Decadal Predictability of Global Sea Surface Temperature Anomalies

Dallas Foster[*,1], N. Urban[2], and D. Comeau[2]

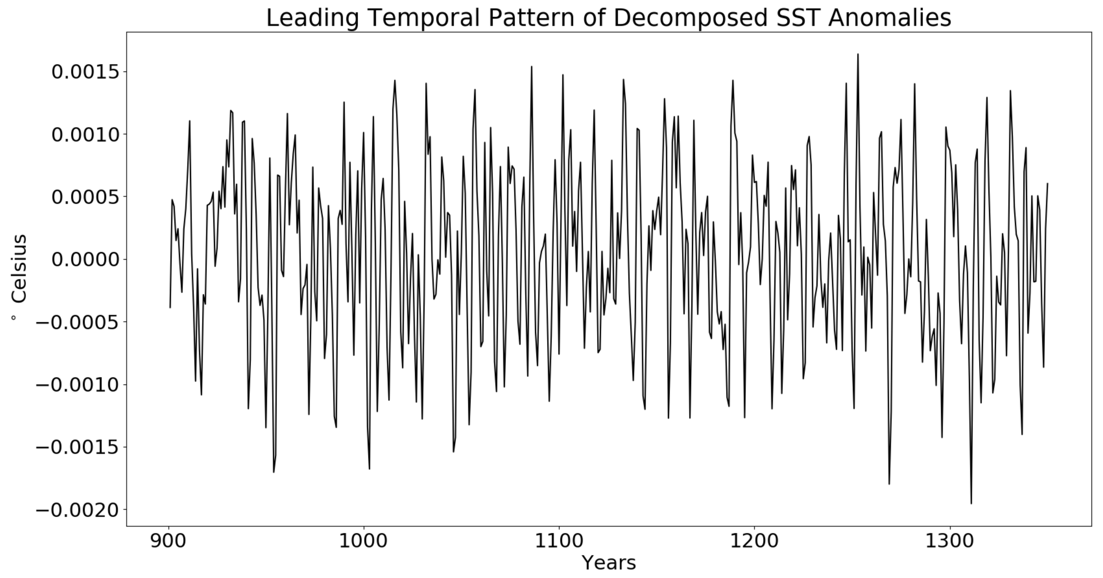
[1]Department of Mathematics, Oregon State UniversityCorvallis, OR, USA and [2]CCS-2, Los Alamos National Laboratory, NM, USA

Abstract

While global climate modeling algorithms are able to capture the mean behavior of the earth's climate system, sub-scale processes are usually unable 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. Historically, these anomalies have been studied in a Linear Inverse Model (LIM) framework with some success in analyzing monthly, regional anomaly predictability. In this framework, parameters and uncertainties have typically been given point estimates. We examine the decadal predictability of global SST anomalies and discuss the ways in which typical point-estimates are unreliable 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

Leading Temporal Pattern of Decomposed SST Anomalies



Problem Description

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

Data

- For methodological concerns, we study CESM Large Ensemble control run data.
- Data is gathered at 1° resolution for 1800 years for a total of roughly 10^9 data points (1GB of disk space!).
- In order to feasibly work with that data we have to reduce the amount of data while keeping as much information as possible.

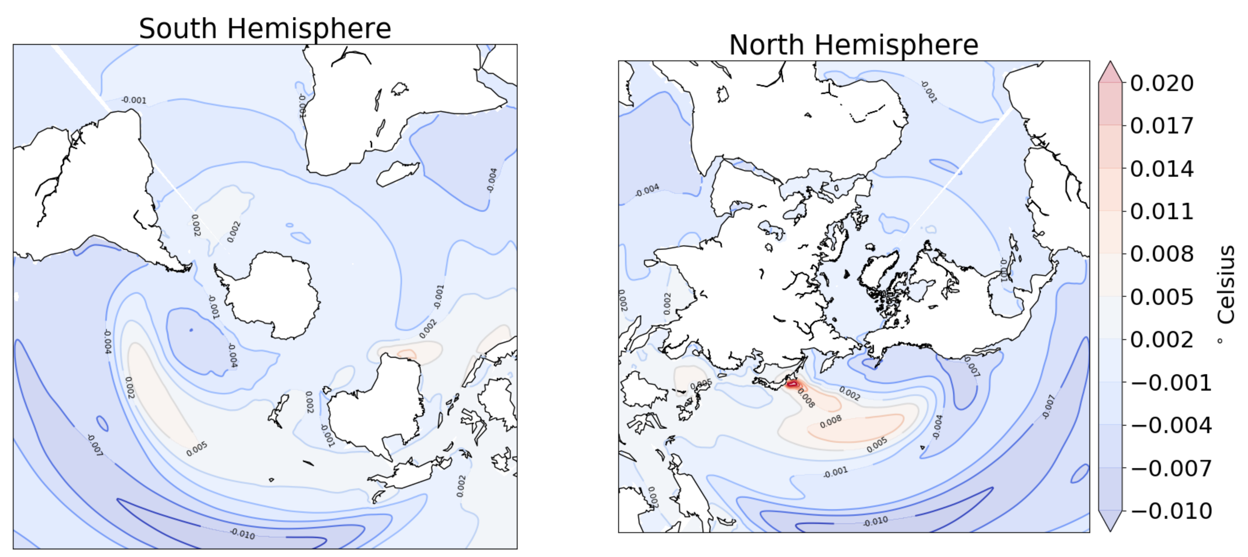
Empirical Orthogonal Functions (EOFs)

