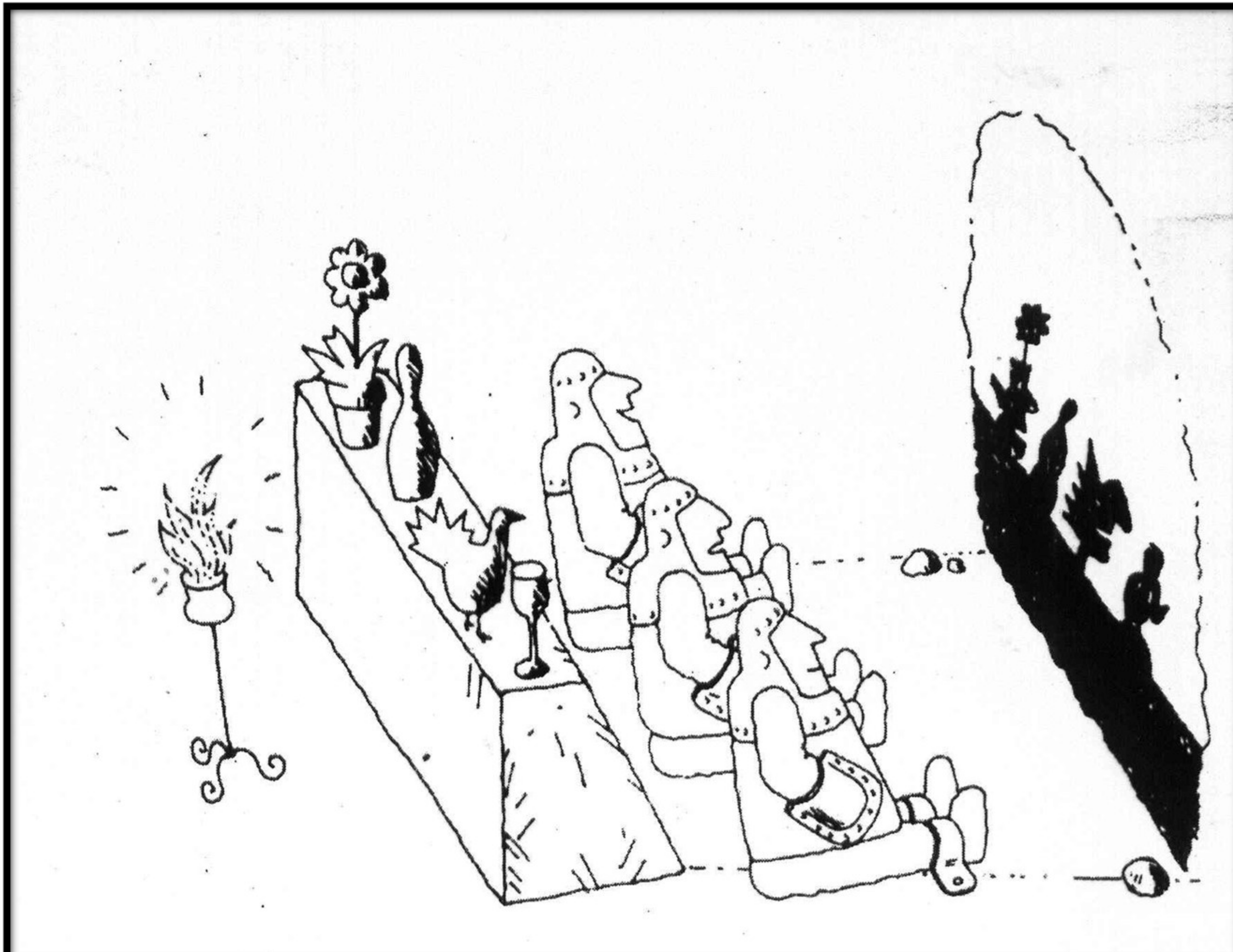
Data analysis and Statistics in paleoclimate

Jessica E. Tierney

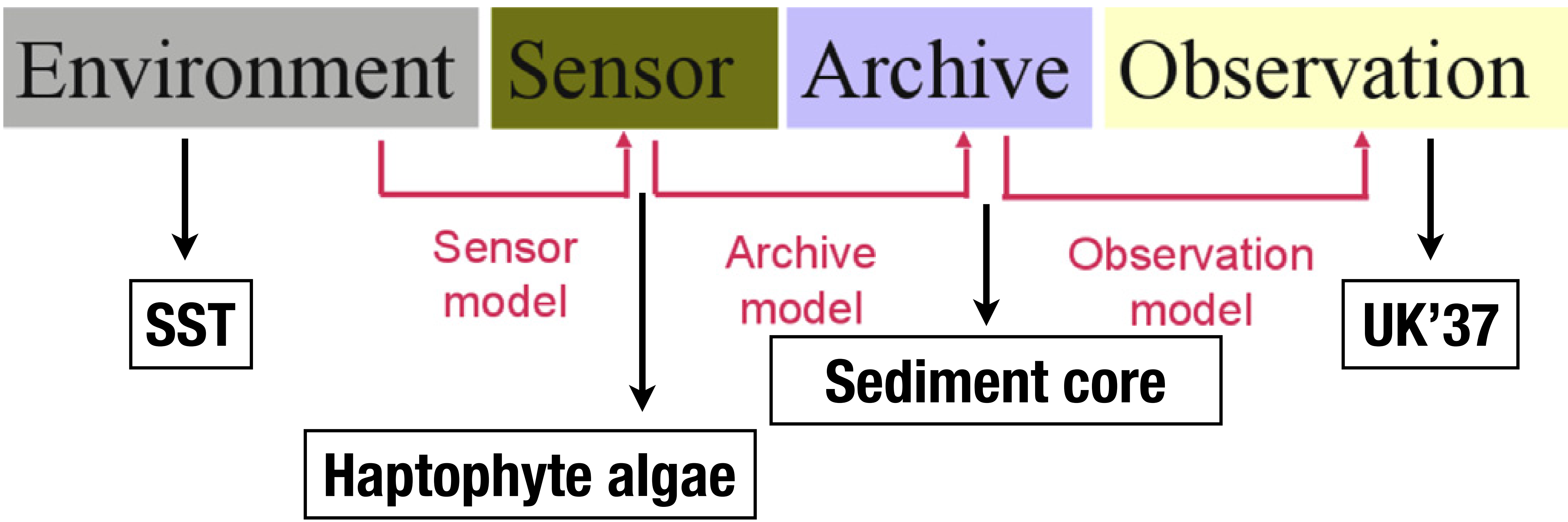


Arc of this lecture

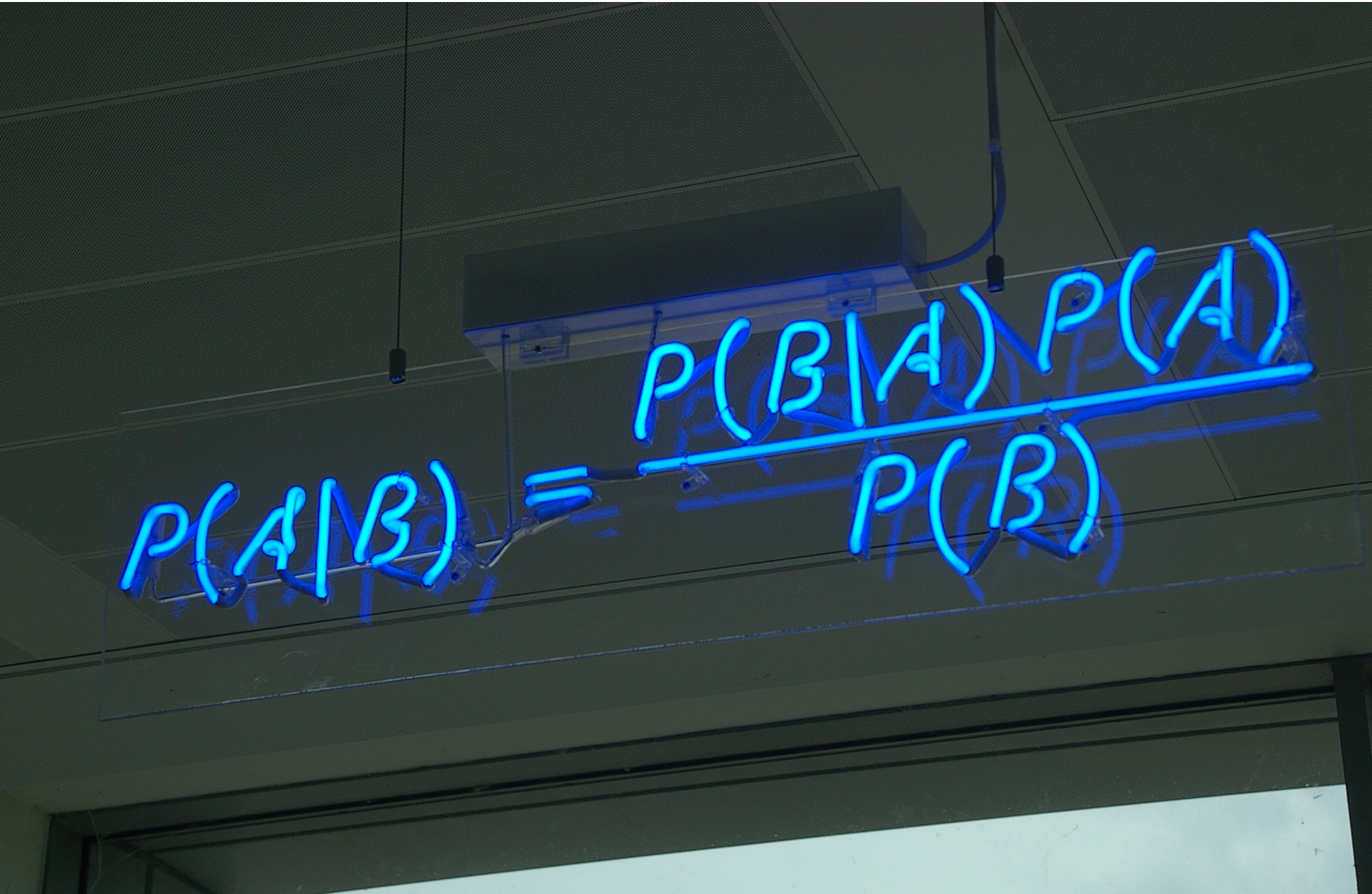
- The “proxy system”: Understanding the nature of the signal.
- Bayesian statistical approaches.
- Reduced space methods.
- Data assimilation.

