Conditions for successful data assimilation

Matthias Morzfeld[*,**], Alexandre J. Chorin[*,**], Peter Bickel[#]

[*]*Department of Mathematics*
University of California, Berkeley

[**]Lawrence Berkeley National Laboratory

[#] *Department of Statistics*
University of California, Berkeley

Probabilistic Approaches to Data Assimilation for Earth Systems, Banff, Canada
Tuesday , 18th of February 2013

Goal

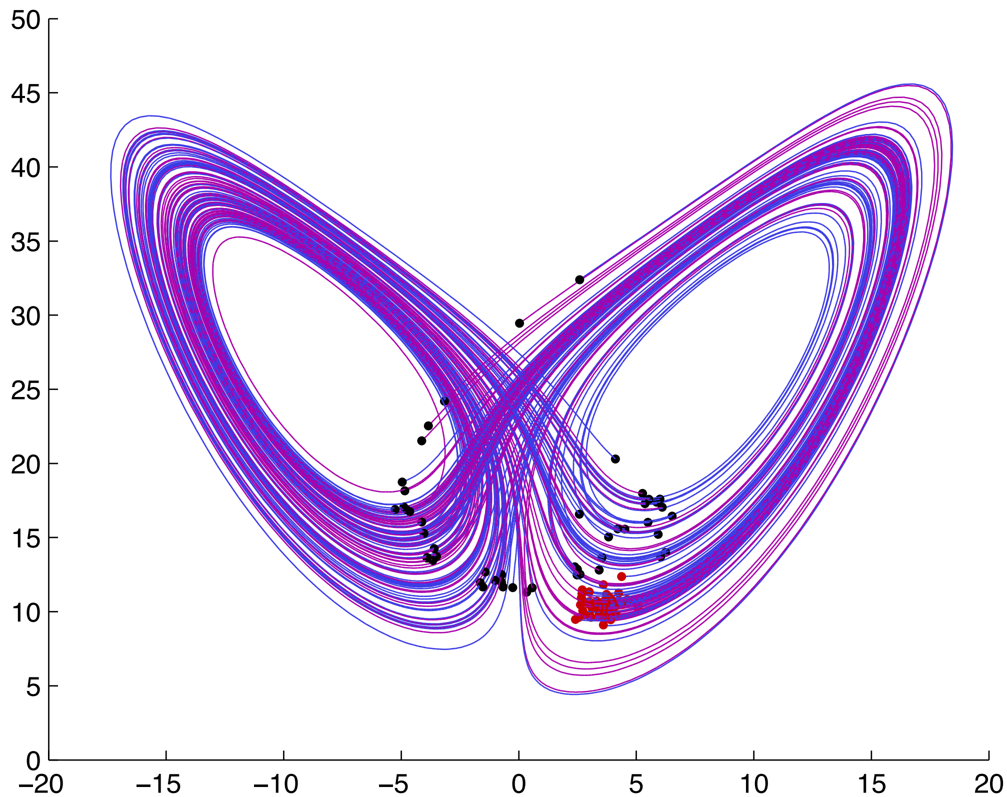
- Use a mathematical model to make predictions about a physical process

Solution

- Use noisy data to update the model

Problem

- Small errors grow quickly and become large errors



Solution of the data assimilation problem

Goal

- Compute the random variable

$$x|z \sim p(x|z)$$

- Conditional mean

$$E(x|z) = \int xp(x|z)dx$$

is the minimum mean square error estimate

Methods

- Kalman filter and its variants for linear Gaussian models
- Variational data assimilation computes the most likely state given the data
- Particle filters (Monte Carlo methods) construct empirical estimate of the conditional pdf

Question: What are the conditions under which DA can be successful?

Question: What are the conditions under which DA can be successful?

To answer the question we analyze the conditional pdf (which depends on errors in model and data)

Assumptions:

- Linear Gaussian synchronous model

$$x^{n+1} = Ax^n + w^n, \quad w^n \sim \mathcal{N}(0, Q), \text{ iid}$$

$$z^{n+1} = Hx^{n+1} + v^{n+1}, \quad v^n \sim \mathcal{N}(0, R), \text{ iid, independent of } w^n$$

- Initial state Gaussian

$$x^0 \sim \mathcal{N}(\mu_0, \Sigma_0)$$

Kalman formalism gives us the conditional pdf

Kalman filter

Model and data:

$$x^{n+1} = Ax^n + w^n, \quad w^n \sim \mathcal{N}(0, Q), \text{ iid}$$

$$z^{n+1} = Hx^{n+1} + v^{n+1}, \quad v^n \sim \mathcal{N}(0, R), \text{ iid, independent of } w^n$$

Kalman update:

$$\begin{aligned} X_n =& AP_nA^T + Q \\ K_n =& X_nH^T(HX_nH^T + R)^{-1} \\ P_{n+1} =& (I - K_nH)X_n \end{aligned}$$

In "steady state":

$$P_{n+1} = P_n = P = (I - KH)X$$

$$X = AXA^T - AXH^T(HXH^T + R)^{-1}HXA^T + Q$$

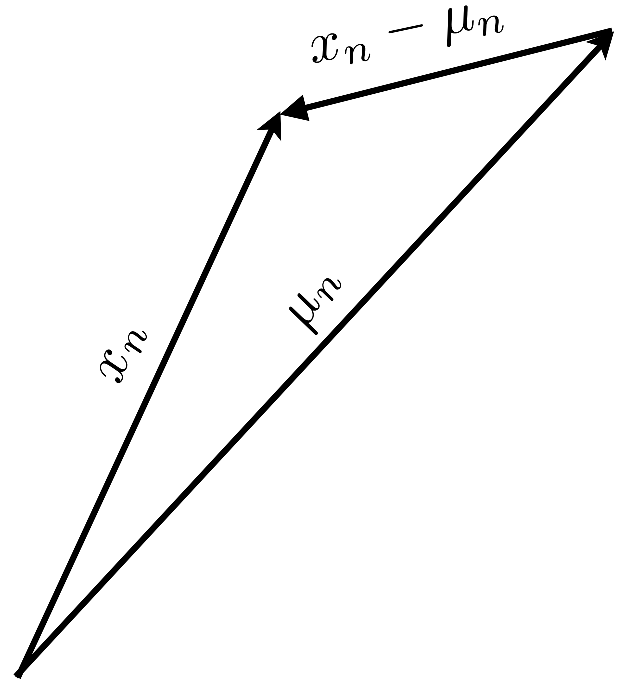
(Discrete Algebraic Ricatti Equation)

