Ensemble data assimilation @ Reading

Peter Jan van Leeuwen and Mel Ades

Data-Assimilation Research Centre DARC

University of Reading

Lorentz Center 2011

Particle filter

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

$\downarrow$ Use ensemble

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

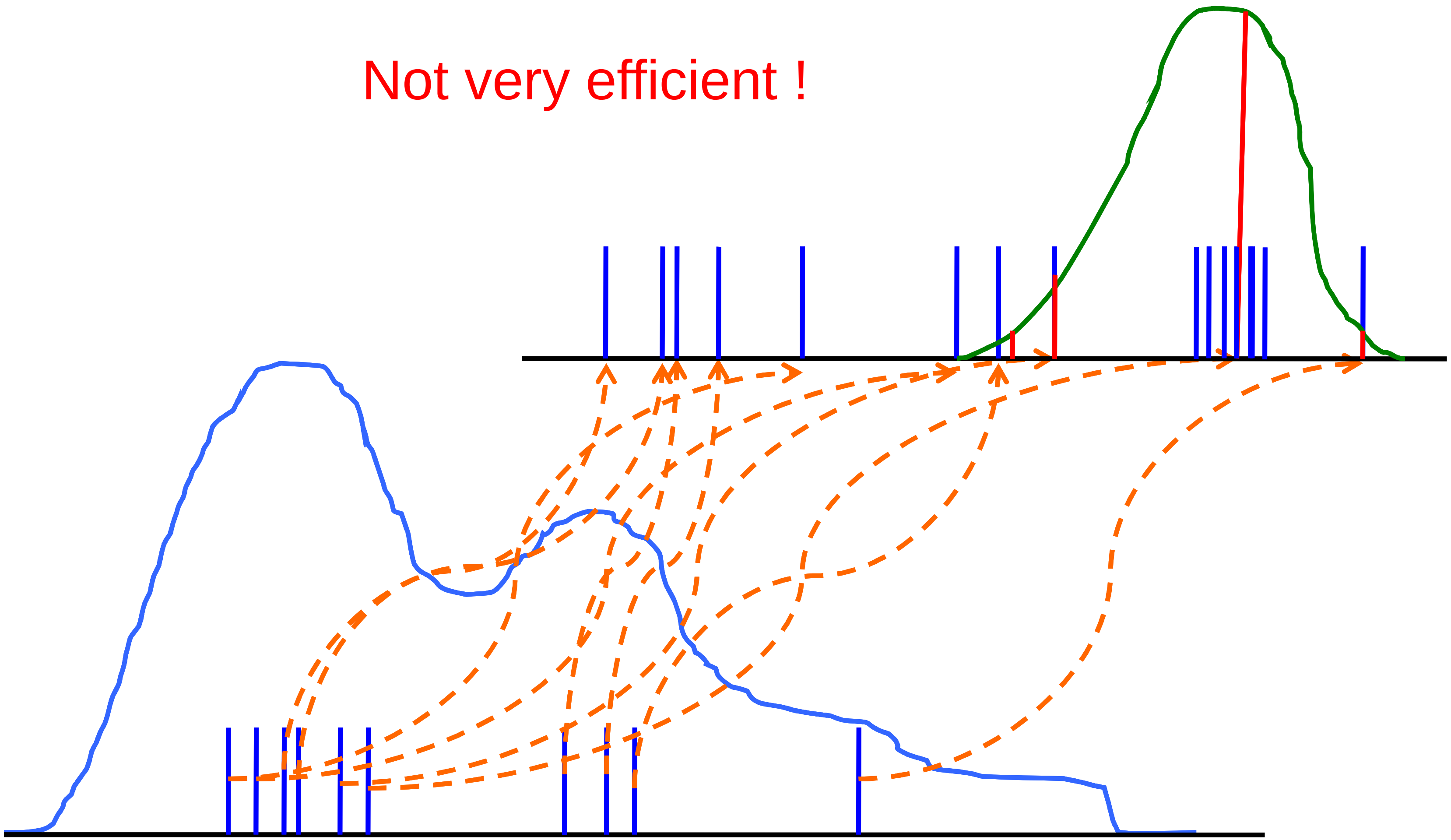
$$\boxed{p(x|y) = \sum_{i=1}^{N} w_i \delta(x - x_i)}$$

