An Introduction to Dynamical Spatio-Temporal Models (DSTMs)

Christopher K. Wikle

Department of Statistics
University of Missouri

9 July 2015

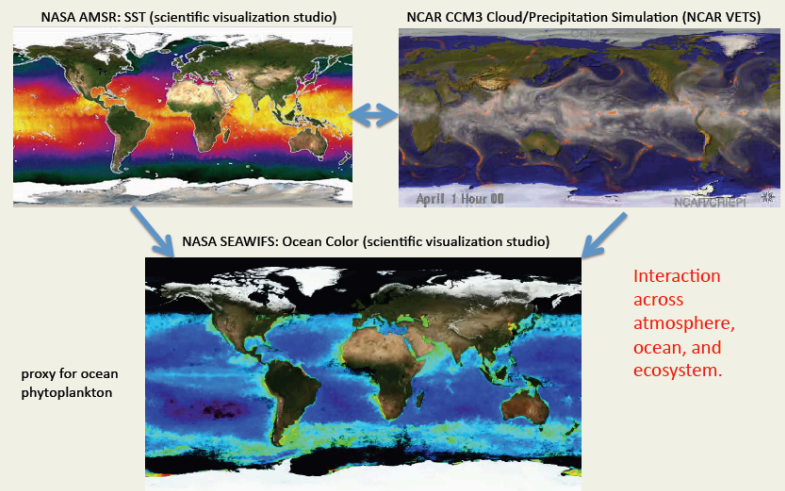
Spatio-Temporal Processes and Data

There is no history without geography (and *vice versa*)! We consider space and time together.

Spatio-temporal data are not new. Consider the digitally restored Lienzo de Quauhquechollan from the indigenous people of Guatemala who documented the spatio-temporal history of the Spanish conquest from 1527 to 1530.

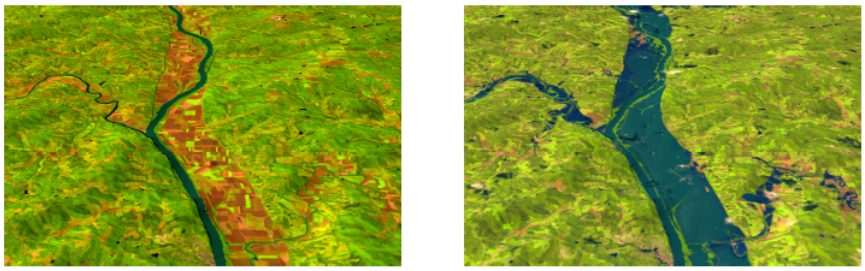
Spatio-Temporal Processes and Data

Data from spatio-temporal processes are common in the real world, representing a variety of interactions across processes and scales of variability. The dynamical evolution (time dimension) of spatial processes means that we are able to reach more forecefully for the "Why" question.



Spatio-Temporal Processes and Data (cont.)

Although it may be informative to see snapshots of spatial events in time (see the Missouri River scene below), to understand the process, we must know something about the behavior from one time-period to the next.



Images from NASA's Landsat Thematic Mapper. Each image shows a segment of the Missouri River near Hermann, MO (mile 96.5, at the bottom of the scene), and Gasconade, MO (mile 104.4, in the "V" in the middle of the scene). The river flows from west (top of the scene) to east (bottom of the scene). Left panel: September 1992, before a major flood event. Right panel: September 1993, after a record-breaking flood event in July 1993.

Learning Points

- Marginal vs. Dynamical Perspective
  - Covariance-based models are limited in their ability to model real-world complexity
  - There are no "magical" covariance functions
  - There are no "magical" PDEs (all models are wrong!)
  - PDEs (and other mechanistic models) **can** provide **motivating structure** for parameterizing spatio-temporal dynamical models in statistics
    - ⋆ We are NOT interested in solving the PDE, just using it to motivate statistical models!
- The essence of dynamical modeling is hierarchical/conditional thinking
  - The primary benefit of hierarchical modeling is through the conditional mean
  - Real-world complexity comes through modeling "parameters as processes"

Learning Points (cont.)

- The real-world is nonlinear!
  - The essence of nonlinear dynamical modeling is understanding interactions
  - Most non-Gaussianity (and extremes) come from nonlinearity
  - Nonlinearity can often be modeled through latent *conditional* Gaussian processes
  - Parameter and/or state reduction is critical for nonlinear dynamical spatio-temporal models
- Be wary of methods that require software that you can't program yourself!

Brief Outline

- Introduction to Spatio-Temporal Processes
  - Descriptive vs. Dynamic Approach
  - Stationarity, separability, and optimal spatio-temporal prediction
- Brief Introduction to Hierarchical Modeling
- Linear DSTMs
  - Mechanistically-motivated models (physical space, spectral space)
- Nonlinear DSTMs
  - General quadratic nonlinearity
  - Mechanistic motivation
  - Importance of basis functions
  - Parameter shrinkage and emulator-assisted prior elicitation
- Comments on Computation
- Environmental and Ecological Examples Throughout

Most of the references and material in this talk can be found in *Statistics for Spatio-Temporal Data* by N. Cressie and C.K. Wikle (2011, John Wiley and Sons).

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 is the temporal domain of interest, $\mathbf{s}$ is a spatial index and t a time index; when we refer to discrete time, we will typically write $Y_t(\mathbf{s})$ (i.e., a subscript t)

It has become customary in hierarchical modeling to denote "probability distributions" using square-bracket notation. For example,

- $[Z]$ - a continuous or discrete distribution for the random variable Z
- $[Z, Y]$ - the joint distribution of the random variables Z and Y
- $[Z|Y]$ - the conditional distribution of Z given Y

We also typically denote vectors and matrices using a bold font: $\mathbf{Y}$, $\boldsymbol{\beta}$; we use a "prime" to represent a vector or matrix transpose: $\mathbf{Y}'$

Spatio-Temporal Statistical Modeling

From a statistician's perspective, what makes a model "statistical"?

- Uncertainty in data, model, and the associated parameters
- Estimation of parameters and prediction of processes

We also often make a distinction between "stochastic" and "statistical"

- The former concerns random structures in models
- The latter concerns estimation and prediction given data

Spatio-Temporal Processes and Data (cont.)

Why spatio-temporal modeling? Characterize processes in the presence of uncertain and (often) incomplete observations and system knowledge, for the purposes of:

- Prediction in space (smoothing, interpolation)
- Prediction in time (forecasting)
- Assimilation of observations with deterministic models
- Inference on parameters that explain the etiology of the spatio-temporal process

Traditionally, there are two approaches to modeling such processes: descriptive and dynamical.

Spatio-Temporal Processes and Data (cont.)

Spatio-Temporal Modeling

Descriptive (marginal) approach: Characterize the first- and second-moment (covariance) behavior of the process

- Several different mechanistic processes could imply the same marginal structure
- Cannot be used for highly complex processes (e.g., nonlinear)
- Most useful when knowledge of the etiology of the process is limited

Spatio-Temporal Processes and Data (cont.)

Spatio-Temporal Modeling

Dynamical (conditional) approach: Current values of the process at a location evolve from past values of the process at various locations

- Conditional models are closer to the etiology of the phenomenon under study
- Most useful if there is some *a priori* knowledge available concerning the process' behavior

Note that, in some simple cases, the descriptive approach and the dynamical approach can be related through their respective covariance functions.

Differential Equations as Motivation for Statistical Models (Temporal)

- Yule (1927) used a 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)*

PDEs as Motivation for Spatial Statistical Models

- In spatial statistics, Heine (1955) used stochastic partial differential equations (PDEs) to develop general classes of spatial covariance functions.
- Most famous example: the "magical" Matèrn class of spatial covariance functions can be derived from a linear stochastic diffusion PDE (e.g., Whittle 1962; see Guttorp and Gneiting (2006) for historical perspective):

$$Y(\mathbf{s}) : \mathbf{s} \in D_s \subset \mathbb{R}^2,$$

$\delta(\mathbf{s})$ is a white noise process, and

$$\left\{ \frac{\partial^2}{\partial s_1^2} + \ldots + \frac{\partial^2}{\partial s_d^2} - \alpha^2 \right\}^p Y(\mathbf{s}) = \delta(\mathbf{s})$$

for $\theta_1 = 1/\alpha > 0$, $\theta_2 = 2p - d/2 > 0$ (i.e., restrict $p > d/4$), the stationary solution of which gives the spatial covariance function:

$$C_Y(\mathbf{h}) \propto \{||\mathbf{h}||/\theta_1\}^{\theta_2} K_{\theta_2}(||\mathbf{h}||/\theta_1).$$

A Simple Example: Motivating a Spatio-Temporal Covariance Function

Consider the deterministic 1-D space $\times$ time, reaction-diffusion equation:

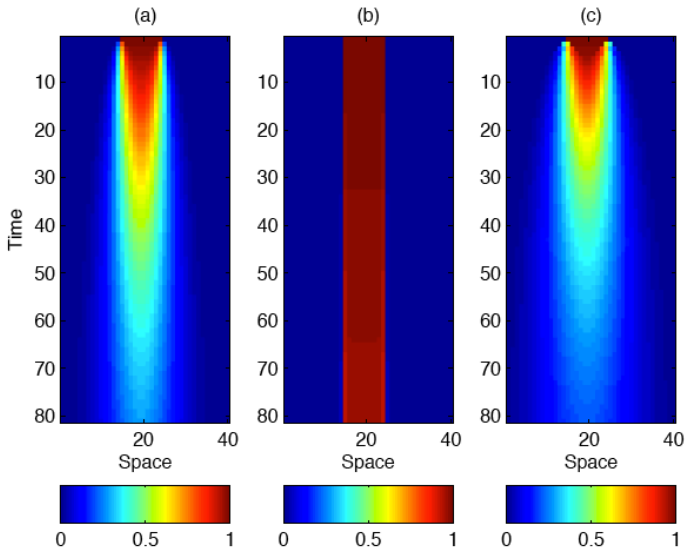
$$\frac{\partial Y(s;t)}{\partial t} = \beta \frac{\partial^2 Y(s;t)}{\partial s^2} - \alpha Y(s;t)\,,$$

for $\{s \in \mathbb{R}, t \geq 0\}$, where β is the diffusion coefficient and α is the "reaction" coefficient.

Meaning of the Equation: The rate of change in Y is equal to the "spread" of Y in space (i.e., diffusion) offset by the "loss" of a certain multiple of Y (i.e., reaction).

Behavior of the Equation: From a given initial condition $Y(s;0)$, the process $Y(s;t)$ dampens as time t increases.

A Simple Example (cont.)



$$Y(s;0) = I(15 \leq s \leq 24)$$

(a) $\alpha = 1$, $\beta = 20$; (b) $\alpha = 0.05$, $\beta = 0.05$; (c) $\alpha = 1$, $\beta = 50$

A Simple Example: Stochastic Version

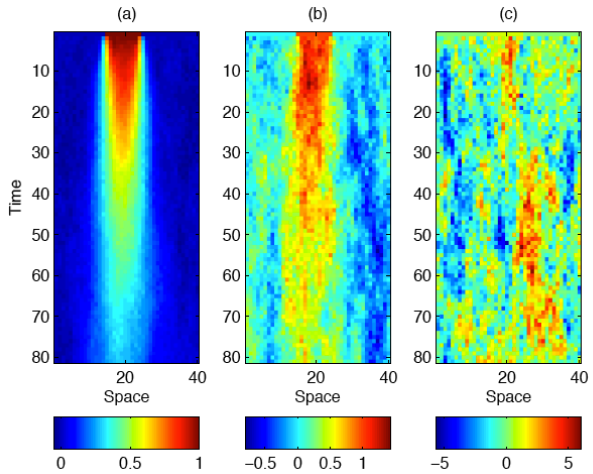
Consider the stochastic version of this PDE:

$$\frac{\partial Y}{\partial t} - \beta \frac{\partial^2 Y}{\partial s^2} + \alpha Y = \eta \,,$$

where $\{\eta(s;t)\colon s \in \mathbb{R}, t \geq 0\}$ is a mean-zero, white-noise process with variance σ^2.

In this case, a statistical balance is reached between the "disturbance" caused by $\eta(\cdot;\cdot)$ and the smoothing effect of the diffusion and loss components. That is, from a given initial condition, the stochastic PDE results in a process that eventually achieves both spatial and temporal stationarity. (The more general case of stochastic PDEs in $\mathbb{R}^d$ is given, e.g., by Brown et al., 2000.)

Stochastic Reaction-Diffusion Simulation Plots



$$Y(s;0) = I(15 \leq s \leq 24)$$

$$\alpha = 1\,, \beta = 20$$

(a) $\sigma = 0.01$; (b) $\sigma = 0.1$; (c) $\sigma = 1$

Spatio-Temporal Covariance Function

- The stochastic reaction-diffusion equation implies a (stationary in space and time) covariance function:

$$C_Y(h;\tau) \equiv \mathrm{cov}(Y(s;t), Y(s+h;t+\tau))$$

and correlation function:

$$\rho_Y(h;\tau) \equiv C_Y(h;\tau)/C_Y(0;0)$$

Heine (1955; *Biometrika*) gives a closed-form solution for $\rho_Y(\cdot;\cdot)$ for spatial lag $h \in \mathbb{R}$ and temporal lag $\tau \in \mathbb{R}$:

$$\rho_Y(h;\tau) = (1/2)\left\{ e^{-h(\alpha/\beta)^{1/2}} Erfc\left(\frac{2\tau(\alpha/\beta)^{1/2} - h/\beta}{2(\tau/\beta)^{1/2}}\right) + e^{h(\alpha/\beta)^{1/2}} Erfc\left(\frac{2\tau(\alpha/\beta)^{1/2} + h/\beta}{2(\tau/\beta)^{1/2}}\right)\right\},$$

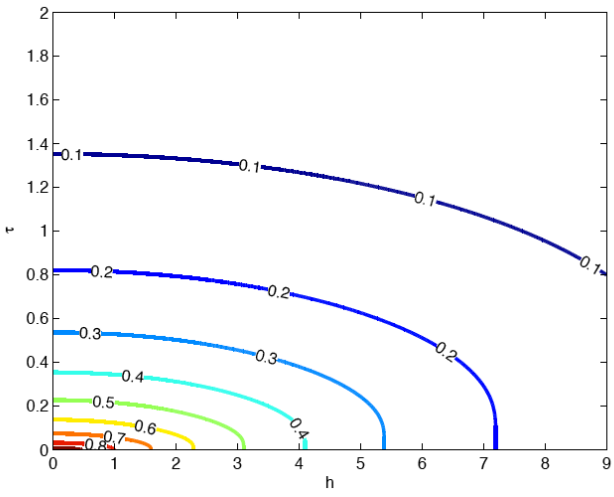
where $Erfc(z)$ is the "complementary error function":

$$Erfc(z) \equiv (2/\pi^{1/2}) \int_z^\infty e^{-v^2}\, dv, \quad z \geq 0;$$

and

$$Erfc(z) = 2 - Erfc(-z), \quad z < 0.$$

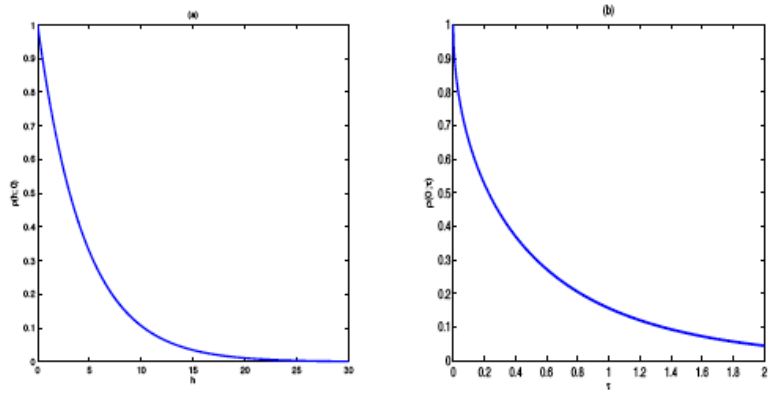
Contour Plot of Spatio-Temporal Correlation Function



The plot shows $\rho_Y(h;\tau)$ for the stochastic reaction-diffusion equation when $\alpha = 1$ and $\beta = 20$

Plots of Marginal Spatial and Temporal Correlation Functions

Special cases include the marginal spatial correlation function at a given time: (a) $\rho_Y(h;0) = \exp\{-h(\alpha/\beta)^{1/2}\},\ h > 0$; and the temporal correlation function at a given spatial location: (b) $\rho_Y(0;\tau) = Erfc(\tau^{1/2}\alpha^{1/2}),\ \tau > 0$.



Spatio-Temporal Stationarity

As with stationarity in space and in time, we can define the usual types of stationarity for spatio-temporal processes.

Definition:

We say that f is a stationary spatio-temporal covariance function on $\mathbb{R}^d \times \mathbb{R}$, if it is nonnegative-definite and can be written as:

$$f((\mathbf{s};t),(\mathbf{x};r)) = C(\mathbf{s}-\mathbf{x};t-r), \quad \mathbf{s},\mathbf{x} \in \mathbb{R}^d,\ t,r \in \mathbb{R}.$$

If a random process $Y(\cdot;\cdot)$ has a constant expectation and a stationary covariance function $C_Y(\mathbf{h};\tau)$, then it is said to be second-order (or weakly) stationary.

