SAND097-1688C
SAND--97-1688C

Travel-Time Correction Surface Generation for the DOE Knowledge Base

CONF-970967--3

Jim Hipp, Chris Young, Ralph Keyser
Sandia National Laboratories

Sponsored by U.S. Department of Energy
Comprehensive Test Ban Treaty Research and Development Program

ABSTRACT

The DOE Knowledge Base data storage and access model consists of three parts: raw data processing, intermediate surface generation, and final output surface interpolation. The paper concentrates on the second step, surface generation, specifically applied to travel-time correction data. The surface generation for the intermediate step is accomplished using a modified kriging solution that provides robust error estimates for each interpolated point and satisfies many important physical requirements including differing quality data points, user-definable range of influence for each point, blend to background values for both interpolated values and error estimates beyond the ranges, and the ability to account for the effects of geologic region boundaries. These requirements are outlined and discussed and are linked to requirements specified for the final output model in the DOE Knowledge Base. Future work will focus on testing the entire Knowledge Base model using the regional calibration data sets which are being gathered by researchers at Los Alamos and Lawrence Livermore National Laboratories.

Keywords: Knowledge Base, kriging, interpolation, delaunay tessellations, natural neighbors

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

OBJECTIVE

The purpose of travel-time correction data in the DOE Knowledge Base is to provide to a client the best estimate possible of the travel-time correction and associated error for any arbitrary point at or near the Earth's surface as quickly as possible. In order to accomplish this, we have developed a three-part model for the storage and access of travel-time correction information in the Knowledge Base, comprised of raw data input (including various validity and consistency checks), conversion to a complete and internally consistent storage representation, and fast output of interpolated data to the clients. The portion of that research discussed in this paper focuses on the intermediate step: the conversion of raw data to the data which will be the basis for the on-the-fly interpolation in the final step. The intermediate step is necessary because the raw data points themselves are not expected to provide a sufficient basis for the final interpolation due to inadequate sampling and a lack of a means to account for other important factors which are expected to influence the interpolated travel-time correction values (e.g. nearby geologic province boundaries).

In our research we have investigated methods of gridding or tessellating data as well as interpolation methods that provide accurate results, including surface error estimates, given the constraints of sparse data and smoothly varying and continuously differentiable surfaces. Our major objectives are to provide a comprehensive set of data representation and interpolation schemes capable of meeting the requirements for providing travel time corrections from the Knowledge Base. Although our discussion here will be limited to travel time correction data, it is expected that the techniques developed will apply to many of the other Knowledge Base data sets as well, such as corrections for amplitude, azimuth, and slowness.

RESEARCH ACCOMPLISHED

Introduction

As stated above, this paper will focus on the intermediate data representation portion of the three-part Knowledge Base model. The technique that we have chosen to use to convert raw data to the data to be stored in the Knowledge Base, is a modification of the interpolation method known as kriging. Kriging can be used to calculate travel-time corrections and also provides associated error estimates, which most other interpolation schemes do not. We can use the modified kriging technique to create a dense grid of interpolated data and errors which will then serve as the basis for the fast interpolation techniques in the final part of the Knowledge Base data storage and access model.

Our technique was modified from ordinary kriging in order to meet certain constraints. In order to fully understand the reasoning behind these constraints imposed on the method used to generate the intermediate data, we must briefly describe the requirements of the final corrections presented to the user. From the perspective of the Knowledge Base, the processed surface data representations presented to any user, either human or an Automated Data Processing (ADP) code, will have to satisfy several basic criteria. These criteria include:

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

2) The representation should be based on a local interpolation scheme to minimize error propagation,

3) The representation should provide accurate results within a specified tolerance,

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

5) The representation should minimize the amount of data required to represent the surface and provide for rapid searching techniques to maintain reasonable performance margins.

