Infilling Sparse Records of Spatial Fields

Craig J. JOHNS, Douglas NYCHKA, Timothy G. F. KITTEL, and Chris DALY

Historical records of weather, such as monthly precipitation and temperatures from the last century, are an invaluable database to use in studying changes and variability in climate. These data also provide the starting point for understanding and modeling the relationship among climate, ecological processes, and human activities. However, these data are observed irregularly over space and time. The basic statistical problem is to create a complete data record that is consistent with the observed data and is useful for other scientific disciplines. We modify the Gaussian-inverted Wishart spatial field model to accommodate irregular data patterns and to facilitate computations. Novel features of our implementation include the use of cross-validation to determine the relative prior weight given to the regression and geostatistical components and the use of a space-filling subset to reduce the computations for some parameters. We feel that the overall approach has merit, treading a line between computational feasibility and statistical validity. Furthermore, we are able to produce reliable measures of uncertainty for the estimates.

KEY WORDS: Bayesian spatial interpolation; Cross-validation; Geostatistics; Prediction.

1. INTRODUCTION

Understanding climate variability and its effect on environmental processes is important not only to increase our scientific understanding of the earth system, but also to assess the impact of a changing climate on society. Historical records of weather, such as monthly temperatures and precipitation from the last century, are an invaluable database to use in studying changes and variability in climate. These data also provide the starting point for understanding and modeling the relationship among climate, ecological processes, and human activities. Unfortunately, even in regions relatively rich in climate data, such as in the United States and Europe, station records often have gaps and extensive missing periods. Many scientific activities related to climate change and variability require temporally-complete climate series that properly represent temporal variability at key time scales (monthly to centennial) and preserve the point location nature of climate records. Thus the basic statistical problem is to create a complete, or an *infilled*, version of the data record that is consistent with the observed data and is useful to researchers in other disciplines.

As an example of the pattern of missing observations over time we plot a typical station in Figure 1. The July time series of total precipitation in millimeters for Arco, in south-central Idaho, (station ID# 100375), is plotted with "× 's" to indicate missing months. In this case there are 67 years of data observed in the 78-year time period of 1920–1997. Our goal is to estimate the total precipitation for each July in the years 1895–1919 and for the 11 years beyond 1920 with missing July observations.

1.1 Scientific Need

An example of the pervasive need for climatological analysis of the United States is the National Assessment (NAST 2001), produced for the U.S. Global Change Research Program. This study was mandated by an act of the U.S. Congress to report on the impacts of climate change on the United States. The assessment's conclusions (p. 537, items 8 and 10) highlight the need for improved long-term datasets and methods that address uncertainty in such data. An infilled record of historical monthly meteorology for the United States is a useful tool for nonstatisticians who examine climatic changes and variability because impacts are assessed at a local level, and so point data are preferred. An infilled time series, even carrying the caveat of being partially estimated as opposed to completely measured, is a still a valuable product. The unique contribution of a statistical approach is the companion quantification of the uncertainties of the infilled portion. At a more technical level, large numerical models that simulate the Earth's climate system, atmosphere/ocean general circulation models (AOGCM's), are tested by their ability to reproduce the past climate. AOGCM's are the primary tool for predicting future changes in climate, and their ability to reproduce past climate is an important measure of their validity. For the purposes of model comparison, an infilled dataset along with measures of uncertainty is a primary reference for evaluating model results.

A more intrinsic need for completed observational data is as the input to ecological and biogeochemical models. To quantify the relationship between climate and ecological processes, numerical models are built to determine the response of vegetation and soil nutrients due to changes in meteorology. As an example, CENTURY (available at *http://www.nrel.colostate.edu/projects/century5/*) is a general model of plant–soil–nutrient cycling used to study the dynamics of carbon and nutrients in a variety of ecosystems. As CENTURY is run, it requires temperature and precipitation on a monthly time scale as a subset of its inputs. Important baseline experiments involve running these models using the observational record as inputs. For example, interannual and decadal variation in climate strongly controls variation in the structure of ecosystems and the movement of carbon and water through the terrestrial biosphere. These processes in turn control more direct influences, such as agricultural production, occurrence of wild fires, carbon sequestration, and urban water supply. Adequate representation of climate variation is crucial to provide confidence in the analysis and modeling of these relationships (e.g., Schimel et al. 2000; Kittel et al. 1997).

Craig J. Johns is Assistant Professor, Mathematics Department, University of Colorado, Denver, CO 80217 (E-mail: *cjohns@math.cudenver.edu*). H. Douglas Nychka is Director, Geophysical Statistics Project and Timothy G. F. Kittel is Scientist, Terrestrial Sciences Section, National Center for Atmospheric Science, Boulder, CO. Chris Daly is Director, Spatial Climate Analysis Service, Department of Geosciences, Oregon State University, Corvallis, OR. This work was supported by National Science Foundation grants DMS 9815344 and DMS 9312686 for the Geophysical Statistics Project, and by NOAA grant NA76GP0558. VEMAP was supported by NASA, EPRI, USDA Forest Service, and the U.S. Department of Energy. The authors thank Nan Rosenbloom (NCAR), Alan McNab (NOAA/National Climactic Data Center), and Wayne Gibson (Oregon State University) for their efforts in compiling and managing climate station datasets used in this study and Richard Jones (University of Colorado, Health Sciences Center) for useful discussions early in the project. The authors also thank the editor, associate editor, and two referees for suggestions that refined and improved the manuscript.

Journal of the American Statistical Association
December 2003, Vol. 98, No. 464, Applications and Case Studies
DOI 10.1198/016214503000000729

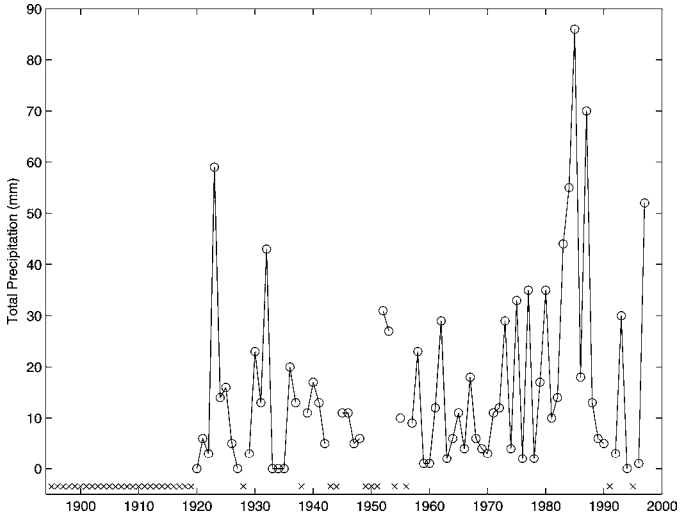

Figure 1. The Record of July Precipitation (in mm) for the Arco, ID Station (100375) in South-Central Idaho. The sign "×" denotes years where the record is missing.

1.2 Creating Complete "Datasets"

We focus on the statistical problem of estimating missing monthly precipitation measurements for an observational network of more than 10,000 stations. Situations similar to this, a large observation network but irregularly observed data on that network, are quite typical in historical records of geophysical data. The main hurdles here are the irregularity of the observations and nonstationary behavior over both space and time. Although these features demand careful statistical modeling, the sophistication is limited of such modeling by the large size of typical geophysical datasets. Thus the challenge and focus of our work is to find a balance between adequate statistical models and efficient methods that enable processing the data in a reasonable amount of time and provide efficient estimators and companion measures of uncertainty. Although this work deals with monthly precipitation, our spatial approach is flexible and will be useful for other meteorological fields, such as temperature, and other large spatial problems.

