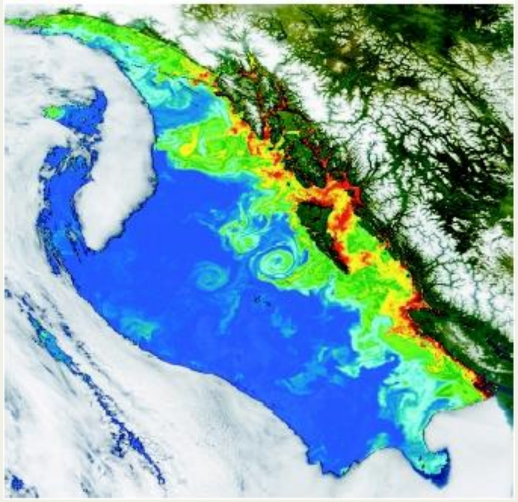
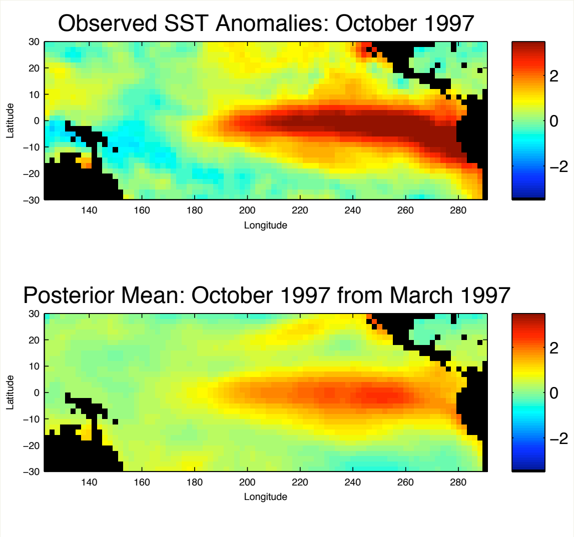
A Hierarchical Approach to Motivate Spatio-Temporal Statistical Models

Christopher K. Wikle

Department of Statistics

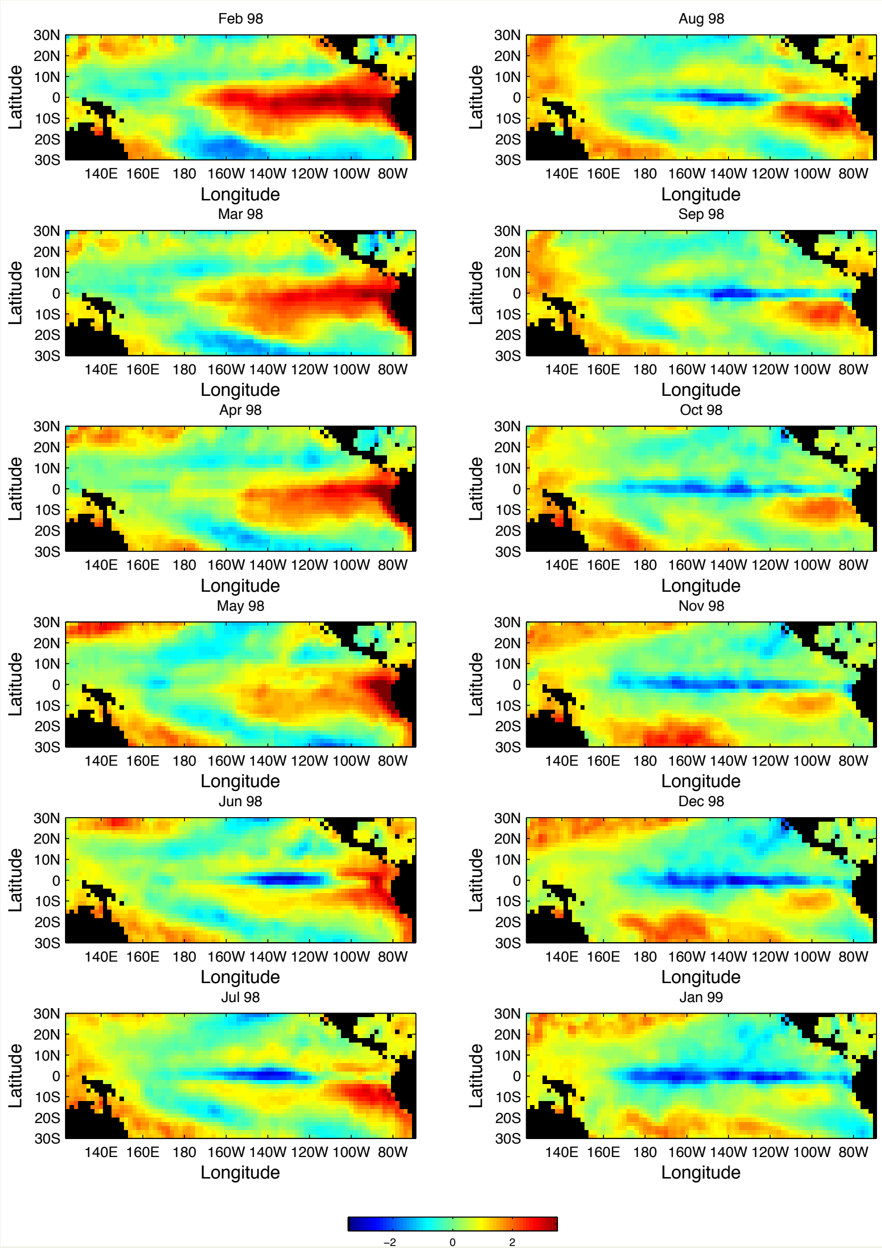
University of Missouri





IPAM, 5/25/10

Spatio-Temporal Data/Processes



SST Anomalies

Complicated interaction across many components and spatio-temporal scales.

Uncertainty:

- Data
- Process
- Parameters

Spatio-Temporal Statistical Modeling

- **Purpose:** Characterize the process in the presence of uncertain and (often) incomplete observations and system knowledge
  - Prediction in space (interpolation)
  - Prediction in time (forecasting)
  - Assimilation of observations and deterministic models
  - Inference on controlling process parameters

Spatio-Temporal Statistical Models

Two Primary Approaches

- Descriptive (joint): Characterize the first and second moment behavior of the process
  - Several different physical processes could imply the same marginal structure; problematic if non-Gaussian
  - Most useful when knowledge of process is limited
- Dynamic (conditional): Current values of the process at a location evolve from past values of the process at various locations
  - Conditional models closer to the etiology of the phenomenon under study
  - Most useful if there is some *a priori* knowledge available concerning the process behavior

(As a group, statisticians don't have a very good understanding of dynamics!)

This Talk

What can statisticians bring to the table?

- Specifically: *using process knowledge as motivation for parameterization and structure of spatio-temporal statistical models*
- Hierarchical Structure:
  - Can facilitate incorporation of scientific information into the statistical model
  - Can facilitate "estimation" of parameters
    - Dependence in parameters
    - Stochastic variable selection

This Talk (cont)

- Focus on conditional/dynamic specifications
- Focus on statistical models that are discrete in time and space, but associated with processes that are continuous in both

Outline:

- Brief comparison between joint/conditional view as a motivation for hierarchical parameterization
- Linear/nonlinear model examples
  - Dimension reduction parameterization
    - Dealing with parameter curse of dimensionality
    - SST long-lead forecasting example

Process Knowledge as Statistical Model Motivation: Temporal

(Not a new idea!)

- Yule (1927) used the differential equation governing pendulum motion as motivation for an AR time series model for the Wolfer sunspot data
- Hotelling (1927) used *approximations* of differential equations to model U.S. population growth

> "Indeed the use of differential equations supplies the statistician with a powerful tool, replacing the purely empirical fitting of arbitrary curves by a reasonable resultant of general considerations with particular data. But this growing use of differential equations must inevitably face the fact that our *a priori* knowledge can never supply us with a definite relation between a variable and its rate of change, but only with a correlation." (Hotelling, 1927, p. 283)

Covariance Model Motivation: Spatio-Temporal

- Heine (1955), Whittle (1986), Jones and Zhang (1997):

$$Y(s;t) : s \in D_s \subset \mathcal{R}, t \in D_t \subset \mathcal{R}$$

Stochastic Injection–Diffusion PDE

$$\frac{\partial Y(s;t)}{\partial t} - \beta \frac{\partial^2 Y(s;t)}{\partial s^2} + \alpha Y(s;t) = \delta(s;t),$$

$\alpha > 0, \;\; \beta > 0$ and δ a random, zero mean error process.

