Data assimilation in high-dimensional highly nonlinear systems

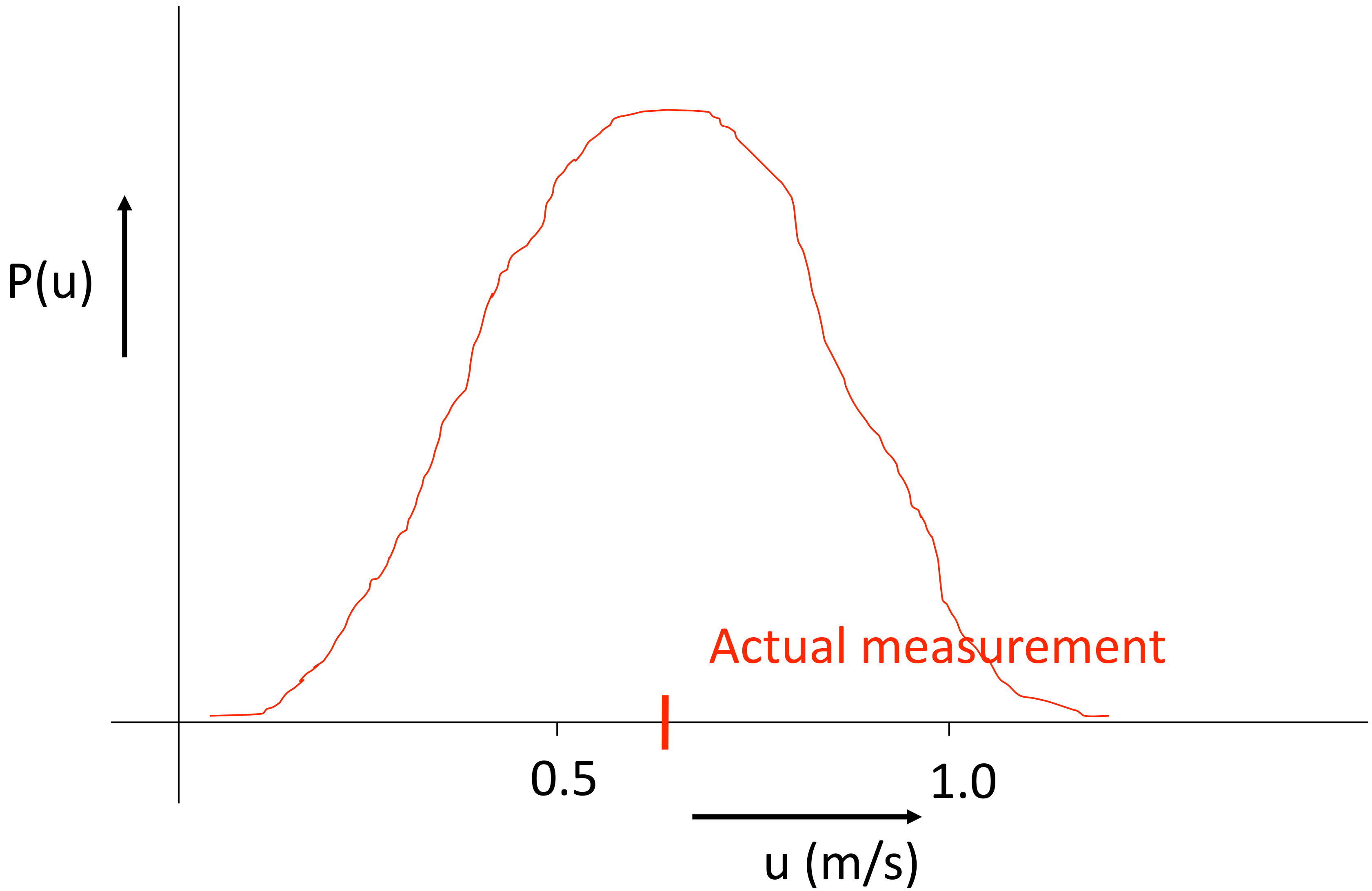
Peter Jan van Leeuwen

Data-Assimilation Research Centre

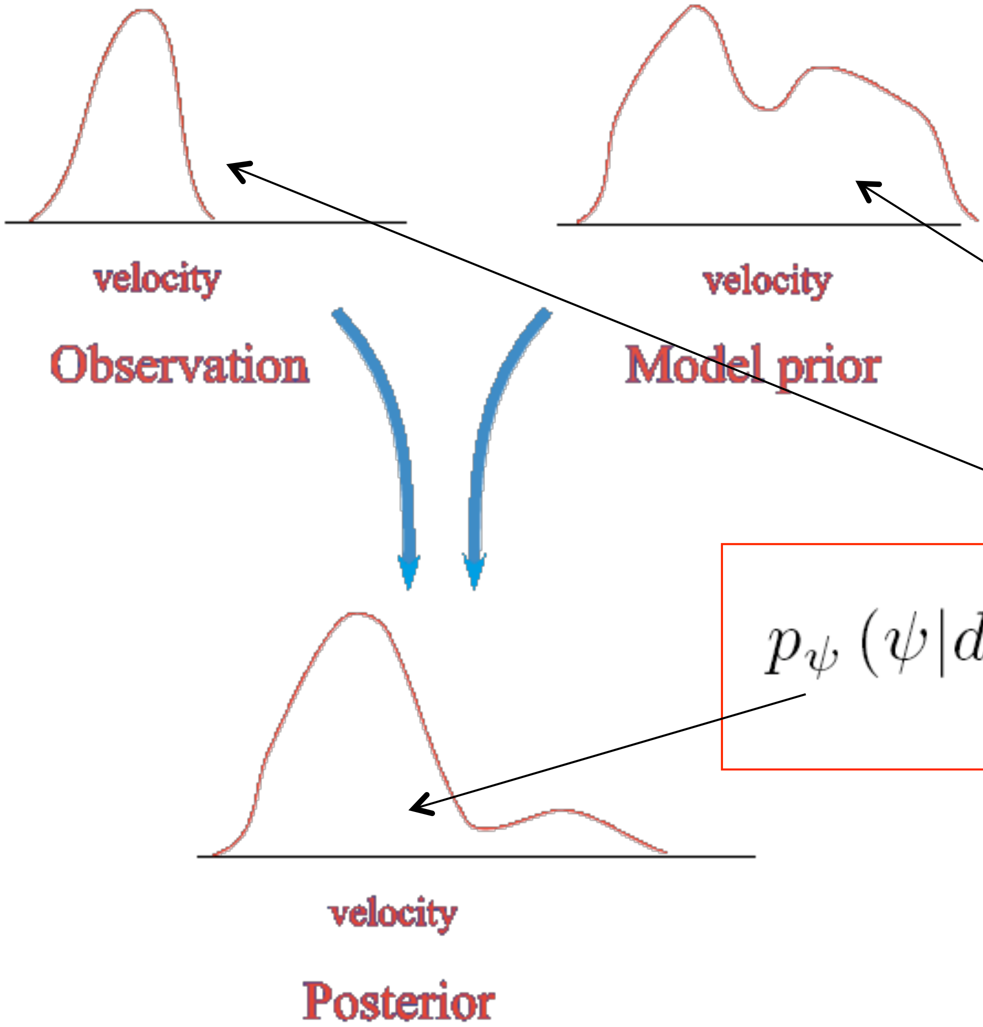
University of Reading

NOTE: THIS PRESENTATION CONTAINS SOME NEW MATERIAL NOT PUBLISHED YET. PLEASE LET ME KNOW IF YOU WANT TO USE IT.

The basics: probability densities



Data assimilation: general formulation



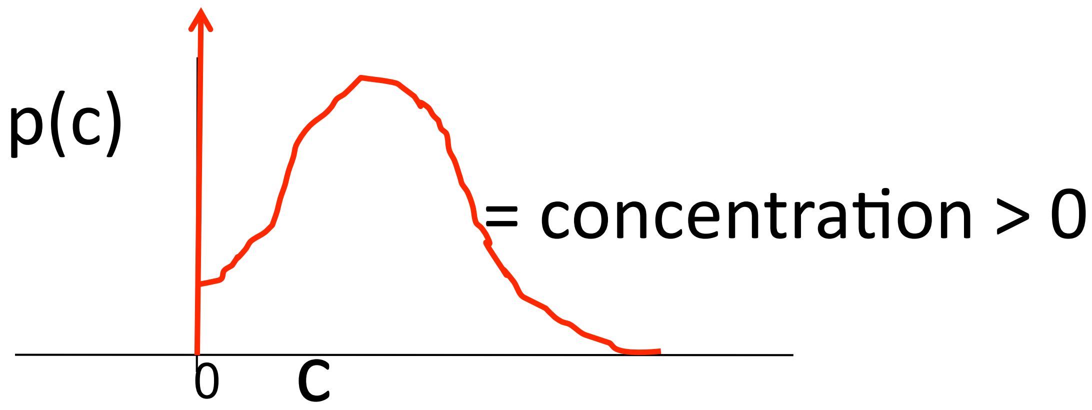
Bayes theorem:

$$p_\psi\left(\psi|d\right) = \frac{p_d\left(d|\psi\right) p_\psi\left(\psi\right)}{\int p_d\left(d|\psi\right) p_\psi\left(\psi\right)\ d\psi}$$

NO INVERSION !!!

Challenges in applying Bayes theorem

- Observation errors? (Gaussian or not? Correlations?)



- Relation observations and model variables?

