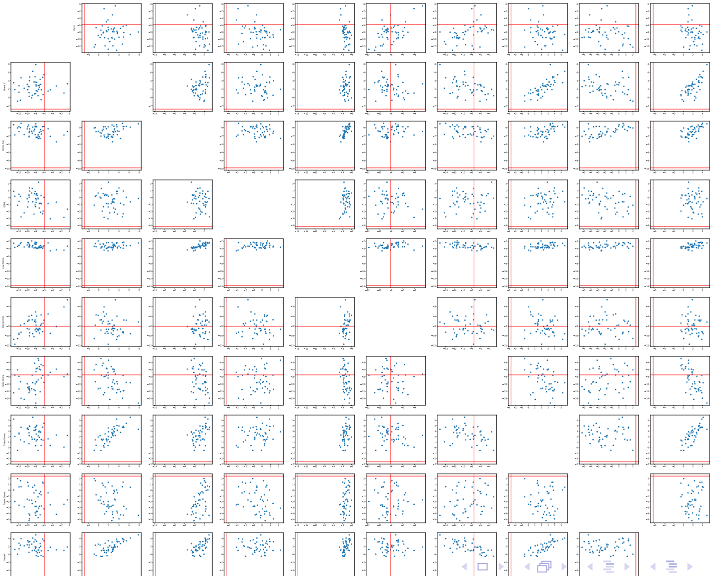
What drives the glacial-interglacial cycle? A Bayesian approach to a long-standing model selection problem

Jake Carson Michel Crucifix Simon Preston
Richard Wilkinson

University of Sheffield

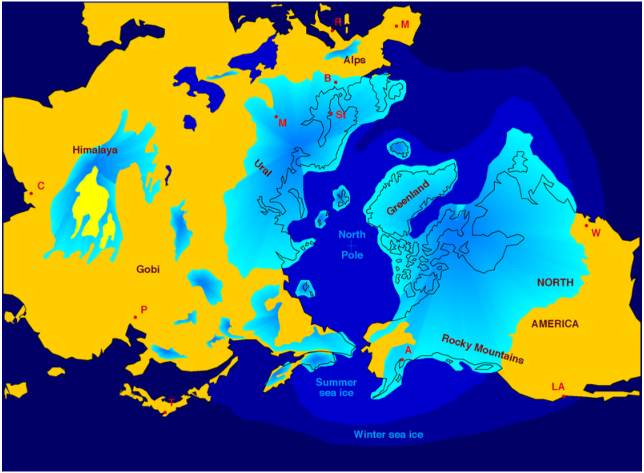
Lancaster 2019

Ice sheet modelling....

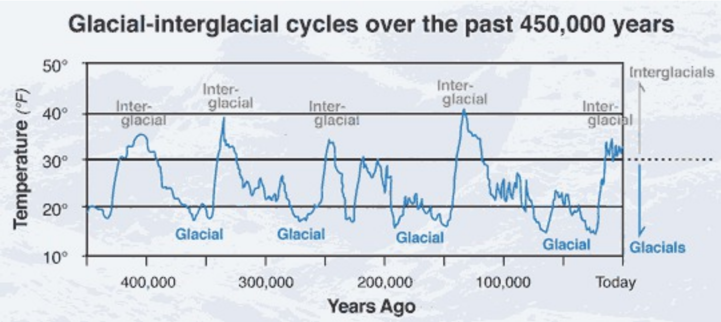


Glacial-Interglacial cycle

We're currently in the quaternary ice age
Last glacial period ended about 10,000 years ago (start of the Holocene)

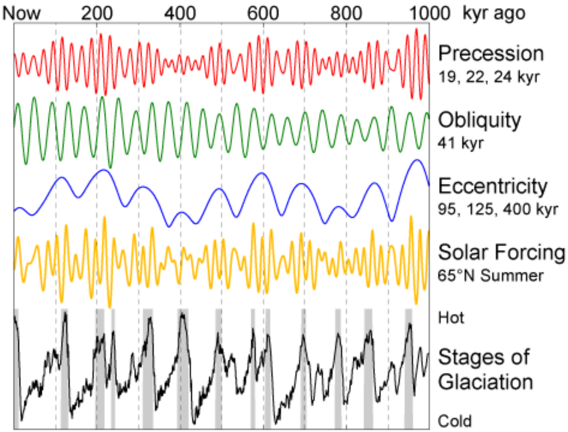


Glacial-Interglacial cycle



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.

Current practice

Climate scientists want to use palaeo-data to gather evidence for different hypotheses. They typically want to

- Compare models (and estimate parameters)
- Compare effects of different aspects of the solar forcing (all components have been argued for)
- Produce climate reconstructions (temperature chronologies)
- ...

Current practice

Climate scientists want to use palaeo-data to gather evidence for different hypotheses. They typically want to

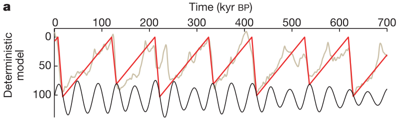
- Compare models (and estimate parameters)
- Compare effects of different aspects of the solar forcing (all components have been argued for)
- Produce climate reconstructions (temperature chronologies)
- ...

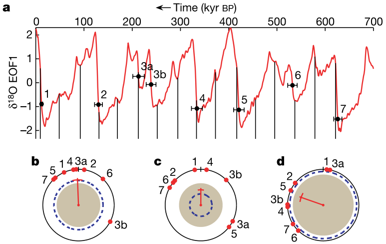
Current approaches tend to be statistically naive

- Models fit by eye,
- Model selection rarely tackled in a statistical manner, and when it is, questionable approaches are taken.

Example

Huybers and Wunsch 2005 argue that obliquity is the primary driver of glacial cycle





- Reduce the dataset to 7 termination times
- Look at the consistency of the phase of each component at terminations
- They propose a random walk model of ice volume with a 100kyr period

$$V_{t+1} = V_t + N(1,2) \text{ and if } V_t > 90, \text{ terminate}$$

and estimate the distribution of the test statistics under H_0

Our aim

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 aim

> *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)*

Can we do any better?