(Strong stationarity implies the equivalence of the two probability measures defining the random process $Y(\cdot;\cdot)$ and $Y(\cdot+\mathbf{h};\cdot+\tau)$, respectively, for all $\mathbf{h} \in \mathbb{R}^d$ and all $\tau \in \mathbb{R}$.)

Separability of Spatio-Temporal Covariance Functions

- Stochastic PDEs are built from dynamical mechanistic considerations, and they imply covariance functions.
- Covariance functions have to be positive-definite (p-d). So, specifying classes of spatio-temporal covariance functions to describe the dependence in spatio-temporal data is not all that easy.
- Suppose the spatial $C^{(1)}(\mathbf{h})$ is p-d and the temporal $C^{(2)}(\tau)$ is p-d. Then the separable class:

$$C(\mathbf{h};\tau) \equiv C^{(1)}(\mathbf{h}) \cdot C^{(2)}(\tau)$$

  is guaranteed to be p-d.
- **Separability is unusual in dynamical models**; it says that temporal evolution proceeds independently at each spatial location. That is, separability comes from a lack of spatio-temporal interaction in $Y(\cdot;\cdot)$.

Stochastic Reaction-Diffusion and Separability

- If $C(h;\tau) = C^{(1)}(h) \cdot C^{(2)}(\tau)$, then

$$\begin{aligned} C(h;0) &= C^{(1)}(h)C^{(2)}(0) \\ C(0;\tau) &= C^{(1)}(0)C^{(2)}(\tau), \end{aligned}$$

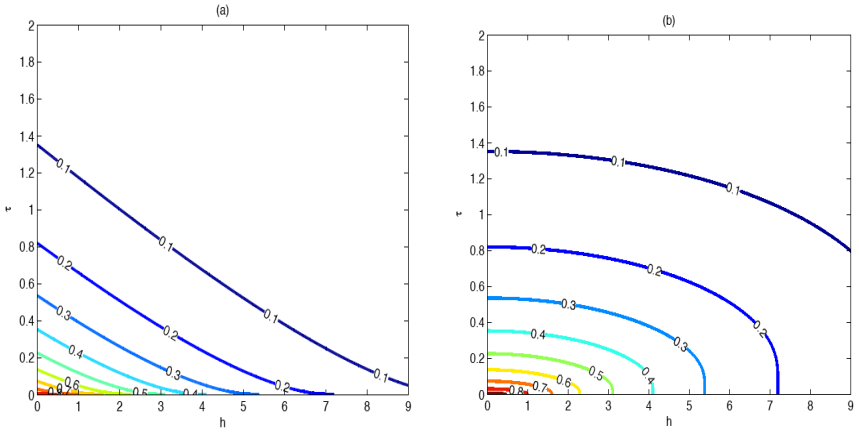
  and hence

$$\begin{aligned} \rho(h;\tau) &= \frac{C^{(1)}(h) \cdot C^{(2)}(\tau)}{C(0;0)} \\ &= \frac{C(h;0) \cdot C(0;\tau)}{C(0;0) \cdot C(0;0)} \\ &= \rho(h;0) \cdot \rho(0;\tau) \end{aligned}$$

- What about the stochastic reaction-diffusion equation for $Y(\cdot;\cdot)$? Plot:

$$\begin{aligned} \rho_Y(h;0) \cdot \rho_Y(0;\tau) &\quad \text{versus} \quad (h,\tau) \\ \rho_Y(h;\tau) &\quad \text{versus} \quad (h,\tau) \end{aligned}$$

Contour Plots of Spatio-Temporal Correlation Functions

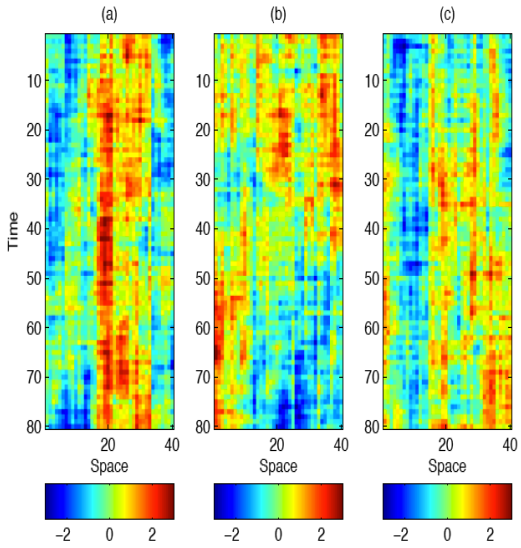


(a) $\rho_Y(h;0) \cdot \rho_Y(0;\tau)$; (b) $\rho_Y(h;\tau)$

The difference in correlation functions is striking. Hence $\rho_Y(\cdot;\cdot)$, for the stochastic reaction-diffusion equation, is non-separable. However, it can be difficult to see the difference between separability and non-separability in realizations from a process.

Non-Separable Realizations in Space-Time

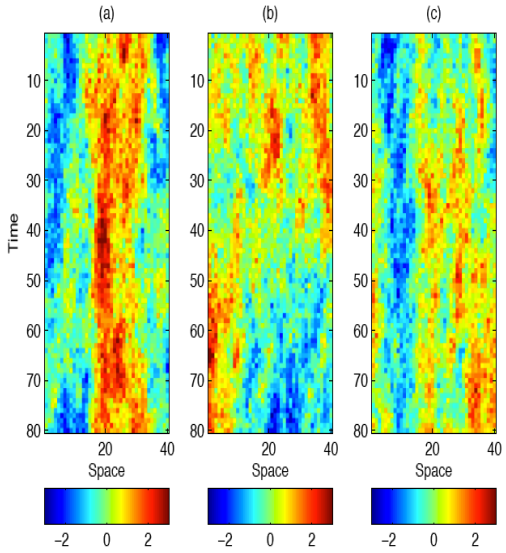
Three realizations of $Y(s;t)$:



Realizations are generated from a stationary Gaussian process with the non-separable, reaction-diffusion correlation function, $\rho_Y(h;\tau)$

Separable Realizations in Space-Time

Three realizations of $Y(s;t)$:



Realizations are generated from a stationary Gaussian process with separable correlation function, $\rho_Y(h;0) \cdot \rho_Y(0;\tau)$

Inference on a Hidden Spatio-Temporal Process

- We could ignore the dynamics and treat time as another "spatial" dimension (i.e., descriptive approach). Write the data as:

$$\mathbf{Z} = (Z(\mathbf{s}_1;t_1),\ldots,Z(\mathbf{s}_m;t_m))',$$

which are observations taken at known space-time "locations."

Note that the data are usually noisy and not observed at all locations of interest.

- Assume a hidden ("true") process, $\{Y(\mathbf{s};t): \mathbf{s} \in D_s \subset \mathbb{R}^d,\ t \geq 0\}$, which is not observable due to measurement error and "missingness." Write

$$\mathbf{Z} = \mathbf{Y} + \boldsymbol{\varepsilon},$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2_\varepsilon \mathbf{I}$.

We wish to predict $Y(\mathbf{s}_0;t_0)$ from data $\mathbf{Z}$

Spatio-Temporal (Simple) Kriging

Predict $Y(\mathbf{s}_0; t_0)$ with the linear predictor, $\boldsymbol{\lambda}'\mathbf{Z} + k$:

For simplicity, assume $E(Y(\mathbf{s}; t)) \equiv 0$. Then $k = 0$, and we minimize w.r.t. $\boldsymbol{\lambda}$, the mean squared prediction error,

$$E(Y(\mathbf{s}_0; t_0) - \boldsymbol{\lambda}'\mathbf{Z})^2 .$$

This results in the simple kriging predictor:

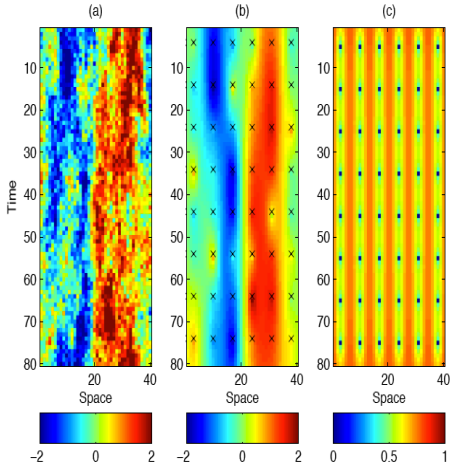
$$\widehat{Y}(\mathbf{s}_0; t_0) = \mathbf{c}(\mathbf{s}_0; t_0)'\boldsymbol{\Sigma}_Z^{-1}\mathbf{Z} ,$$

where $\boldsymbol{\Sigma}_Z \equiv \text{cov}(\mathbf{Z})$, and
$\mathbf{c}(\mathbf{s}_0; t_0)' = \text{cov}(Y(\mathbf{s}_0; t_0), \mathbf{Z}) = \text{cov}(Y(\mathbf{s}_0; t_0), \mathbf{Y})$

The simple kriging standard error (s.e.) is:

$$\sigma_k(\mathbf{s}_0; t_0) = \{\text{var}(Y(\mathbf{s}_0; t_0)) - \mathbf{c}(\mathbf{s}_0; t_0)'\boldsymbol{\Sigma}_Z^{-1}\mathbf{c}(\mathbf{s}_0; t_0)\}^{1/2}$$

Kriging for Stochastic Reaction-Diffusion Equation

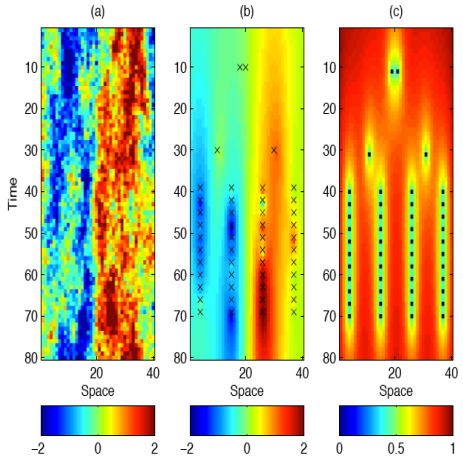


(a) For simplicity, assume no noise in the data $\mathbf{Z}$ (i.e., $\varepsilon = \mathbf{0}$)

(b) Crosses show $\{(\mathbf{s}_i; t_i) : i = 1, \ldots, 48\}$ ("data" locations) superimposed on the kriging predictor map, $\{\widehat{Y}(\mathbf{s}_0; t_0)\}$

(c) Kriging s.e. map, $\{\sigma_k(\mathbf{s}_0; t_0)\}$

Kriging for Stochastic Reaction-Diffusion Equation (cont.)



(a) Same noiseless dataset (i.e., $\varepsilon = \mathbf{0}$)

(b) Crosses show different $\{(\mathbf{s}_i; t_i) : i = 1, \ldots, 48)\}$ superimposed on the kriging predictor map, $\{\widehat{Y}(\mathbf{s}_0; t_0)\}$

(c) Kriging s.e. map, $\{\sigma_k(\mathbf{s}_0; t_0)\}$

Spatio-Temporal Covariance Functions

In practice, one does not typically know the underlying stochastic PDE that governs the system of interest (i.e., most processes are multivariate, nonlinear or have complex parameter dependencies). Even with such knowledge, it may not be possible to find the analytical covariance function.

We saw that the assumption of separability is not very realistic and that covariance functions must satisfy the positive-definiteness property. This suggests the need for realistic classes of spatio-temporal covariance functions.

In recent years, there has been good progress in **developing new classes of spatio-temporal covariance functions** through the use of the spectral-domain representation and Bochner's Theorem (e.g., see C&W 2011, Sec. 6.1.6: Examples include the work of Cressie and Huang, 1999; Gneiting, 2002; Stein, 2005; and many others).

Spatio-Temporal Covariance Functions (cont.)

To date, available classes of (descriptive) S-T covariance functions are not realistic for many complicated phenomena, and there can be serious computational issues with their implementation in traditional kriging formulas due to the dimensionality of the prediction problems of interest.

As an alternative, we can make use of dynamical (conditional) formulations. These simplify the joint-dependence structure. In addition, because conditional models are closer to the process' etiology, it may be easier to incorporate process knowledge directly (e.g., using dynamical models).

Consider again the stochastic reaction-diffusion equation, now from the dynamical perspective.

Emphasize the Dynamics

Approximate the differentials in the reaction-diffusion equation:

$$\frac{\partial Y}{\partial t} = \beta \frac{\partial^2 Y}{\partial s^2} - \alpha Y$$

with differences over the grid from 0 to L at intervals Δ_s:

$$\frac{Y(s; t+\Delta_t) - Y(s;t)}{\Delta_t} = \beta \left\{ \frac{Y(s+\Delta_s;t) - 2Y(s;t) + Y(s-\Delta_s;t)}{\Delta_s^2} \right\} - \alpha Y(s;t)$$

Define $\mathbf{Y}_t \equiv (Y(\Delta_s;t), \ldots, Y(L-\Delta_s;t))'$; $\mathbf{Y}_t^B \equiv (Y(0;t), Y(L;t))'$.

Then the stochastic version of the difference equation above is:

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

where $\mathbf{M}_B\mathbf{Y}_t^B$ represents given boundary effects. The difference equation is a stable explicit approximation to the differential equation, provided $\alpha\Delta_t < 1$ and $2\beta\Delta_t/\Delta_s^2 < 1$.

Emphasize the Dynamics (cont.)

Importantly, the matrix $\mathbf{M}$ is given by

$$\mathbf{M} = \begin{bmatrix} \theta_1 & \theta_2 & 0 & \dots & 0 \\ \theta_2 & \theta_1 & \theta_2 & \dots & \vdots \\ 0 & \theta_2 & \theta_1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \theta_2 \\ 0 & 0 & \dots & \theta_2 & \theta_1 \end{bmatrix},$$

where $\theta_1 = (1 - \alpha\Delta_t - 2\beta\Delta_t/\Delta_s^2)$, $\theta_2 = \beta\Delta_t/\Delta_s^2$.

This can be viewed as the propagator (transition) matrix of a $\mathrm{VAR}(1)$ process. The matrix *is defined* by the dynamics. In other words, in a dynamic model of spatio-temporal dependence, **M** has structure (which is typically sparse).

Emphasize the Dynamics (cont.)

Conditional on the boundary effects, we see that the lagged spatial covariances are given by,

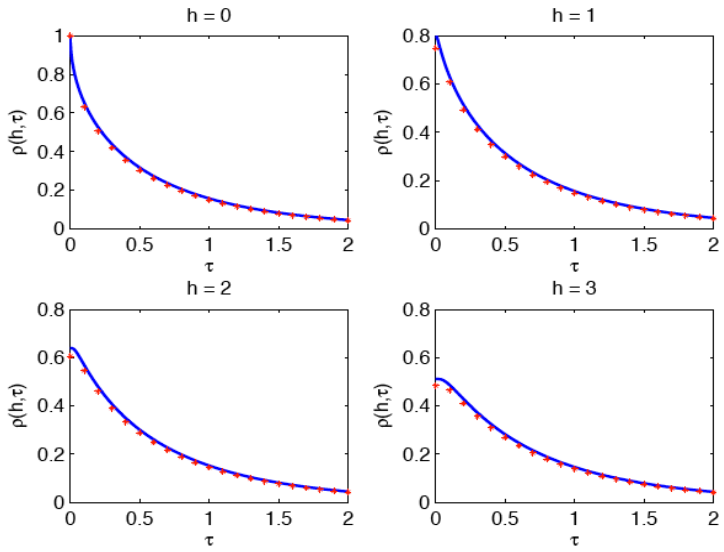
$$\mathbf{C}_Y^{(m)} = \mathbf{M}^m \mathbf{C}_Y^{(0)},$$

where $\mathbf{C}_Y^{(m)} \equiv \mathrm{cov}(\mathbf{Y}_t, \mathbf{Y}_{t+m\Delta_t})$; $m = 0, 1, 2, \dots$, and it can be shown that the lag-0 marginal spatial covariance for Y can be written in terms of the propagator matrix $\mathbf{M}$ and the spatial covariance matrix for the η-process, $\mathbf{C}_\eta^{(0)}$:

$$\mathrm{vec}(\mathbf{C}_Y^{(0)}) = (\mathbf{I} - \mathbf{M} \otimes \mathbf{M})^{-1} \mathrm{vec}(\mathbf{C}_\eta^{(0)}).$$

This suggests that we can compare the spatio-temporal covariance structure for this reaction-diffusion difference equation with the PDE's theoretical form derived by Heine (1955).