(e.g. Radiation at one wavelength is influenced by several processes, so several different model variables)

Challenges in applying Bayes theorem

- Model errors? (Size/shape/biases)

  Should come from neglected physics, but what about neglected turbulence???

- Present-day supercomputers are too small

  (For NWP: joint pdf of 10,000,000 variables so about $100^{10,000,000}$ real numbers)

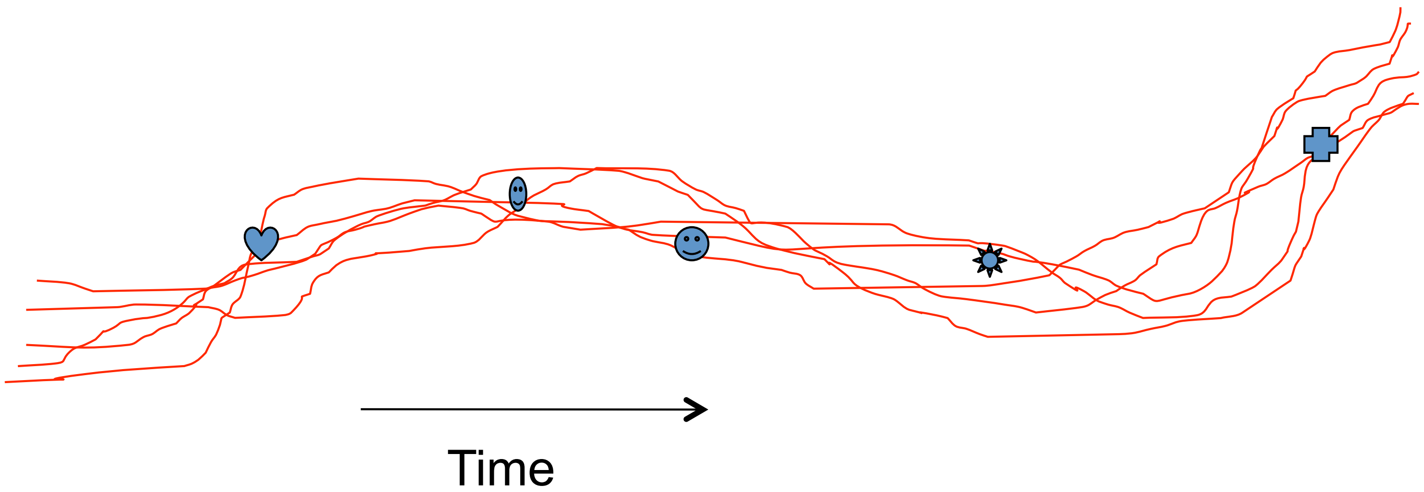
- Present-day supercomputers are too slow

Where are we today?

- All present-day data-assimilation systems are based on linearizations:
- (Ensemble) Kalman filter: assumes Gaussian pdf's
- 4DVar: assumes Gaussian pdf for initial condition and observations (no model errors)
- Representer method: as 4DVar but with Gaussian model errors
- Combinations of these

Where do we want to go?

