

INTERPOLATION OVER LARGE DISTANCES USING SPHEREKIT

Robert Raskin, Associate Scientist, Applied Research Corporation,
Jet Propulsion Laboratory Mail Code 525-389, Pasadena, CA 91109 USA

Chris Funk, Graduate Assistant, Department of Geography,
University of California, Santa Barbara, CA 93106 USA

Cort Willmott, Professor, Department of Geography, University of Delaware,
Newark, DE 19716 USA

ABSTRACT

Spherekit is a spatial interpolation toolkit developed and distributed over the internet by the National Center for Geographic Information and Analysis (NCGIA). A unique feature of the software is its ability to work directly with the spherical geometry of the earth. Thus, distances, areas, and directions are spherically based, and interpolation can be carried out over large distances without distortions induced by the use of planar projections. The user can select from several interpolation methods that have been adapted to the sphere. The package also features "smart interpolation" capabilities to incorporate knowledge of the underlying physical processes that produced some of the spatial variability. Error analysis using cross-validation is built-in to compare the relative performance of interpolation algorithms or parameter settings. The cross-validation errors can themselves be interpolated to a uniform grid to reduce spatial bias. The capabilities of Spherekit are demonstrated using three examples.

OVERVIEW

Spherekit is a spatial interpolation software toolkit developed at NCGIA as part of Initiative 15 (Multiple Roles of GIS in Global Change Research). The source code is available over the internet without charge to the user. The package features several unique capabilities.

Spherekit permits interpolation over continental or global scales because its computations are based upon spherical distances and orientations (Raskin, 1994). Conventional interpolations (Watson, 1992) are based upon planar projections of the earth that produce distortions of some kind over large distances. In Spherekit, projections are applied only for display purposes after

the interpolation has been carried out in spherical geometry. The user can select from several interpolation algorithms that have been adapted to the sphere: inverse distance weighting, thin plate splines, multiquadrics, triangulation, and kriging.

Spherekit permits the user to incorporate knowledge or information about the processes that produced the underlying spatial variations. A built-in equation editor and a collection of nonlinear transforms allows the user to create and experiment with new, physically meaningful variables from the independent and dependent variables available. This "smart" interpolation capability allows Spherekit to intelligently interpolate using auxiliary information. A digital elevation model (DEM) is included with the package. One use of the smart interpolation feature is to incorporate elevation information when interpolating variables that are correlated with height.

Error analysis is an integrated component of Spherekit. This makes the package particularly useful for comparing interpolation methods and parameters. The performance of a method is measured using cross-validation. The cross-validation error is defined at each observation point as the difference between its actual value and its estimated interpolated value using the remaining $n-1$ points. The resulting error field can be displayed either at the data points or interpolated to a regular grid to reduce spatial biases. Error difference fields, comparing a pair of methods and/or parameter settings, can be easily created and displayed.

Spherekit helps the user manage the various files that have been read in or created. The file management window for a sample session is shown in Figure 1. The example shows file listings for observation data, grids, networks, interpolation methods, interpolation results, error fields, and derived variables. Clicking on any field name displays all known metadata for that field. This window also serves as the Spherekit main menu; six main menu options appear along the top of the window.

INTERPOLATION METHODS

Spherekit was designed to be usable by researchers analyzing global scale datasets. Several standard interpolation methods have been modified for use on the sphere by utilizing spherical distance in place of Euclidean distance. Additional modifications for sphericity are used where possible. Most of the interpolation methods can be implemented as either global or local methods. For local methods, a neighborhood size is specified either as a radius, the number of included points, or an average number of points. Overrides are available to bound these values, if desired. Five interpolation methods are available:



Figure 1 Example of the file management window

- * Inverse distance weighting
- * Multiquadric
- * Thin plate spline
- * Kriging
- * Triangulation

For the inverse distance weighting method, the user can choose from three weighting functions: inverse power, Shepard (1968), and a smoothing function. The user also can select levels of anisotropy and gradient correction (Shepard, 1968). The bias correction deweights clustered points to reduce spatial bias. The gradient correction permits extremum values to occur at locations other than the observation points.

The multiquadric and spline methods involve inversion of an $n \times n$ matrix for n data points. For n larger than several hundred, the user should specify a neighborhood for carrying out local fits. This will invert smaller matrices for each interpolation point rather than performing a single large matrix inversion over the entire domain. The user is warned to use a local fit if the storage requirements for carrying out the $n \times n$ inversion exceed the space available.

