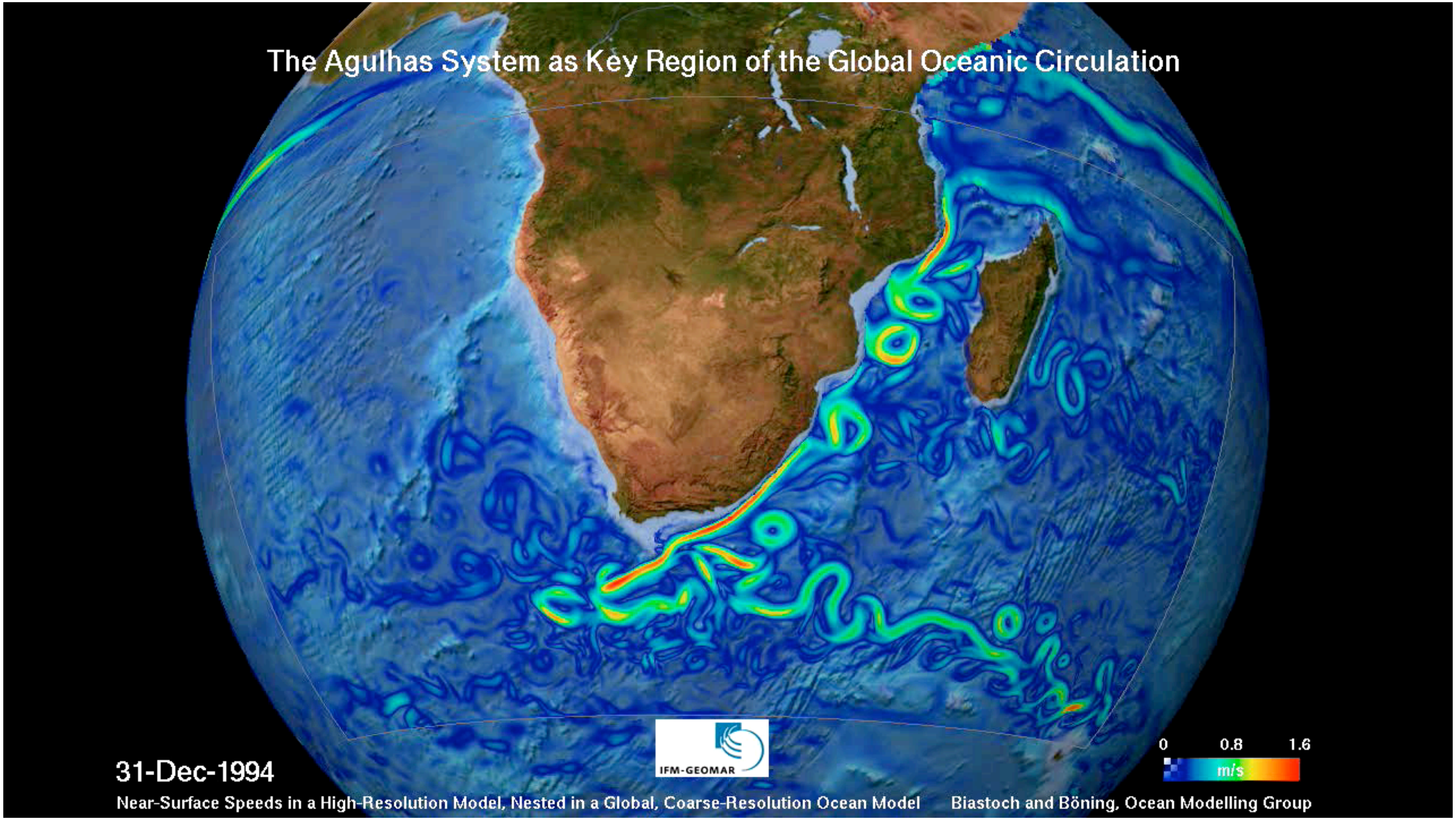
Particle Filters for high-dimensional problems

Peter Jan van Leeuwen and Mel Ades

Data-Assimilation Research Centre DARC

University of Reading

Selbu 2012


The Agulhas System as Key Region of the Global Oceanic Circulation
31-Dec-1994
IFM-GEOMAR
0
0.8
1.6
m/s
Near-Surface Speeds in a High-Resolution Model, Nested in a Global, Coarse-Resolution Ocean Model
Biastoch and Böning, Ocean Modelling Group

How is DA used today in geosciences?

Present-day data-assimilation systems are based on linearizations and state covariances are essential.

4DVar, Representer method (PSAS):

- Gaussian pdf's for the state, solves only for posterior mode, needs error covariance of initial state (B matrix)

(Ensemble) Kalman filter:

- assumes Gaussian pdf's for the state, approximates posterior mean and covariance, doesn't minimize anything in nonlinear systems, needs inflation and localisation

Combinations of these: hybrid methods

Nonlinear filtering: Particle filter

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

$\Big\downarrow$ Use ensemble

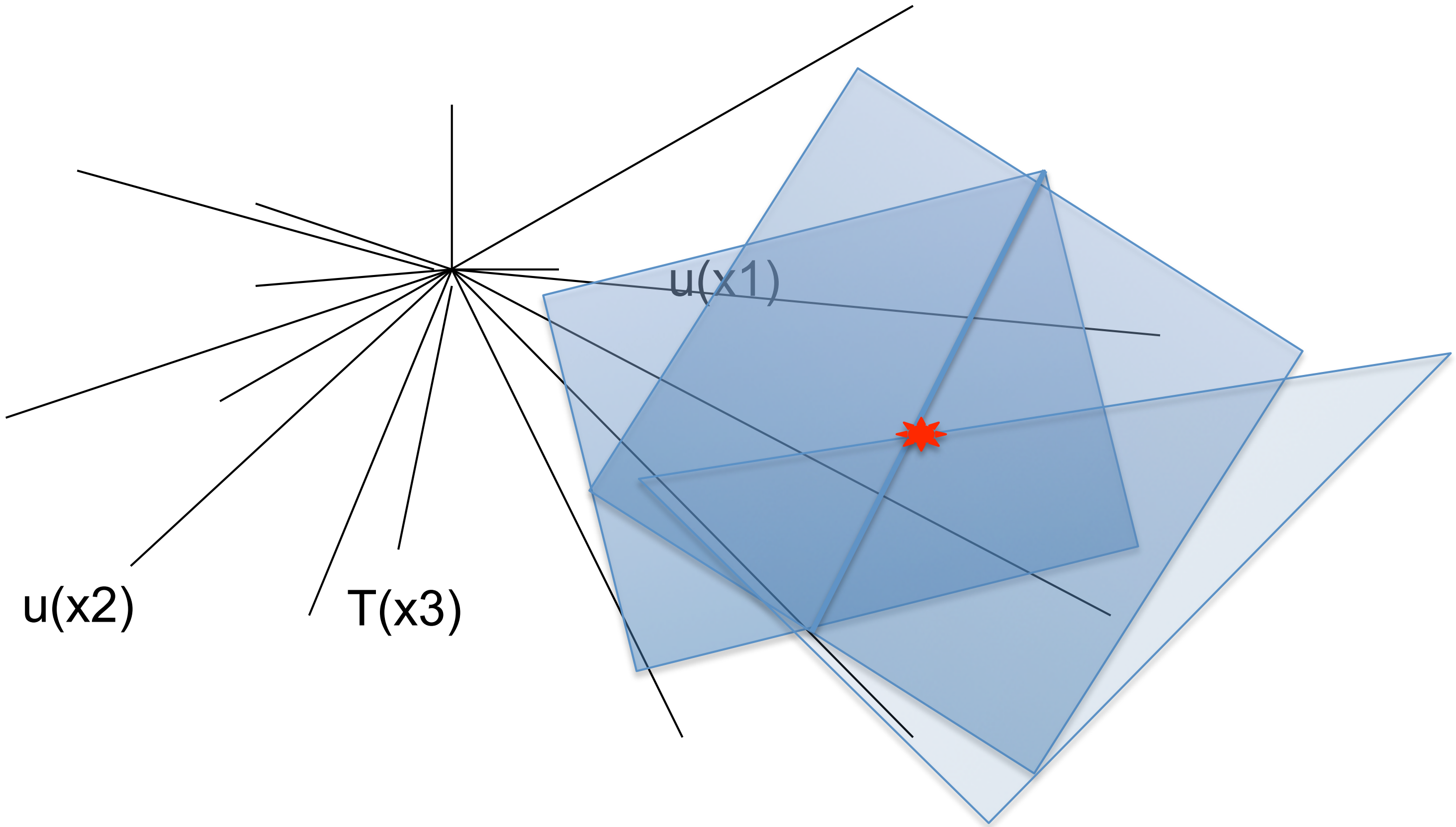
$$\boxed{p(x) = \sum_{i=1}^{N} \frac{1}{N}\delta(x - x_i)}$$

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

with $\boxed{w_i = \dfrac{p(y|x_i)}{\sum_j p(y|x_j)}}$ the weights.

Why are particle filters degenerate I

Probability space in large-dimensional systems is 'empty': the curse of dimensionality



Why are Particle Filters degenerate II

- The volume of a hypersphere of radius *r* in an *M* dimensional space is

$$\boxed{V \propto \frac{r^M}{\Gamma(M/2-1)}}$$

- Taking for the radius $r \approx 3\sigma_y$ we find, using Stirling:

$$\boxed{V \propto \left[\frac{9\sigma_y}{M/2}\right]^{M/2}}$$

- So very small indeed.