High-quality, high-resolution temperature and precipitation data products for the coterminous United States are now available for use in various natural resource, agricultural, and hydrological modeling and assessment activities (USDA-NRCS 1998; Daly et al. 2001). However, these datasets contain climatological means only (typically means from the time period 1961–1990), and they do not adequately reflect the rich structure of climate variability and trends contributing to these means. Earlier investigations (Kittel et al. 1997, 2000) generated a complete spatiotemporal climate dataset for the coterminous United States. However, these authors used fewer stations and a simpler infill method than we present here. Furthermore, this effort did little to quantify infill errors. Although the Historical Climatology Network provides largely complete, long-term station data for the coterminous United States (Karl, Williams, Quinlan, and Boden 1990), it is limited to approximately 1,200 stations, with locations that are biased toward low elevations. Meteorology at higher elevations is important, because it may be more sensitive to a changing climate and also because a significant portion of carbon sequestration and seasonal water storage occurs at higher elevations (Kittel, Thornfon, Royle, and Chase 2002; Schimel et al. 2002).

The construction of "data products" from raw observational data is a common activity in the geophysical community and comes from the recognition that complicated datasets require nontrivial preprocessing, quality control, and analysis to be useful for further scientific investigation. We believe that statisticians can contribute to these efforts not only by helping to transfer advanced and efficient statistical methods, but also by deriving measures of uncertainty when data are interpolated over space or infilled over time. Thus, the motivation behind our work was both to provide good predictions of precipitation and to quantify the uncertainty of the infilled values. A natural way to quantify uncertainty is with the posterior distribution from a Bayes model. However, one goal of our modeling was to ensure that credible intervals derived from the posterior distribution also have good frequentist properties. Beyond standard errors for individual predictions, our methods are also well suited to sampling from the full multivariate posterior. This method, also known as "conditional simulation" in geostatistics, provides a physically realistic field that is consistent with the observations. These conditional simulations can be used to generate an *ensemble* of possible inputs for the ecological and climate models, which will give a better understanding of the variability of predictions obtained from these models.

In this work we focus on spatial prediction at station locations where data are missing. This is a different problem than prediction at arbitrary locations. Although a station location may not have data at a particular time, data at other periods may exist that allow sample estimates of the covariance between that station and its neighbors. For arbitrary locations, one must rely on a model to infer covariance structure. Although we do not address directly the interpolation/extrapolation of precipitation to a regular grid, the basic infilling method is a major step in the process of developing temporally and spatially complete climate datasets (Kittel et al. 1997). The subsequent step of extrapolating station data to a fine grid in this project was accomplished using a terrain-based model. We briefly discuss this extrapolation in Section 2.1, details have been given by Daly, Gibson, Taylor, Johnson, and Pasteris (2002).

1.3 Regression Versus Kriging

As background to the Bayesian approach, we first distinguish between two common methods used to infill (predict) missing values, nearest-neighbor regression (NNR) methods and geostatistical (GS) models. Both of these methods have advantages, and the strength of the Bayesian model used in our work comes from the synthesis of these two methods into a single approach.

Infilling via NNR methods is generally carried out by selecting a set of locations that are close to the locations of interest. Infilled values are typically weighted averages of the observations in the neighborhood. The weights are computed by various methods, including inverse distance (Shepard 1968) and some variants (Legates and Willmott 1990), kriging (Haas 1990), and simple or robust polynomial regression (Rajagopalan and Lall 1998). The regression approach is appealing for its simplicity and predictive ability. One regresses a station's set of measurements on its neighboring values for

time periods when all data are complete. For periods when station data are missing, the measurements are simply predicted from the regression relationship using the neighboring values. The advantage of this approach is that it adapts to each station location. However, in many cases, NNR methods are hindered by ad hoc weighting schemes and short records in the case of regression. Furthermore, the regression techniques are not suited to spatial prediction at locations that are not part of the observation network.

As an alternative, geostatistical methods (Cressie 1993; Stein 1999) rely on a covariance model for the field of interest and derive prediction weights based on the assumed covariance model. Because the predictive weights for the observations depend on the covariance model, infilling efficiency is closely related to accurate modeling of covariances. Typical parametric models for spatial covariances, such as the Matern family, are not expected to hold for a large heterogeneous domain such as the coterminous United States. Nonparametric methods for estimating covariances are also available (Sampson and Guttorp 1992; Higdon, Swall, and Kern 1999); however, these approaches are infeasible for datasets with numerous station locations.

This analysis combines an empirical Bayes implementation of a model similar to that of Brown, Le, and Zidek (1994) mixed with the neighborhood ideas of Haas (1990, 1995). The basic elements of the model include the usual assumptions of a multivariate normal distribution of the field and an inverse Wishart prior on the spatial covariance. The appeal of this model is that the predictor is a mixture of regression and geostatistical methods and allows the model to inherit the strengths of both procedures. Although this basic model is not new, our empirical implementation has several novel features. These include the use of cross-validation to determine the relative weight given to the regression and geostatistical components and the use of a local window that allows for nonstationary fields. We feel the overall approach has merit, marrying computational feasibility and statistical validity. Furthermore, we are able to produce reliable measures of uncertainty for the infilled estimates.

Formally, the reader may identify the infilling problem or the interpolation to a grid as a nonparametric regression. Given irregular data in two dimensions, estimate a smooth surface that describes the full field. With this connection, one could use generic surface-fitting methods, such as kernel estimators or local regression (e.g., loess), to predict the infilled values, and we expect that with suitable tuning, such methods may perform as well as spatial statistical approaches. The disadvantage of these methods is that deriving reliable measures of the estimator's uncertainty is difficult. A hybrid estimator is a thin plate spline interpretable as both a numerical interpolation method and a spatial statistics estimator (Nychka 2000). Although we use thin plate splines in some instances where estimating a full covariance function is difficult, we feel that the generalized covariance model assumed in the thin plate spline formulation is too limited for the precipitation field itself.

The first step in our analysis is to transform and standardize the precipitation totals, which we do in Section 2. We then apply spatial modeling and prediction to these standardized anomalies, with each calendar month considered separately. Our infilling approach relies on a Bayesian model, described in Section 3, and yields an approximate posterior distribution for the missing observations given a small number of neighboring observations. We provide details of the nearest-neighbor criteria that we used in Section 3.1 and outline the infill algorithm in Section 3.2. In Section 4 we discuss specification of the model and prior parameters specific to the analysis of the precipitation data. We present results in Section 5 and a discussion in Section 6.

2. DATA AND PRELIMINARY ANALYSIS

The weather observation record for the United States is based on a combination of different types of stations, ranging from consistent, unbroken recordings at population or scientific centers to often more sporadic measurements made by cooperative observers. We started with 17,405 stations across the contiguous United States providing some records over the period January 1895–December 1997. The variable of interest is total monthly precipitation, recorded in millimeters. Data originally came from the National Climatic Data Center (NCDC) and have passed through their quality control system that includes only monthly data that are completely observed. The data in this study were also rechecked as part of the PRISM analysis system.

Spatial coverage is sparse in the early years. For example, in January 1895, the beginning of our analysis window, only 851 of the stations were reporting, whereas in January 1964 7,921 stations reported. Overall, there is a rapid growth in the size of the observation network in the 1940s with the total number of reporting stations reaching a maximum in the early 1960s. The station density partly reflects the density of population and does not have uniform geographic (or geophysical) coverage.

Many stations have very short periods of record, and we eliminated stations with less than 10 years observed in each calendar month. This reduced the total number of stations to 11,918, but reduced the total number of observations by less than 5%. In part, this reduction improved the chances of overlapping time series between most stations. Although this reduction may seem severe, the relative loss in observations is only about 3% for most months. The exceptions are in the period 1949–1951, where the loss is about 15%. Even though 15% may seem extreme, most of these stations are in locations where the spatial density is already high, so there is little practical loss in the spatial resolution of the dataset.

Eliminating these station records improved the overall quality of the data product, because including these stations would have changed the overall infill percentage (i.e., data in the final product that are estimated, rather than observed) to 69%, compared with 57%. Furthermore, these stations would not be used as neighbors in the infill step (see Secs. 3.2 and 3.3) because of the short time record, and they would have little impact on the final output for the stations included in our analysis. Even with this reduction in stations, this is still a much larger database than is typically used for climate change studies. For example, the historical climate network of Karl et al. (1990) consists of 1,221 stations with long records that are adjusted for urban influences and is a subset of the stations used in our work.

Given the current interest and development of statistical models for space/time processes, it may come as a surprise to the reader that we do not exploit such models for monthly precipitation. Empirical results indicate that on a monthly time scale, there is little dependency between successive monthly totals. For example, average station correlation of the December anomalies with the previous November anomalies is .14, and other monthly pairwise comparisons were even smaller. Furthermore, the spatial distribution of the "large" auto correlations followed no obvious spatial pattern. It did seem evident that some correlations ($\approx .1$), were present, but ad hoc analysis showed that spatiotemporal models would reduce the mean squared error by 1%–2%.

A simple example illustrates this point. Let $\mathbf{w}_t = (x_t, y_t)^T$ be a bivariate random variable under the autoregressive model $\mathbf{w}_t = \delta \mathbf{w}_{t-1} + \mathbf{z}_t$ where $\mathbf{z}_t$ are mean 0 correlated normal vectors such that $\mathbf{z}_{t+k}$ and $\mathbf{z}_t$ are independent for $k \neq 0$. Then it is easy to show that $E(y_t|x_{t-1}, x_t, x_{t+1}) = E(y_t|x_t)$. Furthermore, it is an exercise to show that $E(y_t|y_{t-1}, x_t, y_{t+1}) \approx E(y_t|x_t)$ when $\delta \ll \text{cor}(z_{1t}, z_{2t})$. Thus ignoring the temporal aspects can be justified when strong spatial effects are accounted for. As a practical matter, for many of the infill cases, records from the previous and following months are not available, and therefore a model that includes a temporal component will not improve the estimates. Because of the apparent lack of strong temporal correlation as compared with spatial correlation, we developed the methodology for infilling missing observations using only observations that are contemporaneous in the month of interest. That is, to infill a particular station record for any July in the years 1895–1997, we use the records only from the Julys in those years, not any June or August information. For completeness, in Section 6 we suggest an extension useful for correlated data.

2.1 Estimates of Mean Monthly Total Precipitation Based on PRISM

Precipitation in general is a highly variable spatial process, and spatial predictions benefit from centering predictions about a mean surface. These values referred to in atmospheric science as the "climatology," can be estimated at a station by the month-by-month means of the observations. However, we draw on past research to use a climatological mean field that incorporates more geophysical structure. Daly, Neilson, and Phillips (1994) and Gibson, Daly, and Taylor (1997) have described PRISM (parameter-elevation regressions on independent slopes model), a climate-mapping system that produces high-resolution maps of precipitation and other meteorological variables using point data, a digital elevation model, and other spatial datasets. An extensive PRISM analysis yielded monthly climatological precipitation means for each of the 11,918 stations based on 30-year period 1961–1990. We used this as the baseline climatology. However, because of slowly varying climate patterns and because PRISM did not use station data outside the 1961–1990 time window, we found that in a small number of cases the PRISM means did not correspond closely with the observed means. Such a discrepancy leads to biases that can be detected when cross-validation is used to assess the infilling accuracy. To guard against this problem, we use a simple t statistic to flag suspect PRISM means and adjust them toward the observed station mean.

2.2 Transformation to Anomalies

To simplify the spatial structure of the monthly precipitation fields, we do the spatial infilling after transformation and standardization of the raw observations. We refer to the resulting values as *anomalies*, and unless otherwise specified, perform the statistical analysis in the anomaly scale. The spatial anomaly field has the advantages of being closer to a Gaussian distribution and having no geographical trend. Both of these features facilitate infilling, and the estimates on this scale can be easily transformed back in the scale of the original measurements. Furthermore, evidence suggests that the anomaly field is also closer to being second-order stationary compared with the raw scale (Fuentes, Kelly, Kittel, and Nychka 1998). It is well known that in many cases precipitation amounts can be approximated by a gamma distribution. In this situation, the square root function is quite efficient in transforming gamma distributions to distributions that are approximately normal, and so we used the square root transformation for all locations.

We again emphasize the independence of the monthly datasets, and so suppress dependence on the choice of month (Jan, ..., Dec) in the formulas that follow. To obtain the anomaly at each time point (year of record), the square root of total precipitation is standardized by its climatological mean and standard deviation. If $P(\mathbf{x}, t)$ is precipitation at location $\mathbf{x}$, then let $\theta(\mathbf{x}) = E[\sqrt{P(\mathbf{x}, t)}\,]$ and $\sigma^2(\mathbf{x}) = \text{Var}[\sqrt{P(\mathbf{x}, t)}\,]$. The analysis constructs spatial predictions based on the standardized anomalies, $z(\mathbf{x}, t) = (\sqrt{P(\mathbf{x}, t)} - \theta(\mathbf{x}))/\sigma(\mathbf{x})$.

The most direct way to find the spatial functions θ and σ is based on spatial estimates from the sample means and variances of individual station data. However, additional statistical modeling is advantageous, because short station records may by themselves yield highly variable estimates of local climatology. The spatial analysis used in this work leverages the high-resolution and quality-checked mean precipitation fields from the PRISM analysis, and so the actual smoothing step is nonstandard. Let $\mu(\mathbf{x}) = E[P(\mathbf{x})]$ denote the mean monthly precipitation at a given location, and let $\hat{\mu}$ denote the PRISM mean. From elementary properties of the expectation, $\mu(\mathbf{x}) = \theta(\mathbf{x})^2 + \sigma(\mathbf{x})^2$. Moreover, setting $C(\mathbf{x}) = \sigma(\mathbf{x})^2/(\theta(\mathbf{x})^2 + \sigma(\mathbf{x})^2)$, it follows that $\theta(\mathbf{x}) = \sqrt{\mu(\mathbf{x})(1 - C(\mathbf{x}))}$ and $\sigma(\mathbf{x}) = \sqrt{\mu(\mathbf{x})C(\mathbf{x})}$. Thus once $C(\mathbf{x})$ is known, estimates of θ and σ can be found using the relationships given earlier and substituting $\hat{\mu}$ from the PRISM analysis for μ. We prefer this route because the function C, related to a coefficient of variation, exhibits less spatial dependence than the individual means and variances. As an added benefit, we also found that $C(\mathbf{x})$ does not depend strongly on elevation. By constructing the estimates of θ and σ in this way, the implied estimate of μ will be $\hat{\mu}$, the PRISM mean. In this work $C(\mathbf{x})$ is estimated by smoothing the sample statistics with a kernel estimator. We determined the bandwidths by minimizing the mean squared error (MSE) for a subset of 400 stations reserved for cross-validation. The resulting bandwidths for each of the 12 months were small, followed a seasonal cycle, and ranged from approximately 25 to 35 miles (.4 to .6 degrees of longitude/latitude).

3. APPROXIMATIONS TO A BAYESIAN ANALYSIS

Brown et al. (1994) and Le, Sun, and Zidek (1999) described a hierarchical model used for spatial interpolation based on

Gaussian–inverted Wishart and Gaussian–generalized inverted Wishart models. These models constitute a standard method for Bayesian spatial interpolation. We follow the basic tenets of these models, but make modifications and approximations as necessary.

