Basic concepts in particle filters

Javier Amezcua

NCEO/ECMWF training course

Data Assimilation Research Centre

University of Reading

Data assimilation

$\mathbf{x}^t \in \mathcal{R}^{N_x}$ **Model variables**

$\mathbf{y}^l \in \mathcal{R}^{N_y}$ **Observations**

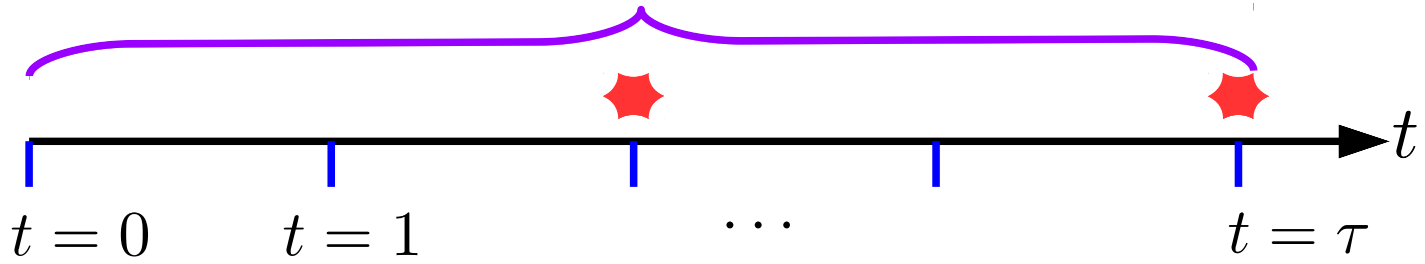
$$\mathbf{x}^t = m^{(t-1)\to t}\left(\mathbf{x}^{t-1}\right) + \boldsymbol{\beta}^t$$

$$\mathbf{y}^l = h^l\left(\mathbf{x}^{t=l}\right) + \boldsymbol{\eta}^l$$

$$\mathbf{x}^0 \; r.v., \; \mathbf{x}^0 \perp \boldsymbol{\beta}^t \perp \boldsymbol{\eta}^l$$

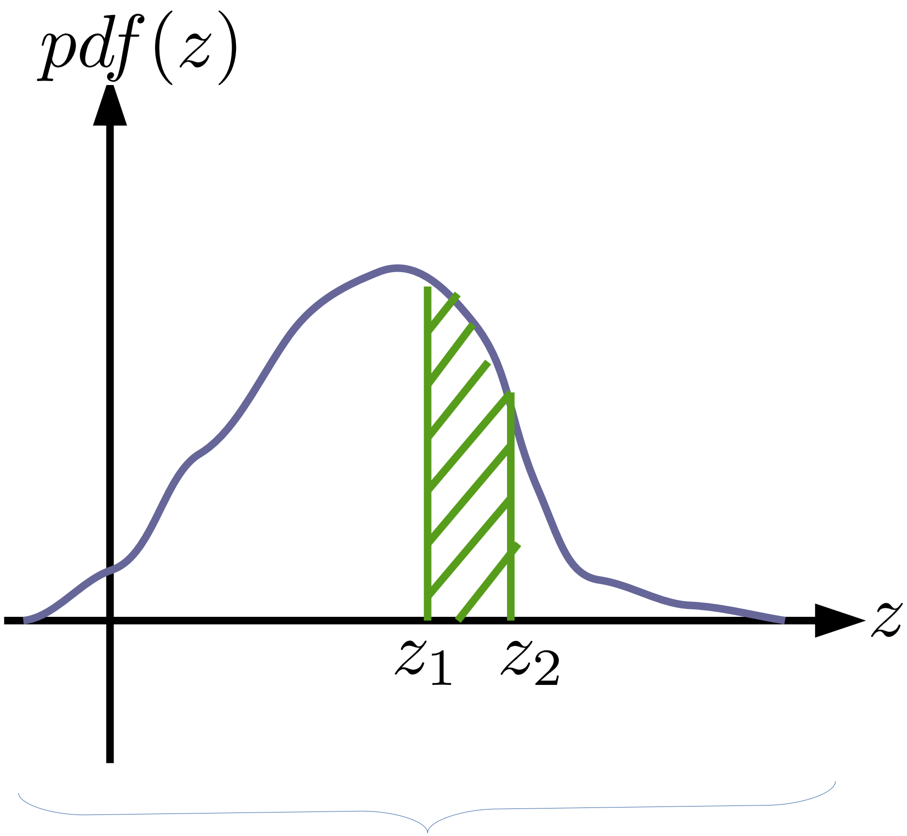
Let us work in the **discrete-time** world.

assimilation window: smoothing



To obtain the posterior pdf we can use **Bayes' theorem**.

Probability density functions



Statistical **support** of the variable

Probability density function:

$$pdf(z) \geq 0 \quad z \in \mathcal{R}^1$$

$$pdf(\mathbf{z}) \geq 0 \quad \mathbf{z} \in \mathcal{R}^N$$

Probability:

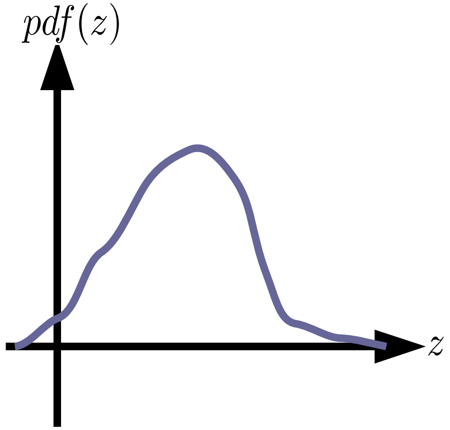
$$p\left(z_1 \leq z \leq z_2\right) = \int_{z_1}^{z_2} pdf(z)dz$$

$$0 \leq p\left(z_i \leq z \leq z_j\right) \leq 1$$

Cumulative density function

$$cdf(z_1) = \int_{-\infty}^{z_1} pdf(z)dz$$

Properties and operations



Normalization:

$$\int_{-\infty}^{+\infty} pdf(z)dz = 1$$

Expected value: **center of mass of the distribution (barycentre/centroid)**

$$\mu_z = E[z] = \int_{-\infty}^{+\infty} z pdf(z)dz$$

$$\mu_g(z) = E[g(z)] = \int_{-\infty}^{+\infty} g(z)pdf(z)dz$$

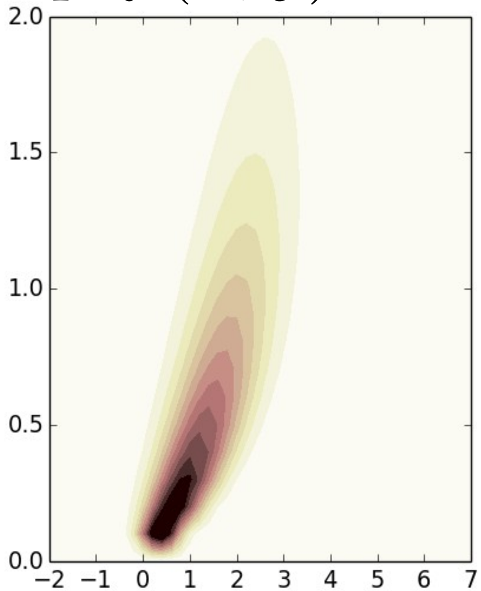
Variance: **mean quadratic deviation from the expected value**

