STATISTICAL INTERPOLATION METHODS

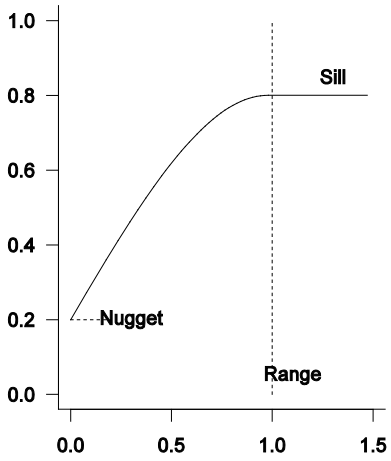
RICHARD SMITH AND NOEL CRESSIE

Statistical methods of interpolation are all based on assuming that the process being reconstructed (for the purpose of this report, a temperature field in space or time or both) can be modeled as a random process. The best-known examples occur in spatial statistics, including the technique widely known as kriging, and in time series analysis. The process may depend on unknown parameters, which have to be estimated as part of the interpolation procedure. Once the process is specified, including any estimated parameters, optimal interpolators are calculated either by finding the best linear interpolator (the linear combination of known observations that minimizes the mean squared prediction error) or by computing a given function of the conditional probability distribution of the predicted quantity conditional on the observations. That function can be determined by appealing to statistical decision theory.

Gaussian processes form a special class of random processes. They are defined by the property that all joint distributions of the process are multivariate Gaussian. In practice, a process is usually assumed Gaussian if histograms and other data-plotting techniques suggest a Gaussian distribution. This is usually considered an acceptable assumption for temperatures and other meteorological variables (e.g., atmospheric pressure) that have smooth continuous distributions. Precipitation, however, is an exception, owing to the many zero values and the tendency of precipitation to come in short, sharp bursts of heavy rainfall interspersed with much lighter or no precipitation. Some quantities may be transformed (e.g., by taking logarithms) to be Gaussian.

There are two reasons why the assumption of a Gaussian process is convenient. First, any Gaussian process is completely specified by its means and its covariances, so we do not need to worry about higher-order moments. Second, for Gaussian processes, it can be shown that the best predictor is linear in the observations, so it suffices to restrict attention to linear predictors.

Kriging is the name commonly given to optimal linear prediction of a spatial process. Although the origins of the method go back to signal processing techniques developed by Kolmogorov and Wiener in the 1940s, it was first developed systematically by Matheron and by Gandin (who called it optimal interpolation) in the 1960s. Modern statistical treatments include the books by Cressie (1993) and Stein (1999). As commonly applied, it contains the following elements:



a. The method requires that we specify either the covariance or the variogram function of the process. These do not have to be spatially stationary, though they are usually assumed so. There are a number of traditional stationary models that are widely used, e.g. linear, exponential, Gaussian, spherical, Matérn. Choice among these models is often based upon which one best fits the data, though it is also argued that the Matérn covariance has desirable properties that make it suitable for a wide range of applications (Stein 1999).

b. Standard models for the covariance or variogram assume that the process is stationary (invariant under translations across space) and isotropic (invariant to rotations). Both assumptions can be tested to some degree, e.g. for isotropy, it is possible to calculate direction-specific variograms, though in practice, it is difficult to be certain that stationarity and isotropy assumptions are valid.

c. The stationary covariance or variogram functions have unknown parameters that must be estimated. For a number of models, the key parameters are the nugget, sill, and range (see diagram). Estimation may be based on either an ordinary or weighted least squares regression applied to the sample variogram, or through likelihood-based methods such as maximum likelihood, restricted maximum likelihood (REML estimation) or Bayesian methods. Variogram methods require less computation and programming effort and are therefore more convenient to apply in practice, but likelihood-based methods are more efficient statistically.

d. Ordinary kriging assumes that the mean of the process is unknown but the same at all sampling locations. Universal kriging is a generalization that allows the mean of the process to be a linear combination of regression functions. In the context of meteorological interpolation, universal kriging should be considered when the mean of the process depends on latitude-longitude coordinates or other measured quantities such as elevation.

e. Once the mean and covariance (or variogram) of a process are specified, kriging consists of fixing a

prediction location or region, and then calculating the weighted linear combination of observations that minimizes the mean squared prediction error (MSPE). This is an automated process and several computer packages are available to do it.

f. The MSPE is often quoted as a measure of the uncertainty of the prediction. It should be used with caution, as it assumes the process has been specified correctly and ignores the error in estimating any process parameters such as sill and range. Nevertheless, provided the caveats are made clear, we would recommend that the MSPE or its square root should be routinely reported when stating the results of a kriging interpolation.

g. Because kriging typically interpolates among available observations, it is not a good way to represent the extremes of the process being predicted. For that purpose, it may be necessary to calculate full conditional distributions, but in that case, the distributional assumptions being made about the process (such as Gaussianity) are much more critical.

h. Other methods of interpolation, such as thin-plate splines, are often regarded as competitors to kriging. In fact, the method of thin-plate splines is a special case of kriging, corresponding to a particular specification of a generalized variogram. Kriging is in principle a more general technique than that of thin-plate splines, because it allows for different covariance or variogram specifications to be tried out and compared based on the data, and it allows for the estimation of covariance or variogram parameters.

