Non-degenerate Particle Filters in high-dimensional systems

Peter Jan van Leeuwen

Javier Amezcua, Mengbin Zhu, Melanie Ades

Data Assimilation Research Centre

University of Reading

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

What is data assimilation ?

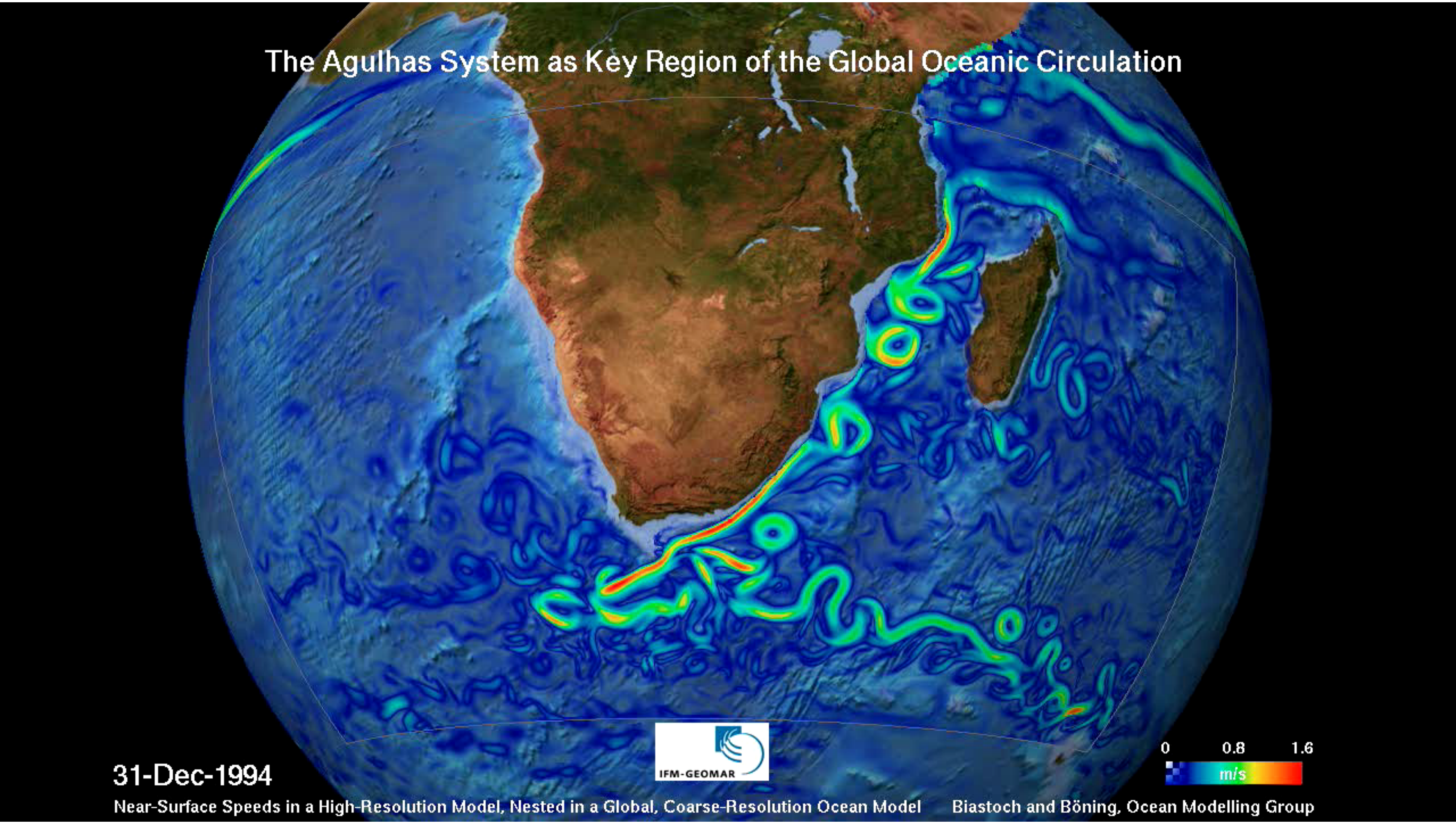
Combine

- a model estimate of a system including
- its uncertainties with
- observations of that system including
- their uncertainties to give
- a new model estimate including
- uncertainties.

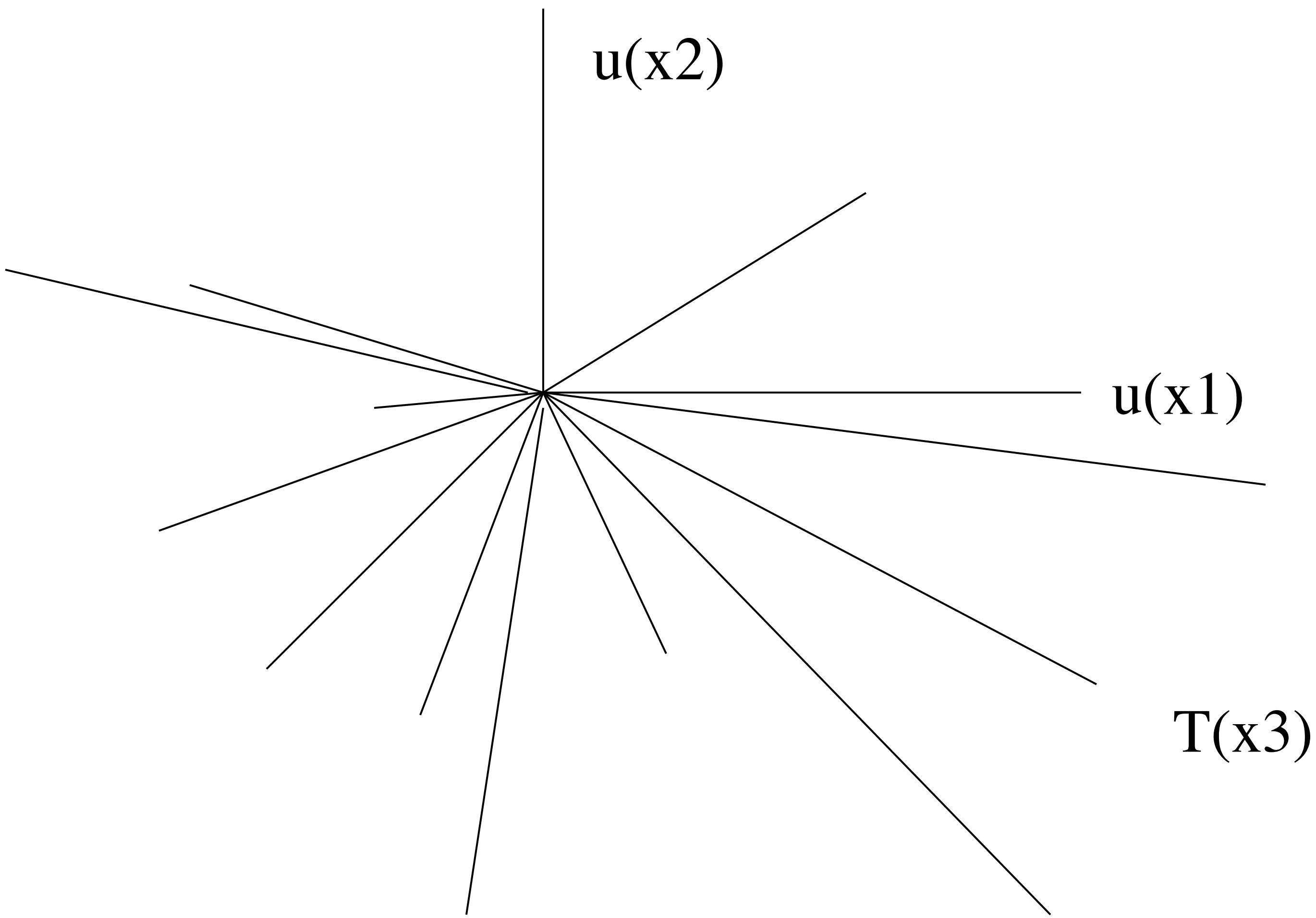
Big Data

- How big is the nonlinear data-assimilation problem?
- Assume we need 10 frequency bins for each variable to build the joint pdf of all variables.
- Let's assume we have a modest model with a million variables.
- Then we need to store $10^{1,000,000}$ numbers.
- The total number of atoms in the universe is estimated to be about 10^{80}.
- So the data-assimilation problem is larger than the universe…