$$\sigma_z^2 = Var[z] = \int_{-\infty}^{+\infty} (z - \mu_z)^2 pdf(z)dz$$

Joint, conditionals, marginals

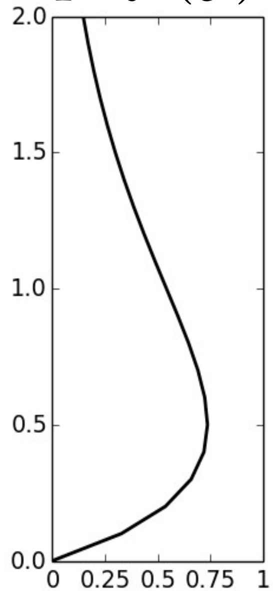
joint

$pdf(x,y)$



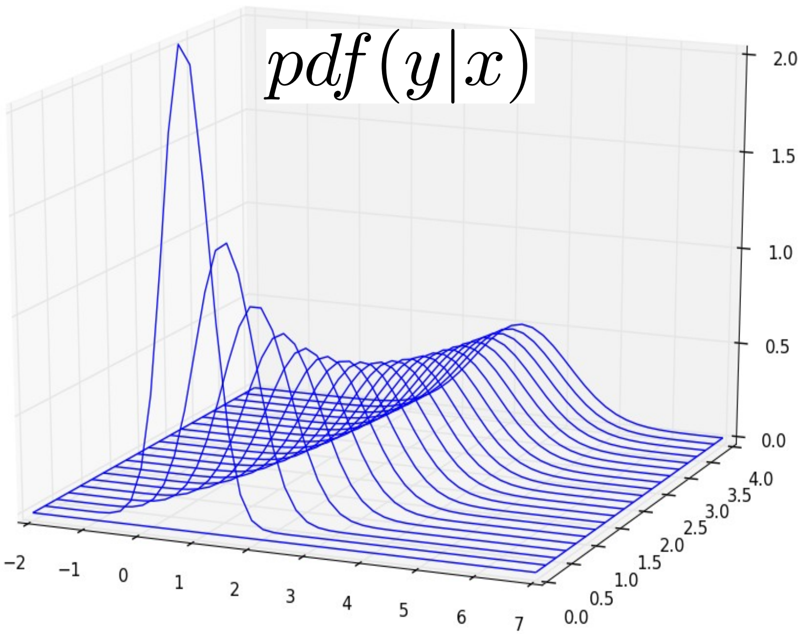
marginal

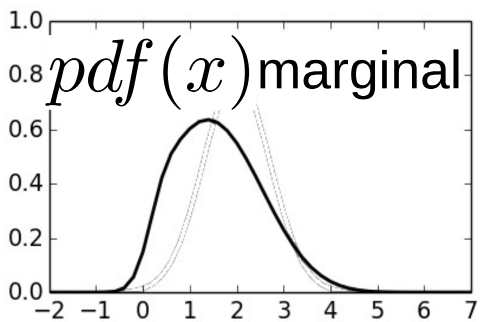
$pdf(y)$



y

conditional





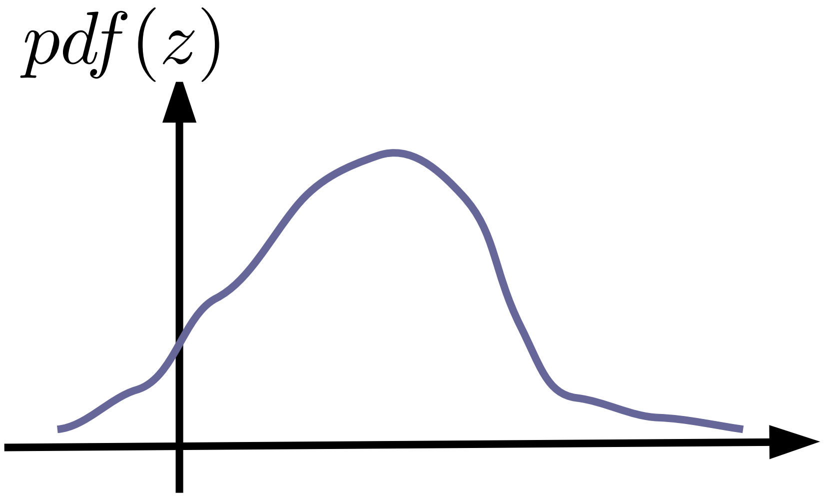
x

$$pdf(y|x) = \frac{pdf(x,y)}{pdf(x)}$$

$$pdf(y) = \int pdf(x,y)dx$$

Representing pdf's

Parametric representation

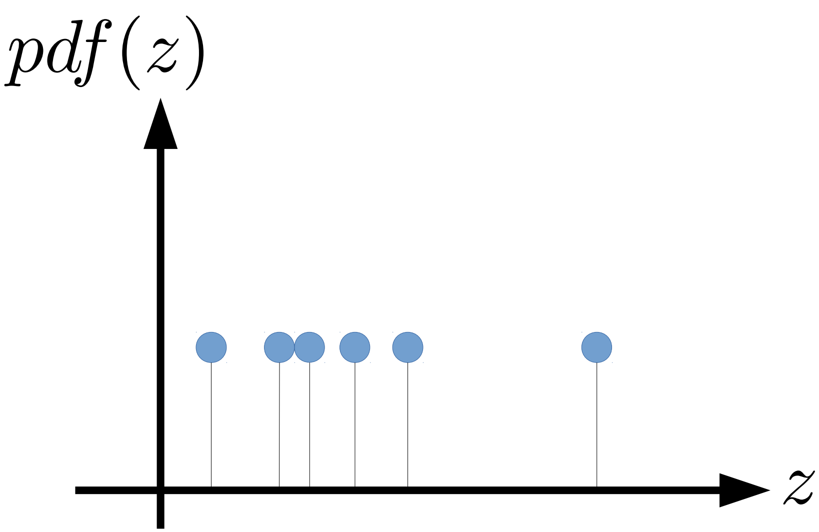


$$pdf(z;\theta)$$

Eg:

$$N(\mu,\sigma) \quad Exp(\theta) \quad \Gamma(\alpha,\beta)$$

Empirical or weak representation (sample)



weight

$$pdf(z) = \sum_{n_e=1}^{Ne} w_{n_e}\delta(z - z_{n_e})$$

Equiprobable sample:

$$pdf(z) = \frac{1}{N_e}\sum_{n_e=1}^{Ne} \delta(z - z_{n_e})$$

The Dirac delta function

$$pdf(z) = \frac{1}{N_e} \sum_{n_e=1}^{Ne} \delta(z - z_{n_e})$$

Properties:

$$\delta(z - z^*) = 0 \; \forall z \neq z^*$$

$$\int_{-\infty}^{+\infty} \delta(z - z^*) dz = 1$$

The Dirac delta 'kills' integrals:

$$\int_{-\infty}^{+\infty} g(z)\delta(z - z^*) dz = g(z^*)$$

Back to data assimilation

$\mathbf{x}^t \in \mathcal{R}^{N_x}$ **Model variables**

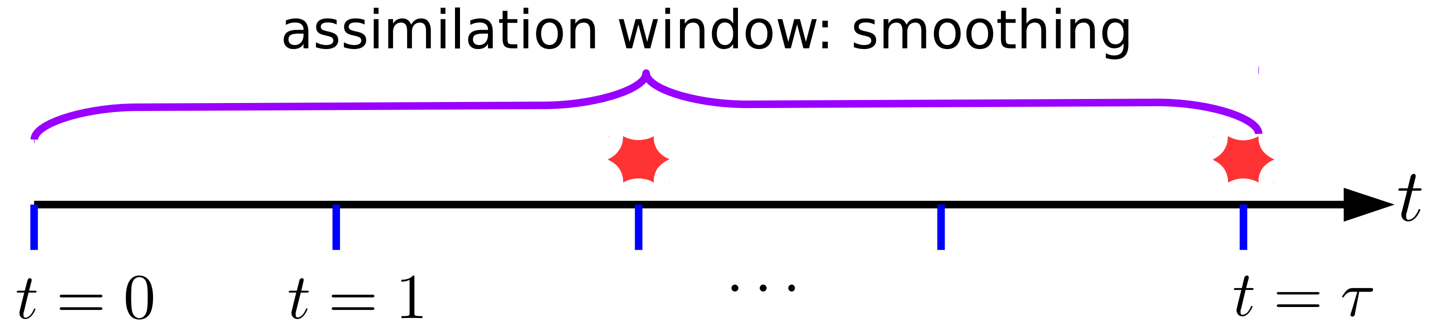
$\mathbf{y}^l \in \mathcal{R}^{N_y}$ **Observations**

$$\mathbf{x}^t = m^{(t-1)\to t}\left(\mathbf{x}^{t-1}\right) + \boldsymbol{\beta}^t$$

$$\mathbf{y}^l = h^l\left(\mathbf{x}^{t=l}\right) + \boldsymbol{\eta}^l$$

$$\mathbf{x}^0 \ r.v., \ \mathbf{x}^0 \perp \boldsymbol{\beta}^t \perp \boldsymbol{\eta}^l$$

Let us work in the **discrete-time Markovian** world.



To obtain the posterior pdf we can use **Bayes' theorem**.

Bayes theorem

Likelihood of the observations over the assimilation window.

Prior distribution of the state variable over the assimilation window.

$$p\left(\mathbf{x}^{0:\tau}|\mathbf{y}^{1,L}\right) = \frac{p\left(\mathbf{y}^{1:L}|\mathbf{x}^{0:\tau}\right) p\left(\mathbf{x}^{0:\tau}\right)}{p\left(\mathbf{y}^{1:L}\right)}$$

Marginal distribution of the **observations**.

Posterior probability distribution of the state variables given the observations over the assimilation window.

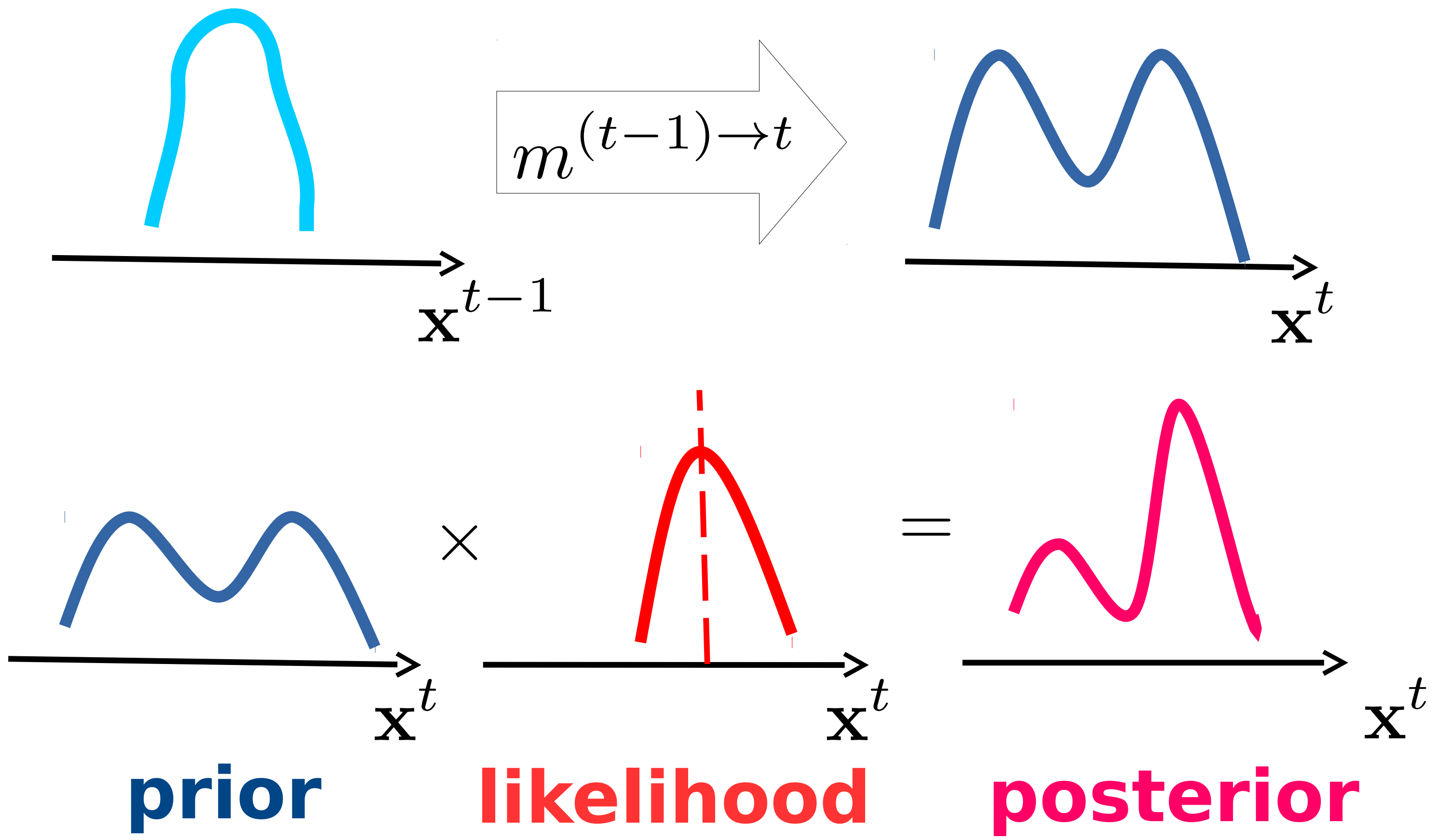
How to get these elements?