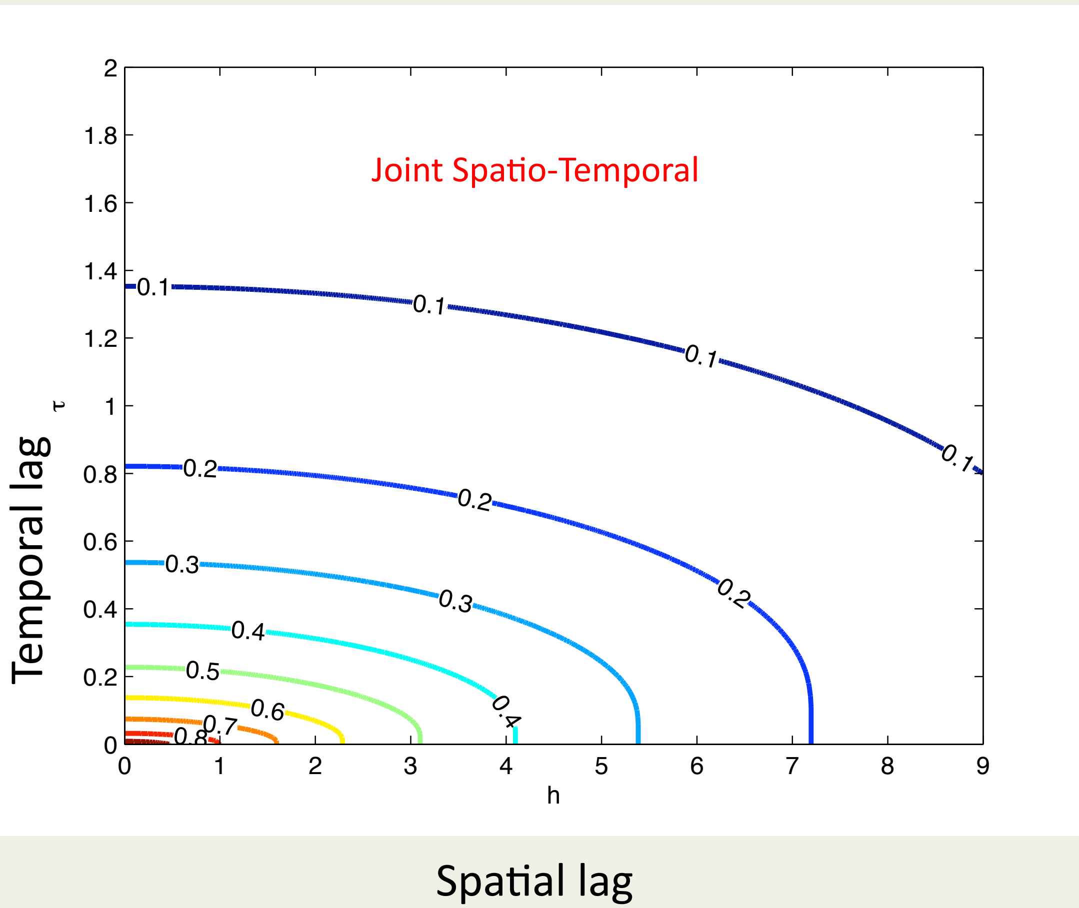
Implied spatio-temporal correlation function

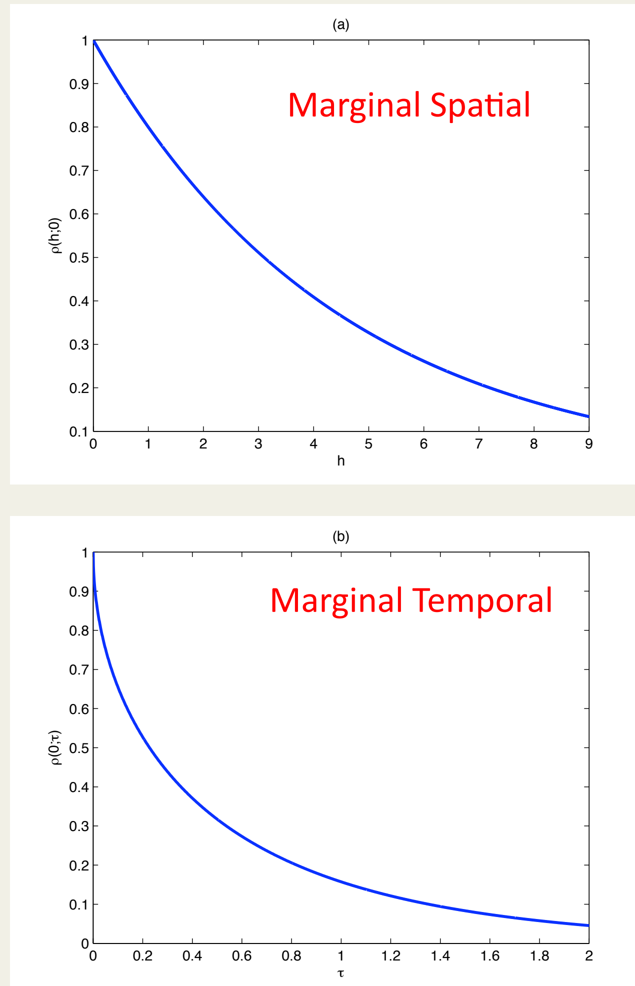
$$\begin{aligned} C(h;\tau)/C(0,0) &\equiv \rho(h;\tau) \\ &= (1/2)\{e^{-h(\alpha/\beta)^{1/2}} Erfc\left(\frac{2\tau(\alpha/\beta)^{1/2} - h/\beta}{2(\tau/\beta)^{1/2}}\right) \\ &\quad + e^{h(\alpha/\beta)^{1/2}} Erfc\left(\frac{2\tau(\alpha/\beta)^{1/2} + h/\beta}{2(\tau/\beta)^{1/2}}\right)\}, \end{aligned}$$

for $h \in \mathcal{R}, \tau \in \mathcal{R}$; $Erfc$ is the "complimentary error function"

Spatio-Temporal Dependence

Correlation function example α=1, β=20





Conditional Perspective

- We have seen that general process knowledge (e.g., injection-diffusion) can be used to develop classes of joint spatio-temporal correlation models
- Typically, such derivations have only been possible for relatively simple processes (although this is an active area of research in Statistics and Applied Math)
- In some cases, since conditional models are closer to the process etiology, it is easier to incorporate process knowledge in that context (e.g., dynamic models)

Statistician's Perspective: Conditional

For a linear process, we might consider a first-order vector autoregressive process:

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t,$$

where $\mathbf{Y}_t \equiv (Y(s_1;t), \ldots, Y(s_n;t))'$

Noise process with some (unknown) covariance, **Q**

When n is large and t=1,…,T with T relatively small, estimation is a problem!

Simple (naïve) parameterizations

- Multivariate random walk
- Common univariate AR models with spatially correlated noise
- Spatially-varying univariate AR models

Issues: easy to implement; often unrealistic (no interaction!); particularly useful when modeling parameter dynamics (rather than process dynamics)

(we'll talk about projections on lower-dim manifolds later)

Common Ground

- Obviously, the SPDE is related to the first-order discrete Markov (AR) model.
- As a toy example, consider a simple finite-difference discretization.

PDE-Motivated Parameterization: Ex

$$\frac{\partial Y(s;t)}{\partial t} - \beta \frac{\partial^2 Y(s;t)}{\partial s^2} + \alpha Y(s;t) = 0,$$

(Deterministic version of the injection-diffusion PDE example from before)

Replacing the first-order derivative with a forward difference and the second-order derivative with a centered difference gives:

$$Y(s;t+\Delta_t) = \theta_1 Y(s;t) + \theta_2 Y(s+\Delta_s;t) + \theta_3 Y(s-\Delta_s;t)$$

where

$$\theta_1 \equiv (1 - \alpha\Delta_t - 2\beta\Delta_t/\Delta_s^2), \;\; \theta_2 = \theta_3 \equiv \beta\Delta_t/\Delta_s^2$$

PDE Ex: cont.

Let $D_s \equiv \{s_0, \ldots, s_{n+1}\}$, where $s_j = s_0 + j\Delta_s$; $j = 0, 1, \ldots, n+1$

(equally spaced grid, with s_0, s_{n+1} boundary locations)

Define $\mathbf{Y}_t \equiv (Y(s_1;t), \ldots, Y(s_n;t))'$ and $\mathbf{Y}_t^B \equiv (Y(s_0;t), Y(s_{n+1};t))'$

Then, we can write the (non-stochastic) vector difference equation:

$$\mathbf{Y}_{t+1} = \mathbf{M}\mathbf{Y}_t + \mathbf{M}_B \mathbf{Y}_t^B,$$

Given initial and boundary conditions, we can use this to get a numerical solution to the PDE.

with

$$\mathbf{M} \equiv \begin{bmatrix} \theta_1 & \theta_2 & 0 & \cdots & \cdots & 0 \\ \theta_3 & \theta_1 & \theta_2 & \cdots & \cdots & 0 \\ 0 & \theta_3 & \theta_1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \theta_1 & \theta_2 \\ 0 & 0 & 0 & \cdots & \theta_3 & \theta_1 \end{bmatrix}, \qquad \mathbf{M}_B \equiv \begin{bmatrix} \theta_3 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & \theta_2 \end{bmatrix},$$

(Henceforth let: $D_t = \{0, 1, 2, \ldots\}$, in units of Δ_t)

PDE Ex: cont.

Stochastic Spatio-Temporal Difference Equation:

$$\mathbf{Y}_{t+1} = \mathbf{M}\mathbf{Y}_t + \mathbf{M}_B \mathbf{Y}_t^B + \boldsymbol{\delta}_{t+1},$$

$$\mathrm{E}(\boldsymbol{\delta}_t) = \mathbf{0}; \quad \mathrm{var}(\boldsymbol{\delta}_t) = \boldsymbol{\Sigma}_\delta$$

Noise term might represent errors due to truncation, model representation, and/or forcing.

