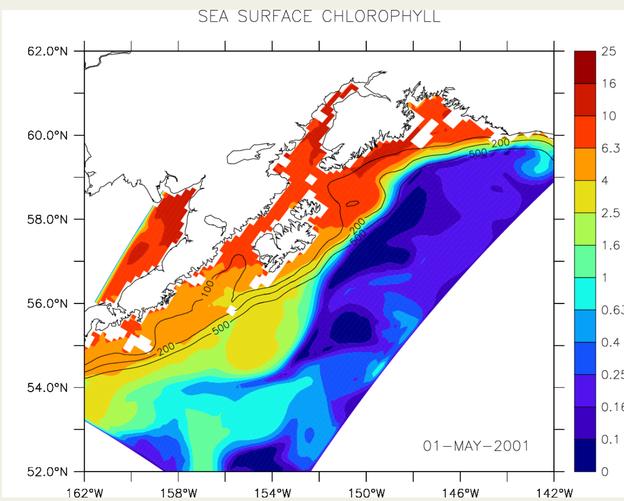
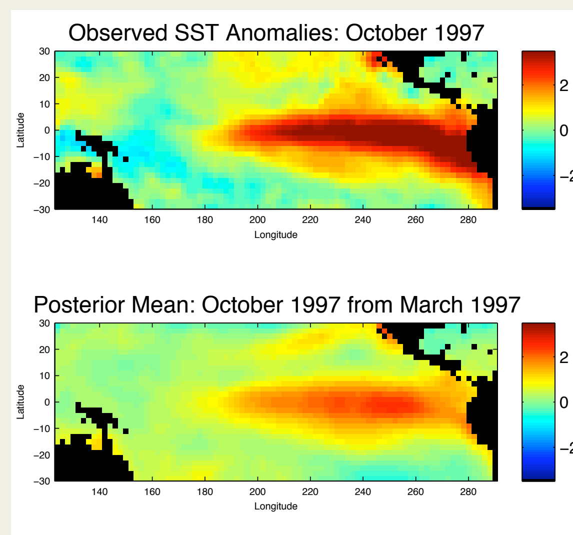
Hierarchical General Quadratic Nonlinear Models for Spatio-Temporal Dynamics

Christopher K. Wikle

Department of Statistics

University of Missouri



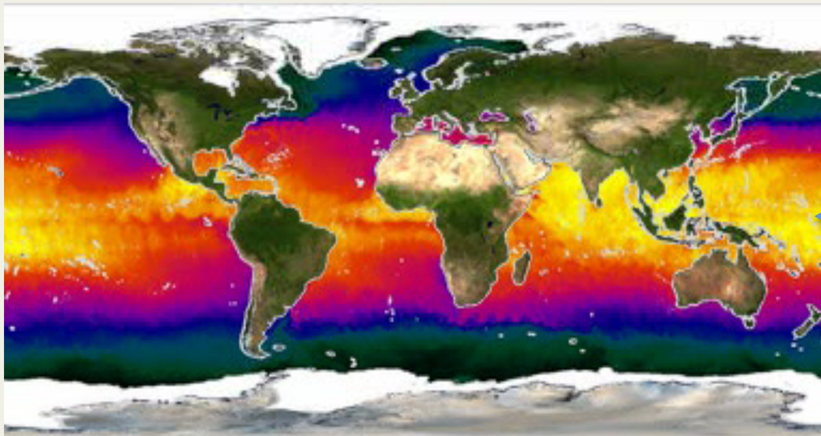


Red Raider 2012

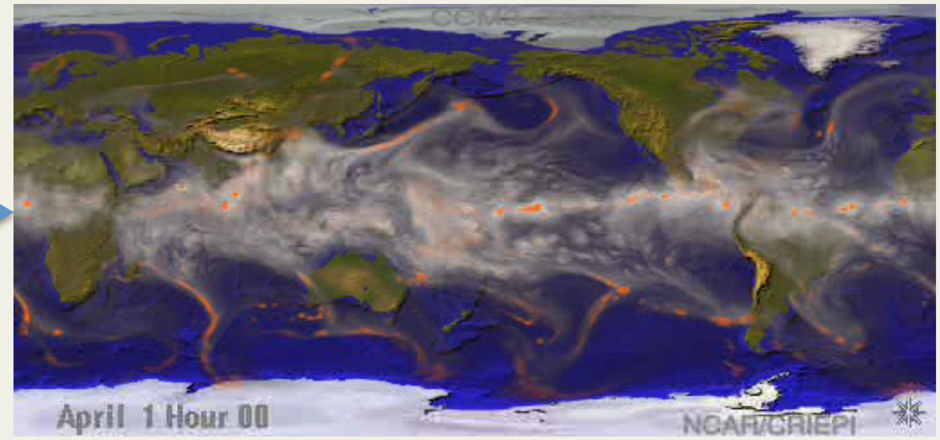
Dynamic Spatio-Temporal Processes

(Sea Surface Temperature, SST)

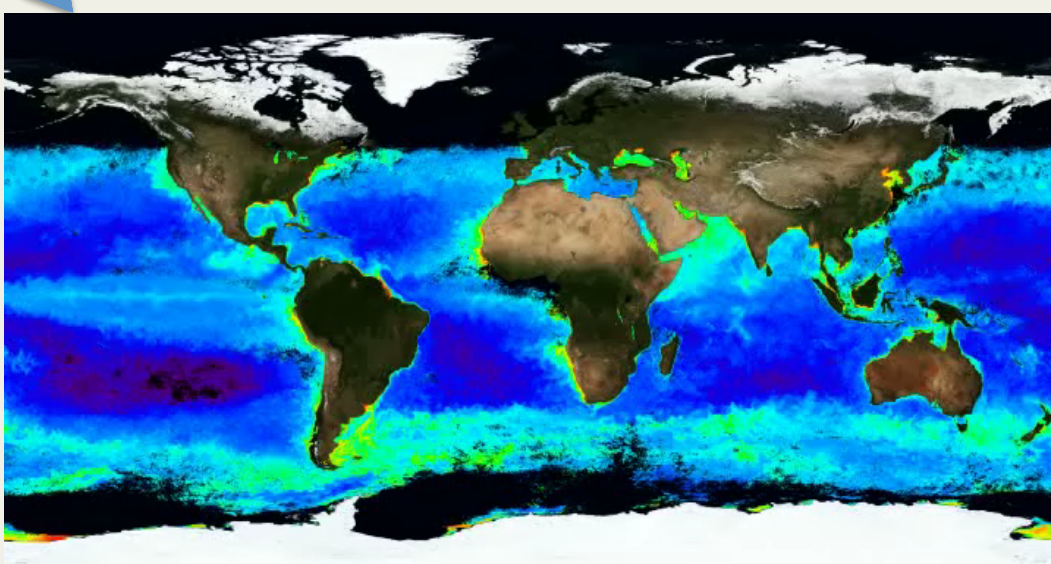
NASA AMSR: SST (scientific visualization studio)



NCAR CCM3 Cloud/Precipitation Simulation (NCAR VETS)



NASA SEAWIFS: Ocean Color (scientific visualization studio)



proxy for ocean phytoplankton

Interaction across atmosphere, ocean, and ecosystem.

multivariate; nonlinear

Complexity

- Spatio-Temporal Dynamics - interacting components of a spatial process (or processes) evolving through time
  - This is how the real world works!
  - E.g., diffusion, advection, vorticity, energy cascades, repulsion, growth, density dependence, infection, waves, fronts
  - Many different scales of interaction, as well as different process components (variables)

Spatio-Temporal Modeling in Statistics

Purpose: Characterize processes in the presence of uncertain and (often) incomplete observations and system knowledge for the purposes of:

- Prediction in space (interpolation)
- Prediction in time (forecasting)
- Assimilation of observations and deterministic models
- Inference on controlling process parameters

* Estimation and prediction in the presence of **uncertainty** in data, process, and the associated parameters

Spatio-Temporal Statistical Models

Traditionally Two Primary Approaches

- Descriptive (marginal): Characterize the second moment (covariance) behavior of the process
  - Several different physical processes could imply the same marginal structure; problematic if non-Gaussian/nonlinear
  - 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
  - Particularly useful if there is some *a priori* knowledge available concerning the process behavior

There is a big gap between how dynamics is viewed in statistics as compared to applied math and geophysics!

This Talk: An Overview

- Focus on the conditional/dynamic perspective
- Specifically: *using process knowledge as motivation for parameterization and structure of spatio-temporal statistical models*
- Focus on statistical models that are discrete in time and space, but may be motivated by processes that are continuous in one or both
- Introduce a class of physically realistic parametric nonlinear dynamical spatio-temporal statistical models
  - General quadratic nonlinearity
    - Parameterization, implementation

Notation

Let $\{Y(\mathbf{s};t) : \mathbf{s} \in D_s \subset \mathbb{R}^d, t \in D_t \subset \mathbb{R}\}$ denote a spatio-temporal random process, where D_s is the spatial domain of interest, D_t the temporal domain of interest, $\mathbf{s}$ is a spatial location and t a time.

