"Visualization '96" – Case Study.

Data Reduction and Interpolation for Visualizing 3D Soil-Quality Data

David C. Banks* (banks) Bernd Hamann† (hamann) Po-Yu Tsai* (ptl)
Robert Moorhead* (rjm) Jonathan Barlow* (barlow)

*NSF Engineering Research Center for Computational Field Simulation
Mississippi State University (erc.msstate.edu)

†Department of Computer Science
University of California-Davis (cs.ucdavis.edu)

Abstract

Sampling and analysis of subsurface contaminants comprise the first steps toward environmental remediation of hazardous spills. We have developed software tools to support the analysis phase, using three different schemes for interpolating scattered 3D soil-quality data onto a grid suitable for viewing in an interactive visualization system. A good interpolation scheme is one that respects the distribution of the original data. We find that the original data can be decimated by up to seventy percent while exhibiting graceful degradation in quality. A prototype software system is being deployed to allow technicians to visually determine, while in the field with their monitoring equipment, where the highest concentrations of contaminants lie. The system is now in use by the U. S. Army Corps of Engineers.

1 Introduction

Remediating environmental damage begins with collecting data from a disturbed region and then analyzing it. Remediation is very expensive. Analysis of subsoil contaminants permits the remediation to be applied only where it is needed.

To sample subsurface contaminants, technicians drill holes (*a.k.a.* "pushes") into the soil and measure contamination values along the trajectory of these pushes. The Site Characterization and Analysis Penetrometer System (SCAPS) is one such measurement device. It reports soil type and contamination value. The SCAPS sensor uses laser-induced flourescence to measure concentrations of petroleum, oil, and lubricants in the soil. Other types of sensors can measure concentrations of volatile organic compounds or explosives.

At a given site under investigation, many locations are chosen at the surface level (parametrized by longitude and latitude) to initiate a push (in the vertical direction). A single push takes up to an hour to initiate and to complete, so SCAPS measures data along only about half a dozen pushes per day at a field site. The push locations are typically several meters apart, spread across a site that may cover 20,000 square meters. Each push, however, may generate hundreds of samples along its path, spaced as closely as 2.5 inches in the vertical direction.

Sublayers under the ground are folded and faulted, so the soil type sometimes changes suddenly along a push. A contaminant may form a pool at the interface between porous and nonporous layers. The contamination values are then discontinuous across the interface, which requires that SCAPS measure them at a very high spatial resolution in the vertical direction.

Typically there are more samples collected in the vertical direction than are actually needed for resampling. The contamination value may be nearly constant over large intervals of the push. We therefore would like to reduce the data in the vertical direction. Section 2 discusses this decimation step. After decimating the data, we resample the set of scattered samples onto a rectilinear grid, to be used by an isosurface and volumetric display system. It is notoriously hard to establish which among various interpolation schemes is the best choice. Section 3 describes the three resampling techniques we applied and identifies the one that appears to be the best performer.

2 Data Reduction

The subsurface contaminant data is finely sampled in the vertical direction in order to capture discontinuities in the concentration that occur along bound-

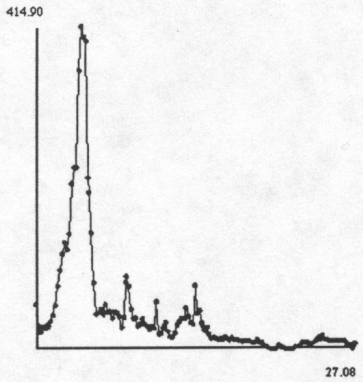

Figure 1: *Original samples along a push. Horizontal axis measures depth of the push. Vertical axis measures concentration of soil contaminant.*

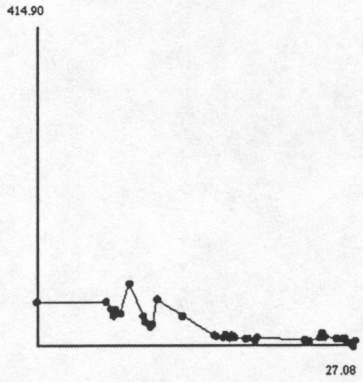

Figure 2: *Data reduced by 70%, based on curvature measured in original graph.*

aries of geological layers. The large number of samples makes the process of resampling to a 3D grid very slow. Since the data values are nearly constant over subintervals of the pushes, we would like to reduce the number of samples while preserving the discontinuities that indicate feature boundaries. Unfortunately, the acquired data is generally very noisy. Filtering out the noise can inadvertently remove the significant discontinuities in the data.

