Assumptions for Precipitation Infilling

1 Spatial Estimates of Mean

Let $P(\mathbf{x}, t)$ denote the precipitaiton at location $\mathbf{x}$ at time t and $\mu(\mathbf{x}) = E[P(\mathbf{x}, t)]$. We assume that $\mu(\mathbf{x})$ does not depend on t. Then, $\mu(\mathbf{x}) = \theta(\mathbf{x})^2 + \sigma(\mathbf{x})^2$ where $\theta(\mathbf{x}) = E[\sqrt{P(\mathbf{x}, t)}]$ and $\sigma(\mathbf{x})^2 = \mathrm{Var}[\sqrt{P(\mathbf{x}, t)}]$. Moreover, setting

$$
\begin{aligned}
C(\mathbf{x}) &= \frac{\sigma(\mathbf{x})^2}{\theta(\mathbf{x})^2 + \sigma(\mathbf{x})^2} \\
&\quad \text{it follows that} \\
\theta(\mathbf{x}) &= \sqrt{\mu(\mathbf{x})(1 - C(\mathbf{x}))} \\
\sigma(\mathbf{x}) &= \sqrt{\mu(\mathbf{x})C(\mathbf{x})}
\end{aligned}
$$

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. We further 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. The function $C(\mathbf{x})$ was 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).

2 Time Independence

We assumed that the monthly observations are independent in time (conditional on nuisance parameters).

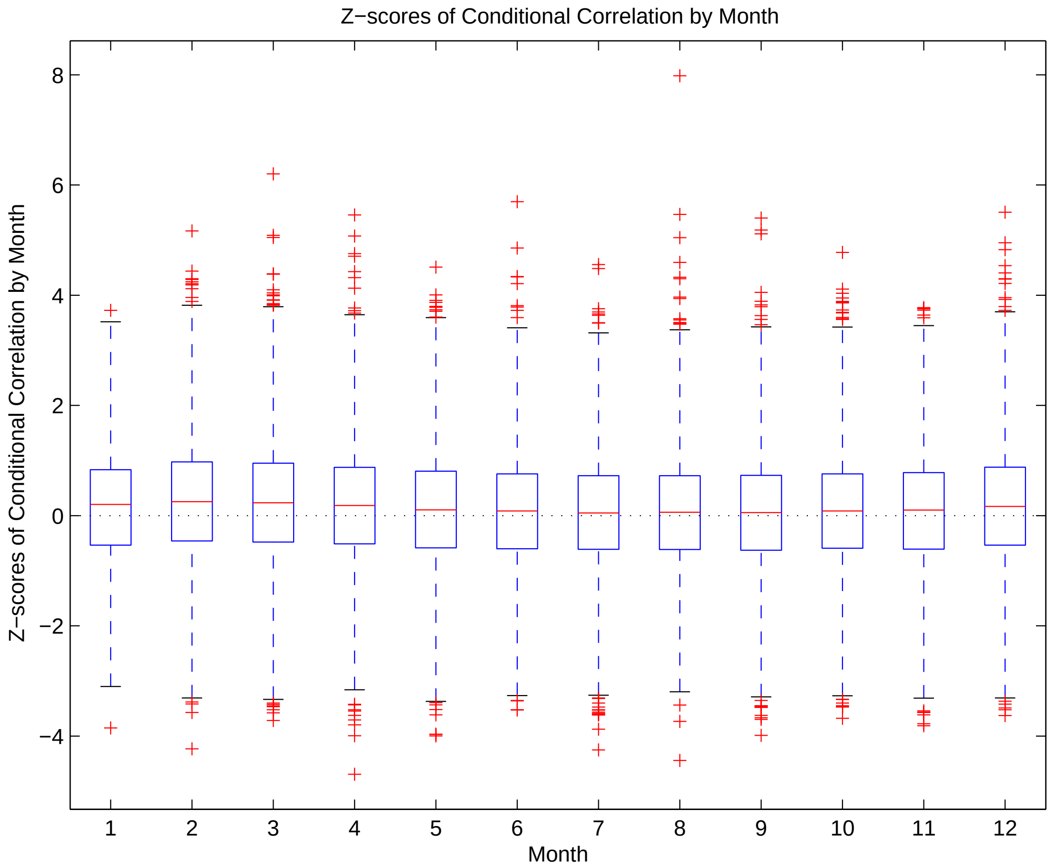

Figure 1: Fisher-transforms of correlation of (Data - Infill)$_i$ with Data$_{i-1}$. The conditional correlations are generally small and suggest that in the presence of spatial informaton from the current time period, there is little predictive ability of the previous time period.

