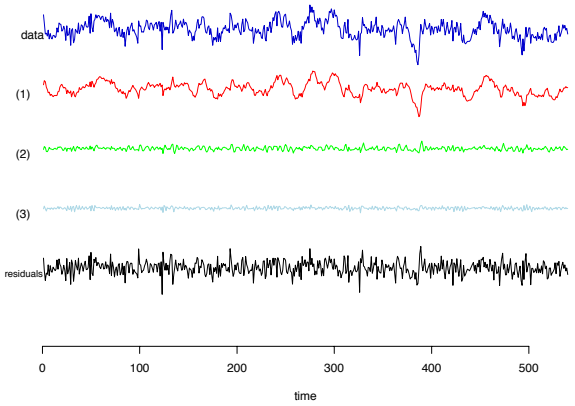
Introduction to a Bayesian Probabilistic Framework

Gabriel Huerta

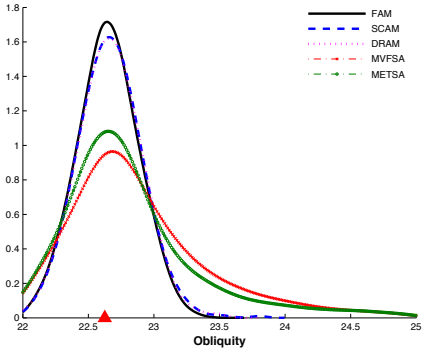
Department of Mathematics and Statistics
University of New Mexico
http://math.unm.edu

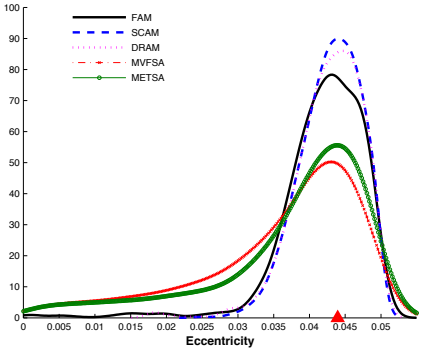
Interests

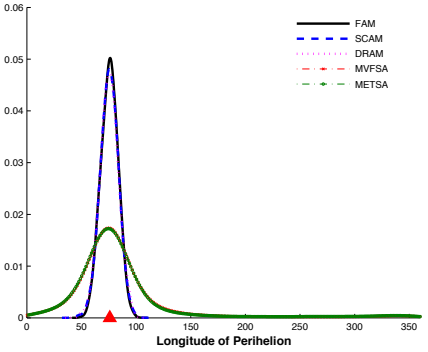
- Bayesian methods, hierarchical models and time series



- *MCMC simulation*.







Thomas Bayes 1701 - 1763

- English mathematician and presbyterian minister.
- Born in Londres y died in Tunbridge Wells Kent.
- Special case,

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}.$$

- Image from the book, ”History of Life Insurance” (1936).

Bayes and the *Binomial* distribution

- X number of times an event has occurred over n trial.
- θ probability of occurrence in one trial.
- If a and b are *degrees of probability*, Bayes required:

$$Pr(a < \theta < b | X = x)$$

- Under *independent* occurrences (postulate),

$$P(X = x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}; x = 0, 1, 2, \ldots, n$$

- **Bayes' billiard table**: Prior distribution $\theta \sim U(0,1)$

$$p(\theta) = 1; 0 < \theta < 1.$$

- If we observe $X = x$,

$$p(\theta|X = p) = \frac{P(X = p|\theta)p(\theta)}{P(X = p)}.$$

- Or,

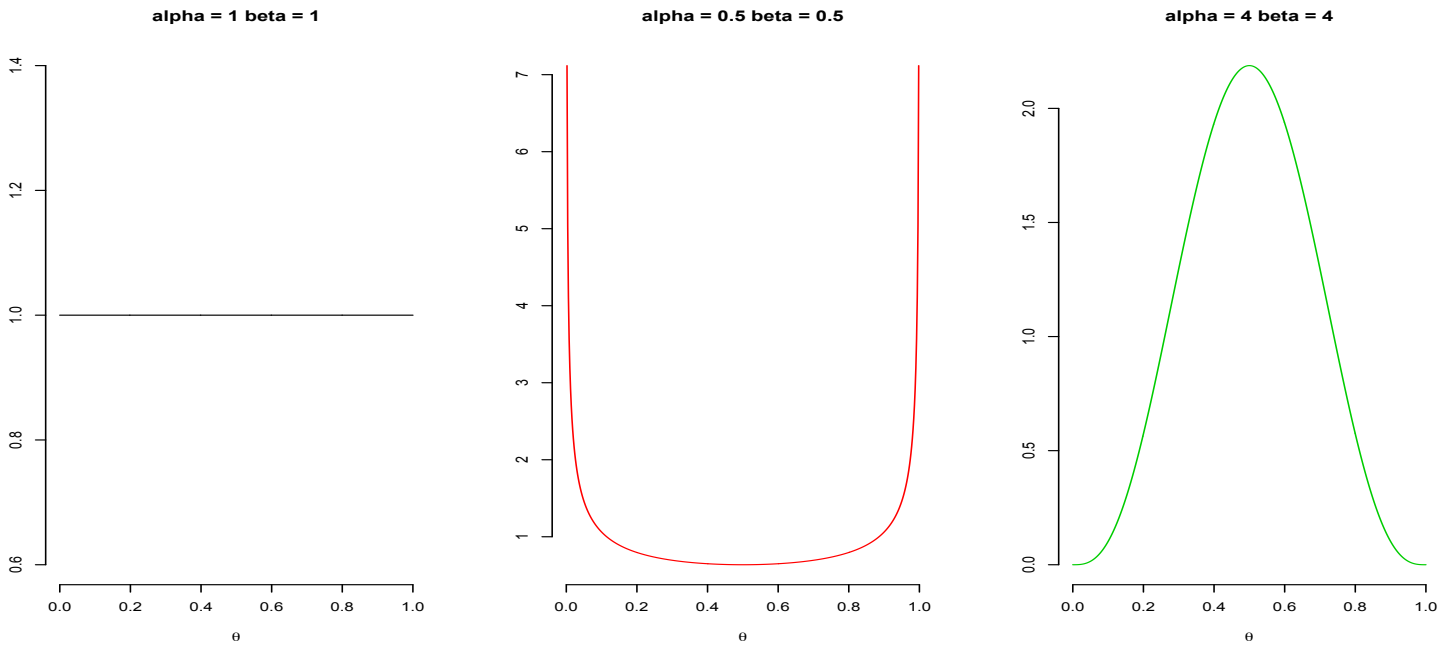
$$p(\theta|X = x) \propto P(X = x|\theta)p(\theta) \propto \theta^x(1-\theta)^{n-x}$$

- **Beta** density,

$$p(\theta|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- with $\alpha = x + 1$, $\beta = n - x + 1$.

- $\theta \sim Beta(\alpha, \beta)$, as a prior $p(\theta)$.

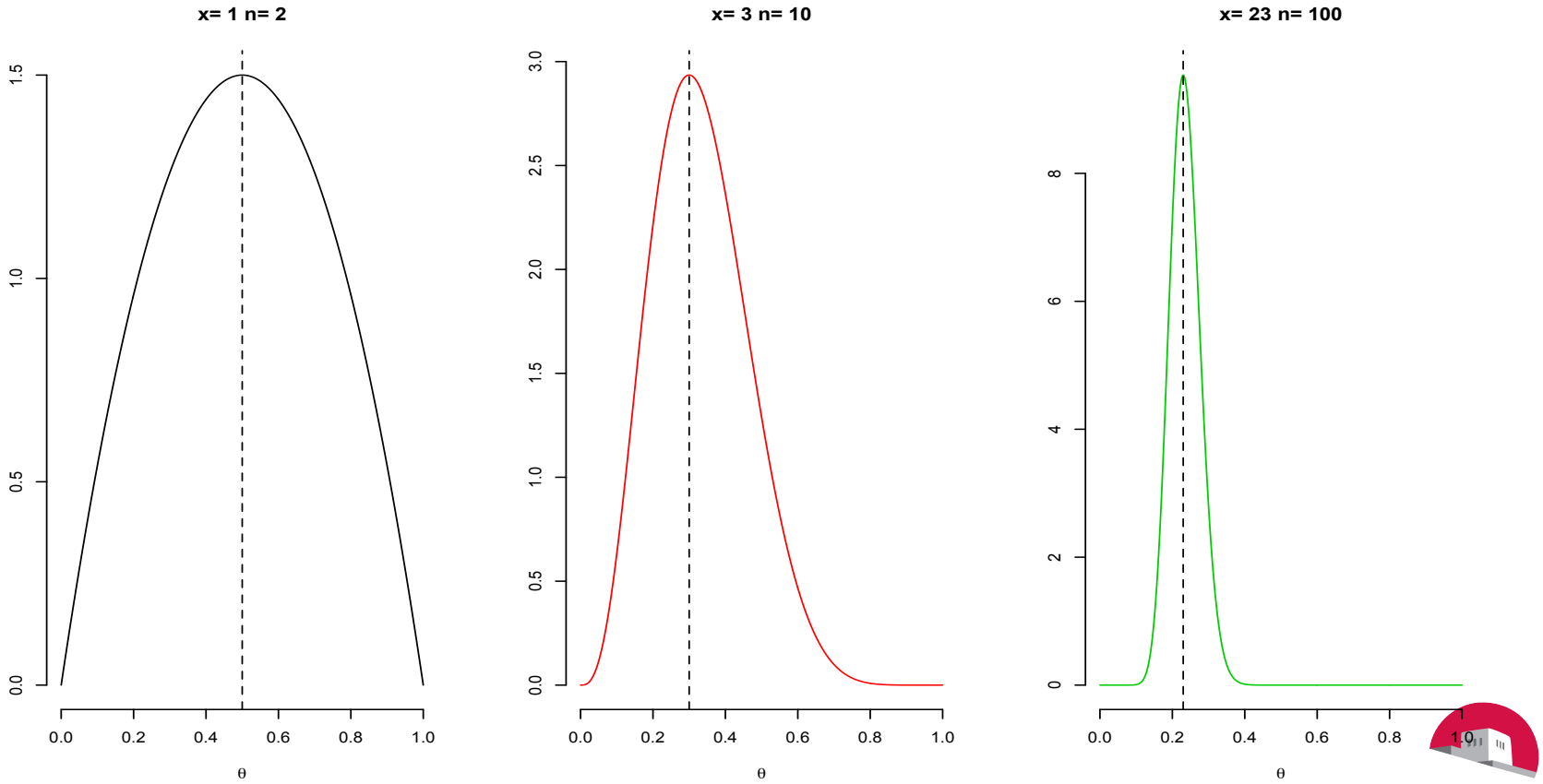


- The posterior distribution,

$$
\begin{aligned}
p(\theta|X = x) &\propto P(X = x|\theta)p(\theta) \\
&\propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}
\end{aligned}
$$