Note: this is just a first-order vector AR process with a highly structured transition matrix. For ease of presentation, assume the boundary process is known. Under temporal stationarity, the lag-m (in time) spatial covariance matrices are:

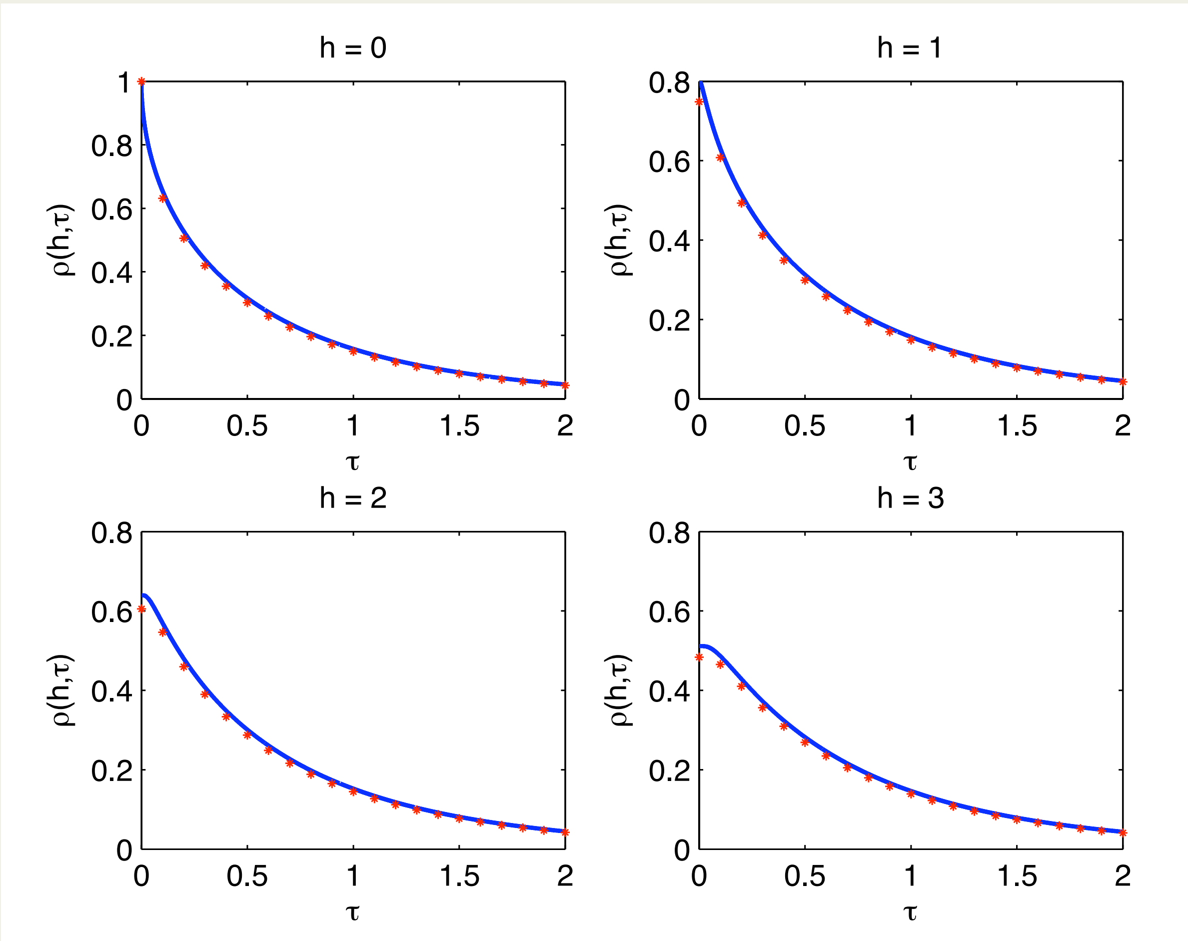
$$\mathbf{C}_Y^{(m)} = \mathbf{M}^m \mathbf{C}_Y^{(0)}$$

How does this marginal S-T covariance compare to the analytical one shown earlier for the continuous stochastic PDE?

where $$\mathrm{vec}(\mathbf{C}_Y^{(0)}) = \{\mathbf{I} - \mathbf{M} \otimes \mathbf{M}\}^{-1} \mathrm{vec}(\boldsymbol{\Sigma}_\delta)$$

PDE Ex: cont.

Comparison of Continuous and Stochastic Difference Equation Correlation Functions



Plots show temporal correlation functions for various spatial lags.

(α=1, β=20, Δ_s = 1, Δ_t = 0.01)

Red dots: discretized correlation values at intervals of $10\Delta_t$ = 0.1;
Blue lines: continuous correlation function

Real World Complexity

- The highly structured **M** could be estimated if we assume **θ** is unknown. But …
- Uncertainty!
  - We may have too little data (or too much) of varying quality and from multiple sources (with different levels of spatial and temporal support)
  - We don't know the true process!
  - We don't know what parameters really are important
- Hierarchical statistical models can help

Basic Hierarchical Model (see Mark Berliner's talk)

Basic rule of probability: [A,B,C] = [A|B,C][B|C][C]

Thus, for complicated problems:

[data, process, parameters] can be factored:

1. Data Model: [data | process, parameters] x
2. Process Model: [process| parameters] x
3. Parameter Model: [parameters]

We are interested in: [process, parameters | data]

(from Bayes' Rule; typically, can't get analytically)

Dynamic Spatio-Temporal Hierarchical Model

Observation Model: $\mathbf{Z}_t = \mathcal{H}(\mathbf{Y}_t; \boldsymbol{\theta}_h; \boldsymbol{\epsilon}_t)$

Data vector

Mapping function

Spatio-temporal process vector of interest (hidden)

Parameters

Errors

Process Model: $\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots; \boldsymbol{\theta}_m; \boldsymbol{\eta}_t)$

Transition function with Markovian dependence structure (conditional independence); **controls dynamics**

Noise

Parameter Models: $[\boldsymbol{\theta}_h][\boldsymbol{\theta}_m][\boldsymbol{\epsilon}_t][\boldsymbol{\eta}_t]$

Adding dependence structure and conditional relationships to these is what gives hierarchical modeling its power!

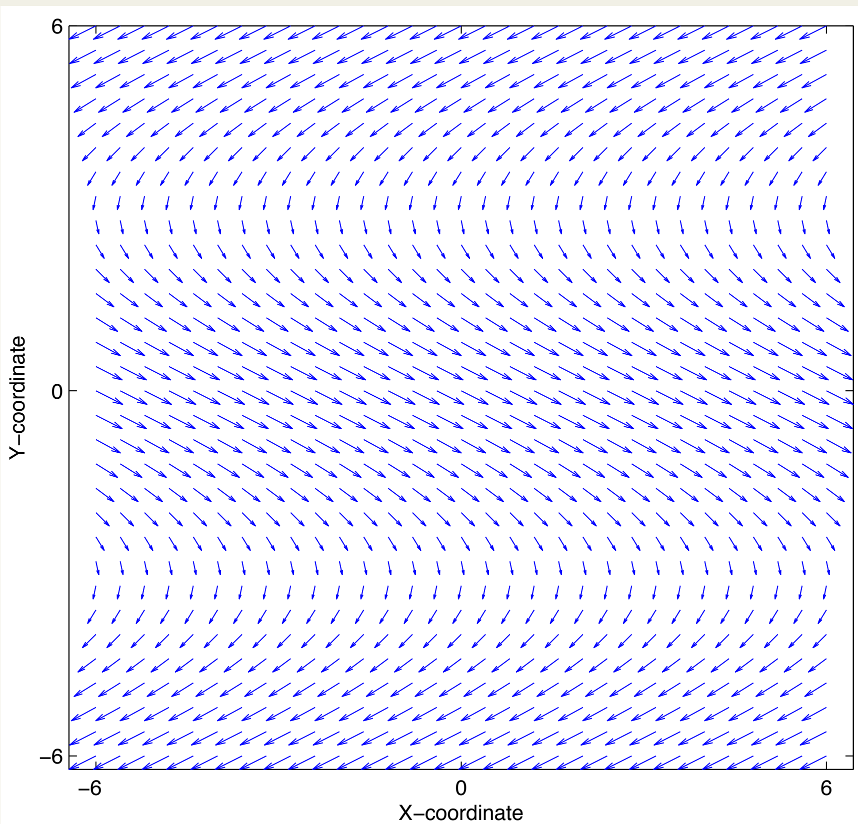
Hierarchical Parameterization

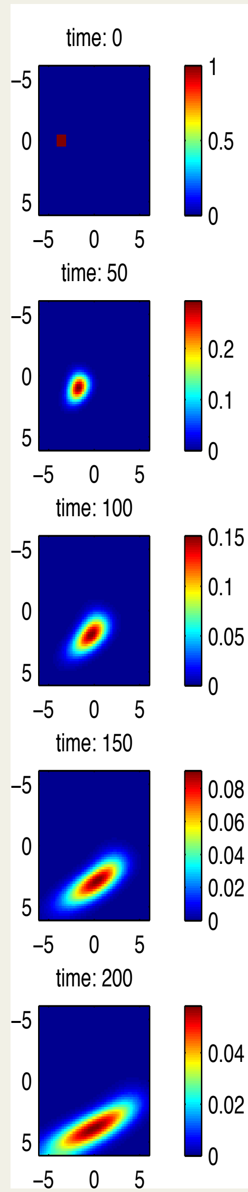
What if we don’t know the exact form of the underlying model, or if the underlying system is too complicated to derive the analytical marginal correlation function?