- Represent pdf by an ensemble of model states
- Fully nonlinear



How do we get there? Particle filter?

$$p_\psi\left(\psi|d\right) = \frac{p_d\left(d|\psi\right)p_\psi\left(\psi\right)}{\int p_d\left(d|\psi\right)p_\psi\left(\psi\right)\ d\psi}$$

↓ Use ensemble

$$p(\psi) = \sum_i \frac{1}{N}\delta(\psi - \psi_i)$$

$$p_\psi(\psi|d) = \sum_i w_i \delta(\psi - \psi_i)$$

with

$$w_i = \frac{p_d\left(d|\psi_i\right)}{\sum_i p_d\left(d|\psi_i\right)}$$

What are these weights?

- The weight *w_i* is the pdf of the observations given the model state *i*.
- For *M* independent Gaussian distributed observation errors:

$$w_i \propto \prod_{j=1}^{M} \exp\left[-\frac{1}{2}\frac{\left(d_j - H_j(\psi_i)\right)^2}{\sigma_j^2}\right]$$

Particle Filter degeneracy: resampling

- With each new set of observations the old weights are multiplied with the new weights.
- Very soon only one particle has all the weight…
- Solution:

  Resampling: duplicate high-weight particles are abandon low-weight particles

Particle filter degeneracy (cont'd)

For large-dimensional systems with lots of independent observations the weights vary too much:

$$w_i \propto \prod_{j=1}^{M} \exp\left[-\frac{1}{2}\frac{\left(d_j - H_j(\psi_i)\right)^2}{\sigma_j^2}\right]$$

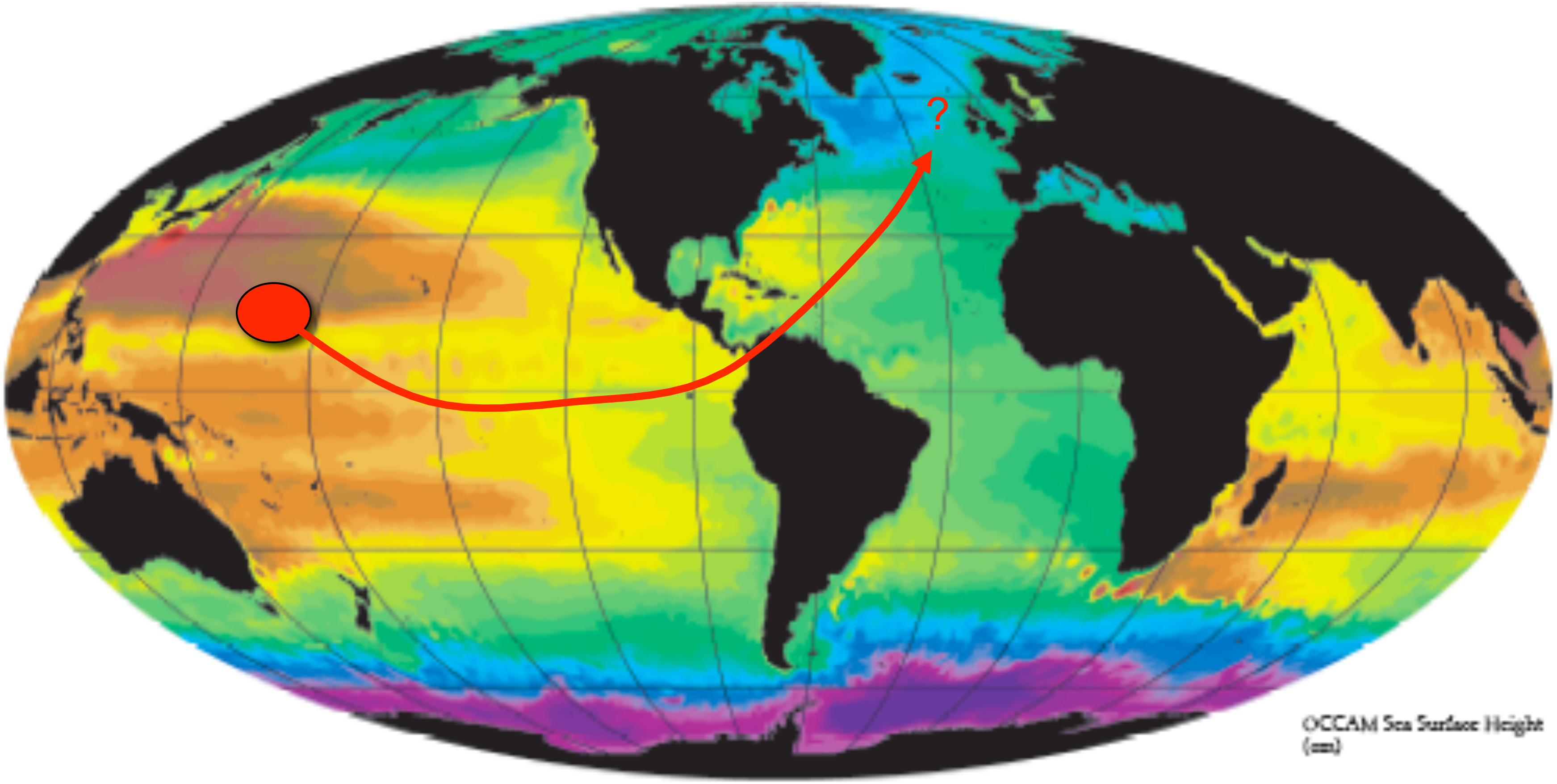
A small difference in the one-observation weights is raised to power M.

(w(1)_1=0.1 w(1)_2=0.09 M=100 w_1/w_2=37648)

Potential solutions

- Increase effective ensemble size by e.g. localization. (Not easy due to resampling)

Solution used in Ensemble Kalman Filter



Local updating

Potential solutions

- Increase effective ensemble size by e.g. localization. (Not easy due to resampling)
- Particles should explore the ‘local attractor’ more efficiently.

