and numerical errors accumulated during the interpolation calculation.

In addition to forming well-shaped triangles, Delaunay triangle sizes are largely determined by the density of the nodal distribution from which they are created. This makes the Delaunay tessellation an ideal selection for tessellating sparse and scattered data.

Many algorithms exist that perform delaunay tessellations. Some of the more popular algorithms include "edge-flipping" (Fortune, 1992), "incremental insertion" (O'Rourke, 1994), and the convex hull algorithms such as Barber's, Dobkin's, and Huhdanpaa's (1993) "quick-hull", to name a few. Delaunay tessellations are widely used in the fields of finite element meshing, computational geometry, and computer graphics. The 4th International Meshing Roundtable (1995), Preparata (1985), and Laszlo (1996), all contain excellent descriptions of the uses of Delaunay triangulations and its dual, voronoi tessellations, in the respective research areas previously mentioned.

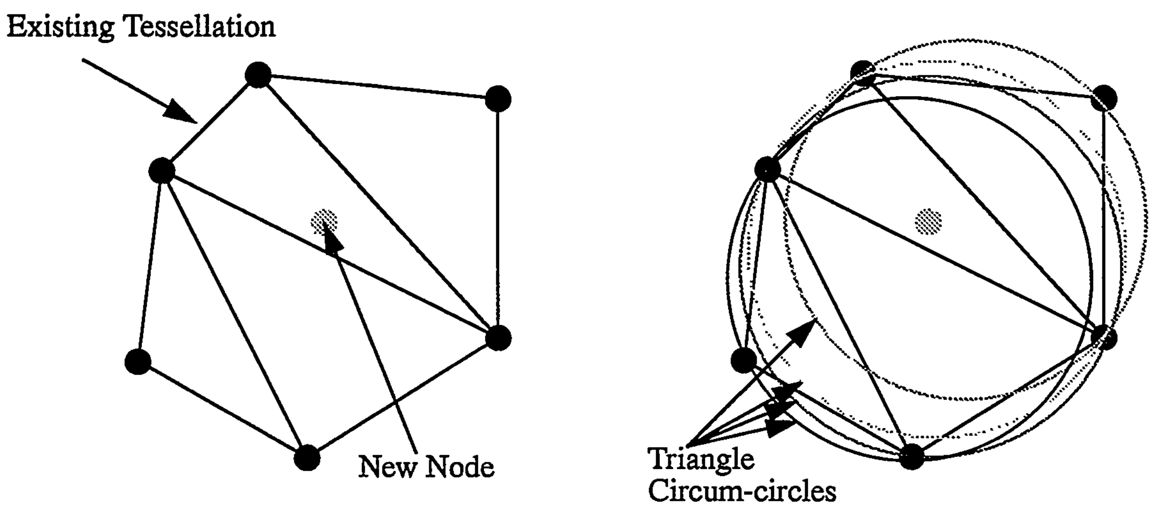
The CTBT knowledge base currently requires a 2-D tessellation on a 3-D spherical surface (the Earth) that must be global in scope and efficient in construction. Since an "incremental" approach can easily be modified to tessellate a spherical surface, that method was selected over other available approaches. Additionally, "incremental" methods allow for easy data insertion and removal that only affects the tessellation locally near the insertion or removal point. Also, the natural-neighbor interpolation method works by assuming that the point to be interpolated is a new data point to be added to the tessellation. This permits a large subset of code used to perform the tessellation to be reused in performing the interpolation.

An incremental Delaunay tessellator works by first enclosing the region to be tessellated with a minimum set of boundary triangles. This set will completely contain all the data points to be inserted in the region. In a 2-D planar context this is accomplished by defining a rectangle that bounds all points to be tessellated and splitting the rectangle into two triangles. The incremental algorithm then proceeds as follows:

For each new node to be inserted

1) **Find all triangles whose circum-circle contains the new node (a circum-circle is a unique circle that passes through the three nodes of its triangle).**
2) **Remove each triangle whose circum-circle contains the new node to form an enclosing polygon that contains the node.**
3) **Connect each node of the enclosing, or "insertion" polygon to the new node to form new triangles that all share the new node.**

This process is illustrated more clearly in Figures 1 and 2 below.



a) b)

Figure 1. Circum-circle containment of a new node to be inserted in an existing tessellation.

