What drives the glacial-interglacial cycle? A Bayesian approach

Jake Carson Michel Crucifix Simon Preston
Richard Wilkinson

Universities of Warwick, Louvain, Nottingham and Sheffield

RSS 2015

Outline

Most simple models of the [...] glacial cycles have at least four degrees of freedom [parameters], and some have as many as twelve. Unsurprisingly [...this is] insufficient to distinguish between the skill of the various models (Roe and Allen 1999)

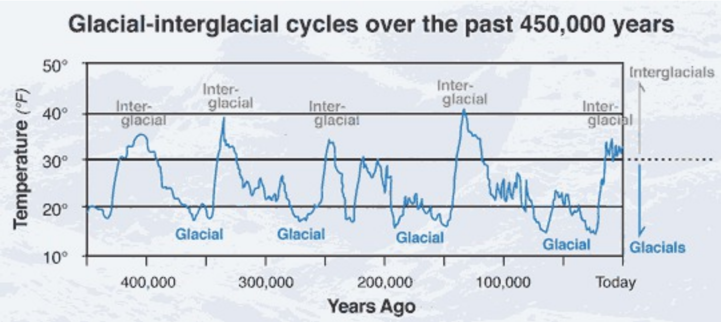
Outline

> *Most simple models of the [...] glacial cycles have at least four degrees of freedom [parameters], and some have as many as twelve. Unsurprisingly [...this is] insufficient to distinguish between the skill of the various models (Roe and Allen 1999)*

Our focus is on statistical computation:

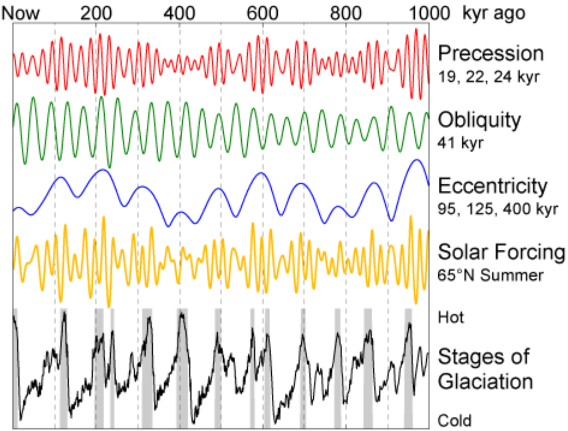
- Can we combine a 'simulator', proxy model, and data in a Bayesian analysis?
  - Data assimilation followed by parameter estimation and model selection
- What can we learn?
- Can we cut feedbacks between climate and age, fitting each component independently?
- Can we ignore dating uncertainties?
- Can we substitute the Kalman filter for the particle filter?

Glacial-Interglacial cycle



Cycle characterised by saw-toothed behaviour: slow accumulation and rapid terminations.
Approx 100 kyr period between cycles, but previously a 40 kyr period was observed.

Milankovitch theory



Eccentricity: orbital departure from a circle, controls duration of the seasons
Obliquity: axial tilt, controls amplitude of seasonal cycle
Precession: variation in Earth's axis of rotation, affects difference between seasons

Insolation at 65° north: combination of these three terms, considered important.

100kyr problem

Spectral analysis suggest the climate response has a period of $\approx$ 100kyr, but the orbital forcing at this period is small.

Eccentricity has 95 and 125kyr periods, but accounts for only 2% of the variation compared to the shifts caused by obliquity (41kyr period) and precession (21kyr period).

