Infilling Sparse Records of Spatial Fields

Craig J. Johns Douglas Nychka Timothy G.F. Kittel Chris Daly *

September 15, 2003

Abstract

Historical records of weather such as monthly precipitation and temperatures from the last century are an invaluable database to study 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 irregularly observed 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 to 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 the overall approach has merit, treading a line along computational feasibility and statistical validity. Furthermore, we are able to produce reliable measures of uncertainty for the estimates.
Keywords: Bayesian Spatial Interpolation, Cross-validation, Prediction, Geostatistics.

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

*Craig J. Johns is Assistant Professor, Mathematics Department, University of Colorado, Denver, CO, 80217-3364, Douglas Nychka is Director, Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO, Timothy G.F. Kittel is Scientist, Terrestrial Sciences Section, National Center for Atmospheric Science, Boulder, CO, and Chris Daly is Director, Spatial Climate Analysis Service, Department of Geosciences, Oregon State University, Corvallis, OR. The authors gratefully acknowledge support from the National Science Foundation under Grants DMS 9815344 and DMS 9312686 for the Geophysical Statistics Project and its research and support from NOAA grant NA76GP0558 . VEMAP was supported by NASA, EPRI, USDA Forest Service, and the US Deptartment of Energy.

changing climate on society. Historical records of weather such as monthly temperatures and precipitation from the last century are an invaluable database to study 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. However, 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 $\times$'s to indicate missing months. In this case there are 67 years of data observed in the 78 year time period 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 US is the National Assessment (NAST, 2001) produced for the US Global Change Research Program. This study was mandated by an act of the US Congress to report on the impacts of climate change on the US. The assessment's conclusions (page 537, items 8 and 10) highlight the need for improved long term data sets and methods that address uncertainty in such data. An infilled record of historical monthly meteorology for the US 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, carrying all the caveats 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), are tested by their ability to reproduce the past climate. AOGCMs 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 to evaluate model results.

A more intrinsic need for completed observational data is as the input to ecological and biogeochemical models. In order 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 (`http://www.nrel.colostate.edu/projects/century5/`) is a general model of plant-soil-nutrient cycling that is 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 are to run 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 have confidence in analysis and modeling of these relationships (e.g., Schimel et al. 2000, Kittel et al. 1997).

1.2 Creating Complete "Data Sets"

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. While these features demand careful statistical modeling, one is limited in sophistication by the large size of typical geophysical datasets. Thus, the challenge and focus of this work is to find a balance between adequate statistical models and efficient methods that allow for processing the data in a reasonable amount of time, provide efficient estimators and companion measures of uncertainty. Although this work deals with monthly precipitation, this 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 a variety of natural resource, agricultural, and hydrological modeling and assessment activities (USDA-NRCS 1998; Daly et al 2001). However, these data sets are climatological means only (typically means from the time period 1961-90), and do not adequately reflect the rich structure of climate variability and trends contributing to these means.

An earlier investigation (Kittel et al. 1997, 2000) generated a complete spatio-temporal climate data set for the coterminous US . However, they used fewer stations and a simpler infill method than presented 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 et al 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 because a significant portion of carbon sequestration and seasonal water storage occurs at higher elevations (Kittel et al. 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 data sets require nontrivial preprocessing, quality control and analysis to be useful for further scientific investigation. We believe that statisticians can contribute to these efforts by not only 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 not only to provide good predictions of precipitation, but also 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 that credible intervals derived from the posterior distribution would 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 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 and details are found in (Daly et al.

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 methods (NNR) and geostatistical models (GS). 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 which are close to the locations of interest. Infilled values are typically weighted averages of the observations in the neighborhood. The weights are computed by a variety of 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, they 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. Non-parametric methods for estimating covariances are also available, (Sampson and Guttorp, 1995, Higdon et al. 1999), however, these approaches are infeasible for data sets with large numbers of station locations.