- Using a simple/approximate model as a template and allowing the parameters (e.g., θ) to be random, and (usually) structured in space (e.g., random fields) and/or time (time series) gives the model flexibility to adapt to the data, but still accommodates the basic process dynamics.

Example: $\frac{\partial Y}{\partial t} = c_1(x,y)\frac{\partial Y}{\partial x} + c_2(x,y)\frac{\partial Y}{\partial y} + \frac{\partial}{\partial x}\left(b_1(x,y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(b_2(x,y)\frac{\partial Y}{\partial x}\right)$

Advection-diffusion simulation with $c_1(x,y)$ and $c_2(x,y)$ given as suggested below (and with constant diffusion parameters.)





In the case where we didn't know these parameters, we could specify a prior distribution for them that might include covariates and/or spatial random fields in the hierarchical framework: e.g.,

$$\mathbf{c}|\boldsymbol{\theta}_c, \boldsymbol{\beta} \sim Gau(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_c))$$

Basic Hierarchical Model

Data: $$\mathbf{Z}_t = \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_r))$$

Process: $$\mathbf{Y}_t = \mathbf{M}(\boldsymbol{\theta}_m)\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_q))$$

Parameters: $\mathbf{M}, \mathbf{R}, \mathbf{Q}$

Critically, these can be structured according to the the science-based models, given the parameters.

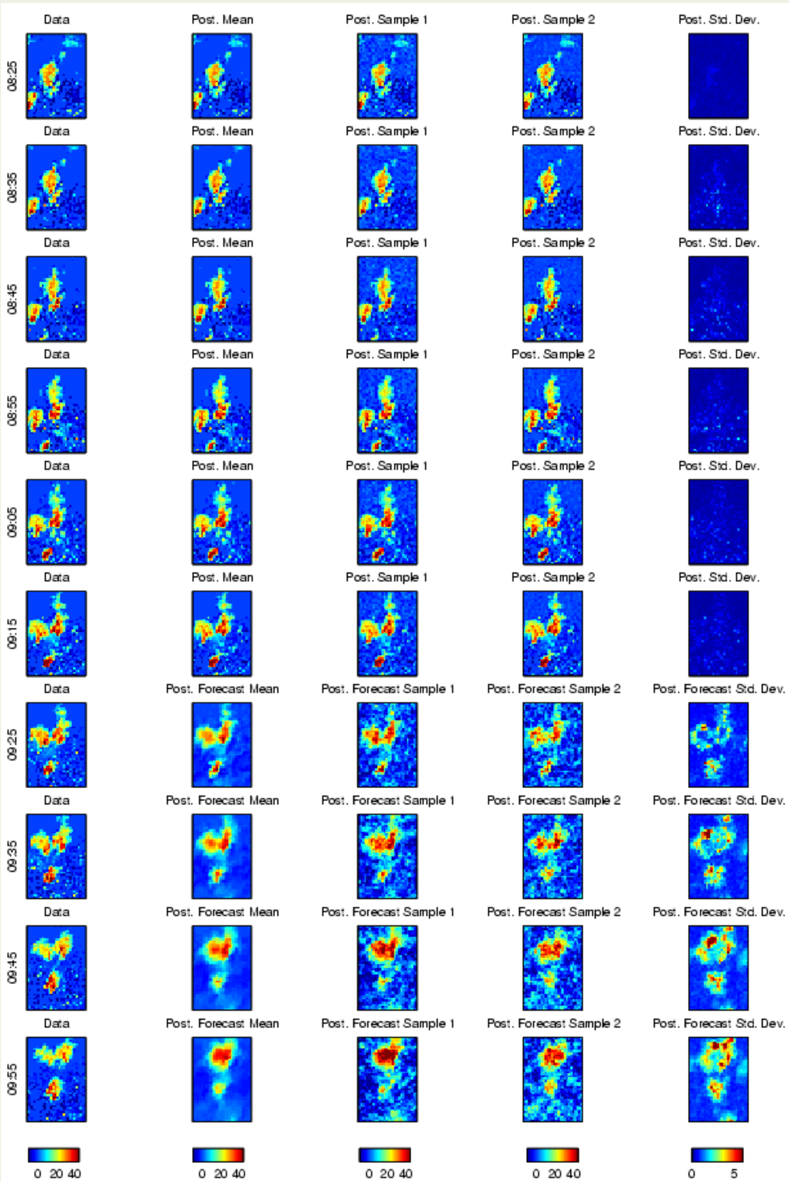
$\boldsymbol{\theta}_m, \boldsymbol{\theta}_r, \boldsymbol{\theta}_q$

These parameters are then given prior distributions, such as Gaussian random processes (that may depend on other variables), and can easily be allowed to vary with time and/or space so as to borrow strength.

Example: Radar Nowcasting

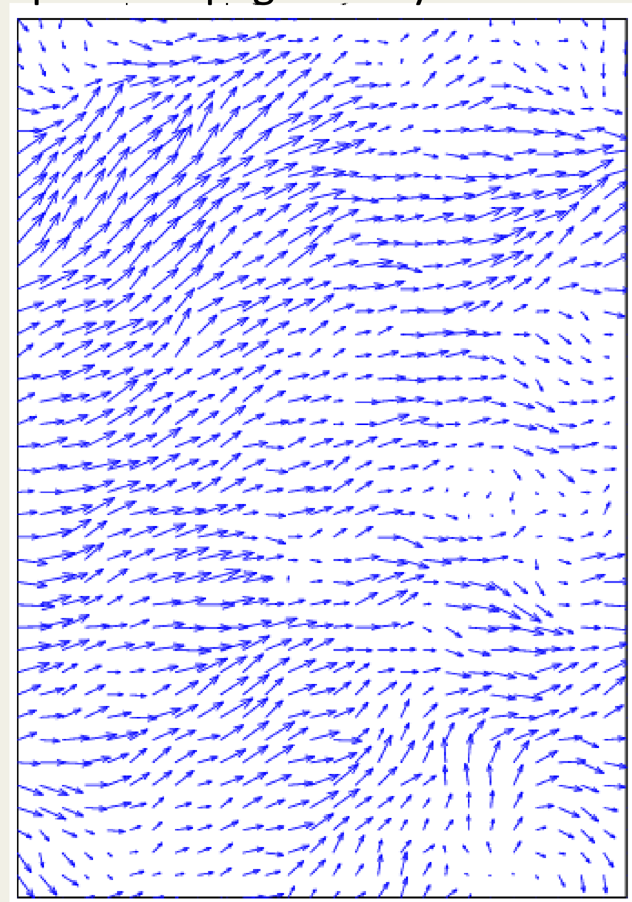
Data Mean Post. Real. Std.Dev.

For.



Statistical model motivated by a linear advection-diffusion process with spatially varying parameters.

Implied Propagation by Post. Parms.



Xu, Wikle, and Fox, 2005; JASA

Nonlinear Spatio-Temporal Statistical Models

- Clearly, models of the form: $\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots; \boldsymbol{\theta}_m)$ are too general.
- A common and useful model in the time-series literature is the state-dependent model:

$$\mathbf{Y}_t = \mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_m)\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t$$

- In the spatio-temporal statistics context, this model is still too general, and we need to think of specific, yet flexible, forms for the transition matrix, $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_m)$

General Quadratic Nonlinearity

We focus on a class of S-T models characterized by what can be termed general quadratic nonlinearity: in scalar form,

$$Y_t(s_i) = \sum_{j=1}^{n} a_{ij} Y_{t-1}(s_j) + \sum_{k=1}^{n} \sum_{l=1}^{n} b_{i,kl} Y_{t-1}(s_k) g(Y_{t-1}(s_l); \boldsymbol{\theta}_g) + \eta_t(s_i),$$

for i=1,…,n.

- Model includes quadratic (dyadic) interactions in random process
- The term "general" refers to the term: $g(Y_{t-1}(s_l); \boldsymbol{\theta}_g)$
- Note that there are $O(n^3)$ parameters in this model!
- Note, if g( ) is the identity function, then there are n(n+1)/2 unique dyadic interactions for each i=1,…,n; otherwise there are n^2.
- This can be recast as a matrix equation: parameters in **A**, **B**, $\boldsymbol{\theta}_g$

General Quadratic Nonlinearity

- There are different ways to write this as a matrix equation. E.g.,

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + (\mathbf{I}_n \otimes g(\mathbf{Y}_{t-1};\boldsymbol{\theta}_g)')\mathbf{B}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t,$$

where the $n^2 \times n$ matrix $\mathbf{B}$ is given by:

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_n \end{pmatrix}, \qquad \mathbf{B}_i \equiv \{b_{i,kl}\}_{k,l=1,\ldots,n}$$

Thus, in terms of the previously presented general state-dependent model: $\mathbf{Y}_t = \mathbf{M}(\mathbf{Y}_{t-1};\boldsymbol{\theta}_m)\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t$, we have

$$\mathbf{M}(\mathbf{Y}_{t-1};\boldsymbol{\theta}_m) = \mathbf{A} + (\mathbf{I}_n \otimes g(\mathbf{Y}_{t-1};\boldsymbol{\theta}_g)')\mathbf{B}$$

with parameters $\boldsymbol{\theta}_m = \{\mathbf{A}, \mathbf{B}, \boldsymbol{\theta}_g\}$.

