What the collapse of the ensemble Kalman filter tells us about particle filters

Matthias Morzfeld

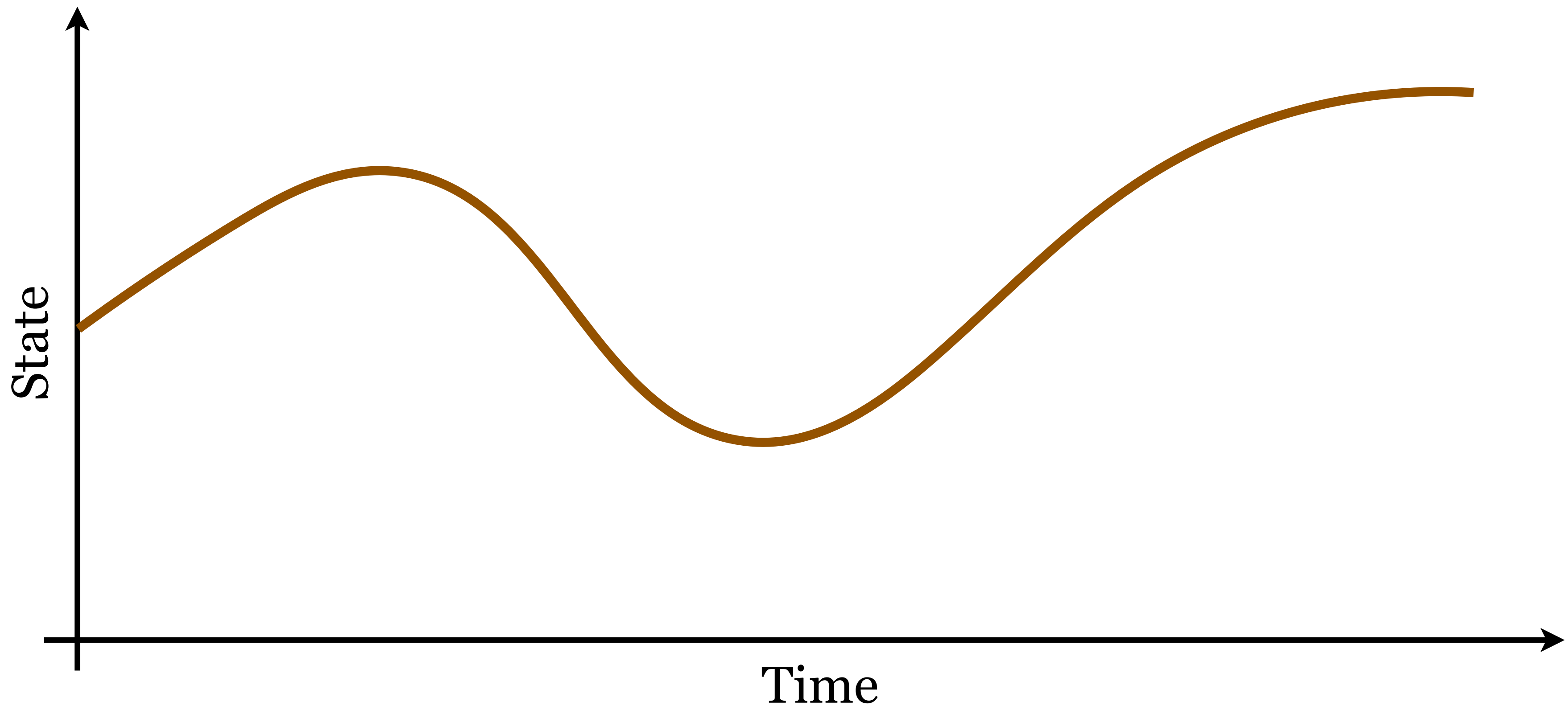
Department of Mathematics
Program in Applied Mathematics
Graduate Interdisciplinary Program in Statistics
University of Arizona

Collaborators: Daniel Hodyss (NRL), Chris Snyder (NCAR)

Supported by:

SIAM Annual Meeting
Tuesday, July 11th, 2017

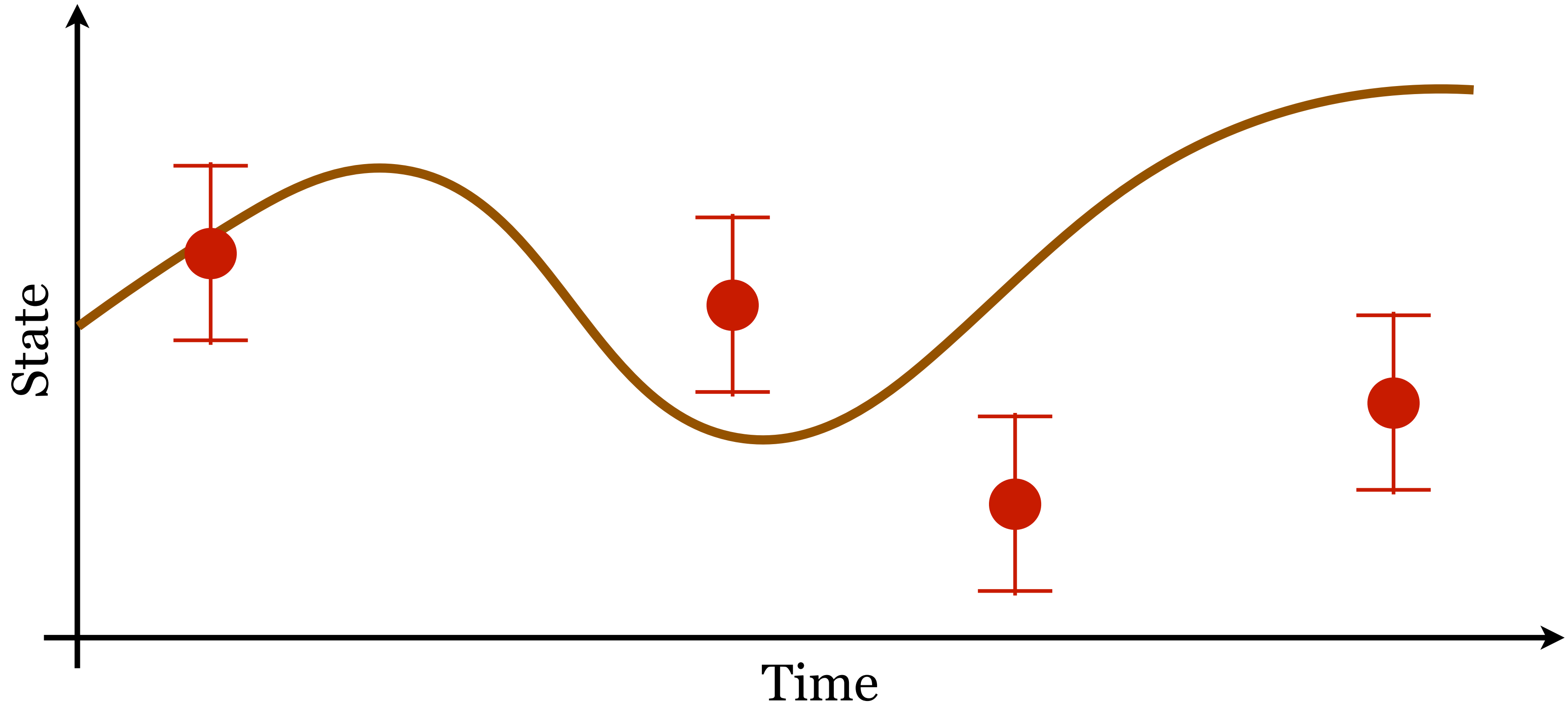
Data assimilation



Mathematical model

$$\frac{dx}{dt} = f(x), \quad x(0) = x_0$$

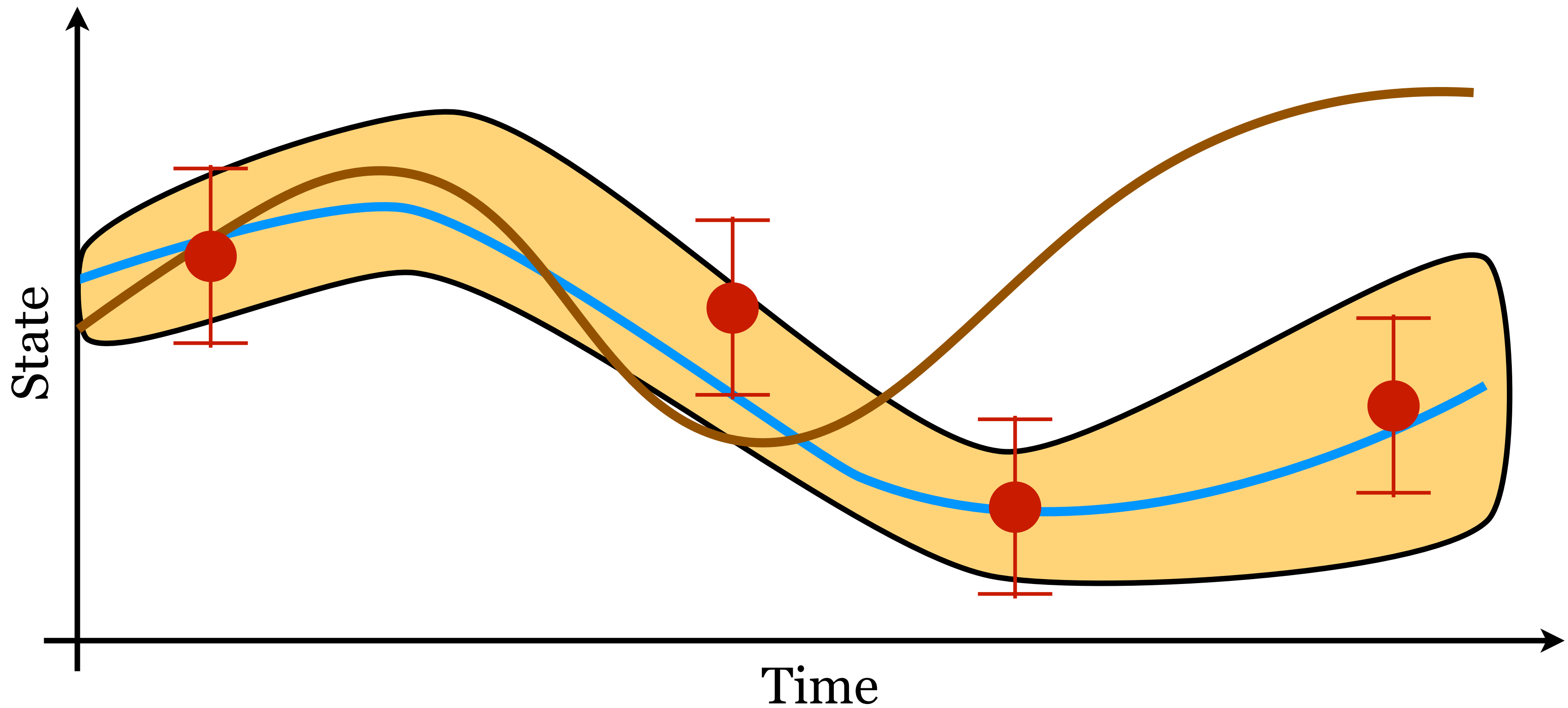
Data assimilation



Mathematical model $\quad \dfrac{dx}{dt} = f(x), \quad x(0) = x_0$

Data $\quad$ sparse and noisy

Data assimilation

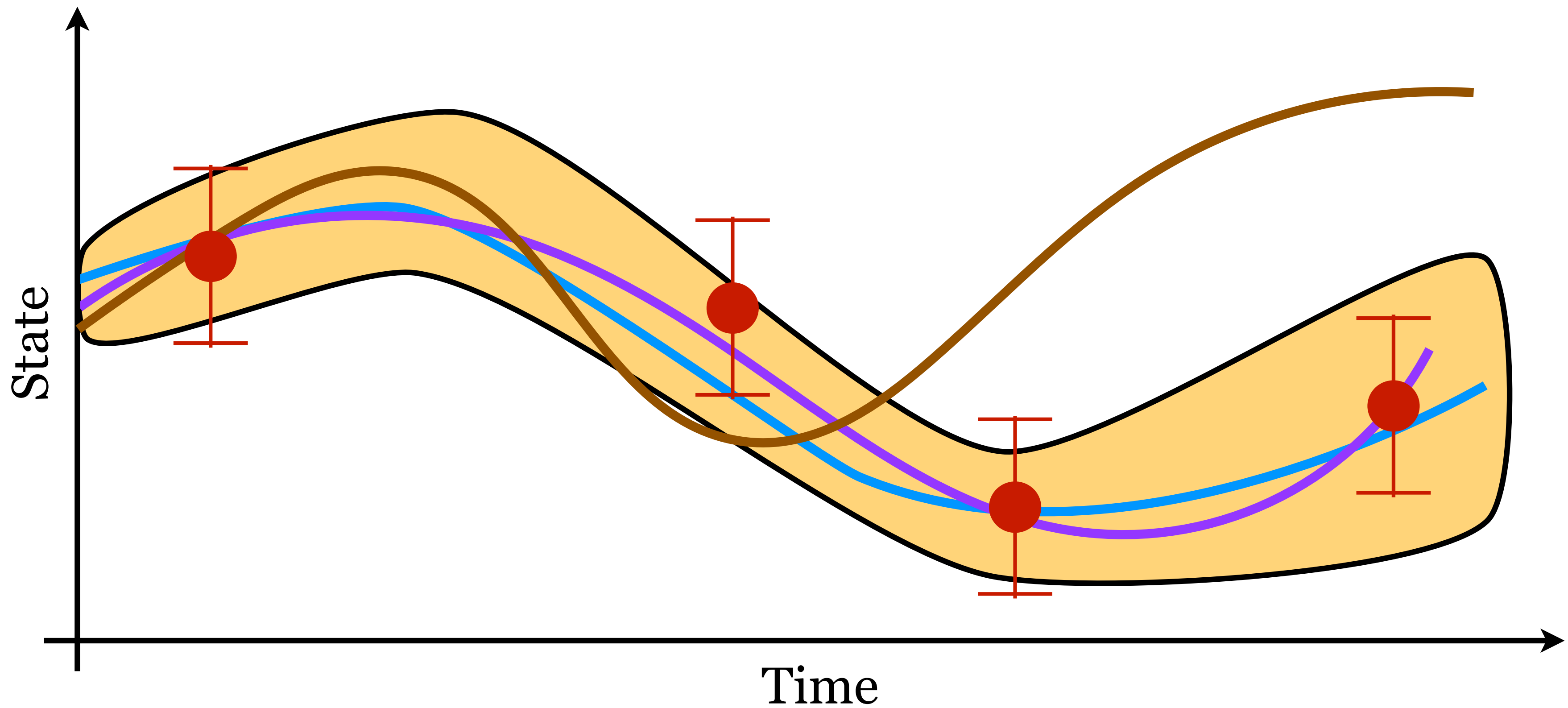


Mathematical model $$\frac{dx}{dt} = f(x), \quad x(0) = x_0$$

Data sparse and noisy

Data assimilation

Data assimilation



Mathematical model $\frac{dx}{dt} = f(x), \quad x(0) = x_0$

Data sparse and noisy

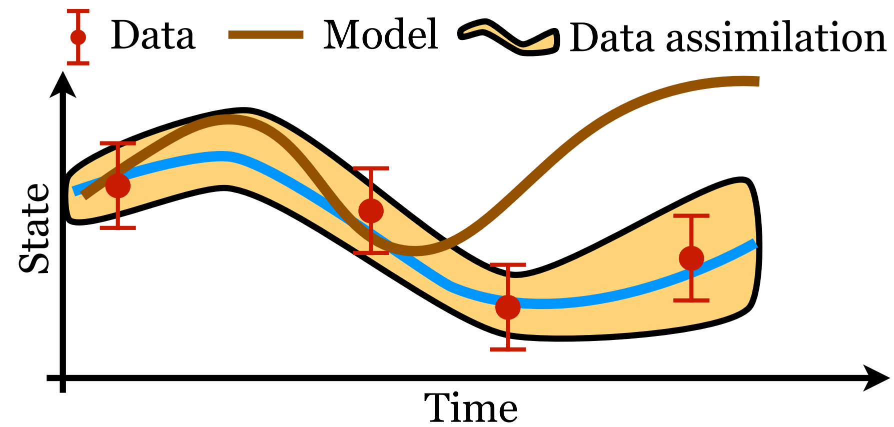
Data assimilation

Physical state

The puzzle

Operational NWP

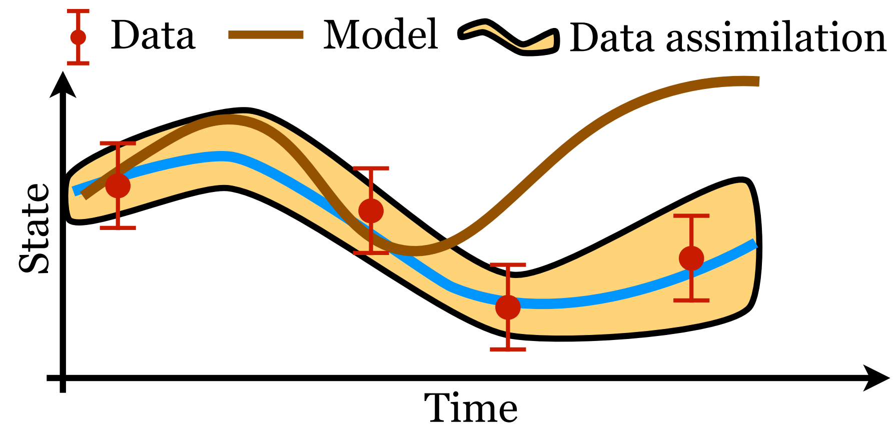
- Data assimilation done every 6 hrs
- EnKF with ensemble size 50—100
- Reported to “work well”
- Typical number of vars.: 650 million
- Typical number of obs.: 2–10 million



The puzzle

Operational NWP

- Data assimilation done every 6 hrs
- EnKF with ensemble size 50—100
- Reported to “work well”
- Typical number of vars.: 650 million
- Typical number of obs.: 2–10 million



Particle filters

- Computational requirements scale exponentially with dimension*,**