100kyr problem

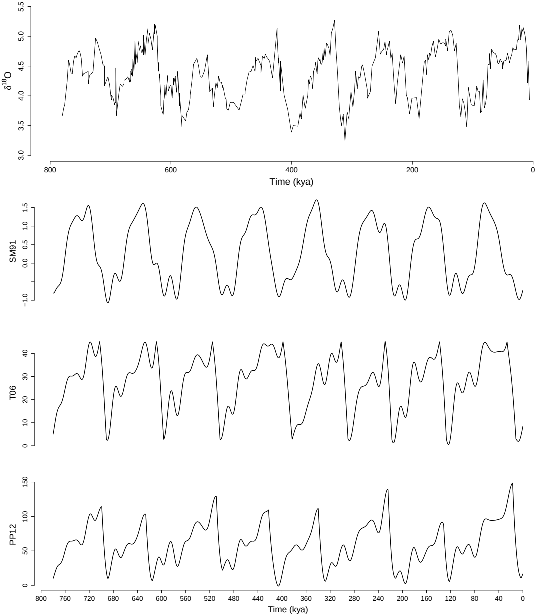
Spectral analysis suggest the climate response has a period of ≈ 100kyr, but the orbital forcing at this period is small.

Eccentricity has 95 and 125kyr periods, but accounts for only 2% of the variation compared to the shifts caused by obliquity (41kyr period) and precession (21kyr period).

Explanatory hypotheses

- Earth's climate may have a natural frequency of 100kyr caused by natural feedback processes
- 100kyr eccentricity cycle acts as a "pacemaker" to the system, amplifying the effect of precession and obliquity at key moments, triggering a termination.
- 21kyr precession cycles are solely responsible, with ice building up over several precession cycles, only melting after four or five such cycles.

Which is the better model?



Models

Use conceptual models based on a few hypothesised relationships that capture some aspect of the climate system, driven by some aspect of the solar forcing

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = g(X_t, \theta) + F(t, \gamma)$$

where $\gamma = (\gamma_P, \gamma_C, \gamma_E)$ controls the combination of precession, obliquity and eccentricity.

$X_t \in \mathbb{R}^p$ denotes the state of the climate at time t.

Models

Use conceptual models based on a few hypothesised relationships that capture some aspect of the climate system, driven by some aspect of the solar forcing

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = g(X_t, \theta) + F(t, \gamma)$$

where $\gamma = (\gamma_P, \gamma_C, \gamma_E)$ controls the combination of precession, obliquity and eccentricity.
$X_t \in \mathbb{R}^p$ denotes the state of the climate at time t.

Embed these simulators within a state space model relating climate to observations

$$\mathrm{d}X_t = g(X_t, \theta)\mathrm{d}t + F(t, \gamma)\mathrm{d}t + \Sigma \mathrm{dW}$$
$$Y_t = d + sX_{1,t} + \epsilon_t$$

The models have 10–15 parameters that need to be estimated

The statistical quantities we would like to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T} | y_{1:T}, \theta, \mathcal{M})$$

where $x_{1:T} = (x_1, \ldots, x_T)$

The statistical quantities we would like to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T} | y_{1:T}, \theta, \mathcal{M})$$

  where $x_{1:T} = (x_1, \ldots, x_T)$
- Model calibration (marginal parameter posterior)

$$\pi(\theta | y_{1:T}, \mathcal{M})$$

The statistical quantities we would like to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T}|y_{1:T}, \theta, \mathcal{M})$$

  where $x_{1:T} = (x_1, \ldots, x_T)$

- Model calibration (marginal parameter posterior)

$$\pi(\theta|y_{1:T}, \mathcal{M})$$

- Model selection (model evidence/Bayes factors)

$$\pi(y_{1:T}|\mathcal{M})$$

The statistical quantities we would like to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T}|y_{1:T}, \theta, \mathcal{M})$$

  where $x_{1:T} = (x_1, \ldots, x_T)$
- Model calibration (marginal parameter posterior)

$$\pi(\theta|y_{1:T}, \mathcal{M})$$

- Model selection (model evidence/Bayes factors)

$$\pi(y_{1:T}|\mathcal{M})$$

These are progressively more difficult to calculate. Moreover,

$$\pi(X_{t+1}|X_t, \theta, \mathcal{M})$$

is unknown ruling out many Monte Carlo approaches.

Bayes factors

Consider comparing two models, $\mathcal{M}_1$ and $\mathcal{M}_2$. Bayes factors (BF) are the Bayesian approach to model selection.