The first criteria is simply a statement that the final surface should not be constrained to a regular grid. Regular grids, in regions of high surface complexity (curvature), are, of course, efficient representations and lend themselves to rapid searching techniques (direct indexing) which are required by the 5th criteria. However, regular grids do not allow thinning of data in regions of simple surface curvature (flat) and are difficult to refine to the surface of a sphere, especially near the poles. If an ADP code constantly accesses a database during data processing, it follows that having more data in the database will result in slower processing times. For these reasons, smooth irregular tessellations such as the Delaunay triangular tessellation (Delaunay, 1934 and Fortune, 1992) are generally thought to be better representations. The mesh density can easily be refined as a function of surface curvature and the surface can be rapidly searched with n log(n) speeds using algorithms such as the "walking triangle" searcher (Lawson, 1977 and Sambridge, 1995).

The second and third criteria are simply a statement that the final representation be an accurate one. By accuracy we mean that the interpolation must yield values at certain known locations which are within a specified range of tolerance; that is, we must have an error specification that accompanies the result. In fact, some ADP codes require an error estimate to function properly. Additionally, if the travel-time correction surface representation is especially erroneous at some arbitrary location, we wish to localize that error and not allow it to propagate to other locations in the surface. These criteria can be satisfied in both the intermediate and final stages of the DOE Knowledge Base model. We shall soon describe the methods for the intermediate stage. In the final stage we simply require that the interpolation scheme that returns results from the intermediate surface to the user be local in nature. That is, that it only uses information near the interpolation point to estimate a result. Many fast interpolation schemes satisfy this result such as linear interpolation or natural-neighbor interpolation (Watson, 1992), to mention just two.

The fourth criteria is necessary because some ADP codes require it. A good example is the event location algorithm used at the USNDC. This algorithm 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.

The last requirement is a performance requirement. Simply stated, with the massive amounts of data that are transferred over the computational data bus from the server to a client we have little time to spend searching for data points and we must minimize the amounts of data transferred over the bus to reduce the total data access times. This criteria also suggests that a moldable unstructured grid that can be tailored to regions of high surface curvature while maintaining only a few points in regions that are relatively flat is highly desirable. To reiterate, this function is easily accomplished with smoothed triangular grids.

All of the requirements stated above for the final surface representation provided to user ADP codes drive the requirements for the generation of the surface in the intermediate representation. The process whereby the intermediate representation is calculated for a data set is the subject of the rest of this paper. The discussion is divided into three main sections including a synopsis of the requirements for the development of the intermediate surface and how they relate to the criteria defined for the final output surface described above; a brief overview of the modified kriging model including those modifications that satisfy the intermediate surface requirements; and lastly, an example of the results of the modified kriging algorithm that illustrates the behavior of the kriged residual and error surfaces for a real data set.

Travel-Time Correction Surface Requirements

Any developed travel-time correction surface (i.e mesh of points) must satisfy several key constraints before the surface can be considered acceptable for the Knowledge Base. These constraints include the following:

1) The surface must be first and second order continuous and be well behaved to properly support ADP codes that require continuity and non-oscillatory behavior,
2) The surface must include associated error estimates,
3) The variance and associated residual at each data point must provide an analyst sufficient capability for modeling the data points range of influence, magnitude, and shape and rate of fall-off (or increase for variance) for each group of data points,
4) The residual surfaces must go to zero, and error surfaces to a specified background, at the extreme range of influence of each data point,
5) The surface must be capable of supporting groups of data with different error estimates that can coexist spatially in neighborhoods whose data density can vary from a single point to hundreds or thousands.
6) At any arbitrary interpolation point, the influence of low variance data (reliable) must be greater than the influence of high variance data (unreliable).
7) At any arbitrary interpolation point, the interpolated value must reflect the influence of known boundaries across which travel-time corrections are expected to change (e.g. geologic provinces).

The first and second constraints above are necessary primarily to satisfy ADP codes such as event locators where continuity and an estimate of error are required to locate an event within a calculated error ellipse. These two requirements translate directly from the 2nd, 3rd and 4th criteria specified in the previous section.

The third constraint is to provide sufficient modeling flexibility so that a seismologist can properly represent the expected physical behavior of a set of data points. Each interpolation may involve different factors such as differing quality of data points (e.g. earthquake versus calibration), or proximity to a tectonic province boundary. This requirement is purely a data modeling constraint and has no equivalent in the output model criteria specifications.