When we refer to discrete time, we will typically write $Y_t(\mathbf{s})$

It has become customary in hierarchical modeling to denote "distributions" using bracket notation. E.g.,

$[Z|Y]$ - a conditional distribution of Z given $Y = y$.

We also typically denote vectors and matrices by a bold font:

e.g., $\mathbf{Y}, \boldsymbol{\theta}$, $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))'$

Marginal/Descriptive Characterization

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

Reaction-Diffusion Eqn.

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

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

Implied spatio-temporal correlation

τ – temporal lag
h – spatial lag

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

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

Conditional Perspective

- Thus, in some cases general process knowledge (e.g., reaction-diffusion) can be used to develop classes of spatio-temporal covariance models
- Typically, analytical derivations in the statistics literature have only been given for relatively simple processes
- In some cases, because conditional models are closer to the process etiology, it is easier to incorporate process knowledge in that context directly (e.g., dynamic models)

Hierarchical Dynamic Spatio-Temporal 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)

Noise

Parameters: $\boldsymbol{\theta}_h, \boldsymbol{\theta}_m, \boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_\eta$

Typically, fixed and unknown in classical statistics ; random and perhaps structure (dependence) if Bayesian

Inference via Bayes' Rule:

$$[\mathbf{Y}, \boldsymbol{\Theta} | \mathbf{Z}] \propto [\mathbf{Z} | \mathbf{Y}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_\epsilon][\mathbf{Y} | \boldsymbol{\theta}_m, \boldsymbol{\theta}_\eta][\boldsymbol{\Theta}]$$

Statistical Dynamic Spatio-Temporal Models?

- Dimensionality can prevent the (efficient) estimation of the full transition operator $\mathcal{M}(\cdot)$
  - Requires sensible parameterizations and/or dimension reduction
  - Some types of parameterizations make sense for some spatio-temporal processes, and some don't (e.g., process knowledge should not be ignored if available)
  - Hierarchical representations can help

Consider a simple linear process model.

Statistician's Conditional Perspective

For a linear process, we might consider a first-order vector autoregressive process with unknown **M**:

Markovian assumption: condition on recent past

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

where

$$\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))'$$

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

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

Traditionally, simple (naïve) parameterizations have been used in statistics (e.g., univariate AR models; random walks, etc.).

It is important to to consider parameterizations that can accommodate the dynamics of the system under consideration.

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

Common Ground Suggests Statistical Model Parameterizations

Consider the Reaction-Diffusion PDE Example:

- We should be able to use our knowledge of the PDE to motivate the parameterization of the VAR model to facilitate its estimation for more complicated processes.
- As a toy example, consider a simple finite-difference discretization. [Note: we have also used more complicated differencing as well as spectral and Galerkin methods in this context; these can motivate different ST models.]

PDE-Motivated Parameterization: Ex

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

(The reaction-diffusion PDE example from before)

Replacing the first-order derivative with a forward difference and the second-order derivative with a centered difference gives a specific parameterization of a first-order vector autoregressive process:

$$\mathrm{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\delta}_t$$

Plus a boundary condition term.

where M is highly structured (tri-diagonal) and dependent on the parameters β, α and discretization constants. Under temporal stationarity, the lag-m (in time) spatial covariance matrices are

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

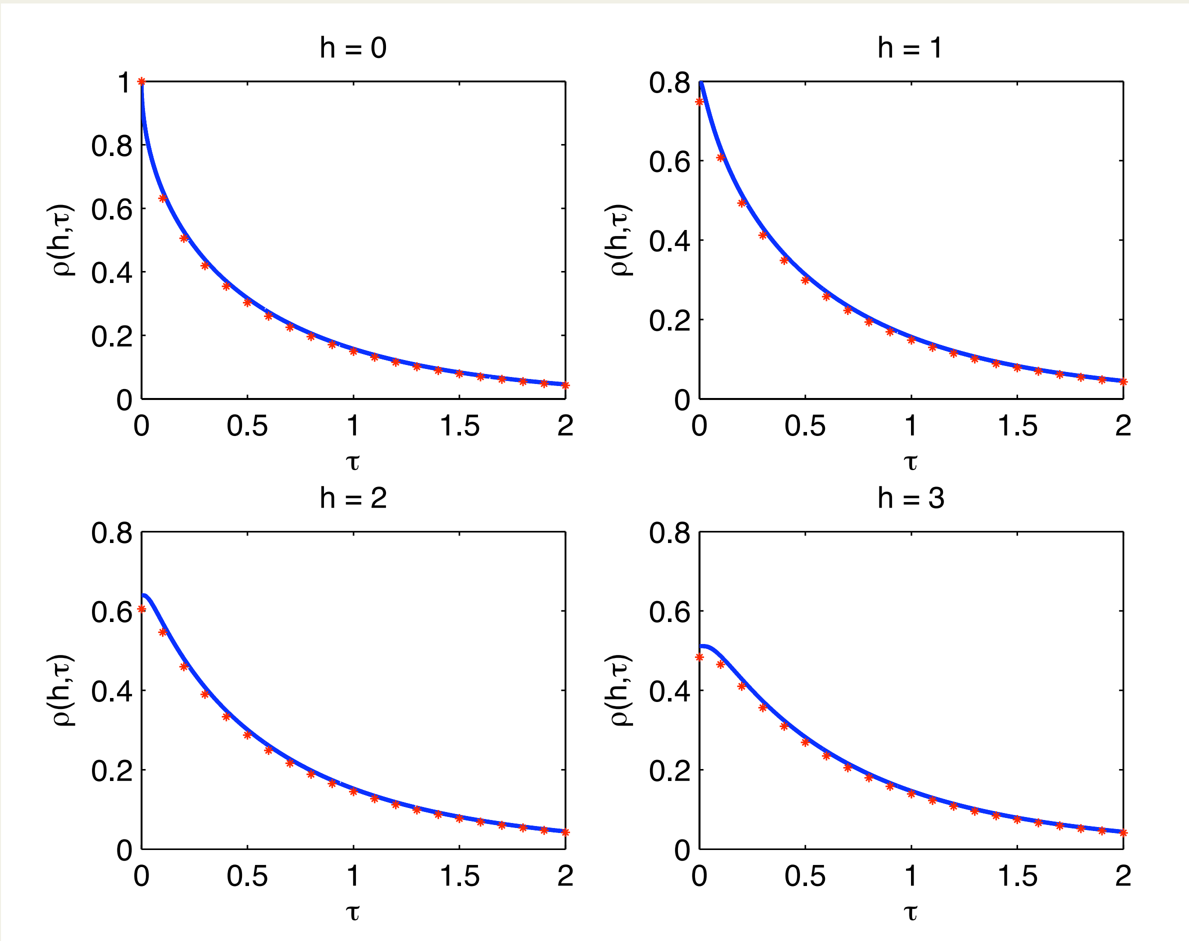
How does this marginal S-T covariance compare to the analytical one shown earlier for the continuous stochastic PDE?

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 10Δ_t = 0.1;
Blue lines: continuous correlation function

What about "real-world" complexity?

What if we don't know the exact form of the underlying process, or if the underlying system is more complicated (e.g., diffusion a spatial process)?

- Using a simpler discretized model as a template and allowing the parameters (e.g., θ) to be random, and (critically) structured in space (e.g., random fields) and/or time (e.g., time series) gives the model more flexibility to adapt to the data
  - This flexibility (through conditioning) is a strength of the hierarchical modeling approach

Basic Linear Gaussian Hierarchical Dynamical Spatio-Temporal Model

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

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

Parameters: $\mathbf{M}, \mathbf{R}, \mathbf{Q}$ — Critically, these can be structured according to the 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

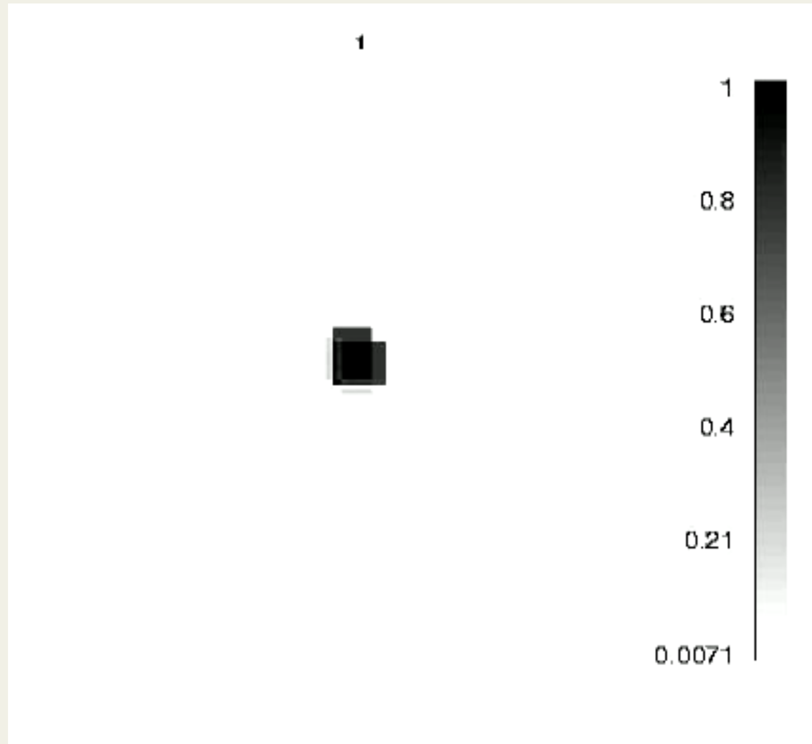
These come from our knowledge of the dynamics.

