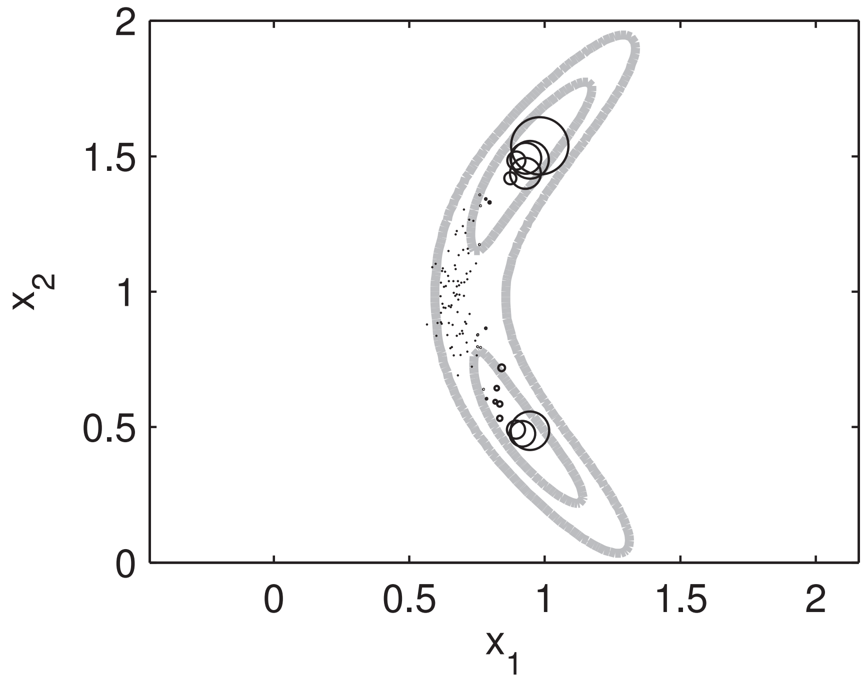
A Brief Review of Localization for Particle Filters



▷ Chris Snyder
National Center for Atmospheric Research*, Boulder Colorado, USA

* NCAR is supported by the US National Science Foundation.

Intro (I): Preliminaries

Wish to estimate state $\mathbf{x}$ given observations $\mathbf{y}$

- ▷ $\mathbf{x} =$ discretized representation of atmospheric state
- ▷ $\mathbf{y} = h(\mathbf{x}) + \epsilon$, where ϵ is random error

Preliminaries

Cannot determine $\mathbf{x}$ precisely

- both observation, forecasts have error
- most that we can know is $p(\mathbf{x}|\mathbf{y})$

$p(\mathbf{x}|\mathbf{y})$ is complete answer to estimation problem

- would like to approximate without assumptions, e.g. of Gaussianity
- fully utilize obs with nonlinear forward operator or non-Gaussian errors
- allow for non-Gaussianity in prior for scales/phenomena that are significantly nonlinear

Preliminaries: Calculating the Conditional PDF

Bayes' rule:

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})p(\mathbf{y}|\mathbf{x})/p(\mathbf{y})$$

Intro (II): Particle Filters

Additional terminology and notation

- particles $\equiv$ ensemble members, sample $\equiv$ ensemble
- $N_y = \dim \mathbf{y}$, $N_e =$ ensemble size (and $N_x = \dim \mathbf{x}$)
- $\mathbf{x}_k^i = i$th particle/member at t_k, the kth observation time
- subscript $k:l$ indicates concatenation of quantities from $t_k, \ldots, t_l$

Particle Filters

Sequential Monte-Carlo method to sample from $p(\mathbf{x}_k|\mathbf{y}_{1:k})$

- fully general approach; converges to Bayes rule as $N_e \to \infty$,
- Large literature for low-dimensional systems, plus recent interest in geophysics (e.g. van Leeuwen 2003, 2010; Morzfeld et al. 2011; Papadakis et al. 2010)

Particle Filters

Generate particles $\mathbf{x}_k^i$ at t_k (based on $\mathbf{x}_{k-1}^i$ and $\mathbf{y}_k$)