The fourth constraint is also a modeling constraint and is necessary to allow interpolated data from different regions to blend together uniformly and continuously. This constraint provides a means to support arbitrary and different error models and yet insure consistency between different regions. It also provides a means to blend to a default model for uncalibrated regions,

The fifth and sixth constraints are necessary to support real world data acquisition including quality and quantity of data, and to take advantage of higher quality data whenever possible.

Since requirements 4 through 7 all result in a continuous surface everywhere they all contribute to satisfying the continuity criteria for the output model data. In addition to providing an accurate physical picture of the travel-time correction surface. In the next section we will briefly outline the changes imposed on an ordinary kriging model that allow surfaces to be generated that satisfy the requirements defined above.

Travel-Time Correction Surface Model

Most surface interpolation schemes can adequately provide accurate, smooth, and continuous surfaces with moderate to high calculational performance margins. However, all but a few are unable to relate any information about error when data error information exists.

One common method of interpolating a surface that will account for error in the raw data is the method of least squares. Least squares interpolation techniques can provide error bounds on the interpolated surface that minimizes the mean squared difference from the regression function chosen to represent the surface. In this sense, the selected function fits the data more closely than any other function of the same type. Unfortunately, least squares techniques have a tendency to over- and under-shoot dramatically near steep slope changes. Additionally, the surface is equally affected by a distant data point, as it is by a near data point which implies that the fit function is not representative of the spatial variation in the data (Watson, 1992). A final undesirable consequence of the method is that even when measurement error on the data is eliminated or inconsequential, the least squares surface still fails to pass through any of the data points any where on the surface. Given the points delineated above one can see that the method of least squares generally fails all of the requirements defined in the previous section with the exception of requirement 2).

Other methods that include error estimates when developing an interpolated surface are termed kriging methods (named after the South African who originally developed the methods) or are strongly related to kriging methods in that the interpolated point is determined as a linearly weighted combination of the data where the weights are evaluated as a best linear unbiased estimate (blue) such that the error of the predicted surface is minimized. Only a modified version of the ordinary kriging method that satisfies the constraints prescribed in the previous section will be described here. The interested reader should consult Wackernagel, 1995, Cressie, 1993, or Isaaks, 1989, for further information concerning kriging and associated derivative methods.

Ordinary kriging methods in and of themselves are unappealing for our problem in that many of the surface constraints specified in the previous section are not met. First, ordinary kriging assumes equal variances for all data points. This assumption violates constraints 5) and 6). Secondly, ordinary kriging does not guarantee that the residual surface will fall away to zero at some range beyond the data or that the variance surface will move to a specified background value. This fact violates constraint 4). Finally, ordinary kriging has no way to account for regional boundaries in the data which violates constraint 7). Boundaries are either completely smooth (ignored) or discontinuous which violates constraint 1).

Since ordinary kriging methods do not completely satisfy the necessary requirements a modified methodology is required. The first major modification of ordinary kriging is to remove the requirement that all data possess the same variance. This means that the traditional variogram approach to kriging is replaced by the covariance definition that includes a correlation coefficient and the variances between data points or a data point and an interpolation point. This reformulation is written as

$$C_{ij} = \rho_{ij}\sigma_i\sigma_j$$

We can now assign different values of variance to each data point which removes the aforementioned restriction. This allows one to incorporate data points that possess variances of varying quality (e.g. explosions vs. earthquakes) such that constraint 5) is now satisfied.

Constraint 4) can be satisfied by allowing the value and variance to transition as a function of distance from the data point. Figure 1) shows a typical transition kernel. Notice that the value (a correction relative to a base model) transitions from its value, Z_i, at range = 0, to 0, at the data point's specified range of influence; while the variance transitions from its value, σ_i, at range = 0, to the background variance, σ_{bg}, at the specified range of influence.

