Decadal Predictability of Global Sea Surface Temperature Anomalies

Dallas Foster\textsuperscript{1,2}, N. Urban, and D. Comeau\textsuperscript{1}

\textsuperscript{1}Department of Mathematics, Oregon State University
\textsuperscript{2}Corvallis, OR, USA and \textsuperscript{3}CCS-2, Los Alamos National Laboratory, NM, USA

Abstract

While global climate modeling algorithms are able to capture the mean behavior of the earth’s climate system, sub-scale processes are usually not able to be resolved. In particular, Sea Surface Temperature (SST) anomalies can represent information that is not captured in model climatologies or stochastic events in climate dynamics. Statistically, these anomalies have been studied in a linear inverse model (LIM) framework with some success in analyzing regional and temporal predictability. In this framework, parameters and uncertainties have typically been given point estimates. We examine the decadal predictability of global SST anomalies and instead the way in which typical point estimates are available in this setting. Then we show how using a Bayesian approach can incorporate model assumptions and improve parameter estimates and forecast skill.

Sea Surface Temperature Anomalies

- For methodological concerns, we study CESM Large Ensemble control run data.
- Statistical models of anomalies can provide information on average behavior as well as inherent uncertainty.
- Modeling anomalies can produce insight into random effects in the climate system and forces not captured by models.

Data

- Sea Surface Temperature Anomalies are the discrepancy between physical models of ocean temperatures and measured SSTs.
- Modeling anomalies can produce insight into random effects in the climate system and forces not captured by models.
- Better understanding of uncertainty can provide more robust forecasts and policy making decisions.

Problem Description

- Leading Temporal Pattern of Decomposed SST Anomalies

- Characteristics of SST Anomalies

- These patterns exhibit persistent anomaly structures in the North Atlantic and North Pacific that appear to be consistent with the Pacific Decadal Oscillation pattern.

Vector Autoregressive Models

If we believe that only previous values of the variable are needed to forecast the future value then we can use a Vector Autoregressive Model (VAR). Suppose that we expect the value at time $t$ of our time series, $X$, depends only on the previous $p$ times, then we can form a VAR(p) model

\[ X_t = A_1 X_{t-1} + \ldots + A_p X_{t-p} + \epsilon_t \]

- $A_1$, $\ldots$, $A_p$ are matrices; $\epsilon_t$ is an error term.
- $A_1$ are matrices that give weighted averages between values at different points and different spatial patterns. These matrices must be estimated for each application. $A_1 \cdots A_p$ is addition noise that capture our uncertainty in predictions.

- The simplest form of the model is $A_1 = 0$, the forecast that the next value is the same as the current value $X_{t+1} = X_t$.

Forecast Skill of 1 year lagged Persistence

Linear Inverse Model

A Linear Inverse Model (LIM) is a mathematical model for the description of processes that vary with random noise. The tests in climate dynamics date from at least the late 1980s and shows that the dynamics of the system can be captured by the matrices $\mathbf{A}$ and $\mathbf{B}$ and the system can be captured by the matrices $\mathbf{A}$ and $\mathbf{B}$ and the LIM (VAR) models.

\[ \dot{X}_t = \mathbf{A} X_t + \mathbf{B} \epsilon_t \]

- In order to use a LIM we must estimate the values of the matrices $\mathbf{A}$ and $\mathbf{B}$. In order to do so, it is often convenient to formulate a LIM as a VAR model and use Maximum Likelihood Estimates (MLE). Suppose we want to forecast $k$ steps ahead, $X_{t+k} = \mathbf{A}^k X_t + \sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B} \epsilon_{t+j}$.

Maximum Likelihood Estimates

Mathematically, we can derive estimates for the parameters $\mathbf{A}$ and $\mathbf{B}$ that have maximum probability of the data occurring. There are several Maximum Likelihood Estimation Methods (MLE). These estimates are single valued called point estimates.

\[ \hat{\mathbf{A}} = \mathbf{B} (\mathbf{X} X^T)^{-1} \]

- By only keeping the patterns that are most significant we can identify the important terms while reducing the size of the problem and sparsity issue.

Leading Spatial Pattern of SST Anomalies

Empirical Orthogonal Functions (EOFs)

- EOFs can be used to study the possible spatial patterns that exist in the data, as well as reduce the total amount of data analyzed.
- Suppose that we have $n$ years of data with $d$ points in space. If our original data, $x_i$, has size $n \times d$, then we can decompose it into temporal patterns, $\mathbf{X}$, with size $n \times d$, spatial patterns, $\mathbf{W}$, with size $d \times d$. and a matrix representing their relative importance, $\lambda$, with size $d \times d$.

\[ X = \mathbf{WX} \]

- Noticeable patterns include possible EOFs and Karhunen-Loeve current related spatially.

Bayesian Estimation

- MLE estimates tend to underestimate the uncertainty in the system and may not be reliable with the amount of data given.
- Further, we don’t get information about the full distribution of the parameters, $\theta$, or the dynamics. Bayes’ Theorem:

\[ P(\theta | \mathbf{X}) \propto P(\mathbf{X} | \theta) P(\theta) \]

- How do we assign prior probabilities, $P(\theta)$? For parameters, we try to infer properties that we think the model should have.
- Inforative Priors $\sim N(\mu, \sigma)$ provide some information about unknown parameters, but only scale linearly with number of parameters.
- Can produce sparsity in the matrix, which we also expect.
- Horseshoe Prior $\sim N(0, \sigma)$, $\lambda \sim \text{Half-Cauchy}(0, \sigma)$.

- Can produce sparsity in matrix.
- Flexible for both variance $\mathbf{A}$ and $\mathbf{B}$.
- Large number of hyper-parameters needed to estimate along with priors. May require more data to be as effective.

- Easy to implement, relatively computationally clean.
- Natural design to sample our correlation matrices.
- Not much further structure imposed, like sparsity.

Comparison of MLE and Bayesian Estimation

Prior for $\mathbf{A}$ and $\mathbf{B}$ were chosen and their performance was tested against test data where the values of the parameters were known a-priori. For Bayesian estimates the numerical errors listed below are based on the Maximum A Posteriori (MAP) value of the parameter.

Numerical Results on Test Data

<table>
<thead>
<tr>
<th>MLE estimate</th>
<th>Bayesian estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point estimate</td>
<td>Posterior median</td>
</tr>
</tbody>
</table>

Forecast Skill 1 year lagged LIM forecast with Minnesota-Horseshoe Prior Estimates

<table>
<thead>
<tr>
<th>South Hemisphere</th>
<th>North Hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>0.050</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Current and future work

- Current estimate should be implemented using MLE-based estimates, more robust data available.
- Different temporal averaging on forecast skill may be necessary.
- Willing to extend on forecast skill is not a measure for sparsity of prior parameters.
- A Bayesian framework can improve forecasts by incorporating prior beliefs about the structure of uncertain parameters.
- Foreword and structure can lead to better approach, and using more data techniques.

References