Inferring likelihoods and climate system characteristics from climate models and multiple tracers

Murali Haran

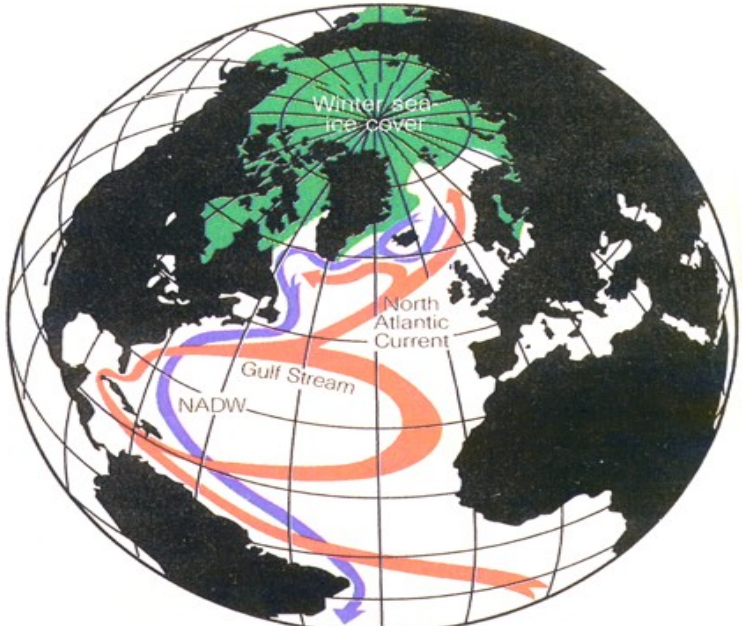
Department of Statistics
Pennsylvania State University

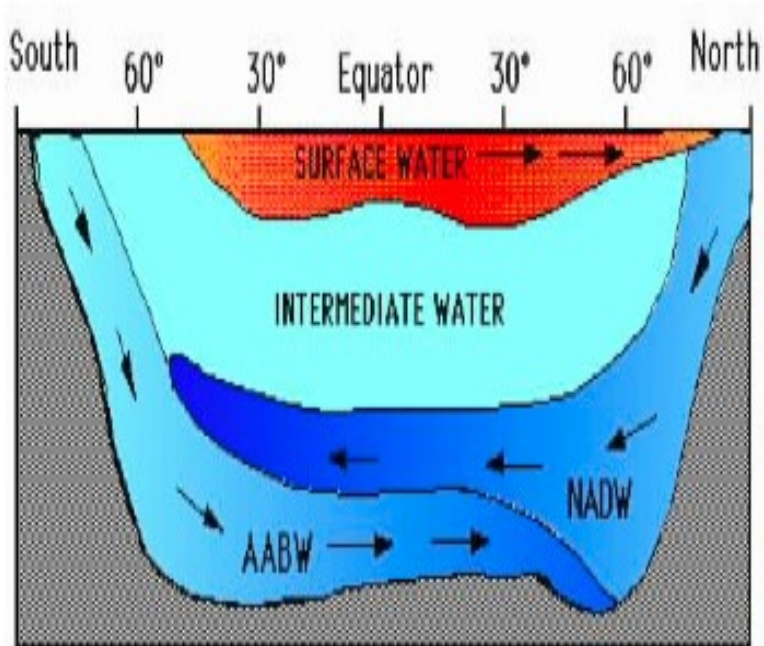
(joint work with Sham Bhat (Statistics), Roman Tonkonojenkov (Geosciences) and Klaus Keller (Geosciences))

Western Michigan University March 2010

The MOC and climate change

- What is the risk of human induced climate change?
- Meridional Overturning Circulation (MOC): Movement of water from equator to higher latitudes, deep water masses created by cooling of water in Atlantic, resulting in sea ice formation, denser salt water, which sinks and causes ocean circulation.





The MOC

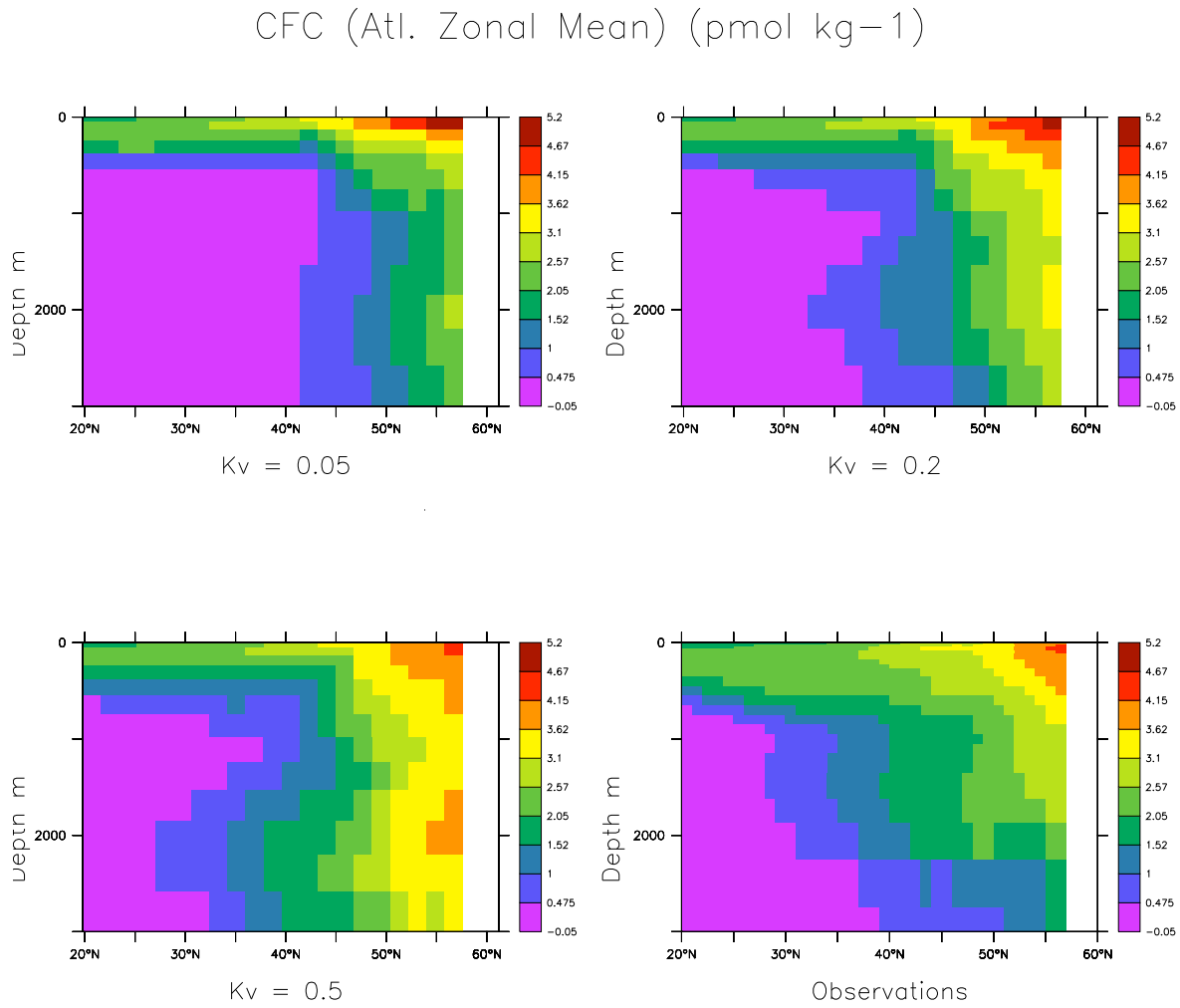
- MOC weakening results in disruptions in the equilibrium state in the climate, may lead to major temperature and precipitation changes and shifts in terrestrial ecosystems.
- The potential collapse of the meridional overturning circulation (MOC) is therefore an example of potentially catastrophic climate change.
- How can we make projections about the MOC? Climate scientists rely on sophisticated *deterministic* climate models to study such phenomena and make projections.
- Climate models have many unknown parameters (inputs).
- A key source of uncertainty in MOC projections is uncertainty about the parameter background ocean vertical diffusivity, $\mathbf{K_v}$.