Posterior distributions for various x and n

- $\alpha = \beta = 1$ (uniform prior)

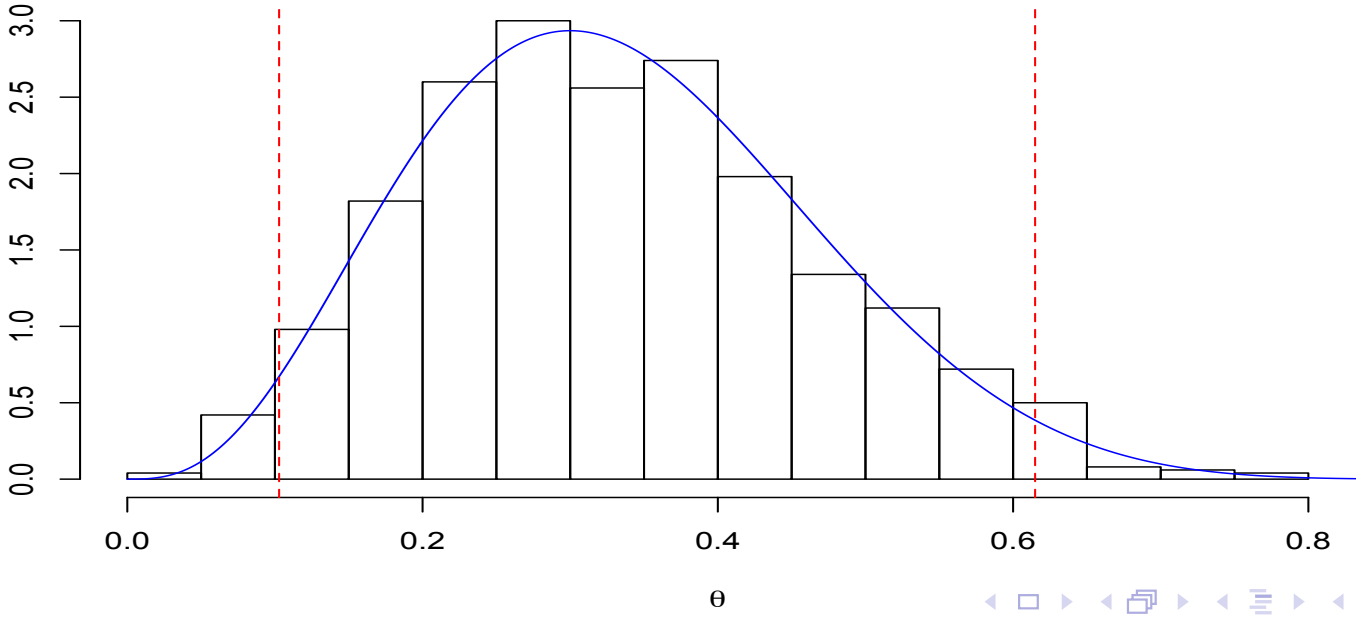


- For n=10, what is $P(0.3 < \theta < 0.6|X = 3)$?
- Answer in R, http://cran.r-project.org/

```
> x=3;n=10
> alpha=beta=c(1,0.5,4)
> alpha=alpha+x
> beta=beta+n-x
> pbeta(0.6,alpha,beta)-pbeta(0.3,alpha,beta)
[1] 0.5402809 0.4918837 0.7404026
# Intervals at 95 \% probability
> qbeta(0.025,alpha,beta)
[1] 0.10926344 0.09269459 0.18443696
> qbeta(0.975,alpha,beta)
[1] 0.6097426 0.6058183 0.6167163
```

Monte Carlo Simulation

```
> x=3;n=10
> alpha=beta=1; alpha=alpha+x; beta=beta+n-x
> theta=rbeta(1000,alpha,beta)
> sum(0.3 < theta & theta < 0.6 )/1000
[1] 0.523
```



Building a Beta prior [1]

- Suppose $\theta \approx 0.6$ (most favorable value)
- Quantile 95% is close to 0.8,

$$P(\theta < 0.8) = 0.95$$

- Is there a Beta(α,β) *matching prior* ?
- Mode: $\frac{\alpha-1}{\alpha+\beta-2} = 0.6$

```
> theta0 <- 0.6; btilde <- 1:100
> a <- (1+theta0*(btilde-2)) / (1-theta0)
> qbeta(0.95,a,btilde)
[1] 0.950 0.887 0.847 0.819 0.800
> a[5]; btilde[5]
7  5
```

[1] Christensen R. et al (2010) . *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians* CRC Press

Bayesian inference

- Interest in the parameter vector (or scalar) $\boldsymbol{\theta}$.
- Prior knowledge expressed as $p(\boldsymbol{\theta})$.
- $\boldsymbol{X} = (X_1, \ldots, X_n)$ with joint distribution $f(\boldsymbol{X}|\boldsymbol{\theta})$.
- Compute $p(\boldsymbol{\theta}|\boldsymbol{X})$, the posterior distribution.
- *Bayes theorem*

$$p(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Theta} f(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

- Exact form only for *conjugate* models.
- Needs stochastic simulation *Markov Chain Monte Carlo* (MCMC).

More comments [2]

- Bayesian ideas not too developed in the 50's.
- Low intensity in the 70's in U.S.
- W. Feller:
  - *Neyman-Pearson theory* capable of solving "all" applied problem. (Normal, Poisson, Binomial, etc.)
  - Best approach to solve engineering problems (type I and II errors).

[2] A. Gelman and C. Robert (2014) Not Only Defended But Also Applied: The Perceived Absurdity of Bayesian Inference *The American Statistician* Vol. 1 pp 1-5

Comments

" We know all use Bayesian inference when its clearly appropriate"

- Random mechanisms: coin toss, dice, gene mixture, roulettes, etc.
- Bayesian adds an extra step and thinks that θ is random, like, $\theta \sim U(0, 1)$.
- Similar to fitting a regression to some data.
- The turning point...MCMC.

Gibbs Sampling [3]

- Suppose the distribution $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ is determined by $\{p_i(\theta_i|\theta_{j\neq i}); i = 1, 2, \ldots, k\}$ (full conditionals).
- Given a starting value, $(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)})$,
  - Sample $\theta_1^{(1)} \sim p_1(\theta_1|\theta_2^{(0)}, \ldots, \theta_k^{(0)})$,
  - Sample $\theta_2^{(1)} \sim p_2(\theta_2|\theta_1^{(1)}, \ldots, \theta_k^{(0)})$,

    $\vdots$
  - Sample $\theta_k^{(1)} \sim p_k(\theta_k|\theta_1^{(1)}, \ldots, \theta_{k-1}^{(1)})$.
- Under general conditions,

$$(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}) \to (\theta_1, \theta_2, \ldots, \theta_k) \sim p(\boldsymbol{\theta})$$

[3] Sampling-Based Approaches to Calculating Marginal Densities. A. E. Gelfand and A.F. M. Smith, JASA. Vol. 85, No. 410 (1990), pp. 398-409.

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

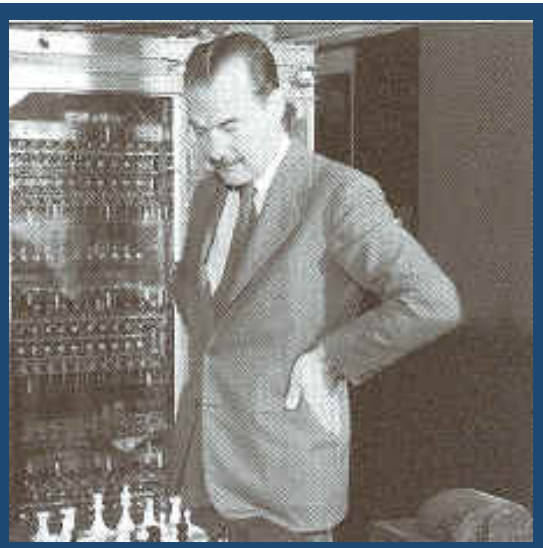
AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Idea of Markov Chains for estimating density of states comes from physics. Experiments determine individual interactions, computers simulate how a system of particles behave.



Metropolis et al. 1953

Metropolis-Hastings [4]

At iteration t,

- Sample $\boldsymbol{z} \sim q(\boldsymbol{z}, \boldsymbol{\theta}^{(t-1)})$.
- Accept the candidate point $\boldsymbol{z}$, with probability

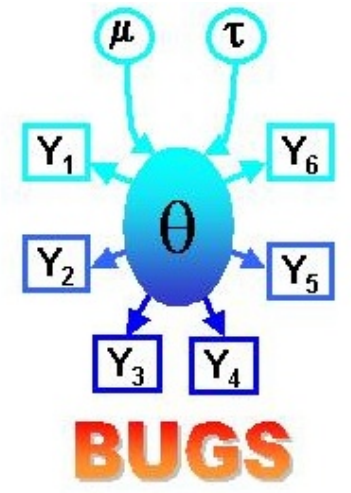
$$\alpha(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{z}) = min\Big(1, \frac{\pi(\boldsymbol{z})q(\boldsymbol{z}, \boldsymbol{\theta}^{(t-1)})}{\pi(\boldsymbol{\theta}^{(t-1)})q(\boldsymbol{z}, \boldsymbol{\theta}^{(t-1)})}\Big)$$

  in which case, $\boldsymbol{\theta}^{(t)} = \mathbf{z}$. Otherwise $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
- Monitor on acceptance rates or trace plots to assess **convergence**.