Experiment Lorenz 1963

$$
\begin{aligned}
dx &= \sigma(y - z)dt + d\beta_x \\
dy &= (\rho x - xz - y)\, dt + d\beta_y \\
dz &= (xy - \beta z)\, dt + d\beta_z
\end{aligned}
$$

Model parameters:

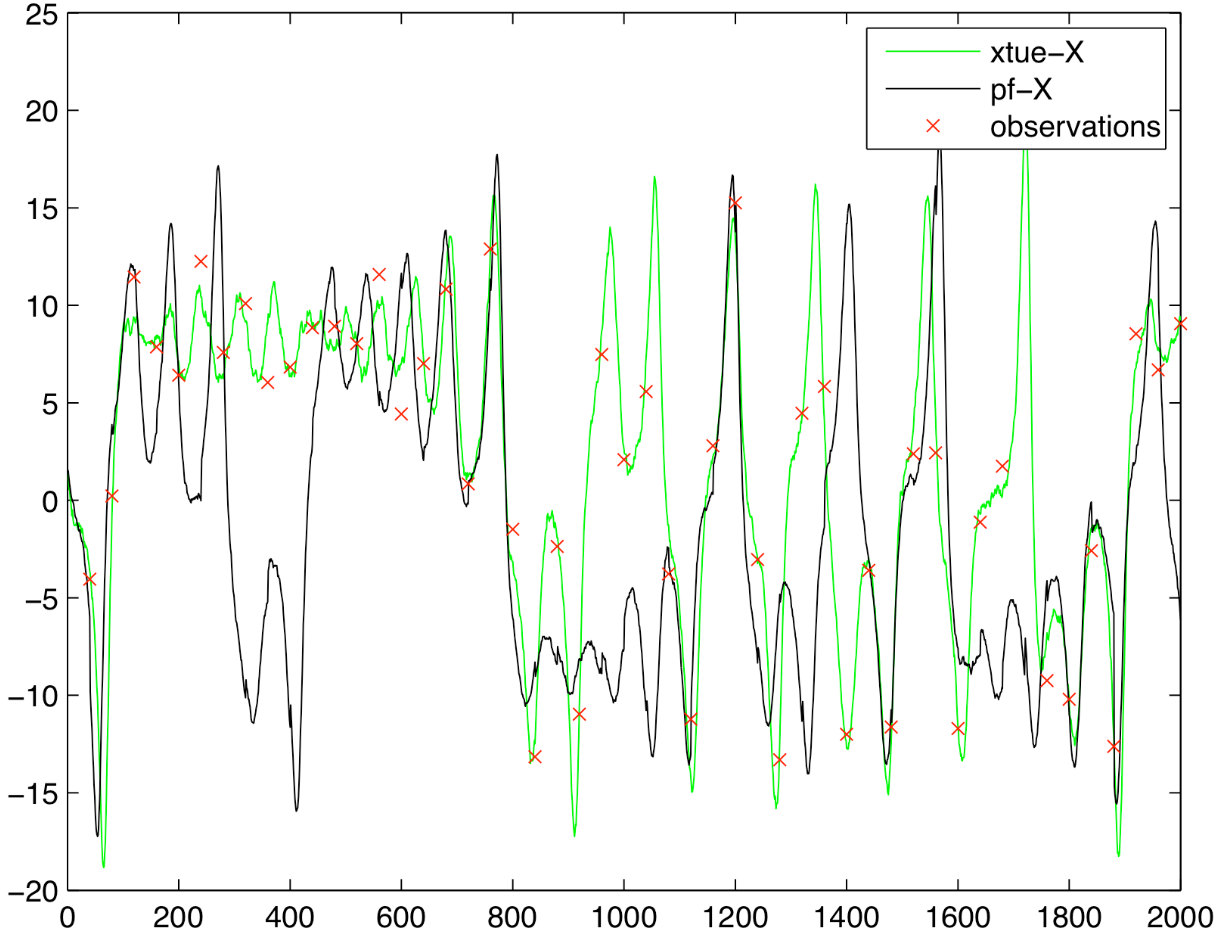
$$dt = 0.01 \quad \Delta t = 40dt \quad \rho = 28 \quad \sigma = 10 \quad \beta = 8/3$$

Statistical parameters:

$$\sigma_{obs} = \sqrt{2} \quad \sigma_{d\beta} = \sqrt{2} \quad \sigma_{init} = \sqrt{2}$$

Measure X only

Particle filter with resampling 20 particles



Particle filter with proposal density

Stochastic model
$$d\psi = f(\psi)dt + d\beta$$

Proposed stochastic model:

$$d\psi = f(\psi)dt + d\beta' - K(d^n - H(\psi))$$

Leads to particle filter with weights

$$w_i = \frac{p(d^n|\psi_i^n)}{\sum_i p(d^n|\psi_j^n)} \frac{p(\psi_i^n|\psi_i^{n-1})}{q(\psi_i^n|\psi_i^{n-1} d^n)}$$

