A Bayesian approach to reconstructing climate fields from proxy data

Analysis of a 600 year multi-proxy temperature data set

Martin Tingley (tingley@fas.harvard.edu), Peter Huybers (Harvard University), Konrad A. Hughen (WHOI)

1 Abstract

We present a Bayesian model to assimilate incompatible (in space and time) instrumental and proxy data sets to estimate, with uncertainty, the evolution of the climate field over a multi-centennial scale. The Bayesian model consists of a prior model (the spatial covariance of the true climate field or a multi-resolution AR1 process with spatially correlated innovations; a data level that specifies how observations are related to the climate parameters; and a forward model that specifies how proxy data are related to the climate parameters) and a multi-dimensional post-processing step that specifies diffuse prior distributions for all unknowns. Multiple draws from the posterior are used to ensemble spin up and temper the posterior distributions of the field, quantify the data-model assumptions, and perform posterior predictive checks. Probability distributions for various statistics can be estimated from this ensemble, from which measures like the time series of spatial means to more exotic quantities like the probability that a given year was the most extreme value of the climate field during the reconstruction.

2 Analysis Model

For ease of description we will assume the field of interest is that of annual mean surface temperature, though the method we have developed is general and in principle applicable to the reconstruction of any climate field.

The evolution of the temperature field sampled at a finite number of spatial locations, \( T \), is assumed to follow a multi-variate first order auto regressive process:

\[ \dot{T}_t = A T_{t-1} + \epsilon_{t}, \]

where \( A \) is the matrix of the process, \( x \) is the AR(1) coefficient, and the subscript \( t \) indicates the year. The innovations, \( \epsilon_t \), are assumed to be independent and identically distributed (iid) normal draws, \( \epsilon_t \sim N(0, \Sigma) \), with spatial covariance structure given by:

\[ \Sigma = \sigma^2 \exp(- (|x_i - x_j|) / a), \]

where \( |x_i - x_j| \) is the distance between the \( i \)th and \( j \)th elements of the field vector \( T \).

It is useful to decompose the vector \( T \) at each year into its ensemble members:

\[ T = (T_1, \ldots, T_n) \]

where \( T_1 \) and \( T_2 \) are the true temperatures at locations where there are instrumental or proxy observations, respectively. If there is no proxy or instrumental observation at a given location, then the true field value at that location appears in both \( T_1 \) and \( T_2 \). The true temperatures at the target locations where there are no observations.

The prior model described above, selects 100 target locations to be the remaining nodes of a regular grid.

The instrumental observations at each year, \( W_{ij} \), are assumed to be noisy versions of the true temperatures at the corresponding locations.

The noise terms are assumed to be iid multivariate normal draws, \( \epsilon_t \sim N(0, \Sigma) \), where \( \Sigma \) is the identity matrix. Note that the instrumental temperature observations are subject to systematic errors (Brockhuysen et al., 2006) which are not modeled.

The proxy observations are assumed to have unobserved, statistically linear relationship to the true temperatures at the corresponding locations, which motivates the instrumental ensemble members:

\[ W_{ij} = A T_{ij} + \epsilon_t \]

The noise terms are once more assumed to be iid normal draws, \( \epsilon_t \sim N(0, \Sigma) \). We specify a separate proxy observation equation for each type of proxy data.

In order to define the ensemble, prior means and standard deviations are specified for the parameter field and the climate forcings for the first year of the analysis. Our approach is to use only informative but poorly prior distributions, and show that the ensemble model is robust to these choices.

A package of Matlab code, which we have dubbed BASSAT (“A Bayesian Algorithm for Reconstructing Spatially-Averaged Temperatures”), is then used to infer the joint distribution of the ensemble members for the instrumental and the field values through time at the target locations. A more thorough description of this algorithm can be found in Tingley and Huybers (2008a,b), and the code package is freely available at pingouin.fas.harvard.edu/tingley/

3 600 year temperature data set

We use BASSAT to infer land surface temperatures between 40°N and 85°N over the last 600 years. The basic result of the analysis is to apply BASSAT to a data set constructed from the CRU compilation.

The instrumental data set consists of 249 temperature time series from the CRU compilation. The proxy data set is composed of 96 tree ring density (TRD) series, 13 annual lake floor sediment deposit thickness series (LAKE), and seven δO18 series from ice cores (ICE).

The proportion of ensemble members for which 1995 and 1998 are the warmest years, and the 1990s the warmest decade, in the 1400-1999 interval. We consider three spatial averages, each confined to the 45°N-85°N latitude band: All land, North America/Greenland, and Eurasia.

Table 1: Warm years and decades, 1400-1999.

<table>
<thead>
<tr>
<th>Region</th>
<th>1995</th>
<th>1998</th>
<th>1990s</th>
</tr>
</thead>
<tbody>
<tr>
<td>All land</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>North America</td>
<td>0.90</td>
<td>0.97</td>
<td>0.92</td>
</tr>
<tr>
<td>Eurasia</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The proportion of ensemble members for which 1995 and 1998 are the warmest years, and the 1990s the warmest decade, in the 1400-1999 interval. We consider three spatial averages, each confined to the 45°N-85°N latitude band: All land, North America/Greenland, and Eurasia.

Table 2: Cold years, 1400-1999.

<table>
<thead>
<tr>
<th>Region</th>
<th>1453</th>
<th>1601</th>
<th>1453-1993</th>
</tr>
</thead>
<tbody>
<tr>
<td>All land</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>North America</td>
<td>0.91</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td>Eurasia</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
</tbody>
</table>

The proportion of ensemble members for which 1995 and 1998 are the warmest years, and the 1990s the warmest decade, in the 1400-1999 interval. We consider three spatial averages, each confined to the 45°N-85°N latitude band: All land, North America/Greenland, and Eurasia.