Accounting for the limitations of quantitative models

Jonathan Rougier

Department of Mathematics
University of Bristol, UK

SECaM, Exeter, Jan 2011

Ab initio modelling of complex systems

Ab initio modelling of complex systems

1. The laws of each of the individual processes are not known,

Ab initio modelling of complex systems

1. The laws of each of the individual processes are not known,
2. And nor is the coupling between them.

Ab initio modelling of complex systems

1. The laws of each of the individual processes are not known,
2. And nor is the coupling between them.
3. We cannot afford to solve these laws at sufficiently high resolution.

Ab initio modelling of complex systems

1. The laws of each of the individual processes are not known,
2. And nor is the coupling between them.
3. We cannot afford to solve these laws at sufficiently high resolution.

(4. Many of the processes are only indirectly relevant for policy-assessment.)

Ab initio modelling of complex systems

1. The laws of each of the individual processes are not known,
2. And nor is the coupling between them.
3. We cannot afford to solve these laws at sufficiently high resolution.

(4. Many of the processes are only indirectly relevant for policy-assessment.)

Fundamental law of complex systems:

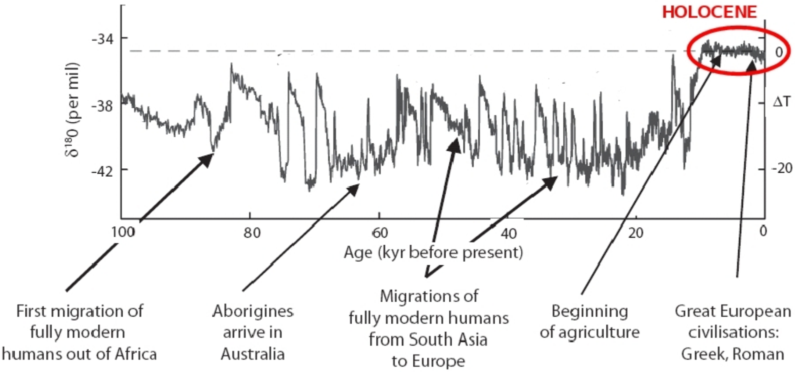
$$\text{Model limitations} = \text{system uncertainty}$$

Unfortunately, however, most *ab initio* models of complex systems are far too large and unwieldy to be embedded within a statistical framework that can defensibly and transparently represent system uncertainty.

Sometimes we are interested in just a few margins of a very complex system.

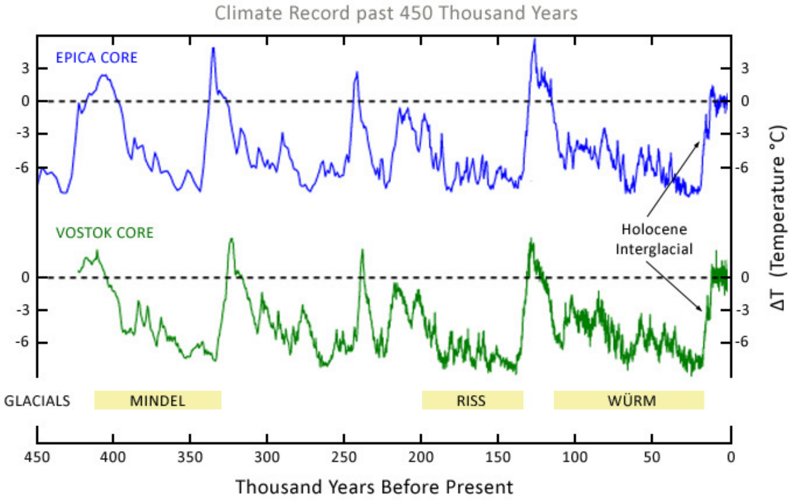
- How can statistics help us to model those margins directly?
- How do we account for the limitations of our model?

Running illustration: Glacial cycles



Source: `http://www.ecologyandsociety.org/vol14/iss2/art32/figure1.html`

Running illustration: Glacial cycles



Source: `http://essayweb.net/geology/quicknotes/iceage.shtml`

Notionally, in a nutshell

We suppose that our system has well-separated ‘slow’ and ‘fast’ variables, where we are interested in modelling the slow variables:

$$\tau_x \frac{\mathrm{d}x}{\mathrm{d}t} = f(x, y), \tag{Slow}$$

$$\tau_y \frac{\mathrm{d}y}{\mathrm{d}t} = g(x, y), \tag{Fast}$$

where $\tau_y \ll \tau_x$.