This analysis combines an Empirical Bayes implementation of a model similar to that in Brown et al. (1994) mixed with the neighborhood ideas in 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. While 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 with suitable tuning such methods may do as well as spatial statistical approaches. The disadvantage of these methods is that it is difficult to derive reliable measures of the estimator's uncertainty. A hybrid estimator is a thin plate spline being interpretable as both a numerical interpolation method and a spatial statistics estimator (Nychka 2000). Although we use thin plate splines in some instances where it is difficult to estimate a full covariance function, we felt 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 (Section 2). Subsequent spatial modeling and prediction is applied 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. Details of the nearest neighbor criteria we used are given in Section 3.1 and the infill algorithm is outlined in Section 3.2. In Section 4 we discuss specification of the model and prior parameters specific to the analysis of the precipitation data. Results and discussion are in Sections 5 and 6 respectively.

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 to 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 a quality control system there 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 are reporting whereas January 1964 has 7921 recorded observations. Overall, there is a rapid growth in the size of the observation network in the 1940's with the total number of reporting stations reaching a maximum in the early 1960's. 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 only 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 data set.

Eliminating these station records improved the overall quality of the data product because inclusion of these stations would have changed the overall infill percentage (data in the final product that are estimated, rather than observed) to 69%, rather than 57%. Furthermore, these stations would not be used as neighbors in the infill step (see Sections 3.2 and 3.3) because of the short time record and 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 data base 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 0.14, 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 (≈ 0.1), were present but *ad hoc* analysis showed that spatio-temporal 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 zero, 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 to spatial correlation, we developed the methodology for infilling missing observations using only observations which are contemporaneous in the month of interest. That is, to infill a particular station record for any July in the years 1895-1997, we only use the records from the Julys in those years, but not any June or August information. For completeness we suggest an extension useful for correlated data in the discussion of Section 6.

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 are referred to in atmospheric science as the climatology and could 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 *et al.* (2002) and Gibson *et al.* (1997) describe 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 data sets. 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 the spatial infilling is done after a transformation and standardization of the raw observations. We will refer to the resulting values as *anomalies* and unless otherwise specified the statistical analysis will be in the anomaly scale. The spatial anomaly field has the advantages that it is closer to a Gaussian distribution and has 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, there is evidence that suggests that the anomaly field is also closer to being second order stationary compared to the raw scale (Fuentes *et al.* 1998). It is well known that precipitation amounts can be approximated by a Gamma distribution in many cases. In this situation, the square root function is quite efficient in transforming gamma distributions to distributions that are approximately normal and so the square root transformation was used for all locations.

We again emphasize the independence of the monthly data sets and so dependence on the choice of month (Jan, ..., Dec) will be suppressed in the formulae 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 $\sigma^2(\mathbf{x}) = Var[\sqrt{P(\mathbf{x}, t)}]$. 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 give 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 $\hat{\mu}$ 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 above 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. The bandwidths were determined by minimizing the mean squared error 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 *et al.* (1999) describe 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 $[\ \mathbf{Z}_t \mid \Omega\ ] \sim N(0, \Omega)$ are independent (conditional on Ω) normally distributed p-dimensional random variables with mean zero and covariance Ω, for $t = 1, \ldots T$ and 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 Ω 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 $E(\Omega) = \mathcal{K}$. 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. (Section 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's 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, $[\ \mathbf{Z}_t \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_{t-1}, \mathbf{Z}_{t+1}, \ldots \mathbf{Z}_T\ ] \sim t_p\,(\nu_c,\ 0,\ Q/\nu_c)$ 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:

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

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 in for Q, the mode of this distribution is

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

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 above. We take this posterior mode to be the point estimate of the infill observation. From a frequentist point of view 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 above, 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) takes advantage of missing observations that can be easily partitioned into a block over time and space, but even so they are only able to obtain an exact analysis because they consider 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) suggests 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 only limited by the number of nearest neighbors. 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) uses nearest neighbor methods extensively but from a different perspective. Haas essentially builds a local spatial model for every prediction point and does 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 $\mathcal{N}(\mathbf{x}_i, t)$, 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 lead to use 6 nearest neighbors for each infill. Figure 2 shows how these three conditions generate neighborhoods for the Arco, ID station. The *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 which are used at some point to do infilling at some point in the 103 years are denoted with a '$*$', while those never used as neighbors are denoted by '$\cdot$'. Those which 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 therefore 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 above 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) discusses 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 neighborhood, $\mathcal{N}(\mathbf{x}_i, t)$