Example of numerical model



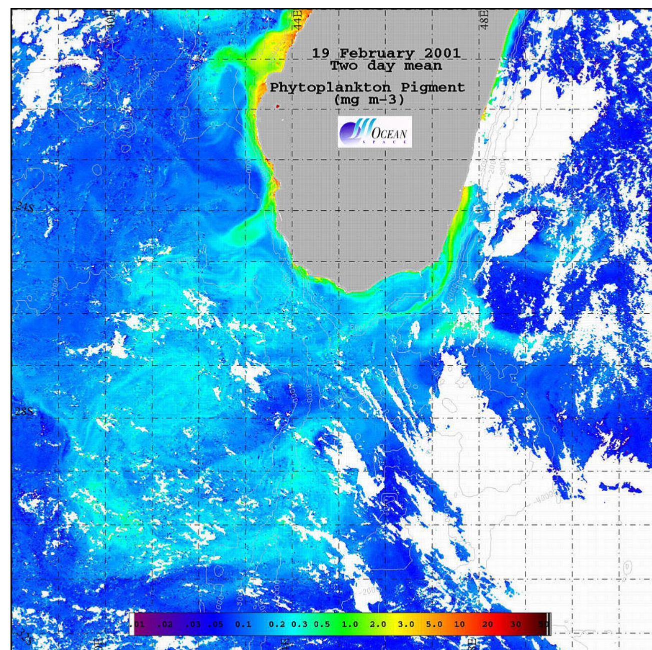
The model state space



Observations

- In situ observations:
  irregular in space and time e.g. sparse hydrographic observations,

- Satellite observations: indirect e.g. of the sea-surface



Notation

- Prior knowledge, the Stochastic model:

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

- Observations: y^n

- Relation between the two:

$$y^n = H(x^n_{truth}) + \epsilon^n$$

- With $x^n,\ x^n_{truth} \in \Re^{N_x}$ and $y^n \in \Re^{N_y}$

Nonlinear filtering: 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.

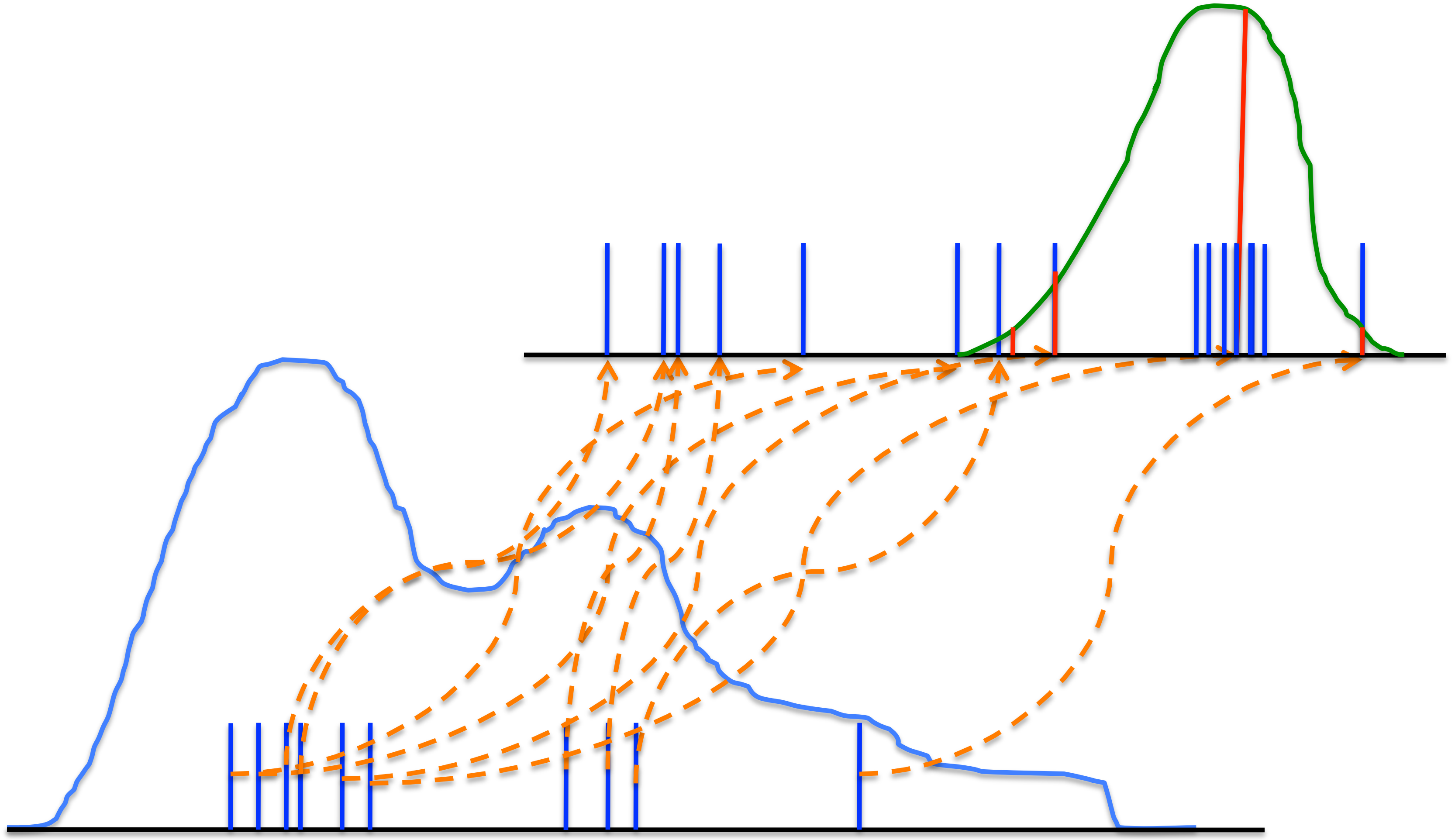
What are these weights?

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

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

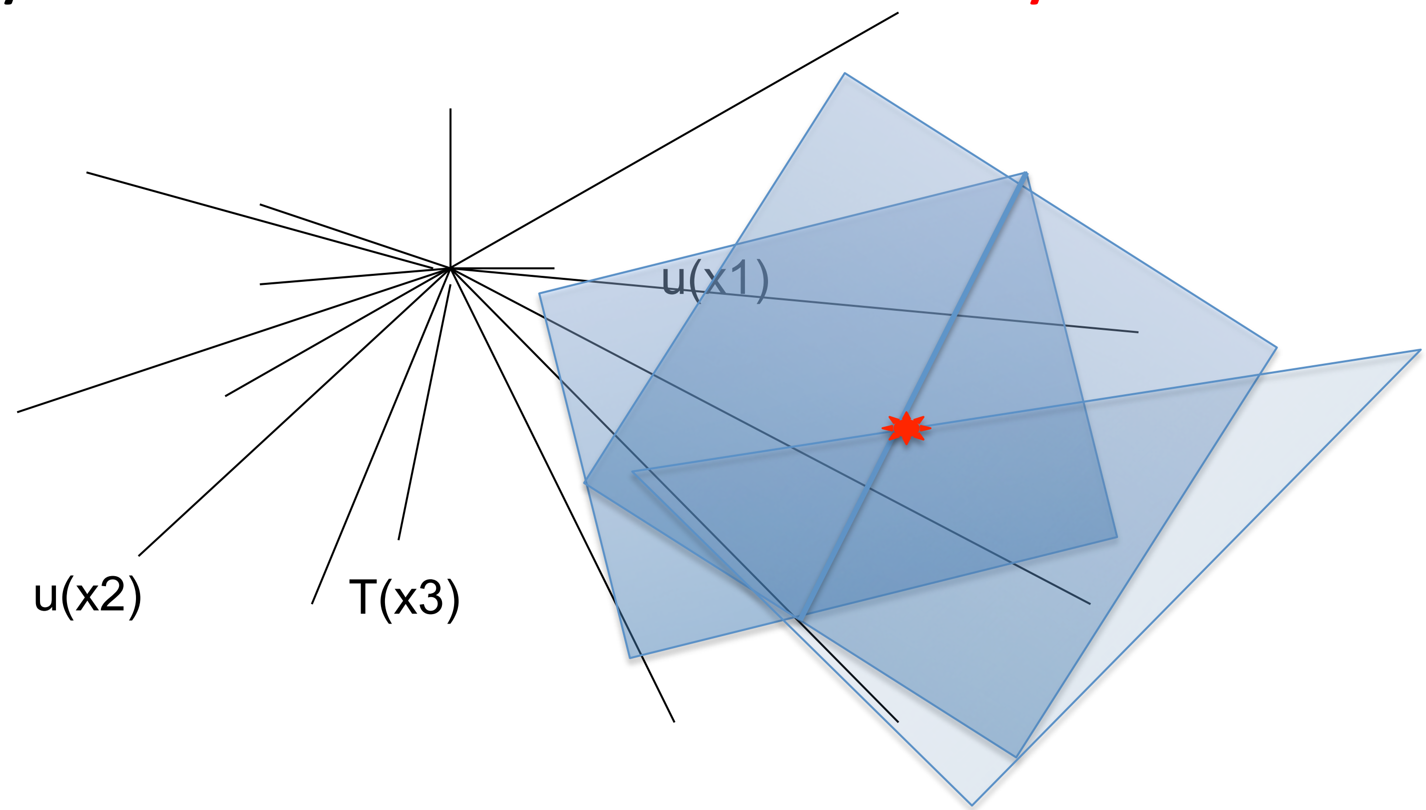
- One can just calculate this value
- That is all !!!

