now P821

KRIGE, SMOOTH, BOTH, OR NEITHER?

BU-1264-MC **January, 2000**

Naomi Altman
Department of Biometrics
Cornell University
Ithaca, NY 14853

Keywords: *nonparametric regression, smooth spline, generalized cross-validation, interpolation, spatial statistics prediction, nonstationary processes.*

Abstract

Both kriging and nonparametric regression smoothing can model a nonstationary regression function with spatially correlated errors. However comparisons have mainly been based on ordinary kriging and smoothing with uncorrelated errors. Ordinary kriging attributes smoothness of the response to spatial autocorrelation whereas nonparametric regression attributes trends to a smooth regression function. For spatial processes it is reasonable to suppose that the response is due to both trend and autocorrelation.

This article reviews methodology for nonparametric regression with autocorrelated errors which is a natural compromise between the two methods. Re-analysis of the one-dimensional stationary spatial data of Laslett (1994) and a clearly nonstationary time series demonstrates the rather surprising result that for these data, ordinary kriging outperforms more computationally intensive models including both universal kriging and correlated splines for spatial prediction. For estimating the regression function, nonparametric regression provides adaptive estimation, but the autocorrelation must be accounted for in selecting the smoothing parameter.

Krige, Smooth, Both or Neither?

Naomi Altman *
Biometrics Unit
Cornell University

January 12, 2000

Abstract

Both kriging and nonparametric regression smoothing can model a nonstationary regression function with spatially correlated errors. However comparisons have mainly been based on ordinary kriging and smoothing with uncorrelated errors. Ordinary kriging attributes smoothness of the response to spatial autocorrelation whereas nonparametric regression attributes trends to a smooth regression function. For spatial processes it is reasonable to suppose that the response is due to both trend and autocorrelation.

This article reviews methodology for nonparametric regression with autocorrelated errors which is a natural compromise between the two methods. Re-analysis of the one-dimensional stationary spatial data of Laslett (1994) and a clearly nonstationary time series demonstrates the rather surprising result that for these data, ordinary kriging outperforms more computationally intensive models including both universal kriging and correlated splines for spatial prediction. For estimating the regression function, nonparametric regression provides adaptive estimation, but the autocorrelation must be accounted for in selecting the smoothing parameter.

AMS subject classification: Primary 62 G05 Secondary 62 G20

*Supported by Hatch Grant 151410 NYF and NSF grant DMS-95-25350. Discussions with Noel Cressie, James Hobert and Jean Opsomer contributed substantially to the the author's understanding of the spline and kriging models.

Keywords and phrases: nonparametric regression; smoothing spline; generalized cross-validation; interpolation; spatial statistics; prediction; nonstationary processes.

1 Introduction

Recently, a number of papers have compared nonparametric regression and kriging as approaches to spatial modeling (Cressie, 1989; Hutchinson and Gessler, 1994; Laslett, 1994; Voltz and Webster, 1990; Wahba, 1990; Yakowitz and Szidarovsky, 1985) with varying conclusions about the merits of each. Both methods are based on the regression model

$$z(s_j) = \mu(s_j) + \epsilon(s_j) \tag{1}$$

where z is a continuous response variable observed at n spatial locations $s_1 \cdots s_n$, μ is the regression function, and $\epsilon(s_j)$ is a second order stationary error term with mean zero, variance σ^2 and covariance matrix Σ. Nonparametric regression (smoothing) models allow the regression function μ to be in a large smooth nonparametric class, such as differentiable functions. Although the errors may be correlated (Altman, 1990, 1992; Diggle and Hutchinson, 1989; Engel, Granger, Rice and Weiss, 1986; Hart, 1991; Opsomer, 1996; Wang, 1998) the usual implementations assume uncorrelated errors with constant variance (Eubank, 1988; Wand and Jones, 1995; Wahba, 1990). Kriging models assume that the regression function is a polynomial (or other orthogonal series) and the errors are second order intrinsically stationary with a parametric correlation structure depending on distance (Cressie, 1993).

The effectiveness of the methods for fitting simulated and real data sets has been compared. There are two notions of "effectiveness" that are relevant. For traditional geostatistical applications, investigators are interested in functions of the current realization of the process, such as $Z(s^*)$ where s^* is an unobserved location. Following current terminology, we will call this the prediction problem. The best predictor of $Z(s_j)$ is the observed value $z(s_j)$, with no prediction error, and, with μ known, the best (in terms of mean squared prediction error) predictor of

$Z(s^*)$ is

$$Z(s^*) = \mu(s^*) + E(\epsilon(s^*)|\epsilon(s_1)\cdots\epsilon(s_n)) \tag{2}$$

with prediction variance $Var(\epsilon(s^*)|\epsilon(s_1)\cdots\epsilon(s_n))$. For most of the data sets analyzed in this paper, ordinary kriging outperforms correlated splines, even when the mean of the spatial series appears to be nonstationary.

For applications such as weather prediction, the investigator may be interested in functions of a new realization of the process, such as $Z^*(s^*)$ where Z^* is independent of the observed z's and s^* may be an observed or new location. The best predictor of $Z^*(s^*)$ is $\mu(s^*)$ with prediction variance $\sigma^2 \geq Var(\epsilon(s^*)|\epsilon(s_1)\cdots\epsilon(s_n))$. We will refer to this problem as the regression problem, but the goal is to produce predictions and prediction bands, rather than regression functions and confidence intervals. Of course, both μ and the error process must be estimated from the sample, and this should be taken into account when estimating prediction variance. The regression function is not explored in detail in this paper. However, if a polynomial trend is fitted, as in universal kriging, the regression estimator is centered around the polynomial "closest" (in the sense of weighted least squares) to the underlying trend while nonparametric regression is asymptotically centered on the true trend.

The purpose of this paper is 3-fold:

1. to review existing methodology and theory for fitting smoothing models with correlated errors and explain the erratic behavior of the smoothing spline (or more correctly, of generalized cross-validation) in Laslett's (1994) study

2. to discuss the difference between prediction and regression as defined above

3. to compare universal kriging and smoothing with correlated errors for spatial prediction.