Learning about K_v

How can we estimate K_v?

- K_v is a model parameter which quantifies the intensity of vertical mixing in the ocean, cannot be measured directly.
- Two sources of indirect information:
  - Observations of two ocean tracers, both provide information about K_v: Carbon-14 (^{14}C) and Trichlorofluoromethane (CFC11) collected in the 1990s (latitude, longitude, depth), zonally averaged.
  - Climate model output at different values of K_v from the University of Victoria(UVic) Earth System Climate Model (Weaver et. al. 2001).
- Latitude between -80 S and 60 N, depths from 0 to 3000m.
- Data size: 3706(observations); 5926(model) per tracers.

CFC example

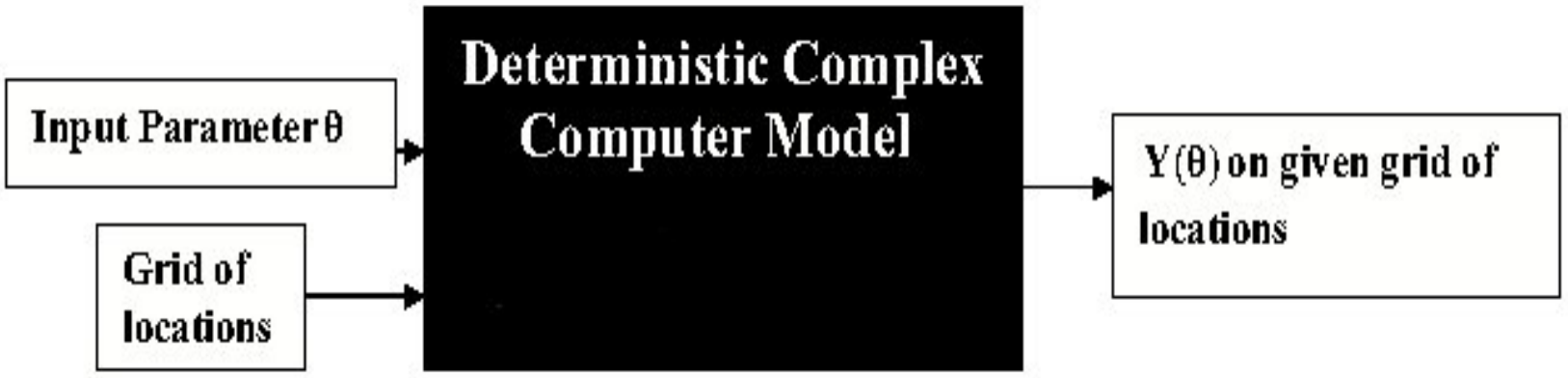


- Bottom right: observations
- Remaining plots: climate model output at 3 settings of K_v.

Challenges

1. No direct connection between observations and climate parameter, need to rely on climate model runs to obtain a probability model connecting observations and climate parameter $\mathbf{K_v}$.
2. The climate model is very computationally intensive. Hence, can only be run at a few different settings. Need some form of interpolation ('emulation') to allow for inference.
3. Large spatial data sets: poses computational challenges for inference.
4. Combining information from multiple tracers, CFC-11, C14: need flexible, computationally tractable models for multivariate spatial data.

Computer model emulation



- **Emulation** involves replacing a complicated computer model with a simpler (usually stochastic) approximation.
- Sacks et. al. (1989) introduced a linear Gaussian process model as an emulator for a complex nonlinear function. Related work by: Currin, Mitchell, Morris, Ylvisaker (1991), Bayarri et al (2007;2008) and many others.