Why are Particle Filters degenerate III

For the optimal proposal density we find, for Gaussian process model and Gaussian observation errors:

$$\boxed{\begin{aligned} w_i &\propto p(y^n|x_i^{n-1}) \\ &\propto \exp\Big[-\frac{1}{2}(y^n - Hf(x_i^{n-1}))(HQH^T + R)^{-1} \\ &\qquad \times (y^n - Hf(x_i^{n-1}))\Big]. \end{aligned}}$$

Ignoring covariances we find:

$$\boxed{\mathrm{var}[-\log(w_i)] \propto \frac{M}{2}\left(\frac{V_x}{V_\beta + V_y}\right)^2 \left(1 + 2\left(\frac{V_y + V_\beta}{V_x}\right)\right)}$$

Why are Particle Filters degenerate?

- ‘Number of particles needed grows exponentially with dimension of the state vector (Bickel et al, 2007).’

- A slightly different view: degeneracy due to number of independent observations.

- This is related to the extremely narrow likelihood, a tiny move of a particle gives a completely different weight.

The statistics

- The Stochastic PDE: $\boxed{x^n = f(x^{n-1}) + \beta^{n-1}}$

- Observations: $\boxed{y^n}$

- Relation between the two: $\boxed{y^n = H(x^n) + \epsilon^n}$

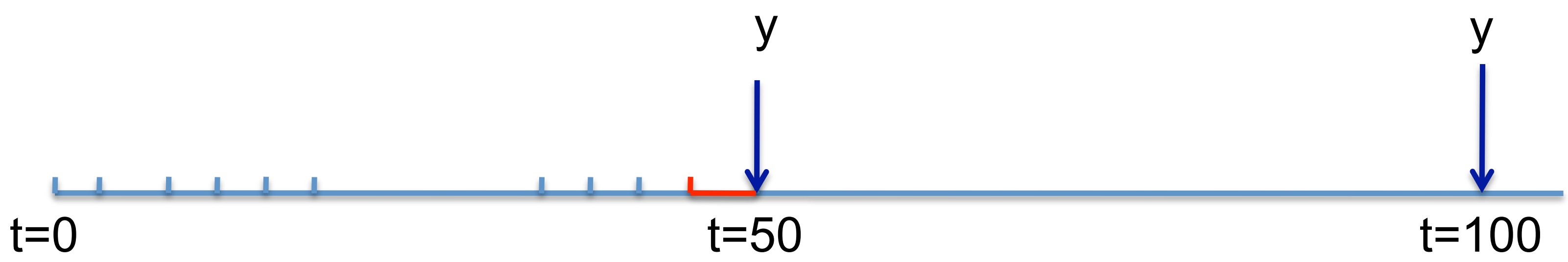
Assume: $\beta \sim N(0, Q)$

$\epsilon \sim N(0, R)$

H is linear

The Equivalent-Weights Particle Filter

- Use simple proposal at each time step, e.g. relaxation to observations.
- Use different proposal at final time step to ensure that weights are very similar.



Transition density for x^n

- Stochastic model

$$\boxed{x^n = f(x^{n-1}) + \beta^{n-1}}$$

- With $\boxed{\beta^{n-1} \sim N(0, Q)}$

- Hence transition density

$$\boxed{p(x^n | x^{n-1}) = N\left(f(x^{n-1}), Q\right)}$$

Bayes Theorem and the proposal density

Bayes Theorem can be written as:

$$
\begin{aligned}
p(x^n|y^n) &= \frac{p(y^n|x^n)p(x^n)}{p(y)} \\
&= \frac{p(y^n|x^n)}{p(y)} \int p(x^n|x^{n-1})p(x^{n-1})\; dx^{n-1}
\end{aligned}
$$

Multiply and divide this expression by a proposal transition density *q*:

$$
p(x^n|y^n) = \frac{p(y^n|x^n)}{p(y)} \int \frac{p(x^n|x^{n-1})}{q(x^x|x^{n-1}, y^n)} q(x^n|x^{n-1}, y^n)p(x^{n-1})\; dx^{n-1}
$$

Proposal transition density

For each particle at time n-1 draw a sample from the proposal transition density *q*, to find:

$$p(x^n|y^n) = \frac{1}{N}\sum_{i=1}^{N} \frac{p(y^n|x_i^n)}{p(y)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)} \delta(x^n - x_i^n)$$

Which can be rewritten as:

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

with weights

$$w_i = \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}$$

Likelihood weight

Proposal weight

Proposal density between observations

We can explore the fact that the model needs several O(100) time steps between observations, e.g. by using a relaxation term in the proposal:

$$\boxed{q(x^n|x_i^{n-1}, y^m) = N\left(f(x_i^{n-1}) + S\left(y^m - H(x_i^{n-1})\right), Q\right)}$$

Corresponding to an evolution equation for each particle

$$\boxed{x_i^n = f(x_i^{n-1}) + \hat{\beta}_i^n + S\left(y^n - H(x_i^{n-1})\right)}$$

Proposal density between observations

- One could also use the ‘optimal proposal density’ between observations.

- This can be implemented as a minimization method for each particle, and is also known as the Implicit Particle Filter.

- This is related to a method called 4DVar in meteorology and oceanography, which explores only the mode of the joint-in-time pdf.

Proposal density at observation time:

the essence of the Equivalent-Weights Particle Filter

The proposal density depends on the maximum weight a particle can achieve using a deterministic time step. It is defined as:

$$\boxed{q(x^n|x_i^{n-1}, y^n) = \begin{cases} q_1(x^n|x_i^{n-1}, y^n) & \text{if} \quad w_i^{max} > w^{target} \\ q_2(x^n|x_i^{n-1}, y^n) & \text{if} \quad w_i^{max} < w^{target} \end{cases}}$$

The target weight is set by the user, as e.g. the weight that 80% of the particles can achieve.

The maximum weights

1. We know:

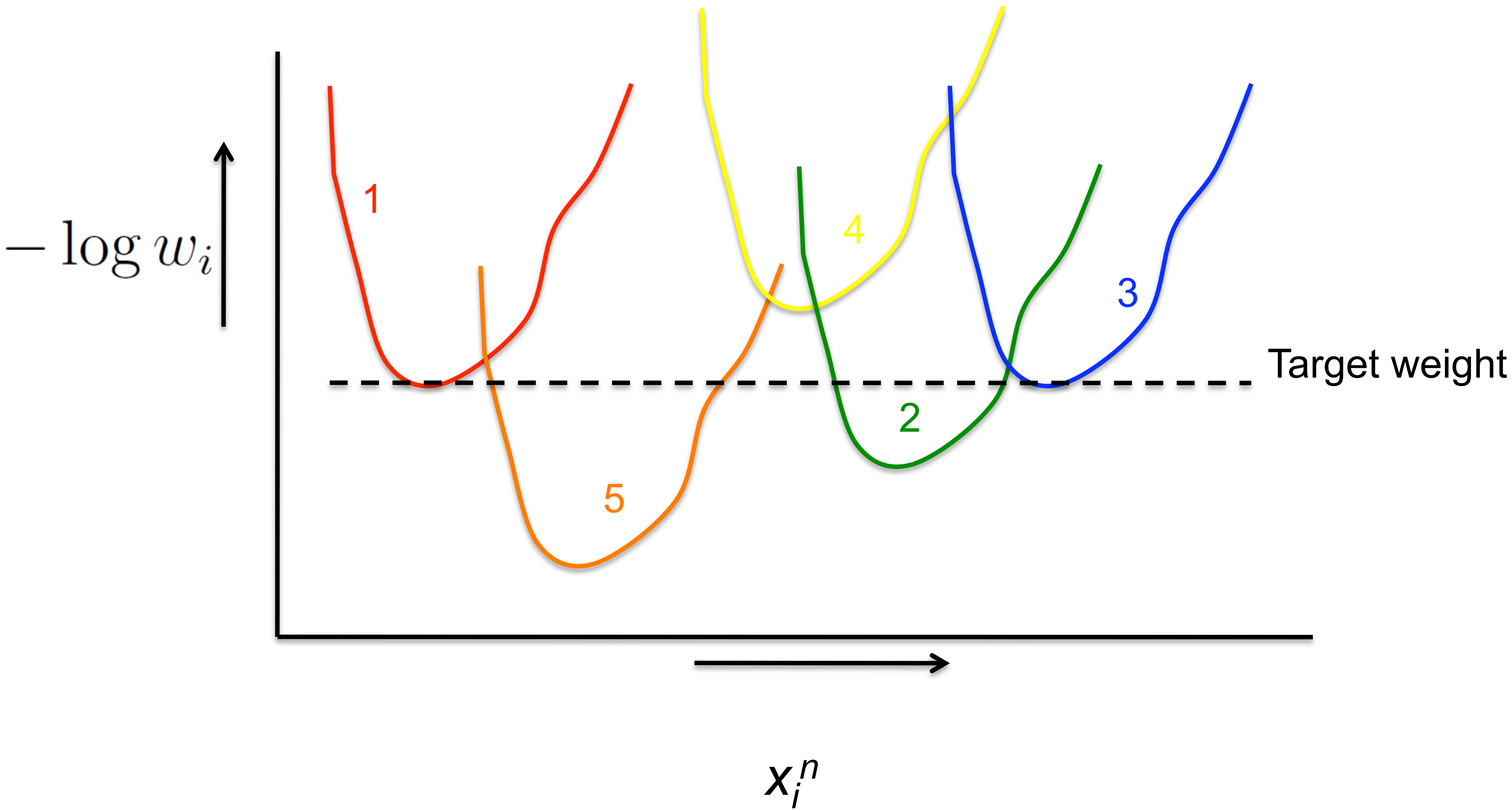
$$\boxed{w_i = \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}}$$