Conditional pdf in steady state

In "steady state":

$$x_n \sim \mathcal{N}(\mu_n, P)$$

λ_j are eigenvalues of P



Distance from mean (most likely state):

$$r = \sqrt{(x_n - \mu_n)^T (x_n - \mu_n)}$$

$$E(r) \approx \frac{4\left(\sum_{j=1}^{m} \lambda_j\right)^2 - 2\sum_{j=1}^{m} \lambda_j^2}{4\left(\sum_{j=1}^{m} \lambda_j\right)^{1.5}} \qquad var(r) \approx \frac{\sum_{j=1}^{m} \lambda_j^2}{2\sum_{j=1}^{m} \lambda_j}$$

Conditional pdf in steady state

$$\lambda = O(1) \rightarrow E(r) = O(m^{1/2}), \quad var(r) = O(1)$$

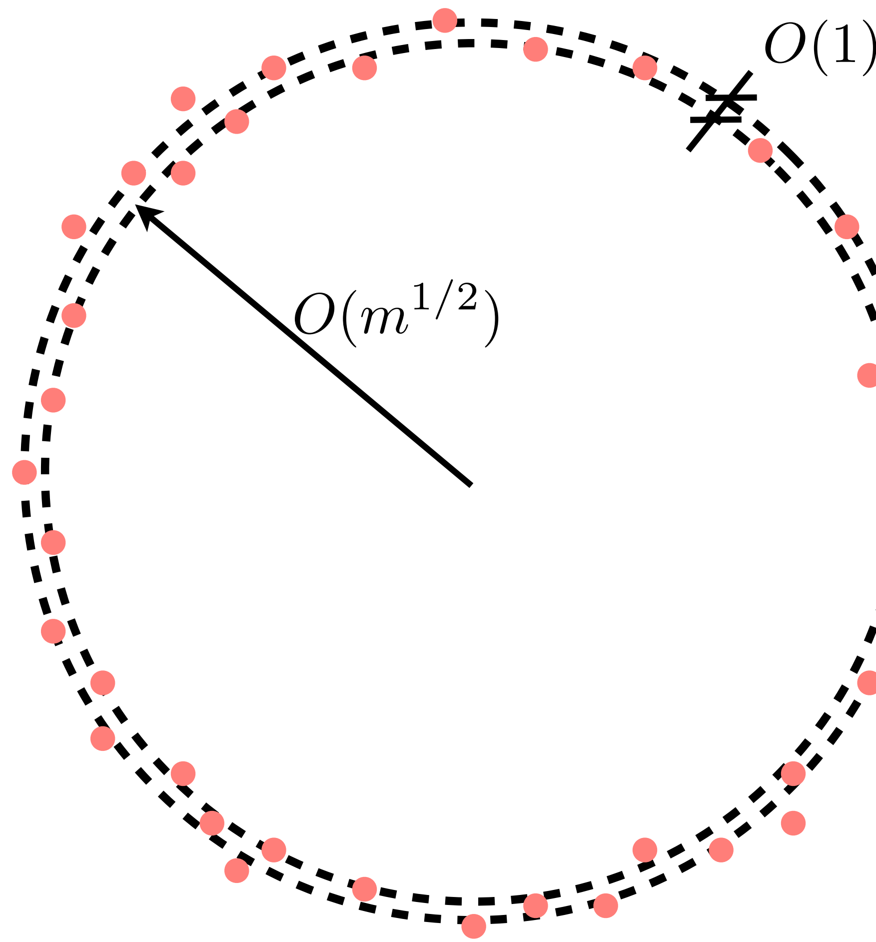
Samples of posterior pdf collect on a thin shell

This is problematic

- There is not enough information in model and data to make reliable conclusions about the state

This is unphysical

- Similar experiments are expected to have similar outcomes



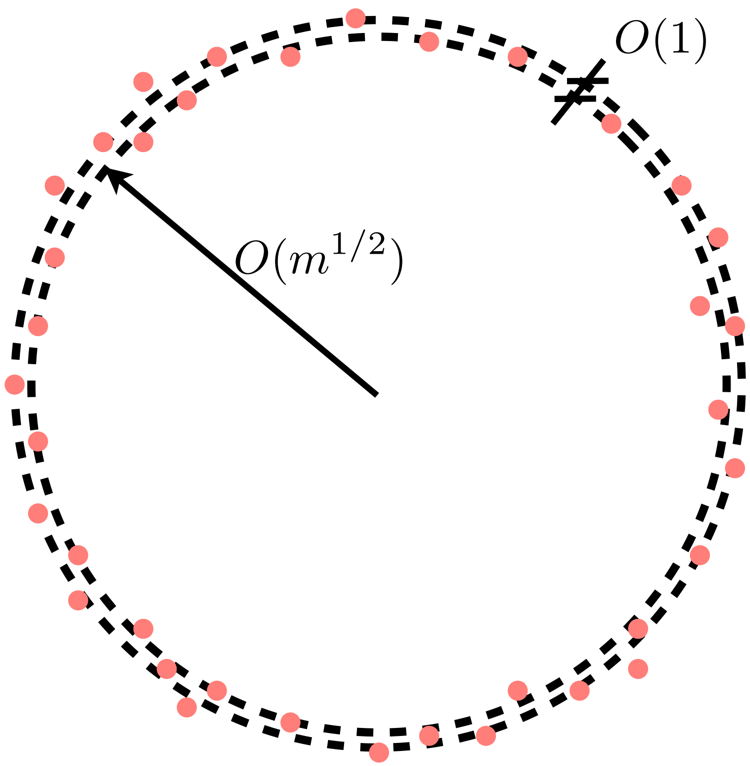
Conditional pdf in steady state

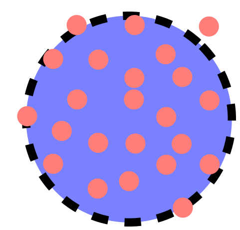
"Large" posterior covariance:

- Sample collect on a thin shell
- There is *not* enough information in model and data to make reliable conclusions about the state

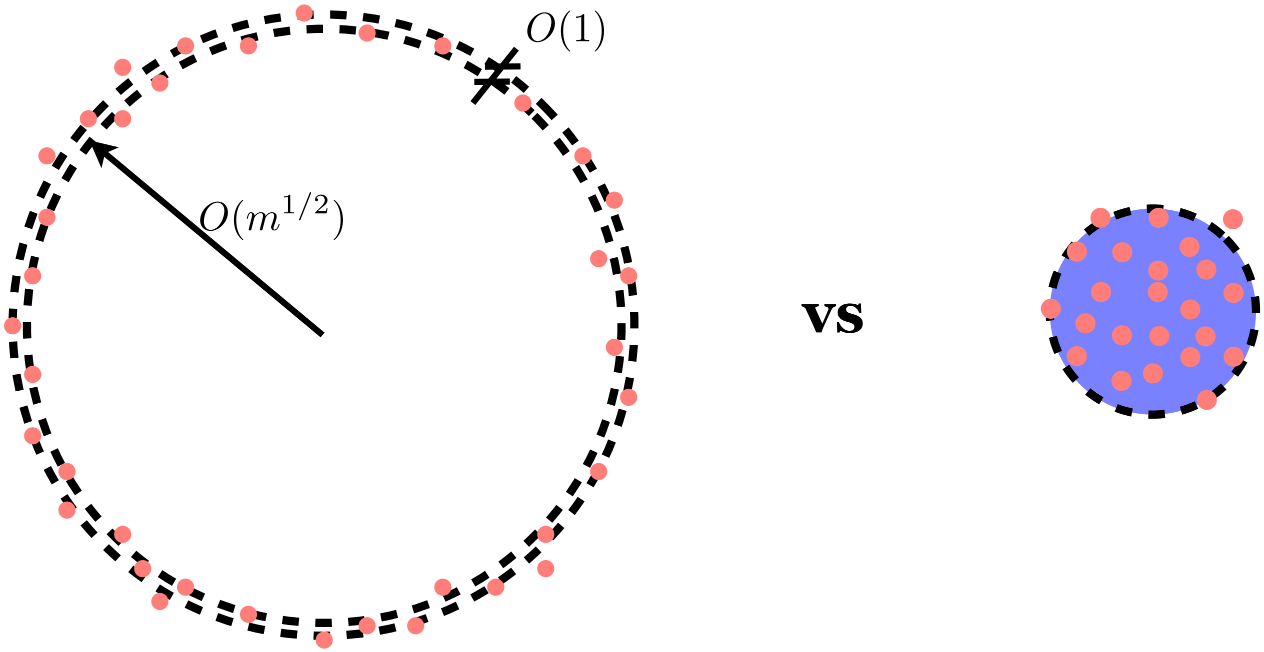
"Small" posterior covariance:

- Samples collect on a low dimensional ball
- There is *enough* information in model and data to make reliable conclusions about the state





The effective dimension



Effective dimension:

- *Definition*: Frobenius norm of steady state covariance matrix *P*
- *Interpretation*: effective dimension must be well bounded or else data assimilation is hopeless (regardless of the algorithm)

$$||P||_F = \sqrt{\sum \lambda^2}$$

Boundedness of the effective dimension: example

Put:

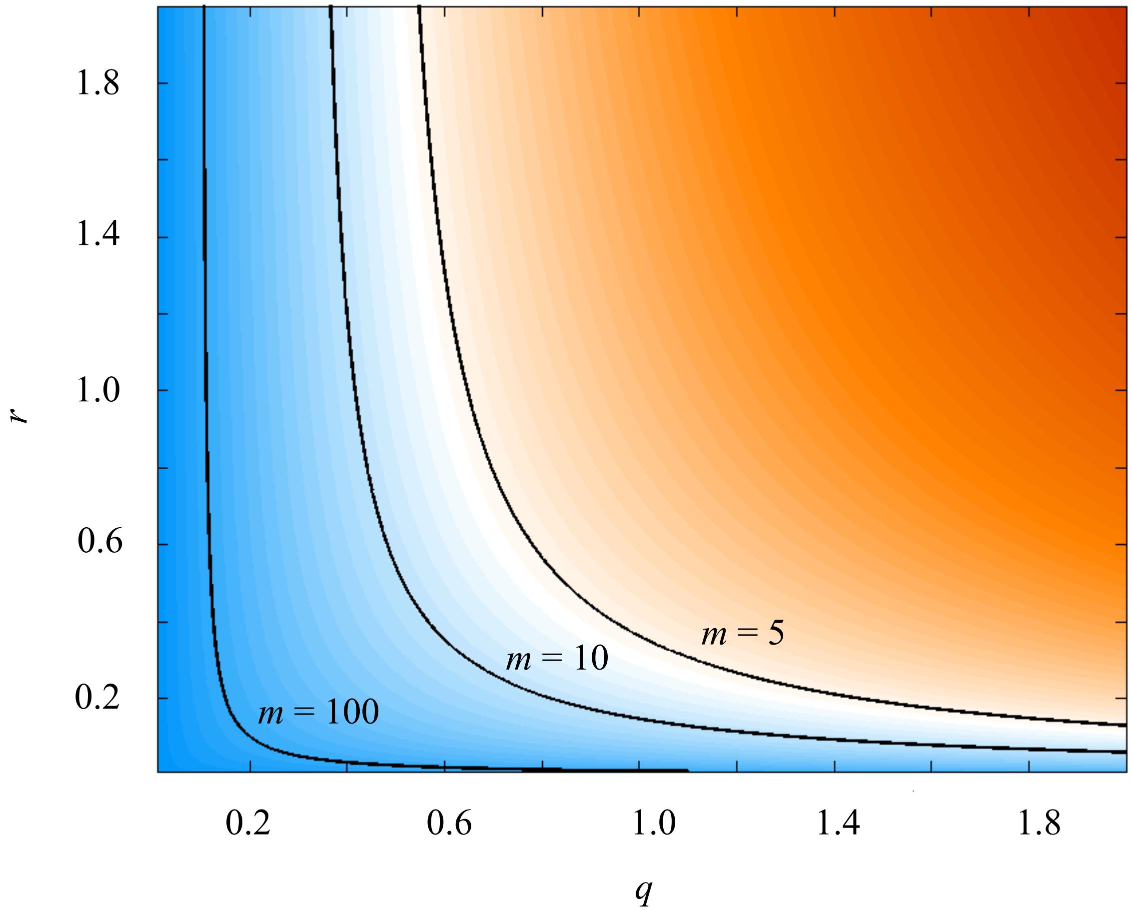
$$A = H = I, \quad Q = qI, \quad R = rI$$