Comparison of Differential and Difference Equations



Spatio-temporal correlations; $\alpha = 1$, $\beta = 20$, $\Delta_s = 1$, and $\Delta_t = 0.01$
Solid blue line: from **differential** equation
Red dots: from **difference** equation

The Dynamics in the Difference Equation

Think of a spatial process at time t rather than a spatio-temporal process. Call it the vector $\mathbf{Y}_t$. Then describe its dynamics by a discrete-time Markov process; e.g., $\mathrm{VAR}(1)$:

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

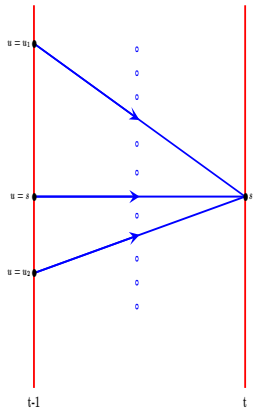
As implied above, the choice of $\mathbf{M}$ is crucial. In particular, we note that $\mathbf{M} \equiv (m_{ij})$ represent "spatial weights" of the process values from the past, e.g.,

$$Y_t(\mathbf{s}_i) = \sum_{j=1}^{n} m_{ij} Y_{t-1}(\mathbf{s}_j) + \eta_t(\mathbf{s}_i).$$

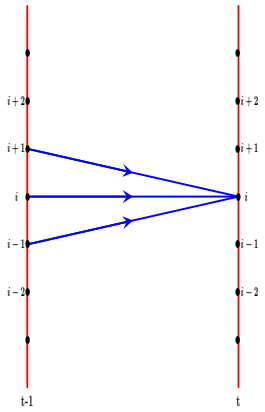
Usually, many of these coefficients have small or zero weight. Typically, the m_{ij} corresponding to nearby locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are non-zero, and they are zero when locations are far apart.

Structure of **M**

These directed graphs show the case of one-dimensional space:



General **M**



M defined "spatially"

Instantaneous Spatial Dependence (ISD)

To capture the process' behavior at small temporal scales between time t and time $t+1$, we need a component of variation that is modeled as **instantaneous spatial dependence (ISD)**:

$$\mathbf{Y}_t = \mathbf{B}_0\mathbf{Y}_t + \mathbf{B}_1\mathbf{Y}_{t-1} + \boldsymbol{\xi}_t\,,$$

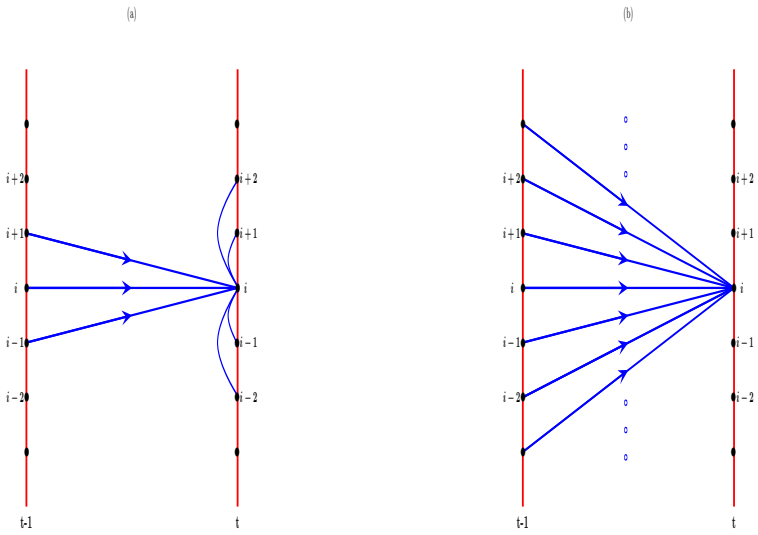
where $\mathbf{B}_0$ has zero down its diagonal. Again, it is useful to model $\mathbf{B}_0$ and $\mathbf{B}_1$ "spatially"; see the figure below. This implies that

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

where $\mathbf{M} = (\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{B}_1$ and $\boldsymbol{\eta}_t = (\mathbf{I} - \mathbf{B}_0)^{-1}\boldsymbol{\xi}_t$.

What use are $\mathbf{B}_0$ and $\mathbf{B}_1$? They imply dynamical structure and are sparse! (But, interestingly, notice that $\mathbf{M}$ is not sparse in this case).

ISD in Graphical Form



(*a*) Graph structure (sparse) showing relationships that are defined "spatially"

(*b*) Equivalent directed graph structure (non-sparse)

Nonstationarity and Dynamical Models

Spatial and temporal stationarity can be an unrealistic assumption. Yet, descriptive approaches to spatio-temporal modeling, expressed in terms of covariance functions, typically demand it.

Dynamical approaches are much more forgiving. Consider the simple (conditionally) nonstationary $\mathrm{VAR}(1)$ process:

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

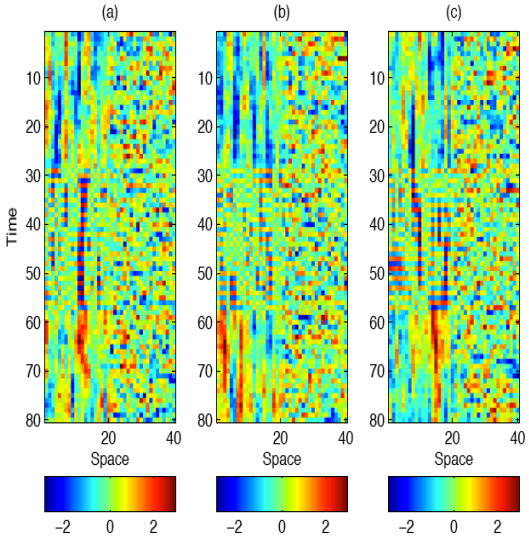
For example, $\mathbf{M}_t = f(t) \cdot \mathbf{M}$, where

$$f(t) = \begin{cases} 1 & 0 \leq t \leq 29 \\ -1 & 30 \leq t \leq 59 \\ 1 & 60 \leq t \leq 79\,, \end{cases}$$

and $\mathbf{M}$ is tridiagonal but has different parameters for $0 \leq s \leq 19$ and for $20 \leq s \leq 39$

Realizations for a Nonstationary Process

Three realizations of $Y(s;t)$:



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

More Realistic PDEs

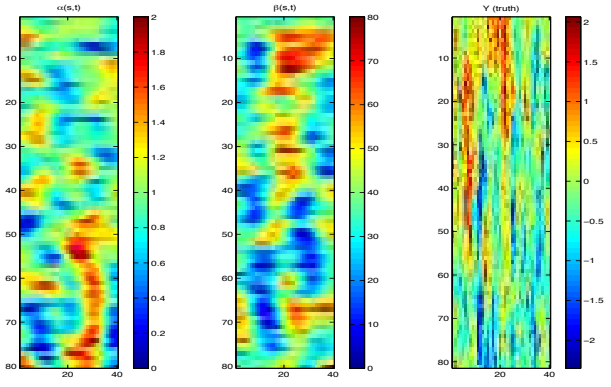
Reaction-Diffusion Ex. Revisted

Assume that the reaction-diffusion equation now has spatially and time-varying parameters, where the parameters are themselves random processes.

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

Assume that the parameters evolve as reduced rank linear DSTMs.

$\alpha(s;t)$ $\beta(s;t)$ $Y(s;t)$



Note the nonlinear interaction between the parameter "processes" and the "Y" process. If this is happening, the process is really multivariate.
Note: no strong indication this is happening visually.

Structure of **M** (cont.)

The importance of the structure of **M** suggests ways in which this matrix can be parameterized.

What is it about the structure of this matrix and the values of these "nearest neighbor" parameters that affect the dynamics? Can we use this sort of scientific process knowledge (in various forms) to help with this parameterization?

In fact, this type of information can help us but we need an efficient framework in which to build it into the model.

The hierarchical modeling framework is quite helpful in this regard.

Towards Hierarchical Spatio-Temporal Statistical Models

- We can motivate dynamical models through mechanistic relationships.
- These models can still be over-parameterized, or too simple for real-world processes.
- We must account for this complexity and our uncertainty in the process and parameters.
- There is also uncertainty in data, and the size of the dataset can be a problem.
- **Hierarchical statistical models** (specifically, Bayesian Hierarchical Models, BHMs) can provide a framework to account for these issues.

Before getting back to the dynamical specifications, consider the following motivating example to illustrate the BHM approach for spatio-temporal modeling.

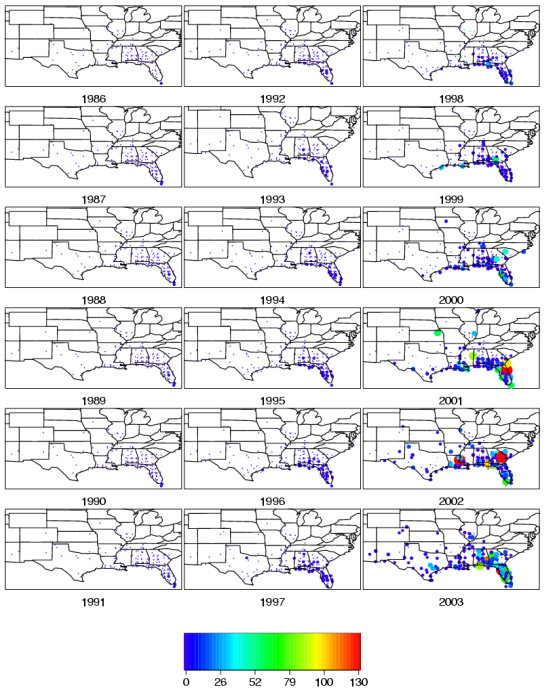
Introduction to Hierarchical Models

Hierarchical Model Motivating Problem: Invasive Species

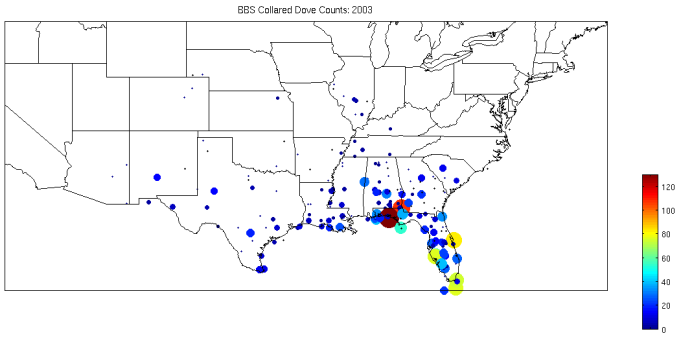
Spread of Eurasian Collared-Doves (ECDs)

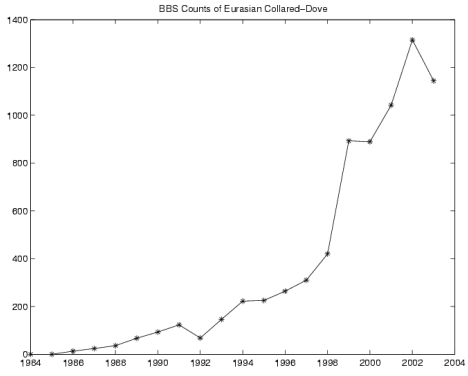
- The Eurasian Collared-Dove (*Streptopelia decaocto*) originated in Asia and, starting in the 1930s, expanded its range into Europe (Hudson, 1965).
- They were first observed in the United States in the mid 1980s after being introduced into the Bahamas in 1974 from a population that escaped captivity (Smith, 1987).
- Since the species' introduction in Florida, its range has been expanding dramatically across North America.

Breeding Bird Survey (BBS) Counts of ECD, 1986-2003



BBS ECD Counts: 2003 and Yearly Totals





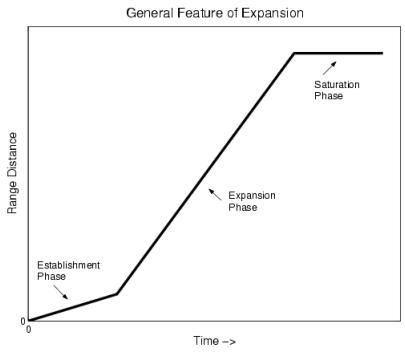
Invasion Impacts

- ECD biological threats (Romagosa and Labisky, 2000): competition for resources with native avifauna; transmission of disease
- "ECD will probably colonize all of North America within a few decades" (Romagosa and Labisky, 2000)

Just how probable is this colonization? The example presented later will answer this question.

The following provides the spatio-temporal hierarchical motivation for such a model.

Typical Invasions



Invasive species phases:

- Introduction
- Establishment
- Range Expansion
- Saturation

Ecological models for invasions involve dispersal and growth

Uncertainty in Spread of Invasives

- Uncertainty in data (e.g, BBS counts)
  - differences in experience and expertise of the BBS volunteer observers leads to differences in probability of detection
  - Eurasian Collared-Dove is similar in appearance to Ringed Turtle-Dove. Although there are fundamental differences, observers routinely mistake these species, especially early in invasion
- Uncertainty and complexity in the underlying spatio-temporal process dynamics
  - "diffusion" (spread) and growth
  - species interactions
  - important exogenous variables
- Uncertainty in parameters
  - diffusion, growth, and carrying capacity vary spatially

Bayesian Hierarchical Models

We temporarily drop the spatio-temporal index and write data as Z, the process as Y, and parameters as θ. The joint uncertainty is expressed through: $[Z, Y, \theta] = [Z|Y, \theta][Y|\theta][\theta]$

Rather than seek to model the complicated joint distribution, we factor it into a product of a sequence of conditional distributions, to which we might be able to apply scientific insight.

Thus, for complicated spatio-temporal processes, we consider the following three-stage factorization of $[data, process, parameters]$ (Berliner, 1996; Wikle et al., 1998):

Stage 1. Data Model: $[data|process, data\ parameters]$

Stage 2. Process Model: $[process|process\ parameters]$

Stage 3. Parameter Model: $[data\ params\ and\ process\ params]$.

Bayesian Hierarchical Models (cont.)

One of the most important points of the Berliner (1996) hierarchical modeling paradigm (*which is often lost on many who claim to do hierarchical modeling*):

- put as much structure as possible in the conditional mean via random effects
- correspondingly, make the conditional dependence as simple as possible

That is, push the dependence down a level of the hierarchy to the mean to **build** dependence; first moments are much easier to model and there is much more scientific knowledge about their specification!

We do this all of the time in linear mixed models: e.g.,

$$\text{Conditional specification: } \mathbf{Z}|\mathbf{Y},\boldsymbol{\beta},\sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}+\mathbf{W}\mathbf{Y},\sigma^2\mathbf{I}),$$
$$\mathbf{Y}|\boldsymbol{\Sigma} \sim N(\mathbf{0},\boldsymbol{\Sigma})$$
$$\text{Marginal specification: } \mathbf{Z}|\boldsymbol{\beta},\boldsymbol{\Sigma},\sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta},\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}'+\sigma^2\mathbf{I}).$$

Data Models

Let Z_a be data observed for some process Y, and let θ_a be parameters.

The data model is written:

$$[Z_a|Y,\theta_a]$$

This distribution is much simpler than $[Z_a|\,\theta_a]$, because most of the complicated structure (spatial and temporal) comes from the process Y.

Data Models (cont.)

Combining data sets: given observations Z_a, Z_b for the same process, Y, often we can write:

$$[Z_a, Z_b|Y, \theta_a, \theta_b] = [Z_a|Y, \theta_a][Z_b|Y, \theta_b] .$$

That is, conditional on the true process, the data are often assumed independent. (Note that they are almost certainly unconditionally dependent!). This hierarchical framework presents a natural way to accommodate data at differing spatial and temporal resolutions and alignments (e.g.,Wikle and Berliner, 2005).

Similarly, for multivariate process (Y_a, Y_b), often we can write:

$$[Z_a, Z_b|Y_a, Y_b, \theta_a, \theta_b] = [Z_a|Y_a, \theta_a][Z_b|Y_b, \theta_b] .$$

Again, conditional on the true processes, the data are often assumed independent.

Process Models

Process models are also often factored into a series of conditional models:

$$[Y_a, Y_b|\theta_Y] = [Y_a|Y_b, \theta_Y][Y_b|\theta_Y]$$

We made such an assumption when using the first-order Markov model for dynamical processes. (In that case, the "a" and "b" subscripts refer to time t and $t-1$, respectively.)

Such factorizations are also important for simplifying multivariate processes; Royle and Berliner (1999) consider such a conditional framework for modeling multivariate spatial processes. For example, consider ozone concentration conditioned on temperature; or consider CO_2 conditioned on potential temperature.

Parameter Models

Parameter models can also be factored into subcomponents. For example, we might assume,

$$[\theta_a, \theta_b, \theta_Y] = [\theta_a][\theta_b][\theta_Y].$$

That is, we often assume that parameter distributions are independent, although subject-matter knowledge may lead to more complex parameter models.

Scientific insight and previous studies can facilitate the specification of these models. For example, measurement-error parameters can often be obtained from previous studies that focused on such issues (this is typically the case for environmental variables and some ecological data such as from the BBS).

Parameter Models (cont.)

Process parameters often carry scientific insight (e.g., spatially dependent diffusion parameters, Wikle, 2003; turbulence parameters, Wikle et al., 2001).

In some cases, we do not know much about the parameters and use vague or non-informative distributions for parameters. Alternatively, we might use data-based estimates for such parameters.

Specification of parameter distributions is often criticized for its "subjectiveness." Such criticism is misguided!