The generalization of multiquadric method to the sphere has been formulated by Pottmann and Eck (1990). The thin plate spline implementation is a spherical extension of the methods described in Franke (1982).

The kriging implementation is that of kriging with a trend (universal kriging). Semivariograms are computed using the GSLIB library software package (Deutsch and Journel, 1992). Exponential, Gaussian, spherical, and linear models are supported, all using spherical distance. A summation of two of these models is permitted. The semivariogram results also can be used for exploratory analysis purposes; an example is provided in the next section.

The triangulation method uses the Delaunay triangulation to identify the nearby observation points to be used in the interpolation. Renka's spherical algorithm (Renka, 1984) is used to carry out the interpolation. The user can choose either a linear interpolation of the values at the triangle vertices or a polynomial fit, obtained by performing an initial cubic spline fit along the edges.

APPLICATIONS

Smart interpolation

"Smart" interpolation improves the performance of traditional interpolations by using knowledge of the processes that produced the spatial variations (Willmott and Matsuura, 1995). In this example, we use the physical law that temperature falls off with altitude, roughly at the environmental lapse rate. Standard and topologically aided interpolations are compared using a sparse network of 160 weather stations in China. The data set is deficient in that high altitude locations in the Himalayan mountains are underrepresented. Figure 2 shows the interpolated temperature field (in °C) using the multiquadric method. The fit based on the sparse dataset fails to take into account the large variations in topography that produce very low temperatures at high altitudes.

Figure 3 shows the corresponding "smart" interpolation based on an interpolation of the derived variable "sea level temperature." In this example, the first-order effect of the temperature- elevation relation has been incorporated into the interpolation. That is, the "smart" interpolation captures the climatological influences of topography. The low temperatures associated with the mountains of western China are now visible, despite the lack of high altitude temperature stations.

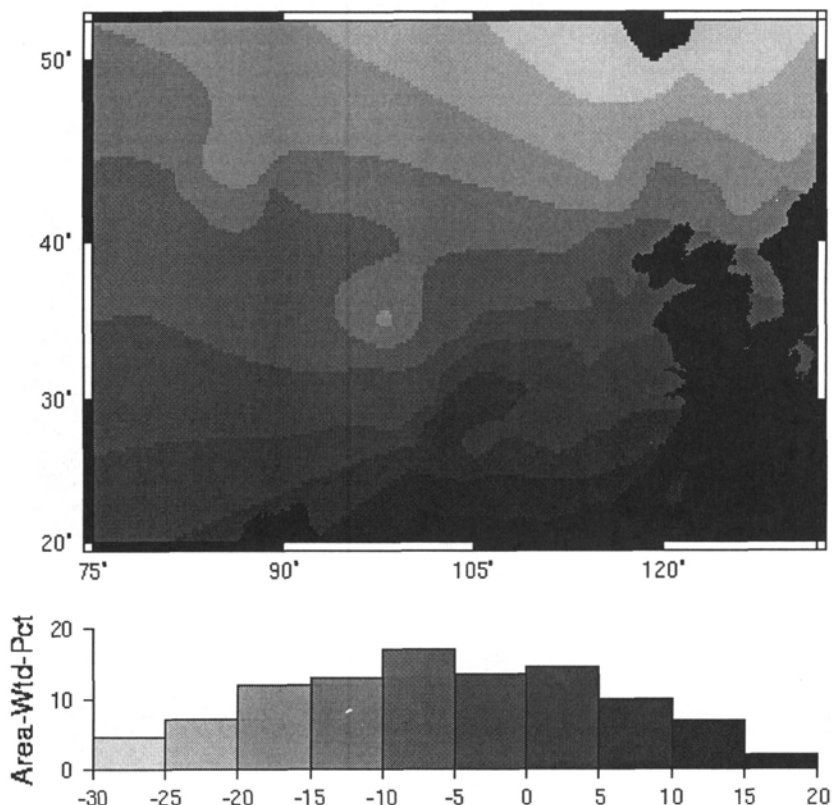


Figure 2 Conventional interpolation (°C)

The interpolation was performed using the following steps:

1. Reduce temperatures to sea level using the environmental lapse rate
 $(\text{SeaLevTemp} = \text{Temp} + \text{EnvLapseRate} * \text{Elevation})$
2. Interpolate the "sea level" temperatures to a one-degree grid using the multiquadric method
3. Reintroduce elevation effect on the interpolated field
 $(\text{Temp} = \text{SeaLevelTemp} - \text{EnvLapseRate} * \text{Elevation})$

This final step (the inversion of the operations inherent in Step 1) is carried out automatically by Spherekit. The user does not have to explicitly return the sea-level temperatures to actual temperatures.