Figure 1 a) shows the original tessellation and the new node to be inserted into the tessellations. Figure 1 b) shows how in this case each circum-circle of the original set of triangles contains the new node. Figure 2 a) depicts the "insertion" polygon after the containing triangles are removed. Finally, Figure 2 b) illustrates the new tessellation after the new node is connected to each node of the insertion polygon.

The incremental tessellator is implemented in a similar fashion on the surface of a sphere except the circum-circles now lie on the spherical surface and are calculated using dot products instead of algebra, as is typically done in the planar 2-D case. Figure 3 shows how the circle is defined by the requirement that the dot-products between the vector to each node of a spherical triangle (V_i) and the vector to the circum-center of the triangle (V_c) must be equal.

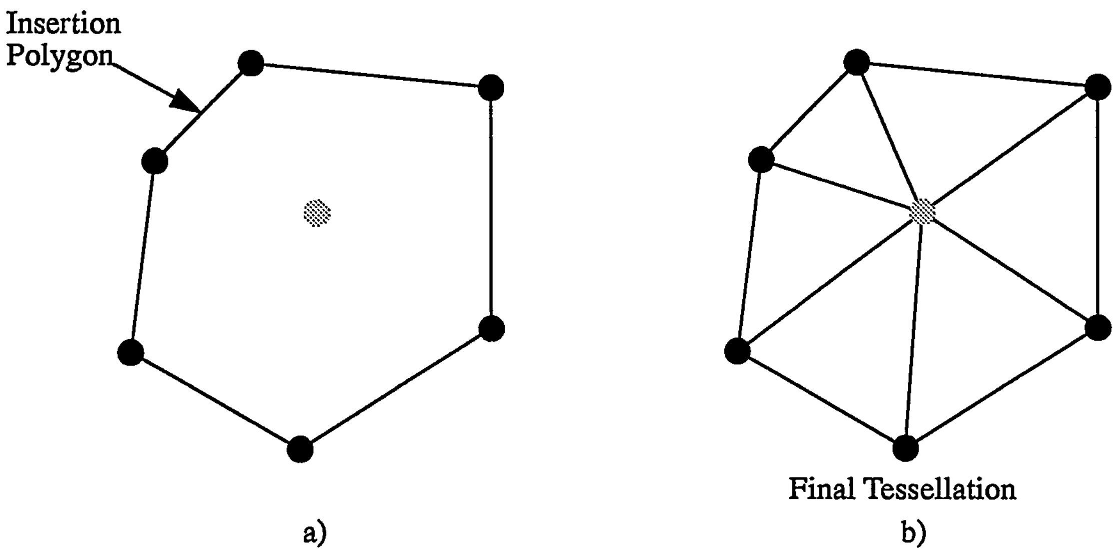


Figure 2. Formation of the insertion polygon resulting in a new tessellation that contains the new node.

