Quantifying uncertainties of climate system properties

Dorin Drignei
Geophysical Statistics Project
National Center for Atmospheric Research

Collaborators:
Chris Forest (MIT), Doug Nychka (NCAR), Bruno Sanso (UCSC), Claudia Tebaldi (NCAR)

Statistics areas involved

- Bayesian modeling
- Design and analysis of computer experiments

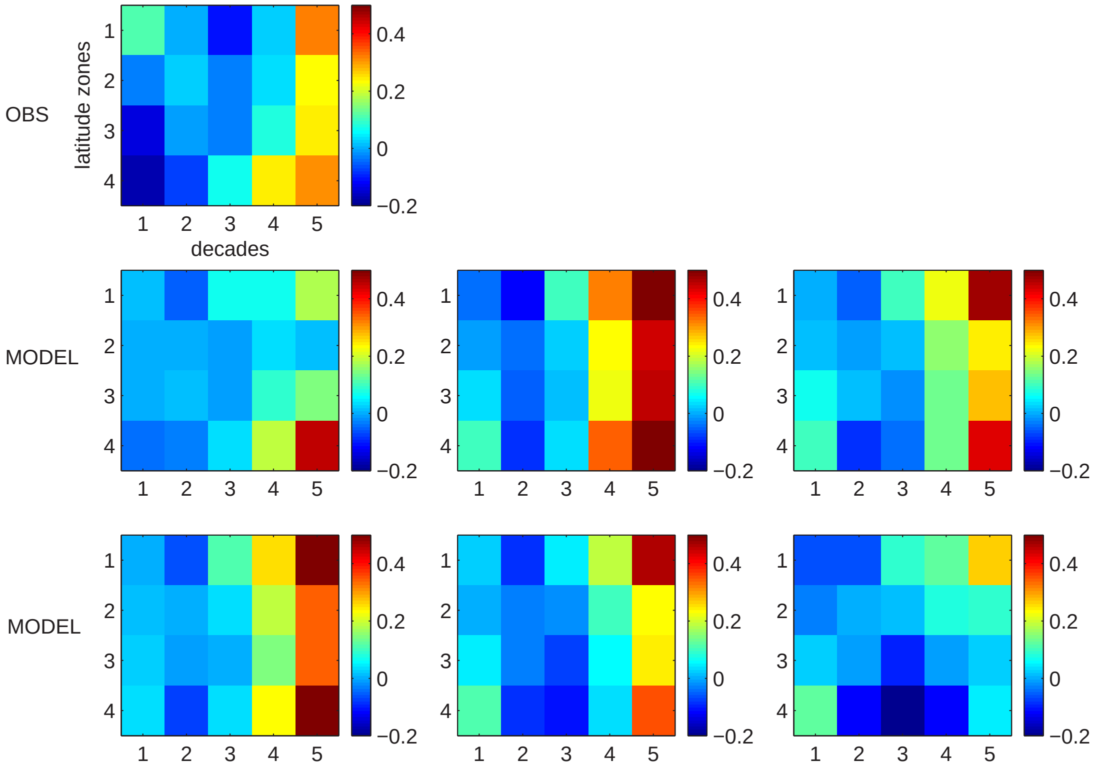
Climate models and observed climate data

- Climate Variables: $\approx$ 30 years averages of meteorological variables.
- Modeled Climate: mathematical model (partial diff. eqns) of weather whose output variables are averaged.
- Observed Climate: averages of observed variables.
- Modeled climate data - on a grid
  Observed climate data - irregular (much sparser than model data)

A particular case

- *Upper-air temperature data:* observed and climate model
- *Model*
  - MIT 2D Land-Ocean Climate Model
- *Parameters of climate model*: $\theta = [S, K_v, F_{aer}]$
  - S: Equilibrium climate sensitivity to a doubling of CO_2
  - K_v: Global mean vertical thermal diffusivity for the mixing of thermal anomalies into the deep ocean
  - F_{aer}: Net aerosol forcing

Goal: Find θ that best fits the observed data.



Observed Temperature (1st row). Modeled Climate Temperature for 6 different θ's (2nd and 3rd rows)

How to Characterize the Uncertainty Associated with the Best Fit?

Naive Approach

- Search the parameter space for θ that minimizes a discrepancy between the observed and model data.
- The climate models are computationally intensive!
- This direct approach is not feasible...

Alternative approach for estimating θ

- Sample a number of parameters θ, run the climate model and obtain the model data
- First Model Fitting:
  - Build a statistical model based on these model data
  - Use the statistical prediction of the model data as surrogate for the model data at new, untried θ parameters
  - The statistical predictor needs to be less computationally intensive than the actual computer model in order to be useful
- Second Model Fitting:
  - Search the parameter space for θ that minimizes a discrepancy between the observed and statistically predicted model data

Some details

- Nonlinear regression in θ: $T_{obs} = \hat{T}_\theta + noise$
  - T_{obs} = observed climate temperature data
  - $\theta = [S, K_v, F_{aer}]$
  - $\hat{T}_\theta$ = point predictor of model climate temperature
  - Σ = noise covariance matrix (currently estimated from control runs, i.e. unforced climate model runs). Note, however, that $\Sigma = \Sigma_1^\theta + \Sigma_2$

    Σ_1^θ: prediction error of model temperature at θ

    Σ_2: observational error (biases between obs and pred model temp).

- Prior distribution:
  $p(\theta) = p(S)p(K_v)p(F_{aer})$ - uniform distributions.

- Posterior distribution: $p(\theta|T_{obs}, \Sigma)$.

Future work

- Current approach for estimating Σ: empirical estimator based on unforced model data (i.e. control runs)
  - This insures nonstationarity, but it does not involve the observed data.

- Large dimensional noise covariance Σ (numerical instabilities)
  - Current approach: truncation of small eigenvalues.

- Parametric models for Σ could solve both problems