Dimension reduced modeling of space-time processes

Downscaling temperatures in Antarctica

Jenný Brynjarsdóttir

Case Western Reserve University

Joint work with L. Mark Berliner, The Ohio State University

MAST Colloquium
November 14, 2014

Dimension-Reduced Modeling of Spatio-Temporal Processes

Jenný BRYNJARSDÓTTIR and L. Mark BERLINER

The field of spatial and spatio-temporal statistics is increasingly faced with the challenge of very large datasets. The classical approach to spatial and spatio-temporal modeling is very computationally demanding when datasets are large, which has led to interest in methods that use dimension-reduction techniques. In this article, we focus on modeling of two spatio-temporal processes where the primary goal is to predict one process from the other and where datasets for both processes are large. We outline a general dimension-reduced Bayesian hierarchical modeling approach where spatial structures of both processes are modeled in terms of a low number of basis vectors, hence reducing the spatial dimension of the problem. Temporal evolution of the processes and their dependence is then modeled through the coefficients of the basis vectors. We present a new method of obtaining data-dependent basis vectors, which is geared toward the goal of predicting one process from the other. We apply these methods to a statistical downscaling example, where surface temperatures on a coarse grid over Antarctica are downscaled onto a finer grid. Supplementary materials for this article are available online.

KEY WORDS: Bayesian hierarchical modeling; Downscaling; Empirical orthogonal functions; Massive datasets; Maximum covariance patterns; Polar MM5.

⋮

Jenný Brynjarsdóttir is Assistant Professor in the Department of Mathematics, Applied Mathematics and Statistics at Case Western Reserve University, Cleveland, OH 44106 (E-mail: *jenny.brynjarsdottir@case.edu*). Mark Berliner is Professor and Chair of the Department of Statistics, The Ohio State University, Columbus, OH 43210 (E-mail: *mb@stat.osu.edu*). This research was supported by the National Science Foundation under grants ATM-07-24403 and DMS-10-49064. The authors thank Peter Craigmile, Noel Cressie, and Steve MacEachern for valuable input during various stages of the development of this work, and the editor, associate editor, and two anonymous reviewers for insightful and helpful comments.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

A class of dimension-reduced alternatives to the DSTM in (1) involves two stages. First, each $\mathbf{Y}_t$ is assumed to have a representation of the form

$$\mathbf{Y}_t = U_t \mathbf{a}_t + \boldsymbol{\eta}_{Y_t}, \quad t = 1, \ldots, \tau, \qquad (2)$$

Journal of the American Statistical Association
xxx 2014, Vol. 00, No. 0, Applications and Case Studies
DOI: 10.1080/01621459.2014.904232

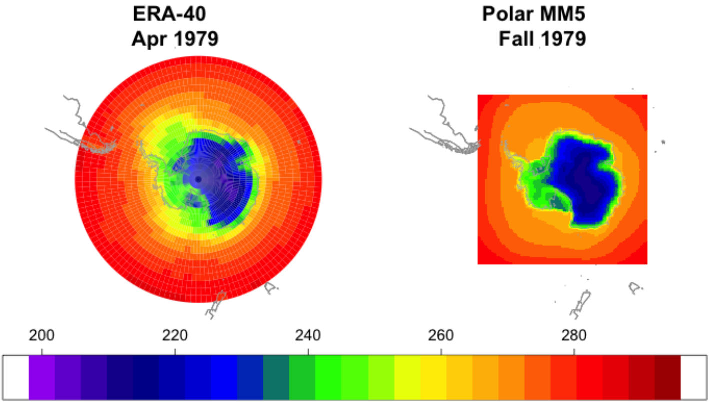
Dimension-reduced modeling – Why?

- Increasingly large datasets in spatial and spatio-temporal statistics
  - Satellite data, global network of weather stations, output from climate models, medical imagery etc.
- Traditional Spatial Statistics methods are computationally expensive for large datasets
- Dimension-reduced modeling is one way of approaching this problem

Our focus:

- Modeling of two spatio-temporal processes, where
  - Both datasets are large
  - The primary goal is to predict one process from the other (e.g. Statistical Downscaling)
- Data-Dependent basis vectors in the spirit of EOFs

Data – 2 meter surface temperatures in the Antarctic



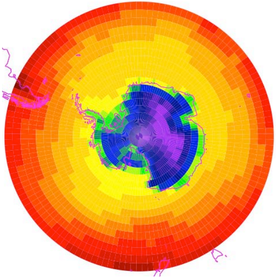
Polar MM5 model output: 14641 locations, 60km resolution
ERA-40 reanalysis data: 2736 locations south of 45°S
Total number of data points: 2,124,957

Statistical Downscaling

- **Downscaling**
  - Data on a low-resolution grid used to infer a process on a high-resolution grid
- **General circulation models (GCMs)**
  - Global model output on a low-resolution grid
- **Regional climate models (RCMs)**
  - Regional model output on a high-resolution grid
  - Deterministic downscaling, driven by GCM output
- **Statistical downscaling**
  - Statistical model high-resolution process given the low-resolution data

