MCMC and Lagrangian data assimilation

Damon McDougall

Mathematics Institute, University of Warwick, United Kingdom

4th April 2011

Overview

- Problem set-up
- The solution
- Computing the solution: optimisation
- Computing the solution: Markov chain Monte Carlo
- Comparison of two problems
- Some insightful visualisations throughout

The forward problem

We start with the linearised shallow water equations

$$\frac{\partial \underline{u}}{\partial t} = J\underline{u} - \nabla h, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$\frac{\partial h}{\partial t} = -\nabla \cdot \underline{u}$$

The forward problem

We start with the linearised shallow water equations

$$\frac{\partial \underline{u}}{\partial t} = J\underline{u} - \nabla h, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$\frac{\partial h}{\partial t} = -\nabla \cdot \underline{u}$$

We consider a truncated Fourier expansion and seek solutions of the form

$$\begin{aligned} u(x, y, t) &= -2\pi \sin(2\pi x) \cos(2\pi y) u_0 + \cos(2\pi y) u_1(t) \\ v(x, y, t) &= 2\pi \cos(2\pi x) \sin(2\pi y) u_0 + \cos(2\pi y) v_1(t) \\ h(x, y, t) &= \sin(2\pi x) \sin(2\pi y) u_0 + \sin(2\pi y) h_1(t) \end{aligned}$$

The forward problem

We start with the linearised shallow water equations

$$\frac{\partial \underline{u}}{\partial t} = J\underline{u} - \nabla h, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$\frac{\partial h}{\partial t} = -\nabla \cdot \underline{u}$$
$$\frac{\mathrm{d}\underline{z}}{\mathrm{d}t} = \underline{u}(\underline{z}(t), t)$$

We consider a truncated Fourier expansion and seek solutions of the form

$$u(x, y, t) = -2\pi \sin(2\pi x)\cos(2\pi y)u_0 + \cos(2\pi y)u_1(t)$$
$$v(x, y, t) = 2\pi \cos(2\pi x)\sin(2\pi y)u_0 + \cos(2\pi y)v_1(t)$$
$$h(x, y, t) = \sin(2\pi x)\sin(2\pi y)u_0 + \sin(2\pi y)h_1(t)$$

The forward problem

- Algebraic wizardry gives

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{1}$$

with initial condition

$$w(0) := (u_0(0), u_1(0), v_1(0), h_1(0), \underline{z}(0)) \tag{2}$$

The forward problem

- Algebraic wizardry gives

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{1}$$

  with initial condition

$$w(0) := (u_0(0), u_1(0), v_1(0), h_1(0), \underline{z}(0)) \tag{2}$$

- The *forward* problem:

  Given (1) *and* (2), *find* $w(t)$, $t > 0$.

Video

Cool video 1

The inverse problem

- Given

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{3}$$

  and initial condition

$$w(0) := (u_0(0), u_1(0), v_1(0), h_1(0), \underline{z}(0)),$$

  find $w(t)$, $t > 0$.

The inverse problem

- Given

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{3}$$

and data

$$\underline{y_{jk}} = \begin{pmatrix} \underline{u}(\underline{x}_j, t_k) \\ \underline{z}(t_k) \end{pmatrix} + \underline{\eta_{jk}}, \quad \underline{\eta_{jk}} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, C),$$

find $w(t)$, $t > 0$.

The inverse problem

- Given

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{3}$$

and data

$$\underline{y_{jk}} = \begin{pmatrix} \underline{u}(\underline{x}_j, t_k) \\ \underline{z}(t_k) \end{pmatrix} + \underline{\eta_{jk}}, \quad \underline{\eta_{jk}} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, C),$$
$$\rightsquigarrow \underline{y} = \mathcal{G}(w(0)) + \underline{\eta}, \quad \underline{\eta} \sim \mathcal{N}(0, B),$$

find $w(t)$, $t > 0$.

The inverse problem

- Given

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{3}$$

and data

$$\underline{y_{jk}} = \begin{pmatrix} \underline{u}(\underline{x}_j, t_k) \\ \underline{z}(t_k) \end{pmatrix} + \underline{\eta_{jk}}, \quad \underline{\eta_{jk}} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, C),$$
$$\rightsquigarrow \underline{y} = \mathcal{G}(w(0)) + \underline{\eta}, \quad \underline{\eta} \sim \mathcal{N}(0, B),$$

find $w(0)$.

The inverse problem

- Given

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{3}$$

  and data

$$\underline{y_{jk}} = \begin{pmatrix} \underline{u}(\underline{x}_j, t_k) \\ \underline{z}(t_k) \end{pmatrix} + \underline{\eta_{jk}}, \quad \underline{\eta_{jk}} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, C),$$
$$\rightsquigarrow \underline{y} = \mathcal{G}(w(0)) + \underline{\eta}, \quad \underline{\eta} \sim \mathcal{N}(0, B),$$

  find $w(0)$.

- This is called the *inverse* problem.

The inverse problem

- Given

$$\dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \tag{3}$$

  and data

$$\underline{y_{jk}} = \begin{pmatrix} \underline{u}(\underline{x}_j, t_k) \\ \underline{z}(t_k) \end{pmatrix} + \underline{\eta_{jk}}, \quad \underline{\eta_{jk}} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, C),$$
$$\rightsquigarrow \underline{y} = \mathcal{G}(w(0)) + \underline{\eta}, \quad \underline{\eta} \sim \mathcal{N}(0, B),$$

  find $w(0)$.

- This is called the *inverse* problem.
- Data assimilation: the act of incorporating $\underline{y}$ into (3) to obtain $w(0)$.

The optimiser's approach

- We want to find the best $w(0)$. What does best mean?
- Define

$$J(\cdot) = \frac{1}{2} \left\| y - \mathcal{G}(\cdot) \right\|^2$$

and find $\operatorname{argmin}_w J(w)$.

The optimiser's approach

- We want to find the best $w(0)$. What does best mean?
- Define

$$J(\cdot) = \frac{1}{2} \left\| y - \mathcal{G}(\cdot) \right\|^2$$

  and find $\text{argmin}_w J(w)$.
- There is a problem with this; ill-posedness.

The optimiser's approach

- We want to find the best $w(0)$. What does best mean?
- Define

$$J(\cdot) = \frac{1}{2}\|y - \mathcal{G}(\cdot)\|^2 + \frac{1}{\gamma}\|\cdot\|^2$$

and find $\operatorname{argmin}_w J(w)$. This is called Tikhonov regularisation.

- I am deliberately not specifying the norms. They depend on the problem.

The optimiser's approach

- We want to find the best $w(0)$. What does best mean?
- Define

$$J(\cdot) = \frac{1}{2}\left\|y - \mathcal{G}(\cdot)\right\|^2 + \frac{1}{\gamma}\left\|\cdot\right\|^2$$

  and find $\operatorname{argmin}_w J(w)$. This is called Tikhonov regularisation.

- I am deliberately not specifying the norms. They depend on the problem.
- Instead of finding $\operatorname{argmin}_w J(w)$, we will sample the probability distribution

$$p(x) = \exp(-J(x))$$

Now we shall tackle the problem from a Bayesian perspective...

The posterior

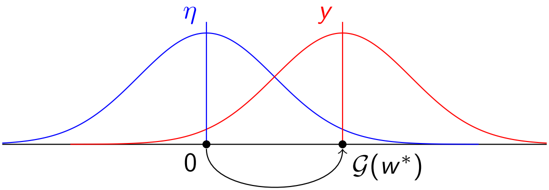
- Given y, find $w^* := w(0)$; i.e., want to know what $\mathbb{P}(w^*|y)$ looks like.
- $\mathbb{P}(w^*|y)$ is called the *posterior* ('after' the data) distribution

$$\begin{aligned}\mathbb{P}(w^*|y) &= \frac{\mathbb{P}(y|w^*)\mathbb{P}(w^*)}{\mathbb{P}(y)} \\ &\propto \mathbb{P}(y|w^*)\mathbb{P}(w^*)\end{aligned}$$

The posterior

- Given y, find $w^* := w(0)$; i.e., want to know what $\mathbb{P}(w^*|y)$ looks like.
- $\mathbb{P}(w^*|y)$ is called the *posterior* ('after' the data) distribution

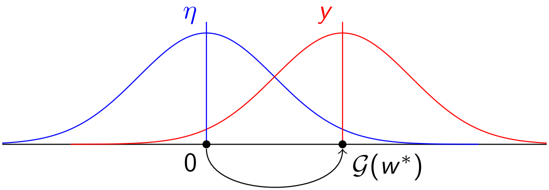
$$\begin{aligned}\mathbb{P}(w^*|y) &= \frac{\mathbb{P}(y|w^*)\mathbb{P}(w^*)}{\mathbb{P}(y)} \\ &\propto \mathbb{P}(y|w^*)\mathbb{P}(w^*)\end{aligned}$$



The posterior

- Given y, find $w^* := w(0)$; i.e., want to know what $\mathbb{P}(w^*|y)$ looks like.
- $\mathbb{P}(w^*|y)$ is called the *posterior* ('after' the data) distribution

$$\begin{aligned}\mathbb{P}(w^*|y) &= \frac{\mathbb{P}(y|w^*)\mathbb{P}(w^*)}{\mathbb{P}(y)} \\ &\propto \mathbb{P}(y|w^*)\mathbb{P}(w^*)\end{aligned}$$



- $\mathcal{G}$ must be linear (see why later).

Gaussians

- Fact: Gaussian probability distribution function (pdf) has the form

$$\exp(-(ax^2 + bx + c))$$

- Note: Product of two Gaussians is Gaussian

$$\begin{aligned} &\exp(-(a_1x^2 + b_1x + c_1))\exp(-(a_2x^2 + b_2x + c_2)) \\ = &\exp(-((a_1 + a_2)x^2 + (b_1 + b_2)x + (c_1 + c_2))) \end{aligned}$$

Gaussians

- Fact: Gaussian probability distribution function (pdf) has the form

$$\exp(-(ax^2 + bx + c))$$

- Note: Product of two Gaussians is Gaussian

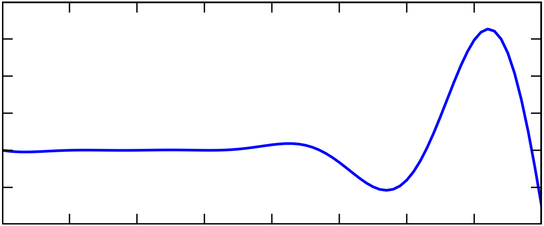
$$\begin{aligned} &\exp(-(a_1x^2 + b_1x + c_1))\exp(-(a_2x^2 + b_2x + c_2)) \\ =&\exp(-((a_1 + a_2)x^2 + (b_1 + b_2)x + (c_1 + c_2))) \end{aligned}$$

- So

$$\mathbb{P}(y|w^*), \mathbb{P}(w^*) \text{ Gaussian} \Rightarrow \mathbb{P}(y|w^*)\mathbb{P}(w^*) \propto \mathbb{P}(w^*|y) \text{ Gaussian}$$

The prior

- We need $\mathbb{P}(w^*)$. Called the *prior* ('before' the data) distribution.
- 'Prior knowledge' means have idea of some property of w^*. E.g., Temperature vs. Time:

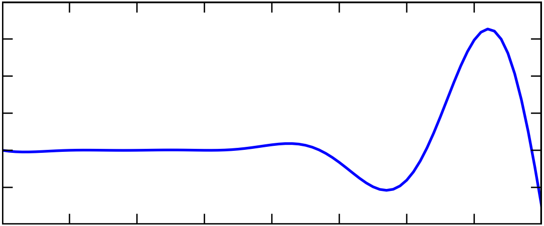


Might expect this to have, say, one derivative.
This 'prior knowledge' is usually either

1. Given to us
2. Obtained from past experience
3. A complete guess

The prior

- We need $\mathbb{P}(w^*)$. Called the *prior* ('before' the data) distribution.
- 'Prior knowledge' means have idea of some property of w^*. E.g., Temperature vs. Time:



Might expect this to have, say, one derivative.
This 'prior knowledge' is usually either

1. Given to us
2. Obtained from past experience
3. A complete guess. Ours will be $\mathbb{P}(w^*) = \mathcal{N}(0, I)$.

Finding 'the answer'

- $f : \mathcal{S} \to \mathbb{R}$ a probability density function if

$$f(x) \geq 0 \quad \forall x \in \mathcal{S} \qquad \text{and} \qquad \int_{\mathcal{S}} f(x)\,\mathrm{d}x = 1$$

Finding 'the answer'

- $f : \mathcal{S} \to \mathbb{R}$ a probability density function if

$$f(x) \geq 0 \quad \forall x \in \mathcal{S} \qquad \text{and} \qquad \int_{\mathcal{S}} f(x)\,\mathrm{d}x = 1$$

- Define $\Phi(\cdot) := \frac{1}{2}\left\|\mathcal{G}(\cdot) - y\right\|_B^2$. Objective is to minimise this.

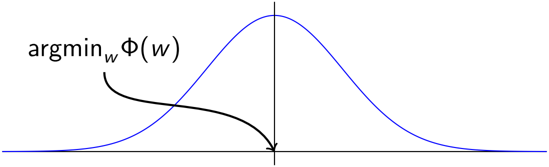
Finding 'the answer'

- $f : \mathcal{S} \to \mathbb{R}$ a probability density function if

$$f(x) \geq 0 \quad \forall x \in \mathcal{S} \qquad \text{and} \qquad \int_{\mathcal{S}} f(x)\,\mathrm{d}x = 1$$

- Define $\Phi(\cdot) := \frac{1}{2}\|\mathcal{G}(\cdot) - y\|_B^2$. Objective is to minimise this.
- Can turn into a probability density function

$$\exp(-\Phi(\cdot)) = \exp\left(-\frac{1}{2}\|\mathcal{G}(\cdot) - y\|_B^2\right).$$



Sampling $\mathbb{P}(w^*|y)$

- Idea: Construct $\{w_j\}_{j=1}^{\infty}$ cleverly such that $\{w_j\}_{j=1}^{\infty} \overset{\text{i.i.d}}{\sim} \mathbb{P}(w^*|y)$
  1. Let w_k be the 'current' state in the sequence and construct a *proposal*, $\hat{w}$

Sampling $\mathbb{P}(w^*|y)$

- Idea: Construct $\{w_j\}_{j=1}^{\infty}$ cleverly such that $\{w_j\}_{j=1}^{\infty} \overset{\text{i.i.d}}{\sim} \mathbb{P}(w^*|y)$
  1. Let w_k be the 'current' state in the sequence and construct a *proposal*, $\hat{w}$
  2. Define $\Phi(\cdot) := \frac{1}{2} \|\mathcal{G}(\cdot) - y\|_B^2$

Sampling $\mathbb{P}(w^*|y)$

- Idea: Construct $\{w_j\}_{j=1}^{\infty}$ cleverly such that $\{w_j\}_{j=1}^{\infty} \overset{\text{i.i.d}}{\sim} \mathbb{P}(w^*|y)$

  1. Let w_k be the 'current' state in the sequence and construct a *proposal*, $\hat{w}$

  2. Define $\Phi(\cdot) := \frac{1}{2} \|\mathcal{G}(\cdot) - y\|_B^2$
  3. Compute $\alpha(w_k, \hat{w}) = 1 \wedge \exp(\Phi(w_k) - \Phi(\hat{w}))$

Sampling $\mathbb{P}(w^*|y)$

- Idea: Construct $\{w_j\}_{j=1}^{\infty}$ cleverly such that $\{w_j\}_{j=1}^{\infty} \overset{\text{i.i.d}}{\sim} \mathbb{P}(w^*|y)$

  1. Let w_k be the 'current' state in the sequence and construct a *proposal*, $\hat{w}$

  2. Define $\Phi(\cdot) := \frac{1}{2} \left\| \mathcal{G}(\cdot) - y \right\|_B^2$

  3. Compute $\alpha(w_k, \hat{w}) = 1 \wedge \exp(\Phi(w_k) - \Phi(\hat{w}))$

  4. Let

$$w_{k+1} = \begin{cases} \hat{w} & \text{with probability } \alpha(w_k, \hat{w}) \\ w_k & \text{with probability } 1 - \alpha(w_k, \hat{w}) \end{cases}$$

Sampling $\mathbb{P}(w^*|y)$

- Idea: Construct $\{w_j\}_{j=1}^{\infty}$ cleverly such that $\{w_j\}_{j=1}^{\infty} \overset{\text{i.i.d}}{\sim} \mathbb{P}(w^*|y)$
  1. Let w_k be the 'current' state in the sequence. Make a draw $\xi \sim \mathcal{N}(0, I)$ and construct a *proposal*, $\hat{w}$

     $$\hat{w} = (1-\beta^2)^{\frac{1}{2}} w_k + \beta\xi, \quad \text{some } \beta \in (0,1)$$