One-dimensional data reduction can be accomplished by selecting a subset of the original points so that curvature is evenly distributed along the selected data [Hamann]. This technique requires that the data is sufficiently smooth in the beginning. For noisy and discontinuous data, we must first smooth

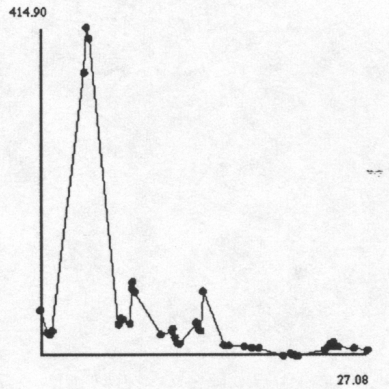

Figure 3: *Data reduced by 70%, based on curvature measured in filtered data.*

the data before we calculate its curvature. We apply an approximate Gaussian filter (having compact support) to smooth the given data. We then select points from the smoothed data set so that the integrated curvature of a curve passing through the data is nearly uniformly distributed. The technique allows us to retain the significant qualitative features (like spikes in concentration) even when retaining only thirty percent of the samples.

Figure 1 shows original data values for samples along a vertical push. Figure 2 shows the result of reducing the data values by seventy percent (based on curvature) without filtering. Note that significant features of the original data are absent from the decimated data.

The unwanted noise is reduced (but the significant spike is also diminished) when smoothing the data with a Gaussian filter. Therefore the filtered data is only consulted to evaluate the curvature function for decimating the *original* data, not the smoothed data. Figure 3 shows the result of reducing the original data by 70% according to the curvature of the filtered data.

3 Resampling to a 3D grid

The data values from all pushes are resampled on a 3D rectilinear grid in order to view isosurfaces of contamination. We first construct a tetrahedral mesh from the scattered data along the pushes, then interpolate the samples onto a grid. We have implemented three interpolation schemes used for resampling scattered data: the global Shepard method, a local Shepard method, and a local Hardy multiquadric method.

Shepard's method computes an approximative function value f at a (varying) location $\mathbf{x} = (x, y, z)$