Assume that $[\mathbf{Z}_t|\boldsymbol{\Omega}] \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$ are independent (conditional on $\boldsymbol{\Omega}$) normally distributed p-dimensional random variables with mean 0 and covariance $\boldsymbol{\Omega}$, for $t = 1, \ldots, T$, where $\mathbf{Z}_t$ represents the distribution of precipitation amounts in the standardized square root scale at all locations for time t. The prior distribution for the variance-covariance matrix $\boldsymbol{\Omega}$ is inverse-Wishart with degrees of freedom $\nu + p + 1$ and symmetric shape matrix $\nu\mathcal{K}$, denoted by $IW(\nu + p + 1, \nu\mathcal{K})$. The parameters are tied together in this way so that $E(\boldsymbol{\Omega}) = \mathcal{K}$. We define $\mathcal{K}$ from a spatial correlation function k; for two locations $\mathbf{x}_i$ and $\mathbf{x}_j$, $E[\Omega_{ij}] = k(\mathbf{x}_i, \mathbf{x}_j)$, and we interpret $\mathcal{K}$ as the expected a priori average correlation function for precipitation anomalies. (Sec. 4 details our choices of $\mathcal{K}$ and ν.) Under these assumptions, the conditional distribution of $\mathbf{Z}_t$ given all other data $(\mathbf{Z}_1, \ldots, \mathbf{Z}_{t-1}, \mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_T)$ follows a multivariate Student t distribution and provides the statistical framework for infilling. A key feature of the model is flexibility in the spatial covariance. Although the prior centers Ω on a particular spatial correlation function, the conditional distribution can modify this form based on the sample correlations from the data. Specifically, $[\mathbf{Z}_t|\mathbf{Z}_1, \ldots, \mathbf{Z}_{t-1}, \mathbf{Z}_{t+1}, \ldots \mathbf{Z}_T] \sim t_p(\nu_c, 0, Q/\nu_c)$, where $Q = \nu\mathcal{K} + \sum_{i \neq t} \mathbf{Z}_i\mathbf{Z}_i^T$ and $\nu_c = \nu + T - 1$.

The posterior distribution of the missing-data $\mathbf{Z}_t^{(0)}$ given the observations $\mathbf{Z}_t^{(1)}$ at time t and observations at all other times is our Bayesian infill distribution. It also follows a multivariate t distribution,

$$\left[\mathbf{Z}_t^{(0)} \,\middle|\, \mathbf{Z}_t^{(1)}, \mathbf{Z}_s, s \neq t\right] \sim t\left(\nu^\star, Q_{01}Q_{11}^{-1}Z_t^{(1)}, \xi[Q_{00} - Q_{01}Q_{11}^{-1}Q_{10}]\right), \quad (1)$$

where $\xi = (1 + \mathbf{Z}_t^{(1)'}Q_{11}^{-1}\mathbf{Z}_t^{(1)})/\nu^\star$, $\nu^\star = \nu + T + p_2 + 1$, p_2 is the number of stations with an observation at time t, and Q_{ij} $(i, j = 0, 1)$ are the appropriate submatrices of Q partitioned by missing (0) and observed (1) blocks of observations. Substituting for Q, the mode of this distribution is

$$\hat{\mathbf{Z}}_t^{(0)} = \left(\nu\mathcal{K}_{01} + n\hat{\Sigma}_{01}\right)\left(\nu\mathcal{K}_{11} + n\hat{\Sigma}_{11}\right)^{-1}\mathbf{Z}_t^{(1)}, \quad (2)$$

where $n\hat{\Sigma} = \sum_{s \neq t} \mathbf{Z}_s\mathbf{Z}_s^T$ and $\mathcal{K}$ and $\hat{\Sigma}$ are partitioned in the same way as Q given earlier. We take this posterior mode to be the point estimate of the infill observation. From a frequentist stand point, this can be identified as a ridge regression, and, letting ν get large, this mode converges to the usual best linear unbiased estimators, assuming the covariance matrix $\mathcal{K}$.

3.1 Nearest Neighbors

Notwithstanding the elegance of the model described earlier, it is impractical to compute the mean of the complete conditional distribution for this problem. The number of computations and storage grow as the square of the number of observed locations, and an exact solution is impractical. This problem is compounded by the long data record and irregular missing patterns. The use of this model by Le et al. (1999) took advantage of missing observations that can be easily partitioned into a block over time and space, but even so, these authors were able to obtain an exact analysis only because they considered a small number of spatial locations.

Here we propose a simple nearest-neighbor (NN) approach that provides an approximate solution. Given a particular station location to infill, we calculate the posterior mode in (2) by restricting to data from a small number of neighboring locations. Stein (1999) suggested that the best linear unbiased prediction at a point of interest, when using a stationary model, depends on local behavior of the random field far more than on behavior at locations far from the point of interest and is one justification for simplifying the infilling over an irregular lattice of time and space. Furthermore, this approximation allows us to reduce the single large spatial prediction problem to a set of many small problems limited only by the number of NN's. We emphasize that the NN strategy here is a computational approximation to the full Bayesian posterior mode. Our approximate solution is still grounded in a single model for the entire spatial field. The work by Haas (1990, 1995) used NN methods extensively, but from a different perspective. Haas essentially built a local spatial model for every prediction point and did not unify these models as being derived from a single spatial field.

The key to making the NN approach work for this problem was to devise a fast but effective rule for determining neighborhoods. Let $\mathcal{N}(\mathbf{x}_i, t)$ denote the neighborhood of locations to infill a station, $\mathbf{x}_i$, at time t. If $\mathbf{x}_j$, is a member of $\mathcal{N}(\mathbf{x}_i, t)$, then it must satisfy three conditions:

- Incidence. The station at $\mathbf{x}_j$ must have an observation at time t.
- Temporal overlap. The periods of record for $\mathbf{x}_j$ must have a sufficient number of observations that overlap in time with the record of $\mathbf{x}_i$.
- Proximity. Given a distance metric, $\mathbf{x}_j$ is close to $\mathbf{x}_i$.

Based on preliminary cross-validation analysis, we were led to use six NNs for each infill. Figure 2 shows how these three conditions generate neighborhoods for the Arco, ID station. The

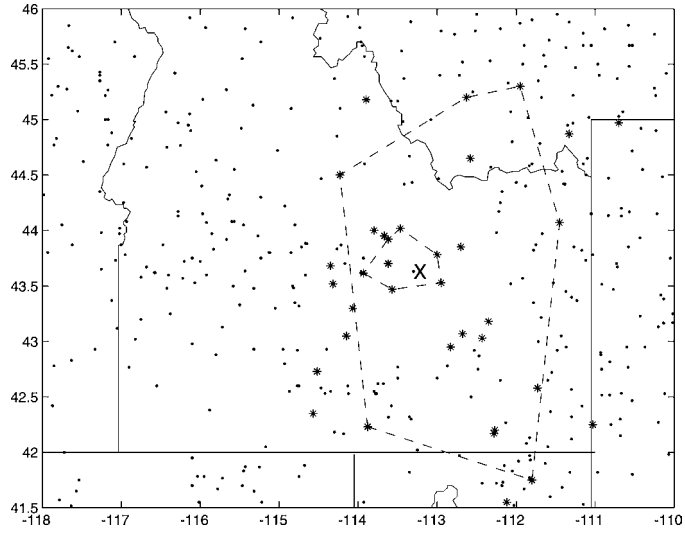

Figure 2. Neighborhoods for the Arco, ID Station. The station of interest is denoted by "×," stations used in infilling are denoted by "∗," and other stations are denoted by "·." Neighborhoods for the infilling of July 1897 and July 1995 are connected with dashed lines; the tighter neighborhood corresponds with July 1995.

incidence condition causes the neighborhood to be quite spread out at the 1895 time point because of the sparsity of operating stations early in the last century. Stations that were used at some point to do infilling at some point in the 103 years are denoted by "∗," whereas those never used as neighbors are denoted by "·." Those that are geographically close but never used as neighbors are eliminated by combinations of the temporal overlap, incidence, and proximity conditions.

Selection of a distance function is an important part of the process, and of course different applications may necessitate different distance measures. In our final analysis, we were led back to a simple form, based on the Bayesian prior correlation function, $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - k(\mathbf{x}_i, \mathbf{x}_j)^2$, where $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kriging correlation model described in Section 4. Although in general d may be different from geographical distance, the anisotropic covariance function used in this work tracks geographical separation.

Finally, we note that there is a trade-off between temporal overlap of neighbors and their proximity. For example, consider a station that has been moved 1 km at time t^* and relabeled as a new station. The old station record and the new station record will not overlap, and thus we have no estimate of covariance based on a sample statistic. However, to infill the old station record after time t^*, it is sensible to use the new station observations because of the close geographical proximity. A station farther away with a long overlapping record may not provide as good a prediction. Specifics of such trade-offs warrant further study.

