Challenges and Recent Advances in Spatio-Temporal Statistical Modeling: An Overview

Christopher K. Wikle

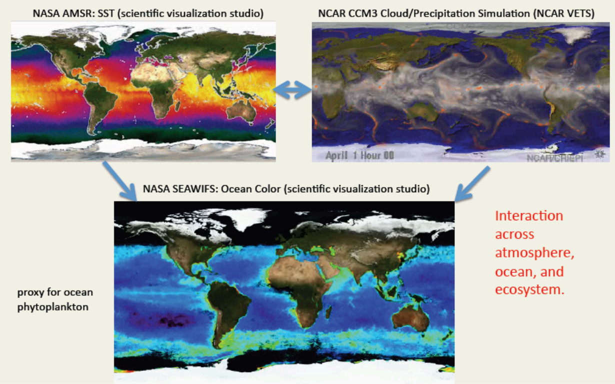
Department of Statistics

University of Missouri

CAPAM: February 27, 2018

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.



Spatio-Temporal Modeling: Challenges

Data:

- some data sets very large (e.g., satellite; model output) and some very small (e.g., *in situ* measurements of zooplankton)
- multiple data sources
- data at different supports in time and space (point measurements, aggregations, satellite footprints, model grid cells; daily data vs monthly data; high frequency data – data storage tags, telemetry)
- unequal and preferential sampling
- data of mixed types (counts, proportions, normal, presence/absence)

Spatio-Temporal Modeling: Challenges (cont.)

Process:

- non-stationary, non-separable, nonlinear
- extremes
- multiple processes (multivariate), potentially at different space and time scales
- scientific realism
- parsimonious parameterizations

Other Important Issues:

- computation
- principled uncertainty quantification
- sampling design
- model-assisted decisions

Spatio-Temporal Modeling: Solutions (Common Themes)

This talk will give an overview of statistical modeling of spatio-temporal processes and emphasize:

- science-motivated (dynamical) models
- Bayesian *hierarchical models* (**the power of conditioning**; deep models)
- parameters as processes
- basis function (functional) representations
- regularization
- emulation/surrogate models: "pre-training" prior elicitation
- change-of-support/high-frequency covariates (if time allows)

Spatio-Temporal Processes and Data

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

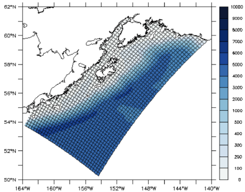
- Prediction in space (smoothing, interpolation)
- Prediction in time (forecasting)
- Assimilation of observations with deterministic models
- Inference on parameters that explain the etiology of the spatio-temporal process
- Design and adaptation of monitoring networks

Motivating Example: Assimilating Ocean Color Observations

(Leeds et al. 2014)

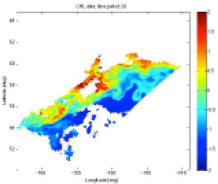
(ocean color: surrogate for phytoplankton)

Coastal Gulf of Alaska

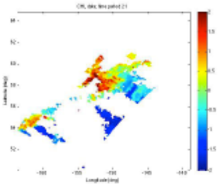


SeaWiFS Ocean Color Satellite Observations (8 day averages)

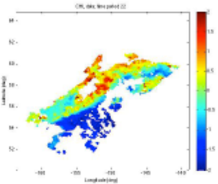
5/19/02 - 5/26/02



5/27/02 - 6/03/02



6/04/02 - 6/11/02



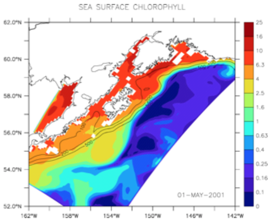
"Gappy" and substantial measurement uncertainty!

We seek to predict at missing locations and filter obs error.

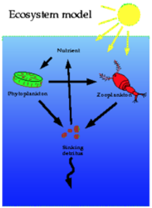
Example: Physical/Biological Interface

- **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)



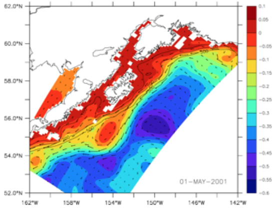
Chlorophyll concentration and bathymetry



(nonlinear)

Coupled!

Sea Surface Height and Currents



Spatio-Temporal (S-T) Models In Statistics

In statistics, historically we are concerned with spatio-temporal models of the form: (simple representation)

$$\begin{aligned} \textit{observations} &= \textit{true process} \;\; \textit{with} \;\; \textit{observation/sampling error} \\ \textit{true (latent) process} &= \text{"}\textit{fixed effects}\text{"} \; + \; \textit{dependent random process} \end{aligned}$$

Challenges:

- data model (likelihood) that represents the generating process
- model for the dependent random process
  - often represents unknown covariates
  - can represent some other scientific process
- computational tractability

Note: dependent processes can also be useful for components of the data model!

Statistical Models for Spatio-Temporal Random Processes

- **Markov random fields (MRFs):** space-time version of auto-Gaussian (CAR), auto-logistic, auto-Poisson, etc.
- **Latent Gaussian Processes:** e.g., generalized linear mixed models (GLMMs) with latent Gaussian S-T processes
- **Latent Conditional Gaussian Processes:** as with the GLMM but with process that has nonlinear evolution and Gaussian errors
- Other "non-traditional" models: e.g.,
    - **Agent (Individual)-Based Models**
    - **Analog ("mechanism free") Models**
    - **Recurrent Neural Network Models**

Example: GLMM with Latent Gaussian S-T Process

Data model: (data: $\{z(\mathbf{s};t)\}$; process: $\{Y(\mathbf{s};t)\}$)

$$z(\mathbf{s};t)|Y(\mathbf{s};t),\gamma \ \sim \ ind.\ EF(Y(\mathbf{s};t),\gamma),$$

where EF is a distribution from the exponential family with scale parameter γ and mean $E(z(\mathbf{s};t)|\gamma) = Y(\mathbf{s};t)$.