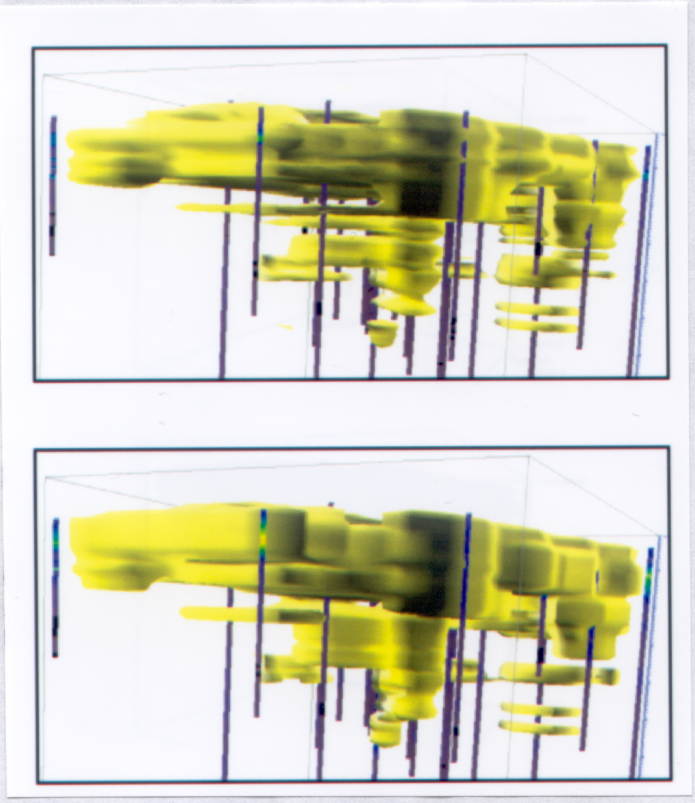

Figure 4: *Local Shepard isosurfaces using 100% (above) and 30% (below) of the original push data. The pushes are indicated by the dark vertical bars. Note that the small, flat, pancake-shaped concentrations of contaminants (lower right) remain visible even in the decimated data.*

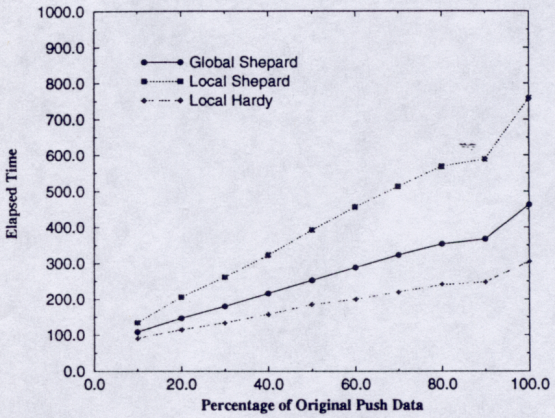

Figure 5: *Comparison of running times for global Shepard, local Shepard, and local Hardy methods. The local methods are based on a 5-point neighborhood. The horizontal axis indicates the percentage of original push samples that were used in the interpolation. The vertical axis indicates execution time (reported by the Unix "time" command) for interpolation to a 3D grid.*

on a rectilinear grid, based on a finite number of samples in a certain neighborhood. If the neighborhood is finite, the method is called local; otherwise it is called global. A sample's influence on the value of f depends on its distance to $\mathbf{x}$: its influence is inversely proportional to the square of its distance to $\mathbf{x}$ [Franke]. Eventually, convex combinations of sample values in a certain neighborhood around $\mathbf{x}$ are used to define the value of f. When applying the local method, we use the underlying tetrahedrization of the (potentially decimated) samples to localize the approximation.

Hardy's multiquadric method [Franke] requires the computation of coefficients for the basis functions defining the approximation. A linear system of n equations must be solved when considering n samples. Again, we use the tetrahedrization of the samples to identify samples in a local neighborhood. The number n should be kept very small in order to compute an approximation efficiently.

In general, it is difficult to determine which interpolation scheme is most applicable for resampling environmental data [Berry]. A good scheme should be a reliable predictor for additional samples that may be acquired from the site. Verifying an interpolation scheme may require prohibitively many samples to be collected. When a commercial analysis product is discussed in the scientific press [Vasilopoulos] [Mahoney], the capabilities of the product are more likely to be featured than are the details of the interpolation being used.

Figure 4 shows the isosurfaces from original samples of soil-quality data (taken from 22 pushes) and from decimated samples. The local Shepard technique produces isosurfaces that respect the discontinuities in the vertical direction: they look like a "stack of pancakes." The global Shepard technique, by comparison, tends to smooth the data too much in the vertical direction.

Figure 5 shows the computation time required to resample a particular data set (at various levels of decimation) using each of the three interpolation schemes. The dataset contains approximately 10,000 samples. The local methods were timed using five neighbors.

It is our experience that as much as 70% of the samples along a push may be discarded while still capturing the essential features and respecting the discontinuities in the data.

