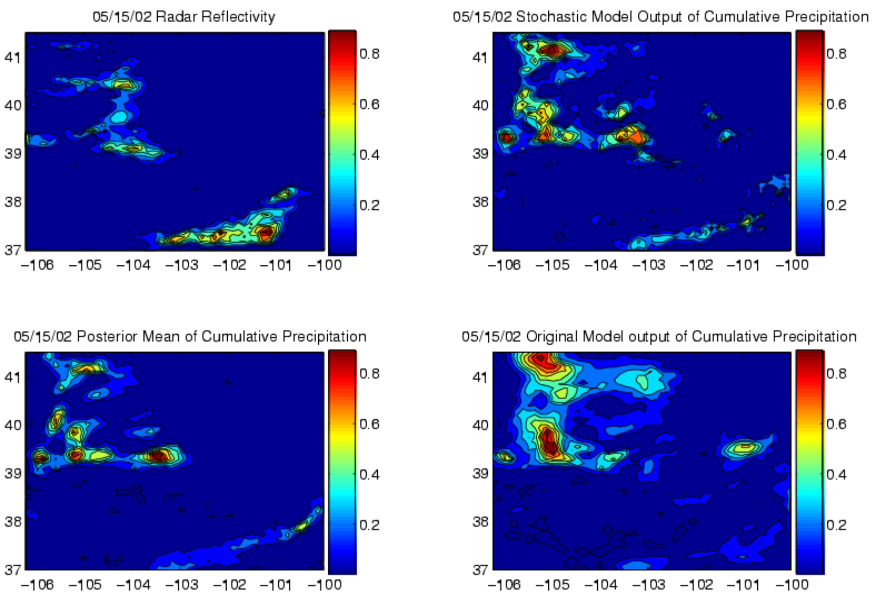
Bayesian Hierarchical Parameterizations

Christopher K. Wikle
Department of Statistics
University of Missouri-Columbia



Outline of Talk

- Bayesian Hierarchical Models (BHMs)
- BHM Parameterizations with Qualitative Dynamics
  - Application: Long Lead Prediction: SST
  - Application: Nowcasting Radar Reflectivities
- BHM Parameterizations in Coupled Systems
  - Application: Air/Sea Experiment
- BHM Parameterizations in Regional Climate Models
  - Application: Stochastic Trigger Function in Climate Model

Bayesian Hierarchical Modeling: Think Conditionally!!

Data $\to D$ Process $\to U$ Parameters $\to \theta$

- Data Model(s): $[D \mid U, \ \theta \,]$

  Simpler dependence structures through conditioning. e.g.,

$$[D_a, D_b|U, \theta] = [D_a|U, \theta][D_b|U, \theta]$$

- Process Model(s): $[U \mid \theta \,]$

  Can build-up complicated dependence by conditional models. e.g.,

$$[U_2, U_1|\theta] = [U_2|U_1, \theta][U_1|\theta]$$

  Incorporate science!! (e.g., PDEs)

- Parameter Model(s): $[\ \theta\ ]$ Further conditioning; Incorporate science

Bayes:

$$[U, \theta|D] \propto [D|U, \theta][U|\theta][\theta]$$

BHM Parameterizations for Qualitative Dynamics

Process Model: (linear in Y)

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

Efficient Parameterization when $\mathbf{M}_t$ is unknown:

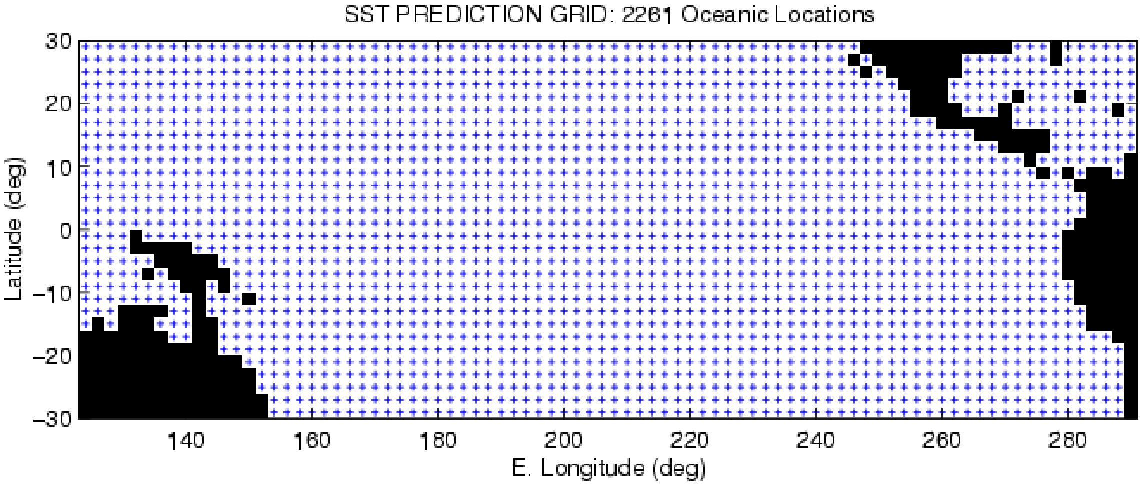
- Dimension reduction in the state process, $\mathbf{Y}_t = \mathbf{\Phi}\mathbf{a}_t + \boldsymbol{\nu}_t$
- Low-dimensional parameterization of $\mathbf{M}_t$

Examples:

- Long Lead Prediction of Pacific SST
- Nowcasting Weather Radar Reflectivities

Application: Long Lead Prediction of SST

Goal: Predict Pacific SST anomalies ($2^\circ \times 2^\circ$ resolution) 7 months in advance while realistically accounting for uncertainty. [Berliner,Wikle, Cressie, 2000, *J. Climate*]



Consider data $\{Z(\mathbf{s}_i;t)\}$ to be *anomalies* from monthly means.

Want to predict $Z(\mathbf{s}_0;T+\tau)$; for $\tau = 7$ months, from space-time data $\mathcal{D}(T) \equiv \{\mathbf{Z}_T, \mathbf{Z}_{T-1}, \ldots, \mathbf{Z}_1\}$.

Dimension Reduction:

$$\mathbf{Z}_t = \mathbf{\Phi}\mathbf{a}_t + \boldsymbol{\nu}_t,$$

where

- $\boldsymbol{\nu}_t \sim Gau(\mathbf{0}, \mathbf{\Sigma}_\nu)$; measurement error and small-scale spatial variability that is uncorrelated in time (includes information related to the ”unresolved” modes in $\mathbf{\Sigma}_\nu$).
- $\mathbf{\Phi}$; truncated **empirical orthogonal function** basis set (not optimal in the dynamical context, but historical precedent), obtained from SVD of $\mathbf{Z} \equiv (\mathbf{Z}_1, \ldots, \mathbf{Z}_T)$.

Process Model

$$\mathbf{a}_{t+\tau} = \boldsymbol{\mu}_t + \mathbf{M}_t\mathbf{a}_t + \boldsymbol{\eta}_{t+\tau},$$

where, $\boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ for all t.

Critical modeling assumption: Let $\mathbf{M}_t$ and $\boldsymbol{\mu}_t$ be dependent on *both the current and future climate regimes*:

$$\begin{aligned}\mathbf{M}_t &= \mathbf{M}(I_t, J_t)\\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}(I_t, J_t),\end{aligned}$$

where,

- I_t - classifies the current regime as “warm” (2), “normal” (1), or “cold” (0)
- J_t - anticipates a transition to one of the three regimes at time $t + \tau$

Climate States

Current, I_t: [Threshold Model, e.g., Tong 1990]

$$\begin{aligned} I_t &= 0, \quad \text{if } SOI_t > \text{low threshold} \\ &= 1, \quad \text{if otherwise} \\ &= 2, \quad \text{if } SOI_t < \text{upper threshold}, \end{aligned}$$

where SOI_t is the Southern Oscillation Index (assumed "known").

Future, J_t: [Latent (hidden) Process Model]

$$\begin{aligned} J_t &= 0, \quad \text{if } W_t > \text{low threshold} \\ &= 1, \quad \text{if otherwise} \\ &= 2, \quad \text{if } W_t < \text{upper threshold}, \end{aligned}$$

where W_t is a latent process which anticipates the future climate regime.