3.2 Infilling

The concise form of the conditional distribution (1) given earlier depends on complete data for time periods other than t, the time point to be infilled. These data are not available in the precipitation record. Although it is possible to work out Bayes estimates with missing observations, this added complexity is not practical for this large problem. Little (1988) discussed the relative merits of various estimates of covariance when data are missing. For simplicity, we settled on using only time periods that were completely observed to estimate Σ as used in (2). The infilling proceeds as follows:

for each station (i)
  for each year to be infilled (t)
    Step 1. Find the neighborhood, $\mathcal{N}(\mathbf{x}_i, t)$.
    Step 2. Find the times (if any) of overlap for all stations in $\mathcal{N}(\mathbf{x}_i, t)$ and station i.
    Step 3. Use the common set of complete observations from step 2 to calculate the degrees of freedom, $\nu^\star(\mathbf{x}_i, t)$, mode (or median) anomaly, $\hat{z}(\mathbf{x}_i, t)$, and dispersion, $\xi(\mathbf{x}_i, t)$, of the conditional distribution (1). The infill is $\widehat{\mathbf{P}(\mathbf{x}, t)} = (\hat{\mathbf{z}}(\mathbf{x}, t)\sigma(\mathbf{x}) + \theta(\mathbf{x}))_+^2$.
  end (t)
end (i).

In Section 5.2 we show that the dispersion, $\xi(\mathbf{x}_i, t)$, is useful in standardizing the prediction errors.

3.3 Approximate Sampling From the Posterior

We note here that the algorithm for finding the posterior mode $\hat{z}(\mathbf{x}_i, t)$ can be reused to sample from the posterior. This justifies spending effort to develop a fast and accurate method of finding the posterior mode in (1). The benefit of a random sample (ensemble) from the posterior distribution is that variation among the members is a comprehensive representation of the uncertainty. At the simplest level, the sample mean across ensemble members approximates the posterior mean, and the ensemble standard deviation would be associated with the posterior standard deviation. Furthermore, ensembles also facilitate inference beyond pointwise prediction.

Let $\mathbf{U}^\star$ be a draw from a multivariate normal distribution with mean 0 and covariance $Q_{00} - Q_{01}Q_{11}^{-1}Q_{10}$, and let s^2 be a chi-squared random variable with $\nu^\star$ df. Then from the properties of the multivariate t, $\mathbf{Z}^\star = Q_{01}Q_{11}^{-1}\mathbf{Z}_t^{(1)} + \sqrt{\nu^\star}\mathbf{U}^\star/s$ is a random observation from the distribution (1). To generate $\mathbf{U}^\star$, first generate a multivariate normal, $\mathbf{U}$, with mean 0 and covariance Q. This random $\mathbf{U}$ corresponds with a complete set of observations. Next, partition $\mathbf{U} = (\mathbf{U}^{(0)}, \mathbf{U}^{(1)})$ into unobserved and observed, mimicking the data at time t, and determine the "infilled" values based on the posterior mode. Setting

$$\mathbf{U}^\star = \mathbf{U}^{(0)} - Q_{01}Q_{11}^{-1}\mathbf{U}^{(1)} \tag{3}$$

gives a random deviate with the desired attributes. We give a brief example of ensemble generation using a subset of the data in Section 5.

4. PRIOR PARAMETER AND MODEL SPECIFICATION

The infilling procedure hinges on specifying a reasonable prior distribution for $\mathbf{\Omega}$. Given the large and rich data record, we derive prior parameters for the inverse-Wishart distribution empirically, estimating the mean for $\mathbf{\Omega}$ using standard correlogram fitting. We determine the degrees of freedom, ν, using cross-validation and functional data analysis.

Recall that the prior distribution for the covariance matrix, $\mathbf{\Omega}$, is inverse-Wishart with degrees of freedom $\nu + p + 1$ and centering matrix $\nu\mathcal{K}$. The relationship between ν and $\mathcal{K}$ forces $E(\mathbf{\Omega}) = \mathcal{K}$. Because we are assuming that climatological standard deviations are known, we model just the correlation structure of the spatial field. We choose to use a stationary model for $\mathcal{K}$ for several reasons. First, the Bayesian structure will blend prior choice of $\mathcal{K}$ with the sample correlations and so result in a nonstationary model that can track observed local effects. Second, current models for nonstationary covariances are computationally intensive and would be overwhelmed by the size of the infill problem. Finally, we believe that the anomaly scale also helps homogenize the correlation structure (Fuentes et al. 1998). Although $\mathcal{K}$ may prescribe a stationary field in the anomaly scale, the implied precipitation field can be nonstationary.

Let $k(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\eta}) = \phi(\|\mathbf{A}(\mathbf{x}_i - \mathbf{x}_j)\|)$, where $\mathbf{A}$ is 2×2 matrix and ϕ is a member from the family of Matern covariances (Stein 1999). The parameters of this model were estimated using weighted least squares, more details were given by Cressie (1993). From extensive data analysis, we found that the model could be productively simplified with the off-diagonal elements in $\mathbf{A}$ set to 0. The scale parameters followed a smooth seasonal cycle, with a mean of 670 km in the east–west direction and 580 km in the north–south direction while the smoothness parameters varied from .58 to .82.

A key parameter in deriving the infilled estimate is ν, the degrees of freedom in the inverse-Wishart prior for the spatial

covariance. An idealized goal is to vary ν for each station location to give the best MSE predictions for infilling. As a practical approximation, we estimate the infill error as a function of ν at a subset of 400 stations and extrapolate these results to the full set of locations. This is done using a combination of smoothing techniques and principle components and provides a balance between using a single value for all stations and a highly variable estimate from single station estimates.

Let $R(\mathbf{x}_i, \nu)$ be the estimated MSE for the ith station using degrees of freedom $\nu + p + 1$ in the infilling estimate and where p is the number of NNs. This MSE is found by infilling anomalies for a station at times when data are observed and then calculating the average squared difference between observed and predicted anomalies. An important assumption here is that the cross-validation function estimated for times when data are observed is comparable to that when station values are missing. For 400 stations in a space-filling subset of the total data record, the MSE is found at a grid of 50 values of ν. Let $\mathbf{R}$ denote this $400\times$ 50 matrix of MSE's. Each row of $\mathbf{R}$ is a curve representing the MSE for a given station as a function of ν. Good values for ν will coincide with the minimums of the MSE curves. To stabilize the estimates, we smoothed these curves. Using singular value decomposition, we have $\mathbf{R} = \mathbf{UDV}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and $\mathbf{D}$ is diagonal with nonnegative values in decreasing order. The smoothing was done by spatial interpolation of the coefficient matrix. The columns of $\mathbf{U}$ are interpreted as weights of the columns of $\mathbf{V}$, which are interpreted as basis functions over ν. $\mathbf{R}$ can be approximated by using only the first three columns of $\mathbf{U}$ and $\mathbf{V}$ and truncating $\mathbf{D}$ to conform. To further smooth this representation over space, we fit (three) smooth surfaces to the first three columns of the coefficient matrix $\mathbf{U}$ using thin plate smoothing splines. The smoothed elements of $\mathbf{U}$ were interpolated to give a stable family of MSE curves for each station in the dataset. These interpolated MSE curves are then minimized to find the ν used at the infill step.

Most of the estimates of the degrees of freedom parameter were approximately 6, with the smallest being 1.5 and the largest being 17.5. Although each month is considered independently, the month-to-month transitions seem quite smooth and provide some support for the observed spatial pattern in ν.

5. RESULTS

The results of the infilling for precipitation (and other datasets) are available at *http://www.cgd.ucar.edu/stats/Data/US.monthly.met/*. The README_precip file available from the same page has an extensive list of the assumptions made in the analysis.