  2. Define $\Phi(\cdot) := \frac{1}{2} \|\mathcal{G}(\cdot) - y\|_B^2$
  3. Compute $\alpha(w_k, \hat{w}) = 1 \wedge \exp(\Phi(w_k) - \Phi(\hat{w}))$
  4. Let

     $$w_{k+1} = \begin{cases} \hat{w} & \text{with probability } \alpha(w_k, \hat{w}) \\ w_k & \text{with probability } 1 - \alpha(w_k, \hat{w}) \end{cases}$$

Sampling $\mathbb{P}(w^*|y)$

- Idea: Construct $\{w_j\}_{j=1}^{\infty}$ cleverly such that $\{w_j\}_{j=1}^{\infty} \overset{\text{i.i.d}}{\sim} \mathbb{P}(w^*|y)$

  1. Let w_k be the 'current' state in the sequence. Make a draw $\xi \sim \mathcal{N}(0, I)$ and construct a *proposal*, $\hat{w}$

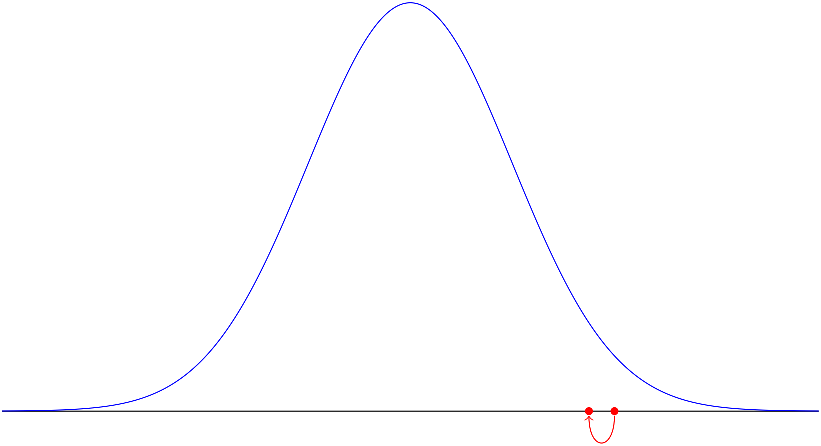
  $$\hat{w} = (1 - \beta^2)^{\frac{1}{2}} w_k + \beta\xi, \quad \text{some } \beta \in (0, 1)$$

  2. Define $\Phi(\cdot) := \frac{1}{2} \|\mathcal{G}(\cdot) - y\|_B^2$
  3. Compute $\alpha(w_k, \hat{w}) = 1 \wedge \exp(\Phi(w_k) - \Phi(\hat{w}))$
  4. Let

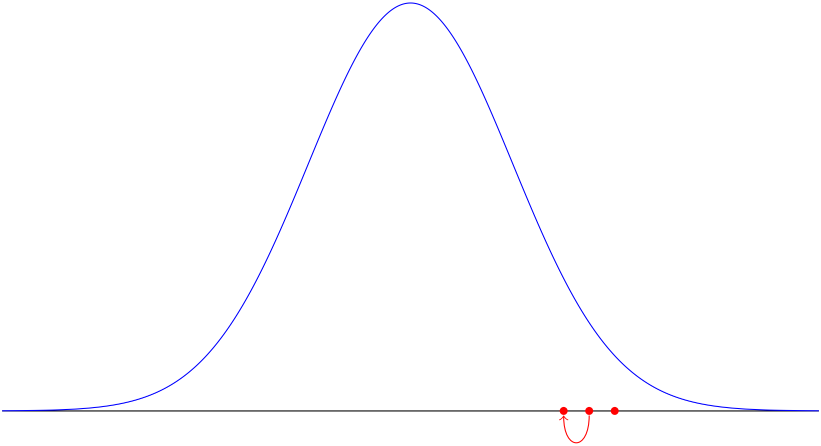
  $$w_{k+1} = \begin{cases} \hat{w} & \text{with probability } \alpha(w_k, \hat{w}) \\ w_k & \text{with probability } 1 - \alpha(w_k, \hat{w}) \end{cases}$$