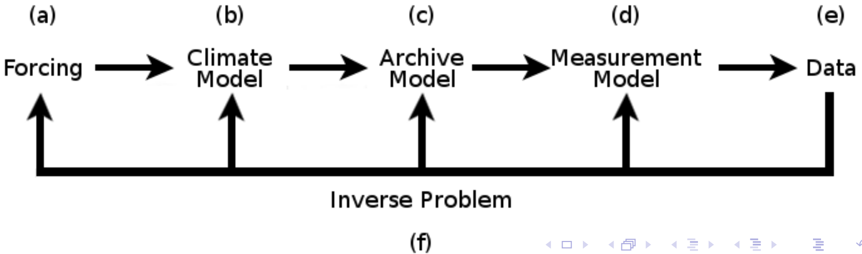
- Aim to demonstrate the power of the Bayesian approach; demonstrate that a full analysis is feasible
- Use all the data, not just the termination times
- Estimate parameters rather than using hand tuned models
- Deal with noisy records and age-model uncertainty

Our aim

> *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)*

Can we do any better?

- Aim to demonstrate the power of the Bayesian approach; demonstrate that a full analysis is feasible
- Use all the data, not just the termination times
- Estimate parameters rather than using hand tuned models
- Deal with noisy records and age-model uncertainty



$\delta^{18}O$ time-series

^{18}O is heavier than ^{16}O, and so its circulation behaviour varies with temperature

Variation in the ratio $\delta^{18}O$ provides information about historic ocean temperatures and ice volume.

The raw measurements are of $\delta^{18}O$ as a function of depth in a core: age must be inferred. The data are noisy, often contain hiatuses, are compacted etc.

Models

A phenomenological approach is taken: idealised simple models based on a few hypothesised relationships that capture some aspect of the climate system.

Let $X_t \in \mathbb{R}^p$ be the state of the climate at time t. Typically $X_{1,t}$ = ice volume, and other components many represent CO_2, ocean temp, etc, or be left undefined.

Models

A phenomenological approach is taken: idealised simple models based on a few hypothesised relationships that capture some aspect of the climate system.

Let $X_t \in \mathbb{R}^p$ be the state of the climate at time t. Typically $X_{1,t} =$ ice volume, and other components many represent CO_2, ocean temp, etc, or be left undefined.

- Oscillators synchronised on the solar forcing (Saltzman and Maasch 1991),

$$\begin{aligned} dX_1 &= -\left(X_1 + X_2 + vX_3 + F(\gamma_P, \gamma_C, \gamma_E)\right) dt + \sigma_1 dW_1 \\ dX_2 &= \left(rX_2 - pX_3 - sX_2^2 - X_2^3\right) dt + \sigma_2 dW_2 \\ dX_3 &= -q\left(X_1 + X_3\right) dt + \sigma_3 dW_3 \end{aligned}$$

- Models with switches in the ice volume (Tziperman 2006)

$$\begin{aligned} dX_1 &= \left((p_0 - KX_1)(1 - \alpha X_2) - (s + F(\gamma_P, \gamma_C, \gamma_E))\right) dt + \sigma_1 dW_1 \\ X_2 &: \quad \text{switches from 0 to 1 when } X_1 \text{ exceeds some threshold } X_u \\ X_2 &: \quad \text{switches from 1 to 0 when } X_1 \text{ decreases below } X_l \end{aligned}$$

- Models with switches dependent upon thresholds in the forcing (Parrenin and Paillard 2012)

Models

A phenomenological approach is taken: idealised simple models based on a few hypothesised relationships that capture some aspect of the climate system.

Let $X_t \in \mathbb{R}^p$ be the state of the climate at time t. Typically $X_{1,t} =$ ice volume, and other components many represent CO_2, ocean temp, etc, or be left undefined.

- Oscillators synchronised on the solar forcing (Saltzman and Maasch 1991),

$$\begin{aligned} dX_1 &= -\left(X_1 + X_2 + vX_3 + F(\gamma_P, \gamma_C, \gamma_E)\right) dt + \sigma_1 dW_1 \\ dX_2 &= \left(rX_2 - pX_3 - sX_2^2 - X_2^3\right) dt + \sigma_2 dW_2 \\ dX_3 &= -q\left(X_1 + X_3\right) dt + \sigma_3 dW_3 \end{aligned}$$

- Models with switches in the ice volume (Tziperman 2006)

$$\begin{aligned} dX_1 &= \left((p_0 - KX_1)(1 - \alpha X_2) - (s + F(\gamma_P, \gamma_C, \gamma_E))\right) dt + \sigma_1 dW_1 \\ X_2 &: \text{ switches from 0 to 1 when } X_1 \text{ exceeds some threshold } X_u \\ X_2 &: \text{ switches from 1 to 0 when } X_1 \text{ decreases below } X_l \end{aligned}$$

- Models with switches dependent upon thresholds in the forcing (Parrenin and Paillard 2012)

The prevailing view is that these should be distinguished between on the basis of scientific principles, not data. Parameters fit by eye.

Statistical model

These models are forced with 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.

Statistical model

These models are forced with 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.

Embed these models within a statistical state space model relating climate to observations

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

where we have 'noised-up' the models turning them into SDEs to account for model discrepancies.
Typically these models have 10-15 parameters that need to be estimated from the data.

Bayesian basics

- We use probability distributions $\pi(\theta)$ to represent our knowledge about all quantities.

Bayesian basics

- We use probability distributions $\pi(\theta)$ to represent our knowledge about all quantities.
- Use conditional probabilities to describe relationships/models, e.g.
  - $\pi(x|\theta)$ = distribution of output x from your simulator when run using parameter θ
  - $\pi(y|x)$ = model describing relationship between observations y and simulator prediction x.

