1-1-2006

Spatial prediction for massive datasets

Noel A. Cressie
Ohio State University, ncressie@uow.edu.au

Gardar Johannesson
Lawrence Livermore National Laboratory

Spatial prediction for massive datasets

Abstract

Remotely sensed spatio-temporal datasets on the order of megabytes to terrabytes are becoming more common. For example, polar-orbiting satellites observe Earth from space, monitoring the Earth's atmospheric, oceanic, and terrestrial processes, and generate massive amounts of environmental data. The current generation of satellites, such as the National Aeronautic and Space Administration's (NASA) Earth Observing System (EOS) Terra and Aqua satellites, generate about 1.5 terrabytes of data per day. In the USA, there are remote-sensing projects in preparation that will dwarf even these datasets. NASA, the National Oceanic and Atmospheric Administration (NOAA), and the Department of Defense (DoD) have created the National Polarorbiting Operational Environmental Satellite System (NPOESS) to provide long-term systematic measurements of Earth's environmental variables beginning about 2009. The precursor of this NPOESS mission, the NPOESS Preparatory Project (NPP), serves as a bridge between NPOESS and NASA's EOS program and is scheduled to launch in Fall 2006. Scalable statistical methods are needed to process and extract information from these massive datasets.

Keywords

prediction, massive, spatial, datasets

Disciplines

Engineering | Science and Technology Studies

Publication Details

Cressie, N. & Johannesson, G. (2006). Spatial prediction for massive datasets. Mastering the Data Explosion in the Earth and Environmental Sciences: Australian Academy of Science Elizabeth and Frederick White Conference (pp. 1-11).

Spatial Prediction for Massive Datasets

Noel Cressie*
The Ohio State University, Columbus, USA
and Gardar Johannesson
Lawrence Livermore National Laboratory, Livermore CA, USA

Background

Remotely sensed spatio-temporal datasets on the order of megabytes to terrabytes are becoming more common. For example, polar-orbiting satellites observe Earth from space, monitoring the Earth's atmospheric, oceanic, and terrestrial processes, and generate massive amounts of environmental data. The current generation of satellites, such as the National Aeronautic and Space Administration's (NASA) Earth Observing System (EOS) Terra and Aqua satellites, generate about 1.5 terrabytes of data per day. In the USA, there are remote-sensing projects in preparation that will dwarf even these datasets. NASA, the National Oceanic and Atmospheric Administration (NOAA), and the Department of Defense (DoD) have created the National Polar-orbiting Operational Environmental Satellite System (NPOESS) to provide long-term systematic measurements of Earth's environmental variables beginning about 2009. The precursor of this NPOESS mission, the NPOESS Preparatory Project (NPP), serves as a bridge between NPOESS and NASA's EOS program and is scheduled to launch in Fall 2006. Scalable statistical methods are needed to process and extract information from these massive datasets.

Of particular interest here is Total Column Ozone (TCO) data from the Total Ozone Mapping Spectrometer (TOMS) instrument (http://toms.gsfc.nasa.gov). Flying on NPP is a whole suite of sensors, including the Ozone Mapping and Profiler Suite instrument used in obtaining measurements of TCO. This will be the next generation of the TOMS instrument that has flown on three satellites since 1978 (Nimbus-7, Meteor-3, and Earth Probe).

In spite of a satellite's regular polar orbit, remotely sensed data yield datasets that are spatially (and temporally) irregular and on occasions are missing whole swaths. Hence, further processing of these data is required to yield a dataset that is regularly

* *Address for correspondence*: Noel Cressie, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus OH 43210-1247 (ncressie@stat.ohio-state.edu)

located on the globe. However, these data are spatially (and temporally) dependent, and they are typically nonstationary in space (and time).

Kriging, or spatial best linear unbiased prediction (spatial BLUP), has become very popular in the earth and environmental sciences, where it is sometimes known as optimum interpolation. With its internal quantification of spatial variability through the covariance function (or variogram), kriging methodology is able to produce maps of optimal predictions and associated prediction standard errors from incomplete and noisy spatial data (e.g., Cressie, 1993, Ch. 3). Solving the kriging equations directly involves inversion of an $n \times n$ variance-covariance matrix $\mathbf{\Sigma}$, where n data may require $O(n^3)$ computations to obtain $\mathbf{\Sigma}^{-1}$. Under these circumstances, straightforward kriging based on massive data is impossible.