The log Bayes factor is the difference between the log-likelihoods for two models

$$\log(BF) = \log \pi(y_{1:T} \mid \mathcal{M}_1) - \log \pi(y_{1:T} \mid \mathcal{M}_2)$$

$$\text{posterior odds} = \text{prior odds} \times \text{Bayes factor}$$

$\log(BF)$ range	$\mathbb{P}(\mathcal{M}_1 \mid D)$	Interpretation
> 5	0.99 - 1	V. strong evidence in favour of model $\mathcal{M}_1$
3 to 5	0.95 - 0.99	Strong evidence in favour of model $\mathcal{M}_1$
-3 to 3	0.05 - 0.95	Grey zone
-3 to -5	0.01 - 0.05	Strong evidence in favour of $\mathcal{M}_2$
< -5	0 - 0.01	V. strong evidence in favour of model $\mathcal{M}_2$

Computational details

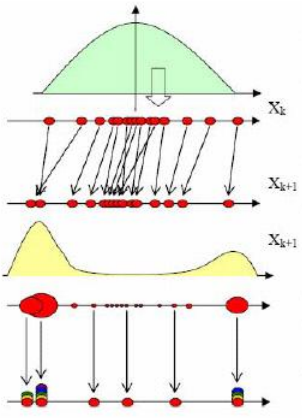
Sequential Monte Carlo (SMC) methods are the natural approach for finding the filtering distributions $\pi(x_{1:T}|y_{1:T}, \theta)$

- Represent all distributions by collection of weighted particles $\{x^{(i)}, w^{(i)}\}$, e.g.,

$$p(x) \approx \sum w_0^{(i)} \delta_{x^{(i)}}(x)$$

- Sequentially build up approximation to $\pi(x_{1:t}|y_{1:t}, \theta)$ one step at a time.

We use Golightly and Wilkinson (2008) style Brownian bridge proposals to guide the particles and reduce degeneracy.



From Pierre Del Moral's webpage

Parameter estimation

SMC provides an unbiased estimate of the marginal likelihood

$$\pi(y_{1:T}|\theta) = \pi(y_1|\theta) \prod_{t=2}^{T} \pi(y_t|y_{1:t-1}, \theta)$$

when we substitute the estimate

$$\tilde{\pi}(y_t|y_{1:t-1}, \theta) = \frac{1}{M} \sum w_t^n$$

for $\pi(y_t|y_{1:t-1}, \theta)$.
We can then use these estimates in a pseudo marginal scheme such as PMCMC (Andrieu *et al.* 2010) to estimate

$$\pi(\theta, x_{1:T}|y_{1:T}) \quad \text{and} \quad \pi(\theta|y_{1:T})$$

SMC^2

We've found that SMC^2 (Chopin *et al.* 2011) works well for our problem

Basic idea:

- Introduce M parameter particles $\theta_1, \ldots, \theta_M$
- For $t = 1, \ldots, T$
  - For each θ_i run a particle filter targeting $\pi(X_{1:t}|y_{1:t}, \theta_i)$
  - Recalculate all the importance weights and resample if necessary

Note that to avoid particle degeneracy, it is still usually necessary to run a PMCMC sampler targeting $\pi(\theta, X_{1:t}|y_{1:t})$ at each resampling step.

This takes 3-4 days on a standard server, or 4-6 hours on a GPU (2500 processors).

SMC^2

We've found that SMC^2 (Chopin *et al.* 2011) works well for our problem

Basic idea:

- Introduce M parameter particles $\theta_1, \ldots, \theta_M$
- For $t = 1, \ldots, T$
  - For each θ_i run a particle filter targeting $\pi(X_{1:t}|y_{1:t}, \theta_i)$
  - Recalculate all the importance weights and resample if necessary