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

No explicit need for state covariances

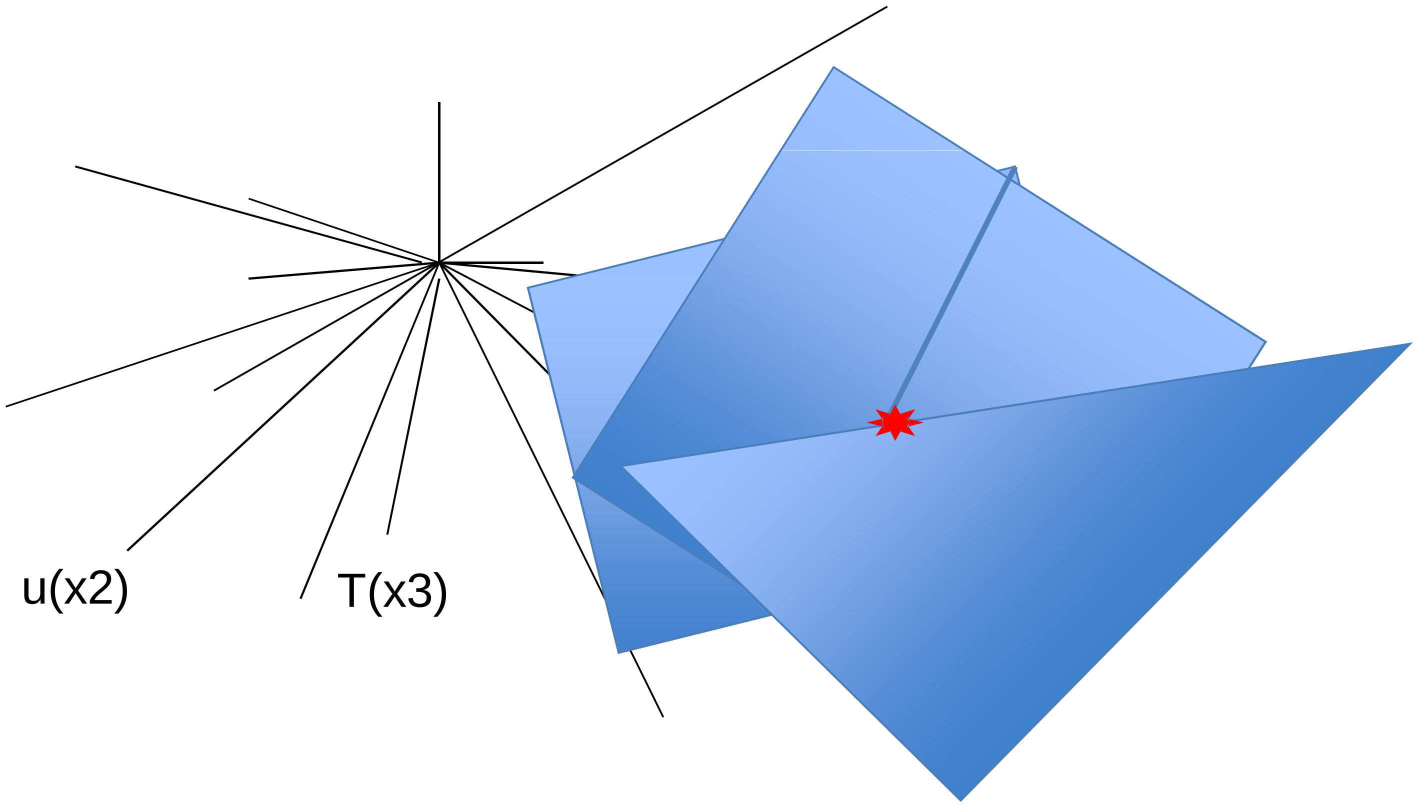
- 3DVar and 4DVar need a good error covariance of the prior state estimate: complicated
- The performance of Ensemble Kalman filters relies on the quality of the sample covariance, forcing artificial inflation and localisation.
- Particle filter doesn't have this problem, but…

Standard Particle filter



A closer look at the weights I

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



Exploring the proposal density

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

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

Which can be rewritten as:

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

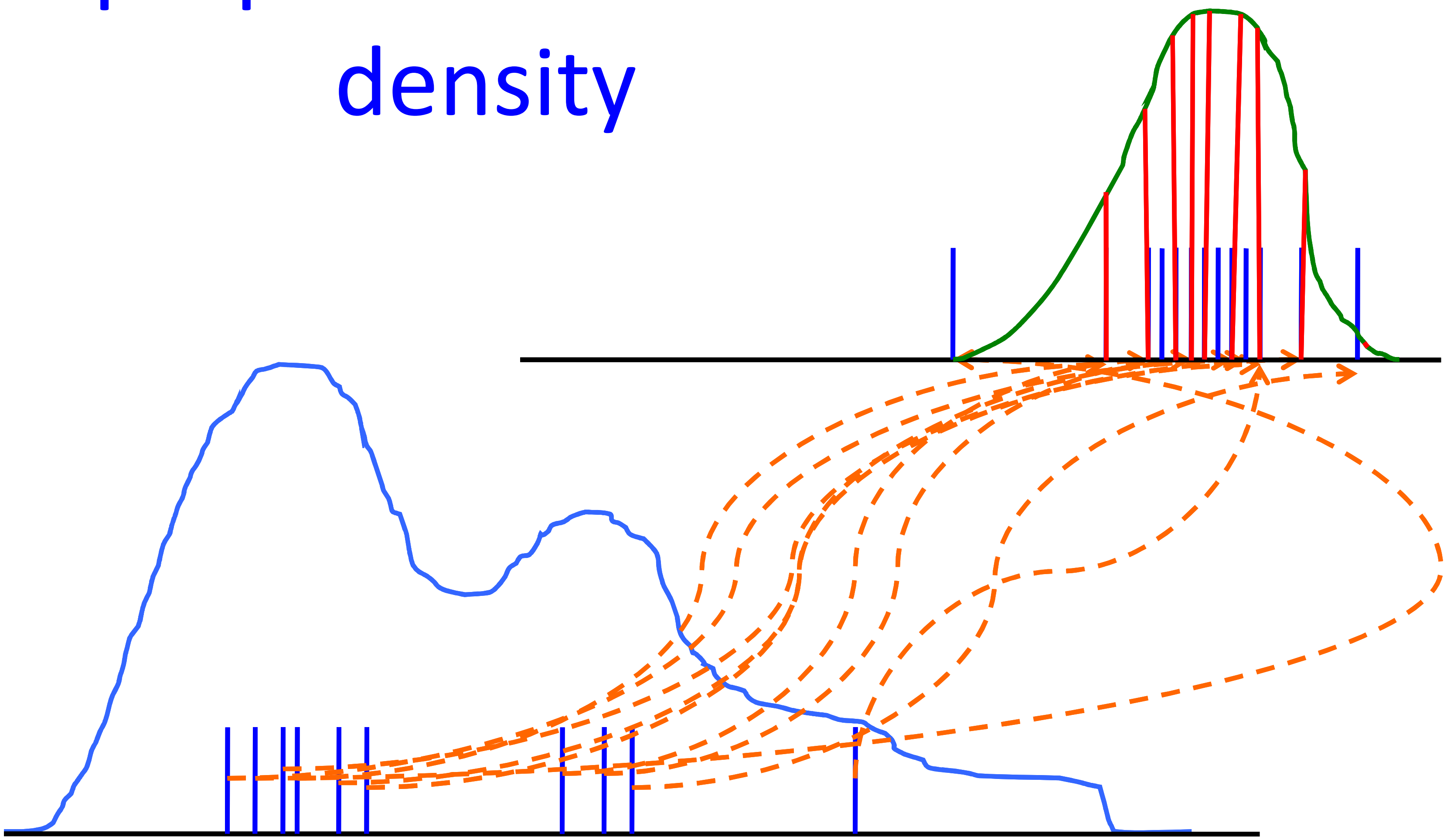
with weights

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

Likelihood weight

Proposal weight

Particle filter with proposal transition density



Equivalent weights I

1. We know:

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

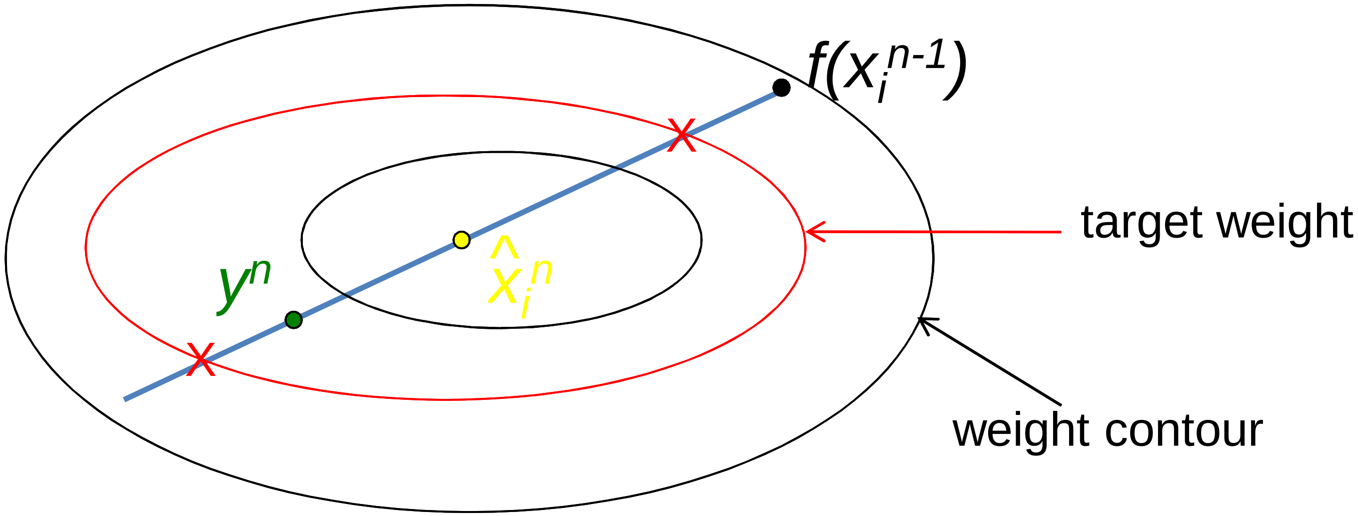
2. Write down expression for each weight ignoring *q* for now:

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

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

Equivalent weights II

5. Set a target weight that 80% of the particles can reach.



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

$$\boxed{x_i^n = f(x_i^{n-1}) + \alpha K \left(y^n - Hf(x_i^{n-1})\right)}$$

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

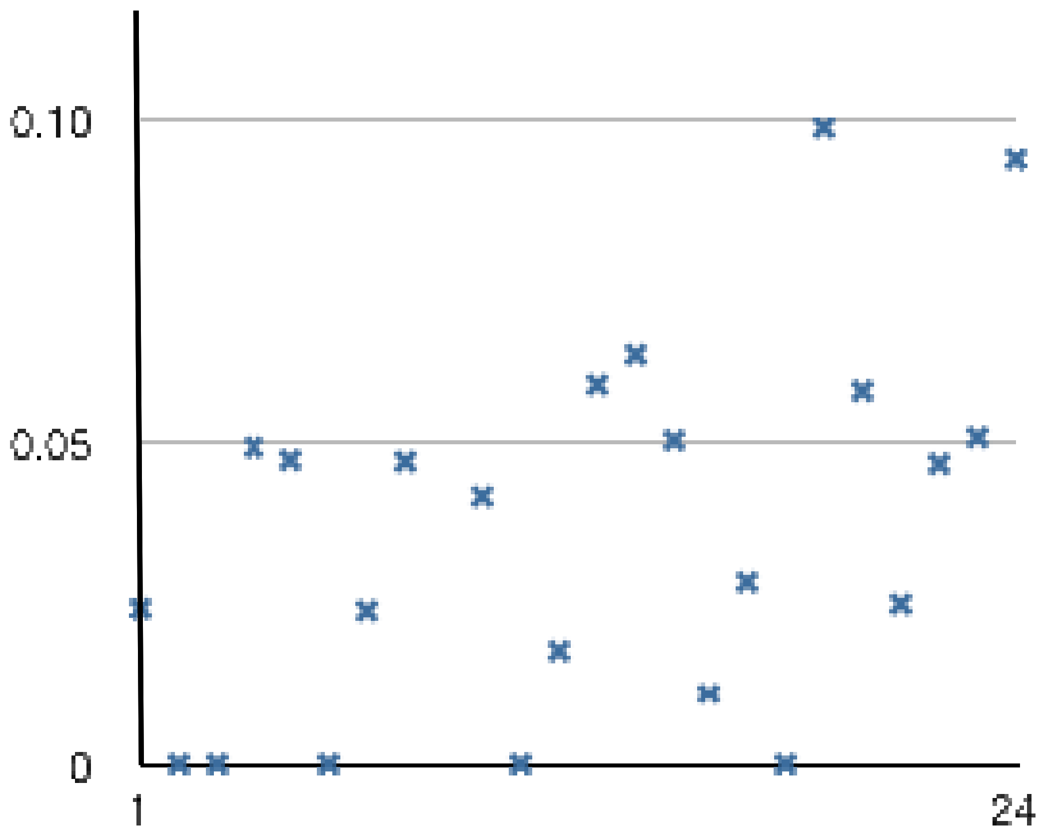
Barotropic vorticity equation

256 X 256 grid points
600 time steps
Typically q=1-3
Decorrelation time scale=
= 25 time steps

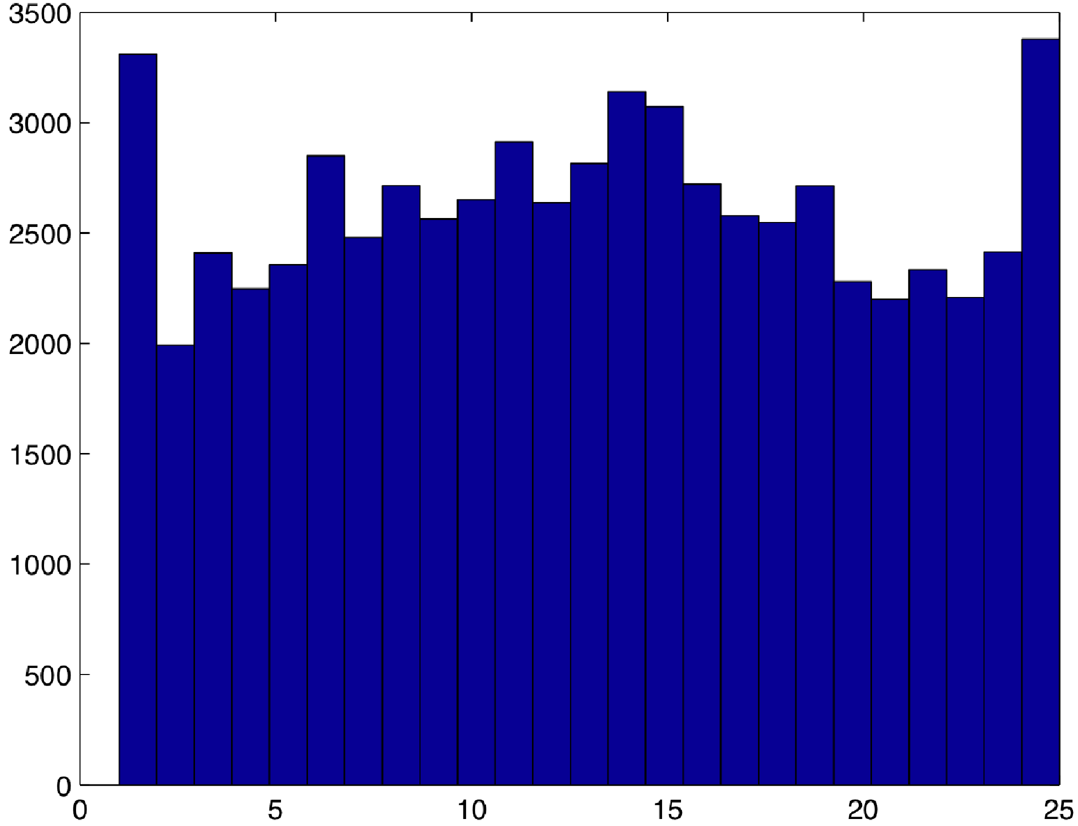
Observations
Every 4th gridpoint
Every 50th time step