It has been realized for some time that even a spatial dataset on the order of several thousand can result in a computational breakdown. Ad hoc methods of subsetting the data were formalized by the moving-window approach of Haas (1995), although it would appear that the local covariance functions fitted within the window may yield incompatible covariances at larger spatial lags. The variance-covariance matrix $\mathbf{\Sigma}$ is typically sparse when the covariance function has a finite range and hence $\mathbf{\Sigma}^{-1}$ can be obtained using sparse-matrix techniques. Barry and Pace (1997) were able to carry out kriging when $n = 916$ using a MATLAB routine that is based on the symmetric minimum degree algorithm. Rue and Tjelmeland (2002) *approximate* $\mathbf{\Sigma}^{-1}$ to be sparse, approximating it to be the precision matrix of a Gaussian Markov random field wrapped on a torus.

When datasets are large (on the order of tens of thousands to hundreds of thousands), the general feeling is that kriging is impossible and ad hoc local kriging neighborhoods are typically used (e.g., Cressie, 1993, pp. 131-134). One avenue of recent research has been to *approximate* the kriging equations (Nychka et al., 1996; Nychka, 2000; Nychka, Wikle, and Royle, 2002; Furrer, Genton, and Nychka, 2005). Suggestions include giving an equivalent representation in terms of orthogonal bases and truncating the bases, doing covariance tapering, using approximate iterative methods such as conjugate-gradient, or replacing the data locations with a smaller set of space-filling locations. Kammann and Wand (2003) take up this latter idea when fitting a class of spatial models they call geoadditive models.

Another approach has been to choose classes of covariance functions for which kriging can be done *exactly*, even though the data are massive. Huang *et al.* (2002) introduced a multi-resolution spatial model (MRSM) that is mass balanced (across resolutions) and designed for processing massive amounts of spatial data. The advantage of the MRSM lies in the fact that it is able to capture nonstationary spatial dependence and to produce fast optimal estimates using a change-of-resolution Kalman-filter algorithm (Chou *et al.*, 1994; Huang and Cressie, 2001). Later developments were given by Johannesson and Cressie (2004a) and Johannesson et al. (2007). In these papers, a multi-resolution spatial (and spatio-temporal) process is constructed explicitly so that (simple) kriging can be computed extremely rapidly,

with computational complexity linear in the size of the data. In the spatial case, Johannesson and Cressie (2004a) achieved speed-ups of the order of 10^8 over direct kriging. They were able to compute optimal spatial predictors and their associated mean squared prediction errors in about 3 minutes for $n \simeq 160,000$. The advantage of having a spatial model that allows exact computations is that there is no concern about how close approximate kriging predictors and approximate mean squared prediction errors are to the corresponding theoretical values.

When kriging using exact methods, an important question is then, how flexible are the spatial covariance functions? For the multi-resolution models referred to above, the implied spatial covariances are nonstationary and "blocky". Cressie and Johannesson (2006) use a different approach to achieve orders-of-magnitude speed-ups for optimal spatial prediction, using covariance functions that are very flexible and can be chosen to be smooth or not, as determined by the type of spatial dependence exhibited by data. They show how to define the $n \times n$ variance-covariance matrix $\mathbf{\Sigma}$ so that $\mathbf{\Sigma}^{-1}$ can be obtained by inverting $r \times r$ matrices, where r is fixed. Then the number of computations per prediction location in the kriging equations is $O(nr^3)$, which increases only linearly with sample size.

Furthermore, suppose that the dataset is the result of remote sensing from a satellite that achieves global coverage. Then any spatial dependencies in the data will almost certainly be heterogeneous across the globe. The methodology in their paper addresses both problems (massiveness and heterogeneity) directly; the result is a spatial BLUP procedure they call Fixed Rank Kriging (FRK).

For completeness, we mention another approach to spatial prediction, based on smoothing splines. In contrast to kriging, smoothing splines do not rely on a spatial stochastic process whose covariance function has to be modeled, fitted, and used for computing the optimal predictor. However, there are knots and a smoothing parameter to be determined and, once again, the massiveness of the data causes computational difficulties. Hastie (1996) and Johannesson and Cressie (2004b) develop low-rank spline smoothers for massive datasets.