Steady state covariance:

$$P = \frac{\sqrt{q^2 + 4qr} - q}{2} I$$

Effective dimension:

$$||P||_F = \sqrt{m}\frac{\sqrt{q^2 + 4qr} - q}{2}$$

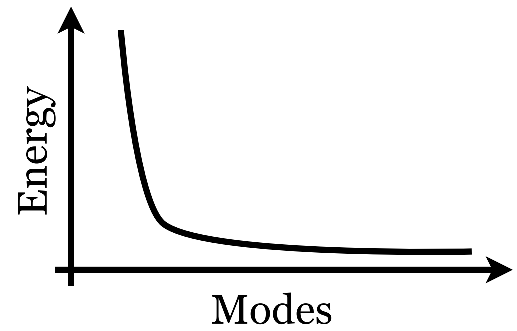


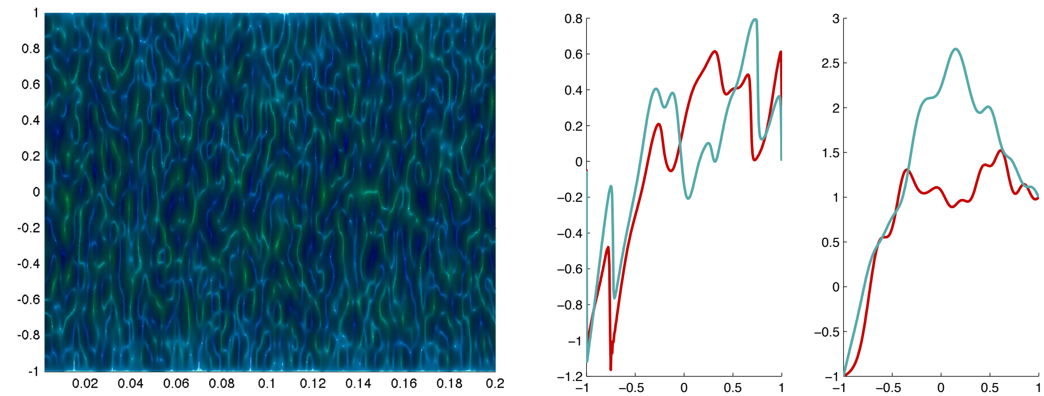
In general: small Frobenius norm of Q and R lead to small effective dimension

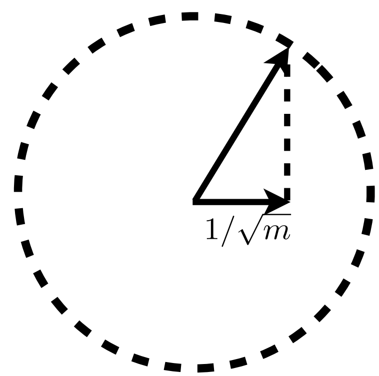
Why is this realistic?

Small effective dimension

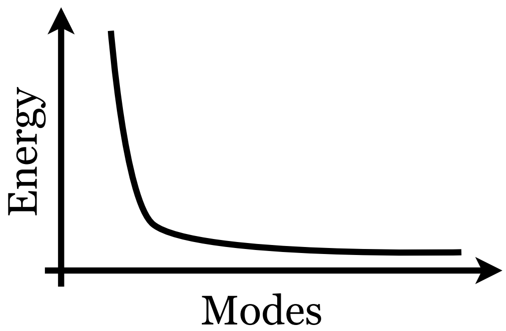
- Khinshin's theorem: bounded energy implies small Frobenius norms of **Q** and **R**
- Smoothness of errors implies small Frobenius norm of **Q** and **R**
- Errors with spherical symmetries (error in one component leads to errors in all other components) lead to small Frobenius norms of **Q** and **R**

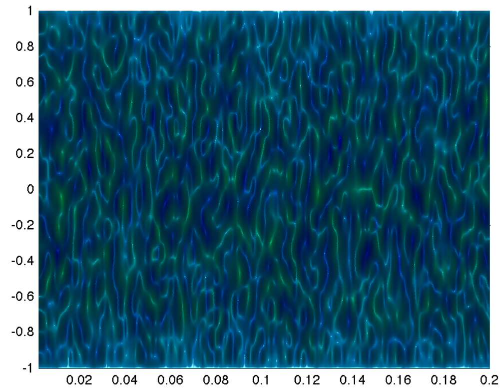


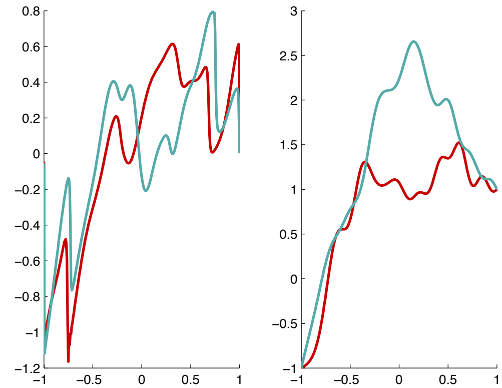


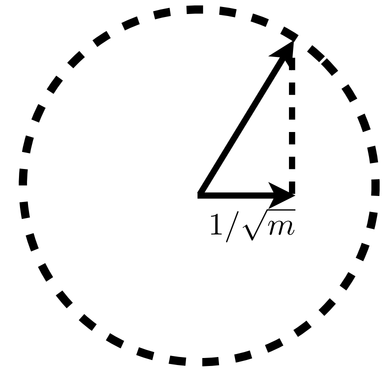


Why is this realistic?







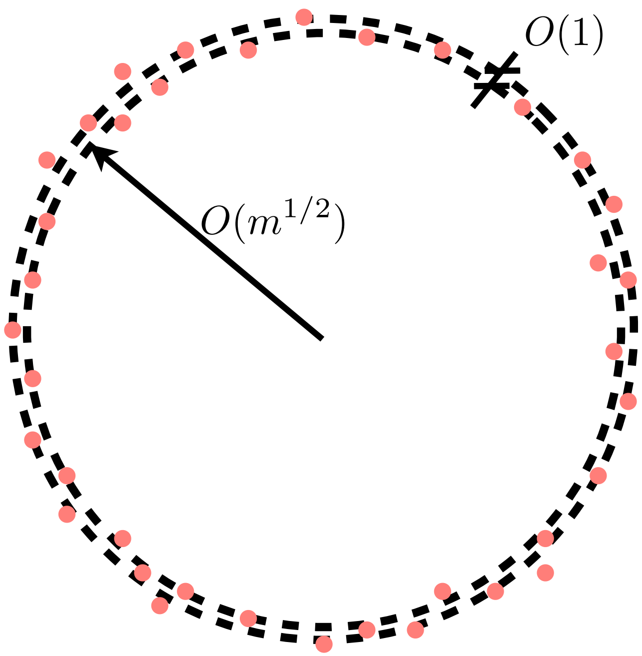


In general: correlations lead to small Frobenius norm of **Q** and **R** and therefore to a small effective dimension