Linear Dynamic Parameterizations

Integro-difference equation (IDE)

$$Y_t(\mathbf{s}) = \gamma \int_D \underbrace{k_s(\mathbf{r};\boldsymbol{\theta}_s)}_{\text{transition kernel}} Y_{t-1}(\mathbf{r}) d\mathbf{r} + \eta_t(\mathbf{s})$$

Simulated advection-diffusion



Linear dynamics can easily (and quite efficiently) accommodate advective and diffusive processes

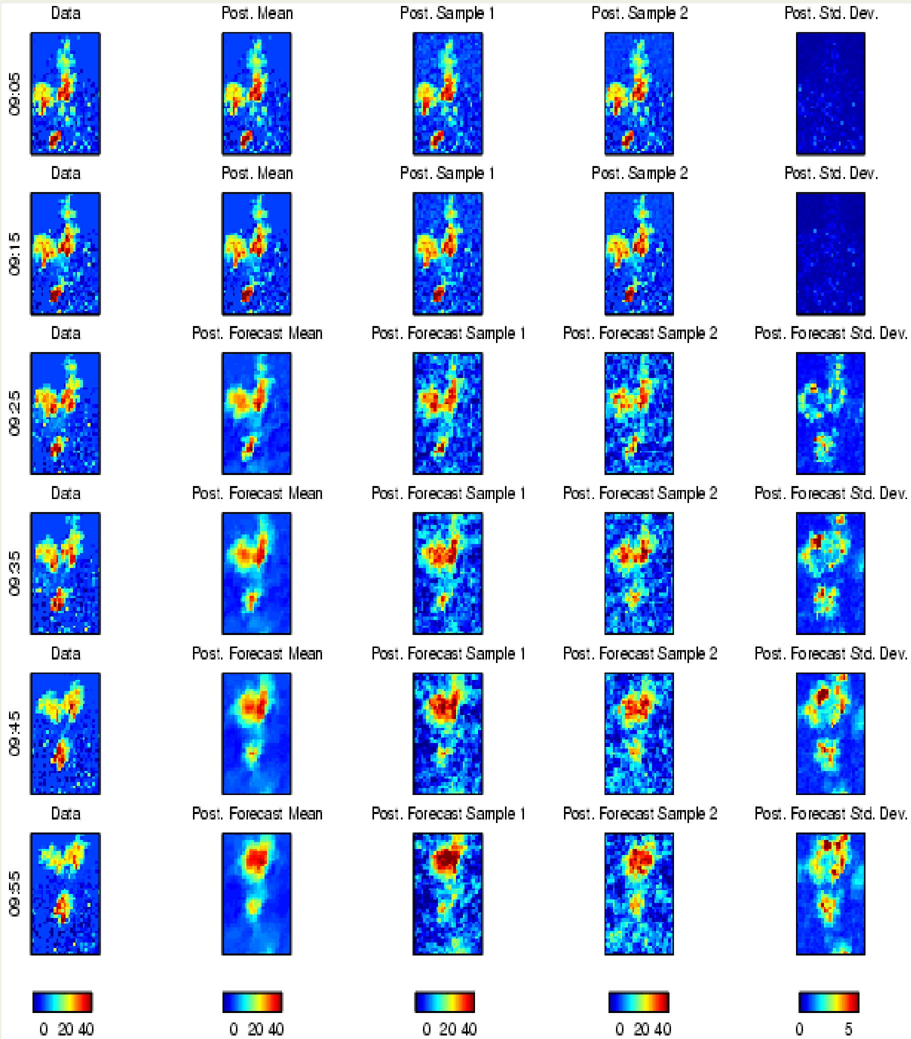
- "width" of the transition kernel controls rate of diffusion
- degree of "asymmetry" in the kernel controls speed of "object" propagation (advection)
- "long range dependence" can be accommodated by "multimodal" kernels

These suggest ways that we might parameterize the transition matrix!

Example: Radar Nowcasting

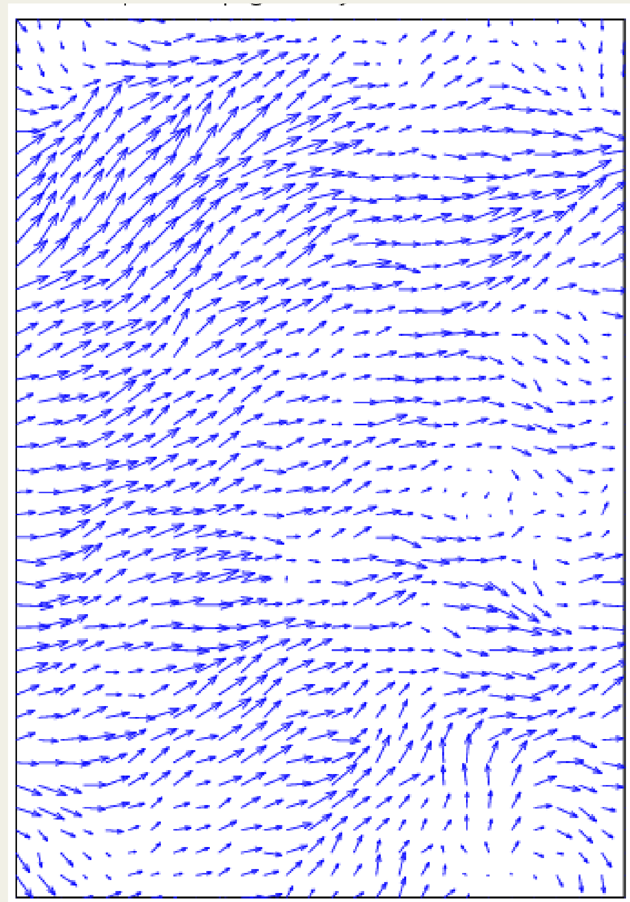
Data Mean Post. Real. Std.Dev.

Forecast



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

Implied Propagation by Post. Parms.



i.e., the advection parameters follow a latent spatial process

Xu, Wikle, and Fox, 2005

Nonlinear Parameterizations

Can the same ideas be used to suggest useful parametric forms for nonlinear dynamic spatio-temporal statistical models?

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

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 and we need to think of classes of parameterizations for this transition function.
- Time varying parameters can accommodate nonlinearity (e.g., regime changes)
- We consider a flexible class of parametric models based on polynomial interactions (recall, "interactions" are the key to spatio-temporal dynamics!)

* Note: nonlinearity here is with respect to the process or parameters; not just the parameters as is the traditional statistics definition

General Polynomial Nonlinear Stochastic IDE Model

Wikle and Holan, 2011

Continuous space and discrete time.

$$
\begin{aligned}
Y_t(\mathbf{s}) &= \int_D k_{\mathbf{s}}^{(1)}(\mathbf{u}_1) Y_{t-1}(\mathbf{u}_1) d\mathbf{u}_1 + \int_D \int_D k_{\mathbf{s}}^{(2)}(\mathbf{u}_1, \mathbf{u}_2) Y_{t-1}(\mathbf{u}_1) g_2(Y_{t-1}(\mathbf{u}_2)) d\mathbf{u}_1 d\mathbf{u}_2 \\
&+ \int_D \int_D \int_D k_{\mathbf{s}}^{(3)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) Y_{t-1}(\mathbf{u}_1) Y_{t-1}(\mathbf{u}_2) g_3(Y_{t-1}(\mathbf{u}_3)) d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 + \ldots \\
&+ \int_D \int_D \cdots \int_D k_{\mathbf{s}}^{(p)}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p) Y_{t-1}(\mathbf{u}_1) Y_{t-1}(\mathbf{u}_2) \cdots g_p(Y_{t-1}(\mathbf{u}_p)) d\mathbf{u}_1 d\mathbf{u}_2 \ldots d\mathbf{u}_p \\
&+ \eta_t(\mathbf{s})
\end{aligned}
$$

"General" because of $g_2()\ldots\ g_p(\ )$

for $\mathbf{s}, \mathbf{u}_1, \ldots, \mathbf{u}_p \in D, t \in \mathcal{T}$

where $\{k_{\mathbf{s}}^l(\mathbf{u}_1, \ldots, \mathbf{u}_l) : l = 1, \ldots, p\}$ are l-dimensional kernel functions and $g_l(\cdot) : l = 2, \ldots, p$ are transformation functions.