Spatial Covariance Function

In order to carry out FRK, we must specify the form of the nonstationary covariance function. In general, the covariance function $C(\mathbf{u}, \mathbf{v})$ has to be positive-definite on $\mathbb{R}^d \times \mathbb{R}^d$. Often $C(\mathbf{u}, \mathbf{v})$ is modeled as being stationary, in which case it has to be a positive-definite function of $(\mathbf{u} - \mathbf{v})$. We take a different approach and instead try to capture the scales of spatial dependence through a set of r (not necessarily orthogonal) basis functions,

$$\mathbf{S}(\mathbf{u}) \equiv (S_1(\mathbf{u}), \ldots, S_r(\mathbf{u}))'; \quad \mathbf{u} \in \mathbb{R}^d,$$

where r is fixed and $\mathbf{S}(\cdot)$ is given. For any $r \times r$ *positive-definite* matrix $\mathbf{K}$, we specify

$$C(\mathbf{u}, \mathbf{v}) = \mathbf{S}(\mathbf{u})'\mathbf{K}\mathbf{S}(\mathbf{v})\,; \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^d\,,$$

which is straightforwardly a positive-definite function, and hence a valid covariance function. The entries of $\mathbf{K}$ are unknown parameters to be estimated.

Kriging: Optimal Linear Spatial Prediction

In this section, we present the notation for, and the definition of kriging, and we equate it with best linear unbiased prediction (BLUP) in a spatial setting. When the datasets are massive, exact computation of the spatial BLUP is generally impossible. However, with the class of nonstationary spatial covariances given above, we can carry out rapid computation of the kriging predictor (spatial BLUP) and the kriging standard error (root mean squared prediction error).

Let $\{Y(\mathbf{s})\colon \mathbf{s} \in D \subset \mathbb{R}^d\}$ be a real-valued spatial process. We are interested in making inference on the Y-process based on data that have measurement error incorporated; consider the process $Z(\cdot)$ of actual and potential observations,

$$Z(\mathbf{s}) \equiv Y(\mathbf{s}) + \varepsilon(\mathbf{s})\,; \quad \mathbf{s} \in D\,,$$

where $\{\varepsilon(\mathbf{s})\colon \mathbf{s} \in D\}$ is a spatial white-noise process with mean 0 and $\mathrm{var}(\varepsilon(\mathbf{s})) = \sigma^2 v(\mathbf{s}) \in [0, \infty)$; $\mathbf{s} \in D$. In fact, the process $Z(\cdot)$ is known only at a finite number of spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$; define the vector of available data to be

$$\mathbf{Z} \equiv (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'\,.$$

The hidden process $Y(\cdot)$ is assumed to have a linear mean structure,

$$Y(\mathbf{s}) = \mathbf{t}(\mathbf{s})'\boldsymbol{\alpha} + \nu(\mathbf{s})\,; \quad \mathbf{s} \in D\,,$$

where $\mathbf{t}(\cdot) \equiv (t_1(\cdot), \ldots, t_p(\cdot))'$ represents a vector process of known covariates; the coefficients $\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_p)'$ are unknown; and the process $\nu(\cdot)$ has zero mean, $0 \le \mathrm{var}(\nu(\mathbf{s})) < \infty$, for all $\mathbf{s} \in D$, and a spatial covariance function,

$$\mathrm{cov}(\nu(\mathbf{u}), \nu(\mathbf{v})) \equiv C(\mathbf{u}, \mathbf{v})\,; \quad \mathbf{u}, \mathbf{v} \in D\,,$$

which for the moment is left as general as possible.