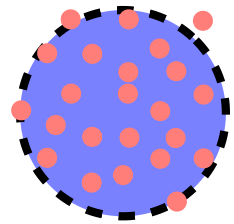
Probability mass is concentrated on a lower dimensional manifold due to correlations in the errors

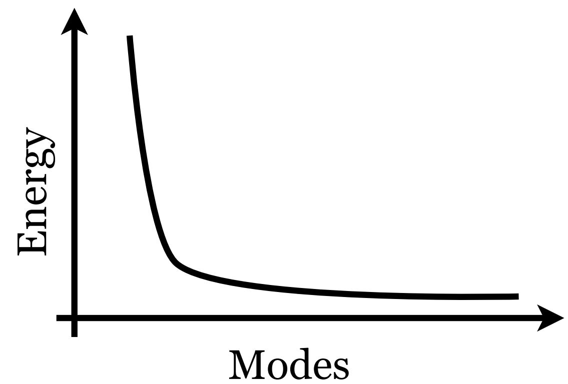
Error models in literature typically show strong correlations

Effective dimension: summary

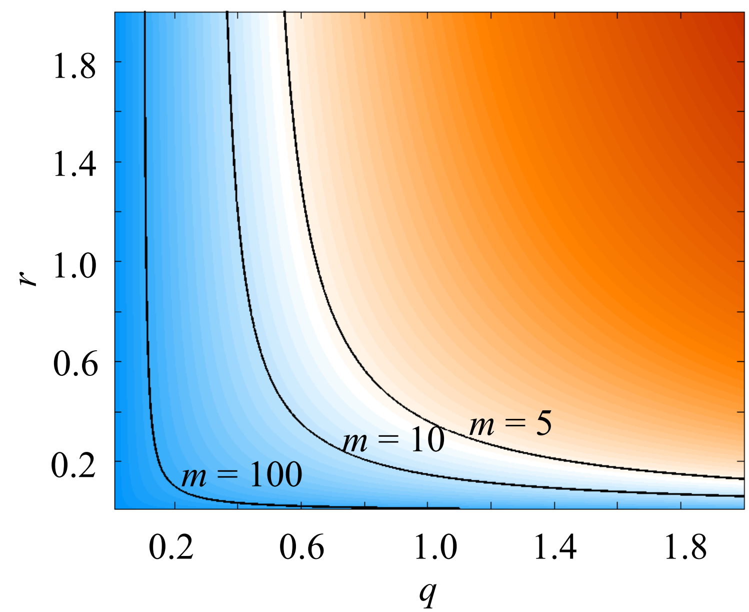


vs





- Effective dimension must be bounded or else data assimilation is hopeless (independent of the DA algorithm)
- Bounded effective dimension induces balance condition between errors in model and data
- In practice, the effective dimension is often small (correlations in errors)



Agenda

1. Introduction
2. What can be expected in general?
3. **How good are particle filters?**
4. How good is 4D-Var or particle smoothing?

Review of importance sampling

Direct Monte Carlo sampling

Suppose we are interested in $x \sim p(x)$ and want to compute the expected value of x. The Monte Carlo approximation is:

Problem: it is not easy to obtain samples directly

$$E(x) = \int xp(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} x_i \quad , \; x_i \sim p(x)$$

Importance sampling

Suppose we can evaluate $p(x)$ and want to compute the expected value of x.

Replace direct samples with weighted samples

$$E(x) = \int xp(x)dx = \int x\frac{p(x)}{\pi(x)}\pi(x)dx \approx \sum_{i=1}^{N} x_i w_i$$

$$w_i \propto \frac{p(x_i)}{\pi(x_i)}, \quad \sum w_i = 1 \quad x_i \sim \pi(x)$$

Particle filters

Apply importance sampling recursively to the conditional pdf

$$p(x^{0:n+1}|z^{1:n+1}) = p(x^{0:n}|z^{1:n})\frac{p(x^{n+1}|x^n)p(z^{n+1}|x^{n+1})}{p(z^{n+1}|z^{1:n})}$$

This requires importance function that factorizes

$$\pi(x^{0:n+1}|z^{0:n+1}) = \pi_0(x^0)\prod_{k=1}^{n+1}\pi_k(x^k|x^{0:k-1}, z^{1:k})$$

Recursion for weights

$$W_j^{n+1} \propto \hat{W}_j^n \frac{p(X_j^{n+1}|X_j^n)p(Z^{n+1}|X_j^{n+1})}{\pi_{n+1}(X_j^{n+1}|X_j^{0:n}, Z^{0:k})}$$

If variance of unnormalized weights is large, then the particle filter collapses (see papers by Snyder, Bickel, Anderson, ...)

The SIR particle filter

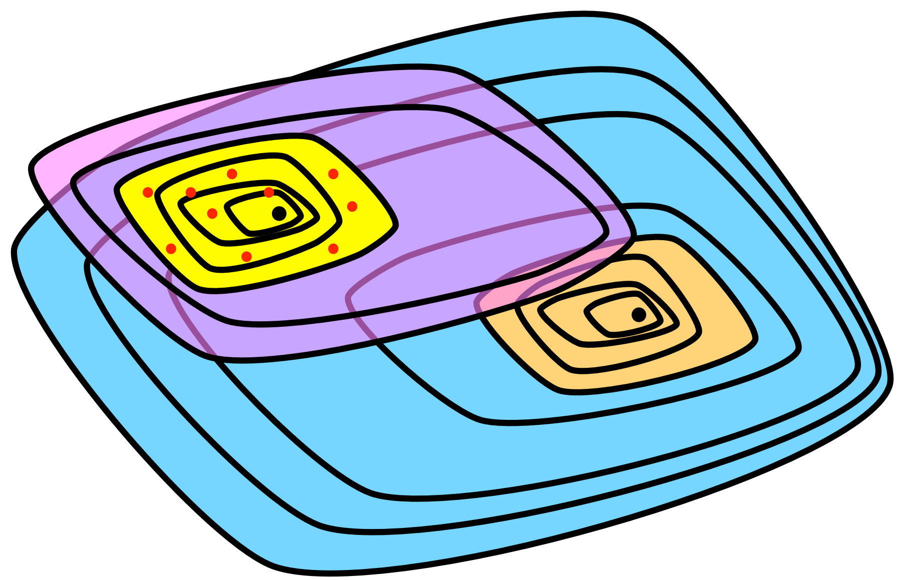
Importance function: $$\pi_{n+1} = p(x^{n+1}|x^n)$$

Weights: $$W_j^{n+1} \propto p(Z^{n+1}|X_j^{n+1}).$$

Condition for collapse in high dimensions is large norm of*:

$$\Sigma = H(Q + APA^T)H^T R^{-1}$$

Collapse can also happen in low dimensions:



*as shown by Snyder, Bickel, Anderson et al.