Parameterizations

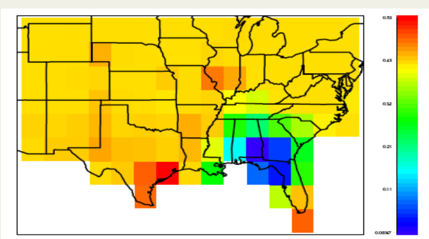
- For most general spatio-temporal processes, **A** and **B** (especially) have too many parameters to estimate reliably.
- Similar to the case with linear spatio-temporal models, knowledge of the process can motivate specific parameterizations.
  - Combined with hierarchical (conditional) specification of parameters, this can provide an effective modeling approach.
    - Ex: Quasi-geostrophy as motivation for a statistical model of ocean streamfunction
    - Ex: Eurasian Collared Dove, Reac.-Diff. Equation

Example: Invasive Species Prediction

Reaction-Diffusion Models: (e.g., density dependent growth for **invasive species**); e.g.,

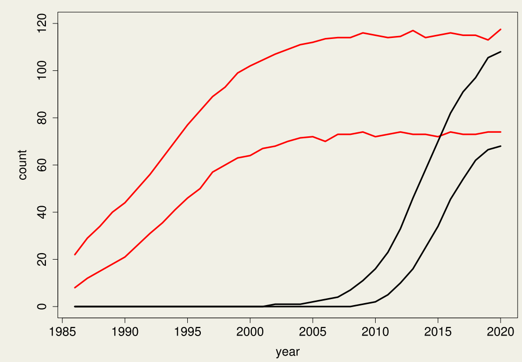
$$\frac{\partial Y}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial Y}{\partial y}\right) + \gamma_0(x,y) Y \exp\left(1 - \frac{Y}{\gamma_1(x,y)}\right)$$

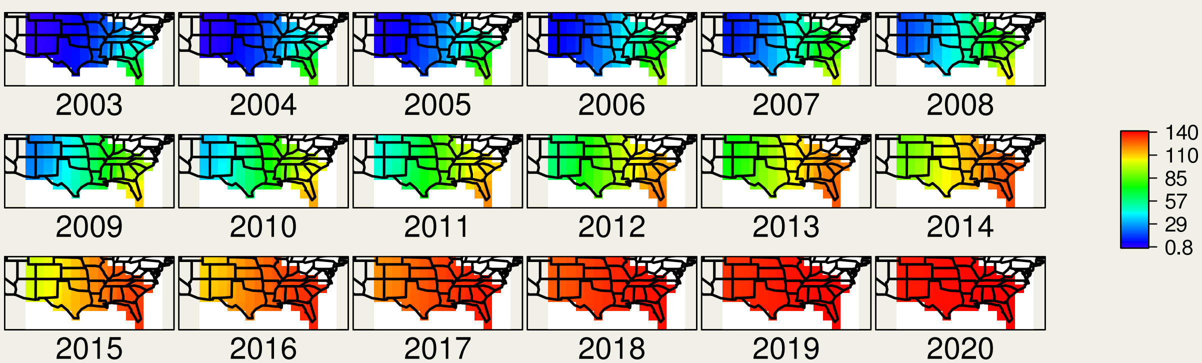
Depends on spatially-explicit random diffusion coefficients $\delta(x,y)$ and carrying capacity $\gamma_1(x,y)$ and growth $\gamma_0(x,y)$ terms specified at a lower level of the model hierarchy (e.g., Hooten and Wikle, 2007; Hooten et al. 2007; **Eurasian Collared Dove Invasion**).



Dispersal δ: post mean

95% PI's for S. Fl and Utah





Forecast relative abundance: post mean

Dimension Reduction

In many cases, the dynamics may not be known or are more complicated than suggested by a single PDE/IDE.

Consider the spectral representation, $\mathbf{Y}_t \approx \mathbf{\Phi}\boldsymbol{\alpha}_t$, where $\boldsymbol{\alpha}_t$ is of dimension $p \times 1$ where $p << N$. We could then model this reduced-dimensional process in terms of quadratic interactions:

$$\alpha_t(i) = \sum_{j=1}^{p} A_{ij}\alpha_{t-1}(j) + \sum_{k=1}^{p}\sum_{l=1}^{k} b_{i,kl}\alpha_{t-1}(k)g(\alpha_{t-1}(l);\boldsymbol{\theta}_g) + \eta_{i,t},$$

Still order p^3 parameters here! Unless p is very small, we still must make some simplifying assumptions and perform model selection.

(Note: choice of ϕ is a very important topic – beyond the scope of this talk!)

Naïve Statistical Simplification by Scale Analysis

Say we can write $\mathbf{Y}_t = \mathbf{\Phi}^{(1)}\boldsymbol{\alpha}_t^{(1)} + \mathbf{\Phi}^{(2)}\boldsymbol{\alpha}_t^{(2)} + \boldsymbol{\nu}_t$, where $\boldsymbol{\alpha}_t^{(i)}$ is of dimension $p_i \times 1$ and where $p_i < N$.

Now, assume that the dyadic interactions between components of $\boldsymbol{\alpha}_t^{(1)}$ are explicit, but those among the "small scale" components $\boldsymbol{\alpha}_t^{(2)}$ are "noise" and the interactions between the components of $\boldsymbol{\alpha}_t^{(1)}$ and $\boldsymbol{\alpha}_t^{(2)}$ imply random coefficients. (motivated by Reynolds averaging)

Although not necessarily physically realistic, this simple procedure illustrates some beneficial features of the hierarchical statistical approach.

As a simple example, consider

$$\boldsymbol{\alpha}_t^{(1)} \equiv (\alpha_{1,t}^{(1)}, \alpha_{2,t}^{(1)})'$$

$$\boldsymbol{\alpha}_t^{(2)} \equiv (\alpha_{1,t}^{(2)}, \alpha_{2,t}^{(2)}, \alpha_{3,t}^{(2)})'$$

Example: Scale Analysis Reduction

$$\boldsymbol{\alpha}_t \equiv \begin{pmatrix} \boldsymbol{\alpha}_t^{(1)} \\ \boldsymbol{\alpha}_t^{(2)} \end{pmatrix}$$

Large Scale Modes:

$$\boldsymbol{\alpha}_t^{(1)} \equiv (\alpha_{1,t}^{(1)}, \alpha_{2,t}^{(1)})'$$

Small Scale Modes:

$$\boldsymbol{\alpha}_t^{(2)} \equiv (\alpha_{1,t}^{(2)}, \alpha_{2,t}^{(2)}, \alpha_{3,t}^{(2)})'$$

Assume g( ) is the identity function here.

All Dyadic Interactions:

$$\left.\begin{array}{l} (\alpha_{1,t}^{(1)}\alpha_{1,t}^{(1)}) \\ (\alpha_{1,t}^{(1)}\alpha_{2,t}^{(1)}) \\ (\alpha_{2,t}^{(1)}\alpha_{2,t}^{(1)}) \end{array}\right\} \text{Resolved dyadic interactions}$$

$$\left.\begin{array}{l} (\alpha_{1,t}^{(1)}\alpha_{1,t}^{(2)}) \\ (\alpha_{1,t}^{(1)}\alpha_{2,t}^{(2)}) \\ (\alpha_{1,t}^{(1)}\alpha_{3,t}^{(2)}) \\ (\alpha_{2,t}^{(1)}\alpha_{1,t}^{(2)}) \\ (\alpha_{2,t}^{(1)}\alpha_{2,t}^{(2)}) \\ (\alpha_{2,t}^{(1)}\alpha_{3,t}^{(2)}) \end{array}\right\} \text{Linear in large scale modes with random coefficients}$$

$$\left.\begin{array}{l} (\alpha_{1,t}^{(2)}\alpha_{1,t}^{(2)}) \\ (\alpha_{1,t}^{(2)}\alpha_{2,t}^{(2)}) \\ (\alpha_{1,t}^{(2)}\alpha_{3,t}^{(2)}) \\ (\alpha_{2,t}^{(2)}\alpha_{2,t}^{(2)}) \\ (\alpha_{2,t}^{(2)}\alpha_{3,t}^{(2)}) \\ (\alpha_{3,t}^{(2)}\alpha_{3,t}^{(2)}) \end{array}\right\} \text{Process noise terms (dependence)}$$

Hierarchical Model

The following hierarchical model is suggested:

$$\mathbf{Z}_t = \mathbf{\Phi}^{(1)}\boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim N(\mathbf{0}, \mathbf{R})$$

$$\boldsymbol{\alpha}_t^{(1)} = \mathbf{A}\boldsymbol{\alpha}_{t-1}^{(1)} + (\mathbf{I}_{p_1} \otimes \boldsymbol{\alpha}_{t-1}^{(1)\prime})\mathbf{B}\boldsymbol{\alpha}_{t-1}^{(1)} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q})$$

Let $\mathbf{R} = \kappa\mathbf{I} + \sum_{k=p_1+1}^{p_1+p_2} \lambda_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k'$ where $\kappa^{-1} \sim Gamma(q_\kappa, r_\kappa)$

$$\mathbf{Q}^{-1} \sim Wishart((\nu\mathbf{S})^{-1}, \nu)$$

$$vec(\mathbf{A}) \sim N(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A)$$

$[\mathbf{B}]$ (see below)

Our choices for these hyperparameters may reflect our prior understanding of the importance of certain modes and their interaction; or, the can be given distributions of their own!