Standard Particle filter



A closer look at the weights I

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



A closer look at the weights II

Assume particle 1 is at 0.1 standard deviations *s* of M independent observations.
Assume particle 2 is at 0.2 *s* of the M observations.

The weight of particle 1 will be

$$w_1 \propto \exp\left[-\frac{1}{2}\ (y - H(x_i))\, R^{-1}\, (y - H(x_i))\right] = exp(-0.005M)$$

and particle 2 gives

$$w_2 \propto \exp\left[-\frac{1}{2}\ (y - H(x_i))\, R^{-1}\, (y - H(x_i))\right] = exp(-0.02M)$$

A closer look at the weights III

The ratio of the weights is

$$\frac{w_2}{w_1} = exp(-0.015M)$$

Take M=1000 to find

$$\boxed{\frac{w_2}{w_1} = exp(-15) \approx 3\ 10^{-7}}$$

Conclusion: the number of independent observations is responsible for the degeneracy in particle filters.

A closer look at the weights IV

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

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

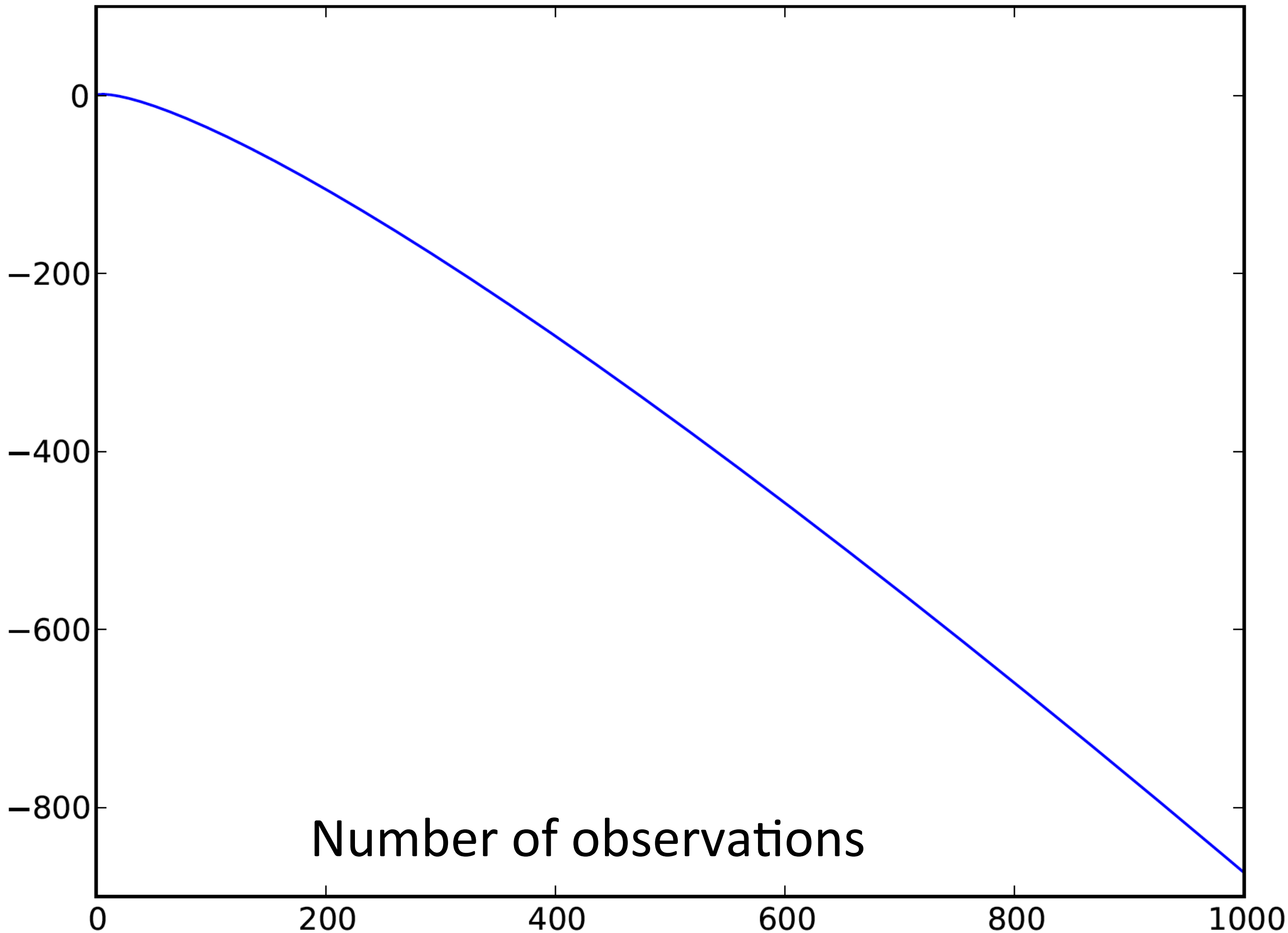
- Taking for the radius $r \approx 3\sigma_y$ we find, using Stirling:

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

- So very small indeed.

The volume in hyperspace occupied by observations

Log$_{10}$ of Volume of hyperball of radius 1



How can we make particle filters useful?

The joint-in-time prior pdf can be written as:

$$p(x^n, x^{n-1}) = p(x^n|x^{n-1})p(x^{n-1})$$

So the marginal prior pdf at time *n* becomes:

$$p(x^n) = \int p(x^n|x^{n-1})p(x^{n-1})\ dx^{n-1}$$

We introduced the **transition densities**

$$p(x^n|x^{n-1})$$

Meaning of the transition densities

Stochastic model:

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

Assume Gaussian distributed model errors:

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

This leads to:

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

Bayes Theorem and the proposal density

Bayes Theorem now becomes:

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

We have a set of particles at time *n-1* so we can write

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

and use this in the equation above to perform the integral:

The magic: the proposal density

Performing the integral over the sum of delta functions gives:

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

The posterior is now given as a sum of transition densities. In the standard particle filter we use these to draw particles at time *n*, which, remember, is running the stochastic model from time *n-1* to time *n*. We know that is degenerate.

So we introduce another transition density, the proposal.

Use a different transition density

Finite dimensional system

$$x^n,\ x^n_{truth} \in \Re^{N_x} \qquad\qquad y^n \in \Re^{N_y}$$

Bayes' Theorem:

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

Transition densities

- Proposal that depends on previous model state:

$$q(x_i^n|...) = p(x_i^n|x_i^{n-1})$$

- Proposal that depends on observations and previous model state:

$$q(x_i^n|...) = q(x_i^n|x_i^{n-1}, y^n)$$

e.g. optimal proposal density: $q(x_i^n|...) = p(x_i^n|x_i^{n-1}, y^n)$

- Proposal that depends on observations and all we know about previous state:

$$q(x_i^n|...) = q(x_i^n|x_{1:N}^{n-1}, y^n)$$

This leads to a whole class of particle filters not hampered by classical proofs of degeneracy.

Implicit Equal-weight Particle Filter

Define an *implicit map* as follows:

$$x_i^n = x_i^a + \sqrt{\alpha_i} P_i^{1/2} \xi_i$$

x_i^a is the mode of the optimal proposal density,

P_i is the covariance of the optimal proposal density.

1) Draw ξ_i from *N(0,I)*
2) Normalise each element $\xi_i^{(j)} = \xi_i^{(j)} \sqrt{N_x / ||\xi_i||}$
3) Calculate α_i such that pre-weights of all particles equal to target weight (see next slides)
4) Draw *k* from [1,..,N_x] with equal probability
5) Draw $\xi_i^{(k)}$ again from *N(0,1)*
6) Recalculate x_i^n using $\xi_i^{(k)}$ and normalised ξ_i for $i \neq k$.

How to find α_i

Remember the new particle is given by:

$$x_i^n = x_i^a + \sqrt{\alpha_i} P_i^{1/2} \xi_i$$

in which x_i^a, P_i, ξ_i are known, ξ_i normalised . Use this in expression for weights and set all weights equal to a target weight:

$$w_i = \frac{p(x_i^n | x_i^{n-1}, y^n) p(y^n | x_i^{n-1})}{q(\xi)} \left|\left| \frac{dx}{d\xi} \right|\right| \cdot w_i^{prev}$$

and solve for α_i . At this stage all weights are equal by construction!