Then, we consider a transformation of the mean response modeled in terms of fixed effects and random processes:

$$g(Y(\mathbf{s};t)) = \mathbf{x}(\mathbf{s};t)'\boldsymbol{\beta} + \delta(\mathbf{s}) + \xi(t) + \nu(\mathbf{s};t),$$

For this talk: focus on $\nu(\mathbf{s};t)$, a spatio-temporal (S-T) Gaussian process (GP) or random field.

Models for the Spatio-Temporal Gaussian Process

The dependent S-T random process can be considered from two perspectives:

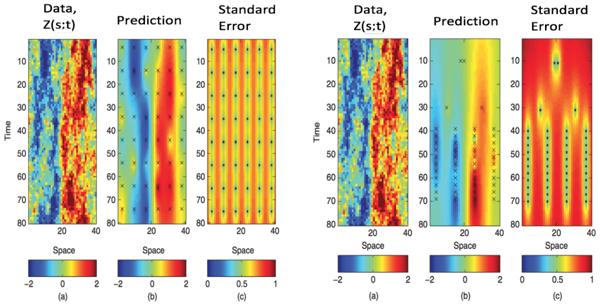
(1) **Descriptive (Marignal) Perspective:** Characterize the first- and second-moment behavior of the process (e.g., kriging)

(2) **Dynamical (Conditional) Perspective:** Spatial process evolves in time (e.g., linear and nonlinear Markov models)

Spatio-Temporal Processes: Descriptive Methods

Descriptive methods: require a valid covariance function for the process $\nu(\mathbf{s}; t)$ at <u>any</u> two locations in space and time!

With that, one can perform optimal prediction and/or account for residual spatio-temporal prediction. E.g., consider simulation example (observation locations given by "x")



Spatio-Temporal Processes: Descriptive Methods

Descriptive S-T models are powerful, but there are issues that make them problematic for many real-world spatio-temporal processes.

- **Dimensionality:** prediction models and/or likelihoods require matrix inverses; can be overcome by various methods:
  - basis function or kernel representations (reduced rank, over complete, full rank)
  - neighbor-based methods
  - covariance tapering
  - moving to a Markov random field framework (precision matrix)
- **Realistic Dependence:** most S-T processes are more complex than can be described by the limited classes of valid S-T covariance functions (non-separable; non-stationary)
  - basis function parameterizations can help here as well
  - still can be problematic for multivariate and nonlinear processes

Basis Function Random Effects Representations for $\nu(\mathbf{s}; t)$

Flexibility through marginalization: usually fixed basis functions and random coefficients. Basis functions can be estimated (factor models) but limits structure that one can place on the coefficients.

- Spatio-Temporal Basis Functions: random effects $\{\alpha_k\}$

$$g(Y(\mathbf{s}; t)) = \mathbf{x}(\mathbf{s}; t)'\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(\mathbf{s}; t)\alpha_k,$$

Basis Function Random Effects Representations for $\nu(\mathbf{s}; t)$

Flexibility through marginalization: usually fixed basis functions and random coefficients. Basis functions can be estimated (factor models) but limits structure that one can place on the coefficients.

- Spatio-Temporal Basis Functions: random effects $\{\alpha_k\}$

$$g(Y(\mathbf{s}; t)) = \mathbf{x}(\mathbf{s}; t)'\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(\mathbf{s}; t)\alpha_k,$$

- Temporal Basis Functions: random effects $\{\alpha_k(\mathbf{s})\}$

$$g(Y(\mathbf{s}; t)) = \mathbf{x}(\mathbf{s}; t)'\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(t)\alpha_k(\mathbf{s}),$$

Basis Function Random Effects Representations for $\nu(\mathbf{s}; t)$

Flexibility through marginalization: usually fixed basis functions and random coefficients. Basis functions can be estimated (factor models) but limits structure that one can place on the coefficients.

- Spatio-Temporal Basis Functions: random effects $\{\alpha_k\}$

$$g(Y(\mathbf{s};t)) = \mathbf{x}(\mathbf{s};t)'\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(\mathbf{s};t)\alpha_k,$$

- Temporal Basis Functions: random effects $\{\alpha_k(\mathbf{s})\}$

$$g(Y(\mathbf{s};t)) = \mathbf{x}(\mathbf{s};t)'\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(t)\alpha_k(\mathbf{s}),$$

- Spatial Basis Functions: random effects $\{\alpha_k(t)\}$

$$g(Y(\mathbf{s};t)) = \mathbf{x}(\mathbf{s};t)'\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(\mathbf{s})\alpha_k(t),$$

Potential issue: confounding

Bayesian Hierarchical Models

Denote the 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**):

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 hierarchical modeling paradigm (*which is often lost on statisticians who equate hierarchical modeling with Bayesian modeling*):

- avoid covariance models as much as possible
- put as much structure as possible in the conditional mean via random effects