Stochastic Search Variable Selection

(George and McCulloch, 1993; 1997)

Without additional information, there are still likely to be too many parameters in B to get reliable statistical "estimates". Again, we can utilize the hierarchical framework to help. Let,

$$\tilde{\mathbf{b}} = (\tilde{b}_1, \ldots, \tilde{b}_{n_b})' \equiv vec(\mathbf{B})$$

$$\tilde{b}_j | \gamma_j \sim \gamma_j N(0, c_j^2 \tau_j^2) + (1 - \gamma_j) N(0, \tau_j^2),$$

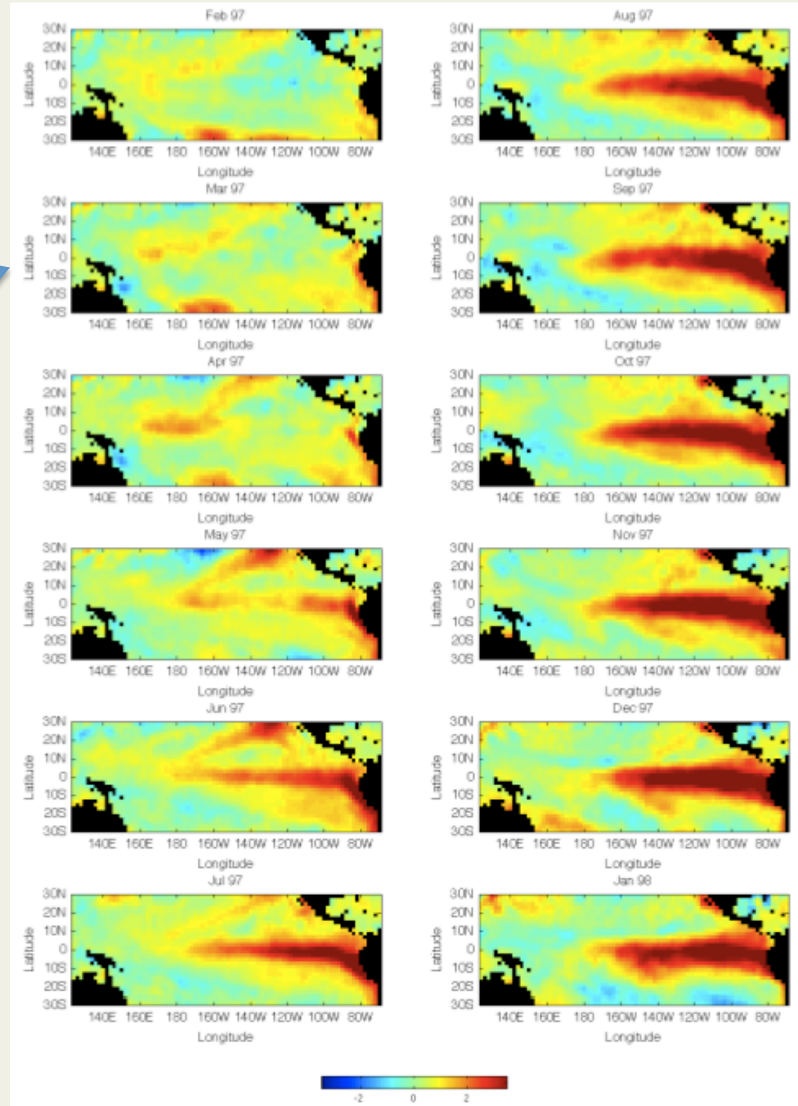
$$\gamma_j \sim Bernoulli(\pi_j),$$

where $\gamma_j = 1$ means that the j-th variable is in the model.

We specify π_j, c_j, τ_j such that c_j is "large" and τ_j is "small" to favor $\tilde{b}_j$ having a small value if it is not "selected" in the model.

Example: Long-Lead Prediction of Tropical Pacific SST

Given SST up to March 1997



(Note: each image contains about 2500 pixels. There are about 300 times (months).)

Forecast SST 7 months later in Oct 1997

Long-Lead Prediction of SST

- SST is a complicated process associated with atmosphere/ ocean interactions on a variety of time and space scales. Its dynamics are not completely understood.
- One of the few situations in oceanography in which ``statistical'' forecast models are often as more skillful than deterministic models (Barnston et al, 1999; van Oldenborgh et al. 2005)
- Linear process models in reduced dimensional space (e.g., Φ are EOFs - spatial principal components) have proven to be pretty effective over the years (e.g., Penland and Magorian, 1993)
- Evidence that ENSO is not linear (e.g., Hoerling, et al. 1997; Burgers and Stephenson 1999)
- A simple nonlinear statistical model can do even better (e.g., Berliner, Wikle, Cressie, 2000; Kondrashov et al. 2005)

SST: Quadratic Nonlinear Hierarchical Model Implementation

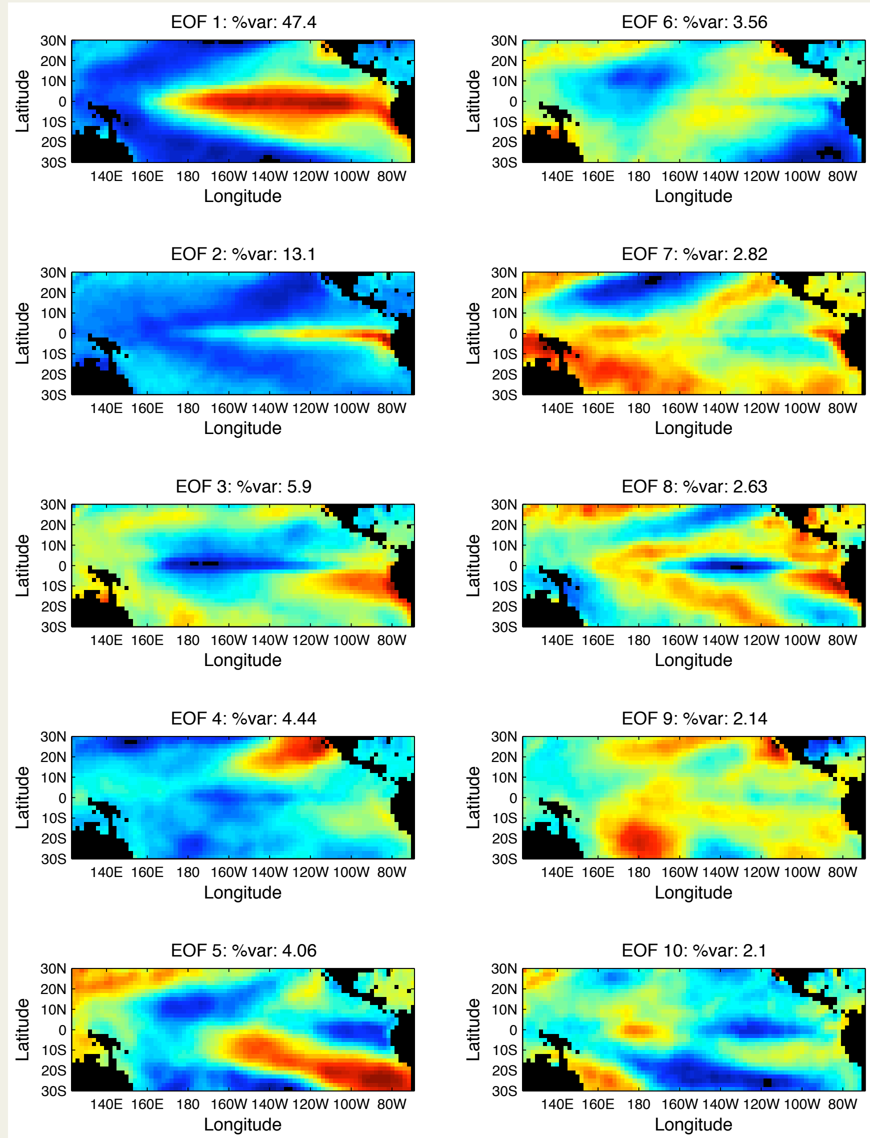
$\mathbf{\Phi}^{(1)}$ - EOF (spatial principal components)

$p_1 = 10$

Data: Monthly Pacific SST anomalies from January 1970 - March 1997 to forecast October 1997

Standard MCMC implementation; vague priors on all parameters except data model variance. (*PRELIMINARY RESULTS*)