Solution for α_i

For Gaussian model errors and observation errors and *H* linear we find for α_i:

$$\alpha_i - \log \alpha_i = C - \phi_i$$

with solution

$$\alpha_i = -W_{0,-1}\left[-e^{-(1+C-\phi_i)}\right]$$

in which *W* is the Lambert W function.
We choose one of the two solutions W_0 or W_{-1} with equal probability.
Target weight chosen equal to lowest weight of all particles.

Resulting scheme:

- Effectively we use N_x-*1* random numbers to calculate α_i such that the weights of the particles are equal.
- Then one random number is chosen to ensure a proper map from ξ_i to x_i^n. Weights will vary, but only slightly.
- This last step also ensures that the proposal has full support.
- Scheme can be seen as adaptation of implicit particle filter (optimal proposal) to avoid weight collapse.
- Scheme is biased because target weight is set to lowest weight of all particles at their highest weight positions.
- The latter are equal to solution of weak-constraint 4Dvar without background term.

Experiments,

model error and observation errors Gaussian, *H* linear

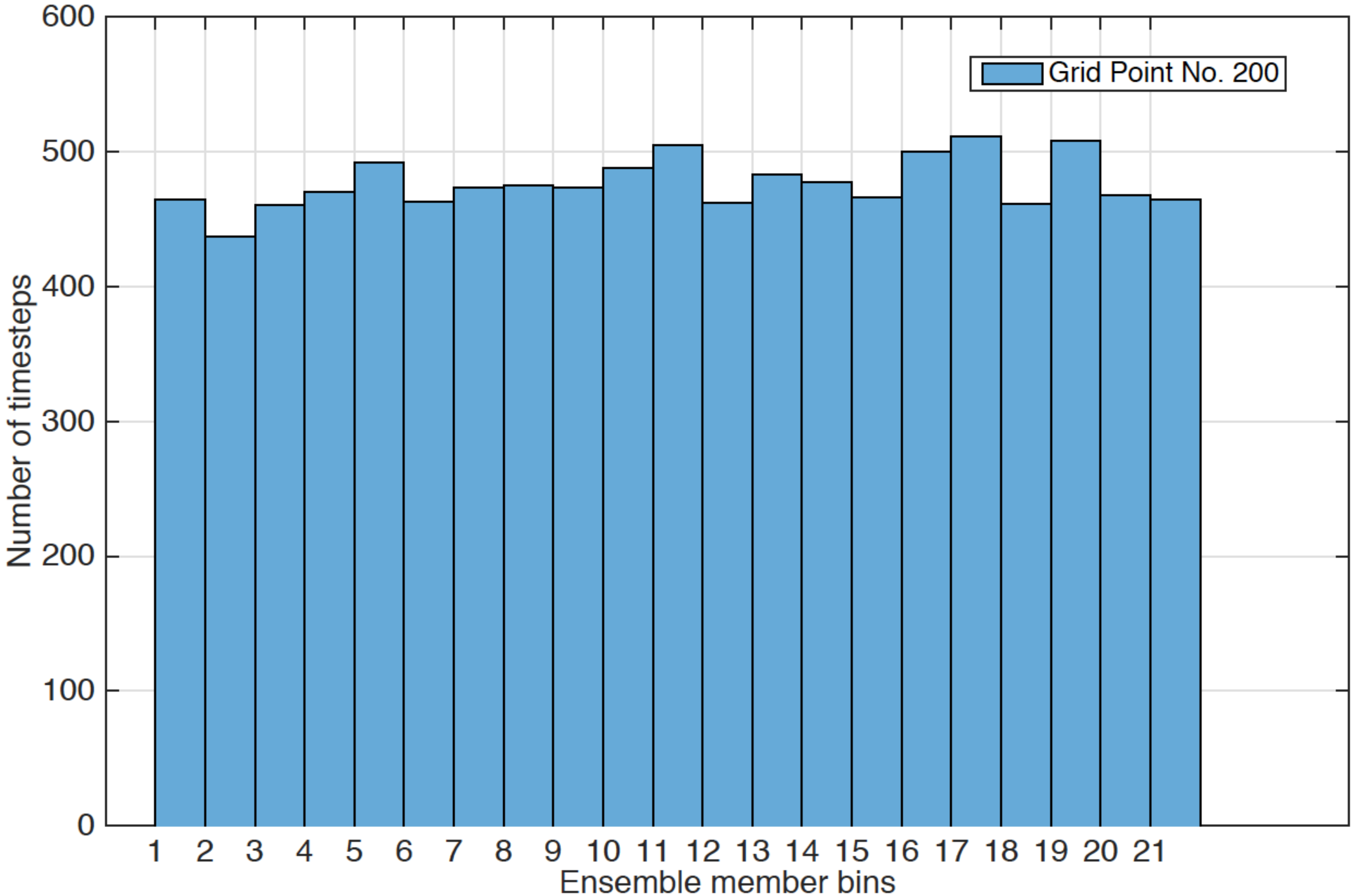
- Linear model of Snyder et al. 2008.
- 1000 dimensional independent Gaussian linear model
- 20 particles
- Observations every time step

- 1000 independent models Lorenz 1963 models
- 3000 variables, 1000 parameters
- 10 particles
- Observations every 10 time steps, first two variables.

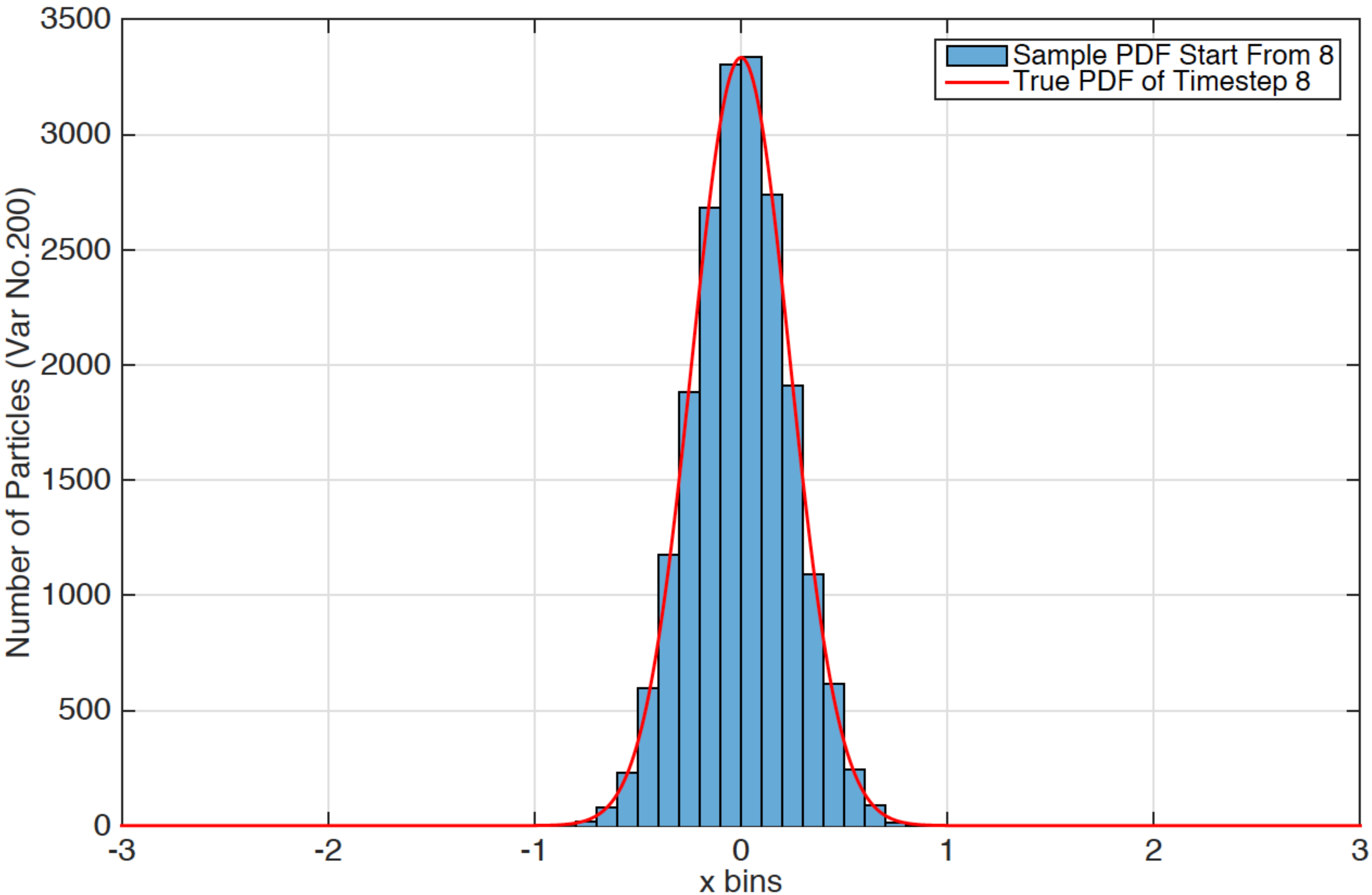
- Climate model 2.3 million variables, observe SST every day