[4] Monte Carlo Sampling Methods Using Markov Chains and Their Applications. W.K. Hastings, Biometrika, Vol. 57, No. 1. (1970), pp. 97-109

Openbugs

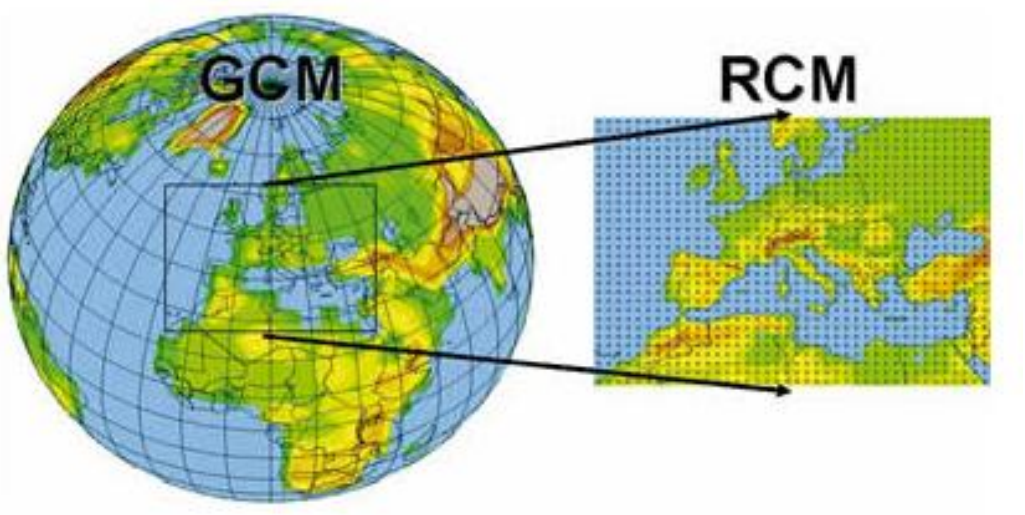
- Generic software to implement Bayesian models.
- Available from www.openbugs.net
- BUGS stands for *Bayesian Inference under Gibbs Sampling*.
- Based on *acyclical graphs*:



- Limited for high-scale applications.
- *JAGS* interfaces with R.

Climate Models

- Defined via differential equations (numerical model).
- *General Circulation Models* (GCM's)
  - cover the Earth. Grids boxes on scale of 100's kms.
- *Regional climate models* (RCM's)
  - resolve processes at smaller scale.



Original question

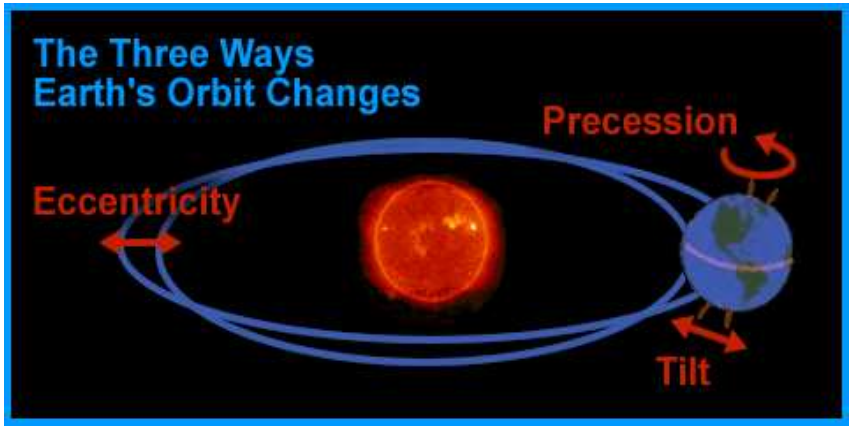
- Approximate

$$\text{posterior}(m) = \frac{\exp[-0.5 * E(m)] \times \text{prior}(m)}{\int exp[-0.5 * E(m)] \times \text{prior}(m)\, dm}$$

- $E(m)$ is a metric of *model skill* (cost function).
- m are *some climate parameters* .
- $E(m)$ considers observations (d_{obs}) and *model runs* $g(m)$.
- Stochastic sampling based on *optimization*.

Surrogate climate model [5]

- **Response:** surface air temperature anomalies.
- Obliquity (Φ'): Earth's axial tilt.
- Eccentricity (e): How elliptical is the Earth's orbit around the Sun.
- Longitud of Perihelion: (λ) Point of closest approach to Sun.



[5] Computational methods for parameter estimation in climate models. A. Villagran et al. Bayesian Anal. Vol. 3, No. 4 (2008), pp. 823-850.

Other aspects

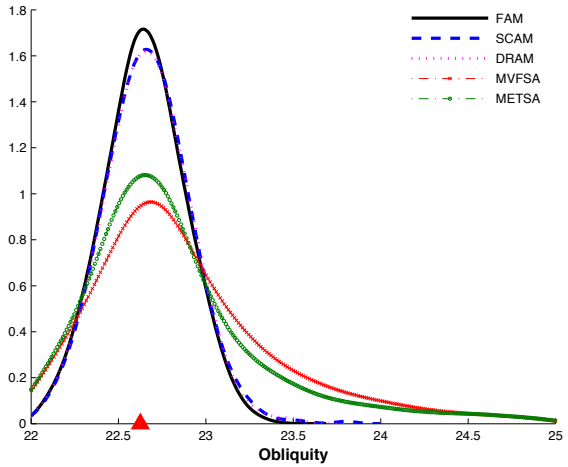
- **Cost function:** Measure of the deviation between the observed data and the model.

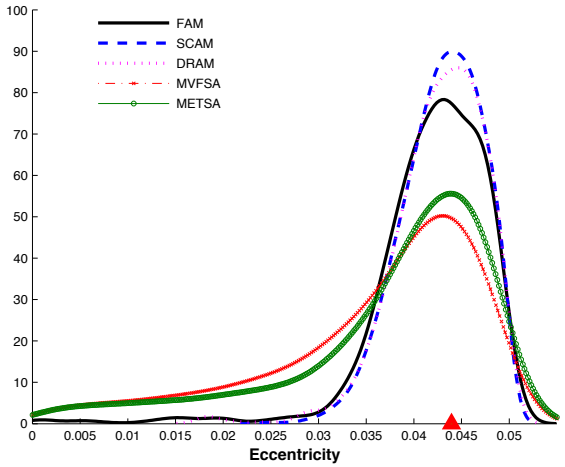
$$E(m) = \sum_{i=1}^{I=40} \sum_{j=1}^{J=48} B_{ij}^{-1} (d_{obs,ij} - g_{ij}(m))^2$$

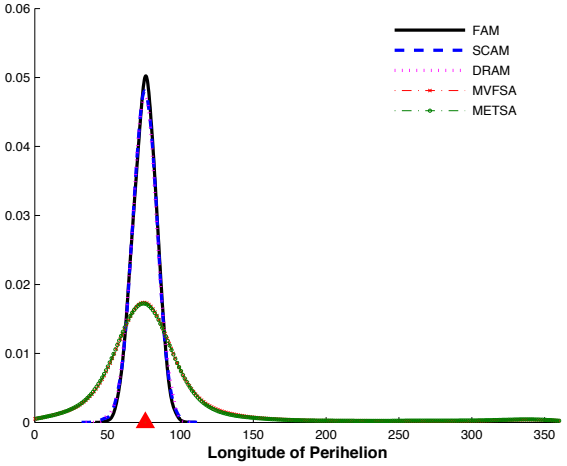
  - d_{obs} are "observations" and $g(m)$ is "climate model output".
  - $m = (\Phi', e, \lambda)$.
  - B_{ij} is the variance of the observations at each grid point.
- **Posterior distribution:** $p(m|d_{obs})$,

$$p(m|d_{obs}) \propto exp\,(-0.5 * E(m))\, p(m)$$

Posterior probability distributions for $m = (\Phi', e, \lambda)$







Multiple Very Fast Simulated Annealing (MVFSA)