- EOF Analysis 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 M years of data with N points in space. If our original data, X, has size $M \times N$ then we can decompose it into temporal patterns, T with size $M \times D$, spatial patterns, W with size $D \times N$ and a matrix representing their relative importance, S with size $D \times D$.

$$X = TSW$$

- By only keeping the patterns that are most significant we can identify the important dynamics while reducing the size of the problem and spurious noise present in the data.

Leading Spatial Pattern of SST Anomalies



Noticeable patterns include possible ENSO and Kuroshio current related variability.

Vector Autoregressive Models

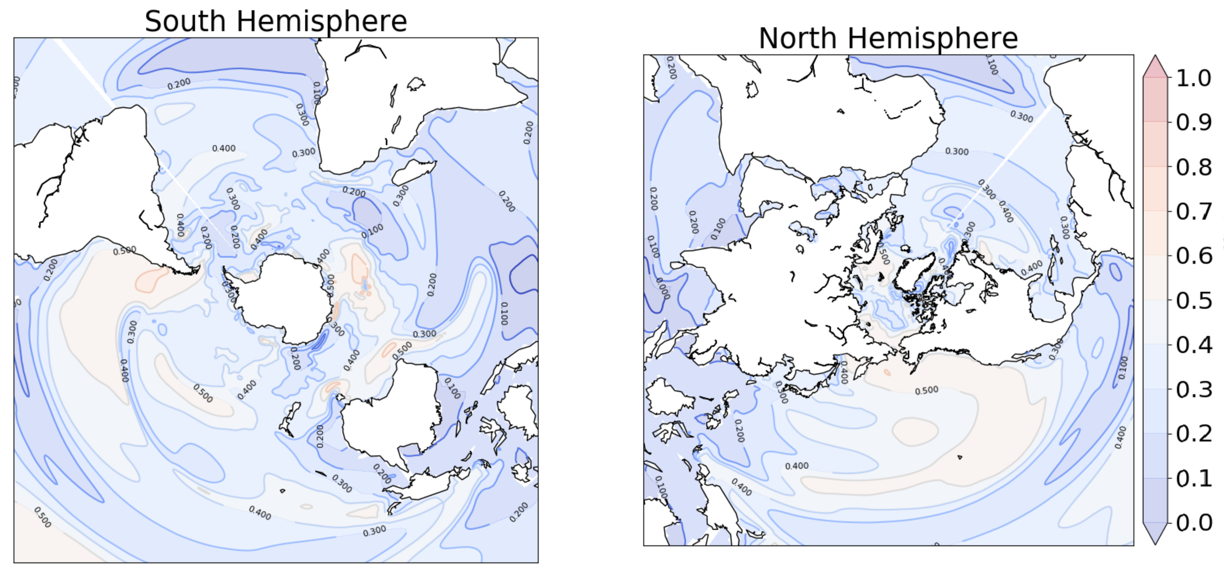
If we believe that only previous values of the variable are needed in order to forecast the future value then we can use a **Vector Autoregression Model** (VAR).

- Suppose that we expect that the value at time t of our timeseries, $\mathbf{X}_t$ depends only on the previous p times, then we can form a $VAR(p)$ model

$$\mathbf{X}_t = A_1 \cdot \mathbf{X}_{t-1} \cdots + A_p \cdot \mathbf{X}_{t-p} + \epsilon$$

- A_i are matrices that give weighted averages between values at different years and different spatial patterns. These matrices must be estimated for each application. $\epsilon \sim N(0, \Sigma)$ is additive noise that captures our uncertainty in predictions.
- The simplest VAR model is **Persistence**, the forecast that the next value is the same as the current value $\mathbf{X}_{t+1} = \mathbf{X}_t$.