Discretizing and Truncating:

$$\begin{aligned} Y_t(\mathbf{s}_i) &= \sum_{j_1=1}^{n} k_{i,j_1}^{(1)} Y_{t-1}(\mathbf{s}_{j_1}) + \sum_{j_2=1}^{n}\sum_{j_1=1}^{n} k_{i,j_1 j_2}^{(2)} Y_{t-1}(\mathbf{s}_{j_2}) g_2(Y_{t-1}(\mathbf{s}_{j_1})) \\ &+ \sum_{j_3=1}^{n}\sum_{j_2=1}^{n}\sum_{j_1=1}^{n} k_{i,j_1 j_2 j_3}^{(3)} Y_{t-1}(\mathbf{s}_{j_3}) Y_{t-1}(\mathbf{s}_{j_2}) g_3(Y_{t-1}(\mathbf{s}_{j_1})) + \ldots \\ &+ \sum_{j_p=1}^{n} \cdots \sum_{j_2=1}^{n}\sum_{j_1=1}^{n} k_{i,j_1 j_2 \cdots j_p}^{(p)} Y_{t-1}(\mathbf{s}_{j_p}) \cdots Y_{t-1}(\mathbf{s}_{j_2}) g_p(Y_{t-1}(\mathbf{s}_{j_1})) \\ &+ \eta_t(\mathbf{s}_i), \end{aligned}$$

Alternatively, we can consider a basis function expansions of the kernels and process, which leads to a polynomial interaction model on functional coefficients (e.g., see Wikle and Holan, 2011).

Polynomial to Quadratic

- This polynomial interaction model is quite powerful.
- However, there are on the order of n^{p+1} parameters!
- Fortunately, a general <u>quadratic</u> model can accommodate a very large class of real-world phenomena.

General Quadratic Nonlinearity (GQN)

(Wikle and Hooten, 2010)

In scalar form,

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

for i=1,…,n.

- Model includes quadratic (dyadic) interactions in random process Y
- The term "general" refers to the term: $g(Y_{t-1}(s_l); \boldsymbol{\theta}_g)$
- There are $O(n^3)$ parameters in this model!
- This can be recast as a matrix equation: parameters in **A**, **B**, **$\boldsymbol{\theta}_g$**

e.g.,

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{A}\mathbf{Y}_{t-1} + (\mathbf{I}_n \otimes \mathbf{Y}'_{t-1})\mathbf{B}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \\ &= (\mathbf{A} + (\mathbf{I}_n \otimes \mathbf{Y}'_{t-1})\mathbf{B})\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t \end{aligned}$$

Examples of Quadratic Nonlinearity

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

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

Wind-Driven Ocean Circulation: (Quasigeostrophy; Ψ is the streamfunction)

$$(\nabla^2 - \frac{1}{r^2})\frac{\partial \psi}{\partial t} = -J(\psi, \nabla^2\psi) - \beta\frac{\partial \psi}{\partial x} + \frac{1}{\rho H}\,\mathrm{curl}_z\tau - \gamma\nabla^2\psi - a_H\nabla^4\psi$$

where $J(a,b) = \partial a/\partial x \partial b/\partial y - \partial b/\partial x \partial a/\partial y$ is the Jacobian (**nonlinear** in ψ), τ the wind-stress, and r, β, ρ, H, γ, a_H are parameters.

Epidemic Dynamics (SIR; Susceptible, Infected, Recovered)

$$\frac{\partial S}{\partial t} = \nu - \beta SI - \mu S + \omega R + D_S\nabla^2 S$$
$$\frac{\partial I}{\partial t} = \beta SI - \mu I - \gamma I + D_I\nabla^2 I$$
$$\frac{\partial R}{\partial t} = \gamma I - \mu R - \omega R + D_R\nabla^2 R$$

GQN Parameterization/Implementation

- Note: the process can consist of multiple component processes: e.g., $\mathbf{Y}_t = (\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t}, \mathbf{Y}'_{3t})'$
- A substantial problem with these models is that they have **too many parameters** to estimate reliably without extra information; to help, we can use:
  - Mechanistically-motivated parameterizations
  - Stochastic search variable selection (SSVS)
  - Rank reduced "spectral" models
  - Emulator-based priors

Example: Invasive Species Prediction

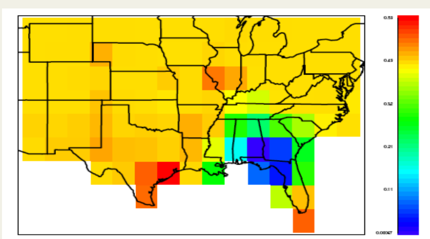
(e.g., Wikle 2003; Hooten and Wikle, 2007; Hooten et al. 2007)

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

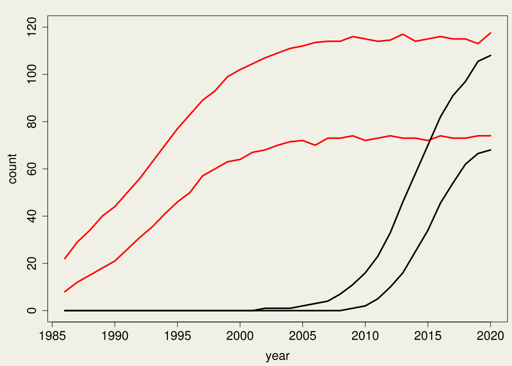
$$\frac{\partial Y}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial Y}{\partial y}\right) + \gamma_0(x,y)Y\exp\left(1 - \frac{Y}{\gamma_1(x,y)}\right)$$

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.

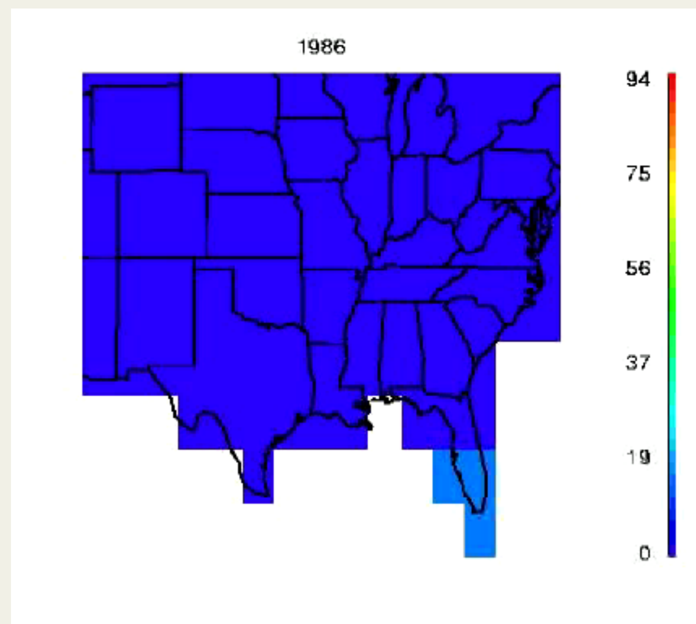
Eurasian Collared Dove Invasion



Dispersal δ: post mean



95% PI's for S. Fl and Utah



Relative abundance: post mean

GQN: Rank Reduction

We can consider the essential dynamics on a manifold of lower rank. That is, we will consider the dynamical process after projection into a lower dimensional space.

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 we have order p^3 parameters here! Unless p is very small, we may need to make simplifying assumptions, perform stochastic search model selection, or develop mechanistic-based priors in order to do estimation.