Section 2 of this article summarizes the computational methodology for fitting model (1) by universal kriging and correlated smoothing splines. Section 3 is a summary of results for nonparametric regression with time series errors which apply to the regular one-dimensional spatial series of Laslett. The results extend to higher dimensional spatial samples with arbitrary sampling distribution. Section 4 is an analysis of Laslett's data sets and a clearly nonstationary

time series. The regression problem is discussed in Section 5. Section 6 is general comparison of the two approaches to fitting model (1).

2 Fitting the Model for Prediction

In this Section we outline methodology for fitting (2). Estimation of the spatial correlation function

$$\rho(d) = Corr\left(\epsilon(s), \epsilon(s')\right) \qquad |s - s'| = d$$

or alternatively, the variogram

$$\Gamma(d) = E\left((\epsilon(s) - \epsilon(s')\right)^2 \qquad |s - s'| = d.$$

is an essential step. If the errors ϵ of the observed values and the correlation function or variogram were known, then the best linear unbiased estimator (in the sense of mean squared error) of the error at location s^* would be

$$\begin{aligned} \hat{E}(\epsilon(s^*)|\epsilon(s_1)\cdots\epsilon(s_n)) &= R_{12}R_{11}^{-1}\epsilon \\ &= G_{12}G_{11}^{-1}\epsilon \end{aligned} \tag{3}$$

where R_{11} is the matrix of correlations among the observations, R_{12} is the matrix of correlations between the observations and the unobserved value at location s^*, G_{11} is the matrix with entries $\Gamma(s_i - s_j)$ and G_{12} is the matrix with entries $\Gamma(s_i - s^*)$. Generally, $\rho(d)$ or equivalently $\Gamma(d)$ is assumed to have a parametric form which is estimated from the data. Inferences are then made conditionally on the estimated parameter values.

When $\rho(d)$ (or $\Gamma(d)$) has a discontinuity at zero, the variance or variogram is said to have a nugget. When there is no nugget, (3) interpolates the errors. When there is a nugget, (3) smooths the errors.

Under the ordinary kriging model, μ is estimated by the sample mean. Under the universal kriging model, μ is estimated by a low order polynomial. In either case, the parameters of $\rho(d)$ or $\Gamma(d)$ may then be estimated by restricted maximum likelihood (REML), following which the

regression is fitted by weighted least squares, to obtain $\hat{\mu}$. Details are in Searle, Casella and McCulloch, (1992, Chap. 6) and Cressie (1993, Chap 2).

Under the correlated smoothing model, μ is estimated by a nonparametric regression estimator. Laslett (1994) used a cubic smoothing spline with generalized cross-validation (GCV) for smoothing parameter estimation. The smoothing spline estimator is defined as the minimizer of

$$\sum_{i=1}^{n}(z(s_i) - \mu(s_i))^2 + \lambda \int \mu''(s)^2 ds. \tag{4}$$

where λ is a smoothing parameter estimated from the data. Details are in Wahba (1990) and Eubank (1988, Chap. 5).

Appropriate choice of smoothing parameter is critical to nonparametric regression. Selection methods for correlated splines are discussed in Diggle (1989) and Wang (1998); for kernel regression in (Altman, 1990; Chiu, 1989; Hart, 1991); and for local polynomials in (Opsomer, 1996). Bandwidth choice for kernel regression is discussed in the next section. In order to understand the poor behavior of GCV in Laslett's study, illustrated here in Figure 1a, we shall need to consider correlation-adjusted GCV as in Altman (1990).

A linear mixed models approach to joint estimation of the spline and error correlation is described in Wang (1998) and is simple to use in application. In that approach, the smoothing parameter is a ratio of variance components, and GCV is not required. Joint estimation can be handled simply using off-the-shelf software for linear mixed models.

Although the degree of the polynomial (universal kriging), the smoothing parameter (smoothing spline regression) and the parameters of the correlation function are generally estimated from the data (or determined subjectively) prediction intervals are generally computed as if these were known a priori. For all three estimators, once these parameters have been estimated, $\hat{\mu}$ is linear in the data: that is it has the form $\hat{\mu}(s^*) = F(s)'z(\mathbf{s})$ where $F(s)$ is a vector. So the prediction variance for the regression problem is

$$\sigma^2 + F(s)'\Sigma F(s)$$

where Σ is the variance matrix of the observations. Note that $F(s)'\Sigma F(s)$ is typically $o(\frac{1}{n^d})$ where p¿0, so the major contribution to prediction error is the error variance.

The predictor of $Z(s)$ is

$$\hat{Z}(s^*) = \hat{\mu}(s^*) + \hat{E}(\epsilon(s^*)|\epsilon(s_1)\cdots\epsilon(s_n)). \tag{5}$$

This is also linear in the data, $\hat{Z}(s^*) = P(s)'z(\mathbf{s})$, so that the prediction variance for the prediction problem is

$$P(s)'\Sigma P(s). \tag{6}$$

3 Nonparametric Regression with Autocorrelated Errors

The theory for nonparametric regression with autocorrelated errors is most fully developed for kernel and local polynomial regression, but similar results should hold for smoothing splines. All 3 estimators have the linear form

$$\hat{\mu}(s) = F_\lambda(s)'z(s) \tag{7}$$

where the vector of weights F_λ depends on a smoothing parameter λ which controls the bias and variance of the estimator.