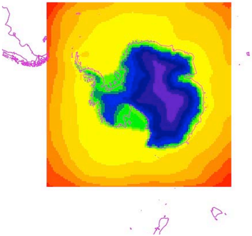
Low resolution

ERA-40 data January 1986

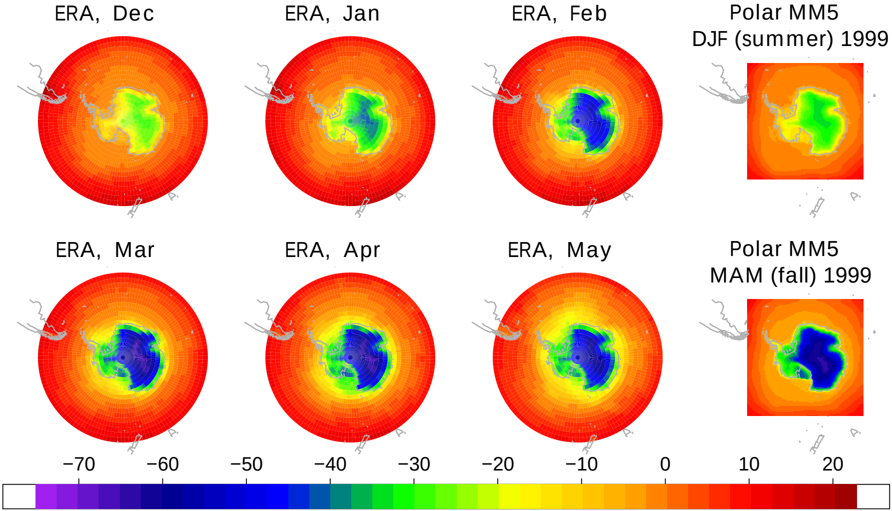


High resolution

Polar MM5 data Summer 1986



Polar MM5 and ERA-40 data [°C]



Outline

Spatial Statistics

- Observations $Y_1, \ldots, Y_N$ at spatial locations $s_1, \ldots, s_N$
- Typical statistical model:

$$\mathbf{Y} = (Y_1, \ldots, Y_N)^T \sim N(\boldsymbol{\mu}, \Sigma)$$

- Premise: Two Y's at close locations are more likely to be similar than two Y's far apart.
- Σ modelled via a **covariance function** $\mathrm{Cov}(Y_i, Y_j) = c(s_i, s_j|\theta)$
  - Usually a function only of the *distance* between s_i and s_j
  - The point: Given the locations (and θ) we can calculate Σ
- The likelihood:

$$f(\mu, \theta|\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma(\theta)|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma(\theta)^{-1}(\mathbf{y} - \boldsymbol{\mu})\right\}$$

Spatial Statistics

- Observations $Y_1, \ldots, Y_N$ at spatial locations $s_1, \ldots, s_N$
- Typical statistical model:

$$\mathbf{Y} = (Y_1, \ldots, Y_N)^T \sim N(\boldsymbol{\mu}, \Sigma)$$

- Premise: Two Y's at close locations are more likely to be similar than two Y's far apart.
- Σ modelled via a **covariance function** $\mathrm{Cov}(Y_i, Y_j) = c(s_i, s_j|\theta)$
  - Usually a function only of the *distance* between s_i and s_j
  - The point: Given the locations (and θ) we can calculate Σ
- The likelihood:

$$f(\mu, \theta|\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma(\theta)|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\Sigma(\theta)^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}$$

Dimension-reduced spatio-temporal modeling

- Add time: $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_N))'$, $t = 1, \ldots, T$
- Linear dynamical spatio-temporal model:

$$\mathbf{Y}_t = M\mathbf{Y}_{t-1} + \epsilon_{Y_t}, \quad \epsilon_{Y_t} \sim N(\mathbf{0}, \Sigma)$$

  - Problematic if N is large

Dimension-reduced spatio-temporal modeling

- Add time: $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_N))'$, $t = 1, \ldots, T$
- Linear dynamical spatio-temporal model:

$$\mathbf{Y}_t = M\mathbf{Y}_{t-1} + \epsilon_{Y_t}, \quad \epsilon_{Y_t} \sim N(\mathbf{0}, \Sigma)$$

  - Problematic if N is large

Dimension-reduced spatio-temporal modeling

- Add time: $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_N))'$, $t = 1, \ldots, T$
- Linear dynamical spatio-temporal model:

$$\mathbf{Y}_t = M\mathbf{Y}_{t-1} + \epsilon_{Y_t}, \quad \epsilon_{Y_t} \sim N(\mathbf{0}, \Sigma)$$

  - Problematic if N is large

Dimension-reduced approach:

$$\mathbf{Y}_t = U\mathbf{a}_t + \eta_t, \qquad \eta_t \sim N(\mathbf{0}, \Sigma_\eta)$$
$$\mathbf{a}_t = H\mathbf{a}_{t-1} + \xi_t, \qquad \xi_t \sim N(\mathbf{0}, \Sigma_\xi)$$

Where

- U includes K pre-specified basis vectors,
- $\mathbf{a}_t$: K_Y-dimensional **amplitude** vector, $K_Y << N_Y$.
- Spatio-Temporal Random Effects model of Cressie et al. (2010)

Dimension-reduced spatio-temporal modeling

$$\mathbf{Y}_t = U\mathbf{a}_t + \eta_t, \qquad \eta_t \sim N(\mathbf{0}, \Sigma_\eta)$$
$$\mathbf{a}_t = H\mathbf{a}_{t-1} + \xi_t, \qquad \xi_t \sim N(\mathbf{0}, \Sigma_\xi)$$

- Implies a covariance matrix for **Y**:

  $$\mathrm{Var}(\mathbf{Y}) = U\Sigma_\xi U' + \Sigma_\eta$$

  Get a rich spatial covariance structure without having to specify a covariance function
- Need to estimate H, Σ_ξ and Σ_η
- Need to specify the basis vectors, U. Examples:
  - EOFs: Wikle & Cressie (1999), Berliner et al. (2000)
  - Multi-resolution basis functions (wavelets): Wikle et al. (2001), Kang et al. (2010), Katzfuss & Cressie (2011, 2012)
  - Process convolution methods: Higdon (1998), Calder (2007)