Gaussian processes: basics

- Model random variable at location $\mathbf{s}$ by

$$Z(\mathbf{s}) = X(\mathbf{s})\beta + w(\mathbf{s}), \text{ for } \mathbf{s} \in D \subset \mathbb{R}^d$$

- $\{w(\mathbf{s}), \mathbf{s} \in D\}$ is (infinite dimensional) Gaussian process.
- Let $\mathbf{w} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))^T$, $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$.
  Predictions at new locations: $\mathbf{Z}^* = (Z(\mathbf{s}_1^*), \ldots, Z(\mathbf{s}_m^*))^T$.

$$\mathbf{w} \mid \xi \sim N(0, \Sigma(\xi)), \quad \xi \text{ are covariance parameters}$$

- $\mathbf{Z}^* \mid \mathbf{Z}$ is normal ($_1$,$_2$ correspond to mean, var of $\mathbf{Z}, \mathbf{Z}^*$):

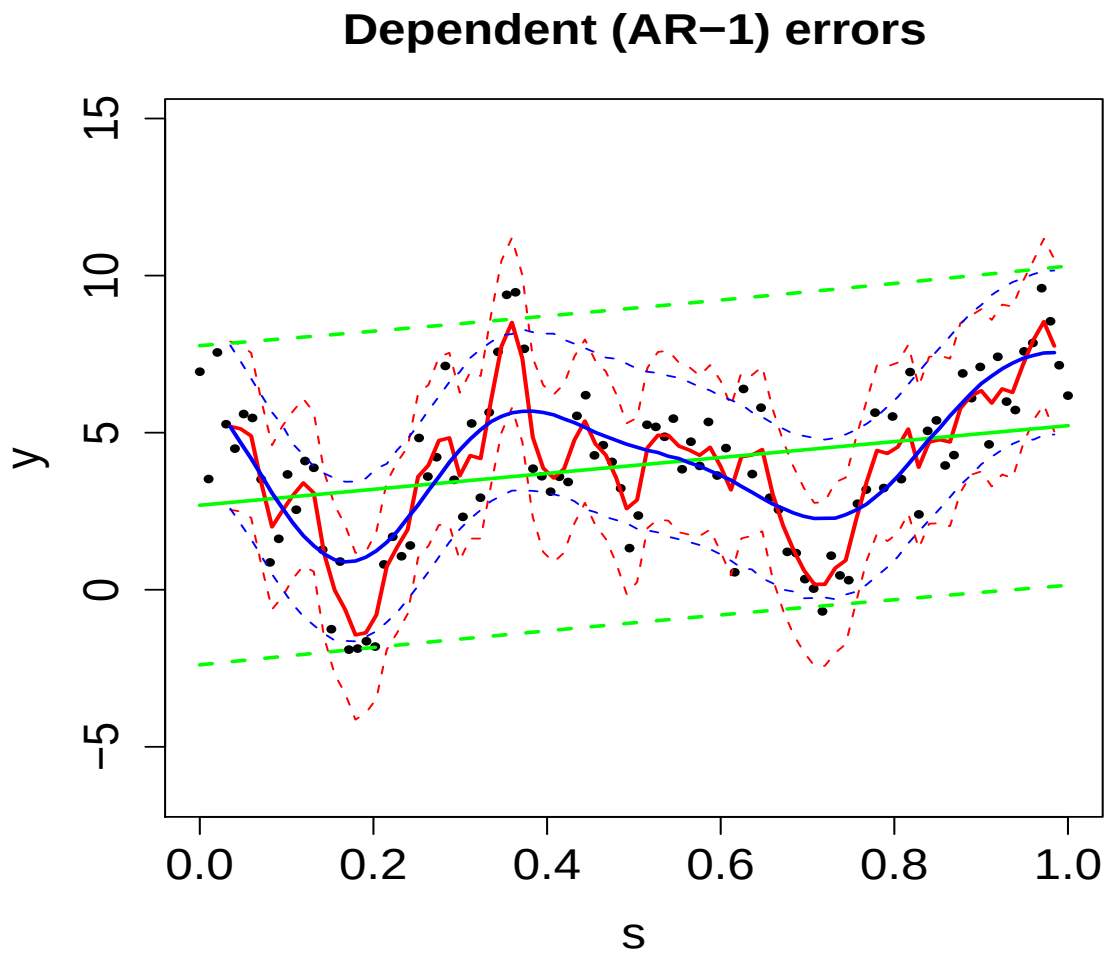
$$E(\mathbf{Z}^* \mid \mathbf{Z}, \beta, \xi) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{Z} - \mu_1)$$

$$\text{Cov}(\mathbf{Z}^* \mid \mathbf{Z}, \beta, \xi) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Gaussian processes (contd)

- Standard assumption: Assume process is stationary and covariance function that determines $\Sigma(\boldsymbol{\xi})$ belongs to Matérn family. Important special cases: gaussian (infinitely differentiable), exponential (no derivatives).
- Predictions: obtain estimates $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}$.
  - ML inference: plug $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\beta}}$ into conditional distribution $\mathbf{Z}^* \mid \mathbf{Z}$.
  - Bayesian inference: find posterior $\pi(\boldsymbol{\xi}, \boldsymbol{\beta} \mid \mathbf{Z})$ and obtain *posterior predictive distribution* $\pi(\mathbf{Z}^* \mid \mathbf{Z})$, integrating with respect to $\boldsymbol{\beta}, \boldsymbol{\xi}$ over $\pi(\boldsymbol{\xi}, \boldsymbol{\beta} \mid \mathbf{Z})$.
  - Very convenient and very flexible models for both spatially dependent processes and complicated functions.

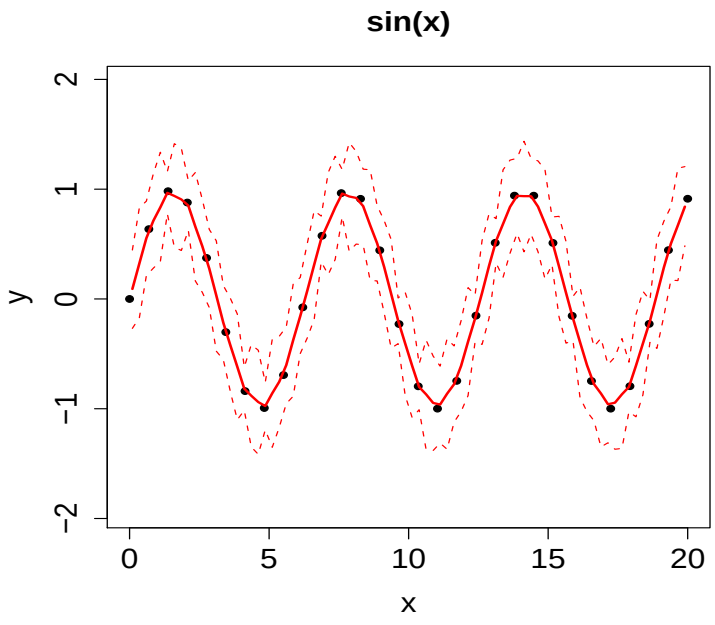
GP model for dependence: toy 1-D example