It can readily be shown for kernel and local polynomial regression (e.g. Altman, 1990; Opsomer, 1996) that the bias of the estimator does not depend on the correlation of the errors, and but that for equally spaced data the variance of the regression estimator asymptotically has the form

$$Var(\hat{\mu}(s)) = \frac{\sigma^2}{n\lambda} A\,(1 + 2C(\epsilon)) + o(\frac{1}{n\lambda})$$

where A depends on the form of the estimator, but not on the regression or correlation functions and $C(\epsilon) = \sum_{d=1}^{\infty} \rho(d)$. This shows that as long as $C(\epsilon) < \infty$ the estimator has the same rate of convergence. However, if goodness of fit is measured by mean integrated squared error

$$\begin{aligned} MISE(\lambda) &= E\left(\int [\hat{\mu}(s) - \mu(s)]^2 ds\right) \\ &= \int (Bias(\hat{\mu}(s))^2 ds + \int Var(\hat{\mu}(s)) ds \end{aligned} \tag{8}$$

then when $C(\epsilon) > 0$, which is generally the case for spatially correlated data, MISE is minimized by a larger bandwidth than in the uncorrelated case.

A number of data-adaptive techniques for selection of λ have been devised, many of which are summarized in Wand and Jones (1995). As Laslett (1994) used generalized cross-validation (GCV) to choose the smoothing parameter for the smoothing spline, the properties of this selector under model (1) will be discussed below. The erratic behavior of the smoothing spline observed by Laslett when using subsequences of the data was primarily due to the poor behavior of ordinary GCV for correlated data.

The GCV criterion of Craven and Wahba (1979) has the form:

$$GCV(\lambda) = \frac{\frac{1}{n}[\mathbf{z} - \hat{\mu}(\mathbf{s})]'[\mathbf{z} - \hat{\mu}(\mathbf{s})]}{[1 - \frac{1}{n}trF_\lambda(\mathbf{s})]^2} \tag{9}$$

where tr denotes matrix trace . Note that $\frac{1}{n}trF_\lambda(\mathbf{s}) = O(1)$. For a kernel regression estimator, the expectation of $GCV(\lambda)$ is

$$E[GCV(\lambda)] = \sigma^2(\epsilon) + MISE(\lambda) - \frac{4\sigma^2(\epsilon)}{n\lambda}\frac{1}{n}trF_\lambda(\mathbf{s})C(\epsilon) + o\left(\frac{1}{n\lambda}\right) + o\left(\lambda^p\right) \tag{10}$$

where MISE is $O(\frac{1}{n\lambda})+O(\lambda^p)$ and p depends on the kernel (equivalently, the degree of the spline) and the number of derivatives of μ (Altman, 1990). When $C(\epsilon) = 0$, the extra $O(1/n\lambda)$ term following $MISE(\lambda)$ disappears, so that asymptotically the expectation differs from $MISE(\lambda)$ only by a constant not depending on λ. From (10) we can see that when $C(\epsilon) > 0$, $GCV(\lambda)$ tends to pick smaller bandwidths than in the independent case, although our analysis of $MISE(\lambda)$ above shows that a larger bandwidth is needed. In fact, Chu and Marron (1991) and Hart (1991) show that a related criterion, cross-validation, selects arbitrarily small bandwidths with probability approaching 1, when the autocorrelation is greater than a threshold that depends on the kernel.

Chu and Marron (1991) show that if data at a distance greater than k are uncorrelated, and the subsequences using every k^{th} point are used for bandwidth selection with CV, the resulting "partitioned" CV estimator is consistent for the asymptotically optimal bandwidth for $C(\epsilon) = 0$, which is smaller than the required bandwidth. This result is important in understanding the behavior of the spline smoother in Laslett's analysis. Laslett used the subsequences of every k^{th} point, and noted that the selected smoothing parameters increased with k.