Working with pdf's

Consider the following 1-step scenario:

Forecast:

Analysis:



Solutions:

Forecast:

Continuous system:

- With model error (Wiener process): **Fokker-Plank equation**.
- Without model error: **Liouville equation**

Discrete system:

- With model error. Transition probabilities. **Chapman-Kolmogorov equation.**
- Without model error. **Chapman-Kolmogorov equation using Dirac deltas**.

Analysis:

Bayes Theorem.

The simplest particle filter

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\ dx}$$

Use ensemble

$$p(x) = \sum_{i=1}^{N} \frac{1}{N}\delta(x - x_i)$$

$$p(x|y) = \sum_{i=1}^{N} w_i \delta(x - x_i)$$

with

$$w_i = \frac{p(y|x_i)}{\sum_j p(y|x_j)}$$

the weights.

The weights

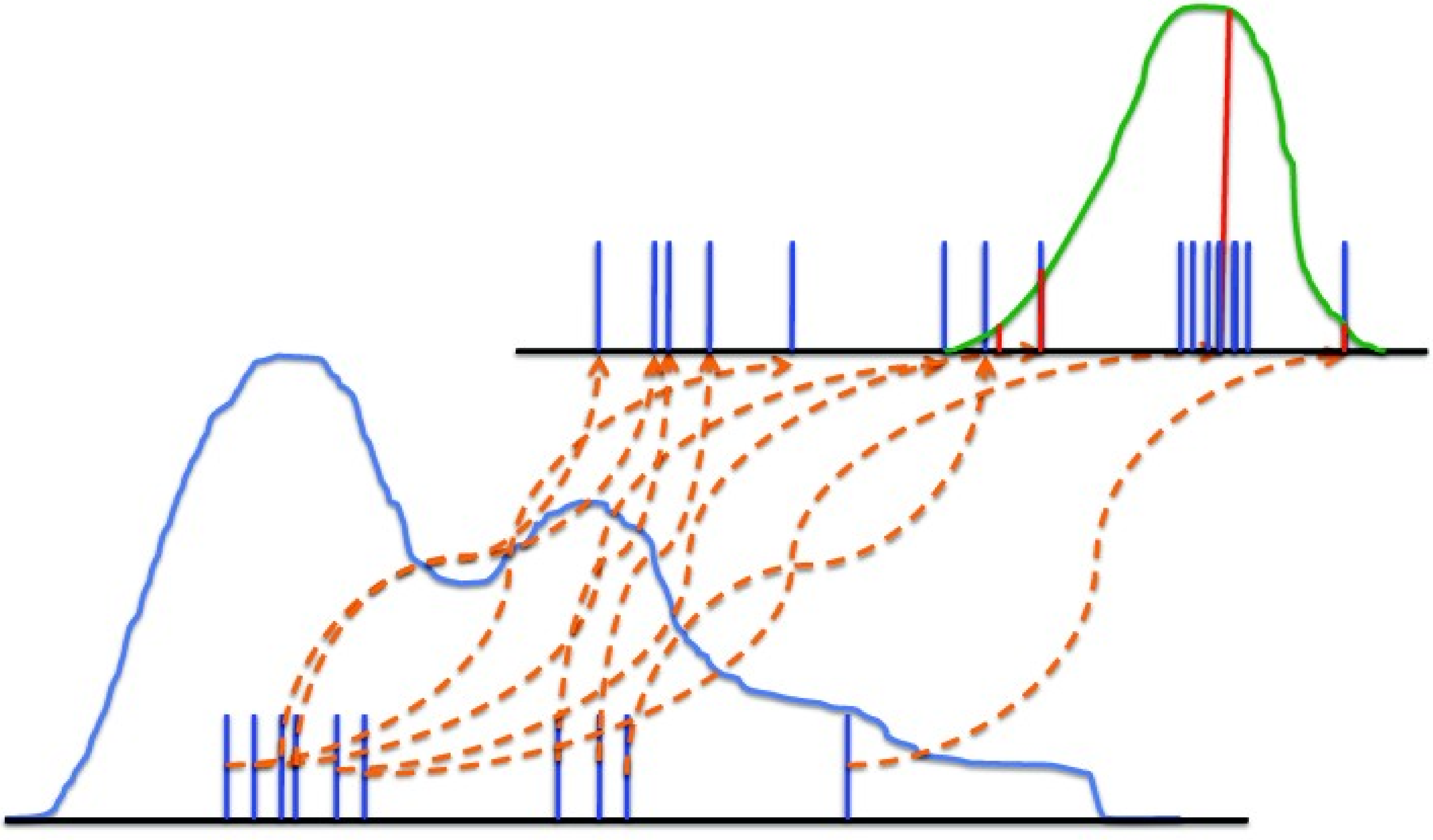
- The weight w_i is the normalised value of the pdf of the observations given model state x_i.
- For Gaussian distributed variables it is given by:

$$\begin{aligned} w_i &\propto p(y|x_i) \\ &\propto \exp\left[-\frac{1}{2}\left(y - H(x_i)\right)^T R^{-1}\left(y - H(x_i)\right)\right] \end{aligned}$$

- That is all !!!

We can use the weights to compute any sample statistic.