The optimal particle filter

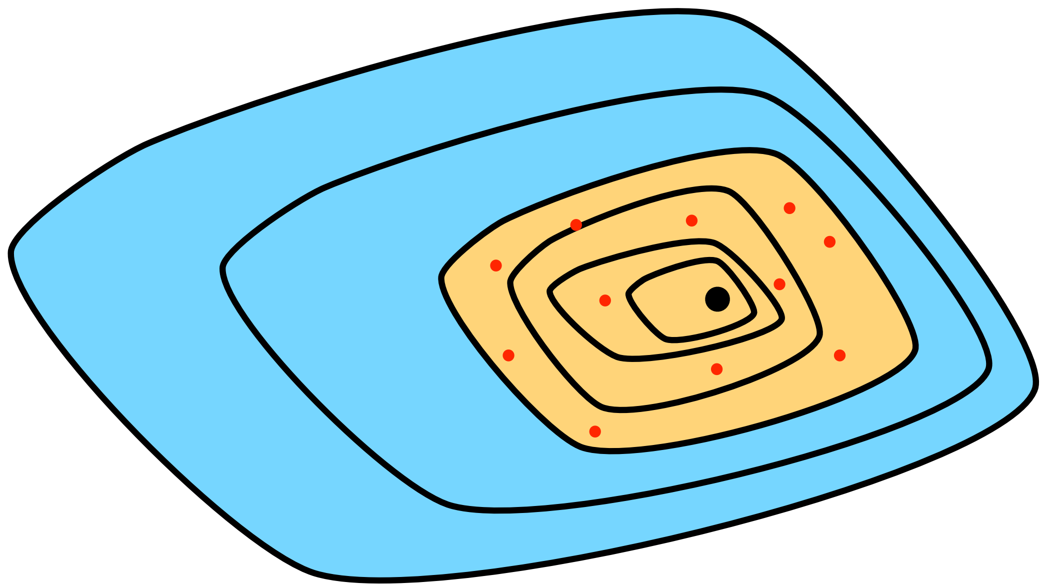
Importance function: $\pi_{n+1} = p(x^{n+1}|x^n, z^{n+1})$

Weights: $W_j^{n+1} \propto p(Z^{n+1}|X_j^n)$

Condition for collapse in high dimensions is large norm of*:

$$\Sigma = HAPA^T H^T \left(HQH^T + R\right)^{-1}$$

Collapse is avoided in low dimensions:



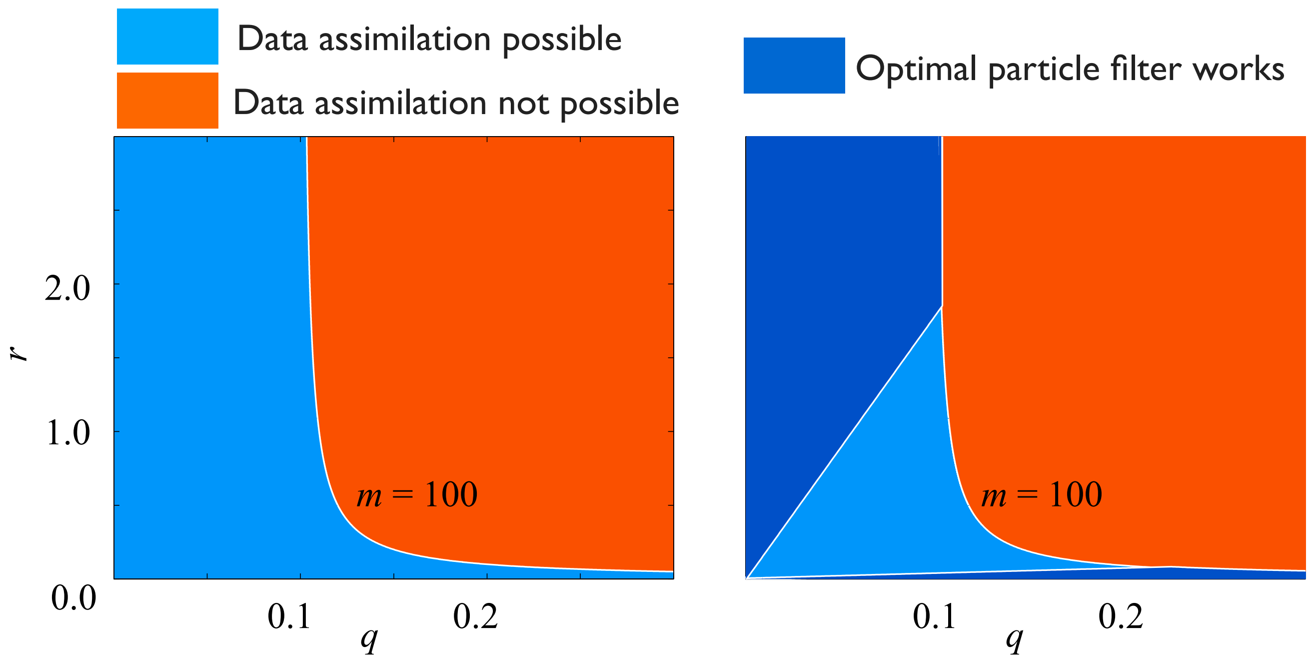
*as shown by Snyder

How good can we get with particle filters?