Hierarchical modeling of two spatial processes

- $Y(s)$ and $X(c)$: spatial processes at locations s and c
- Let $\mathbf{Y} = (Y(s_1), \ldots, Y(s_{N_Y}))'$ and $\mathbf{X} = (X(c_1), \ldots, X(c_{N_X}))'$

Hierarchical Bayesian model:

$$\left[\mathbf{Y}, \mathbf{X} | \theta\right] \left[\theta\right] = \left[\mathbf{Y} | \mathbf{X}, \theta\right] \left[\mathbf{X} | \theta\right] \; \left[\theta\right]$$

- Natural when we want to predict **Y** from **X**
- Allows incorporation of knowledge of physical dependence
- Avoids specification of the joint covariance model for **Y** and **X**

Dimension-reduced modeling of two spatial processes

- A linear model for $[\mathbf{Y}|\mathbf{X},\theta]$:

$$\mathbf{Y} = F\mathbf{X} + \varepsilon, \quad E(\varepsilon) = \mathbf{0}, \; Cov(\varepsilon) = \Sigma_\varepsilon$$

  - Problematic if N_Y and N_X are big

Dimension-reduced modeling of two spatial processes

- A linear model for $[\mathbf{Y}|\mathbf{X},\theta]$:

$$\mathbf{Y} = F\mathbf{X} + \varepsilon, \quad E(\varepsilon) = \mathbf{0}, \; Cov(\varepsilon) = \Sigma_{\varepsilon}$$

  - Problematic if N_Y and N_X are big

Dimension-reduced modeling of two spatial processes

- A linear model for $[\mathbf{Y}|\mathbf{X}, \theta]$:

$$\mathbf{Y} = F\mathbf{X} + \varepsilon, \quad E(\varepsilon) = \mathbf{0}, \; Cov(\varepsilon) = \Sigma_\varepsilon$$

  - Problematic if N_Y and N_X are big
- We consider using Dimension reduction models for both **Y** and **X**:

$$\mathbf{Y} = U\mathbf{a} + \eta_Y \quad \text{and} \quad \mathbf{X} = V\mathbf{b} + \eta_X$$

  - **a** and **b**: K_Y and K_X dimensional unknown vectors of amplitudes
  - Need to specify U and V

Dimension-reduced modeling of two spatial processes

- A linear model for $[\mathbf{Y}|\mathbf{X},\theta]$:

$$\mathbf{Y} = F\mathbf{X} + \varepsilon, \quad E(\varepsilon) = \mathbf{0},\ Cov(\varepsilon) = \Sigma_\varepsilon$$

  - Problematic if N_Y and N_X are big
- We consider using Dimension reduction models for both **Y** and **X**:

$$\mathbf{Y} = U\mathbf{a} + \eta_Y \quad \text{and} \quad \mathbf{X} = V\mathbf{b} + \eta_X$$

  - **a** and **b**: K_Y and K_X dimensional unknown vectors of amplitudes
  - Need to specify U and V
- This implies a linear model for the amplitude vectors:

$$\mathbf{Y} = F\mathbf{X} + \varepsilon \ \Rightarrow \ \mathbf{a} = U'FV\mathbf{b} + U'F\eta_X + U'\varepsilon - U'\eta_Y$$

Or: $\mathbf{a} = H\mathbf{b} + \mathbf{e}$

  - $H = U'FV$ is of dimension $K_Y \times K_X$, much lower dimension than F

Dimension-reduced model of two space-time processes

- Let $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_{N_Y}))'$ and $\mathbf{X}_t = (X_t(c_1), \ldots, X_t(c_{N_X}))'$

Data Model

$$\mathbf{Y}_t = U\mathbf{a}_t + \eta_{tY} \quad \text{and}$$
$$\mathbf{X}_t = V\mathbf{b}_t + \eta_{tX} \quad \text{for } t = 1, \ldots, T$$

Process Model

$$\left[\begin{array}{c|c} \mathbf{a}_t & \mathbf{a}_{t-1} \\ \mathbf{b}_t & \mathbf{b}_{t-1} \end{array} \right] \quad \text{for } t = 1, \ldots, T$$

Dimension-reduced model of two space-time processes

- Let $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_{N_Y}))'$ and $\mathbf{X}_t = (X_t(c_1), \ldots, X_t(c_{N_X}))'$

Data Model

$$\mathbf{Y}_t = U\mathbf{a}_t + \eta_{tY} \quad \text{and}$$
$$\mathbf{X}_t = V\mathbf{b}_t + \eta_{tX} \quad \text{for } t = 1, \ldots, T$$

Process Model

$$\mathbf{a}_t = H\mathbf{b}_t + \mathbf{e}_{1t} \quad \text{and}$$
$$\mathbf{b}_t = B\mathbf{b}_{t-1} + \mathbf{e}_{2t} \quad \text{for } t = 1, \ldots, T$$

MSE of dimension reduced predictors

- Dimension-reduced predictors may be sub-optimal
- The optimal predictor in terms of MSE is $E(\mathbf{Y}|\mathbf{X})$:

$$E\left(||\mathbf{Y} - h(\mathbf{X})||^2\right) \geq E\left(||\mathbf{Y} - E(\mathbf{Y}|\mathbf{X})||^2\right)$$

- A predictor under the dimension-reduced hierarchical model:

$$E(\mathbf{Y}|\mathbf{b}) = E\big(E(\mathbf{Y}|\mathbf{a}, \mathbf{b})|\mathbf{b}\big) = E\big(U\mathbf{a}|\mathbf{b}\big)$$

MSE of dimension reduced predictors

- Dimension-reduced predictors may be sub-optimal
- The optimal predictor in terms of MSE is $E(\mathbf{Y}|\mathbf{X})$:

$$E\left(||\mathbf{Y} - h(\mathbf{X})||^2\right) \geq E\left(||\mathbf{Y} - E(\mathbf{Y}|\mathbf{X})||^2\right)$$