Notionally, in a nutshell

We suppose that our system has well-separated 'slow' and 'fast' variables, where we are interested in modelling the slow variables:

$$\tau_x \frac{\mathrm{d}x}{\mathrm{d}t} = f(x, y), \tag{Slow}$$

$$\tau_y \frac{\mathrm{d}y}{\mathrm{d}t} = g(x, y), \tag{Fast}$$

where $\tau_y \ll \tau_x$.

1. Replace the fast variable y with an ensemble average, and retain a fluctuation term,

$$\tau_x \frac{\mathrm{d}x}{\mathrm{d}t} = \langle f(x, Y) \rangle + \xi(x).$$

Notionally, in a nutshell

We suppose that our system has well-separated ‘slow’ and ‘fast’ variables, where we are interested in modelling the slow variables:

$$\tau_x \frac{\mathrm{d}x}{\mathrm{d}t} = f(x, y), \tag{Slow}$$

$$\tau_y \frac{\mathrm{d}y}{\mathrm{d}t} = g(x, y), \tag{Fast}$$

where $\tau_y \ll \tau_x$.

1. Replace the fast variable y with an ensemble average, and retain a fluctuation term,

$$\tau_x \frac{\mathrm{d}x}{\mathrm{d}t} = \langle f(x, Y) \rangle + \xi(x).$$

2. Replace the fluctuation term with a Brownian motion,

$$\tau_x \, \mathrm{d}x = \tilde{f}(x) \, \mathrm{d}t + \sigma(x) \cdot \mathrm{d}W(t),$$

where $\tilde{f}(x) := \langle f(x, Y) \rangle$.

The glacial cycle model

Stochastic forced van der Pol oscillator (model due to Michel Crucifix)

Has *slow* and *medium* variables represented explicitly, with 'ensemble averaging' over *fast* variables:

$$\tau_x \,\mathrm{d}x = -\big(y + \beta + \gamma F(t)\big)\mathrm{d}t \qquad \text{(Slow)}$$

$$\tau_y \,\mathrm{d}y = -\big(\psi'(y) - x\big)\mathrm{d}t + \sigma \cdot \mathrm{d}W(t) \qquad \text{(Medium)}$$

where $\psi'(y) := y^3/3 - y$. It is convenient to write $\tau_x = \tau$, and $\tau_y = \tau/\alpha$, where $\alpha \gg 1$. F is orbital forcing.

Crudely –

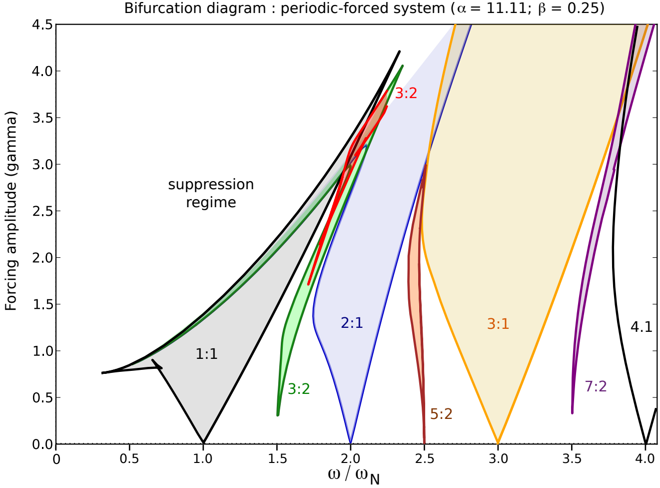
Slow: Ice volume,

Medium: Climate (e.g., Atlantic ocean circulation),

Fast: Weather.

The glacial cycle model . . . is subtle

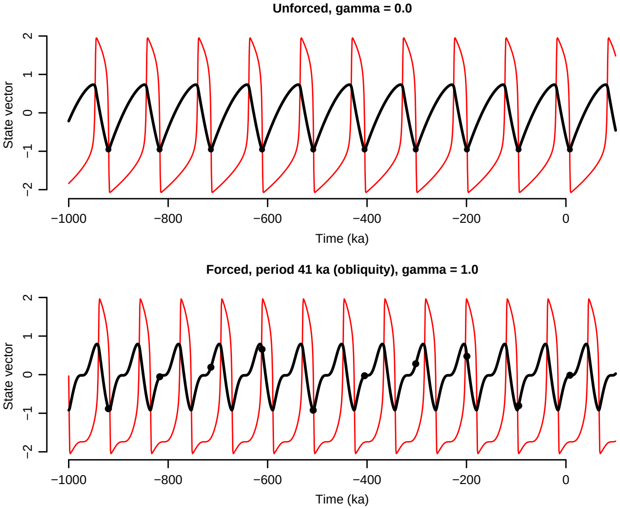
Deterministic model ($\sigma = 0$), periodic behaviour:



Source: Crucifix et al, 2010, Synchronisation on Astronomical Forcing, Isaac Newton Preprint, NI10044-CLP

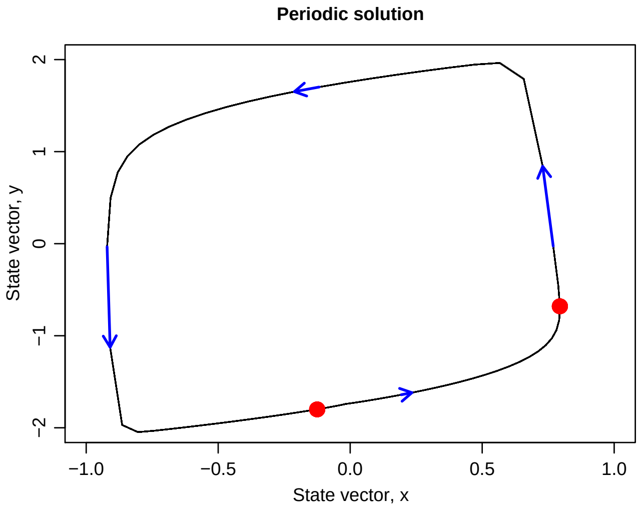
The glacial cycle model . . . is subtle

Deterministic model ($\sigma = 0$), periodic behaviour:



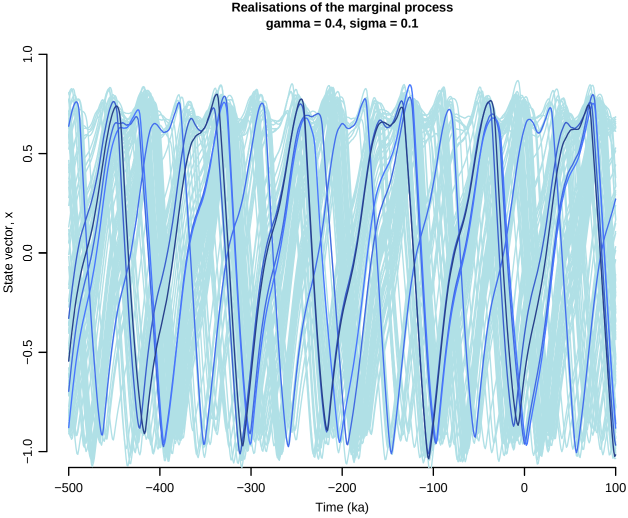
The glacial cycle model . . . is subtle

Deterministic model ($\sigma = 0$), periodic behaviour:



The glacial cycle model . . . is subtle

Phase-slip in the stochastic model:



Calibration and prediction

Denote the state vector as $\mathbf{x}_t := (x_t, y_t)$, the whole trajectory as $\mathbf{x} := (\mathbf{x}_0, \ldots, \mathbf{x}_T)$, and the parameters as $\theta := (\alpha, \beta, \gamma, \tau, \sigma)$. Denote observations on the state vector as $\mathbf{z} := (z_{t_1}, \ldots, z_{t_n})$, $0 \leq t_1 \leq t_2 \cdots \leq T$

1. Suppose that there are specified likelihood functions such that

$$L(\theta, \mathbf{x}) := \Pr(\mathbf{z} \mid \theta, \mathbf{x}) = \prod_{i=1}^{n} \Pr(z_{t_i} \mid \mathbf{x}_{t_i}).$$

2. Denote the specified marginal distribution of θ as $\Pr(\theta)$.
3. The model defines a stochastic process from which we can sample realisations from $\Pr(\mathbf{x}_0 \mid \theta)$, and from $\Pr(\mathbf{x}_t \mid \theta, \mathbf{x}_0, \ldots, \mathbf{x}_{t-1})$ for $t = 1, \ldots, T$.