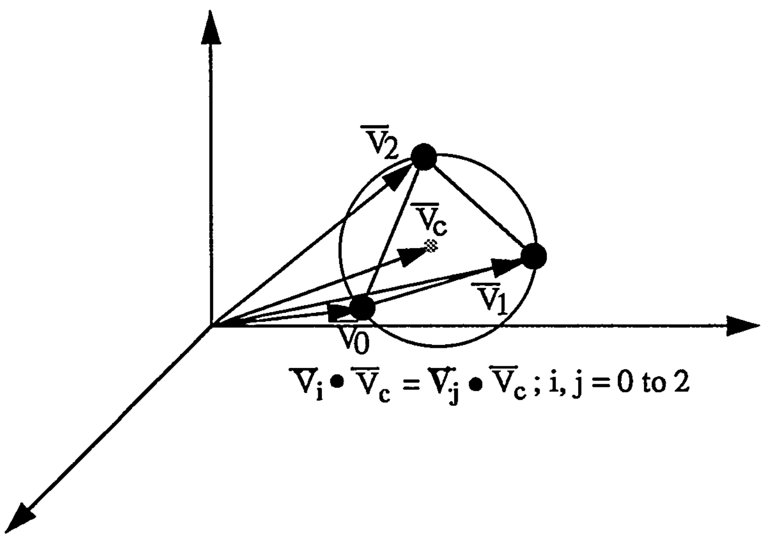


Figure 3. Circum-circle calculation for a triangle whose nodes lie on the surface of a sphere.

Before the data can be tessellated the spherical surface (the Earth) must be divided up into bounding triangles. This is accomplished by forming eight triangles, 4 in each hemisphere, that are comprised of six control nodes located at the poles and 4 equidistant locations about the equator. This provides a complete and bounded surface to begin the insertion algorithm. Once all data points have been inserted the control points can easily be removed and the local neighborhood re-tessellated using "edge-flipping" or any other appropriate approach.

Searching

If the Knowledge Base is to be used in a real time context then finding each interpolation point in the tessellation could be extremely time consuming, especially if a large data set is used. Many search algorithms could be modified to perform the search but most would have to be used in 3-space as the tessellation is a 3D tessellation (a 3D surface). However, one particular method known as the Walking Triangle Algorithm (Lawson, 1977 and Sambridge, 1995) allows the data set to be searched as a 2D data set in linear time. This avoids a global search (typically proportional to the square of the number of nodes) and makes this method ideal for searching a large data set.

The method requires that the edges of each triangle be oriented such that a counter-clockwise traverse of the edges produces an outward-pointing normal when the right-hand rule is used. Figure 4. below illustrates the 'left', 'right' test. The method works by initially guessing a start triangle and subsequently testing each side in turn. If the interpolation point lies to the 'right' of the edge being tested then the interpolation point is not contained by the triangle and the algorithm moves to the triangle shared by the edge just tested. If, however, the interpolation point is to the 'left' of the edge, then the next edge is tested. If all edges of a triangle result in a 'left' test then that triangle contains the interpolation point. The actual test is performed by taking the dot-product of the interpolation point vector with the normal produced by crossing the node 0 vector of the edge being tested with node 1 vector. In this context the direction from node 0 to node 1 is in the counter-clockwise direction around the triangle. If the dot-product is greater than zero then the interpolation point lies to the 'left' of the edge, otherwise it lies to the 'right'.