Bayesian basics

- We use probability distributions $\pi(\theta)$ to represent our knowledge about all quantities.
- Use conditional probabilities to describe relationships/models, e.g.
  - $\pi(x|\theta)$ = distribution of output x from your simulator when run using parameter θ
  - $\pi(y|x)$ = model describing relationship between observations y and simulator prediction x.
- Use Bayes theorem ('calculus of probabilities') to describe our knowledge about quantities of interest

$$\text{posterior } = \pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \propto \text{likelihood} \times \text{prior}$$

Bayesian basics

- We use probability distributions $\pi(\theta)$ to represent our knowledge about all quantities.
- Use conditional probabilities to describe relationships/models, e.g.
  - $\pi(x|\theta)$ = distribution of output x from your simulator when run using parameter θ
  - $\pi(y|x)$ = model describing relationship between observations y and simulator prediction x.
- Use Bayes theorem ('calculus of probabilities') to describe our knowledge about quantities of interest

$$\text{posterior} = \pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \propto \text{likelihood} \times \text{prior}$$

$$= \frac{\int \pi(y|x)\pi(x|\theta)\mathrm{d}x \; \pi(\theta)}{\iint \pi(y|x)\pi(x|\theta)\pi(\theta)\mathrm{d}x\mathrm{d}\theta}$$

Added difficulty: $\pi(x|\theta)$ is usually unknown!

Bayesian basics

The quantities we need to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T}|y_{1:T},\theta_m,\mathcal{M}_m) \propto \pi(x_{1:T-1}|y_{1:T-1},\theta)\pi(x_T|x_{T-1},\theta)\pi(y_T|x_T)$$

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

Bayesian basics

The quantities we need to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T}|y_{1:T}, \theta_m, \mathcal{M}_m) \propto \pi(x_{1:T-1}|y_{1:T-1}, \theta)\pi(x_T|x_{T-1}, \theta)\pi(y_T|x_T)$$

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

- Model calibration (marginal parameter posterior)

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

Bayesian basics

The quantities we need to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T}|y_{1:T}, \theta_m, \mathcal{M}_m) \propto \pi(x_{1:T-1}|y_{1:T-1}, \theta)\pi(x_T|x_{T-1}, \theta)\pi(y_T|x_T)$$

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

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

- Model selection (model evidence/Bayes factors)

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

Bayesian basics

The quantities we need to calculate are

- Climate reconstruction (filtering)

$$\pi(x_{1:T}|y_{1:T}, \theta_m, \mathcal{M}_m) \propto \pi(x_{1:T-1}|y_{1:T-1}, \theta)\pi(x_T|x_{T-1}, \theta)\pi(y_T|x_T)$$

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

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

- Model selection (model evidence/Bayes factors)

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

These are progressively more difficult to calculate, particularly as

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

is unknown.

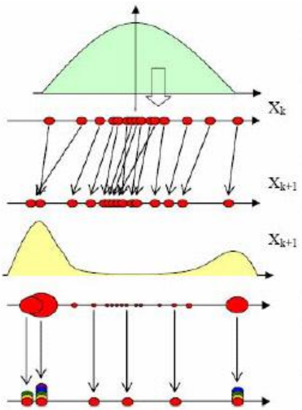
Filtering

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.



SMC

At time $t-1$, suppose $(X^n_{1:t-1}, W^n_{t-1})_{n=1}^N$ is a collection of weighted particles approximating $\pi(X_{1:t-1}|Y_{1:t-1},\theta)$

- Sample ancestor particle index $\mathcal{A}^n_{t-1} \sim \mathcal{F}(W^n_{t-1})$
- Propagate state particles $X^n_t \sim q_t(\cdot|X^{\mathcal{A}^n_{t-1}}_{t-1}, \theta, Y_t)$
- Weight state particles

$$w^n_t(X^n_{1:t}) = \frac{\pi(X^n_t|X^{\mathcal{A}^n_{t-1}}_{t-1},\theta)\pi(Y_t|X^n_t)}{q_t(X^n_t|X^{\mathcal{A}^n_{t-1}}_{t-1},\theta, Y_t)}, \qquad W^n_t = \frac{w^n_t(X^n_{1:t})}{\sum_n w^n_t(X^n_{1:t})}$$

SMC

At time $t-1$, suppose $(X_{1:t-1}^n, W_{t-1}^n)_{n=1}^N$ is a collection of weighted particles approximating $\pi(X_{1:t-1}|Y_{1:t-1},\theta)$

- Sample ancestor particle index $\mathcal{A}_{t-1}^n \sim \mathcal{F}(W_{t-1}^n)$
- Propagate state particles $X_t^n \sim q_t(\cdot|X_{t-1}^{\mathcal{A}_{t-1}^n},\theta,Y_t)$
- Weight state particles

$$w_t^n(X_{1:t}^n) = \frac{\pi(X_t^n|X_{t-1}^{\mathcal{A}_{t-1}^n},\theta)\pi(Y_t|X_t^n)}{q_t(X_t^n|X_{t-1}^{\mathcal{A}_{t-1}^n},\theta,Y_t)}, \qquad W_t^n = \frac{w_t^n(X_{1:t}^n)}{\sum_n w_t^n(X_{1:t}^n)}$$

We need $\pi(X_t|X_{t-1},\theta)$ to cancel, but setting $q=\pi$ can lead to extreme degeneracy, as too many proposals are in regions of low-posterior probability

We use the Golightly and Wilkinson (2006) approach to nudge the proposals towards the data.

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 $\pi(y_t|y_{1:t-1}, \theta)$.

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

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

and

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

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.

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 with 1000 θ particles and 1000 X particles.