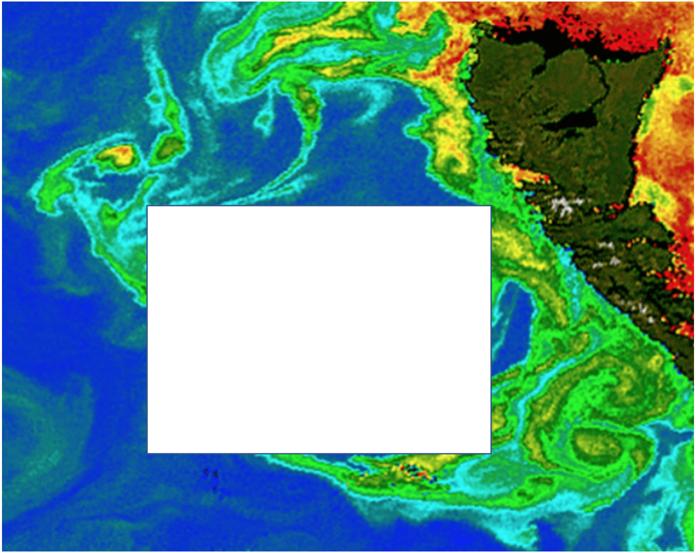
That is, push the dependence down a level of the hierarchy to the mean to **build** dependence through marginalization;

- first moments are much easier to model and there is much more scientific knowledge about their specification!

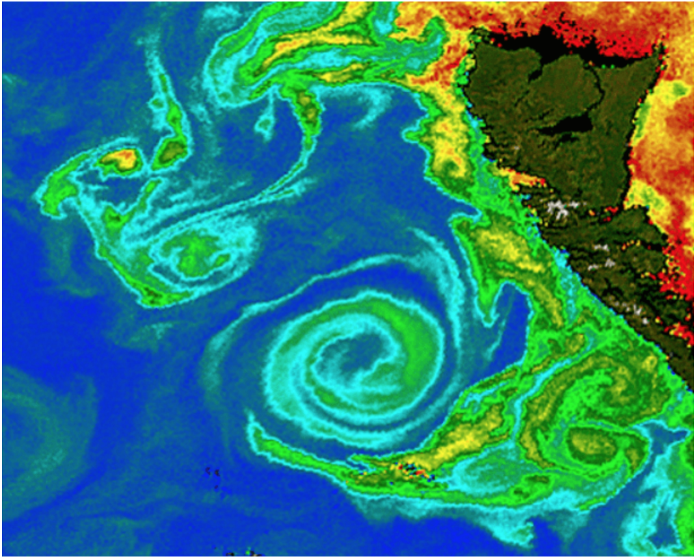
These are "deep" models! (Paves the way for dynamical models)

Linear Dynamical Spatio-Temporal Models (DSTMs)

Dynamics: Why?



Dynamics: Why?



Dynamical Spatio-Temporal Models (DSTMs)

There are two critical assumptions for DSTMs: (state-space formulation)

- Data conditioned on the latent process can be factored into the product of independent distributions; e.g.,

$$[\mathbf{z}_T, \ldots, \mathbf{z}_1 | \mathbf{Y}_T, \ldots, \mathbf{Y}_1, \boldsymbol{\theta}_d] = \prod_{t=1}^{T} [\mathbf{z}_t | \mathbf{Y}_t, \boldsymbol{\theta}_d]$$

- The joint distribution of the latent process can be factored into conditional (in time) models; e.g.,

$$[\mathbf{Y}_T, \ldots, \mathbf{Y}_1, \mathbf{Y}_I | \boldsymbol{\theta}_p] = \prod_{t=1}^{T} [\mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots; \boldsymbol{\theta}_p][\mathbf{Y}_I | \boldsymbol{\theta}_p]$$

Challenge: specification of the models associated with these component distributions, particularly: $[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots; \boldsymbol{\theta}_p]$

Linear DSTM: Process Model

In the linear process case, the conditional distribution $[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_p]$ implies a first-order Markov model of the form:

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

These equations imply a matrix model (a VAR(1) model):

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

However, the difficulty here is dimensionality! It can be difficult to get stable estimates of the parameters $\{m_{ij}, \ \ i, j = 1, \ldots, n\}$ in the spatio-temporal case (requires $T \gg n$).

We must parameterize $\mathbf{M}$ in these settings! (how?)

Linear DSTM: Efficient Parameterization of $\mathbf{M}$

- Spatio-Temporal Random Walk: $\mathbf{M} = \mathbf{I}$

$$\mathbf{Y}_t = \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta), \quad \mathbf{C}_\eta \text{ spatial dependence}$$

Linear DSTM: Efficient Parameterization of $\mathbf{M}$

- Spatio-Temporal Random Walk: $\mathbf{M} = \mathbf{I}$

$$\mathbf{Y}_t = \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta), \;\; \mathbf{C}_\eta \text{ spatial dependence}$$

- Common AR with Spatial Errors: $\mathbf{M} = m\mathbf{I}$

$$\mathbf{Y}_t = m\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta),$$

Linear DSTM: Efficient Parameterization of $\mathbf{M}$

- Spatio-Temporal Random Walk: $\mathbf{M} = \mathbf{I}$

$$\mathbf{Y}_t = \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta), \;\; \mathbf{C}_\eta \text{ spatial dependence}$$

- Common AR with Spatial Errors: $\mathbf{M} = m\mathbf{I}$

$$\mathbf{Y}_t = m\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta),$$

- Spatially-Varying AR Models: $\mathbf{M} = \text{diag}(\mathbf{m})\mathbf{I}$

$$\mathbf{Y}_t = \text{diag}(\mathbf{m})\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta).$$

Linear DSTM: Efficient Parameterization of $\mathbf{M}$

- Spatio-Temporal Random Walk: $\mathbf{M} = \mathbf{I}$

$$\mathbf{Y}_t = \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta), \;\; \mathbf{C}_\eta \text{ spatial dependence}$$