2. Write down expression for each weight ignoring proposal:

$$\boxed{\begin{aligned} w_i \propto \; w_i^{rest} \quad \exp \quad & \left[-\frac{1}{2} \left(x_i^n - f(x_i^{n-1}) \right)^T Q^{-1} \left(x_i^n - f(x_i^{n-1}) \right) \right. \\ & \left. - \frac{1}{2} (y^n - H(x_i^n))^T R^{-1} (y^n - H(x_i^n)) \right] \end{aligned}}$$

3. When H is linear this is a quadratic function in x_i^n for each particle. Otherwise linearize.

The target weight



The Equivalent-Weights Particle Filter

The proposal density is chosen as:

$$\boxed{q(x^n|x_i^{n-1},y^n) = \begin{cases} q_1(x^n|x_i^{n-1},y^n) & \text{if} \quad w_i^{max} > w^{target} \\ q_2(x^n|x_i^{n-1},y^n) & \text{if} \quad w_i^{max} < w^{target} \end{cases}}$$

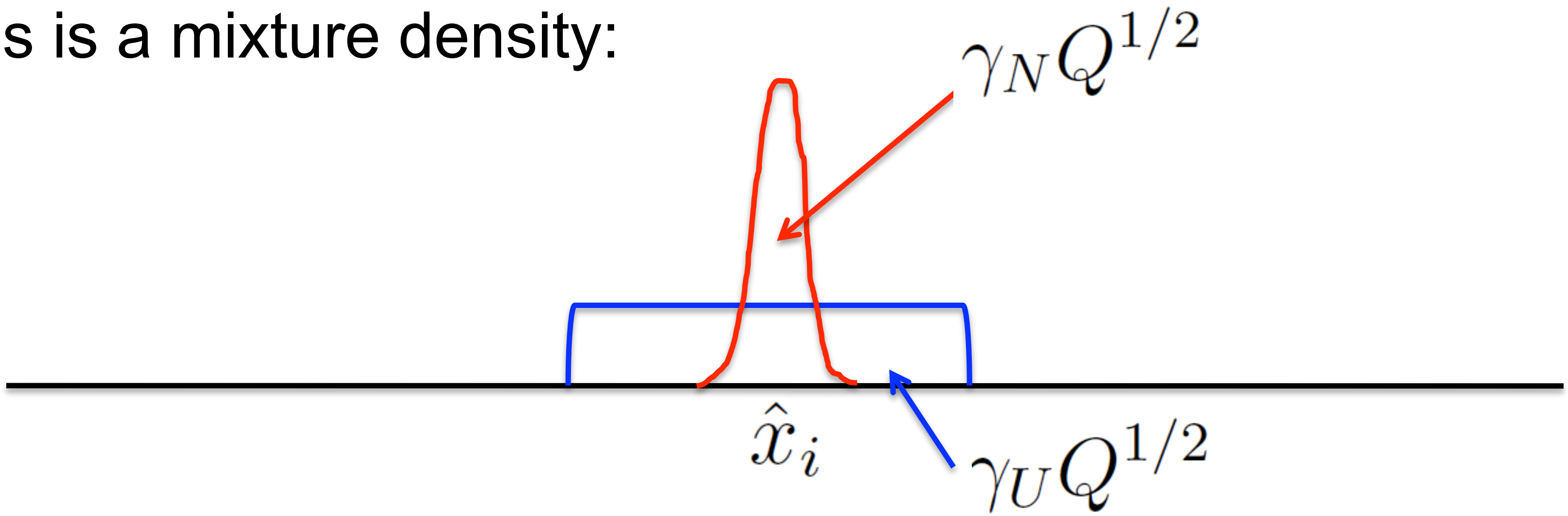
The target weight is set by the user, as e.g. the weight that 80% of the particles can achieve.

The two proposal densities

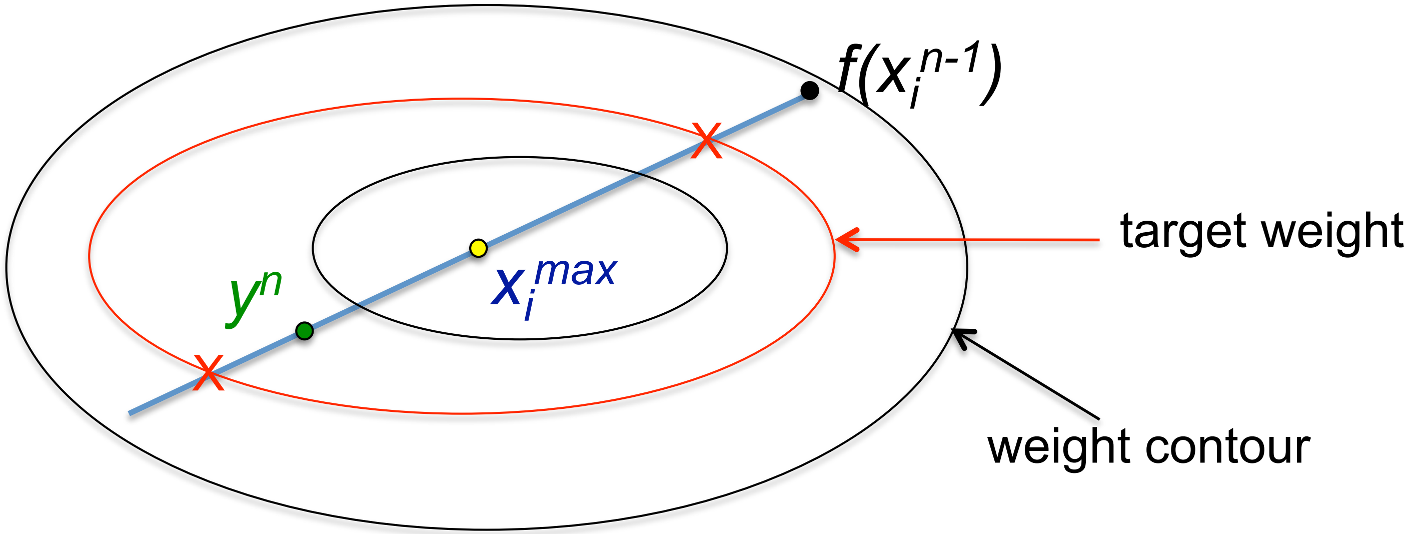
For particles that can reach the target weight we use:

$$\boxed{\begin{aligned} q_1(x^n|x_i^{n-1}, y^m) \quad &= \quad (1-\epsilon)U\left(\hat{x}_i - \gamma_U Q^{1/2}\mathbf{1}, \hat{x}_i + \gamma_U Q^{1/2}\mathbf{1}\right) \\ &+ \quad \epsilon \quad N\left(\hat{x}_i, \gamma_N^2 Q\right) \end{aligned}}$$

This is a mixture density:



The deterministic move



Determine α at crossing of line with target weight contour in:

$$\boxed{\hat{x}_i = f(x_i^{n-1}) + \alpha_i K \left(y^n - H(f(x_i^{n-1})) \right)}$$

with

$$\boxed{K = QH^T(HQH^T + R)^{-1}}$$

The stochastic part of the proposal

A draw from the uniform density gives:

$$\boxed{w_i = \frac{|Q|^{1/2}(2\gamma_U)^k}{1-\epsilon} w_i^{rest} p(x_i^n | x_i^{n-1}) p(y^n | x_i^n)}$$

A draw from the Gaussian density gives:

$$\boxed{w_i = \frac{w_i^{rest} p(x_i^n | x_i^{n-1}) p(y^n | x_i^n)}{\frac{\epsilon}{(2\pi)^{k/2} |\gamma_N^2 Q|^{1/2}} \exp(-\frac{1}{2}\gamma_U d\beta_i^n (\gamma_U^2 Q)^{-1} \gamma_U d\beta_i^n)}}$$

The ratio between the two is (ignoring the exp part):

$$\boxed{\frac{(2\pi)^{k/2} |\gamma_N^2 Q|^{1/2}}{\epsilon} \frac{(1-\epsilon)}{|Q|^{1/2}(2\gamma_U)^k}}$$

which can be made equal to one when:

$$\boxed{\gamma_N = \frac{2^{k/2}\epsilon}{\pi^{k/2}(1-\epsilon)} \gamma_U^k}$$

Equivalent-Weights Particle Filter

- Use relaxation-term proposal up to last time step
- Calculate w_i^{max} and target weight (e.g. 80%)
- Calculate deterministic moves for high-weight particles:

$$\boxed{\hat{x}_i = f(x_i^{n-1}) + \alpha_i K \left(y^n - H(f(x_i^{n-1}))\right)}$$

- Determine stochastic move

$$\boxed{p(\hat{\beta}_i^{n-1}) \propto (1-a)U[-b,b] + aN(0,\hat{Q})}$$

- Calculate new weights and resample 'lost' particles

How essential are Gaussian assumptions?

- Allows for analytical expressions.
- But no real need.
- w_i^{max} calculations do not have to be very accurate.
- Same for $w^{target.}$
- Deterministic move has to be very accurate, good iterative schemes should be used.