- Particle filters are *not/not often* used in NWP

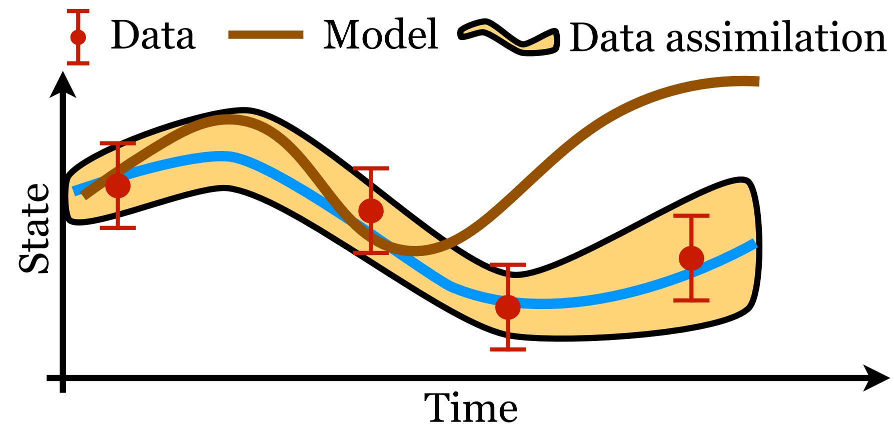
* *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015

***Bickel et al*., 2008, *Bengtsson et al*. 2008, *Snyder et al*. 2008, *Snyder* 2011

The puzzle

Operational NWP

- Data assimilation done every 6 hrs
- EnKF with ensemble size 50—100
- Reported to “work well”
- Typical number of vars.: 650 million
- Typical number of obs.: 2–10 million



Particle filters

- Computational requirements scale exponentially with dimension*,**
- Particle filters are *not/not often* used in NWP

Puzzle

- EnKF *can be interpreted as* a particle filter
- *It should not work in theory, so why does it work in practice?*

* *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015

** *Bickel et al.*, 2008, *Bengtsson et al.* 2008, *Snyder et al.* 2008, *Snyder* 2011

Agenda

1. Problem formulation
2. **Background**
   - Ensemble Kalman filter
   - Particle filters
   - Limitations of particle filters
3. Why can EnKF “work” when ensemble size is small

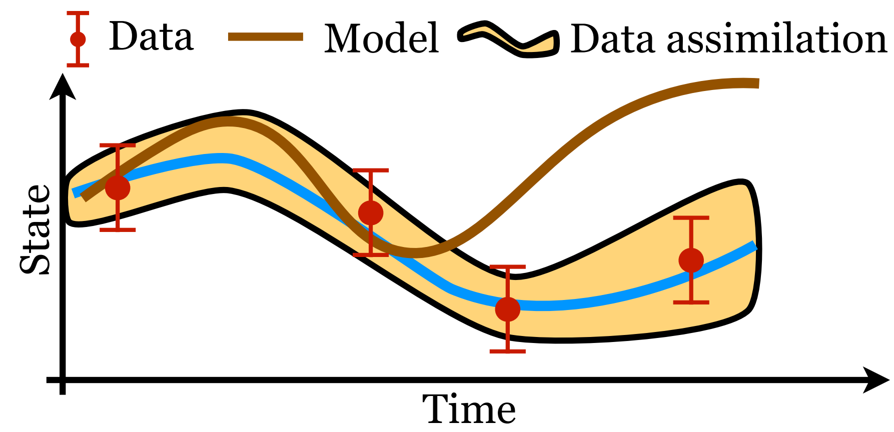
Ensemble Kalman filter

Model

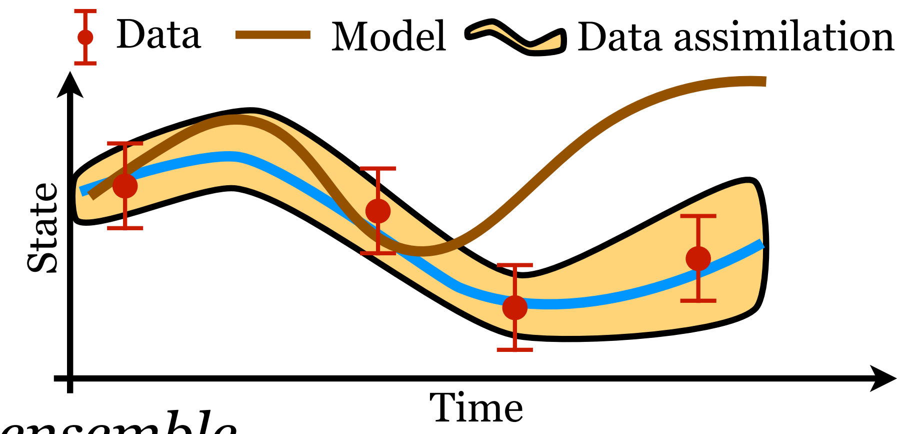
$$x^k = f(x^{k-1}) + \varepsilon^k, \quad \varepsilon^k \sim \mathcal{N}(0, Q)$$