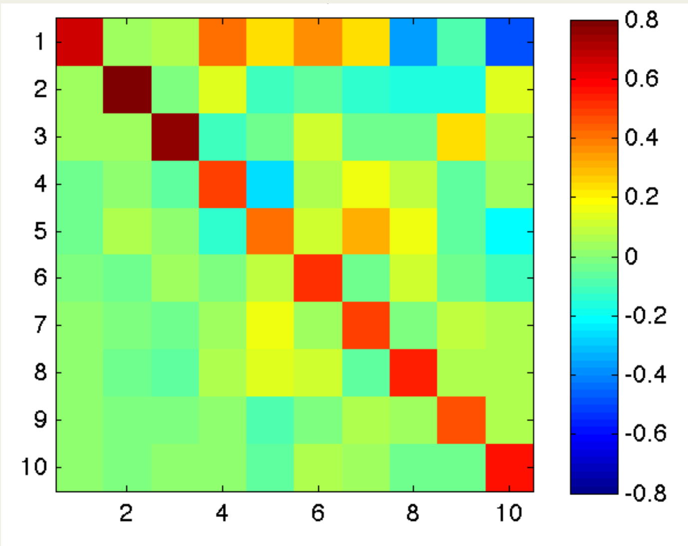
First 10 EOF Patterns



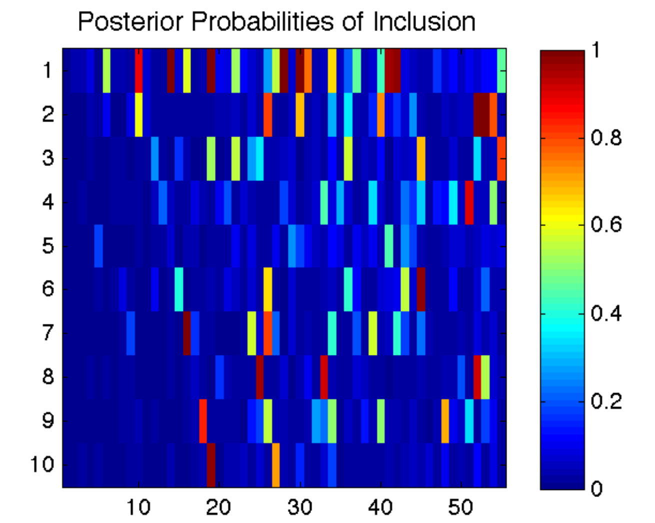
Posterior Means: Parameters

$$\boldsymbol{\alpha}_t^{(1)} = \mathbf{A}\boldsymbol{\alpha}_{t-1}^{(1)} + (\mathbf{I}_{p_1} \otimes \boldsymbol{\alpha}_{t-1}^{(1)\prime})\mathbf{B}\boldsymbol{\alpha}_{t-1}^{(1)} + \boldsymbol{\eta}_t,$$

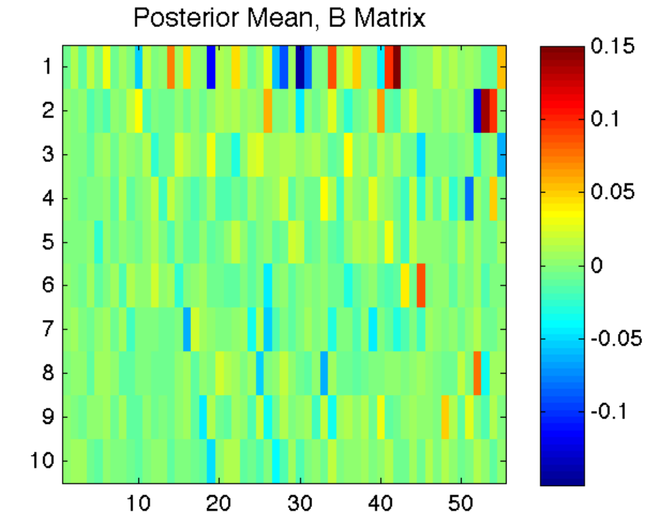
Posterior Mean: A matrix (linear term)