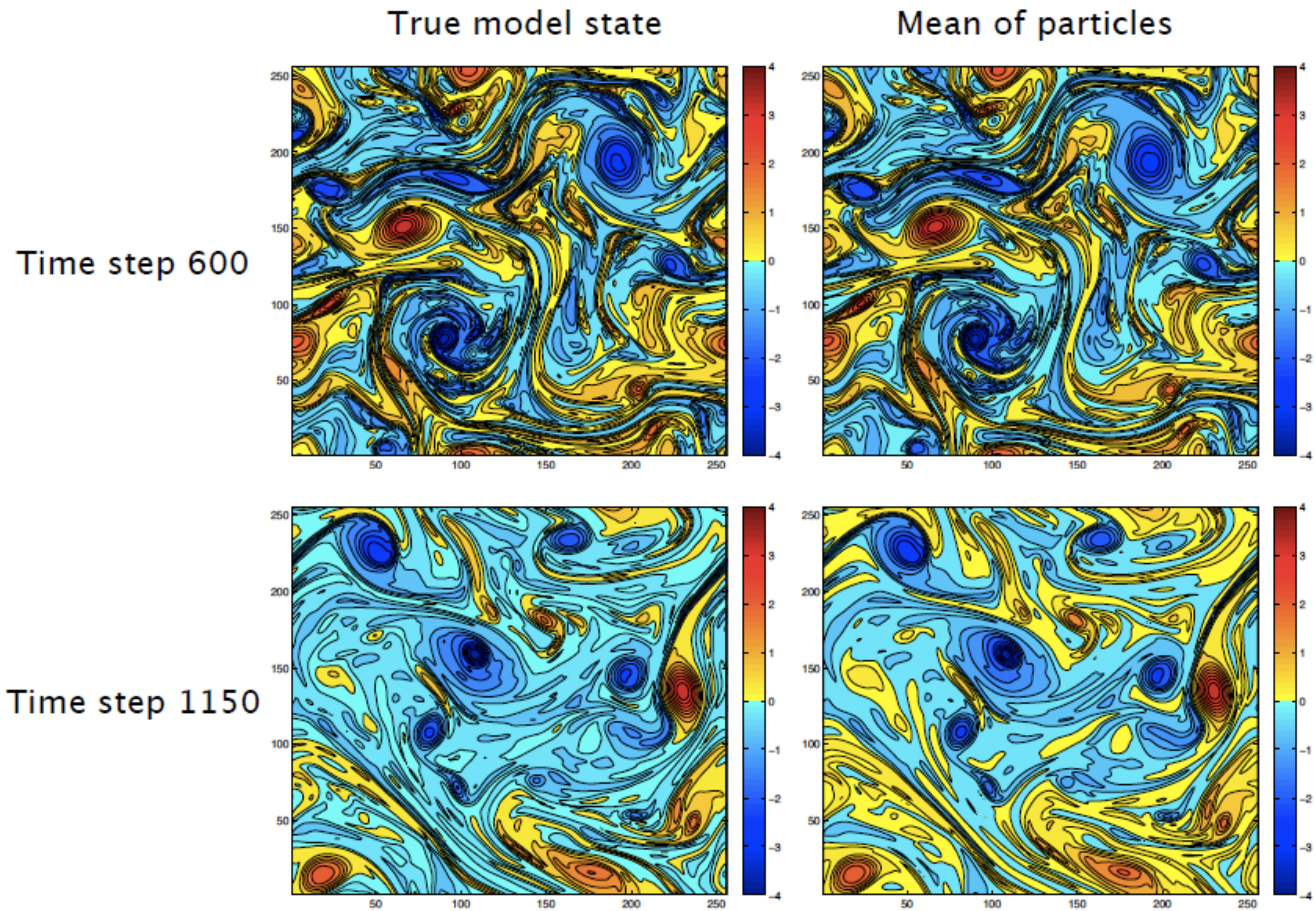
Application: the barotropic vorticity equation

- Stochastic barotropic vorticity equation:

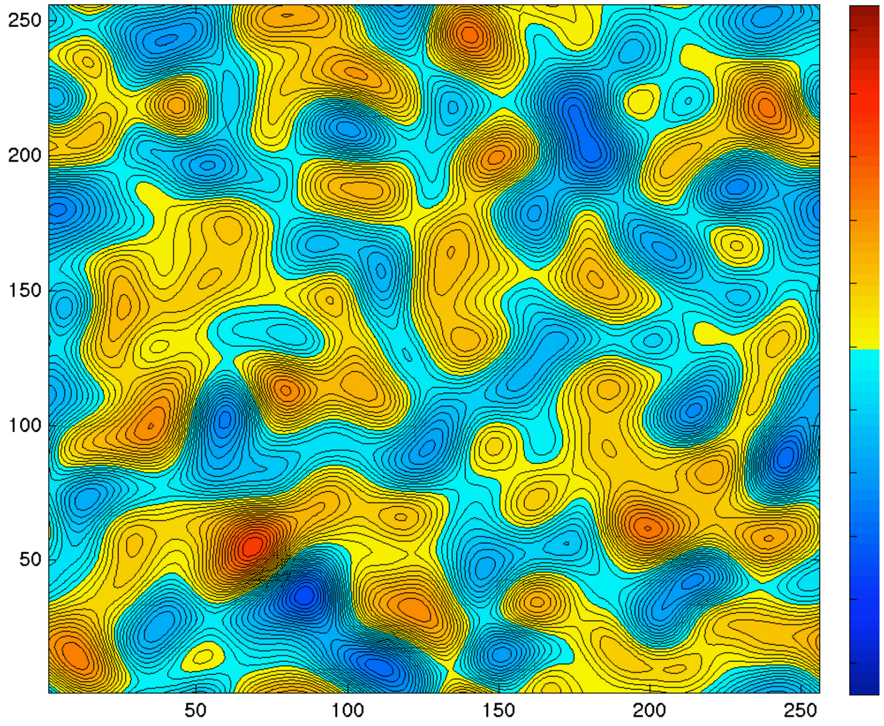
$$\boxed{\frac{\partial q}{\partial t} + u \cdot \nabla q = F}$$

- 256 by 256 grid - 65,536 variables
- Double periodic boundary conditions
- Semi-Langrangian time stepping scheme
- Twin experiments
- Observations every 50 time steps – decorrelation time of 42
- 32 particles