Time series analysis is appropriate when observations are taken sequentially in time, usually at equally spaced time intervals, without a spatial component. The mathematical theory is based on similar principles to kriging, in particular, both using optimal linear combinations of existing observations as predictors of unobserved data. However, because of the long history of time series analysis as an independent discipline, the statistical models used are somewhat different, with particular emphasis being placed on autoregressive and moving-average (ARMA) models, or alternatively, time series models that use the Kalman filter (Brockwell and Davis 1991). It is possible to treat time series analysis as a one-dimensional version of kriging, using models such as the Matérn covariance function, though this is not the way time series analysis is usually done.

Spatio-temporal analysis is used when data are collected in both space and time. Since the majority of meteorological datasets indeed have both a spatial and temporal component, it might be thought that this would be the preferred method of data interpolation in meteorology. In practice, however, statistical methods that fully respect both the spatial and temporal components are much less developed than those that focus on one or the other. Therefore, a fully spatio-temporal analysis is rarely used in the practical analysis of meteorological data. Nevertheless, there have been considerable advances in the statistical theory of spatio-temporal models over the past decade (for example, Wikle and Cressie 1999), and we would view this as an important area for future research.

Applications of Spatial Interpolation Methods to Meteorological Data

The two papers by Haylock et al. (2008) and Hofstra et al. (2008) provided an up-to-date review of how these methods have been applied to actual meteorological data.

Haylock et al. proposed a three-step process for interpolation of temperature fields. In the first step, monthly mean temperatures were interpolated using thin-plate splines. In the second step, anomalies from the monthly means were interpolated using kriging. The third step consisted of combining the interpolated monthly means and anomalies, and calculating an overall uncertainty estimate.

For the first step, they used thin-plate splines based on the ANUSPLIN package described in Hutchinison (1995), which includes the use of generalized cross-validation to optimize the interpolator (Craven and Wahba 1979). They interpolated monthly means to a 0.1^{o} grid, the intention being to perform the initial interpolation on a very fine spatial scale and to aggregate into large grid boxes later.

The second stage of the analysis was to interpolate the anomalies (or residuals) from the monthly means using kriging. Parameters were estimated by using ordinary least squares fitting of the theoretical variogram function to the empirical variogram. For the choice of variogram model, the authors considered five possibilities (Gaussian, exponential, spherical, hole effect, and power), but used the spherical model for all the temperature series on the grounds of best fit as measured by a chi-squared statistic. They used a single variogram for the entire time period after establishing that two alternative methods, one based on a different variogram for each day and the other based on a different variogram for each month, performed less well in a cross-validation comparison. In interpolating at a given location, they used only stations within a certain search radius of that location, fixed at 500 km. for temperature and at 450 km. for precipitation, again using cross-validation to make that determination. They considered the possibility of direction-dependent variograms but concluded that this did not improve the results. They also incorporated elevation as an external covariate.

The third stage of the analysis was to combine the monthly and daily interpolations and calculate a combined measure of uncertainty. Separate uncertainty variances were computed for the monthly and daily components and combined by adding them together, in effect, assuming that the monthly and daily interpolations are statistically independent. For the daily kriging variances, the authors did not use the standard formula for the kriging MSPE but instead they used an alternative "interpolation variance" method due to Yamamoto (2000). Yamamoto's method is apparently designed to allow for the possibility of a process variance (or sill) that varies locally over the region of interpolation, but it might be better to incorporate this feature directly into the covariance model rather than the rather ad hoc correction proposed by Yamamoto.

The final step of the uncertainty calculation is to combine the estimates at individual locations into an estimate of uncertainty for a grid-cell average. For this calculation, it is necessary not only to know the variance of the prediction at individual locations, but also covariances of the predictors between different locations. For the latter, it appears that the authors used the fitted variogram model to determine the correlations between different locations, but there is also a direct formula for the covariance of two kriging predictors which would be a more precise calculation.

The description so far applies to the method of interpolating temperature data. It would be reasonable to apply a similar method for the interpolation of other variables obeying approximately Gaussian distributions, such as pressure, but different methods are needed for precipitation because of the highly non-normal distributions in that case. Haylock et al. used the method of "indicator kriging" to model the probability of precipitation. They first reclassified the precipitation variable as binary based on exceedances of a threshold, taken as 0.5 mm. The binary (0 or 1) variable was interpolated using standard kriging, the resulting variable being interpreted as a probability of precipitation at each location. Then, locations with a probability of precipitation below 0.4 were classified as having no rain; for the rest, a kriging method based on observed rainfall amounts was applied.

