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

Murali Haran

Department of Statistics

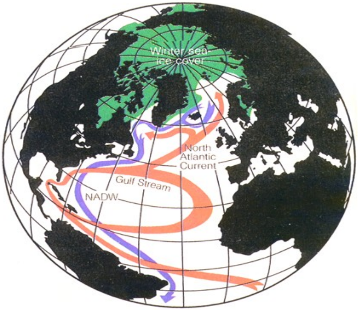
Penn State University

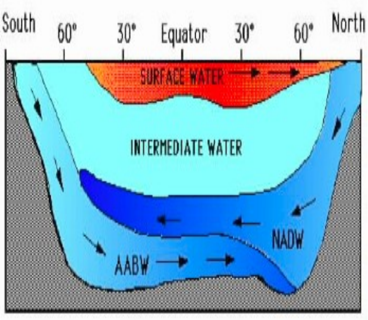
(joint work with Sham Bhat (Statistics), Roman Tonkonojenkov (Geosciences) and Klaus Keller (Geosciences))

George Washington University, April 2009

Motivation

- What is the risk of human induced climate change?
- Example of climate change: potential collapse of meridional overturning circulation (MOC).
- Early and accurate predictions of the risk of MOC collapse would save billions of dollars (Keller and McInerney, 2007)





(plots: Rahmstorf (Nature, 1997) and Behl and Hovan)

Motivation-MOC

- MOC phenomenon: Movement of water from equator to higher latitudes, deep water masses created by cooling of water in Atlantic, resulting in sea ice formation. Result is denser salt water, which sinks, causing ocean circulation.
- MOC weakening results in disruptions in the equilibrium state in the climate leading to non-trivial changes.
- Predictions of MOC strength can be made for particular climate parameter settings, e.g. vertical diffusivity.
- These climate parameter values are difficult to measure directly. Two sources of indirect information:
  - Climate models: output at different parameter settings.
  - ‘Tracers’ of climate parameters: spatio-temporal data.

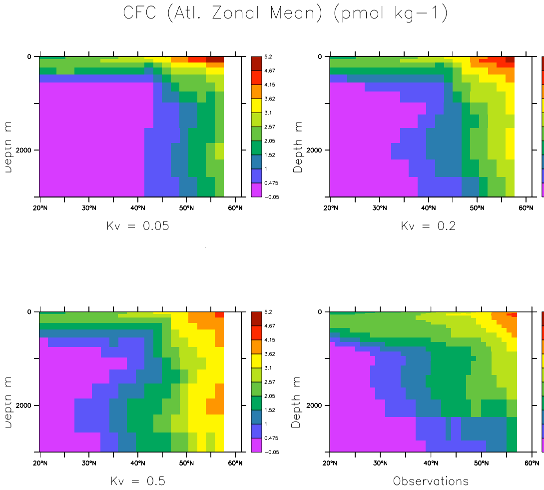
MOC via CFC-11

- Trichlorofluoromethane (CFC11), is often used as a tracer to understand deep ocean behavior such as MOC strength in the Atlantic Ocean.
- CFC11 data collected in the 1990s: spatial field across latitude and depth, averaged over longitudes.
- CFC-11 has high signal-to-noise ratio, is considered a stable tracer; unaffected by physical conditions and not produced in nature.
- Prominent articles in Science and Nature (e.g. Knutti et al 2002) use ad-hoc, non-stochastic approaches to obtain best guesses of climate parameters. Scientists want pdfs for climate parameters based on existing information: use this to project future MOC strength.

Statistical Inference

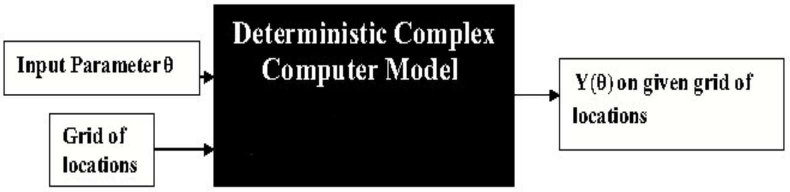
- **Goals**: Infer important climate characteristics (parameters) that drive major climate systems; study sources of variability, model-data discrepancies.
- Sources of information
  - Physical observations of tracers. e.g. CFC
  - Output from complex climate models at several different climate parameter settings. e.g. CFC at different values of vertical diffusivity.
- Rely on climate models to provide connection between tracer data and climate parameter. Climate models take weeks or months to run at each setting.
- Let $\mathbf{s}, t$ be space, time; θ be climate parameters.
  - $Z(\mathbf{s}, t)$: physical observations
  - $Y(\mathbf{s}, t, \theta)$: climate model output

CFC example



- Climate model is run at several parameter settings of vertical diffusivity (K_v). Above: 3 settings; observations.