Latent Process to Assess "Future"

$$W_t | \boldsymbol{\beta}_w, \sigma_w^2 \sim Gau(\mathbf{x}_t' \boldsymbol{\beta}_w, \sigma_w^2)$$

where,

$$\mathbf{x}_t = (1, U_t, U_t sin(\frac{2\pi m_t}{12}), U_t cos(\frac{2\pi m_t}{12}), U_t^2)',$$

where

- U_t - the lowpass filtered E-W component of the wind at 10 meters above the surface at $5°$ N and $157°$ E. (This is absolutely critical!)
- m_t - index of month (0 - 11) at time t

Priors on Model Parameters

At next stage of hierarchy:

$$\begin{aligned} \boldsymbol{\beta}_w &\sim Gau(\hat{\boldsymbol{\beta}}_{SOI}, c_\beta \mathbf{var}(\hat{\boldsymbol{\beta}}_{SOI})), \\ \sigma_w^2 &\sim IG(q, r) \end{aligned}$$

where the mean and variance and IG ("Inverse Gamma") parameters are obtained from fitting

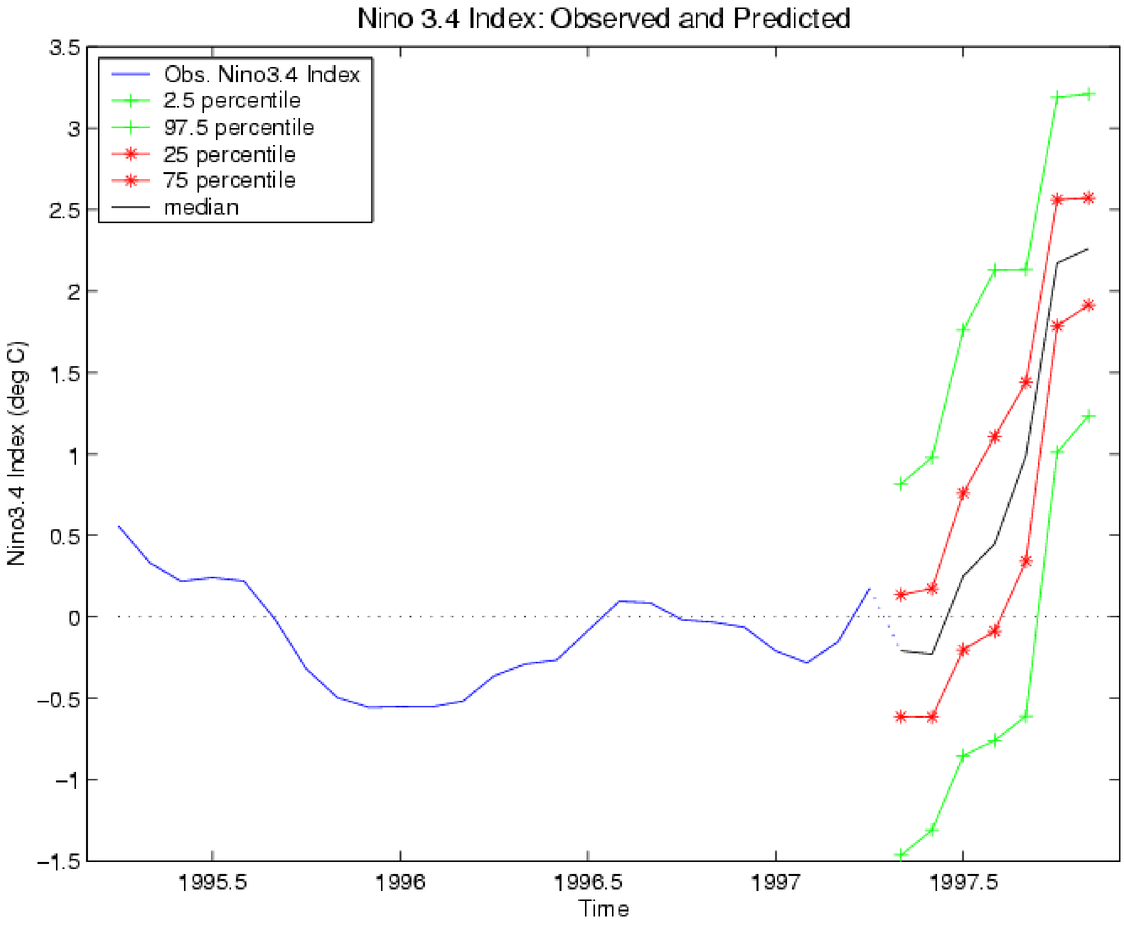
$$SOI_{t+\tau} = \mathbf{x}_t' \boldsymbol{\beta}_{SOI} + e_t$$

(gives $R^2 \approx .7$!)

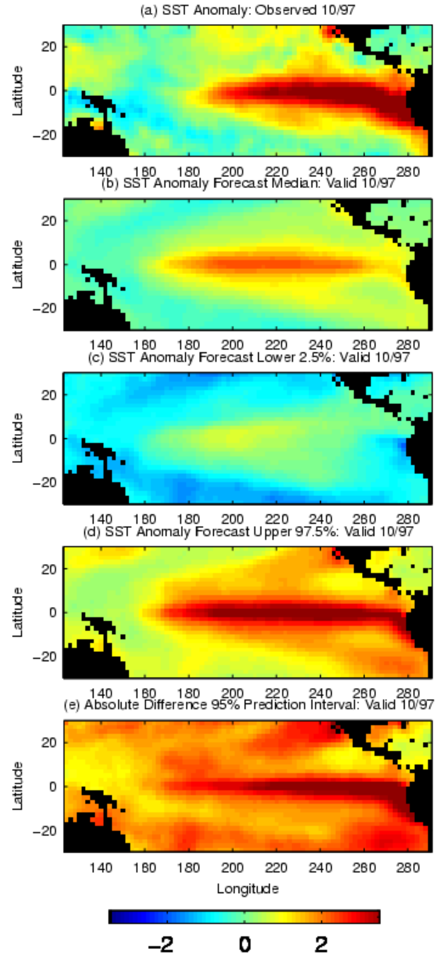
- $\mathrm{vec}(\mathbf{H}_j), j = 1, 2, 3$ - Multivariate Normal distributions
- $\boldsymbol{\mu}_j, j = 1, 2, 3$ - Multivariate Normal distributions
- Covariance matrices - Wishart distributions

("Empirical Bayes" priors)

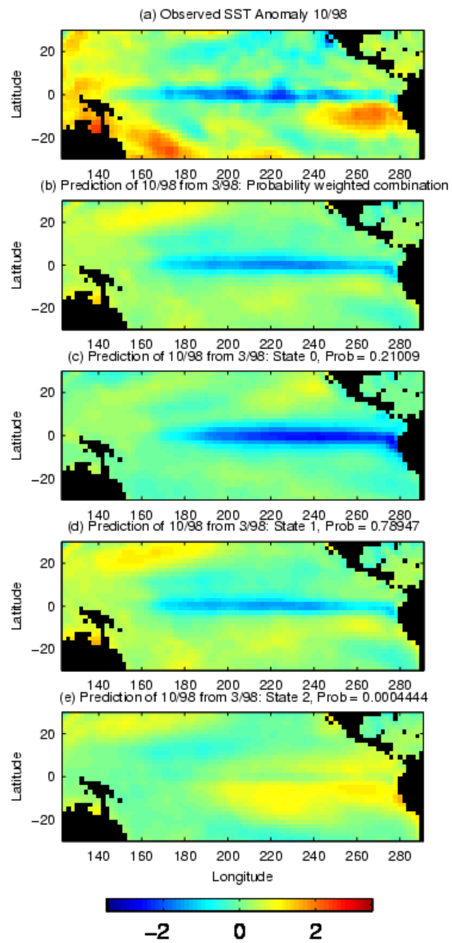
Results: Niño 3.4 Prediction of 10/97 from 3/97



Results: Posterior Predictions for 10/97



Results: Posterior Predictions for 10/98



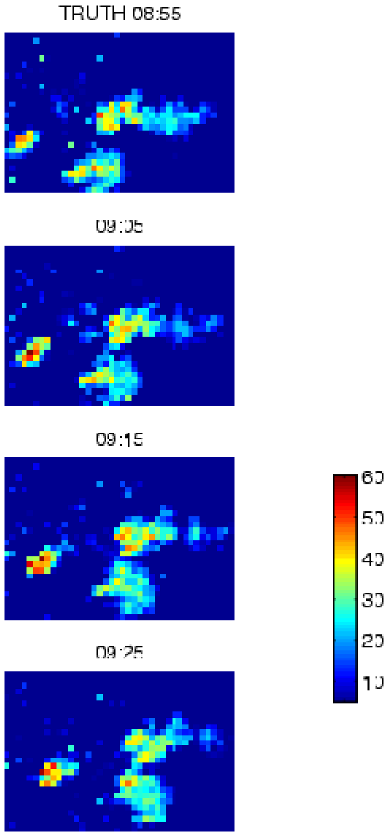
Operational Forecasts Generated Each Month:

Spatial Statistics and Environmental Sciences (SSES) program at Ohio State University

`http://www.stat.ohio-state.edu/~sses/collab_enso.php`

Application: Radar Nowcasting

Short-term forecast (nowcast) of radar reflectivities based on recent past.
(Xu, Wikle, and Fox, 2005; *J. Amer. Stat. Assoc.*)



Parameterization of Process Model

Model Assumption: linear at short time scales; spatially explicit

Ideally, would like to estimate the propagator matrix $\mathbf{M}$ in:

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

where $\mathbf{Y}_t$ is n-dimensional and $t = 1, \ldots, T$ corresponds to the data periods.