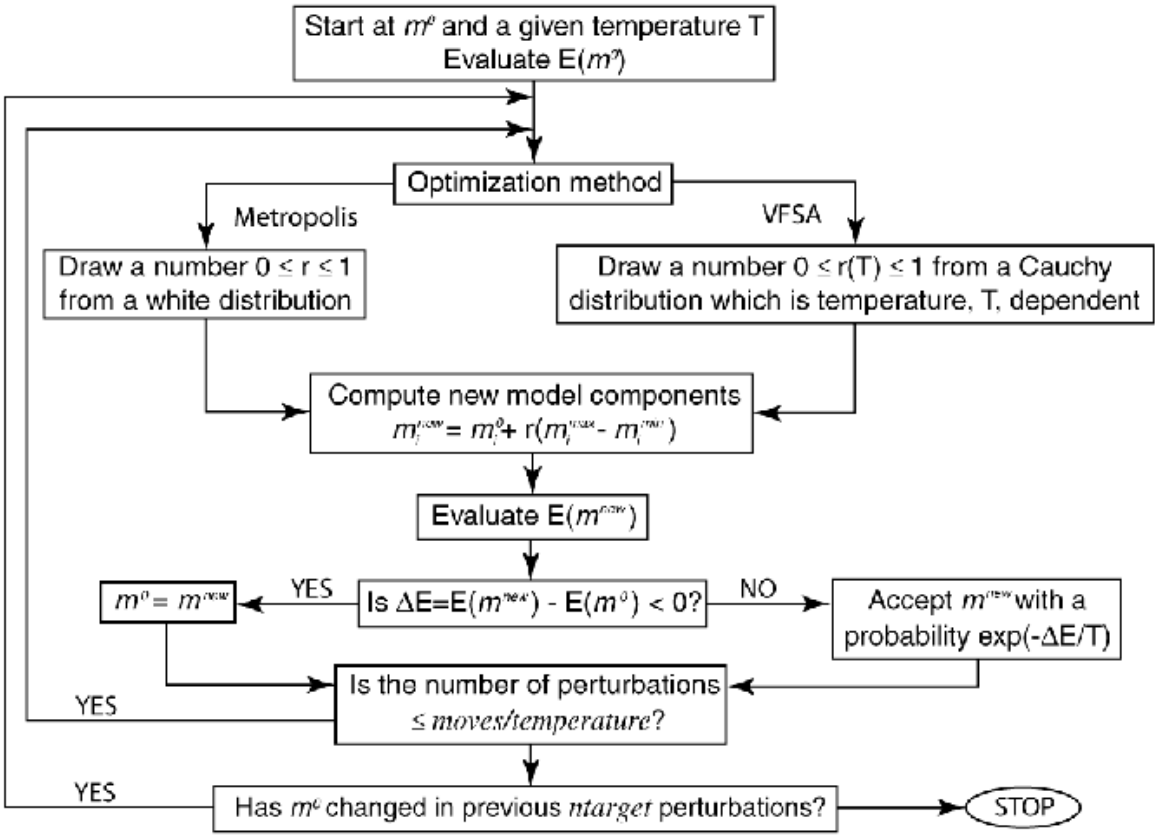
- Approach for *Bayesian inversion* used in Geosciences.
- Depends on a *cooling schedule* and *Metropolis* rule.
- Ingber (1989),

$$m_i^{(k+1)} = m_i^{(k)} + y_i(m_i^{max} - m_i^{min})$$

- y_i from some *proposal distribution*.
- Sen and Stoffa (1996), suggest multiple repetitions.
- ”fast, requires few iterations”.
- Compare to Adaptive Metropolis (AM) and *Delayed rejection* (DRAM).

How Charles thinks about MVFSA

wpe18.gif (GIF Image, 710 × 523 pixels) http://www.ig.utexas.edu/people/staff/charles/images/wpe18.gif



Adaptive Metropolis

Proposed by **Haario et. al. (2001)**

Suppose thus that at time $t-1$ we have sampled the states $m^{(0)}, ..., m^{(t-1)}$.

- Update the covariance matrix C_t of the proposal distribution based on sampled states.
- Sample $\mathbf{z} \sim N_d(m^{(t-1)}, C_t)$.
- Accept the candidate point $\mathbf{z}$ with probability:

$$\alpha(m^{(t-1)}, \mathbf{z}) = min\Big(1, \frac{\pi(\mathbf{z})}{\pi(m^{(t-1)})}\Big)$$

in which case we set $m^{(t)} = \mathbf{z}$, and otherwise $m^{(t)} = m^{(t-1)}$.

Delayed Rejection Adaptive Metropolis

DRAM (Haario et al., 2006)

At current position $m^{(t-1)}$. A new candidate z_1 is generated from a proposal $q_1(m^{(t-1)}, \cdot)$ and accepted with probability

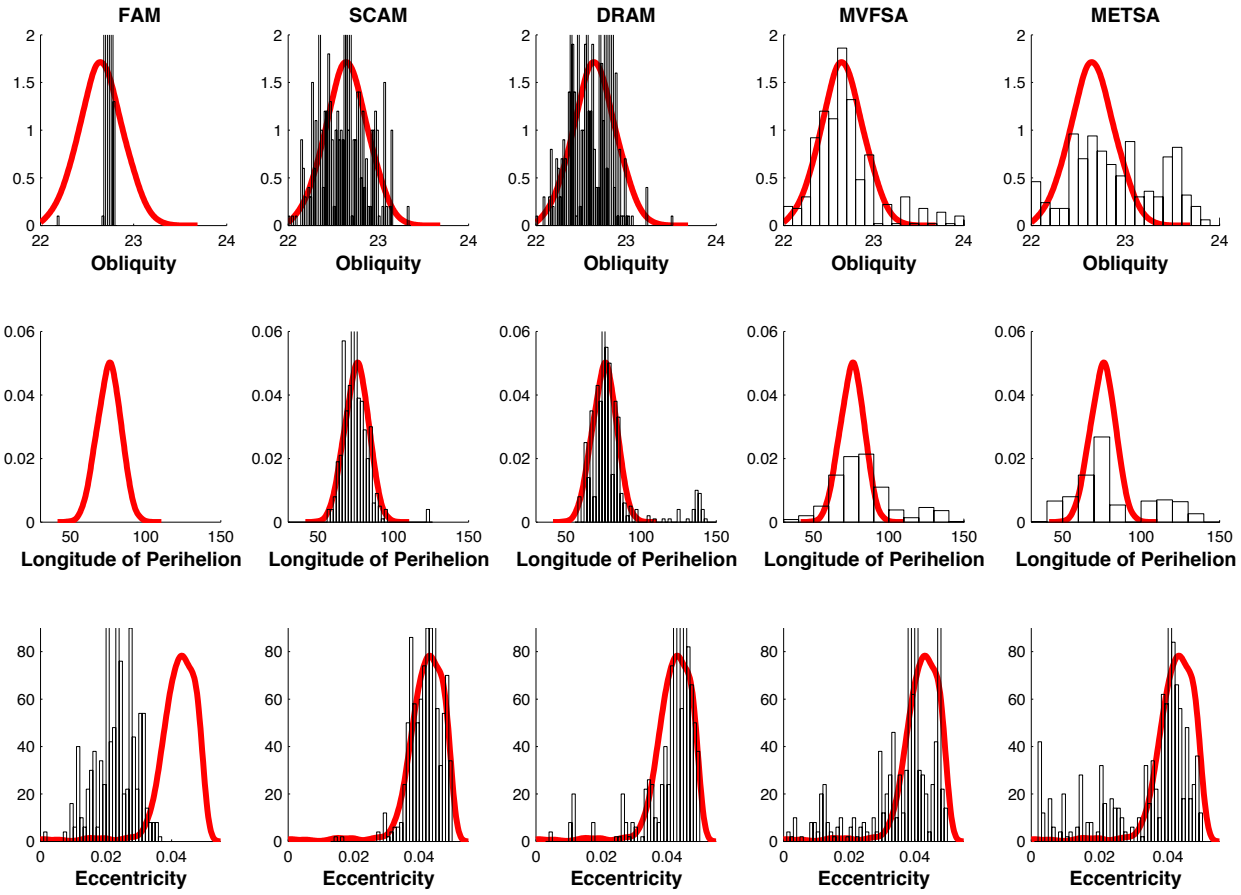
$$\alpha_1(m^{(t-1)}, z_1) = min\left(1, \frac{\pi(z_1)q_1(z_1, m^{(t-1)})}{\pi(m^{(t-1)})q_1(m^{(t-1)}, z_1)}\right).$$

- Upon rejection, instead of retaining $m^{(t)} = m^{(t-1)}$.
- Propose a second state move z_2 from $q_2(m^{(t-1)}, z_1, \cdot)$.
- z_2 is accepted with probability $\alpha_2(m^{(t-1)}, z_1, z_2)$.

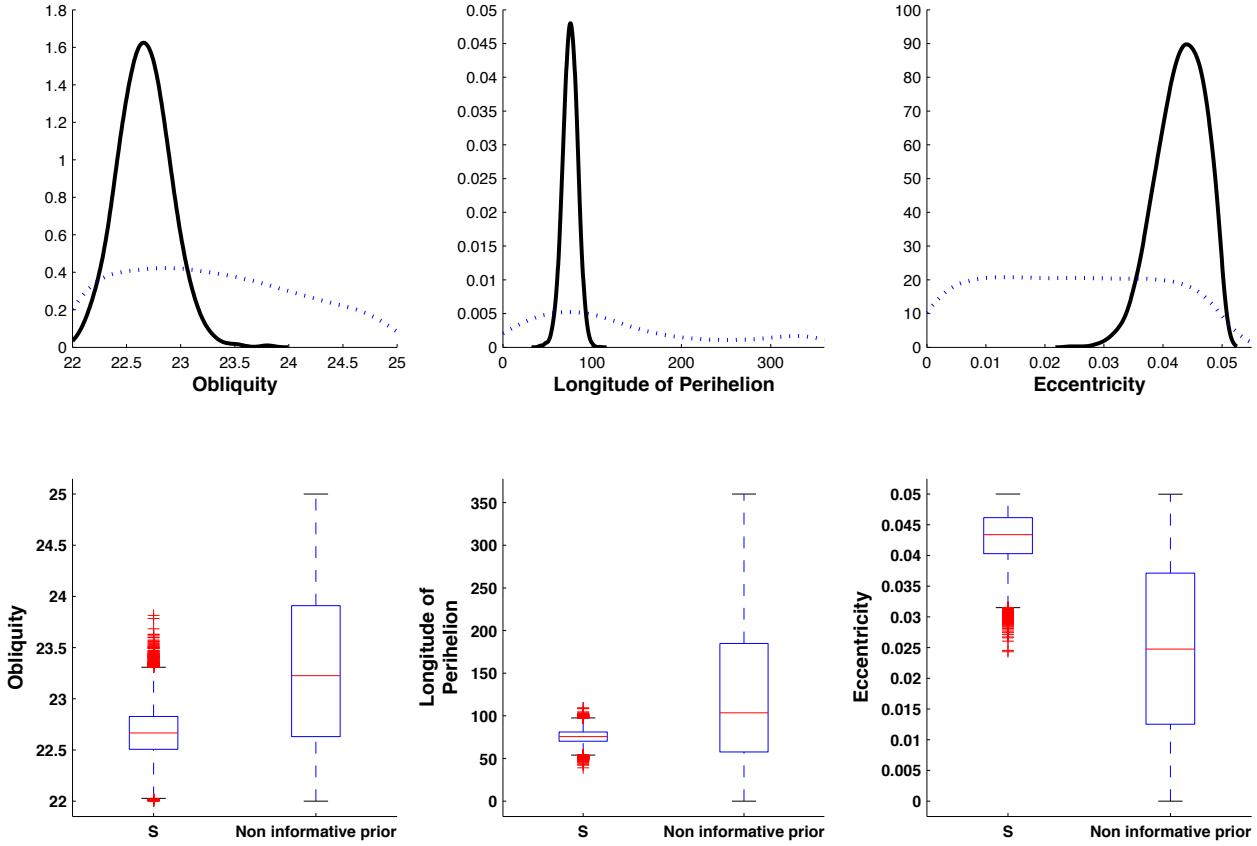
Comments:

- Different strategies to implement it.
- It is a way of combining different proposals.