Weight degeneracy



Simple resampling

1. Put all weights after each other on the unit interval:

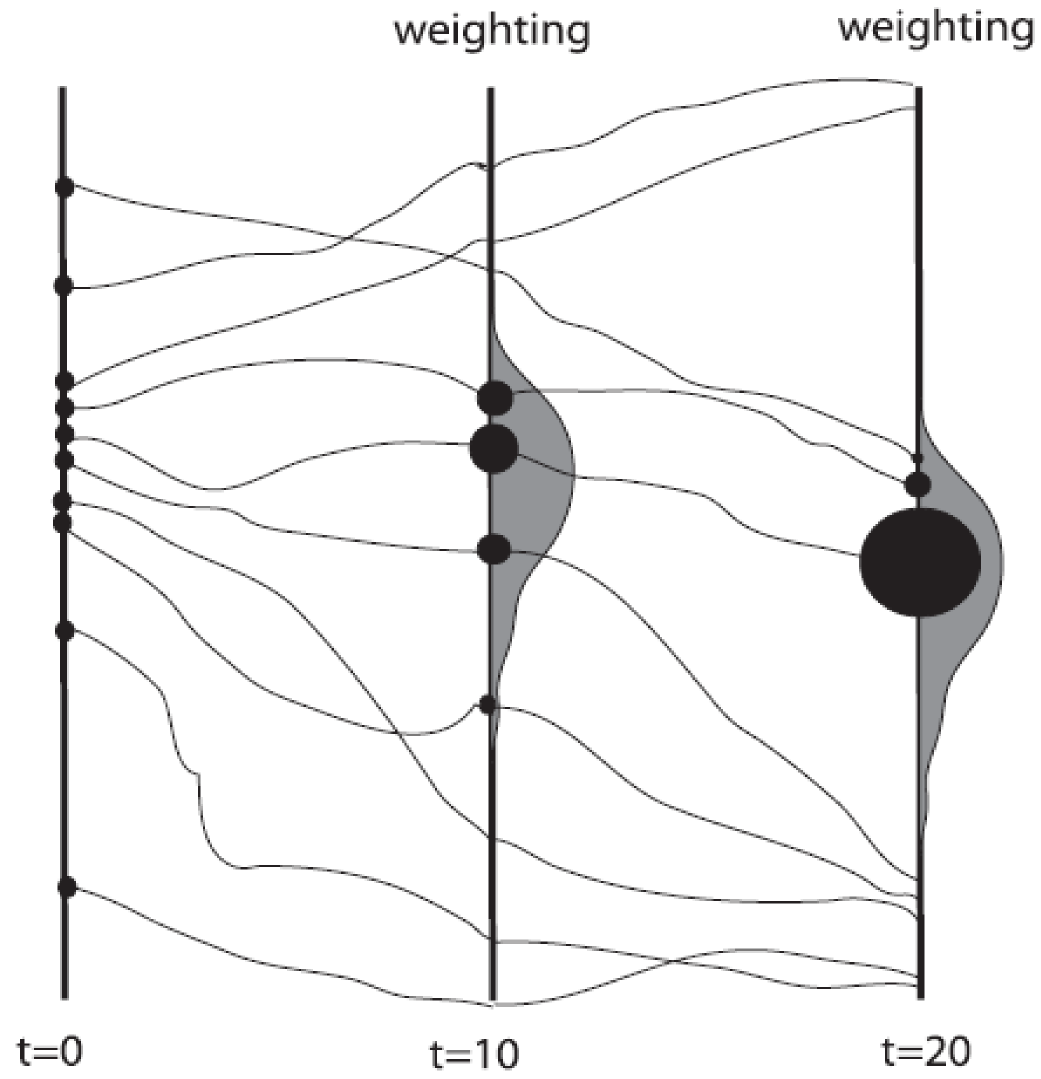
2. Draw a random number from the uniform distribution over [0,1/N], in this case with 10 members over [0,1/10].
3. Put that number on the unit interval: this points to the first member

4. Add 1/N to the end point: the new end point is our second member. Repeat this until N new members are obtained.

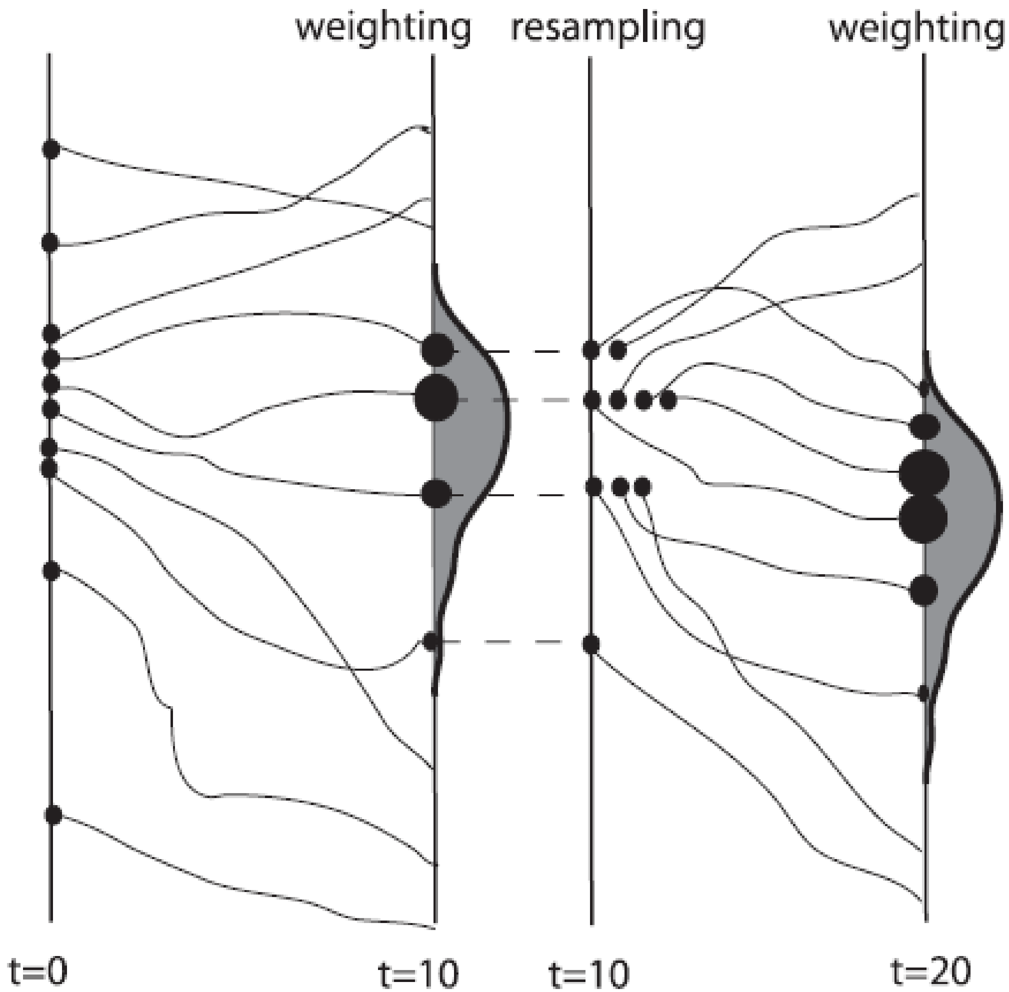
5. In our example we choose m1 2 times, m2 2 times, m3, m4, m5 2 times, m6 and m7.

Particle filtering in time

If we iterate this in time we need to weight every time there is an observation.