In principle, this is easy. In practice, it is NOT (Large n, small T problem)!

Integro Difference Equation (IDE) Parameterization

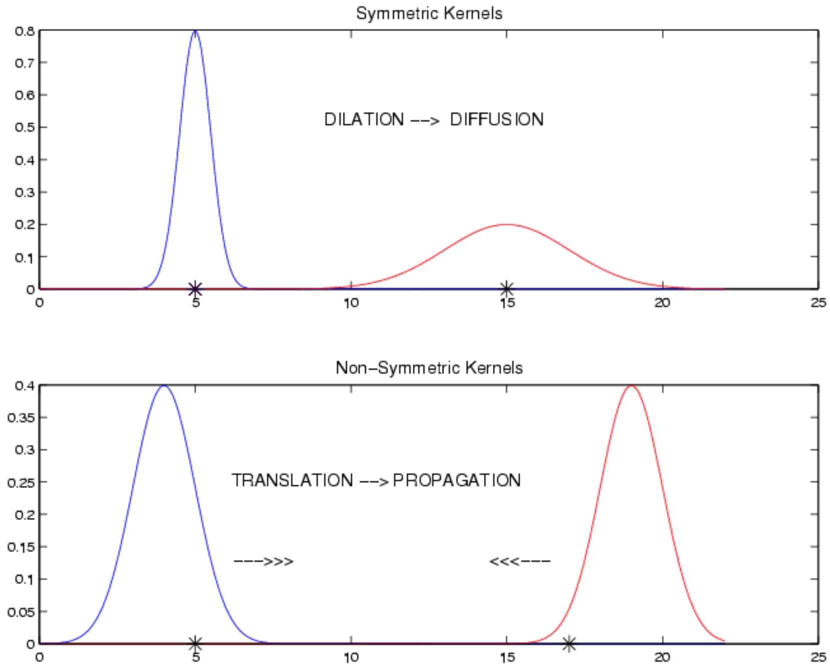
Consider the linear IDE model:

$$Y_{t+1}(s) = \gamma \int k_s(r; \theta_s) Y_t(r) dr$$

The key to modeling dynamical processes is $k_s(r; \theta_s)$, the redistribution kernel.

- Has been shown to model diffusive wave fronts (e.g., Kot et al. 1996); shape and speed of diffusion depends on kernel width and tail behavior (dilation)
- Can also model non-diffusive propagation via the relative displacement of kernel (translation) (Wikle, 2001; 2002)

Kernel Properties and Dynamics



NOTE: If dilation and translation parameters are allowed to vary with space, then complicated (advection/diffusion) dynamics can be modeled with relatively few parameters.

IDE Process: Spectral Representation

Stochastic IDE:

$$Y_{t+1}(s) = \gamma \int k_s(r;\theta_s) Y_t(r) dr + \tilde{\eta}_t(s), \quad \tilde{\boldsymbol{\eta}}_t \sim \; N(\mathbf{0}, \boldsymbol{\Sigma}_{\tilde{\eta}})$$

Consider the spectral expansion of the kernel and process in terms of orthonormal spectral basis functions, $\phi_j(s)$:

$$k_s(r;\theta_s) = \sum_j b_j(s;\theta_s)\phi_j(r)$$

$$Y_t(s) = \sum_j a_j(t)\phi_j(s)$$

Substituting these into the IDE for spatial locations $\mathrm{s}_1, \ldots, \mathrm{s}_n$ leads to a spectral representation of the model:

$$\boldsymbol{a}_{t+1} = \gamma \boldsymbol{\Phi}' \mathbf{B}'_{\boldsymbol{\theta}} \boldsymbol{a}_t + \boldsymbol{\eta}_{t+1}, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$$

where $[\boldsymbol{\Phi}]_{ij} = \phi_j(\mathbf{s}_i)$, $\;[\mathbf{B}_{\boldsymbol{\theta}}]_{kl} = b_k(\mathbf{s}_l;\boldsymbol{\theta}_{s_l})$, and $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\tilde{\eta}}\boldsymbol{\Phi}$.

Comments on Hierarchical Parameterization

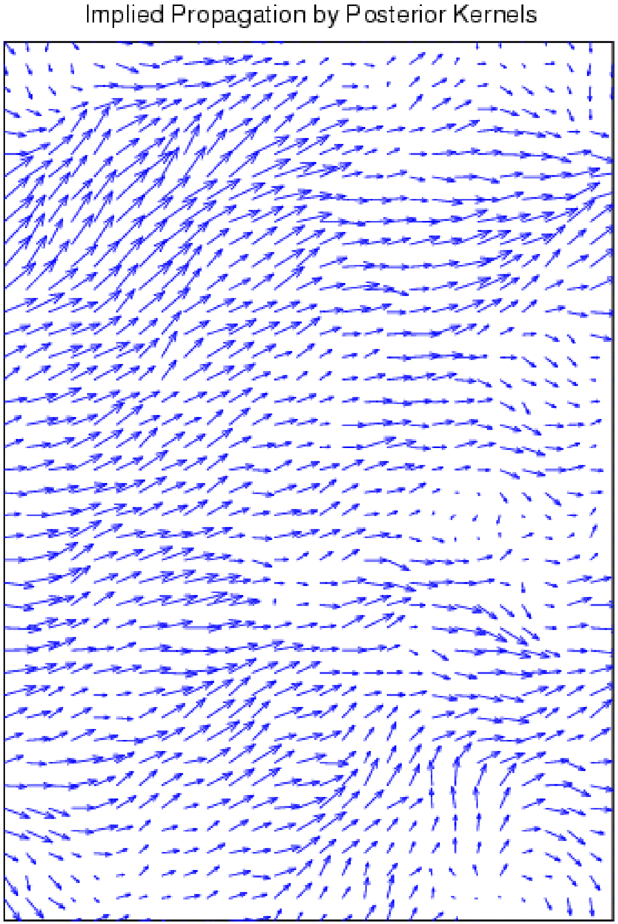
Let $\phi_i(\mathbf{s})$ be Fourier basis functions.

The kernel spectral coefficients, $b_j(\mathbf{s}; \boldsymbol{\theta}_s)$ are then known if the kernel is a pdf since the Fourier transform of the pdf kernel is its characteristic function. (Can lead to dimension reduction)

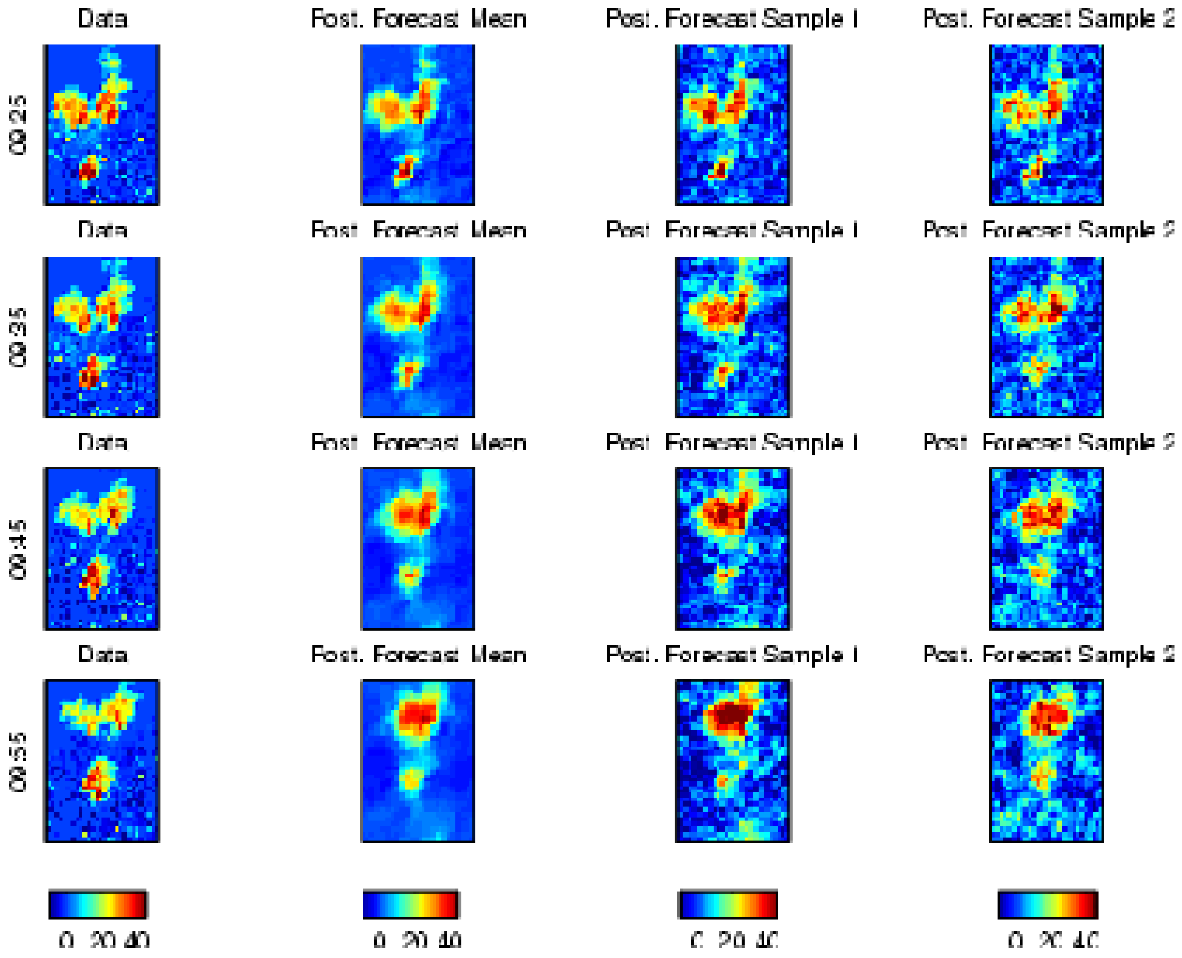
Thus, $\mathrm{B}_{\boldsymbol{\theta}}$ is completely defined if we know the kernel translation and dilation parameters ($\boldsymbol{\theta}_{\mathrm{s}}$).

