Dorin Drignei (*Short bio*)

2004: PhD, Statistics Department, Iowa State University
2000: MSc, Statistics Department, Iowa State University
1995: BS, Mathematics Department, Univ. of Craiova, Romania

NCAR projects and collaborators:

Estimation of Climate Model Parameters

Doug Nychka (NCAR), Chris Forest (MIT)

Statistical Tests for Climate Similarity

Doug Nychka (NCAR), William Collins (NCAR), Phil Rasch (NCAR)

Climate Detection and Attribution

Doug Nychka (NCAR), Tom Wigley (NCAR)

Estimation of Climate Model Parameters

Dorin Drignei
Geophysical Statistics Project*
National Center for Atmospheric Research

Collaborators: Doug Nychka (NCAR), Chris Forest (MIT)

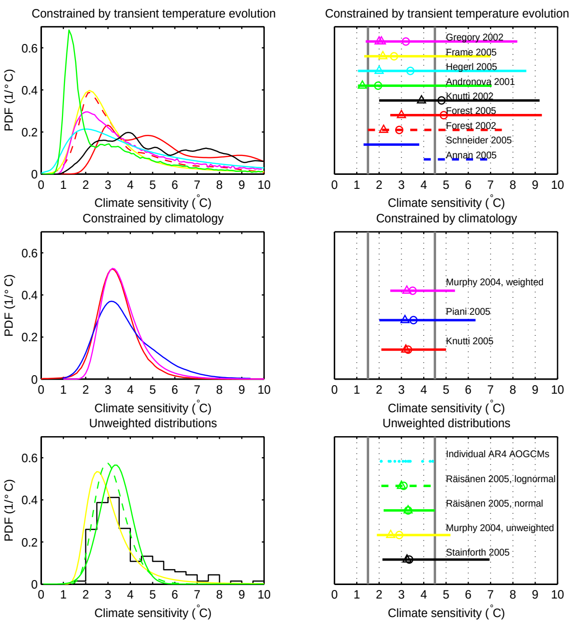
*Supported by the U.S. National Science Foundation DMS

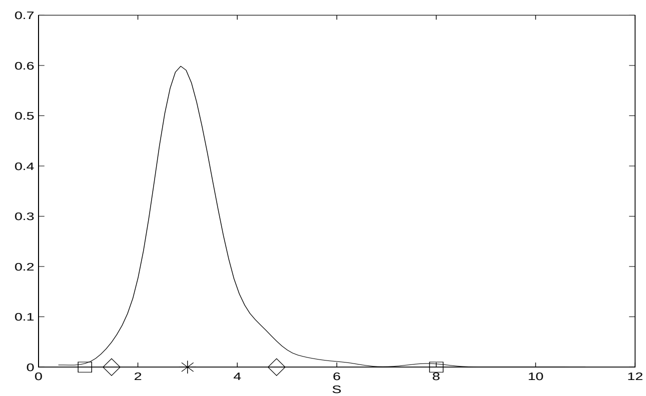
Scientific Motivation: *Estimation of climate sensitivity S*

S = global-mean surface temperature change when doubling CO_2.

Various research methods

to estimate the pdf of $\hat{S}$ *Our method belongs to the class of methods that provide shorter confidence intervals*





- **Ongoing research on ‘constraining’ the right tail of the pdf of $\widehat{S}$: how large will be the increase of global-mean surface temperature change when doubling CO_2?**

The statistical problem and our approach

- *STATISTICAL PROBLEM*: **Estimation of parameters θ in the nonlinear**
  regression model
  $$\mathrm{Y} = f(\mathrm{X}, \theta) + \epsilon$$
  when computing f requires a great computational effort.
  (Y observations, X covariates, ϵ random errors)

- *OUR APPROACH*: **Construct a computationally faster approximation for the computationally intensive nonlinear function f, and account for the approximation error.**

  This approximation is based on statistical Design and Analysis of Computer Experiments (DACE) methodology.

What is methodologically new?

- The use of DACE in the context of model calibration
- DACE multidimensional (mostly DACE scalar in literature)
- Our approach is statistically more rigorous:
  - it accounts for various sources of uncertainty;
  - it includes space-time correlation;
- New space-time covariance for output data

A naive nonlinear regression model

- The simplest model tried: nonlinear regression

$$\text{Observed Climate} = \text{Modeled Climate}\ ([S, K_v, F_{aer}]) + \epsilon.$$

- Observed Climate: averages of observed climate variables (e.g. temperature, precipitation) over long time periods.
- Modeled Climate: output variables (e.g. temperature, precipitation) from a numerical model, which are averaged over long time periods.
- Climate model parameters $\theta = [S, K_v, F_{aer}]$
  - S: Equilibrium climate sensitivity: global-mean surface temperature change if doubling CO_2 (^oC)
  - K_v: Global-mean vertical thermal diffusivity for the mixing of thermal anomalies into the deep ocean (cm^2/sec)
  - F_{aer}: Net aerosol forcing (W/m^2)
- *Computational challenge !!!* 'Modeled Climate' requires 4 hours computational time for each θ => Iterative likelihood maximization not feasible!