24 particles
sigma_model=0.03
sigma_obs=0.01

Posterior weights



Rank histogram:
How the truth ranks in the ensemble



Equivalent Weights Particle Filter

Recall:

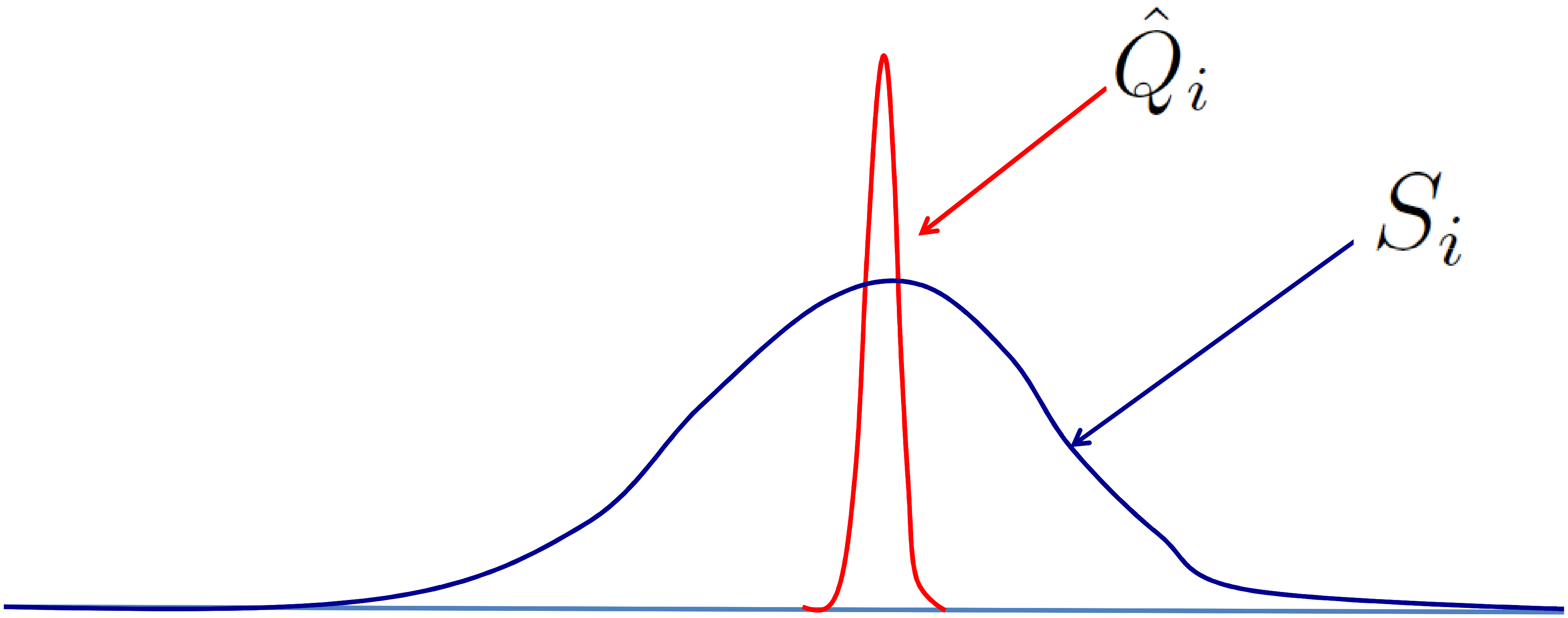
$$w_i = \frac{p(y|x_i)}{p(y)} \frac{p(x_i)}{q(x_i)}$$

Assume

$$\boxed{\begin{aligned} -2\log\left(p(y|x_i)p(x_i)\right) \quad \propto \quad & \left(x_i - x_0\right)^T P^{-1} \left(x_i - x_0\right) \\ & + \left(y - H(x_i)\right)^T R^{-1} \left(y - H(x_i)\right) \end{aligned}}$$

Find the minimum for each particle by perturbing each observation, gives x_i^{3DVAR}