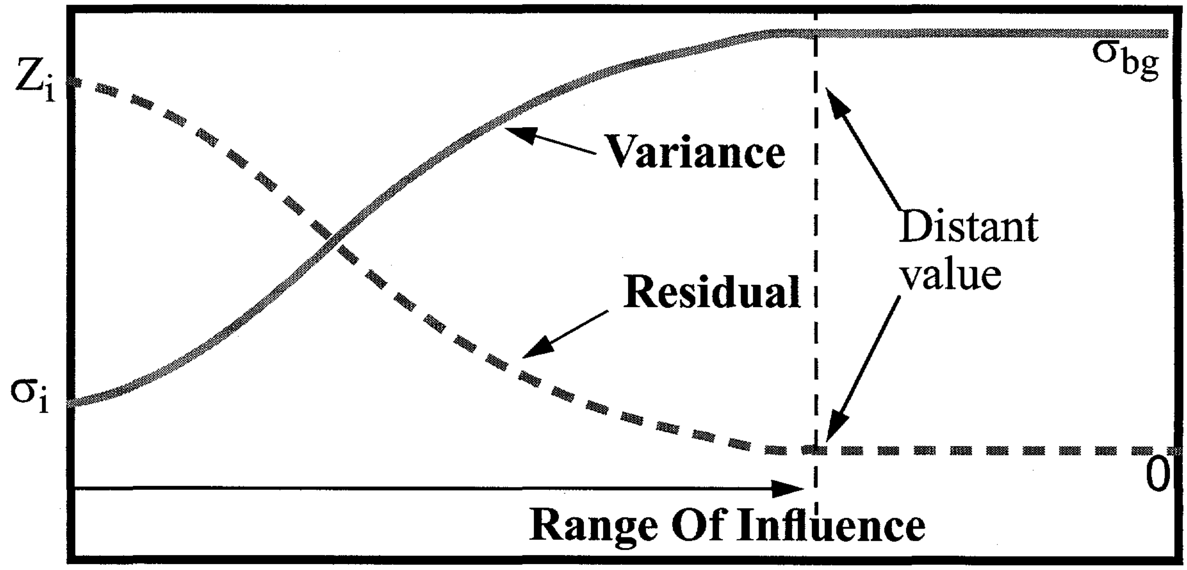

Figure 1). Typical Transition Kernels for Residuals and Variances

Constraint 6) can be satisfied by specifying the interpolation point variance prior to estimating the surface at that point. By forcing the interpolation point variance σ_p to be equal to the minimum of all other variances evaluated at the interpolation point, we can always force the solution to utilize the best available error in the local neighborhood about the interpolation point. We can write the above statement mathematically as

$$\sigma_p = min(\sigma_i(h_{ip}); i = 1,N)$$

where h_{ip} is the range between the ith data point and the interpolation point and the minimization is carried out over all N data points.

Finally, if the distance between a data point and the interpolation point is intersected by a region boundary we can correct the interpolated value to reflect this by forcing the transition function of a data point to move to zero (for the residual or σ_{bg} for the variance) earlier than it would normally. Figure 2) shows how the transition is patched in the influence region of a boundary for a residual blending kernel.

Every boundary can have a different width for the transition zone and can be arranged in arbitrary open or closed segment groupings. This provides a wealth of flexibility while making transitions smooth and continuous.