B matrix inclusion probabilities



B matrix



Forecast: October 1997 from March 1997

Nonlinear Model

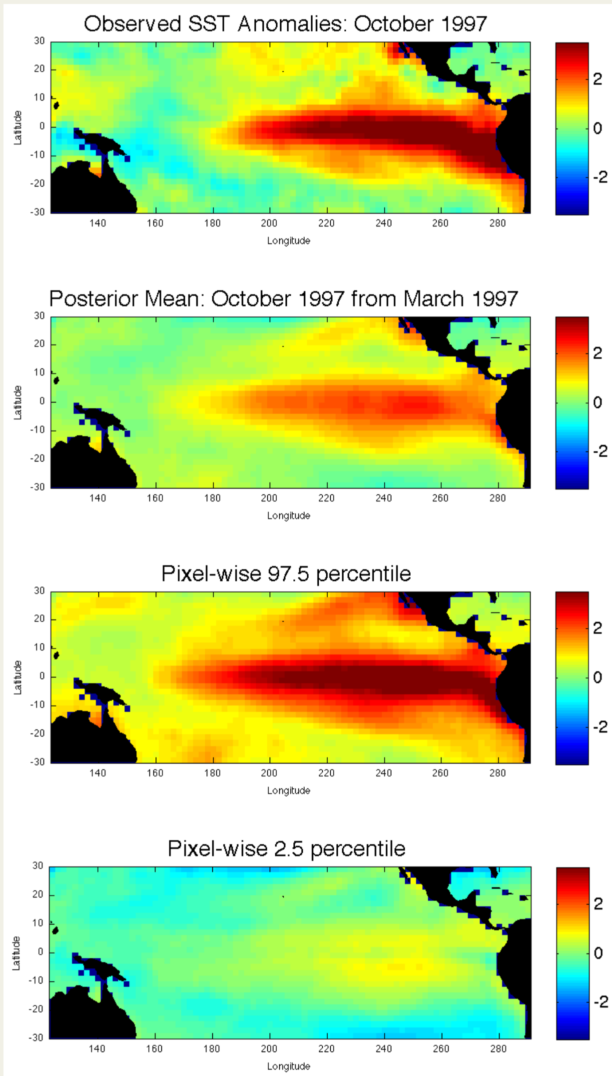
Linear (VAR) Model

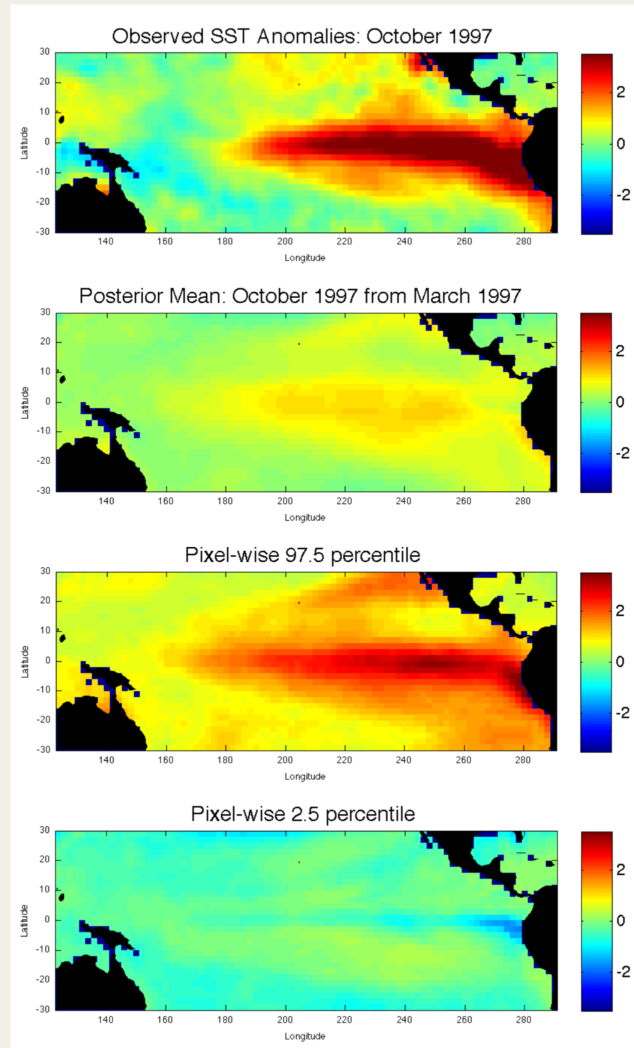
Obs

Post. Mean

Post. Pixel 97.5%-tile

Post. Pixel 2.5%-tile





Forecast: October 1998 from March 1998

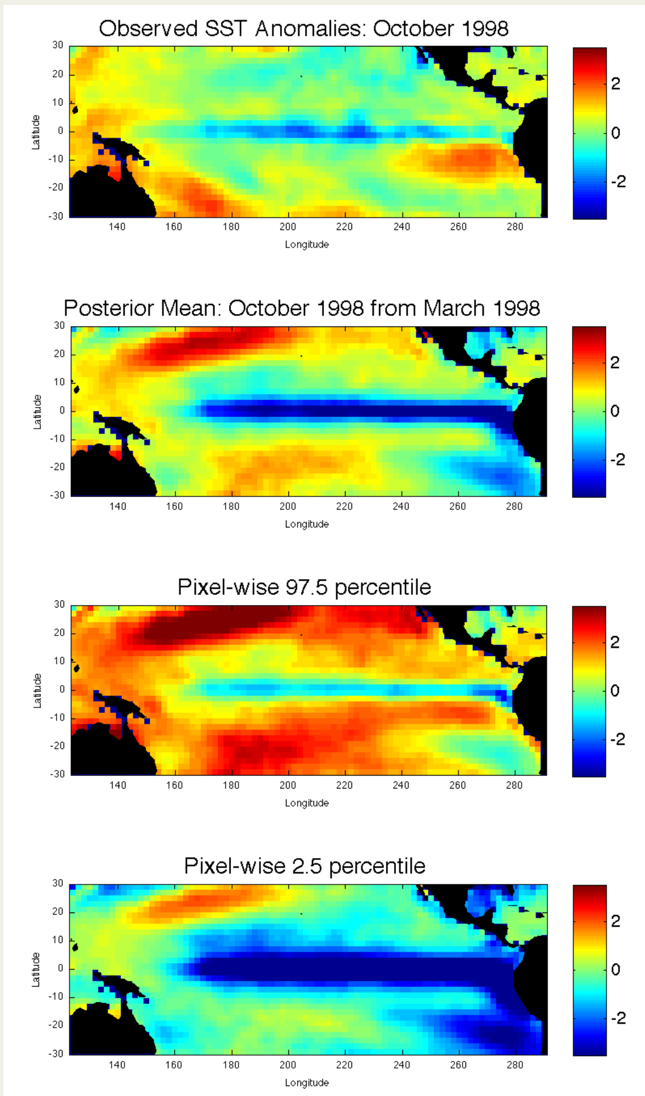
Obs

Post. Mean

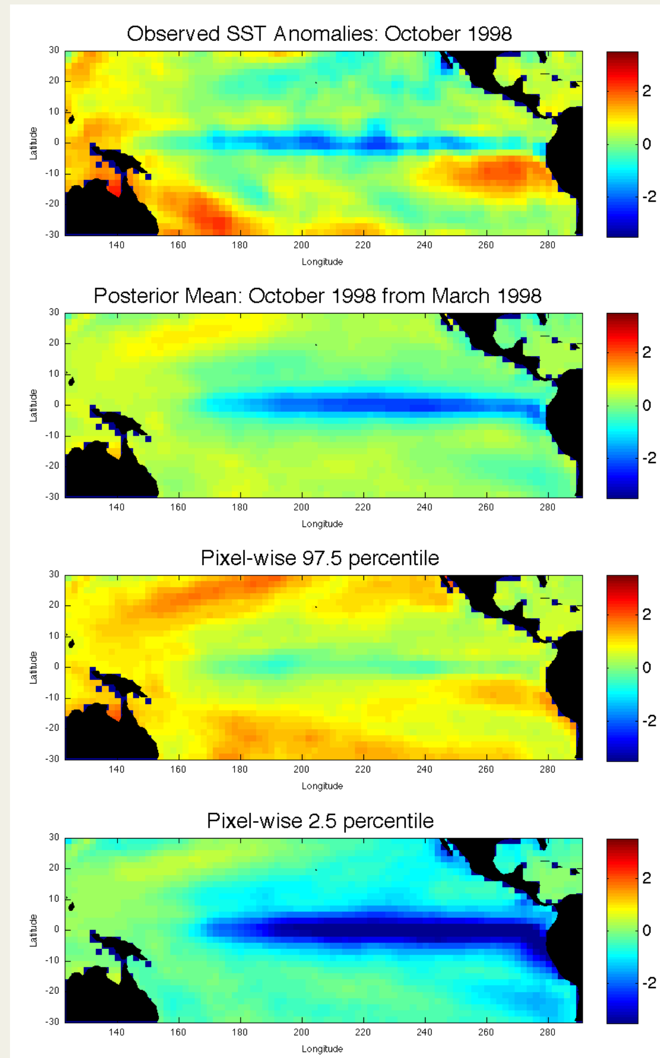
Post. Pixel 97.5%-tile

Post. Pixel 2.5%-tile

Nonlinear Model



Linear (VAR) Model



Extensions

It is relatively simple to add other types of dependence as well. For example, say we want to allow the dynamics to change with time: e.g., (threshold AR model)

$$\mathbf{B}_t = \begin{cases} \mathbf{B}_0, & I_t = I_0 \\ \mathbf{B}_1, & I_t = I_1 \\ \mathbf{B}_2, & I_t = I_2 \end{cases}$$

Where I_t is either a "known" index (e.g., SOI, ONI) or it might be random as well and can be related to other features of the atmosphere/ocean system (e.g., see Berliner, Wikle, Cressie, 2000.)

Non-Gaussian Data/Process/Parameters

- These methods have been applied to non-Gaussian situations as well. E.g.,
  - Eurasian Collared Dove invasive species example
    - Data from a Poisson process
  - Tornado counts related to climate indices
    - Data from a zero-inflated Poisson process
  - Spread of rabies in New England
    - Data from a Bernoulli model; process model motivated by stochastic cellular automata
  - Models of the lower-trophic ecosystem (N,P,Z, etc.) in coupled ocean-biogeochemical models
    - Data truncated normal; process and parameters constrained to have non-negative support

Conclusion

- Physical models can provide motivation for statistical parameterizations of linear and non-linear dynamic models for spatio-temporal processes.
- Critical that we don't expect the process to follow the physical models exactly, but expect the implied statistical model to be flexible enough to accommodate realistic dynamics.
- The statistical hierarchical model allows one to incorporate additional information (data and science) into parameter structures
- Statistical models for many spatio-temporal processes suffer from the curse of dimensionality.
- Science-based dimension reduction and hierarchical implementation of variable selection can help

Cited and Relevant Literature

Berliner, L.M., Wikle, C.K., and N. Cressie, 2000: Long-lead prediction of Pacic SSTs via Bayesian Dynamic Modeling. Journal of Climate, 13, 3953-3968.

Cressie, N. and C.K. Wikle, 2011: *Statistics for Spatio-Temporal Data.* John-Wiley & Sons. New York.

Hooten, M.B. and C.K. Wikle, 2010: Statistical agent-based models for discrete spatio-temporal systems*. Journal of the American Statistical Association*, 105: 236-248.

Berliner, L.M., Milli, R.F., and C.K. Wikle, 2003: Bayesian hierarchical modeling of air-sea interaction. Journal of Geophysical Research - Oceans, 108(C4).

Hooten, M.B. and C.K. Wikle (2008). A Hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*, 15(1): 59-70.

Hooten, M.B., Wikle, C.K., Dorazio, R.M., and J.A. Royle, 2007: Hierarchical matrix models for characterizing invasions, *Biometrics*, 63, 558-567.

Song, Y., Wikle, C.K., Anderson, C.J., and S.A. Lack, 2007: Bayesian estimation of stochastic parameterizations in a numerical weather forecasting model, Monthly Weather Review. 135, 4045-4059.

Wikle, C.K. and M.B. Hooten, 2010: A general science-based framework for spatio-temporal dynamical models. Invited discussion paper for *Test*. (forthcoming)

Wikle, C.K., and L. M. Berliner, 2007: A Bayesian tutorial for data assimilation*, Physica D*, 230, 1-16.

Wikle, C.K., 2003: Hierarchical models in environmental science. *International Statistical Review*, 71, 181-199.

Wikle, C.K., 2003: Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84, 1382-1394.

Wikle, C.K., L.M. Berliner, and R.F. Milli, 2003. Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. Monthly Weather Review, 131, 1051-1062.

Wikle, C.K. and C.J. Anderson, 2003: Climatological analysis of tornado report counts using a hierarchical Bayesian spatio-temporal Model. Journal of Geophysical Research – Atmospheres, 108(D24), 9005, doi: 10.1029/2002JD002806.

Wikle, C.K., Milli, R.F., Nychka, D., and L.M. Berliner, 2001: Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds, Journal of the American Statistical Association, 96, 382-397.

Xu, K., Wikle, C.K., and N.I. Fox, 2005: A kernel-based spatio-temporal dynamical model for nowcasting radar refectivities. Journal of the American Statistical Association, 100, 1133-1144.

Cited and Relevant Literature (cont.)

Barnston, A.G. and C. F. Ropelewski, 1992. Prediction of ENSO episodes using canonical correlation analysis. *Journal of Climate, 5:1316–1345.*

Burgers, G. and D. B. Stephenson, 1999. The “normality” of El Ni ˜no. *Geophysical Research Letters* 26:1027–1030.

George, E.I. and R. E. McCulloch, 1997. Approaches for Bayesian variable selection. Statistica Sinica ,**7**, 339-373.

George, E.I. and McCulloch, R.E., 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association. 88, 881-889.

Heine, V., 1955. Models for two-dimensional stationary stochastic processes. Biometrika, 42:170178.

Hoerling,M.P., A. Kumar, and M. Zhong, 1997. El Ni˜no, La Ni˜na, and the nonlinearity of their teleconnections. *Journal of Climate, 10:1769–1786.*

Hotelling, H., 1927. Differential equations subject to error, and population estimates. Journal of the American Statistical Association, 22:283–314.

Jones, R.H., Zhang, Y.,1997. Models for continuous stationary spacetime processes. In: Gregoire, T.G., Brillinger, D.R., Diggle, P.J., Russek-Cohen, E., Warren, W.G., Wolfinger, R.D. (Eds.), Modelling Longitudinal and Spatially Correlated Data, Lecture Notes in Statistics, Vol. 122. Berlin: Springer, 289298.

Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil, 2005. A hierarchy of databased ENSO models. *Journal of Climate, 18:4425–4444.*

Penland, C. and T. Magorian, 1993. Predicton of Ni˜no 3 sea surface temperatures using linear inverse modeling. *Journal of Climate, 6:1067–1076.*

van Oldenborgh, G.J., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. T. Anderson, 2005. Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *Journal of Climate, 18:3240–3249.*

Whittle, P. *Systems in Stochastic Equilibrium. John Wiley & Sons, Ltd., Chichester, 1986.*

P. Whittle, P. Stochastic process in several dimensions. *Bulletin of the International Statistical Institute, 40:974–994, 1963.*

Yule, G.U., 1927. On a method of investigating periodicities in disturbed series, with reference to Wolfer’s sunspot numbers. Philosophical Transactions of the Royal Society of London, Series A, 226:267–298.