5.1 Computing Efficiency

Given the prior information regarding the means θ, standard deviations σ, stationary covariance model, and the collection of degrees of freedom parameters ν, the infilling step across the 11,918 stations and 103 time points for a month was implemented in Matlab and can be executed in a matter of hours on a Linux PC with CPU speed of 4.0 Ghz. This included infilling the observed values to do cross-validation. Timing tests suggest that the increase in execution time due to doubling the number of NN's to 12 increased the computation time by a factor of 1.20. Using 24 neighbors rather than 6 increased computation time by a factor of approximately 1.80. Cross-validation studies on the 400 stations of the space-filling subset show that the decrease in cross-validation sums of squares by using more than 6 neighbors is marginal and can be detrimental for some stations.

Finding neighborhoods is a significant portion of the computation. This comes primarily from calculating and sorting the geographical distances. Because distances depend on the anisotropic correlation model and the anisotropy is different for each month, little information about neighbors can be carried over from the infilling of the collection of Januarys to the infilling of the collection of Februarys. Furthermore, because neighborhoods depend on irregular observation patterns, each station can have a distinct neighborhood for each year that needs to be infilled, and thus neighborhood information cannot be easily shared across years. Finally, paucity of data in the first 20 years eliminates some of the benefit that could be gained by limiting the search for neighbors to small geographic regions.

5.2 Quantifying Infill Errors

Although the value of the infilling algorithm is for observations that are missing, the same procedure can be applied to all times. The comparison between the infilled prediction and the actual observation can be used to assess the accuracy of the infilling methods and also the adequacy of model-derived standard errors.

Figure 3 shows the standardized prediction standard error where the standardizing coefficient is the station standard deviation from Section 2.2. This represents the proportion of variation in the data (in the anomaly scale) not explained by the infilled values. Values near 0 represent near-perfect infilling, and a value of 1 implies that the infilling procedure is no better than using the station mean.

If the spatial model used for infilling was correct, then one might expect $(z(\mathbf{x}, t) - \hat{z}(\mathbf{x}, t))/\sqrt{\xi(\mathbf{x}, t)}$ to approximately follow a standard normal distribution. Based on cross-validation, we found the standardization useful for setting confidence intervals. Figure 4 shows the quantile-quantile plot of these standardized prediction residuals. Slightly more than the expected number of standardized residuals fall within the typical approximate 95% confidence intervals. In essence, a scientist who wishes to find an approximate 95% confidence interval for the infilled value of a particular station and time combination with the formula *infill* ± 2 *standard errors* will find that the interval estimate is conservative. Outside of 2 standard errors, the residuals are heavier-tailed relative to a Gaussian; this is discussed further in Section 6.

5.3 Comparison With Other Methods

The MSE curves for the special subset of 400 stations and the interpolated MSE curves as described in Section 4 were typically convex, with unique minima at levels of ν that were not endpoints of the testing grid. These unique minima at values $0 < \nu < \infty$ imply that the approximate Bayes method (AB) that we describe is more effective, in terms of prediction error, than either kriging (with the specified covariance function) or NNR methods (using only six neighbors).

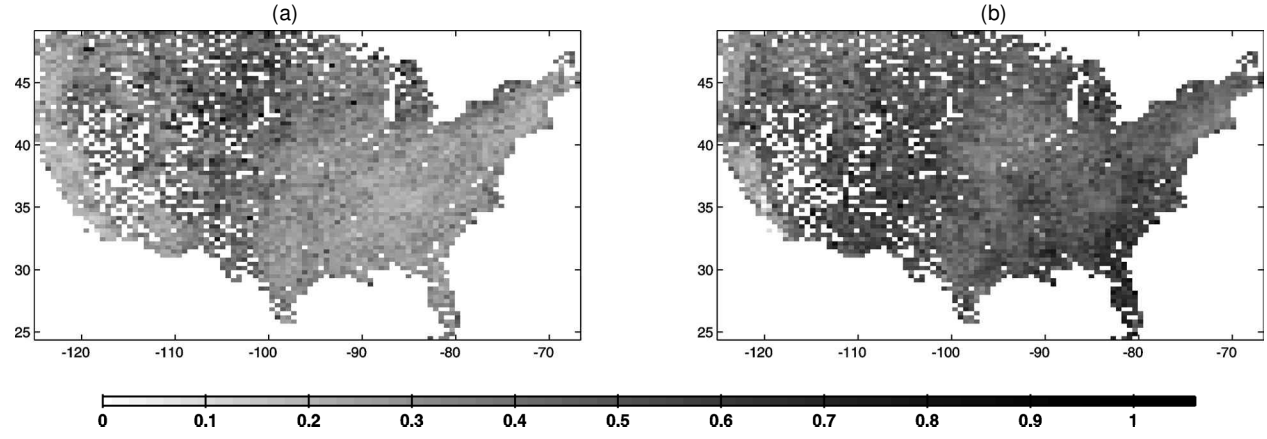

Figure 3. Square Root of the Standardized Prediction Error for (a) January and (b) July. Values near 0 correspond with infills that are near the observations; values near 1 suggest that using the climatological mean would be as effective as the infilling.

To calibrate this infilling method with other standard techniques in the statistical literature, we investigated the performance of a windowed kriging (WK) model on the subset of stations in Colorado. The topography of Colorado is sufficiently diverse to represent many of the potential problems that could be encountered over the continental United States. The local covariance model used in WK was a Matern correlation model fit to correlations estimated over time using a nonlinear least squares algorithm. The WK used a window with radius of 150 km. Qualitatively, there is little difference in the infilled values from the two methods (correlation coefficient = .99); however, infilling via the AB method was computationally four to five times faster. Part of the computational bottleneck in WK came from the need to use large windows to enable infilling of some of the early years. This slows down the nonlinear optimizer that estimates parameters for the correlation function. Both the WK and AB methods include approximate standard errors for the infilled values. Comparing the cross-validation standardized residuals from the two models, the standardized AB residuals more closely approximated a standard normal distribution than did those from WK.

In summary, we found that a WK approach is comparable to our approach but requires more computation.

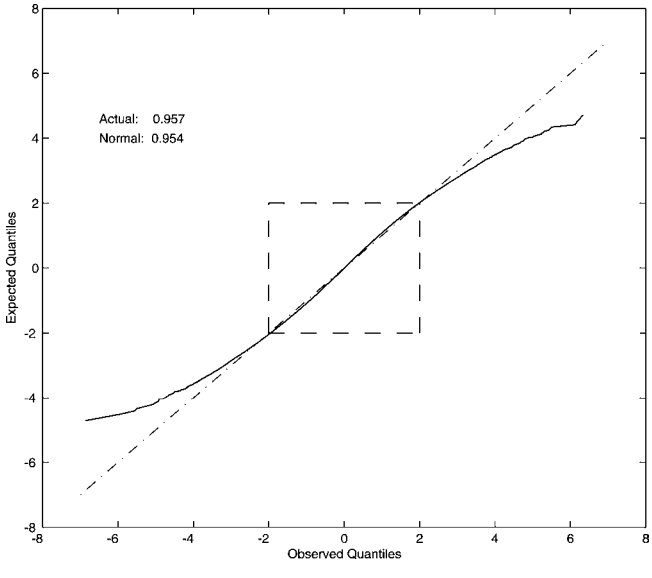

Figure 4. Q–Q Plot of the Cross-Validation Standardized Residuals (actual, .957; normal, .954). Approximately 500,000 residuals for July are depicted. The box shows a 96.4% Gaussian confidence level. Because 97.9% of the residuals fall in the region, intervals at typical confidence levels will be slightly conservative.

5.4 Sampling From the Posterior

In many applications it may be more appropriate to generate an ensemble of infilled observations or gridded fields that not only reflect the pointwise uncertainty, but also preserve the correlations among the locations. To showcase the utility of the infilling algorithm for sampling the posterior conditional distributions, we again consider the subregion centered on Colorado, an area with highly varied terrain and average precipitation that contains 391 stations. We emphasize that this method, although reported for a small portion of the United States, can be scaled to the full problem.