Comparison of sampling methods



Comparison of $p(m|d_{obs})$

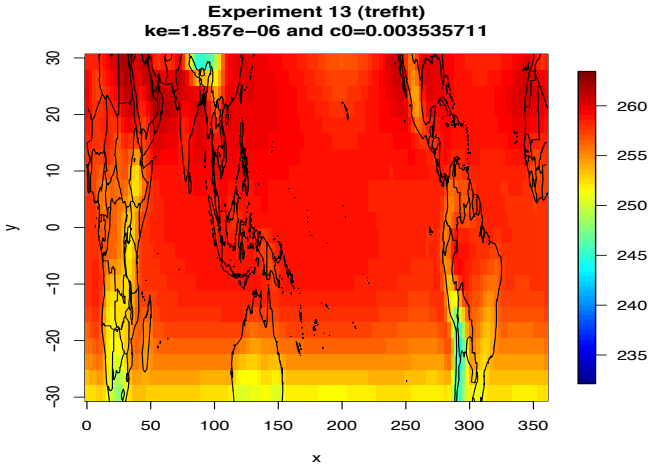


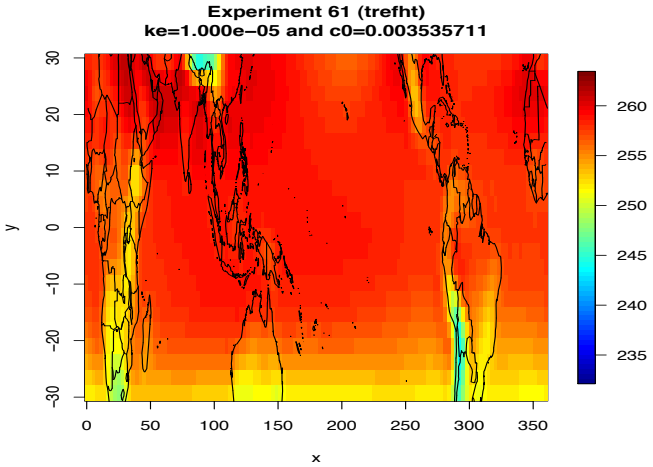
Calibration of Parameters for a Community Atmosphere Model [6]

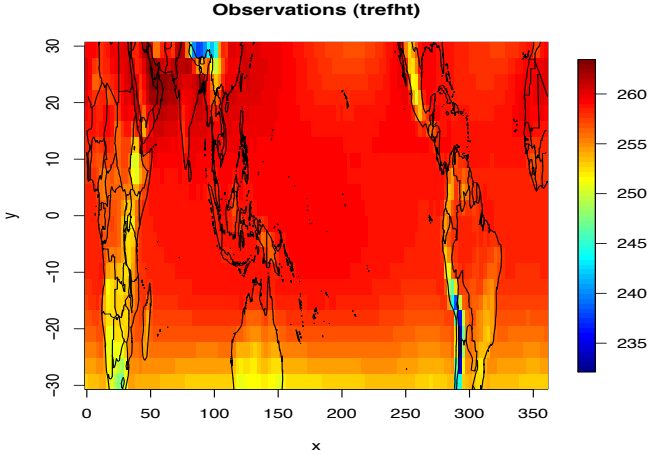
- NCAR Community Atmosphere Model (CAM) version 3.1.
- Model runs of climate model simulation with physical observations.
- Inputs or parameters related to clouds and ice.
- *2-m air temperature* 'field T' or TREFHT.
- Other outputs: *shortwave cloud forcing*, *precipitation over ocean*, *vertically averaged relative humidity*. *latent heat flux over ocean*.
- 7 spatial fields on a grid of 128×64.

[6] C.S. Jackson et al Error Reduction and Convergence in Climate Prediction. Journal of Climate. Vol. 21 (2008), 24, pp. 6698-6709.

Model runs and observations for field ”TREFHT”

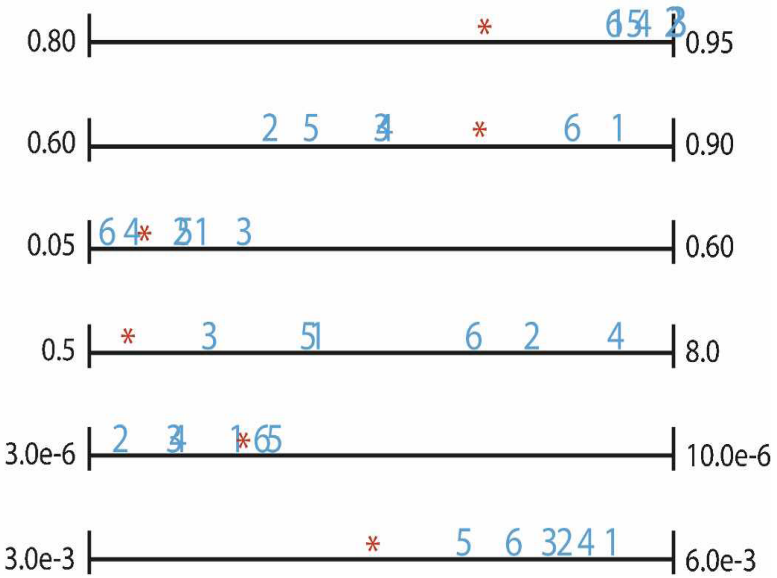




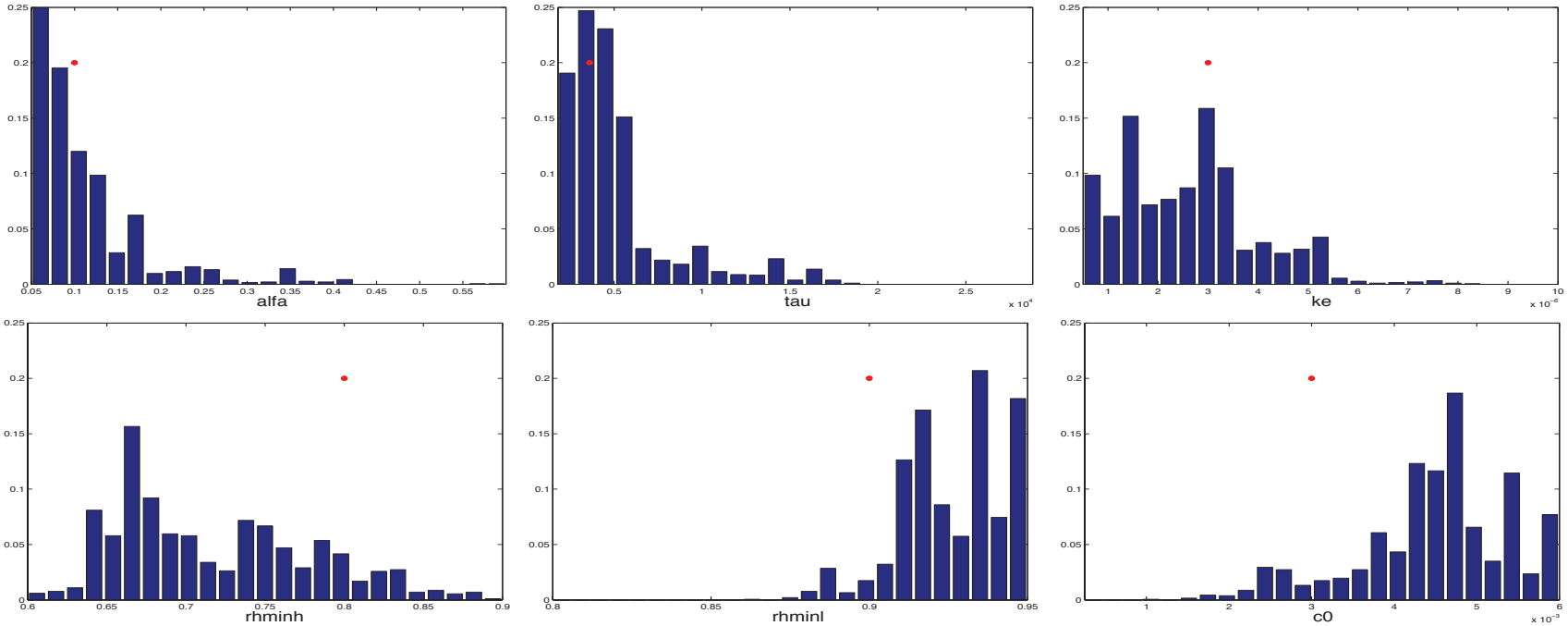


6 parameters from CAM 3.1

Parameter	Definition	Value Ranges
RHMINL [%/100]	Low cloud critical relative humidity	0.80 — 0.95
RHMINH [%/100]	High cloud critical relative humidity	0.60 — 0.90
ALFA [fraction]	Initial cloud downdraft mass flux	0.05 — 0.60
TAU [hours]	Consumption rate of CAPE	0.5 — 8.0
ke $[(\mathrm{kg\ m^{-2}\ s^{-1}})^{-1/2}\mathrm{s}^{-1}]$	Environmental air entrainment rate	3.0e-6 — 10.0e-6
c0 $[\mathrm{m}^{-1}]$	Precipitation efficiency	3.0e-3 — 6.0e-3



Samples for 6 climate parameters



- $\approx$ 1500 experiments (or runs).
- (*) denotes *default values*.