Step 2. Find 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 to find the posterior mode $\hat{z}(\mathbf{x}_i, t)$, can be reused to sample from the posterior. This justifies spending effort on 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 zero and covariance $Q_{00} - Q_{01}Q_{11}^{-1}Q_{10}$ and s^2 a χ^2 random variable with $\nu^\star$ degrees of freedom. 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 zero and covariance Q. This random $\mathbf{U}$ corresponds with a complete set of observations. Next, partition $\mathbf{U} = \left(\mathbf{U}^{(0)}, \mathbf{U}^{(1)}\right)$ 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 Ω. Given the large and rich data record, we derive prior parameters for the Inverse-Wishart distribution empirically, estimating the mean for Ω 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, Ω, is inverse-Wishart with degrees of freedom $\nu + p + 1$ and centering matrix $\nu\mathcal{K}$. The relationship between ν and $\mathcal{K}$ forces $E(\Omega) = \mathcal{K}$. Because we are assuming that climatological standard deviations are known, we just model 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. Secondly, 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 in homogenizing 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 $k(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\eta}) = \phi(||A(\mathbf{x}_i - \mathbf{x}_j)||)$ where 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 and more details can be found in Cressie (1993). From extensive data analysis we found that the model could be productively simplified with the off-diagonal elements in A set to zero. The scale parameters followed a smooth seasonal cycle with a mean of 670 km in the E-W direction and 580 km in the N-S 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 mean squared error 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 mean squared error (MSE) for the i^{th} station using degrees of

freedom $\nu+p+1$ in the infilling estimate and where p is the number of nearest neighbors. 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 CV 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 400×50 matrix of MSEs. Each row of $\mathbf{R}$ is a curve representing the mean squared error for a given station as a function of ν. Good values for ν will coincide with the minimums of the MSE curves. In order to stabilize the estimates, we smoothed these curves. Using singular value decomposition, $R = UDV^T$ where U and V are orthogonal and D is diagonal with non-negative values in decreasing order.The smoothing was done by spatial interpolation of the coefficient matrix The columns of U are interpreted as weights of the columns of V, which are interpreted as basis functions over ν. $\mathbf{R}$ can be approximated by using only the first three columns of U, and V and truncating D to conform. To further smooth this representation over space, we fit (three) smooth surfaces to the first three columns of the coefficient matrix U using thin plate smoothing splines. The smoothed elements of U were interpolated in order to give a stable family of MSE curves for each station in the data set. 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 data sets) are available via the internet 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 in order to do cross-validation. Timing tests suggest that the increase in execution time due to doubling the number of nearest neighbors 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 for some stations can be detrimental.

Finding neighborhoods is a significant portion of the computation. This primarily comes from calculating and sorting the geographical distances. Since 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, since neighborhoods depend on irregular observation patterns, each station can have a distinct neighborhood for each year that needs to be infilled, 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 had 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 zero represent near perfect infilling and a value of one implies that the infilling procedure is no better than using the station mean.

If the spatial model used for infilling was correct, one might expect that $(z(\mathbf{x},t)-\hat{z}(\mathbf{x},t))/\sqrt{\xi(\mathbf{x},t)}$ approximately follows a standard normal distribution. Based on cross-validation we found the standardization to be 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 the residuals are heavier tailed relative to a Gaussian and this will be 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 which were not endpoints of the testing grid. These unique minima at values between $0 < \nu < \infty$ imply that the approximate Bayes method (AB) we describe is more effective, in terms of prediction error, than either kriging (with the specified covariance function) or nearest-neighbor regression methods (using only 6 neighbors).

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 which could be encountered over the continental US. The local covariance model used in WK was a Matern correlation model and was fit to correlations estimated over time using a nonlinear least squares algorithm. The WK used a window with radius 150 km. Qualitatively, there is little difference in the infilled values from the two methods (correlation coefficient = 0.99), however infilling via the AB method was computationally 4-5 times faster. Part of the computational bottleneck in WK came from the need to use large windows in in order for some of the early years to be infilled. 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.