3 Spatial Estimate of α

We assumed that α, the degrees of freedom in the inverse-Wishart model, depended on location. We used leave-one-out prediction on the special subset of stations to obtain RMSE curves on the 400 stations in 4. Principle Components and smoothing was then used to estimate RMSE curves for all stations. Minima of these curves are shown in Figure 2.

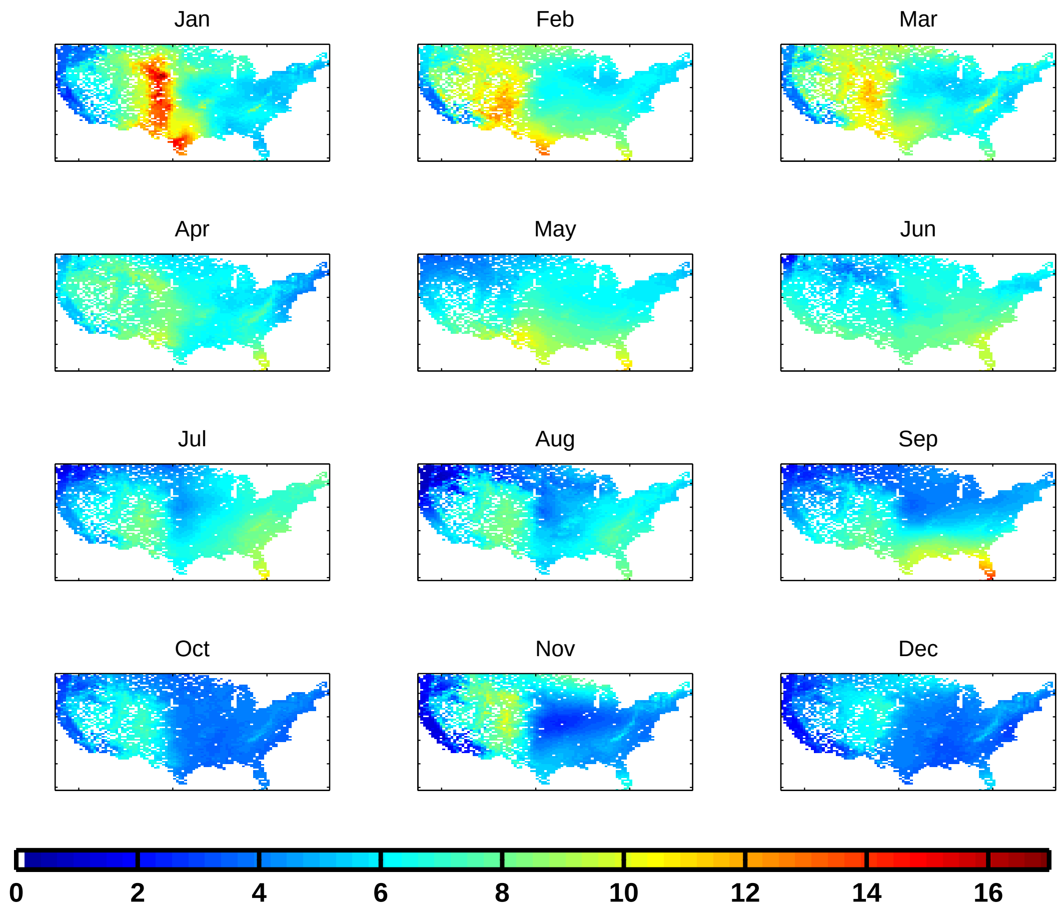

Figure 2: The degrees of freedom (α) estimated by Cross-validation and spatial smoothing.

4 Special Subset

We chose a subset of the stations and used data from these stations to 1) estimate long-range covariance and 2) approximate degrees of freedom over space. These 400 stations were chosen from stations with at least 50 observations in such a way that the spacing is as uniform as possible over the domain. Figure 4 shows the locations of these stations.

