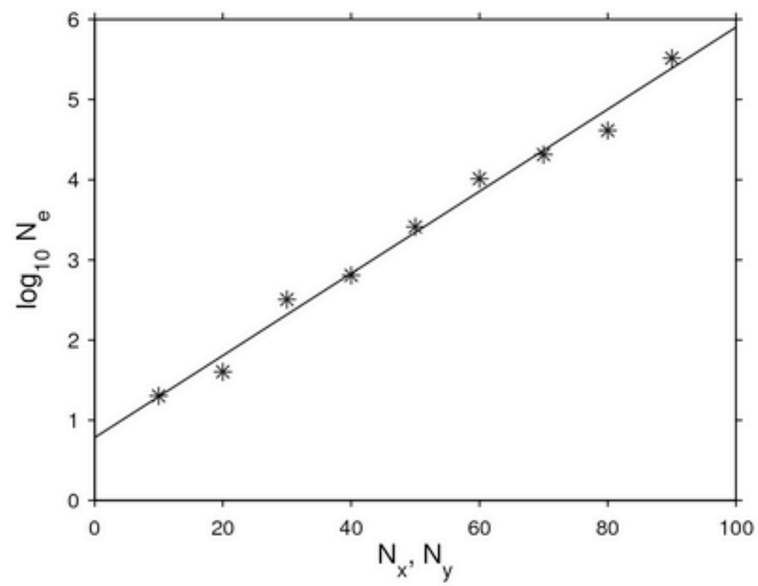
Obstacles to High-Dimensional Particle Filtering



Snyder et al. 2007: *Mon. Wea. Rev.*, submitted.

- Chris Snyder (NCAR)
- Thomas Bengtsson (Bell Labs)
- Peter Bickel (UC Berkeley)
- Jeff Anderson (NCAR)

Preliminaries

Exceptions from Ide et al. (1997) notation:

- $\mathbf{x}$ is the (true) system state to be estimated
- $\dim(\mathbf{x}) = N_x$, $\dim(\mathbf{y}) = N_y$

Other stuff:

- ensemble size $= N_e$
- Assume $\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon$

The Particle Filter (PF)

Non-parametric Monte-Carlo method

- given forecast ensemble, $\{\mathbf{x}_i^f, i = 1, \ldots, N_e\}$
- approximate posterior pdf as sum of point masses,

$$p(\mathbf{x}|\mathbf{y}) \approx \sum_{i=1}^{N_e} w_i \delta(\mathbf{x} - \mathbf{x}_i^f)$$

- weights w_i given by

$$w_i = \frac{p(\mathbf{y}|\mathbf{x}_i^f)}{\sum_{j=1}^{N_e} p(\mathbf{y}|\mathbf{x}_j^f)}$$

- Exact implementation of Bayes' rule as $N_e \to \infty$

Widely applied, and effective, in low-dim'l systems

What happens for high-dim'l systems?

Collapse of Weights and Resampling

Ensemble methods generally underestimate uncertainty of posterior

- for PF, $\max w_i \to 1$ as N_x, N_y increase with N_e fixed
- when cycling over multiple observation times, tendency for collapse increases with t

Crucial to resample from PF posterior distribution

- members with small weights ignored, those with large weights resampled multiple times