The proposed statistical model

- Our statistical model will be a *modification* of the previous nonlinear regression

$$\text{Observed Climate} = \text{Modeled Climate } (\theta) + \epsilon.$$

$$\mathbf{Y} \quad = \quad f_\theta + \epsilon$$

f may also depend on covariates X (e.g. precipitation).

Modification of the nonlinear regression:

$$\mathbf{Y} \quad = \quad \tilde{f}_\theta + (f_\theta - \tilde{f}_\theta) + \epsilon$$

$$\mathbf{Y} \quad = \quad \tilde{f}_\theta + E + \epsilon$$

- $\tilde{f}_\theta$ computationally faster *surrogate* (i.e. approximation) for f_θ.
- E and ϵ normal errors.

Method $(\mathbf{Y} = \tilde{f}_\theta + E + \epsilon)$

DACE – Design and Analysis of Computer Experiments ($\tilde{f}_\theta$ and Error E)

- **Sample a number of parameters θ, run the climate model and obtain the output data.**

- **Construct a statistical model for the output data.**

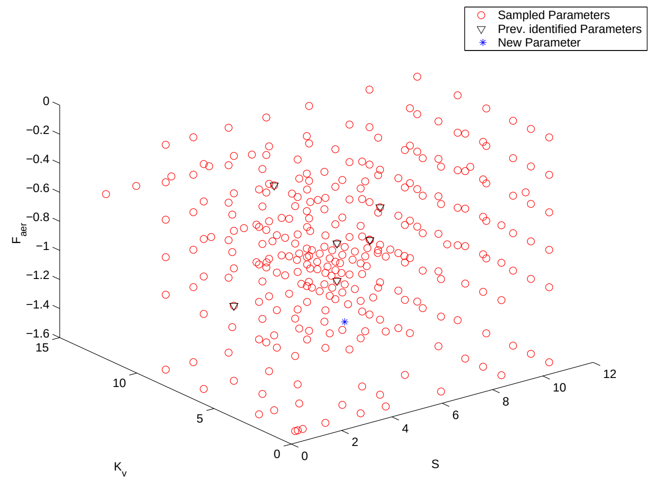
- **Build a statistical surrogate to predict the climate model output data at new, not-sampled θ parameters.**

FIT THE OBSERVED DATA (Error ϵ)

- **Use the above DACE model to find θ that best fits the observed data, and characterize its uncertainty.**

Sampled parameters θ and the final data sets

$D = 306$ **sampled parameters** θ



Temperature output data sets:

- **Surface (5 decades $\times$ 4 latitude bands $\times$ 306 parameters)**
- **Deep-ocean (linear trend $\times$ 306 parameters)**
- **Upper-air (26 latitudes $\times$ 8 pressure levels $\times$ 306 parameters)**

Temperature observed data sets:

- **Surface (5 decades $\times$ 4 latitude bands)**
- **Deep-ocean (linear trend: scalar)**
- **Upper-air (26 latitudes $\times$ 8 pressure levels)**

Data sets are vectorized

Statistical model for output data ($\mathbf{Y} = \tilde{f}_\theta + E + \epsilon$)

The output data set at sampled parameters (surface temperatures)

$$f_S = f_{S,s} + f_{S,n}$$

$f_{S,s}$ **climate signal,** $f_{S,n}$ **climate model internal variability.**

$$f_S \sim \mathbf{N}(\mu\mathbf{1}, \sigma_S^2(C_\ominus \otimes C_z \otimes C_t) + \nu_S^2\mathbf{I} \otimes \Gamma)$$

- $C_\ominus, C_z, C_t$ **matrices of power exponential correlations.**
- Γ **estimated from ensemble members**

- **If** $\Sigma_\ominus = \sigma_S^2(C_\ominus \otimes C_z \otimes C_t) + \nu_S^2\mathbf{I} \otimes \Gamma$**,**
  the likelihood for output data

$$L(f_S) = (\frac{1}{\sqrt{2\pi}})^{N_Y} \frac{1}{\sqrt{det\Sigma_\ominus}} exp(-\frac{1}{2}(f_S - \mu\mathbf{1})'\Sigma_\ominus^{-1}(f_S - \mu\mathbf{1}))$$

  is maximized and the statistical parameters will be fixed at their point estimate values.