- Common AR with Spatial Errors: $\mathbf{M} = m\mathbf{I}$

$$\mathbf{Y}_t = m\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta),$$

- Spatially-Varying AR Models: $\mathbf{M} = \text{diag}(\mathbf{m})\mathbf{I}$

$$\mathbf{Y}_t = \text{diag}(\mathbf{m})\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta).$$

- STAR Parameterization: $\mathbf{M} = (\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{B}_1$, and $\boldsymbol{\eta}_t = (\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{E}_0\boldsymbol{\varepsilon}_t$

$$\mathbf{Y}_t = \mathbf{B}_0\mathbf{Y}_t + \mathbf{B}_1\mathbf{Y}_{t-1} + \mathbf{E}_0\boldsymbol{\varepsilon}_t \;\; \Longrightarrow \;\; \mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t,$$

Linear DSTM: Efficient Parameterization of $\mathbf{M}$

- Spatio-Temporal Random Walk: $\mathbf{M} = \mathbf{I}$

$$\mathbf{Y}_t = \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta), \;\; \mathbf{C}_\eta \text{ spatial dependence}$$

- Common AR with Spatial Errors: $\mathbf{M} = m\mathbf{I}$

$$\mathbf{Y}_t = m\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta),$$

- Spatially-Varying AR Models: $\mathbf{M} = \text{diag}(\mathbf{m})\mathbf{I}$

$$\mathbf{Y}_t = \text{diag}(\mathbf{m})\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim \text{ind. } Gau(\mathbf{0}, \mathbf{C}_\eta).$$

- STAR Parameterization: $\mathbf{M} = (\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{B}_1$, and $\boldsymbol{\eta}_t = (\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{E}_0\boldsymbol{\varepsilon}_t$

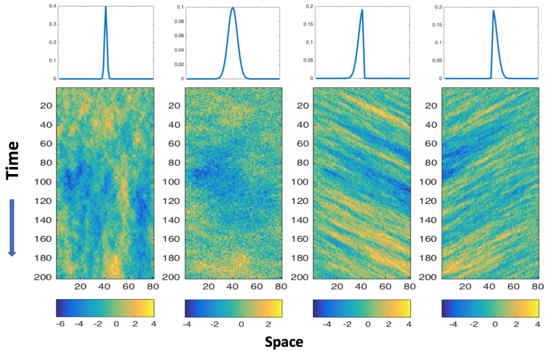
$$\mathbf{Y}_t = \mathbf{B}_0\mathbf{Y}_t + \mathbf{B}_1\mathbf{Y}_{t-1} + \mathbf{E}_0\boldsymbol{\varepsilon}_t \;\; \Longrightarrow \;\; \mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t,$$

- ***Lagged-Nearest Neighbor:** $\mathbf{M}$ (sparse and banded) Let $\mathbf{M}$ be parameterized such that only the m_{ij} corresponding to the location $\mathbf{s}_i$ and the nearest neighbors, say $\{\mathbf{s}_j : j \in \mathcal{N}_i\}$, at the previous time are important.

Linear DSTM: Lagged Nearest Neighbor Process Behavior

Dynamical behavior is implied by changes in the transition operator (kernel) "shape": e.g., 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)



Basic "Deep" Hierarchical Linear DSTM

Data:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim Gau(\mathbf{0}, \mathbf{C}_{\epsilon,t}(\boldsymbol{\theta}_d))$$

Process:

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

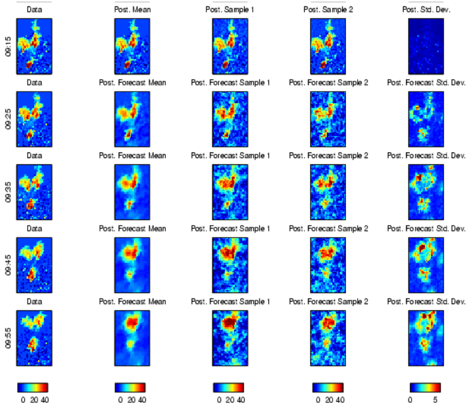
Parameters:

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

These parameters may be estimated empirically, but we get more flexibility if they are given dependent 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.

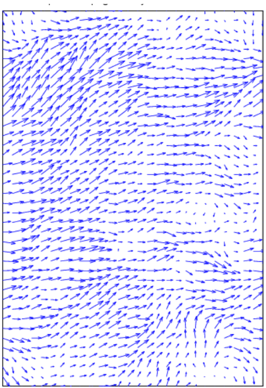
Example: Radar Nowcasting (Sydney, pre 2000 Olympics)

Data | Mean | Post. Real. | Std.Dev.



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

Implied Propagation by **M**



Xu et al. (2005; JASA)

Depicts skewness and relative decay of transition weights

Basis Function Representation

As with descriptive models, one can parameterize the dynamical spatio-temporal process in terms of basis function expansions:

Example "Deep" Hierarchical Basis Expansion DSTM

- Stage 1:

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

- Stage 2:

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{C}_\nu)$$

- Stage 3:

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

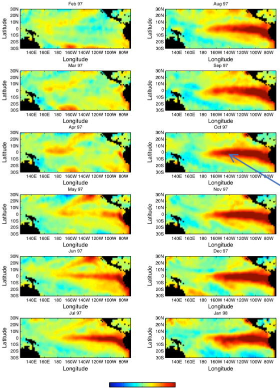
- Stage 4: Structure on the parameters (typically, problem specific, but important for modeling complex processes).