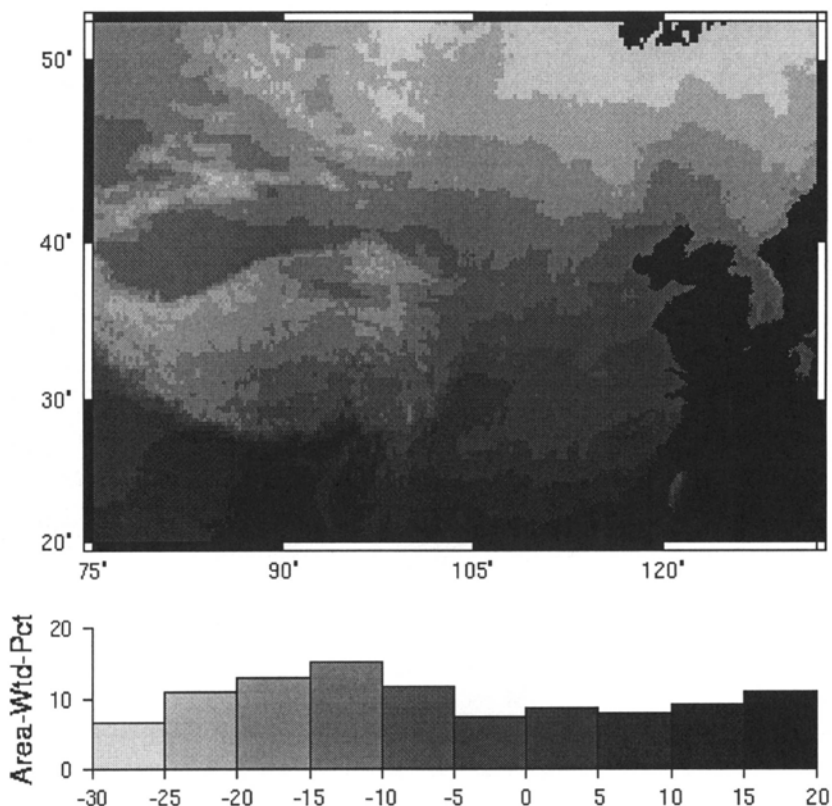


Figure 3 Smart interpolation (°C)

Error analysis

This example demonstrates the cross-validation analysis capabilities of Spherkite. In cross-validation, a data point is removed and its value is interpolated using the remaining $n-1$ points. The difference (actual - predicted) is the interpolation error at that point. Spherkite provides the option of interpolating the errors to a regular grid to reduce the spatial bias. Figure 4 shows the cross-validation error of a temperature dataset for Australia. Thin-plate splines were used as the interpolation method; the gridded plot reveals the one-degree granularity of the interpolation. Errors are reported in terms of three measures: mean average, mean bias, and root mean square.