Although the "indicator kriging" method appears popular with practitioners, it does not have very good statistical properties. The choices of thresholds are ad hoc; the method may well produce predicted probabilities of rainfall that are not between 0 and 1; and since it is not based on any well-defined statistical model, its properties are impossible to determine (for example, we cannot conduct a simulation because we don't know what stochastic process to simulate). Although more complicated to apply, methods for interpolating rainfall fields that are based on stochastic process models for rainfall, such as Sansó and Guenni (2004), are in principle much more desirable.

The parallel paper by Hofstra et al. (2008) compared local kriging (LK) and global kriging (GK) with a number of alternative interpolation techniques. GK is kriging with a single variogram model used for all stations; LK is kriging restricted to a finite search radius around the location being predicted. The other methods included two forms of angular distance weighting (a variant on the better known method of inverse distance weighting; in both IDW and ADW, the interpolation weights at a given location depend only on the distances of predictor stations from that location); a "natural neighbor interpolation" (which also depends solely on the geometry of the predictor locations); thin-plate splines in either two or three dimensions, the third dimension being elevation; and a regression method using latitude, longitude, elevation and distance to coast as predictors. In the case of precipitation, a method of "conditional interpolation", based on a classification of rainfall into synoptic states, was also considered. For each method, a variety of skill scores was used to assess its predictive ability; these included mean absolute deviation, root mean squared error and Pearson correlation, besides several others. Although the results varied by skill score and variable being predicted, the authors concluded that GK was consistently the best of the methods, with the sole exception of maximum temperature for which three-dimensional thin plate splines performed better.

This paper is interesting because it provides evidence that a kriging approach is overall superior to its competitors. However, it should be pointed out that apart from the GK versus LK comparison, the paper largely leaves open the question of what form of kriging method is best. Thus, more refined questions concerning such aspects as the choice of spatial covariance function (including the possibility of a nonstationary covariance) or parameter estimation technique are still unresolved at the present time. Also, LK has an unresolved theoretical issue, that it may not be derived from an underlying positive-definite covariance structure for the whole process. Finally, the paper did not provide any validation for the various uncertainty measures.

Is Kriging an Exact Interpolator?

Kriging was developed by Matheron with the idea that data are observed precisely and one only needs to interpolate between these observations. In reality, the data are observed with error and the scientific goal should be to predict the true process, absent this measurement error. Consequently, kriging done properly should *not* generally be an exact interpolator. The role of kriging should be to filter out the measurement error and to make inference on the true (hidden) spatial process, including at locations where

there *are* data. This was not well understood by Matheron and Gandin, and it is not well understood to this day. For example, indicator kriging should make inference on the hidden process, not on the process that includes measurement error. This would lead to intractable indicator-kriging equations, and for this and other reasons leads one to conclude that indicator kriging as currently practiced is not a statistically meaningful interpolation method. The presence of measurement error requires extra knowledge as to its magnitude, or equivalently as to its percentage of the total nugget effect. Matheron's kriging equations assume 0%, and the default value in ESRI's Geostatistical Analyst is 0% (ESRI 2001, p. 169). That software does allow the user to input a measurement-error percentage or it will calculate the percentage from multiple observations at identical locations. Many public-domain and commercial kriging packages do not handle measurement-error properly and/or fail when there are multiple observations at individual locations.

There are kriging equations that have been derived in the presence of measurement error (Cressie 1988), but a more general approach is through the hierarchical statistical model (HM). Under this scheme, the data and the true process are considered random (as in geostatistics), but their joint distribution is modeled through a sequence of conditional probabilities, captured by the Data model and the Process model. In the case where both the Data and Process models are Gaussian, this allows for a quick derivation of kriging equations in the measurement error case without Monte Carlo simulation (Holland et al. 2000). The general formulation of HM, however, allows distributions to be anything (e.g., non Gaussian) and predictors to be nonlinear; see Diggle, Tawn, and Moyeed (1998). There is an extra component, the Parameter model, that turns the HM into a Bayesian HM or BHM, but it is not necessary to add this extra component. Under the HM approach, optimal statistical interpolators do not have to be linear, so they generalize statistical interpolation beyond ordinary kriging and universal kriging. Furthermore, the problem of predicting an indicator function of the true process can now be formulated and solved in an optimal, statistically coherent manner. References for the HM approach are Banerjee, Carlin, and Gelfand (2004) and Cressie and Wikle (2011, Ch. 4).

Alternative Methods of Spatial and Spatio-temporal Statistics

This section briefly touches on a number of modern topics in spatial statistics that have potential relevance for the problem of interpolating meteorological data.