Basis Example: Long-Lead SST Forecasting

Long-lead (7 mo) Prediction of Tropical Pacific Sea Surface Temperature (SST)

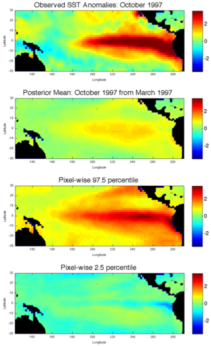
SST Feb 1997-Jan 1998

3/97



El Nino

Model Prediction for 10/97 starting from 3/97



mean prediction

Upper 97.5% Uncertainty

Lower 2.5% Uncertainty

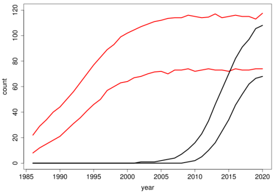
Non-Linear Dynamical Spatio-Temporal Models

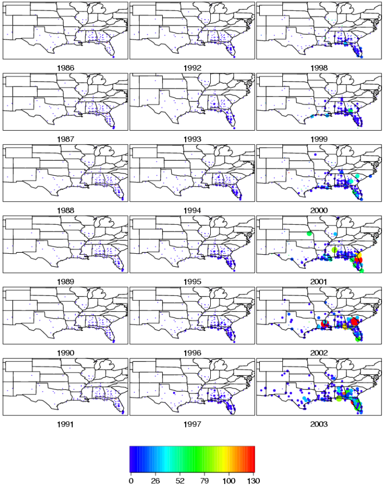
What About Nonlinear DSTMs? e.g., Density dependent growth

Eurasian Collared Dove Invasion: Introduced in South Florida in 1980s

Can we forecast this invasion into the future and accommodate density dependent growth?

95% PI's for S. Fl and Utah





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

Statistical dynamical models of the following form are too general:

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

In statistics, we might then consider:

- Time-varying transition operators in linear Markov models/switching models
- Mechansitic-Motivated Conditional Markov models

Nonlinearity Through Time-Varying Parameters: Threshold Models

Consider the "regime-dependent" switching vector (spatial) process, $\mathbf{Y}_t$:

$$\mathbf{Y}_t = \begin{cases} \mathbf{M}_1 \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, & \text{if } \ell_1, \\ \mathbf{M}_2 \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, & \text{if } \ell_2, \\ \vdots & \vdots \\ \mathbf{M}_J \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, & \text{if } \ell_J, \end{cases}$$

where the threshold criterion for the jth regime, ℓ_j, can depend on the process $\mathbf{Y}_t$, and/or some covariates, $\mathbf{X}_t$.

Note that we typically assume $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{C}_{\eta,j})$ (i.e., the forcing variance-covariance matrix may change by regime.)

(E.g., see Berliner et al. 2000; Wu et al., 2013, and many others for examples)

These models can be inefficient for certain processes.

Scientific Motivation: Nonlinearity

Simple Coupled Ecosystem

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

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

(nonlinear interaction)

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

Shallow Water Equations

$$\begin{aligned}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv &= -g\frac{\partial h}{\partial x} \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu &= -g\frac{\partial h}{\partial y} \\
\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} &= -h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) .
\end{aligned}$$

Nonlinear interactions (advection terms)

Statistical Parameterization: General Nonlinear DSTM

What do these processes have in common? linear terms, quadratic nonlinear terms, exogenous input variables and/or multiviarate interaction (and stochastic forcing/noise)

They also should allow for **linear and nonlinear transformations.** Why?

- mechanistic processes: e.g., density dependent growth, competition (predator-prey), dimension reduction
- stability: e.g., preventing explosive behavior

It is useful to think about a general form for this type of model.

Statistical Parameterization: General Quadratic Nonlinearity (GQN)

General quadratic nonlinearity (GQN) process model: (scalar form for ith component, $i = 1, \ldots, n$; Wikle and Hooten, 2010; Wikle and Holan, 2011):

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

- $\{\eta_t(\cdot)\}$ – dependent Gaussian noise
- "General" because of the $g(Y_{t-1}(\ell); \boldsymbol{\theta}_g)$ term.
- Exogenous inputs can go into $g(\ )$ or enter into the $\{a\}$ and $\{b\}$ parameters (hierarchically)

Recall, this model is usually considered as a latent (conditionally Gaussian) process; there is an observation equation that maps Y to the observations

Quadratic Nonlinearity

Major Problem: There are too many parameters ($O(n^3)$) to estimate directly in typical spatio-temporal applications without extra information or regularization!

Solutions:

- Mechanistic (science)-based parameter shrinkage (hard shrinkage – many parameters are set to zero) [invasive species reaction-diffusion eqn.]
- Reduced dimension basis-function random effects representation (with multiscale basis functions)
- Regularization priors (soft shrinkage; e.g., stochastic search variable selection (SSVS), lasso, elastic net, horseshoe priors; deterministic model output to inform priors – "pre-training")

In practice, we combine these!

Soft Shrinkage: Model Assisted Prior Elicitation (Pre-Training)

In many cases, one has information about parameters from other sources.

E.g., in the physical and biological sciences, this information can be from large (typically) deterministic simulation models.

- **Deterministic Model Output as "Data":** The hierarchical modeling framework easily allows one to condition the various sources of information on the true process of interest.
- **Model Output to Elicit Priors Through Surrogates:** Fit the process model (surrogate or emulator) to the deterministic model output to get an *a priori* understanding of parameter importance. Thus, one builds informative priors that help regularize an over-parameterized model.
  - This can be used in conjunction with SSVS (or other shrinkage priors).