Observations

$$z^k = Hx^k + \eta^k, \quad \eta^k = \mathcal{N}(0, R)$$



Ensemble Kalman filter



Model

$$x^k = f(x^{k-1}) + \varepsilon^k, \quad \varepsilon^k \sim \mathcal{N}(0, Q)$$

Observations

$$z^k = Hx^k + \eta^k, \quad \eta^k = \mathcal{N}(0, R)$$

Idea: Monte Carlo

- Represent posterior distribution by an *ensemble*

$$x_i^k, \quad i = 1, 2, \ldots, N_e$$

- *Ensemble average ≈ posterior mean*
- *Ensemble covariance ≈ posterior covariance*

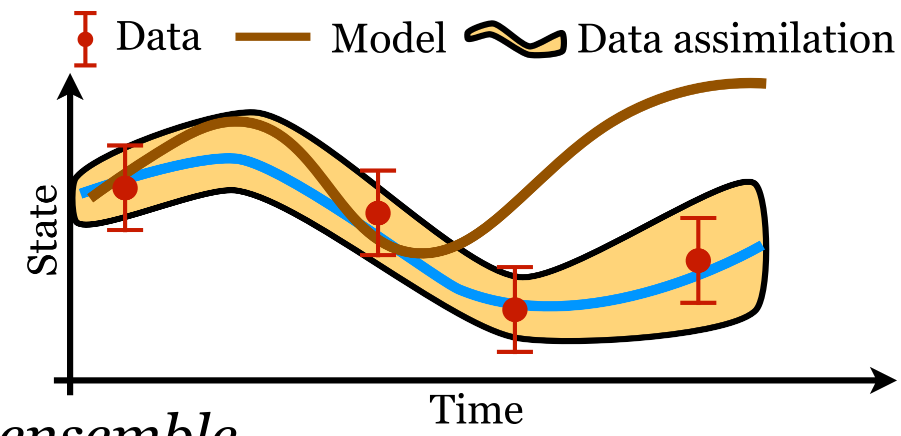
Ensemble Kalman filter

Model

$$x^k = f(x^{k-1}) + \varepsilon^k, \quad \varepsilon^k \sim \mathcal{N}(0, Q)$$

Observations

$$z^k = Hx^k + \eta^k, \quad \eta^k = \mathcal{N}(0, R)$$



Idea: Monte Carlo

- Represent posterior distribution by an *ensemble*

$$x_i^k, \quad i = 1, 2, \ldots, N_e$$

- *Ensemble average ≈ posterior mean*
- *Ensemble covariance ≈ posterior covariance*

EnKF

Forecast step:

$$x_i^f = f(x_i^{k-1}) + w_i^k$$
$$P^f = cov(x_i^f)$$

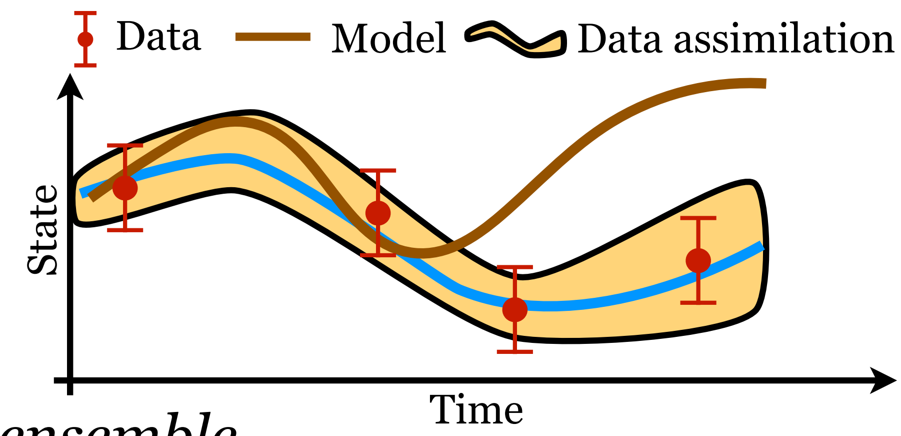
Kalman gain:

$$K = P^f H^T (HP^f H^T + R)^{-1}$$

Analysis ensemble:

$$x_i^k = x_i^f + K(z^k - Hx_i^f + v_i)$$
$$P_a = cov(x_i^k)$$

Ensemble Kalman filter



Model

$$x^k = f(x^{k-1}) + \varepsilon^k, \quad \varepsilon^k \sim \mathcal{N}(0, Q)$$

Observations

$$z^k = Hx^k + \eta^k, \quad \eta^k = \mathcal{N}(0, R)$$

Idea: Monte Carlo

- Represent posterior distribution by an *ensemble*

$$x_i^k, \quad i = 1, 2, \ldots, N_e$$

- *Ensemble average ≈ posterior mean*
- *Ensemble covariance ≈ posterior covariance*

EnKF

Forecast step:

$$x_i^f = f(x_i^{k-1}) + w_i^k$$

$$P^f = cov(x_i^f)$$

Kalman gain:

$$K = P^f H^T (HP^f H^T + R)^{-1}$$

Analysis ensemble:

$$x_i^k = x_i^f + K(z^k - Hx_i^f + v_i)$$

$$P_a = cov(x_i^k)$$

Localization & inflation (tuning)

- *Delete spurious correlations*
- *Inflate to counteract sampling error*

Agenda

1. Problem formulation

2. **Background**

   Kalman filter and ensemble Kalman filter

   Particle filters

   Limitations of particle filters

3. Why can EnKF “work” when ensemble size is small

Review of importance sampling

Compute the expected value of $x \sim p(x)$ by Monte Carlo:

$$E[x] = \int xp(x)\mathrm{d}x \approx \frac{1}{N_e}\sum_{i=1}^{N_e} x_i, \quad x_i \sim p(x)$$

Compute the expected value of $x \sim p(x)$ by Monte Carlo:

$$E[x] = \int x p(x) \mathrm{d}x \approx \frac{1}{N_e} \sum_{i=1}^{N_e} x_i, \quad x_i \sim p(x)$$

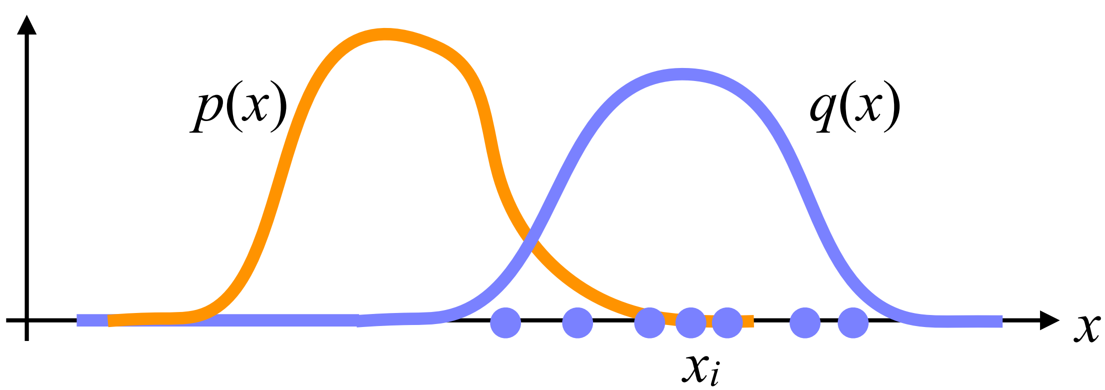
Difficult unless p(x) is elementary

Importance sampling

- Convert averaging into weighted averaging by replacing target density $p(x)$ with a simpler version (proposal distribution $q(x)$)

$$E[x] = \int x p(x) \mathrm{d}x \int x \frac{p(x)}{q(x)} q(x) \mathrm{d}x \approx \frac{1}{N_e} \sum_{i=1}^{N_e} x_i w_i, \quad x_i \sim q(x)$$