- A predictor under the dimension-reduced hierarchical model:

$$E(\mathbf{Y}|\mathbf{b}) = E\big(E(\mathbf{Y}|\mathbf{a},\mathbf{b})|\mathbf{b}\big) = E\big(U\mathbf{a}|\mathbf{b}\big)$$

- We can show:

$$E\left(||\mathbf{Y} - E(U\mathbf{a}|\mathbf{b})||^2\right) = E_{\mathbf{a},\eta,\mathbf{b}}\left(||U\mathbf{a} + \boldsymbol{\eta}_Y - E(U\mathbf{a}|\mathbf{b})||^2\right)$$
$$= E\left(||\mathbf{a} - E(\mathbf{a}|\mathbf{b})||^2\right) + \sum_{i=1}^{N_Y} Var(\eta_{Yi})$$

  - ⇒ **Need good prediction of the a amplitudes**
  - ⇒ **Need good dimension-reduced representation of Y (U)**

Outline

EOFs, MCA, CCA – Popular methods in climate science

One climate field:

- Empirical Orthogonal Functions (EOFs) = Principal components
- Used to study “modes of variation”

Two climate fields:

- Canonical Correlation Analysis (CCA)
- Maximum Covariance Analysis (MCA), also known as “SVD-Analysis”
- Have also been used for downscaling, e.g. van Storch et al. (1993) and Widman et al. (2003).

Empirical Orthogonal Functions (EOFs)

- $\mathbf{Y}$: N_Y-dimensional random vector, mean $\mathbf{0}$, covariance Σ_Y

EOFs

Linear combinations (principal components)

$$A_1 = \mathbf{u}_1'\mathbf{Y}, \quad A_2 = \mathbf{u}_2'\mathbf{Y}, \quad \ldots, \quad A_{N_Y} = \mathbf{u}_{N_Y}'\mathbf{Y}$$

- where $Var(A_i)$ is maximized subject to being uncorrelated with the first $i-1$ principal components
- and EOF patterns have unit length, $\mathbf{u}_i'\mathbf{u}_i = 1$ for all i

- Solution: Spectral decomposition $\Sigma_Y = \mathcal{U}\Lambda\mathcal{U}'$
- Dimension reduction model: $\mathbf{Y} = U\mathbf{a} + \eta_Y$

Maximum covariance analysis (MCA) = "SVD analysis"

- $\mathbf{Y}$ and $\mathbf{X}$: N_Y and N_X dimensional random vectors

MCA

$$\begin{aligned} A_1 &= \tilde{\mathbf{u}}_1'\mathbf{Y} \quad \text{and} \quad B_1 = \tilde{\mathbf{v}}_1'\mathbf{X} \\ &\vdots \qquad\qquad\qquad\quad \vdots \\ A_d &= \tilde{\mathbf{u}}_d'\mathbf{Y} \quad \text{and} \quad B_d = \tilde{\mathbf{v}}_d'\mathbf{X} \qquad (d = \min\{N_Y, N_X\}) \end{aligned}$$

- where $Cov(A_i, B_i)$ are maximized and
- the *MC patterns* $\tilde{\mathbf{u}}_i$ and $\tilde{\mathbf{v}}_i$ are orthonormal

- Solution: Singular decomposition of the cross-covariance matrix, $\Sigma_{YX} = \tilde{\mathcal{U}}\tilde{D}\tilde{\mathcal{V}}'$
- Dimension reduction models: $\mathbf{Y} = U\mathbf{a} + \eta_Y$ and $\mathbf{X} = V\mathbf{b} + \eta_X$

Predicting one process (**Y**) from another (**X**)

$$\mathbf{Y} = U\mathbf{a} + \eta_Y \quad \text{and} \quad \mathbf{X} = V\mathbf{b} + \eta_X$$

Could use EOFs for each process

- Appeal: Good representation of each individual field
- Concern: May not have a strong relationship between amplitudes **a** and **b**

Could use MC patterns (or CCA)

- Appeal: First few pairs of patterns capture most of the cross-covariance structure (amplitudes have high covariance)
- Concern: No guaranty that the patterns give a good representation of each individual field

Predicting one process (**Y**) from another (**X**)

Dimension reduction:

$$\mathbf{Y} = U\mathbf{a} + \eta_Y \quad \text{and} \quad \mathbf{X} = V\mathbf{b} + \eta_X$$

Model for amplitudes:

$$\mathbf{a} = H\mathbf{b} + \mathbf{e}$$

We want U to give a good representation of **Y**.

- E.g: Use EOFs that capture a high proportion of the total variance of **Y**.

We want **b** to be a good predictor for **a**

- Given basis vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{K_Y}$ (e.g. EOFs),
  we want to find N_X-dimensional vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_{K_Y}$, so that

$$Cov(A_k, B_k) = Cov(\mathbf{u}_k'\mathbf{Y}, \mathbf{v}_k'\mathbf{X})$$

  is maximized for each $k = 1, \ldots, K_Y$

Maximum Covariance Patterns

Maximum Covariance Patterns (MCPs)

- The vectors

$$\mathbf{v}_k = \frac{\Sigma_{XY}\mathbf{u}_k}{||\Sigma_{XY}\mathbf{u}_k||} \quad k = 1, \ldots, K_Y$$

  maximize $Cov(\mathbf{u}_k'\mathbf{Y}, \mathbf{v}_k'\mathbf{X})$ for each k
- We call the $\mathbf{v}_k$ vectors Maximum Covariance Patterns (MCPs)

- Sample MCPs: Replace Σ_{XY} with S_{XY}

Proposed basis vectors:

- Use the K first EOFs, $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K$, of the $\mathbf{Y}$ field as basis vectors for U
- Use the K MCPs, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K$, as basis vectors for V

Proportion of total variance explained

- **Y** and **X**: N_Y and N_X dimensional random vectors
- **u** and **v**: N_Y and N_X dimensional known vectors
- Let $A = \mathbf{u}'\mathbf{Y}$ and $B = \mathbf{v}'\mathbf{X}$

Proportion of total variance explained by a basis vector **u**:

$$\frac{Var(A)}{\mathrm{tr}(\Sigma_{YY})}$$

We define the proportion of the total variance in **Y** that is explained by **v** through **u** as

$$p_{\mathbf{u}}(\mathbf{v}) = \frac{Var(A)}{\mathrm{tr}(\Sigma_{YY})} Cor(A,B)^2 = \frac{Cov(A,B)^2 / Var(B)}{\mathrm{tr}(\Sigma_{YY})}$$

Proportion of total variance explained

Theorem

Y and **X**: N_Y and N_X dimensional random vectors
u: N_Y dimensional known vector
v: MCP of **X** with respect to **u**
e: The first EOF of **X**. Then

$$p_{\mathbf{u}}(\mathbf{v}) \geq p_{\mathbf{u}}(\mathbf{e})$$

- I.e. When explaining the variability in **Y** it is better to use MCPs as basis vectors for **X** rather than the EOFs.

Orthogonal Maximum Covariance Patterns (OMCPs)

- Potential Problem: MCPs are not necessarily linearly independent

$$\mathbf{v}_k = \frac{\Sigma_{XY}\mathbf{u}_k}{||\Sigma_{XY}\mathbf{u}_k||} \quad k = 1, \ldots, K_Y$$

Orthogonal Maximum Covariance Patterns (OMCPs)

OMCPs are vectors $\mathbf{v}_k$ that maximize $Cov(\mathbf{u}_k'\mathbf{Y}, \mathbf{v}_k'\mathbf{X})$ with the constraint that $\mathbf{v}_1, \ldots, \mathbf{v}_K$ are orthogonal

- Same as Gram-Schmidt orthogonalization of the MCPs

- OMCPs eliminate concerns about MCPs possibly being linearly dependent
- Orthonormal basis vectors are computationally convenient for dimension-reduced modeling

Outline

Polar MM5 – Antarctic hindcast project

Polar MM5 model output

- PSU/NCAR Fifth-generation Mesoscale Model (MM5)
- Modified for polar regions by the Byrd Polar Research Center, OSU (Monaghan et al. 2006)
- Antarctic hindcast project:
  - Polar MM5 used to model climate over Antarctica 1979-2001
  - Data used here: Seasonal mean 2-meter temperature fields

ERA-40 reanalysis data

- ECMWF re-analysis project (http://data-portal.ecmwf.int/)
- Used for initial and boundary conditions for the Polar MM5 simulations.
- Cover the period from mid-1957 to mid-2002
  - Data used here: Monthly averages of daily means of 2-meter temperatures

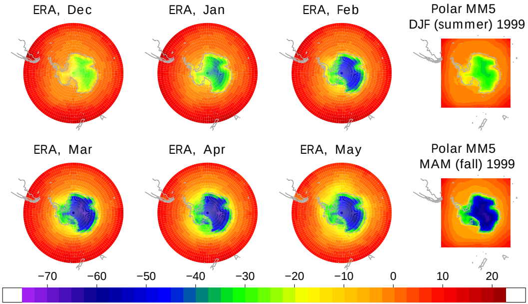
Polar MM5 and ERA-40 surface temperature data

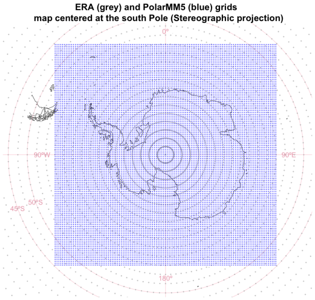
Polar MM5 model output

- Regional climate model, seasonal 2-meter temp.
- 14641 spatial locations (blue grid)

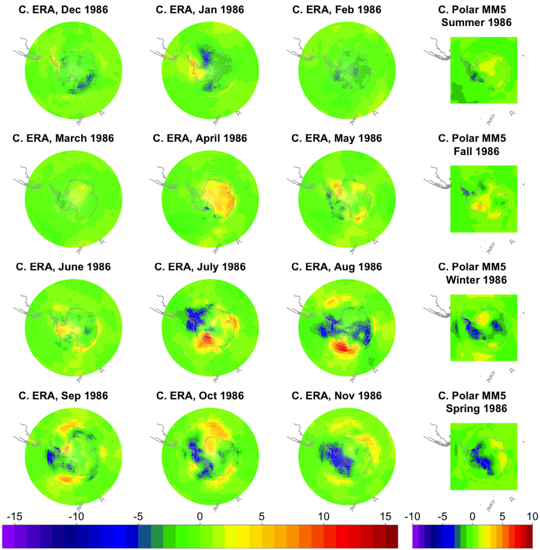
ERA-40 data from ECMWF

- Data product, monthly 2-meter temperatures
- 2736 spatial locations south of 45°S (grey grid)