Compute weights

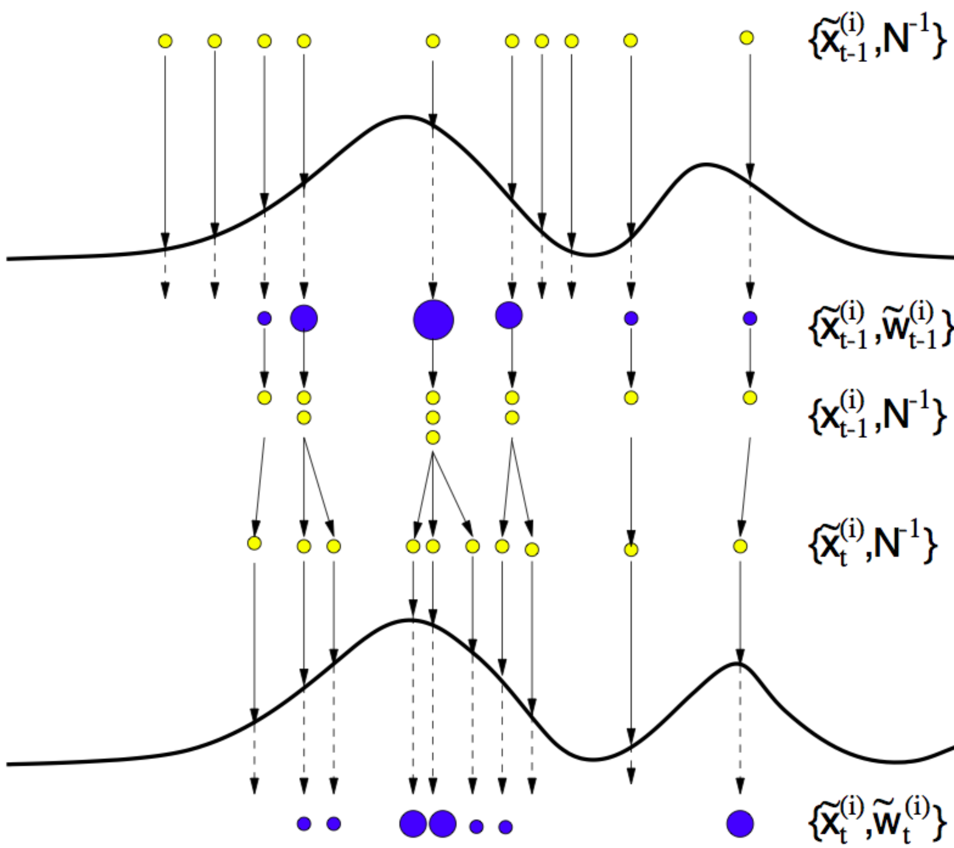
Resample

Repeat

Particle Filters

Begin with members $\mathbf{x}_{k-1}^i$ drawn from $p(\mathbf{x}_{k-1}|\mathbf{y}_{k-1})$

- ▷ generate $\mathbf{x}_k^i$ by evolving each member to t_k under the system dynamics
- ▷ compute weight, given new obs $\mathbf{y}_k$: $w_k^i \propto p(\mathbf{y}_k|\mathbf{x}_k^i)$
- ▷ resample



Doucet, de Freitas and Gordon: 2001

Particle Filters

Recall Bayes' rule,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{x}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_k|\mathbf{x}_k)$$

If $p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = {N_e}^{-1} \sum_{i=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}_k^i)$, then

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto \sum_{i=1}^{N_e} p(\mathbf{y}_k|\mathbf{x}_k^i)\delta(\mathbf{x} - \mathbf{x}_k^i)$$

Degeneracy of PF Weights

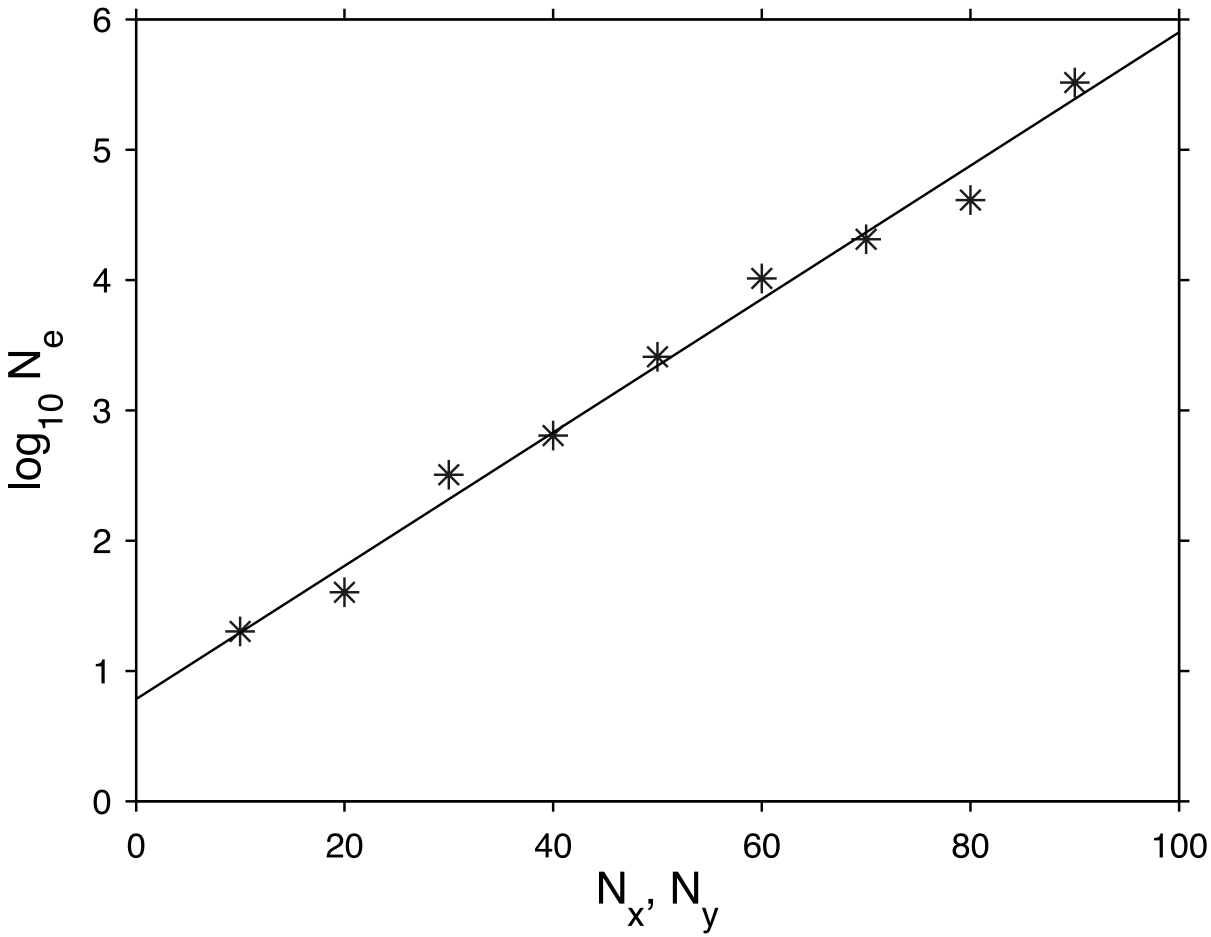
Key issue is tendency for $\max_i w_k^i \to 1$