But resampling begins from estimated posterior distribution

- does not improve quality of estimate of $p(\mathbf{x}|\mathbf{y})$

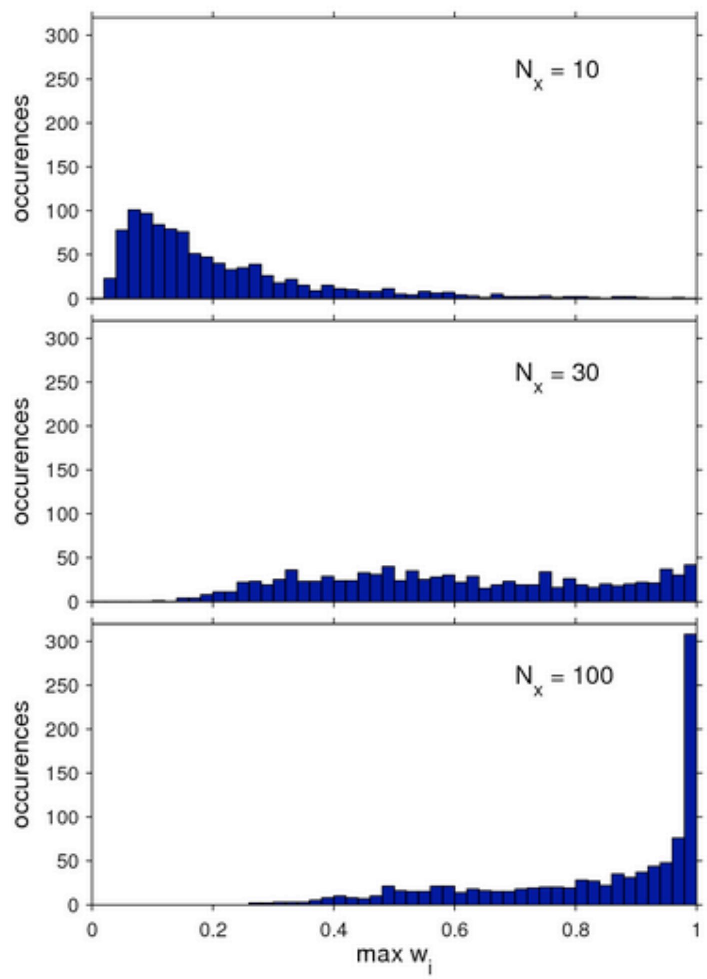
Simple Example

Gaussian state and obs, identity covariances

- prior: $\mathbf{x} \sim N(0, \mathbf{I})$
- identity observations: $N_y = N_x$, $\mathbf{H} = \mathbf{I}$
- observation error: $\epsilon \sim N(0, \mathbf{I})$