Bayesian Hierarchical Model (BHM): Schematic Example

- [data | process,parameters]: uncertainty in observations. For example,

  [bird-count observations | true bird counts, data parameters]

- [process | parameters]: science (diffusion PDEs); partitioned into subcomponents (first-order Markov process); uncertainty (additive noise, random effects). For example,

  [true bird counts | diffusion and growth processes, process params]

- [parameters]: prior scientific understanding. For example,

  [diffusion parameters | habitat covariates]

Empirical Hierarchical Model (EHM)

- [data | process, parameters]: For example,

  [bird-count observations | true bird counts, data parameters]

- [process | parameters]: For example,

  [true bird counts | diffusion and growth processes, process params]

- data parameters and process parameters are assumed fixed but unknown. They are typically estimated based on the marginal distribution,

  [data | parameters]

This framework is common in traditional state-space models where one might use an E-M algorithm for parameter estimation.

Inference for Hierarchical Statistical Models

Bayesian Hierarchical Model (BHM): Use Bayes' Theorem to derive the posterior distribution,

$$[\text{process}, \text{parameters} \mid \text{data}]$$
$$\propto \text{Data Model} \times \text{Process Model} \times \text{Parameter Model}$$

The normalizing constant is [data]

Empirical Hierarchical Model (EHM): Use Bayes' Theorem to derive the predictive distribution,

$$[\text{process} \mid \text{data}, \text{parameters}]$$
$$\propto \text{Data Model} \times \text{Process Model}$$

The normalizing constant is [data | parameters]. The unknown parameters are replaced with estimates.

Bayesian Hierarchical Modeling Issues

- **Data Models:**
  - Gaussian/non-Gaussian
  - Discrete/continuous
  - Measurement error
  - Change of support and alignment
  - Multiple data sources
- **Process Models:**
  - Prior scientific knowledge (e.g., PDEs)
  - Dimensionality
  - Boundary conditions
  - Distributional form
- **Parameter Models:**
  - Prior scientific knowledge (e.g., habitat suitability)
  - Dimensionality
  - Distributional form

Computation of the posterior distribution can be a challenge!

General Dynamic Spatio-Temporal Model (DSTM)

Data 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**

Process Error

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

The ability to incorporate dependence structure and conditional relationships in these models is what gives hierarchical modeling its power!

General DSTM (Data Models)

The general DSTM data model typically makes the same assumption as in generalized linear mixed models (GLMMs): conditioned on the mean response, the observations are independent. This makes a dramatic simplification in the case of non-Gaussian likelihoods. Note, in the hierarchical framework, this is not limited to the exponential family.

In the context of DSTMs, conditioned on the spatio-temporal process(es), the observations are assumed to be independent. The focus is then on modeling the **latent spatio-temporal process(es).**

General DSTM (Data Models)

In most cases, a transformation of the underlying latent spatio-temporal process is assumed to be conditionally Gaussian – this is then where we put our modeling effort.

E.g., one could imagine this corresponding to the underlying intensity of a spatio-temporal (log-Gaussian Cox) point process or the logit of the probability of presence in an occupancy model.

We consider this conditional Gaussian latent process approach here. *It is important to note that we do NOT assume in general that the latent process is marginally Gaussian! Such an assumption is too limiting for real-world processes.*

NOTE: although this conditional likelihood perspective is quite general and effective, there are some alternative approaches (e.g., spatio-temporal auto-logistic models (Zheng and Zhu, 2008); spatio-temporal stochastic agent-based models (Hooten and Wikle, 2010), etc.).

Statistical DSTMs: Process Modeling

- Spatio-temporal dynamics are due to the interaction of the process across space and time and/or across scales of variability
  - Some types of interaction make sense for some processes, and some don't (e.g., process knowledge should not be ignored if available)
  - Statisticians have often ignored such knowledge!
- Dimensionality can prevent the (efficient) estimation of model parameters, e.g., $\mathcal{M}(\cdot)$ or $\mathbf{M}$
  - Requires sensible science-based parameterizations and/or dimension reduction; sparse structures
  - Hierarchical representations can help here as well

Linear DSTM Process Models

Linear Dynamical Spatio-Temporal Process Models

Models for physical, environmental, and ecological data: the process at the current time is related to the process at a previous time (or times).

We refer to such a process as a dynamical process.

Consider n finite spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and discrete times $t \in \{0, 1, 2, \ldots, T\}$. Define $\mathbf{Y}_t \equiv (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))'$.

The joint distribution of the spatio-temporal process can be factored as follows:

$$[\mathbf{Y}_0, \ldots, \mathbf{Y}_T] = [\mathbf{Y}_T|\mathbf{Y}_{T-1}, \ldots, \mathbf{Y}_0][\mathbf{Y}_{T-1}|\mathbf{Y}_{T-2}, \ldots, \mathbf{Y}_0] \ldots [\mathbf{Y}_2|\mathbf{Y}_1, \mathbf{Y}_0][\mathbf{Y}_1|\mathbf{Y}_0][\mathbf{Y}_0]$$

We can simplify this greatly by a first-order Markov assumption:

$$[\mathbf{Y}_t|\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots] = [\mathbf{Y}_t|\mathbf{Y}_{t-1}]$$

Linear Dynamical Spatio-Temporal Process Models (cont.)

Thus, the joint distribution can be written:

$$[\mathbf{Y}_0, \ldots, \mathbf{Y}_T] = [\mathbf{Y}_T|\mathbf{Y}_{T-1}][\mathbf{Y}_{T-1}|\mathbf{Y}_{T-2}] \ldots [\mathbf{Y}_2|\mathbf{Y}_1][\mathbf{Y}_1|\mathbf{Y}_0][\mathbf{Y}_0].$$

Critical to the specification of this distribution is a model for:

$$[\mathbf{Y}_t|\mathbf{Y}_{t-1}]$$

In general, we write this in terms of some function,

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

where the parameters $\boldsymbol{\theta}_m$ describe the dynamical propagation, and $\boldsymbol{\eta}_t$ represents process uncertainty. This function can be nonlinear, and the associated distribution can be Gaussian or non-Gaussian. For now, consider a linear, Gaussian model.

Statistical DSTMs: Linear Process Modeling

Consider a mechanistically-motivated parameterization. E.g., parameterizations suggested by linear models of the form:

- PDE:
$$\mathcal{M}Y(\mathbf{s}; t) = \eta(\mathbf{s}; t)$$
- Integro-difference equation (IDE):
$$Y_t(\mathbf{s}) = \gamma \int_D m_s(\mathbf{r}; \boldsymbol{\theta}_s) Y_{t-1}(\mathbf{r}) d\mathbf{r} + \eta_t(\mathbf{s})$$
- Difference equation:
$$Y_t(\mathbf{s}_i) = \sum_{j=1}^{n} m_{ij}(\boldsymbol{\theta}_i) Y_{t-1}(\mathbf{s}_j) + \eta_t(\mathbf{s}_i),$$

In discretized form, they all can be written (or, as higher-order AR models):

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

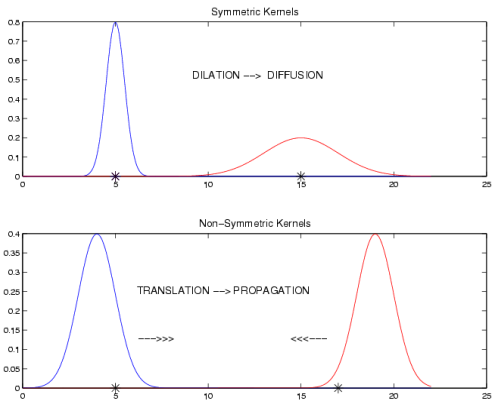
DSTMs: Linear Processes

The transition operators/kernels/weights correspond to the redistribution of the process from the previous time at all potential spatial locations in the domain. Typically, these can be quite sparse and/or we can parameterize $\mathbf{M}$ in terms of $\boldsymbol{\theta}_s$ as suggested by the form of the mechanistic models. How?

First-Order Linear DSTM Process

Linear spatio-temporal processes often exhibit advective and diffusive behavior:

- "width" (decay rate) of the transition operator neighborhood controls the rate of spread (diffusion)
- degree of "asymmetry" in the transition operator controls the speed and direction of propagation (advection)
- "long range dependence" can be accommodated by "multimodal" operators and/or heavy tails



This suggests ways that we might parameterize the transition operator and/or induce sparse structure.

Basic Hierarchical Linear DSTM

Data:

$$\mathbf{Z}_t = \mathbf{H}_t(\boldsymbol{\theta}_{h,1})\mathbf{Y}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim Gau(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{h,2}))$$

Process:

$$\mathbf{Y}_t = \mathbf{M}(\boldsymbol{\theta}_{m,1})\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_{m,2}))$$

Parameters:

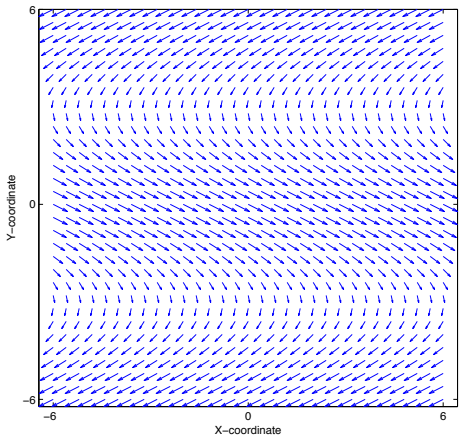
$\boldsymbol{\theta}_{h,1},\ \boldsymbol{\theta}_{h,2},\ \boldsymbol{\theta}_{m,1},\ \boldsymbol{\theta}_{m,2}$

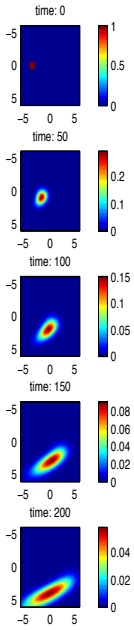
These parameters may be estimated empirically, or they can be given prior distributions, such as Gaussian random process priors (that may depend on other variables), and they can easily be allowed to vary with time and/or space so as to borrow strength.

First-Order Linear DSTM Process

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.)





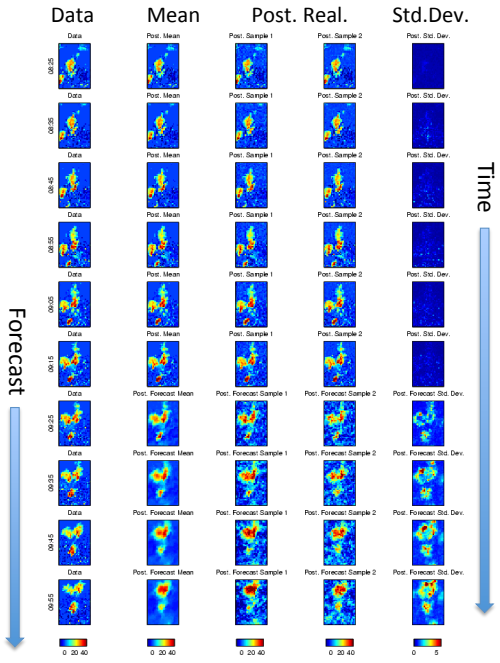
In the case where we didn't know these parameters, we could specify a prior distribution for them that might include covariates and/or spatial random fields in the hierarchical framework: e.g.,

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

September 25, 2010: Radar Nowcasting Motivation!

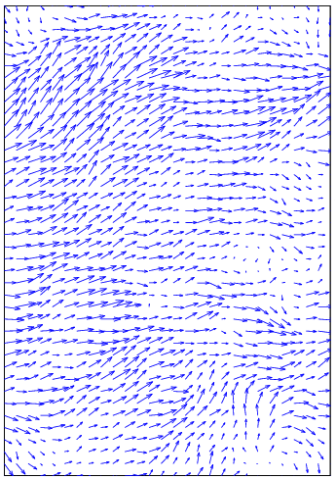
It may look like something out of a movie, but it's not — this was the scene from early Saturday afternoon in Kansas. (AP Photo/Charlie Riedel)

Example: Radar Nowcasting (Sydney, pre 2000 Olympics)



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

Implied Propagation by Post. Parms.



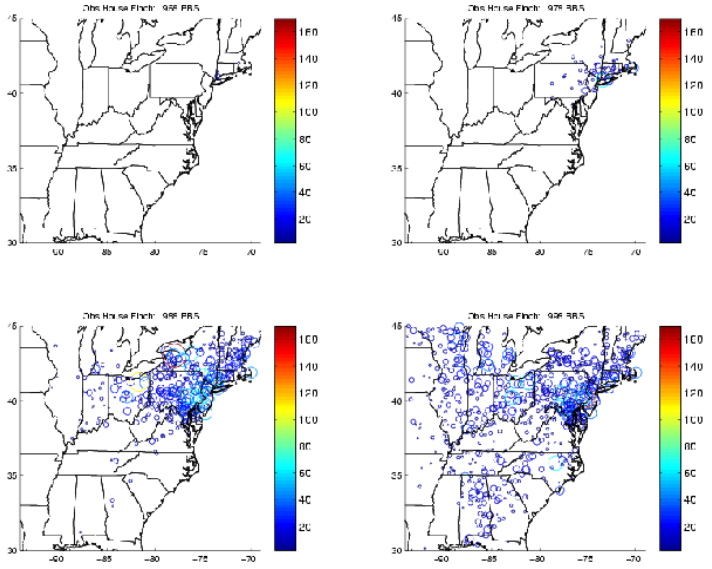
i.e., the advection and diffusion parameters follow a latent spatial process

Xu, Wikle, and Fox, 2005; *JASA*

Diffusion Application: (Wikle 2003; Ecology)

Spread of invasive species on landscape scale:

Breeding Bird Survey (BBS) counts for house finch (*Carpodacus mexicanus*) [substantial observer error and bias!]



Ecology Example: Process Model Motivation

Skellam's (1951) Model (Diffusion plus Malthusian growth):

$$\frac{\partial u}{\partial t} = \delta \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + \alpha u.$$

Over large spatial scales, constant diffusion rate is not realistic. Spread will not be spatially homogeneous! e.g., a better model (but, still, an incorrect one) might be:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(\delta(x,y) \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(\delta(x,y) \frac{\partial u}{\partial y} \right) + \alpha u,$$

Ecology Example: Process Model Motivation

Implementation Issues:

- No general analytical solution
- Although this can be solved numerically, we still don't expect it to be exactly appropriate
- Growth process naive
- How do we estimate the spatial process $\delta(x, y)$?
- How do we account for the measurement/sampling errors (e.g., BBS data)?

Ecology Example: Hierarchical Model (Sketch)

- **Data Model:** $Z_t(\mathbf{s}_i)$ observed BBS count

$$Z_t(\mathbf{s}_i)|\lambda_t(\mathbf{s}_i) \sim \textit{iid Poisson}(\lambda_t(\mathbf{s}_i))$$

- **Process Models:**

$$\log(\lambda_t(\mathbf{s}_i)) = \mu_t + \mathbf{h}_{it}'\mathbf{u}_t + \eta_t(\mathbf{s}_i), \quad \eta_t(\mathbf{s}_i) \sim \textit{iid } N(0, \sigma_\eta^2)$$

$$\mu_t = \mu_{t-1} + \epsilon_t, \quad \epsilon_t \sim \textit{iid } N(0, \sigma_\epsilon^2)$$

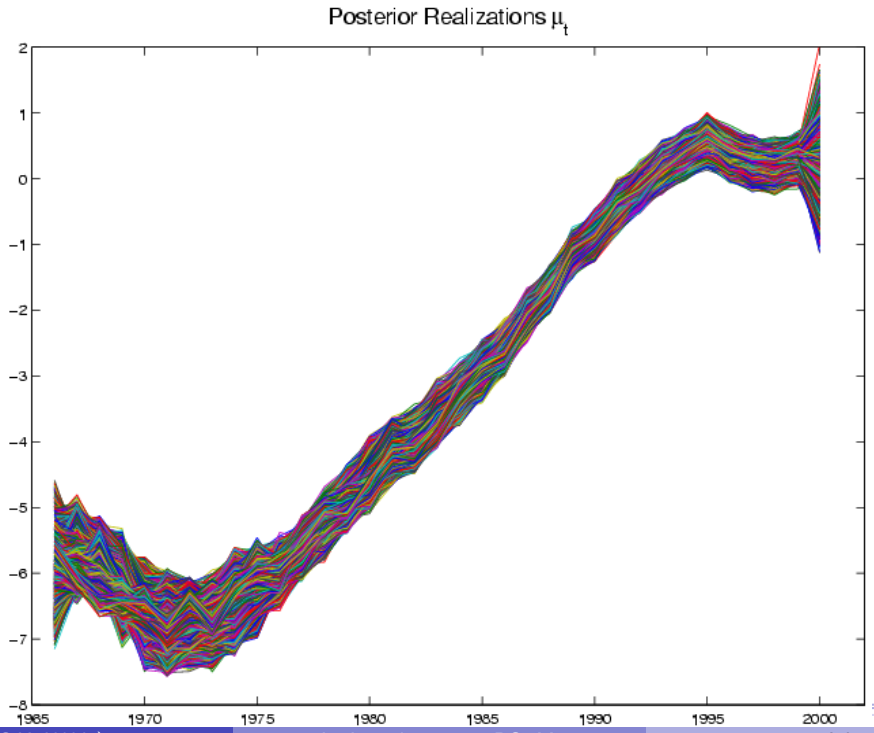
$$\mathbf{u}_t = \mathbf{M}(\boldsymbol{\delta}, \alpha)\mathbf{u}_{t-1} + \boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim \textit{iid } N(\mathbf{0}, \boldsymbol{\Sigma}(\theta_\gamma))$$