Understanding the nature of the signal:
"The proxy system"

Evans et al., 2013, QSR



The World of Bayes



Bayesian vs. Frequentist

<u>Frequentist:</u> data are a series of events, with a frequency, and parameters are set. The probability of occurrence is immutable.

<u>Bayesian:</u> data are fixed and parameters must be estimated. Probability is mutable and <u>changes with the addition of information.</u>

Why go Bayesian?

1) Allows one to model the proxy behavior and uncertainties in an etiologically correct way:

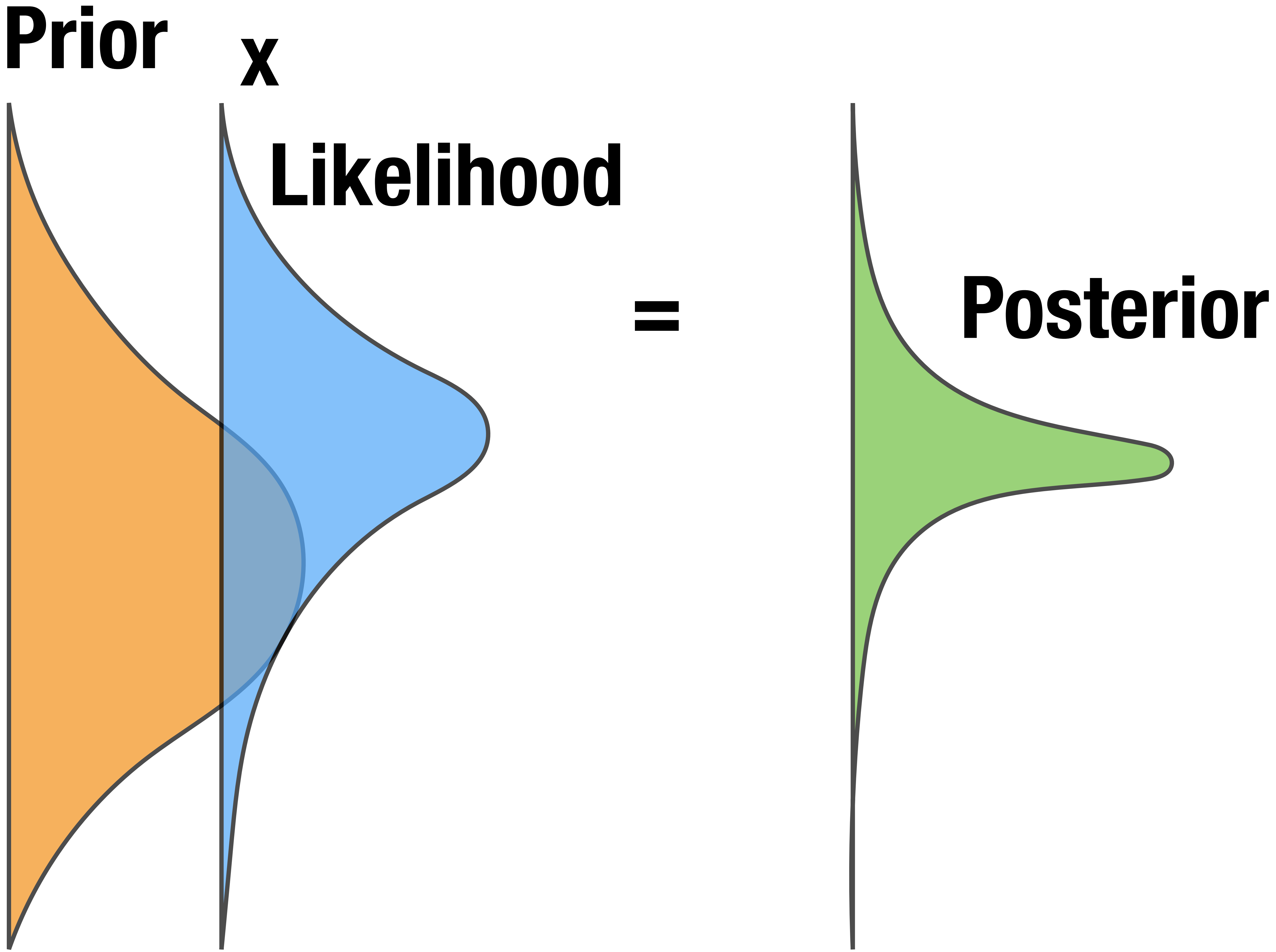
$$TEX_{86} = \alpha_s + \beta_s * SST + \epsilon$$

2) Then we can use Bayes' rule to predict SST from TEX_{86}:

$$P(SST \mid TEX_{86}) \propto P(TEX_{86} \mid SST) * P(SST)$$

3) P(SST) is the prior. It can be relatively uninformative if the model is robust.

4) From this process, we get meaningful uncertainty propagation, plus ensembles of results.


Prior
x
Likelihood
=
Posterior

BAYSPAR and BAYSPLINE

(or any type of Bayesian regression designed for paleoclimate!)

involves two applications of Bayes' Rule.

Step 1: estimate parameters: beta coefficients and tau^2 error term:

$$p(\beta, \tau^2 | \mathrm{proxy}, \mathrm{SST}) \propto p(\mathrm{proxy} | \mathrm{SST}, \beta, \tau^2) \cdot p(\beta, \tau^2).$$

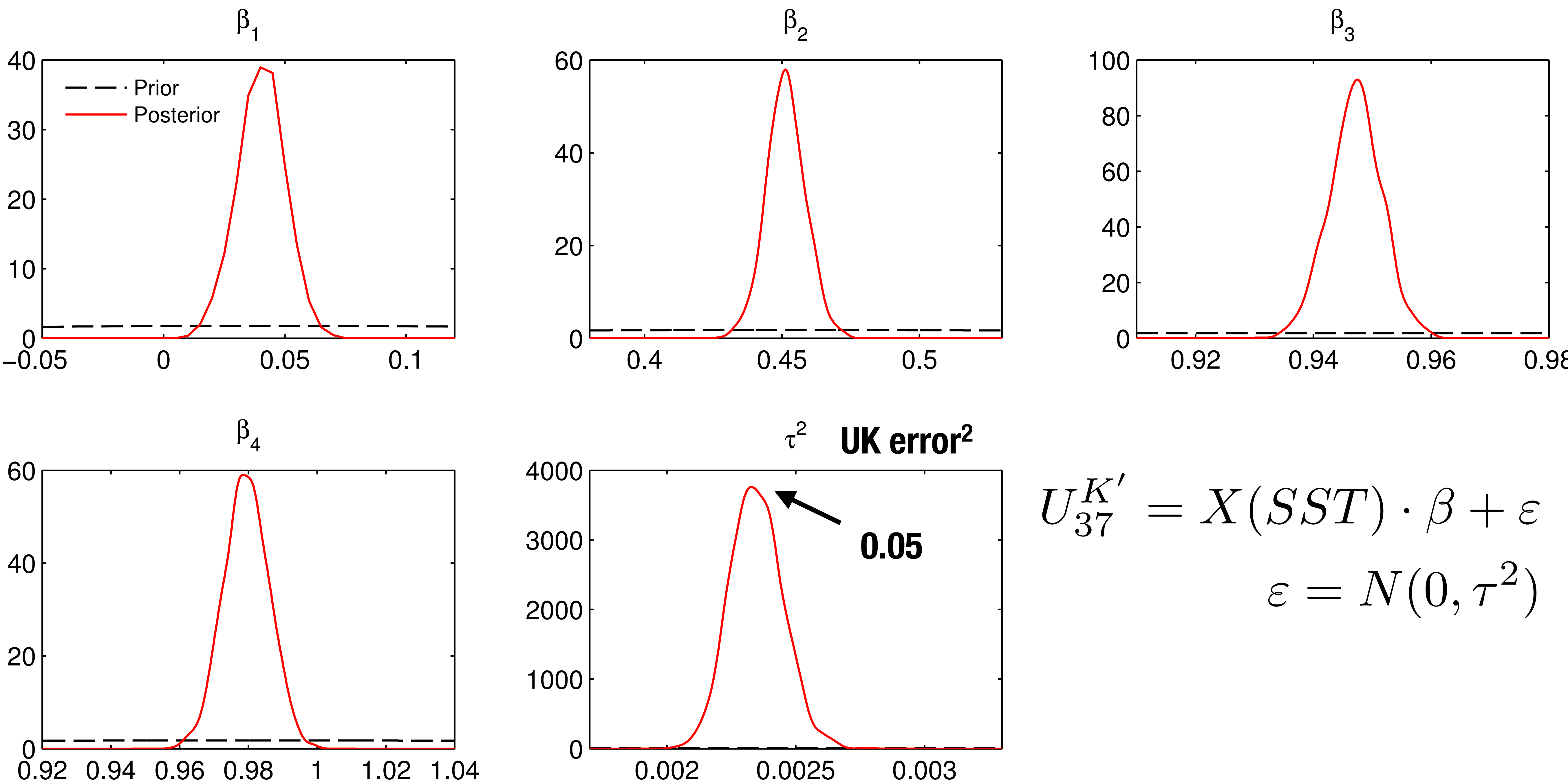
Step 2: estimate SST from new proxy data. Need to invert Step 1:

$$p(\mathrm{SST}_{new} | \mathrm{proxy}_{new}) \propto p(\mathrm{proxy}_{new} | \mathrm{SST}, \beta, \tau^2) \cdot p(\mathrm{SST}_{new}).$$

Matlab code:
bayes_regression_demo.m
predictfromreg_demo.m

Step 1: Parameter estimation

Example from BAYSPLINE. Priors are really flat, which means likelihood dominates.

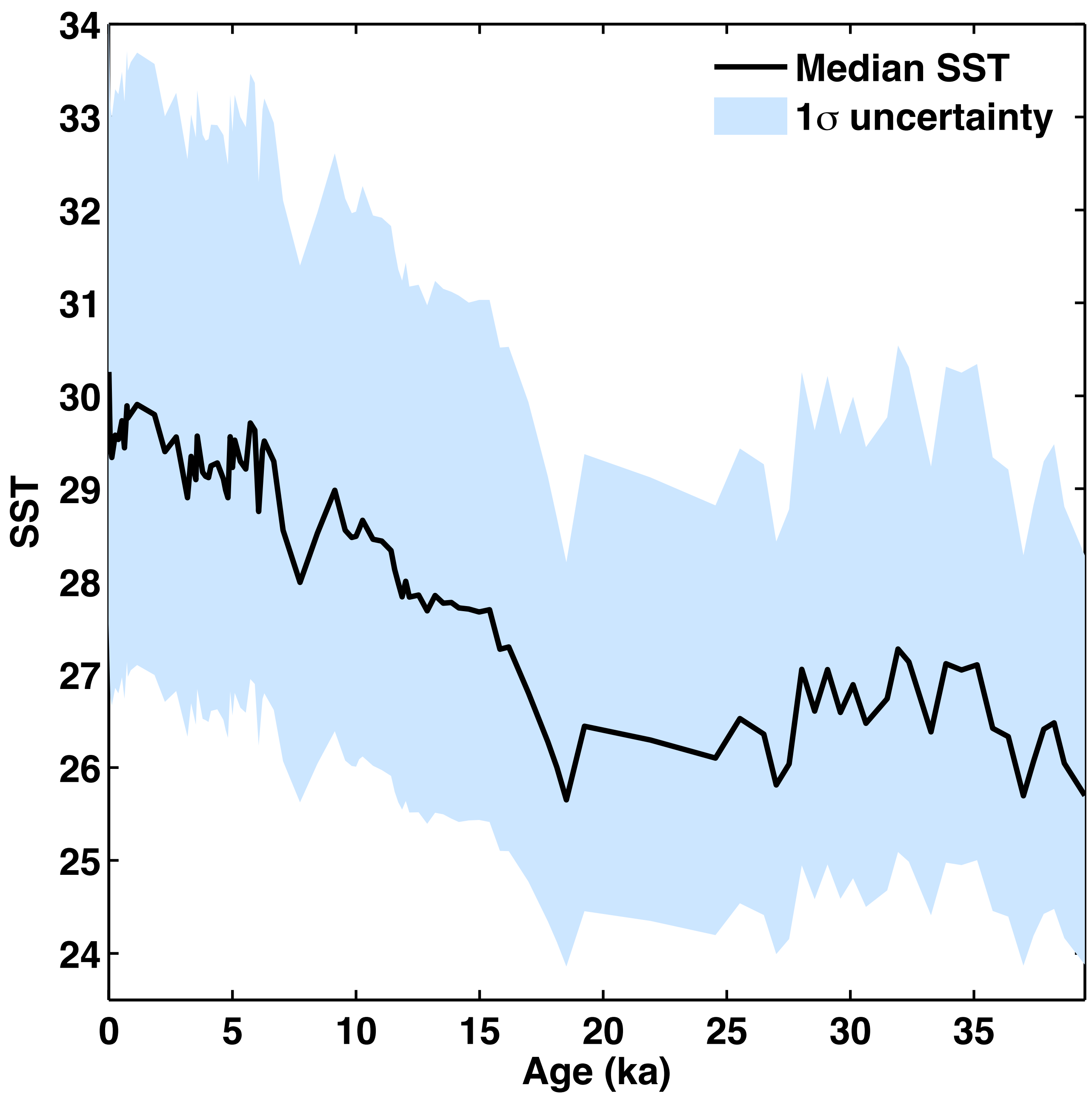


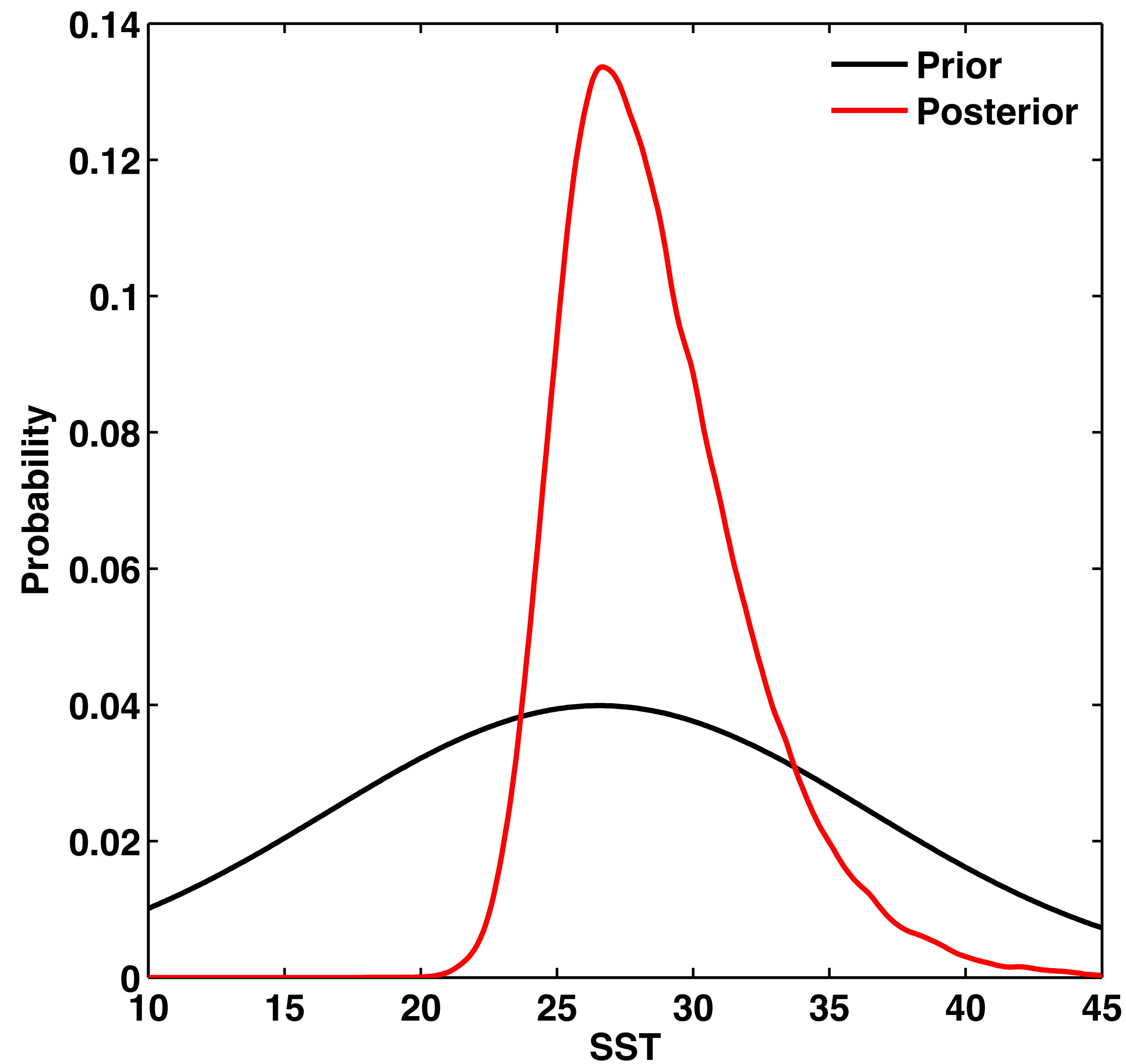
$$U_{37}^{K'} = X(SST) \cdot \beta + \varepsilon$$