Test problem: $\mathbf{x} \sim N(0, \mathbf{I})$, $\mathbf{y} = \mathbf{x} + \epsilon$, $\epsilon \sim N(0, \mathbf{I})$

- ignore forecast and resampling steps

Degeneracy of PF Weights

- ensemble size N_e s.t. $E(\max\ w^i) \approx 0.7$, as N_y varies



- N_e must grow exponentially with N_y to avoid degeneracy
  . . . need a better approach

Asymptotics and Bounds

Asymptotics for maximum weight

- let $\tau^2 = \text{var}(-\log \tilde{w}_k^i)$. As τ^2 increases, $\max_i w_k^i \to 1$ unless

$$N_e \sim \exp(\tau^2/2)$$

- applies to both standard and optimal proposals
- valid only if $\log \tilde{w}_k^i$ has $\approx$ Gaussian distribution
- occurs for well observed, high-dimensional systems having many independent and comparably important degrees of freedom.

See:

Bengtsson, T., P. Bickel and B. Li, 2008: in *Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds.

Bickel, P, B. Li, and T. Bengtsson, 2008: in *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, B. Clarke and S. Ghosal Eds.

Snyder, C., T. Bengtsson, P. Bickel and J. Anderson, 2008: *Mon. Wea. Rev.*, **136**, 4629–4640.

Snyder, C., 2012: Proceedings, ECMWF Seminar on Data Assimilation for Atmosphere and Ocean., September 2011.

Snyder, C., T. Bengtsson and M. Morzfeld, 2015: *Mon. Wea. Rev.*, **143**, 4750–4761.

Asymptotics and Bounds

Performance bounds from optimal proposal

- all PF employing SIS must have degeneracy bounded below by that of optimal proposal (Snyder et al. 2015)
- all PF employing SIS will need exponential large N_e to avoid degeneracy in "large" problems