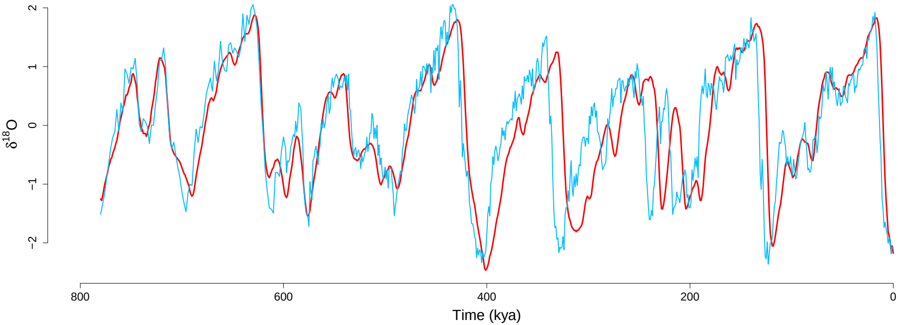
Note that to avoid particle degeneracy, it is still usually necessary to run a PMCMC sampler targeting $\pi(\theta, X_{1:t}|y_{1:t})$ at each resampling step.

This takes 3-4 days on a standard server, or 4-6 hours on a GPU (2500 processors).

- The results here used $\sim 10^8$ simulator runs!

We tried using the Kalman filter instead of the PF, but the results were unreliable

Results: ODP677



We use the ODP677 stack (a composite record from multiple cores), which has been dated by two authors:

- Lisiecki and Raymo (2005) used orbital tuning
- Huybers 2007 used a depth-derived age model (no orbital tuning)

Do we get the same results using the two dating estimates?

Results: ODP677

Model		Dataset	
		H07(unforced)	LR04(forced)
SM91	Forced	-8.4	
	Unforced	-3.9	
T06	Forced	-6.2	
	Unforced	0	
PP12	Forced	-13.9	

- Using Huybers' non-orbitally tuned data, we find evidence in favour of the unforced T06 model.

Results: ODP677

Model		Dataset	
		H07(unforced)	LR04(forced)
SM91	Forced	−8.4	−14.3
	Unforced	−3.9	−37.0
T06	Forced	−6.2	−10.6
	Unforced	0	−29.4
PP12	Forced	−13.9	0

- Using Huybers' non-orbitally tuned data, we find evidence in favour of the unforced T06 model.
- Using Lisiecki's orbitally tuned data, we find strong evidence for forced models (PP12)

The dating method applied changes the answer
Orbitally tuned data leads us to strongly prefer the orbitally tuned version of each model (and vice versa)

The age model used to date the stack (often taken as a given) has a strong effect on model selection conclusions

Age model

Can we also quantify chronological uncertainty?

Age model

Can we also quantify chronological uncertainty?

Target

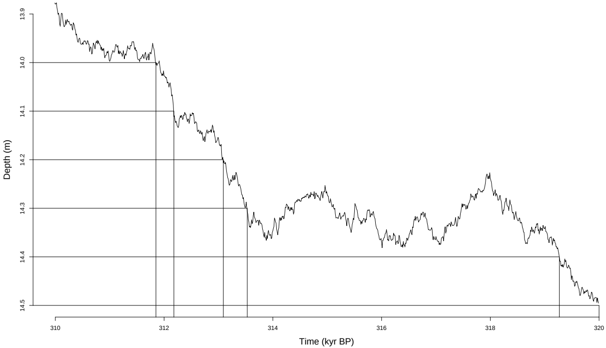
$$\pi(\theta, T_{1:N}, X_{1:N} | y_{1:N})$$

where $T_{1:N}$ are the times of the observation $Y_{1:N}$, which were previously taken as given.

Propose a simple age model for sediment accumulation:

$$\mathrm{d}S = -\mu_s \mathrm{d}T + \sigma_s \mathrm{d}W_s \qquad S(T = 0) = 0$$

- Sample the core at depths H_m, $m = 1 \ldots M$.
- When the core is sampled at depth H_m, the observation relates to the climate at the most recent time at which $S = -H_m$.