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:

$$\mathbf{Y}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \;\; \boldsymbol{\xi}_t \sim Gau(\mathbf{0}, \mathbf{R})$$

$$\boldsymbol{\alpha}_t = \mathbf{A}\boldsymbol{\alpha}_{t-1} + (\mathbf{I}_p \otimes \boldsymbol{\alpha}'_{t-1})\mathbf{B}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \mathbf{Q})$$

$$\mathbf{R} = \kappa\mathbf{I} + \sum_{k=p+1}^{p+n_p} \lambda_k \phi_k \phi'_k$$ where $$\kappa^{-1} \sim Gamma(q_\kappa, r_\kappa)$$

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

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

Interaction of small-scale modes parameterizes cov matrix

$$[\mathbf{B}]$$ (see below; hierarchical stochastic search variable selection)

Hierarchical Stochastic Search Variable Selection

(George and McCulloch, 1993; 1997; Wikle and Holan 2011)

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,

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

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

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

(note, prior knowledge can also be placed on π_j)

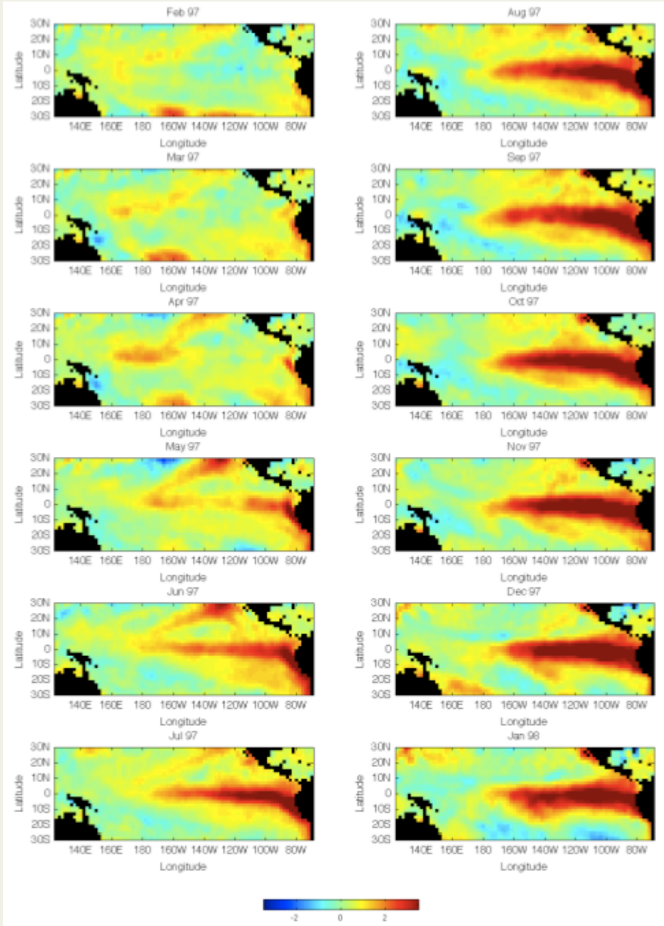
where $\gamma_j = 1$ means that the j-th variable is in the model.

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.

Note: we can do the same thing for the linear transition matrix, **A**.

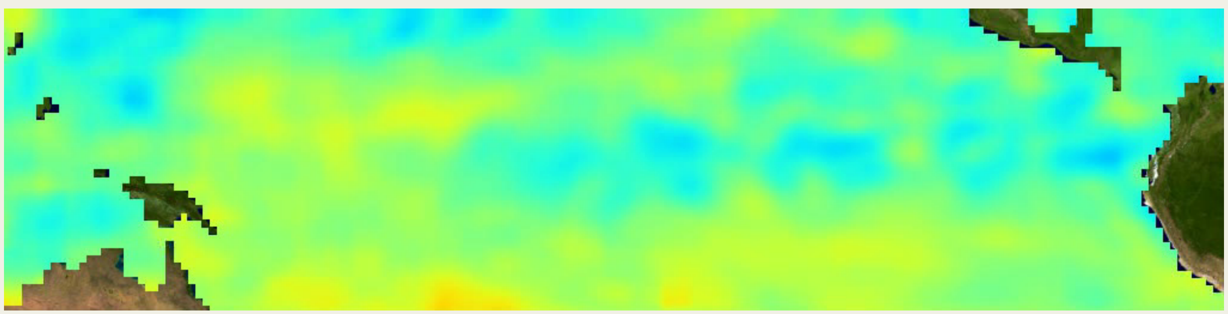
Example: Long-Lead Prediction of Tropical Pacific SST

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



Let

$$\mathbf{\Phi} - \text{ EOFs}$$

$$p = 10, \; n_p = 10$$

Standard MCMC implementation; vague priors on all parameters except data model variance.

SST: Quadratic Nonlinear Hierarchical Model Implementation

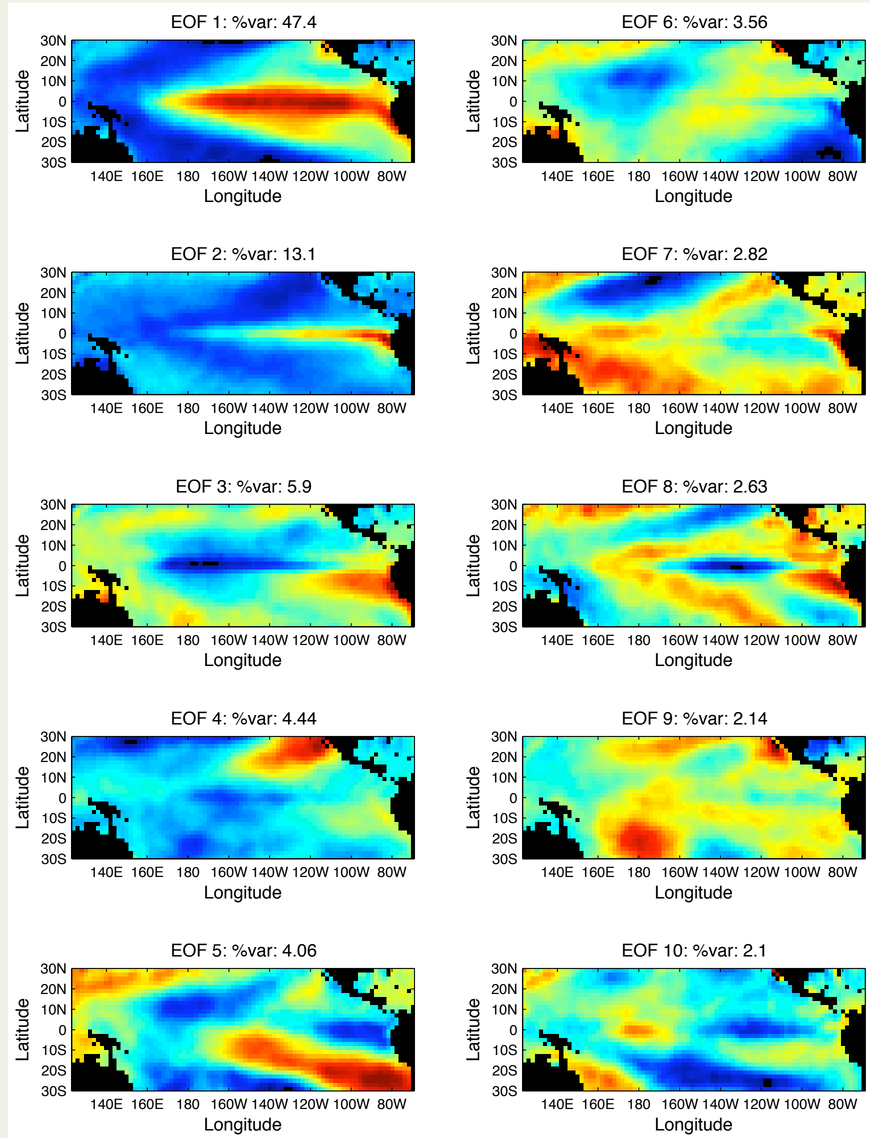
$\boldsymbol{\Phi}^{(1)}$ - EOFs

$p_1 = 10$

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.