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

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

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

- $\{w(\mathbf{s}), \mathbf{s} \in D\}$ 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: $\mathbf{Z}^* = (Z(\mathbf{s}_1^*), \ldots, Z(\mathbf{s}_m^*))^T$.

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

- $\mathbf{Z}^* \mid \mathbf{Z}$ is multivariate normal with mean and covariance:

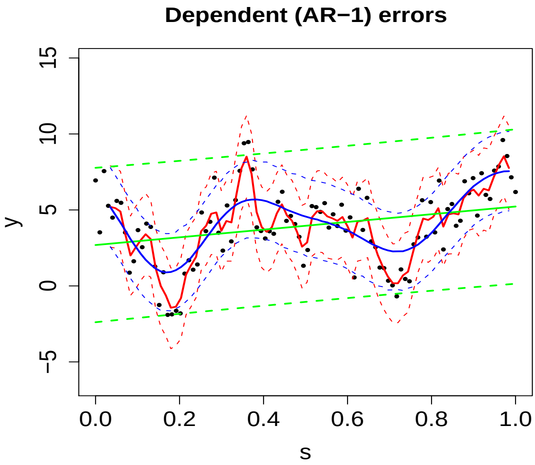
$$E(\mathbf{Z}^* \mid \mathbf{Z}, \boldsymbol{\beta}, \xi) = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{Z} - \boldsymbol{\mu}_1)$$

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

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.

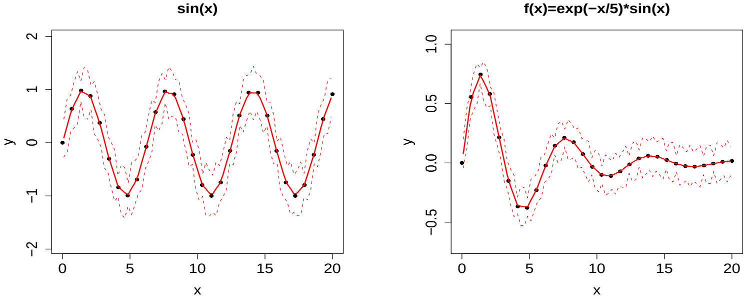
GP model for dependence: toy 1-D example



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.

A joint model

- Want to determine parameter settings that are 'most likely' given **Y**, **Z** (vector obtained by stacking columns of matrix of $Y(\mathbf{s}, t, \theta)$, $Z(\mathbf{s}, t)$ 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 $\boldsymbol{\theta}^*$ exists.

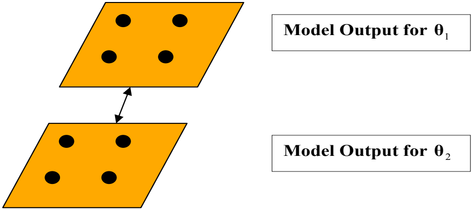
$$Z(\mathbf{s_i}, t_i) = Y(\mathbf{s_i}, t_i, \boldsymbol{\theta}^*) + \epsilon_i.$$

  Note: there is no true $\boldsymbol{\theta}^*$, 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 ($\boldsymbol{\theta}$) space.
- Separable covariance between $\mathbf{s}, t, \boldsymbol{\theta}$ dimensions.

Joint model (contd)

- Let observation error, $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^T$. 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 $(Y(\mathbf{s}_i, t_i, \boldsymbol{\theta}_{i'}), Y(\mathbf{s}_j, t_j, \boldsymbol{\theta}_{j'})) = \kappa\, \Sigma_{ij}\, r(\boldsymbol{\theta}_{i'}, \boldsymbol{\theta}_{j'})$.
- $\phi_c = (\phi_{c1} \cdots \phi_{ck})$ are the climate covariance parameters.

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



Joint model (contd)

- Hence the joint distribution of **Z** and **Y** is a multivariate normal, and

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

- Inference for $\boldsymbol{\theta}^*$, $\boldsymbol{\xi}_{st}$, etc is based on the posterior distribution $\pi(\boldsymbol{\theta}^*, \boldsymbol{\xi}_{st}, \boldsymbol{\phi}_c, \boldsymbol{\beta} | \mathbf{Z}, \mathbf{Y})$

$$\pi(\boldsymbol{\theta}^*, \boldsymbol{\xi}_{st}, \boldsymbol{\phi}_c, \boldsymbol{\beta} | \mathbf{Z}, \mathbf{Y}) \propto \mathcal{L}(\mathbf{Z}, \mathbf{Y} \mid \boldsymbol{\theta}^*, \boldsymbol{\xi}_{st}, \boldsymbol{\phi}_c, \boldsymbol{\beta}) \\ \times p(\boldsymbol{\theta}^*) p(\boldsymbol{\xi}_{st}) p(\boldsymbol{\phi}_c) p(\boldsymbol{\beta})$$

