On Some Challenges of Statistical Analysis in Climate Change Study

Jingfeng Wang
School of Civil and Environmental Engineering
Georgia Institute of Technology

MathGeo CMG Conference, Princeton, NJ
1 October 2012

Education Background

- **BS** (1984) Mechanics and Applied Mathematics
  (dynamics, fluid and solid mechanics)

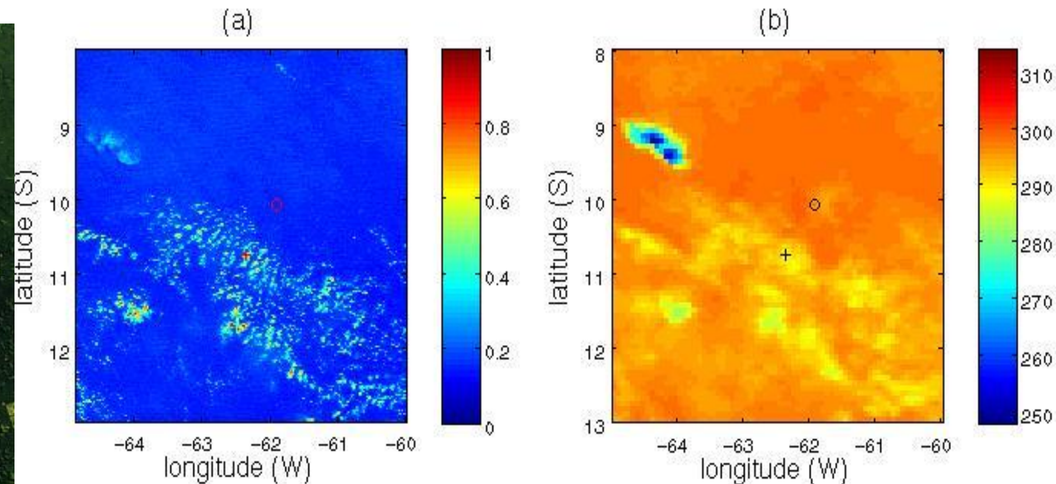
- **MS** (1987) Computation Fluid Mechanics

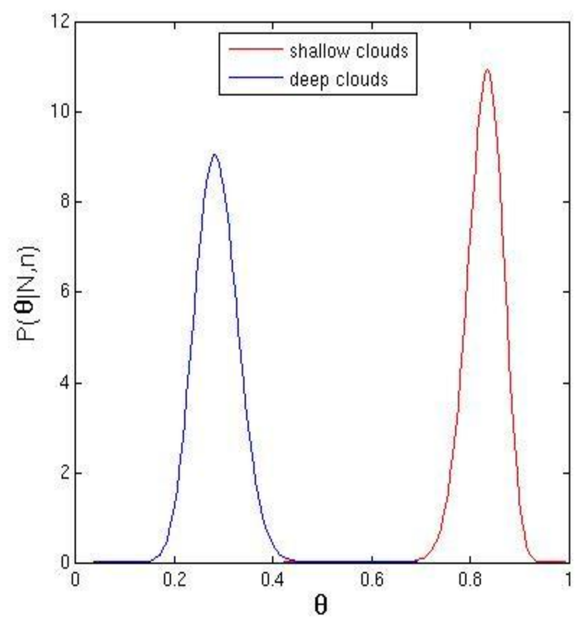
- **Sc.D.** (1997) Hydro-meteorology
  (statistical and dynamic modeling of land-atmosphere systems)