Meaning of the transition densities

$$p(\psi_i^n|\psi_i^{n-1}) = p(d\beta_i)$$

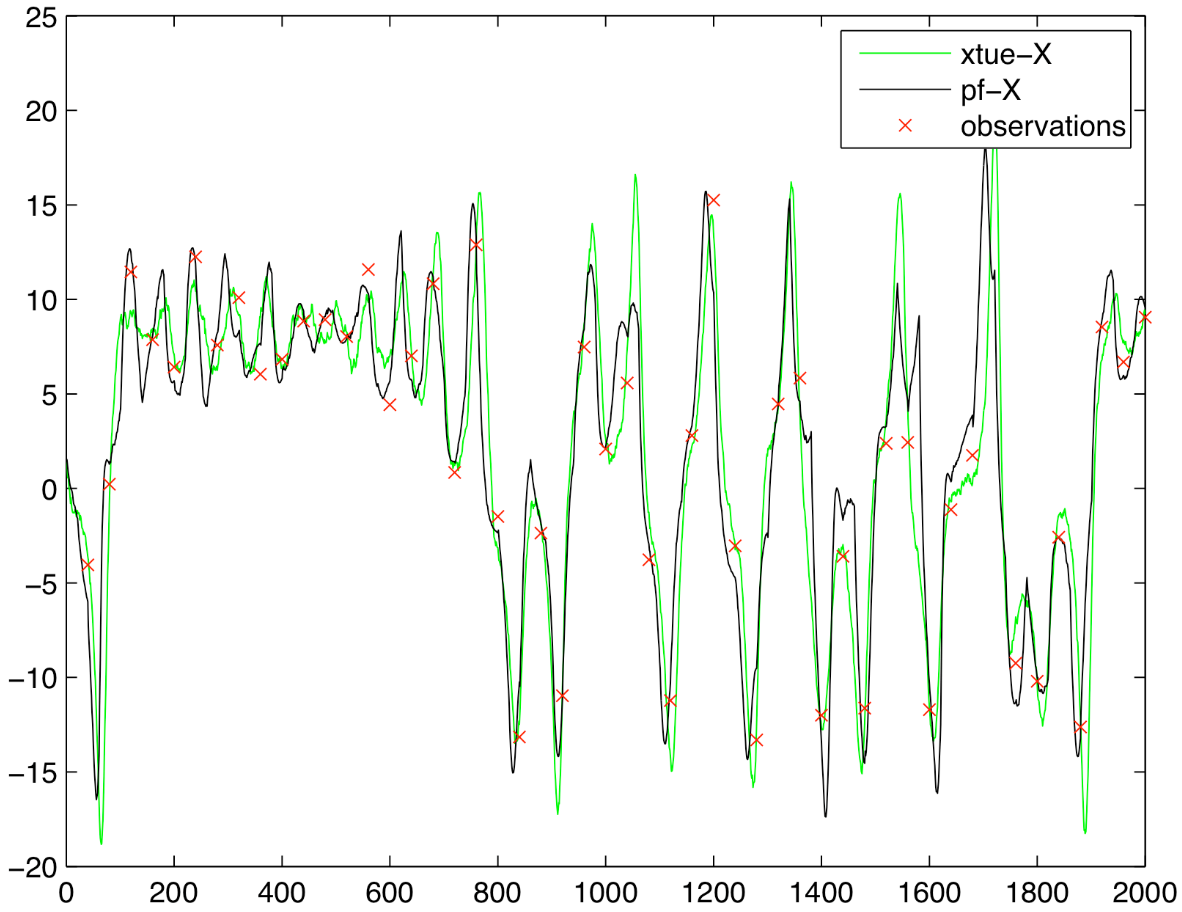
= the probability of this specific value for the model error.

For Gaussian model errors we find:

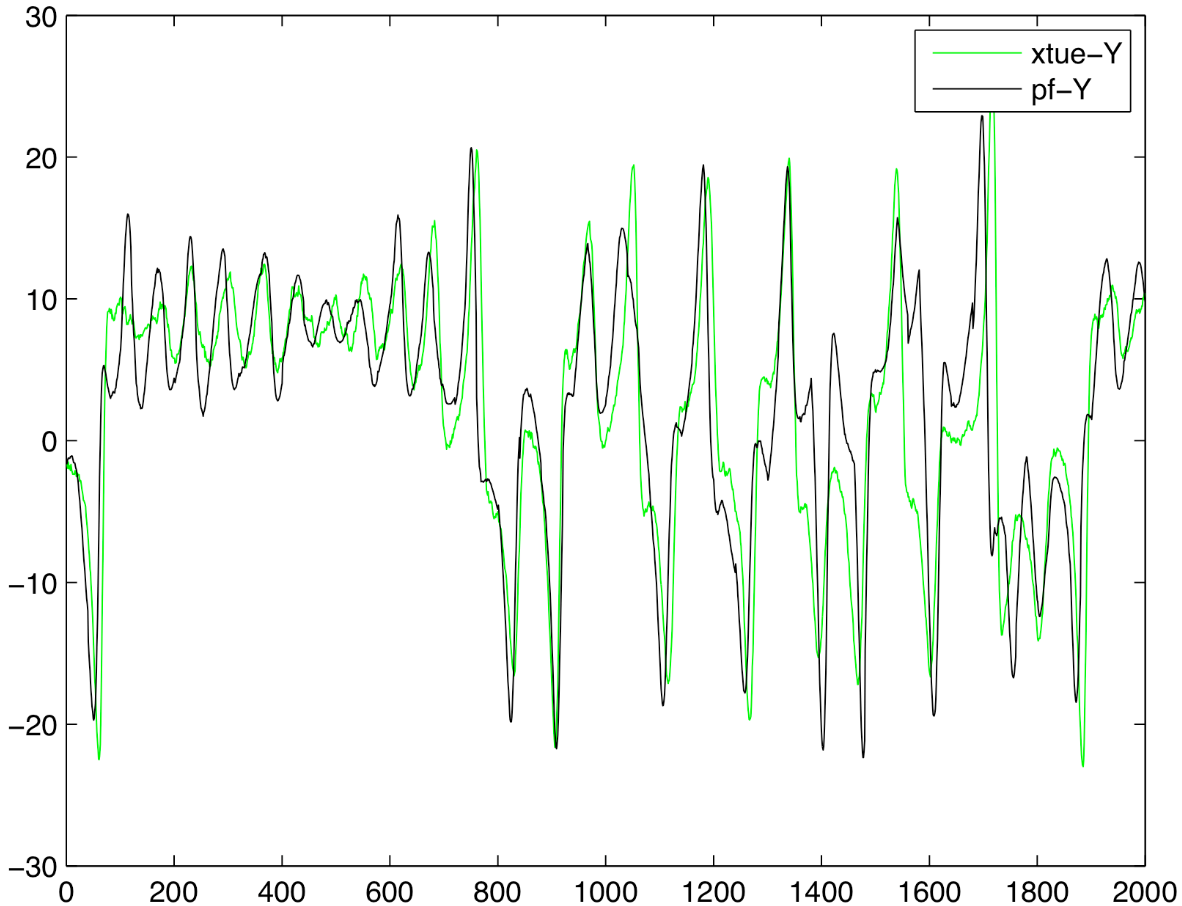
$$\propto \exp\left[-\frac{1}{2}\left(\psi_i^n - \psi_i^{n-1} + f(\psi_i^{n-1})dt\right) Q^{-1} \left(\psi_i^n - \psi_i^{n-1} + f(\psi_i^{n-1})dt\right)\right]$$

A similar expression is found for the proposal transition

Particle filter with proposal density
3 particles X variable



Particle filter with proposal density 3 particles, Y variable (not observed)



Remarks

- Note the freedom in the proposal density!
- Some localization through the model error covariance.
- It is possible to chose the proposal density such that all posterior weight equal.