$$w_i = \frac{p(x^i)}{q(x^i)}$$

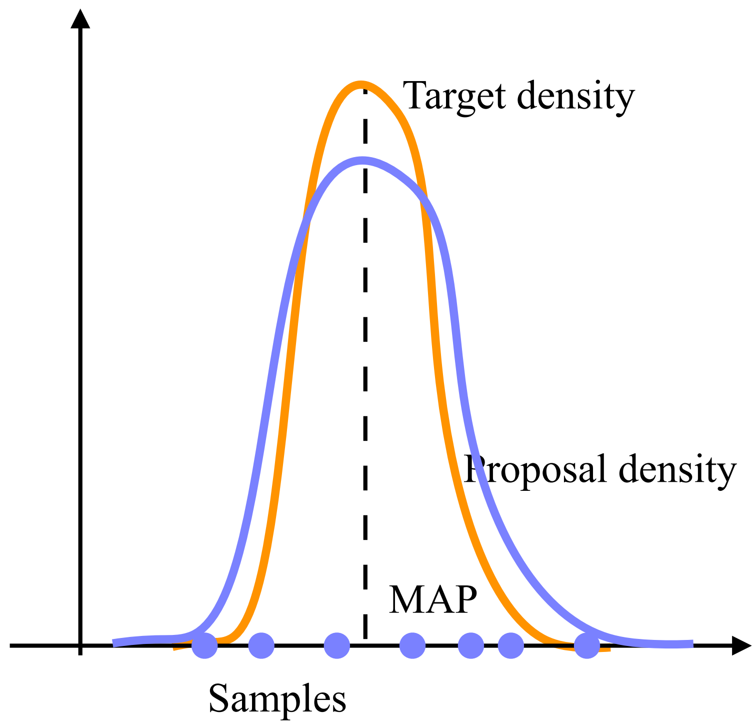


Efficiency of importance sampling

Effective sample size

- Weights describe differences between the target distribution and the proposal distribution
- Effective number of samples

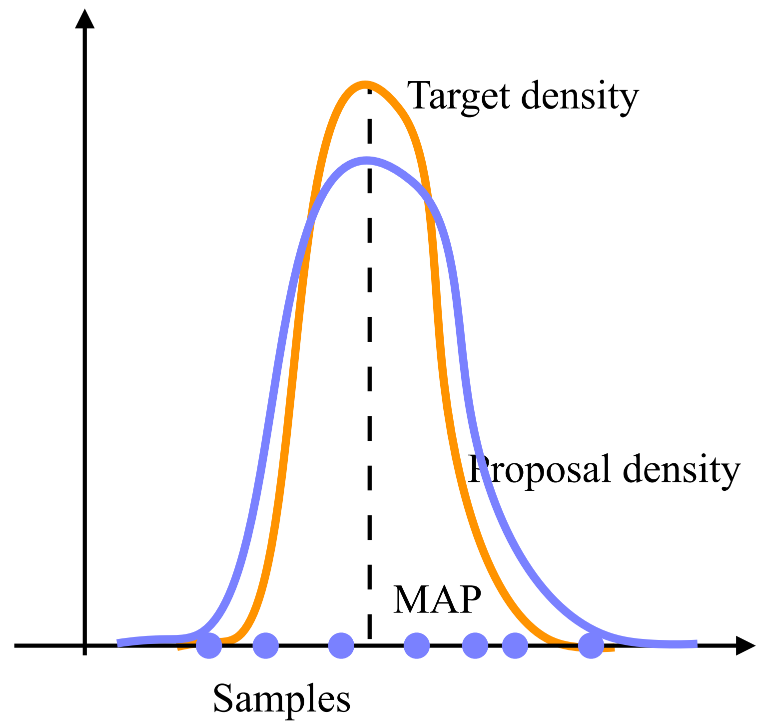
$$N_{\mathrm{eff}} = \frac{N_e}{G}, \quad G = \frac{E[w^2]}{E[w]^2}$$



Effective sample size

- Weights describe differences between the target distribution and the proposal distribution
- Effective number of samples

$$\boxed{N_{\text{eff}} = \frac{N_e}{G}, \quad G = \frac{E[w^2]}{E[w]^2}}$$



Efficient importance sampling

- An efficient sampling algorithm must have small G
- Variance of the weights must be small

$$\frac{var[w]}{E[w]^2} = \frac{E[w^2] - E[w]^2}{E[w]^2} = G - 1$$

Optimal particle filter

Posterior: $$p(x^{0:k}|z^{1:k}) \propto p(x^{0:k-1}|z^{1:k-1})\; p(x^k|x^{k-1})p(z^k|x^k)$$

Proposal distribution: $$q(x^{0:k}; z^{1:k}) \propto q_0(x^0) \prod_{j=1}^{k} q_j(x^j; x^{0:j-1}, z^{1:j})$$

Weights: $$w^k = w^{k-1} \frac{p(x^k|x^{k-1})p(z^k|x^k)}{q_k(x^k; x^{0:k-1}, z^{1:k})}$$

Optimal particle filter

Posterior: $$p(x^{0:k}|z^{1:k}) \propto p(x^{0:k-1}|z^{1:k-1})\; p(x^k|x^{k-1})p(z^k|x^k)$$

Proposal distribution: $$q(x^{0:k};z^{1:k}) \propto q_0(x^0)\prod_{j=1}^{k} q_j(x^j;x^{0:j-1},z^{1:j})$$

Weights: $$w^k = w^{k-1}\frac{p(x^k|x^{k-1})p(z^k|x^k)}{q_k(x^k;x^{0:k-1},z^{1:k})}$$

Optimal importance function minimizes variance of weights*

$$q_j^{\mathrm{opt}}(x^j;x^{0:j-1},z^{1:j}) = p(x^j|x^{j-1},z^j)$$

* *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015

Agenda

1. Problem formulation

2. **Background**

   Kalman filter and ensemble Kalman filter

   Particle filters

   Limitations of particle filters

3. Why can EnKF "work" when ensemble size is small

Standard example[*]

Model

$$x^k = x^{k-1} + \varepsilon^k, \quad \varepsilon \sim \mathcal{N}(0, I)$$

Observations

$$z^k = x^k + \eta^k, \quad \eta \sim \mathcal{N}(0, I)$$

[*] *Bickel et al.*, 2008, *Bengtsson et al.* 2008, *Snyder et al.* 2008, *Snyder* 2011

How good are particle filters

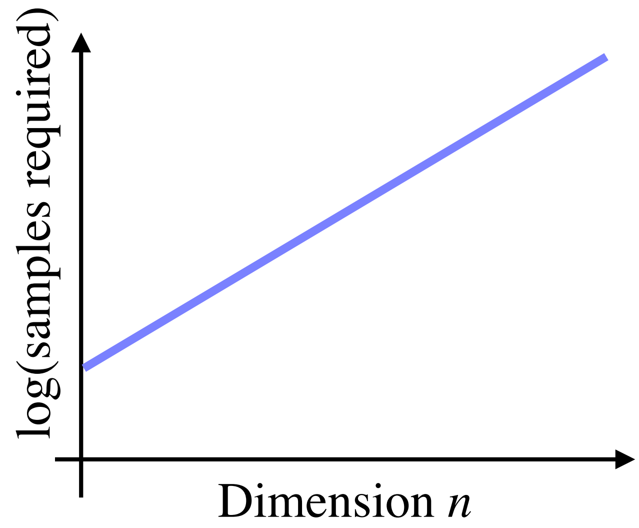
Standard example *

Model

$$x^k = x^{k-1} + \varepsilon^k, \quad \varepsilon \sim \mathcal{N}(0, I)$$

Observations

$$z^k = x^k + \eta^k, \quad \eta \sim \mathcal{N}(0, I)$$



- Variance of weights scales exponentially with the dimension*
- Ensemble size scales exponentially with dimension ("**collapse**")

$$N_{\text{eff}} = \frac{N_e}{G}, \quad G = \exp(n), \quad N_e \propto \exp(n) N_{\text{eff}}$$