$$\varepsilon = N(0, \tau^2)$$

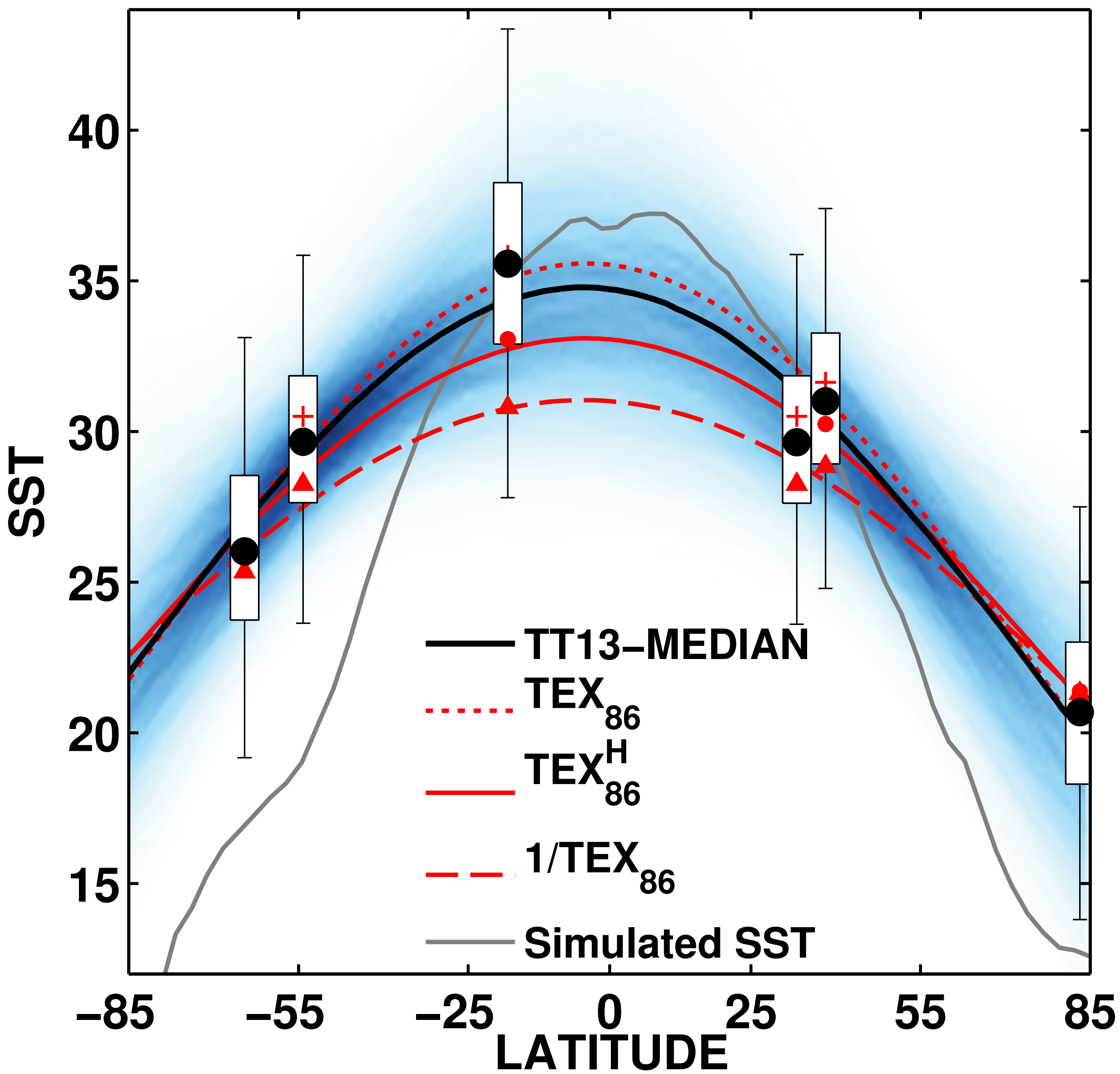
Step 2: Temperature estimation

Example: applying BAYSPLINE to UK37 data from the Gulf of Aden.





Probabilistic Statements



There is a high (78%) probability that the pole-to-equator gradient exceeded 10°C, but it is unlikely (14%) that it exceeded 20°C.

There is a very high (93%) probability that tropical SSTs exceeded 30°C, but it is unlikely (6%) that they exceeded 40°C.

Envisioning a hierarchical model for sedimentary records

Level 1: Spatiotemporal process model (describes SST field)

$$SST_{t+1} - \mu = \alpha \cdot (SST_t - \mu) + \epsilon_t$$

$$\epsilon_t \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma_{i,j} = \sigma^2 exp(-\phi|x_i - x_j|)$$

Level 2: Proxy forward model, can be updated to account for age model uncertainty **(Werner & Tingley, 2015, Clim. Past)**

$$U_{37}^{K'} = X(SST) \cdot \beta + \varepsilon$$

$$\varepsilon = N(0, \tau^2)$$

Level 3: Archive model (bioturbation or other sedimentary features)

$$U_{37obs}^{K'} | U_{37}^{K'} = \sum_{t_n} U_{37t_1+t_n}^{K'} \cdot g(t_n) + \epsilon(t_1),$$