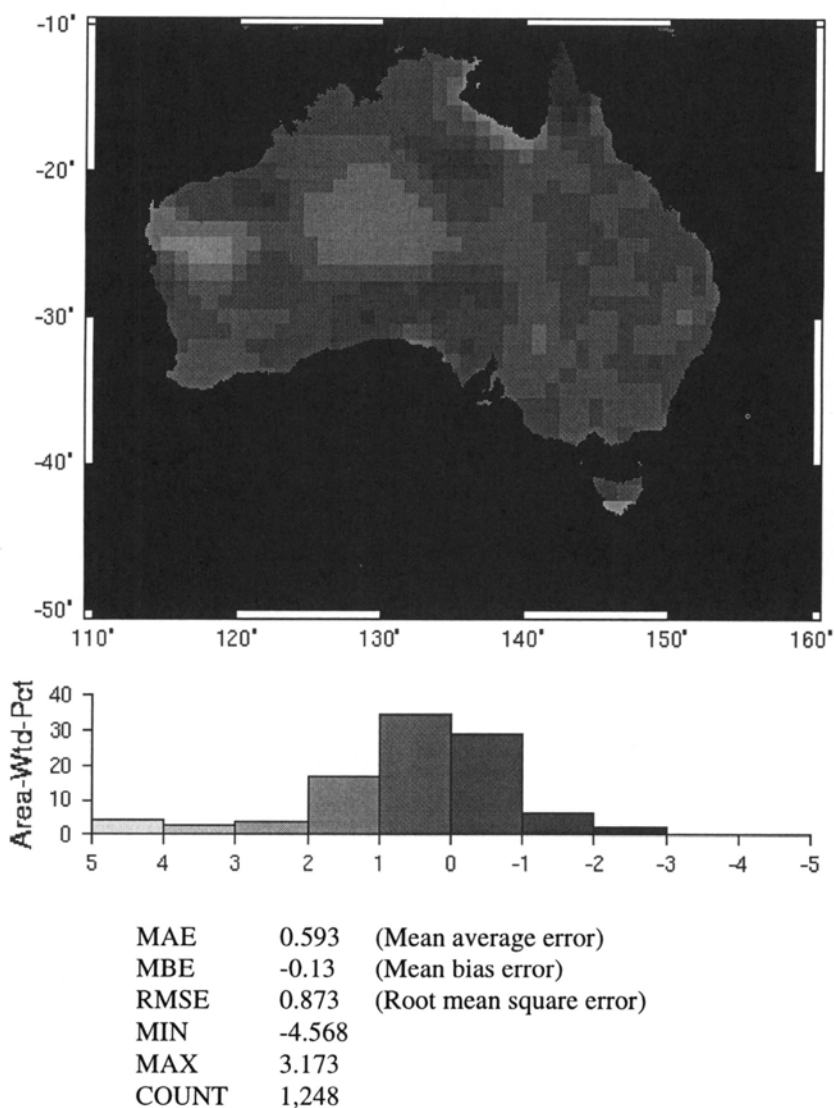


Figure 4 Thin-plate spline cross-validation errors (°C)

Spatial variability

A final example demonstrates the exploratory analysis capabilities of the package. A global temperature dataset is used to demonstrate long-distance correlations present in climate data. As Spherkit computes distances using

great circle distances, distances at continental and global scales are computed correctly.

Figure 5 shows an isotropic semivariogram of the dataset. There is a plateau in the semivariogram in the 2000-4000 km range and a sharp rise thereafter. This calculation is repeated using anisotropic semivariograms in the east-west and north-south directions. Figure 6 (the east-west semivariogram) displays the plateau more prominently. This characteristic corresponds to the common notion that zonal variations are relatively small. The north-south variations in Figure 7 vary at shorter distances, as would be expected. Interestingly, the semivariogram falls after reaching a peak; presumably this is due to a return to the same latitude zone at these distances.

AVAILABILITY

Spherekit runs on most UNIX-based machines. The source code can be downloaded from the Spherekit home page at:

www.ncgia.ucsb.edu/pubs/spherekit/main.html

The code uses Tcl/Tk for its Graphical User Interface (GUI), Generic Mapping Tools (GMT) for display of output fields, and netCDF for storing the DEM data. All of these auxiliary packages are required and can be downloaded together with Spherekit.