- True for all particle filters**

All particle filters collapse in high-dimensional problems

* *Bickel et al.*, 2008, *Bengtsson et al.* 2008, *Snyder et al.* 2008, *Snyder* 2011
** *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015

Agenda

1. Problem formulation

2. Background

   Kalman filter and ensemble Kalman filter
   Particle filters
   Limitations of particle filters

3. **Why can EnKF “work” when ensemble size is small**

Summary

Particle filters

- Particle filters fail unless effective dimension is "small"

EnKF

- Uses Monte Carlo step to approximate forecast covariance
- "Works well"in high-dimensions

Summary

Particle filters	***EnKF***
• Particle filters fail unless effective dimension is "small"	• Uses Monte Carlo step to approximate forecast covariance • "Works well"in high-dimensions

EnKF can be interpreted as a particle filter*

EnKF proposal distribution: $q_{\text{EnKF}}(x^{0:k}; z^{1:k}) \propto q_0(x^0) \prod_{i=1}^{k} q_{i,\text{EnKF}}(x^i; x^{0:i-1}, z^{1:i})$

$$q_{i,\text{EnKF}}(x^i; x_j^{1:i-1}, z^{1:i}) = \mathcal{N}\left(\mu_j^i, \Sigma_k\right),$$

$$\mu_j^i = (I - KH)f(x_j^{i-1}) + Ky^i, \quad \Sigma_k = (I - KH)Q(I - KH)^T + KRK^T$$

* *Papadakis et al.*, Tellus 2010

PF vs. EnKF vs. PF-EnKF

Particle filters

- Particle filters fail unless effective dimension is “small”

EnKF

- Uses Monte Carlo step to approximate forecast covariance
- “Works well”in high-dimensions

Puzzle

- Unweighted EnKF ensemble is “good”
- Weighted PF-EnKF ensemble is “bad”
- Why?

PF vs. EnKF vs. PF-EnKF

Particle filters

- Particle filters fail unless effective dimension is “small”

EnKF

- Uses Monte Carlo step to approximate forecast covariance
- “Works well”in high-dimensions

Puzzle

- Unweighted EnKF ensemble is “good”
- Weighted PF-EnKF ensemble is “bad”
- Why?

Solutions

- Effective dimension is small?

PF vs. EnKF vs. PF-EnKF

Particle filters

- Particle filters fail unless effective dimension is “small”

EnKF

- Uses Monte Carlo step to approximate forecast covariance
- “Works well”in high-dimensions

Puzzle

- Unweighted EnKF ensemble is “good”
- Weighted PF-EnKF ensemble is “bad”
- Why?

Solutions

- ~~Effective dimension is small?~~
- Typical investigations of PF are missing something?

What does “work” mean?

Mean square error

- Definition

$$\boxed{\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left((\bar{x}^k)_j - (x^{\mathrm{true},k})_j\right)^2, \quad \bar{x}^k = \frac{1}{N_e}\sum_{i=1}^{N_e} x_i^k}$$

EnKF "*works*" means "*MSE is small*"

Mean square error

- Definition

$$\boxed{\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left((\bar{x}^k)_j - (x^{\mathrm{true},k})_j\right)^2, \quad \bar{x}^k = \frac{1}{N_e}\sum_{i=1}^{N_e} x_i^k}$$

- Small MSE means "small errors at each grid point"
- MSE is small if MSE is approximately equal to average variance
- EnKF is tuned (localization & inflation) such that:

$$n \cdot \frac{\mathrm{MSE}}{\mathrm{trace}(P)} \approx 1$$

EnKF “*works*” means “*MSE is small*”

Standard example

Model

$$x^k = x^{k-1} + \varepsilon^k, \quad \varepsilon \sim \mathcal{N}(0, I)$$

Observations

$$z^k = x^k + \eta^k, \quad \eta \sim \mathcal{N}(0, I)$$

Result: $E\,(\mathrm{MSE}) = 1 + O(N_e^{-1}) + O(N_e^{-3/2})$

MSE can be small even if N_e is moderate and even if PF-EnKF collapses

Global vs. local assessment of errors/weights

- Small MSE is local assessment of error: data assimilation is useful if errors in each dimension are small

  Example: If one makes small errors in a weather forecast in various locations around the globe, then one would declare success, not failure

Global vs. local assessment of errors/weights

- Small MSE is local assessment of error: data assimilation is useful if errors in each dimension are small

  Example: If one makes small errors in a weather forecast in various locations around the globe, then one would declare success, not failure
- During weight calculation, small errors in each dimension add up and cause the collapse of all particle filters

Global vs. local assessment of errors/weights

- Small MSE is local assessment of error: data assimilation is useful if errors in each dimension are small

  Example: If one makes small errors in a weather forecast in various locations around the globe, then one would declare success, not failure
- During weight calculation, small errors in each dimension add up and cause the collapse of all particle filters
- Weights may turn a useful ensemble into one that is not useful (collapse)

Why does PF-EnKF collapse?

So far

- Unweighted *unlocalized* EnKF ensemble is “bad”
- Unweighted *localized* EnKF ensemble is “good”
- Weighted *localized* EnKF ensemble is “bad”

Why does PF-EnKF collapse?

So far

- Unweighted *unlocalized* EnKF ensemble is “bad”
- Unweighted *localized* EnKF ensemble is “good”
- Weighted *localized* EnKF ensemble is “bad”

Solution of puzzle: weight localization

- Weights of PF-EnKF are *not localized*, but *ensemble is localized*
- PF-EnKF, and other PF, fails because weights are *not* localized

Why does PF-EnKF collapse?

So far

- Unweighted *unlocalized* EnKF ensemble is “bad”
- Unweighted *localized* EnKF ensemble is “good”
- Weighted *localized* EnKF ensemble is “bad”

Solution of puzzle: weight localization

- Weights of PF-EnKF are *not localized*, but *ensemble is localized*
- PF-EnKF, and other PF, fails because weights are *not* localized
- Localization exploits *banded* problem structure
- Similar to numerical linear algebra:
    - *Matrix computations in high dimension are difficult in general*
    - *Feasible if matrix is low-rank -> small effective dimension*
    - *Feasible if matrix is banded -> localization*

Standard example

Model

$$x^k = x^{k-1} + \varepsilon^k, \quad \varepsilon \sim \mathcal{N}(0, I)$$

Observations

$$z^k = x^k + \eta^k, \quad \eta \sim \mathcal{N}(0, I)$$

- Problem can be decoupled into *n* independent scalar sub-problems
- Apply scalar particle filter to each sub-problem independently
- Exponential scaling with (effective) dimension disappears

Standard example

Model

$$x^k = x^{k-1} + \varepsilon^k, \quad \varepsilon \sim \mathcal{N}(0, I)$$

Observations

$$z^k = x^k + \eta^k, \quad \eta \sim \mathcal{N}(0, I)$$

- Problem can be decoupled into n independent scalar sub-problems
- Apply scalar particle filter to each sub-problem independently
- Exponential scaling with (effective) dimension disappears

- "Easy" for diagonal or linear problems
- Difficult for non-diagonal and nonlinear problems (complex, multivariate relationships, or "balance")
- Weights/importance sampling only useful in NWP, probably many other problems, if localized
- Localization of PF in nonlinear problems is "hot topic"

Summary

- Localization is key to make EnKF feasible in large dimensions
- Localization is key to make importance sampling/particle filters feasible in large dimension
- Same as numerical linear algebra in large dimensions: problems must be sparse (low effective dimension) or sparse/banded (localization)

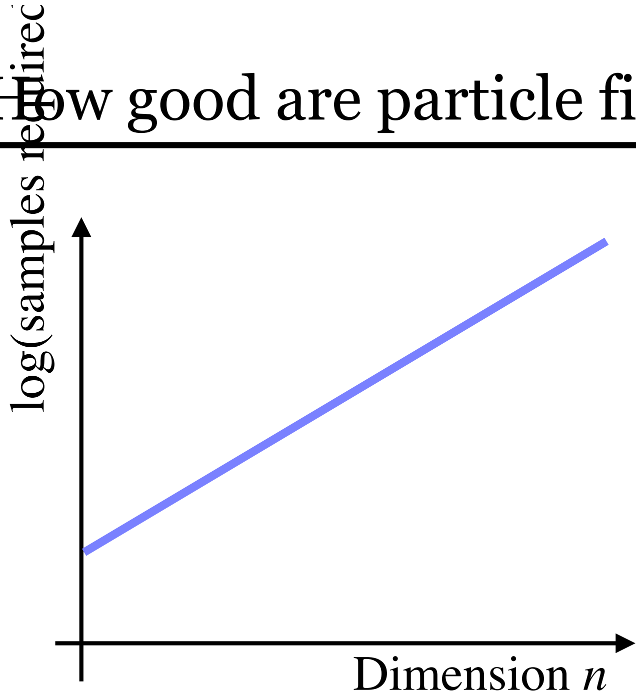
Thank you.