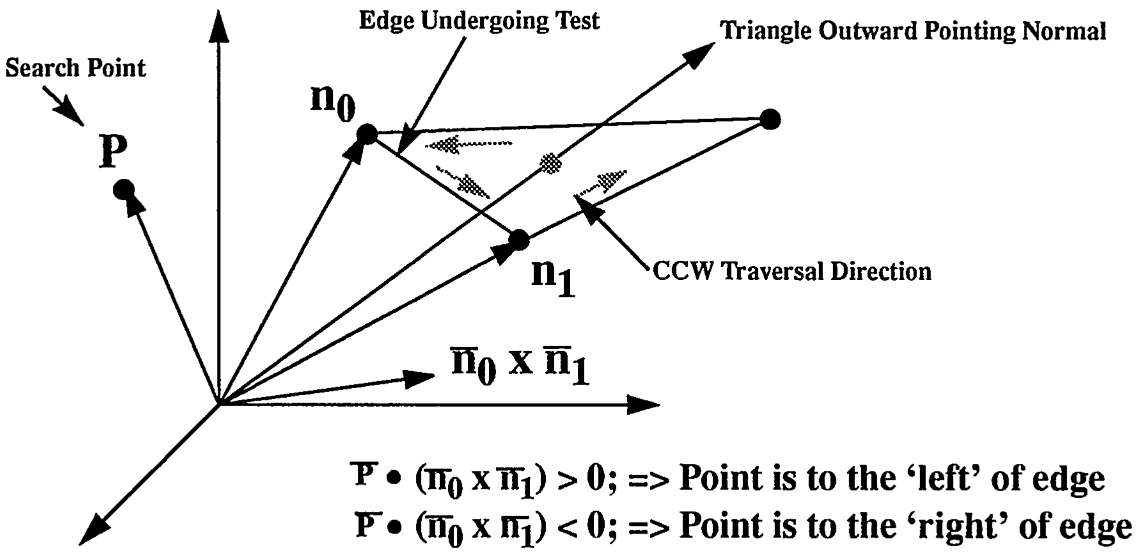


Figure 4. Walking Triangle Search Test.

This method is ideal for use by the current USNDC event location algorithm which is iterative since the previous containing triangle can be used as a new guess for the next event-location search. The searching performance of the algorithm was tested by tessellating 12,000 event locations (provided by 12 years of NEIC seismic event data) with a 26,000 triangle arbitrary tessellation and then searching for the triangles containing the event locations. Typically between 50 and 100 triangles would be traversed to find each interpolation point. On a 167 MHz Sun Ultra the average search time per interpolation point was approximately 0.007 seconds.

Natural Neighbor Interpolation

Natural-neighbor interpolation, as discussed above, is a local weighted average interpolation method. These methods (in two dimensions) have an interpolation function of the form

$$f(x, y) = \sum_{i=1}^{n} w_i(x, y) f_i$$

where i is summed over all node neighbors that surround the interpolation point (x, y), f_i is the function value at the nodes, and $w_i(x,y)$ is the normalized 'influence' weight attached to neighbor i that influences the interpolation result at (x, y). In natural-neighbor interpolation these weights are referred to as the natural-neighbor coordinates of point (x,y) (Watson, 1992). A brief synopsis of the method by which the natural-neighbor coordinates are calculated will be given here. Refer to Watson (1992) and Sambridge (1995) for excellent treatments of the full procedure.

The natural-neighbor coordinates are best described using a geometric definition. Examine the five nodes in Figure 5a. The dashed lines connecting the nodes represent the Delaunay triangles connecting the nodes whose circum-circles include the interpolation point, X (see **Data Tessellations** section for a description of circum-circle containment).