- **Parameter Models:**

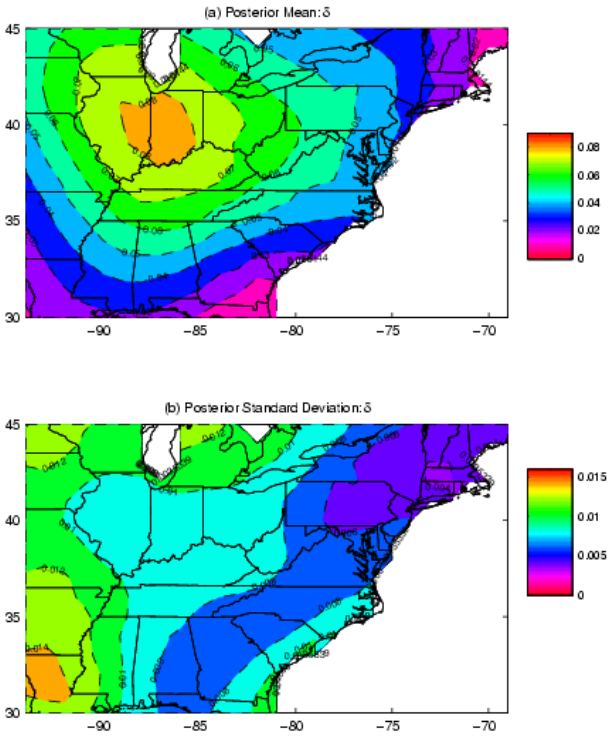
$$\boldsymbol{\delta}|\boldsymbol{\beta}, \sigma_\delta^2, \mathbf{R}_\delta \sim N(\boldsymbol{\Phi}\boldsymbol{\beta}, \sigma_\delta^2\mathbf{R}_\delta),$$
$$\alpha \sim N(\tilde{\alpha}_0, \tilde{\sigma}_\alpha^2)$$

Others: $\mathbf{u}_0$, $\boldsymbol{\beta}$, variances

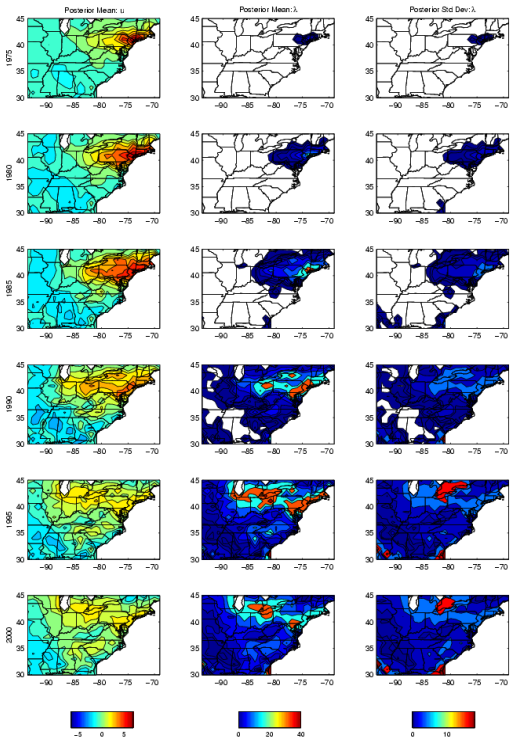
Ecological Example Results: (log) Growth Posterior Samples



Ecological Example Results: Spatial Diffusion Parameter



Ecological Example Results: Process



"Spectral" Representation of Dynamics

It can be useful to parameterize the science-based dynamical spatio-temporal process in terms of basis function (spectral) expansions:

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

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

where $\boldsymbol{\Phi}$ is a basis function matrix and $\boldsymbol{\alpha}_t$ are associated expansion coefficients.

In this case, either the dimension of the dynamical process $\boldsymbol{\alpha}_t$ is much lower than n, reducing the number of parameters in $\mathbf{M}_\alpha$ and $\mathbf{Q}_\alpha \equiv \text{cov}(\boldsymbol{\eta}_{\alpha,t})$, many elements of $\mathbf{M}_\alpha$ are zero, and/or the reduction acts as a decorrelator, which reduces the complexity of these matrices.

One can still get science-based parameterizations when working in "spectral" space. In particular, many PDEs and IDEs are amenable to spectral and Galerkin-based representations (e.g., see Wikle et al. 2001; Arab 2007; C&W 2011, pp. 396-402).

Advection/Diffusion Example Revisited: Spectral

Again, consider the linear advection/diffusion equation (in 1-D):

$$\frac{\partial u}{\partial t} = -u_0 \frac{\partial u}{\partial x} + A \frac{\partial^2 u}{\partial x^2}$$

Assume solutions are a superposition of wave modes of the form:

$$u_t(x) = \sum_j [a_{1j}(t)\cos(\omega_j x) + a_{2j}(t)\sin(\omega_j x)]$$

where $\omega_j = 2\pi j / D_x$ is the spatial frequency of a wave with wave number j over domain D_x.

Thus, for all spatial locations of interest, $\mathbf{u}_t = \mathbf{\Phi}\mathbf{a}_t$ where $\mathbf{\Phi}$ is made up of the Fourier basis functions, and $\mathbf{a}_t$ is the collection of all wave-mode coefficients.

ADE Spectral (cont.)

The deterministic solution gives formulae for $a_{1j}(t), a_{2j}(t)$ (exponentially decaying sinusoids in time):

$$\begin{aligned} a_{1j}(t) &= \exp(-A\omega_j^2 t)\sin(u_0\omega_j t) \\ a_{2j}(t) &= \exp(-A\omega_j^2 t)\cos(u_0\omega_j t) \end{aligned}$$

Note time evolution:

$$\begin{bmatrix} e^{-A\omega_j^2(t+\delta)}\sin\{\omega_j(t+\delta)\} \\ e^{-A\omega_j^2(t+\delta)}\cos\{\omega_j(t+\delta)\} \end{bmatrix} = \mathbf{G}_j \begin{bmatrix} e^{-A\omega_j^2 t}\sin\{\omega_j t\} \\ e^{-A\omega_j^2 t}\cos\{\omega_j t\} \end{bmatrix}$$

where

$$\mathbf{G}_j = \begin{bmatrix} e^{-A\omega_j^2\delta}\cos\{\omega_j\delta\} & e^{-A\omega_j^2\delta}\sin\{\omega_j\delta\} \\ -e^{-A\omega_j^2\delta}\sin\{\omega_j\delta\} & e^{-A\omega_j^2\delta}\cos\{\omega_j\delta\} \end{bmatrix}$$

ADE Spectral (cont.)

Thus (deterministic) linear wave theory suggests:

$$\mathbf{a}_j(t+\delta) = \mathbf{G}_j\mathbf{a}_j(t)$$

However, we don't expect the true process to behave *exactly* as the linear wave theory suggests!

- Let $a_{1j}(t), a_{2j}(t)$ be *stochastic*
- Add noise term $\boldsymbol{\eta}_j(t)$ to account for uncertainty
- Let the propagator be $\mathbf{M}_j$ with *prior mean* $\mathbf{G}_j$

$$\mathbf{a}_{t+\delta} = \mathbf{M}\mathbf{a}_t + \boldsymbol{\eta}_{t+1}$$

where $\mathbf{a}_t \equiv [\mathbf{a}_1(t)' \ \ldots \ \mathbf{a}_J(t)']'$ and $\mathbf{M}$ is block diagonal with blocks $\mathbf{M}_j$, $j = 1, \ldots, J$, where J is number of wave modes.

ADE Spectral (cont.)

Hierarchical Spectral Spatio-Temporal Model

- Stage 1:

$$\mathbf{Z}_t = \mathbf{H}_t\mathbf{u}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma^2_\epsilon\mathbf{I})$$

- Stage 2: (truncated modes)

$$\mathbf{u}_t = \boldsymbol{\Phi}\mathbf{a}_t + \boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$$

- Stage 3:

$$\mathbf{a}_t = \mathbf{M}\mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$$

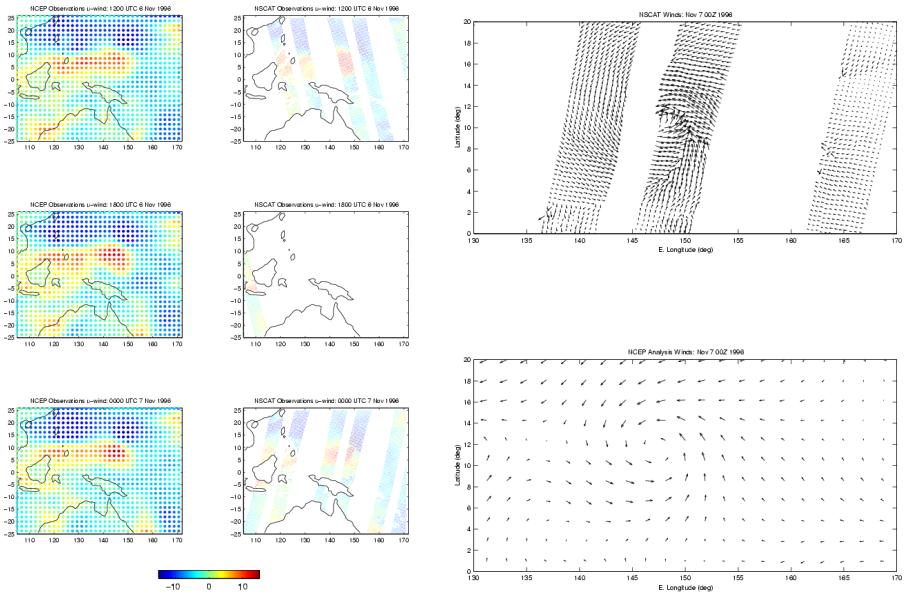
- Stage 4: (Other priors could be specified)

$$\text{vec}(\mathbf{M}_j) \sim N(\text{vec}(\mathbf{G}_i), \boldsymbol{\Sigma}_m)$$
$$\boldsymbol{\Sigma}_\eta^{-1}(j) \sim W((\nu_j S_j)^{-1}, \nu_j)$$
$$\sigma^2_\epsilon \sim IG(q_\epsilon, r_\epsilon)$$

$\boldsymbol{\Sigma}_\gamma(\theta)$is a stationary spatial cov matrix

Spectral Application: Tropical Ocean Winds

Problem: Blending tropical surface winds given high-resolution satellite scatterometer observations and low-resolution assimilated model output. (Wikle, Milliff, Nychka, Berliner, 2001; *JASA*)



Wind Problem: Process Model Motivation

Dynamics motivated by the linear shallow water equations (on an equatorial beta-plane):

$$\frac{\partial u}{\partial t} - \beta_0 yv + g\frac{\partial h}{\partial x} = 0$$

$$\frac{\partial v}{\partial t} + \beta_0 yu + g\frac{\partial h}{\partial y} = 0$$

$$\frac{\partial h}{\partial t} + \bar{h}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0$$

- u, v: east-west, north-south wind components at location (x,y)
- h: deviation of the fluid depth about its mean; $\bar{h}$
- β_0: constant related to rotation of earth
- g: gravitational acceleration

This linear system can be solved analytically, giving a series of traveling waves (*equatorial normal modes*) - **Observed in tropics!**

Wind Problem: Process/Parameter Model Motivation (cont.)

Empirical results suggest turbulent scaling behavior for near surface tropical winds (e.g., Wikle, Milliff and Large, 1999; *JAS*)

That is, there is multi-resolution/scale behavior to accommodate energy transfer across scales that are not directly accommodated by the large-scale equatorial wave dynamics.

We model this as an additional spectral process that has multiresolution dynamics.

Wind Model: Hierarchical Model Sketch (v-Component)

- **Data Model:** Change of support

$$[\mathbf{Z}_{st}' \;\; \mathbf{Z}_{at}']' = \mathbf{H}_t \mathbf{v}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$$

- **Process Model:**

$$\mathbf{v}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{a}_t + \boldsymbol{\Psi}\mathbf{b}_t,$$

$\boldsymbol{\mu}$ - mean vector reflecting "climatological" winds
$\boldsymbol{\Phi}$ - matrix containing "important" linear shallow-water normal mode basis functions (importance determined from empirical studies, e.g., Wheeler and Kiladis, 1999)
$\mathbf{a}_t$ - time-varying spectral coefficients
$\boldsymbol{\Psi}$ - matrix containing multiresolution (wavelet) basis functions (representing small/meso-scale variability)
$\mathbf{b}_t$ - time-varying multiresolution coefficients

Process Model (cont.)

$$\boldsymbol{\mu}|\boldsymbol{\beta}, \boldsymbol{\Sigma}_\mu \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\mu)$$

$\boldsymbol{\beta}$ - "regression" coefficients for climatological covariates $\mathbf{X}$

$$\mathbf{a}_t = \mathbf{M}_a(\boldsymbol{\theta})\mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$$

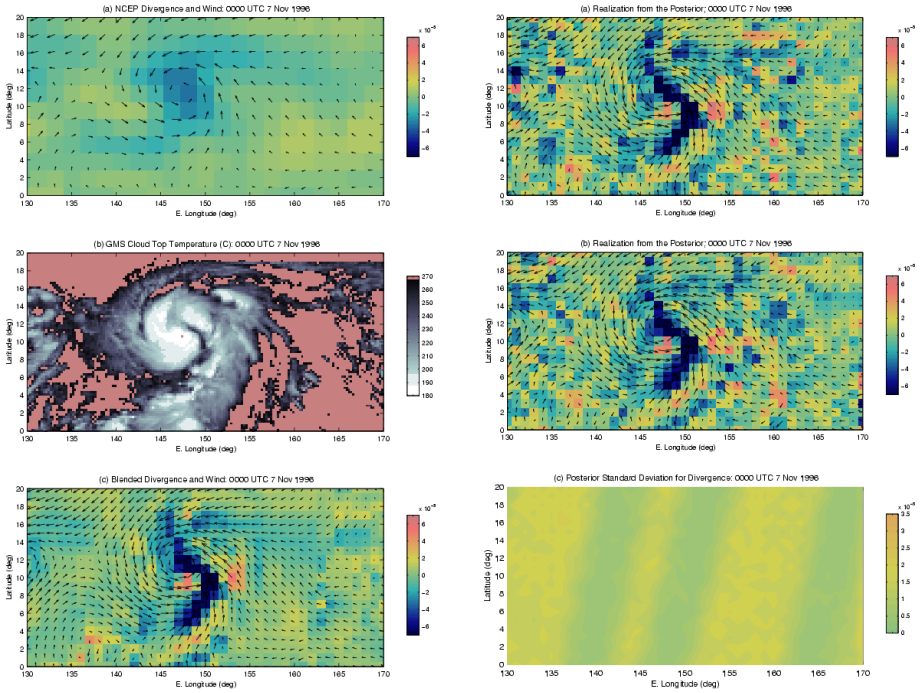
$\boldsymbol{\theta}$ - informative priors based on equatorial wave theory and empirical studies

$$\mathbf{b}_t = \mathbf{M}_b(\boldsymbol{\theta}_b)\mathbf{b}_{t-1} + \boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$$

$\mathbf{M}_b(\boldsymbol{\theta})$ - diagonal propagator (AR(1))
$\boldsymbol{\Sigma}_\gamma$ - diagonal, but informative priors to suggest empirical power-law scaling relationships

- **Parameter Models:**
  Parameter distributional choices reflect theoretical and empirical science related to equatorial normal modes and spectral scaling relationships.

Wind Problem Results: Tropical Cyclone Dale



A Word About Spatial Basis Expansion Models

In recent years, spatial basis-expansion models have become quite popular approaches to model spatial (and spatio-temporal) processes in high dimensions. **Why?** Induced sparseness, low-rank matrix inverses, non-stationarity, multi-resolution, etc.

In general, we can write (leaving off the time-index for now)

$$Y(\mathbf{s}) \approx \sum_{k=1}^{p} \phi_k(\mathbf{s})\alpha(k),$$

where $\{\phi_k(\mathbf{s}), \alpha(k), k = 1, \ldots, p\}$ are spatial basis functions and associated random expansion coefficients (weights), respectively.