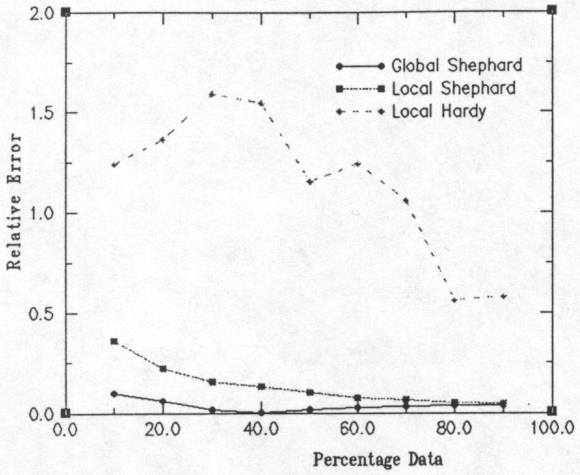

Figure 6: *Average relative error for the three interpolation methods. For each method, the relative error on the decimated data is measured against its evaluation (onto the same 3D grid) of the complete dataset.*

It is our experience that, due to the discontinuous nature of the soil data, the local methods (with appropriate neighborhood sizes) perform better than the global method. The local Hardy method is faster than the local Shepard method for a fixed number of neighbors being interpolated. On the other hand, the Shepard method can, with only 5 nearest neighbors, produce 3D interpolations that the Hardy doesn't match with fewer than 50 neighbors. The undecimated data is so dense in the vertical direction that the Hardy approximant needs to be allowed to consider many neighbors simply in order to find a nearby sample from a different push. We could compensate for the effect by scaling the data in the vertical direction, but this is a somewhat *ad-hoc* technique without genuine knowledge of the soil types at the site. We note that the same system for interpolating and visualizing soil-quality data can also be used for site characterization: determining and visualizing the soil types at different layers.

The Hardy method is more sensitive to the reduction of data along the pushes. The graph in figure 6 shows how the relative error of each method degrades as the original data is decimated. For each method, we interpolated the original data onto a 3D grid and considered that to be the "ideal" result for that method. Then we decimated the data and applied the method again, calculating the relative error against the "ideal." While the global Shepard technique did not produce satisfying results (based on the push data we tested), its results were fairly insensitive to data reduction than. The results of the local Shepard interpolation were considered (by an expert at the Army Corps of Engineers Waterway Experiment Station) to be the best; they also are less sensitive to data reduction than the local Hardy method is. The local Hardy method produces dramatically different interpolated values, with average relative errors of about 1.0 (100 percent) when the amount of push data is reduced by only 30 percent.

4 Conclusion

Visualization is being used to analyze environmental data and guide remediation efforts at hazardous sites. Fast and accurate conversion of scattered sample data into 3D images will allow technicians in the field to collect additional samples more efficiently. In our experience, as much as 70% of the soil data can be discarded while retaining significant features. This decimation allows faster resampling onto a 3D grid. We have found that local interpolation schemes are more effective than global ones for analyzing soil quality data, and that the local Shepard method is less sensitive to decimation the local Hardy method. While visualization does not eliminate the need for continued monitoring (the Environmental Protection Agency requires the presence of monitoring wells at impacted sites), it can serve to guide in the placement of monitors and in the choice of the number of monitors needed.

References

[Berry] Joseph Berry, *Spatial Reasoning: Concepts and Considerations in Effective GIS Solutions,* GIF World, Fort Collins, CO, 1995.

[Franke] Richard Franke, "Scattered Data Interpolation: Tests of Some Methods" *Mathematics of Computation*, Vol. 38, No. 157, 1982, pp. 181-200.

[Hamann] Bernd Hamann and Jiann-Liang Chen, "Data point selection for piecewise linear curve approximation," *Computer Aided Geometric Design 11(3),* pp. 289-301, 1994.

[Mahoney] Diana Phillips Mahoney, "Mapping Toxic Spills," *Computer Graphics World,* PennWell Publishing Company, Jan 1991.

[Vasilopoulos] Audrey Vasilopoulos, "A New World View," *Computer Graphics World,* PennWell Publishing Company, Apr 1991, pp. 63-72.