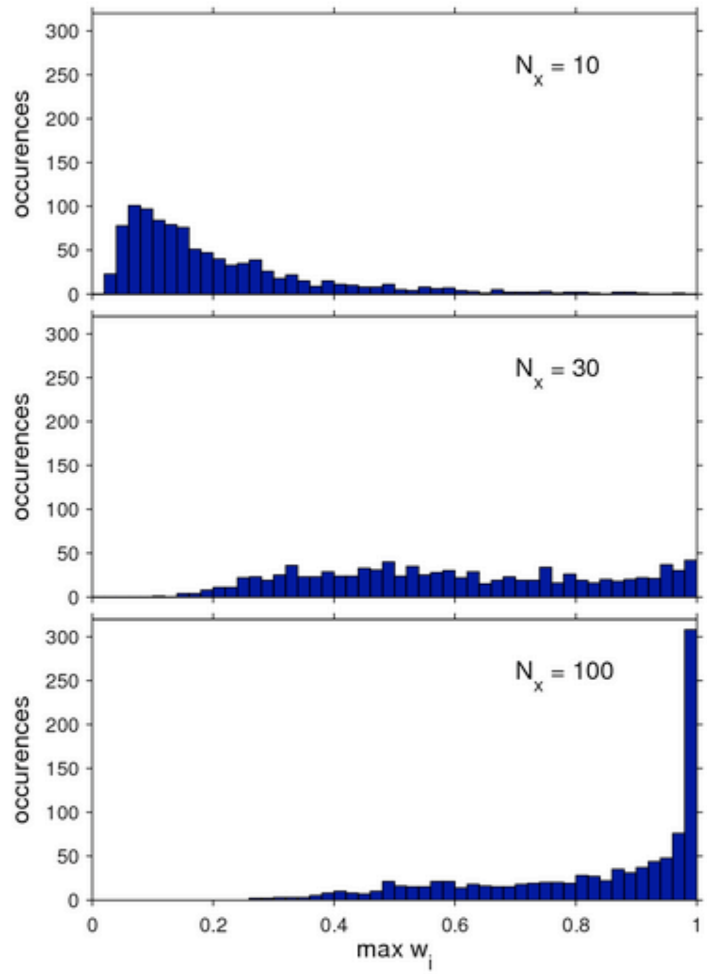
Behavior of $\max w_i$

- $N_e = 10^3$; $N_x = 10, 30, 100$; 10^3 realizations



Behavior of $\max w_i$

- $N_e = 10^3$; $N_x = 10, 30, 100$; 10^3 realizations



av. squared error of posterior mean

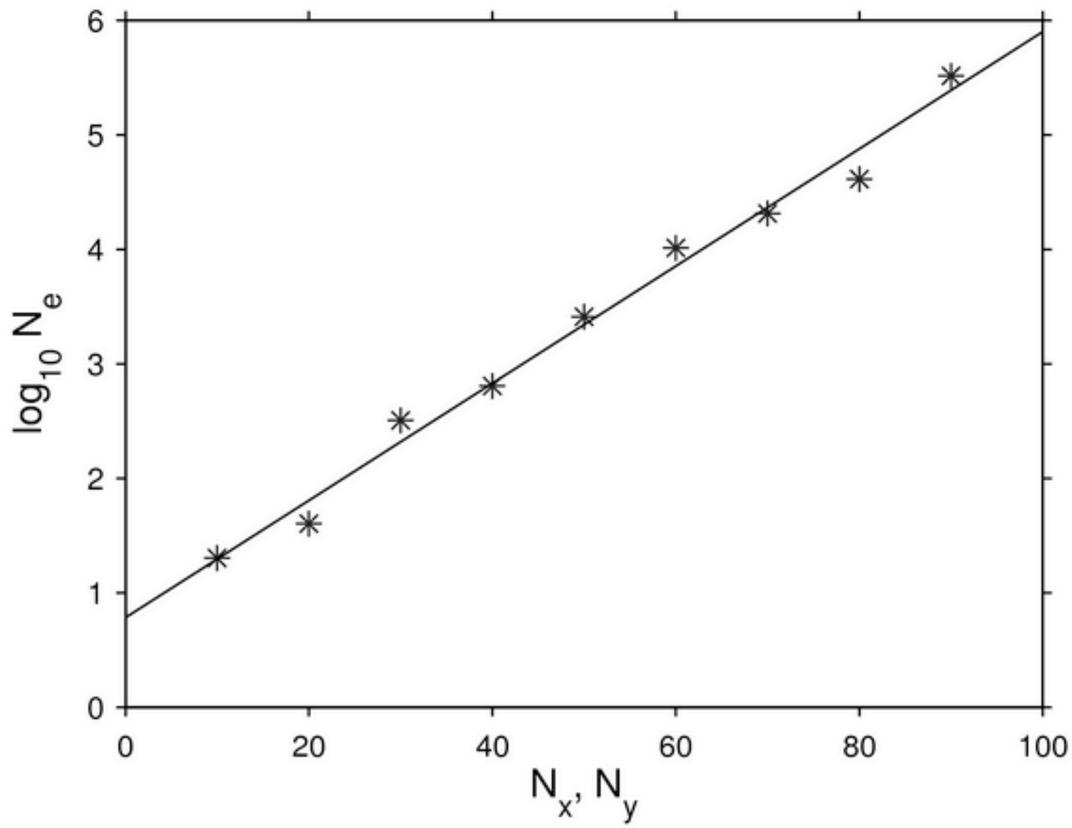
= 5.5

= 25

= 127

Required Ensemble Size

- N_e s.t. PF mean has expected error less than obs



Asymptotic Conditions for Collapse

$w_i \propto p(\mathbf{y}^o|\mathbf{x}_i^f)$

- So $\max w_i \approx 1$ if one $p(\mathbf{y}^o|\mathbf{x}_i^f)$ is much larger than all others

Consider i.i.d. observation error

- Each component ϵ_j of ϵ is i.i.d. with density $f(\cdot)$.
- Likelihood factors as

$$p(\mathbf{y}^o|\mathbf{x}_i^f) = p(\mathbf{y}^o|\mathbf{x}^f{}_i) = \prod_{j=1}^{N_y} f\left(y_j^o - (\mathbf{H}\mathbf{x}^f{}_i)_j\right)$$

Likelihood depends only on N_y, $f(\cdot)$ and the prior as reflected in the observed variables $\mathbf{Hx}$. No direct dependence on N_x.

Asymptotic Conditions for Collapse (cont.)

Rewrite observation likelihood

- Let $V_{ij} = \psi\left(y^o_j - (\mathbf{Hx}^f_i)_j\right)$, where $\psi(\cdot) = \log f(\cdot)$.
- So likelihood given by

$$p(\mathbf{y}^o|\mathbf{x}^f_i) = \exp\left(\sum_{j=1}^{N_y} \psi\left(y^o_j - (\mathbf{Hx}^f{}_i)_j\right)\right) = \exp\left(-\sum_{j=1}^{N_y} V_{ij}\right).$$

Asymptotic Conditions for Collapse (cont.)

Rewrite observation likelihood again

- Let

$$S_i = (\sum_{j=1}^{N_y} V_{ij} - \mu)/\tau, \quad \mu \equiv \sum_{j=1}^{N_y} E\left(V_{ij}\right), \quad \tau^2 \equiv \mathrm{var}\left(\sum_{j=1}^{N_y} V_{ij}\right)$$

- Then likelihood becomes

$$p(\mathbf{y}^o|\mathbf{x}_i^f) = \exp\left(-\mu - \tau S_i\right)$$

Wish to analyze statistics of $\min\{S_i, i = 1, \ldots, N_e\}$.

Analytic progress possible when S_i is approximately Gaussian and τ^2, N_e are large.

Asymptotic Conditions for Collapse (cont.)