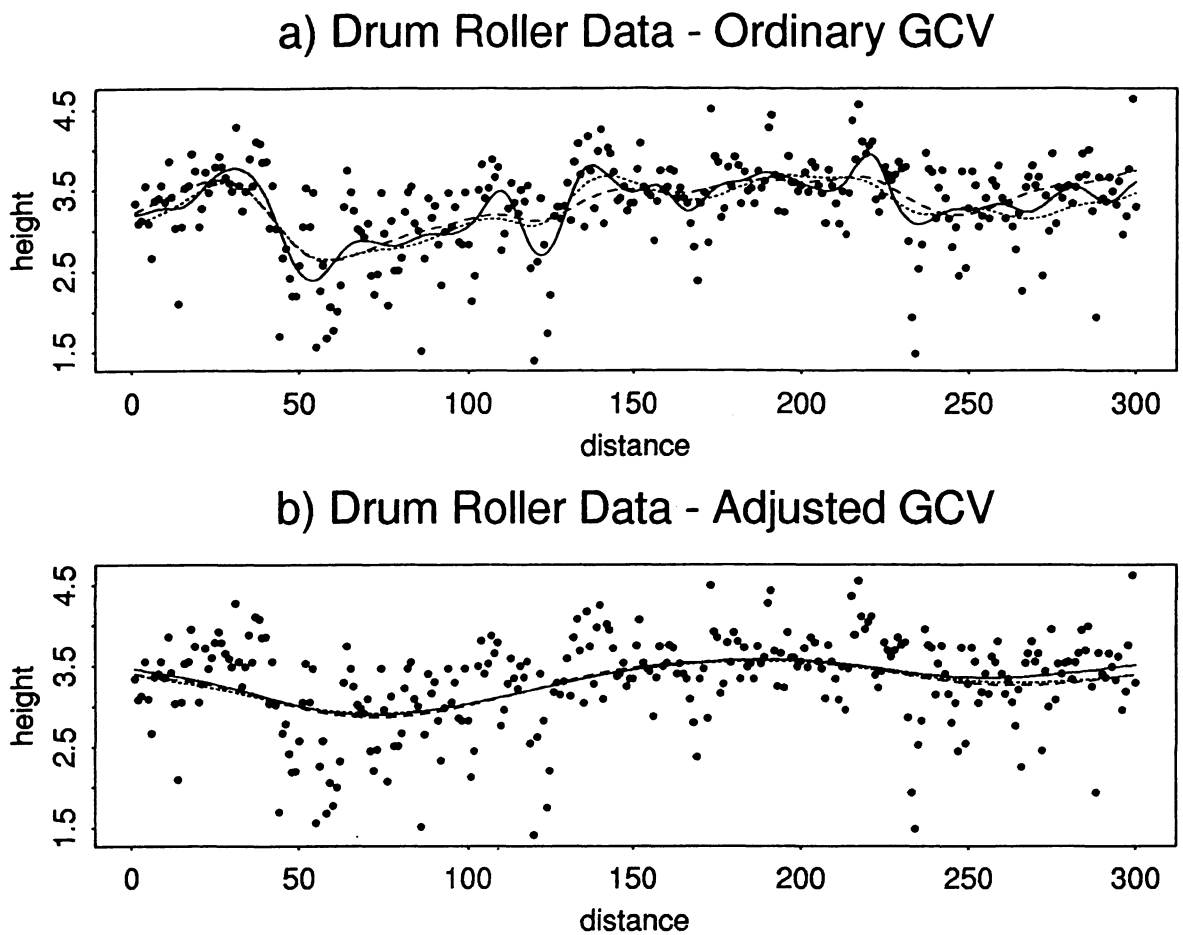

Figure 1: Smoothing spline estimator of the mean surface height for the drum roller data (first 300 data points). The solid line is the estimator based on the entire sequence. The dashed lines are the estimators based on the subsequences of every second point. a) Smoothing parameter selected by GCV without correlation adjustment. Notice that the bandwidth chosen for the full data set produces a very wiggly estimate, and that the estimates based on the subsequences are substantially smoother. b) Smoothing parameter selected by GCV adjusted for residual correlation. Notice that the estimate for the full sequence is very smooth and quite similar to the estimates based on the subsequences.

Figure 1a demonstrates Chu and Marron's effect for Laslett's drum roller data. The solid curve is the value of $\hat{\mu}$ when the smoothing parameter was estimated using GCV and the entire series. The two dashed curves are the values of $\hat{\mu}$ based on the subsequences $\{z(s_1), z(s_3), z(s_5)\cdots\}$ and $\{z(s_2), z(s_4), z(s_6)\cdots\}$. A substantially larger smoothing parameter was selected for each of the subsequences than for the entire sequence. Picking every third point leads to selection of yet a larger smoothing parameter. The reason for this is that as the distance between the points in the subsequence increases, the autocorrelation decreases and $C(\epsilon)$ approaches zero.

A number of methods have been suggested for adjusting GCV to select the bandwidth or smoothing parameter in the presence of autocorrelated errors (Altman, 1990; Chiu, 1989; Hart, 1991; Wang, 1998). A suitably adjusted GCV criterion is,

$$GCV_\rho(\lambda) = \frac{\frac{1}{n}\sum_{i=1}^{n}[z_i - \hat{\mu}(s_i)]^2}{[1 - \frac{1}{n}trF_\lambda(\mathbf{s})(1 + 2C(\epsilon))]^2}. \tag{11}$$

To compute this criterion, an estimator of the multiplier $[1+2C(\epsilon)]$ is required and a number of approaches have been suggested. Truong (1991) shows that if the errors come from an autoregressive process with finite order, the process can be consistently estimated from the residuals from a kernel smoother. The estimated process can then be used to obtain an estimator of $C(\epsilon)$. Noting that $\sigma^2(1 + 2C(\epsilon))$ is the spectral density of the error process at 0, Chiu (1989), Hart (1991) and Opsomer (1996) estimate the spectral density and then extrapolate to zero. Herrmann, Gasser and Kneip (1992) estimate $C(\epsilon)$ by the sum of the first few empirical autocorrelations of residuals. Other approaches to adjusting GCV for autocorrelation are discussed in Altman (1992), Diggle and Hutchinson (1989) and Engle et al (1986). Simulations in Altman (1992) suggest that the approaches work about the same in practice.

Figure 1b displays the fits from the partitioned data when GCV_ρ, rather than ordinary GCV was used to select the smoothing parameter. The fitted curves are much smoother than those fitted under the assumption of uncorrelated errors, and the effects of partitioning the data are (visually) negligible. The resulting smooth supports Laslett's claim that these data are close to stationary in the mean. Note, however, that under model (1), the fitted curves are estimates of the regression function, not predictors of unobserved values of the process. Comparison of these fits with the kriged values would be inappropriate as they do not estimate

the same quantities.

When the observations become dense on a compact region ("in-fill asymptotics") nonparametric estimation of both the regression and correlation functions leads to lack of identifiability. However, this confounding, which makes estimation of $\mu(s)$ difficult, may actually assist in prediction. In particular, choice of λ too large will have the effect of attributing more of the surface trend to the error correlation, but the kriging estimator of the error process should recover this. In practise, however, the effect of smoothing is to smooth at the observed data, which is similar to introducing a nugget into a kriging estimator.

4 Reanalysis of Laslett's Data Sets and the Sea Surface Temperature Data

Laslett (1994) compared the use of ordinary kriging and spline smoothing with independent errors for prediction in a number of one-dimensional spatial data sets and concluded that kriging provides better prediction for these data. Because he assumed that the errors were uncorrelated in the spline model, he used $\hat{\mu}$ as the predicted value for this model. To assess the squared prediction error, Laslett partitioned each set of data into predictor subsets P consisting of every k^{th} point and predicted at the observed locations not in P. That is:

$$\widehat{MSEP} = \frac{\sum_{s_i \notin P} (\hat{Z}(s_i) - z(s_i))^2}{\#(s_i \notin P)}. \tag{12}$$

The use of GCV with data partitioning used by Laslett is almost identical to the "partitioned cross-validation" idea of Chu and Marron (1991). Laslett's observation that the bandwidth increases dramatically when the data are partitioned was the motivation for re-analysis of these data.

Reanalysis of Laslett's data sets is undertaken in Section 4.1. The results suggest that these series are indeed close to stationary in the mean. The Müller and Stadtmüller (1988) test for detection of correlation in the presence of smooth trend was applied to determine if it could be used to choose to determine the appropriateness of model (1). All but 4 of the data sets have highly significant values of the test statistic. However, when applied to the partitioned data sets, the test was not sufficiently sensitive to pick up the moderate residual autocorrelations.