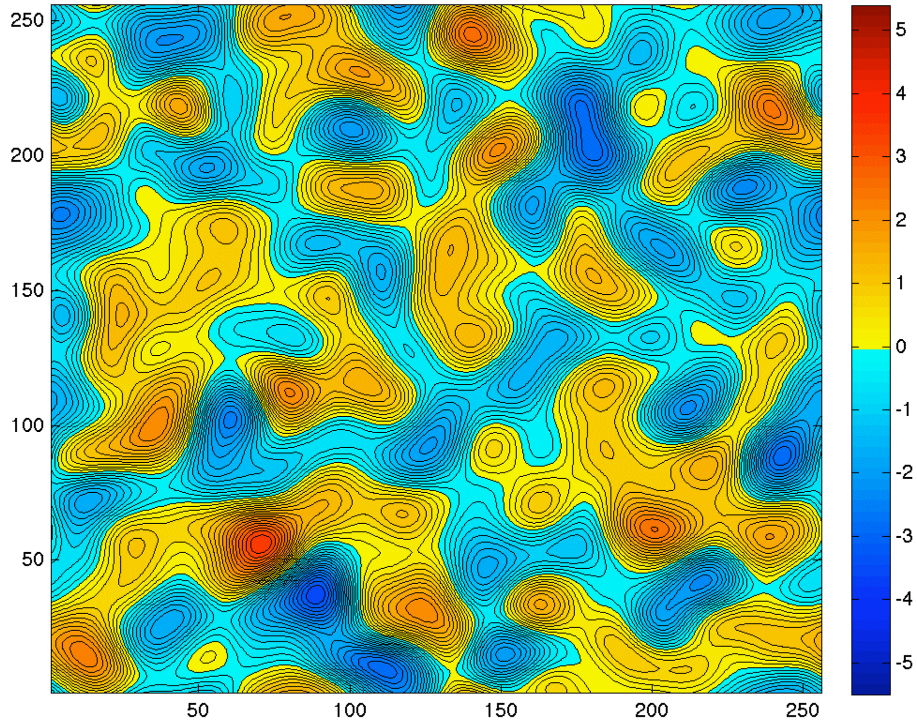
Fully observed system



¼ Observations over half of state

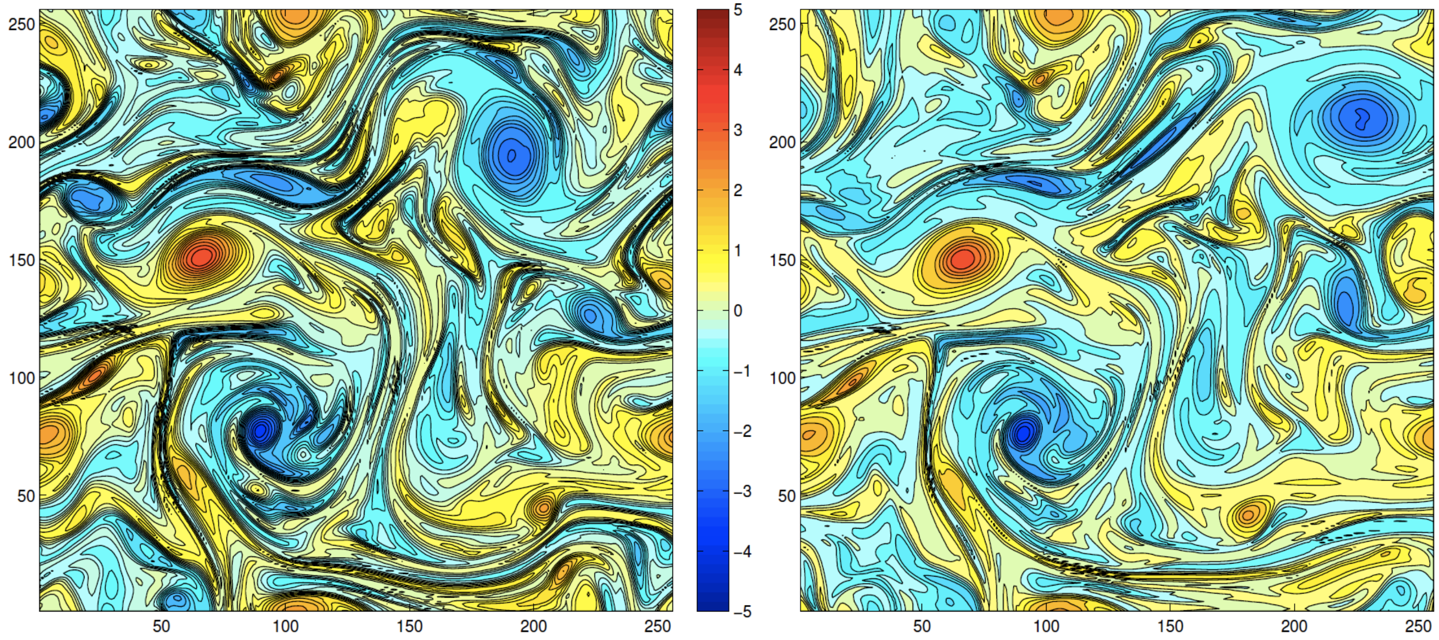


Truth

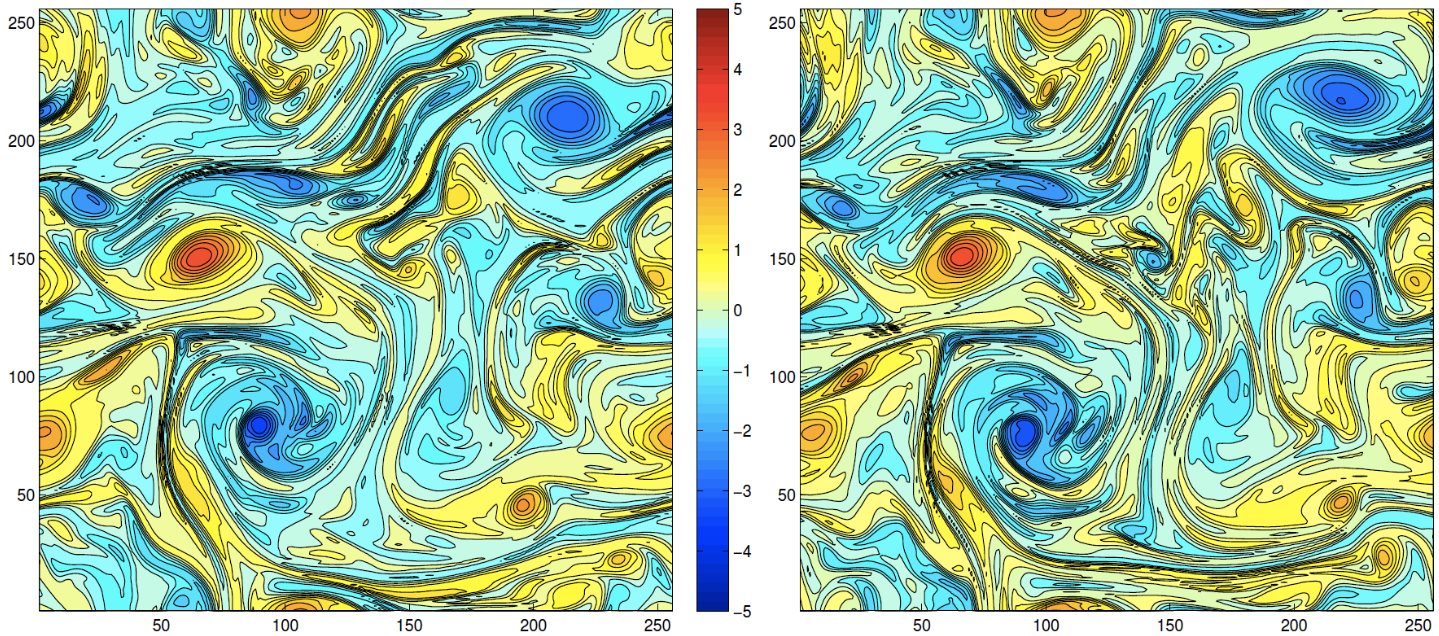


Mean of particle filter ensemble

Individual particles are not smooth.