For S_i to be approximately Gaussian, need

- $N_y \to \infty$, so that $S_i =$ sum of many random variables.
- V_{ij} must be "sufficiently" independent given $\mathbf{y}^o$ and have "sufficiently" similar distributions as j varies.
- A necessary requirement for the above is $N_x \to \infty$.

Then can show

- When $N_e \to \infty$ and $\tau^2 / \log N_e \to \infty$

$$E\left(1/\max w_i\right) \approx 1 + \frac{\sqrt{2 \log N_e}}{\tau}$$

- See Bengtsson et al. 2007, Bickel et al. 2007
- Thus, weights collapse $(\max w_i \to 1)$ unless N_e scales as $\exp(\tau^2/2)$.

The Gaussian-Gaussian Case

Suppose prior and obs error are Gaussian (and **H** is linear)

- let $\mathbf{x} \sim N(0, \mathbf{P})$ and $\epsilon \sim N(0, \mathbf{R})$
- V_{ij} $(j = 1, \ldots, N_y)$ not independent when elements of **x** are correlated

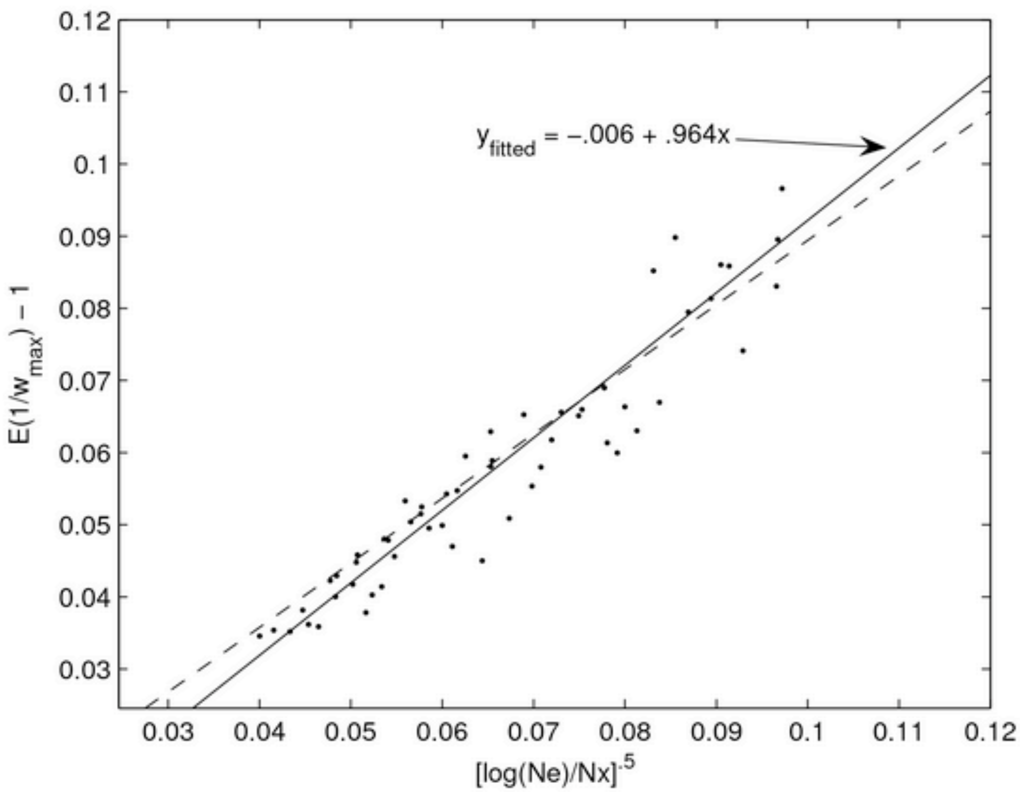
W.l.o.g., $\mathbf{R} = \mathbf{I}$ and $\mathbf{HPH}^T = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_{N_y}^2)$. Then:

- $V_{ij} = -\frac{1}{2}\left(y_j^o - (\mathbf{Hx}^f{}_i)_j\right)^2 + c$
- $\tau^2 = \mathrm{var}\left(\sum_{j=1}^{N_y} V_{ij}\right) = \sum_{j=1}^{N_y} \lambda_j^2 \left(\frac{1}{2}\lambda_j^2 + y_j^{o2}\right)$
- $E(\tau^2) = \sum_{j=1}^{N_y} \lambda_j^2 \left(1 + \frac{3}{2}\lambda_j^2\right)$

Pdf of S_i converges to Gaussian as $N_y \to \infty$ iff $\sum_{j=1}^{N_y} \lambda_j^2 \to \infty$.