First 10 EOF Patterns



Forecast: October 1997 from March 1997

Nonlinear Model

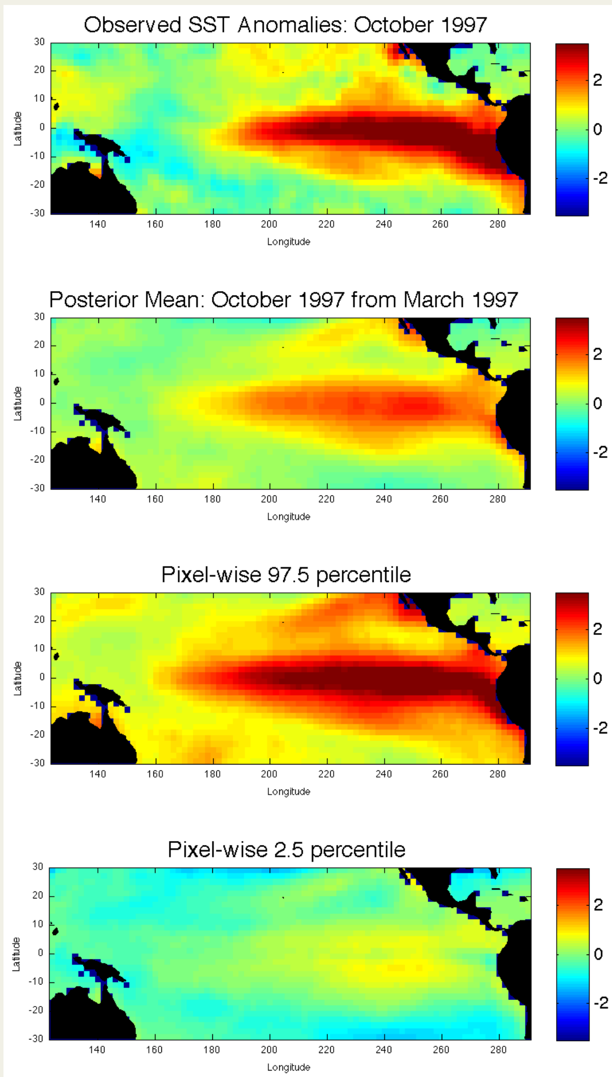
Linear (VAR) Model

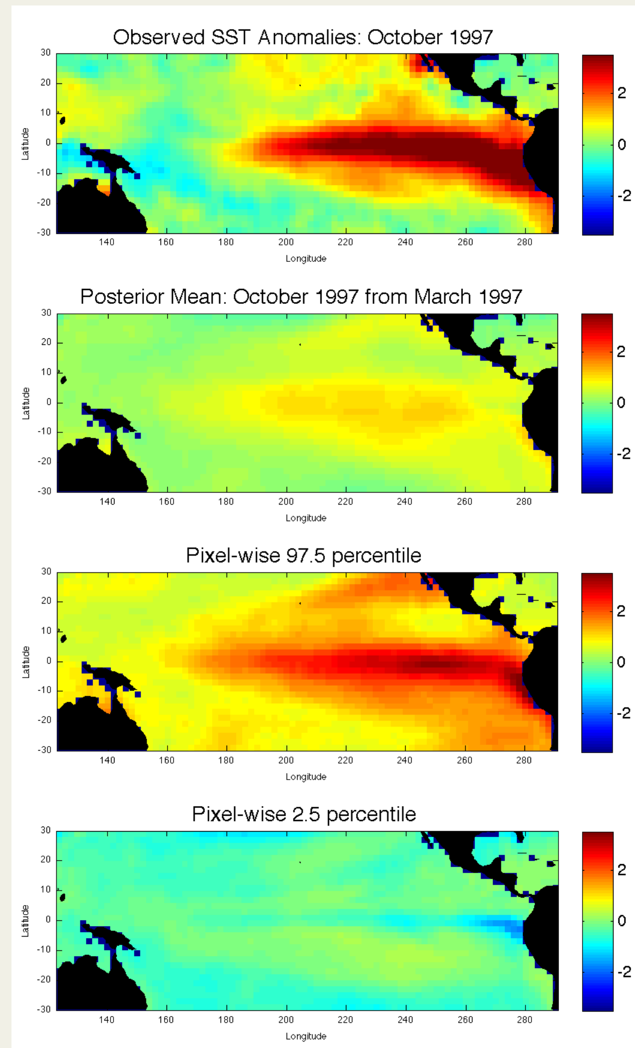
Obs

Post. Mean

Post. Pixel 97.5%-tile

Post. Pixel 2.5%-tile

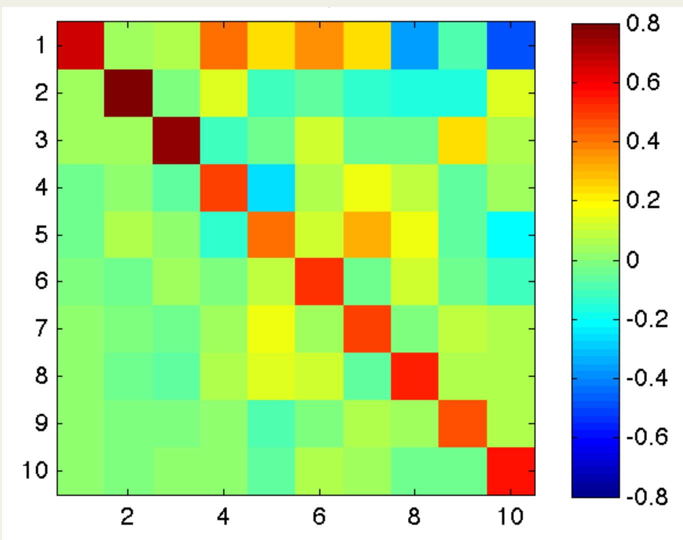




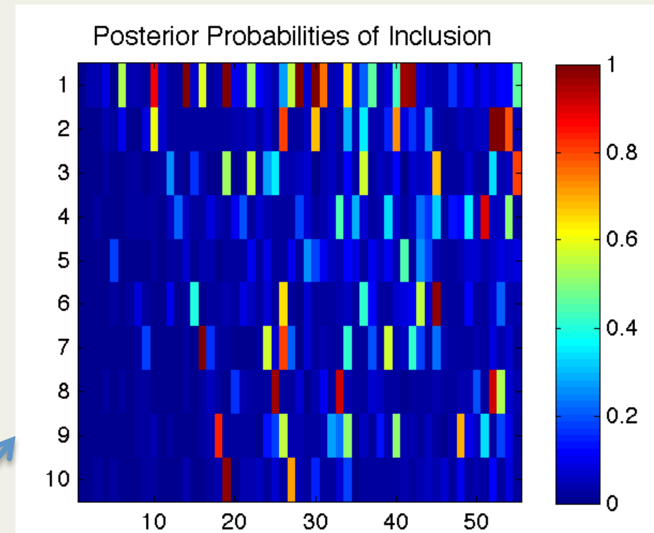
Posterior Means: Parameters

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

Posterior Mean: A matrix (linear term)

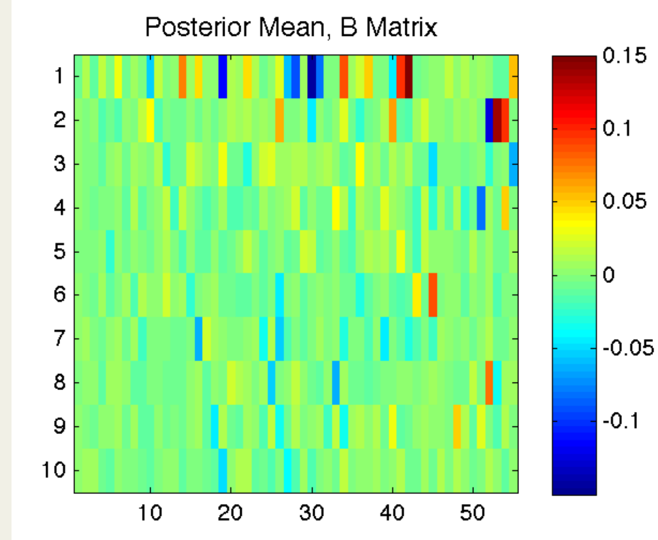


B matrix inclusion probabilities



Interpretation?

B matrix



Informative Priors

If information is available from other sources, we may be able to combine information to improve estimation.

One approach for data assimilation applications is to develop a surrogate parametric model (or emulator) for a mechanistic simulator and use that to develop informative prior distributions.

Statistical Emulators (Surrogates)

- With very complex nonlinear multivariate S-T processes for which there exists large (typically, mechanistic) simulators ("computer models"), one can use statistical models to "emulate" the simulator. [e.g., Sacks et al., 1989; Kennedy and O'Hagan, 2001; Higdon et al. 2004,2008; and MANY more!!]
- In statistics, the tradition has been to use "second-order emulators", based on Gaussian processes with the emphasis on covariance to model a response surface (e.g., Kennedy and O'Hagan, 2001).
- As an alternative, one can consider modeling the dependence through a first-order linear or nonlinear model (e.g., van der Merwe et al. 2007; Hooten et al. 2011; Margvelashvili and Campbell, 2012), which is more suited to dynamic processes.