Referring to the algorithm outlined in Section 3.3, the random field $\mathbf{U}$ was generated from a multivariate normal with mean 0 and covariance Q. Here Q is a sum of the stationary correlation matrix and a nonstationary, short-range correlation matrix derived from station sample correlations. This second matrix involved some statistical estimates due to missing data. Specifically, for station pairs with less than 10 years of common data, the correlations were estimated using a thin plate spline using available correlations. Also, this correlation matrix, consisting of some estimated and sample quantities, was tapered to a range of approximately 70 miles. Finally, the tapered matrix was projected onto a nonnegative definite matrix to ensure that the resulting Q was indeed positive definite. (A total of 88% of the variability in the short-range, nonstationary covariance matrix is explained by the dominating 50 positive eigenvalues.)

Figure 5 depicts the posterior mean field and three ensemble members for April 1948. To simplify the statistical interpretation, the ensemble members are plotted in the anomaly scale. The cost for computing the (approximate) conditional field from the unconditional field, $\mathbf{U}^\star$ via (3), was negligible, because of the NN approach.

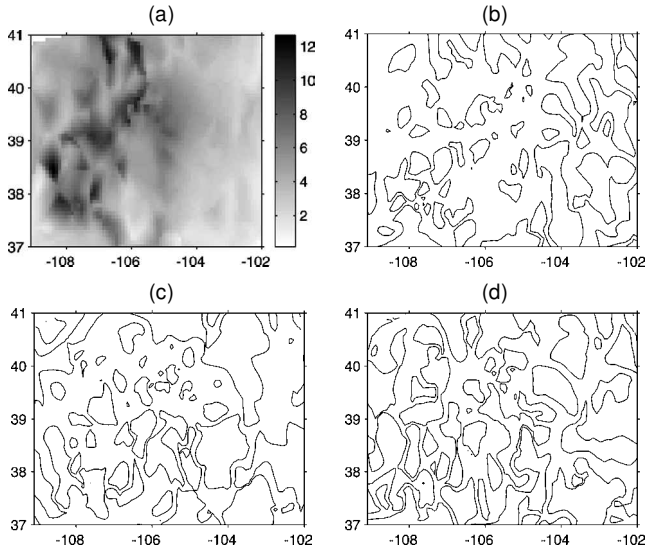

Figure 5. Posterior Samples. (a) The infilled mean for stations in Colorado in April 1948. (b, c, and d) Contours of random ensemble anomalies (deviations) conditioned on observations from April 1948. An ensemble member would be the pixel-wise sum of panel (a) with any of the other panels. The contour levels are at −.67 and .67 standard deviations.

6. DISCUSSION

This work has shown how local spatial methods combined with large-scale models can be effective in the prediction of nonstationary fields. The accuracy of the infilling, as measured by cross-validation, is high in areas that are data rich or spatially homogeneous. Larger MSE's are often found in regions of varied topography, such as the Rocky Mountain states, or regions that have frequent localized summer storms, such as the Florida peninsula. This poorer performance is expected, and it is not clear that more sophisticated spatial modeling will circumvent these physical limitations.

An important aspect of this work is the validity of the model-based standard errors when judged by cross-validation. A standard principle in spatial statistics is that although many estimators may yield comparable accuracy, their internal measures of uncertainty may vary widely. Here we have found that approximating the posterior distribution by a t distribution and fixing some of the parameters at estimated values yielded useful standard errors.

The potential for ensemble generation is particularly useful for numerical models, such as biogeochemical simulations, that require meteorological inputs. Separate runs based on the ensemble members facilitate error estimation of regional and intraregional effects. This is achieved by propagating the spread in the ensemble members to a subsequent spread in the individual model results.

Another concrete measure of the success of this methodology is the ability to process the 103 years of monthly precipitation records for the coterminous United States on a standard computing platform (e.g., Linux-based PC) and a high level language (Matlab). Typically, the final production run must be executed several times as models are refined and unusual cases are handled. Therefore, the time for production must be much shorter than the total time allocated for computing. There is a tendency in the atmospheric sciences to rely on lower-level languages (e.g., FORTRAN) for processing large datasets. Although this does become necessary at some size, we found that the benefits of flexibility in algorithm development in Matlab far exceeded the possible speedup by coding in a compiled language.

The heavy tails in the cross-validation residuals are likely due to departures from Gaussianity of the transformed data. It can also be partly attributed to modeling small amounts of precipitation with a continuous distribution. For very dry areas, it is not reasonable to assume that the anomalies will be normally distributed. Moreover, small mean levels also have associated small standard deviations (σ), and so small differences in the observed precipitation will be magnified on the anomaly scale. At some point the continuous model for monthly precipitation amount will break down, and a discrete model explicitly including zero incidence must be entertained. For stations where the precipitation values are a mixture of 0's and some positively valued random variable, conditioned on rainfall, we found that the choice of transformation is of little consequence in a practical setting. Typically, if a station has climatological mean near 0, then neighboring stations likewise will have means and observations near 0. On the positive side, these discrepancies for arid locations are small in an absolute scale and will have little effect in models used for studying such processes as vegetation.

There is simple way to check that our analysis reproduces the correct (non-Gaussian) distribution for precipitation. One simply observes the plots of the data versus the infills. These plots complement the information in Figure 3 and show strong agreement between infill and observed in most cases. For April, the correlation between infill and observed was greater than .9 for more than 10,500 of the 11,918 stations. We trace this infill robustness to two factors. First, the spatial prediction, being based only on second moments, is robust to moderate departures from normality, and in fact the square root transformation and standardization have done a good job of transforming the distributions closer to Gaussianity. We therefore expect the posterior means to be efficient. Second, because of the strong dependence among neighboring stations, the posterior mean will explain a large fraction of the variance, and the prediction weights favor the NNs. Any non-Gaussian distribution in the neighboring anomalies will be transferred to the infilled value, because this value is linear combination of a few neighboring values.

Another systematic error is misspecifying the climatology at a location to center the transformed data (θ). There are some cases where PRISM means deviate significantly from earlier historical data. Part of the difficulty in this case is to avoid introducing spurious temporal effects by using means from an early time period and transferring them to a later time period. We have taken the approach of adjusting the PRISM means when there are gross differences. Just as important, the infilled data product will include the climatological means and standard deviations, in anomaly scale, as part of its meta data, also allowing users to diagnose the problems of centering.

The subset selected for estimating prior parameters can also potentially bias the model. We chose a subset of stations to fill space because we wanted to include long-range correlations when estimating the global correlation function. We

could have used other schemes for selecting the subset, such as an "interest-filling subset" based on the regional variability of $C(\mathbf{x})$ (Sec. 2.2). For example, it makes sense to choose a more spatially dense set of stations in the Rocky Mountains than in the Great Plains region to help account for orographic effects. Regardless, it is not clear what changes to the final data product a different subset might produce because the method uses local information to the greatest extent possible. Although more study is merited, we suspect that the optimal subsets will be different depending on whether the optimality criteria is related to estimating the global spatial correlation function or to determining degrees of freedom parameters.

In this work we found that a simple distance-based criterion for selecting neighborhoods was effective. However, infilling could be improved using more geographic or climatological information. A classical example involves two mountain stations on different ridges that have common precipitation patterns but are not considered neighbors because valley stations are geographically closer. The PRISM model has a sophisticated scheme for associating neighborhoods based on such covariate information that could be incorporated into the neighborhood selection criteria.

A delicate balance must be struck when finding neighborhoods, especially when there are many missing values. On one hand, a large set of overlapping observations is good for reducing prediction variance in the regression setting. On the other hand, having neighbors that are close geographically is also useful for reducing prediction variance, because correlation between two stations is strictly decreasing as the separation distance increases. The nuances of this trade-off merit more study.

An obvious extension to this analysis is to incorporate a more flexible covariance model in the prior. One compromise is to consider a slowly varying parametric model that approximates the variable convolution ideas of Higdon et al. (1999). Ideally, components of the PRISM system, including dependence of the covariance on elevation and aspect (the direction a slope faces), could be matched in this model. A more realistic covariance model would also facilitate interpolating the field to grid locations off the observation network.