References

- M. Morzfeld, D. Hodyss, and C. Snyder, *What the collapse of the ensemble Kalman filter tells us about localization of particle filters,* Tellus A: Dynamic Meteorology and Oceanography 69:1, 1283809 (2017).
- C. Snyder, T. Bengtsson and M. Morzfeld, *Performance bounds for particle filters using the optimal proposal*, Monthly Weather Review 143, 4750—4761 (2015).
- A.J. Chorin and M. Morzfeld, *Conditions for successful data assimilation*, Journal of Geophysical Research - Atmospheres 118, 11522-11533 (2013).

Supported by: NSF, Alfred P. Sloan Foundation, U.S. Naval Research Laboratory

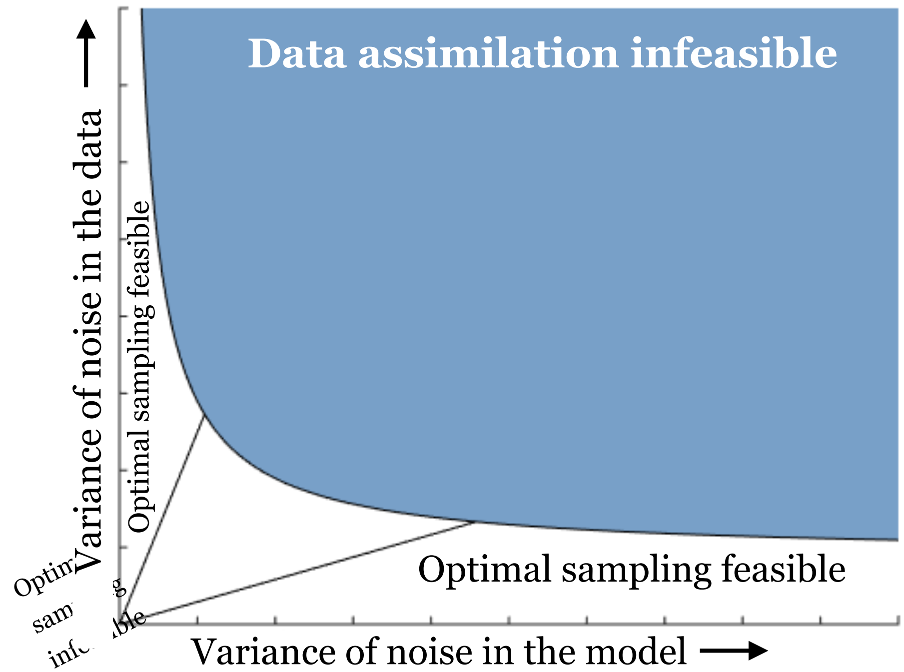
How good are particle filters



- In simple examples, *variance of weights* scales exponentially with the dimension[*]
- Ensemble size scales exponentially with dimension (“**collapse**”)

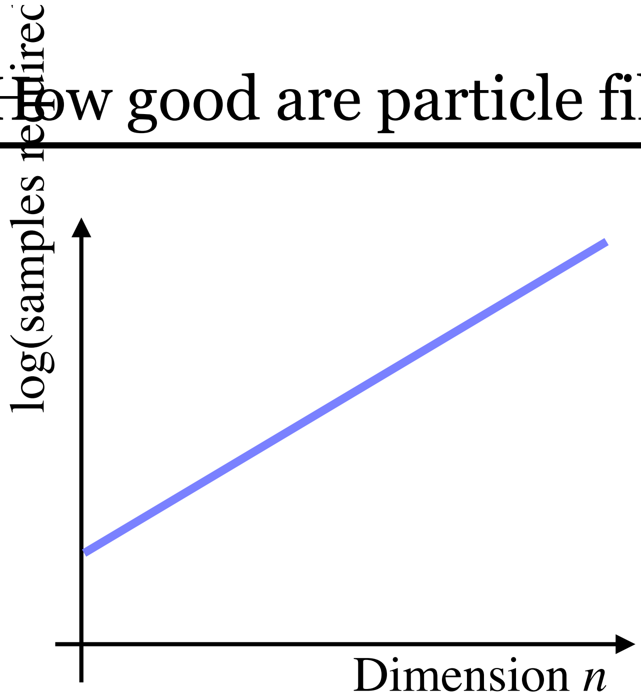
$$N_{\mathrm{eff}} = \frac{N_e}{G}, \quad G = \exp(n), \quad N_e \propto \exp(n) N_{\mathrm{eff}}$$

- Correlations reduce the “effective dimension” of the problem[**]
- Ensemble size scales exponentially with *effective* dimension



[*] *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015
[**] *Chorin & Morzfeld*, Journal of Geophysical Research , 2013

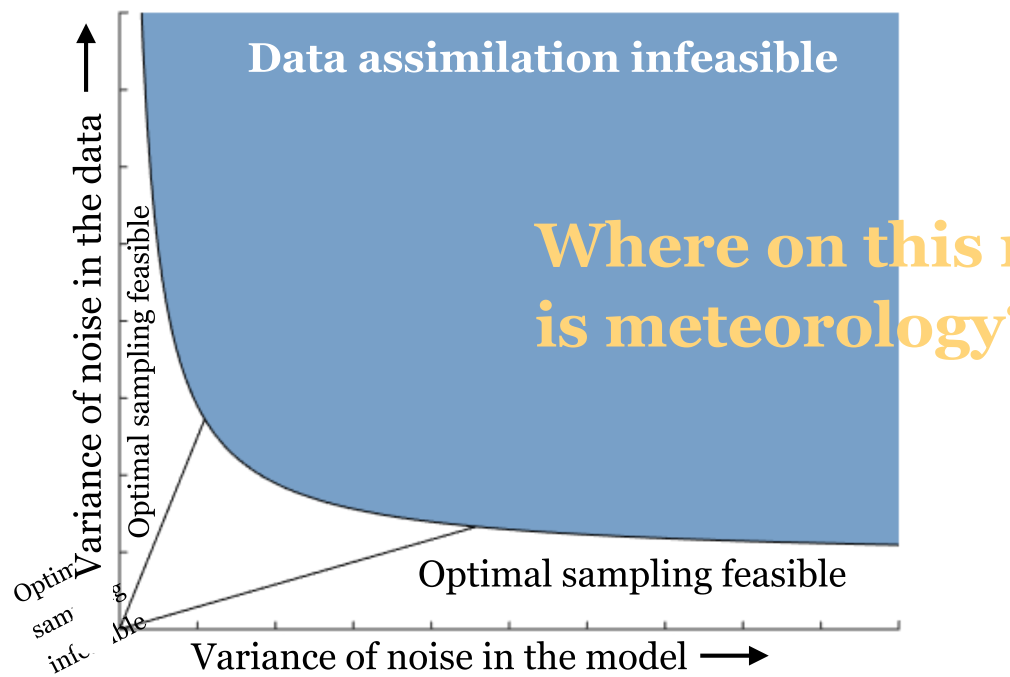
How good are particle filters



- In simple examples, *variance of weights* scales exponentially with the dimension*
- Ensemble size scales exponentially with dimension ("**collapse**")