If we define $\boldsymbol{\varepsilon}$, $\mathbf{Y}$, and $\boldsymbol{\nu}$ in an analogous manner to $\mathbf{Z}$, then the preceding equations imply a general linear mixed model,

$$\begin{aligned} \mathbf{Z} &= \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\delta}\,, \\ \boldsymbol{\delta} &= \boldsymbol{\nu} + \boldsymbol{\varepsilon}\,, \end{aligned}$$

where $\mathbf{T}$ is an $n \times p$ matrix of covariates $(\mathbf{t}(\mathbf{s}_1), \ldots, \mathbf{t}(\mathbf{s}_n))'$. Observe that the error term $\boldsymbol{\delta}$ is made up of two zero-mean components, resulting in $E(\boldsymbol{\delta}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma} \equiv (\sigma_{ij})$, where

$$\sigma_{ij} = \begin{cases} C(\mathbf{s}_j, \mathbf{s}_j) + \sigma^2 v(\mathbf{s}_j) & ; \quad i = j \\ C(\mathbf{s}_i, \mathbf{s}_j) & ; \quad i \neq j \,. \end{cases}$$

Upon writing $\mathbf{C} \equiv (C(\mathbf{s}_i, \mathbf{s}_j))$ and $\mathbf{V} \equiv \mathrm{diag}(v(\mathbf{s}_1), \ldots, v(\mathbf{s}_n))$, it is easily seen that

$$\boldsymbol{\Sigma} = \mathbf{C} + \sigma^2 \mathbf{V} \,.$$

No assumptions of stationarity or isotropy of the covariance functions have been made.

Interest is in inference on the Y-process, not the noisy Z-process. For point prediction, we wish to predict the Y-process at a location $\mathbf{s}_0$; $\mathbf{s}_0 \in D$, regardless of whether $\mathbf{s}_0$ is or is not an observation location. Cressie (1993, Section 3.4.5) shows that one formula for the kriging predictor of $Y(\mathbf{s}_0)$ is:

$$\widehat{Y}(\mathbf{s}_0) \;=\; \mathbf{t}(\mathbf{s}_0)'\widehat{\boldsymbol{\alpha}} + \mathbf{k}(\mathbf{s}_0)'(\mathbf{Z} - \mathbf{T}\widehat{\boldsymbol{\alpha}}) \,, \tag{1}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= (\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1}\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} \,, \\ \mathbf{k}(\mathbf{s}_0)' &= \mathbf{c}(\mathbf{s}_0)'\boldsymbol{\Sigma}^{-1} \,, \end{aligned}$$

and $\mathbf{c}(\mathbf{s}_0) \equiv (C(\mathbf{s}_0, \mathbf{s}_1), \ldots, C(\mathbf{s}_0, \mathbf{s}_n))'$. The kriging standard error is the root mean squared prediction error of $\widehat{Y}(\mathbf{s}_0)$, given by:

$$\begin{aligned} \sigma_k(\mathbf{s}_0) \;=\; \{&C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{k}(\mathbf{s}_0)'\boldsymbol{\Sigma}\mathbf{k}(\mathbf{s}_0) \\ &+ (\mathbf{t}(\mathbf{s}_0) - \mathbf{T}'\mathbf{k}(\mathbf{s}_0))'(\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1}(\mathbf{t}(\mathbf{s}_0) - \mathbf{T}'\mathbf{k}(\mathbf{s}_0))\}^{1/2} \,. \end{aligned} \tag{2}$$

As the prediction location $\mathbf{s}_0$ varies over D, a kriging-prediction map and a kriging-standard-error map, respectively, are generated. (In practice, prediction locations are finite in number and typically taken as nodes of a fine-resolution grid superimposed on D.)

Inspection of the kriging equations shows $\boldsymbol{\Sigma}^{-1}$ to be an essential component and the most obvious place where a computational bottleneck could occur. Cressie and Johannesson (2006) show that for the given class of covariance functions given in the previous section,

$$\boldsymbol{\Sigma} = \mathbf{S}\mathbf{K}\mathbf{S}' + \sigma^2 \mathbf{V} \,,$$

and hence,

$$\boldsymbol{\Sigma}^{-1} = (\sigma^2\mathbf{V})^{-1} - (\sigma^2\mathbf{V})^{-1}\mathbf{S}\{\mathbf{K}^{-1} + \mathbf{S}'(\sigma^2\mathbf{V})^{-1}\mathbf{S}\}^{-1}\mathbf{S}'(\sigma^2\mathbf{V})^{-1} \,.$$

Notice that this formula involves inverting either *fixed-rank* $r \times r$ positive-definite matrices or the $n \times n$ *diagonal* matrix $\mathbf{V}$. Inspection of the kriging equations reveals that for a fixed number of regressors p and a fixed rank r of the covariance model, the computational burden is only linear in n. Thus, it becomes feasible to construct maps of kriging predictors and kriging standard errors based on massive amounts of data.