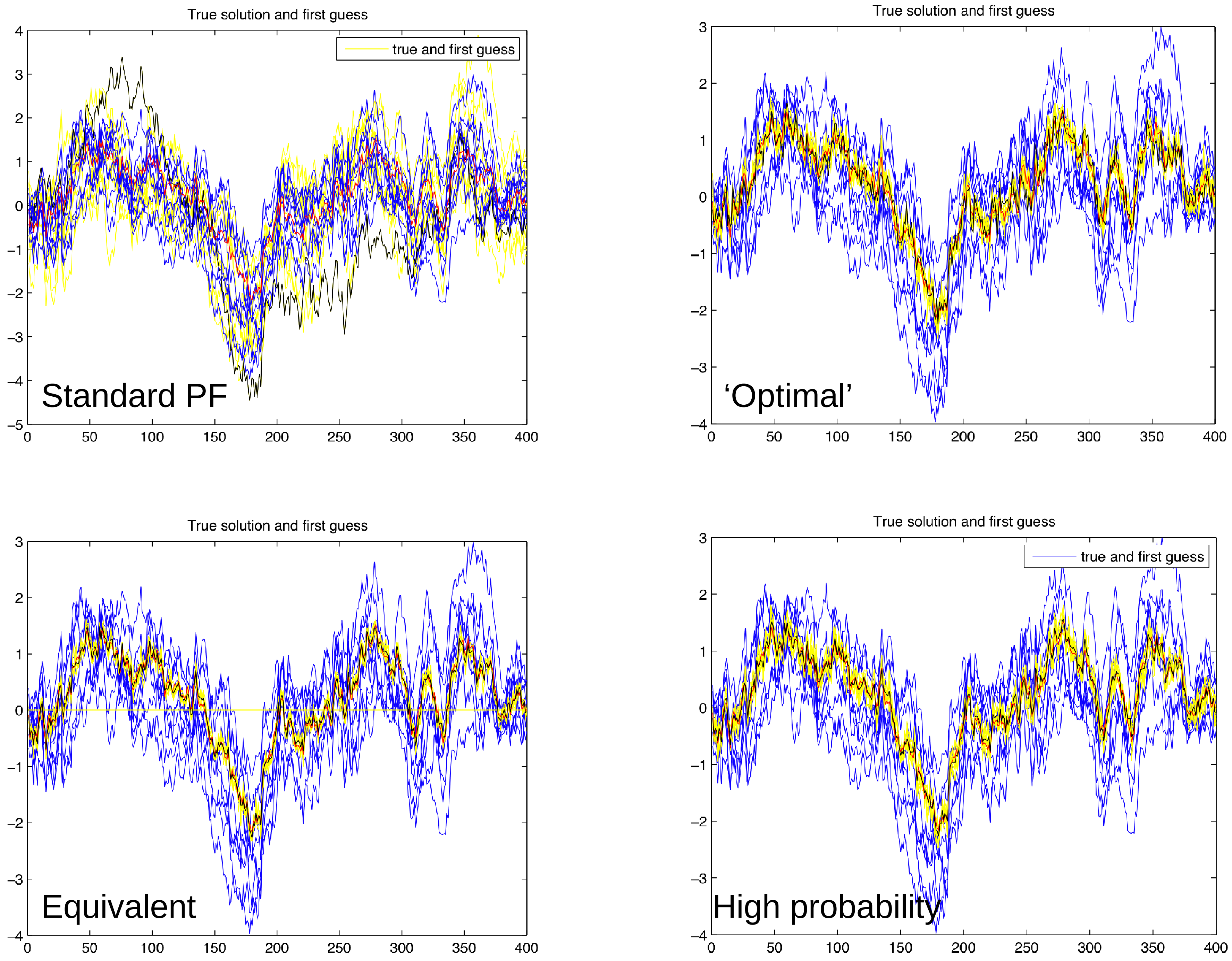
High-probability-weights Particle Filter



So, the idea is to draw from $N(0,\hat{Q}_i)$ and the weights come out as drawn from $N(0,S_i)$.

Example: one step, with equal weight ensemble at time *n-1*

- 400 dimensional system, Q = 0.5
- 200 observations, sigma = 0.1
- 10 particles
- Four Particle filters:
  - Standard PF
  - ‘Optimal’ proposal density
  - Almost equal weight scheme
  - Gaussian-peak weight scheme



True solution and first guess
true and first guess
Standard PF
True solution and first guess
'Optimal'
True solution and first guess
Equivalent
True solution and first guess
true and first guess
High probability

Performance measures

Effective ensemble size

$$\boxed{N_{eff} = \frac{1}{\sum_{i=1}^{N} w_i^2}}$$

Filter:	Squared difference from truth:	Effective ensemble size:
PF standard error	1.3931	1
PF-'optimal' error	0.10889	1
PF-Almost equal error	0.073509	8.8
PF-Gaussian Peak error	0.083328	9.4

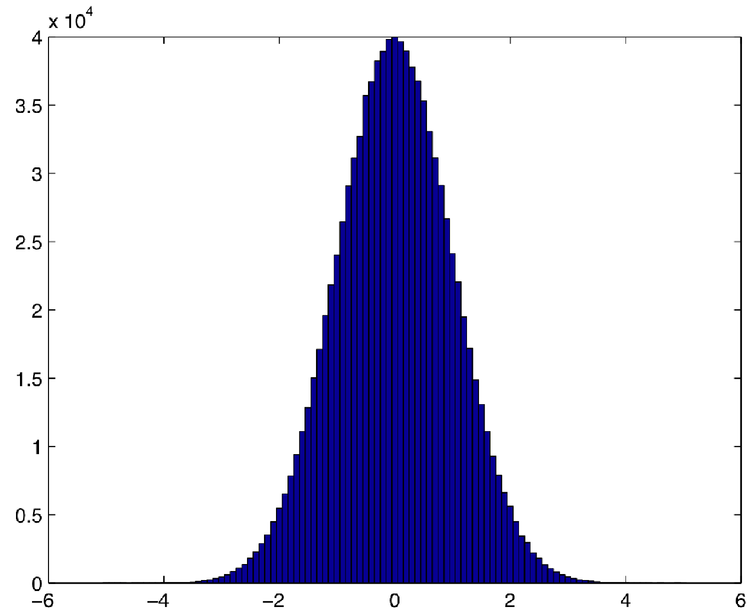
'Optimal' proposal density has no pdf information,
new schemes performing well.

