Uncertainty Quantification for Large Multivariate Spatial Computer Model Output with Climate Applications

K. Sham Bhat
Statistical Sciences Group, Los Alamos National Laboratory
(collaborators include Jim Gattiker, Murali Haran, Dave Higdon, Klaus Keller, and many others)

SIAM Uncertainty Quantification Mini-Symposium, April 2012
Climate models often fail to reproduce certain physical processes (e.g. eddies, mixing).

Parameterizations approximates unresolved processes.

**Goal:** Characterize the uncertainty due to these parameters by calibrating the model to observations.
Introduction: Computer Models

- Computer model calibration - determine parameter that best fits observations.
- Incorporation of dependence and sources of uncertainty, i.e. model discrepancy and observation error.
- Parameter inference using Bayesian methods and MCMC.
Calibration—$\Delta^{14}$C

- $K_v$ quantifies the intensity of vertical mixing in the ocean, a model-driven parameter and cannot be measured directly.

- Indirect information about $K_v$: observations and output from multiple tracers (or metrics, attributes, variables).
Calibration—$\Delta^{14}C$

**Observations**

![Observations Graph]

**Kv=0.05**

![Kv=0.05 Graph]
Calibration—$\Delta^{14}C$

**Observations**

**Kv=0.2**
Overview

1. Model is complex and takes a long time to run.
   - Gaussian process emulator (surrogate model).

2. Data and output are multiple spatial fields, with complex underlying behavior.
   - Flexible emulator to combine information from multiple tracers.
   - Different dependence patterns and non-linear relationships between tracers, nonseparability.
   - Nonstationary dependence patterns.
   - Use mean and covariance function of Gaussian process.

3. Large data and output.
   - Dimension reduction to improve computational tractability.

4. Information from multiple models for prediction.
   - Bayesian model averaging.
Combining Information from Multiple Tracers

- Flexible Gaussian process emulator.
- Independence - not a realistic assumption here.
- Separability - tracers have similar dependence patterns.
  - Computationally efficient and easy to interpret.
  - Assumes same spatial range and tracers are linearly related.
  - While these assumptions may be satisfactory in some cases, for many situations they are too strong.
Combining Information from Multiple Tracers

- $Y_1, Y_2, \cdots, Y_k$ are output for multiple tracers.
- Conditional hierarchical model.
  - Hierarchically model tracers ($Y_1, Y_2 | Y_1$, etc) as Gaussian processes allowing for nonlinear mean functions.
  - $Y_2 | Y_1$ and $Y_1$ have different covariance functions, and thus have different spatial dependence patterns.
  - Inference may be affected by choice of ordering.
- Linear Model of Coregionalization.
  - Tracers to "covary" together in a region through a latent process. $Y(s) = A(s)w(s)$.
  - Flexibility in dependence patterns and relationships among tracers.
  - Can be utilized for nonstationary and dimension reduction.
Posterior distribution of parameter for single tracers and for information from tracers combined using flexible nonstationary and nonseparable emulators.

Crossvalidation, both spatial and leave-one-out.

LMC-based and conditional hierarchical approaches have lower BIC values than less flexible emulators.
Dimension Reduction

Kernel mixing: continuous process created by convolving a discrete white noise process (J knots on a lattice) with a kernel function $k$, can induce nonstationarity.

- Low rank matrix approaches, integrate out kernel process using special matrix structure and matrix identities.
- Principal components or basis methods.
"Channel model" describes mixing, analog to Antarctic circumpolar current.

Goal is to approximate fine resolution eddy-resolving runs with non-resolving models with GM parametrizations, part of work in multi-parameter POP calibration.

Fully Bayesian calibration approach using GP emulator, and principal components for dimension reduction.
POP Heat Transport (GM) Parameterization

- Three metrics: potential temperature (above), vertical heat transport, and forward heat transport.
- 11 runs of 0.8° model of PT, GM range \((0.15,1.0) \times 10^7\).
- Related testbeds will be key to the future of ocean model tuning and validation.
Goal: Projections of surface temp (and precip) in 2100 using large spatial output from 20 GCMs and data
Bayesian Model Averaging

- BMA derives probability weights for models based on hindcast skill, while incorporating model uncertainty.
- Approach includes model bias, space-time dependence, and computation for large data.
BMA- Space-Time Projections
Summary

- Quantify uncertainty due to parameterizations by calibrating model to observations.
  - Gaussian process emulator as surrogate model.
  - Combine information from multiple tracers in a flexible manner.
    - Different nonstationary dependence patterns.
    - Nonseparable relationships between the tracers.
  - Dimension reduction to improve computational tractability.
  - Incorporation of model discrepancy and observation error.
  - Bayesian inference using MCMC.

- Combine information from multiple models using BMA.
Acknowledgments and References

- Financial support from CSSEF, USGS and the NSF-HSD.
Thank You!!!
Backup Slides
What is the MOC?