To determine the efficacy of universal kriging as a predictor for model (1), a time series which is clearly nonstationary in the mean, the sea surface temperature data of Breaker and Lewis (1988), is analyzed in the same way. Against the intuition of the author, ordinary kriging performed as well as correlated splines for point prediction of these data. However, for all data sets, a much simpler and model-free method, linear interpolation, provided very similar prediction error to the statistical methods.

4.1 Data Analysis

In this section, kriging and correlated splines are assessed as predictors for the 11 one-dimensional spatial data sets of Laslett (1994, data available from STATLIB), which appear to be generated by stationary spatial processes and for a nonstationary time series, the sea surface temperature data of Breaker and Lewis, 1988. As well, two other predictors are used, the mean of the data, and the simple linear interpolator of the adjacent pairs. The data mean is the appropriate predictor when the data are stationary and the errors are independent. When the data are generated by model (1) with no nugget, all methods interpolate - the linear interpolating spline is used as a computationally simpler alternative. The goodness of prediction is assessed by (12).

Following Laslett (1994) the error covariance is modeled by the spherical function

$$
\begin{aligned}
C(d) &= E(\epsilon(s)\epsilon(s')) && |s-s'| = d \\
&= \begin{cases} c_0 + c_1 & \text{if } d = 0 \\ c_1(1 - 1.5(d/\gamma) + .5(d/\gamma)^3) & 0 < d < \gamma \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

where c_0 is the nugget, $c_0 + c_1$ is the error variance, and γ is the range. (Values at distances greater than the range are independent.) Use of the simpler exponential covariance function had negligible effect on the predictions. Modeling without the nugget parameter c_0 improved the prediction for many of the data sets. However, the estimated nugget parameter was zero for only a few series (no attempt was made to test for this) and the analyses reported below included an estimator of the nugget.

For the universal kriging predictor, REML is used to estimate Σ and the regression function

is then fitted using general least squares with the estimated covariance function. The prediction is computed by (5). The correlated spline predictor is fitted in 4 steps.

1. A regression function is fitted by the kernel estimator of Altman (1992) and $C(\epsilon)$ is estimated from the first order autocorrelation.

2. The regression function is fitted using a cubic smoothing spline, with smoothing parameter selected by GCV_ρ and $C(\epsilon)$ estimated in Step 1.

3. The spherical correlation function is fitted to the residuals from Step 2, using the maximum likelihood equations (i.e. ignoring the fact that these are residuals).

4. The fitted values are computed by (5).

A REML-type approach was also tried in Step 3. For most series, this made the fit slightly worse, and led to some numerical difficulties.

4.2 Drum Roller Data

The data are 1150 observations of adjusted height measured along the drum of a roller, more fully described in Laslett (1994). Following Laslett, the modeling set is chosen to be the observations at $2, 4, 6, \cdots 1150$ microns and the interpolation set is chosen to be $3, 5, \cdots 1149$ microns. Then the roles of the odd and even numbered observations are switched. Prediction is done only at interior points, to avoid extrapolation.

The empirical autocorrelations are all positive but die away somewhat more slowly than they should if the exponential or spherical models suggested by Laslett (1994) are correct. Nonetheless, we assume that the data come from an spherical process.

The $\widehat{MSEP}$ from all the fits are displayed in Table 1. The most notable thing is that linear interpolation outperforms the other methods for the "odd" and "even" subsequences. This is particularly interesting, because the statistical methods all indicated the presence of a substantial "nugget effect" which suggests that smoothing should provide better prediction than interpolation. The kriging fits were comparable to one another and better than the correlated spline fits. For the subsequence consisting of every 10th point, the spline and kriging fits were better than linear interpolation, but only a small improvement over fitting the sample mean.

4.3 Gilgai Data Sets

Laslett (1994) also modeled 9 spatial series taken along on a gilgai, a gentle depression. The data are 3 different measures (pH, log(chloride) and log(electrical conductivity)) at 3 depths, (0-10 cm., 30-40 cm. and 80-90 cm.) on a transect. We follow Laslett in modeling these multivariate data as 9 univariate series. The results are displayed in Table 2.

The most notable feature of Table 2 is that for 7 of the 9 series, linear interpolation provides the best prediction or is within 10% of the best, the exceptions being pH at the shallow depths. For these 2 series, the sample mean is at least as good as the statistical methods. When the Müller and Stadtmüller test did not indicate significant residual autocorrelation, the correlated spline model performed well, but was not as good as the kriging methods. When the test indicated significant residual autocorrelation, there was often a statistical model which performed clearly worst, but there was no clear "winner" over all the data sets.

4.4 Sea Surface Temperature Data

To determine the relative merits of kriging and correlated spline modeling for nonstationary data, the sea surface temperature data of Breaker and Lewis (1988) was modeled. This data set consists of 12 years of daily measurements of sea surface temperatures collected daily at Granite Canyon, but to keep the REML computations reasonable, only the first 3 years of data were analyzed for this study. The data show pronounced seasonal effects, and the effects of a large "El Nino" warming event in 1972 is evident. The full data were previously analyzed with kernel regression estimators in Altman (1990, 1992) with a view to estimation of μ.

Although ordinary and universal kriging models do not appear to be appropriate for these data, stationary spherical errors were assumed, and the odd numbered data were predicted from the even numbered data and vice versa as described above. Although the smoothing spline estimator picks up considerable nonstationarity (Figure 3b, solid line) predictions from the kriging models were better, and linear interpolation performed comparably. This is likely because the kriging variance estimators did not indicate a nugget effect, and thus kriging predictors were interpolators, whereas the variance estimator from the spline model had a substantial nugget. The correlated spline, which follows the seasonal trend, performed better

when every 10th data point is used for prediction. The mean, not surprising performs very poorly.