Construction of a Reduced-Rank First-Order Emulator

- Generate inputs: $\mathbf{W}_{q \times K} = (\mathbf{w}_1, \ldots, \mathbf{w}_K)$
- Generate vector realizations from computer model output given the inputs, $y_i = f(w_i)$: $\mathbf{Y}_{n_y \times K} \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_K)$
- Consider the SVD of the computer model output:

$$\mathbf{Y} = \mathbf{UDV}'$$

- Approximate the SVD by keeping only the first p left and right singular vectors: $\mathbf{Y} \approx \tilde{\mathbf{U}}_{n_y \times p} \tilde{\mathbf{D}}_{p \times p} \tilde{\mathbf{V}}'_{p \times K}$
- Model the right singular vectors as a nonlinear function of inputs: $\tilde{\mathbf{V}} \sim g(\mathbf{W}, \boldsymbol{\theta})$ (select favorite nonlinear model)
- Thus, for an n_y-dim response, **y**, and input, **w** :

$$\mathbf{y} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\mathbf{v}(\mathbf{w}, \boldsymbol{\theta}) + \boldsymbol{\eta} \equiv \mathbf{F}\mathbf{v}(\mathbf{w}, \boldsymbol{\theta}) + \boldsymbol{\eta} \equiv m(\mathbf{w}, \boldsymbol{\theta}) + \boldsymbol{\eta}$$

First-Order Emulator (process)

- Typically **w** corresponds to parameters in the mechanistic model.
- It is also the case that **w** may correspond to forcings (e.g., climate drivers), initial conditions, or the previous values of the state process (1-step ahead emulator): e.g.,

$$\mathbf{y}_t = m(\mathbf{w}_t, \boldsymbol{\theta}) + \boldsymbol{\eta}_t \equiv m(\boldsymbol{y}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{\eta}_t$$

- When "trained" on the mechanistic model output **Y**, we get:

$$\mathbf{y}_t = m(\mathbf{y}_{t-1}, \hat{\boldsymbol{\theta}}) + \boldsymbol{\eta}_t$$

Thus, parameters of m( ) are found "off-line".

- **This can be the basis for a prior model on the dynamics within a hierarchical nonlinear DSTM.**

Hierarchical Reduced-Rank Emulator-Assisted DSTM

(Leeds, Wikle, Fiechter, 2012)

$$\mathbf{Z}_t = \mathbf{H}_t \boldsymbol{\Phi} \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\Psi} \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim Gau(\mathbf{0}, \mathbf{R}_t)$$

$$\boldsymbol{\alpha}_t = m(\boldsymbol{\alpha}_{t-1}; \boldsymbol{\theta}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \mathbf{Q})$$

quadratic nonlinear model

$$\boldsymbol{\beta}_t \sim Gau(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}))$$

$$[\mathbf{R}_t, \mathbf{Q}, \boldsymbol{\tau}]$$

$$\boldsymbol{\theta} \sim (\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_\theta)$$

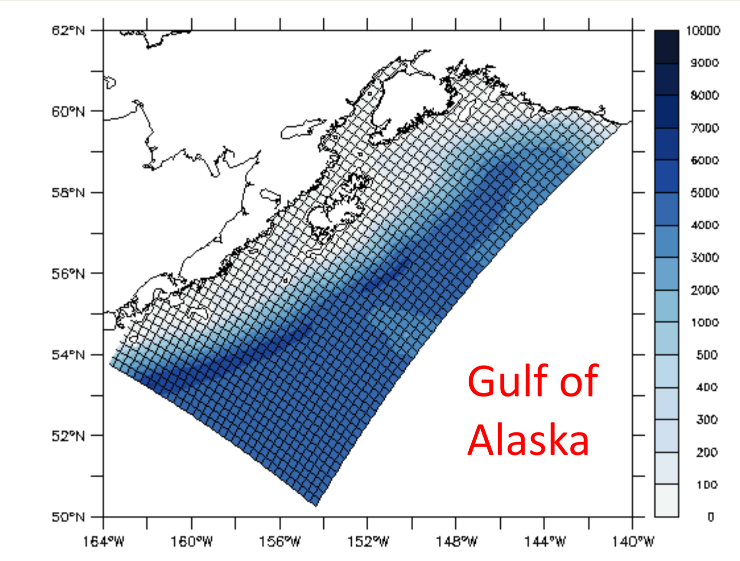
Note: $\mathbf{Y}_t = \boldsymbol{\Phi} \boldsymbol{\alpha}_t + \boldsymbol{\Psi} \boldsymbol{\beta}_t$

KEY POINT: The mean is from a parametric statistical emulator; estimated "off-line"

- $\mathbf{H}_t$ is an $m_t \times n$ mapping matrix.
- $\boldsymbol{\Phi}$ is an $n \times p$ matrix of basis functions obtained from the first p left-singular vectors from the SVD of the mechanistic model output
- $\boldsymbol{\Psi}$ is an $n \times q$ matrix of basis functions obtained from the remaining q left-singular vectors from the SVD of the mechanistic model output
- * $m(\cdot)$ is the multivariate DSTM propagator, depending on parameters $\boldsymbol{\theta}$
- $\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t$ are analogous to the first p and remaining q right singular vectors of the mechanistic model output
- Choices made for $(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_\theta), [\mathbf{R}_t, \mathbf{Q}, \boldsymbol{\tau}]$ depend on application

EXAMPLE: Spatio-temporal prediction of primary production (chlorophyll) in the Coastal Gulf of Alaska (GOGA)

- **Data Assimilation:** Combine primary production data and mechanistic computer model for a coupled ocean and ecosystem model (ROMS-NPZDFe; Fiechter et al. 2009)
- **Emulator:** quadratic nonlinear emulator for coupled model: Phytoplankton, SSH (sea surface height), and SST (sea surface temperature) model output
- **Predict/Assimilate:** Primary Production given high-dimensional ocean color (SeaWiFS) satellite data and ocean model physical output



•Train based on 4 years (1998-2001), 8 day averages

•Predict/Assimilate for 2002

Proof of Concept Experiment

- In this case:

$$\mathbf{Z}_t = \begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \mathbf{Z}_{3,t} \end{pmatrix}$$ $m_{i,t} (i = 1, 2, 3)$ - dimensional data vectors for Chlorophyll, SSH, SST

$$\mathbf{Y}_t = \begin{pmatrix} \mathbf{Y}_{1,t} \\ \mathbf{Y}_{2,t} \\ \mathbf{Y}_{3,t} \end{pmatrix}$$ $p_i (i = 1, 2, 3)$ - dimensional reduced rank process vectors for Chlorophyll, SSH, SST

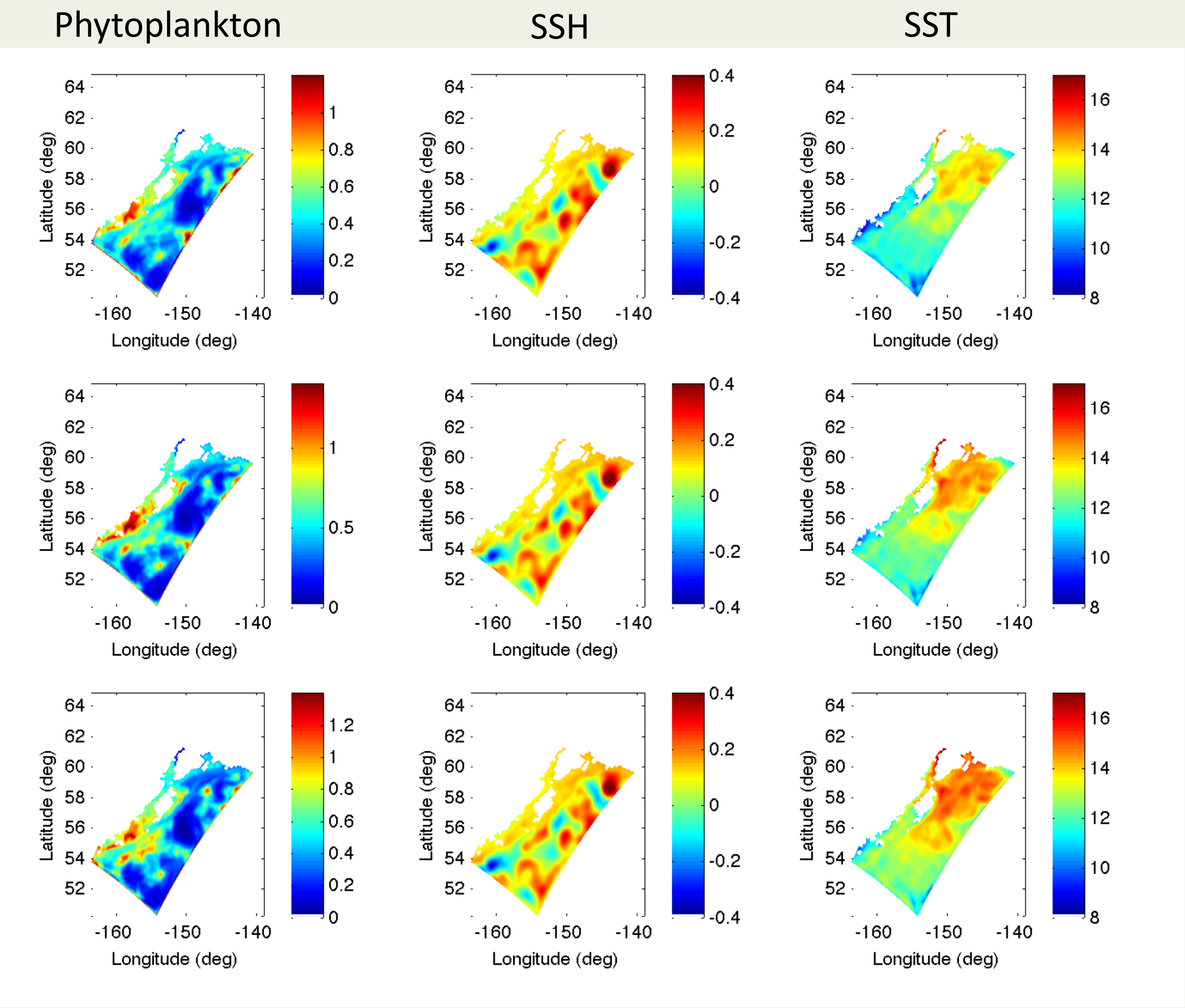
(Work in log space for Chlorophyll)