  - $\mathcal{L}(\mathbf{Z}, \mathbf{Y} \mid \boldsymbol{\theta}^*, \boldsymbol{\xi}_{st}, \boldsymbol{\phi}_c, \boldsymbol{\beta})$: likelihood(multivariate normal)
  - $\boldsymbol{\xi}_{st} = (\psi, \kappa, \phi_s, \phi_t)$: covariance parameters.
- Priors: $\boldsymbol{\theta}^*$ based on scientific knowledge, other parameters are low precision priors (critical to do sensitivity analysis).

Computation

- $\pi(\boldsymbol{\theta}^*, \boldsymbol{\xi}_{st}, \phi_c, \boldsymbol{\beta} | \mathbf{Z}, \mathbf{Y})$ 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 reduces the computational burden. *Important*: This is brought about by assuming the same covariance Σ in modeling dependence among observations (**Z**), computer model output (**Y**) and in the block cross-covariance.
- Multimodality issues: used slice sampling (even more expensive).

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 leads to identifiability issues. Also adds computational burdens. 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 $\boldsymbol{\theta}$:
  - Model the **Y**'s stochastically to 'infer a likelihood', connecting $\boldsymbol{\theta}$ to **Y**.
  - Model **Z** using fitted model from above, with additional errors, biases, to infer $\boldsymbol{\theta}$ (along with errors, biases.)
- Model **Y** as a Gaussian process emulator, with mean a linear function of $\boldsymbol{\theta}$.

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

- $\boldsymbol{\xi}$ is the set of covariance parameters, covariance function assumed to be separable among **s**, t, and $\boldsymbol{\theta}$.
- Covariance parameters:
  - Maximum likelihood estimates by optimization.
  - Bayesian approach: obtain posterior via MCMC.

Two stage approach (contd)

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

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

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

Observations

- Our approach is perhaps counter to standard Bayesian modeling philosophy: instead of a coherent joint model, we are fitting models stagewise.
- 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).
- 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.])

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: flatness of likelihood surfaces lead to fairly similar prediction intervals for frequentist and Bayesian inference.

Large spatial data sets

- Critical to use as much information as possible as this can inform discrepancies between the climate model and reality (based on the observations). Scientists are really interested in learning about these discrepancies; they can also inform decisions about model choice/averaging.
- Computational problems due to large climate model output and tracer observations: tens of thousands to millions.
- Choices for large spatial data: approximate likelihood (Vecchia, 1988; Caragea and Smith, 2002; Stein et al., 2004), kernel mixing (Higdon, 1998,2002; Paciorek and Schervish, 2006), frequency domain (Fuentes, 2007), sparse matrix approaches (Cornford et al. 2005), patterned covariances (Cressie and Johannesson, 2008).

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.

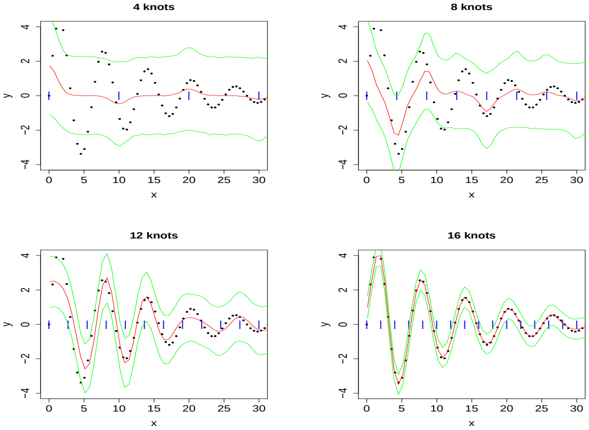
$$w(\mathbf{s}) = \int_D k(\mathbf{u} - \mathbf{s}) z(\mathbf{u}) d\mathbf{u}.$$

- Replace z 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.