Overview of Research

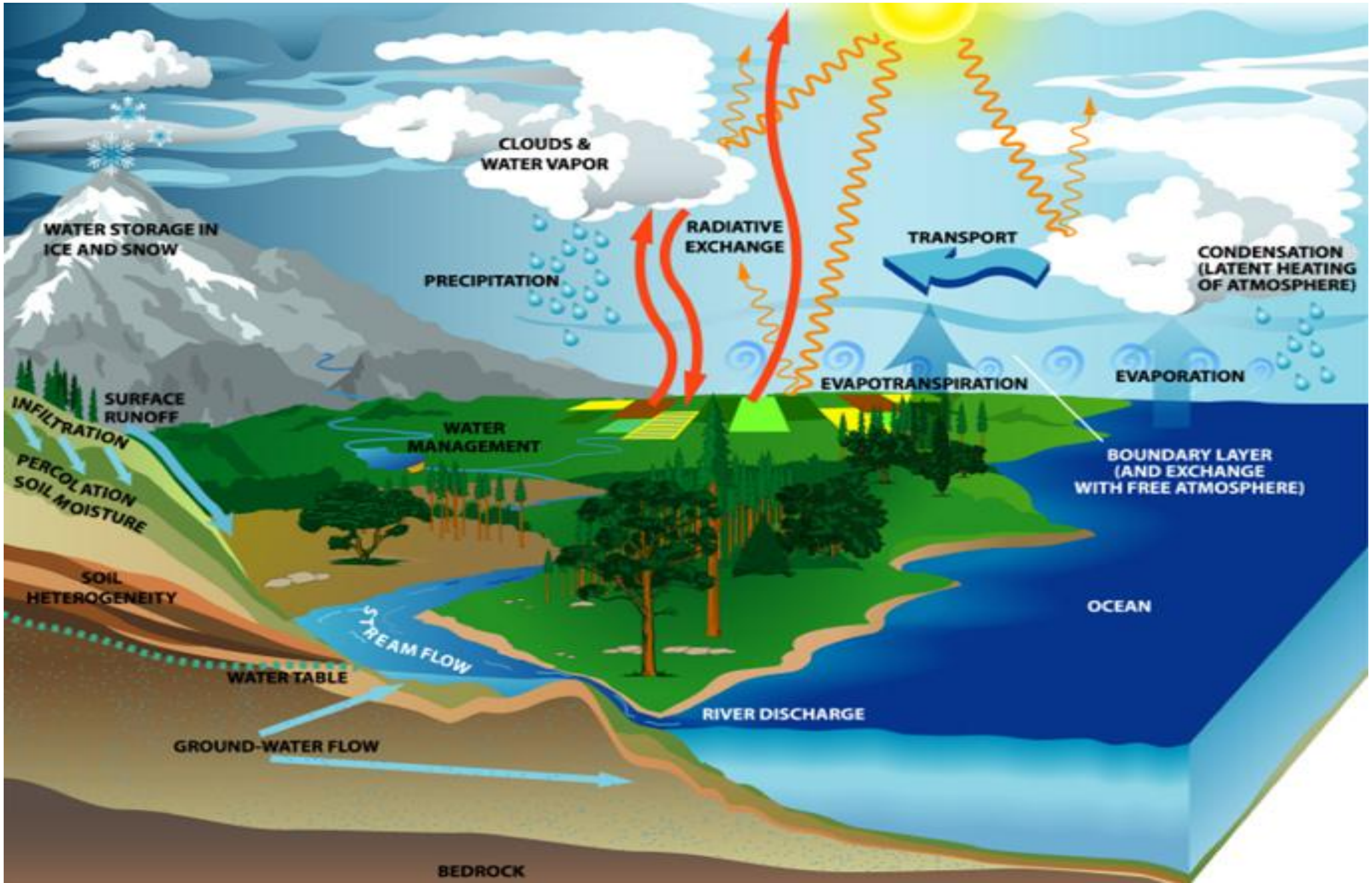
- Numerical simulation of Navier-Stokes equation for industrial applications,
- Chaotic dynamics of large-scale soil moisture,
- Deforestation and regional climate in the Amazon,
- Exchange of water and heat over the Earth surface (fundamental physics, algorithm development, field experiment),
- Bayesian Statistics.

❑ Statistical analysis of Amazon deforestation and cloud climatology [*Wang et al*, 2009, PNAS]