Statistical Emulation with Principal Component Analysis

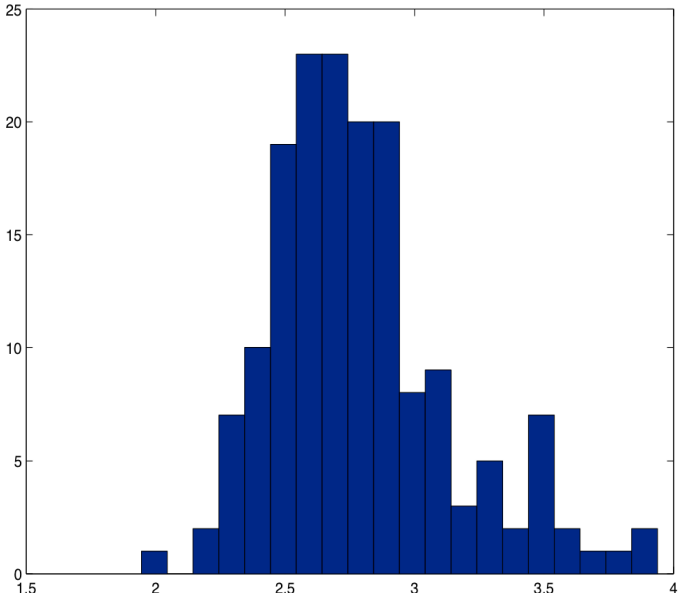
- Emulation with *Gaussian Process*: $g(\cdot) \sim N(\mu(\cdot), C(\cdot,\cdot))$
- Use runs to estimate an *orthogonal* basis.
- Look PCA variability across seasons.
- Model output and observations are summarized with small number of *scores*.
- Follows the approach of Higdon et al. (2012).

$$y_{pc} = g_{pc} + \epsilon_{clim} + \epsilon_{discrep}$$

- $g_{pc} \sim N(\mu_g, \Sigma_g)$; $\epsilon_{clim} \sim N(0, \sigma^2_{clim} I)$; $\epsilon_{discrep} \sim N(0, \Sigma_\delta)$

Global warming experiments

- *Coupled* CAM 3.1 to a ”slab” ocean.
- 165 experiments (runs) were completed.
- *Control* and $2 \times CO2$.
- Climate sensitivity: ”change in global mean temperature after doubling CO2”.



Relating Sensitivity to Climate Fields [7]

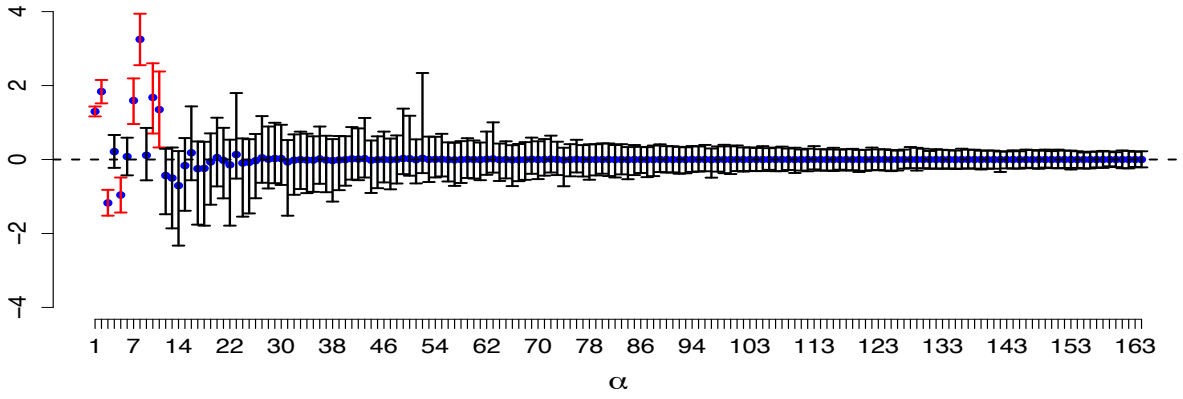
- Regression problem with $n = 165$ and $p = 129 \times 22 \times 7$.
- Perform a *principal component regression* (PCR)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}_n = \boldsymbol{W}\alpha + \boldsymbol{\epsilon}_n$$

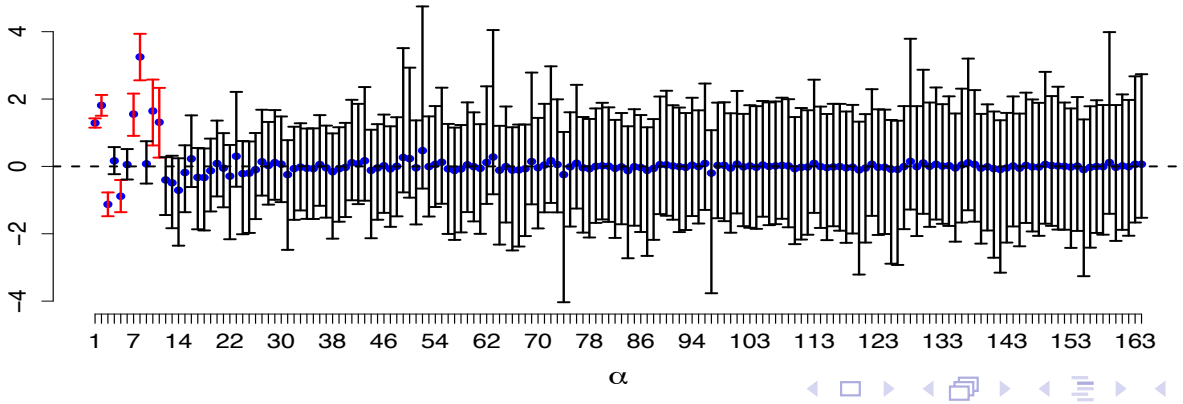
- $\boldsymbol{W}$ denotes a $n \times k$ matrix, with the first k PCs as columns.
- Mapping of α's back to β's.
- Studying *Bayesian* solution under two priors for α:
  - $\alpha_i \sim N(0, (g \times i^{-2})/\phi_i)$ and $\alpha_i \sim N(0, g/\phi_i)$; $i = 1, 2, \ldots, k$.
  - $\phi_i \sim Gamma(\delta, \delta)$ and $g \sim Unif(0, a)$.

[7] M.H. Hattab et al. (2015) A regression between bias and climate sensitivity within a perturbed physics ensemble of CAM3.1 (work in progress)