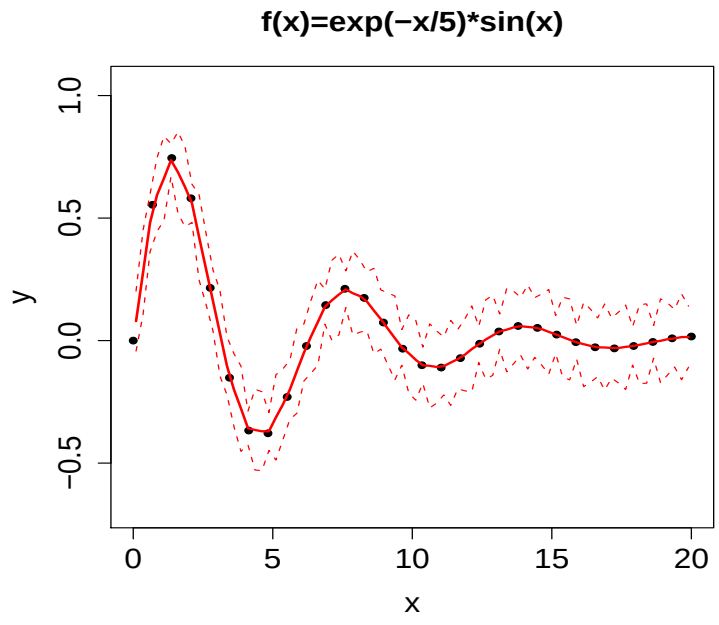


Black: 1-D AR-1 process simulation. Green: independent error.

Red: GP with exponential, Blue: GP with gaussian covariance.

GP model for emulation





Functions: $f(x) = \sin(x)$ and $f(x) = \exp(-x/5)\sin(x)$.

Both were fit with linear GP model, $f(x) = \alpha + \epsilon(x)$, where $\{\epsilon(x),\ x \in (0, 20)\}$ is a GP, α is just a constant mean.

Statistical inference for climate model

- Notation: $Z(\mathbf{s})$: physical observations, $Y(\mathbf{s}, \theta)$: model output at location $\mathbf{s}$=(latitude, depth), and climate parameter θ.
- Climate model: Complex and requires long time to run.
- This is a computer model calibration problem.
- **Data Sources**: Observations for ^{14}C/CFC11: $\mathbf{Z}_1, \mathbf{Z}_2$.
- Climate model runs at several values of θ ($\mathbf{K_v}$): $\mathbf{Y}_1, \mathbf{Y}_2$.
- **Goal**: Inference for climate parameter θ.

Bayesian model calibration

- Want to determine parameter settings that are 'most likely' given **Y**, **Z** (vector obtained by stacking columns of matrix of $Y(\mathbf{s}, \theta)$, $Z(\mathbf{s})$ respectively).
- Kennedy and O'Hagan (2001) developed a fully Bayes approach for 'computer model calibration'. Sanso et al. (2007) used a variant for climate parameter inference.
- Assumption: a "true" set of climate parameters θ^* exists.

$$Z(\mathbf{s_i}) = Y(\mathbf{s_i}, \theta^*) + \epsilon_i.$$

  Note: there is no true θ^*, so perhaps more appropriate to think of it as a fitted value (Bayarri, Berger et al. 2007).
- Model **Y** and **Z** jointly. Model **Y** as a Gaussian process, with dependence in climate parameter (θ) space.
- Separable covariance between $\mathbf{s}, \theta$ dimensions.

GP for climate model emulation

- Unlike the toy example, the output from the climate model is much more complicated — for each $\mathbf{K_v}$ we have two related spatial fields (not a single point). We fit a Gaussian process model to the climate model output: covariance depends on both distance in physical space ($\|\mathbf{s}_1 - \mathbf{s}_2\|$) as well as in climate parameter space ($\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$).
- We can now use this GP model instead of the very complicated climate model — this provides a connection between $\mathbf{K_v}$ and the climate model output, in this case the tracers CFC-11 and C-14.
- Model for the observed CFC-11 and C-14: can use the GP model + allow for additional sources of structural uncertainty and bias.

Calibration with multiple spatial fields

- How can we combine information from multiple tracers (^{14}C, CFC11) in a flexible manner to infer $\mathbf{K_v}$?
- Two stage approach to obtain posterior of $\boldsymbol{\theta}$:
  - Model relationship between $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ and $\boldsymbol{\theta}$ via emulation of model output $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$.
  - Use observations $\mathbf{Z}$ to infer $\boldsymbol{\theta}$ (parameter of interest).
- Model $(\mathbf{Y}_1, \mathbf{Y}_2)$ as a hierarchical model: $\mathbf{Y}_1|\mathbf{Y}_2$ and $\mathbf{Y}_2$ as Gaussian processes. (following Royle and Berliner (1999)).

$$\mathbf{Y}_1 \mid \mathbf{Y}_2, \beta_1, \xi_1, \gamma \sim N(\mu_{\beta_1}(\boldsymbol{\theta}) + \mathbf{B}(\gamma)\mathbf{Y}_2, \Sigma_{1.2}(\xi_1))$$

$$\mathbf{Y}_2 \mid \beta_2, \xi_2 \sim N(\mu_{\beta_2}(\boldsymbol{\theta}), \Sigma_2(\xi_2))$$

- $\mathbf{B}(\gamma)$ is a matrix relating $\mathbf{Y}_1$ and $\mathbf{Y}_2$, with parameters γ.
- βs, ξs are regression, covariance parameters.