Critically, the kernel parameters are assumed to be spatially-varying, and are assigned spatial random field priors at the next level of the hierarchy.

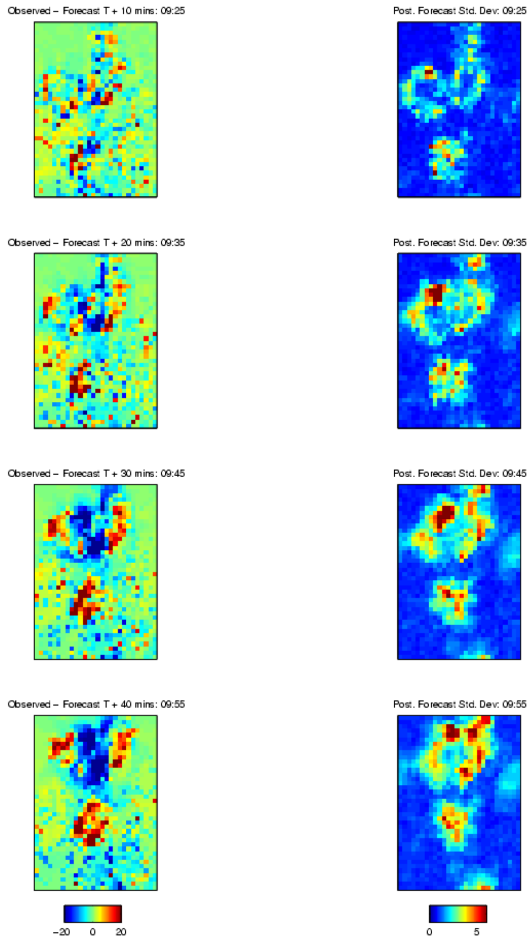
Example Posterior Mean: Kernel Translation (Implied Propagation)



Nowcast Results: Data, Posterior Mean, Samples



Uncertainty Characterization: Posterior Standard Deviation



Example: Bayesian Hierarchical Modeling of Air-Sea Interaction

(Berliner, Milliff, Wikle, 2003: *Journal of Geophysical Research: Oceans*)

- Couple models of interacting spatio-temporal processes (atmosphere and ocean)
  - Hierarchical coupling of complicated systems; each of which is also modeled hierarchically
  - Use approximate dynamics; physical-statistical models
- Incorporate diverse datasets
- Include stochastic elements to adjust for model uncertainty, un-modeled components, etc.
- Quantify uncertainty in each phase

Stochastically Coupled Air/Sea Model

Data

* D_a Atmospheric data (scatterometer)
* D_o Ocean data (altimeter)

HBM Skeleton

1. $[D_a, D_o | \text{Atm}, \text{Ocean}, \theta_a, \theta_o]$
2. $[\text{Atm}, \text{Ocean} | \eta_a, \eta_{o|a}]$
3. $[\theta_a, \theta_o, \eta_a, \eta_{o|a}]$

Parameters

$\theta_a, \theta_o, \eta_a, \eta_{o|a}$

BHM Keys

1. $[D_a, D_o|\text{Atm}, \text{Ocean}, \theta_a, \theta_o] = [D_a|\text{Atm}, \theta_a][D_o|\text{Ocean}, \theta_o]$

2. $[\text{Atm}, \text{Ocean}|\eta_a, \eta_{o|a}] = [\text{Ocean}|\text{Atm}, \eta_{o|a}][\text{Atm}|\eta_a]$

- Atm & Ocean data are conditionally independent
- Parameterized air-sea model is stochastic atmospheric model coupled to stochastic Ocean-given-Atmospheric model

* Posterior: $[\text{Atm}, \text{Ocean}, \eta_a, \eta_{o|a}|D_a, D_o]$

“Full” coupling of Atmosphere and Ocean

Observation Simulation System Experiment (OSSE)

- Ocean truth simulation driven by idealized wind fields
  - primitive equation, shallow water approximation (Milliff and McWilliams, 1994) [This is more complicated than the PDE prior we will use!]
- 10-day forcing with intense idealized atmospheric cyclone (emulating polar low)
- Sample winds (scatterometer sampling) and altimetry (e.g., TOPEX) corrupted with noise
- Compare BHM based on simple ocean and atmosphere models and sparsely sampled, corrupted data to shallow-water truth

Wind Driven Ocean Process Model

Prior Based on "Simple" Physics: Quasigeostrophy (Ψ - streamfunction)

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

where J is the Jacobian, τ the wind-stress, and r, β, ρ, H, γ, a_H are parameters. Using traditional finite difference approximations to time- and space-derivatives:

$$
\begin{aligned}
\mathbf{\Psi}^I_{t+1} &= \{\mathbf{I} + \Delta\,\tilde{\mathbf{G}}\,(-\beta\,\mathbf{D}_x - \gamma\mathbf{G} - a_H\,\mathbf{G}^2)\}\,\mathbf{\Psi}^I_t \\
&\quad + \Delta\,\tilde{\mathbf{G}}\,(-\mathcal{J} + \frac{1}{\rho H}\mathcal{C}(\mathbf{U}_t, \mathbf{V}_t)) + \mathbf{B}\mathbf{\Psi}^B_{t+1}
\end{aligned}
$$

where: $\mathbf{\Psi}^I_t$ - vectorization of interior streamfunction; $\mathbf{\Psi}^B_t$- boundary streamfunction values; Δ- time step; $\mathbf{G}$ - discretized 2-d Laplacian; $\mathcal{J}$ - discretized Jacobian; $\mathcal{C}(\mathbf{U}_t, \mathbf{V}_t)$- discretized wind stress curl; $\tilde{\mathbf{G}} = (\mathbf{G} - r^{-2}\mathbf{I})^{-1}$

Stochastic Streamfunction Model

$$\boldsymbol{\Psi}_{t+1} = \mathbf{P}(\mathbf{L})\,\boldsymbol{\Psi}_t - j\,\tilde{\mathbf{G}}\,\mathcal{J} + c\,\tilde{\mathbf{G}}\,\mathcal{C}(\mathbf{U}_t, \mathbf{V}_t) + b\,\mathbf{B}\boldsymbol{\Psi}^B_{t+1} + \mathbf{e}_{t+1},$$

where

$$\mathbf{e}_{t+1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_e),$$

and

$$\mathbf{P}(\mathbf{L}) = l_1\mathbf{I} + l_2\tilde{\mathbf{G}}\mathbf{D}_x + l_3\tilde{G}\mathbf{G}^2.$$

Critically: $\mathbf{L} = (l_1, l_2, l_3)'$ and (j, c, b) are random parameters with prior means are suggested by the deterministic model, but they can be informed by the data! For example:

Prior mean of l_1: $1 - \gamma$

Also, we have hierarchical boundary conditions (e.g., Wikle, Berliner, Milliff, 2003): $[\Psi_I, \Psi_B] = [\Psi_I | \Psi_B][\Psi_B]$