4.5 Other Subsequences

Above we have followed Laslett in interpolating data to a finer scale than the observed prediction set - e.g. when predicting from the even numbered observations, the finest scale for which it is possible to estimate the correlation is $\rho(2)$, but the interpolation requires an estimate of $\rho(1)$. It is clearly difficult to statistically resolve features finer than the scale of the prediction set.

When the data provide resolution finer than the scale for which predictions are to be made, we might expect better behavior of estimator (5). Suppose, for example, that of the 364 observations on a gilgai, 10 consecutive observations are removed. $\rho(i)$, $i = 1 \cdots 10$ are required for prediction of these data, but due to the assumption of intrinsic stationarity, these are available from the remainder of the data. The kriging estimates of the error can therefore be quite wiggly. By contrast, the spline estimate of the mean is a cubic polynomial between knots (which are all at the data locations) and so is very smooth in gaps. Universal kriging assumes a polynomial regression function. If a low order polynomial is used, the regression function is forced to be smooth whether or not there are gaps in the data locations. However, if high order polynomials are used, spurious bumps can occur in data gaps as the weighted least squares measures the fit only at the observed locations.

To understand the differences between methods better, we removed the data at 101-110, 161-170 and 261-260 from the gilgai series pH 30, and then predicted at these locations from the remaining 334 observations. The MSEP was 0.139, 0.149 and 0.161, for ordinary kriging, universal kriging and correlated splines, respectively and 0.210 for the mean. These results are perhaps not surprising, since the series is close to stationary in the mean, so that fitting the extra parameters of the smoothing spline and cubic regression add variance to the predictions without bias reduction.

4.6 Conclusions from Data Analysis

Ordinary kriging provided very good prediction for these data, even when the data were highly nonstationary in the mean. Linear interpolation also provided very good prediction when the data grid was fairly dense compared the scale of interpolation, but did not do well when the grid was sparse. Both methods are computationally simpler and require less modeling than universal kriging and correlated splines. Ordinary kriging has the additional advantages over linear interpolation that it adapts to sparseness of the grid, and that prediction intervals can be computed.

The poor performance of the correlated spline for prediction was somewhat surprising, since both the ordinary and universal kriging models can be shown to be limiting cases. The variance parameters for the odd sequences of the gilgai data are displayed in Table 6. Examination of the estimated variance parameters shows that in most cases, the estimated range decreases and the estimated nugget increases as we go from ordinary kriging to universal kriging to correlated splines. Accordingly, there is progressively more smoothing and less interpolation as we go from the most constrained to the least constrained regression estimator. Not surprisingly, the correlated spline does worst when linear interpolation does best.

5 Regression

Much of the work in spatial statistics has focused on prediction. However, occasionally the goal is to predict the spatial distribution of an independent realization of the process (e.g. Hobert, Altman and Schofield, 1997, predicts the number of species that could be supported by lakes in the Adirondacks if acid deposition is reduced). In this case, the prediction errors are independent of the observed data, so that the best point predictor is the regression estimator.

Figure 2 displays the predictions and regression estimates for the sea surface temperatures. The prediction set was the even numbered subsequence. The solid, dashed and dotted lines are the predictions from the smoothing spline, ordinary kriging and universal kriging respectively. Figure 2a displays the predictions. These are very wiggly as they track both the variation in the mean and the variation in the correlated errors. All three predictors produce very similar