Calibration with multiple spatial fields [cont'd]

- Based on fitted GP, obtain predictive distribution at locations of observations. This serves as the emulator.
- We then model the observations by adding measurement error and a model discrepancy term to the GP emulator:

$$\mathbf{Z} = \eta(\mathbf{Y}, \boldsymbol{\theta}) + \delta(\mathbf{Y}) + \epsilon$$

  where $\delta(\mathbf{Y}) = (\delta_1 \ \delta_2)^T$ is the model discrepancy, $\epsilon = (\epsilon_1 \ \epsilon_2)^T$ is the observation error.
- Inference on $\boldsymbol{\theta}$ performed using Markov chain Monte Carlo (MCMC) to estimate $\pi(\boldsymbol{\theta} \mid \mathbf{Z}, \mathbf{Y})$, integrating over remaining parameters.

Computational issues

- Matrix computations are $\mathcal{O}(N^3)$, where N is the number of observations. If we are not careful about modeling, N could be on the order of tens of thousands.
- Need long MCMC runs since there may be multimodality issues, and the algorithm mixes slowly.
- Used reduced rank approach based on kernel mixing (Higdon, 1998): continuous process created by convolving a discrete white noise process with a kernel function.
- Special structure + Sherman-Woodbury-Morrison identity used to reduce matrix computations.
- In MLE (optimization) step: take advantage of structure of hierarchical model to reduce computations.

Kernel mixing for spatial processes

- Model spatial dependence terms ($w(\mathbf{s})$) via kernel mixing of white noise process (Higdon, 1998, 2001).
- New process created by convolving a continuous white noise process with a kernel, k, which is a circular normal.

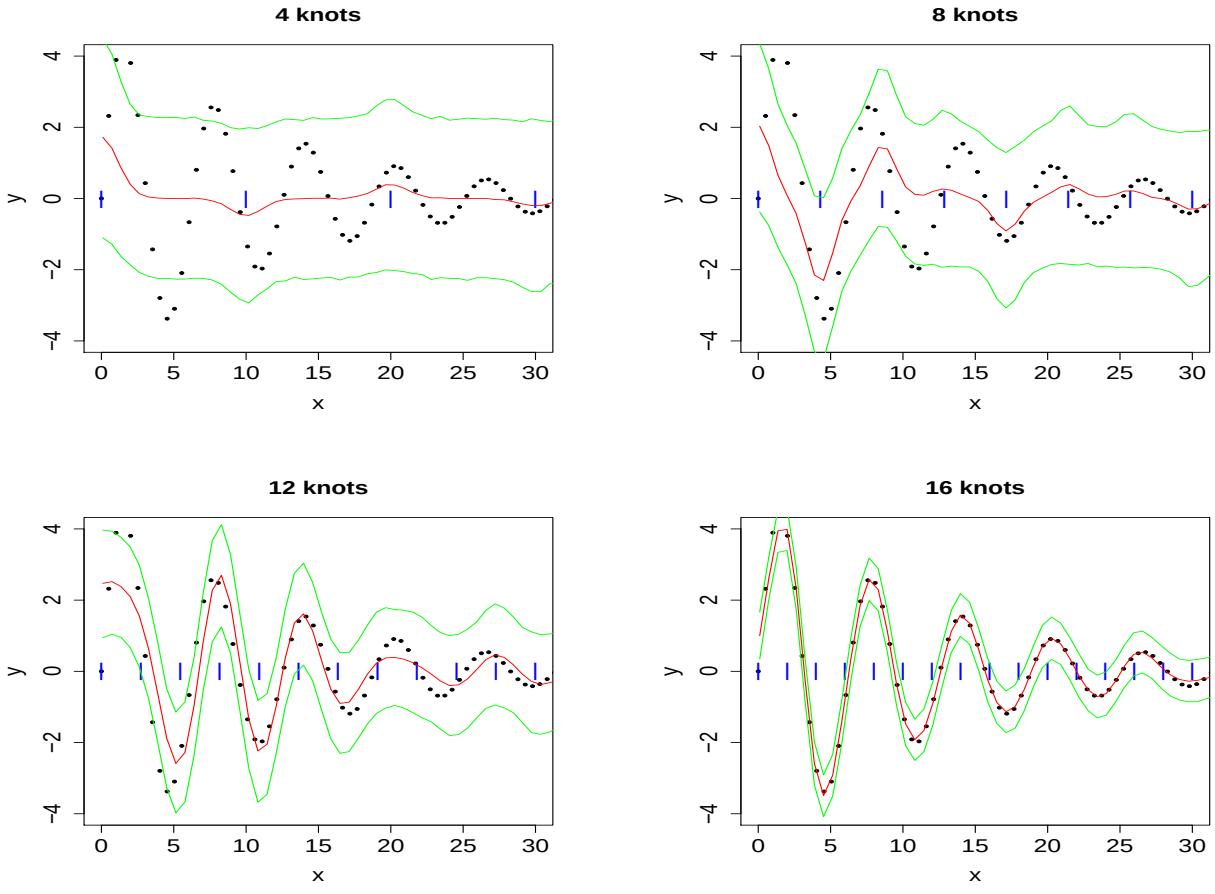
$$w(\mathbf{s}) = \int_D k(\mathbf{u} - \mathbf{s}) z(\mathbf{u}) d\mathbf{u}.$$

- Replace original GP by a finite sum approximation $\mathbf{z}$ defined on a lattice $\mathbf{u}_1, \ldots, \mathbf{u}_J$ (knot locations).

$$w(\mathbf{s}) = \sum_{j=1}^{J} k(\mathbf{u}_j - s) z(\mathbf{u}_j) + \mu(\mathbf{s}),$$

- Flexible: easily allows for non-stationarity and nonseparability. e.g. if k varies in space, have non-stationary process.

Kernel mixing for spatial processes (cont'd)



- Dimension reduction: Computation involves only the J random variables $z_1, \ldots, z_J$ at the locations $\mathbf{u}_1, \ldots, \mathbf{u}_J$.
- Figures are for 4, 8, 12, and 16 knots.