Total Column Ozone Over the Globe

The problem of measuring total column ozone (TCO) has been of interest to scientists for decades. Ozone depletion results in an increased transmission of ultraviolet radiation (290-400 nm wavelength) through the atmosphere. This is mostly deleterious due to damage to DNA and cellular proteins that are involved in biochemical processes, affecting growth and reproduction.

Relatively few measurements of TCO were taken in the first quarter of the twentieth century. Subsequently, with the invention of the Dobson spectrophotometer, researchers gained the ability to measure efficiently and accurately TCO abundance (London, 1985). A system of ground-based stations has provided important TCO measurements for the past 40 years; however, the ground-based stations are relatively few in number and provide poor geographic coverage of the earth. The advent of polar-orbiting satellites has dramatically enhanced the spatial coverage of measurements of TCO.

The Nimbus-7 polar-orbiting satellite was launched on October 24, 1978, with the Total Ozone Mapping Spectrometer (TOMS) instrument aboard. The TOMS instrument scans in three-degree steps to an extreme of 51 degrees on each side of nadir, in a direction perpendicular to the orbital plane (McPeters *et al.*, 1996). Each scan takes roughly eight seconds to complete, including one second for retrace (Madrid, 1978). The altitude of the satellite and scanning pattern of the TOMS instrument are such that consecutive orbits overlap, with the area of overlap depending on the latitude of the measurement. The TOMS instrument covers the entire globe in a 24-hour period. NASA receives the data, calibrates it ("level 1"), and pre-processes it to yield spatially and temporally irregular TCO measurements ("level 2"). The level-2 data are subsequently processed to yield a spatially and temporally uniform data product that is released widely to the scientific community ("level 3"). The level-3 data product uses 1 degree latitude by 1.25 degrees longitude ($1^\circ \times 1.25^\circ$) equiangular grid cells (McPeters *et al.*, 1996, p. 44).

Level-2 TCO data and the level-3 data product released by NASA were obtained from the Ozone Processing Team of NASA/Goddard, Distributed Active Archive Center, and were stored in Hierarchical Data Format as developed by the National Center for Supercomputing Applications at the University of Illinois. The goal is to produce a level-3 data product for all $1^\circ \times 1.25^\circ$ grid cells, on a daily basis, from the spatially irregular level-2 data referred to above. Based on the development in the previous section, FRK-based optimal spatial predictions of TCO can be used as a level-3 data product.

In what follows, we have used the $173,405$ level-2 TCO data available for October 1, 1988; see Figure 1.

The basis functions we chose are made up of three scales of variation. Each scale has respectively 32, 92, and 272 functions associated with them, corresponding to the center points of a discrete global grid (DGG); see Figure 2.