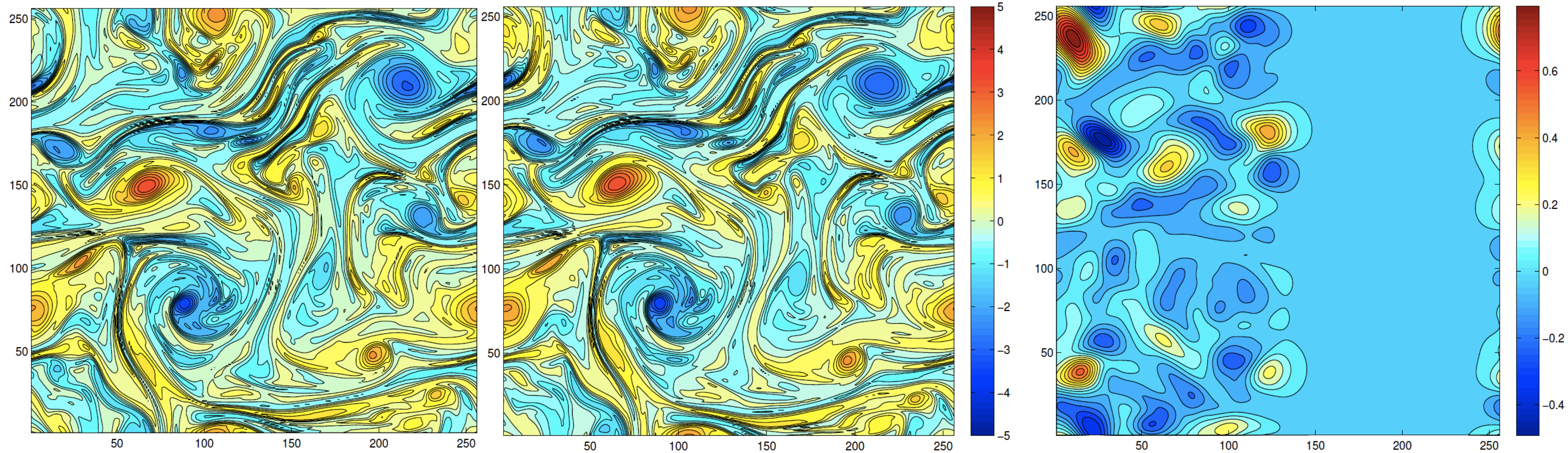


(a) truth (b) Particle 2



(c) Particle 23 (d) Particle 28

The update of the unobserved part

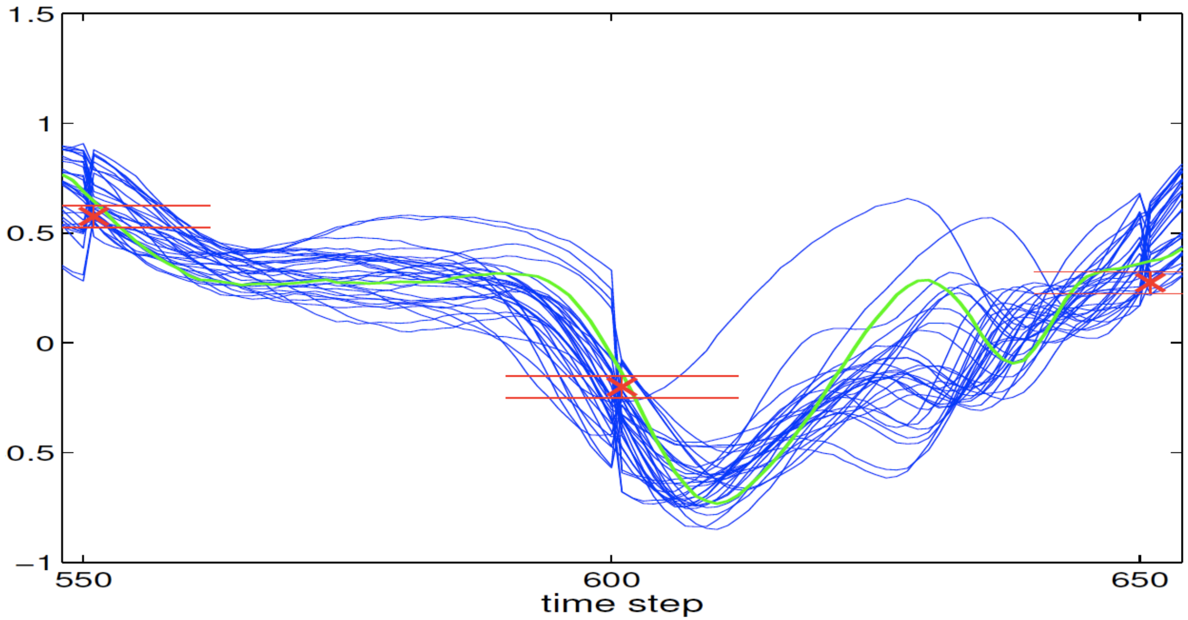


Particle 23 before update

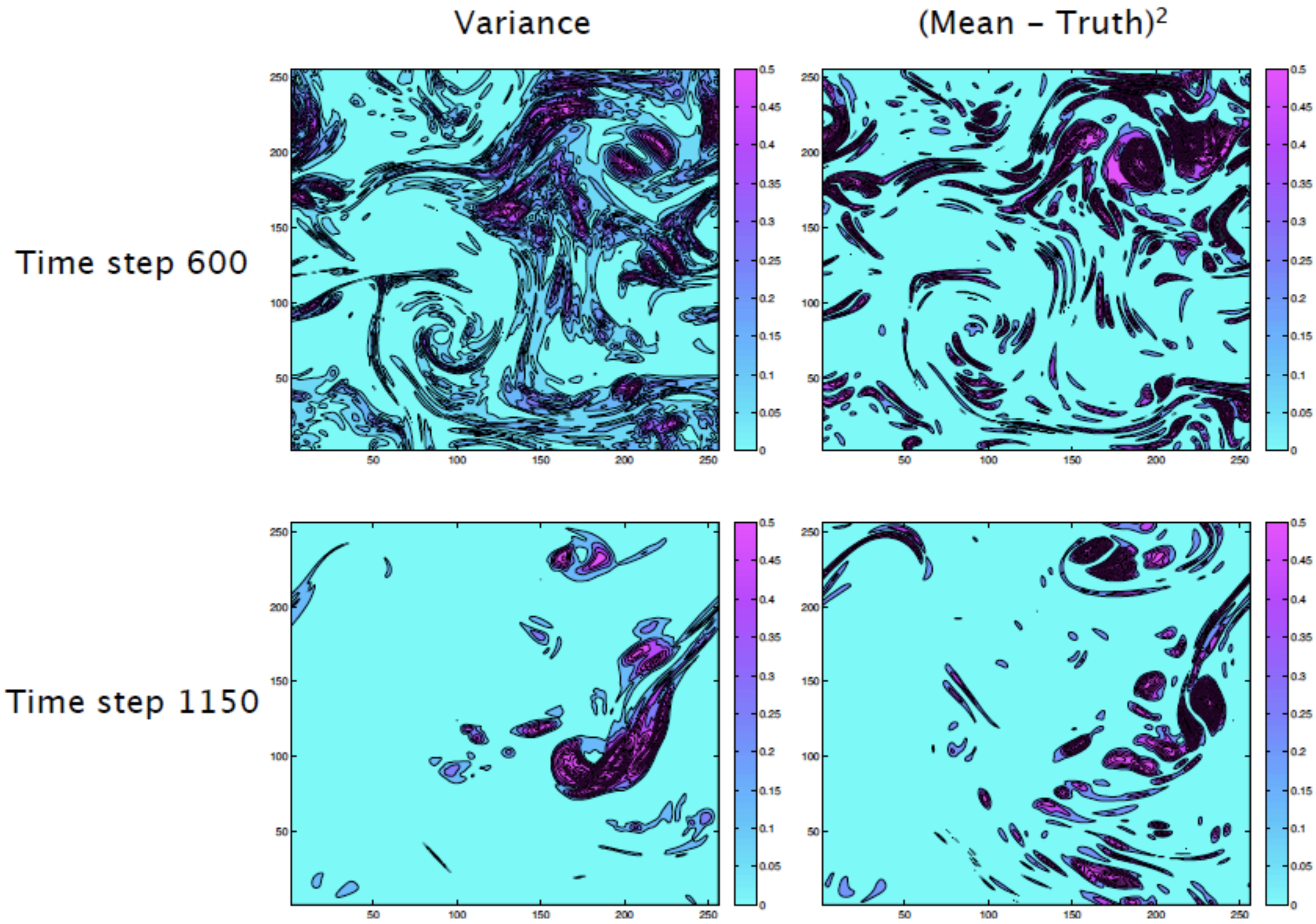
Particle 23 after update

Difference

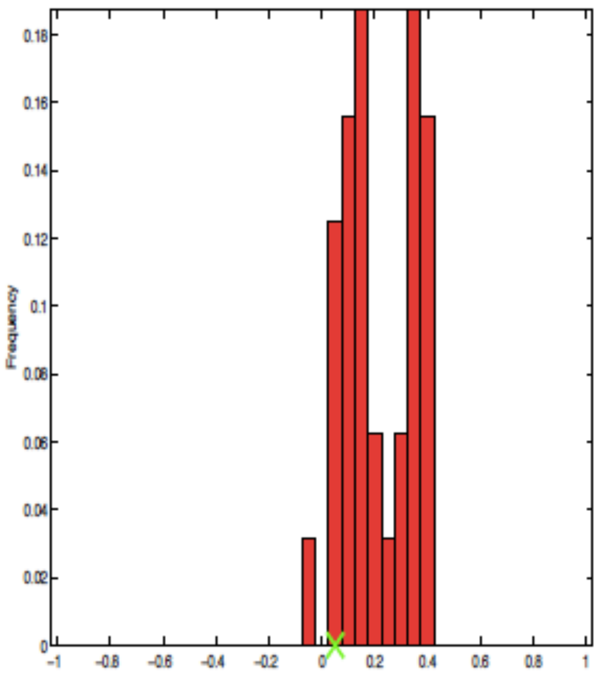
Time evolution at a specific grid point



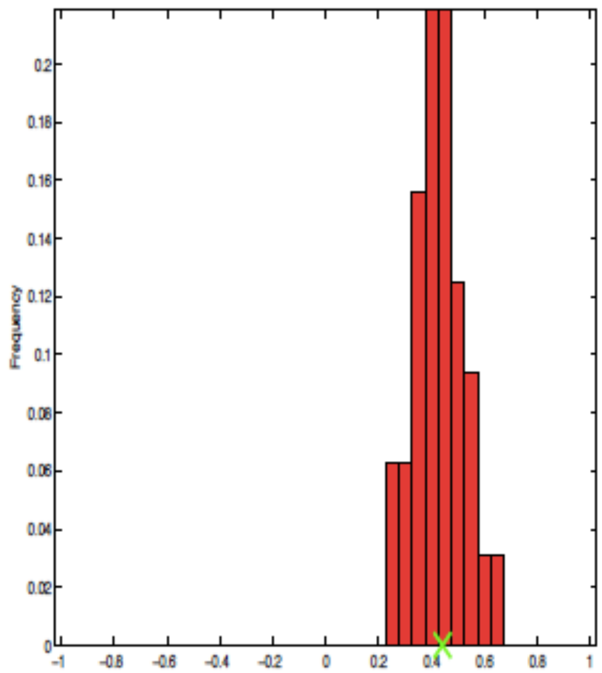
¼ observations over half of state



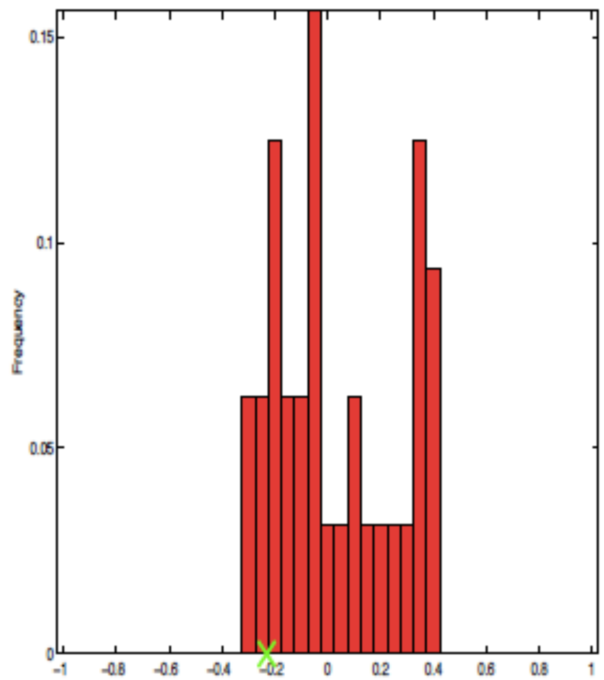
Marginal posterior probability densities



variable (20,64)