Matrix identities

- Kernel mixing can be used to induce special matrix forms that permit very fast computations. In fact, we ignore the latent variables and simply use the kernel mixing formulation to obtain matrices of special forms.
- Sherman-Woodbury-Morrison identity: Suppose a matrix can be written in the form $A + UCV$, where A is of dimension $N \times N$, U is dimension $N \times J$, V is dimension $J \times N$, and C is dimension $J \times J$. Its inverse is rewritten as:

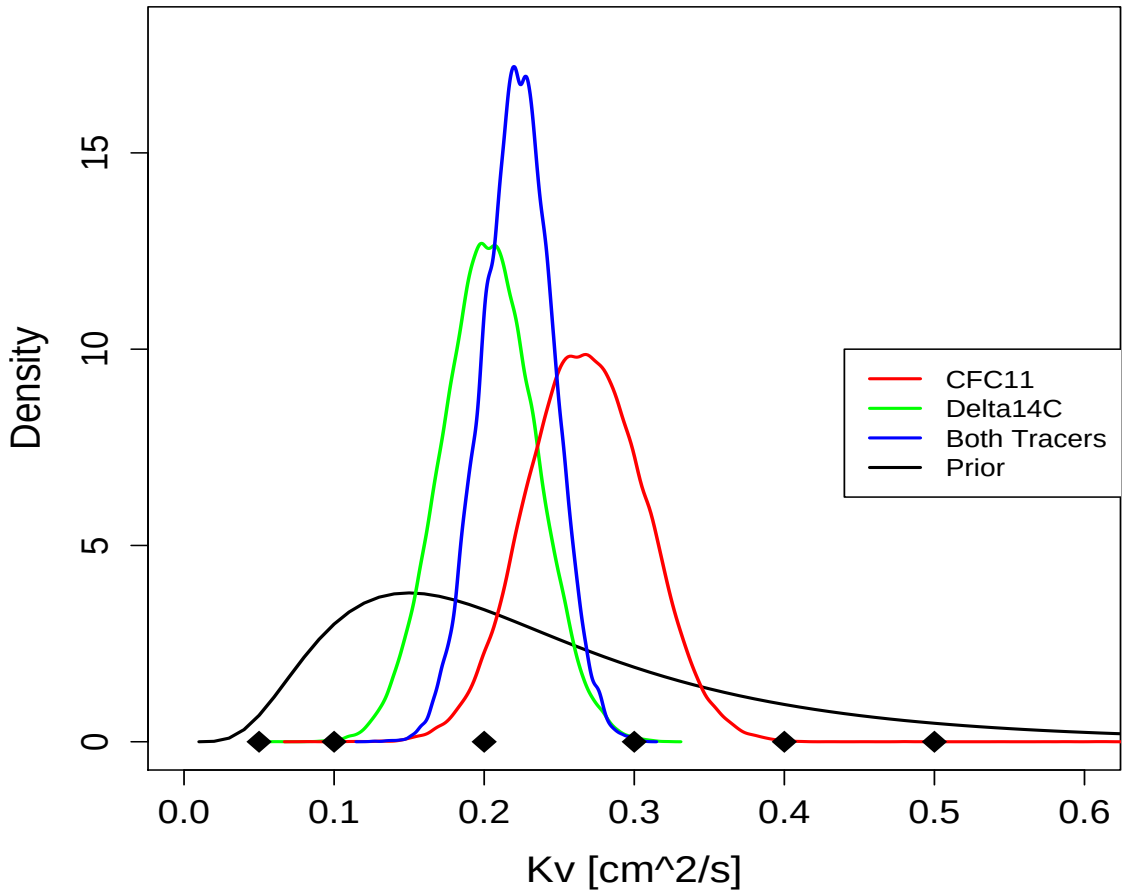
$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}.$$

This involves inversions of matrices of dimension $J \times J$ rather than $N \times N$. (our e.g. $J = 190$ versus $N =$4,500.)

K_v inference: summary

- Taken the climate model output, CFC-11 and C14 spatial fields at several values for $\mathbf{K_v}$, and fit a very flexible GP model for bivariate spatial fields.
- Now, assume this GP model + model for error, discrepancy is the model for the observations of CFC-11 and C14.
- Since we have a probability model, we can perform inference for $\mathbf{K_v}$ based on the data. That is, we can use statistical techniques to learn about the values of $\mathbf{K_v}$ most compatible with all the information we have. This information will be in the form of a (posterior) probability distribution for $\mathbf{K_v}$.
- Computational considerations are important in modeling (hierarchical structure + kernel mixing approach).

Results for K_v inference



Probability density functions (pdfs): the prior pdf (assumption *before* using data), and posterior pdfs (*after* using the tracers.)

Summary

1. Our approach is to perform inference in two stages:
   - Obtain a probability model connecting CFC-11, C-14 tracer observations to $\mathbf{K_v}$ by fitting a Gaussian process model to climate model runs.
   - Using this probability model, infer a posterior density for $\mathbf{K_v}$ from the observations.
2. We model multivariate spatial data via a flexible hierarchical structure.
3. We use kernel mixing to obtain patterned covariances, making computations tractable for large data sets.

Our approach allows us to infer $\mathbf{K_v}$ based on all the climate model output and observations, modeling the tracers jointly. We can use $\mathbf{K_v}$ in computer models to project the MOC.

Future work

- Many open problems, research avenues including:
  - Combining information from multiple climate models: Multiresolution/multiscale modeling ideas, Bayesian model averaging.
  - Flexible covariance functions, non-stationarity.
  - Combining information from several tracers (e.g. 5–10).
- Other projects that can potentially borrow some of this methodology:
  - Atmospheric Science: Estimating mean temperature fields over the past millenia using proxies and climate models.
  - Infectious disease: inferring infectious disease dynamics from sparse observations and dynamic models.

Some references

- Kennedy, M.C. and O'Hagan, A.(2001), Bayesian calibration of computer models, *JRSS(B)*.
- Sanso, B. and Forest, C.E. and Zantedeschi, D (2008) , Inferring Climate System Properties Using a Computer Model, *Bayesian Analysis (with discussion)*.
- Higdon (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean, *Environmental and Ecological Statistics*.
- Royle, J.A. and Berliner, L.M. (1999) A hierarchical approach to multivariate spatial modeling and prediction, *Journal of Agricultural, Biological, and Environmental Statistics*.
- Bhat, K.S., Haran, M., Tonkonojenkov, R., Keller, K. (2009) "Inferring likelihoods and climate system characteristics using climate models and multiple tracers."