Nonstationary spatial covariances. Originally, kriging methods were developed for mining applications, where by the nature of the problem, only one sample of data was available. Consequently, in order to estimate covariances, simplifying assumptions had to be made, the most popular being stationarity of either the covariance or the variogram function. In meteorology, there is in effect a new random field observed every day, that provide the *replications* needed to estimate a general spatial covariance matrix. In practice, stationarity is often assumed, either for computational convenience, or because of a generic belief that such an assumption is reasonable. Nevertheless, alternative methods are available. One possibility that has been mentioned already is to assume a locally varying variance (or sill) while still assuming that the underlying correlation function is stationary. More refined models, in which both the correlation and variance functions vary locally, have been proposed by various authors (e.g., Sampson and Guttorp 1992, Higdon et al. 1998, Fuentes 2002, Paciorek and Schervish 2006). The spatial random effects model of Cressie and Johannesen (2008) is both nonstationary and computationally advantageous when datasets are large. As methods for meteorological interpolation become more refined, we would recommend that nonstationary models be considered as alternatives to standard kriging.

Alternative estimation techniques. Estimating spatial models by variogram techniques has a number of ad hoc features, e.g. the distance intervals for which the variogram is computed. In the case of Gaussian processes, it is generally recognized that likelihood-based methods are superior, either traditional maximum likelihood or the refinement known as restricted maximum likelihood (see Cressie 1993 or Stein 1999). Another issue related to parameter estimation is that the traditional kriging formula for uncertainty ignores the uncertainty associated with the parameter estimation; this can be corrected either through asymptotic approximations (e.g. Zimmerman and Cressie 1992) or through Bayesian methods. Bayesian techniques have in any case become very popular because of the rich possibility of incorporating spatial statistics into hierarchical models (Banerjee, Carlin and Gelfand 2004).

Improving computation by dimension reduction: fixed rank spatial models. Kriging does not require stationary covariance functions; the kriging equations rely on knowing the covariance between any two spatial locations, regardless of whether that function depends only on the relative displacement of the two locations. However, kriging does require the (possibly nonstationary) covariance functions to be positive-definite. The idea behind fixed-rank models is to express the covariance function in terms of a fixed-rank positive-definite matrix, pre-multiplied by a row vector of spatial basis functions evaluated at the first spatial location and post-multiplied by a column vector of the same spatial basis functions evaluated at the second spatial location. The resulting nonstationary spatial model is positive-definite; Cressie and Johannesson (2008) call this the Spatial Random Effects (SRE) model. The other main advantage of the SRE model is that it leads to highly efficient computations. The inverse of the covariance matrix of the data

is typically the computational bottleneck in kriging. Cressie and Johannesson (2008) show that for the SRE model the computations are scalable in sample size, and they are able to produce kriging maps from hundreds of thousands of (global remote-sensing) data in a matter of seconds of CPU time; they call the method Fixed Rank Kriging. The SRE model can be embedded in a BHM, and that leads to enormous speed-ups that make massive data within the realm of Bayesian posterior analysis. For more details on fixed rank representations, see Wikle (2010).

Spatio-temporal methods. So far as we can tell, current methods of interpolating meteorological data largely ignore the temporal component of variation, in effect treating the temperatures (or their anomalies from a monthly mean) as independent for each day. Covariance models that include both the spatial and temporal components have been widely studied (e.g., Christakos 1992, Cressie and Huang 1999, Gneiting 2002, Stein 2005) but have had limited application to large datasets, partly for computational reasons. Nevertheless, methods are being developed that incorporate both spatial and temporal dependence and improve greatly on the computational limitations of earlier methods. For example, Cressie, Shi and Kang (2010) process 120,000 datapoints in a spatio-temporal Kalman filter with computation time on the order of 2 minutes.

Summary and Conclusions

Methods of interpolation of meteorological data have advanced considerably in recent years. Statistical methods based on kriging have been extensively applied, and in at least one major study, proved superior to most alternatives. New mathematical techniques based on dimension reduction have the potential for much improvement, including nonstationary models and rapid computing in large datasets. We also see a number of research questions for future investigation. Methods of assessing uncertainty are still rather crude and need to be assessed more rigorously. Incorporation of the temporal component in spatial analysis is an area undergoing considerable activity. Statistical and computational efficiencies are needed both for kriging and for Bayesian spatial interpolations.

Technical Appendix

Initially we review basic definitions. Some references where these are explained in more detail include Cressie (1993), Stein (1999) and Smith (2001).

In general, capital letters such as X,Y,Z denote random variables, on their own or as part of a random field, e.g. $Z(\mathbf{s})$ may indicate the value of a spatial field Z at a spatial location $\mathbf{s}$, where $\mathbf{s}$ is a vector in 2- or higher-dimensional space; in a spatio-temporal setting, we may write $Z(\mathbf{s},t)$ to denote the value of the field at location $\mathbf{s}$ at time t. The letter E generally denotes expectation or mean value.