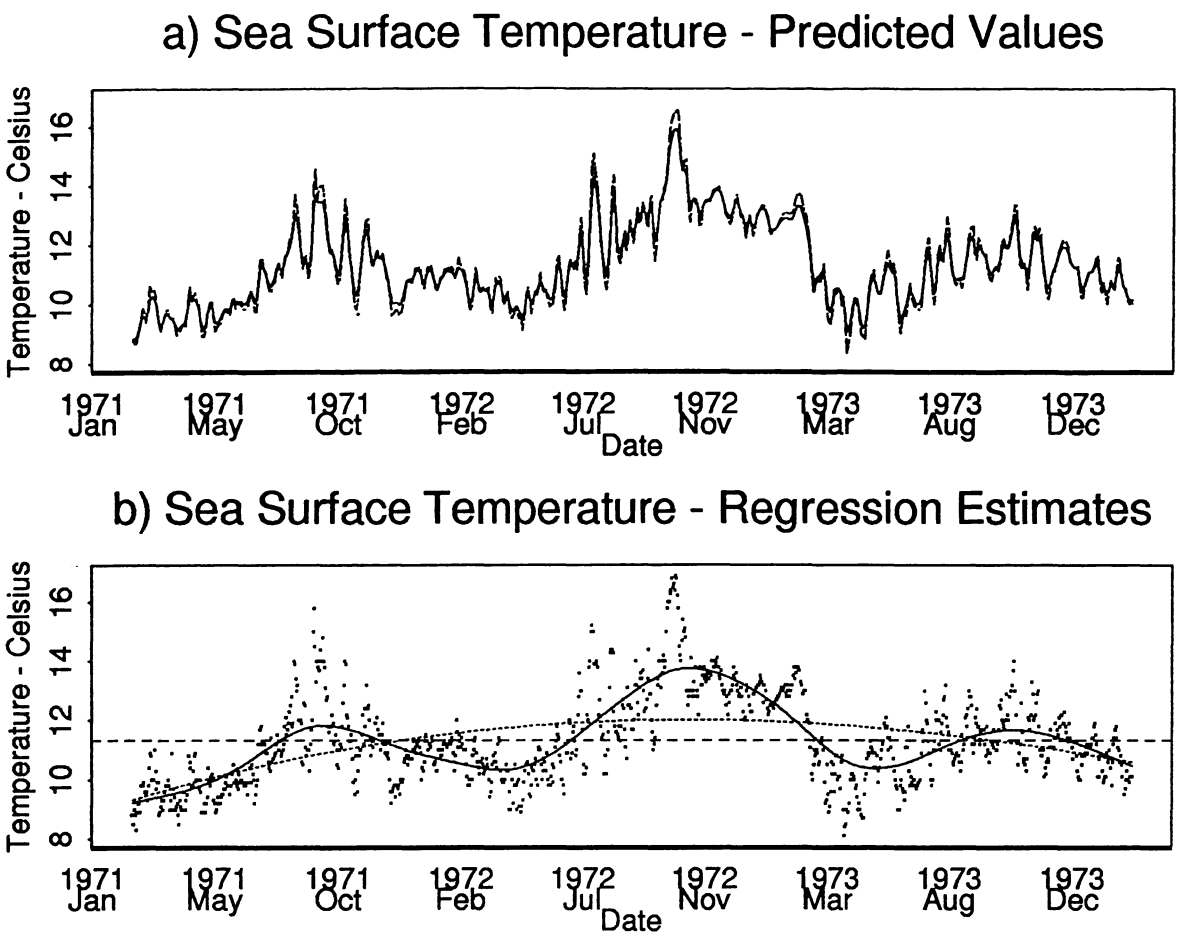

Figure 2: Predicted sea surface temperatures at Granite Canyon for the odd numbered days based on the data from the even numbered days. a) Predicted values for the odd days. The solid lines, dashed lines and dotted lines are the predictions from the smoothing spline estimator, ordinary kriging and universal kriging respectively. Notice that all three models produce predictions which are extremely close. b)Predicted values (regression estimates) for a hypothetical independent realization of the process. Notice that the smoothing spline tracks the seasonal effects of the data.

predictions. The poorer performance of the correlated spline model is clearly seen to be caused by the tendency of the predictions from this model to be less extreme in the peaks and dips, due to the non-zero nugget.

The regression estimates for the sea surface temperature data are displayed in Figure 2b. As we can see from this extreme example, point predictions for an independent realization vary considerably between the models. Fitting a sinusoid, rather than a cubic polynomial, would of course be preferable for the universal kriging model. However, the interesting point is that choice of the model has little effect on prediction, but a great deal of effect for regression.

The estimated variance of the data from the ordinary kriging model is 3.75, from universal kriging model is 3.02, and from the correlated spline estimator is 0.29. These differences are due the amount of variation in the data which each estimator attributes to the regression function. The kriging models attribute the seasonal effect to residual variance. Although all three models produce similar predictions for interpolating the current realization, the predictions for an independent realization would be very different, and the prediction bands for the spline model would be much narrower, due to the small estimated population variance. The appropriateness of the various models for regression cannot be determined from the fit to a single realization of the process, but must be determined from subject area knowledge and analysis objectives.

6 Summary

Reanalysis of Laslett's data was motivated by Figure 1a, in which partitioning the data caused GCV to select a much larger smoothing parameter for subsequences than for the full data set. Such behavior is indicative that at least some of the smoothness of the data is due to autocorrelation rather than trend. The test statistic suggested by Müller and Stadtmüller (1988), although useful for detecting high autocorrelation, does not appear to be sufficiently sensitive for detecting error autocorrelation if GCV or related techniques are to be used for selecting smoothing parameters, because of the extreme sensitivity of these techniques to even small positive autocorrelation.

The smoothing spline model is related to universal kriging in a number of ways. The spline estimator of the regression function tends to the least squares line as $\lambda \to \infty$. An alternative

formulation of the smoothing spline as a penalized maximum likelihood estimator for model 1 leads to use of the estimator which minimizes

$$(z(\mathbf{s}) - \mu(\mathbf{s}))'\Sigma^{-1}(z(\mathbf{s}) - \mu(\mathbf{s})) + \lambda \int \mu''(s)^2 ds. \tag{13}$$

(Diggle and Hutchinson, 1988; Engle et al, 1986). In this case, as $\lambda \to \infty$ the smoothing spline estimator tends to a linear fit by weighted least squares. Another connection between universal kriging and smoothing splines comes from the observation that $\hat{\mu}(\mathbf{s})$ can be obtained as a solution to the linear mixed model

$$Z = X\beta + Z\gamma + \eta \tag{14}$$

where X is the matrix of cubic polynomials in location, β is the vector of regression coefficients, Z is a set of truncated polynomials determined by the knots (observed locations), γ is a mean zero random vector with covariance structure determined by σ^2, λ and the definition of the smoothing spline (equation 4 or 13) and η is an uncorrelated error with mean zero and variance σ^2. Thus the smoothing spline model can be seen to be a special case of universal kriging with covariance structure determined by the variance of $Z\gamma + \eta$ (Wang, 1998). However, interpretation differs. The solution to the mixed model equations is an estimate of the regression function but the appropriate predictor is (5).

Note, however, that Model 14 can also be used for correlated splines, assuming a spatial (or other) covariance structure for η (Wang, 1998). This is a convenient formulation for handling correlated splines via the linear mixed model and is readily implemented in standard mixed model software.

Comparisons of nonparametric smoothing and kriging currently in the literature (Cressie, 1989; Hutchinson and Gessler, 1994; Laslett, 1994; Voltz and Webster, 1990; Wahba, 1990; Yakowitz and Szidarovsky, 1985) typically assume that the errors in the smoothing model are uncorrelated. Thus comparison between fits are really comparisons between models. We have seen in this article that ordinary kriging can predict well even in the presence of a nonstationary mean, and that there is little improvement (or even deterioration) in point prediction when universal kriging or correlated splines are fitted. Differences in estimating the regression function may be large, since kriging models make very restrictive shape assumptions. Typically,

the correlated spline model typically produces a smaller estimate of σ^2.

This article has focused primarily on the 1-dimensional equally spaced case which was the focus of the data analyses in Laslett (1994). However, advances in linear mixed models methodology makes it a simple task to use either kriging or correlated splines for higher dimensional, irregularly spaced data.