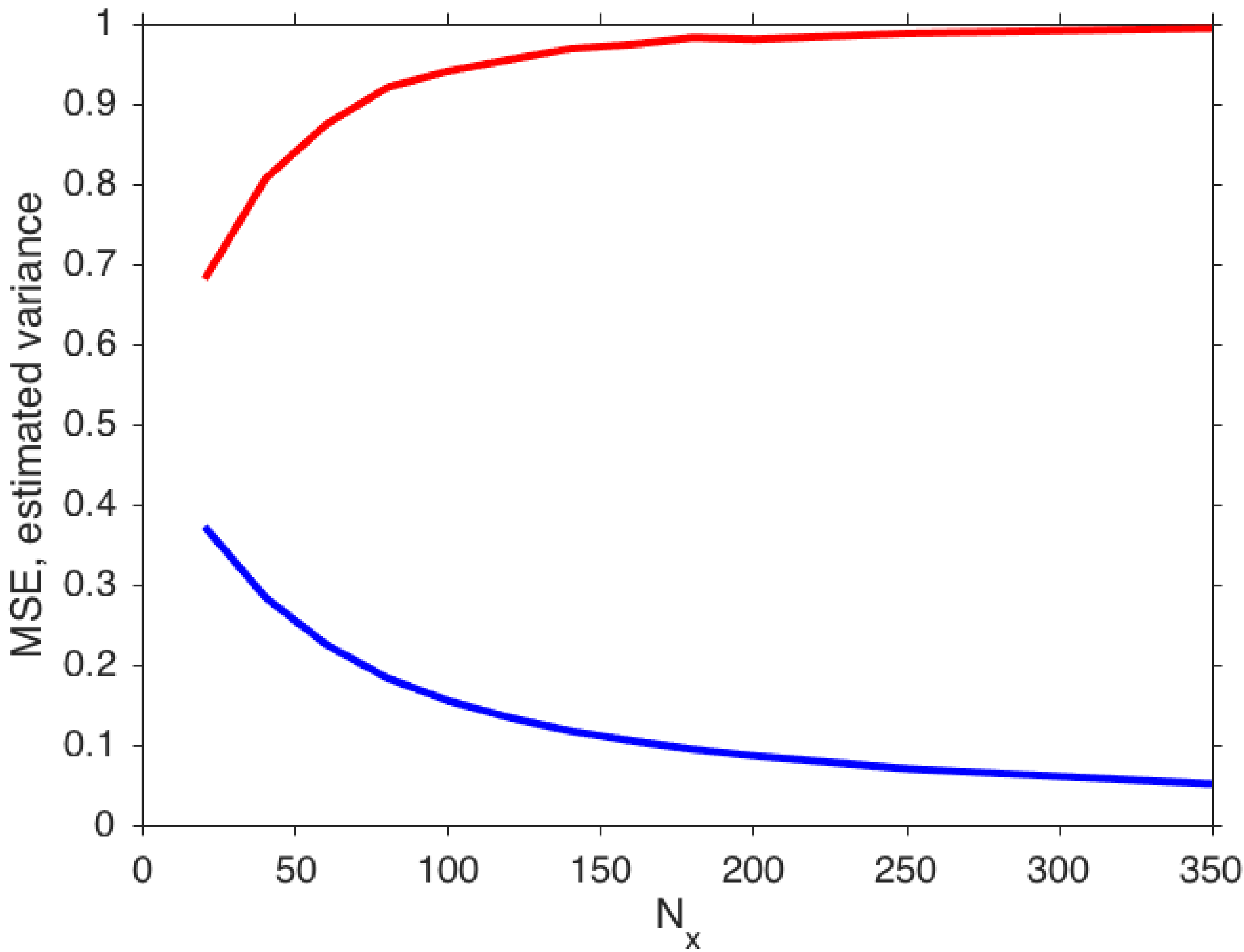
EnKF in High Dimensions

EnKF works when N_x is large. Why?

Return to (first) example: $\mathbf{x} \sim N(0, \mathbf{I})$, $\mathbf{y} = \mathbf{x} + \epsilon$, $\epsilon \sim N(0, \mathbf{I})$

EnKF in High Dimensions

- posterior MSE, predicted posterior variance for $N_e = 20$ as fn of N_x



EnKF in High Dimensions

Sampling error

- $O({N_e}^{-1/2})$ in each element of $\hat{\mathbf{P}}$
- errors correlated across elements, so poor global properties of $\hat{\mathbf{P}}$

Key idea: localization

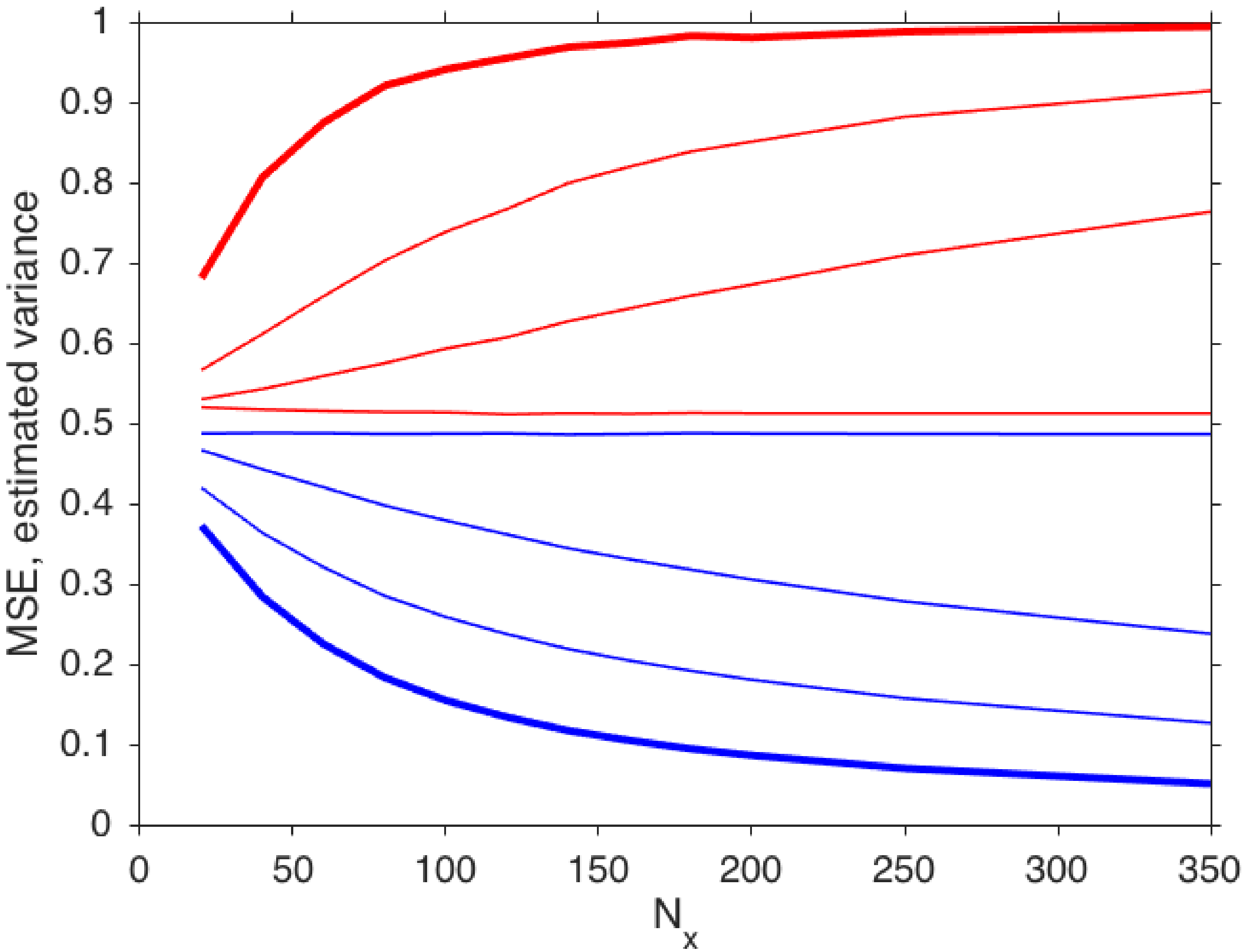
- assume state variables separated by sufficient distance are uncorrelated
- replace noisy estimated covariances by zero at large separation, $\hat{\mathbf{P}} \leftarrow \mathbf{C} \circ \hat{\mathbf{P}}$