The *variance* of a random variable X, often written σ^2_X, is defined as $E\{(X-E(X))^2\}$. The *covariance* of two random variables X,Y, often written σ_{XY}, is defined as $E\{(X-E(X))(Y-E(Y))\}$.

If $Z(\mathbf{s})$ is a spatial random field, its *covariance function* is $C(\mathbf{s}_1,\mathbf{s}_2)$, the covariance of Z at two spatial locations $\mathbf{s}_1$ and $\mathbf{s}_2$. If $C(\mathbf{s}_1,\mathbf{s}_2)$ depends on $\mathbf{s}_1$ and $\mathbf{s}_2$ only through their vector difference, $\mathbf{s}_1-\mathbf{s}_2$, and if, in addition, the mean is constant, then the process is defined to be *second-order stationary* (often abbreviated to just *stationary*). If the covariance depends only on the scalar distance between $\mathbf{s}_1$ and $\mathbf{s}_2$, so $C(\mathbf{s}_1,\mathbf{s}_2)=C_0(||\mathbf{s}_1-\mathbf{s}_2||)$ for some scalar function C_0, the process is called *isotropic*. Sometimes, a process that is both stationary and isotropic is called *homogeneous*.

An alternative formulation is to use the *dispersion function* $D(\mathbf{s}_1,\mathbf{s}_2)=E\{(Z(\mathbf{s}_1)-Z(\mathbf{s}_2))^2\}$. If $D(\mathbf{s}_1,\mathbf{s}_2)=D_0(\mathbf{s}_1-\mathbf{s}_2)$ for some function D_0, and if the mean is constant, the process is called intrinsically stationary. If, further, $D_0(\mathbf{s}_1-\mathbf{s}_2)=2\gamma(||\mathbf{s}_1-\mathbf{s}_2||)$ for some scalar function γ, the process is again called isotropic and γ is called the *semivariogram* function (or just the *variogram*; the prefix "semi" is increasingly ignored).

It's possible for a process to be intrinsically stationary without being second-order stationary; for example, the *linear variogram,* $\gamma(h)=a+bh$, does not correspond to any second-order stationary process.

In practice, most variograms tend to an asymptote (known as the *sill*) as distance h tends to infinity; the *range* corresponds to the distance required for the variogram to reach a certain fraction (e.g. 100% or 95%) of the sill; the *nugget* (when present) represents a discontinuity at the origin; this leads to the characteristic shape depicted earlier. Examples of variograms of this shape include the *spherical* ($\gamma(h)=\alpha+\beta(1.5(t/R)-0.5(t/R)^3)$ for $0<t<R$, $\gamma(h)=\alpha+\beta$ for $t>R$), *exponential* ($\gamma(h)=\alpha+\beta(1-\exp(-t/R))$) and *Gaussian* ($\gamma(h)=\alpha+\beta(1-\exp(-t^2/R^2))$); in each case the range is some multiple of R, α is the nugget and $\alpha+\beta$ the sill. The *Matérn* model is especially widely used because of its mathematical flexibility (Stein 1999); it is usually expressed in terms of its covariance function, given as $C_0(h)=(2h\sqrt{v}/R)^v\ K_v(2h\sqrt{v}/R)/2^{v-1}\Gamma(v)$. Here $\Gamma(v)$ denotes the gamma function while K_v is the modified Bessel function of the third kind of order v; here v is an adjustable shape parameter. Note that as given here, there is no separate nugget parameter, though it is easy also to include a nugget in the Matérn covariance function.

In practice, all of these models must be estimated from observational data. The most widely used

estimator of the variogram is the *method of moments* (Journel and Huijbregts 1978) – pairs of spatial locations ($\mathbf{s}_1$,$\mathbf{s}_2$) are binned according to the distance $||\mathbf{s}_1-\mathbf{s}_2||$, then the values of $(Z(\mathbf{s}_1)-Z(\mathbf{s}_2))^2$ are averaged within each bin. The *robust variogram estimator* (Cressie and Hawkins 1980) uses an intermediate fourth-root transformation and is less sensitive to outliers.

To estimate the parameters of a variogram model, Journel and Huijbregts (1978) recommended using nonlinear least squares regression to the estimated variogram function. An improvement is the approximate weighted least squares method of Cressie (1985): if $\gamma(h)$ is estimated by $\gamma\hat{}(h)$ for a series of distances $h_1,\ldots,h_m$ (typically corresponding to a binned estimator), and if $\gamma(h)$ is represented by a parametric function $\gamma(h;\theta)$ say, where θ is the vector of unknown parameters, then the estimator is derived by minimizing the weighted least squares function, $\sum N_j(\gamma\hat{}(h_j)/\gamma(h_j;\theta)-1)^2$, where N_j is the number of pairs of locations in the j^{th} bin centered on h_j.