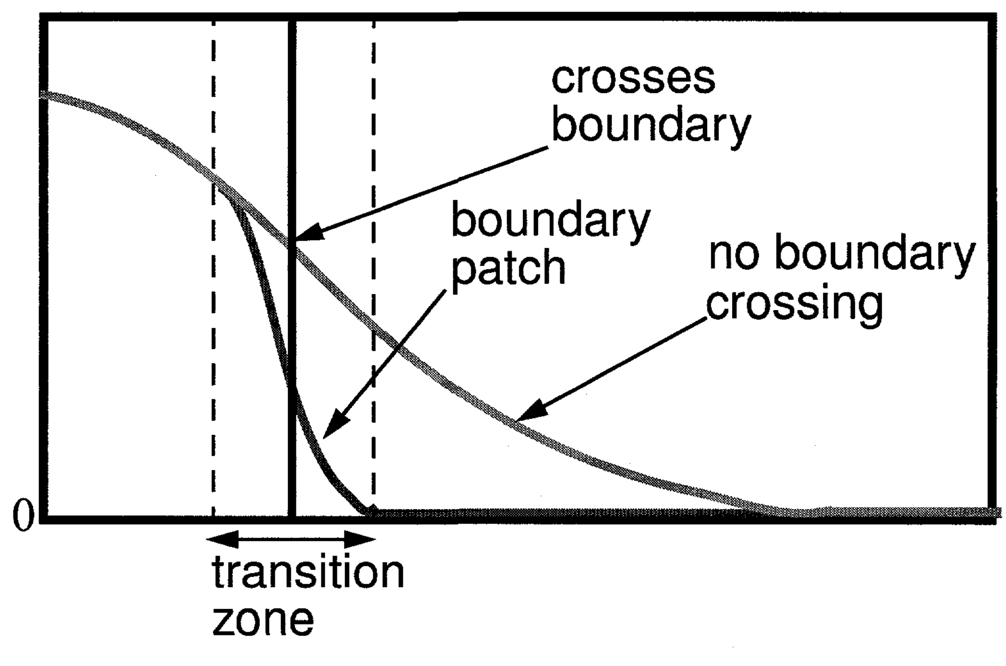

Figure 2). Typical Transition Patch of a Residual Kernel as it Crosses a Region Boundary.

In the next section we shall give an example problem that utilizes the modified kriging method defined above. The resulting solution shows that surface continuity exists everywhere and that both the residual and error surfaces migrate to their respective background values as the interpolation point leaves the influence range of the data.

Example of a Modified-Kriged Surface

The following illustrations depict travel time residual and error surfaces taken from regional travel time correction data calculated for seismic station LZH from a set of events located in China. The data was provided by Allen Cogbill of Los Alamos National Laboratories.

The surfaces are calculated using the residual and error blending kernels depicted in Figure 3). The range selected for this example is 10 degrees and the variance for all the data was set to zero. The background variance was arbitrarily chosen to be 2 sec^2. The error blending kernel is the y-axis flip of the residual blending kernel as shown in Figure 1) above.

Figure 4) shows the residual surface using the modified kriging solution described in the previous section. Notice the smooth behavior throughout the region and the way that the residual surface falls back to zero outside the range of the data. Figure 5) shows a top down view of Figure 4) with the input data points indicated by small circles.

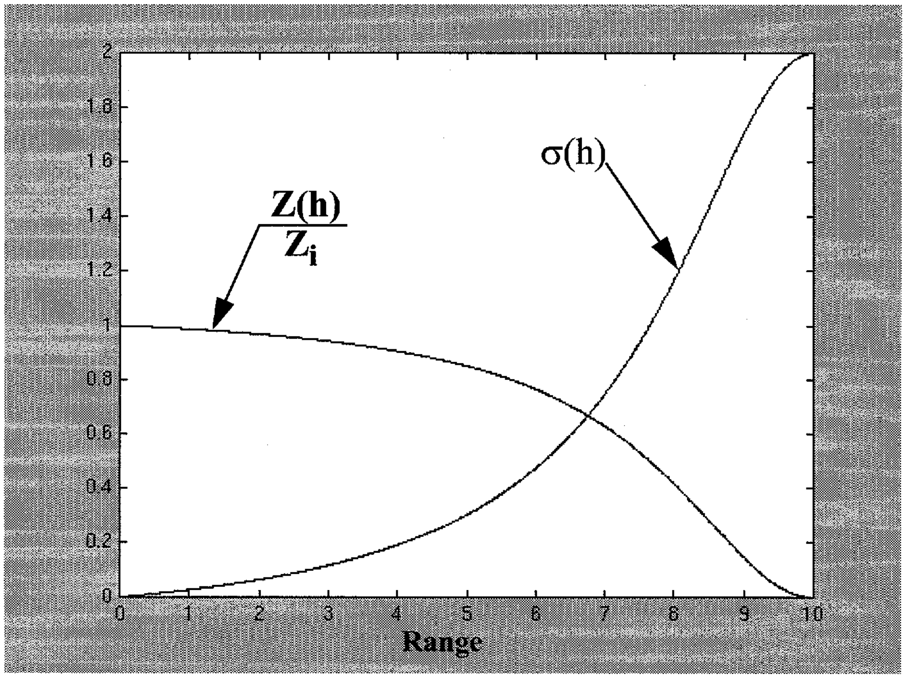

Figure 3). Residual and Error Transition Kernels Used by the Example Problem in this Section.