Issues in high dimensions

Assume variables are iid Gaussian:

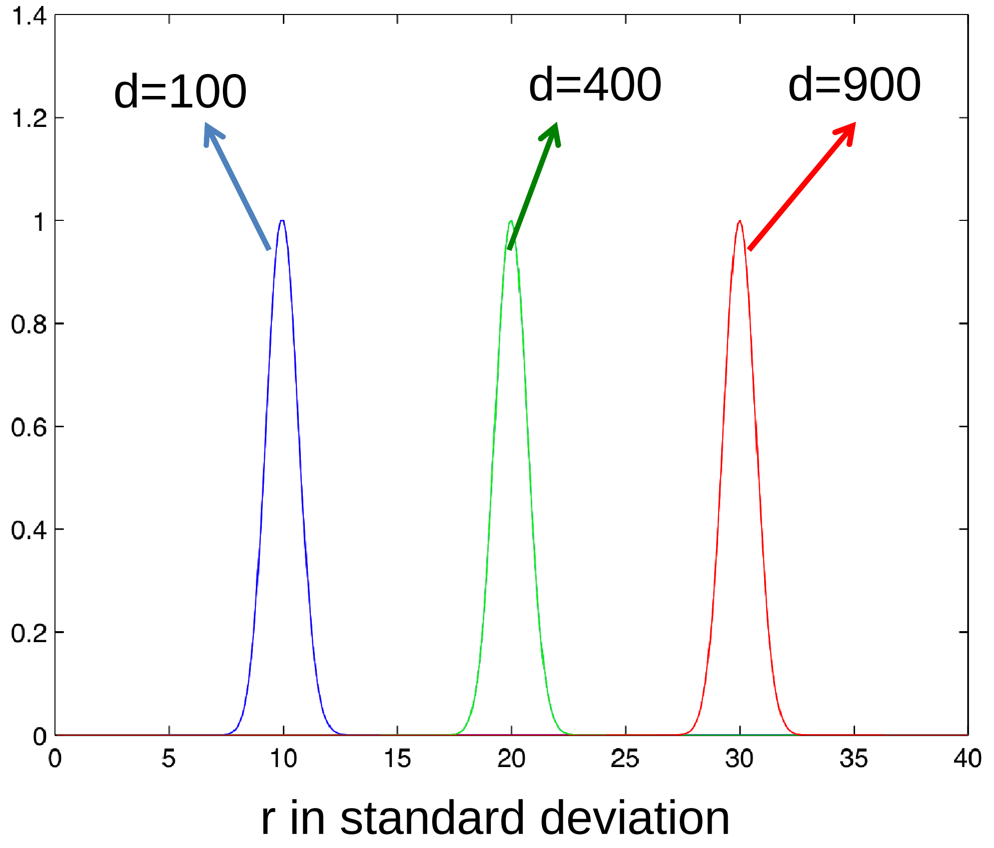
$$\boxed{p(x) = \frac{1}{\pi^{d/2}} \exp\left[-\sum_{i=1}^{d} x_i^2\right]}$$

Along each if the axes it looks like a standard Gaussian:

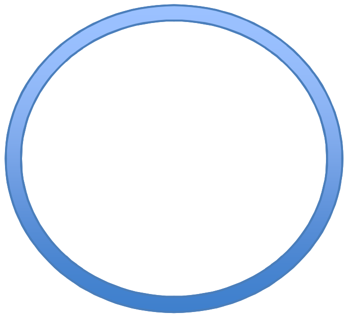


However, the probability mass as function of the distance to the centre is given by:

$$\boxed{mass(r) = e^{-r^2} r^{d-1} \int d\Omega}$$



The so-called
Important Ring



Why?

In distribution
$$\boxed{r^2 = \sum_{i=1}^{d} x_i^2 \quad \propto \quad \chi_d^2}$$