The main alternative approach to estimation is to assume the process is Gaussian and work directly with the likelihood function. Suppose we have a vector of observations $\mathbf{Z}$ consisting of $Z(\mathbf{s}_i)$, $i=1,\ldots,n$, where n is the number of spatial locations. Suppose $\mathbf{Z}$ has mean $\mathbf{X}\beta$ ($\mathbf{X}$ a known matrix of covariates, β an unknown vector of linear regression coefficients) and covariance matrix $\Sigma(\theta)$. This formulation corresponds to the slight generalization of ordinary kriging known as *universal kriging*, whereby an assumed constant unknown mean μ is replaced by a vector of unknown regression coefficients β. Note, however, that ordinary kriging may be recovered by replacing $\mathbf{X}\beta$ with $\mathbf{1}\mu$ ($\mathbf{1}$ a vector of n ones, μ a scalar).

For a given set of values of the covariance parameters θ, we define the generalized least squares regression estimator, $\beta\hat{}=(\mathbf{X}^T\Sigma(\theta)^{-1}\mathbf{X})^{-1}(\mathbf{X}^T\Sigma(\theta)^{-1}\mathbf{Z})$, and the generalized residual sum of squares $G^2(\theta)=(\mathbf{Z}-\mathbf{X}\beta\hat{})^T\Sigma(\theta)^{-1}(\mathbf{Z}-\mathbf{X}\beta\hat{})$. The maximum likelihood estimator (MLE) chooses θ to minimize the negative log likelihood function $\frac{1}{2}[n\log(2\pi)+G^2(\theta)+\log|\Sigma(\theta)|]$. Alternatively, the restricted maximum likelihood estimator (REMLE) minimizes $\frac{1}{2}[(n-q)\log(2\pi)+G^2(\theta)+\log|\Sigma(\theta)|+\log|\mathbf{X}^T\Sigma(\theta)^{-1}\mathbf{X}|-\log|\mathbf{X}^T\mathbf{X}|]$, where q is the number of linearly independent regressors. These estimation procedures have been implemented in a number of statistical computing packages, for instance *fields* and *geoR* which are both downloadable packages for the R program (R Core Development Team 2010). REML estimation is usually considered superior to the usual maximum likelihood procedure; for instance, the REMLE is approximately unbiased whereas the MLE often exhibits considerable bias. A third possibility is to implement a fully Bayesian procedure, which is computationally feasible thanks to modern methods of Monte Carlo simulation, such as the Metropolis-Hastings algorithm (Gelman et al. 2003, Banerjee et al. 2004).

Next we turn to kriging, which is more precisely described as optimal unbiased linear prediction. Suppose, as above, $\mathbf{Z}$ is a vector of observations from a spatial process with mean $\mathbf{X}\beta$ and covariance matrix Σ. Although Σ will still (in most cases) depend on some unknown parameter θ, in the traditional formulation of kriging, θ is treated as known, so we do not indicate it explicitly. Suppose we want to predict a scalar z_0 which has mean $x_0^T\beta$ (x_0 known, β the same as previously) and variance σ_0^2. We also assume the vector of covariances between $\mathbf{Z}$ and z_0 is written as ρ. A few comments about this formulation:

1. Usually z_0 is the value of the random field at a single unobserved location, but it may also be the average (i.e. spatial mean) of the random field over some region. Usually, predicting such a spatial mean in a single step is simpler (especially for the MSPE calculation) than calculating the kriging predictor point by point and then averaging.

2. Ordinary kriging, where both the observed and predicted quantities have a single unknown mean μ, is a special case of this: simply replace $\mathbf{X}$ with $\mathbf{1}$.

3. There is no explicit assumption of a Gaussian process here, though this is often assumed for convenience.

The kriging problem is to find a predictor of the form, $z_0\hat{}=\lambda^T\mathbf{Z}$, that minimizes the mean squared prediction error (MSPE), given by $E\{(z_0\hat{}-z_0)^2\}$, under the unbiasedness constraint $\lambda^T\mathbf{X}=x_0^T$. The solution is

$$\lambda=\Sigma^{-1}\rho+\Sigma^{-1}\mathbf{X}(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}(x_0-\mathbf{X}^T\Sigma^{-1}\rho)$$

and the resulting mean squared prediction error is

$$\text{MSPE}=\sigma_0^2-\rho^T\Sigma^{-1}\rho+(x_0-\mathbf{X}^T\Sigma^{-1}\rho)^T(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}(x_0-\mathbf{X}^T\Sigma^{-1}\rho).$$

Two further comments about these formulas:

4. Nowhere in our discussion of maximum likelihood estimation and kriging did we assume that the covariance matrix Σ arises from a stationary (or isotropic) process. That is an assumption often made for convenience, because of the existence of a number of variogram or covariance models that happen to be stationary and isotropic, but it is in no way required for kriging.