Centered Polar MM5 and ERA-40 data



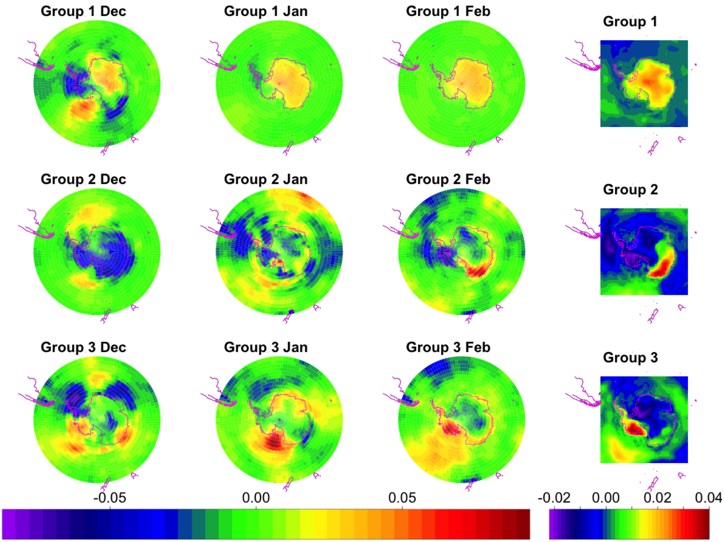
Construct EOFs of centered Polar MM5 data

- Separately for each season

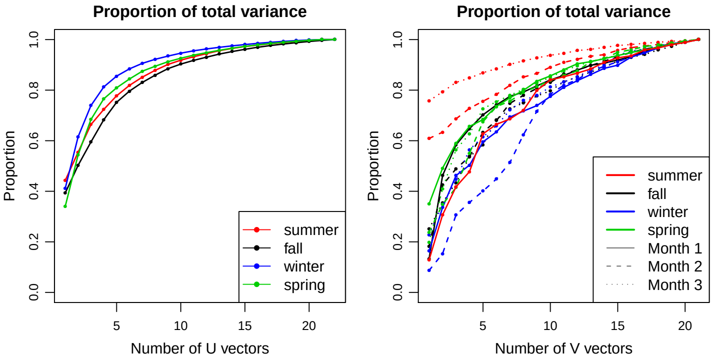
Then construct OMCPs of centered ERA-40 data

- For every EOF we obtain three OMCPs, $\rightarrow$ one for each month of ERA-40 data

Basis Vectors (summer)



Basis Vectors



- Chose $K_Y = 4$ and $K_X = 4$
- First 4 EOFs explain 68% - 82% of total sample variance of Polar MM5 data
- First 4 OMCPs explain 36% - 85% of total sample variance of ERA-40 data

Bayesian Hierarchical Model

$\mathbf{Y}_{l,t}$: N_Y-dim. vector of centered Polar MM5 data, season l, year t
$\mathbf{X}_{m,t}$: N_X-dim. vector of centered ERA-40 data, month m, year t

Data Model

For every t, m and l:

$$\left[\mathbf{Y}_{l,t}|\mathbf{a}_{l,t}, R_l\right] = N\left(U_l\mathbf{a}_{l,t}, R_l\right) \text{ and } \left[\mathbf{X}_{m,t}|\mathbf{b}_{m,t}, S_m\right] = N\left(V_m\mathbf{b}_{m,t}, S_m\right)$$

U_l and V_m: first four EOFs and OMCPs for season l and month m

Process Model

$$\left[\mathbf{a}_{l,t}|\mathbf{b}_{m_{l1},t}, \mathbf{b}_{m_{l2},t}, \mathbf{b}_{m_{l3},t}H_l, C_l\right] = N\left(H_l\begin{bmatrix}\mathbf{b}_{m_{l1},t}\\ \mathbf{b}_{m_{l2},t}\\ \mathbf{b}_{m_{l3},t}\end{bmatrix},\ C_l\right)$$

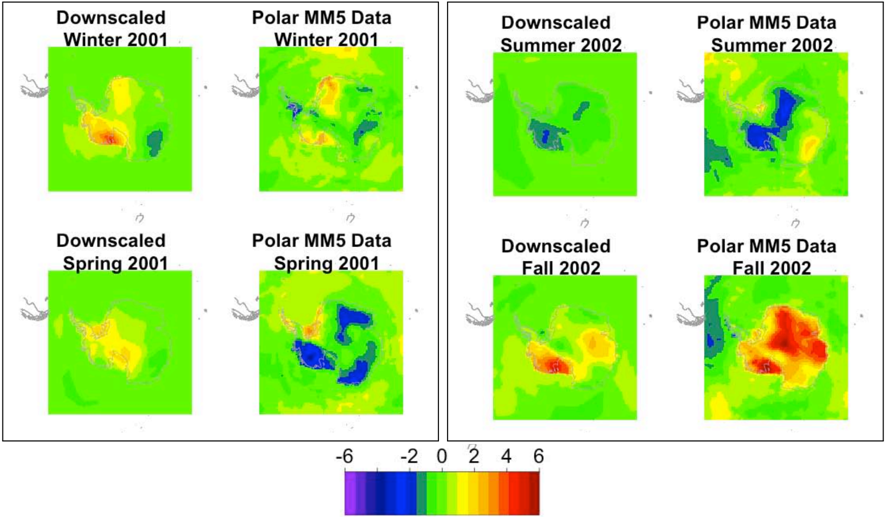
$$H_l = \begin{pmatrix}H_{m_{l1}} & H_{m_{l2}} & H_{m_{l3}}\end{pmatrix}, \quad H_m = diag(\mathbf{h}_m)$$

$$\left[\mathbf{b}_{m,t}|\mathbf{b}_{m-1,t}, B_m, D_m\right] = N\left(B_m\mathbf{b}_{m-1,t},\ D_m\right) \quad \text{and } \left[\mathbf{b}_{2,1}\right] = N\left(\mu_b, \Sigma_b\right)$$