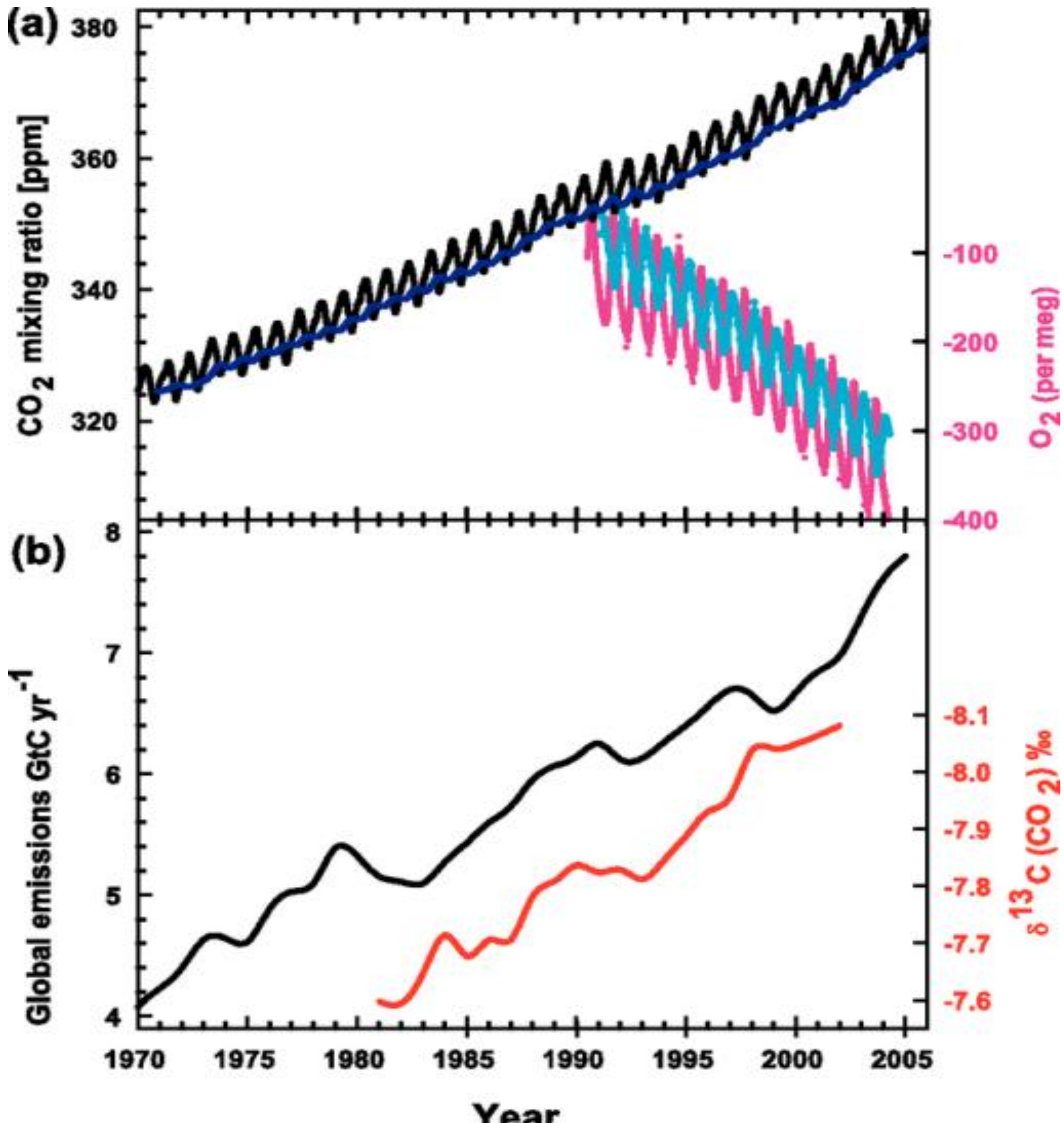


- Application of maximum entropy theory in modeling surface fluxes --- non-bulk transfer equation based model [*Wang et al*, 2009, 2011, WRR]