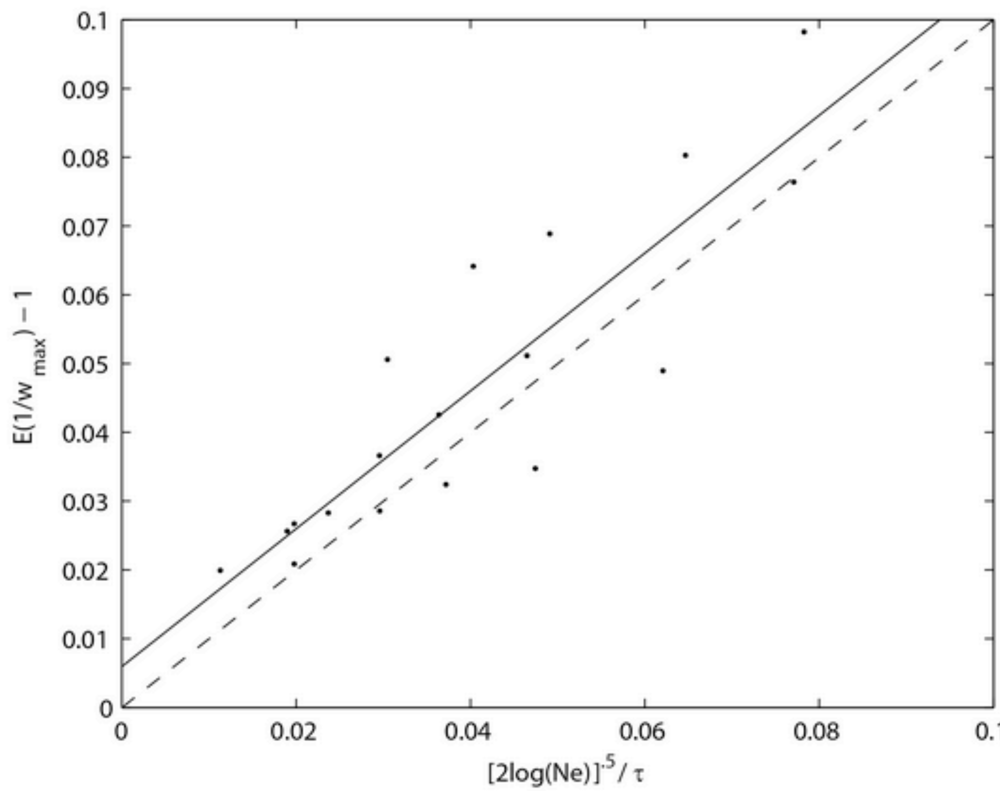
Simulations vs. Asymptotics (i)

- $\lambda_j^2 = 1$, 60 (N_y, N_e) pairs, 10^3 realizations
- Asymptotic behavior dashed



Simulations vs. Asymptotics (ii)

- $\lambda_j^2 = cj^{-\theta}$, with various θ and c; fix $N_y = 4 \times 10^3$ and $N_e = 10^5$; 10^3 realizations
- Asymptotic behavior dashed



Two Views of the Obstacle

Curse of dimensionality

- Prior and posterior pdfs are $\sim$ mutually singular in high dimensions
- Sample from prior has v. low probability under posterior
- E.g., high-dim'l Gaussians are restricted to shells of hyper-spheres

Weights apply globally

- For finite N_e, assimilation of single observation changes estimate for all state variables
- Bad in typical case that most state variables are nearly independent of observation prior

Potential Fixes for High-Dim'l PF

Localization

- Allow any observation to influence posterior estimate only locally in space (c.f. EnKF)
- Not yet clear how to achieve; some ideas in Bengtsson et al. (2003)
- Spatial smoothness is problematic: How to patch together local, non-Gaussian estimates?

Resampling

- Not clear this can remove detrimental effects of collapse
- Changes multiplier but not requirement that $N_e \sim \exp(\tau^2)$?

Summary

PF for high-dim'l system may require v. large ensemble

- "Problem size" not N_x per se; instead determined by τ^2, variance of observation log-likelihood.
- When log-likelihood has approximately Gaussian pdf, required ens. size $N_e \sim \exp(\tau^2/2)$.
- Approximation as Gaussian holds when N_y and N_x are large and obs priors are not too dependent.
- Can think of τ^2 as dimension of identity-covariance, Gaussian example with same collapse properties.

Resampling is not an obvious fix

Further work

- Test predictions from Gaussian-Gaussian case on non-Gaussian examples
- Quantitative, empirical guidelines for required N_e
- Algorithm for spatially local, non-Gaussian update