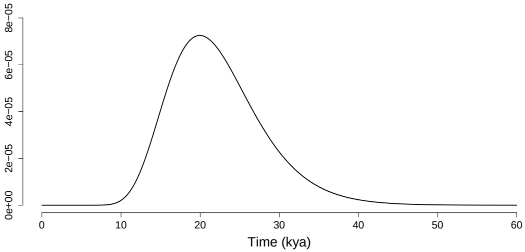


Sediment Model

- Under a time reversal this becomes a first passage time problem:

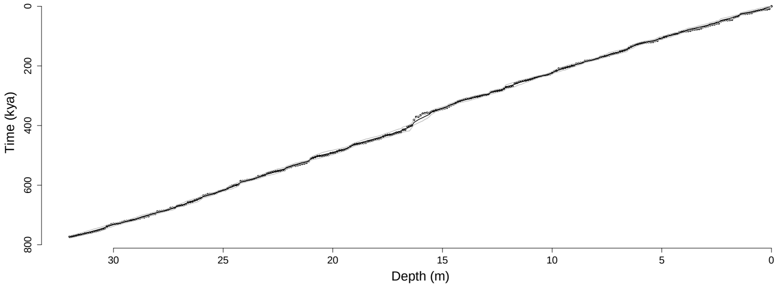
$$T_{m-1} \mid T_m \sim T_m - IG\left(\frac{H_{m-1} - H_m}{\mu_s}, \frac{(H_{m-1} - H_m)^2}{\sigma_2^2}\right), \quad T(H = 0) = 0.$$

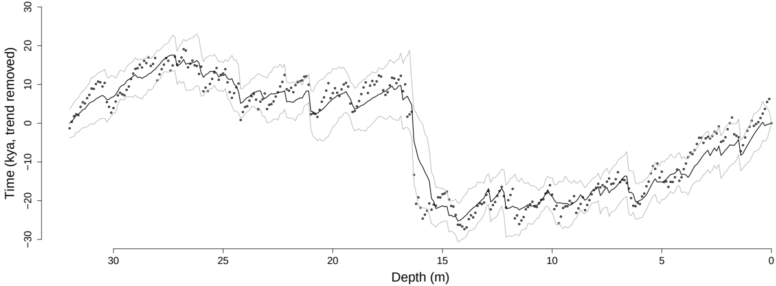
- $T_m \mid T_{m-1}$ is given by Bayes theorem.
- We account for compaction by using *uncompacted* depth, $\hat{H}_m$ (Huybers 2007), giving 2 more parameters.



Simulation Study

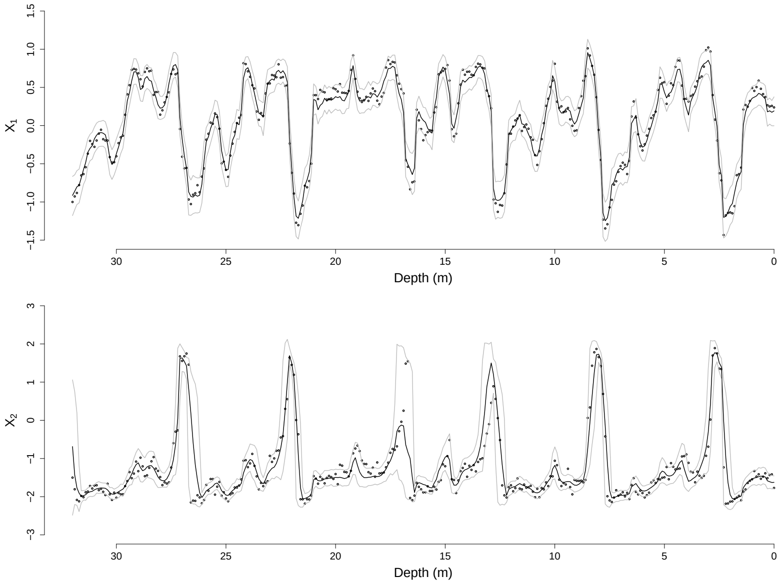
Dots = truth, black line = estimate, grey = 95% CI