The rather surprising (at least to the author) result of assessing the methodology on real data is that, even in the presence of obvious nonstationarity, the extra effort of regression function estimation does not seem to improve the predictions over ordinary kriging, particularly if the autocorrelations are strong and the scale at which interpolation is required is not much finer than that at which the data were collected. However, in this case linear interpolation of the data performs as well as kriging, so that the effort of modeling the covariance or variogram also seems superfluous. Of course, more realistically a prediction interval, rather than a point prediction, is desired and kriging provides a more statistically valid interval.

Overall, for prediction, kriging provides a simpler estimator than correlated splines, and appears to be quite robust to misspecification of a smooth regression function. For regression, correlated splines provide an adaptive fit to the regression function which is not provided by parametric kriging. However, it is critical to account for the correlation when choosing the smoothing parameter.

References

Altman, N.S. (1990) "Kernel Smoothing of Data with Correlated Errors," *Journal of the American Statistical Association* **85** 749-759.

Altman, N. S. (1992) "An Iterated Cochrane-Orcutt Procedure for Nonparametric Regression," *Journal of Statistical Computing and Simulation* **40**, 93-108.

Breaker, L.C. and Lewis P.A.W. (1988) "A 40-50 Day Oscillation in Sea-Surface Temperature Along the Central California Coast," *Estuarine Coast and Shelf Science* **26** 395-408.

Chiu, S. (1989) "Bandwidth Selection for Kernel Estimate with Correlated Noise," *Statistics & Probability Letters* **8** 347-354.

Chu, C.-K. and Marron, J. S. (1991) "Comparison of Two Bandwidth Selectors with Dependent Errors," *Annals of Statistics*, **19**, 1906-1918.

Craven, P. and Wahba, G. (1979) "Smoothing Noisy Data with Spline Functions". *Numerische Mathematik* **31**, 377-403.

Cressie, N. (1989) "Geostatistics," *The American Statistician*, **43**, 256-258.

Cressie, N. (1993) *Statistics for Spatial Data (Revised Edition).* John Wiley & Sons: New York.

Diggle, P. G. and Hutchinson, M. F. (1989) "On Spline Smoothing with Autocorrelated Errors," *Australian Journal of Statistics*, **31**, 166-182.

Engle, R. Granger C.W.J. Rice, J. Weiss A. (1986) "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association* **81** 310-320.

Eubank, R. L., (1988) *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, Inc., New York.

Hart, J.D. (1991) "Kernel Regression with Time Series Errors," *Journal of the Royal Statistical Society Series B*, **53**, 173-187.

Hart, J.D. and Wehrly, T.E. (1986) "Kernel Regression Estimation Using Repeated Measurements Data," *Journal of the American Statistical Association* **81** 1080-1088.

Herrmann, E., Gasser, T., Kneip, A. (1992) "Choice of bandwidth for kernel regression when residuals are correlated." *Biometrika*, **79**, 783- 795.

Hobert, J. P., Altman, N. S. and Schofield, C. L. (1997) *Spatial Analysis of Fish Species Richness of Adirondack Lakes: Applications of Geostatistics and Nonparametric Regression.* (to appear in Journal of the American Statistical Association).

Hutchinson, M. F., and Gessler, F. R. (1994) "Splines – More Than Just a Smooth Interpolator," *Geoderma*, **62**, 45-67.

Laslett, G. M. (1994), "Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications," (with discussion) *Journal of the American Statistical Association*, **89**, 392-409.

Müller, H.-G. and Stadtmüller, U. (1988) "Detecting Dependencies in Smooth Regression Models," *Biometrika*, **75**, 639-650.

Opsomer, J. (1996) "Estimating an Unknown Function by Local Linear Regression when the Errors are Correlated." Iowa State University preprint 95-42.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992) *Variance Components.* John Wiley & Sons: New York.

Silverman, B.W. (1984) "Spline Smoothing: The Equivalent Variable Kernel Method," *Annals of Statistics* **12** 898-916.

Truong, Y.K. (1991) "Nonparametric curve estimation with time series errors," *Journal of Statistical Planning and Inference*, **28**, 167-183.

Voltz, M. and Webster, R. (1990), "A Comparison of Kriging, Cubic Splines and Classification for Predicting Soil Properties From Sample Information," *Journal of Soil Science*, **41**, 473-490.

Wahba, G. (1990), "Comment on Cressie," *The American Statistician*, **94**, 255-256.

Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing.* Chapman and Hall: London.

Wang, Y. (1998) "Smoothing Spline Models with Correlated Random Errors", *Journal of the American Statistical Association*, **93**, 341-348.

Yakowitz, S. J. and Szidarovsky, F. (1985), "A Comparison of Kriging with Nonparametric Regression Methods," *Journal of Multivariate Analysis*, **16**, 21-53.

Drum Roller Data						
		Mean Squared Error of Prediction				
Estimation Set	Prediction Set	Correlated Spline	Ordinary Kriging	Universal Kriging	Linear Interpolation	Mean
odd	even	.356	.252	.260	.237	.460
even	odd	.293	.222	.218	.208	.387
ending in 0	not ending in 0	.406	.398	.390	.500	.423

Table 1: Mean squared error of prediction for partitions of the drum roller height data using various predictors.

Gilgai Data Sets					
	Mean Squared Error of Prediction				
Data Set	Correlated Spline	Ordinary Kriging	Universal Kriging	Linear Interpolation	Mean
Cl 0	1.41	1.14	1.14	1.23	1.57
Cl 30*	2.38	1.01	1.02	0.900	2.56
Cl 80*	0.656	0.383	0.359	0.341	0.729
Ec 0	0.454	0.416	0.421	0.452	0.562
Ec 30*	0.897	0.517	0.493	0.459	1.03
Ec 80*	0.389	0.283	0.263	0.242	0.542
Ph 0	0.567	0.569	0.560	0.700	0.574
Ph 30	0.149	0.143	0.142	0.171	0.156
Ph 80*	0.219	0.214	0.180	0.176	0.415

Table 2: Mean squared error of prediction averaged over even and odd partitions of the gilgai data sets using various predictors. Data sets which had significant residual autocorrelation as determined by the Müller and Stadtmüller test are flagged with an asterisk (*).