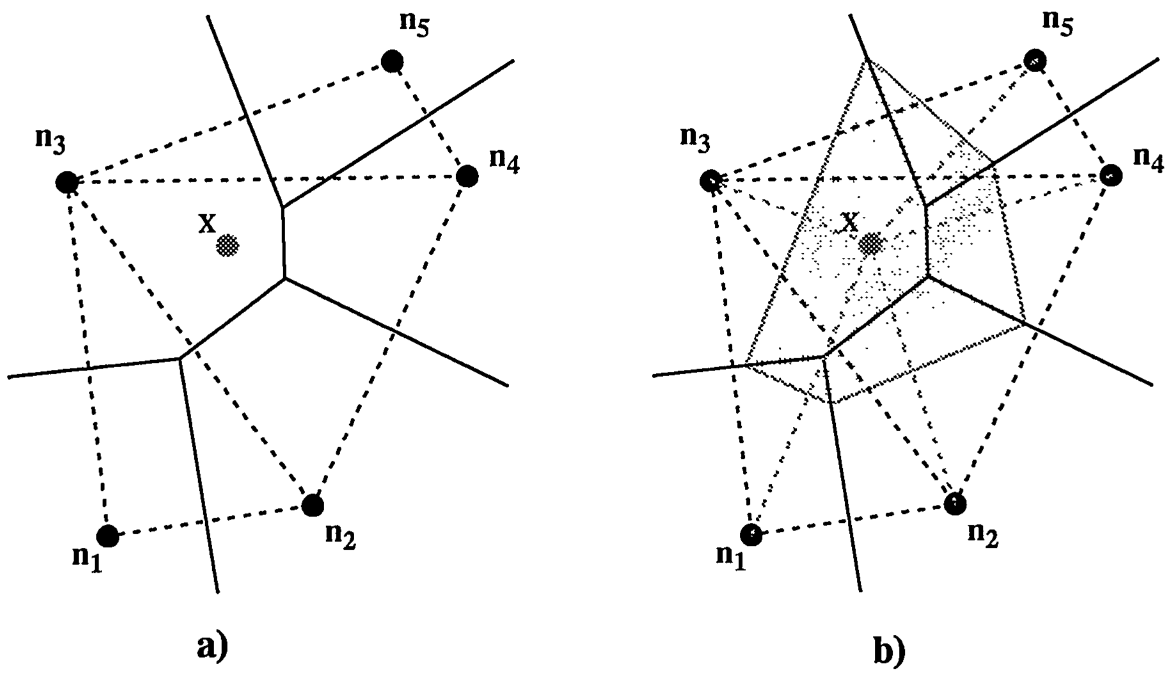


Figure 5. Natural-neighbor coordinate definitions.

The solid lines are the voronoi edges of the nodes, the complete set of which is referred to as a voronoi polygon (Preparata, 1985). Each node in a Delaunay tessellation has exactly one convex voronoi polygon that completely surrounds it. The voronoi polygon can be constructed by passing a perpendicular line through the midpoints of each edge of the Delaunay tessellation. The intersection of those lines form the vertices of the voronoi polygon. In Figure 5a, we have removed outlying edges for clarity.

Figure 5b depicts the result of adding the interpolation point, X, to the tessellation. The shaded portion of Figure 5b represents the new voronoi polygon for the interpolation point. The new voronoi cell about X overlaps all of the original cells of its natural-neighbors. The natural-neighbor coordinate, or weight, for a neighbor node acting on the interpolation point is defined as the ratio of the area of overlap with the neighbor nodes original voronoi cell and the new voronoi cell about X to the total area of the new voronoi cell about X.

Since the nearest-neighbor coordinates are normalized they always lie between 0 and 1. This results in the property that if an interpolation falls exactly on top of a node the function value of the node is returned. The other important property to note is that the interpolation is entirely local and only depends on the nearest-neighbor set of the interpolation point.

This type of interpolation is linear and results in a surface that has continuous first order derivatives throughout the tessellation except at the data points themselves. A completely first and second order continuous interpolation can be evaluated by blending the linear form with the natural-neighbor gradient evaluated at the interpolation point. Watson (1992) describes some simple parametric blending functions that account for the linear interpolant, gradient, and the variability in the data to control the surface tautness.

Results

At present the research has succeeded in implementing the tessellation, smoothing, searching, and interpolation functionality. However, no testing of the algorithms on true Knowledge Base data has been performed as the data is not yet available. This has led us to improvise testing of the tessellation, smoothing, and searching functionality on analogous data sets which are available. As an example, seismicity travel time correction factors were taken from a single station (Albuquerque) to model a mesh density that is smoothed on the travel-time correction gradient. The correction factors are modeled by using a tessellation of all of the seismic event points. Each event point has a nearest neighbor set (voronoi set) whose corrections are weighted by distance to provide an arbitrary field function that approximates the correction factor variability over the Earth.

Once the field function is defined a coarse mesh is prescribed and smoothed over the Earth. Then new points are added and smoothed using the field function as a partial weight. This results in a mesh that densifies within regions of large correction-factor gradient. Figure 4. below illustrates the resulting mesh.