Example revisited:

Optimal filter: $A = H = I, \quad Q = qI, \quad R = rI$

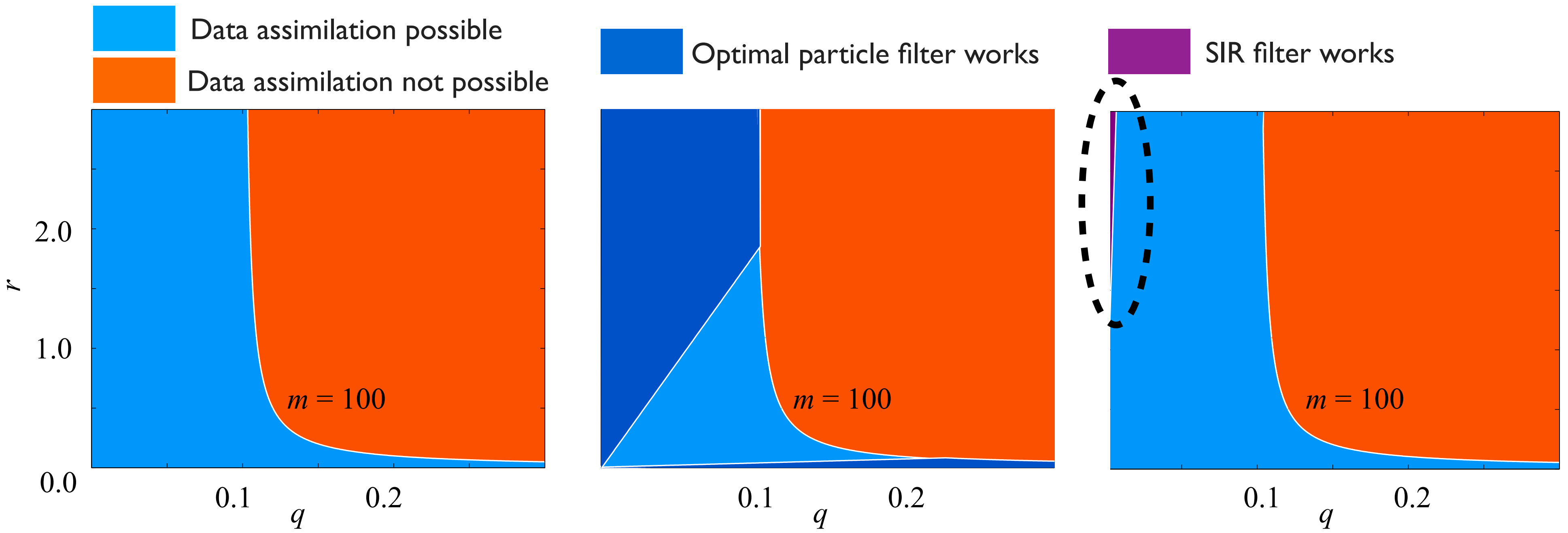
$$||\Sigma||_F = \sqrt{m}\frac{\sqrt{q^2 + 4qr} - q}{2(q + r)}$$



How good can we get with particle filters?

Example revisited:

Optimal filter vs. SIR filter



How good can we get with particle filters?

The general linear case:

- Using matrix bounds we find the balance conditions:

$$||A||_F^2||H||_F^2||P||_F \leq ||H||_F^2||Q||_F + ||R||_F$$ Optimal filter

$$\frac{1}{2}||H||_F^2\left(||Q||_F + ||A||_F^2||P||\right) \leq ||R||_F$$ SIR filter

- Balance condition is easy in simple cases, but delicate in general

The nonlinear/non-Gaussian case:

- Correlations can be expected for realistic noise models
- Balance conditions must be worked out in each particular case
- Optimal filter hard/impossible to implement, while SIR remains easy to use

Agenda

1. Introduction
2. What can be expected in general?
3. How good are particle filters?
4. **How good is 4D-Var or particle smoothing?**

Strong constraint 4D-Var

Perfect model assumption:

$$x^{n+1} = Ax^n$$

$$z^{n+1} = Hx^{n+1} + v^{n+1}, \quad v^n \sim \mathcal{N}(0, R)$$

Conditional pdf:

$$p(x^0|z^{1:n}) \propto \exp\left(-\frac{1}{2}\left(x^0 - \mu_0\right)^T \Sigma_0^{-1}\left(x^0 - \mu_0\right)\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left(z^j - HA^jx^0\right)^T R^{-1}\left(z^j - HA^jx^0\right)\right)$$

Covariance:

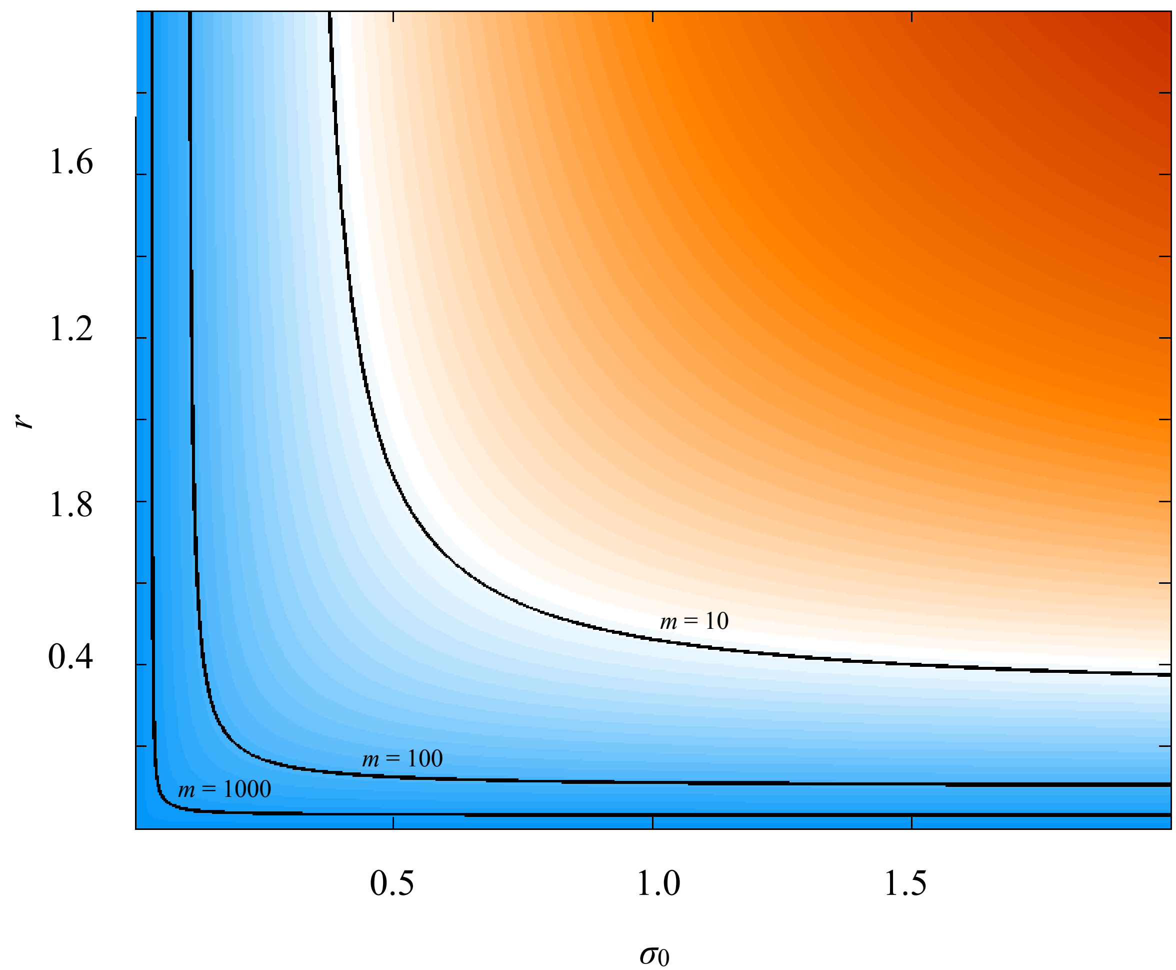
$$\Sigma = \Sigma_0^{-1} + \sum_{j=1}^{n}(A^j)^T H^T R^{-1} H A^j$$