- The Atlantic Meridional Overturning Circulation (MOC) is the ocean circulation due to the movement of water masses from the tropics to the North Atlantic.
- These water masses cool during movement, releasing heat and forming sea ice and denser salt water, which sinks and returns southward as cold deep water masses.

(plot: IPCC (2001))
Is a Model Discrepancy Term Necessary?

Not including discrepancy may result in a bias in inferring parameters (Bayarri et. al, 2007 and Bhat, Haran, and Goes, 2010).
Tracers

- Key source of uncertainty in AMOC projections is uncertainty about background ocean vertical diffusivity ($K_v$).
- $K_v$ is a model parameter which quantifies the intensity of vertical mixing in the ocean, cannot be measured directly.
- Observations of two tracers: Carbon-14 ($^{14}$C) and Trichlorofluoromethane (CFC11) collected in the 1990s (latitude, longitude, depth), zonally averaged.
- Second source of information: climate model output at six different values of $K_v$.
- Latitude between 80 S and 60 N, depths from 0 to 3000m.
- $^{14}$C and CFC11 are stable tracers, enter ocean by air-sea gas exchange, transported by advection/diffusion.
- Data size: 3706(observations); 5926(model) per tracers.
Two-stage approach in Bhat et. al. (2011) to obtain probability distribution of $\theta$:

1. Model relationship between $Z$ and $\theta$ via model output $Y$.
2. Use observations $Z$ to infer $\theta$.

We model $Y$ (model output) as a Gaussian process (emulation): $Y \mid \beta, \xi \sim N(\mu_\beta(\theta), \Sigma(\xi))$.

$\beta$: regression parameters, $\xi$: covariance parameters.

Estimate $\beta$ and $\xi$ using ML or Bayesian methods.

$\Sigma(\xi)$, the covariance matrix of $Y$, is separable between space and calibration parameter.
Computer Model Calibration: Our Approach

\[ \text{Cov}(Y(s_i, \theta_{i'}), Y(s_j, \theta_{j'})) = \kappa \sum_{ij} r(\theta_{i'}, \theta_{j'}) \]

▶ \( \Sigma_{ij} \) describes the spatial dependence between \( Y(s_i) \) and \( Y(s_j) \).

▶ \( r(\theta_{i'}, \theta_{j'}) \) describes the dependence due to calibration parameter with range parameters \( \phi_c = (\phi_{c1} \cdots \phi_{ck}) \), e.g.

\[
r(\theta_{i'}, \theta_{j'}) = \prod_{m=1}^{k} \exp \left( -\frac{((\theta_{i'})_m - (\theta_{j'})_m)^2}{\phi_{cm}^2} \right)
\]
Computer Model Calibration

1. Inferring a likelihood between $Z$ and $\theta$:
   - $\eta(Y, \theta)$ (a random variable) is the prediction at a new $\theta$ and locations $S$; obtained by using the standard kriging framework and has a multivariate normal distribution.
   - Model for the observations: $Z = \eta(Y, \theta) + \delta(S) + \epsilon$.
   - $\eta(Y, \theta)$ is the emulator prediction, $\epsilon$ is observation error and $\delta(S)$ is model discrepancy.

2. Bayesian inference based on inferred likelihood:
   - Inference on $\theta$ performed using MCMC to estimate $\pi(\theta | Z, Y)$, integrating over remaining parameters.

3. Model Discrepancy: discrepancy between reality and the computer model at “best input” $\theta^*$, $Z^R = Y(\theta^*)$.
   - $Z^R = Y(\theta^*) + \delta + \epsilon$.
   - A Gaussian process is used to model $\delta$.
   - There is known confounding between $\delta$ and $\theta$ (Liu et. al., 2009).
Conditional Hierarchical Model

- Model the emulator for $(Y_1, Y_2)$ as a hierarchical model (following Royle and Berliner (1999)):
  - Model $Y_1 | Y_2$ as a Gaussian process allowing for a flexible non-linear relationship in mean function.
  - Model $Y_2$ as a univariate Gaussian process.

\[
Y_1 \mid Y_2, \beta_1, \xi_1, \gamma \sim N(\mu_{\beta_1}(\theta) + B(\gamma)Y_2, \Sigma_{1.2}(\xi_1))
\]

\[
Y_2 \mid \beta_2, \xi_2 \sim N(\mu_{\beta_2}(\theta), \Sigma_2(\xi_2))
\]

- $Y_1 | Y_2$ and $Y_2$ have different covariance functions, and thus have different spatial ranges.
- $B(\gamma)$ describes nonlinear relationship between $Y_1$ and $Y_2$.
- $\beta$’s, $\xi$’s are regression and covariance parameters.
- Likelihoods for $Y_1 | Y_2$ and $Y_2$ have no common parameters to be estimated; maximized separately to obtain MLE.
Computational Issues