For precipitation, it was not productive to model temporal dependence, but for other variables, this may be an important component. A simple approach that builds on these models is to model the field as an autoregressive process in time with innovations that have spatial correlation. Such a model has been useful in capturing the space/time structure of wind fields and filling in sparse measurements (Wikle et al. 2001). Given that $z(\mathbf{x}, t)$ are transformed and standardized meteorological anomalies, one would start with the first-order autoregression,

$$z(\mathbf{x}, t) = a(\mathbf{x})z(\mathbf{x}, t-1) + u(\mathbf{x}, t),$$

where $a(\mathbf{x})$ are autoregressive parameters varying in space and possibly over season and $u(\mathbf{x}, t)$ are Gaussian fields that are independent in time. Given the autoregressive model, the spatial prediction now involves infilling the shock $u(\mathbf{x}, t)$ for missing times and locations. This operation will be computationally similar to the one presented earlier but will inherit some added uncertainty from estimating the autoregressive parameter surface. Of course, there is the added difficulty inherent in time series models of specifying the initial time vector ($z(\mathbf{x}, 1)$), and this problem is compounded by the sparsity of data in the earliest period.

In closing, we have produced a useful analysis of the U.S. precipitation record. However, additions to our research suggest a fertile area of statistical research.

[Received November 2001. Revised September 2003.]

REFERENCES

Brown, P. J., Le, N. D., and Zidek, J. V. (1994), "Multivariate Spatial Interpolation and Exposure to Air Pollutants," *Canadian Journal of Statistics*, 22, 489–510.

Cressie, N. C. (1993), *Statistics for Spatial Data* (2nd ed.), New York: Wiley.

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., and Pasteris, P. (2002), "A Knowledge-Based Approach to the Statistical Mapping of Climate," *Climate Research*, 22, 99–113.

Daly, C., Neilson, R. P., and Phillips, D. L. (1994), "A Statistical-Topographic Model for Mapping Climatological Precipitation Over Mountainous Terrain," *Journal of Applied Meteorology*, 33, 140–158.

Daly, C., Taylor, G. H., Gibson, W. P., Parzybok, T. W., Johnson, G. L., and Pasteris, P. (2001), "High-Quality Spatial Climate Data Sets for the United States and Beyond," *Transactions of the American Society of Agricultural Engineers*, 43, 1957–1962.

Fuentes, M., Kelly, R., Kittel, T. G. F., and Nychka, D. (1998), "Spatial Prediction of Climate Fields for Ecological Models," technical report, National Center for Atmospheric Research, Boulder, CO.

Gibson, W. P., Daly, C., and Taylor, G. H. (1997), "Derivation of Facet Grids for Use With the PRISM Model," in *Proceedings of the 10th American Meteorological Society Conference on Applied Climatology*, pp. 208–209.

Haas, T. C. (1990), "Lognormal and Moving Window Methods of Estimating Acid Deposition," *Journal of the American Statistical Association*, 85, 950–963.

——— (1995), "Local Prediction of a Spatio-Temporal Process With an Application to Wet Sulfate Deposition," *Journal of the American Statistical Association*, 90, 1189–1199.

Higdon, D., Swall, K., and Kern, J. (1999), "Non-Stationary Spatial Modeling," in *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press.

Karl, T. R., Williams, C. N., Jr., Quinlan, F. T., and Boden, T. A. (1990), *United States Historical Climate Network (HCN) Serial Temperature and Precipitation Data*, Environmental Science Division Publication No. 3404, Oak Ridge, TN: Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory.

Kittel, T. G. F., Rosenbloom, N. A., Kaufman, C., Royle, J. A., Daly, C., Fisher, H. H., Gibson, W. P., Aulenbach, S., Yates, D. N., McKeown, R., Schimel, D. S., and VEMAP2 Participants (2000), "VEMAP Phase 2 Historical and Future Scenario Climate Database for the Conterminous U.S.," available online from the ORNL Distributed Active Archive Center, Oak Ridge National Laboratory, *http://www.daac.ornl.gov/*.

Kittel, T. G. F., Royle, J. A., Daly, C., Rosenbloom, N. A., Gibson, W. P., Fisher, H. H., Schimel, D. S., Berliner, L. M., and VEMAP2 Participants (1997), "A Gridded Historical (1895–1993) Bioclimate Dataset for the Conterminous United States," in *Proceedings of the 10th American Meteorological Society Conference on Applied Climatology*, pp. 219–222.

Kittel, T. G. F., Thornton, P. E., Royle, J. A., and Chase, T. N. (2002), "Climates of the Rocky Mountains: Historical and Future Patterns," in *Rocky Mountain Futures: An Ecological Perspective*, Covelo, CA: Island Press, pp. 59–82.

Le, N. D., Sun, L., and Zidek, J. V. (1999), "Bayesian Spatial Interpolations and Backcasting Using Gaussian–Generalized Inverted Wishart Model," Technical Report 185, University of British Columbia, Statistics Dept., *http://www.stat.ubc.ca/research/tr99.html*.

Legates, R., and Willmott, C. J. (1990), "Mean Seasonal and Spatial Variability in Global Surface Air Temperature," *Theoretical and Applied Climatology*, 41, 11–21.

Little, R. J. A. (1988), "Robust Estimation of the Mean and Covariance Matrix From Data With Missing Values," *Applied Statistics*, 37, 23–38.

NAST (National Assessment Synthesis Team) (2001), *Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change* (report for the U.S. Global Change Research Program), Cambridge, U.K.: Cambridge University Press.

Nychka, D. (2000), "Spatial Process Estimates as Smoothers," in *Smoothing and Regression: Approaches, Computation and Application*, ed. M. G. Schimek, New York: Wiley, pp. 393–424.

Nychka, D., Meiring, W., Fuentes, M., and Royle, J. A. (2000), "FIELDS 5.2 Statistical Tools for Spatial Data," available at *http://www.cgd.ucar.edu/stats/software*.

Parton, W. J. (1996), "The CENTURY Model," in *Evaluation of Soil Organic Matter Models* (NATO ASI Series 138), eds. D. S. Powlson, P. Smith, and J. U. Smith, Berlin: Springer-Verlag, Reprint 759, p. 283–291.

Rajagopalan, B., and Lall, U. (1998), "Locally Weighted Polynomial Estimation of Spatial Precipitation," *Journal of Geographic Information and Decision Analysis*, 2, 48–57.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.

Schimel, D., Kittel, T. G. F., Running, S., Monson, R., Turnipseed, A., and Anderson, D. (2002), "Carbon Sequestration in the Mountains of the Western U.S.," *Eos, Transactions of the American Geophysical Union*, 83, 445–449.

Schimel, D., Melillo, J., Tian, H., McGuire, A. D., Kicklighter, D., Kittel, T., Rosenbloom, N., Running, S., Thornton, P., Ojima, D., Parton, W., Kelly, R., Sykes, M., Neilson, R., and Rizzo, B. (2000), "Contribution of Increasing CO_2 and Climate to Carbon Storage by Ecosystems of the United States," *Science*, 287, 2004–2006.

Shepard, D. (1968), "A Two-Dimensional Interpolation Function for Irregularly Spaced Data," in *Proceedings of the Twenty-Third ACM Conference*, pp. 517–524.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag.

USDA-NRCS (principal investigator, C. Daly) (1998), "PRISM Climate Mapping Project Precipitation," mean monthly and annual precipitation digital files for the continental U.S., USDA NRCS National Cartography and Geospatial Center, Ft. Worth, TX, CD-ROM.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), "Spatiotemporal Hierarchical Baysian Modeling: Tropical Ocean Surface Winds," *Journal of the American Statistical Association*, 96, 382–397.

Wood, A. T. A., and Chan, G. (1994), "Simulation of Stationary Gaussian Processes in $[0, 1]^d$," *Journal of Computational and Graphical Statistics*, 3, 409–432.