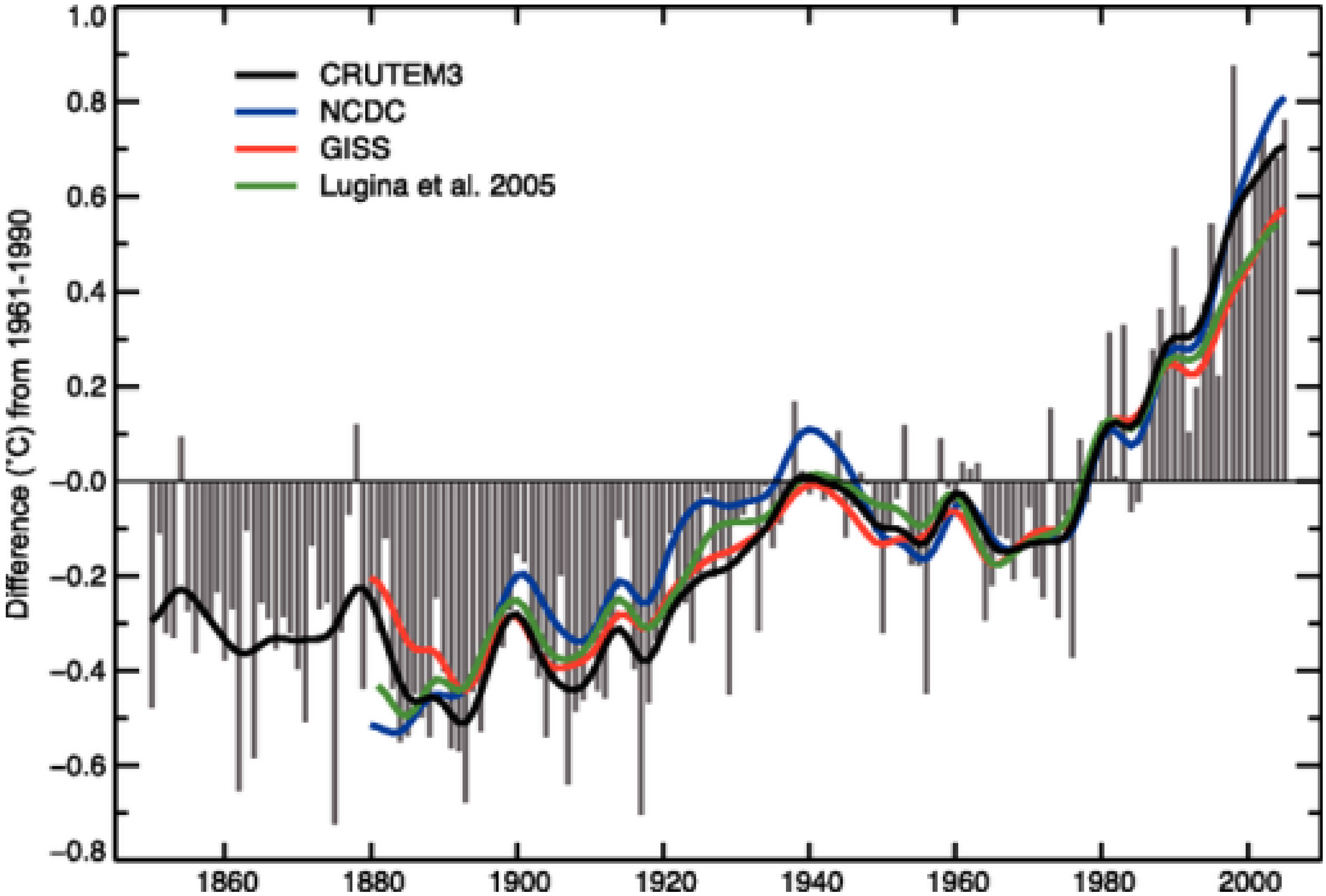


- detecting trends and cycles in climate data,

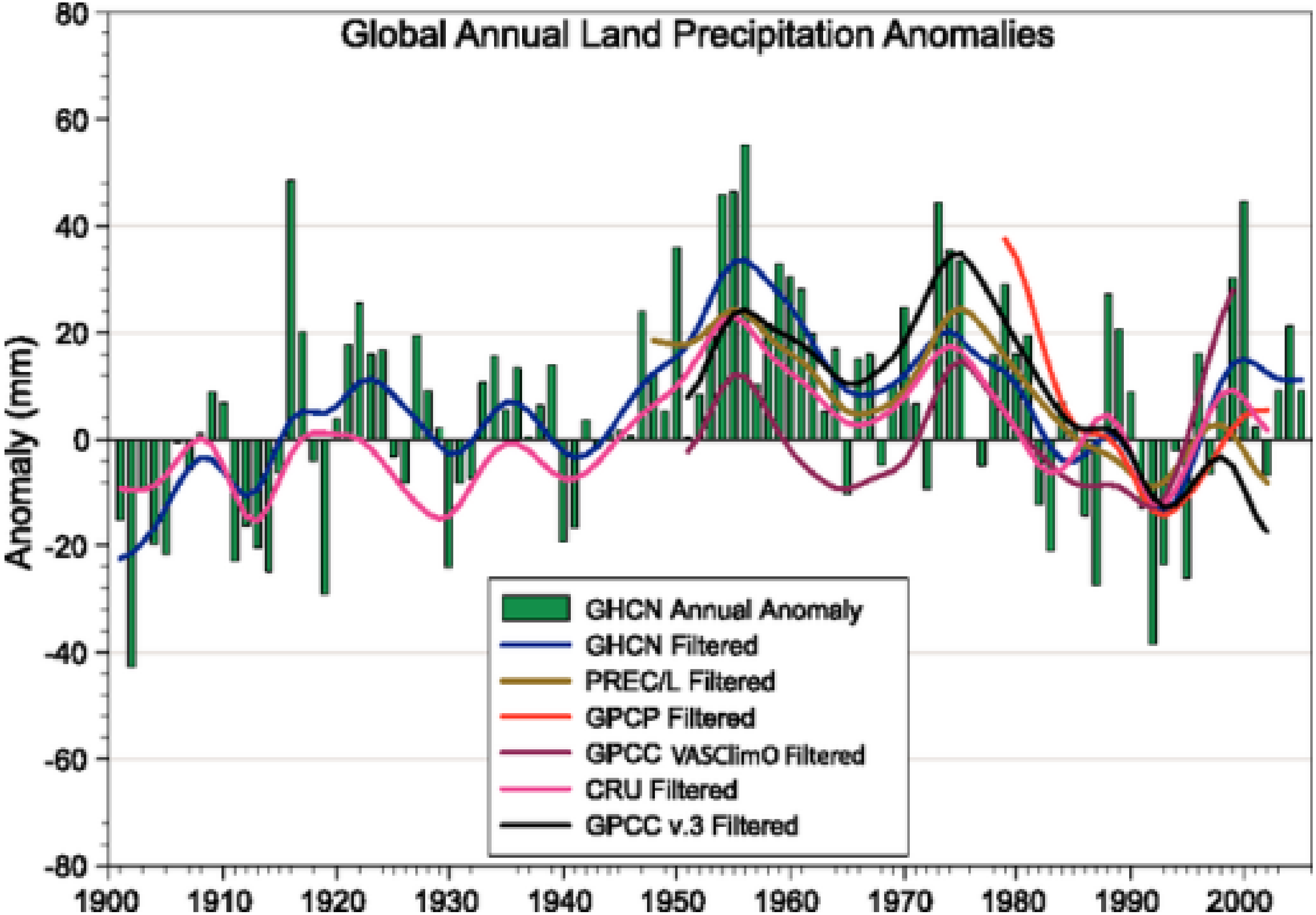
trends/cycles in climate data



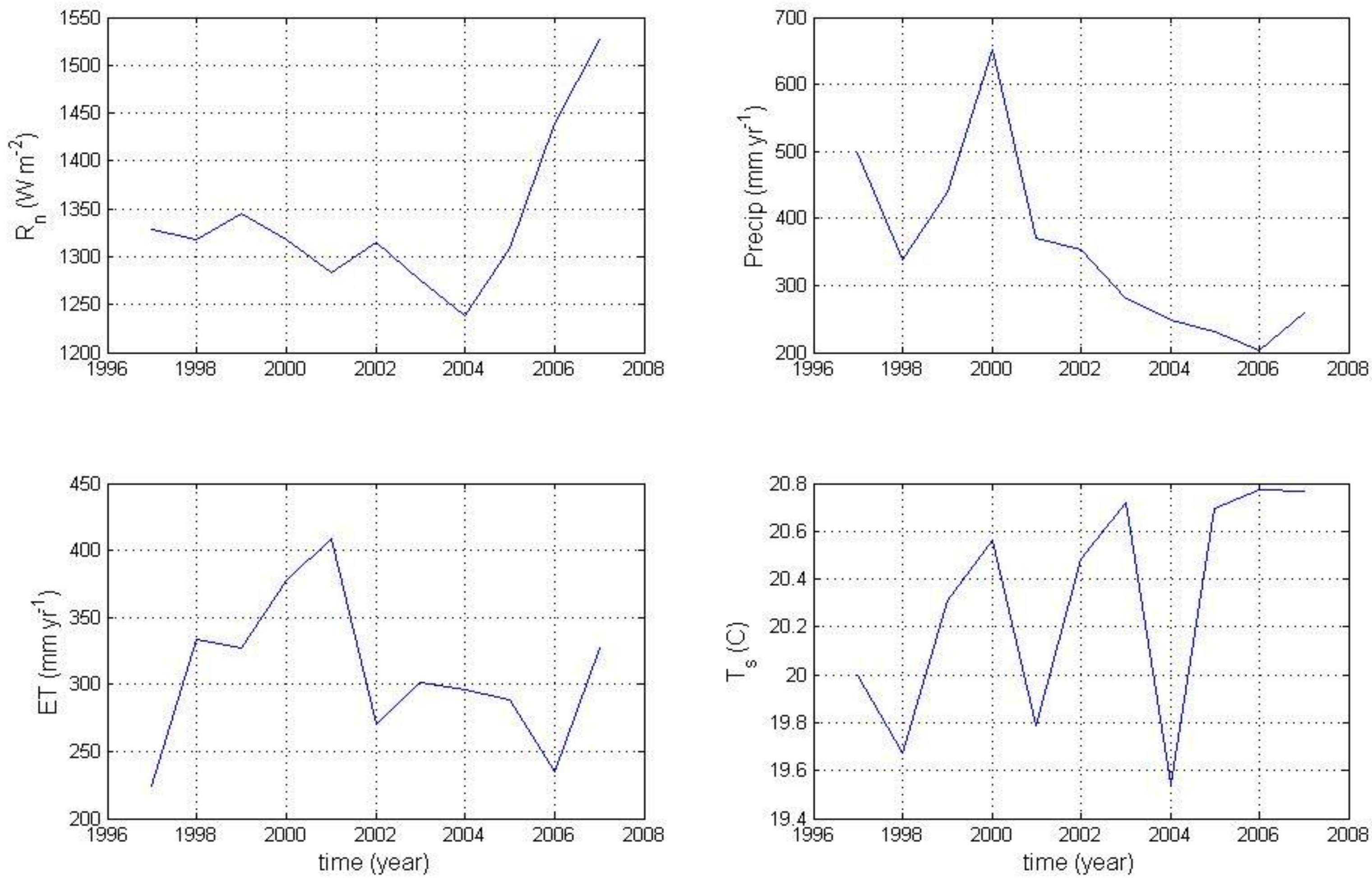
IPCC AR4 Report



Annual anomalies of global land-surface air temperature (°C), 1850 to 2005, relative to the 1961 to 1990 mean for CRUTEM3 updated from Brohan et al. (2006). The smooth curves show decadal variation. (IPCC AR4 Report)



Time series for 1900 to 2005 of annual global land precipitation anomalies (mm) from GHCN with respect to the 1981 to 2000 base period.



Field observations of net radiation R_n, precipitation, evapotranspiration ET and surface temperature T_s in Southern Arizona.

C: To detect trend/cycles from noisy data with space-time coverage and resolution.

C: To detect trend/cycles from noisy data with space-time coverage and resolution.

Q: Are the classical statistical methods, e.g. linear and nonlinear regression/curve-fitting, spectral analysis methods, sufficient in detecting trends/cycles in climate data?