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 consider again the subregion centered on Colorado, an area with highly varied terrain and average precipitation that contains 391 stations. We emphasize that this method,

while reported for a small portion of the US, 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 zero 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 insure that the resulting Q was indeed positive definite. (88% of the variability in the short range, nonstationary covariance matrix is explained by the dominating 50 positive eigenvalues.)

The posterior mean field and three ensemble members for April 1948 are depicted in Figure 5. 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 one, $\mathbf{U}^\star$ via (3), was negligible because of the nearest neighbor approach.

6 DISCUSSION

This work has shown how local spatial methods combined with large scale models can be effective for 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 mean squared errors are often in regions of varied topography, such as the Rocky Mountain states, or regions that often have 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 while many estimators may give 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 intra-regional effects. This is achieved by prop-

agating 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 US 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 data sets. While this becomes necessary at some size, we found that the benefits of flexibility in algorithm development in Matlab far exceeded possible speedup by coding in a compiled language.

The heavy tails in the CV residuals are likely due departures from Gaussianity of the transformed data. It can also be attributed partly to modeling small amounts of precipitation with a continuous distribution. For areas that are very dry it is not reasonable to assume 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 zeros 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 that is near zero, then neighboring stations will likewise have means and observations which are near zero. 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 processes such as vegetation.

There is is simple way to check that our analysis reproduces the correct (nonGaussian) distribution for precipitation. One simply observes the plots of the data vs. the infills. These plots compliment 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 0.9 for more that 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 nearest neighbors. Any nonGaussian distribution in the neighboring anomalies with be transfered to the infilled value, as it is linear combination of a few neighboring values.

Another systematic error is mis-specifying 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 period in time and transferring them to a later 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 and so the users can also diagnose the problems of centering.

The selection of the subset used 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. Other schemes for selecting the subset, such as an "interest-filling subset" based on the regional variability of $C(\mathbf{x})$ (Section 2.2), could have been used. 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 a different subset might have on the final data product, as the method uses local information to the extent possible. Though 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 determining degrees of freedom parameters.

In this work we found that a simple distance based criterion for selecting neighborhoods was effective. However, the infilling could be improved using more geographic or climatological information. A classical example are 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 regime for associating neighborhoods based on such covariate information. This scheme could be incorporated into the neighborhood selection criteria.

There is a delicate balance that must be struck when finding neighborhoods, especially when there are many missing values. A large set of overlapping observations is good for reducing prediction variance in the regression setting. On the other hand, having neighbors which are close geographically is also useful for reducing prediction variance, since 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 could be matched in this model including dependence of the covariance on elevation and aspect (the direction a slope faces). 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 off of 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 auto-regression

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

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 above, but will inherit some added uncertainty due to estimating the autoregressive parameter surface. Of course, there is the added difficulty inherent with time series model 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 US precipitation record. However, additions to our research suggest a fertile area of statistical research.

Acknowledgments

The authors are indebted to Nan Rosenbloom (NCAR) and Alan McNab (NOAA/NCDC) and Wayne Gibson (Oregon State University) for their efforts in compiling and managing climate station datasets used in this study and to Richard Jones (University of Colorado, Health Sciences Center) for useful discussion early in the project. The authors also thank the Editor, Associate Editor and two referees for suggestions that refined and improved the paper.

References

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

Chan, G. and A.T.A. Wood, (1997). An Algorithm for Simulating Stationary Gaussian Random Fields. *A*pplied Statistics, **46**, 171-181.

Cressie, N.C., 1993, *Statistics for Spatial Data,* 2nd Ed., Wiley, New York.

Daly, C., W. P. Gibson, G.H. Taylor, G. L. Johnson, P. Pasteris. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research*, **22**, pp. 99-113.

Daly, C., R.P. Neilson, and D.L. Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* **33**, pp 140-158.

Daly, C., G.H. Taylor, W. P. Gibson, T.W. Parzybok, G. L. Johnson, P. Pasteris. 2001. High-quality spatial climate data sets for the United States and beyond. *Transactions of the American Society of Agricultural Engineers*, **43**, pp. 1957-1962.

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

Gibson, W. P., C. Daly, and G. H. Taylor, 1997. Derivation of Facet Grids for Use with the PRISM Model. In *Proceedings, 10^{th} AMS Conference on Applied Climatology, American Meteorological Society,* Reno, NV, October 20-23, 1997, pp 208-209.