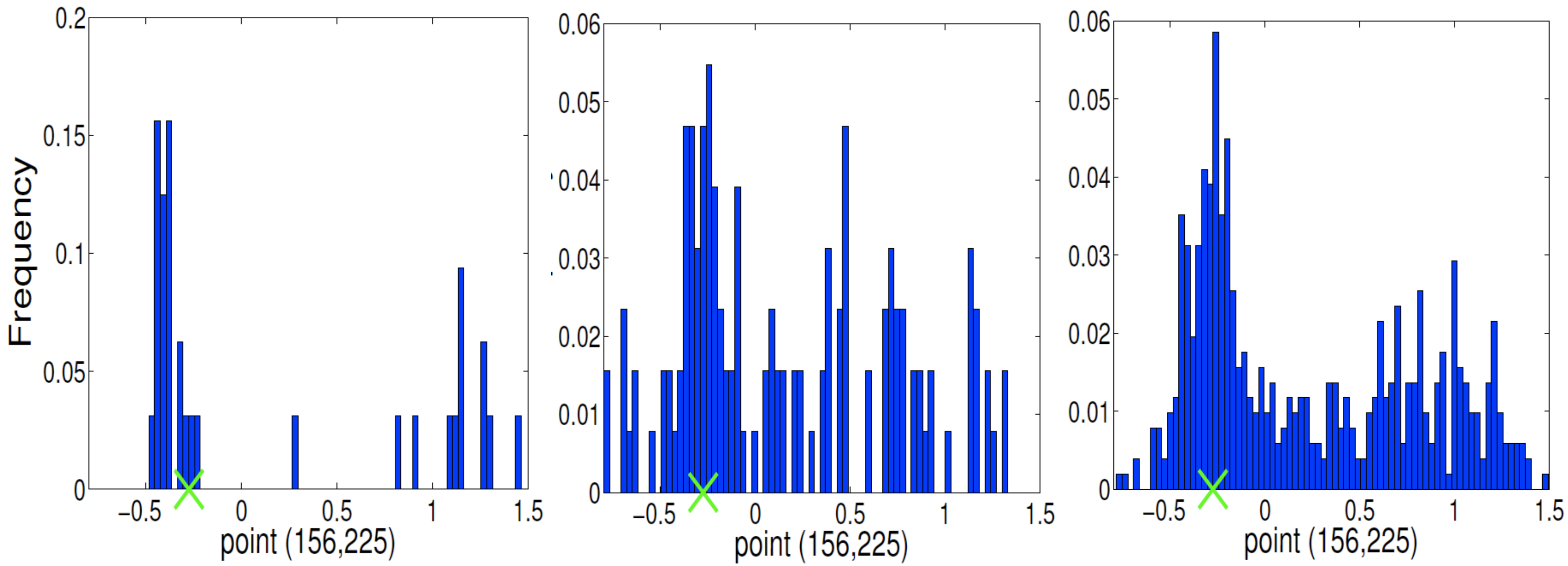


variable (56,64)



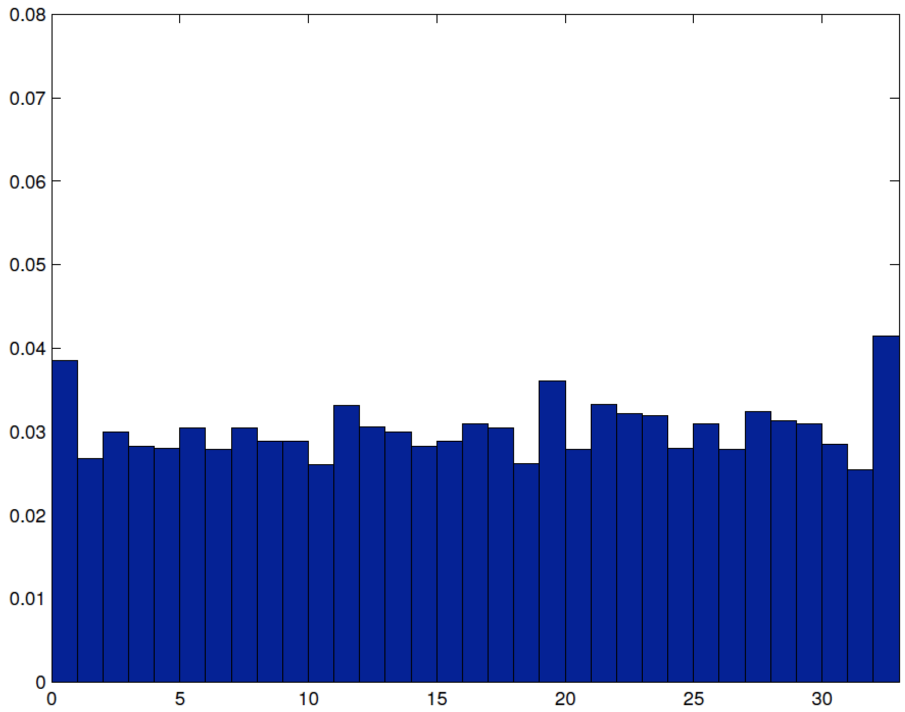
variable (197,64)