$$\epsilon(t_1) \sim \mathcal{N}(0, \tau^2) \text{ IID.}$$

COMING SOON

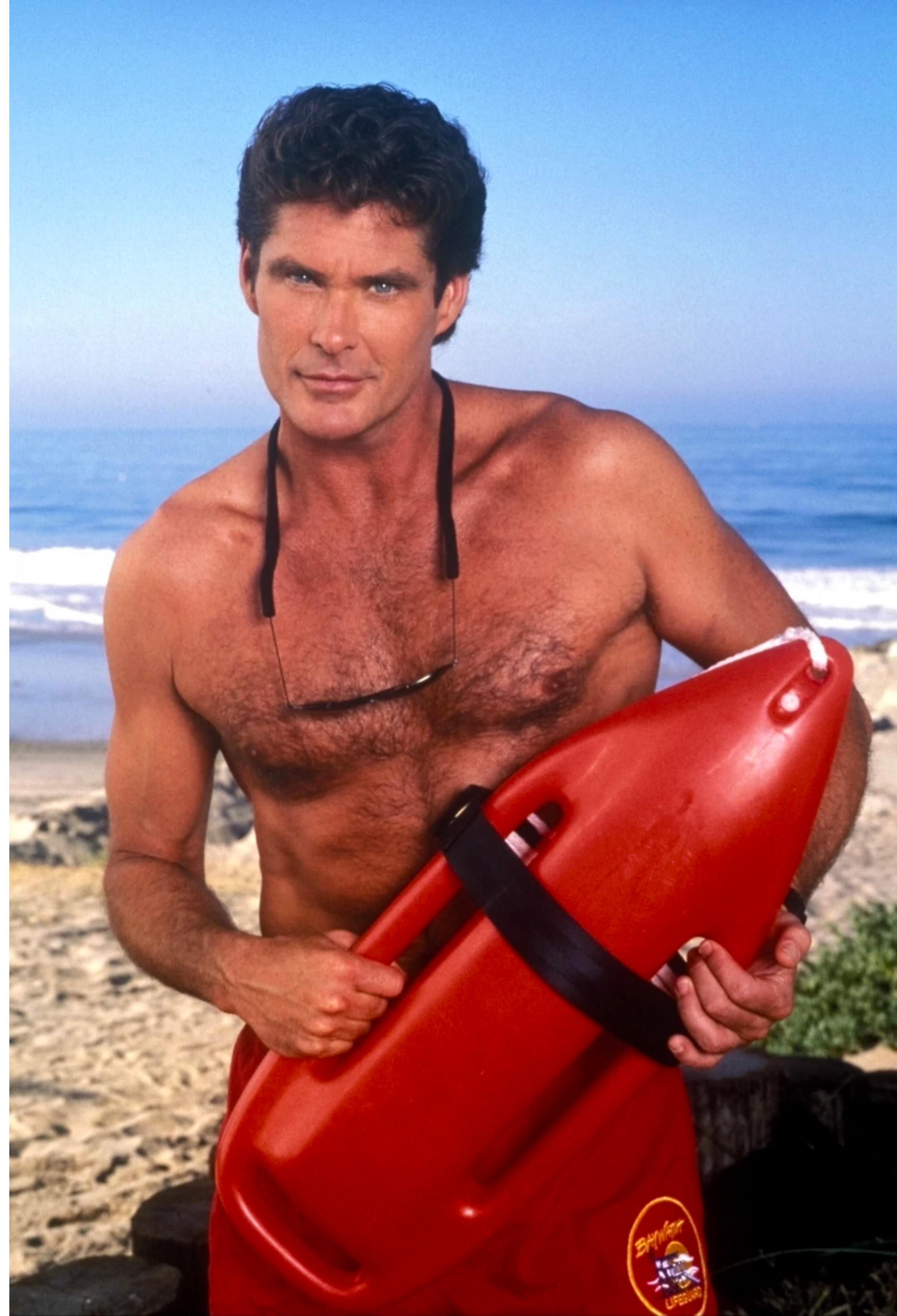
The BaySST package

BAYSPAR

BAYSPLINE

BAYFOX

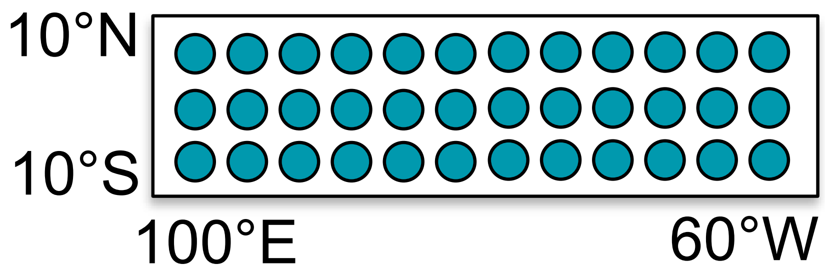
BAYMAG



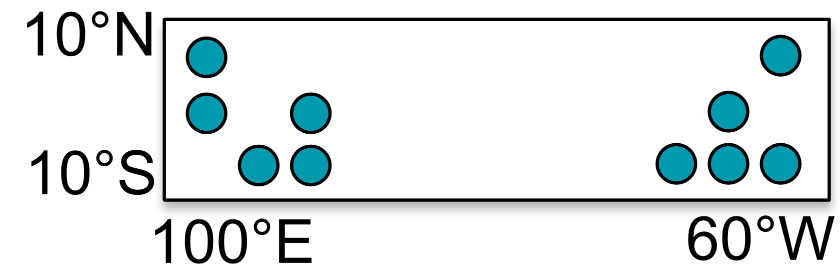
Reduced space approach

After Gill et al., 2016, Paleoceanography

ERSST v5, 1854-2017 Pacific full field



ERSST v5, 1854-2017 limited field



T = N (years) x G (locations)

L = N (years) x P (locations)

SVD

$$T_{[N\times G]} = U_{[N\times N]}\Sigma_{[N\times G]}V_{[G\times G]}$$

$$L_{[N\times P]} = U_{[N\times N]}\Sigma_{[N\times P]}V_{[P\times P]}$$

$$PCs_{[N\times G]} = T_{[N\times G]}V_{[G\times G]} \longleftrightarrow PCs_{[N\times P]} = T_{[N\times P]}V_{[P\times P]}$$

Regression to predict full PCs from limited PCs

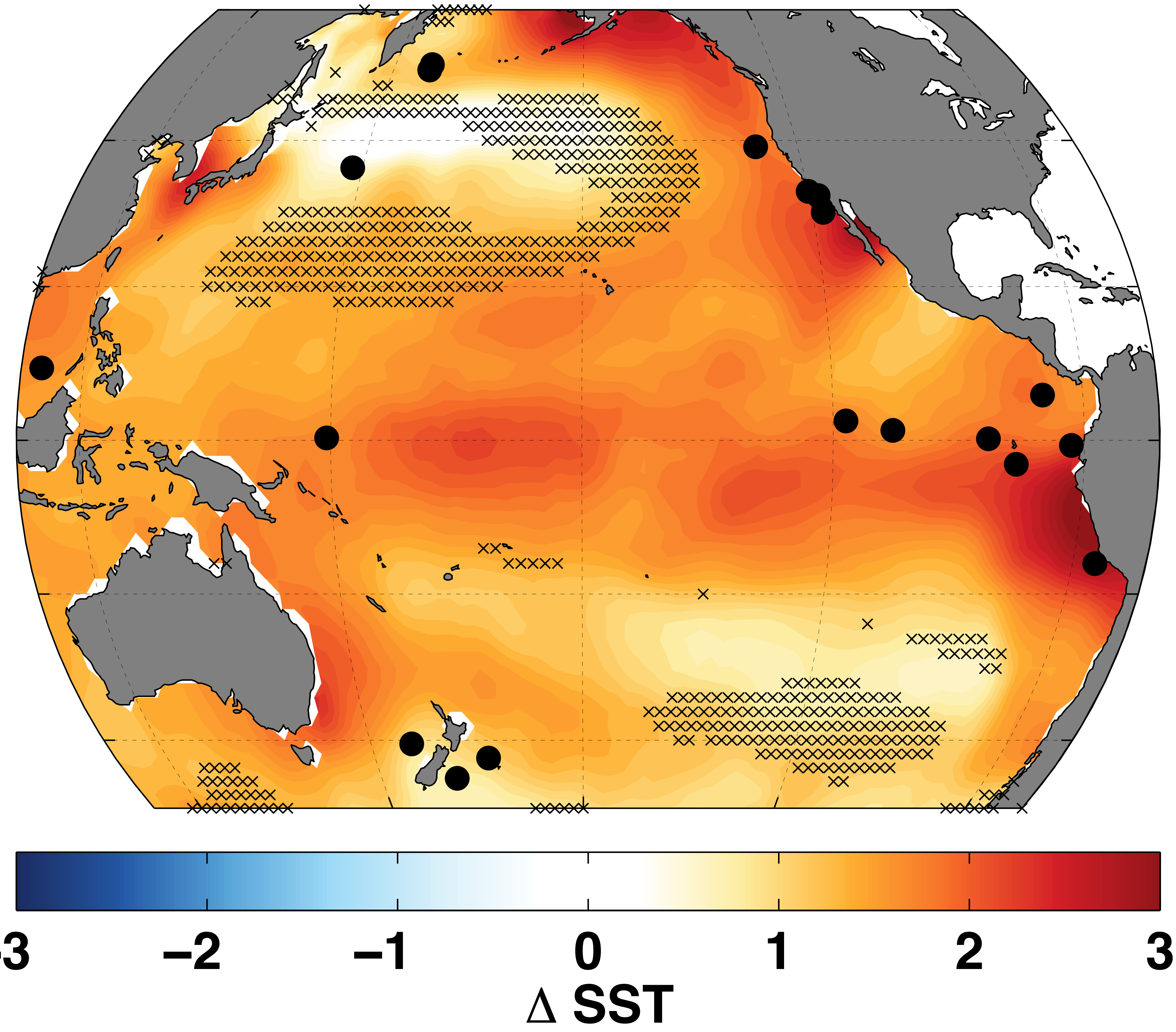
Calculate Pliocene limited field PCs: $PCs_{Plio[1\times P]} = T_{[1\times P]}V_{[P\times P]}$

Use regression to predict first few G PCs, rest of G is zeroes or means:

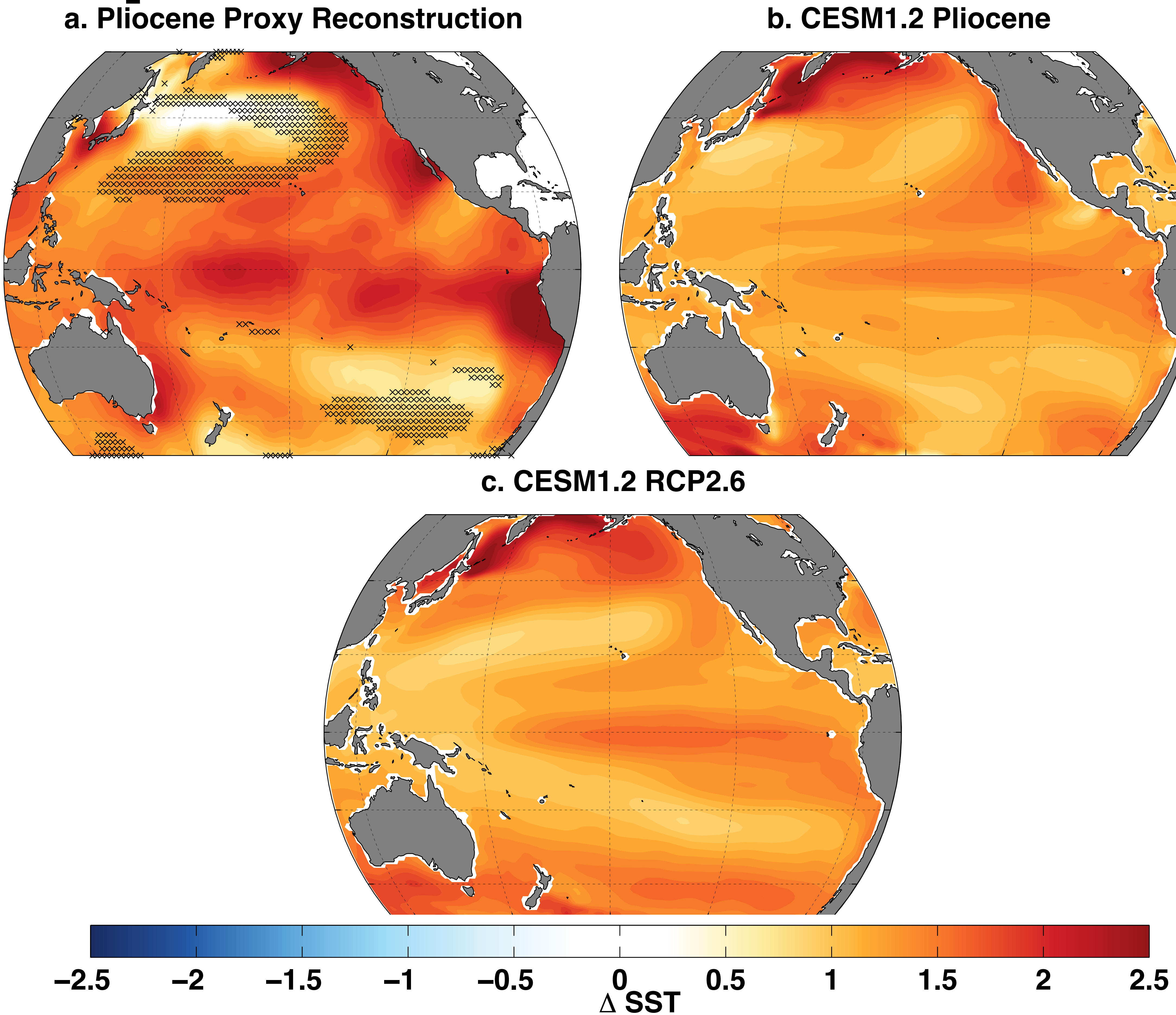
$$PCs_{Plio[1\times G]}$$

Expand back to full field: $T_{Plio[1\times G]} = PCs_{Plio[1\times G]}V^{T}_{[G\times G]}$

mid-Pliocene full field reconstruction



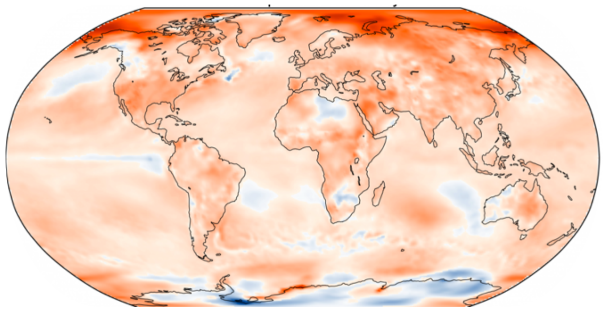
Comparison to model simulations



Paleoclimate data assimilation

- Blends climate model output and proxy data.
- Relies on the model for the relationship between climate variables
- Can recover full field of many different variables
- Uncertainties in proxies and models can be built into the method

Ensemble Kalman Filter Approach



Kalman gain

$$x_a = x_b + K[y - \hat{y}]$$

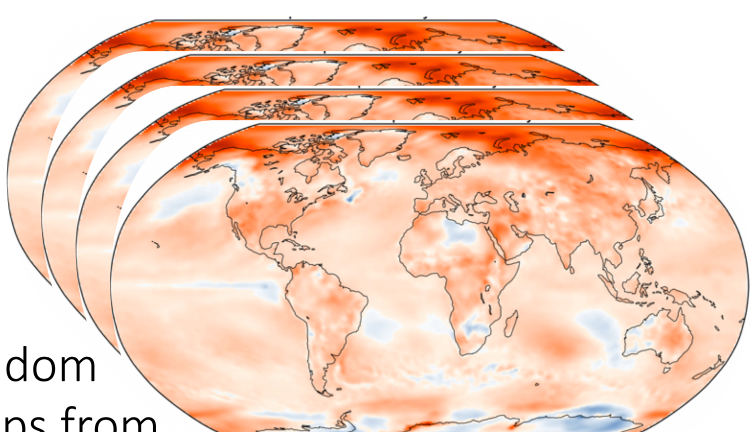
Reconstruction (posterior)