Analysis increment

- w/o localization, linear combination of ensemble members
- w/ localization, "spatially varying" linear combination

EnKF in High Dimensions

- as previous, but using various localizations of sample covariance



Localization for Particle Filters

0. Local weights

Easy to compute local weights from spatially local subset of obs;
$\{w^i, i = 1, \ldots, N_e\}$ replaced by $\{\mathbf{w}^i, i = 1, \ldots, N_e; \dim(\mathbf{w}) \leq N_x\}$

Localization for Particle Filters

1. The obvious approach

If we follow usual PF, need to resample locally based on $\{\mathbf{w}^i\}$. Retains fully general, non-parametric algorithm—but need not respect spatial continuity etc. between local regions.

One option is conditional resampling to maintain continuity. Recent quasi-deterministic techniques are also promising.

Bengtsson et al. 2003, *J. Geophys. Res.*, **62**(D24), 8775–8785

Penny and Miyoshi 2016, *Nonlinear Proc. Geophys.*, **23**, 391–405.

Robert and Künsch 2016, arXiv:1605.05476v2.

Localization for Particle Filters

2. Moment matching

With local weights, can estimate posterior mean locally:

$$\hat{\mathbf{x}} = \sum \mathbf{w}^i \circ \mathbf{x}^i$$

If weights vary smoothly, no issues with continuity. Sliding spatial window, as in LETKF, is useful.

Also correct deviations from $\hat{\mathbf{x}}$ consistent w/ local, weighted-sample $\hat{\mathbf{P}}$.

Not a full solution: approximates only the first two moments of $\mathbf{x}|\mathbf{y}$.

Lei and Bickel 2011 *Mon. Wea. Rev.*, **39**, 3964–3973

Tödter and Ahrens 2015 *Mon. Wea. Rev.*, **143**, 1347–1367

Poterjoy 2016 *Mon. Wea. Rev.*, **144**, 59–76

Localization for Particle Filters

3. Optimal coupling

Seek a map that, when applied to random variable with density $p(\mathbf{x})$, yields r.v. with density $p(\mathbf{x}|\mathbf{y})$. For discrete r.v., becomes a linear transformation between prior and weighted posterior ensembles minimizing an expected distance.

Again natural to compute weights and transform from local subset of obs. Transform again varies smoothly with location, with benefit that transform is designed to minimize change necessary in each $\mathbf{x}_k^i$.

Expensive to compute transform, but can approximate while retaining first and second moments of posterior.

Cheng and Reich 2015, in *Nonlinear Data Assimmilation*, Springer-Verlag.