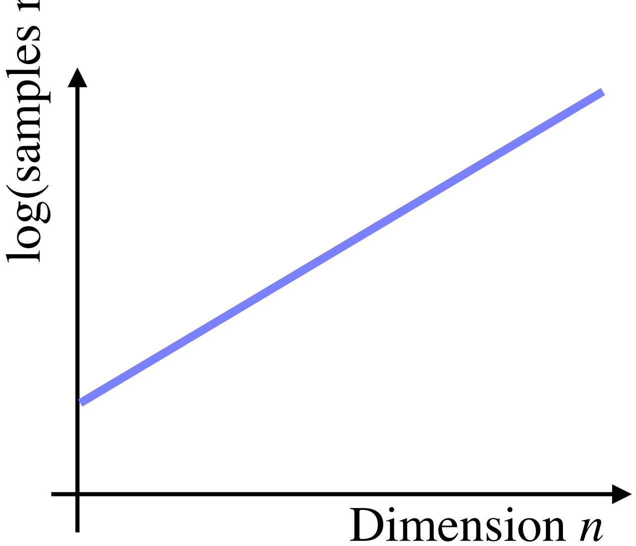
$$N_{\text{eff}} = \frac{N_e}{G}, \quad G = \exp(n), \quad N_e \propto \exp(n) N_{\text{eff}}$$

- Correlations reduce the "effective dimension" of the problem**
- Ensemble size scales exponentially with *effective* dimension



* *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015
** *Chorin & Morzfeld*, Journal of Geophysical Research , 2013

How good are particle filters

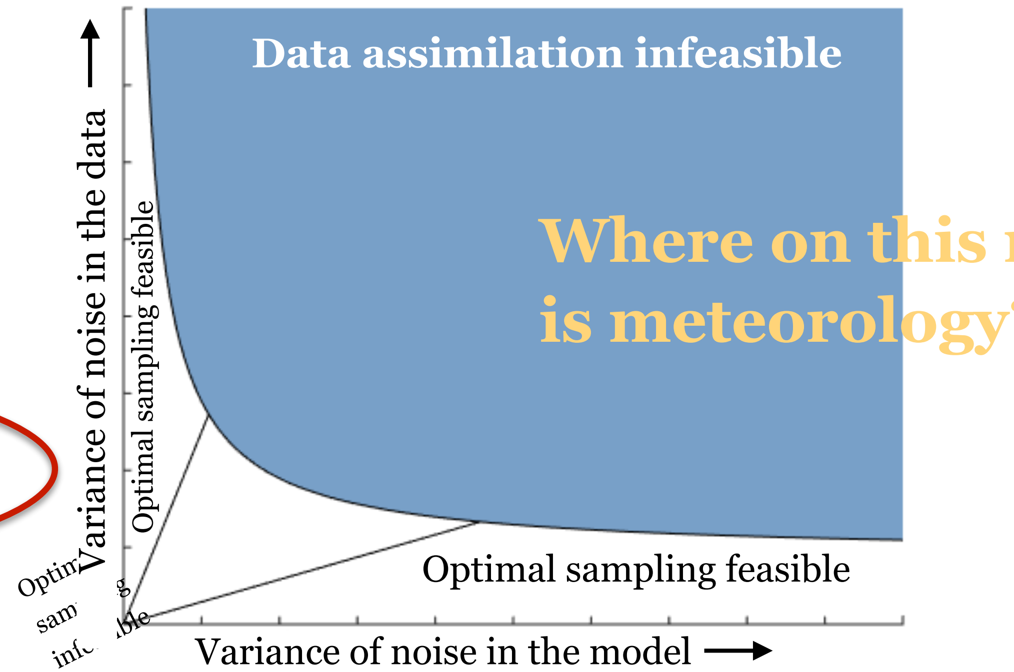


- In simple examples, *variance of weights* scales exponentially with the dimension*
- Ensemble size scales exponentially with dimension ("**collapse**")

$$N_{\text{eff}} = \frac{N_e}{G}, \quad G = \exp(n), \quad N_e \propto \exp(n) N_{\text{eff}}$$

- Correlations reduce the "effective dimension" of the problem**
- Ensemble size scales exponentially with *effective* dimension

What if the effective dimension of NWP is not small?



* *Snyder, Bengtsson, Morzfeld*, Monthly Weather Review, 2015
** *Chorin & Morzfeld*, Journal of Geophysical Research , 2013

Subtleties of different densities

Smoothing distribution

- Probability distribution of *trajectories* conditioned on data

$$p^k \propto p^{k-1}\ p(x^k|x^{k-1})p(z^k|x^k)$$

- Particle filters usually applies to this distribution

Filtering distribution

- Probability distribution of state conditioned on data

$$p(x^k|z^{1:k}) \propto p(z^k|x^k)p(x^k|z^{1:k-1})$$

- EnKF usually applies to this distribution

Subtleties of different densities

Smoothing distribution

- Probability distribution of *trajectories* conditioned on data

$$p^k \propto p^{k-1}\; p(x^k|x^{k-1})p(z^k|x^k)$$

- Particle filters usually applies to this distribution

Filtering distribution

- Probability distribution of state conditioned on data

$$p(x^k|z^{1:k}) \propto p(z^k|x^k)p(x^k|z^{1:k-1})$$

- EnKF usually applies to this distribution

Question

- Weights are computed with respect to *smoothing* distribution?
- What happens when we compute weights with respect to *filtering* distribution?

Subtleties of different densities

Standard example

Model

$$x^k = x^{k-1} + w^k, \quad w^k \sim \mathcal{N}(0, I)$$

Data (observations)

$$z^k = x^k + v^k, \quad v^k \sim \mathcal{N}(0, I)$$

nKF as importance sampler for $p(x^k|z^{1:k}) \propto p(z^k|x^k)p(x^k|z^{1:k-1})$

---

nsity: $p(x^k|z^{1:k}) = \mathcal{N}(0, I)$

sampling error

al distribution: $q_{\text{EnKF}}(x^k) = \mathcal{N}(0, \sigma^2 I)$, $\sigma^2 = 1 + \beta, \quad \beta \propto 1/\sqrt{N_e}$

Standard example

Model

$$x^k = x^{k-1} + w^k, \quad w^k \sim \mathcal{N}(0, I)$$

Data (observations)

$$z^k = x^k + v^k, \quad v^k \sim \mathcal{N}(0, I)$$

nKF as importance sampler for $\quad p(x^k|z^{1:k}) \propto p(z^k|x^k)p(x^k|z^{1:k-1})$

---

nsity: $\quad p(x^k|z^{1:k}) = \mathcal{N}(0, I)$

al distribution: $\quad q_{\text{EnKF}}(x^k) = \mathcal{N}(0, \sigma^2 I), \quad \sigma^2 = 1 + \beta, \quad \beta \propto 1/\sqrt{N_e}$

Result: Effective sample size scales linearly with dimension

Standard example

Model

$$x^k = x^{k-1} + w^k, \quad w^k \sim \mathcal{N}(0, I)$$

Data (observations)

$$z^k = x^k + v^k, \quad v^k \sim \mathcal{N}(0, I)$$

nKF as importance sampler for $\quad p(x^k|z^{1:k}) \propto p(z^k|x^k)p(x^k|z^{1:k-1})$

---

nsity: $\quad p(x^k|z^{1:k}) = \mathcal{N}(0, I)$

al distribution: $\quad q_{\mathrm{EnKF}}(x^k) = \mathcal{N}(0, \sigma^2 I), \quad \sigma^2 = 1 + \beta, \quad \beta \propto 1/\sqrt{N_e}$

Result: Effective sample size scales linearly with dimension

Idea: Effective dimension ≈ EnKF ensemble size ≈ 50 - 100

Standard example