1000-dimensional linear model: Rank histogram

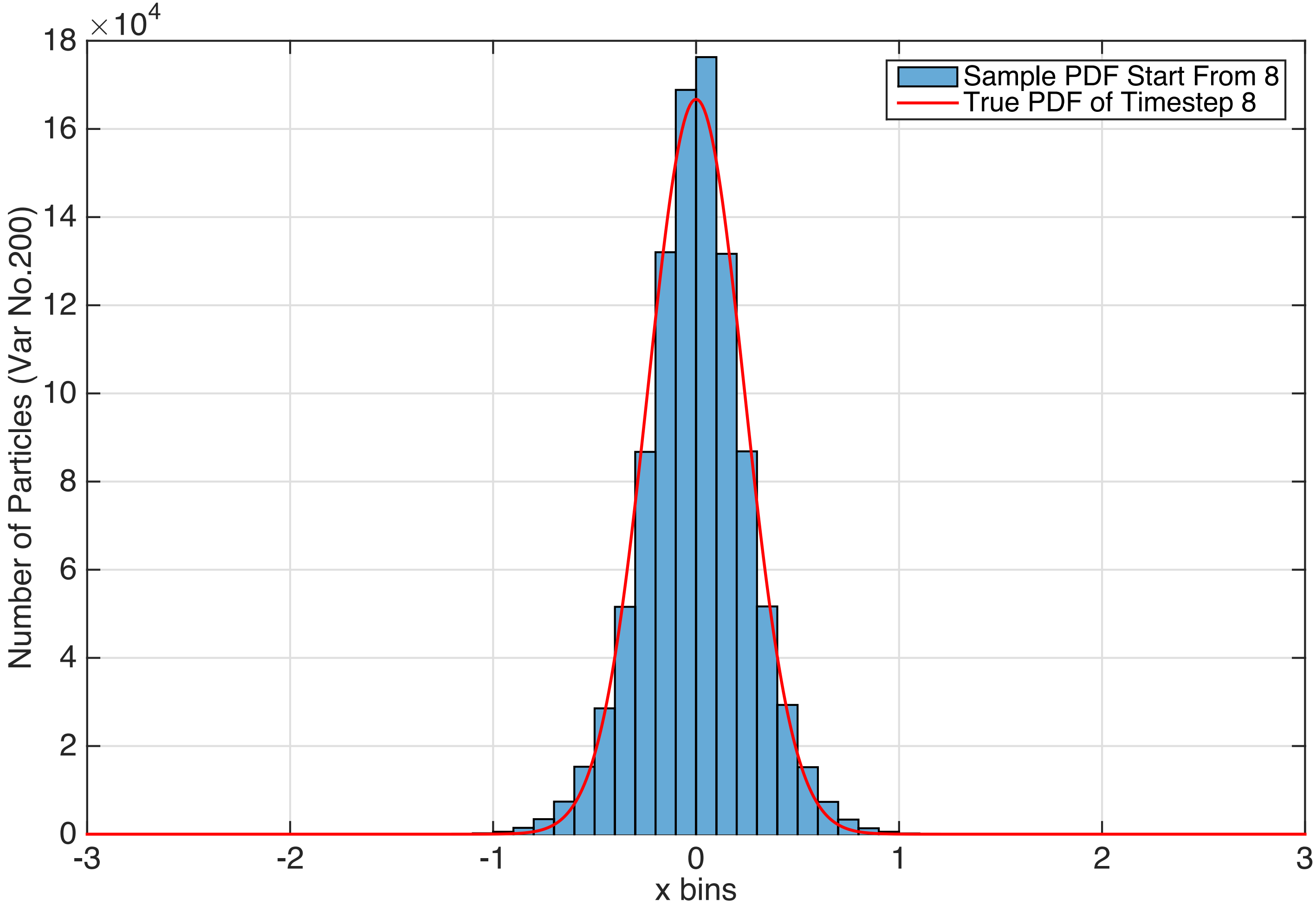
1000 time steps, 20 particles



Normalised pdf 1000 time steps
20 particles



Normalised pdf 1000 time steps 1000 particles -> small bias visible



Sequential parameter estimation

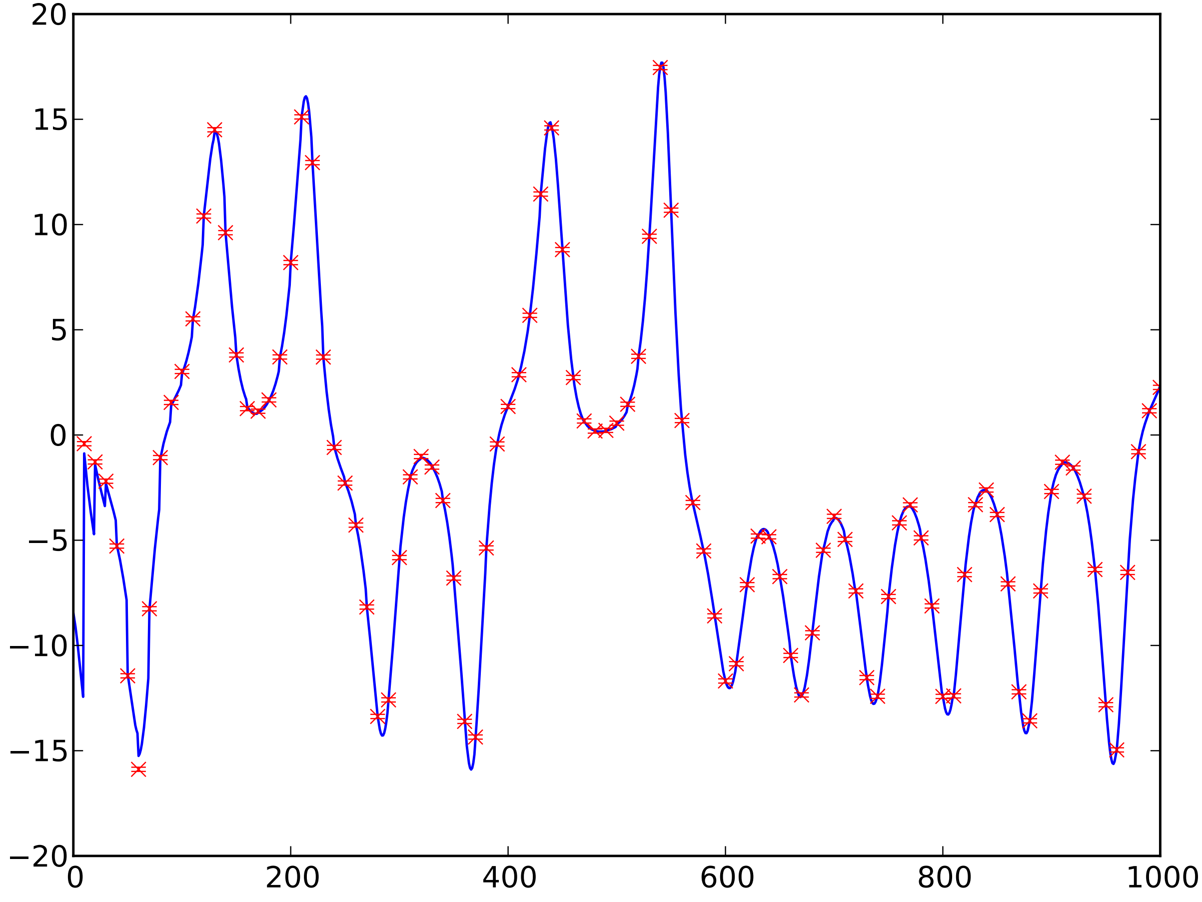
- SPDE $$x^n = f(x^{n-1}, \theta) + \beta^n$$
- Unknown parameter
$$x^n = f(x^{n-1}, \theta_0) + \frac{\partial f}{\partial \theta}(\theta - \theta_0) + \beta$$
- Model as $$\theta^n = \theta^{n-1} + \eta^n$$

and $$Q_{xx} = Q_\beta + \frac{\partial f}{\partial \theta} Q_\eta \frac{\partial f}{\partial \theta}^T$$

4000 dimensional system (3000 variables, 1000 parameters).