The objective is to sample from the conditional distribution

$$\Pr(\theta, \mathbf{x} \mid \mathbf{z}) \propto L(\theta, \mathbf{x}) \Pr(\mathbf{x} \mid \theta) \Pr(\theta).$$

PMMH

PMMH = Particle Marginal Metropolis-Hastings.

1. Random walk in the parameter space, $q(\theta \to \theta')$;
2. Use an N particle filter to propose $\mathbf{x}' \sim \Pr(\mathbf{x} \mid \mathbf{z}, \theta')$, and to approximate the marginal likelihood $\hat{p}' \approx \Pr(\mathbf{z} \mid \theta')$;
3. Accept or reject $\{\theta', \mathbf{x}', \hat{p}'\}$ according to

$$\frac{\hat{p}' \Pr(\theta')}{\hat{p} \Pr(\theta)} \frac{q(\theta' \to \theta)}{q(\theta \to \theta')}$$

PMMH

PMMH = Particle Marginal Metropolis-Hastings.

1. Random walk in the parameter space, $q(\theta \rightarrow \theta')$;
2. Use an N particle filter to propose $\mathbf{x}' \sim \Pr(\mathbf{x} \mid \mathbf{z}, \theta')$, and to approximate the marginal likelihood $\hat{p}' \approx \Pr(\mathbf{z} \mid \theta')$;
3. Accept or reject $\{\theta', \mathbf{x}', \hat{p}'\}$ according to

$$\frac{\hat{p}' \Pr(\theta')}{\hat{p} \Pr(\theta)} \frac{q(\theta' \rightarrow \theta)}{q(\theta \rightarrow \theta')}$$

Theorem (Andrieu et al, 2010)

The stationary distribution of this chain is $\Pr(\theta, \mathbf{x} \mid \mathbf{z})$, even though N, the number of particles, may be small.

Making it work in practice

One needs to use every trick in the book in order to make this inference run on a laptop in a few hours.

1. One or two pilot studies to approximate the conditional variance matrix of θ, in order to set the proposal increment in a symmetric random walk (transformed parameter space). Can use a reduced set of measurements for greater speed.

Making it work in practice

One needs to use every trick in the book in order to make this inference run on a laptop in a few hours.

1. One or two pilot studies to approximate the conditional variance matrix of θ, in order to set the proposal increment in a symmetric random walk (transformed parameter space). Can use a reduced set of measurements for greater speed.

2. Tuning the number of particles may be extremely risky if done adaptively in the chain, but can be done in the pilot study. Speed of convergence to the region of high likelihood from different initial values is a useful guide.

Making it work in practice

One needs to use every trick in the book in order to make this inference run on a laptop in a few hours.

1. One or two pilot studies to approximate the conditional variance matrix of θ, in order to set the proposal increment in a symmetric random walk (transformed parameter space). Can use a reduced set of measurements for greater speed.

2. Tuning the number of particles may be extremely risky if done adaptively in the chain, but can be done in the pilot study. Speed of convergence to the region of high likelihood from different initial values is a useful guide.

3. Timid proposals along the eigenvectors of the conditional variance, seem to provide better mixing and have a strong psychological benefit.

Making it work in practice

One needs to use every trick in the book in order to make this inference run on a laptop in a few hours.

1. One or two pilot studies to approximate the conditional variance matrix of θ, in order to set the proposal increment in a symmetric random walk (transformed parameter space). Can use a reduced set of measurements for greater speed.

2. Tuning the number of particles may be extremely risky if done adaptively in the chain, but can be done in the pilot study. Speed of convergence to the region of high likelihood from different initial values is a useful guide.

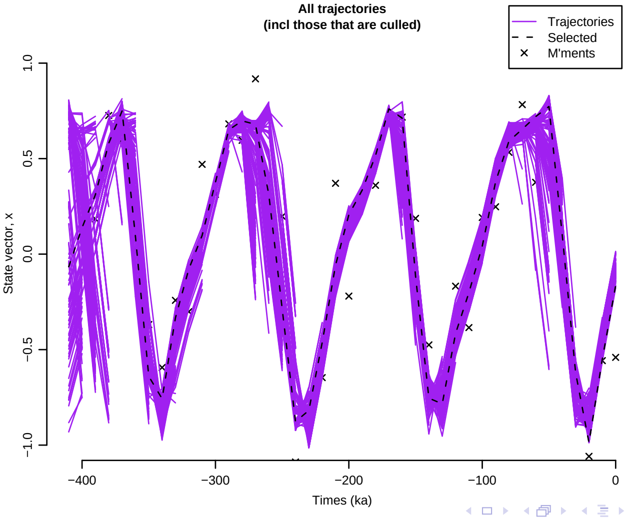
3. Timid proposals along the eigenvectors of the conditional variance, seem to provide better mixing and have a strong psychological benefit.