Model prior

Proxy observations

Model estimate of proxy observations

$$U_{37}^{K'} = X(SST) \cdot \beta + \varepsilon$$



200 random timesteps from the 1000 year model run

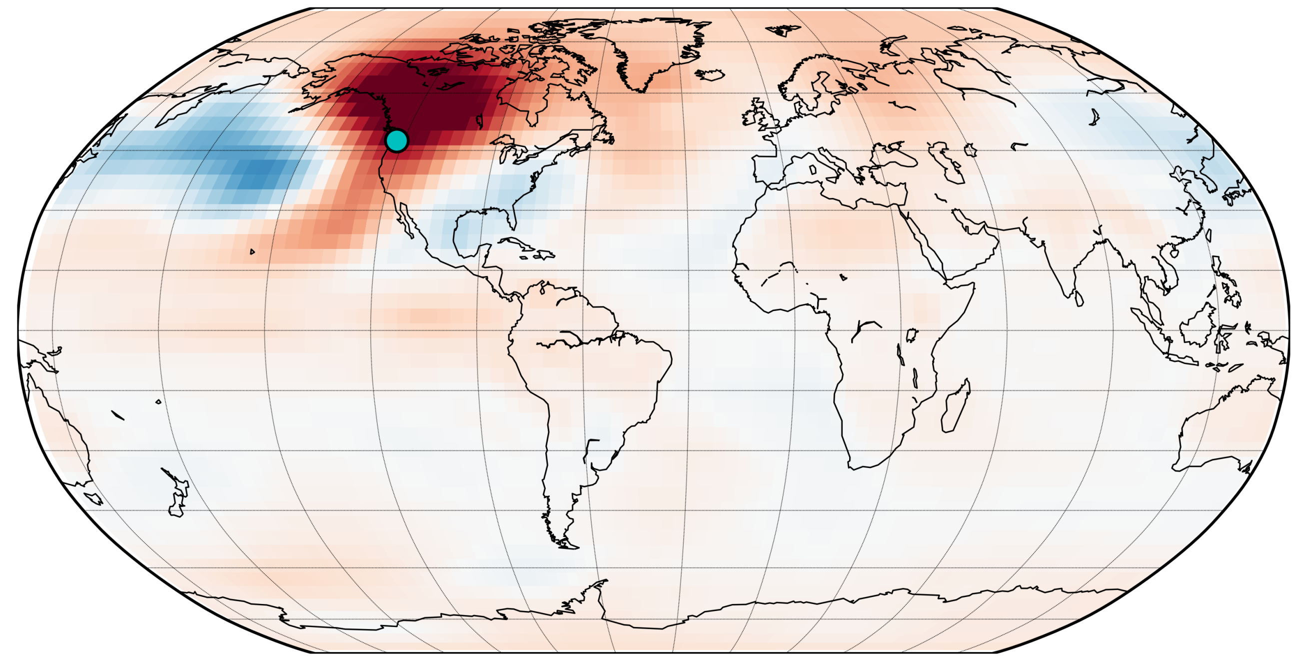
$$U_{37}^{K'}, Mg/Ca, \delta^{18}O$$

Kalman Gain

Covariance b/t the fields in the prior ensemble and forward proxy estimates

$$K = \frac{cov(x_b, \hat{y})}{var(\hat{y}) + R}$$

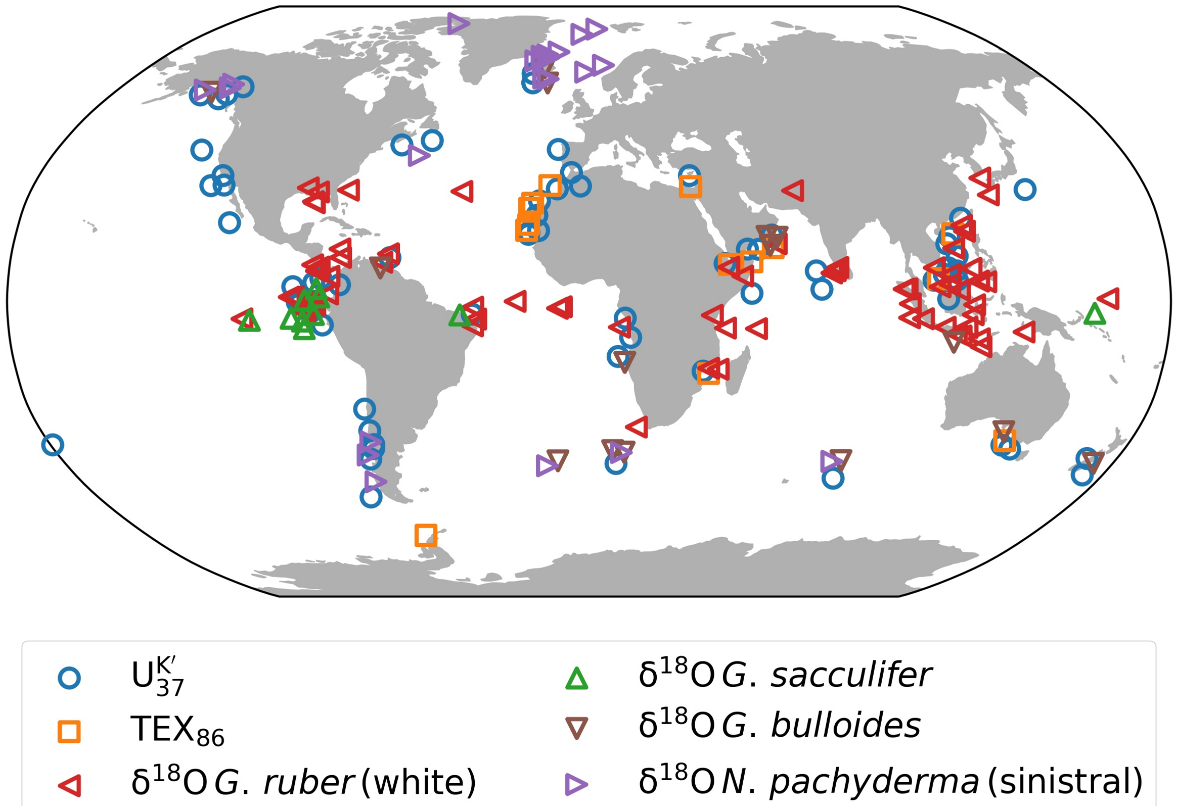
Covariance (temperature, $\delta^{18}O$)

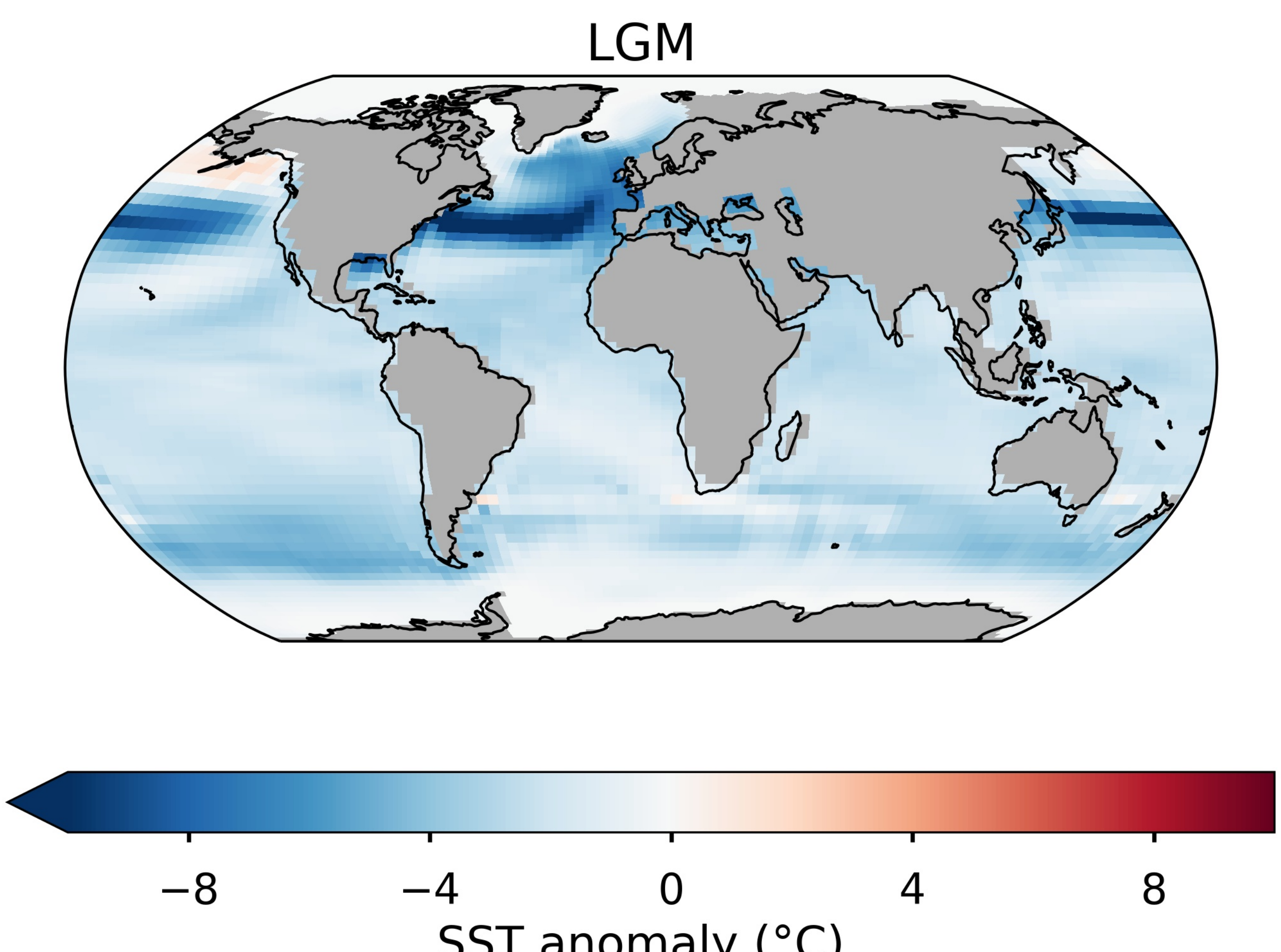


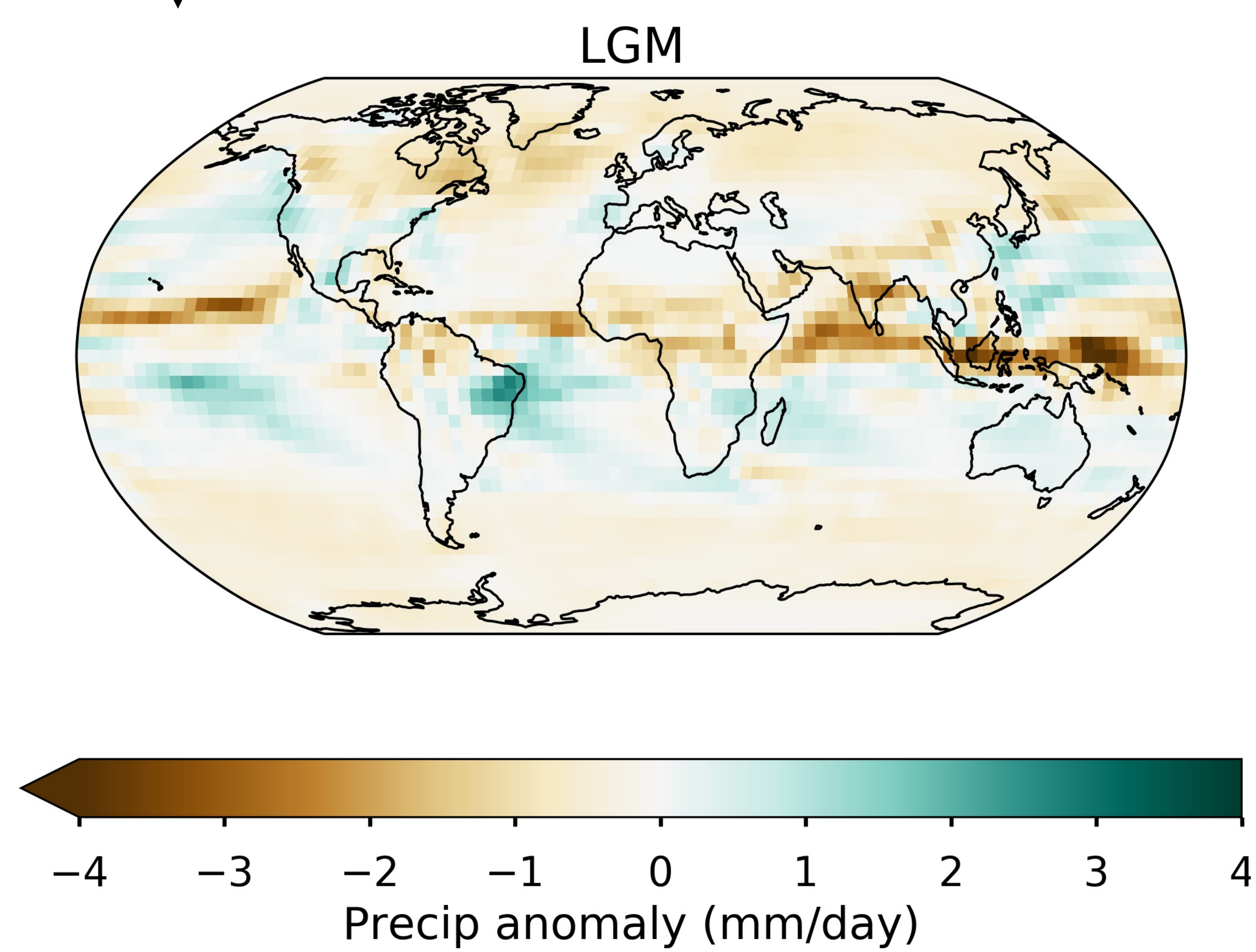
Uncertainty in the model prior (in proxy space)