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

Haas, T.C., 1995, Local Prediction of a Spatio-Temporal Process with an Application to Wet Sulfate Deposition, *Journal of the American Statistical Association*, Vol. **90**, pp. 1189-1199.

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

Karl, T.R., C.N. Williams, Jr., F.T. Quinlan, and T.A. Boden, 1990. *United States Historical Climate Network (HCN) Serial Temperature and Precipitation Data*, Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

Kittel, T.G.F., N.A. Rosenbloom, C. Kaufman, J.A. Royle, C. Daly, H.H. Fisher, W.P. Gibson, S. Aulenbach, D.N. Yates, R. McKeown, D.S. Schimel, and VEMAP2 Participants. 2000. VEMAP Phase 2 Historical and Future Scenario Climate Database for the Conterminous US. Available

online from the ORNL Distributed Active Archive Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee http://www.daac.ornl.gov/

Kittel, T.G.F., J.A. Royle, C. Daly, N.A. Rosenbloom, W.P. Gibson, H.H. Fisher, D.S. Schimel, L.M. Berliner, and VEMAP2 Participants. 1997. A gridded historical (1895-1993) bioclimate dataset for the conterminous United States. Pages 219-222, in: *Proceedings of the 10th Conference on Applied Climatology,* 20-24 October 1997, Reno, NV. American Meteorological Society, Boston.

Kittel, T.G.F., P.E. Thornton, J.A. Royle, and T.N. Chase. 2002. Climates of the Rocky Mountains: Historical and Future Patterns. Pp. 59-82 (Chapter 4), in: J.S. Baron (ed.). Rocky Mountain Futures: An Ecological Perspective. Island Press, Covelo, CA. 325 p.

Le, N. D., L. Sun, and J.V. Zidek, 1999, Bayesian Spatial Interpolations and Backcasting using Gaussian-Generalized Inverted Wishart Model, Technical Report 185, Statistics Department, University of British Columbia, http://www.stat.ubc.ca/research/tr99.html.

Legates, R. and C.J. Willmott, 1990, Mean Seasonal and Spatial Variability in Global Surface Air Temperature, *Theoretical and Applied Climatology*, **41**, pp 11-21, Springer-Verlag.

Little, R.J.A., 1988. "Robust Estimation of the Mean and Covariance Matrix from Data with Missing Values," Applied Statistics, **37**, No. 1, pp. 23-38.

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

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

Nychka, D., W. Meiring, M. Fuentes, and J. A. Royle, (2000) FIELDS 5.2 statistical tools for spatial data. http://www.cgd.ucar.edu/stats/software

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

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

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

Schimel, D., T.G.F. Kittel, S. Running, R. Monson, A. Turnipseed, and D. Anderson. 2002. Carbon sequestration in the mountains of the Western US. *Eos*, Transactions, American Geophysical Union,

83(40): 445–449.

Schimel, D., J. Melillo, H. Tian, A.D. McGuire, D. Kicklighter, T. Kittel, N. Rosenbloom, S. Running, P. Thornton, D. Ojima, W. Parton, R. Kelly, M. Sykes, R. Neilson, and B. Rizzo. 2000. Contribution of increasing CO2 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-Space Data. *Proceedings of the Twenty-Third ACM Conference,* pp 517-524.

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

USDA-NRCS (C. Daly, principle investigator). 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. December, CD-ROM.