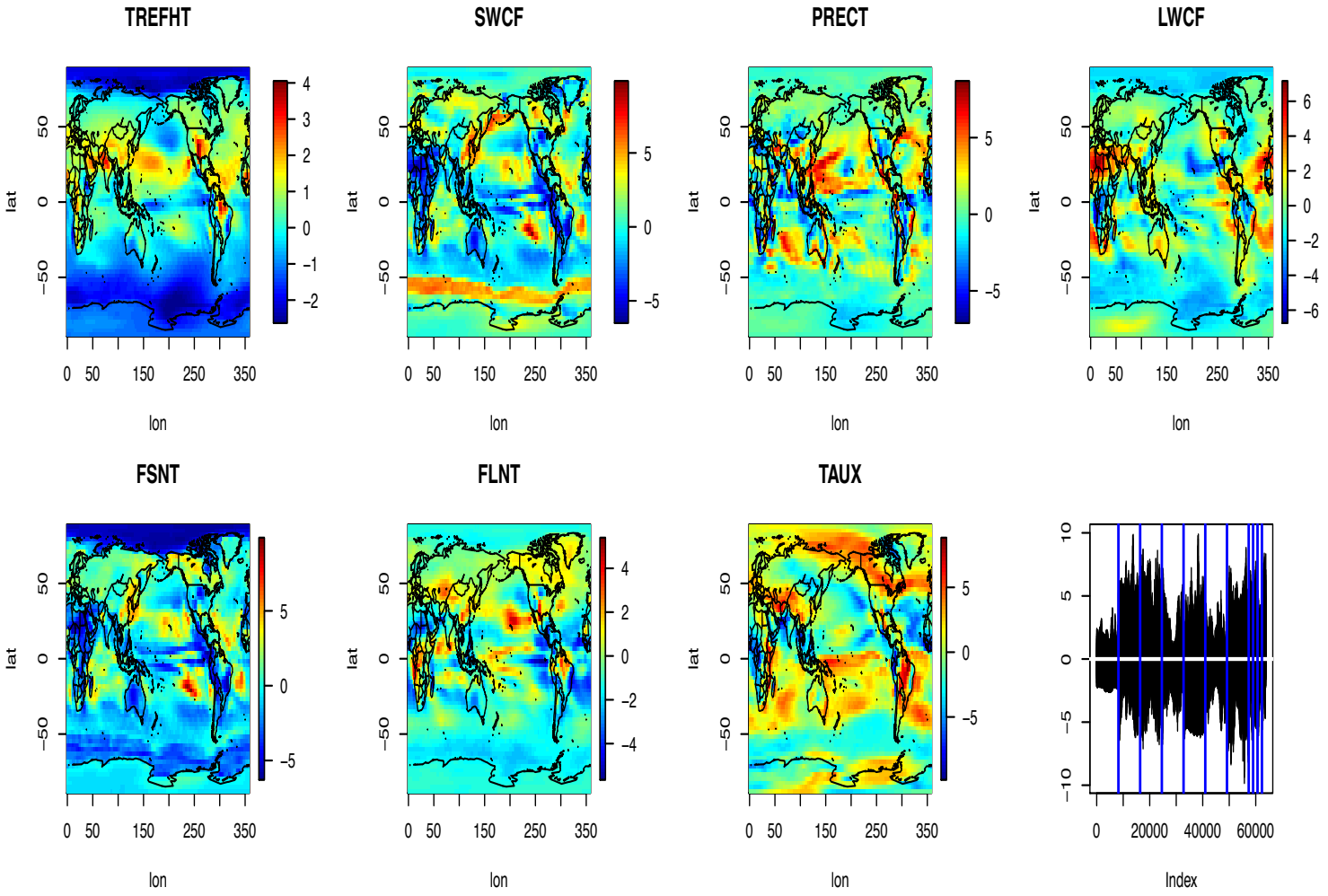
Posterior for α_i under two priors



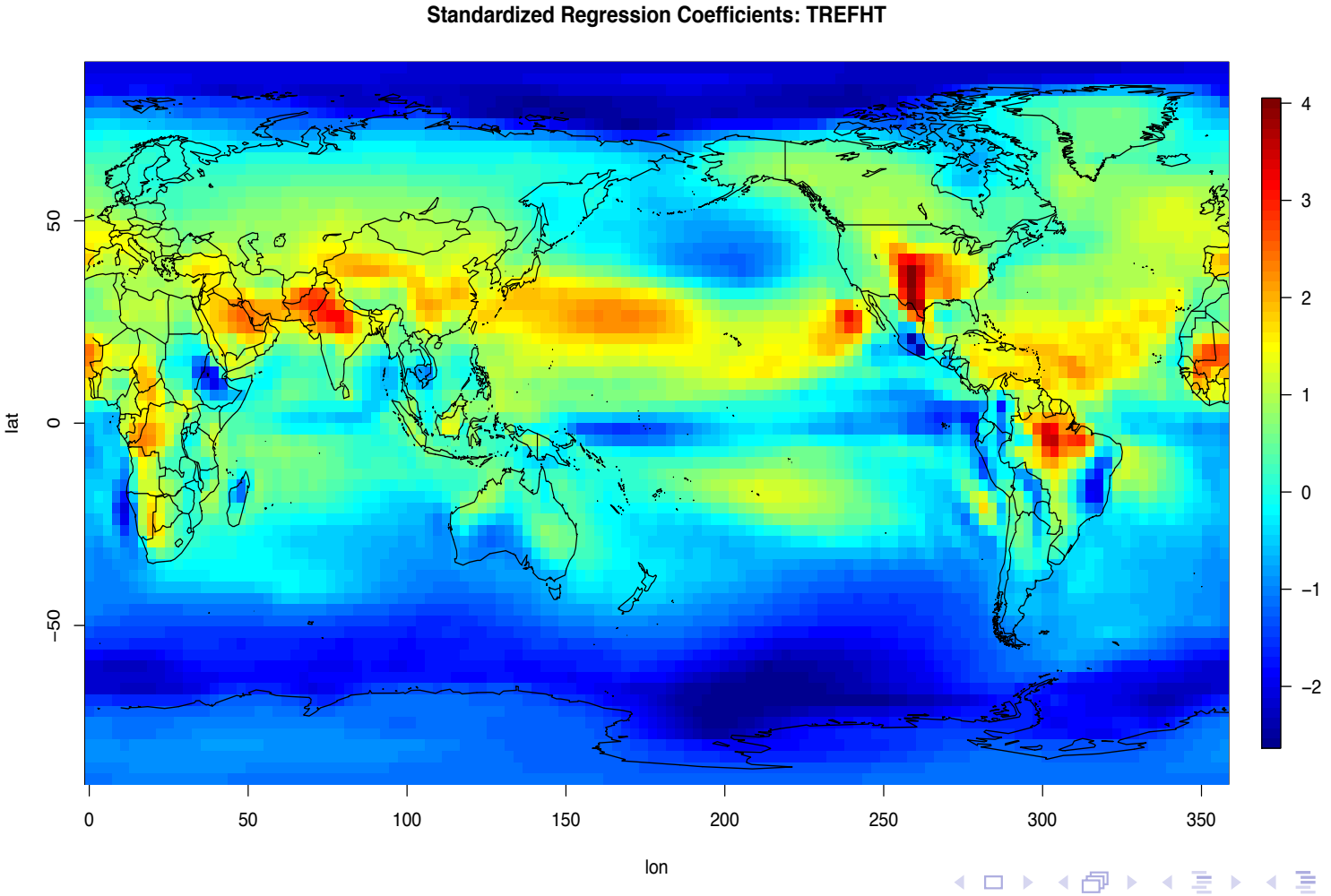
(a) M_1



Standardized Regression Coefficients



TREFHT β


Standardized Regression Coefficients: TREFHT
lat
50
0
−50
0
50
100
150
200
250
300
350
lon
4
3
2
1
0
−1
−2

Computer Lab session

- **Example 1:** Binomial-Beta prior. Files: *binomial.R*. JAGS/Openbugs model: *model-bin.txt*.
- **Example 2:** Regression model with time trend. Files: *regression-example.R*. JAGS/Openbugs model: *model-regression.txt*, *model-regression2.txt*.
- **Example 3:** Calibration of Mg/Ca model. Files: *calibration-script.R*. Dataset: *calibration.txt*. JAGS/Openbugs model: *model-khider3.txt*, *predictive1.txt*. This version and a Matlab version will be at https://github.com/khider/

Thank you.

Bayesian Calibration of *Globigerinoides ruber* Mg/Ca [8]

- Estimate probability distributions of Mg/Ca sensitivity to SST, sea level salinity (SSS) and deep water ΔCO_3^{2-}.
- Data set of 186 core top-samples with "global coverage".

$$\pi(unknows \mid data) \propto f(data \mid unknows) \cdot \pi(unknowns)$$

- $\pi(unknowns|data)$ "probability of unknowns given data" (posterior).
- $f(data|unknowns)$ "likelihood of the data given the unknowns".
- $\pi(unknowns)$ "prior probability of unknowns".

[8] D. Khider et al. in review for G^3 (2015).

Calibration equation for Mg/Ca

- A *piecewise regression:*

$$\text{if } \Delta CO_{3\,(i)}^{2-} \geq 21\,\mu mol/kg,\ Mg/Ca_{(i)} = \left(\exp\left(\alpha_1 T_{(i)} + \alpha_2 S_{(i)} + \alpha_0\right) + \alpha_3 21\right) \Big/ \left(1 + \alpha_4 C_{(i)}\right) + \varepsilon_{(i)}$$

$$\text{if } \Delta CO_{3\,(i)}^{2-} < 21\,\mu mol/kg,\ Mg/Ca_{(i)} = \left(\exp\left(\alpha_1 T_{(i)} + \alpha_2 S_{(i)} + \alpha_0\right) + \alpha_3 \Delta CO_{3\,(i)}^{2-}\right) \Big/ \left(1 + \alpha_4 C_{(i)}\right) + \varepsilon_{(i)}$$

$$\varepsilon_{(i)} \sim N\left(0, \tau^2\right)$$

- $\mathbf{\Phi} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \tau^2)$. By Bayes theorem,

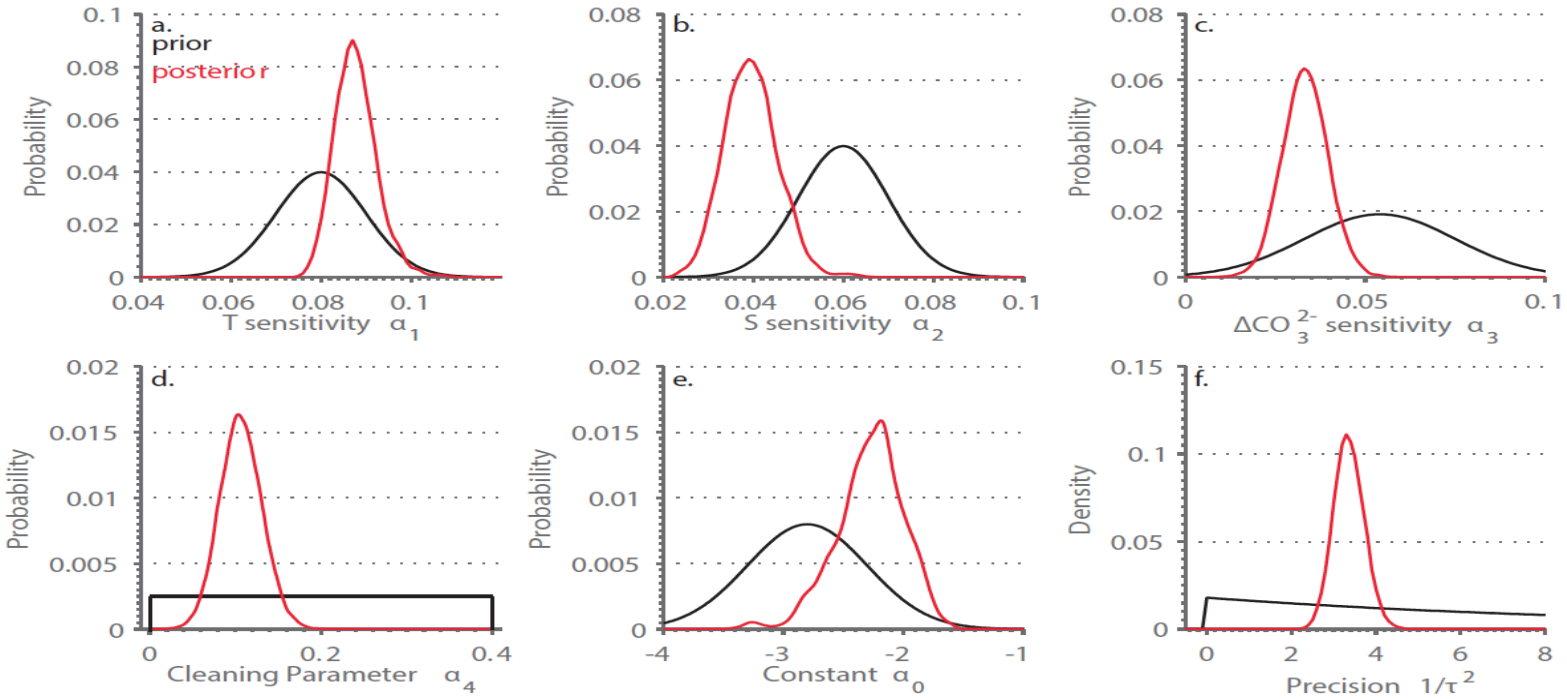
$$\pi\left(\Phi \middle| Mg/Ca, T, S, \Delta CO_3^{2-}, C\right) \propto f\left(Mg/Ca \middle| T, S, \Delta CO_3^{2-}, C, \Phi\right) \cdot \pi(\Phi)$$