Equal weights

The weights can be written as:

$$\begin{aligned} w_i &\propto \exp\left[-\frac{1}{2}\left(\psi^n - \psi^{n-1} - f(\psi^{n-1})dt\right) Q^{-1} \left(\psi^n - \psi^{n-1} - f(\psi^{n-1})dt\right)\right. \\ &+ \frac{1}{2}\left(\psi^n - \psi^{n-1} - f(\psi^{n-1})dt - K(d - H(\psi^{n-1}))\right) Q^{-1} \\ &* \left(\psi^n - \psi^{n-1} - f(\psi^{n-1})dt - K(d - H(\psi^{n-1}))\right) \\ &- \left.\frac{1}{2}(d - H(\psi^n))R^{-1}(d - H(\psi^n))\right] \end{aligned}$$

or

$$\boxed{\log(w_i) = \psi_i^n Q^{-1} \psi_i^n + \alpha_i^{n-1}\psi_i^n + \beta_i^{n-1} - \frac{1}{2}(d^n - H(\psi_i^n))R^{-1}(d^n - H(\psi_i^n))}$$

This quadratic equation has an infinite number of solutions.
We just have to find one …

Equal weights (cont'd)

Take the weight as small as possible:

Minimizing the weights as function of the state at *n* gives:

$$Q^{-1}\psi_i^n + \alpha_i^{n-1} - \frac{\partial H(\psi_i^n)}{\partial \psi_i^n} R^{-1}(d^n - H(\psi_i^n)) = 0$$

Which, for linear measurement operators reduces to:

$$\psi_i^n = (1 - QH^T[HQH^T + R]^{-1})(Q\alpha_i^{n-1} + QH^T R^{-1} d)$$

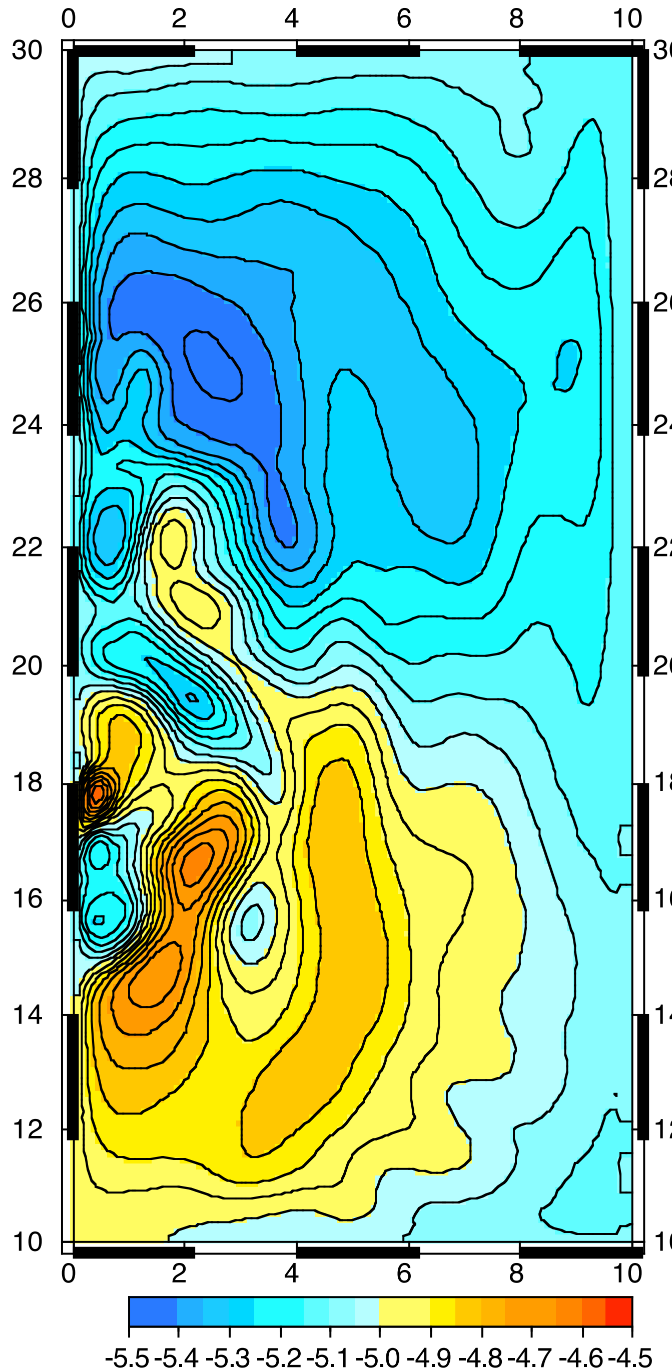
And make weights equal by some iterative method…

Preliminary example large system

Two-layer primitive equation model of a double gyre.

Lx=2000 km, Ly = 4000 km
Delta x=Delta y = 20 km
H1=1000 m, H2=4000 m
Wind profile 0.6 cos (y/L)