At this point it is not clear how viable mesh smoothing will be. For hydroacoustic interpolation the technique might be very useful because acoustic travel times can be interpolated directly from an already defined mesh. Accuracy and efficiency could be increased by densifying the mesh in regions of large gradients. Seismic data, on the other hand, is more problematic. Regional characterization efforts will yield sets of control points with fixed locations, and these may be embedded in a global coarse mesh which may have null values at many locations. Under these circumstances the global connectivity of the mesh will be used to search for the triangle containing the point to be interpolated but the interpolation will only be performed if one or more nodes of the containing triangle have non-zero data.

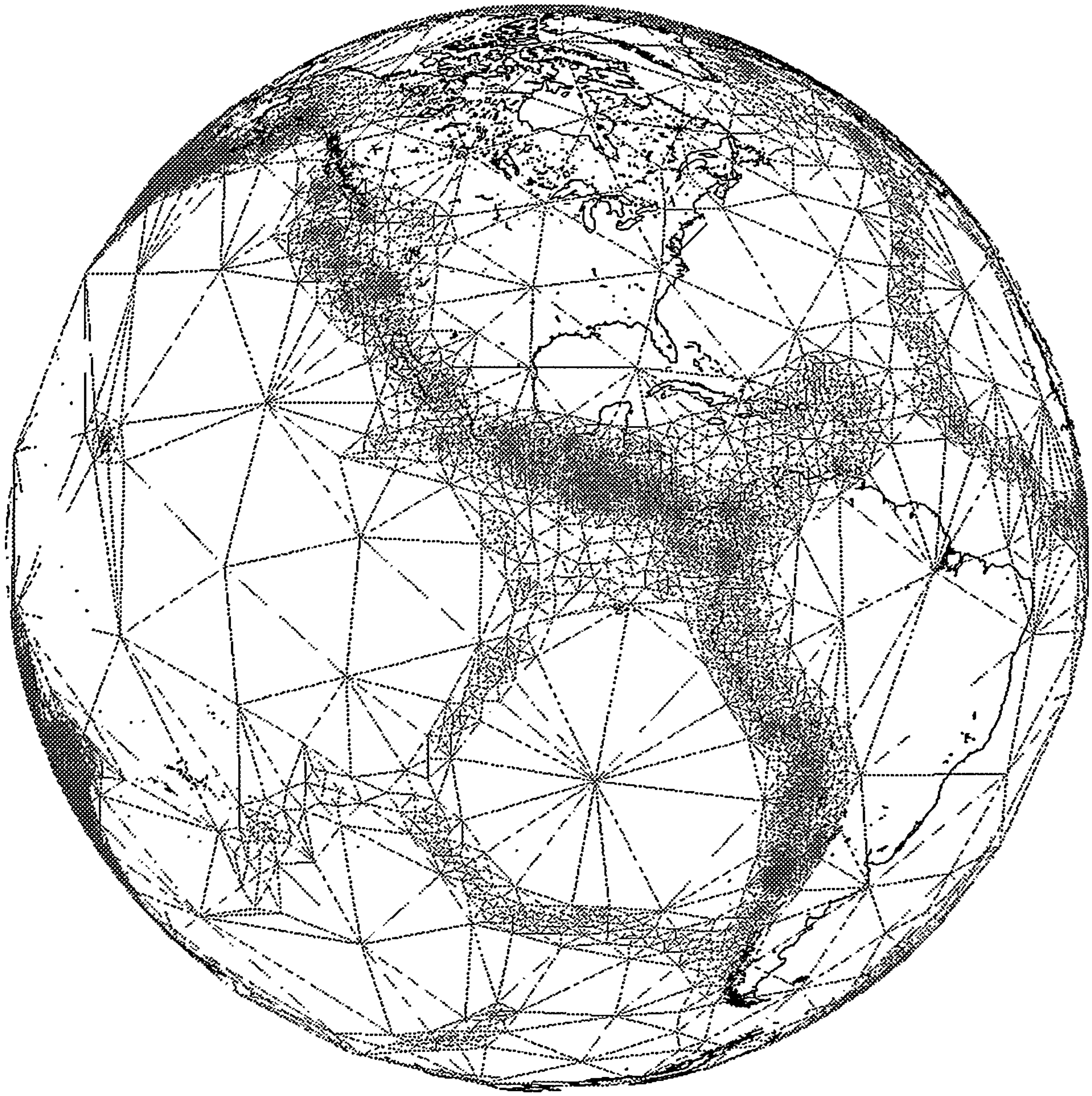


Figure 4. Seismicity map illustrating correction-factor gradient smoothing.

CONCLUSIONS AND RECOMMENDATIONS

This paper described the requirements of the CTBT Knowledge Base relative to seismic event processing and interpolation. We have shown that the methods of Delaunay tessellations and natural-neighbor interpolation more than meet the requirements of the Knowledge Base. The method of Delaunay triangulization is extremely flexible and provides the perfect connectivity network to perform natural-neighbor interpolation. In addition, natural-neighbor interpolation satisfies the Knowledge Base requirements of accuracy, continuity, and the ability to model sparse and irregular data. Moreover, Watson (1992) has shown that the method of natural-neighbor interpolation meets all the conditions of an ideal interpolator and may be the only means of interpolating sparse and irregular data accurately.

We have briefly described the Delaunay tessellation and nearest-neighbor interpolation methods applied to interpolation on the spherical Earth. We have described an appropriate triangle search technique that is efficient and avoids the use of complex search algorithms and their associated additional memory requirements. Finally, we have shown an example of the meshing technique to model the travel-time correction factor gradient in terms of mesh density.