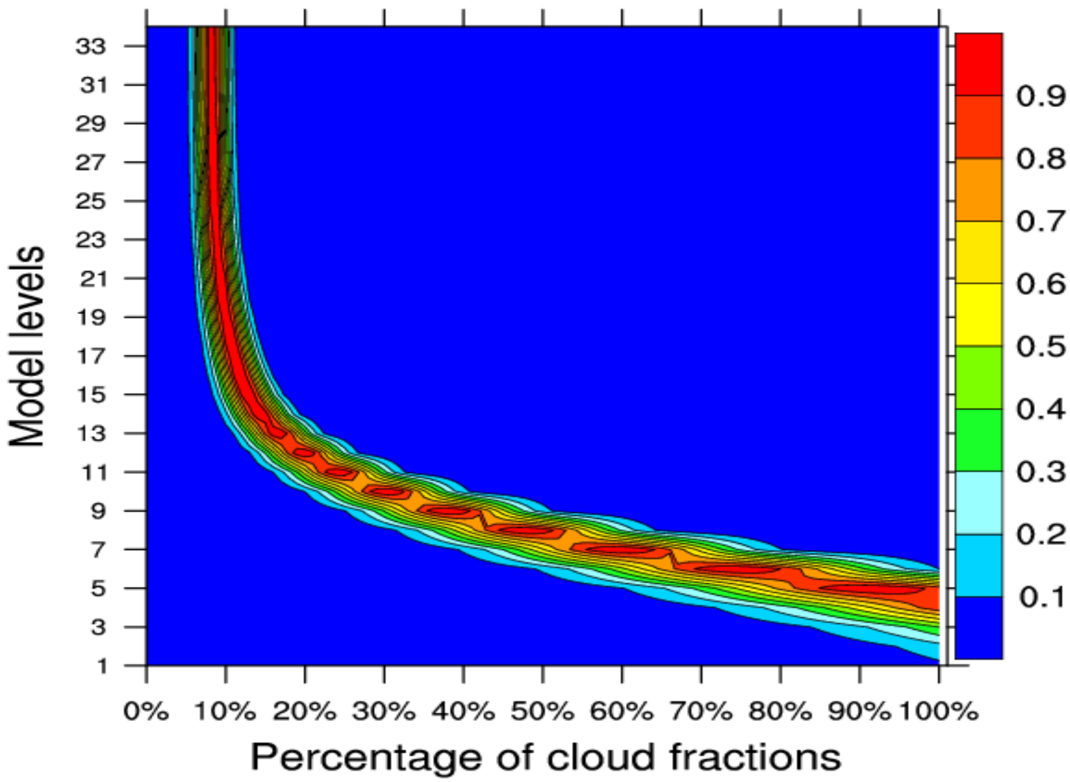
de Wiljes, Acevedo and Reich 2016, arXiv:1608.08179v2.

Summary

- Naive PF and EnKF fail similarly when $N_e \ll N_x$, N_y: poor use of obs and hugely optimistic estimates of posterior uncertainty
- PF (using sequential importance sampling) requires exponentially large N_e to avoid this failure in many problems of interest
- Localization, à la EnKF, likely essential to PF for large problems. Harder, however; must maintain continuity during local resampling.
- Many promising options on the table for localization. Most other enhancements (hybrids, splitting likelihood, beter proposals) will help in any specific case, but cannot cure $\exp(N_e)$ issue.

Why $p(\mathbf{x}|\mathbf{y})$?

- $\mathbf{x} =$ fraction c of opaque, black cloud at each model level
- prior $p(\mathbf{x})$: c nonzero at no more that one model level
- $p(\mathbf{x}|$GOES radiance, single channel and single pixel$)$, unnormalized

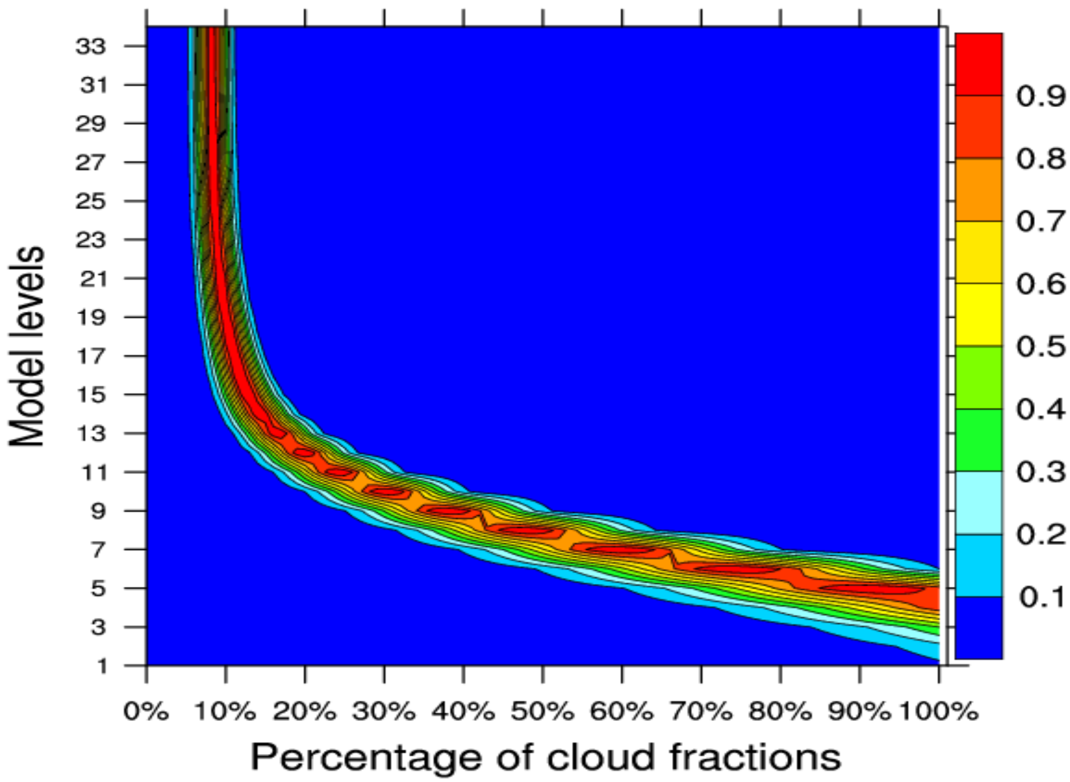


courtesy Dongmei Xu, NCAR

Why $p(\mathbf{x}|\mathbf{y})$? (cont.)