SUNDRIES

Joint modeling approach: pros and cons

- Bayesian machinery and MCMC makes it relatively easy to write down a reasonable joint model.
- Modelers (especially Bayesians) often argue that having a joint model is critical. Pragmatic argument: propogation of uncertainty through the model.
- However, joint model adds computational burdens. Also leads to identifiability issues. Hence, in order to build a joint model: have to resort to unrealistic covariance assumptions and heavy spatial and temporal aggregation of both observations and model output.

Alternative: Two stage approach

- Two stage approach to obtain posterior of θ:
  - Model the **Y**'s stochastically to 'infer a likelihood', connecting θ to **Y**.
  - Model **Z** using fitted model from above, with additional errors, biases, to infer θ (along with errors, biases.)
- Model **Y** as a Gaussian process emulator, with mean a linear function of θ.

$$\mathbf{Y} \mid \beta, \xi \sim N(\mu_\beta(\theta), \Sigma(\xi)),$$

- ξ is the set of covariance parameters, covariance function assumed to be separable among **s**, t, and θ.
- Covariance parameters:
  - Maximum likelihood estimates by optimization.
  - Bayesian approach: obtain posterior via MCMC.

Two stage approach (cont'd)

- For location **s** at a given value of θ, we can then obtain the predictive distribution $\pi(\mathbf{Z}(\theta)^* \mid \mathbf{Y})$, multivariate normal for a *given* $\hat{\xi}, \hat{\beta}$ (MLE or posterior mean/mode). Otherwise this is not in closed form.
- This multivariate normal is our approximate probability model $\hat{\eta}$, written explicitly with mean and variance as functions of θ from conditional distribution.

$$\mathbf{Z} = \hat{\eta}(\mathbf{Z}^* \mid \theta^*, \mathbf{Y}) + \delta + \epsilon,$$

- where δ is the model error term and ϵ is observation error.
- $\epsilon \sim N(0, \psi I)$ and δ is modeled as a Gaussian process, ϵ and δ are assumed to be independent. Strong prior information for ϵ can help identify the errors.
- We can now perform inference on θ^*.

Observations

- Our approach is perhaps counter to standard Bayesian modeling philosophy: instead of a coherent joint model, we are fitting models stagewise.
- Principle: If we had a likelihood, $\mathcal{L}(\mathbf{Z}; \theta)$, we could perform inference for θ based on data $\mathbf{Z}$.
- Here: We are using climate model output ($\mathbf{Y}$) to 'infer' this likelihood and then perform standard likelihood-based inference. Intuitively: separate problems (see "Subjective likelihood" [Rappold, Lavine, Lozier, 2005.])
- Our approach can be seen as a way of 'cutting feedback' (Best et al. 2006; Rougier, 2008). Advantages:
  - Protecting emulator from a poor model of climate system.
  - Modeling emulator separately to facilitate careful evaluation of emulator. (Rougier, 2008).

More advantages

- Computational advantages allow for relaxing unreasonable assumptions, e.g. no need to assume same covariance for both spatiotemporal dependence and observation error.
- Potentially helps with identification of variance/covariance components since not all parameters are being estimated/sampled at once; parameters estimated from first stage are fixed.
- Concern: are we ignoring crucial variability in parameter estimates by not propogating it as in the Bayesian formulation? Data sets/problems considered so far: not obvious that this is the case. (Also, cannot compare results for the large multivariate spatial data since cannot fit the joint model.)

Kernel mixing for climate model output

- Extend kernel and knot process $\mathbf{z}$ to t and θ dimensions:

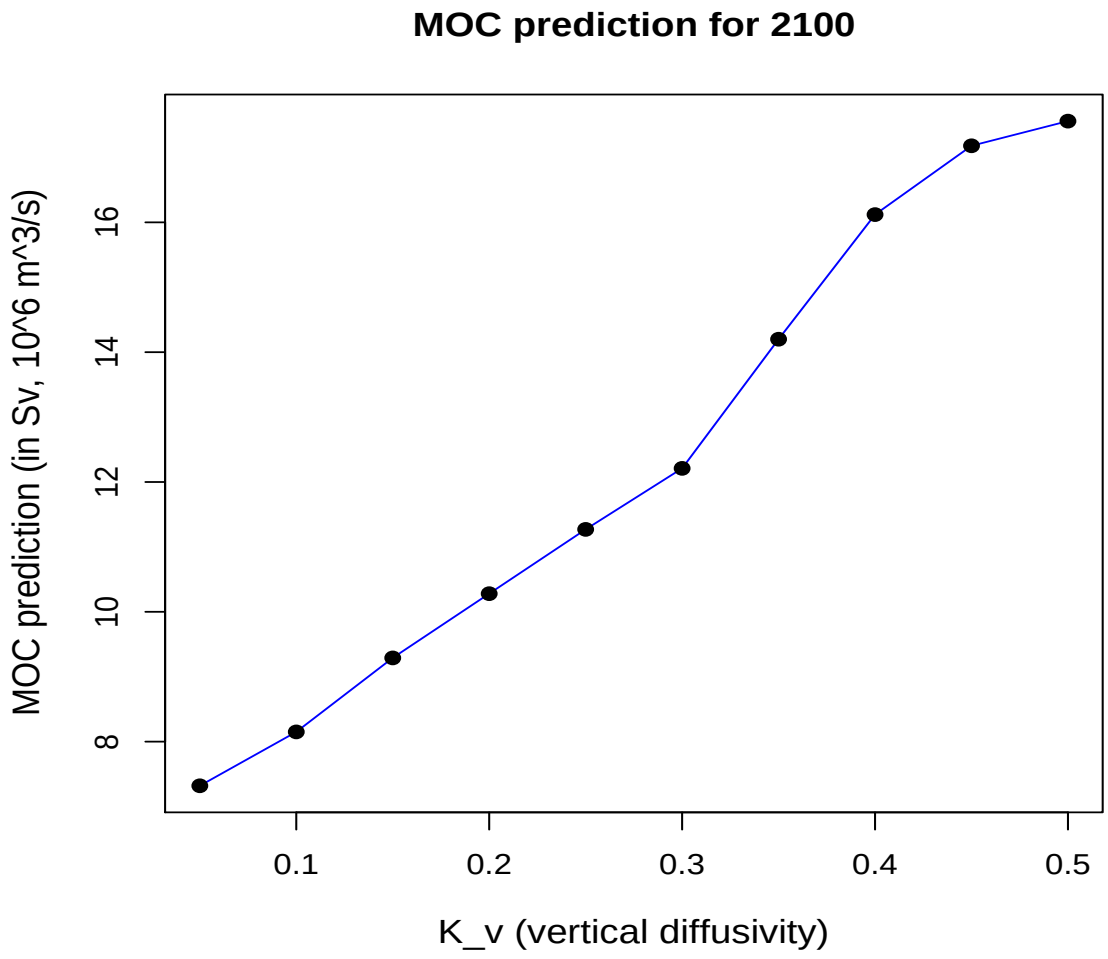
$$Y(\mathbf{s}, t, \theta) = \sum_{j=1}^{J} k(\mathbf{u}_j - \mathbf{s}; v_j - t, \ell_{1j} - \theta_1, \cdots \ell_{kj} - \theta_k) w(\mathbf{u}_j, v_j, \ell_j) + \mu(\theta)$$