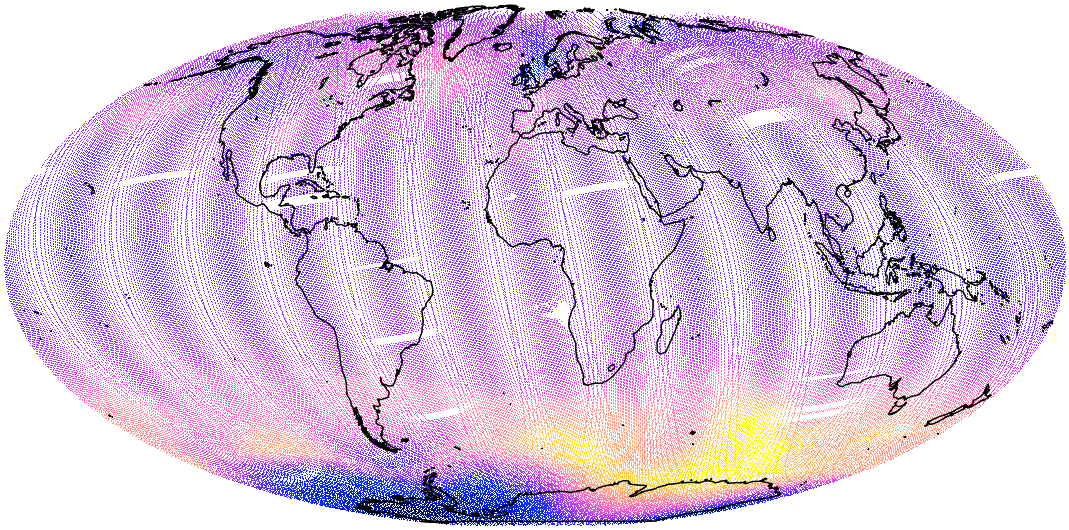

Figure 1: 10/1/88, level-2 TCO data

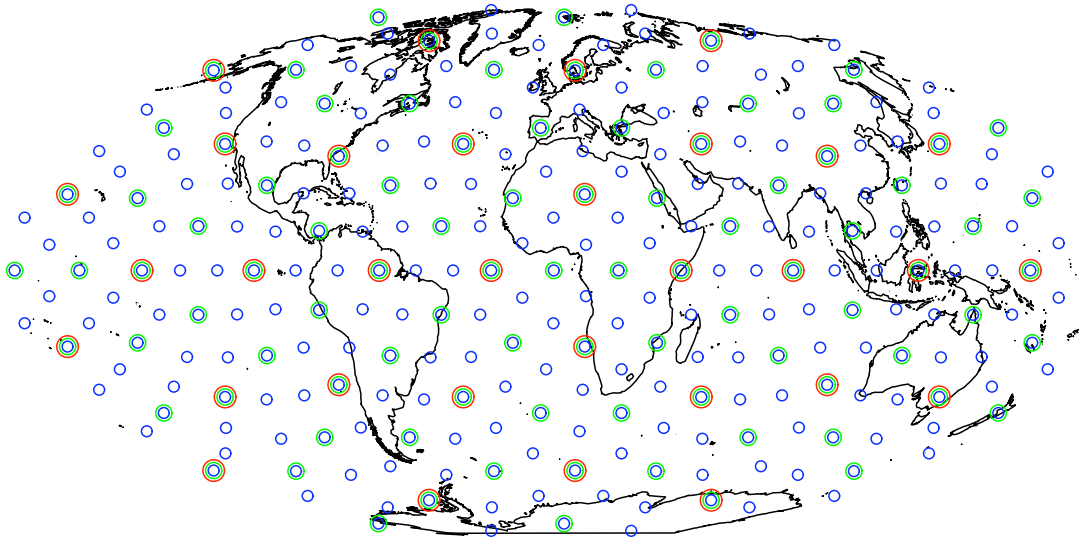

Figure 2: Center points of 3 resolutions of DGG

A generic basis function is:

$$S_j(\mathbf{u}) \equiv \begin{cases} (1 - (\|\mathbf{u} - \mathbf{v}_j\|/r_j)^2)^2\,; & \|\mathbf{u} - \mathbf{v}_j\| \le r_j \\ 0\,; & \text{otherwise,} \end{cases}$$

where $\mathbf{v}_j$ is one of the center points of Figure 2 and

$$r_j = (1.5) \times (\text{shortest distance between like-resolution center points}) .$$

For example, if $\mathbf{v}_j$ is from resolution 1, the shortest distance is $4,165$km and $r_j = 6,747.5$; the distances between center points from resolution 2 and 3 are $1,610$km and $1,435$km, respectively. Notice that there are a total of $r = 32 + 92 + 272 = 396$ basis functions.

Cressie and Johannesson (2006) give a method for estimating $\mathbf{K}$ and σ^2 in

$$\boldsymbol{\Sigma} = \mathbf{SKS} + \sigma^2\mathbf{V} ,$$

which results in excellent fits of the theoretical, non-stationary variograms to the empirical variograms; see Figure 3.