- Take w_1 to be a draw from $\mathcal{N}(0, I)$

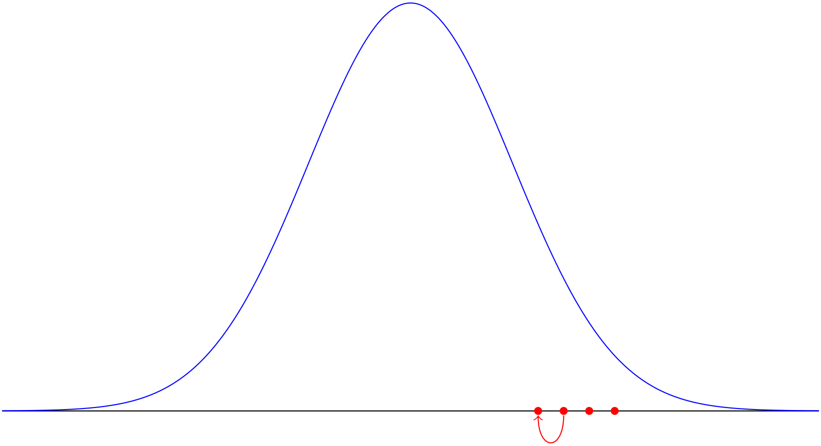
Sampling $\mathbb{P}(u|y)$



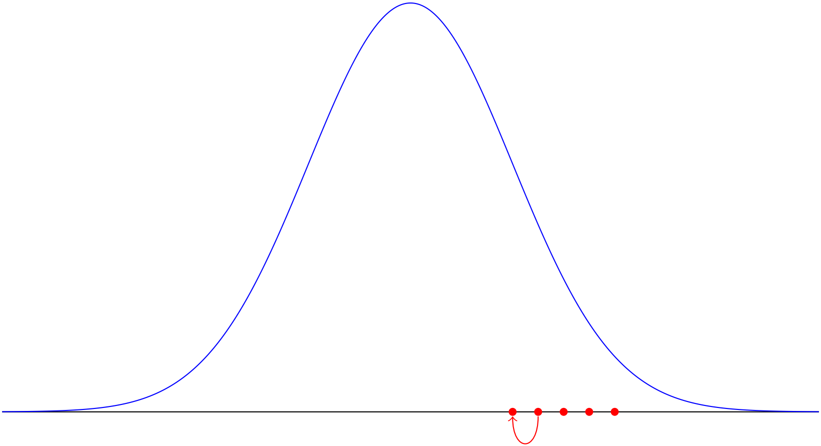
Sampling $\mathbb{P}(u|y)$



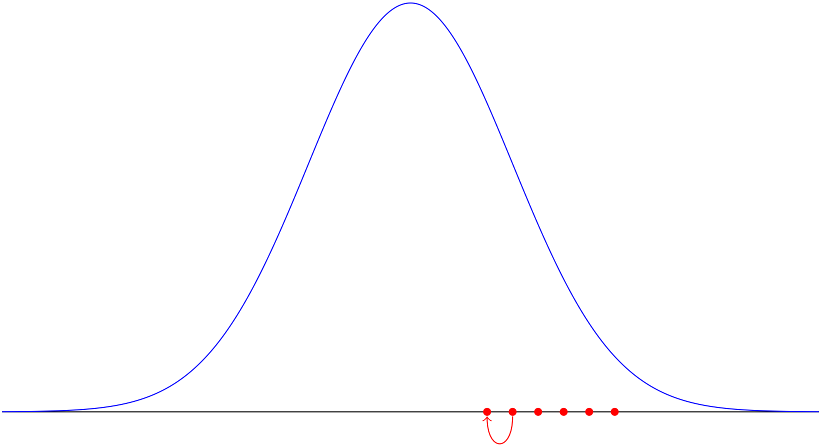
Sampling $\mathbb{P}(u|y)$



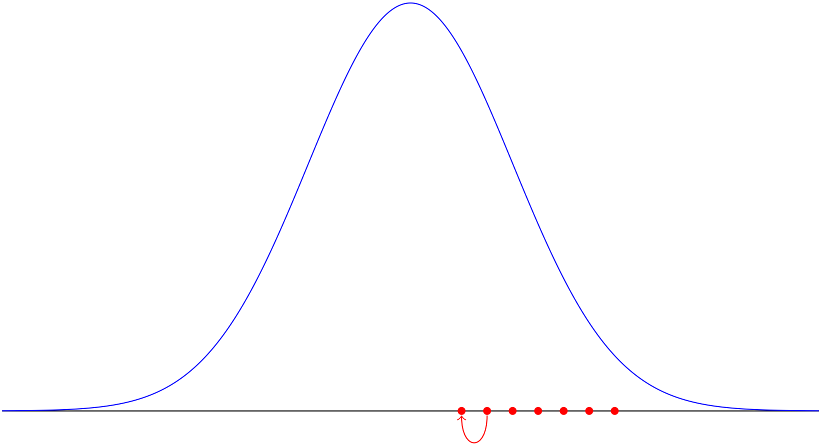
Sampling $\mathbb{P}(u|y)$



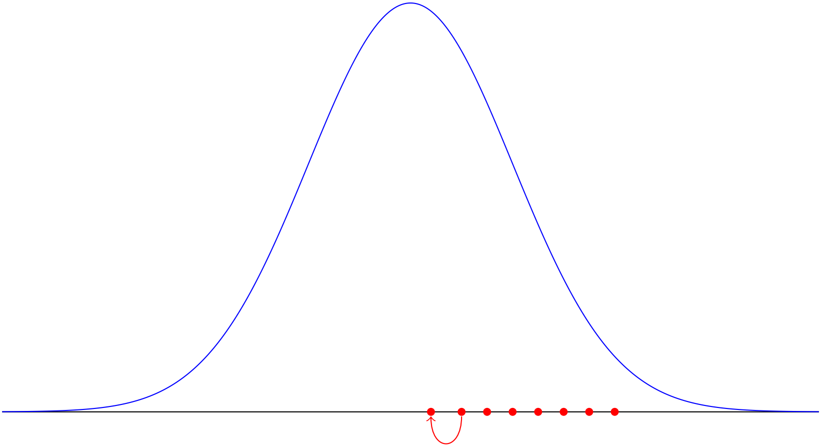
Sampling $\mathbb{P}(u|y)$



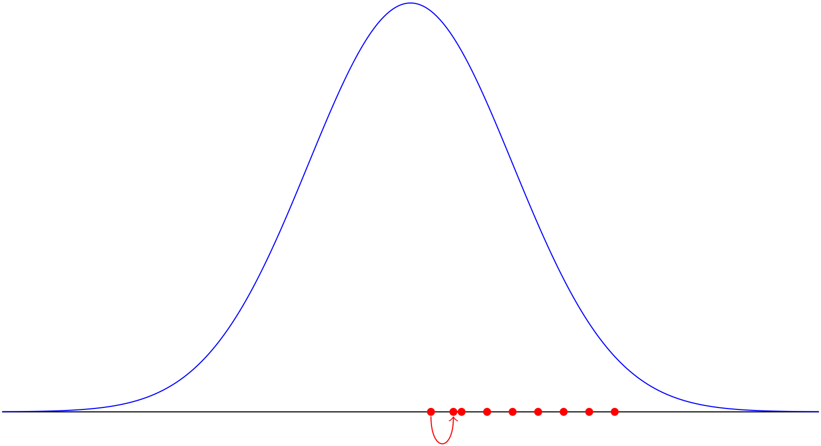
Sampling $\mathbb{P}(u|y)$



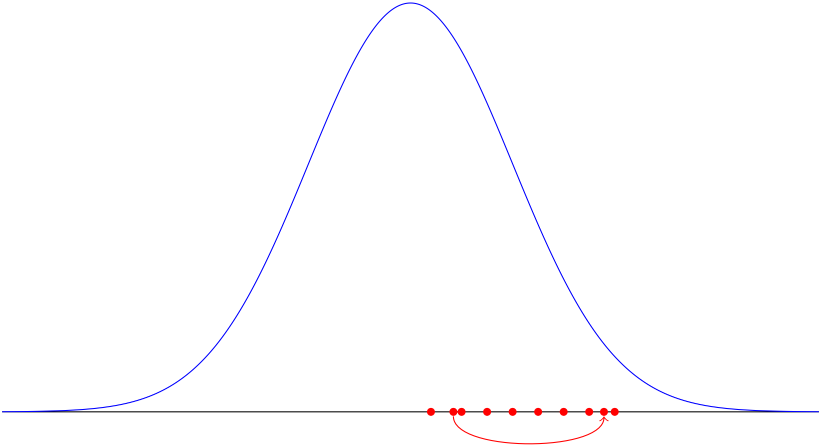
Sampling $\mathbb{P}(u|y)$



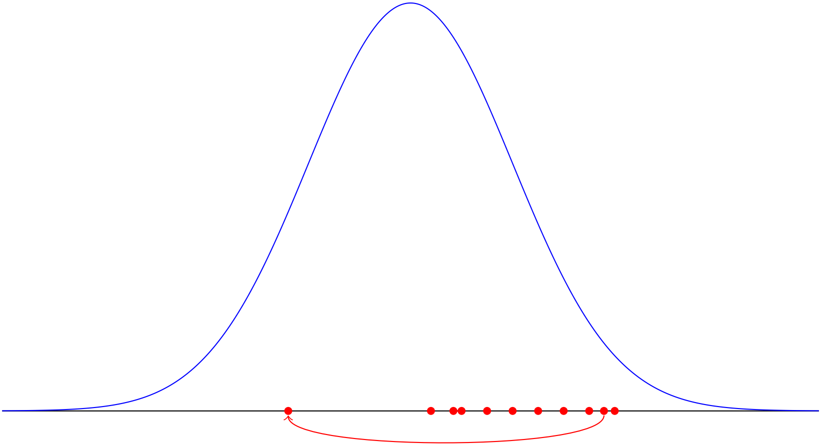
Sampling $\mathbb{P}(u|y)$



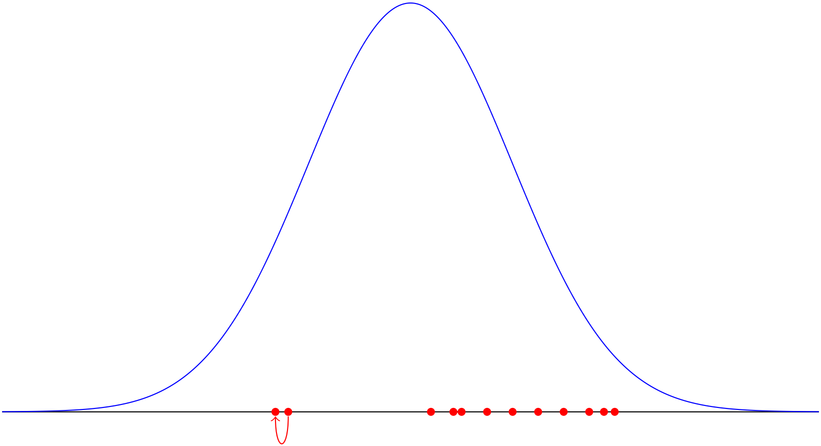
Sampling $\mathbb{P}(u|y)$