C: To detect trend/cycles from noisy data with space-time coverage and resolution.

Q: Are the classical statistical methods, e.g. linear and nonlinear regression/curve-fitting, spectral analysis methods, sufficient in detecting trends/cycles in climate data?

A: Bayesian method.

Bayesian Linear Regression

(Jaynes, 1990)

$$Y = \alpha + \beta X$$

$$D \equiv \{(x_1, y_1), \cdots, (x_n, y_n)\}$$

$$x_i = X_i + \delta x_i, \qquad y_i = Y_i + \delta y_i, \qquad 1 \le i \le n$$

$$Y_i = \alpha + \beta X_i$$

$$\delta x_i \sim N(0, \sigma_{x_i}), \qquad \delta y_i \sim N(0, \sigma_{y_i})$$

In the Bayesian linear regression analysis where α, β are the parameter of interest, the posterior distribution $\boldsymbol{p(\alpha, \beta | D)}$ can be derived from the "full" posterior distribution with $3n$ nuisance parameters $\{X_1, \cdots, X_n, \sigma_{x_1}, \cdots, \sigma_{x_n}, \sigma_{y_1}, \cdots, \sigma_{y_n}\}$.

$$p(\alpha,\beta|D)=\int\cdots\int \underbrace{p\left(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots|D\right)}_{\uparrow} dX_i\, d\sigma_{x_i}\, d\sigma_{y_i}$$

"full" posterior distribution

$$p(\alpha,\beta|D)=\int\cdots\int \underbrace{p\left(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots\middle|D\right)}_{\uparrow}dX_i\,d\sigma_{x_i}\,d\sigma_{y_i}$$

“full” posterior distribution

$$p\left(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots\middle|D\right)$$

$$\propto \underbrace{p_0(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots)}_{\uparrow}\underbrace{L(D|\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots)}_{\uparrow}$$

prior **likelihood**

$$p(\alpha,\beta|D) = \int\cdots\int \underbrace{p\left(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots \middle| D\right)}_{\text{"full" posterior distribution}} dX_i\, d\sigma_{x_i}\, d\sigma_{y_i}$$

"full" posterior distribution

$$p\left(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots \middle| D\right)$$