For a set of spatial locations,

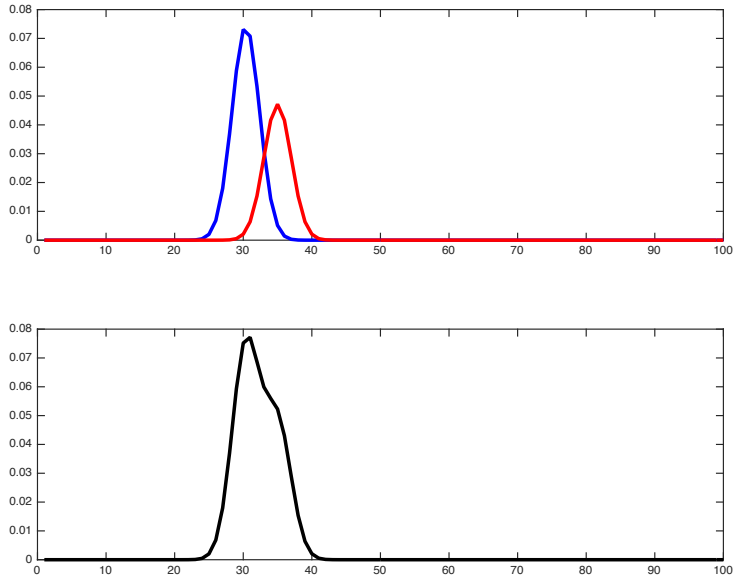
$$\mathbf{Y} \approx \mathbf{\Phi}\boldsymbol{\alpha}.$$

Basis Expansion Models (cont.)

- Many choices for basis functions: e.g., orthogonal polynomials, wavelets, splines, Wendland, Galerkin (finite element), EOFs, discrete kernel convolutions, "factor" loadings, "predictive processes", Moran's I bases, etc.
- Basis function decisions:
  - Fixed or "estimated" (parameterized)
  - reduced rank ($p << n$), complete ($p = n$), or overcomplete ($p > n$)
  - expansion coefficients in physical space or "spectral" space (see examples below)
  - discrete or continuous space
  - stationary or non-stationary
  - truncation error?
  - what kind of distribution on the random effects ($\boldsymbol{\alpha}$; iid, dependence via covariance matrix or precision matrix)

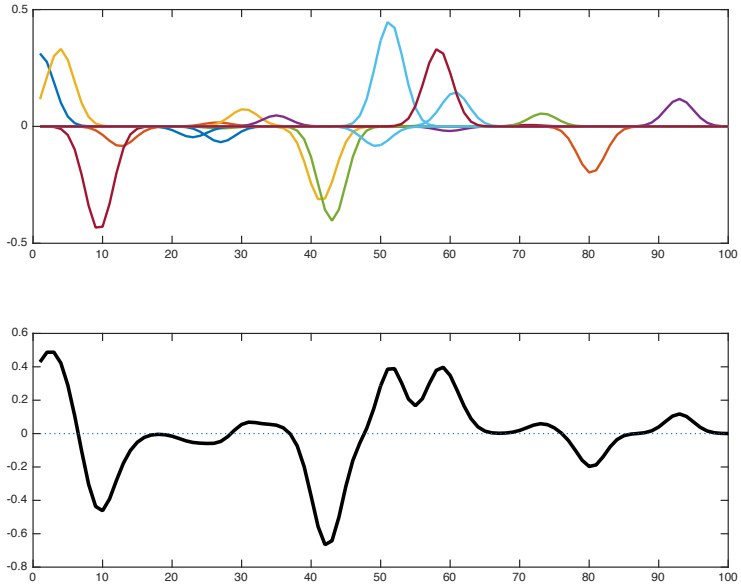
Basis Expansion Models (cont.)

Consider the case in 1-D space where the basis functions are "bumps" (e.g., Gaussian kernels); in this case, the random weights are spatially-referenced and give the height of the bumps (2 bumps)



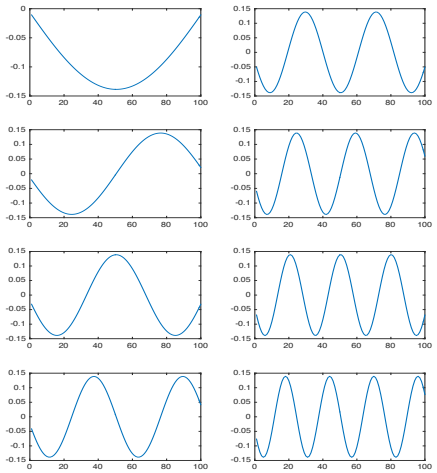
Basis Expansion Models (cont.)

Consider the case in 1-D space where the basis functions are "bumps" (e.g., Gaussian kernels); in this case, the random weights are spatially-referenced and give the height of the bumps (21 bumps)



Basis Expansion Models (cont.)

Now consider the case in 1-D space where the basis functions are curves; in this case, the random weights are not spatially-referenced but give the importance of the curves (8 curves)



Basis Expansion Models (cont.)

There is very little guidance on which bases to select!

- people have their favorites and there is a lot of dogma
- practical experience and recent research suggests that in most spatial cases, it probably doesn't matter much! (e.g., Bradley et al. 2014)
- in linear DSTM settings, the mechanistic process often suggests an appropriate basis set (e.g., the shallow-water normal modes)
- as will be discussed later, for nonlinear DSTMs there are important reasons why it **can** make a difference what basis functions are chosen

Also, note that in spatio-temporal processes, there may be good reason why the basis functions should be temporal and the random coefficients are spatial processes; e.g., $Y(\mathbf{s};t) = \sum_i \psi_i(t)\alpha_i(\mathbf{s})$.

An Introduction to Nonlinear DSTM Process Models

Christopher K. Wikle

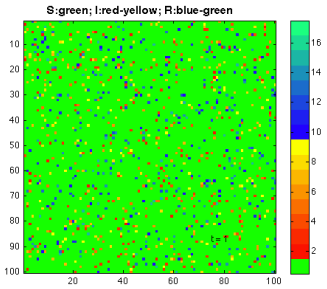
Nonlinear Spatio-Temporal Processes

Few environmental processes are linear (e.g., growth, nonlinear advection, density dependence, shock waves, repulsion, predator-prey, etc.)

- Nonlinear dynamical behavior arises from the complicated **interactions** across spatio-temporal scales of variability and interactions across multiple processes
- Examples in mechanistic models across many disciplines

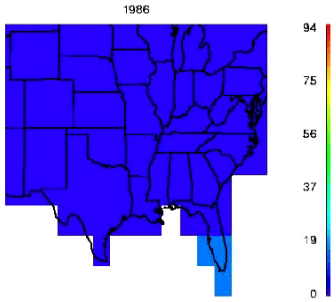
Nonlinear Processes

SIR Epidemic: simulated

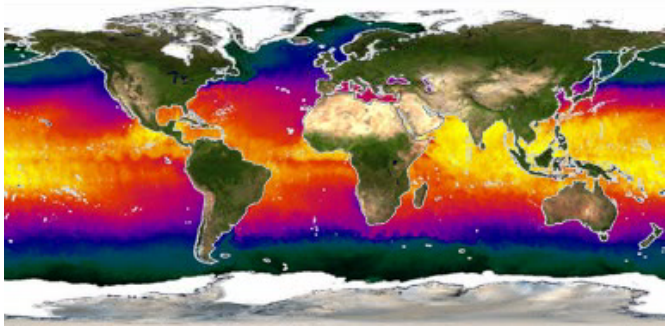


Susceptible
Infected
Recovered

Invasive Species: Spread of the Eurasian Collared Dove



NASA AMSR: Sea Surface Temperature
(scientific visualization studio)



Nonlinear* Spatio-Temporal Statistical Models

- Clearly, statistical 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 statistical 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, $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_m)$

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

Nonlinear DSTMs: Approaches

Over the last few years, several parametric approaches have been considered that have been useful for hierarchical formulations of nonlinear DSTMS.

- **Time-Varying Parameters**
  - E.g., Threshold models
- ***General Quadratic Nonlinearity (Class of Parametric Models)** (Wikle and Hooten, 2010; Wikle and Holan, 2011)
  - Mechanistically motivated structure
  - Shrinkage-based parameter estimation
  - Emulator-assisted priors
- **Individual (Agent)-Based Models**
  - Hierarchical stochastic cellular automata (Hooten and Wikle, 2010)

Nonlinearity Through Time-Varying Parameters

- Long-Lead Forecasting of Pacific Sea Surface Temperature (Berliner, Wikle, Cressie, 1999)
  - Dynamics on low-dimensional spectral coefficients
  - Linear propagator conditioned on current state and anticipated future state;
- Forecasting Waterfowl Settling Patterns (Wu et al. 2013)
  - See below
- Modeling the transition operator as a function of covariates in very high dimensions (Bradley et al. 2014)

Forecasting Waterfowl Settling Patterns

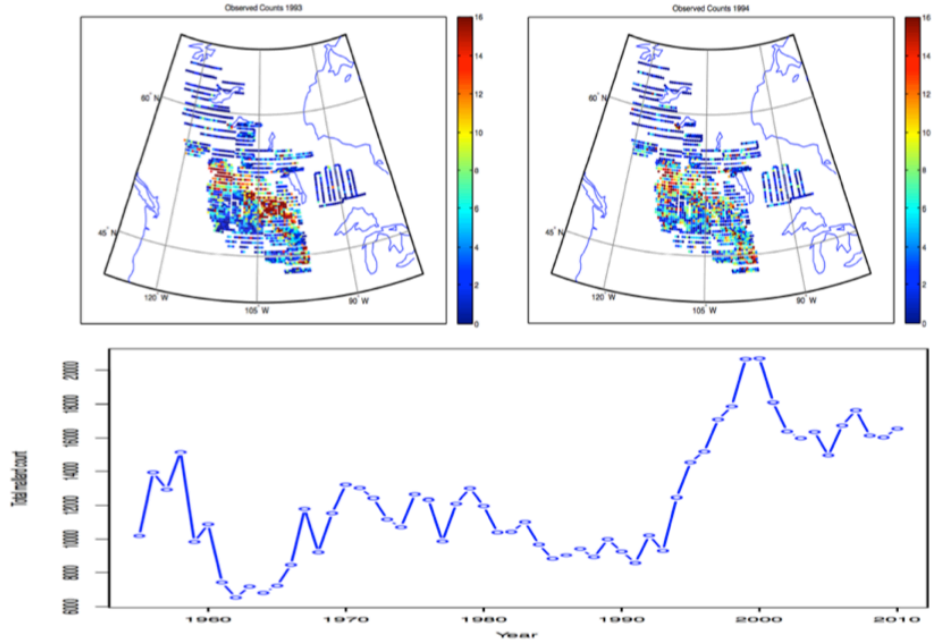
Biological Considerations:

- Variability in productivity across landscape
- Local population persistence ("site philopatry")
- Overflight ("flexible settling pattern") during years of poor home site conditions (e.g., Johnson and Grier, 1988)
- Can affect productivity and management decisions

Data

- Breeding Population Survey (BPS); 1955-2010
- US FWS, Canadian WS
- 2,171 segments; 29km long, 400m wide (treated as points)
- Pilot and observer count and speciate
- We consider Mallard raw indicated pair counts
- Detection probability < 1
- Relative abundance

Observed Counts: Spatial (1993-94); Total (1955-2010)



Hierarchical Model

- **Data Model:**

$$\mathbf{Z}_t|\mathbf{Y}_t,\nu_t \overset{ind.}{\sim} \text{CMP}(\mathbf{H}_t\mathbf{Y}_t,\nu_t),\ t=1,\ldots,T$$

Conway-Maxwell Poisson

- **Process Model:**

$$\log(\mathbf{Y}_t) = \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\gamma}_t,\ \ \boldsymbol{\gamma}_t \overset{iid}{\sim} \text{Gau}(\mathbf{0},\sigma^2_\gamma\mathbf{I})$$

  - Assumption: The CMP intensity process is governed by a dynamical process $\boldsymbol{\alpha}_t$ that exists on a low-dimensional manifold.
  - $\boldsymbol{\mu}$: spatially-referenced mean intensity
  - $\boldsymbol{\alpha}_t$: low-dimensional random dynamical process with dimension $p \ll n$ ($p = 6$ here after model selection)
  - $\boldsymbol{\Phi}$: spatial basis function matrix of dimension $n \times p$. (Kernel PCA Bases)
  - $\boldsymbol{\gamma}_t$: small-scale (assumed uncorrelated) spatio-temporal noise.

Threshold Dynamical Model (Conditionally Linear)

Based on the notion that settling pattern dynamics could change depending on habitat conditions, we consider the following (nonlinear) threshold autoregressive (TVAR) model for $\boldsymbol{\alpha}_t$ conditioned on a climate index:

$$\boldsymbol{\alpha}_t = \begin{cases} \mathbf{M}_1\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t & \text{if } c_t < d_L, \\ \mathbf{M}_2\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t & \text{if } d_L \le c_t \le d_U, \\ \mathbf{M}_3\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t & \text{if } c_t > d_U, \end{cases}$$

where $\boldsymbol{\eta}_t \overset{iid}{\sim} \text{Gau}(\mathbf{0},\boldsymbol{\Sigma}_\eta)$, and d_L, d_U are the threshold values that govern the switch from one climate regime to another.

In our case, c_t corresponds to the May value of the Palmer drought severity index (PDSI) over the Prairie Pothole Region of North America.

Parameter Models

- Spatial Mean:

$$\boldsymbol{\mu}|\boldsymbol{\beta}, \sigma_\mu^2 \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta}, \sigma_\mu^2 \mathbf{I}_n),$$

where $\mathbf{X}$ is a matrix of covariates with the i-th row being $\mathbf{x}_i' = [1, \text{lon}_i, \text{lat}_i, \text{lon}_i \times \text{lat}_i, \text{lat}_i^2, \text{lon}_i^2]$.

- Time-Varying Dispersion:

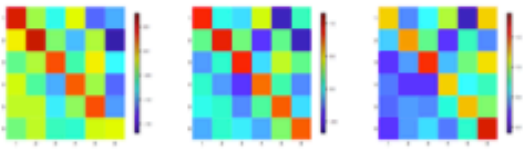
$$u_t = a_0 + a_1 u_{t-1} + \xi_t, \ t = 2, \ldots, T.$$

where $u_t = \log(\nu_t)$, $t = 2, 3, \ldots, T$, and $u_1 \sim \text{Unif}(u_l, u_h)$.

- Hyperparameters: vague

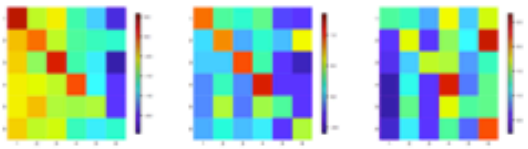
Posterior Summary: Threshold Transition Matrices

$\mathbf{H}_1$



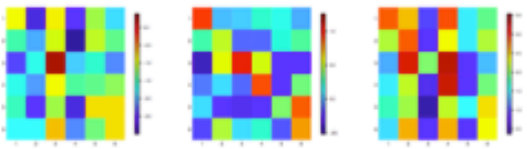
(a) $\mathbf{H}_1$: Lower CI (b) $\mathbf{H}_1$: Posterior Mean (c) $\mathbf{H}_1$: Upper CI

$\mathbf{H}_2$



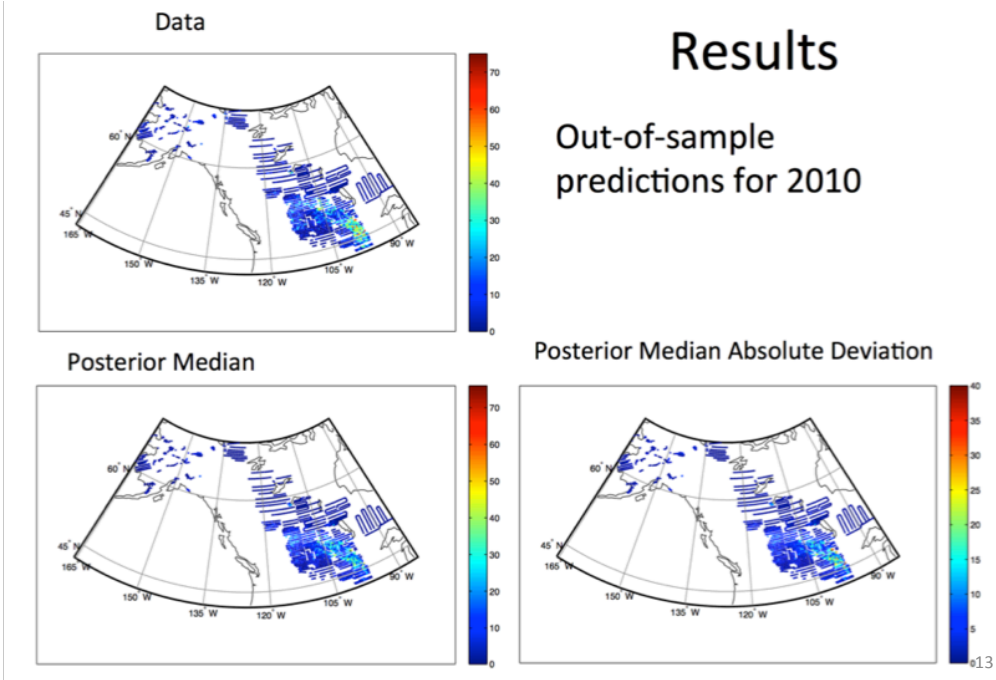
(d) $\mathbf{H}_2$: Lower CI (e) $\mathbf{H}_2$: Posterior Mean (f) $\mathbf{H}_2$: Upper CI

$\mathbf{H}_3$



(g) $\mathbf{H}_3$: Lower CI (h) $\mathbf{H}_3$: Posterior Mean (i) $\mathbf{H}_3$: Upper CI

Out of Sample Prediction



A Useful Class of Parametric Models for Nonlinear DSTMs

- Although the time-varying models are useful for many situations, they do not necessarily account for the types of nonlinearity that are present in a large number of environmental and ecological processes.
- Recall, we seek to parameterize the transition operator in: $\mathbf{Y}_t = \mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_m)\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t$
  - Consider the motivation from mechanistic PDEs

Examples: Nonlinearity

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

$$\frac{\partial Y}{\partial t} = \underbrace{\frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial Y}{\partial y}\right)}_{\text{(linear in Y)}} + \underbrace{\gamma_0(x,y)Y\exp\left(1-\frac{Y}{\gamma_1(x,y)}\right)}_{\text{(nonlinear in Y and a function of Y)}}$$

Epidemic Dynamics (SIR; Susceptible, Infected, Recovered)

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

(nonlinear in S, I interaction)

$$\frac{\partial I}{\partial t} = \beta SI - \mu I - \gamma I + D_I\nabla^2 I$$

$$\frac{\partial R}{\partial t} = \gamma I - \mu R - \omega R + D_R\nabla^2 R$$

Examples (cont.)
Simple Coupled Ecosystem

$$\frac{\partial Y_1}{\partial t} = \frac{NY_1}{k_s+N}e^z - R_m Y_2(1-e^{-\lambda Y_1}) - mY_1 + K_\nu\frac{\partial^2 Y_1}{\partial z^2}$$

$$\frac{\partial Y_2}{\partial t} = (1-\gamma)R_m\underbrace{Y_2(1-e^{-\lambda Y_1})}_{Y_2\, g(Y_1;\lambda)} - gY_2 + K_\nu\frac{\partial^2 Y_2}{\partial z^2}$$

(nonlinear interaction)

Simple 1-D biogeochemical model (Y_1: phytoplankton; Y_2: zooplankton; N - constant)

"Shallow Water Equations"

State variables: u,v – velocity in x and y directions; h – height deviation of surface
Constants: g – gravity, f – Coriolis, b – drag coefficient, H – mean surface height

Velocity in x:

$$\frac{\partial u}{\partial t} + \underbrace{u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}}_{\text{(Nonlinear advection terms)}} - fv = -g\frac{\partial h}{\partial x} - bu$$

Velocity in y:

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fv = -g\frac{\partial h}{\partial y} - bv$$

Surface height (deviation) :

$$\frac{\partial h}{\partial t} = \underbrace{-\frac{\partial}{\partial x}\left(u(H+h)\right) - \frac{\partial}{\partial y}\left(v(H+h)\right)}_{\text{(Nonlinear interactions, u and v with h)}}$$

Commonality?