Observed proxy uncertainty (also a variance term)

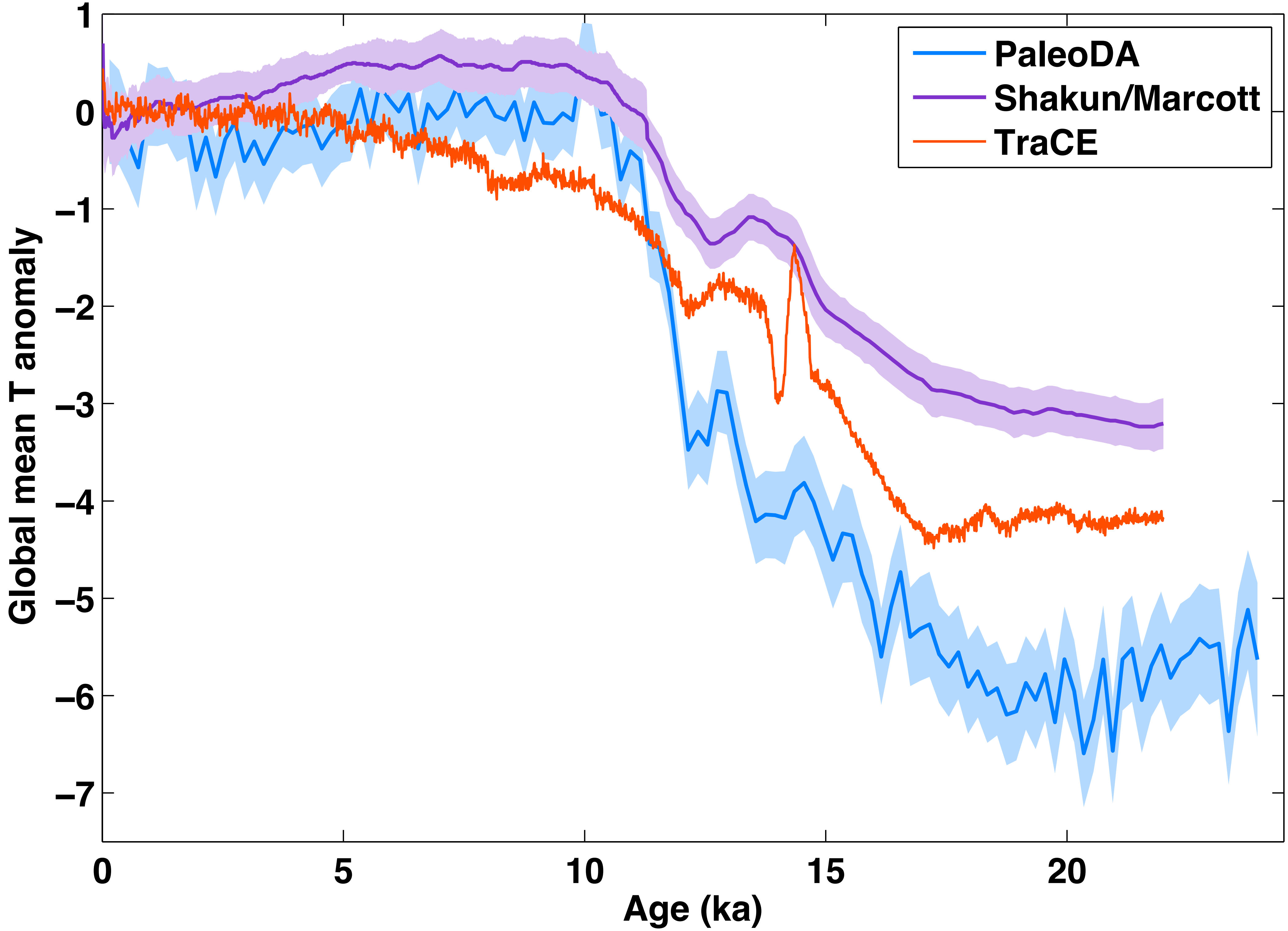
Example: SST proxies + TraCE



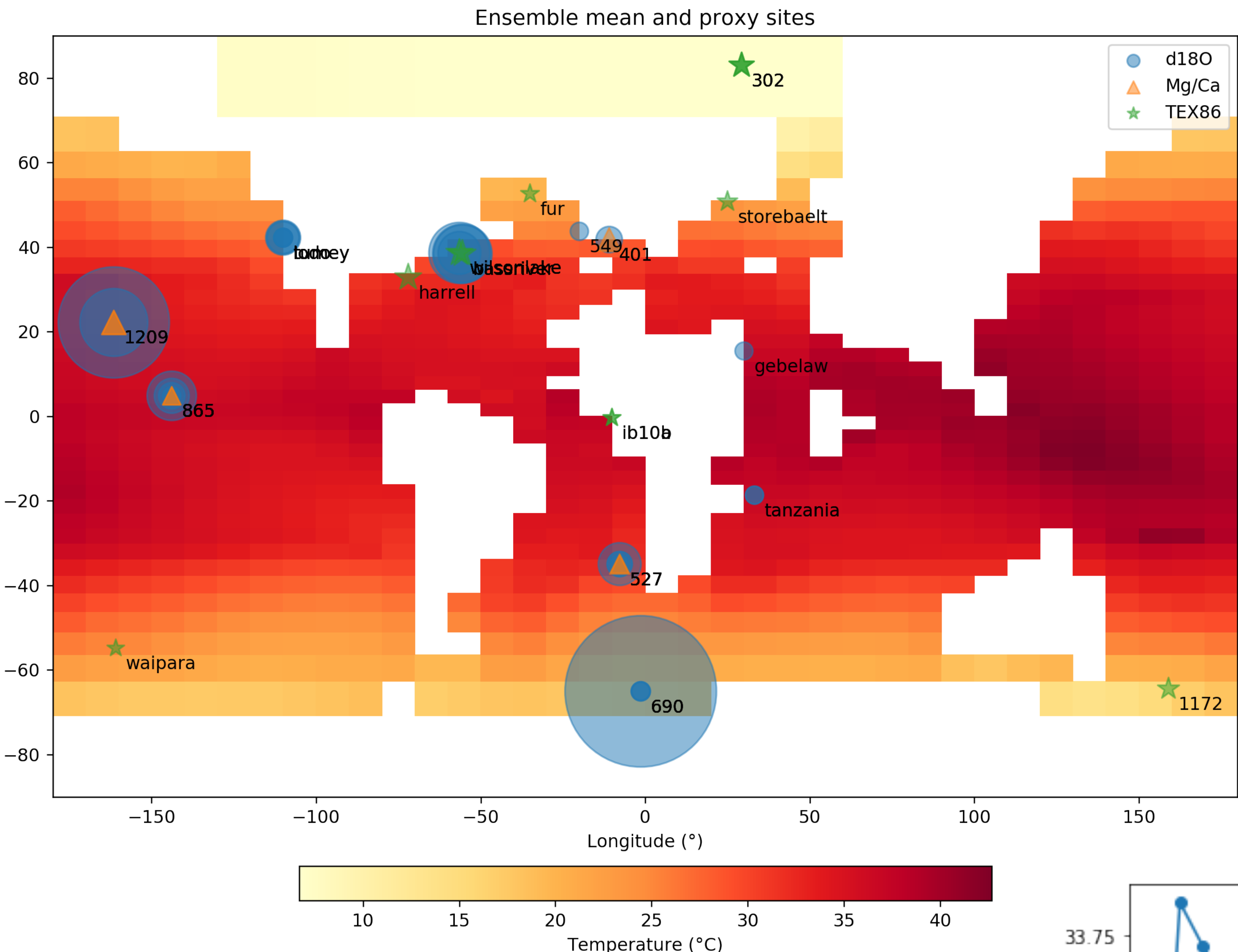




Compare with model- and data-only estimates



PETM example



Serial assimilation of $\delta^{18}O$ data