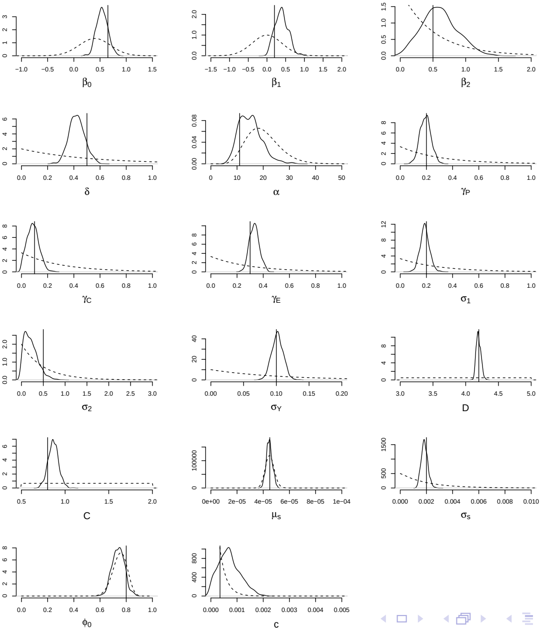


Simulation Study

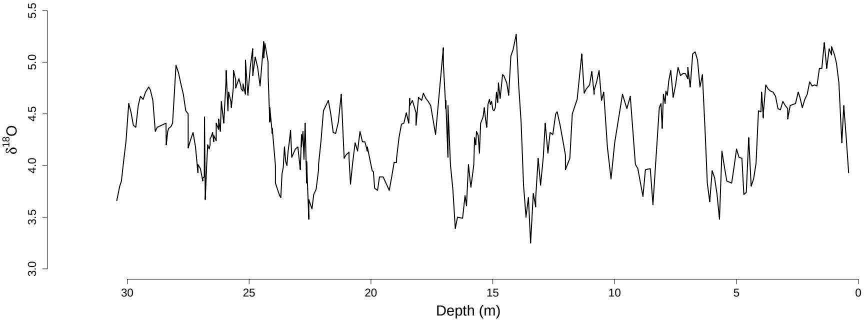
Dots = truth, black line = estimate, grey = 95% CI