- Matrix computations are $\mathcal{O}(N^3)$, where $N$ is the number of observations (could be tens of thousands).
- Used kernel mixing (Higdon, 1998): continuous process created by convolving a discrete white noise process ($J$ knots on a lattice) with a kernel function.
- We can write covariance matrix as: $\mathbf{A} + \mathbf{KCK}^T$, $\mathbf{K}$ kernel matrix with rank $J$.
- Special matrix structure and Sherman-Woodbury-Morrison identity can be used to reduce matrix computations (Cressie and Johannesson, 2008).
- The parameter space is greatly reduced, it is not necessary to estimate the kernel process.
- Matrix inversions are $\mathcal{O}(J^3)$ and mult are $\mathcal{O}(NJ^2)$. 
Inference and AMOC Strength Projections

Left: Inference of $K_v$ when combining both tracers. Right: Distribution of projected AMOC strength (Sv) in 2100 (left) given posterior distributions of $K_v$ for both single tracers and multiple tracers.
Combining Multiple Models Using BMA

How can we combine spatial field output from multiple climate models?

Bayesian Model Averaging (BMA) is a formal statistical approach for combining multiple models while accounting for model uncertainty.

\[ f(Z | Y^h_1 \cdots Y^h_k) = \sum_{k=1}^{K} w_k \ast g_k(Z | Y^h_k). \]

\[ g_k(Z | Y^h_k) \sim a_k 1 + b_k Y^h_k + \delta_k + \epsilon_k. \]

Weight \( w_k \) assigned to model \( M_k \) based on hindcast skill.

Contributions:

1. Incorporation of spatiotemporal dependence.
2. Kernel-mixing approaches for large data sets.
3. Bayesian inference on weights using a Dirichlet prior.
Combining Information from an Multi-Model Ensemble

ECHO - 2100

ECHAM - 2100

GISS - 2100
Crossvalidation
Selected References


- University of Victoria (UVic) Earth System Climate Model (Weaver et. al. 2001).
Linear Model of Coregionalization

- Linear Model of Coregionalization: model for multiple ($q \geq 3$) spatial fields using a linear combination of independent latent processes.
  - Allows for a non-separable covariance function.
  - Reduces data-specific modeling decisions.

- Model output $Y(s)$ expressed as follows:
  \[ Y(s) = (Y_1(s) \cdots Y_q(s))^T = \mu(s) + A(s)w(s) + e. \]

- $w(s) = (w_1(s) \cdots w_r(s))^T, \ r \leq q$ (allows for dim reduction).

- Latent processes $w_i$’s are independent zero mean GP with different covariance functions, $\rho_i$.

- $A(s)$ is a $q \times r$ matrix describing local non-spatial effects.

- Flexibility to allow for nonstationary models if $A(s)$ varies by location.
Calibration using LMC

- We develop the emulator for $Y$ as follows:

$$Y = \mu_\beta(\theta) + A_y w + e.$$ 

$$Y | \beta, \theta, \xi_y, A \sim N \left( P_y^T \operatorname{vec}(X_y \beta), \Sigma_y(\xi_y) \right).$$ 

$$\Sigma_y(\xi_y) = A_y P_y^T \mathcal{H}_y(\phi) P_y A_y^T + I_M \otimes D(\zeta).$$

- $A_y = \bigoplus_{i=1}^M A_m(s_i) = I_M \otimes A$, $\mathcal{H}(\phi) = \bigoplus_{j=1}^q H_j(\phi)$. 

- $H_j(\phi)$ represents spatial dependence for $w_j$, $A$ represents the local non-spatial cross covariance, $D(\zeta)$ represents microscale error, $P_y$ is a permutation matrix.

- For large datasets, we replace $w$ with a kernel process to improve computational tractability.
Calibration using LMC-cont’d

- We infer a likelihood between $Z$ and $\theta$.

- Model the observations: $Z = \eta(Y, \theta) + \delta(S) + \epsilon,$
  $\epsilon \sim N(0, \Sigma_{\epsilon}), \Sigma_{\epsilon} = I_N \otimes D_d(\psi)$, with $\psi = (\psi_1, \cdots, \psi_q)$.

- $\delta(S) = (I_N \otimes A_d)w_d \sim N(0, \Sigma_d(\xi_d))$,

  $\Sigma_d(\xi_d) = (I_N \otimes A_d)P_z^T \mathcal{H}_d(\phi_d)P_z(I_N \otimes A_d^T)$.

- More flexible $\delta(S)$; we do not assume independence among model discrepancies from different tracers.

- Use MCMC to sample from $\pi(\theta \mid Z, Y)$.

- Priors: lognormal for $\theta$, inverse Wishart for $T_d = A_dA_d^T$, inverse gamma for variances.