Forecast Skill of 1 year lagged Persistance



Linear Inverse Model

A Linear Inverse Model (LIM) is a mathematical model for the description of processes that vary with random motion. The term in climate dynamics dates from at least the late 1980's and assumes that the dynamics of the process can mostly be expressed by the matrix B and the noise can be captured by the matrix Q and Brownian motion $d\mathbf{W}_t$.

$$d\mathbf{X}_t = B\mathbf{X}_t dt + Qd\mathbf{W}_t$$

- In order to use a LIM we must estimate the values of the matrices B and Q. 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 τ years in advance,

$$X_{t+\tau} = e^{B\tau} X_t + \epsilon, \qquad \epsilon \sim N(0, \Sigma)$$

Maximum Likelihood Estimates

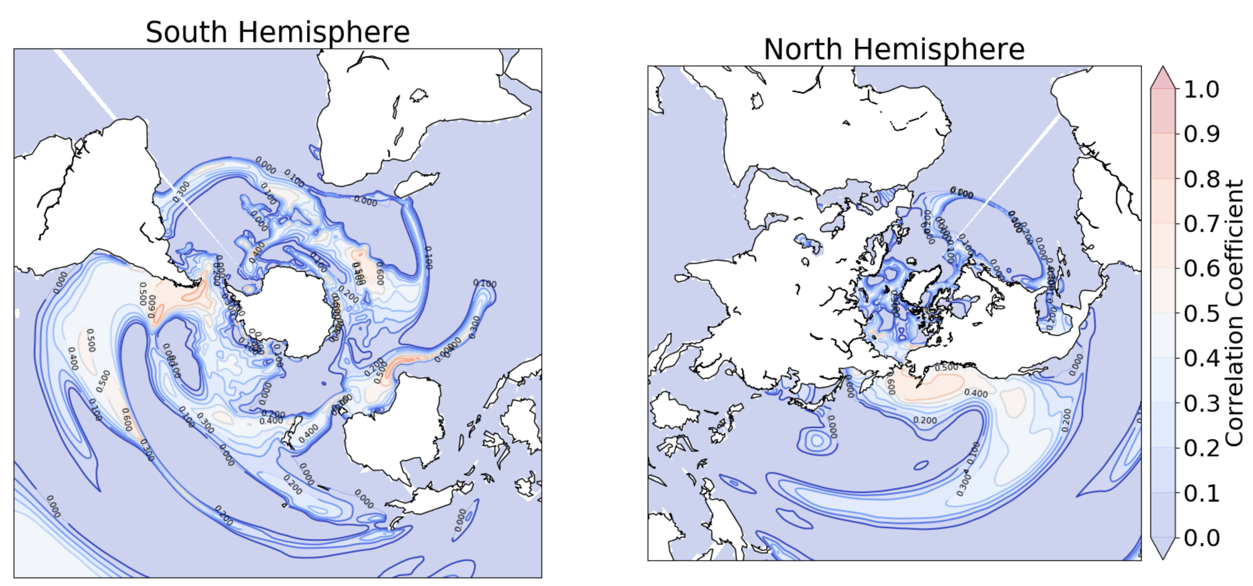
Mathematically we can derive estimates for the parameters B and Q that have maximizes the probability of the data occurring. These are called Maximum Likelihood Estimates (MLEs). These estimates are single values called **point-estimates**.

$$\Lambda = \mathbb{E}[\mathbf{X}_t^2] \qquad \Lambda_\tau = \mathbb{E}[X_t X_{t+\tau}^T] \qquad \Lambda_\tau = e^{\hat{B}\tau}\Lambda$$

$$\hat{\Sigma} = \frac{1}{N}\left(\mathbf{X}_{t+\tau} - e^{\hat{B}\tau}\mathbf{X}_t\right)\left(\mathbf{X}_{t+\tau} - e^{\hat{B}\tau}\mathbf{X}_t\right)^T \qquad \hat{Q} - e^{\hat{B}\tau}\hat{Q}e^{\hat{B}^T\tau} = -\hat{B}\hat{\Sigma} - \hat{\Sigma}\hat{B}^T$$

These formulas converge to the actual values of the parameters as the amount of data increases. Their usefulness depends on the amount of data (in time) that we have. 1800 yearly data points may not be enough to ensure good estimates.

Forecast Skill of 1 year lagged LIM forecast with MLE Estimates



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, $\theta = [B, Q]$, or the dynamics. **Bayes' Theorem**:

$$\pi(\theta|\mathbf{X}_t) = \frac{\pi(\mathbf{X}_t|\theta)\pi_0(\theta)}{\pi(\mathbf{X}_t)} \tag{1}$$

- How do we assign prior probabilities, $\pi_0(\theta)$ for parameters? We try to enforce properties that we think the matrices should have.
  - Uninformative Priors $\theta \sim N(0,1)$
    - Simple to use and computationally inexpensive.
    - Does not use knowledge of the system, requires more data to be effective.
  - Minnesota Prior for matrix B
    - $B_{ii} \sim N(0,\lambda),\ B_{ij} \sim N\left(0, \lambda\theta\frac{Q_{ii}}{Q_{jj}}\right),\ \lambda \sim$ Half-C$(0,1),\ \theta \sim \mathcal{U}(0,1)$
    - Assumes a random walk nature in the dynamics, which we observe in the data.
    - Is more computationally expensive than uninformative priors, but only scales linearly with number of parameters.
    - Can produce sparsity in the matrix, which we also expect.
  - Horseshoe Prior $\theta_{ij} \sim N(0, \tau\lambda_{ij}), \quad \tau, \lambda_{ij} \sim$ Half-C$(0,1)$
    - Can produce sparsity in matrices.
    - Flexible for both matrices B and Q.
    - Large number of hyper-parameters needed to estimate along with matrices. May require more data to be as effective.
  - LKJ Prior for matrix Q
    - Easy to implement, relatively computationally cheap.
    - Naturally designed to sample over correlation matrices.
    - Not much further structure imposed, like sparsity.

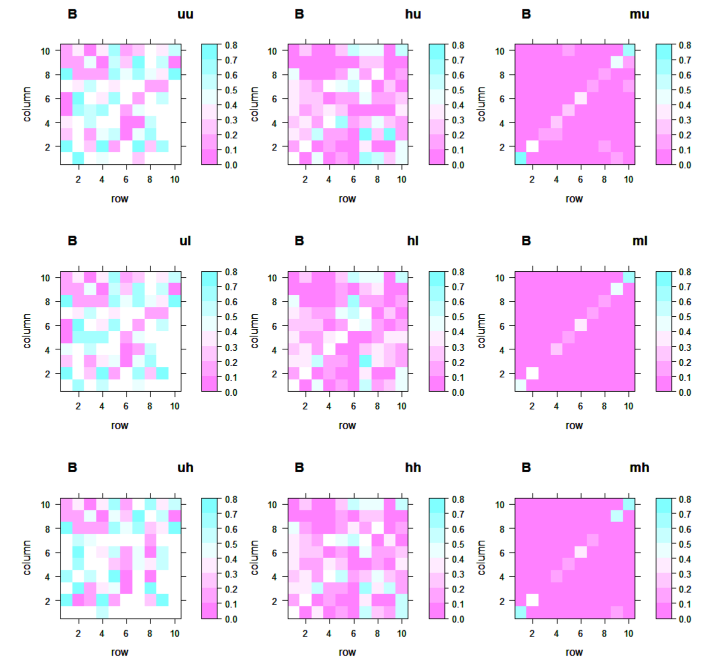


Figure: Matrix plots for errors of MAP estimates of B matrix using various prior distributions π_0 for parameters. These priors can influence the structure of the matrix by helping to promote sparsity.

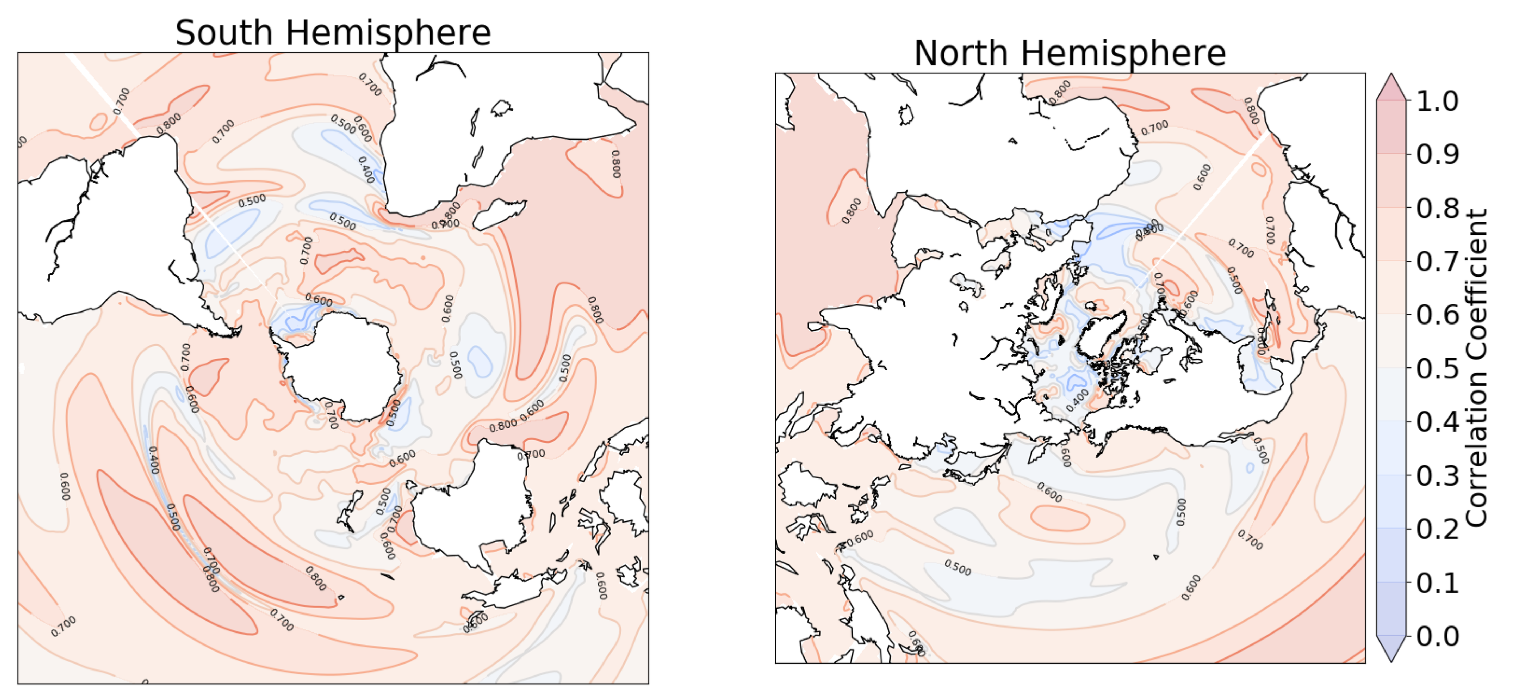
Numerical Results on Test Data

Comparison of MLE and Bayesian Estimation

Priors for B - Q were chosen and their performance was tested again 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 parameters.

Relative Error in Estimating B Matrix							
N	**MLE**	unif-unif	unif-horseshoe	horseshoe-unif	horseshoe-horseshoe	**minn-unif**	**minn-horseshoe**
50	**0.993**	1.953	1.971	3.000	7.00	**0.574**	**0.543**
100	**0.999**	1.607	2.371	2.242	2.295	**0.298**	**0.300**
1000	**1.000**	1.415	1.412	0.500	0.503	**0.288**	**0.280**
Relative Error in Estimating Q Matrix							
50	**0.536**	2.157	1.511	3.403	4.405	**0.710**	**0.240**
100	**0.391**	1.214	0.890	1.349	**0.711**	**0.761**	**0.383**
1000	**0.098**	0.168	0.103	0.158	**0.088**	**0.110**	**0.075**

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



Current and future work

- Climate Scientists should be skeptical when using MLE based estimates, even when abundant data is available.
  - Theoretical convergence may be slow to realize.
  - Judging convergence on forecast skill is not a measure for spectra of the parameters.
- A Bayesian framework can improve forecasts by incorporating prior beliefs about the structure of estimated parameters.
  - Spectra and sparsity can both be better approximated using monte carlo techniques.

References

[1] C. Penland, P. Sardeshmukh
The Optimal Growth of Tropical Sea Surface Temperature Anomalies
Journal of Climate, (8):1999–2023, 1995.

[2] J.E. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J.M. Arblaster, S.C. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J.F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, M. Vertenstein
The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability
Bulletin of the American Meteorological Society, doi: 10.1175/BAMS-D-13-00255.1, 96, 1333-1349

[3] J. Piironen, A. Vehtari
Sparsity information and regularization in the horseshoe and other shrinkage priors
Electronic Journal of Statistics, Volume 11, Number 2 (2017), 5018-5051

[4] M. Alexander, L. Matrosova, C. Penland, J. Scott, P. Chang
Forecasting Pacific SSTs: Linear Inverse Model Predictions of the PDO
J. Climate, 21, (2007), 385-402

*Dallas Foster <fostdall@oregonstate.edu>