- State Rank Reduction: $O(10^5)$ to $O(10)$
- Nonlinear emulator: a quadratic nonlinear model based on the first 7 singular vectors (97.5% of the variation) of the ROMS-NPZDFe output SVD for 1998-2001
  - the non-dynamic small-scale components were based on the next 10 singular vectors (over 99% of variation in model output total)

Coupled Dynamics: Example from Coupled Ocean Model

Example training data

Time: three consecutive 8-day periods

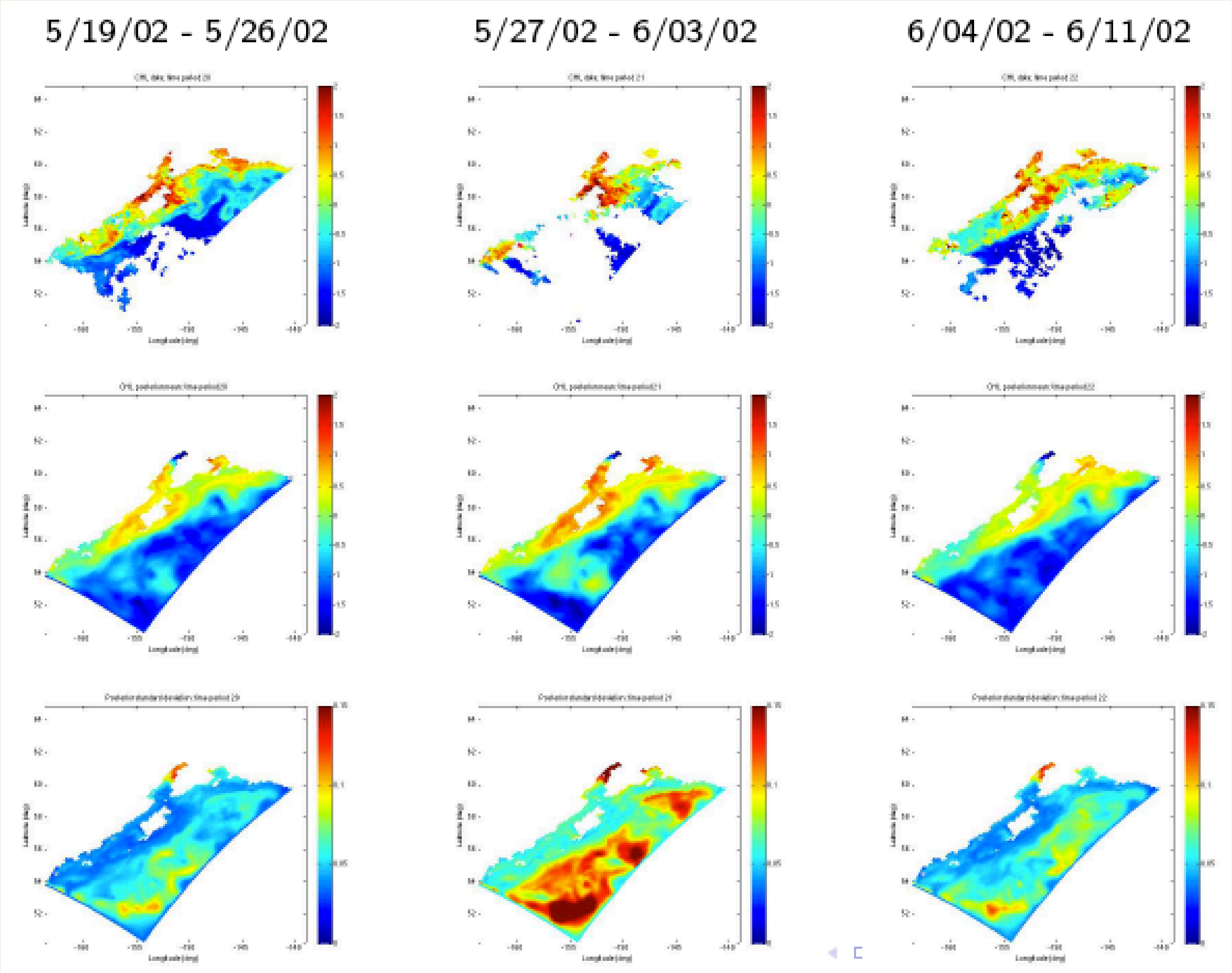


Results: log(CHL)

Data

Posterior Mean

Posterior STD



Results (cont.)

More prominent chlorophyll eddie in posterior mean (middle) than ROMS-NPZDFe output (right):

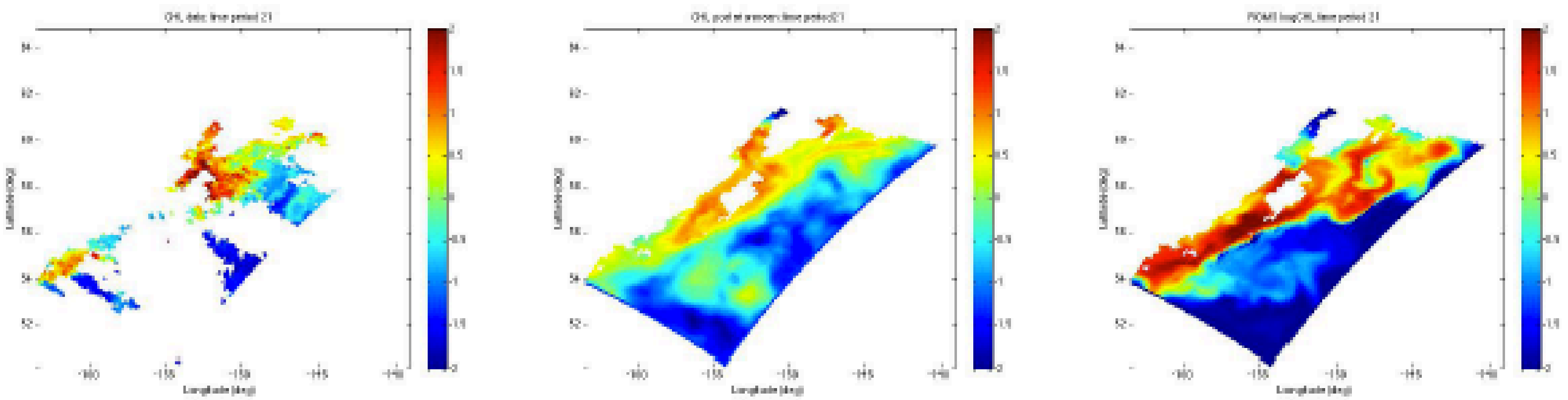


Figure: 8 day composite of SeaWiFS observations (left), ROMS-NPZDFe P output (right) and the posterior mean for phytoplankton (center) from May 27, 2002 to June 3, 2002.

Conclusion

- Environmental processes involve interactions across space and time and multiple variables
- These interactions are governed by nonlinear processes
- Statistical models for nonlinear dynamic spatio-temporal processes typically can benefit from incorporation of scientific information, while considering the dimensionality of the state variables and parameters
- Helpful approaches:
  - Generalized quadratic nonlinearity
    - Mechanistic motivation; rank reduction; stochastic search variable selection; emulator-assisted models
- Much work to be done in nonlinear S-T model development, computation and theory!

Collaborators and Sources

- Bill Leeds (University of Chicago)
- Mevin Hooten (Colorado State University)
- Scott Holan (University of Missouri)

Primary Sources:

• Wikle, C.K. and Hooten, M.B. (2010) A general science-based framework for dynamical spatio-temporal models. *Test* (discussion paper), 19, 417-451.
• Cressie, N. and Wikle, C.K. (2011) *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Chaps 6, 7.
• Wikle, C.K. and S. H. Holan (2011) Polynomial nonlinear spatio-temporal integro-difference equation models. *Journal of Time Series Analysis*, 32, 339-350.
• Leeds, W.B., Wikle, C.K., and J. Fiechter (2012) Emulator-assisted reduced-rank ecological data assimilation for nonlinear multivariate dynamical Spatio-temporal processes. (In review)