Standard estimators

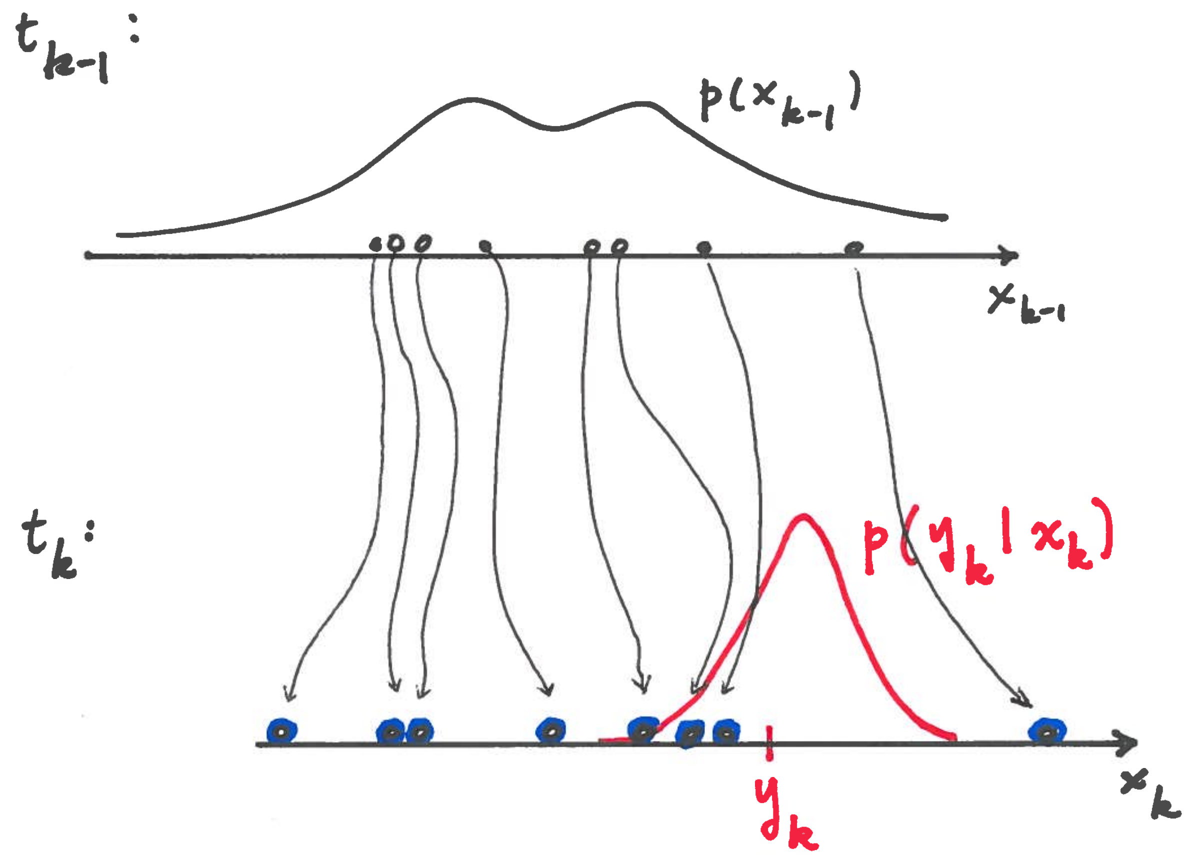
- $E(\mathbf{x}|\mathbf{y})$, state that minimizes expected squared error
- $\arg\max p(\mathbf{x}|\mathbf{y})$, most likely state



courtesy Dongmei Xu, NCAR

Intuition for Weight Degeneracy

- 1D intuition: $p(\mathbf{y}_k|\mathbf{x}_k^i)$ is narrow relative to spread of prior particles