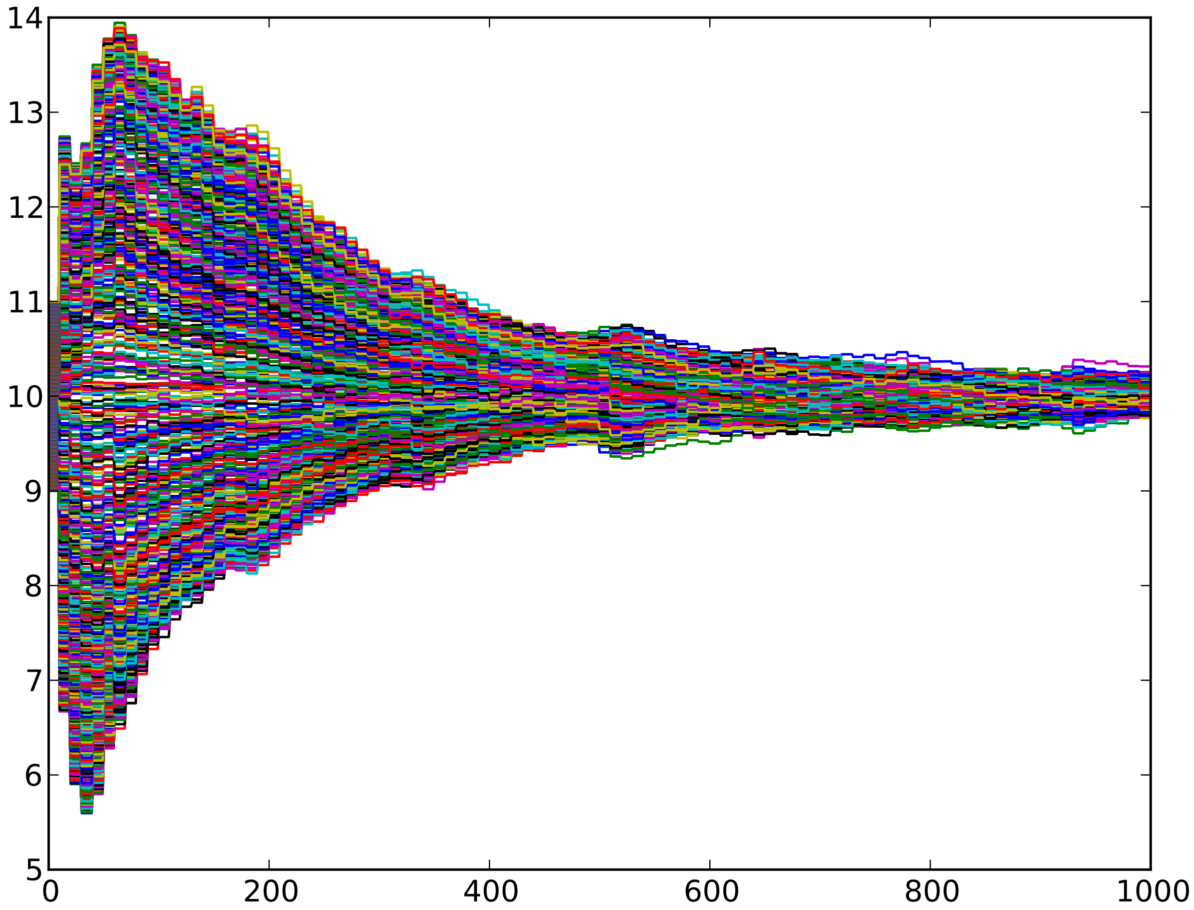
Evolution of first variable

Initial value 10 lower than true value.



Parameter mean values (dim=1000)

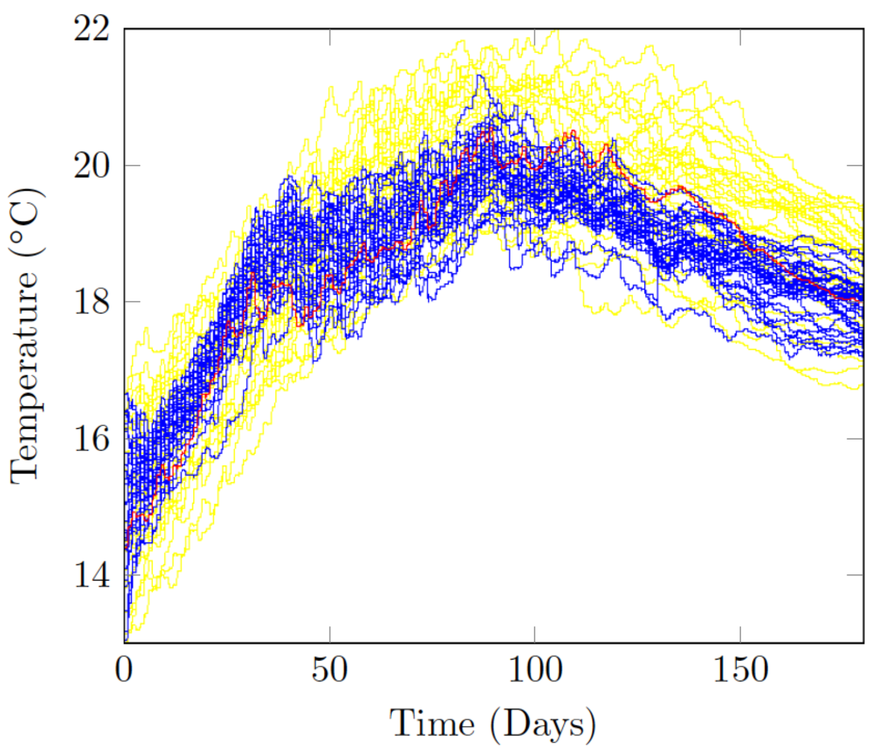
Initial values between 9 and 11, true value 10.

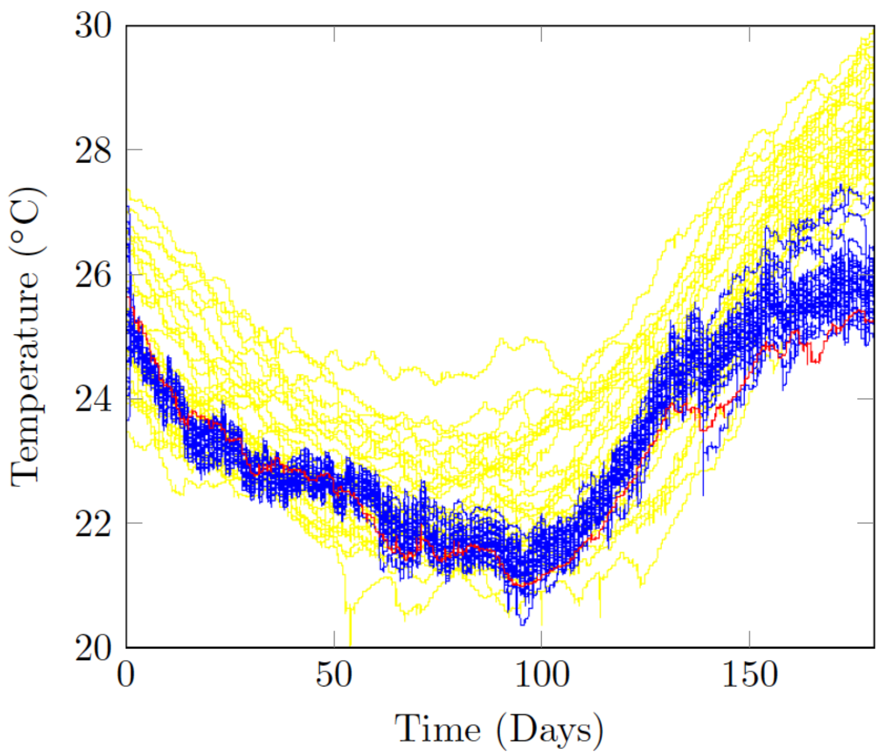


Climate model HadCM3

- Identical twin experiment
- 32 particles
- 2.3 million variables
- Daily observations of Sea-Surface Temperature with uncertainty 0.55 K
- Model errors smaller than 0.1 times deterministic model update
- Correlation structure from snapshots of long model run.