Simple Importance Resampling



Particle filters degenerate

$$N_e \propto exp[D_{eff}^2]$$

Snyder et al 2008

$$D_{eff} \propto N_{y,indep}$$

Ades and Van Leeuwen 2013

Imperfect models

$$p(\mathbf{x}^t|\mathbf{y}^{1:t}) = \frac{p(\mathbf{y}^t|\mathbf{x}^t)}{p(\mathbf{y}^t)} p(\mathbf{x}^t|\mathbf{y}^{1:t-1})$$

$$\frac{p(\mathbf{y}^t|\mathbf{x}^t)}{p(\mathbf{y}^t)} \int \underbrace{p(\mathbf{x}^t|\mathbf{x}^{t-1})}_{\text{transition}} \underbrace{p(\mathbf{x}^{t-1}|\mathbf{y}^{1:t-1})}_{\text{prior}} d\mathbf{x}^{t-1}$$

Giving a **particle representation to the prior**:

$$p(\mathbf{x}^{t-1}|\mathbf{y}^{1:t-1}) = \frac{1}{M} \sum_{m-1}^{M} \delta(\mathbf{x}^{t-1} - \mathbf{x}_m^{t-1})$$

Some assumptions

Conditional independence of the observations.

$$p\left(\mathbf{y}^{1:L}|\mathbf{x}^{0:\tau}\right) = \prod_{l=1}^{L} p\left(\mathbf{y}^{l}|\mathbf{x}^{t=l}\right)$$

Likelihoods at different observational times

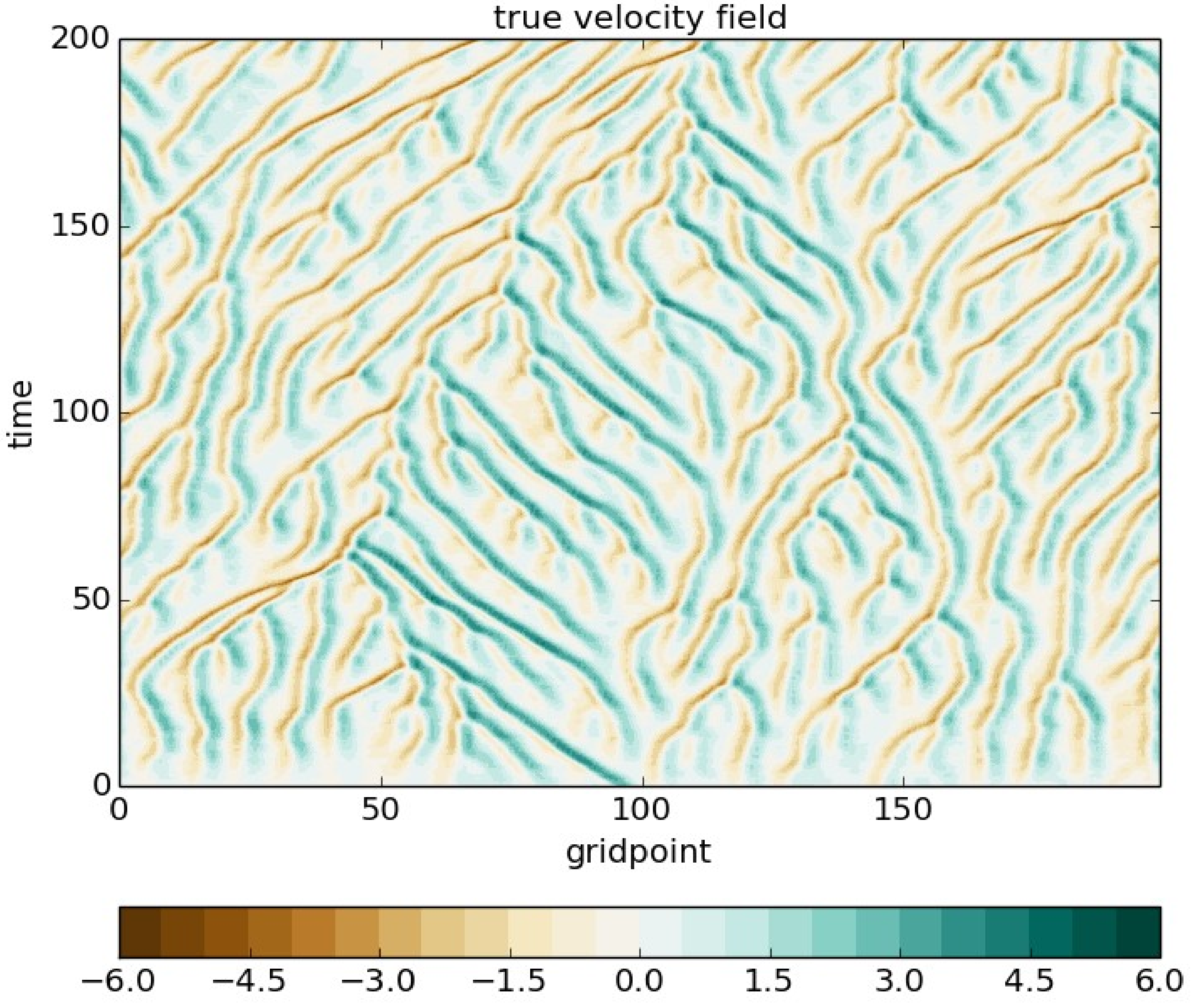
Markovian system. No memory.

$$p\left(\mathbf{x}^{0:\tau}\right) = \prod_{t=1}^{\tau} p(\mathbf{x}^{t}|\mathbf{x}^{t-1})\ p(\mathbf{x}^{0})$$