4. Interventions in the chain suggests carefully structuring the code to have specified break-points, and cold-, warm-, and hot-starts (standard MCMC programming).

The particle filter in action

The particle filter allows us to sample from $\Pr(\mathbf{x} \mid \mathbf{z}, \theta)$.

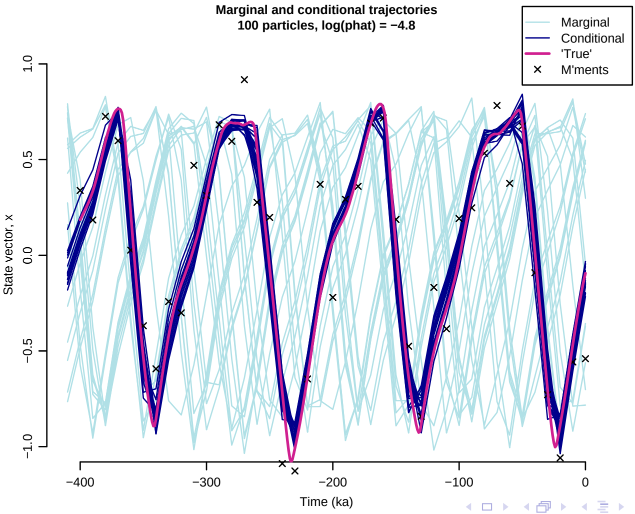
One run of the filter:



The particle filter in action

The particle filter allows us to sample from $\Pr(\mathbf{x} \mid \mathbf{z}, \theta)$.

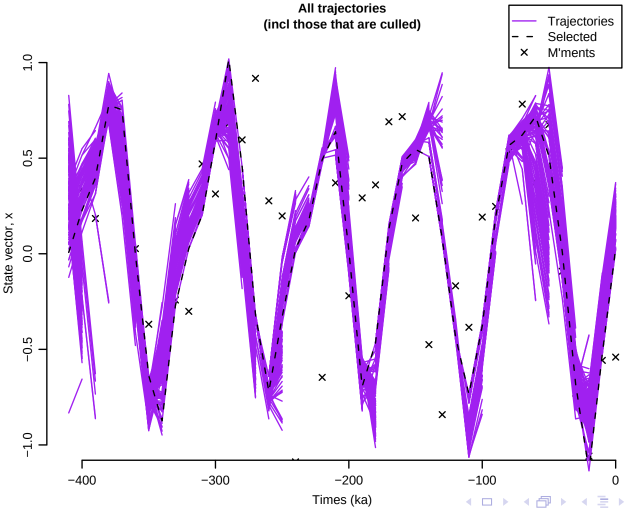
Sampling from the conditional distribution:



The particle filter in action

The particle filter allows us to sample from $\Pr(\mathbf{x} \mid \mathbf{z}, \theta)$.

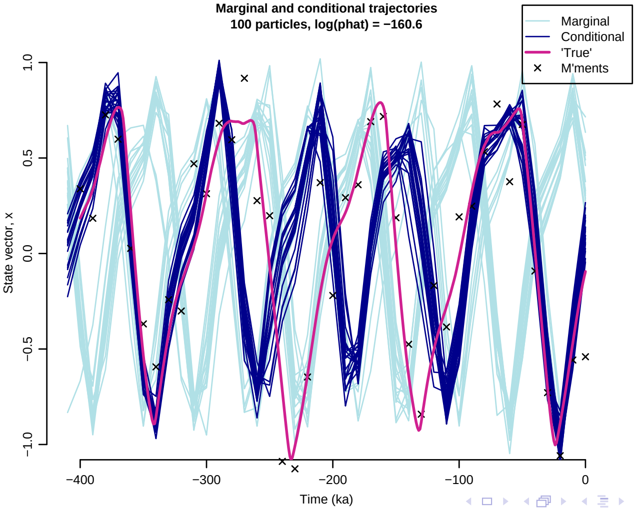
One run of the filter (WRONG θ):