SMC^2 to sample from $\pi(\theta, X_{1:T}|y_{1:T})$

Assume that at stage t we have particles $\{\theta^i, X_{1:t}^{1:N_x,i}\}_{i=1}^{N_\theta}$ with weights $\{W_t^i\}$ that approximates $\pi(\theta, X_{1:t}|y_{1:t})$
For $t = 1, \ldots, T$:

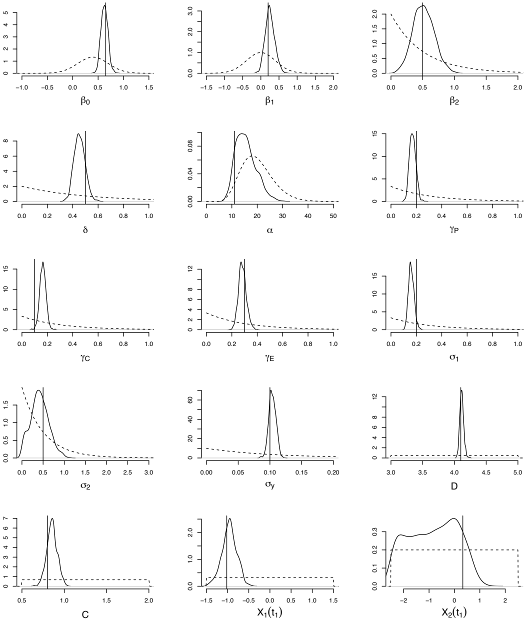
- If effective sample size is too small, resample by running a PMCMC algorithm targeting $\pi(\theta, X_{1:t}|y_{1:t})$
- Sample $\{\theta^i, X_{1:t+1}^{1:N_x,i}\}$ by performing iteration $t+1$ of the PF
- Estimate $\hat{\pi}(y_{t+1}|y_t, \theta^i)$
- Reweight by setting

$$w_{t+1}^i = w_t^i \hat{\pi}(y_{t+1}|y_t, \theta^i)$$

and $W_{t+1}^i = \frac{w_{t+1}^i}{\sum_i w_{t+1}^i}$

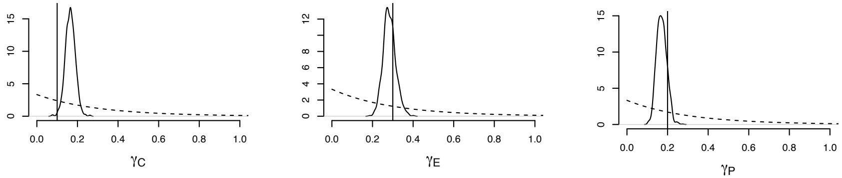
In total, this requires the use of $N_\theta \times N_x$ particles.

Results - simulation study



Results - simulation study

$\gamma = (\gamma_P, \gamma_E, \gamma_C)$ controls the relative contribution of the three components of the orbital variations in the forcing.



Bayes factors

Consider comparing two models, $\mathcal{M}_1$ and $\mathcal{M}_2$.

Bayes factors (BF) are the Bayesian approach to model selection.

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})} = \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_2)} \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)}$$

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

where

$$B_{12} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} = \frac{\int \pi(\theta_1|\mathcal{M}_1)\mathbb{P}\ (\mathcal{D}|\theta_1, \mathcal{M}_1) d\theta_1}{\int \pi(\theta_2|\mathcal{M}_2)\mathbb{P}(\mathcal{D}|\theta_2, \mathcal{M}_2) d\theta_2}$$

Bayes factors

Consider comparing two models, $\mathcal{M}_1$ and $\mathcal{M}_2$.

Bayes factors (BF) are the Bayesian approach to model selection.

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})} = \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_2)} \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)}$$

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

where

$$B_{12} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} = \frac{\int \pi(\theta_1|\mathcal{M}_1)\mathbb{P}\ (\mathcal{D}|\theta_1, \mathcal{M}_1) d\theta_1}{\int \pi(\theta_2|\mathcal{M}_2)\mathbb{P}(\mathcal{D}|\theta_2, \mathcal{M}_2) d\theta_2}$$

B_{12} range	$\mathbb{P}(\mathcal{M}_1\|D)$ range	Interpretation
1–3	0.5-0.75	Barely worth mentioning
3–10	0.75 - 0.91	Substantial
10–30	0.91-0.97	Strong
30–100	0.97- 0.99	Very strong
> 100	0.99-1	Decisive

Model selection

To compare models $\mathcal{M}_1$ and $\mathcal{M}_2$, we want to find the Bayes factor

$$B_{12} = \frac{\pi(y_{1:T}|\mathcal{M}_1)}{\pi(y_{1:T}|\mathcal{M}_2)}$$

Values of $B_{12} > 100$ indicate 'decisive' evidence in favour of $\mathcal{M}_1$.

Model selection

To compare models $\mathcal{M}_1$ and $\mathcal{M}_2$, we want to find the Bayes factor

$$B_{12} = \frac{\pi(y_{1:T}|\mathcal{M}_1)}{\pi(y_{1:T}|\mathcal{M}_2)}$$

Values of $B_{12} > 100$ indicate 'decisive' evidence in favour of $\mathcal{M}_1$.

SMC^2 can be used to provide an unbiased estimate of

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

for any model.

However, the variance of our estimates are typically an order of magnitude, so don't consider B_{12} to be large until we see values > 1000.

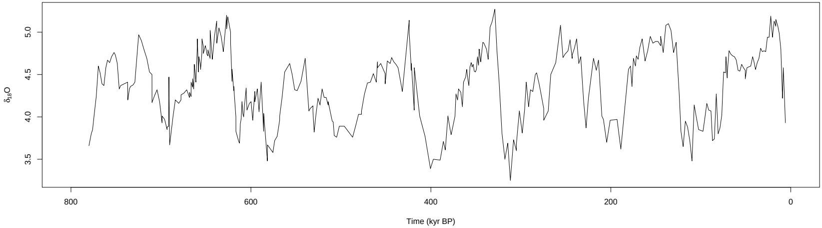
Results

We generate simulated data from SM91, using both the astronomically forced and unforced version of the model

Model		Evidence $\pi(y_{1:N} \mid \mathcal{M}_m)$	
		SM91-unforced	SM91-forced
SM91	Forced	5.6×10^{28}	1.4×10^{41}
	Unforced	1.1×10^{30}	2.4×10^{18}
T06	Forced	3.6×10^{20}	2.6×10^{30}
	Unforced	1.1×10^{22}	2.9×10^{14}
PP12	Forced	2.8×10^{8}	2.1×10^{18}

- Strongest evidence for the true model found each time
- Unforced model is special case of forced model with 3 parameters set to zero, so we expect it to be harder to select the unforced model.
- For the data generated from the forced model, the forced version of the wrong model is preferred.