Strong constraint 4D-Var can only be successful if the Frobenius norm of the covariance is small

How good can we get with 4D-Var?

Example revisited:

Norm of covariance matrix: $$||\Sigma||_F = \sqrt{m}\frac{\sigma_0 + r}{\sigma_0 r}$$



Weak constraint 4D-VAr

Model and data:

$$x^{n+1} = Ax^n + w^n, \quad w^n \sim \mathcal{N}(0, Q), \text{ iid}$$

$$z^{n+1} = Hx^{n+1} + v^{n+1}, \quad v^n \sim \mathcal{N}(0, R), \text{ iid, independent of } w^n$$

Conditional pdf:

$$p(x^{0:n} | z^{1:n}) \propto \exp\left(-\frac{1}{2}\left(x^0 - \mu_0\right)^T \Sigma_0^{-1} \left(x^0 - \mu_0\right)\right) \times \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left(z^j - Hx^j\right)^T R^{-1}\left(z^j - Hx^j\right)\right)$$

Covariance:

$$\Sigma = \begin{pmatrix} \Sigma_0^{-1} + A^T \Sigma_1^{-1} A & -A^T Q^{-1} & \cdots & 0 \\ -Q^{-1} A & Q^{-1} + A^T Q^{-1} A + H^T R^{-1} H & -A^T Q^{-1} & \\ 0 & \ddots & \ddots & \ddots \\ \vdots & & & -A^T Q^{-1} \\ 0 & \cdots & -Q^{-1} A & Q^{-1} + H^T R^{-1} H \end{pmatrix}.$$

How good can we get with 4D-Var?

Strong constraint 4D-Var

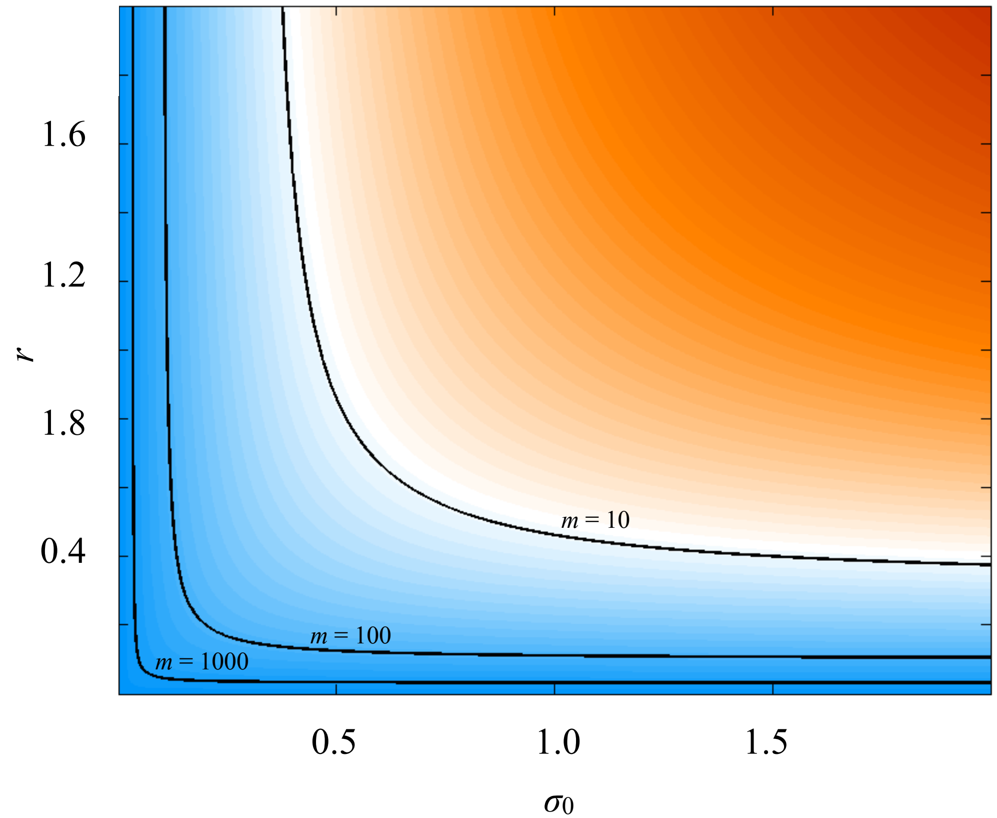
- Boundedness of covariance matrix induces balance condition between errors in prior and in the data

Weak constraint 4D-Var

- Boundedness of covariance matrix induces balance condition between errors in prior, in the model and in the data

Particle smoothing

- Same balance condition as in 4D-Var, but choice of importance function is critical



Conclusions

- Numerical data assimilation hopeless unless effective dimension is small (probability mass is concentrated on a low dimensional manifold)
- Boundedness of effective dimension induces balance condition between errors in model and data
- In practice, effective dimension often small because correlations in errors
- Particle filters can work in high dimensions, provided their implementation is sound
- Variational data assimilation requires well boundedness of covariance matrix, i.e. balance condition between errors in prior, model and data

Analysis for linear Gaussian case only. Nonlinear/non-Gaussian problems must be analyzed in each particular case.

Thank you!

How good can we get with particle filters?

Example revisited:

Optimal filter vs. SIR filter