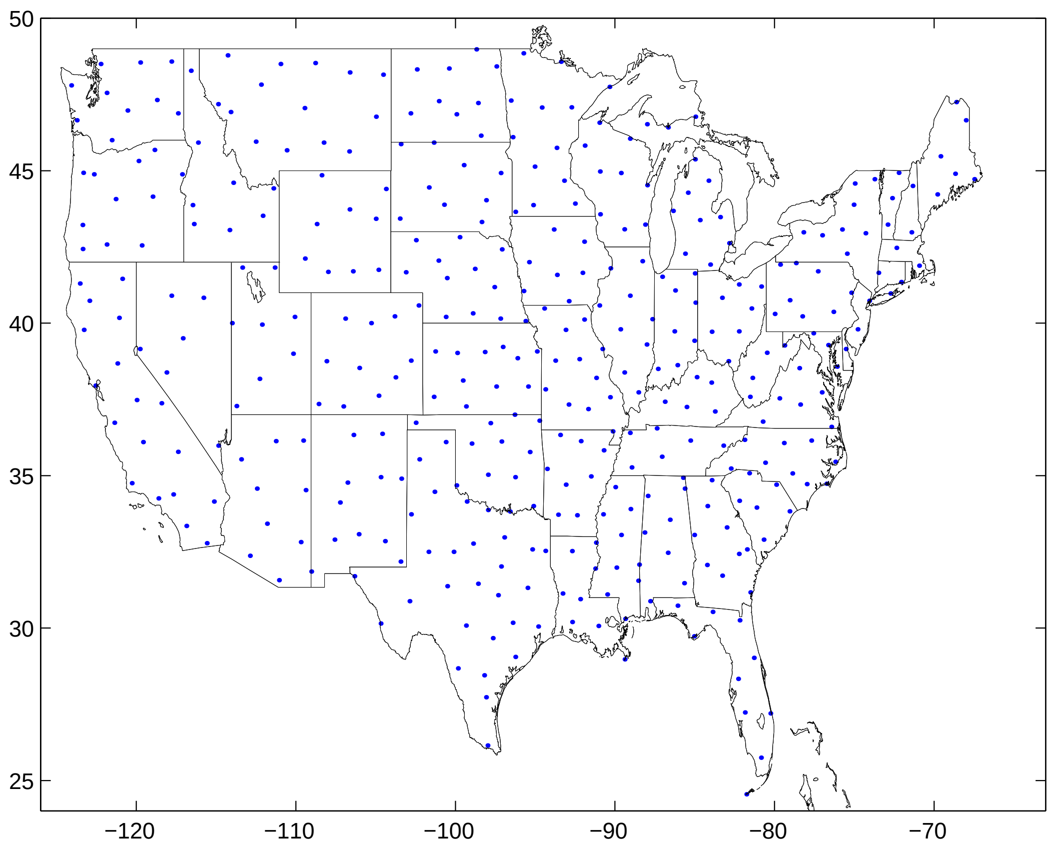

Figure 3: The subset of 400 stations used for estimating the long-range correlation and the degrees of freedom (see Section 3)

5 Covariance function

$$
\begin{aligned}
k(\mathbf{x}_i, \mathbf{x}_j) &= \exp\left\{ -\left(\frac{d_1(\mathbf{x}_i, \mathbf{x}_j)}{\eta_1} + \frac{d_2(\mathbf{x}_i, \mathbf{x}_j)}{\eta_2} \right)^{\eta_3} \right\} \\
\text{where} & \\
d_1(\mathbf{x}_i, \mathbf{x}_j) &= \text{east-west distance between points } \mathbf{x}_i \text{ and } \mathbf{x}_j \\
d_2(\mathbf{x}_i, \mathbf{x}_j) &= \text{north-south distance between points } \mathbf{x}_i \text{ and } \mathbf{x}_j
\end{aligned}
$$

Month	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
η_1	736	680	771	715	676	499	372	363	786	947	786	686
η_2	618	566	624	588	577	425	300	308	661	908	723	614
η_3	0.81	0.79	0.75	0.72	0.63	0.58	0.56	0.61	0.63	0.71	0.80	0.83