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)

Results: ODP677

Model		Evidence	
		ODP677: H07(unforced)	ODP677: LR04(forced)
SM91	Forced	4.0×10^{24}	1.1×10^{28}
	Unforced	3.5×10^{26}	1.6×10^{18}
T06	Forced	3.3×10^{25}	4.5×10^{29}
	Unforced	1.7×10^{28}	3.3×10^{21}
PP12	Forced	1.5×10^{22}	1.8×10^{34}

The dating method applied changes the answer

- 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 PP12 a tuned model (PP12)

Moreover, orbitally tuned data leads us to strongly prefer the orbitally tuned version of each model (and vice versa)

Results: ODP677

Model		Evidence	
		ODP677: H07(unforced)	ODP677: LR04(forced)
SM91	Forced	4.0×10^{24}	1.1×10^{28}
	Unforced	3.5×10^{26}	1.6×10^{18}
T06	Forced	3.3×10^{25}	4.5×10^{29}
	Unforced	1.7×10^{28}	3.3×10^{21}
PP12	Forced	1.5×10^{22}	1.8×10^{34}

The dating method applied changes the answer

- 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 PP12 a tuned model (PP12)

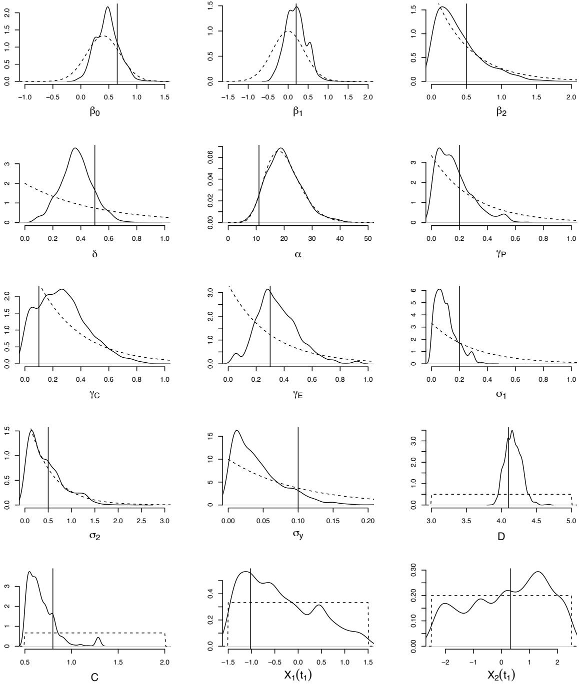
Moreover, orbitally tuned data leads us to strongly prefer the orbitally tuned version of each model (and vice versa)

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

Alternative approaches

ABC

- Instead of approximating the likelihood (as in SMC^2), we try to find θ that give good match between observed and simulated data
- Allows us to calibrate on carefully chosen aspects of the system (period, volatility, etc), rather than just on the data.
- The loss of information from the ABC approximation is large, so the posteriors are usually much wider than with SMC^2.

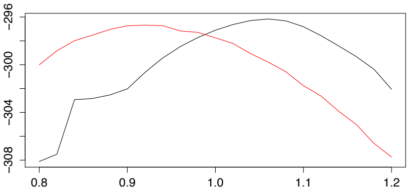


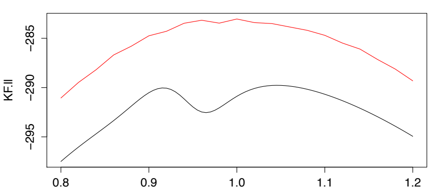
β_0
β_1
β_2
δ
α
γ_P
γ_C
γ_E
σ_1
σ_2
σ_y
D
C
$X_1(t_1)$
$X_2(t_1)$

Alternative approaches

Instead of using the particle filter (SMC) to do the filtering, we would like to use the unscented Kalman filter (UKF) or EnKF.

- Assumes $\pi(x_t|y_{1:t})$ is Gaussian and uses Sigma-point particles to estimate mean and variance.
- Much cheaper than *SMC* or *MCMC* approaches.
- We found the UKF works well for filtering (location), less well for parameter estimation, and terribly for model selection.

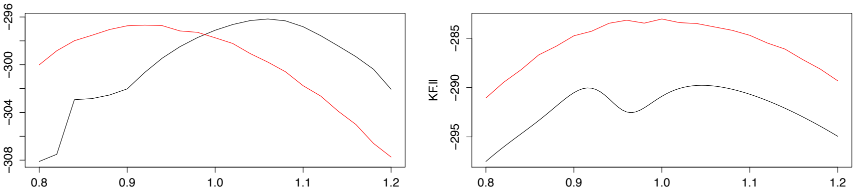




Alternative approaches

Instead of using the particle filter (SMC) to do the filtering, we would like to use the unscented Kalman filter (UKF) or EnKF.

- Assumes $\pi(x_t|y_{1:t})$ is Gaussian and uses Sigma-point particles to estimate mean and variance.
- Much cheaper than *SMC* or *MCMC* approaches.
- We found the UKF works well for filtering (location), less well for parameter estimation, and terribly for model selection.



A further discredited alternative is the idea of augmenting the state vector with the parameter, and inferring the joint distribution using the particle filter.

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.

Age model

Can we also quantify chronological uncertainty?

Target

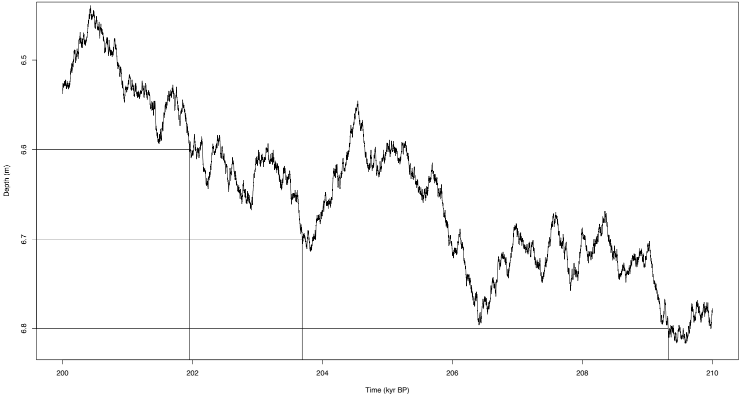
$$\pi(\theta, T_{1:N}, X_{1:N} | y_{1:N})$$