5. The formula we have given for MSPE is the correct formula if the model is correctly specified. However, like all mathematical formulas, it is valid only if the assumptions that have been made (in this case, on the means and covariances of the process) are correct. In cases where it has been found that the formula does not work well in practice, this is evidence of model misspecification; our advice would be to look for alternative models that better describe the data, rather than make ad hoc adjustments to the formulas, as in

Yamamoto (2000).

Spatial Random Effects (SRE) Model. A very general model for a nonstationary spatial model is of the form, $\mathbf{Z}=\mathbf{Y}+\boldsymbol{\varepsilon}$, where $\mathbf{Z}$ is a vector of observations and $\boldsymbol{\varepsilon}$ is a vector of measurement errors with mean 0 and covariance matrix $\sigma_0^2\,\mathbf{V}$ ($\mathbf{V}$ diagonal). Here $\mathbf{Y}$ is a sample of n locations from a process model of form

$$Y(\mathbf{s}) = \mathbf{T}(\mathbf{s})^{\mathsf{T}}\,\alpha + \mathbf{U}(\mathbf{s})^{\mathsf{T}}\,\eta + \boldsymbol{\xi}(\mathbf{s})$$

where $\mathbf{T}(\mathbf{s})^{\mathsf{T}}\,\alpha$ are fixed effects ($\mathbf{T}(\mathbf{s})$ a known vector of covariates, α an unknown parameter vector), $\mathbf{U}(\mathbf{s})^{\mathsf{T}}\,\eta$ are random effects ($\mathbf{U}(\mathbf{s})$ a known r-dimensional vector of spatial basis functions for each $\mathbf{s}$, η an r-dimensional random vector with mean 0 and covariance matrix $\mathbf{K}_{rxr}$) and $\boldsymbol{\xi}(\mathbf{s})$ is a white-noise process (mean 0, uncorrelated, common variance τ^2) that represents micro-spatial variation. Here $\mathbf{U}(\mathbf{s})=(U_1(\mathbf{s}),\ldots,U_r(\mathbf{s}))^{\mathsf{T}}$ represent *r basis functions* that could arise, for instance, from a Fourier or wavelet representation, though there is no requirement that the basis functions be orthogonal.

For such a process, we have the covariance function,

$$C(\mathbf{s}_1,\mathbf{s}_2) = \mathbf{U}(\mathbf{s}_1)^{\mathsf{T}}\mathbf{K}\,\mathbf{U}(\mathbf{s}_2) + \tau^2\, I(\mathbf{s}_1=\mathbf{s}_2),$$

which is non-negative definite and nonstationary; the parameters to be estimated are $\mathbf{K}$ and τ^2.

Spatial Fixed Rank Kriging (FRK). The covariance matrix of $\mathbf{Z}$, which we write as $\boldsymbol{\Sigma}$, is of the form $\mathbf{UKU}^{\mathsf{T}}+\mathbf{D}$, where $\mathbf{U}$ is the matrix formed by the vectors $\mathbf{U}(\mathbf{s}_1),\ldots,\mathbf{U}(\mathbf{s}_n)$ and $\mathbf{D}=\tau^2\mathbf{I}_n+\sigma_0^2\mathbf{V}$ ($\mathbf{V}$ is a diagonal matrix and $\mathbf{I}_n$ is the nxn identity matrix). The main computational difficulty in kriging is the calculation of the inverse matrix $\boldsymbol{\Sigma}^{-1}$ when n is large. The problem may, however, be much simplified by the application of the *Sherman-Morrison-Woodbury* identity: for any nxr matrix $\mathbf{P}$, this gives

$$(\mathbf{I}_n + \mathbf{PKP}^{\mathsf{T}})^{-1} = \mathbf{I}_n - \mathbf{P}(\mathbf{K}+\mathbf{P}^{\mathsf{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathsf{T}}.$$

Hence

$$\boldsymbol{\Sigma}^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{U}\,(\mathbf{K}^{-1} + \mathbf{U}^{\mathsf{T}}\mathbf{D}^{-1}\mathbf{U})^{-1}\,\mathbf{U}^{\mathsf{T}}\mathbf{D}^{-1},$$

which involves inversion of the fixed rank matrix $\mathbf{K}_{rxr}$ and the diagonal matrix $\mathbf{D}_{nxn}$.

Substituting this into the kriging equations yields the kriging map $\{\hat{Y}(\mathbf{s}_0) : \mathbf{s}_0 \in D\}$ and the associated kriging standard error map $\{\sigma_k(\mathbf{s}_0) : \mathbf{s}_0 \in D\}$, where D is the domain of interest (Cressie and Johannesson 2008).

The covariance parameters can be estimated using a binned method-of-moments and then minimizing the Frobenius norm between the empirical (binned method-of-moments) covariance and the theoretical covariance implied by the SRE model (Cressie and Johannesson 2008).

References

Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2004). *Hierarchial Modeling and Analysis for Spatial Data.* Monographs on Statistics and Applied Probability 101. Chapman & Hall/CRC, Boca Raton.