Statistical surrogate for the climate model ($\mathbf{Y} = \tilde{f}_\theta + E + \epsilon$)

For θ arbitrary (sampled or not) in the parameter space

$$E(f_{\theta,s}|f_S)$$

$f_{\theta,s}$ climate signal, f_S climate model output data.

$$\tilde{f}_\theta = \mu 1 + \tilde{\Sigma}_{\theta\ominus}\Sigma_{\ominus}^{-1}(X_S - \mu 1)$$

$$E \sim \mathrm{N}(\mathbf{0}, V_\theta), \qquad V_\theta = \sigma_S^2(C_z \otimes C_t) - \tilde{\Sigma}_{\theta\ominus}\Sigma_{\ominus}^{-1}\tilde{\Sigma}'_{\theta\ominus},$$

where

$$\tilde{\Sigma}_{\theta\ominus} = \sigma_S^2(C_{\theta\ominus} \otimes C_z \otimes C_t),$$

and $C_{\theta\ominus}$ gives the correlation between the new parameter θ and the set of sampled parameters $\ominus$.

Nonlinear statistical model for observations ($\mathbf{Y} = \tilde{f}_\theta + E + \epsilon$)

$$L(Y|\theta) := L(Y|\theta, \text{other stat parameters}) =$$

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{N_Y} \frac{1}{\sqrt{det(V_\theta + \tau^2 R_z \otimes R_t)}} exp\left(-\frac{1}{2}(Y - \tilde{f}_\theta)'(V_\theta + \tau^2 R_z \otimes R_t)^{-1}(Y - \tilde{f}_\theta)\right)$$

R_z, R_t matrices of exponential correlations.

Y_S observed surface temperature change
Y_K observed deep ocean temperature trend
Y_F observed upper air temperature change

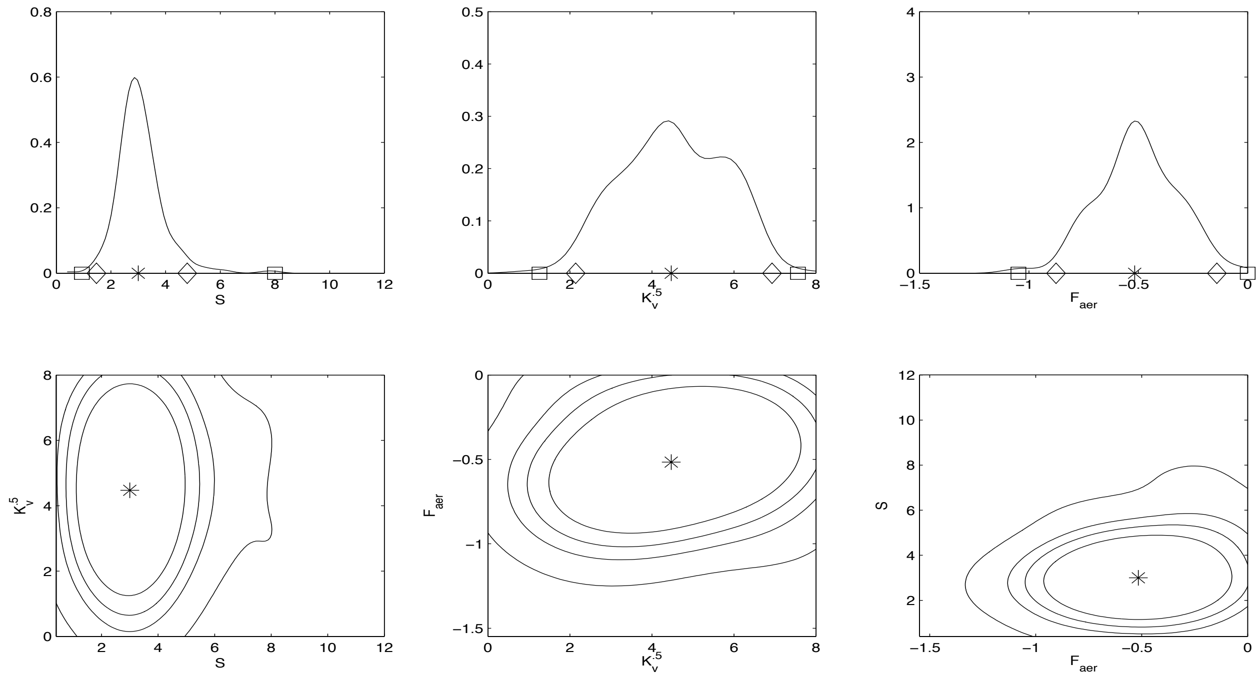
Overall likelihood to be optimized (conditional independence)

$$L(Y_S, Y_K, Y_F|\theta) = L(Y_S|\theta)L(Y_K|\theta)L(Y_F|\theta).$$

A single likelihood evaluation takes about 10 sec.

Results

- Parametric bootstrap MLE sample of size 300.
- Nonparametric kernel density estimation of MLE pdf.



Future work

- **Optimum design: how can we choose the model runs (sampled parameters) to minimize the volume of the confidence region?**

- **Bayesian model based on our likelihood development for a direct comparison with previous Bayesian methods for estimating pdf of S.**

- **Analyze other climate data sets (e.g. precipitation);**

- **Theoretical study: are the bootstrap MLEs of the unknown parameter "attracted" by the sampled (design) parameters?**

Paper to be submitted to
Journal of the Royal Statistical Society: Series C (Applied Statistics)