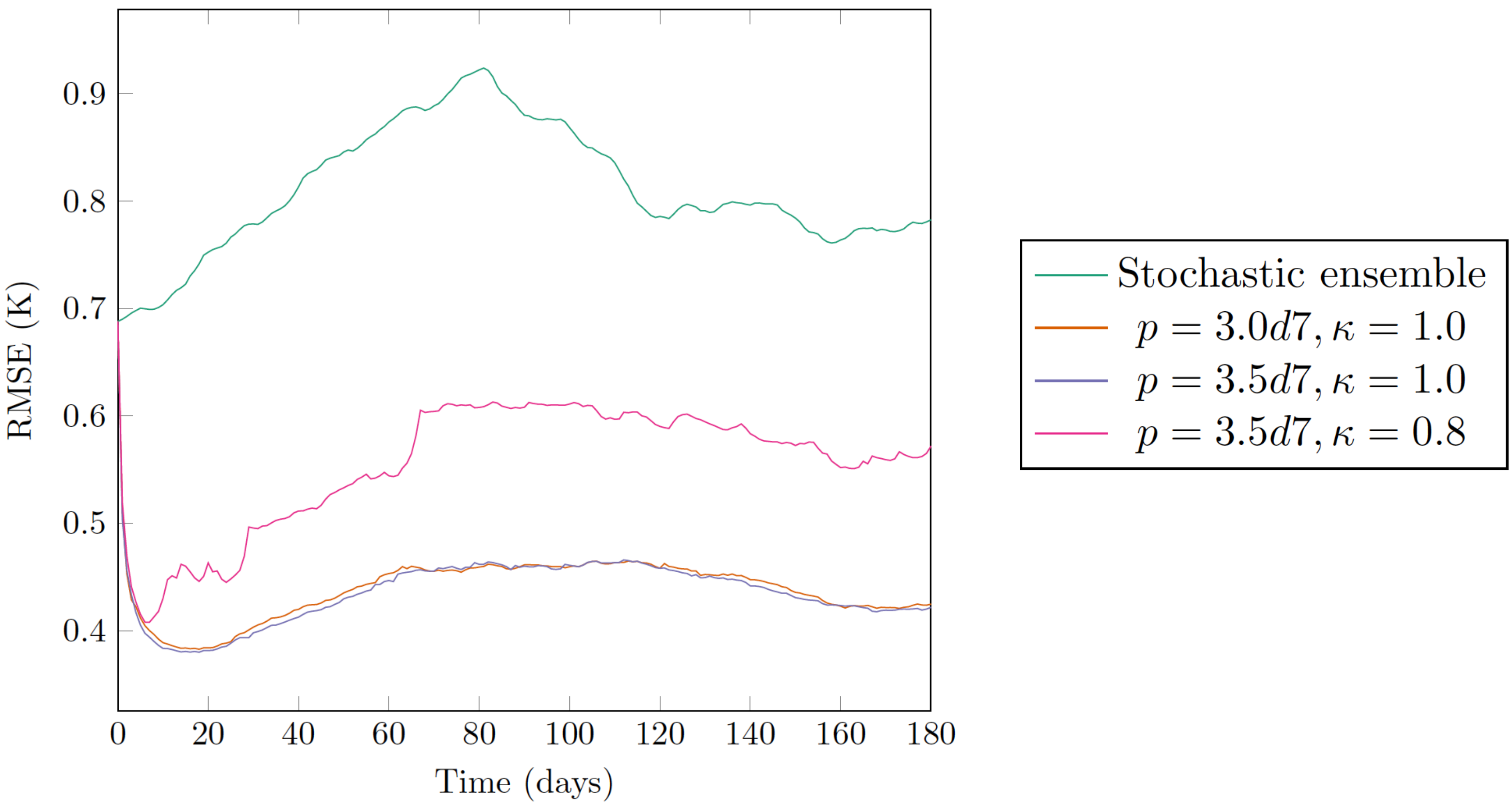
Time evolution of particles



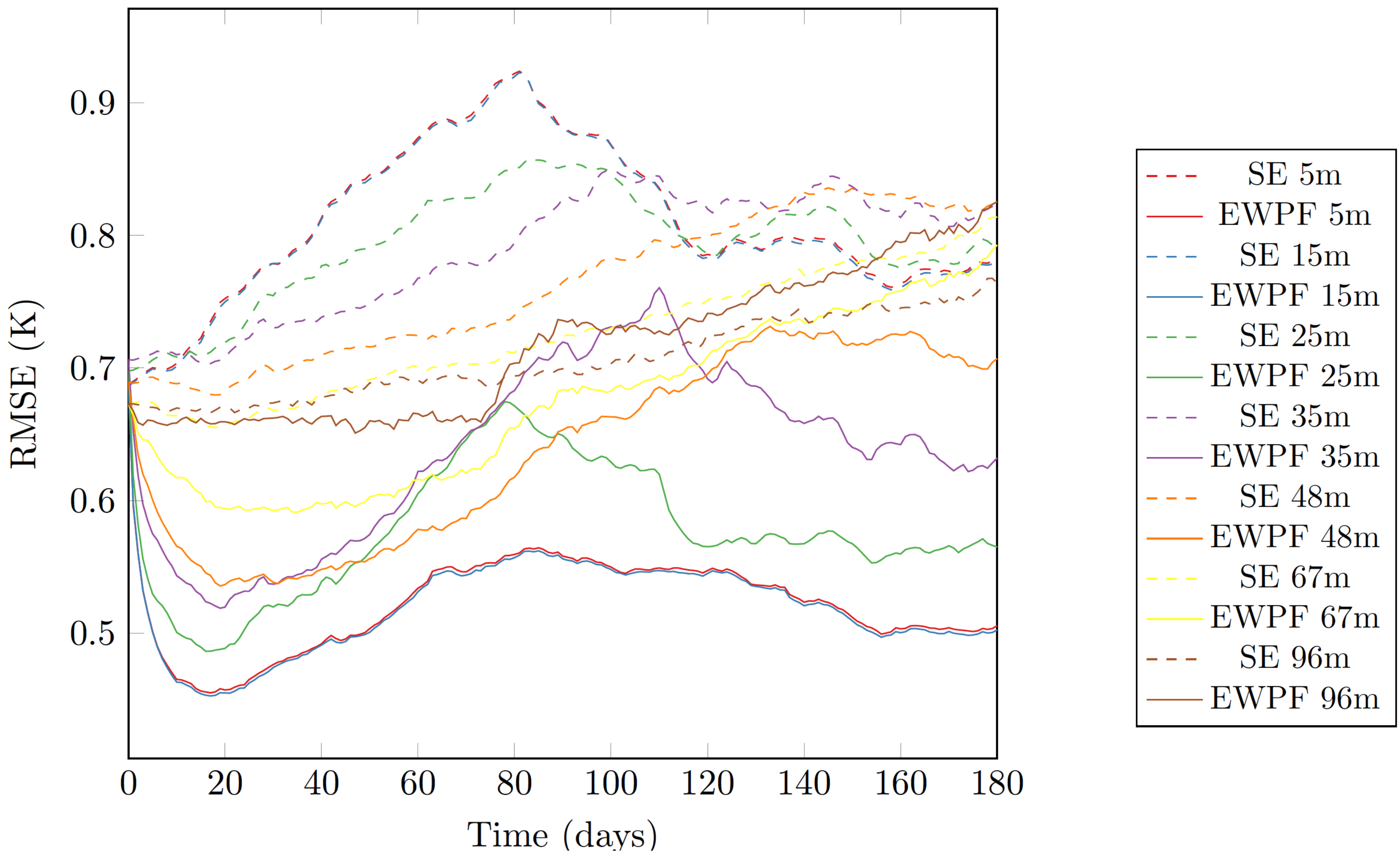


Prior ensemble (yellow), posterior ensemble (blue), truth (red), for SST in two grid points