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

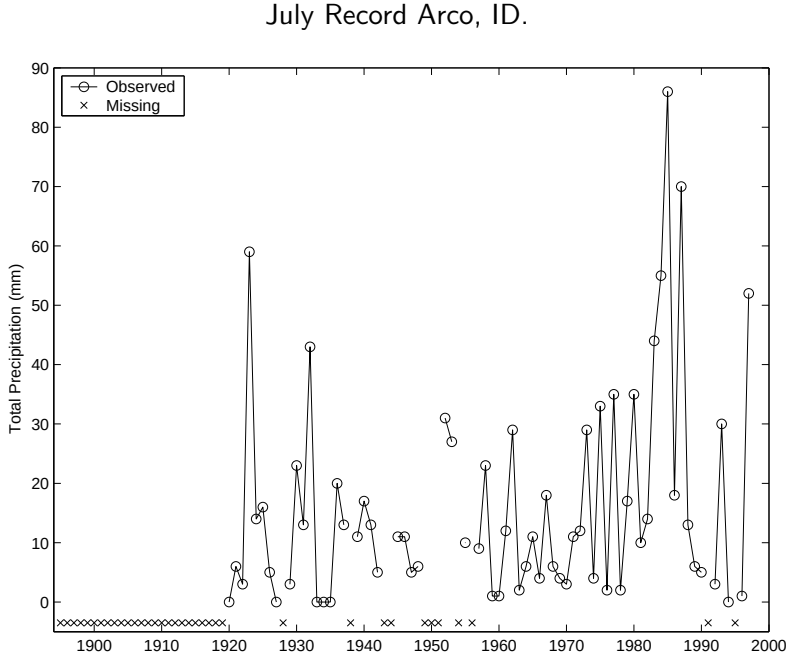

Figure 1: The record of July precipitation (mm) for the Arco, ID Station (100375) in south-central Idaho. The × denotes years where the record is missing.

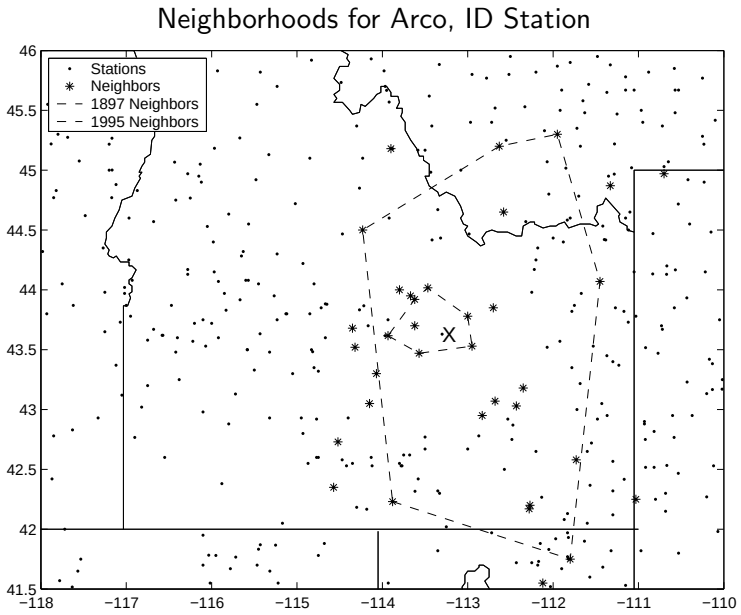

Figure 2: The station of interest is denoted by '×', stations used in infilling are denoted with '*', and other stations are denoted with by '·'. Neighborhoods for the infilling of July 1897 and July 1995 are connected with dashed lines; the tighter neighborhood corresponds with July 1995.