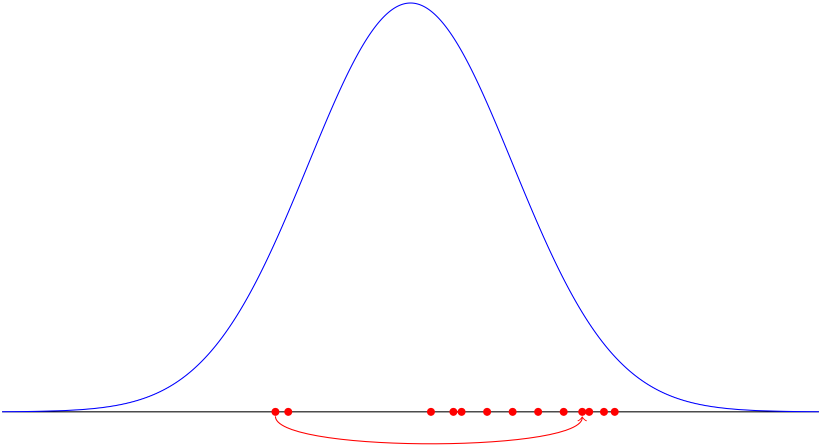
Sampling $\mathbb{P}(u|y)$



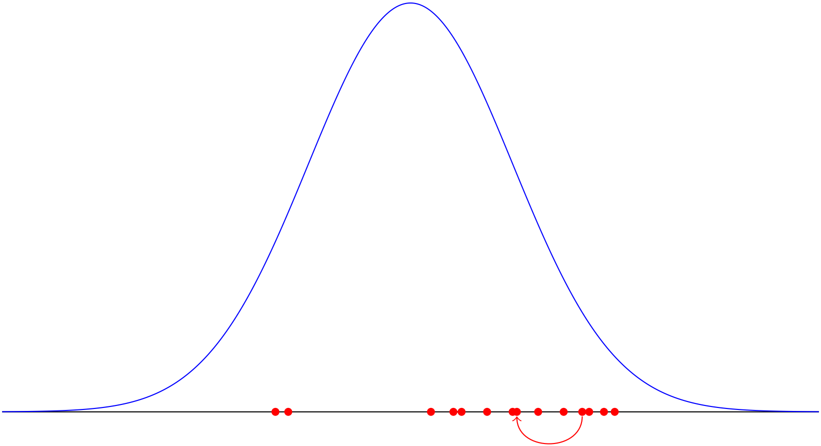
Sampling $\mathbb{P}(u|y)$



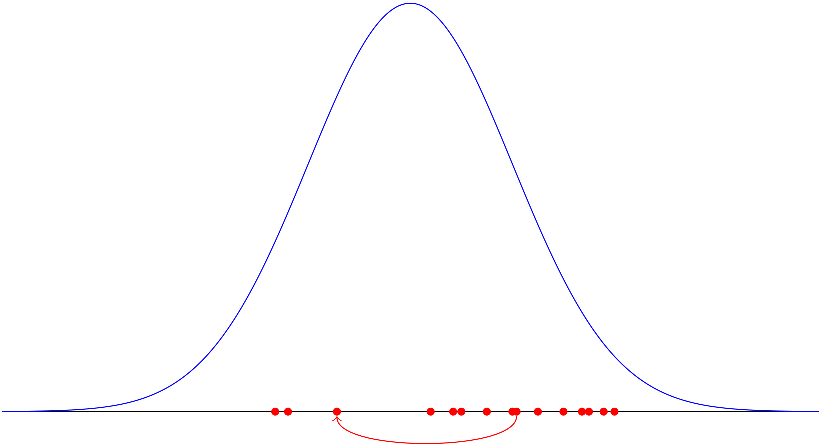
Sampling $\mathbb{P}(u|y)$



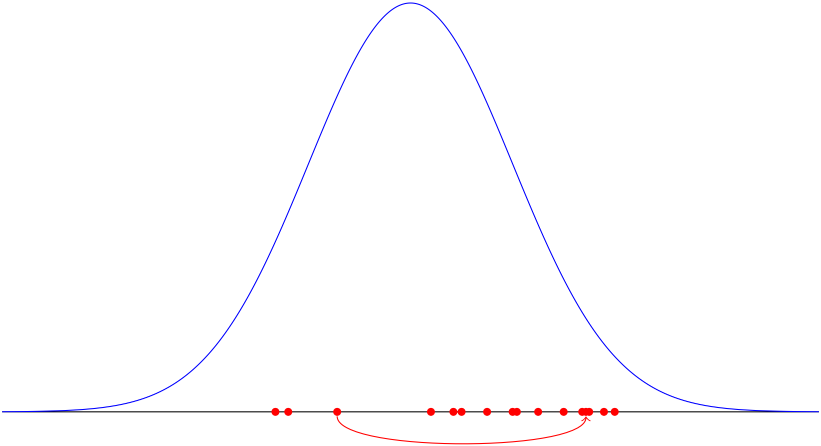
Sampling $\mathbb{P}(u|y)$



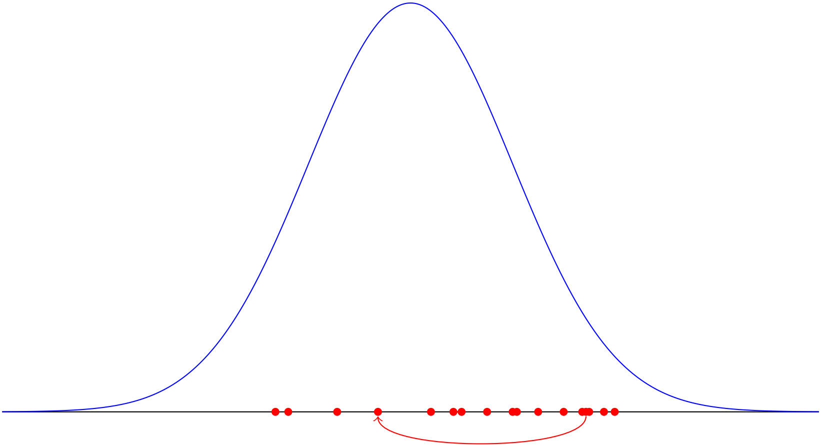
Sampling $\mathbb{P}(u|y)$



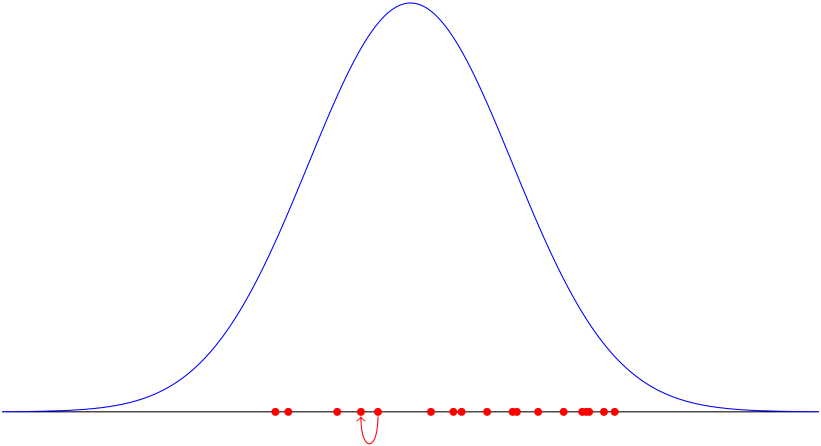
Sampling $\mathbb{P}(u|y)$



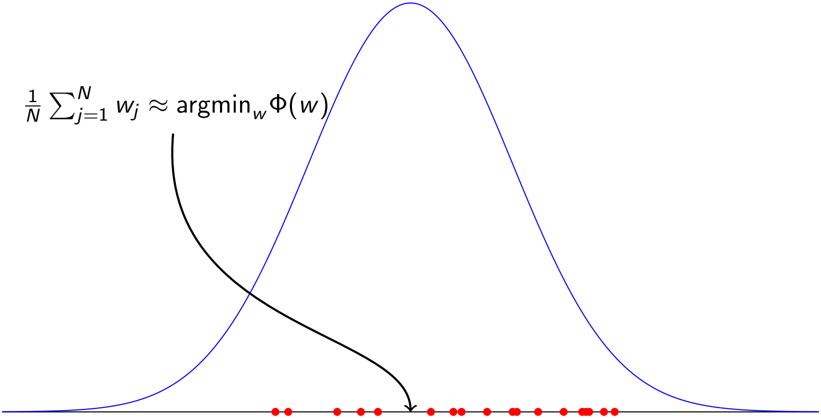
Sampling $\mathbb{P}(u|y)$



Sampling $\mathbb{P}(u|y)$



Sampling $\mathbb{P}(u|y)$



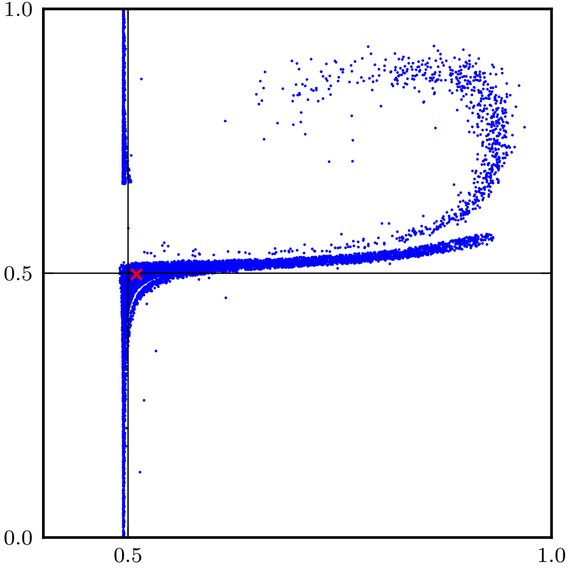
Sampling $\mathbb{P}(\underline{z}(0)|y)$

- Back to our model

$$
\begin{aligned}