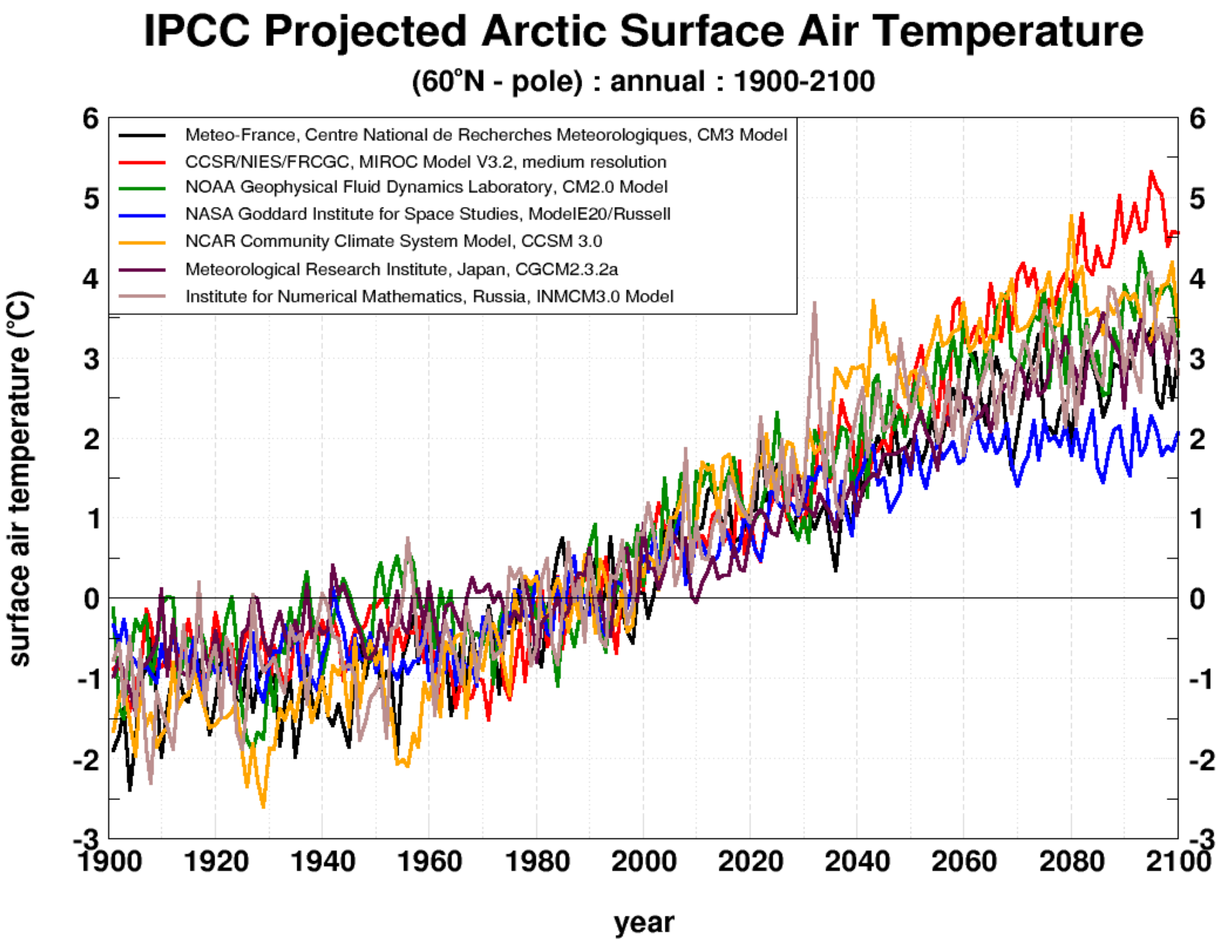
$$\propto \underbrace{p_0(\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots)}_{\text{prior}}\underbrace{L(D|\alpha,\beta,\cdots,X_i,\sigma_{x_i},\sigma_{y_i},\cdots)}_{\text{likelihood}}$$

prior **likelihood**

p_0 and ***L*** may be derived using the *Maximum Entropy* method.

- detecting trends and cycles in climate data,
- ensemble simulations for forecast and data assimilation.