Brockwell, P.J. and Davis, R.A. (1991), *Time Series: Theory and Methods*. Second Edition. Springer Verlag, New York.

Christakos, G. (1992), *Random Field Models in Earth Sciences*. Academic Press, San Diego.

Craven, P. and Wahba, G. (1979), Smoothing noisy data with spline functions. *Numerische Mathematik* **31**, 377-403.

Cressie, N. (1985), Fitting variogram models by weighted least squares. *Mathematical Geology* **17**, 563-586.

Cressie, N. (1988). Spatial prediction and ordinary kriging. *Mathematical Geology*, **20**, 405-421. [Erratum, *Mathematical Geology*, **21**, 493-494 (1989).]

Cressie, N. (1993), *Statistics for Spatial Data*. Second edition, John Wiley, New York.

Cressie, N. and Hawkins, D.M. (1980), Robust estimation of the variogram I. *Mathematical Geology* **12**, 115-125.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B*, **70**, 209-226.

Cressie, N. and Huang, H.-C. (1999), Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**, 1330-1340.

Cressie, N., Shi and Kang (2010), *Journal of Computational and Graphical Statistics*, to appear.

Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-temporal Data*. John Wiley and Sons, Inc., New York.

Diggle, P.J., Tawn, J.A., and Moyeed, R.A. (1998). Model-based geostatistics. *Journal of the Royal*

412 *Statistical Society, Series C*, **47**, 299-350.
413 ESRI (2001). *Using ArcGIS Geostatistical Analyst*. GIS by ESRI, Redlands, CA.
414 Fuentes, M. (2002), Periodogram and other spectral methods for nonstationary spatial processes.
415 *Biometrika*, **89**, 197-210.
416 Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003), *Bayesian Data Analysis*. Second
417 Edition. CRC Press, London.
418 Gneiting, T. (2002), Nonseparable, stationary covariance functions for space-time data. *Journal of*
419 *the American Statistical Association* **97**, 590-600.
420 Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. and New, M. (2008), A
421 European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006.
422 *Journal of Geophysical Research*, **113**, D20119, doi:10.1029/2008JD010201.
423 Higdon, D., Swall, J. and Kern, J. (1999), Non-stationary spatial modeling. In *Bayesian Statistics 6*,
424 eds. J.M. Bernardo et al., Oxford University Press, pp. 761--768.
425 Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008), Comparison of six methods for the
426 interpolation of daily, European climate data. *Journal of Geophysical Research*, **113**, D21110,
427 doi:10.1029/2008JD010100.
428 Holland, D.M., De Oliveira, V., Cox. L.H. and Smith, R.L. (2000), Estimation of regional trends in
429 sulfur dioxide over the eastern United States. *Environmetrics* **11**, 373-393.
430 Hutchinson, M.F. (1995), Interpolating mean rainfall using thin plate smoothing splines. *International*
431 *Journal of Geographic Information Science* **9**:4, 385-403.
432 Journel, A.G. and Huijbregts, C.J. (1978), *Mining Geostatistics*. Academic Press, London.
433 Paciorek, C.J. and Schervish, M.J. (2006), Spatial modeling using a new class of nonstationary
434 covariance functions. *Environmetrics* **17**, 483-506.
435 R Core Development Team (2010), *R: A Language and Environment for Statistical Computing*. R
436 Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. http://www.R-project.org.
437 Sampson, P.D. and Guttorp, P. (1992), Nonparametric estimation of nonstationary spatial
438 covariance structure. *Journal of the American Statistical Association* **87**, 108-119.
439 Sansó, B. and Guenni, L. (2004), A Bayesian approach to compare observed rainfall data to
440 deterministic simulations. *Environmetrics* **15**, 597-612.
441 Smith, R.L. (2001), *Environmental Statistics*. Unpublished course notes, available from
442 http://www.stat.unc.edu/postscript/rs/envnotes.pdf
443 Stein, M.L.(1999), *Interpolation of Spatial Data: Some Theory of Kriging.* Springer Verlag, New York.
444 Stein, M. (2005), Space-time covariance functions. *Journal of the American Statistical Association*
445 **100**, 310-321.
446 Wikle, C.K. (2010). Low rank representations as models for spatial processes. Chapter 8 in
447 *Handbook of Spatial Statistics*, A.E. Gelfand, M. Fuentes, P. Guttorp, and P. Diggle (eds). Chapman &
448 Hall/CRC, Boca Raton.
449 Wikle, C.K. and Cressie, N. (1999), A dimension-reduced approach to space-time Kalman filtering.
450 *Biometrika* **86**, 815-829.
451 Yamamoto, J.K. (2000), An alternative measure of the reliability of ordinary kriging estimates.
452 *Mathematical Geology* **32**, 489-509.
453 Zimmerman, D.L. and Cressie, N. (1992), Mean squared prediction error in the spatial linear model
454 with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics* **44**, 27-43.
455