Atmospheric Process/Parameter Models/Data Models

- **Atmospheric Process:** Stochastic Geostrophy (Royle et al. 1999)

$$\begin{aligned} \mathbf{U}_t, \mathbf{V}_t | \mathbf{P}_t, \boldsymbol{\theta}_w &\sim N(\mathbf{K}(\boldsymbol{\theta}_w)\mathbf{P}_t, \boldsymbol{\Sigma}_w) \\ \mathbf{P}_t &\sim N(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p) \quad \text{(hidden process!)} \\ \boldsymbol{\theta}_w &\sim N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) \end{aligned}$$

- **Data Models:**
  - Scatterometer:

$$\begin{pmatrix} \mathbf{D}_u^t \\ \mathbf{D}_u^t \end{pmatrix} = \mathbf{K}_w^t \begin{pmatrix} \mathbf{U}^t \\ \mathbf{V}^t \end{pmatrix} + \boldsymbol{\epsilon}_w^t,$$

  - Altimetry:

$$\mathbf{D}_\Psi^t = \mathbf{K}_o^t \boldsymbol{\Psi}^t + \boldsymbol{\epsilon}_o^t,$$

Posterior Computation

For ease of notation, let
$\mathbf{u} = \{\mathbf{U}^t : t \in \mathcal{T}\}$, $\mathbf{v} = \{\mathbf{V}^t : t \in \mathcal{T}\}$, $\boldsymbol{\psi} = \{\boldsymbol{\Psi}^t : t \in \mathcal{T}\}$

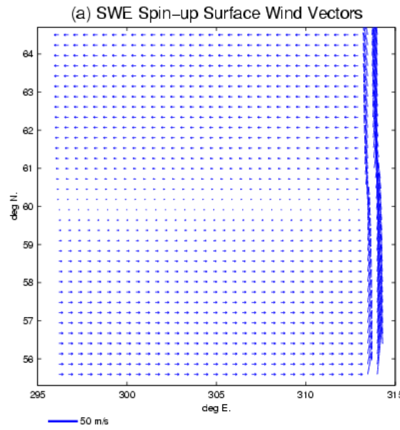
Posterior:

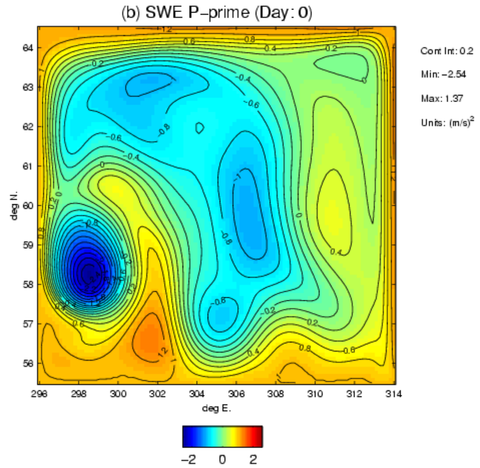
$$
\begin{aligned}
[\mathbf{u}, \mathbf{v}, \boldsymbol{\Psi}, \theta_w, \theta_\psi, \eta_w, \eta_\psi | D_\psi, D_w] \;\propto\; & [D_\psi | \boldsymbol{\psi}, \theta_\psi] \\
& \prod [\boldsymbol{\Psi}_{t+1} | \boldsymbol{\Psi}_t, \mathbf{u}, \mathbf{v}, \eta_\psi][\Psi_1 | \mathbf{u}, \mathbf{v}, \eta_\psi][\eta_\psi, \theta_\psi] \\
& [\mathbf{u}, \mathbf{v}, \theta_w, \eta_w | D_w].
\end{aligned}
$$

Since the proportionality constant intractable we use a **combination MCMC- Importance Sampling MC (ISMC) approach**.

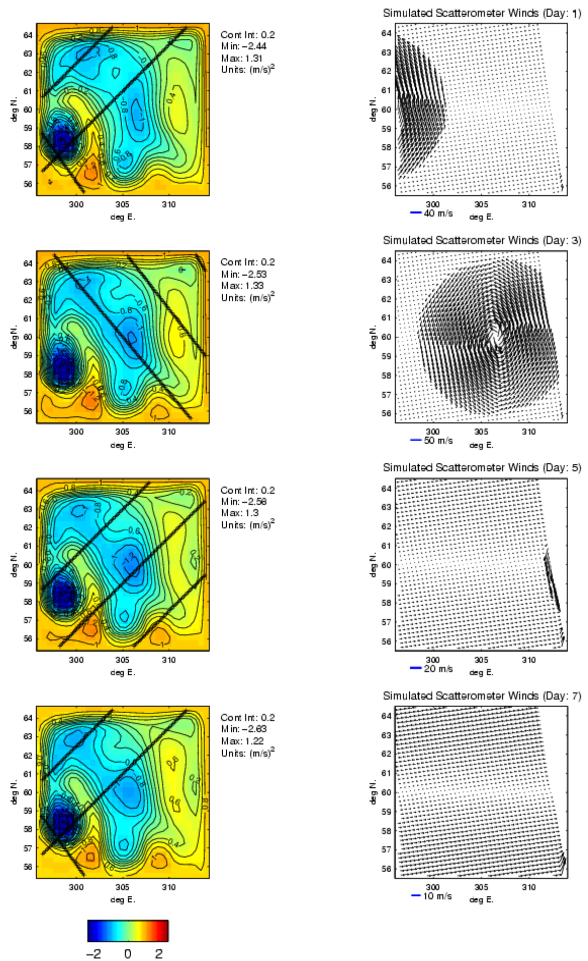
- MCMC Atmospheric Model: $[\mathbf{u}, \mathbf{v}, \theta_w, \eta_w | D_w]$
- Use MCMC samples, and samples from prior dist. on parameters to get MC samples for streamfunction: $\prod [\boldsymbol{\Psi}^{t+1} | \boldsymbol{\Psi}^t, \mathbf{u}, \mathbf{v}, \eta_\psi][\Psi^1 | \mathbf{u}, \mathbf{v}, \eta_\psi][\eta_\psi, \theta_\psi]$
- Use Importance Sampling with weights proportional to $[D_\psi | \boldsymbol{\psi}, \theta_\psi]$

