A Bayesian Algorithm for Reconstructing Spatially Averaged Temperatures

Martin Tingley ( tingley@fas.harvard.edu ) and Peter Huybers, Harvard University

Abstract

The determination of spatial average temperature from point estimates is a non-trivial statistical problem. In the paleo-climate context, the additional need to convert proxy time-series into temperature estimates makes the problem still more challenging. We present a hierarchical Bayesian algorithm for estimating spatial average temperature from sets of overlapping and incomplete instrumental and proxy records. Our method explicitly accounts for both the spatial covariance and temporal auto-covariance of temperature data.

1. INTRODUCTION

Most reconstructions of spatial average temperature at paleo-climate time-scales proceed in two steps: the proxy values are first averaged through space, and these estimates are then transformed onto the temperature scale via some form of regression (Jansen et al., 2007). Such approaches fix the spatial covariance structure provided by the Bayesian approach. Instead, we explicitly consider spatial covariances. This formulation resolves several issues with the block-averaging approach, and any distribution of instrumental and proxy observations.

Our approach is to model the relationship between the true, space-time co-varying temperature field and the noisy, localized measurements of it using a hidden Markov model (Wikle and Berliner, 2004; Wikle and Berliner, 2005). We assume proper but weakly informative, and, wherever possible, conjugate priors for all unknowns (Gelman et al., 2003), and sample from the posterior using a hybrid Gibbs-Metropolis algorithm.

A major benefit of this Bayesian approach is that, by drawing repeatedly from the full posterior distribution, we obtain an estimate of the uncertainty covariance structure.

2. PARAMETER ESTIMATES

We present our assimilation model and results of testing the MCMC sampler on a surrogate data model, and to explicitly consider spatial covariances. This formulation resolves several issues with the block-averaging approach, and any distribution of instrumental and proxy observations.

We present a hierarchical Bayesian algorithm for reconstructing spatially averaged temperatures. For a more detailed discussion of the Bayesian approach to model the relationship between the true, space-time co-varying temperature field and the noisy, localized measurements of it using a hidden Markov model, see Wikle and Berliner (2004, 2005). We assume proper but weakly informative, and, wherever possible, conjugate priors for all unknowns (Gelman et al., 2003), and sample from the posterior using a hybrid Gibbs-Metropolis algorithm.

A major benefit of this Bayesian approach is that, by drawing repeatedly from the full posterior distribution, we obtain an estimate of the uncertainty covariance structure. This allows us to make quantitative statements about both the relative contributions of the different proxies to the spatial average, and the extent to which the model can constrain the various parameters. In particular, the model outputs the uncertainty in the coefficients of the transformation linking the proxy values to temperature units. The model can easily be generalized to accommodate different categories of proxy data assumed to have different uncertainty properties. We present our assimilation model and results of testing the MCMC sampler on a surrogate data model. In the future, we will apply this model to instrumental data, and then to proxy data as well.

2.1 Model Parameters

We represent the underlying true temperature values by \( \tilde{F} \), which includes the true temperature values at a number of random locations, from which we estimate the block average. Observations of subsets of \( \tilde{F} \) will be denoted by \( \tilde{W} \), while the locations of the points corresponding to the entries of \( \tilde{F} \) are given by \( L \). Subscripts \( I \) and \( P \) indicate subsets of vectors corresponding to the instrumental and proxy observations.

Forms of the priors and posteriors, as well as the true parameter values used in constructing the surrogate data, and the prior parameters used to assimilate this data, are shown in the table below. \( \mu \) denotes the mean of all surrogate instrumental observations, \( \sigma \) denotes the variance of all surrogate instrumental observations. A factor of 0.16.

### Table 1

<table>
<thead>
<tr>
<th>Surrogate Value</th>
<th>Prior Conditional Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>Uniform(0, 1)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>Inverse-Gamma(1, 1)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>Log-Normal(-7.5, 4)</td>
</tr>
<tr>
<td>( \tau^2 )</td>
<td>Inverse-Gamma(1, 1)</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>Inverse-Gamma(1, 1)</td>
</tr>
<tr>
<td>( s )</td>
<td>Normal(-1, 0.75)</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
<tr>
<td>( \delta )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
<tr>
<td>( \phi )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
<tr>
<td>( \omega )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Normal(0.5, 0.5)</td>
</tr>
</tbody>
</table>

3. MODEL DESCRIPTION

Surface temperatures are assumed to follow a multivariate first order auto-regressive process, with mean value a linear function of latitude:

\[
\tilde{t}_i = \left( \mu + \alpha L \right) + \beta \tilde{t}_{i-1} + \varepsilon_i
\]

where \( \alpha \) and \( \beta \) are the locations of the \( \alpha \) and \( \beta \) elements of the temperature vector \( \tilde{T} \).

The instrumental observations, at each year, are assumed to be noisy versions of the true temperature:

\[
\tilde{W}_{ij} = \tilde{T}_j + \varepsilon_{ij}
\]

The observation equation, at each year, is more complex. Proxy records (presumably scaled in some way specific to our analysis) are assumed to have an unknown, linear, statistical relationship to the true temperature, and we generalize to assume that the constant term depends on latitude. This suggests the regression equation:

\[
W_{ij} = \frac{1}{\alpha} - W_{ij} + \left( \tilde{T}_j + \beta \tilde{T}_{i-1} \right)
\]

4. SURROGATE DATA CONSTRUCTION AND ANALYSIS

We randomly select a large number of points in the polar region, north of 70°, and simulate temperatures at these locations for 60 years according to Equation 1, using the parameters in Table 1. We select out 15 of these locations to correspond to instrumental observations, and add noise according to Equation 2. Similarly, we select 10 locations to correspond to proxy observations, and add noise according to Equation 2. In all cases, the posterior relative to the priors, is well constrained by the data. Those priors not visible in Figure 3 have zero probability in the region of high posterior probability. Note that the AR(1) coefficient, is not well identified by the model: 0.8, the true value, falls in the upper tail of the posterior. However, it all parameters are fixed at their true values, the sampler does correctly identify \( \mu \).

Figure 4 shows posterior draws for several parameters, illustrating the full uncertainty structure provided by the Bayesian approach. Note in particular the strong correlation between \( \beta \) and \( \mu \) (which define the mean of the AR(1) process), and the strong non-linear relationship between \( \beta \) and \( \mu \). The following notation is used to shorten the equations:

\[
J_T = \left( \tilde{T}_{ij} - \beta \tilde{T}_{i-1,j} \right) = \beta \left( \tilde{T}_{ij} - \tilde{T}_{i-1,j} \right)
\]

The probability of the data, given the true temperatures and all parameters, can be factored to give:

\[
P \left( \tilde{W}_1, \ldots, \tilde{W}_n, \tilde{T}_{1}, \ldots, \tilde{T}_n \right) = P \left( \tilde{W}_1, \tilde{T}_{1} \right) \prod \left( P \left( \tilde{W}_i | \tilde{T}_{i} \right) \right)
\]

5. SUMMARY

We present a hierarchical Bayesian algorithm for reconstructing spatially averaged temperatures. Our approach is to model the relationship between the true, space-time co-varying temperature field and the noisy, localized measurements of it using a hidden Markov model, and to explicitly consider spatial covariances. This formulation resolves several issues involved in paleo-climate reconstructions: the model solves explicitly for the parameters linking the proxy scale to temperatures, and, by considering the underlying temperature field, we arrive at an estimate of the regional mean that accounts for the spatial covariance between the records.

REFERENCES: