Supplementary Material with the article "Dimension-Reduced Modeling of Spatio-Temporal Processes"

Jenný Brynjarsdóttir and L. Mark Berliner

March 11, 2014

1 Introduction

This is supplementary material to the article "Dimension-reduced modeling of spatio-temporal processes with an application to statistical downscaling". In section 2 we provide details regarding the application found in Section 4 in that article. We show exploratory data analysis (Section 2.1), a detailed description of the Bayesian model and prior selection in (Section 2.2) and address model the issue of model adequacy (Section 2.3). Section 3 outlines the Gibbs sampler.

2 Statistical downscaling of temperatures over the Antarctic

2.1 Exploratory data analysis

Let $Y_{l,t,i}$ be the (temporally) centered 2-meter temperature model output of the Polar MM5 from season l, year t and location i and let $X_{m,t,j}$ be the centered ERA-40 2-meter temperatures for month m, year t and location j. The number of spatial locations are $N_Y = 14641$ for the Polar MM5 data and $N_X = 2736$ for the ERA-40 data.

We have data from the fall season (March) of 1979 through the fall season (May) of 2002; a total of 93 seasons (279 months). We leave out the last year of data so that we can compare them to our predictions for that period. We therefore obtain the basis vectors using only the first 89 seasons (267 months), i.e. through fall (May) of 2001.

Note that seasons and months start and end at different years since we have no summer data for the first year (1979) and no winter or spring data the 23$^{\text{rd}}$ year (2001). Let τ_l and T_l be the the first and last year we have available data for season l, $l = s, f, w, p$. For example, we have summer data for years $t = 1980, \ldots, 2001$ so $\tau_s = 1980$ and $T_s = 2001$. Similarly, let τ_m and T_m be the first and last year we have data for month m, $m = 1, \ldots, 12$.

We define data vectors and data matrices as follows:

$\mathbf{Y}_{l,t}$ is an N_Y-dimensional vector of centered seasonal Polar MM5 2-meter surface temperatures, indexed by year $t = \tau_l, \ldots, T_l$ and season $l = s, f, w, p$.

Y_l is a $N_Y \times (T_l - \tau_l + 1)$ dimensional data matrix with columns $\mathbf{Y}_{l,\tau_l}, \ldots, \mathbf{Y}_{l,T_l}$.

$\mathbf{X}_{m,t}$ is an N_X-dimensional vector of centered monthly ERA 2-meter surface temperatures, indexed by year $t = \tau_m, \ldots, T_m$ and month $m = 1, \ldots, 12$.

X_m is a $N_X \times (T_m - \tau_m + 1)$ dimensional data matrix with columns $\mathbf{X}_{m,\tau_m}, \ldots, \mathbf{X}_{m,T_m}$.

Centered data are obtained by subtracting the seasonal (Polar MM5) or monthly (ERA-40) means at each location. This means that

$$\sum_{t=\tau_l}^{T_l} Y_{l,t,i} = 0 \quad \text{and} \quad \sum_{t=\tau_m}^{T_m} X_{m,t,j} = 0 \quad \text{for all } l, m, i, j \; . \tag{1}$$

Figure 1 shows the centered temperature fields for fall and winter 1986. Examining these maps for all 23 years (not shown) we noticed that most centered temperatures over the ocean are close to 0K, indicating that temperatures at these locations do not vary much over time. Locations on land/permanent ice vary more, especially those close to the coast of Antarctica. Furthermore, the spatial patterns in the centered data seem to be more alike within seasons and within months (i.e. across years) than across seasons and months.

To allow for different spatial structures between seasons in the centered Polar MM5 data, we construct EOF basis vectors for the centered Polar MM5 data separately for each season. This means that for each season the 22 years (23 for fall season) are treated as repeated measurements as opposed to treating all 89 seasons as repetitions of one single process. We obtain matrices of four EOFs, $\mathcal{U}_l$, by performing singular value decomposition (SVD) on the data matrices Y_s, Y_f, Y_w and Y_p.

We have three months of centered ERA-40 data for every season. We obtain OMCPs separately for each month, using EOFs from the corresponding season. For example,

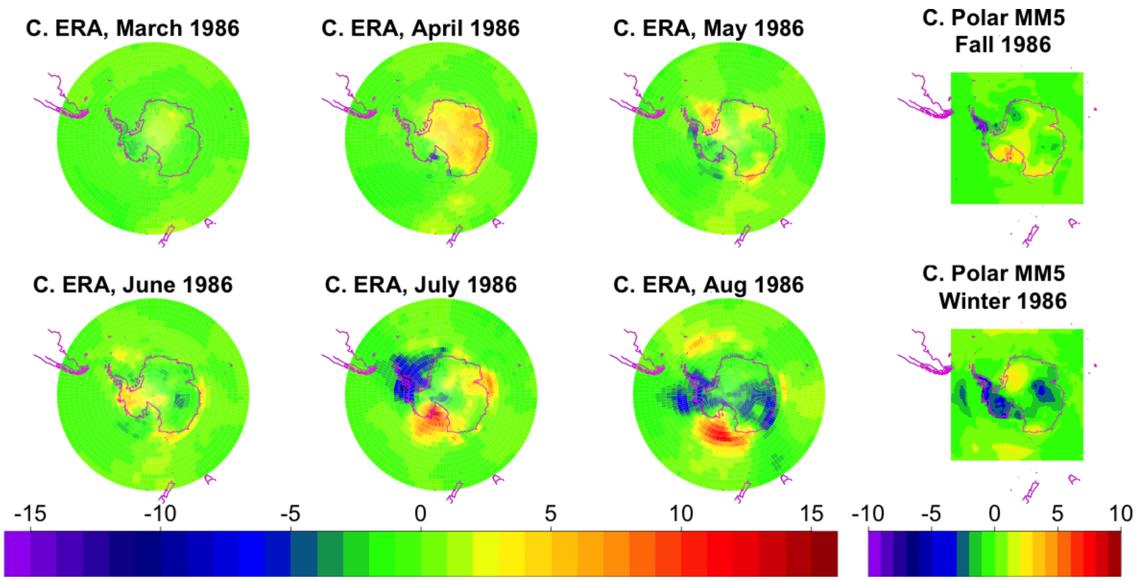

Figure 1: Centered temperature fields, ERA-40 (circles) and Polar MM5 (squares) in Kelvin, fall and winter season 1986. The color palettes are separate for the ERA-40 and Polar MM5 data.

the first OMCPs for March, April and May are

$$\mathbf{v}_{3,1} = \frac{X_3 Y_f' \mathbf{u}_{f,1}}{||X_3 Y_f' \mathbf{u}_{f,1}||}, \quad \mathbf{v}_{4,1} = \frac{X_4 Y_f' \mathbf{u}_{f,1}}{||X_4 Y_f' \mathbf{u}_{f,1}||} \quad \text{and} \quad \mathbf{v}_{5,1} = \frac{X_5 Y_f' \mathbf{u}_{f,1}}{||X_5 Y_f' \mathbf{u}_{f,1}||} \tag{2}$$

where $\mathbf{u}_{f,1}$ is the first EOF for the fall season. The patterns $\mathbf{v}_{3,k}$, $\mathbf{v}_{4,k}$, $\mathbf{v}_{5,k}$ and $\mathbf{u}_{f,k}$ are referred to as *group* k for fall season. The rationale for modeling each month separately as opposed to using seasonally averaged ERA-40 data was twofold. Firstly, there are stronger relationship between monthly temperatures than seasonal, that hopefully would result in a better model of the temporal evolution. Secondly, this way one could do downscaling (prediction) of the Polar MM5 seasonal data even if only one or two of the ERA-40 months had been observed.

Figure 2 shows the proportion of the sample covariance between the amplitudes is lost by orthogonalizing the MCPs. For most of the first three patterns we lose less than 20% of the sample covariance while some of the 4th and 5th patterns lose as

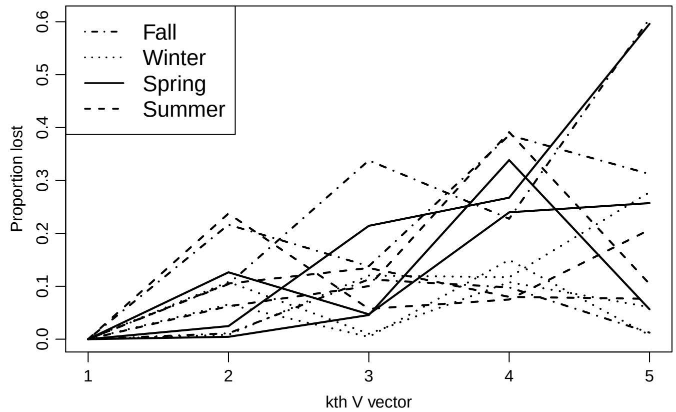

Figure 2: Proportion of the sample covariance between amplitudes lost by orthogonalizing the MCPs.

much as 60%.

2.1.1 Exploratory analysis of amplitudes

Before we build a model of the amplitudes we perform some exploratory data analysis. We estimate amplitude vectors using regular least squares estimates. Let U_l be a matrix that contains the first K_Y EOFs and let V_m be a matrix that contains the first K_X OMCPs. Note that since the basis vectors are orthonormal, we have $U_l'U_l = I$ and $V_m'V_m = I$. The estimated amplitude vectors are calculated as follows:

$$\hat{\mathbf{a}}_{l,t} = (U_l'U_l)^{-1}U_l'\mathbf{Y}_{l,t} = U_l'\mathbf{Y}_{l,t} \quad \text{and} \quad \hat{\mathbf{b}}_{m,t} = (V_m'V_m)^{-1}V_m'\mathbf{X}_{m,t} = V_m'\mathbf{X}_{m,t} \tag{3}$$

We have 23 estimates of each amplitude vector and we can explore the relationships between them via simple scatterplots.

First, we consider the relationship between an amplitude vector $\hat{\mathbf{a}}_{l,t}$ and each of the three $\hat{\mathbf{b}}_{m,t}$ amplitude vectors from the same season. The goal here is to examine possible models of the form

$$\mathbf{a}_{l,t} = H_l \begin{bmatrix} \mathbf{b}_{m_{l1},t} \\ \mathbf{b}_{m_{l2},t} \\ \mathbf{b}_{m_{l3},t} \end{bmatrix} + \mathbf{e}_Y \tag{4}$$

where m_{l1}, m_{l2} and m_{l3} are the three months in season l. We plotted each of the seasonal $\hat{\mathbf{a}}_{l,t}$ against each of the three monthly $\hat{\mathbf{b}}_{m_{li},t}$. Since we chose to use four basis vectors we plotted the four amplitudes $\hat{a}_{l,t,k}$, $k = 1, \ldots, 4$, against each of the four $\hat{b}_{m_{li},j}$, $j = 1, \ldots, 4$, amplitudes (192 scatterplots in total). Examples of these plots are shown in Figure 3. In general we concluded that amplitudes from the same group i.e., when $k = j$, have strong linear relationships (e.g. first row in Figure 3). This is not surprising since OMCP basis vectors are designed so that these amplitudes would have high covariance. When we plotted higher order EOFs versus lower order OMCPs ($k < j$), we detected virtually no relationships (e.g. $k = 1$, $j = 2$ and $k = 2$, $j = 3$ in Figure 3). In some cases the lower order EOFs versus higher order OMCPs ($k > j$) show linear relationships but most are weaker than the ones on the diagonal (e.g. $k = 2$, $j = 3$ in Figure 3). These exploratory analyses indicate that when we model the amplitude a_{l,t,k_Y} conditional on the three amplitude vectors $\mathbf{b}_{m_1,t}$, $\mathbf{b}_{m_2,t}$ and $\mathbf{b}_{m_3,t}$ we can make do with only the amplitudes within the same group, i.e. $k = j$.

We also examined multiple linear regression of the estimated amplitudes EOF $\hat{a}_{l,t,k}$ on estimated OMCP amplitudes from the same group. For example, for fall season and fixed k we fitted the following regression model:

$$\hat{a}_{f,k,t} = H_{3,k}\hat{b}_{3,k,t} + H_{4,k}\hat{b}_{4,k,t} + H_{5,k}\hat{b}_{5,k,t} + \epsilon_t \tag{5}$$

The $\hat{b}_{m,t,k}$ amplitudes explain a big part of the variation in $\hat{a}_{l,t,k}$, R^2 of these regressions range from about 56% to 86%. Furthermore, in many cases the last month of the season is the most significant month. The estimated coefficients, $\hat{H}_{m,k}$, where mostly

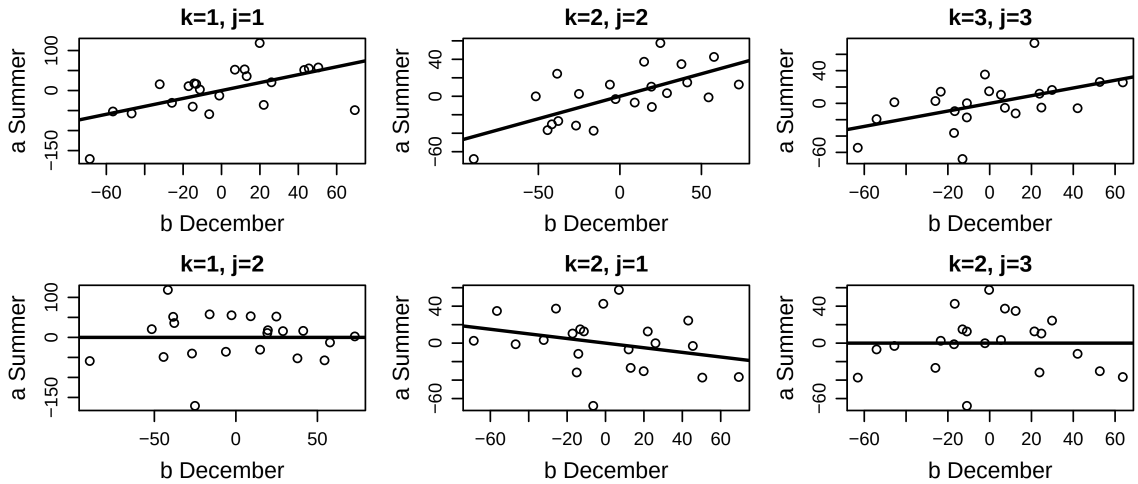

Figure 3: Estimated Polar MM5 EOF amplitudes for summer, $\hat{a}_{s,t,k}$, plotted against estimated ERA-40 OMCP amplitudes for December, $\hat{b}_{12,t,j}$. Each point on the graphs represents one year t (22 years). The lines show the simple regression line though the points.

between 0 and 1.

Next we consider the relationship between the OMCP amplitudes and the OMCP amplitude vectors from the previous month. The goal here is to explore the temporal aspect, i.e. the distribution $\left[\mathbf{b}_{m,t}|\mathbf{b}_{m-1,t},\boldsymbol{\theta}\right]$. Figure 4 shows examples of OMCP amplitudes plotted against OMCP amplitudes for the month before. A few of these graphs showed a moderately strong linear relationship between amplitudes $\hat{b}_{m,t,k}$ and $\hat{b}_{m-1,t,j}$, but we did not detect any pattern of some combinations of k and j consistently showing stronger linear relationships than others. It therefore seems that when modeling amplitudes $\hat{b}_{m,t,k}$ given amplitudes the month before that there is reason to include all the $\hat{b}_{m-1,t,j}$, $j = 1, \ldots, 4$, amplitudes.

We also considered multiple linear regression of the estimated OMCP amplitudes $\hat{b}_{m,t,k}$ on the first four OMCP amplitudes from the month before, i.e. $\hat{b}_{m-1,t,1}$, $\hat{b}_{m-1,t,2}$, $\hat{b}_{m-1,t,3}$ and $\hat{b}_{m-1,t,4}$. There seem to be some predictive potential in the previous month's amplitudes, although in seems rather weak in some cases. The range of R^2

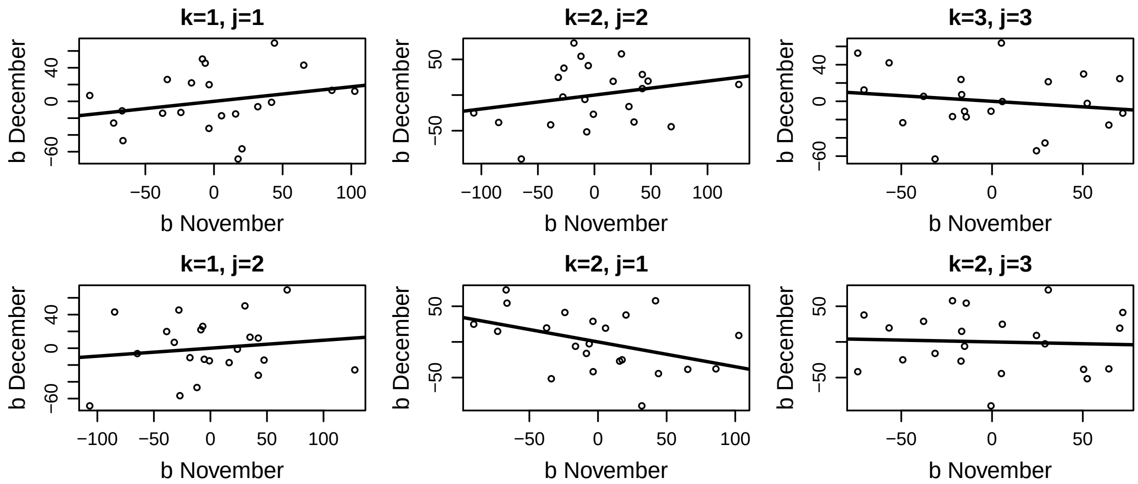

Figure 4: Estimated December OMCP amplitudes, $\hat{b}_{12,t,k}$, plotted against estimated November OMCP amplitudes, $\hat{b}_{11,t,j}$. The lines show the simple regression line though the points.

is quite wide, from 4% to 73%, and most of the R^2 values are under 50%.

Finally, in an effort to improve prediction of OMCP amplitudes, we investigate whether there could be an autoregressive relationship within the same month but between years. We regressed estimated OMCP amplitudes $\hat{b}_{m,t,k}$ on both the whole OMCP amplitude vector from the month before ($\hat{b}_{m-1,t,1}$, $\hat{b}_{m-1,t,2}$, $\hat{b}_{m-1,t,3}$ and $\hat{b}_{m-1,t,4}$) and the same amplitude one year earlier ($\hat{b}_{m,t-1,k}$). Only in few cases is the added (previous year) amplitude significant, most notably for March amplitudes. Note that March is a beginning of a new season and so this suggests that prediction of March amplitudes from February amplitudes (different season) could be improved by also including March amplitudes form the year before.

2.2 Bayesian hierarchical model

We discuss separately the three layers of the hierarchical model: *data model*, *process model* and *parameter model*.

2.2.1 Data model

The data model defines the distribution of the centered temperature vectors given the basis vectors and their amplitudes. We chose to model only the centered data. We could include a parametric function to model the mean, e.g. by using the elevation data. However, our focus in this paper is on the dimension reduction techniques so we leave that part for future work. We assume that the observations are conditionally independent given the amplitude vectors $\mathbf{a}_{l,t}$ and $\mathbf{b}_{m,t}$ and that

$$\begin{aligned} \mathbf{Y}_{l,t}|\mathbf{a}_{l,t}, R_{l,t} &\sim N\left(U_l\mathbf{a}_{l,t}, R_{l,t}\right) \quad \text{for} \;\; l \in \{s,f,w,p\},\; t = \tau_l,\ldots,T_l \quad \text{and} \\ \mathbf{X}_{m,t}|\mathbf{b}_{m,t}, S_{m,t} &\sim N\left(V_m\mathbf{b}_{m,t}, S_{m,t}\right) \quad \text{for} \;\; m = 1,\ldots,12,\; t = \tau_m,\ldots,T_m+1\,. \end{aligned} \tag{6}$$

Note that we include the last year of ERA-40 data but not the last year of the Polar MM5 data, since we wish to obtain predictions of the Polar MM5 fields for that time period. The covariance matrices $R_{l,t}$ and $S_{m,t}$ are unknown and U_l and V_m are the first four EOFs and OMCPs for season l and month m.

2.2.2 Process model

The process model defines the joint distribution of the amplitude vectors $\mathbf{a}_{l,t}$ and $\mathbf{b}_{m,t}$. We assume that the Polar MM5 amplitude vectors, $\mathbf{a}_{l,t}$, are conditionally independent given the ERA amplitude vectors, $\mathbf{b}_{m_{l1},t}$, $\mathbf{b}_{m_{l2},t}$ and $\mathbf{b}_{m_{l3},t}$, where the index m_{lk} denotes the k^{th} month within season l. For all seasons we assume a normal distribution for the Polar MM5 amplitudes given the ERA amplitudes:

$$\mathbf{a}_{l,t}|\mathbf{b}_{m_{l1},t}, \mathbf{b}_{m_{l2},t}, \mathbf{b}_{m_{l3},t}H_l, C_l \sim N\left(H_l \begin{bmatrix} \mathbf{b}_{m_{l1},t} \\ \mathbf{b}_{m_{l2},t} \\ \mathbf{b}_{m_{l3},t} \end{bmatrix},\; C_l\right) \tag{7}$$

for $l \in \{s, f, w, p\}$ and $t = \tau_l, \ldots, T_l + 1$, where C_l are unknown 4×4 dimensional covariance matrices and

$$H_l = \begin{pmatrix} H_{m_{l1}} & H_{m_{l2}} & H_{m_{l3}} \end{pmatrix} . \tag{8}$$

Each of the H_m matrices are 4×4 dimensional. Based on the preliminary data analysis in Section 2.1 we make the H_m matrices diagonal. Note that we include the amplitude vectors for the last year $\mathbf{a}_{l,T_l+1}$ to facilitate prediction of the Polar MM5 process of the last year.

The $\mathbf{b}_{m,t}$ amplitude vectors are modeled by a first order Markov process with normally distributed errors:

$$\mathbf{b}_{m,t} | \mathbf{b}_{m-1,t}, B_m, D_m \sim N\big(B_m \mathbf{b}_{m-1,t},\ D_m\big) \tag{9}$$

for $m = 1, \ldots, 12$ and $t = \tau_m, \ldots, T_m + 1$ where D_m is an unknown 4×4 dimensional covariance matrix. Note that here we use the notation $\mathbf{b}_{0,t}$ for $\mathbf{b}_{12,t-1}$. We let the B_m matrices have full structure, that is we have K_X^2 unknowns for each B_m matrix.

Exploratory analysis of estimated amplitudes (see Section 2.1) indicated that, at least for some months, it might be beneficial to model each $\mathbf{b}_{m,t}$ amplitude vector dependent on both the month before and also the same month one year earlier. However, this only applied for a few months and amplitudes, so for the sake of parsimony, we do not include such dependence structure here.

Finally, we use a normal prior on the first (i.e. February 1979) ERA-40 amplitude vector,

$$\big[\mathbf{b}_{2,1979}\big] = N\big(\boldsymbol{\mu}_b, \Sigma_b\big) . \tag{10}$$

We actually have ERA-40 data from this month and we set the prior mean equal to the estimated amplitudes,

$$\boldsymbol{\mu}_b = V_2 X_{2,1979} = \hat{\mathbf{b}}_{2,1979} = (-26.784, 12.094, -5.186, 5.562)' \,. \tag{11}$$

The estimated standard error of $\hat{\mathbf{b}}_{2,1979}$ is 0.72 (same for all k). In order to avoid being too concentrated in our prior selection, we use the considerably larger prior standard deviation and se $\Sigma_b = 10^2 I_{K_X}$.

2.2.3 Parameter model

The parameter model defines prior distributions on unknown parameters introduced in the data and process models. There are three groups of parameters we need to consider: (1) the transition matrices H_m and B_m; (2) the process model covariance matrices C_l and D_m; and (3) the data model covariance matrices $R_{l,t}$ and $S_{m,t}$.

We do not have apriori information about hyperparameters so we use the exploratory analysis in Section 2.1 to guide our selection. The situation where hyperparameters are estimated from data is referred to as *empirical* Bayes (Berger (1985); Section 4.5) and is not strictly a fully Bayesian approach. Applying Bayes theorem with hyperparameters that are estimated from data ignores errors introduced in the estimation. These errors will therefore not be reflected in the posterior distribution. To partially mitigate this problem we selected prior densities wider than what estimated hyperparameters would have suggested. Ideally we should perform sensitivity analysis on the selected hyperparameters, i.e. examine how sensitive our conclusions are to the values of the hyperparameters.

(1) Transition matrices H_m and B_m: Let $\mathbf{h}_m = (H_{m,1}, \ldots, H_{m,K_Y})'$ be a vector that contains the diagonal elements of H_m (the non-zero elements). Let $vec(B_m)$ be the vectorization of B_m, i.e. $vec(B_m)$ is a column vector with the columns of B_m stacked on top of each other. We assume that the H_m and B_m matrices are apriori independent and we use the following prior distributions for the elements of the H_m and B_m matrices:

$$\begin{aligned} \left[\mathbf{h}_m\right] &= N(\boldsymbol{\mu}_{1,m}, \Sigma_{1,m}), \quad m = 1, \ldots, 12 \quad \text{and} \\ \left[vec(B_m)\right] &= N(\boldsymbol{\mu}_{2,m}, \Sigma_{2,m}), \quad m = 1, \ldots, 12\ . \end{aligned} \tag{12}$$

The parameters $\boldsymbol{\mu}_{1,m}$, $\Sigma_{1,m}$, $\boldsymbol{\mu}_{2,m}$ and $\Sigma_{2,m}$ are constant *hyperparameters* and need to be specified. There is no prior information available on the $H_{m,k}$ parameters and without seeing any examples of the $a_{f,k,t}$ and $b_{m,k,t}$ amplitudes, it is difficult to even postulate about the scale of $H_{m,k}$. We therefore allow ourselves a peak at the data to guide the selection of $\boldsymbol{\mu}_{1,m}$ and $\Sigma_{1,m}$. We saw in Section 2.1 that estimates of the elements of H were almost all between 0 and 1. With that in mind we set the prior mean equal to 0.5 for all m and k and the prior standard deviation equal to 4, i.e. $\boldsymbol{\mu}_{1,m} = 0.5\mathbf{1}_{K_Y}$ and $\Sigma_{1,m} = 4^2 I_{K_Y}$ for all m. We note that the prior is concentrated on a considerably wider than the range of $H_{m,k}$ values than was indicated in the EDA in Section 2.1. Similar considerations led us to set $\boldsymbol{\mu}_{2,m} = \mathbf{0}_{K_X}$ and $\Sigma_{2,m} = 4^2 I_{K_X}$ for all m.

(2) Process model covariance matrices C_l and D_m: We use Inverse-Wishart priors for C_l and D_m:

$$C_l \sim IW_{\nu_l}(W_l), \quad l = s, f, w, p, \quad \text{and} \quad D_m \sim IW_{\nu_m}(W_m), \quad m = 1, \ldots, 12 \tag{13}$$

The expected value of an $IW_\nu(W)$ distributed, $k \times k$ matrix is $\frac{1}{\nu-k-1}W$. Furthermore, the larger ν is the tighter the distribution is around its mean. To select a vague prior, we select $\nu_m = K_X + 2$ and $\nu_l = K_Y + 2$ which means that the mean of the inverse-Wishart distributions will be W_l and W_m. From the regressions in Section 2.1 we found that the residual errors are no larger than 60 so we select $W_l = 60^2 I_{K_Y}$ for all seasons l. For same reasons we select $W_m = 70^2 I_{K_Y}$ for all seasons m.

(3) Data model covariances $R_{l,t}$ and $S_{m,t}$: The covariance matrices in the data model are very large; $R_{l,t}$ are $(N_Y \times N_Y)$-dimensional and $S_{m,t}$ are $(N_X \times N_X)$-dimensional. They account for many sources of variation that arise when modeling the observations given the coefficient vectors $\mathbf{a}_{l,t}$ and $\mathbf{b}_{m,t}$. These include measurement error, small non-time-varying variation in the processes and error due to the dimension reduction itself. One way to avoid the high dimensional covariance matrices is to assume (conditional) independence and take $R_{l,t}$ and $S_{m,t}$ as diagonal matrices. But in light of the many complex sources of variation in $R_{l,t}$ and $S_{m,t}$ that seems too simplistic. Instead we apply an approach used in Berliner et al. (2000) which is both computationally appealing and attempts to account for some of the structure left over after the dimension reduction. The idea is to form $R_{l,t}$ and $S_{m,t}$ by using a few of the next basis vectors to capture some of what is left of the spatial structure in $\mathbf{Y}_{l,t}$ and $\mathbf{X}_{m,t}$. In principle $R_{l,t}$ and $S_{m,t}$ could be time-varying (different for each year t), but here we assume that they only vary between seasons or months and can therefore be denoted as R_l and S_m.

Let $\tilde{U}_l$ be a matrix that contains columns $K_Y+1, \ldots, L_Y$ of $\mathcal{U}_l$ and let $\tilde{D}_l$ be a diagonal matrix containing the corresponding eigenvalues, $d_{l,j}$ (i.e. the variances associated with these EOFs). As in Berliner et al. (2000) we set

$$R_l = r_l \left(c_l I_{N_Y} + \sum_{k=K_Y+1}^{L_Y} d_{l,k} \mathbf{u}_{l,k} \mathbf{u}'_{l,k} \right) = r_l \left(c_l I_{N_Y} + \tilde{U}_l \tilde{D}_l \tilde{U}'_l \right) \equiv r_l \tilde{R}_w \ . \tag{14}$$

Here, c_l is a constant set to be the total sample variance left after accounting for the first L_Y EOFs. That is,

$$c_l = \sum_{k=L_Y+1}^{N_Y} d_{l,k} \quad \text{for } l = s, f, w, p \ . \tag{15}$$

where $d_{l,k}$ are the eigenvalues of $S_Y = \frac{1}{T_l - 1} Y_l Y'_l$. Note that, although $\tilde{U}_l \tilde{D}_l \tilde{U}'_l$ is not of full rank, by adding the diagonal matrix $c_l I_{N_Y}$ the covariance matrix R_l is guaranteed to be non-singular. Using this approach we have reduced the unknown $N_Y \times N_Y$ covariance matrix R_l down to a single unknown scalar r_l while still capturing extra spatial structure through the extra basis vectors. Furthermore, computations never require one to actually store or do calculations with R_l, we only need to work with $\tilde{U}$ and the $d_{l,k}$ scalars.

We use a similar approach for S_m. Let $\tilde{V}_m$ be a matrix that contains columns $K_X + 1$ through L_Y of $\mathcal{V}_m$. The $\mathcal{V}_m$ are not eigenvectors but we can still obtain the variance explained by each basis vector $\mathbf{v}_{m,k}$ which we call $d_{m,k}$. With that in mind we set the diagonal elements of $\tilde{D}_m$ equal to the sample variance explained by the corresponding basis vector and c_m to be the total variance left in X after accounting for the first L_X basis vectors. We then we set

$$\begin{aligned} S_m &= s_m \left(c_m I_{N_X} + \sum_{j=K_X+1}^{L_X} d_{m,j} \mathbf{v}_{m,j} \mathbf{v}'_{m,,j} \right) \\ &= s_m \left(c_m I_{N_X} + \tilde{V}_m \tilde{D}_m \tilde{V}'_m \right) \equiv s_m \tilde{S}_m \ . \end{aligned} \tag{16}$$

This approach reduces the $N_X \times N_X$ dimensional matrix S_m down to one scalar, s_m.

The unknown scalars r_l and s_m are assumed to be apriori independent and are assigned conjugate inverse gamma priors:

$$
\begin{aligned}
\left[r_l\right] &= IG(\alpha_1, \beta_1) \text{ for } l = s, f, w, p \quad \text{ and} \\
\left[s_m\right] &= IG(\alpha_2, \beta_2) \text{ for } m = 1, \ldots, 12 \; .
\end{aligned} \tag{17}
$$

We set proper, but vague priors on r_l and s_m. In particular we select $\alpha_1 = \alpha_2 = 2$ and $\beta_1 = \beta_2 = 1$ which means that the prior mean is 1 but the prior distribution has an infinite variance. The reason for picking the mean as 1 is that we have included the total variance of the data within the data model (through the c_l and c_m in R_l and S_m).

The computational advantages of this approach, stems from the basis vectors being orthogonal and the inverses of R_l and S_m can be easily obtained. For example,

$$
R_l^{-1} = \frac{1}{r_l}\left(\frac{1}{c_l} I_{N_Y} - \tilde{U}_l \Lambda_l \tilde{U}_l'\right) \tag{18}
$$

where Λ_l is a diagonal matrices with diagonal elements

$$
\frac{d_{l,k}}{c_l(c_l + d_{l,k})} \text{ for } k = K_Y + 1, \ldots, L_Y \; . \tag{19}
$$

Furthermore, we note that R_l appear in the Gibbs sampler as $U_l' R_l^{-1} \mathbf{Y}_{l,t}$, $U_l' R_l^{-1} U_l$ which can be written as

$$
U_l' R_l^{-1} \mathbf{Y}_{l,t} = U_l' \frac{1}{r_l}\left(\frac{1}{c_l} I_{N_Y} - \tilde{U}_l \Lambda_l \tilde{U}_l'\right) \mathbf{Y}_{l,t} = \frac{1}{r_l c_l} U_l' \mathbf{Y}_{t,l}, \quad \text{ and} \tag{20}
$$

$$
U_l' R_l^{-1} U_l = U_l' \frac{1}{r_l}\left(\frac{1}{c_l} I_{N_Y} - \tilde{U}_l \Lambda_l \tilde{U}_l'\right) U_l = \frac{1}{r_l c_l} I_{K_Y} \; . \tag{21}
$$

In the calculations above we used the fact that $U_l' \tilde{U}_l$ is a matrix with all elements

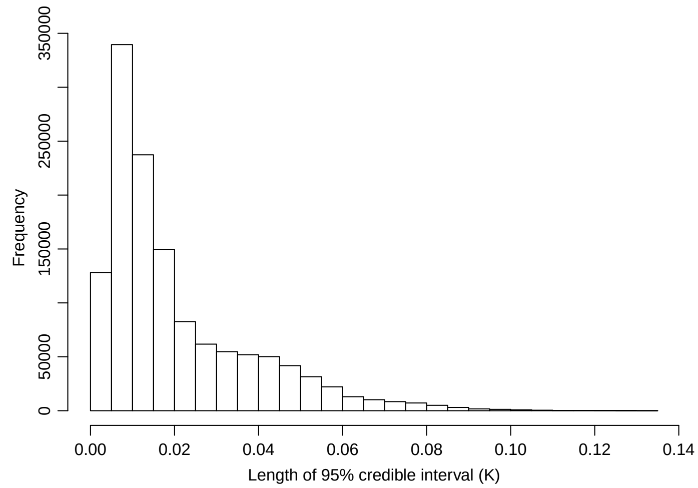

Figure 5:

equal to zero due to the orthogonality of the basis vectors.

2.3 Model Adequacy

In the paper we address the issue of model adequacy by obtaining point wise credible intervals for each $Y_{l,t,i}$ and compared to the observations, i.e. for observations that were use to fit the model. To shine a light on how much uncertainty reduction happens from the prior to the posterior we look at the lengths of the 95% point-wise credible intervals and compare to the prior standard deviation. A histogram of the lengths of the 95% point-wise credible intervals is shown in Figure 5. We note that the lengths of the point-wise prediction intervals are all less than about 0.14 K.