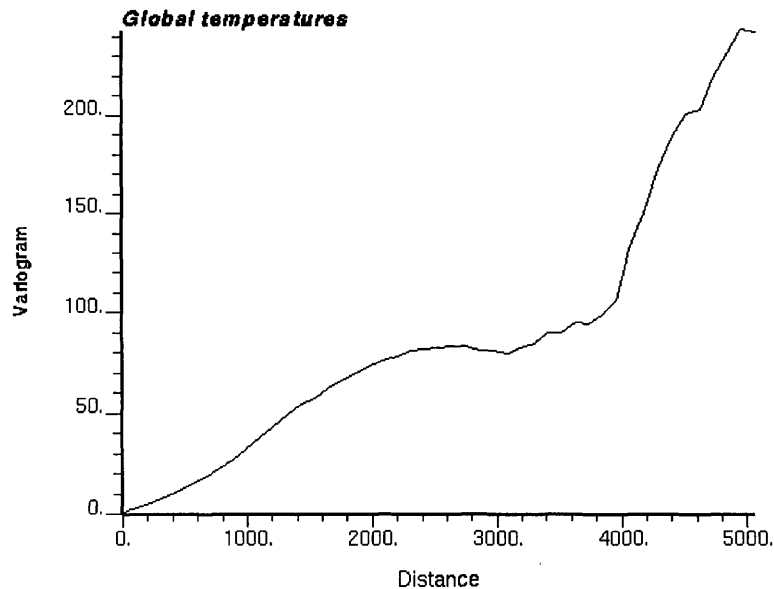


Figure 5 Isotropic semivariogram ($^{\circ}\text{C}$)²

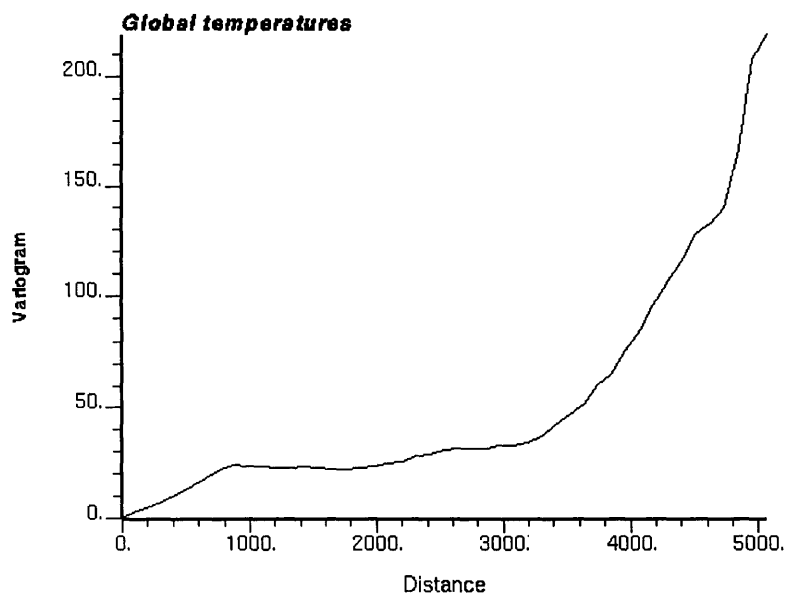


Figure 6 East-West semivariogram $(^{\circ}\text{C})^2$

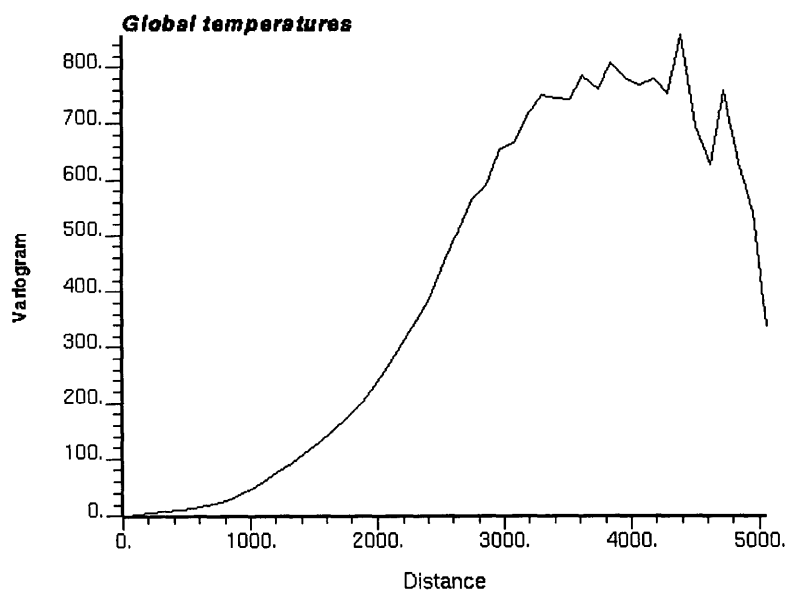


Figure 7 North-South semivariogram $(^{\circ}\text{C})^2$

REFERENCES

- Deutsch, C. V. and A. G. Journel (1992). *GSLIB: Geostatistical Software Library and User's Guide*, New York, Oxford University Press.
- Franke, R. (1982). Scattered data interpolation: Tests of some methods, *Math. Comp.*, 46: 181-200.
- Pottmann, H. and M. Eck (1990). Modified multiquadric methods for scattered data interpolation over a sphere, *Computer Aided Design*, 7: 313-321.
- Raskin, R. G. (1994). Spatial analysis on the sphere, Technical Report 94-7, National Center for Geographic Information and Analysis, 44 pp.
- Renka, R. J. (1984). Interpolation of data on the surface of a sphere, *ACM Transactions on Mathematical Software*, 10: 417-436.
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data, Proc. 23rd National Conference ACM, ACM, pp. 517-524.
- Watson, D. F. (1992). *Contouring: A Guide to the Analysis and Display of Spatial Data*, Pergamon Press, 321 pp.
- Willmott, C.J. and K. Matsuura (1995). Smart interpolation of annually averaged air temperature in the United States, *Journal of Applied Meteorology*, 34(12), 2577-2586.