In the future we intend to use data from the regionalization efforts under way at Los Alamos and Lawrence Livermore Laboratories to begin testing our interpolation and database connectivity algorithms. A major part of this testing will involve connecting the current event location algorithm to the searching and interpolation algorithms to demonstrate both the accuracy of the collected data and the effectiveness of the interpolation.

REFERENCES

Barber, B., Dobkin, D.P., and Huhdanapaa, H., 1993. **"The Quickhull Algorithm for Convex Hull,"** The Geometry centre technical report GCG53. The Geometry Centre, Univ. of Minnesota, Minneapolis, MN 55454.

Braun, J., and Sambridge, M., 1995, **"A Numerical Method for Solving Partial Differential Equations on Highly Irregular Evolving Grids,"** Nature, Vol 376.

Delaunay, B.N., 1934. **"Sur la sphere vide,"** Bull. Acad. Science USSR VII: Class. Sci. Math., 793-800.

Fortune, S., 1992. **"Voronoi diagrams and Delaunay triangulations,"** Computing in Euclidean Geometry, eds. Du, D.Z., and Hwang, F., World Scientific.

Laszlo, M. J., 1996, **"Computational Geometry and Computer Graphics in C++,"** Prentice Hall.

Lawson C. L., 1977, **"Software for C1 surface interpolation,"** Mathematical Software III, ed. Rice. J., Academic Press, New York.

Preparata, F. P., and Shamos, M. I., 1985, **"Computational Geometry: an Introduction,"** Springer-Verlag, New York.

Sambridge, M., Braun, J., and McQueen, H., 1995, **"Geophysical parameterization and interpolation of irregular data using natural neighbors,"** Geophys. J. Int.

Shepherd-Dombroski, E., D. Baur., J., Hipp, J.R., Keyser, R.G., and Young, C. J., 1996 (Draft), **"CTBT Knowledge Base Initial Design"**, Sandia National Laboratories.

Watson, D. F., 1992, **"Contouring: A Guide to the Analysis and Display of Spatial Data,"** Pergamon, Oxford.

"4th International Meshing Roundtable," 1995, Sandia National Laboratories, Sand95-2130.