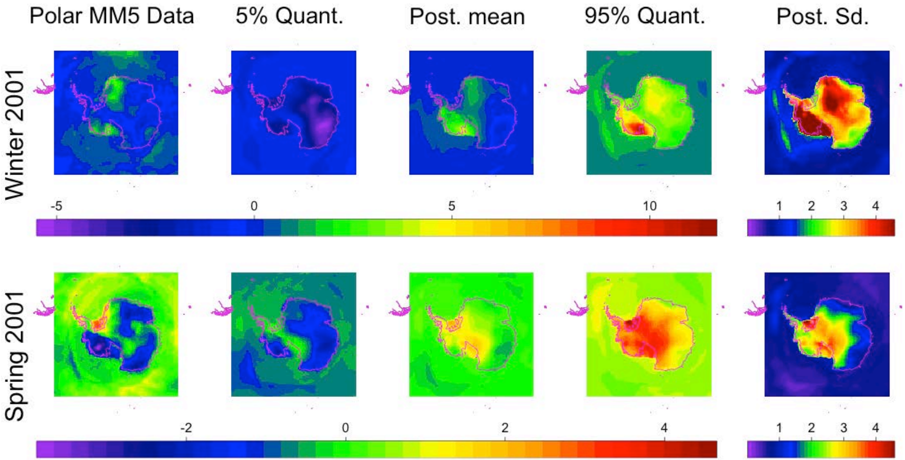
Bayesian Hierarchical Model – Parameter Model

- Normal priors for elements of transition matrices B_m and H_m
- Inverse-Wishart priors for process-model covariance matrices, C_l and D_m
- Data-model covariances matrices R_l and S_m (Berliner et al. 2000):
  - Modeled using the next six basis vectors to account for some of the leftover structure
  - Each covariance matrix has one unknown scalar with an Inverse-Gamma prior
  - Computationally very effective
- Obtained samples from the posterior distribution via a Gibbs Sampler

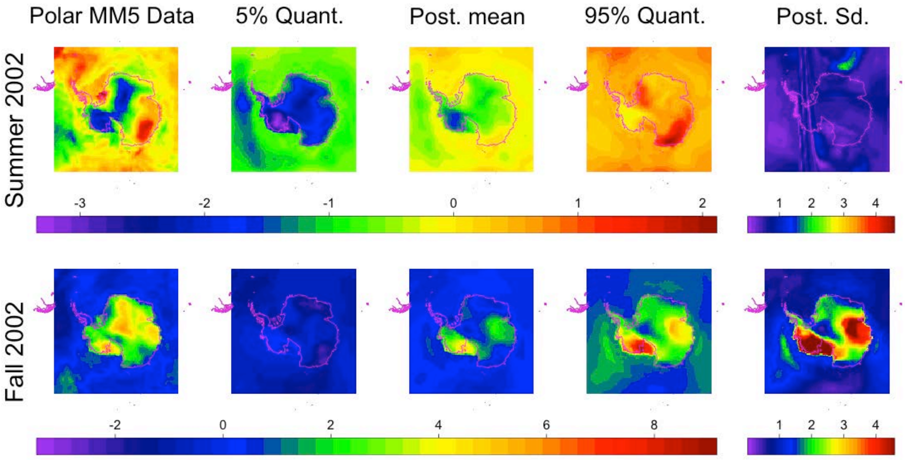
Downscaling winter 2001 - fall 2002



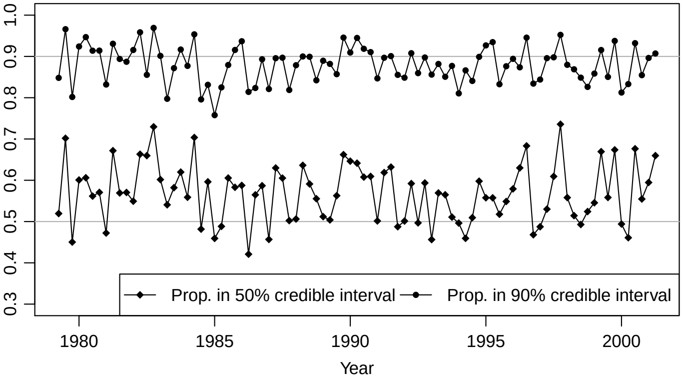
Downscaling winter 2001 - fall 2002



Downscaling winter 2001 - fall 2002



Coverage of credible intervals



- Proportion of datapoints y_{lti} that fall within the 90% (and 50%) posterior credible interval of Y_{lti}

Downscaling winter 2001 - fall 2002

Point wise prediction intervals

- Obtained samples from the marginal posterior predictive distribution of $Y_{l,t,j}$, for each location j
- Percentage of locations where the actual Polar MM5 temperature data fall within the 50%, 90% and 95% prediction intervals:

	Winter 2001	Spring 2001	Summer 2002	Fall 2002
50% Pr. Int.	70.72%	38.19%	52.52%	50.67%
90% Pr. Int.	97.40%	77.58%	89.39%	80.60%
95% Pr. Int.	98.60%	86.45%	94.58%	85.31%

Conclusions

- A dimension reduced modeling approach can be very useful when faced with large spatial or spatio-temporal datasets
- The Orthogonal Maximum Covariance Patters are a promising choice when the goal is to predict one space-time process from another

Brynjarsdóttir, J. & Berliner, L.M. (2014).
Dimension-reduced modeling of spatio-temporal processes.
Journal of the American Statistical Association, Accepted.

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

Calder, C. A. (2007).
Dynamic factor process convolution models for multivariate space- time data with application to air quality assessment.
Environmental and Ecological Statistics , 14, 229–247.

Higdon, D. (1998).
A process-convolution approach to modelling temperatures in the North Atlantic Ocean.
Environmental and Ecological Statistics , 5, 173–190.

Kang, E. L., Cressie, N., and Shi, T. (2010).
Using temporal variability to improve spatial mapping with application to satellite data.
The Canadian Journal of Statistics , 38(4), 271–289.