RESULTS: Testbed Initial Conditions

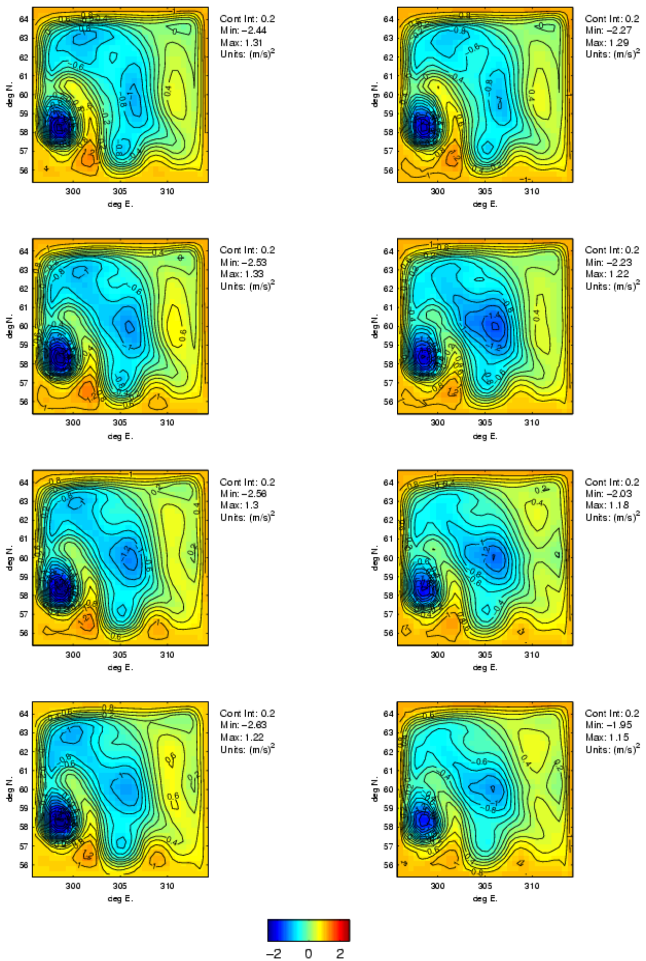




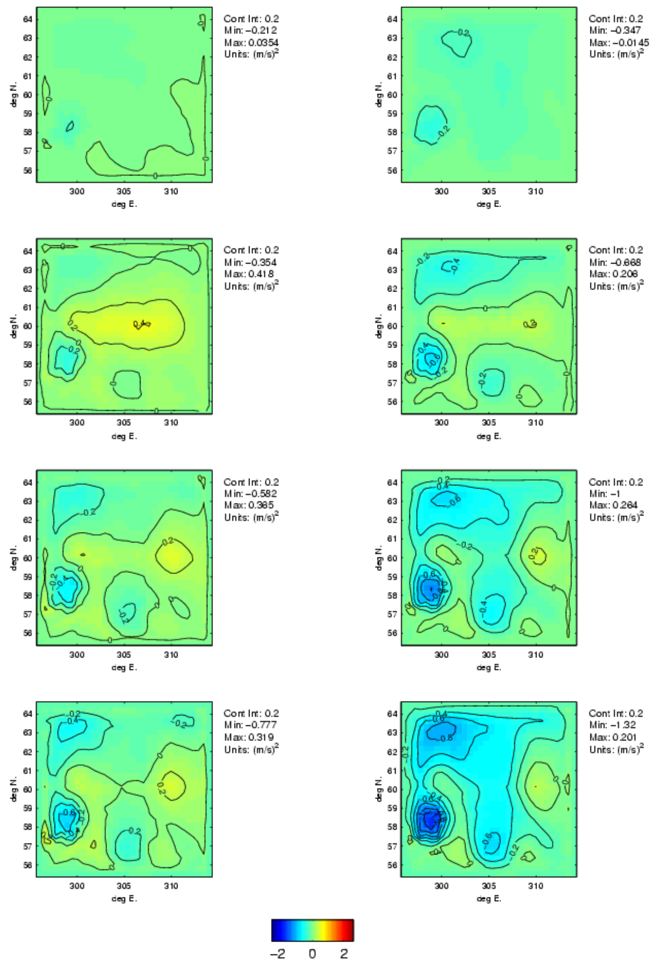
RESULTS: Simulated Data



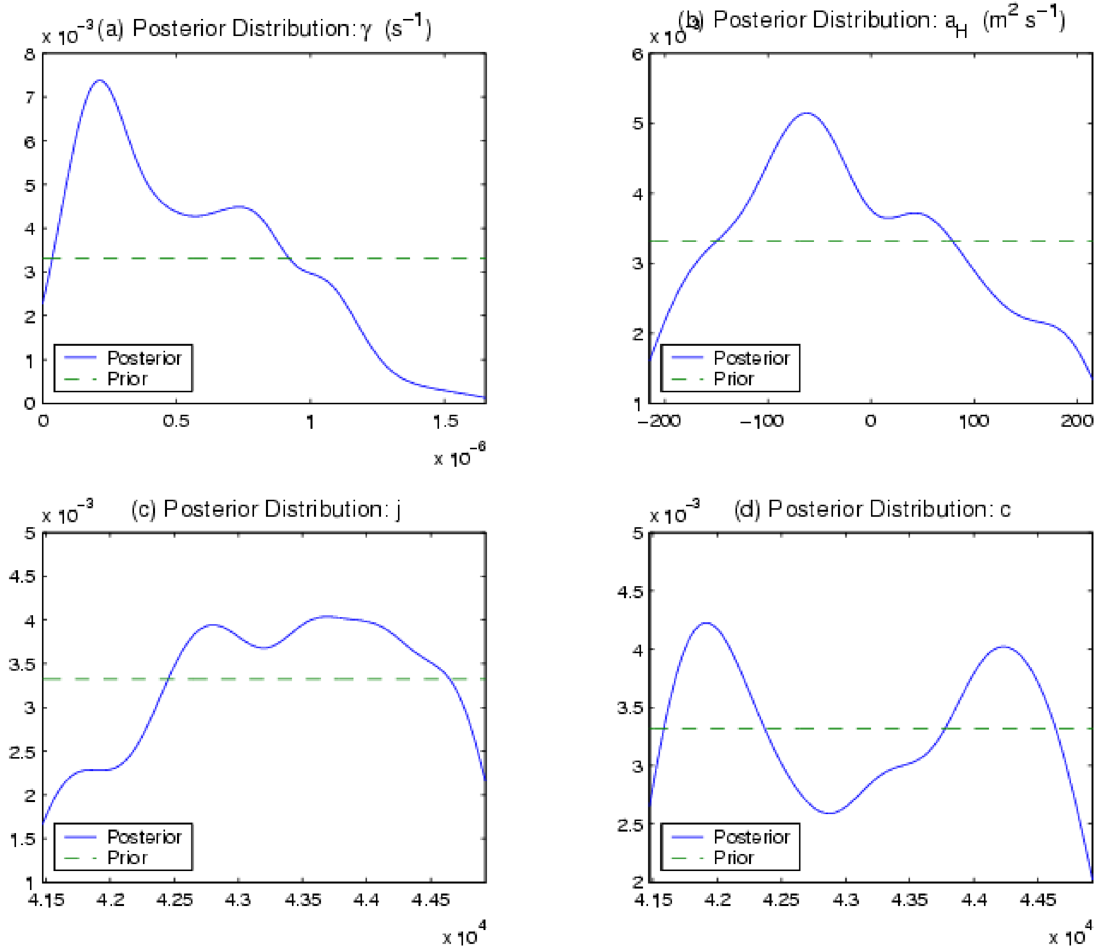
RESULTS: Comparison Between Truth and Posterior Mean



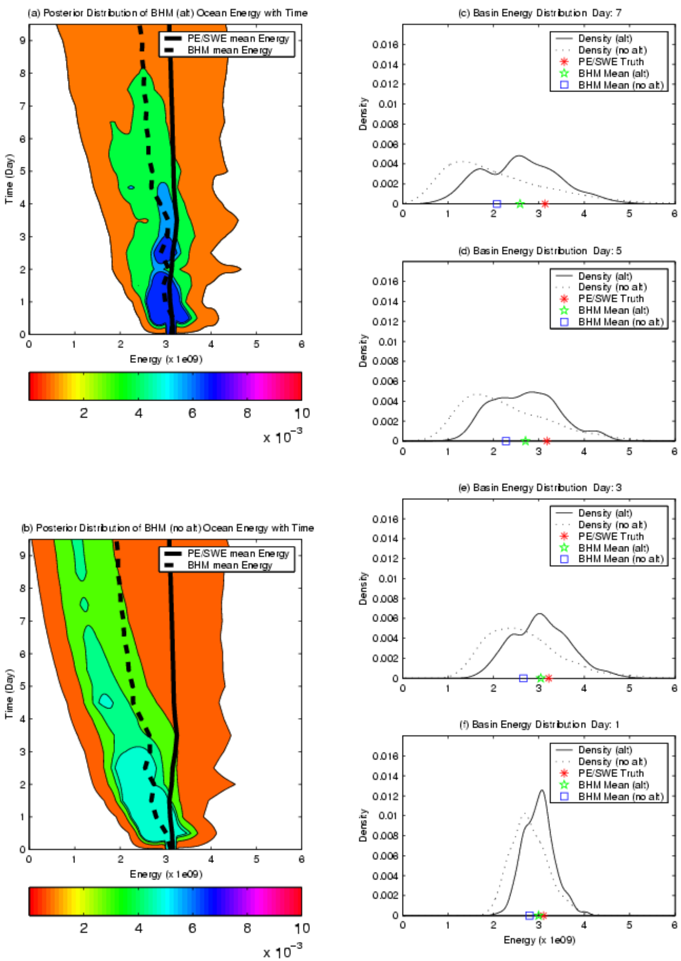
RESULTS: Difference Between Truth and Posterior Mean



RESULTS: Posterior Distributions of Parameters



RESULTS: Posterior Distribution of Kinetic Energy



Joint work with Yong Song (UMC-Stat), Chris Anderson (NOAA, ESRL)

Stochastic Convective Initiation in MM5

Kain-Fritsch (KF) Trigger Function:

$$(T_{i,LCL} + \alpha_i) - T_{i,ENV} > 0 \longrightarrow \text{convective initiation}$$

where for the i-th horizontal grid box,

- $T_{i,LCL}$ - parcel temperature at lifted condensation level (LCL)
- $T_{i,ENV}$ - ambient environmental temperature
- $\alpha_i = 4.64\bar{w}_i^{1/3}$, where $\bar{w}_i$ is the average vertical velocity in the i-th column

Stochastic Trigger Function

Assume that convective initiation follows an independent Bernoulli process in each (horizontal) grid box:

$$y_i = \begin{cases} 1, & \text{if convection initiates} \\ 0, & \text{otherwise} \end{cases}$$

$$p(y_i = 1) = p_i \qquad p(y_i = 0) = 1 - p_i$$

where

$$p_i = \Phi(\theta_0 + \theta_1 T_{i,LCL} + \theta_2 \bar{w}_i^{\theta_3} - \theta_4 T_{i,ENV}),$$

where Φ is the standard normal cumulative distribution function (i.e., probit), and $\boldsymbol{\theta}$ are random parameters:

$$\theta_k \sim N(\mu_k, \sigma_k^2), \quad k = 0, \ldots, 4$$

where $\boldsymbol{\mu} = (-2,\ 1,\ 4.64,\ 1/3,\ 1)'$ (based on climatology and original KF), with σ_k^2 chosen to give fairly vague prior distributions.