To find the prior covariance matrix of $\mathbf{Y}_{l,t}$ implied by the hierarchical model we can utilize iterative expectations:

$$\begin{aligned}\mathrm{Cov}(\mathbf{Y}_{l,t}) &= \mathrm{Cov}\left(E(\mathbf{Y}_{l,t}|\boldsymbol{\theta})\right) + E\left(\mathrm{Cov}(\mathbf{Y}_{l,t}|\boldsymbol{\theta})\right) \\ &= \mathrm{Cov}\left(U_l\mathbf{a}_{l,t}|C\right) + E\left(R_{l,t}|r_l\right) \\ &= U_lCU_l' + r_l\tilde{R}_{l,t}\end{aligned}$$

if C and r_l are known. Note that U_l and $\tilde{R}_{l,t}$ are constant. Plugging in the prior means for C and r_l:

$$C = 60^2 I_4 \quad \text{and} \quad r_l = 1$$

we can obtain the diagonal of $\mathrm{Cov}(\mathbf{Y}_{l,t})$. The average implied prior standard deviations for the four season are

$$34.96,\ 27.93,\ 25.91,\ \text{ and }\ 35.79 \text{ Kelvin.}$$

So clearly the posterior intervals are much smaller than the prior intervals showing that reduction of uncertainty from prior to posterior is substantial.

3 Gibbs Sampler

We provide full conditional distributions for Gibbs samplers used to obtain MCMC samples for inference for the models in Section 2.2. Table 1 gives an overview of the unknown parameters in the model in Section 2.2. In the following use the notation "$\big[X|\mathrm{rest}\big]$" to denote the conditional distribution of a random variable X given both the data and all other unknown parameters in the model.

Parameter		Dimension	
$\mathbf{a}_{l,t}$	seasonal Polar MM5 amplitude vectors	$93 \times K_Y$	$= 372$
$\mathbf{b}_{m,t}$	monthly ERA amplitude vectors	$93 \times 3 \times K_X$	$= 1116$
$\mathbf{b}_{2,1}$	Initial ERA amplitude vector	K_X	$= 4$
$\mathbf{h}_m$	Diagonal of H_m, from the process model	$12 \times K_Y$	$= 48$
B_m	Transition matrices of the process model	$12 \times K_X^2$	$= 192$
C_l	variances for $\mathbf{a}_{l,t}$ vectors	$4 \times K_Y(K_Y+1)/2$	$= 49$
D_l	variances for $\mathbf{b}_{m,t}$ vectors	$12 \times K_X(K_Y+1)/2$	$= 120$
r_l	Parameters the data model covariance matrices	4	$= 4$
s_m	Parameters the data model covariance matrices	12	$= 12$
	Total		$= 1908$

Table 1: List of unknown parameters in the model. The last column shows the number of parameters for the chosen number of basis vectors, $K_Y = 4$ and $K_X = 4$. Note that, in comparison, the total number of observations is 14641*89 + 2736*93*3 = 2,066,393.

Polar MM5 amplitudes $\mathbf{a}_{l,t}$:

Let $\mathbf{b}_{l,t}$ denote the three ERA amplitude vectors that correspond to season l. The full conditional distribution $\left[\mathbf{a}_{l,t}|\text{rest}\right]$ for each year and season is $N(A^{-1}\mathbf{d}, A^{-1})$ with

$$A = U_l' R_l^{-1} U_l + C_l^{-1} = \frac{1}{r_l c_l} I_{K_Y} + C_l^{-1} \quad \text{and}$$
$$\mathbf{d} = U_l' R_l^{-1} \mathbf{Y}_{l,t} + C_l^{-1} H_l \mathbf{b}_{l,t} = \frac{1}{r_l c_l} U_l' \mathbf{Y}_{l,t} + C_l^{-1} H_l \mathbf{b}_{l,t} \tag{22}$$

where U_l, c_l and $\mathbf{Y}_{l,t}$ are constants and $\frac{1}{c_l} U_l' \mathbf{Y}_{l,t}$ can be calculated (and stored) outside of the MCMC iterations.

Polar MM5 amplitudes $\mathbf{a}_{l,t}$, the last year:

The full conditional distribution of $\mathbf{a}_{l,t}$ for timepoints where we do not have Polar MM5 data is simply the process model for $\mathbf{a}_{l,t}$:

$$\left[\mathbf{a}_{l,t}|\text{rest}\right] = \left[\mathbf{a}_{l,t}|\mathbf{b}_{l,t}, H_l, C_l\right] = N(H_l \mathbf{b}_{l,t}, C_l)\ . \tag{23}$$

ERA amplitudes $\mathbf{b}_{m,t}$:

The full conditional distribution $\left[\mathbf{b}_{m,t}|\text{rest}\right]$ is $N(A^{-1}\mathbf{d}, A^{-1})$ with

$$
\begin{aligned}
A &= V_m' S_m^{-1} V_m + B_{m+1}' D_{m+1}^{-1} B_{m+1} + H_m' C_l^{-1} H_m + D_m^{-1} \\
\mathbf{d} &= V_m' S_m^{-1} \mathbf{X}_{m,t} + B_{m+1}' D_{m+1}^{-1} \mathbf{b}_{m+1,t} + H_m' C_l^{-1} (\mathbf{a}_{l,t} - \mathbf{b}) + D_m^{-1} B_m \mathbf{b}_{m-1,t} \ .
\end{aligned} \tag{24}
$$

Some of the calculations can be simplified by noting that we have

$$
V_m' S_m^{-1} V_m = \frac{1}{s_m c_m} I_{K_X} \quad \text{and} \quad V_m' S_m^{-1} \mathbf{X}_{m,t} = \frac{1}{s_m c_m} V_m' \mathbf{X}_{l,t} \ . \tag{25}
$$

V_m, c_m and $\mathbf{X}_{m,t}$ are constants and $\frac{1}{c_m} V_m' \mathbf{X}_{m,t}$ can be calculated (and stored) outside of the MCMC iterations.