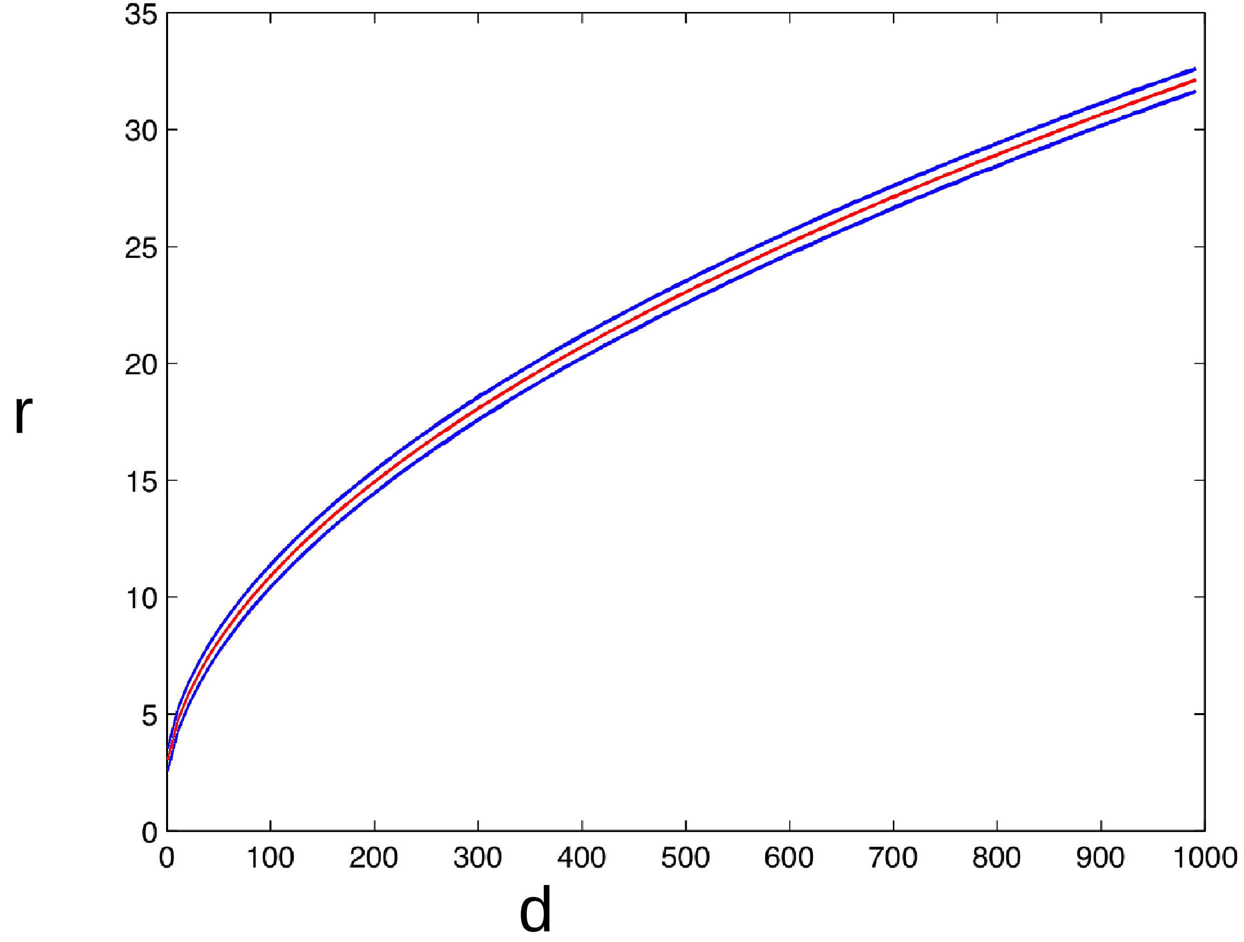
Fisher has shown
$$\boxed{\sqrt{2\chi_d^2} \quad \propto \quad N(\sqrt{2d-1}, 1)}$$

So we find

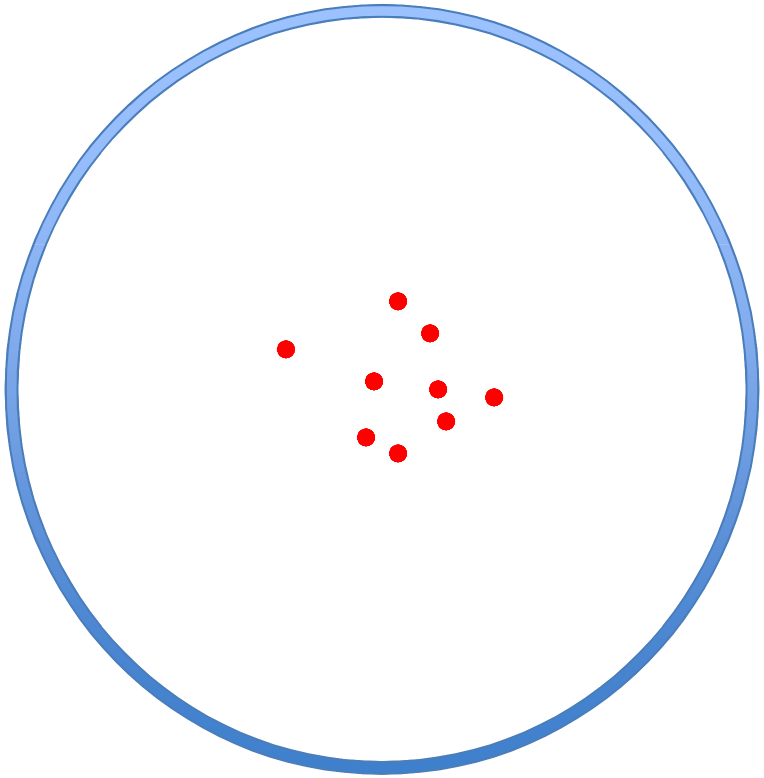
$$\boxed{r \quad \propto \quad N(\sqrt{d-1/2}, 1/2) \approx N(\sqrt{d}, 1/2)}$$

Importance Ring

Experimental evidence, sums of d squared random numbers:



Given this what do these efficient particles represent???



Nonlinear observation impact

- Degrees of freedom when prior is non-Gaussian
- Sensitivity matrix when prior is non-Gaussian
- Relative entropy
- Mutual information
- New measures…

A few conclusions

1. Particle filters with proposal transition density:
- solve for fully nonlinear posterior pdf
- very flexible, much freedom
- scalable => high-dimensional problems
- extremely efficient
- But what do the particles represent?

2. What information is present in ensemble covariances?

3. Non-linear observation impact needs non-linear DA method