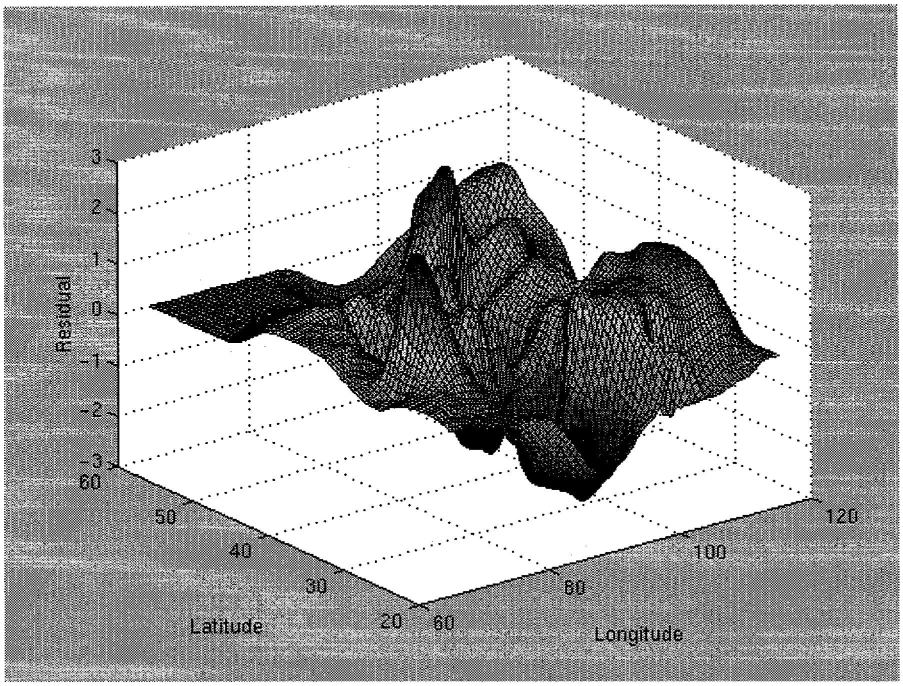

Figure 4). Resulting Residual Surface for the Example Problem in this Section.

Figure 6) illustrates the accompanying error surface to the residual surface shown in Figure 4). Notice that the error drops to essentially zero throughout the region of influence of the data and rises sharply to the background variance outside of the data points influence region.

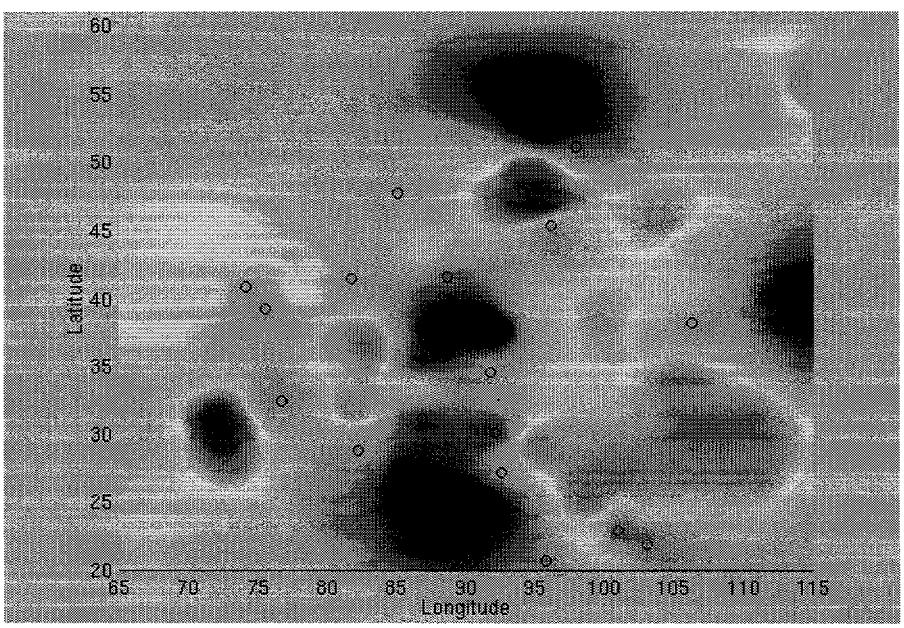

Figure 5). Top View of the Residual Surface. Data points are indicated by the small circles.