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:
Let H be the depth in the core, with $H_N = 0$ at $T_N = 0$

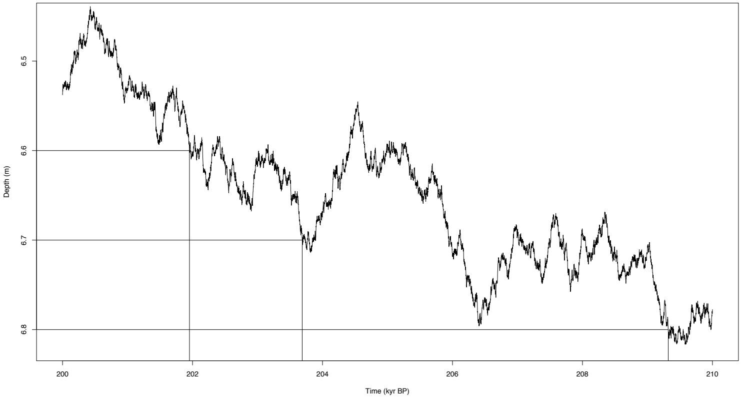
$$\mathrm{d}H = -\mu_s \mathrm{d}T + \sigma \mathrm{d}W$$

Slices are then taken through the core at specific depths $H_1, \ldots, H_N$.



There may have been multiple times when a certain depth was reached: the most recent time is the age of that slice, i.e., it is a first passage problem. Given (T_m, H_m), then T_{m-1} is the first passage time of H_{m-1} with

$$T_{m-1}|T_m \sim IG\left(T_m - \frac{H_{m-1} - H_m}{\mu_s}, \frac{(H_{m-1} - H_m)^2}{\sigma_s^2}\right)$$



There may have been multiple times when a certain depth was reached: the most recent time is the age of that slice, i.e., it is a first passage problem. Given (T_m, H_m), then T_{m-1} is the first passage time of H_{m-1} with

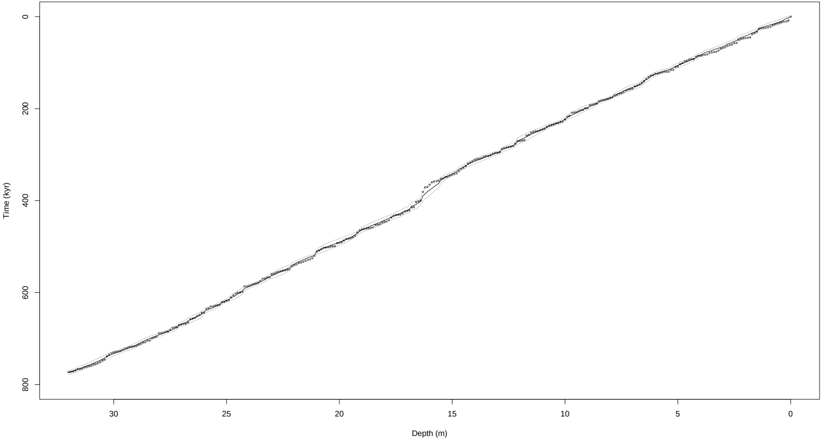
$$T_{m-1}|T_m \sim IG\left(T_m - \frac{H_{m-1} - H_m}{\mu_s}, \frac{(H_{m-1} - H_m)^2}{\sigma_s^2}\right)$$

We then add a model to account for compaction in the core, and apply Bayes theorem to find $\pi(T_m|T_{m-1})$ so that we can run the model forward in time

Simulation study results ($n = 321$) - age vs depth

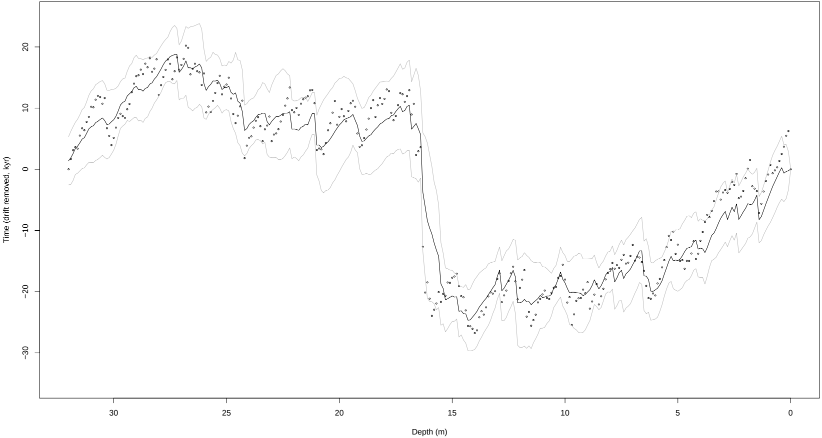
Dots = truth, black line = estimate, grey = 95% CI

We use simulated data from the CR12 model, with parameter values, and initial conditions comparable to real data. We consider the period 780 kyr to the present.



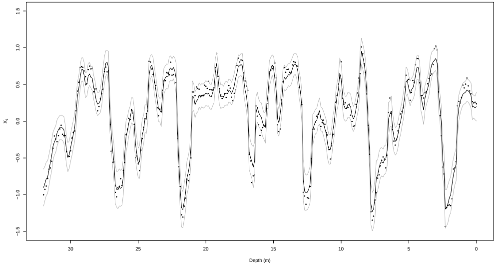
Simulation study results - age vs depth (trend removed)

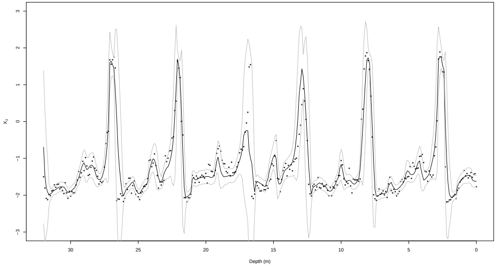
Dots = truth, black line = estimate, grey = 95% CI