Kernel mixing for climate model output

- Extend kernel and knot process **z** to t and θ dimensions:

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

- where the set of knots are $\mathbf{u}_j$, v_j, $\boldsymbol{\ell}_j$ for $j = 1, \ldots, J$. $w(\mathbf{u}_j, v_j, \boldsymbol{\ell}_j)$ is the process at the jth knot.
- The random field for $\mathbf{Y}(\mathbf{s}_i, t_i, \boldsymbol{\theta}_i)$ is

$$\mathbf{Y}(\mathbf{s}_i, t_i, \boldsymbol{\theta}_i) \mid \mathbf{w}, \psi, \kappa, \boldsymbol{\beta}, \phi_s, \phi_t, \boldsymbol{\phi}_c$$

$$\sim N\left(\mathbf{X}(\boldsymbol{\theta}_i)\boldsymbol{\beta} + \sum_{j=1}^{J} K_{ij}(\phi_s, \phi_t, \boldsymbol{\phi}_c) w(\mathbf{u}_j, v_j, \boldsymbol{\ell}_j), \psi\right)$$

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

Linear algebra identities

- Kernel mixing can be used to induce special matrix forms that permit very fast computations. In fact, may often ignore the latent variable formulation and simply use matrix identities to speed up computation.
- For example, 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 can be written as:

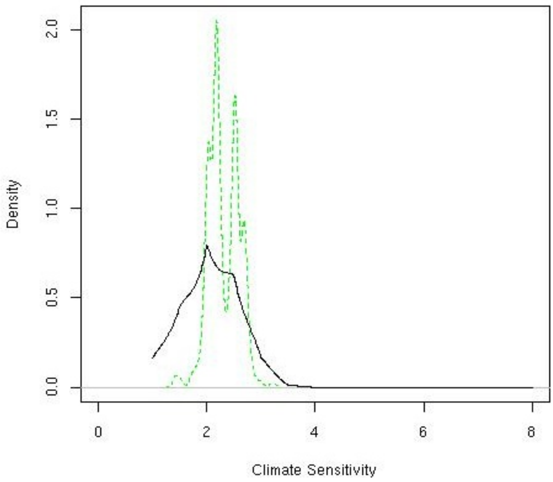
$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}.$$

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

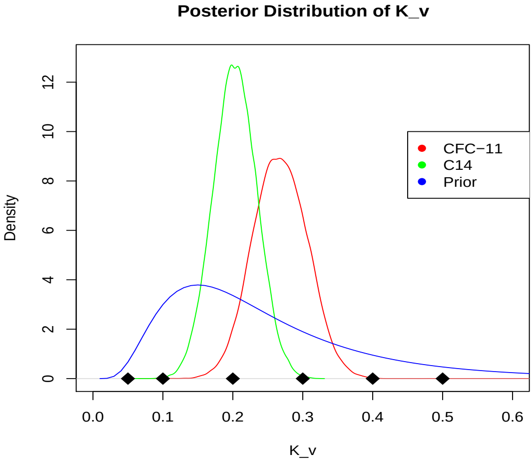
Example: Ocean heat anomalies

- Small data set from Levitus (2005): global ocean heat anomalies observed over 40 years.
- Goal: infer the distribution of climate sensitivity, $\mathcal{S}$.
- From climate model: global ocean heat anomalies generated for 40 years at each of several ($\mathcal{S}$) settings: between 1 and 8 at intervals of 0.5.
- Can use both joint and two stage approaches for this data.

Ocean heat anomalies: posterior distribution



- Dotted green lines: Joint model. Note the bimodality.
- Solid black lines: Two stage (ran in less than half the time).



- UVIC model run at six different values of K_v ($\mathrm{cm}^2\mathrm{s}^{-1}$).
- Small $K_v \Rightarrow$ low MOC, sensitivity to anthropogenic forcings.
- Different (but overlapping) pdfs based on different tracers!

Joint inference based on multiple tracers

- Want single pdf based on both tracers.
- Scientifically sound: model the non-linear relationship between CFC and C14 and perform inference for K_v based on *both* CFC and C14.
- Outline of our approach: Utilize Royle and Berliner (1999) approach to modeling CFC and C14 jointly — treat them as bivariate spatial fields with a non-linear relationship.
  - Model $\mathbf{Y}_1 \mid \mathbf{Y}_2$, that is, climate output for one tracer given the other. 'Infer' a likelihood based on both tracers.
  - Model $(\mathbf{Z}_1, \mathbf{Z}_2)$, the joint set of physical observations of both tracers, using above. Now obtain posterior for K_v based on $(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Z}_1, \mathbf{Z}_2)$.
- Analogous to univariate case but computing is even more challenging... [ongoing work].

Summary and related work

- Two stage 'inferred likelihood' approach: computational complexity reduced, allowing for added flexibility. Potentially allows for sharper inference.
- Kernel mixing, patterned covariances help make computations tractable. With thousands of data points, for instance, simple Gaussian process-based approaches are computationally infeasible.
- We now have an approach for inference based on two space-time tracers, allowing for non-linear relationships between the tracers.

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.10-20). We have preliminary results based on a simple separable cross-covariance.
- 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.

Key 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.*