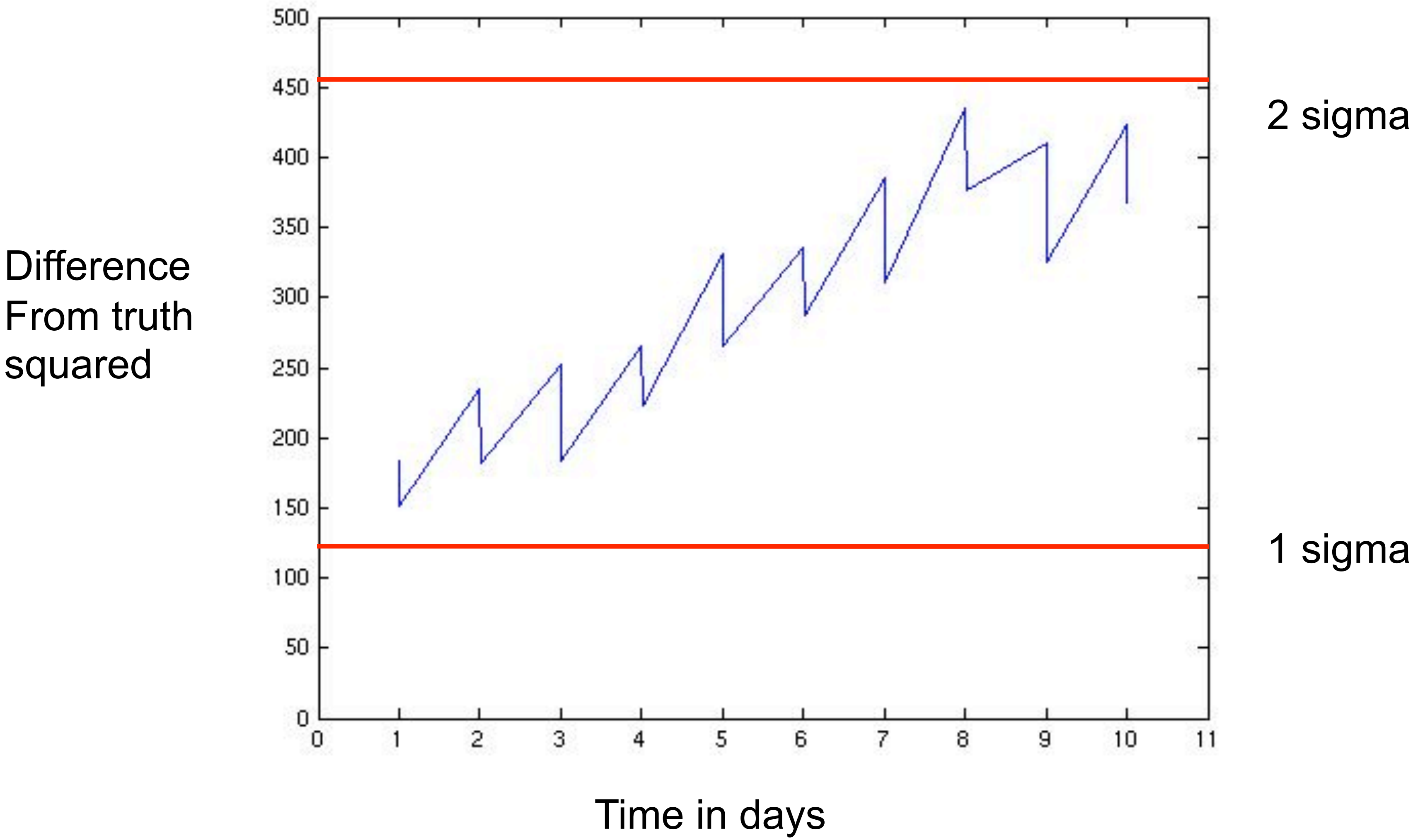
Observations
sea-surface height
Delta x = 80 km,
s= 2 cm
Interval: 1 days



Statistics

- About 100,000 variables
- About 1100 observations each day
- Initial error sigma 10 m
- Model error sigma 1 m in layer thickness each BC time step
- 256 particles

Quality of the particle filter



Intelligent monitoring

Relative entropy

- For a highly-nonlinear system the analysis is a pdf, not a single best estimate.
- A measure of the information content of a pdf is given by the relative entropy:

$$E[p|\mu] = -\int p(\psi) \log\left(\frac{p(\psi)}{\mu(\psi)}\right) d\psi$$

Mutual information

Mutual information measures the change in entropy of a pdf:

$$MI = E[p(\psi)] - E[p(\psi|d)] = -\int p(\psi|d) \log \frac{p(\psi)}{p(\psi|d)} \, d\psi$$

Use a particle filter:

$$MI = \sum_i w_i \log N w_i$$

Mutual Information

- Or, more generally, the mutual information of a new set of extra observations given the existing ones:

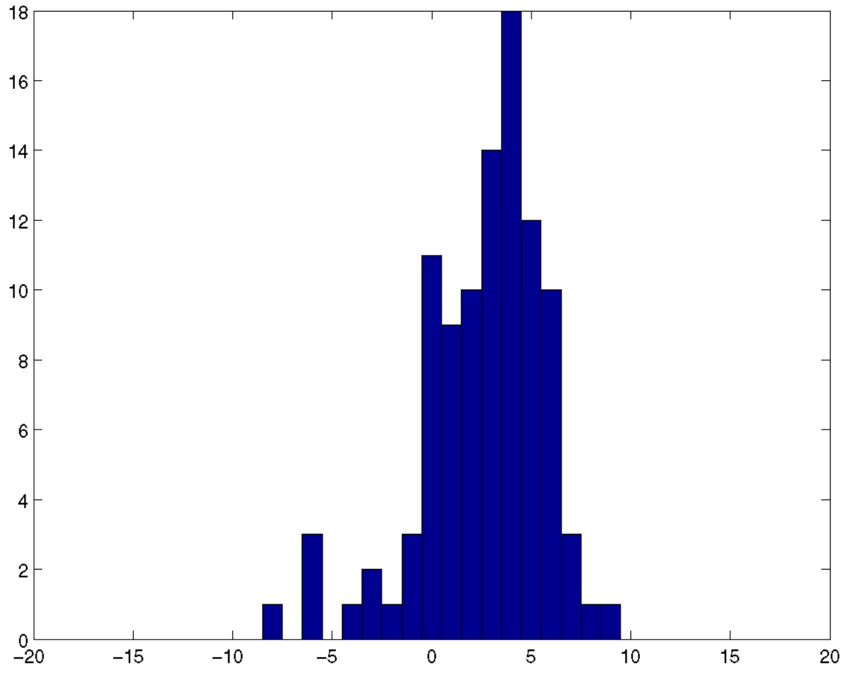
$$MI = \sum_i w_i^{new} \left(\log w_i^{new} - \log w_i^{old} \right)$$

- Theoretical bounds:

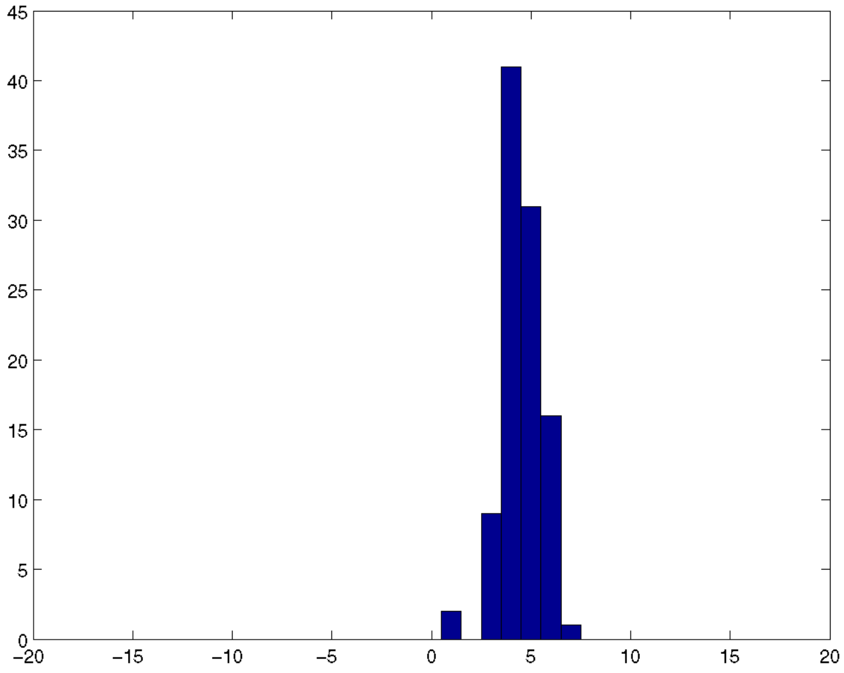
  - no influence (no change in weights) *MI = 0*

  - maximal influence (one weight is 1) *MI = log N*

Example with Lorenz 1963 model
100 particles (log N=4.6)



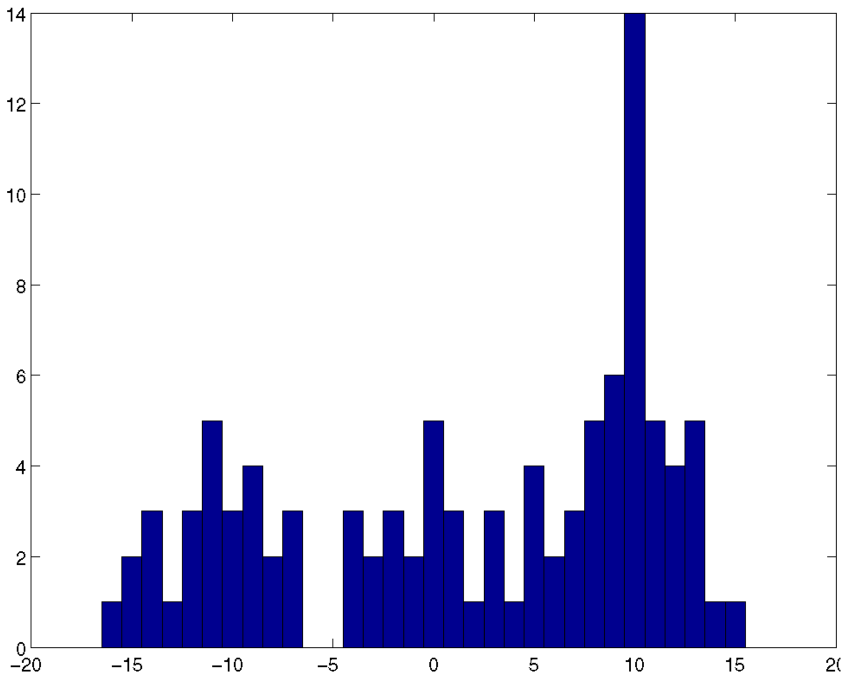
Prior pdf



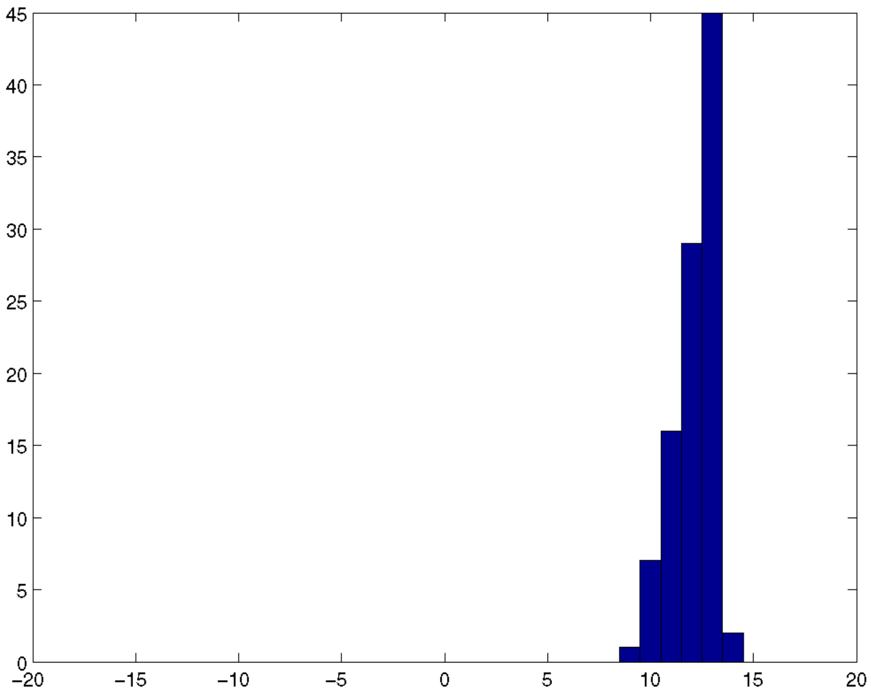
Posterior pdf

Mutual Information is 0.6984

Same example, another observation time



Prior pdf



Posterior pdf

Mutual Information is 1.7307

Conclusions particle filter

- Use of proposal density helps avoiding degeneracy.
- Some localization through the model error covariance.
- It is possible to chose the proposal density such that all posterior weights are equal.
- Particle filters start to look promising for large-scale models.

Conclusions observation influence

- Mutual Information is a fully nonlinear measure for observation influence
- It is extremely easy to calculate for a particle filter