Bayesian Estimation

- $\mathbf{z}_p = (z_p(s_1), \ldots, z_p(s_n))'$ - observed cumulative "precipitation" (normalized) over domain of interest with n gridboxes.
- $\mathbf{x} = (x(s_1), \ldots, x(s_n))'$ - true (model) cumulative precipitation (normalized) at n gridboxes.

Interest in Posterior Distribution:

$$[\mathbf{x}, \boldsymbol{\theta} | \mathbf{z}_p] \propto [\mathbf{z}_p | \mathbf{x}][\mathbf{x} | \boldsymbol{\theta}][\boldsymbol{\theta}]$$

We can't evaluate this posterior analytically, nor can we do MCMC due to the complicated non-linear dependencies and high-dimensionality.

Use Importance Sampling Monte Carlo (ISMC)

Monte Carlo

Interest in $f(\boldsymbol{\theta}, \mathbf{x})$

Note:

$$E(f(\boldsymbol{\theta}, \mathbf{x})|\mathbf{z}_p) = \int f(\boldsymbol{\theta}, \mathbf{x}) p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}_p) d\boldsymbol{\theta} d\mathbf{x}$$

Monte Carlo Estimate:

Sample $\boldsymbol{\theta}^j$, $\mathbf{x}^j$, $j = 1, \ldots, N$ from $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}_p)$, then

$$\hat{E}_N(f(\boldsymbol{\theta}, \mathbf{x})|\mathbf{z}_p) = (1/N) \sum_{j=1}^{N} f(\boldsymbol{\theta}^j, \mathbf{x}^j)$$

.

In our example, we can't sample from this distribution!

Importance Sampling Monte Carlo

Sample from proposal distribution: $q(\mathbf{x}, \boldsymbol{\theta}|\mathbf{Z}_p)$, then

$$E(f(\boldsymbol{\theta}, \mathbf{x})|\mathbf{z}_p) = \int f(\boldsymbol{\theta}, \mathbf{x}) \frac{p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z})}{q(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z})} q(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}_p) d\boldsymbol{\theta} d\mathbf{x}$$
$$= \int f(\boldsymbol{\theta}, \mathbf{x}) \; w \; q(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}_p) d\boldsymbol{\theta} d\mathbf{x},$$

where unnormalized importance weights are:

$$w = p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}_p)/q(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}_p)$$

Taking samples $\mathbf{x}^j, \boldsymbol{\theta}^j$ from the proposal, the ISMC estimate is

$$\hat{E}_N(f(\mathbf{x}, \boldsymbol{\theta})|\mathbf{z}_p) = \frac{1}{N} \sum_{j=1}^{N} \tilde{w}_j f(\mathbf{x}^j, \boldsymbol{\theta}^j),$$

where normalized ISMC weights are:

$$\tilde{w}_j \equiv \frac{w_j}{\Sigma_{k=1}^{N} w_k}$$

ISMC for Stochastic Parameterization

Analogous to particle filtering, We choose proposal distribution:

$$q(\boldsymbol{\theta}, \mathbf{x}|\mathbf{z}_p) = p(\mathbf{x}, \boldsymbol{\theta}) = m(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- Sample $\boldsymbol{\theta}^j, \quad j = 1, \ldots, N$ from prior $p(\boldsymbol{\theta})$
- Obtain samples $\mathbf{x}^j$ by running MM5 with parameters $\boldsymbol{\theta}^j$
- Unnormalized ISMC weights are given by data model: $w_j = p(\mathbf{z}_p|\mathbf{x}^j)$

  In our stochastic model, data model is:

$$w_j \propto \exp\{-\frac{1}{2\tau} tr\{(\mathbf{Z}_p - \mathbf{X}_j\mathbf{Q}_j)'(\mathbf{Z}_p - \mathbf{X}_j\mathbf{Q}_j)\}\}$$

  where

  – $\mathbf{Z}_p$, $\mathbf{X}_j$ are matrix forms of $\mathbf{z}_p$, $\mathbf{x}^j$

  – $\mathbf{Q}_j$ is the Procrustes transformation matrix

Posterior Distributions and Implementation

With ISMC weights, we can obtain estimates of moments of the posterior distribution of $\boldsymbol{\theta}$ and $\mathbf{x}$: e.g., $\hat{E}(\theta_i|\mathbf{z}_p) = \frac{1}{N}\Sigma_{j=1}^{N}\tilde{w}_j\theta_i^j$

Or, we can use kernel density estimation to approximate the posteriors: e.g.,

$$p(\theta_i|\mathbf{z}_p) \approx \frac{1}{N}\sum_{j=1}^{N}\tilde{w}_j k(\theta_i^j, \gamma)$$

where $k(\theta_i^j, \gamma)$ is a kernel function centered at θ_i^j with kernel bandwidth γ.

Two approaches for utilizing these posteriors:

- brute force: Run model many times, one for each sample $\boldsymbol{\theta}^k \sim [\boldsymbol{\theta}|\mathbf{z}_p]$ (measures of uncertainty/ expensive)
- time step: Sample $\boldsymbol{\theta}^k$ from posterior at each time step in a single model run. (no measure of uncertainty/ cheap)

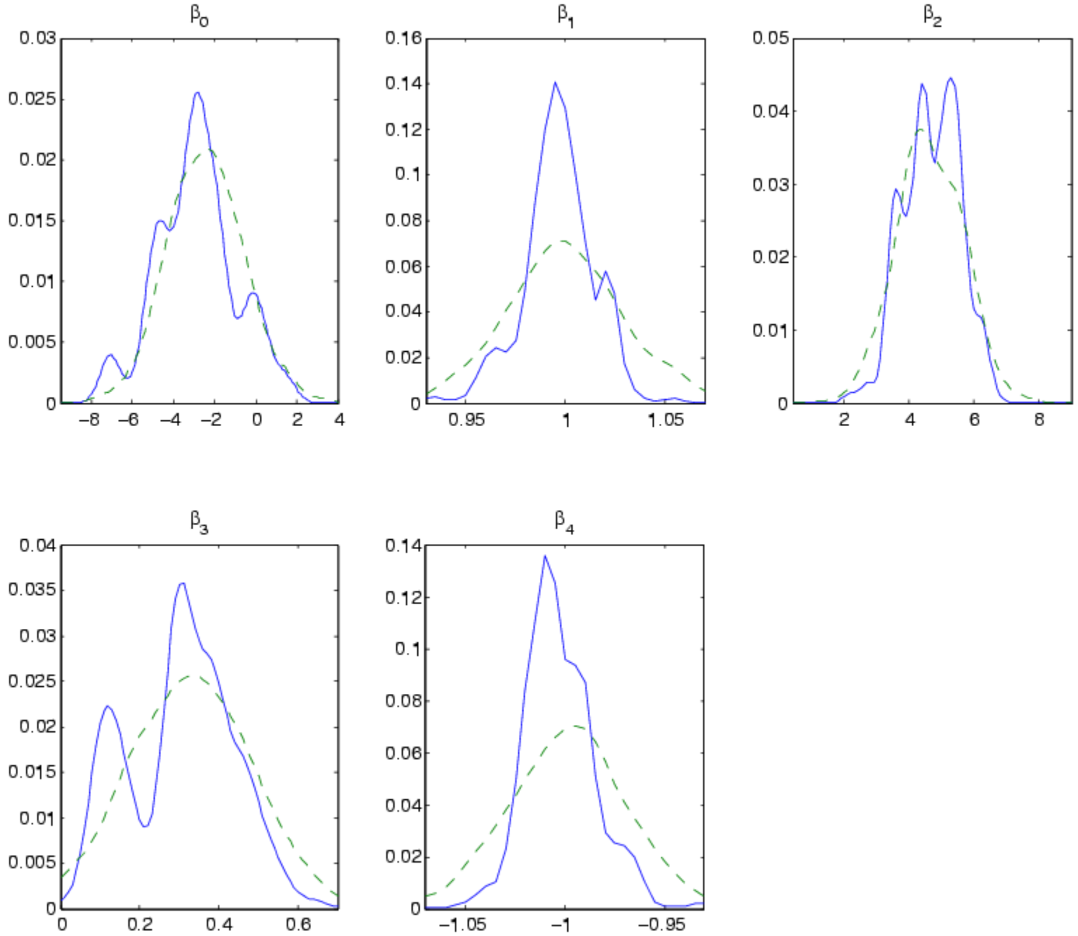
Experiment

MM5:

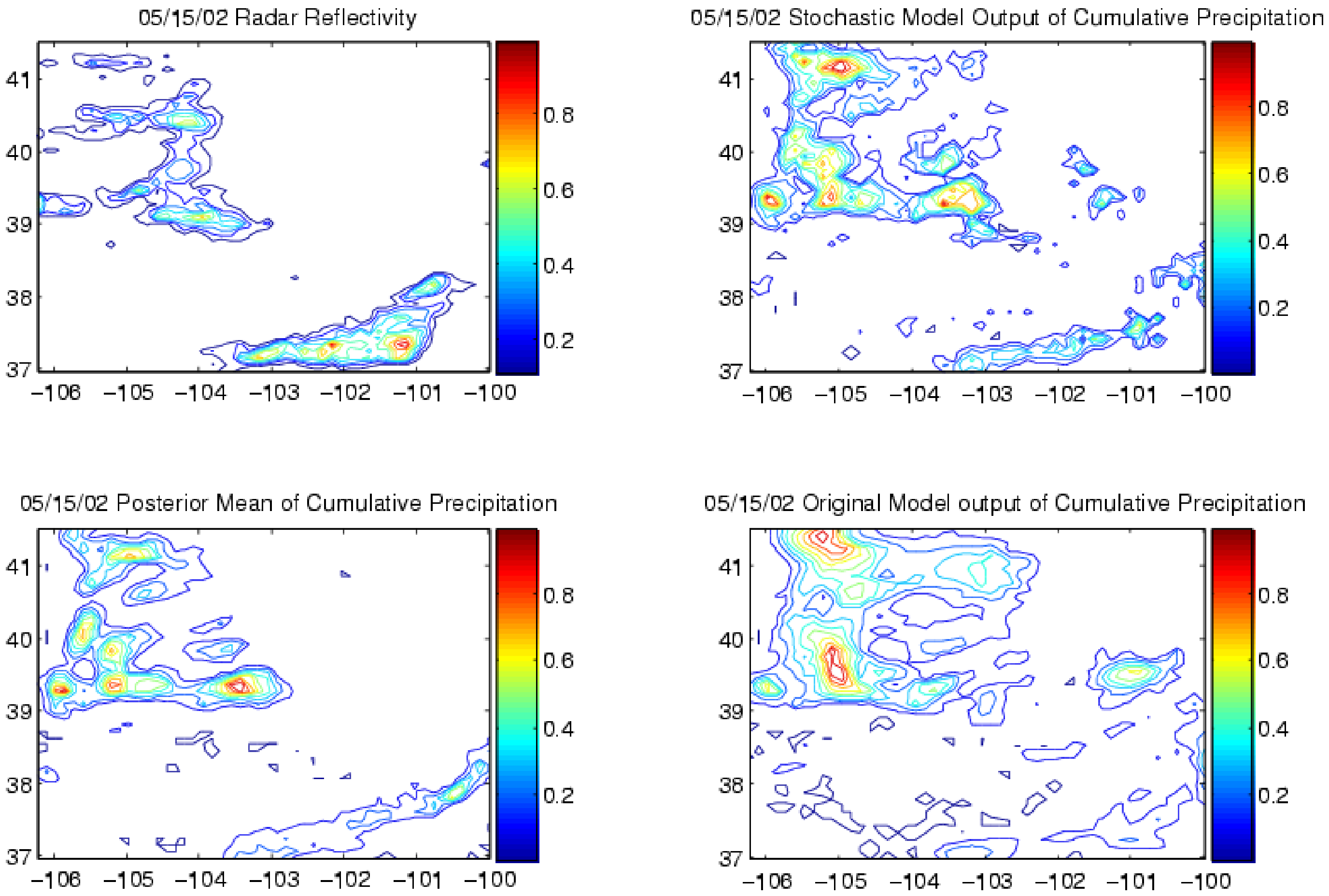
- Initialization: IHOP (International H2O Project), May 15, 2002
- Centered on Goodland, KS
- Domain: 70×70 (10-km resolution)
- Vertical: 38 sigma levels
- 4 hour run, 40 s time steps
- Simple moisture scheme

Radar Data: NWS Goodland, KS

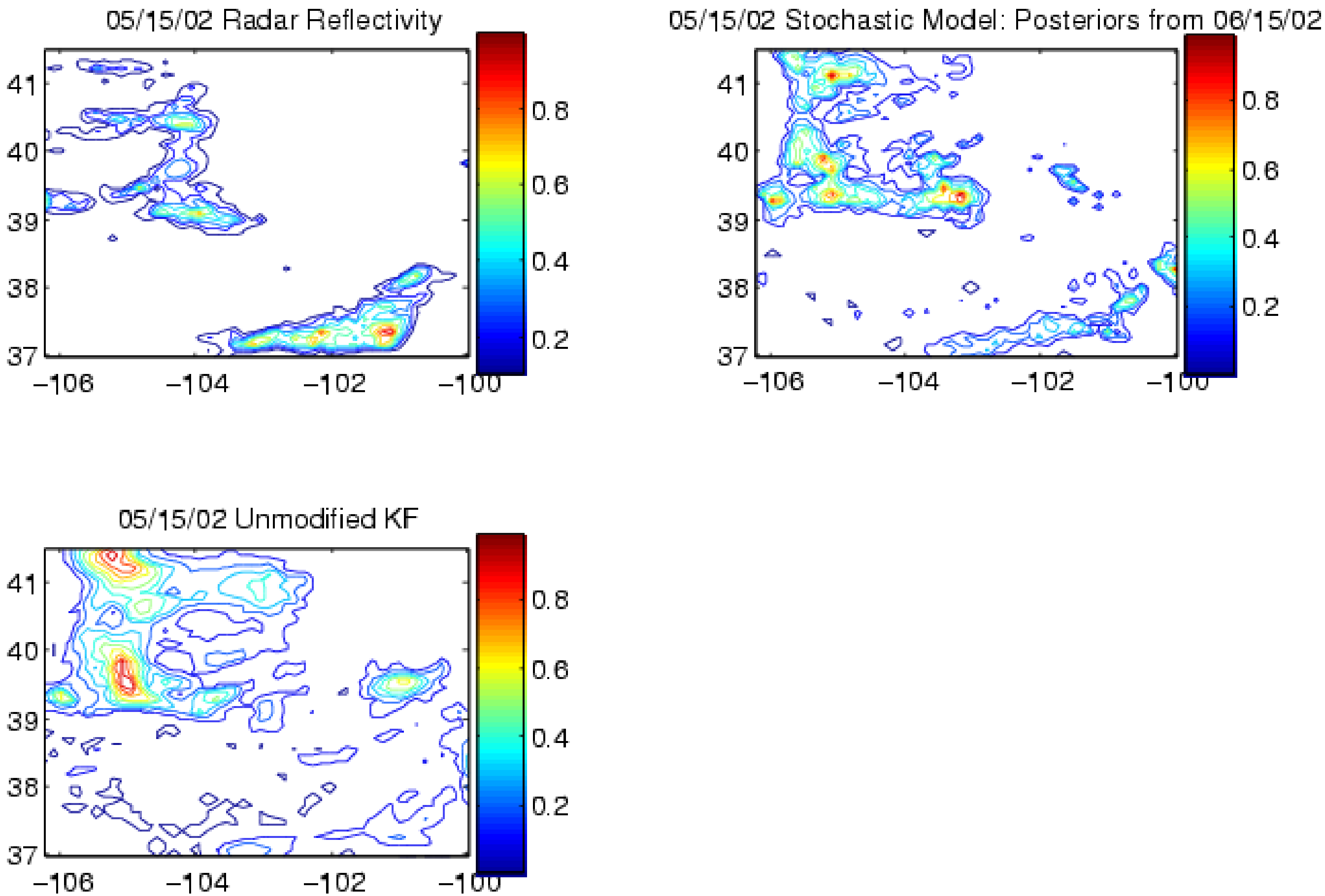
Prior and Posterior Parameter Distributions



Radar, Posterior Mean, Simulation, Original Model



Forecast (“Hindcast”) Example



Conclusion

- The hierarchical Bayesian paradigm is ideal for managing uncertainty in data, process and parameters.
- Over the last 5-10 years we have demonstrated that, in the presence of data, HBM can be an effective approach for accounting for statistical parameterization in dynamical models.
- We are just now being able to apply this methodology to “realistic” problems in the atmospheric (and other) sciences.
- Much work needs to be done with regards to:
  - extensions to 4-d domains
  - multivariate systems
  - parallelization
  - Importance Sampling degeneracy problem
- The future of HBM is bright!