Prior distribution on coefficients

- Based on ”culturing experiments” and expert knowledge (Khider).
- *Apriori independent:*
  $\pi(\mathbf{\Phi}) = p(\alpha_0)p(\alpha_1)p(\alpha_2)p(\alpha_3)p(\alpha_4)p(\tau^2)$

$$\alpha_0 \sim N(-2.8, 0.5)$$
$$\alpha_1 \sim N(0.08, 0.01)$$
$$\alpha_2 \sim N(0.06, 0.01)$$
$$\alpha_3 \sim N(0.054, 0.019)$$
$$\alpha_4 \sim U(0, 0.4)$$
$$1/\tau^2 \sim Ga(1.0, 0.1)$$

Posterior and prior distribution for Φ.



- Required a large number of MCMC iterations.
- Implementation in *rjags* and *matjags*.

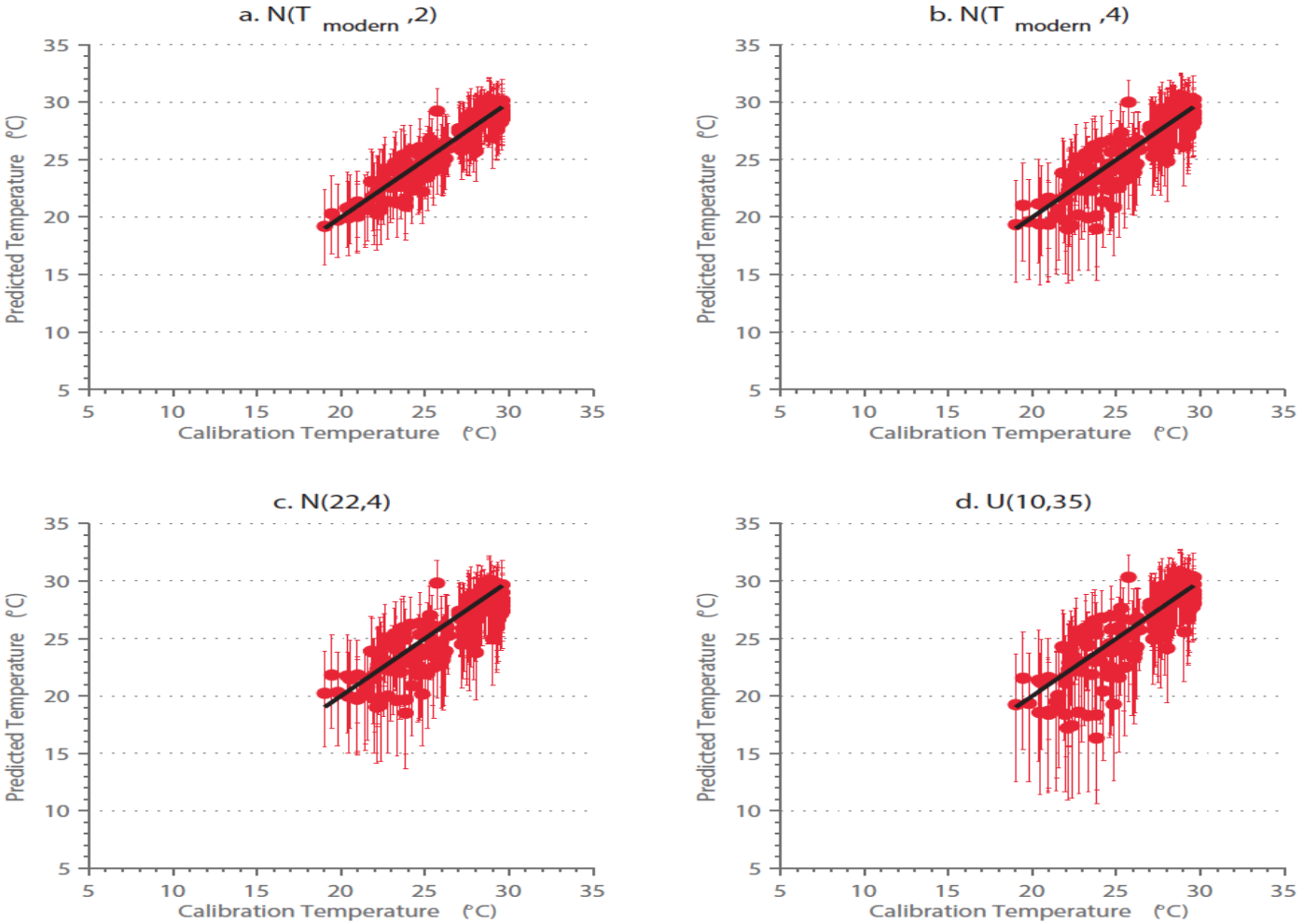
Predicitive framework

- Goal: provide predictions of (say) T.
- Just another application of Bayes theorem and MCMC!

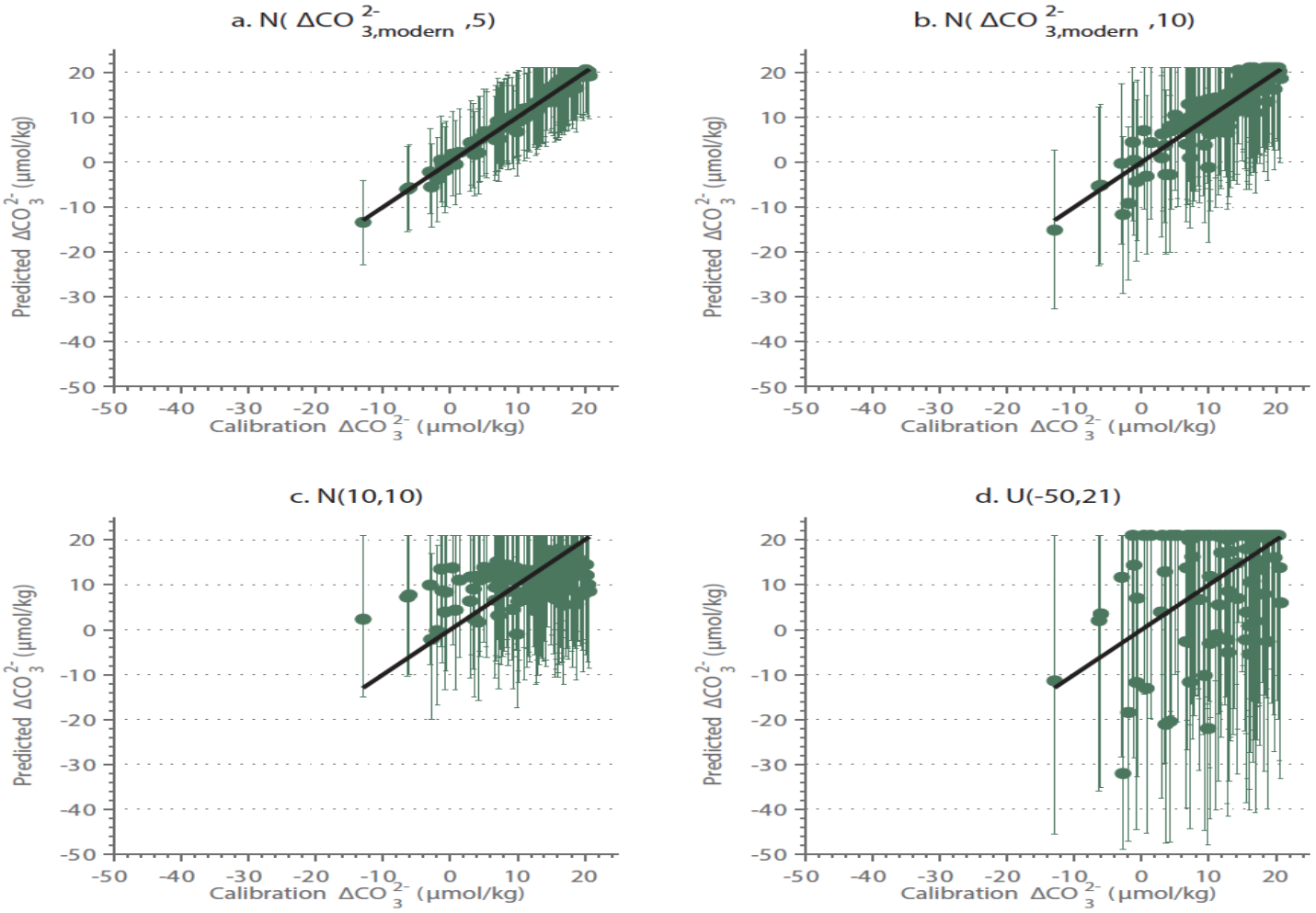
$$\pi\left(T \middle| Mg/Ca, S, \Delta CO_3^{2-}, C, \Phi\right) \propto f\left(Mg/Ca \middle| T, S, \Delta CO_3^{2-}, C, \Phi\right) \cdot \pi\left(T\right)$$

- Requires prior on T.
- Notice conditioning on $\mathbf{\Phi}$.
- Post-processing of posterior samples of $\mathbf{\Phi}$.
- Done also for ΔCO_3^{-2} an S.

Prediction results for T.



Prediction results for ΔCO_3^{-2}



Prediction results for *S*.