What do all of these processes have in common?
Quadratic nonlinearity!

This suggests a class of useful statistical models for nonlinear DSTM processes:

General Quadratic Nonlinearity (GQN)

Statistical Parameterization: General Quadratic Nonlinearity (GQN)

(Wikle and Hooten, 2010)

In scalar form (discrete space),

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

for i=1,…,n.

- Model includes quadratic (dyadic) interactions in random process Y
- The term "general" refers to the term: $g(Y_{t-1}(s_l); \boldsymbol{\theta}_g)$

(Can also write in continuous space form and with higher level interactions; e.g., Wikle and Holan, 2011).

GQN: Matrix Notation

There are different ways to write this in matrix form. E.g.,

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

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

The n^2 x n matrix **B** is given by:

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

This is a special case of a state-dependent transition matrix; i.e.,

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

GQN: Curse of Dimensionality

- **Major Problem:** There are too many parameters ($O(n^3)$) to estimate in typical spatio-temporal applications without extra information!
- Minor Problem: potential for finite-time "blow-up" (Majda and Yuan, 2010)
- **We must reduce the state/parameter space:**
  - Mechanistically-motivated parameterizations (e.g., reaction-diffusion, invasive species, Wikle 2003)
  - Rank reduced functional "spectral" models
  - Parameter shrinkage (stochastic search variable selection, SSVS) (Wikle and Holan, 2011)
  - Emulator (surrogate)-based priors (Leeds et al. 2013)

Example: Invasive Species Prediction

Mechanistic Model Motivation: Most parameters are structurally 0! (i.e., inducing mechanistic sparseness)

Reaction-Diffusion Models: (e.g., density dependent); e.g.,

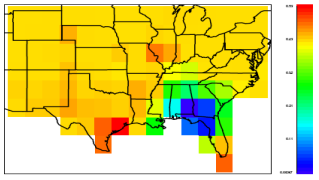
$$\frac{\partial Y}{\partial t} = \underbrace{\frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial Y}{\partial y}\right)}_{\text{diffusion portion suggests nearest neighbor structure in the linear operator}} + \underbrace{\gamma_0(x,y)Y\exp\left(1-\frac{Y}{\gamma_1(x,y)}\right)}_{\text{nonlinear ("reaction") portion suggests only nonlinear interactions at the same spatial location (but those vary by space)}}$$

Thus, only $O(n)$ parameters need be estimated. These are dependent spatially and can be modeled as spatial random fields at the next level of the hierarchical model; related to the diffusion and growth process parameters.

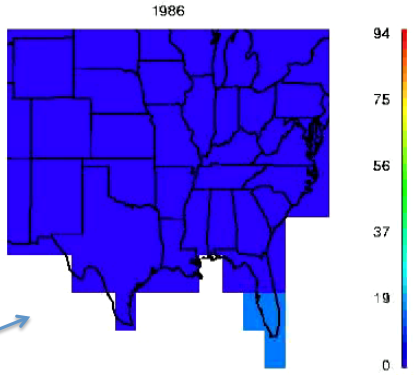
Example: Invasive Species Prediction

Eurasian Collared Dove Invasion: Introduced in S. Fl in 1980s

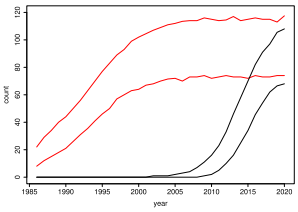
Note: data model is Poisson, conditioned on nonlinear spatio-temporal process with conditional Gaussian errors.



Dispersal δ: post mean



Relative abundance: post mean



95% PI's for S. Fl and Utah

Hooten and Wikle (2007)
Hooten et al. (2007)

GQN: Rank Reduction

In most situations, the motivating dynamics may not be known in detail or are much more complicated than suggested by a single mechanistic model, or are simply too high-dimensional.

As before, we can **consider the essential dynamics on a manifold of lower rank:**

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

Spatial basis functions (n x p)

Low-dimensional Dynamic process (p x 1)

GQN for Spectral Coefficients

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

Still we have order p^3 parameters here! Unless p is very small, we may need to make simplifying assumptions or utilize additional approaches to facilitate estimation (e.g., variable selection/shrinkage/sparcity).

***Critically, we have to pay more attention to the implications of our basis function and parameterization choices relative to the dynamic process.**

Special Considerations

There are important issues that should be considered for many physical and biological processes.

These relate to how information is transferred across spatio-temporal scales in nonlinear systems and has implications in terms of our choice of basis functions and nonlinear transition operators parameters. (Wikle 2015)

Illustration: Resonance

Consider a simple nonlinear advection equation (i.e., an inviscid Burgers equation) for state variable, u(x,t) in one spatial dimension (x):

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0$$

Seek a solution as an expansion of the state variable in terms of (spatial) sinusoidal basis functions:

$$u(x,t) = \sum_k \alpha_k(t)\sin(\omega_k x)$$

where ω_k is the "spatial frequency" and k = 1,2,… corresponds to the spatial "wave number" (number of complete sine waves over the spatial domain).

Note: this is just $\mathbf{u}_t = \mathbf{\Phi}\,\boldsymbol{\alpha}_t$ with the spatial basis functions in $\mathbf{\Phi}$ as sine functions.

Resonance Illustration (cont.)

Now, consider the nonlinear term after substitution of this spectral representation for u(x,t):

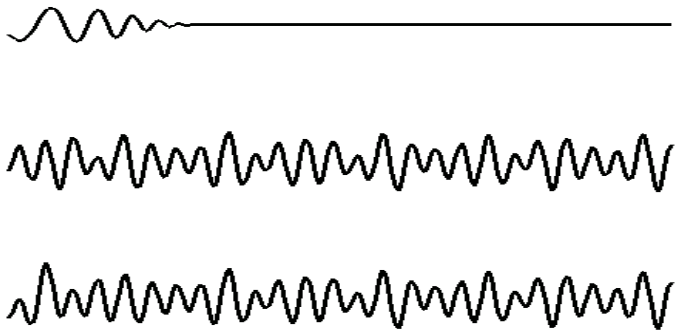
$$\begin{aligned} u\frac{\partial u}{\partial t} &= \sum_{k_1}\alpha_{k_1}(t)\sin(\omega_{k_1}x)\sum_{k_2}\omega_{k_2}\alpha_{k_2}(t)\cos(\omega_{k_2}x) \\ &= \sum_{k_1}\sum_{k_2}\omega_{k_2}\alpha_{k_1}(t)\alpha_{k_2}(t)\sin(\omega_{k_1}x)\cos(\omega_{k_2}x) \\ &= \sum_{k_1}\sum_{k_2}\omega_{k_2}\alpha_{k_1}(t)\alpha_{k_2}(t)(\sin\{(\omega_{k_1}+\omega_{k_2})x\} + \sin\{(\omega_{k_1}-\omega_{k_2})x\} \end{aligned}$$

Thus, nonlinear interactions of spatial patterns with frequencies ω_{k1} and ω_{k2} can excite coefficients associated with spatial patterns of frequencies $\omega_{k1} + \omega_{k2}$ and $\omega_{k1} - \omega_{k2}$.

There are certain conditions when such resonance happens.

Illustration of Resonance

The two propagating 1-d oscillations below excite the propagating oscillation above.



Source: http://www.theartofdredging.com/

Why is this important?

Natural Modes of Variability:

In physical/environmental systems, there are typically "normal modes" of variability

In biological systems, there are often modes of "spatial synchrony"

When the one of these natural modes correspond to $\omega_{k1} + \omega_{k2}$ or $\omega_{k1} - \omega_{k2}$ then there is often nonlinear growth and/or significant information exchange across spatial scales.

Why is this important? (cont.)

Basis Functions: As mentioned, there are many "fashionable" basis functions used in spatial statistics. These imply either:

- α_t is spatially referenced (e.g., Φ is based in some sense on "knots" as with certain splines, kernel convolutions, predictive processes, etc.)
- α_t is not spatially referenced (e.g., Φ corresponds to known spatial basis functions such as orthogonal polynomials, wavelets, empirical orthogonal functions, etc.)

**** The choice should be able to easily allow multiple spatial scales to interact**

Why is this important? (cont.)

Many scales interacting for real world processes:

- Difficult to know *a priori* which ones will be interacting unless there is strong knowledge of the mechanistic process
- Suggests that we should allow flexibility and utilize shrinkage priors (with the possibility of additional scientific information)
  - Wikle and Holan (2011) use stochastic search variable selection.
  - Gladish and Wikle (2014) use information concerning energy propagation such that medium scale structures influence larger scales, but not vice-versa

Additional Dimension Reduction

- When fitting high-dimensional DSTMs (e.g., GQN models) to data or model output, it is important to regularize the estimation by shrinking the parameter space towards zero.
- It also helps with prediction by implicit model averaging.
- We have found that stochastic search variable selection (SSVS) works well in this context (but other shrinkage approaches could be used).

Hierarchical Stochastic Search Variable Selection

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

Even after basis expansion, 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,

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

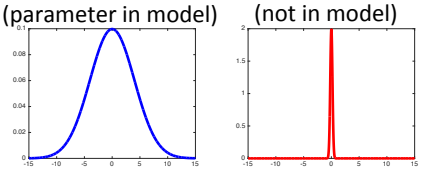
$$b_j|\xi_j \sim \xi_j N(0, c_j^2\tau_j^2) + (1-\xi_j)N(0, \tau_j^2)$$

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

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

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

(parameter in model) (not in model)



i.e., "mixture of normals" prior

Hierarchical Nonlinear DSTM: Summary

Multi-level GLMM w/ **latent conditional Gaussian process dynamical process**

$$\mathbf{Z}_t|\mathbf{Y}_t,\boldsymbol{\theta}_h \sim \mathcal{D}(\mathbf{H}_t\mathbf{Y}_t;\boldsymbol{\theta}_h)$$

$$\mathbf{Y}_t|\boldsymbol{\mu},\boldsymbol{\alpha}_t,\boldsymbol{\beta}_t,\boldsymbol{\Sigma}_\gamma \sim Gau(\boldsymbol{\mu}+\boldsymbol{\Phi}\boldsymbol{\alpha}_t+\boldsymbol{\Psi}\boldsymbol{\beta}_t,\boldsymbol{\Sigma}_\gamma)$$

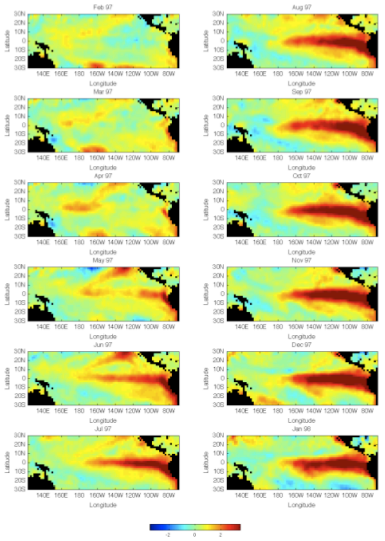
$$\alpha_t|\alpha_{t-1},\theta_A,\theta_B,\theta_g,\Sigma_\eta \sim Gau(\mathrm{A}(\theta_A)\alpha_{t-1}+(\mathrm{I}_p\otimes g(\alpha_{t-1};\theta_g)')\mathrm{B}(\theta_B)\alpha_{t-1},\Sigma_\eta)$$

$$[\boldsymbol{\theta}_A,\boldsymbol{\theta}_B|\boldsymbol{\xi}]$$ (SSVS priors)

$$[\boldsymbol{\theta}_h,\boldsymbol{\theta}_g,\boldsymbol{\xi},\boldsymbol{\pi},\boldsymbol{\Sigma}_\gamma,\boldsymbol{\Sigma}_\eta]$$ (hyperparameters – problem specific)

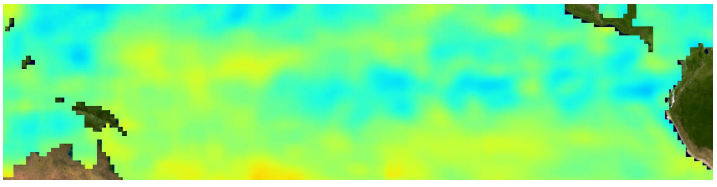
Example: Long-Lead Prediction of Tropical Pacific SST

Given SST up to March 1997



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

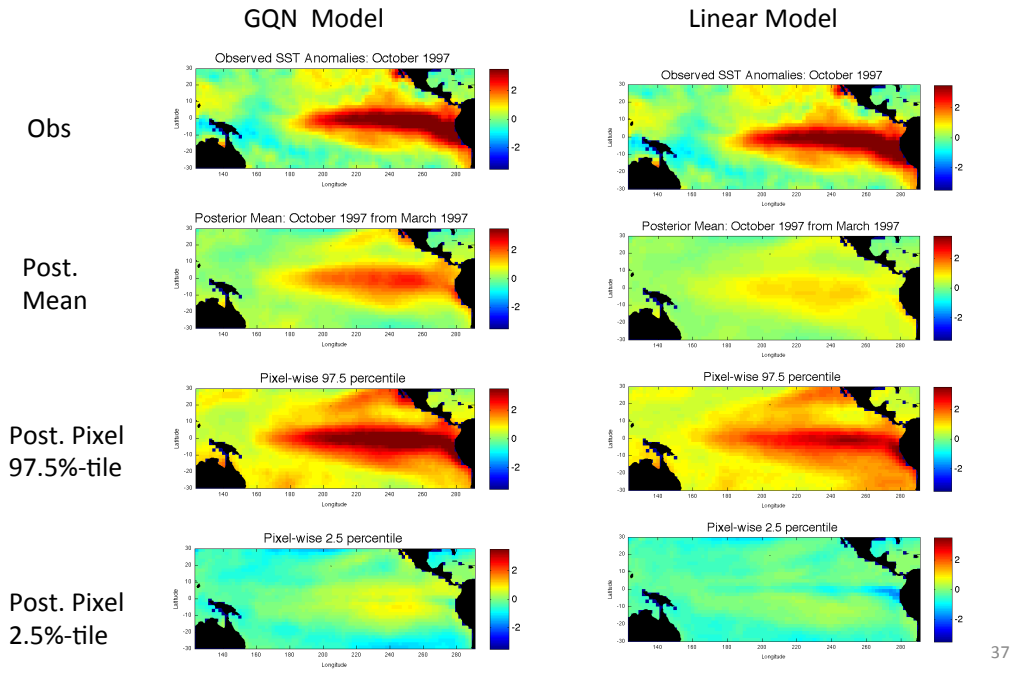
Forecast SST 7 months later in Oct 1997



Let **Φ** be spatial (functional) principal components (EOFs), p=10.