Convergence of the pdf's

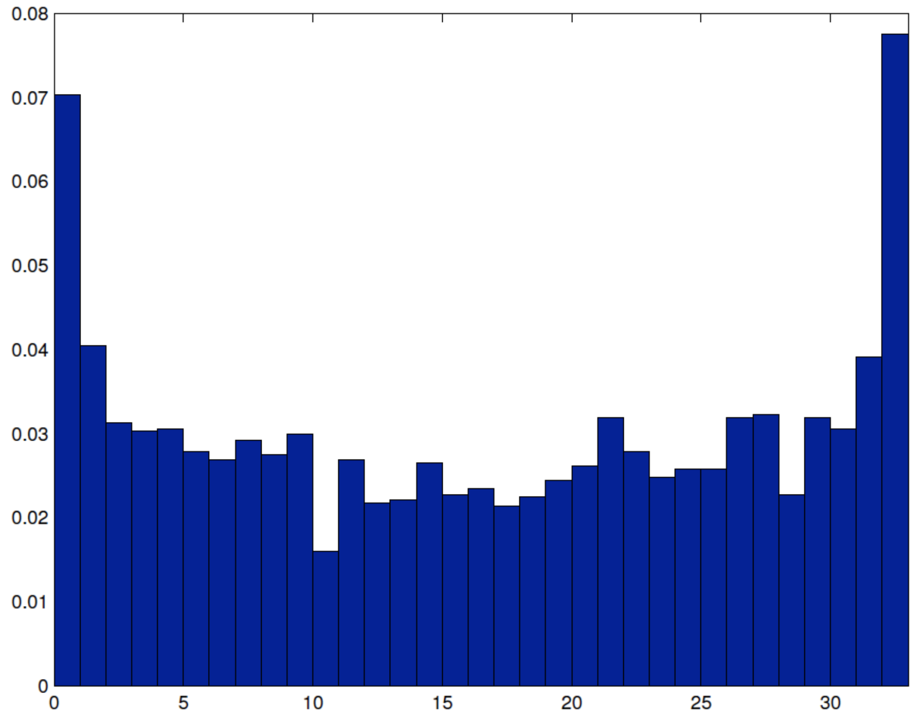


32 particles　　128 particles　　512 particles

Rank histograms

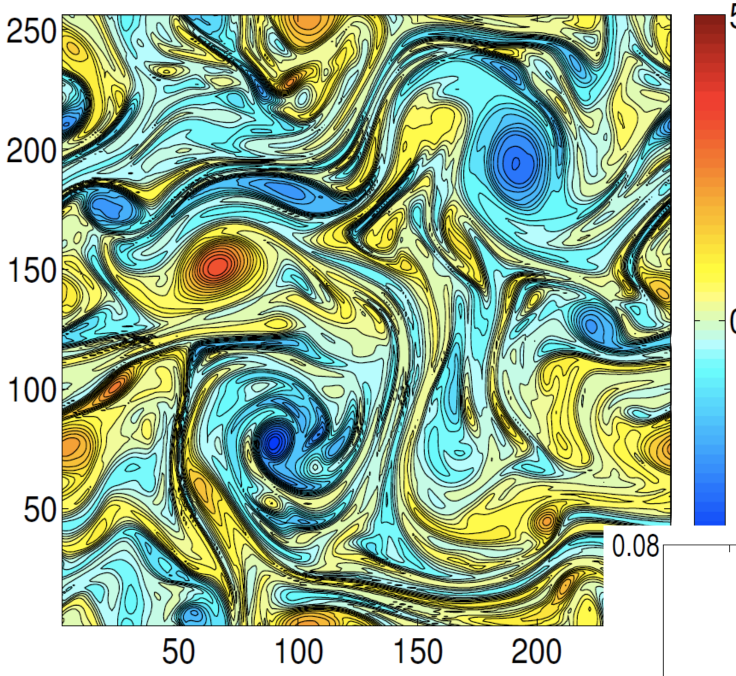


Full state observed

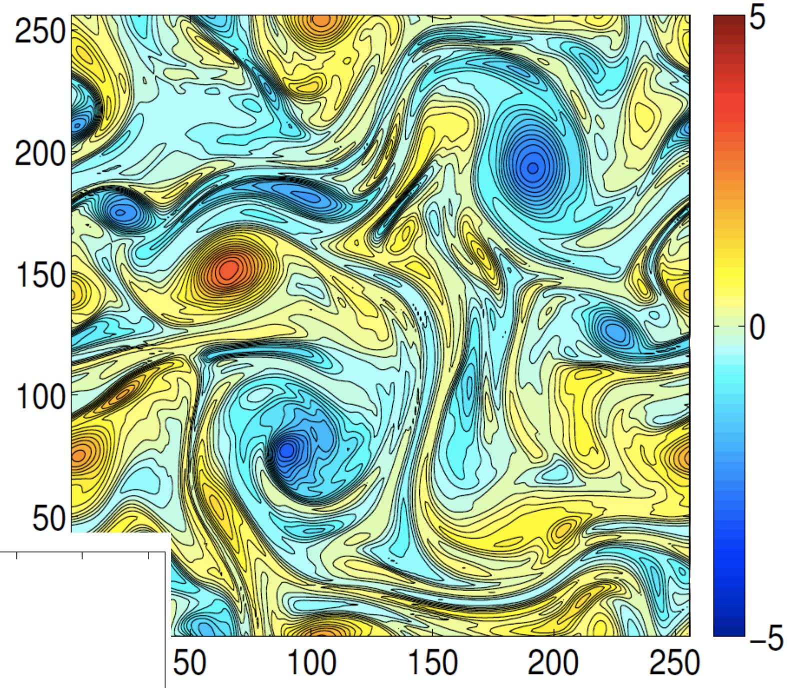


¼ of half state observed

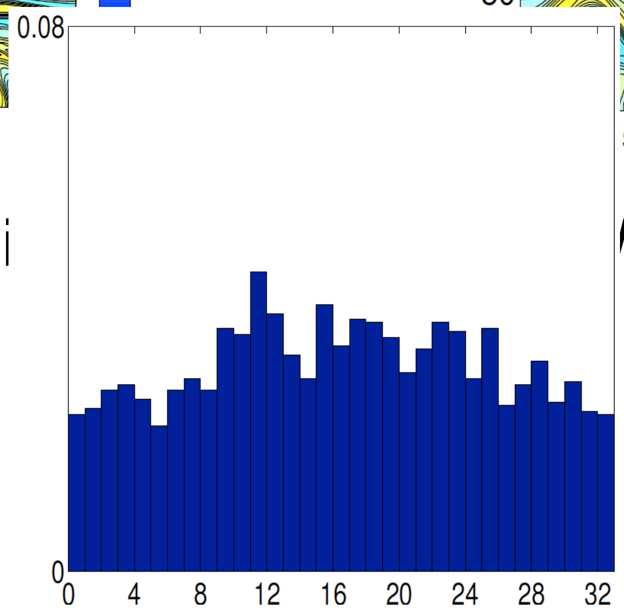
Miss-specification of process noise



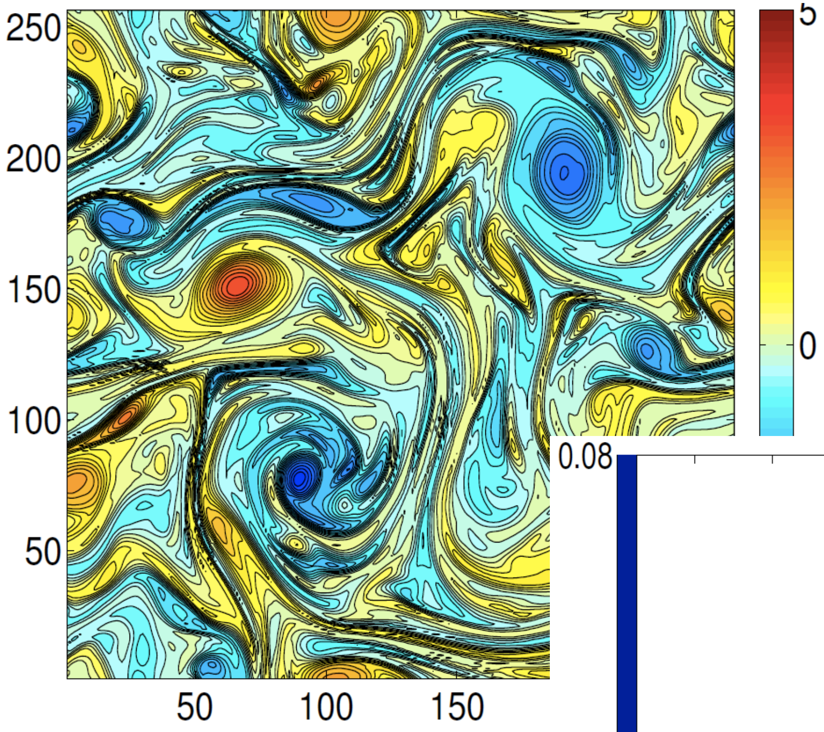
Truth: L_Q=5 gridpoi



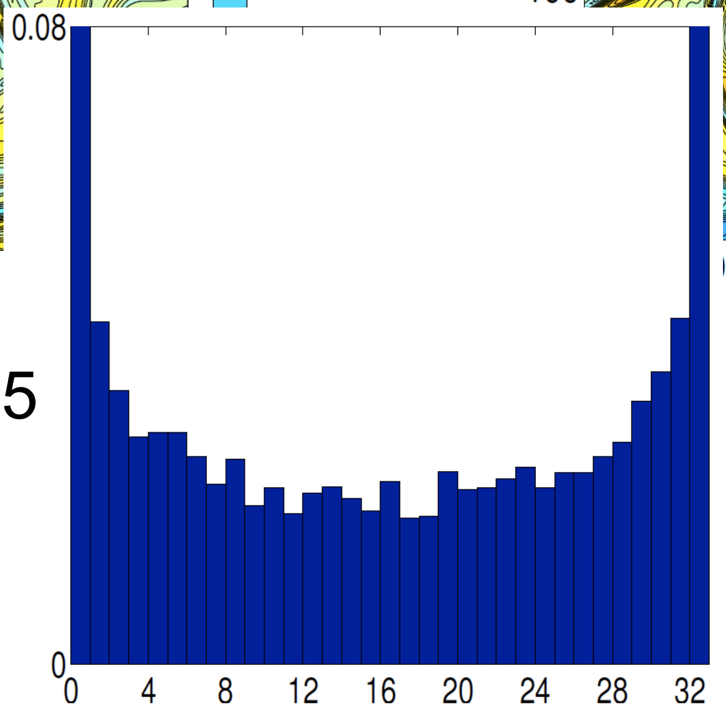
/-PF: L_Q=9 gridpoints

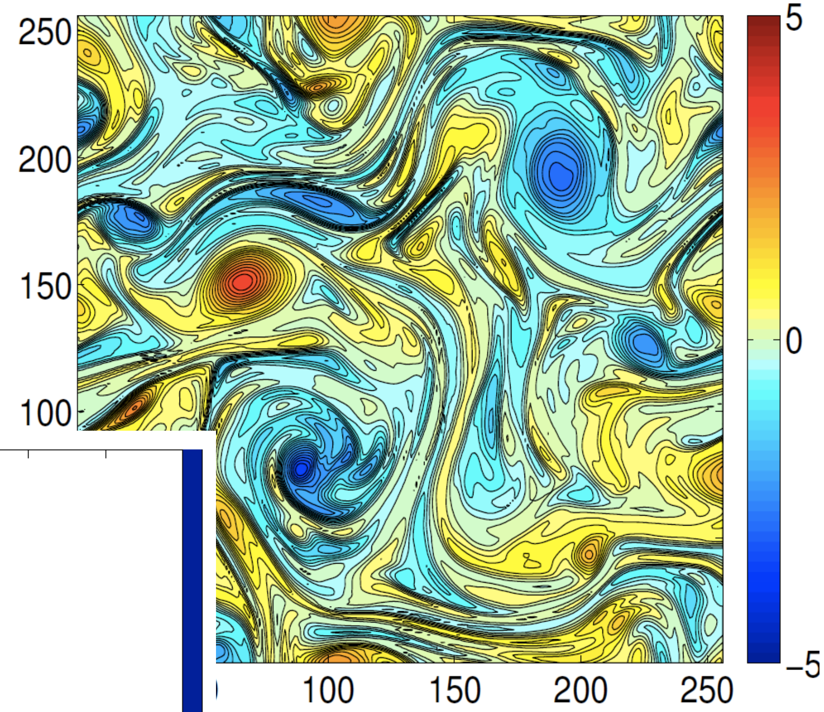


Miss-specification of process noise



Truth: Q=0.025





EW-PF: Q=0.1

Conclusions

- Particle filters do not need state covariances.
- Degeneracy is related to number of observations, not to size of the state space.
- Proposal density allows enormous freedom
- Equivalent-weights scheme solves dimensionality problem?
- Other efficient schemes are being derived.
- Present work: numerical weather prediction, climate forecasting

We need more people !

- The Data Assimilation group at reading consists of 30 scientists
- We still have room in the

Data Assimilation and Inverse Methods

in Geosciences MSc programme

References

All references can be found on my website
http://www.met.reading.ac.uk/~xv901096/research/publications.html

- Nonlinear Data Assimilation in geosciences: an extremely efficient particle filter Van Leeuwen, P.J., Q.J.R.Meteorol.Soc., 136, 1991-1996, doi:10,1002/qj699, 2010.
- Efficient non-linear Data Assimilation in Geophysical Fluid Dynamics Van Leeuwen, P.J., Computers and Fluids, doi:10,1016/j.compfluid.2010.11.011, 2011.
- Efficient non-linear Data Assimilation for ocean models of intermediate complexity Van Leeuwen, P.J., IEEE Statistical Signal Processing workshop, Nice, 28-30 June, doi: 10.1109/SSP.2011.5967700, 2011.
- Efficient fully nonlinear data assimilation for geophysical fluid dynamics Van Leeuwen, P.J. and M. Ades, In Press, Computers and Geosciences, http://dx.doi.org/10.1016/j.cageo.2012.04.015, 2012.
- An exploration of the Equivalent-Weights Particle Filter Ades, M., and P.J. Van Leeuwen, In Press, Q. J. Royal Meteorol. Soc., 2012.