"Deep" Hierarchical QGN DSTM: Overview

The hierarchical nonlinear DSTM is essentially a multi-level ("deep") GLMM w/ a latent conditional Gaussian dynamical process and regularization priors. Requires MCMC implementation.

Data Model: $\mathbf{z}_t | \mathbf{Y}_t, \boldsymbol{\theta}_h \sim EF(\mathbf{H}_t \mathbf{Y}_t; \boldsymbol{\theta}_h)$,

Conditional Mean: $f(\mathbf{Y}_t) = \boldsymbol{\mu}_t + \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\Psi}\boldsymbol{\beta}_t$

Nonlinear Dynamical Process: $\boldsymbol{\alpha}_t \sim GQN(\boldsymbol{\alpha}_{t-1}, \mathbf{A}(\boldsymbol{\theta}_A), \mathbf{B}(\boldsymbol{\theta}_B), \boldsymbol{\theta}_g; \mathbf{C}_\eta)$

Process 2 (problem specific): $[\{\boldsymbol{\beta}_t\} | \boldsymbol{\theta}_\beta]$

Process Mean: $[\{\boldsymbol{\mu}_t\} | \boldsymbol{\theta}_\mu]$ (can depend on covariates)

Regularization Priors: $[\boldsymbol{\theta}_A, \boldsymbol{\theta}_B | \zeta]$

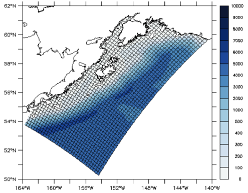
Problem Specific Hyperparameters: $[\boldsymbol{\theta}_h, \boldsymbol{\theta}_g, \boldsymbol{\theta}_\beta, \boldsymbol{\theta}_\mu, \zeta, \mathbf{C}_\eta]$

Recall: Assimilating Ocean Color Observations

(Leeds et al. 2014)

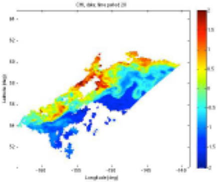
(ocean color: surrogate for phytoplankton)

Coastal Gulf of Alaska

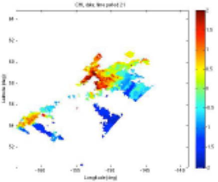


SeaWiFS Ocean Color Satellite Observations (8 day averages)

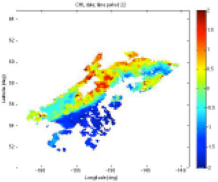
5/19/02 - 5/26/02



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.

Prediction of Primary Production (Phytoplankton) in Coastal Gulf of Alaska

- Data Assimilation: combine satellite ocean color with information in mechanistic computer model for a coupled ocean and ecosystem model (ROMS-NPZDFe; Fiechter et al. 2009)
- Multivariate Quadratic Nonlinear Model: (surrogate dynamical model) on coefficients from basis function expansion on phytoplankton, sea surface height (SSH), and sea surface temperature (SST) (multivariate)
- Priors: pretraining on mechanistic model output (4 years, 1998-2001; 8 day averages) [e.g., **emulator-based prior elicitation** for regularization]
- Implementation: MCMC (fairly time consuming)
- Predict: use satellite ocean color observations for phytoplankton, and mechanistic model output for SSH and SST; predict for 2002

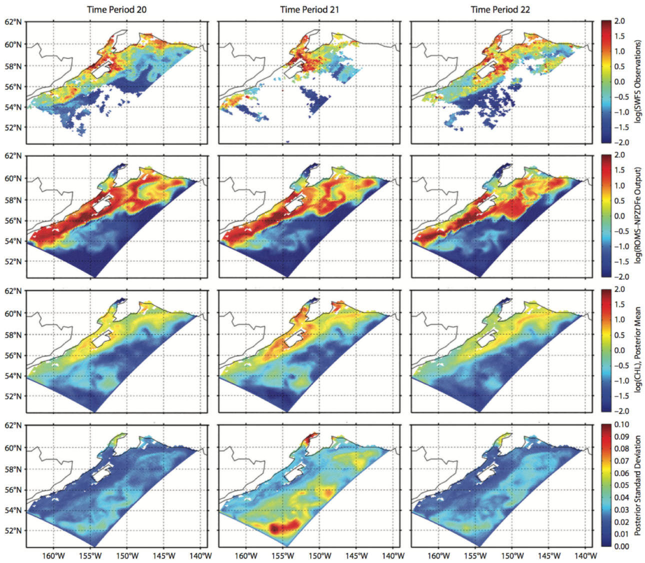
Model Predictions of log(CHL) and Uncertainty

Observations

Ocean Model

Post. Mean

Post. Std. Dev.



Robust (But Expensive)

- The parametric nonlinear GQN model described above is flexible, interpretable, and can accommodate many different types of dynamical processes and uncertainty quantification
- But, this model
  - can be difficult to implement computationally due to the high-dimensionality of the hidden states and parameters
  - typically requires sophisticated regularization (and/or lots of information)

Alternative Non-Linear Models

Alternative Nonlinear S-T Models?

Alternatives? Other flexible methods that don't require us to estimate so many parameters; consider approaches from outside of statistics but include formal uncertainty quantification. Some we have been working on:

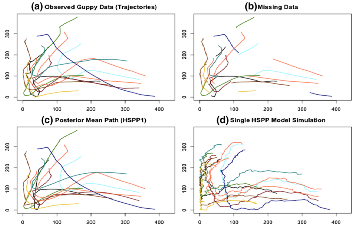
- **Individual-based models** (simple bottom-up rules; highly interpretable, computationally challenging)
- **Analog methods** ("mechanism free models")
- **Recurrent Neural Networks**

In some cases, can be very parsimonious and flexible, but often require a lot of data and traditionally don't include principled uncertainty quantification.