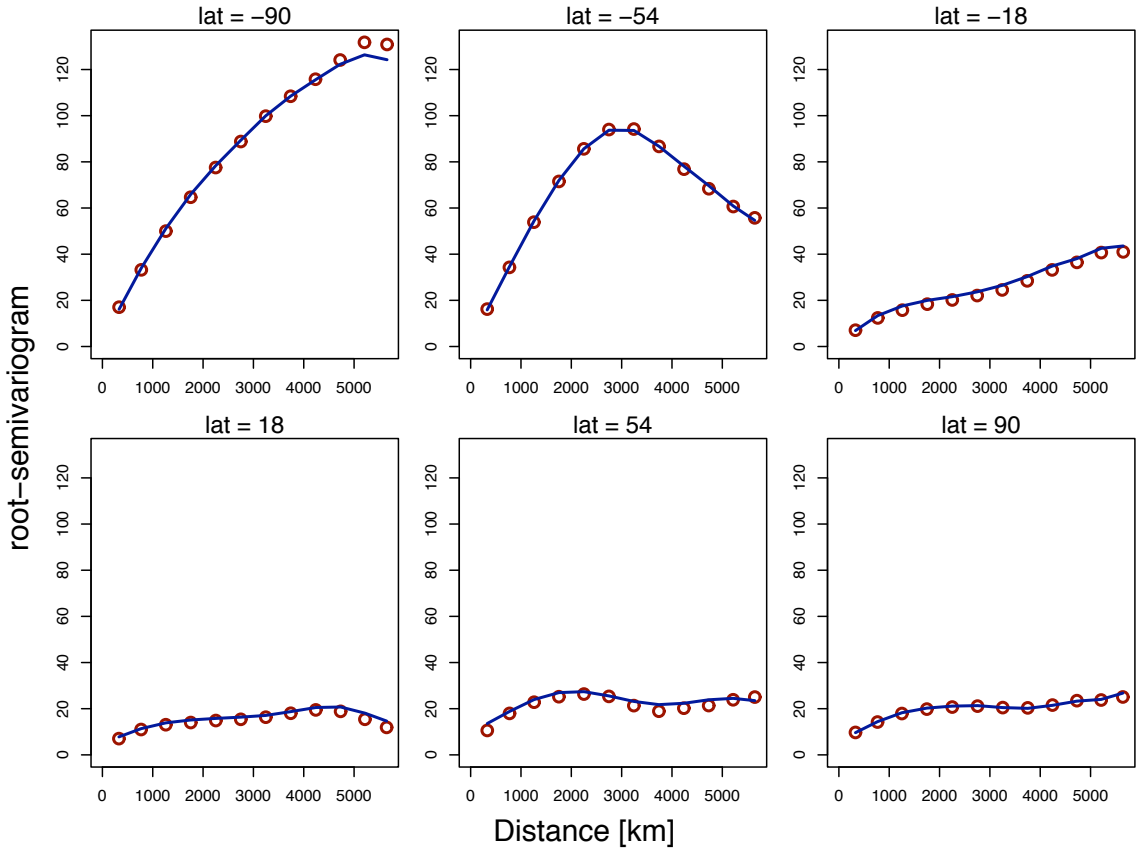

Figure 3: Semivariograms (on square root scale) for different locations

Finally, assuming a constant mean (i.e., $\mathbf{t}(\mathbf{s}) \equiv 1$) and using the kriging predictor (1), we obtain the level-3 data product shown in Figure 4.