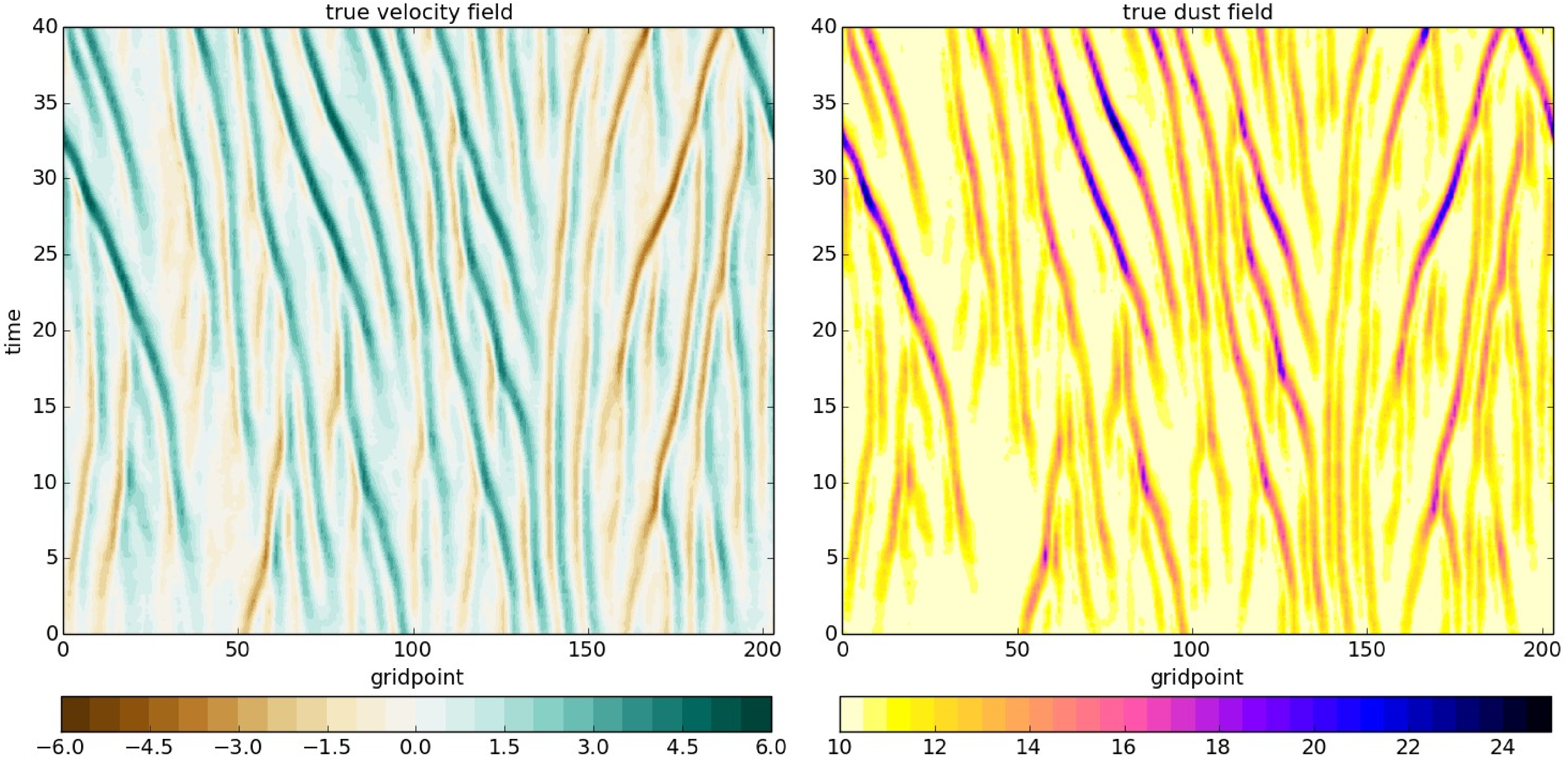
Transition densities

Prior at the beginning of window

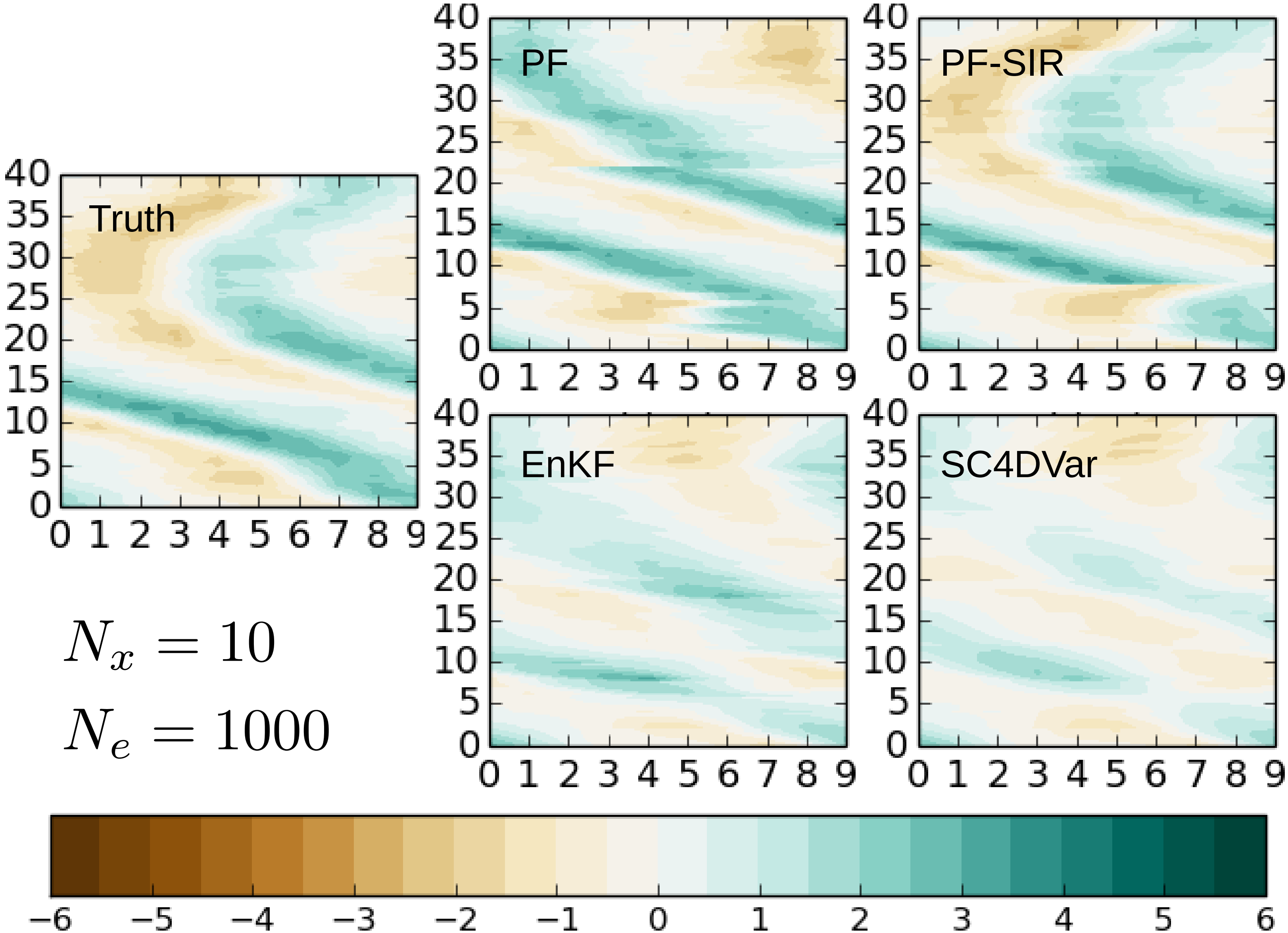
Example: Kuramoto-Shivashinski



Non-linear observations



Example: Kuramoto-Shivashinski



$$N_x = 10$$

$$N_e = 1000$$

Proposal densities

$$p(\mathbf{x}^t|\mathbf{y}^t) = \frac{1}{p(\mathbf{y}^t)} \frac{p(\mathbf{x}^t)p(\mathbf{y}^t|\mathbf{x}^t)}{q(\mathbf{x}^t)} q(\mathbf{x}^t) = \sum_{m=1}^{M} w_m \delta(\mathbf{x}^t - \mathbf{x}_m^t)$$