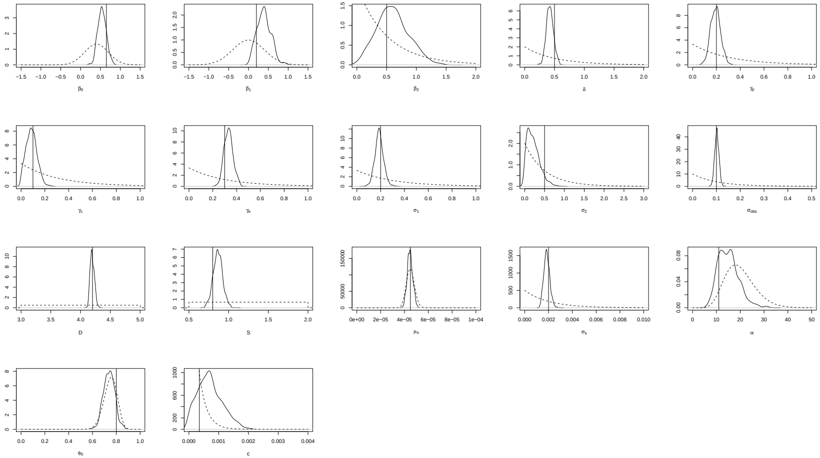
Simulation study results - climate reconstruction

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





Simulation study results - parameter estimation

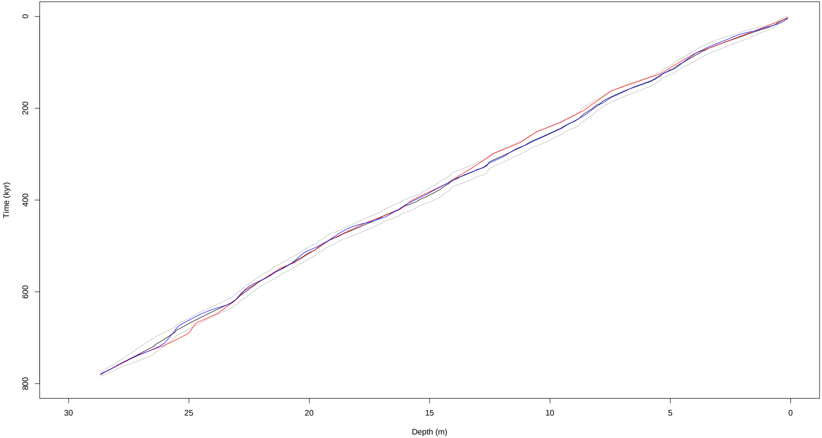


Results for ODP846 and ODP677 cores

- Cores contain markers for the Brunhes-Matuyama magnetic reversal at 780kyr, allowing us to give a strong prior for T_1 ($\pm$2kyr).
- Date estimates provided by two groups
  - Lisiekci and Raymo (LR04): graphical correlation of 57 cores. The stack is then orbitally tuned
  - Huybers and Wunsch 2004 (HW04) use a depth-derived age model. They decompact each core, fit a linear age model, then average over many several realisations and to get a distribution for 17 age control points(ACPS) , such as terminations. Average ages for the the ACP events are then found, and a linear age model is fitted between consecutive ACPs

Results for ODP846 - age vs depth

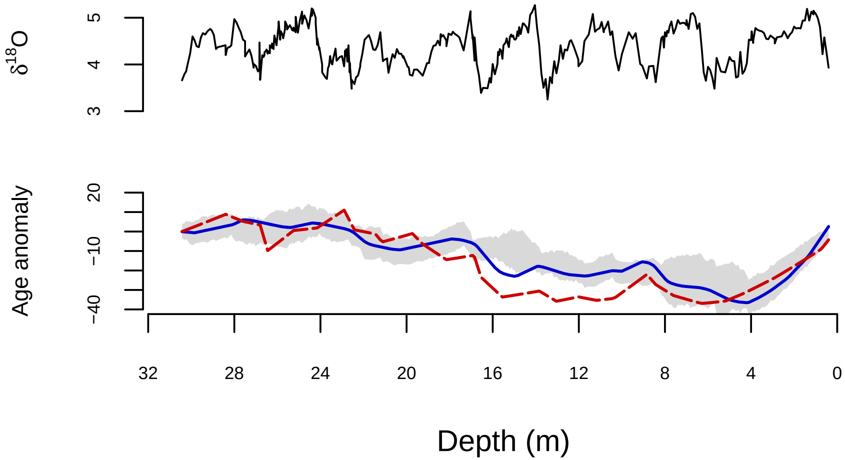
Black = posterior mean, grey = 95%CI, red = H07, blue = LR04



Our results come with uncertainty bounds (HW04 estimate accuracy of $\pm$9kyr for all ages). Moreover, the full joint distribution for all quantities is available if required.

Results for ODP677 - age vs depth (trend removed)

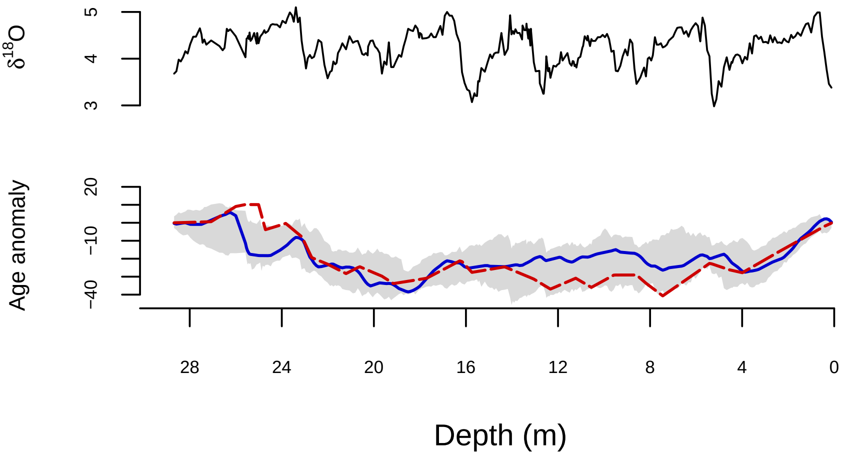
Grey = 95%HDR for observation ages, red = H07, blue = LR04