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

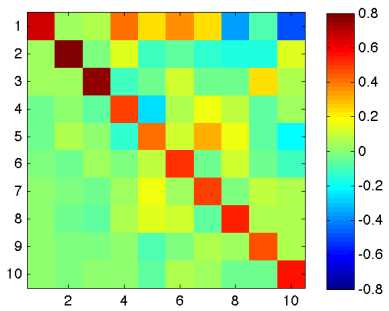
Forecast: October 1997 from March 1997



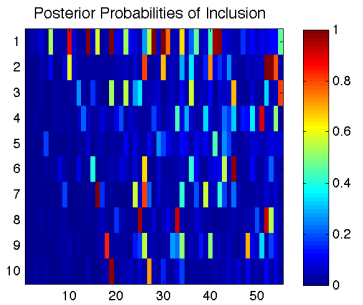
Posterior Means: Parameters

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

Posterior Mean: A matrix (linear term)

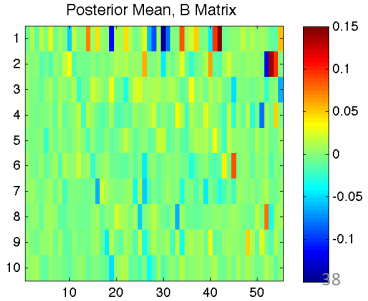


B matrix inclusion probabilities



Interpretation?

B matrix



Mechanistic Models: Emulator-Based Prior Elicitation

- In some cases, mechanistic computer models may be available for processes
- We may wish to use this information – particularly for data assimilation applications
  - One way to use this information is to emulate the dynamics of this mechanistic model with a statistical surrogate
  - **It is important that this model respect the potential for nonlinear interactions as described previously**
  - If this surrogate model is reduced rank GQN; the mechanistic model provides informative priors

Statistical Emulators (Surrogates)

- With very complex nonlinear processes for which there exists large (typically, mechanistic) simulators ("deterministic 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" that are based on Gaussian processes with the emphasis on covariances 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 locally 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.

Reduced-Rank First-Order Emulator (e.g.,Hooten et al. 2011)

- Generate inputs: $\mathbf{W}_{q\times K} = (\mathbf{w}_1, \ldots, \mathbf{w}_K)$ — Inputs can be parameters or previous values of process
- Generate vector realizations from computer model output given the inputs, $\mathbf{y}_i = f(\mathbf{w}_i)$: $\mathbf{Y}_{n_y \times K} \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_K)$
- Consider the SVD of the computer model output:

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

- 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: $\tilde{\mathbf{V}} \sim h(\mathbf{W}, \boldsymbol{\theta})$
- Thus, for an n_y-dim response, $\mathbf{y}^*$, and input, $\mathbf{w}^*$:

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

First-Order Emulator (process)

- Typically the "inputs" **w** correspond to parameters in the mechanistic model (e.g., calibration).
- 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.,

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

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

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

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

- **Critically, this can be the basis for a prior distribution on the parameters within a hierarchical nonlinear DSTM; also suggests form of basis functions, Φ, ψ, from SVD.**

Hierarchical Stochastic Search Variable Selection

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

Even after basis expansion, there are still likely to be too many parameters in $\boldsymbol{\theta}_m$ to get reliable statistical estimates in GQN. Again, we can utilize the hierarchical framework to help. Let,

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

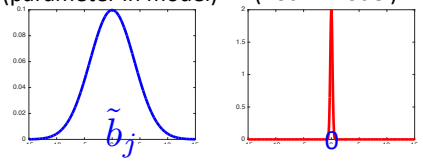
$\tilde{b}_j$ from emulator or 0

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

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

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

(parameter in model) (not in model)



i.e., "mixture of normals" prior; use the prior information from the emulator

Hierarchical Reduced-Rank Emulator-Assisted DSTM

Leeds et al. (2013)

Data Model: (conditional independence; need not be Gaussian)

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

Process Model:

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

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

* m() quadratic nonlinear model

$$\boldsymbol{\beta}_t \sim Gau(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_\beta))$$

Parameter Models:

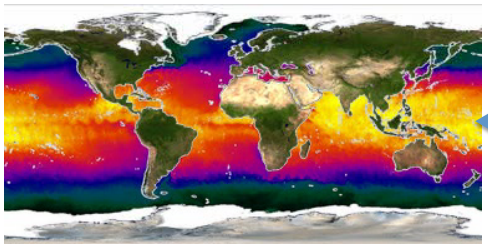
$$[\boldsymbol{\theta}_R], [\boldsymbol{\theta}_Q], [\boldsymbol{\theta}_m|\boldsymbol{\gamma}], [\boldsymbol{\theta}_\beta] \quad [\boldsymbol{\gamma}|\boldsymbol{\pi}]$$

Key Point: Prior distributions based on emulator fits to mechanistic model and/or stochastic search mixture priors.
Important: parameters can still "learn" from data via posterior.

Ocean Biogeophysical Coupling

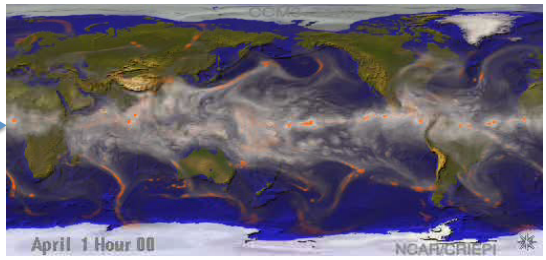
(Sea Surface Temperature, SST)

NASA AMSR: SST (scientific visualization studio)

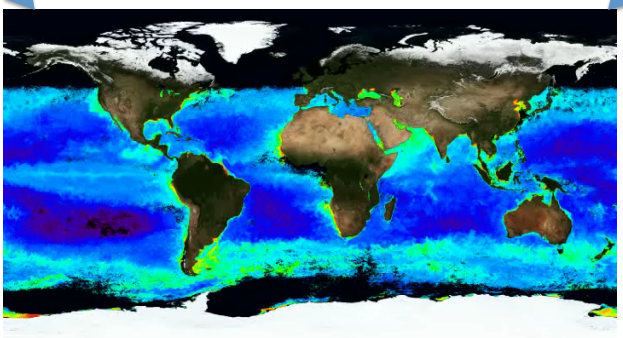


(Clouds/Precipitation)

NCAR CCM3 Cloud/Precipitation Simulation (NCAR VETS)



NASA SEAWIFS: Ocean Color (scientific visualization studio)



Proxy for ocean primary production (i.e., chlorophyll/ phytoplankton)

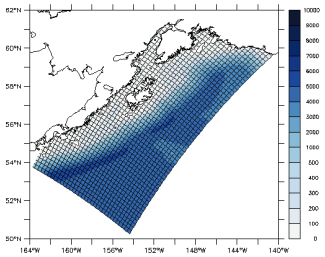
Interaction across atmosphere, ocean, and ecosystem.

Where is the uncertainty?

Ocean Color Observations

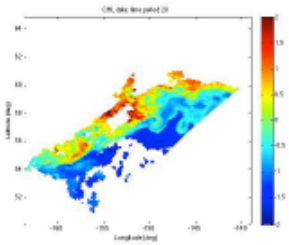
(ocean color: surrogate for phytoplankton)

Coastal Gulf of Alaska

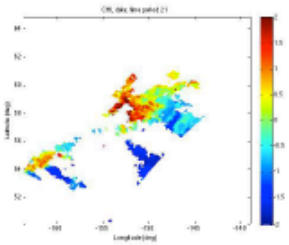


SeaWiFS Ocean Color Satellite Observations (8 day averages)

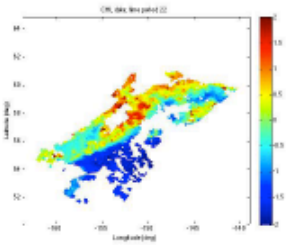
5/19/02 - 5/26/02



5/27/02 - 6/03/02



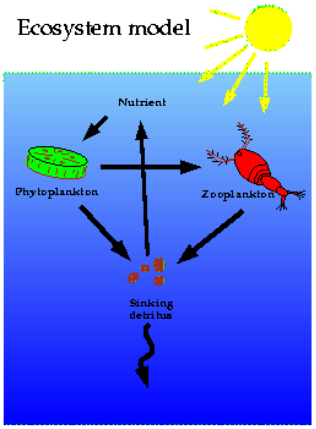
6/04/02 - 6/11/02



"Gappy" and substantial measurement uncertainty!
We seek to predict at missing locations and filter obs error.

Complicated Model Components

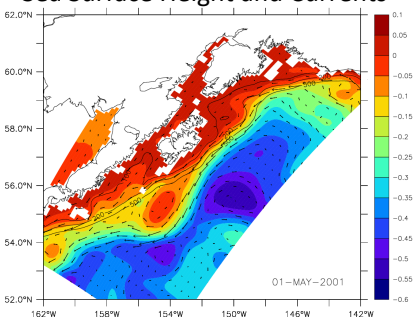
- **Lower Trophic Ecosystem**
  - Essentially a complicated multicomponent predator-prey system influenced by the environment (highly nonlinear)
- **Physical Ocean**
  - Navier-Stokes fluid dynamic process across multiple state variables (highly nonlinear)



Coupled!

(nonlinear)

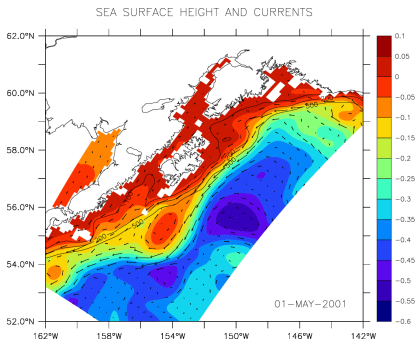
Sea Surface Height and Currents



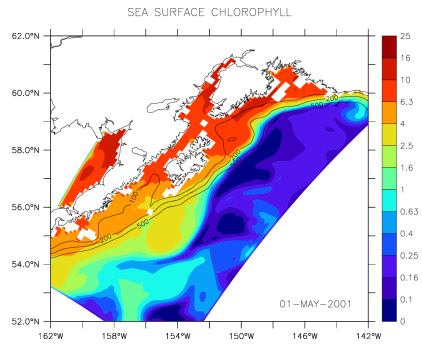
The process of interest is multivariate, nonlinear and spatio-temporal.

Physical-Biological Interface

Sample output from a coupled ocean-ecology model in the coastal Gulf of Alaska for May 1, 2001 (Fiechter et al. 2008) (A Deterministic Model)



Sea surface height (SSH) and currents



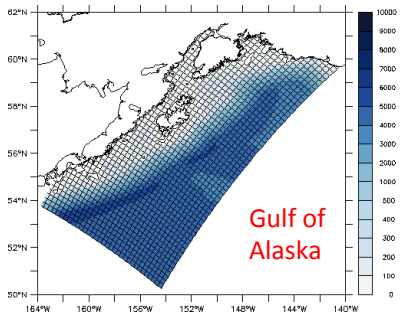
Chlorophyll concentration and bathymetry

Strong association!

Note: we can learn about the biology by knowing something about the physics!

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

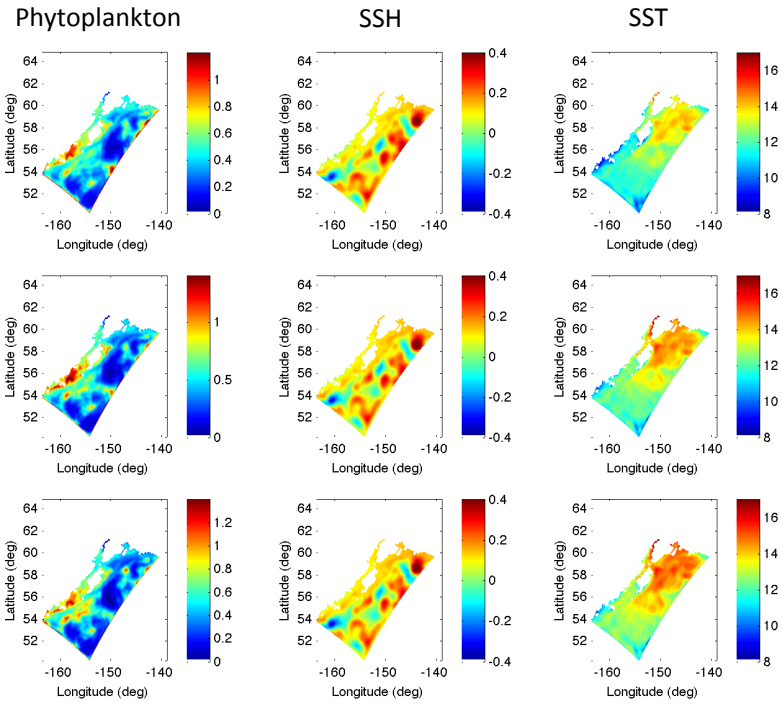
•Predict/Assimilate for 2002

•MCMC implementation

Coupled Dynamics: Example from Coupled Ocean Model

Example training data

Time: consecutive 8-day periods



Proof of Concept Experiment

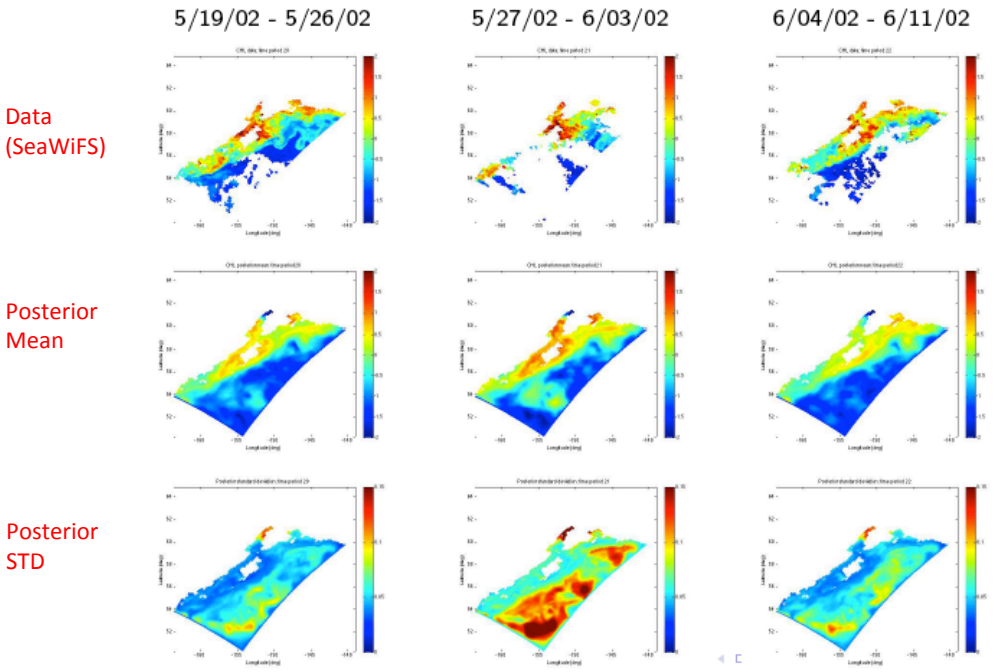
- In this case: $\mathbf{Z}_t = \begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \mathbf{Z}_{3,t} \end{pmatrix}$ $m_{i,t}(i = 1, 2, 3)$ - dimensional data vectors for Chlorophyll, SSH, SST

$\mathbf{Y}_t = \begin{pmatrix} \mathbf{Y}_{1,t} \\ \mathbf{Y}_{2,t} \\ \mathbf{Y}_{3,t} \end{pmatrix}$ $p_i(i = 1, 2, 3)$ - 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)

Results: log(CHL)



Alternatives?

- Methods described above are flexible and can accommodate many different types of dynamical processes
- Alternatives include:
  - Agent-based/individual-based DSTMs
  - Functional-based DSTMs (especially for high-frequency observations)
- Major challenges:
  - Multivariate processes
  - Handling very large data sets

Brief Word on Computation

- There are a growing number of R-packages that one can use to do spatio-temporal modeling in statistics: e.g.,
  - *SpatioTemporal, spTimer, spBayes, gstat, R-INLA, spate*, etc.
  - None can do the full general hierarchical DSTM models or nonlinear models I have talked about here!
- Regardless of whether you need such generality, you should program your own solution (MCMC) at least once! (R, Matlab, WinBUGS, JAGS, STAN, etc.)
  - See Wikle and Hooten (2006) for a complete example of the Eurasian Collared Dove solution
  - Be wary of packages that you can't program yourself! Do you really know what it is doing? When does it fail?

Learning Points Redux

- Marginal vs Dynamic Models
  - Rather than use in appropriate covariance functions or SPDEs, motivate conditional dynamical models by mechanistic processes and let the data help
- Key to Dynamical/Hierarchical Modeling
  - Place structure on the conditional mean
  - Least amount of structure possible in 2nd moment
  - Let random parameters have structure at the next level
- Don't Ignore the Nonlinearity!
  - Use the science to motivate the interactions
  - Use parameter shrinkage and informative priors to help with estimation

THANK YOU!!

If you have any questions or would like me to send a specific reference, please feel free to contact me at:

wiklec@missouri.edu