Particle representation

Point-wise evaluation

$$p(\mathbf{x}^t|\mathbf{y}^{1:t}) = \frac{1}{M} \frac{p(\mathbf{y}^t|\mathbf{x}^t)}{p(\mathbf{y}^t)} \sum_{m=1}^{M} \frac{p(\mathbf{x}^t|\mathbf{x}_m^{t-1})q(\mathbf{x}^t|\mathbf{x}_m^{t-1}, \mathbf{y}^t)}{q(\mathbf{x}^t|\mathbf{x}_m^{t-1}, \mathbf{y}^t)}$$

Proposal densities

Proposal densities:

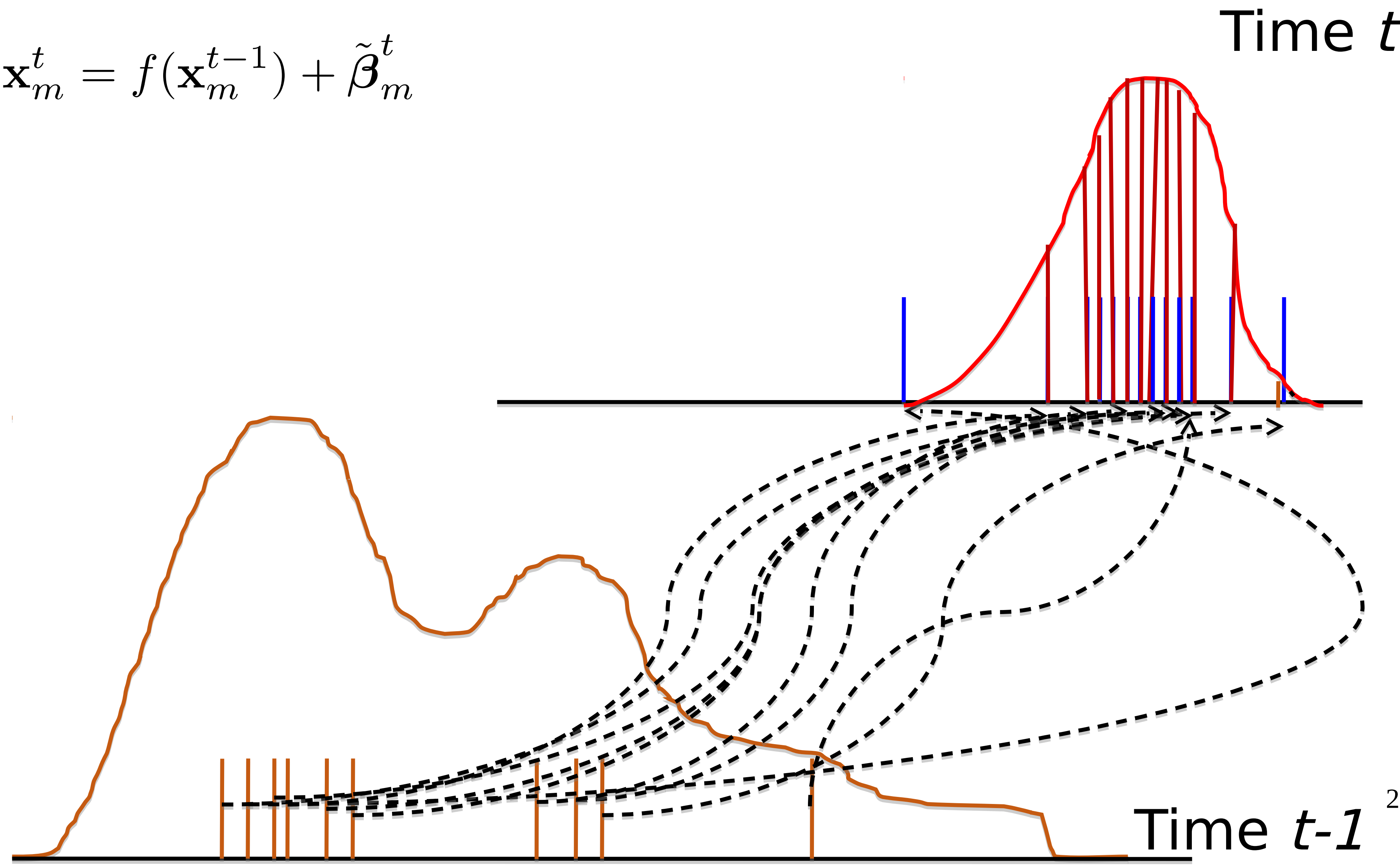
- Simple nudging (e.g. van Leeuwen 2010)

- Optimal proposal density

- **Equal-weight implicit sampling**

'Informed' particles

$$\mathbf{x}_m^t = f(\mathbf{x}_m^{t-1}) + \tilde{\boldsymbol{\beta}}_m^t$$



Optimal proposal density

$$p(\mathbf{x}^t|\mathbf{y}^{1:t}) = C\sum_{m=1}^{M}\frac{p(\mathbf{y}^t|\mathbf{x}_m^t)p(\mathbf{x}^t|\mathbf{x}_m^{t-1})}{q(\mathbf{x}^t|\mathbf{x}_m^{t-1},\mathbf{y}^t)}\delta(\mathbf{x}^t-\mathbf{x}_m^t)$$

When using the **optimal proposal density**:

$$q(\mathbf{x}^t|\mathbf{x}_m^{t-1},\mathbf{y}^t) = p(\mathbf{x}^t|\mathbf{x}_m^{t-1},\mathbf{y}^t)$$

The problem reduces to:

$$p(\mathbf{x}^t|\mathbf{y}^{1:t}) = C\sum_{m=1}^{M}p(\mathbf{y}^t|\mathbf{x}_{m-1}^t)\delta(\mathbf{x}^t-\mathbf{x}_m^t)$$

But it is **not** that **simple** to **sample** from the **OPD**.

Targeting

Problem: inspite of the use of proposal densities **degeneracy can occur**. However, for a given particle the proposal only depends on that particle.

$$p(\mathbf{x}^t|\mathbf{y}^{1:t}) = C \sum_{m=1}^{M} \frac{p(\mathbf{y}^t|\mathbf{x}^t_m)p(\mathbf{x}^t|\mathbf{x}^{t-1}_m)}{q(\mathbf{x}^t|\mathbf{x}^{t-1}_m, \mathbf{y}^t)} \delta(\mathbf{x}^t - \mathbf{x}^t_m)$$

Solution: force **equal* weights**. For each particle, use a **proposal density that depends on the whole sample**.

$$p(\mathbf{x}^t|\mathbf{y}^{1:t}) = C \sum_{m=1}^{M} \frac{p(\mathbf{y}^t|\mathbf{x}^t_m)p(\mathbf{x}^t|\mathbf{x}^{t-1}_m)}{q(\mathbf{x}^t|\underline{\{\mathbf{x}^{t-1}_m\}_m}, \mathbf{y}^t)} \delta(\mathbf{x}^t - \mathbf{x}^t_m)$$