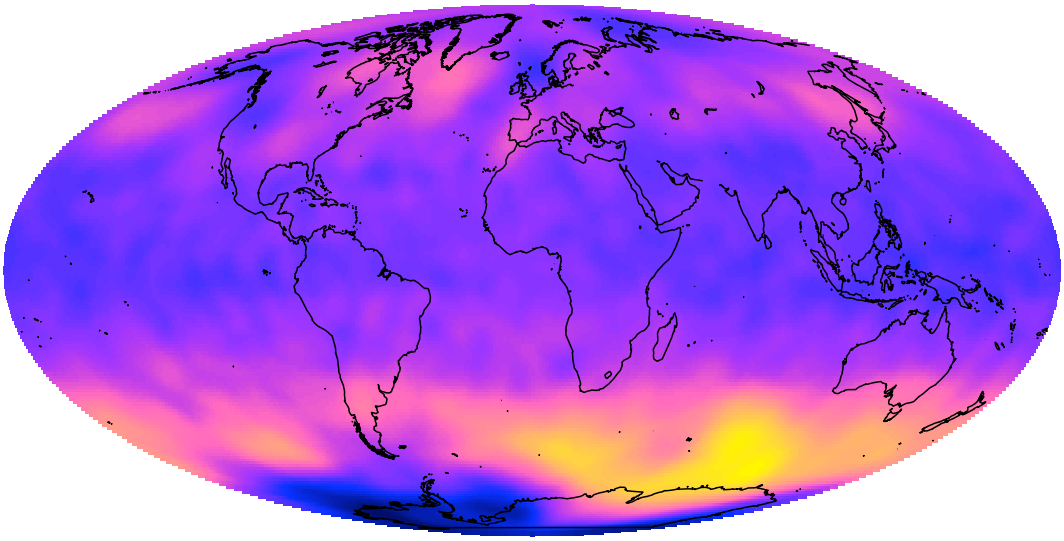

Figure 4: 10/1/88, kriging predictor of TCO

We would like to emphasize that *all* the 173,405 data were used to produce Figure 4, the covariance function we used is *nonstationary*, a matrix inversion of only a 396×396 positive-definite matrix was needed to produce Figure 4, and the map in the figure is the *optimal* predictor (for squared-error loss) of TCO on the $1^\circ \times 1.25^\circ$ grid.

Acknowledgment

This research was supported by the Office of Naval Research under grant number N00014-05-1-0133.

References

Barry, R. P. and Pace, R. K. (1997). Kriging with large data sets using sparse matrix techniques. *Communications in Statistics: Simulation and Computation*, **26**, 619-629.

Chou, K. C., Willsky, A. S., and Nikoukhah, R. (1994). Multiscale systems, Kalman filters, and Riccati equations. *IEEE Transactions on Automatic Control*, **39**, 479-492.

Cressie, N. (1993). *Statistics for Spatial Data, rev. edn.* Wiley: New York.

Cressie, N. and Johannesson, G. (2006). Fixed rank kriging for massive datasets. Manuscript.

Furrer, R., Genton, M. G., and Nychka, D. (2005). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, submitted.

Haas, T. C. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association*, **90**, 1189-1199.

Hastie, T. (1996). Pseudosplines. *Journal of the Royal Statistical Society B*, **58**, 379-396.

Huang, H.-C. and Cressie, N. (2001). Multiscale graphical modeling in space: applications to command and control. In *Spatial Statistics: Methodological Aspects and Some Applications*, M. Moore (ed). Springer Lecture Notes in Statistics (Vol. **159**), Springer: New York, pp. 83-113.

Huang, H.-C., Cressie, N., and Gabrosek, J. (2002). Fast, resolution-consistent spatial prediction of global processes from satellite data. *Journal of Computational and Graphical Statistics*, **11**, 63-88.

Johannesson, G. and Cressie, N. (2004a). Variance-covariance modeling and estimation for multi-resolution spatial models. In *geoENV IV - Geostatistics for Environmental Applications*, X. Sanchez-Vila, J. Carrera, and J. Gomez-Hernandez (eds). Kluwer Academic Publishers: Dordrecht, 319-330.

Johannesson, G. and Cressie, N. (2004b). Finding large-scale spatial trends in massive, global, environmental datasets. *Environmetrics*, **15**, 1-44.

Johannesson, G., Cressie, N., and Huang, H.-C. (2007). Dynamic multi-resolution spatial models. *Environmental and Ecological Statistics*, forthcoming.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Applied Statistics*, **52**, 1-18.

London, J. (1985). The observed distribution of atmospheric ozone and its variations. In *Ozone in the Free Atmosphere*, R. C. Whitten and S. S. Prasad (eds). Van Nostrand Reinhold Company: New York, 11-80.

Madrid, C. R. (1978). *The Nimbus-7 User's Guide*. Goddard Space Flight Center, NASA: Greenbelt, MD.

McPeters, R. D., Bhartia, P. K., Krueger, A. J., Herman, J. R., Schlesinger, B. M., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Taylor, S. L., Swissler, T., Torres, O., Labow, G., Byerly, W., and Cebula, R. P. (1996). *The Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide*. NASA Reference Publication 1384, Greenbelt, MD.

Nychka, D. (2000). Spatial-process estimates as smoothers. In *Smoothing and Regression: Approaches, Computation, and Application*, M. G. Schimek (ed). Wiley: New York, 393-424.

Nychka, D., Bailey, B., Ellner, S., Haaland, P., and O'Connell, M. (1996). *FUNFITS: Data analysis and statistical tools for estimating functions*. North Carolina Institute of Statistics Mimeoseries, No. 2289, North Carolina State University, Raleigh, NC.

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modeling*, **2**, 315-331.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, **29**, 31-49.