The last ERA amplitude $\mathbf{b}_{m,t}$:

The full conditional distribution of the last $\mathbf{b}_{m,t}$ amplitude vector is different from others because there is no $\mathbf{b}_{m+1,t}$ amplitude after it. Hence we get that $\left[\mathbf{b}_{m,t}|\text{rest}\right]$ is $N(A^{-1}\mathbf{d}, A^{-1})$ with

$$
\begin{aligned}
A &= V_m' S_m^{-1} V_m + H_m' C_l^{-1} H_m + D_m^{-1} \\
\mathbf{d} &= V_m' S_m^{-1} \mathbf{X}_{m,t} + H_m' C_l^{-1} (\mathbf{a}_{l,t} - \mathbf{b}) + D_m^{-1} B_m \mathbf{b}_{m-1,t}
\end{aligned} \tag{26}
$$

The first ERA amplitude $\mathbf{b}_{2,1}$:

The full conditional distribution of the first ERA amplitude is different from the others because there is neither data, nor a $\mathbf{a}_{l,t}$ amplitude at that timepoint and no $\mathbf{b}_{m-1,1}$ amplitude before it. Hence we get that $\left[\mathbf{b}_{2,1}|\text{rest}\right]$ is $N(A^{-1}\mathbf{d}, A^{-1})$ with

$$
A = B_3' D_3^{-1} B_3 + \Sigma_b^{-1} \quad \text{and} \quad \mathbf{d} = B_3' D_3^{-1} \mathbf{b}_{3,1} + \Sigma_b^{-1} \boldsymbol{\mu}_b \tag{27}
$$

Transition matrices H_m:

Recall that the transition matrices H_m are diagonal matrices and that $\mathbf{h}_m = (H_{m,1}, \ldots, H_{m,K_Y})'$ is a vector of the diagonal elements of H_m. Also recall the $K_Y \times 3K_X$ dimensional matrices H_l from (8). Let $\mathbf{h}_l$ be the $3K_Y$ dimensional vector of non-zero elements of H_l:

$$\mathbf{h}_l = (\mathbf{h}'_{m_{l1}}, \mathbf{h}'_{m_{l2}}, \mathbf{h}'_{m_{l3}})' \tag{28}$$

where $m_{li},\ i = 1, 2, 3$ indicate the three months of season l. We note that we can write $H_l \mathbf{b}_{l,t}$ as a linear function of $\mathbf{h}_l$:

$$H_l \mathbf{b}_{l,t} = \begin{pmatrix} b_{m_{l1},t,1} & & & b_{m_{l2},t,1} & & & b_{m_{l3},t,1} & & \\ & \ddots & & & \ddots & & & \ddots & \\ & & b_{m_{l1},t,K_X} & & & b_{m_{l2},t,K_X} & & & b_{m_{l1},t,K_X} \end{pmatrix} \mathbf{h}_l \equiv J_{l,t}\mathbf{h}_l \tag{29}$$

We then get that $\left[\mathbf{h}_l | \text{rest}\right]$ is $N(A^{-1}\mathbf{d}, A^{-1})$ with

$$\begin{aligned} A &= \Sigma_{l,1}^{-1} + \sum_{t=\tau_l}^{T_l+1} J'_{l,t} C_l^{-1} J_{l,t} \quad \text{and} \\ \mathbf{d} &= \sum_{t=\tau_l}^{T_l+1} J'_{l,t} C_l^{-1} \mathbf{a}_{l,t} + \Sigma_{l,1}^{-1} \boldsymbol{\mu}_{l,1} \end{aligned} \tag{30}$$

Transition matrices B_m:

We note that

$$B_m \mathbf{b}_{m-1,t} = (\mathbf{b}'_{m-1,t} \otimes I_{K_X}) vec(B_m) \equiv J_{m,t-1}\, vec(B_m) \tag{31}$$

The full conditional distribution of the K_X^2 dimensional vector $vec(B_m)$, $\left[vec(B_m)|\text{rest}\right]$, is $N(A^{-1}\mathbf{d}, A^{-1})$ with

$$A = \sum_{t=\tau_m}^{T_m+1} J'_{m,t-1} D_m^{-1} J_{m,t-1} + \Sigma_{2,m}^{-1}$$
$$\mathbf{d} = \sum_{t=\tau_m}^{T_m+1} J'_{m,t-1} D_m^{-1} \mathbf{b}_{m,t} + \Sigma_{2,m}^{-1} \boldsymbol{\mu}_{2,m} \tag{32}$$

Covariance Matrices of the data model (r_l and s_m):

Recall that $R_l^{-1} = \frac{1}{r_l} E_l$ and $S_m^{-1} = \frac{1}{s_m} E_m$ where $E_l = \left(\frac{1}{c_l} I_{N_Y} - \tilde{U}_l \Lambda_l \tilde{U}_l'\right)$ and $E_m = \left(\frac{1}{c_m} I_{N_X} - \tilde{V}_m \Lambda_m \tilde{V}_m'\right)$

$$\left[r_l | rest\right] = \mathrm{IG}\left(T_l N_Y/2 + \alpha_1,\ \beta_1 + \frac{1}{2} \sum_{t=\tau_l}^{T_l} (\mathbf{Y}_{l,t} - U_l \mathbf{a}_{l,t})' E_l (\mathbf{Y}_{l,t} - U_l \mathbf{a}_{l,t})\right) \tag{33}$$

$$\left[s_m | rest\right] = \mathrm{IG}\left(T_m N_X/2 + \alpha_2,\ \beta_2 + \frac{1}{2} \sum_{t=\tau_m}^{T_m} (\mathbf{X}_{m,t} - V_m \mathbf{b}_{m,t})' E_m (\mathbf{X}_{m,t} - V_m \mathbf{b}_{m,t})\right) . \tag{34}$$

Covariance Matrices of the process model (C_l and D_m):

The full conditional distributions of C_l and D_m are

$$\left[C_l | rest\right] = \mathrm{IW}\left(\nu_Y + T_l - \tau_l + 1, W_l + \sum_{t=\tau_l}^{T_l} (\mathbf{a}_{l,t} - H_l \mathbf{b}_{l,t})(\mathbf{a}_{l,t} - H_l \mathbf{b}_{l,t})'\right) \tag{35}$$

$$\left[D_m | rest\right] = IW\left(\nu_X + T_m - \tau_m + 1,\ W_m + \sum_{t=\tau_m}^{T_m} (\mathbf{b}_{m,t} - B_m \mathbf{b}_{m-1,t})(\mathbf{b}_{m,t} - B_m \mathbf{b}_{m-1,t})'\right) \tag{36}$$

References

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag, 2nd ed.

Berliner, L. M., Wikle, C. K., and Cressie, N. (2000), "Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling," *Journal of Climate*, 13, 3953–3968.