&\text{(PDE)} \qquad \dot{w} := \begin{cases} \dot{u}_0 = 0 \\ \dot{u}_1 = v_1 \\ \dot{v}_1 = -u_1 - 2\pi h_1 \\ \dot{h}_1 = 2\pi v_1 \\ \dot{\underline{z}} = \underline{u}(\underline{z}(t), t) \end{cases} \\
&\text{(IC)} \qquad w(0) = (u_0(0), u_1(0), v_1(0), h_1(0), \underline{z}(0))
\end{aligned}
$$

- For demonstration, try to find $\underline{z}(0)$ and assume $u_0(0), u_1(0), v_1(0)$, and $h_1(0)$ are known.
- Make observations $y_k = \underline{z}(t_k) + \eta_k$.
- Now pretend we don't know $\underline{z}(0)$.
- See if we can recover $\underline{z}(0)$ from y by sampling $\mathbb{P}(\underline{z}(0)|y)$.

Sampling $\mathbb{P}(\underline{z}(0)|y)$



Video

Cool video 2

The glider problem

- Use autonomous gliders instead of passive tracers,

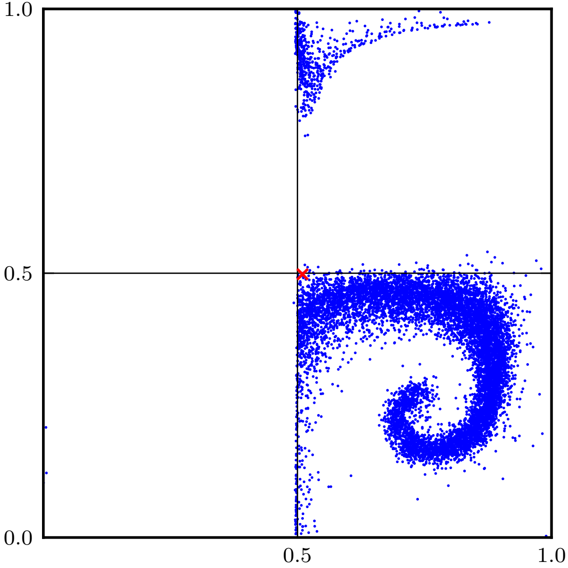
$$\dot{\underline{z}} = \underline{u}(\underline{z}(t), t) + \underline{f}(\underline{z}(t)).$$

- Force it to stay in the bottom-right cell. How? Good question.
- Observe

$$\underline{y_k} = \underline{z}(t_k) + \underline{\eta_k}$$

- What does the posterior look like?

The glider problem



Video

Cool video 3

Summary

I have told you

- what an inverse problem is;
- the link between inverse problems and DA for a toy ocean model;
- how to pose the problem in a Bayesian framework;
- how to solve the problem: optimisation;
- how to solve the problem: Markov chain Monte Carlo;
- a comparison between this model and a forced model.

Thank you