- where the set of knots are $\mathbf{u}_j$, v_j, ℓ_j for $j = 1, \ldots, J$.
  $w(\mathbf{u}_j, v_j, \ell_j)$ is the process at the jth knot.
- The random field for $\mathbf{Y}(\mathbf{s}_i, t_i, \theta_i)$ is
  $\mathbf{Y}(\mathbf{s}_i, t_i, \theta_i) \mid \mathbf{w}, \psi, \kappa, \beta, \phi_s, \phi_c$

$$\sim N\left(\mathbf{X}(\theta_i)\beta + \sum_{j=1}^{J} K_{ij}(\phi_s, \phi_c) w(\mathbf{u}_j, v_j, \ell_j), \psi\right)$$

- Linear mean trend on θ and kernel is separable covariance function over $\mathbf{s}$, t, θ.

MOC predictions versus K_v

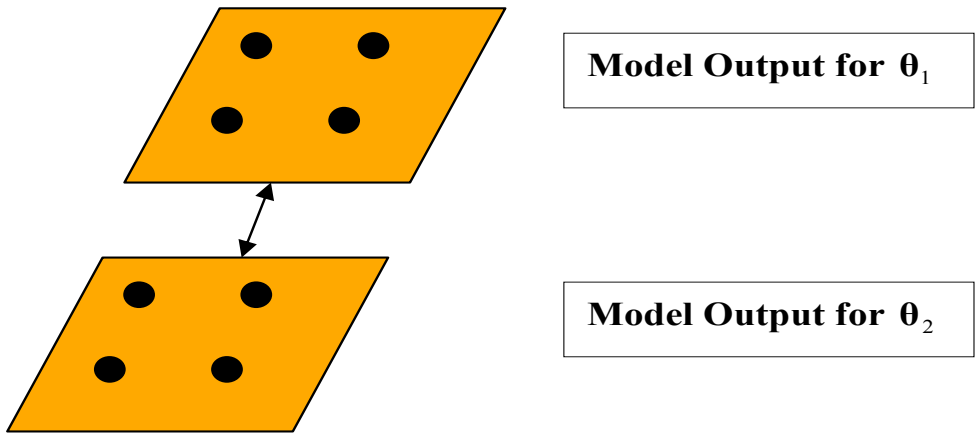


- MOC predictions are clearly much lower as K_v values get small.

Bayesian model calibration (cont'd)

- Let observation error, $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$. Modeled as Normal $(0, \psi\Sigma)$, where Σ is estimated from other model runs (different runs from the ones used here; for e.g. 'control' runs that exclude human intervention/forcings.)
- Cov $(Y(\mathbf{s}_i, \theta_{i'}), Y(\mathbf{s}_j, \theta_{j'})) = \kappa\, \Sigma_{ij}\, r(\theta_{i'}, \theta_{j'})$.
- $\phi_c = (\phi_{c1} \cdots \phi_{ck})$ are the climate covariance parameters.

$$r(\theta_{i'}, \theta_{j'}) = \prod_{m=1}^{k} \exp\left(-\frac{|\theta_{i'm} - \theta_{j'm}|}{\phi_{cm}}\right)$$



Bayesian model calibration: inference

- Hence the joint distribution of $\mathbf{Z}$ and $\mathbf{Y}$ is a multivariate normal, and

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{Y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{M}(\theta^*) \\ \mathbf{M} \end{bmatrix} \beta, \begin{bmatrix} (\psi + \kappa) \otimes \Sigma & r(\theta^*)^T \otimes \Sigma \\ r(\theta^*) \otimes \Sigma & \mathbf{R} \otimes \Sigma \end{bmatrix} \right)$$

- Inference for θ^*, ξ_s, etc is based on the posterior distribution $\pi(\theta^*, \xi_s, \phi_c, \beta | \mathbf{Z}, \mathbf{Y})$

$$\pi(\theta^*, \xi_s, \phi_c, \beta | \mathbf{Z}, \mathbf{Y}) \propto \mathcal{L}(\mathbf{Z}, \mathbf{Y} \mid \theta^*, \xi_s, \phi_c, \beta) \times p(\theta^*) p(\xi_s) p(\phi_c) p(\beta)$$

  - $\mathcal{L}(\mathbf{Z}, \mathbf{Y} \mid \theta^*, \xi_s, \phi_c, \beta)$: likelihood(multivariate normal)
  - $\xi_s = (\psi, \kappa, \phi_s)$: covariance parameters.

- Priors: θ^* based on scientific knowledge, other parameters are low precision priors (critical to do sensitivity analysis).

Computation

- $\pi(\boldsymbol{\theta}^*, \boldsymbol{\xi}_s, \phi_c, \beta | \mathbf{Z}, \mathbf{Y})$ is intractable, so rely on sample-based inference: Markov Chain Monte Carlo (MCMC).
- Computational bottleneck: matrix computations (e.g. Choleski factors) are of order N^3, where N is the number of observations.
- Kronecker products greatly reduce the computational burden. *Important*: This is brought about by assuming the same covariance Σ in modeling dependence among observations ($\mathbf{Z}$), computer model output ($\mathbf{Y}$) and in the block cross-covariance.