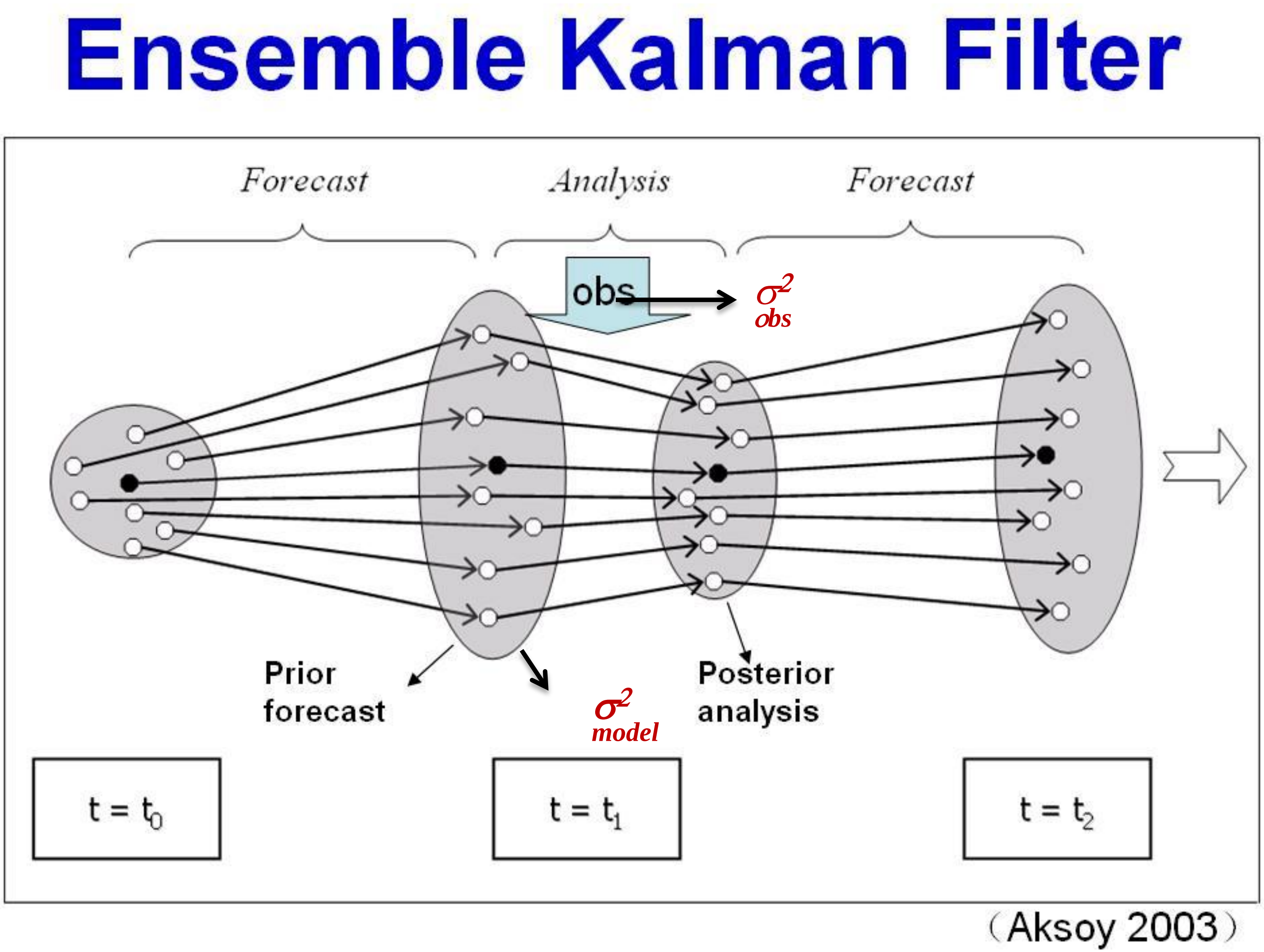
Ensemble Simulations



Multi-model ensemble
Perturbed-physics ensemble
Multi-scenario ensemble
….

statistics of uncertainty

Ensemble Kalman Filter Method of Data Assimilation



posterior **prior** **data**

$$X^{forecast} = \frac{\sigma_{model}^{-2}}{\sigma_{model}^{-2} + \sigma_{obs}^{-2}} X^{model} + \frac{\sigma_{obs}^{-2}}{\sigma_{model}^{-2} + \sigma_{obs}^{-2}} X^{obs}$$

statistics from ensembles

C: The ensemble simulations that take long computing time are not fully utilized.

C: The ensemble simulations that take long computing time are not fully utilized.

Q: Are there more efficient methods to characterize uncertainty in model projected future climate without using ensembles?

C: The ensemble simulations that take long computing time are not fully utilized.

Q: Are there more efficient methods to characterize uncertainty in model projected future climate without using ensembles?

A: **?**