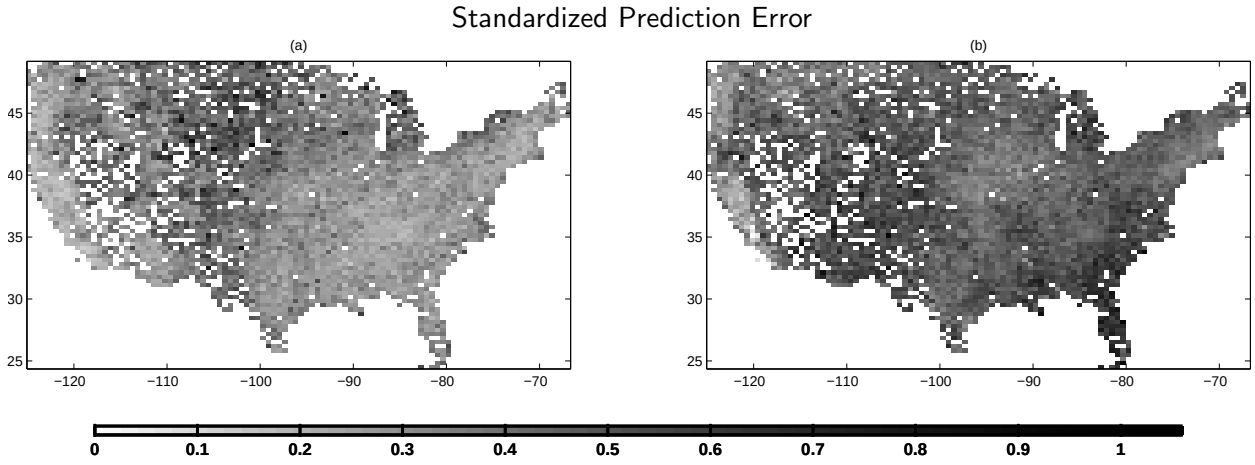

Figure 3: Square root of the standardized prediction error for (a) January and (b) July. Values near zero correspond with infills that are near the observations, while values near one suggest that using the climatological mean would be as effective as the infilling.

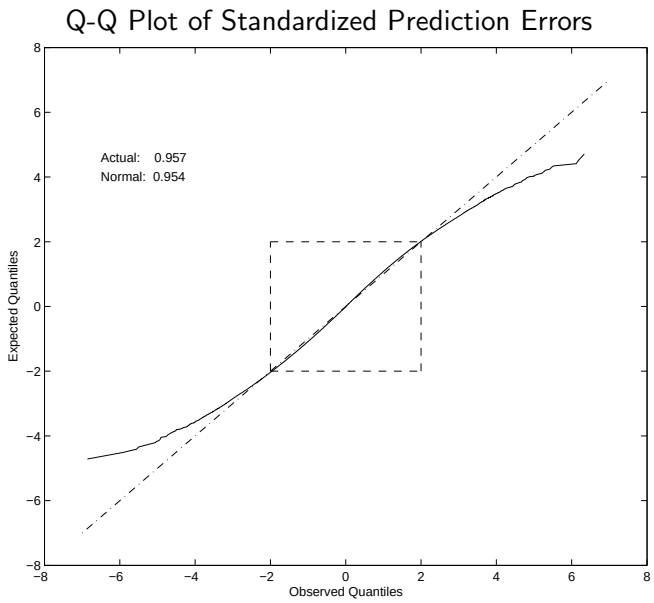

Figure 4: Q-Q plot of the cross-validation standardized residuals. Approximately 500,000 residuals for July are depicted. The box shows a 96.4% Gaussian confidence level. Since 97.9% of the residuals fall in the region, intervals at typical confidence levels will be slightly conservative.

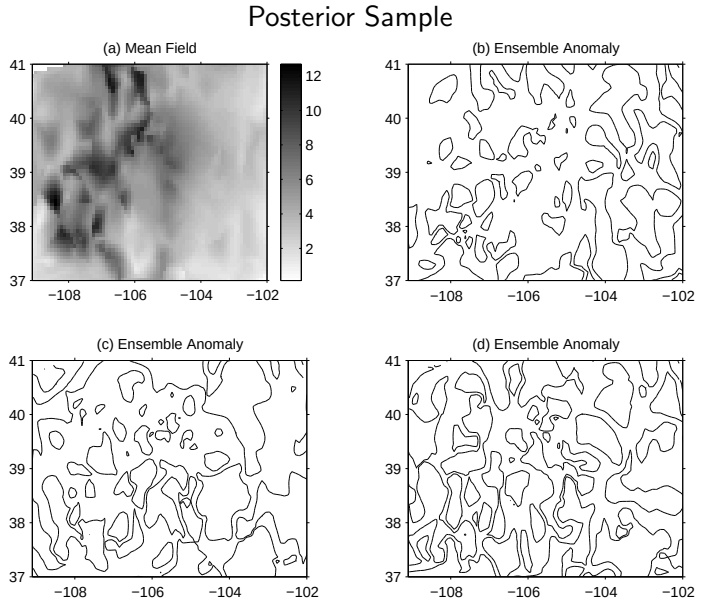

Figure 5: Panel (a) shows the infilled mean for stations in Colorado, April 1948. Panels (b-d) show contours of random ensemble anomalies (deviations) conditioned on observations from April 1948. An ensemble member would be the pixel-wise sum of panel (a) with any of the other panels. The contour levels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