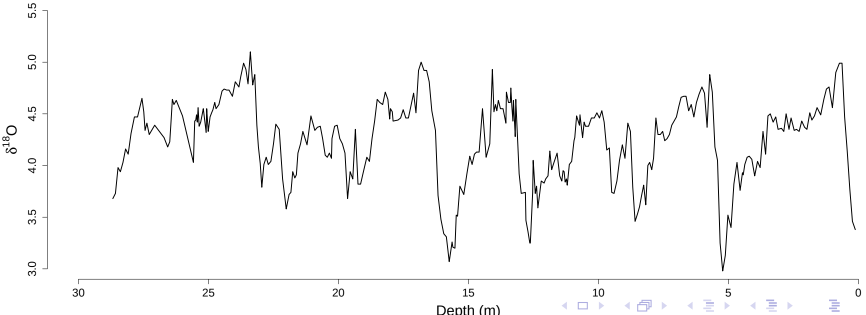


Simulation Study



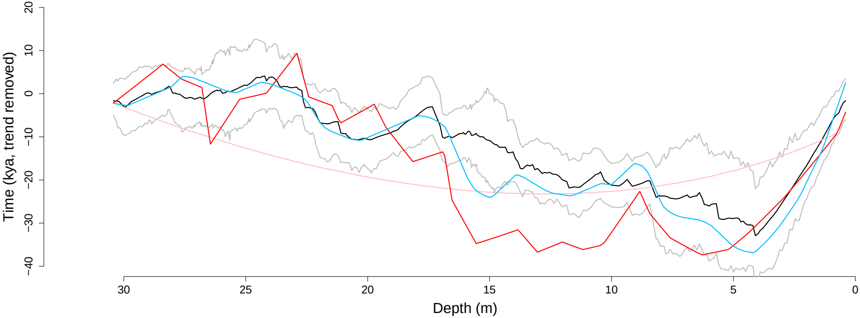
Real Data: ODP677 (top), ODP846

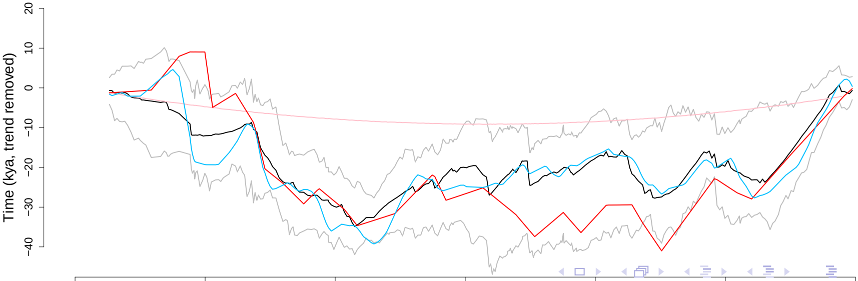




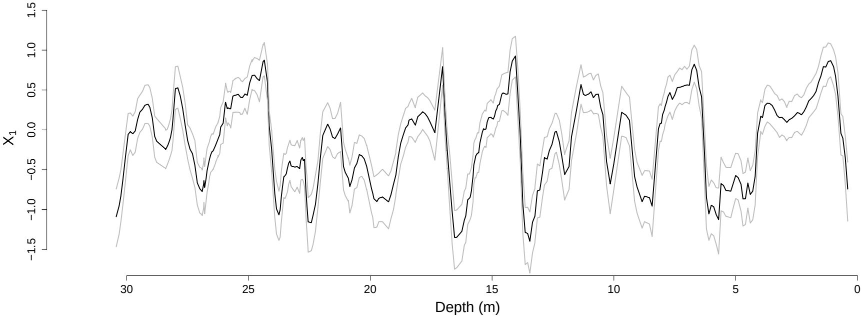
Real Data: ODP677(top), ODP846

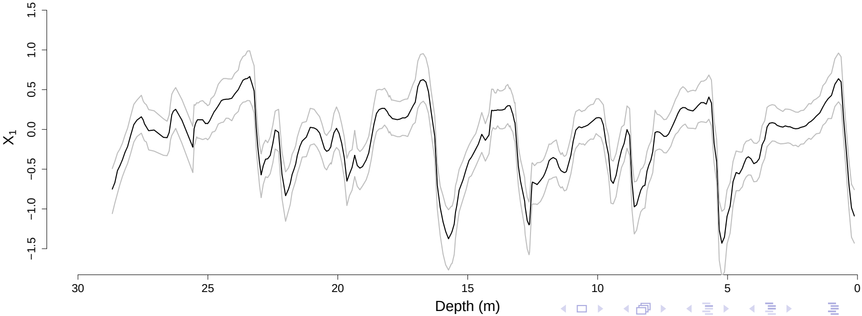
Posterior mean, 95%CI, Huybers 2007, Lisieki and Raymo 2005