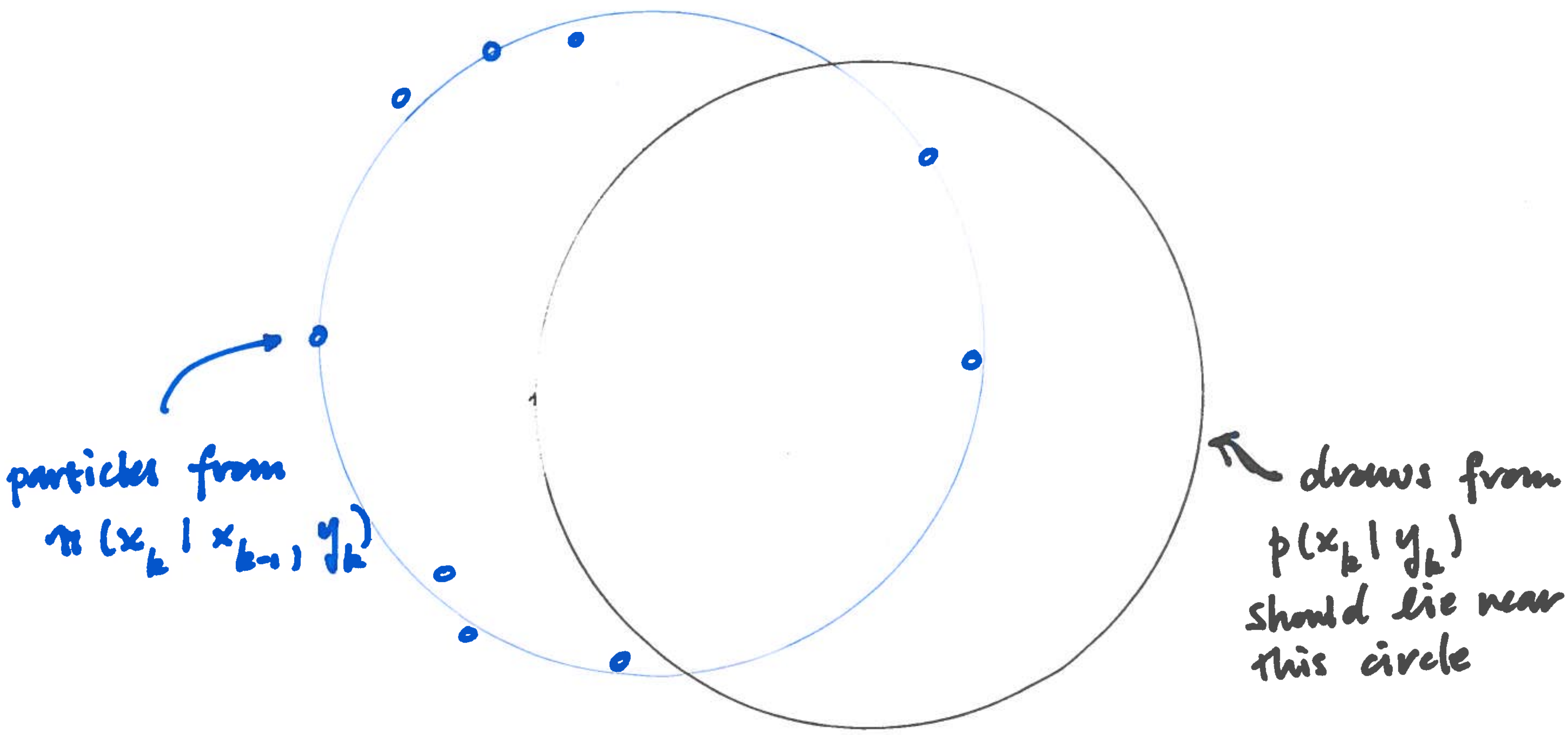
- Optimal proposal pulls new particles toward obs. Why doesn't this fix degeneracy?

Samples From High-Dimensional PDFs

- In high dimensions, sample is almost surely confined to small subset of total possible outcomes
- e.g., drawing from $\mathbf{x} \sim N(0, \mathbf{I})$, particles lie in "hypershell" of radius $\approx \sqrt{N_x}$ and thickness $1/\sqrt{2}$

Intuition for Weight Degeneracy

- high-dimensional intuition: overlap between proposal and desired posterior pdf is very small



Heuristics for Optimality

Begin with the identity

$$p(\mathbf{x}_{k-1}, \mathbf{x}_k | \mathbf{y}_k) = p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) p(\mathbf{x}_{k-1} | \mathbf{y}_k)$$

- allows sequential sampling: starting from $\mathbf{x}_{k-1}^i$, generate draw for $\mathbf{x}_k^i$
- $\mathbf{x}_{k-1}^i$ should be drawn from $p(\mathbf{x}_{k-1} | \mathbf{y}_k)$

Optimal proposal is given by

$$\pi(\mathbf{x}_{k-1}, \mathbf{x}_k | \mathbf{y}_k) = p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) p(\mathbf{x}_{k-1})$$

- improvements can only come from condition $\mathbf{x}_{k-1}^i$ on $\mathbf{y}_k$, i.e. improving the sample at t_{k-1} given new obs $\mathbf{y}_k$

τ^2 and $\dim(\mathbf{x})$

In simple examples, $\tau^2 \propto N_x,\ N_y$

General linear, Gaussian case:

$$\tau^2 = \sum_{j=1}^{N_y} \lambda_j^2\,(3\lambda_j^2/2+1),$$

where λ_j^2 are eigenvalues of

$$\mathbf{A} = \begin{cases} \mathrm{cov}\left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{x}_k\right), & \text{std. proposal} \\ \mathrm{cov}\left(\left(\mathbf{H}\mathbf{Q}\mathbf{H}^T+\mathbf{R}\right)^{-1/2}\mathbf{H}\mathbf{M}\mathbf{x}_{k-1}\right), & \text{opt. proposal.} \end{cases}$$

τ^2 and $\dim(\mathbf{x})$ (cont.)

τ^2 is an observation-space quantity

Large τ^2 follows from:

- many observations, N_y large
- relatively accurate observations
- relatively small system noise, for optimal proposal

NWP has many and accurate observations

- τ^2 is large and PFs require very large ensembles