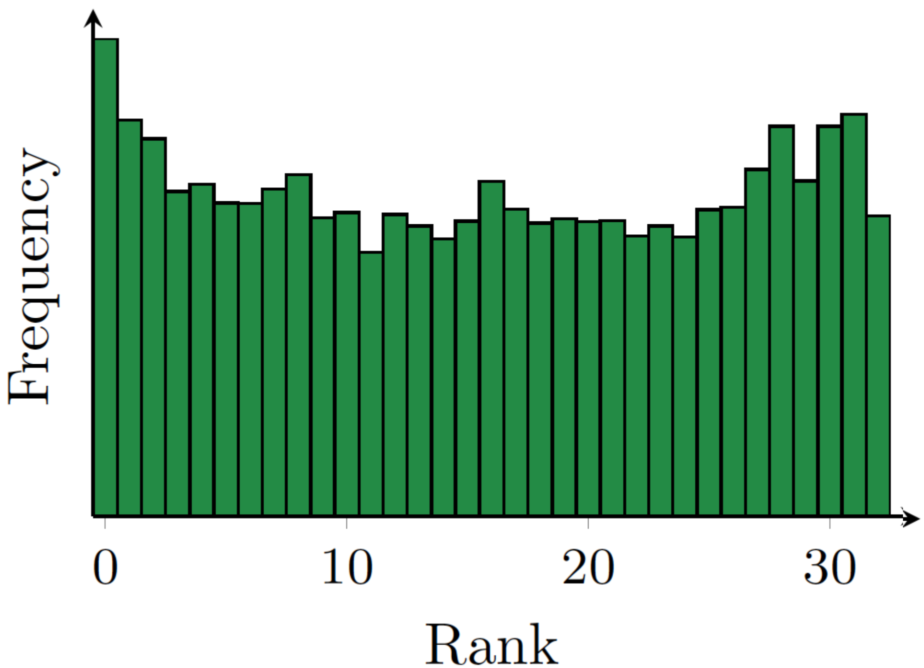
Results: Observed variable SST



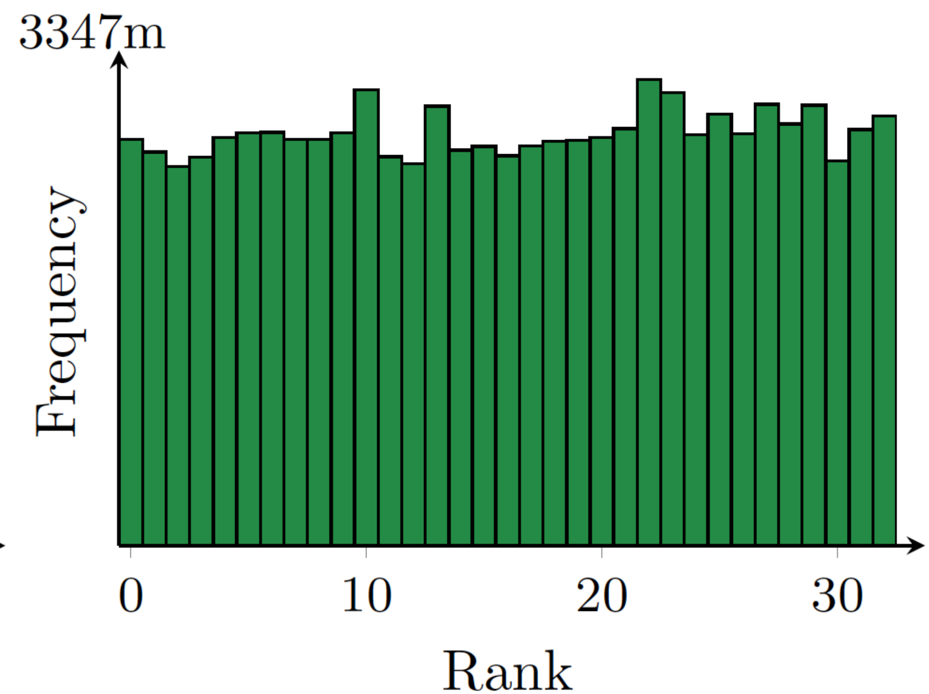
Results: Ocean Temperature



Rank Histograms

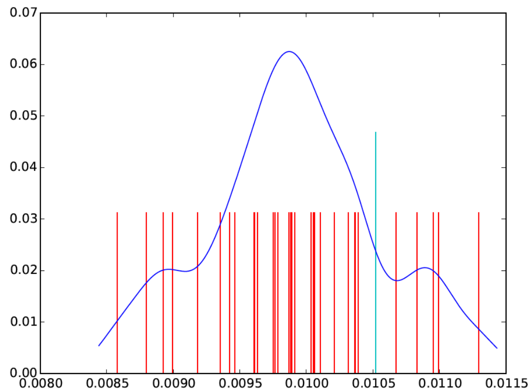


SST (observed)

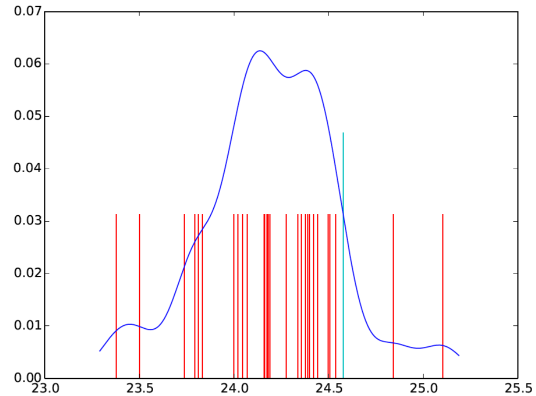


Meridional wind high up in Atmosphere (unobserved)

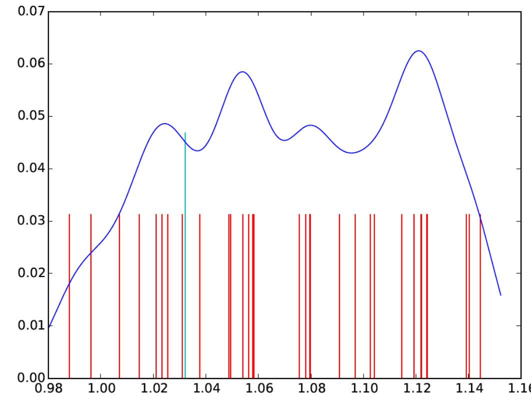
Estimated pdfs



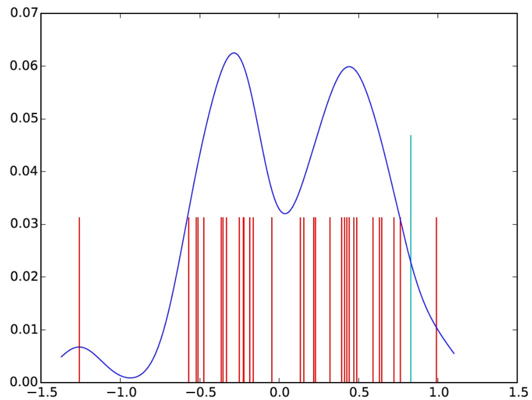
(a) Surface level specific humidity at (20S,270E)



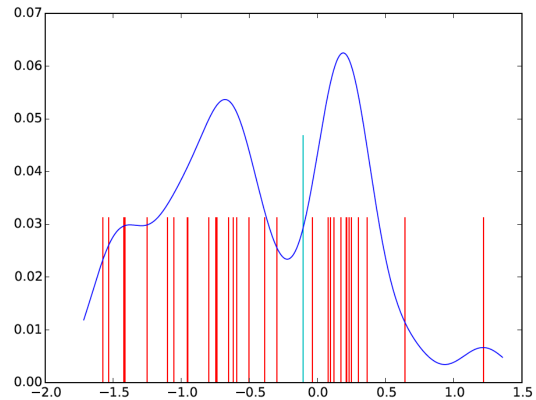
(b) Seawater temperature at (1.875S,213.75E) at a depth of 48m



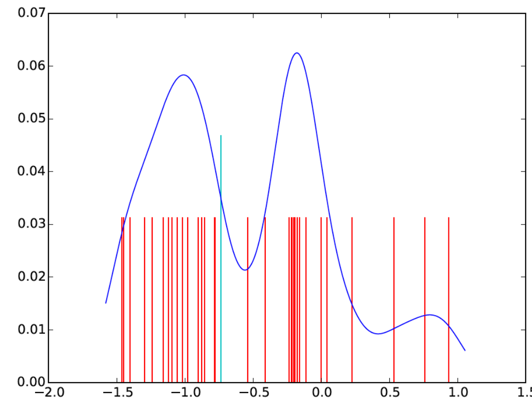
(c) Seawater temperature at (51.875S,228.75E) at a depth of 3347m



(d) Meridional seawater flow at (15S,100.625E) at 3347m depth on day 75



(e) Meridional seawater flow at (15S,100.625E) at 3347m depth on day 161



(f) Meridional seawater flow at (15S,100.625E) at 3347m depth on day 177

Conclusions

- Fully nonlinear non-degenerate particle filters for systems with arbitrary dimensions with small bias have been derived.
- Proposal-density freedom needs further exploration
- Examples shown for 1000 dimensional linear system, high-dimensional parameter estimation, 2.3 million dimensional climate model.
- Need to explore bias versus degeneracy issues.

- Implicit Equal-weights Particle Filter Zhu, M, P.J. van Leeuwen, and J. Amezcua, submitted to Q J Royal Meteorol. Soc., 2015
- Aspects of Particle Filtering in high-dimensional spaces. Van Leeuwen, P.J., in: Dynamic Data-Driven Environmental System Science, LNCS 8964, 251, Springer, doi:10.1007/978-3-319-25138-7, 2015.
- Nonlinear Data Assimilation. Van Leeuwen, P.J., Y. Cheng, and S. Reich., Springer, doi:10,1007/978-3-319-18347-3, 2015.
- Twin experiments with the Equivalent-Weights Particle Filter and HadCM3. Browne, P.A., and P.J. van Leeuwen, Q.J.Royal.Meteorol. Soc., 141, doi: 10.1002/qj.2621, 2015.
- The effect of the Equivalent-weights particle Filter on model balances in a primitive equation model Ades M., P.J. van Leeuwen Monthly Weather Rev. 143, 581-596, doi:10.1175/MWR-D-14-00050.1, 2015.
- The Equivalent-weights Particle Filter in a high-dimensional system Ades M., P.J. van Leeuwen Q.J.Royal.Met.Soc.,141,484-503,doi:10.1002/qj.2370, 2015.