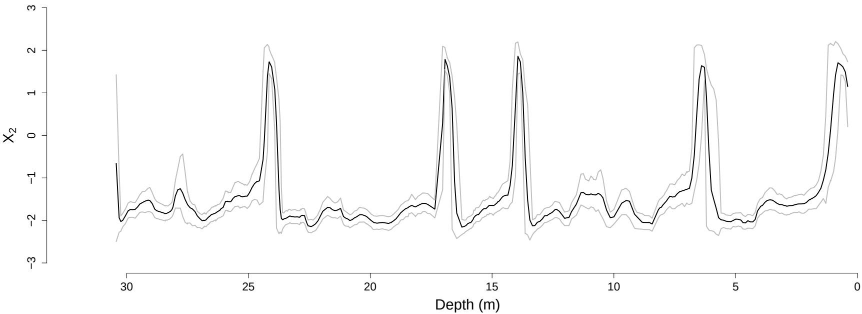


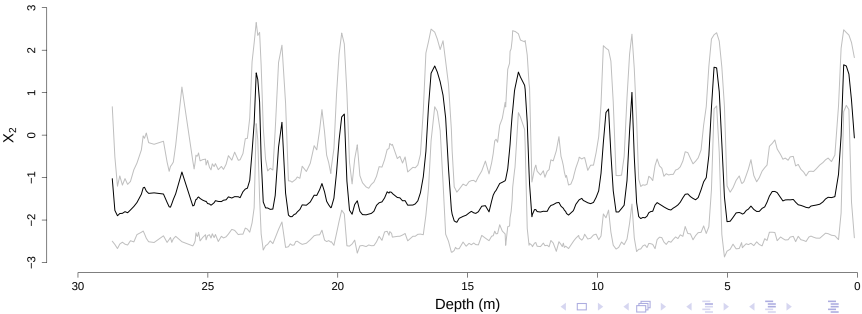
Real Data: ODP677(top), ODP846



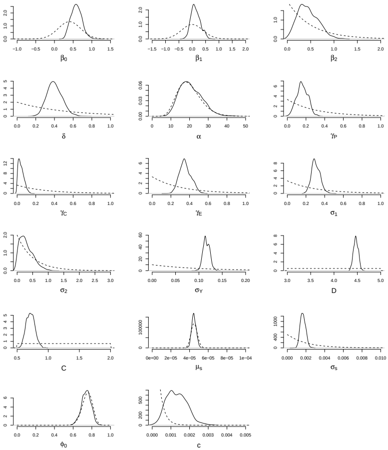


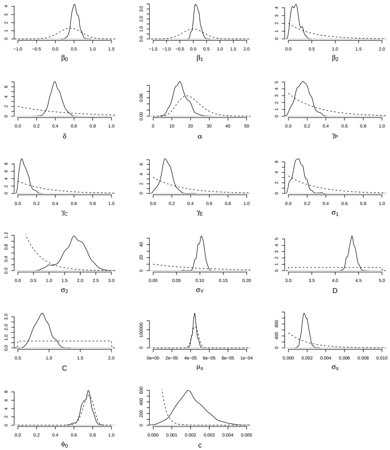
Real Data: ODP677(top), ODP846





Real Data: ODP677(left), ODP846





Model Comparison

- Selecting between the forced CR14 model and the unforced CR14 model, the $\log_{10}$ BFs are
  - Simulation Study: $\sim$9.
  - ODP 677: $\sim$5.
  - ODP 846: $\sim$0.

Model Comparison

- Selecting between the forced CR14 model and the unforced CR14 model, the $\log_{10}$ BFs are
  - Simulation Study: $\sim$9.
  - ODP 677: $\sim$5.
  - ODP 846: $\sim$0.
- Sampling four chronologies from the ODP 846 analysis then treating the ages as fixed, the $\log_{10}$ BFs are $\sim$0, $\sim$3, $\sim$9, $\sim$6.
- **Using a single fixed chronology leads to unreliable conclusions!**

Conclusions

- The data do contain enough information to discriminate between models and fit parameters
- Monte Carlo methodology and computer power are now sufficiently advanced that we can tackle the joint reconstruction, age model, and model selection problems for simple models
    - Proxy combination
    - but it remains computationally expensive ($\sim$ 1 week to compute per model here)
- Dating uncertainty matters
    - Independently calculating age estimates and fitting climate models can be a bad idea

Conclusions

- The data do contain enough information to discriminate between models and fit parameters
- Monte Carlo methodology and computer power are now sufficiently advanced that we can tackle the joint reconstruction, age model, and model selection problems for simple models
  - Proxy combination
  - but it remains computationally expensive ($\sim$ 1 week to compute per model here)
- Dating uncertainty matters
  - Independently calculating age estimates and fitting climate models can be a bad idea

Still to do/issues:

- ...

Conclusions

- The data do contain enough information to discriminate between models and fit parameters
- Monte Carlo methodology and computer power are now sufficiently advanced that we can tackle the joint reconstruction, age model, and model selection problems for simple models
  - Proxy combination
  - but it remains computationally expensive ($\sim$ 1 week to compute per model here)
- Dating uncertainty matters
  - Independently calculating age estimates and fitting climate models can be a bad idea

Still to do/issues:

- ...

Thank you for listening!

Bayes factors

Advantages:

- Can provide evidence for and against a model
- Penalises for complexity (Occam's razor).
- Asymptotic consistency

Disadvantages

- Hard to calculate
- Sensitive to choice of prior
- Integrated likelihood may not be desirable treatment
  - predictive evaluation via scoring rules? (not p-values)