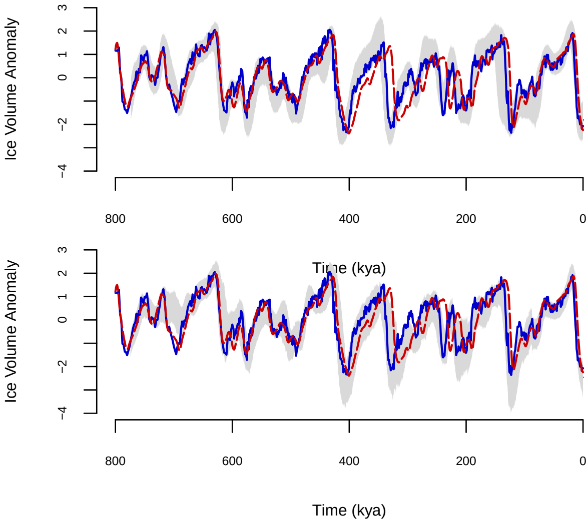
Note that these age estimates now depend explicitly on the model CR12.

Results for ODP846 - age vs depth (trend removed)

Grey = 95%HDR for observation ages, red = H07, blue = LR04

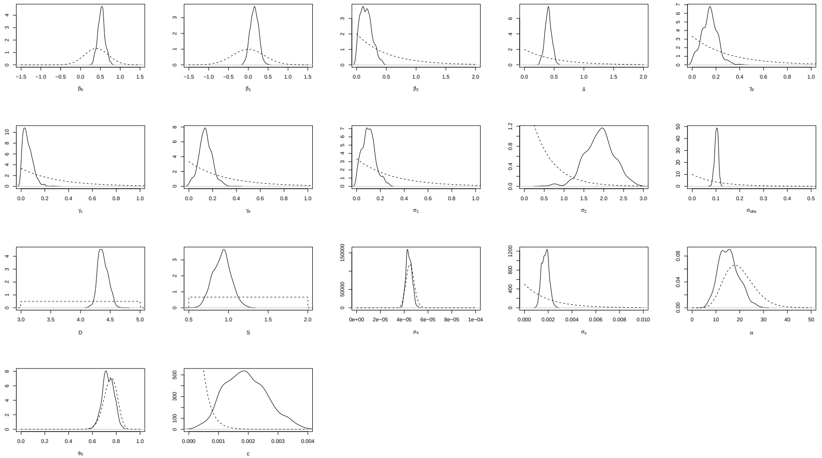


Climate reconstructions that account for age uncertainty



95% HDRs for the normalised ice volume over time for ODP677 (top) and ODP846 (bottom). LR04 (blue) H07 (red).

Results for ODP846 - parameter estimates



Interpretation of Bayes factors

Advantages:

- Provide evidence in favour of a model
- Provides an automatic form of Occam's razor.
- Do not require models to be nested
- Asymptotic consistency

Disadvantages

- **Hard** to calculate
- Sensitive to choice of prior
- Integrated likelihood may not be desirable treatment

Interpretation of Bayes factors

Advantages:

- Provide evidence in favour of a model
- Provides an automatic form of Occam's razor.
- Do not require models to be nested
- Asymptotic consistency

Disadvantages

- **Hard** to calculate
- Sensitive to choice of prior
- Integrated likelihood may not be desirable treatment

I believe the statistical estimates, but what is the correct scientific interpretation of a BF of 10^9?

Interpretation of Bayes factors

Advantages:

- Provide evidence in favour of a model
- Provides an automatic form of Occam's razor.
- Do not require models to be nested
- Asymptotic consistency

Disadvantages

- **Hard** to calculate
- Sensitive to choice of prior
- Integrated likelihood may not be desirable treatment

I believe the statistical estimates, but what is the correct scientific interpretation of a BF of 10^9?

Bayes factors are an *M-closed* tool - this is an *M-open* problem.

- None of the models is correct
- Are the estimates of the relevant astronomical forcing meaningful?
- Is the phenomenological motivation of the preferred model relevant?

Using GPs

Latent force models.
Mention Dennis Prangle's AISTATS paper
Using GCMs?

Conclusions

- The data do contain enough information to partially discriminate between models
- However, the results are very sensitive to the age model applied to the data.
  - If we don't do joint estimation of all uncertain quantities, the results are over confident and can lead to contradictory conclusions.
- Methodology and computer power now sufficiently advanced that we can tackle the joint reconstruction, age model, and model selection problems in a fully Bayesian manner
  - but it remains computationally expensive. The age model results take ~ 1 week to compute per model.

Conclusions

- The data do contain enough information to partially discriminate between models
- However, the results are very sensitive to the age model applied to the data.
  - If we don't do joint estimation of all uncertain quantities, the results are over confident and can lead to contradictory conclusions.
- Methodology and computer power now sufficiently advanced that we can tackle the joint reconstruction, age model, and model selection problems in a fully Bayesian manner
  - but it remains computationally expensive. The age model results take $\sim$ 1 week to compute per model.

Thank you for listening!

References

- Carson, Crucifix, Preston, W *Quantifying Age and Model Uncertainties in Palaeoclimate Data and Dynamical Climate Models with a Joint Inferential Analysis*, arXiv:1803.08482.
- Carson, Crucifix, Preston, W. *Bayesian model selection for the glacial-interglacial cycle*. J. Roy. Stat. Soc. C 2018.
- Chopin et al. *SMC^2 an efficient algorithm for sequential analysis of state-space models.* J. Roy. Stat. Soc. C 2013.
- Holden, Edwards, Ridgwell, W, et al. *Climate-carbon cycle uncertainties and the Paris Agreement*, Nature Climate Change, 2018.
- Holden, Edwards, Garthwaite, W., *Emulation and interpretation of high-dimensional climate model outputs*. J. App. Stat., 2015.
- Bounceur, Crucifix, W., *Global sensitivity analysis of the climate vegetation system to astronomical forcing: an emulator-based approach.* Earth Syst. Dynam. Discuss, 2015.