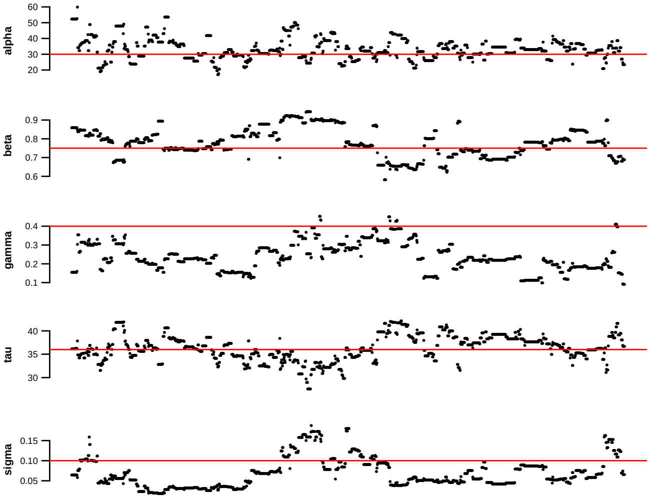
The particle filter in action

The particle filter allows us to sample from $\Pr(\mathbf{x} \mid \mathbf{z}, \theta)$.

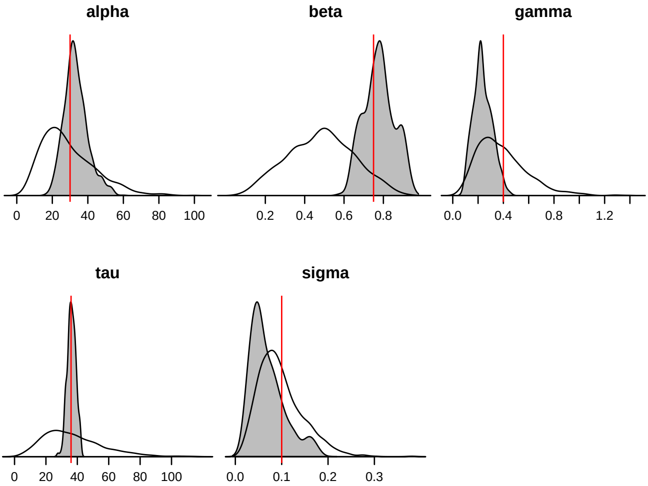
Sampling from the conditional distribution (WRONG θ):



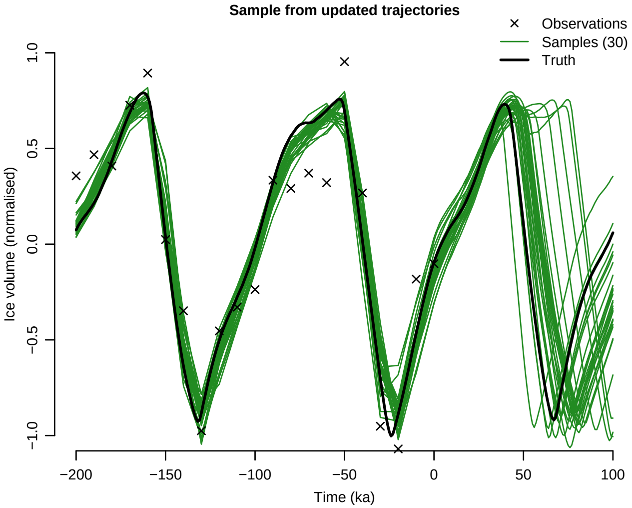
PMMH seems to be working in a toy problem



PMMH seems to be working in a toy problem



Conditional distribution of the trajectory (toy problem)



Summary

We now have a viable statistical framework for *data assimilation with uncertain static parameters.* This is exactly what we require when modelling margins of complex systems.

1. Phenomenological models represented as stochastic differential equations, using arguments such as separation of scale.
2. These models have uncertain parameters, and these can include the nature and size of the stochastic contribution.
3. Implementation can still be demanding. What PMMH (and its variants) offers is the replacement of an unfeasibly large calculation with a tediously long one.

With thanks to Michel Crucifix and the scientists associated with the ITOP project, Christophe Andrieu, and the Isaac Newton Institute CLP programme.