Example: Statistical Individual-Based Models (IBMs)

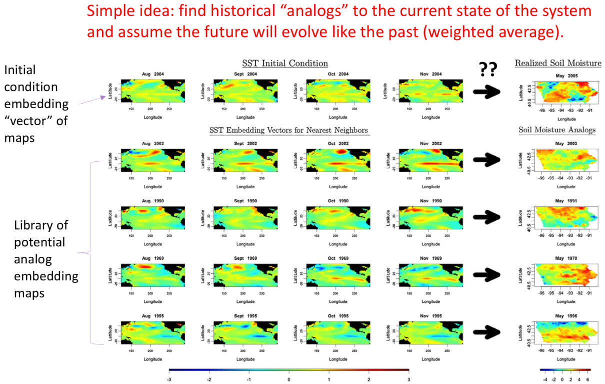
The biggest challenges with modeling IBMs are the estimation of parameters and, more critically, the realistic quantification of uncertainty associated with the data, model and parameters.

- In McDermott et al. (2017) we consider a hierarchical extension of a stochastic self-propelled particle (SPP) model (Vicsek et al. 1995) to represent the collective movement of guppies in a tank.
- **Important Model Rules:** fish movement based on behavior of neighbors; important - covariates can influence behavior (e.g., shelter)
- **Computation:** hybrid ABC-MCMC algorithm



A Computationally Efficient Alternative: S-T Analog Models

- Analog forecasting was invented by meteorologists in the 1920s (similar notions in ecology: e.g., Sugihara and May, 1990)
- Simple idea: find historical "analogs" to current situation and assume the new future will evolve like the past (e.g., find closest sets of evolving maps and weight them)

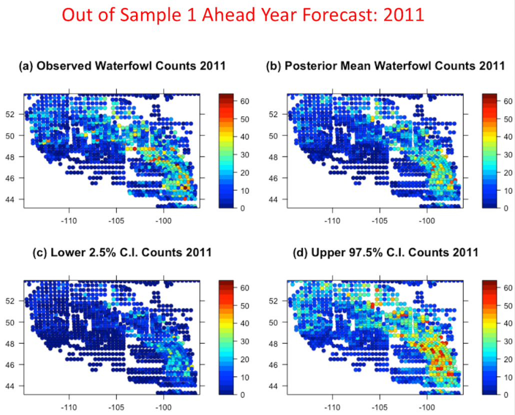


Example: Mallard Duck Settling Pattern Prediction

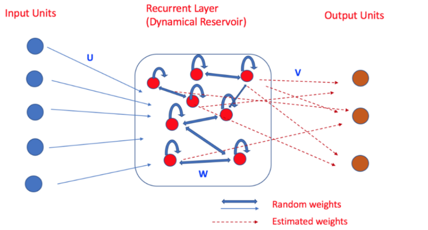
McDermott et al. (2017): Bayesian Non-Gaussian S-T Analog Model

Breeding Population Survey

- Spring relative abundance counts from aerial survey (1955-2014); 2171 spatial segments
- Production varies across landscape
- "site philopatry" unless habitat conditions are bad; then overflight
- Climate conditions affect productivity and management decisions
- Analog forecasts based on Pacific SST analogs



A Computationally Efficient Alternative: Recurrent Neural Networks (Echo State Networks, ESNs)



response	$\mathbf{Y}_t = g_o(\mathbf{o}_t)$
output	$\mathbf{o}_t = \mathbf{V}\mathbf{h}_t$
hidden state	$\mathbf{h}_t = g_h(\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\mathbf{x}_t)$ ← external inputs
weight matrices	$\mathbf{V}, \mathbf{W}, \mathbf{U}$
activation functions	$g_o(\cdot), g_h(\cdot)$ ← (e.g., identity, softmax, tanh, etc.)

Crucially: $\mathbf{W}$ and $\mathbf{U}$ are chosen at random then assumed to be fixed (subject to a spectral radius constraint for $\mathbf{W}$)!

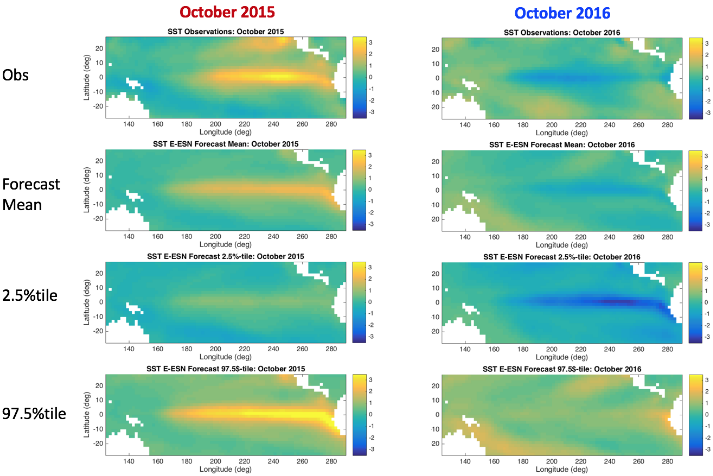
Huge cost saving in terms of training – only the output weights $\boldsymbol{V}$ need be estimated (usually with a ridge penalty)!

So, it is a nonlinear model but, depending on $g_o(\cdot)$, can be fit with regression or glm methods.