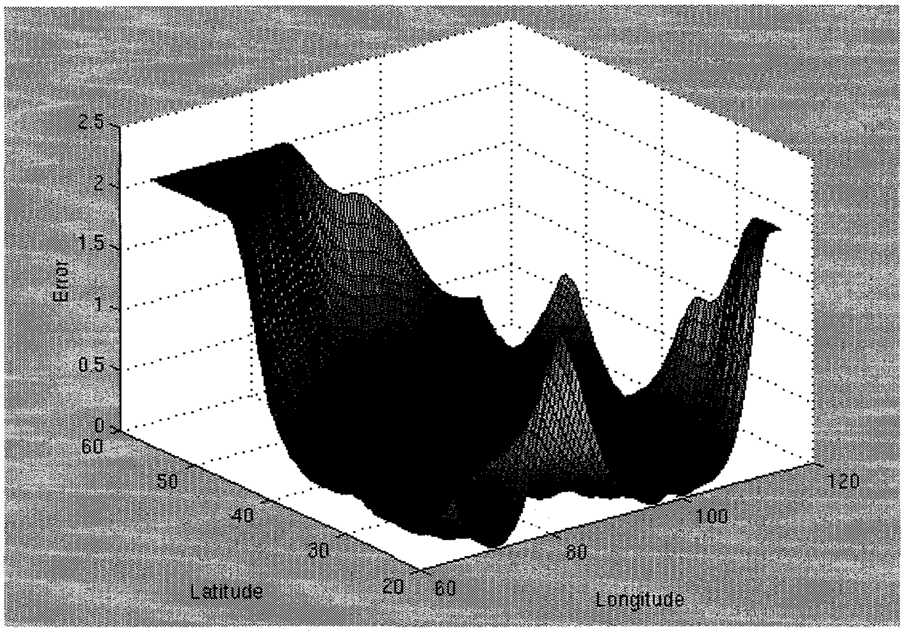

Figure 6). Resulting Error Surface for the Example Problem in this Section.

CONCLUSIONS AND RECOMMENDATIONS

The research described here has focused on the intermediate portion of the three part DOE Knowledge Base data storage and access model. We defined the requirements of the intermediate surface representation and used them to develop a modified kriging approach to generate the surface. We have shown that the modified-kriging approach satisfies all of the requirements of the intermediate surface representation which in turn solves all of the requirements specified for the final data product provided to Knowledge Base clients. These requirements provide for smooth continuous residual and error surfaces that mold easily to background values. Additionally, they allow for data of differing quality to exist simultaneously while forcing the highest quality data in the interpolation neighborhood to control the resulting interpolated surface values. Regional boundaries are also accounted for by forcing the solution to background values when boundaries are intercepted between data points and interpolation points.

Future work will include using data from the regionalization efforts under way at Los Alamos and Lawrence Livermore Laboratories to begin testing the entire Knowledge Base model, from input of raw data to output of travel-time corrections for clients. A major part of this testing will involve connecting the current event location algorithm to the Knowledge Base to demonstrate both the accuracy of the collected data and the speed of the final interpolation step. Our work with the example data set described in the previous section represents a first step in this process. Performance issues will also become increasingly important as we attempt to demonstrate that the overall performance of the final output step is acceptable from the perspective of the ADP codes which are expected to request travel-time correction data from the Knowledge Base.

REFERENCES

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

Cressie, N. A. C., 1993. **"Statistics for Spatial Data,"** John Wiley & Sons.

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

Isaaks, E. H., Srivastava, R. M., 1989, **"An Introduction to Applied Geostatistics,"** Oxford University Press, New York.

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

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

Wackernagel, H., 1995, **"Multivariate Geostatistics: An Introduction With Applications,"** Springer-Verlag, Berlin.

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