Katzfuss, M. and Cressie, N. (2011).
Spatio-temporal smoothing and EM estimation for massive remote-sensing data set.
Journal of Time Series Analysis , 32, 430–446.

Katzfuss, M. and Cressie, N. (2012).
Bayesian hierarchical spatio-temporal smoothing for very large datasets.
Environmetrics , 23, 94–107.

Monaghan, A.J., Bromwich D.H., & Wang, S.H. (2006).
Recent trends in Antarctic snow accumulation from Polar MM5 simulations.
Phil. Trans. R. Soc. A, 364, 1683–1708.

http://polarmet.osu.edu/PolarMet/pmm5.html

von Storch, H., Zorita, E., and Cubasch, U. (1993).
Downscaling of global climate change estimates to regional scales: An application to Iberian Rainfall in wintertime
Journal of Climate, 6, 1161–1171.

Wikle, C.K., & Cressie, N. (1999).
A dimension-reduced approach to space-time Kalman filtering.
Biometrica, 86, 815–829.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001).
Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds.
Journal of the American Statistical Association, 96, 382–397.

Extra slides

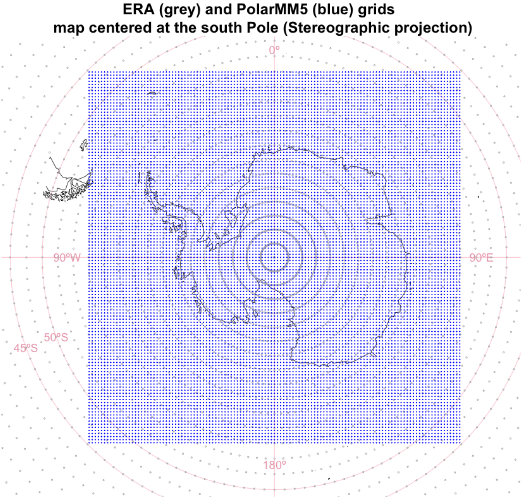
Polar MM5 and ERA-40 grids

Polar MM5 model output

- 121 × 121 polar stereographic grid
- Model resolution: 60km in each horizontal direction
- 14641 spatial locations

ERA-40 reanalysis data

- 2.5° × 2.5° latitude-longitude grid
- 2736 spatial locations south of 45°S



Data model covariance matrices R_l and S_m

- Balance modeling the structure of the data with computational feasibility
- Idea from Berliner et al. 2000:

$$R_l = r_l \left(c_l I_{N_Y} + \tilde{U}_l \tilde{D}_l \tilde{U}_l' \right) \equiv r_l \tilde{R}_w \ . \qquad (1)$$

  where
  - $\tilde{U}_l$: next few eigen vectors from $\mathcal{U}_l$
  - $\tilde{D}_l$: the corresponding eigenvalues (diagonal
  - and

$$c_l = \sum_{k=L_Y+1}^{N_Y} d_{l,k} \quad \text{for } l \in \{1, 2, 3, 4\} \ . \qquad (2)$$

- $\Rightarrow$ data model incorporates additional spatial structure beyond that represented in the leading basis vectors used to specify the means.
- Same approach was used to treat S_m.

Assessing Conditional Independence Assumptions

We assume that *given* the EOF amplitudes the Polar MM5 temperatures are independent of the ERA-40 temperatures, i.e.

$$\left[\mathbf{Y}|\mathbf{a}, \mathbf{b}, \theta, \mathbf{X}\right] = \left[\mathbf{Y}|\mathbf{a}, \theta\right], \tag{3}$$

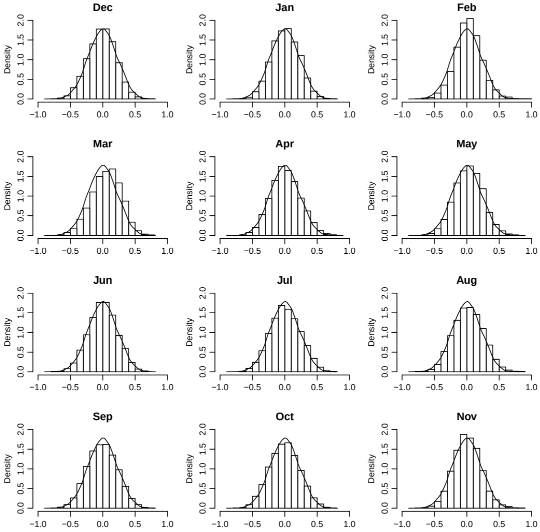
If we believe that the EOFs capture most of the structure in the Polar MM5 temperatures this assumption makes intuitive sense.

Heuristic assessment

- Sample correlations between the residuals $\mathbf{R} = \mathbf{Y} - U\mathbf{a}$ and $\mathbf{X}$ (or between $\mathbf{R}$ and $\mathbf{X} - V\mathbf{b}$)
- For each season-month combination we have > 150 million such pairs
  - Randomly sampled 10,000 pairs of R_i and X_j
- "null-density": sampled 10,000 pairs of 22-dimensional uncorrelated normal random variables

The same analysis for $\mathbf{R}$ and $\mathbf{X} - V\mathbf{b}$ gave similar results.

Assessing Conditional Independence Assumptions



Comparisons with EOFs only

- We claimed that using OMCPs as basis vectors for the **X** is better than using their EOFs
- Fitted the same model with EOFs for as *V* instead of OMCPs

Mean square prediction error of the Polar MME temperatures using OMCP vectors vs. EOF vectors for the ERA-40 temperatures:

	Winter 2001	Spring 2001	Summer 2002	Fall 2002
OMCP	**0.420**	0.947	**0.266**	1.533
EOF	0.826	**0.912**	0.275	**1.331**

Yearly variations