Very fast! (But, requires a lot of data)

Statistical ESN Example: 6 Month SST Predictions for 10/2015 and 10/2016

Need to add uncertainty quantification: McDermott and Wikle (2017; parametric bootstrap); McDermott and Wikle (2018; fully Bayesian)



Spatio-Temporal Support of Data and Processes

Spatio-Temporal Support of Data and Processes

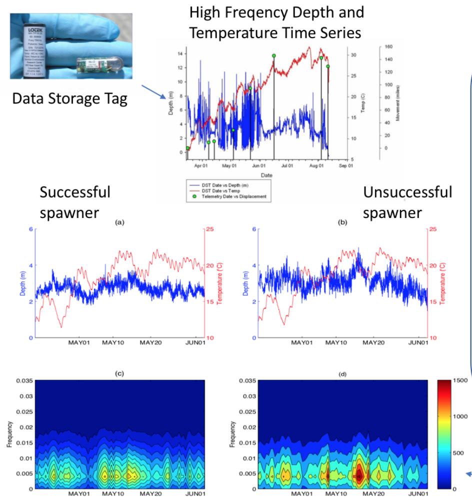
With technological advancement, we expect ever increasing data volume – particularly, "high-frequency" and remotely-sensed data.

A major challenge for statisticians and scientists is how to address likely disparities between the scales at which the data were collected and those at which we would like to do inference or prediction, as well as to remove redundant information.

Areas of current research:

- High-frequency information as covariates
- Change-of-support (resolution) in time and space (particularly challenging for non-Gaussian data with uncertainty)
- Optimal resolution/data reduction

High Frequency Covariates Example: Spawning Success of Missouri River Sturgeon



Spectrogram (time-frequency) "Image"

How to incorporate into a GLMM?

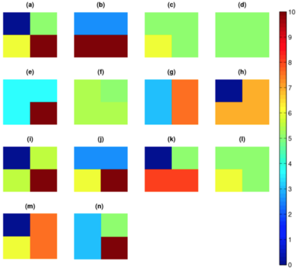
- Convert time series into time-frequency spectrogram
- Treat spectrogram as an "image" or spatial process
- Decompose image into a basis expansion
- Expansion coefficients used as covariates in the model
- Requires regularization/shrinkage (e.g., lasso, stochastic search variable selection)
- Can include interactions among multiple input signals (e.g., depth and temperature)
- Yang et al. 2013

Aggregation Error and Change-of-Support

Aggregation Error: Ecological Fallacy/Modifiable Areal Unit Problem (MAUP): when inference on the aggregate scale of spatial support differs from inference on another distinct spatial support.

Example: (a) truth; (b)-(n) various 2-3 group realizations

We seek to:
(i) quantify regionalization error,
(ii) select optimal regionalizations that minimize this error!

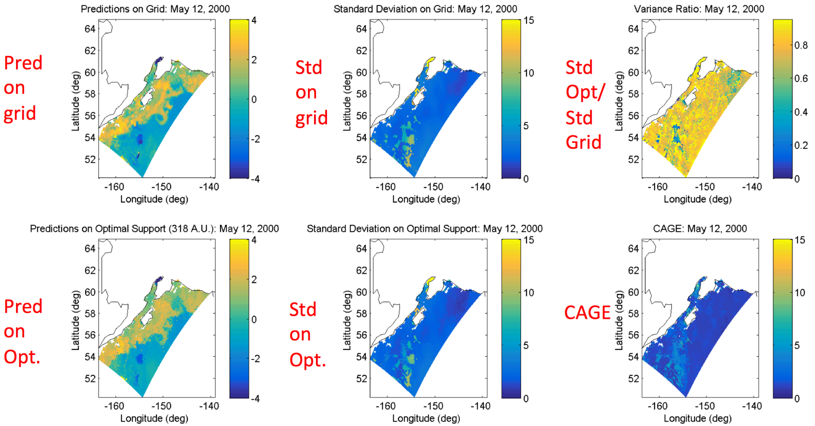


- We would like to use models to predict at any level of spatial support, taking into account measurement/sampling uncertainty.
- We would like to choose the "optimal" spatial support:
  - Gives lower prediction variances
  - Gives essentially the same inference as finer resolutions
  - Is "optimal" in the sense that it minimizes a formal aggregation error criterion

By using a spatial basis function exapansion, we develop a "Criterion for Spatial Aggregation Error (CAGE)" (Bradley et al. 2017) that allows us to develop optimal aggregations and change support.

Ex: Optimizing Support Relative to CAGE

Consider just one (active) time period from the Chlorophyll example; CAGE algorithm; **318 clusters were optimal (reduced from 5255 grid cells)**



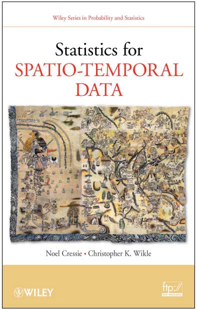
Some important things I didn't get to talk about

- multivariate processes and interactions
- multi-type data
- adaptive design of monitoring systems
- extremes
- model selection and evaluation
- diagnostics and verification
- sampling issues
- S-T point processes/ S-T random sets
- Computation! Approximate and distributed computing are increasingly important.

THANK YOU! (and Shameless Promotion)

THANK YOU!! If you have questions or would like references to anything I talked about, please feel free to contact me: wiklec@missouri.edu

Existing Book:



New Book – Coming: Spring 2018

"Spatio-Temporal Statistics with R"

by

Christopher K. Wikle
Andrew Zammit-Mangion
Noel Cressie

(Chapman and Hall/CRC)