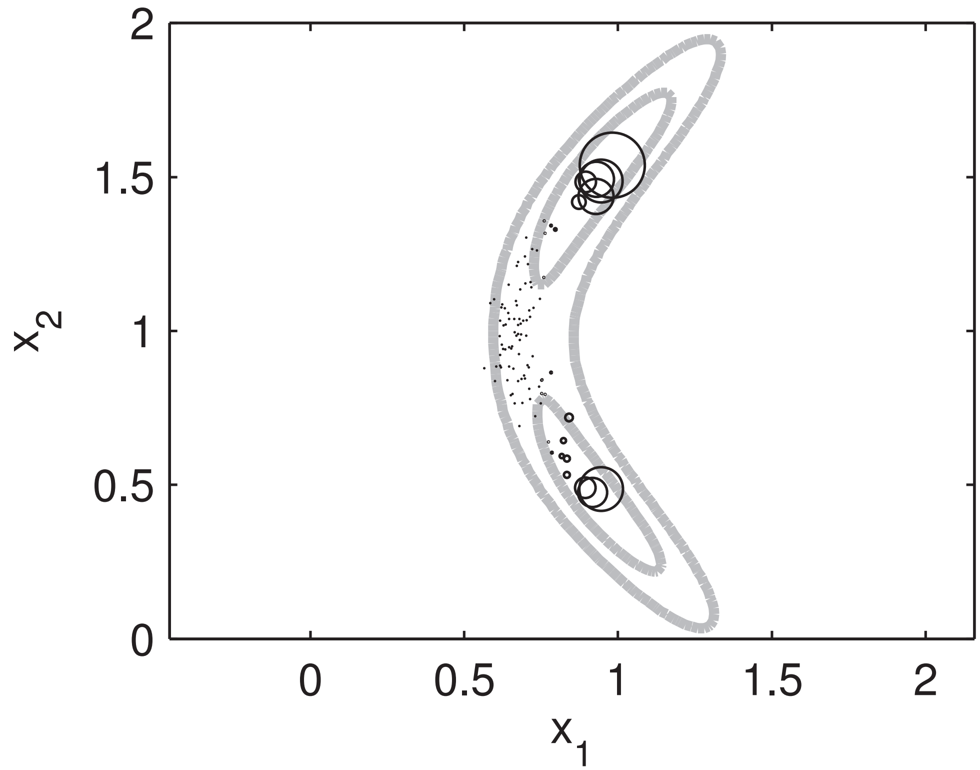
Performance Bounds for Particle Filters in High Dimensions



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

* NCAR is supported by the US National Science Foundation.

Preliminaries

Notation

- state evolution: $\mathbf{x}_k = M(\mathbf{x}_{k-1}) + \eta_k$, where $\mathbf{x}_k = \mathbf{x}(t_k)$
- observations: $\mathbf{y}_k = H(\mathbf{x}_k) + \epsilon_k$
- superscript i indexes ensemble members
- $\dim(\mathbf{x}) = N_x$, $\dim(\mathbf{y}) = N_y$, ensemble size $= N_e$

Interchangeable terms

- particles $\equiv$ ensemble members
- sample $\equiv$ ensemble

Preliminaries (cont.)

State $\mathbf{x}_k$ is a random variable

- goal is to estimate pdf $p(\mathbf{x}_k|\mathbf{y}^o)$ of this "true" state given obs $\mathbf{y}^o$

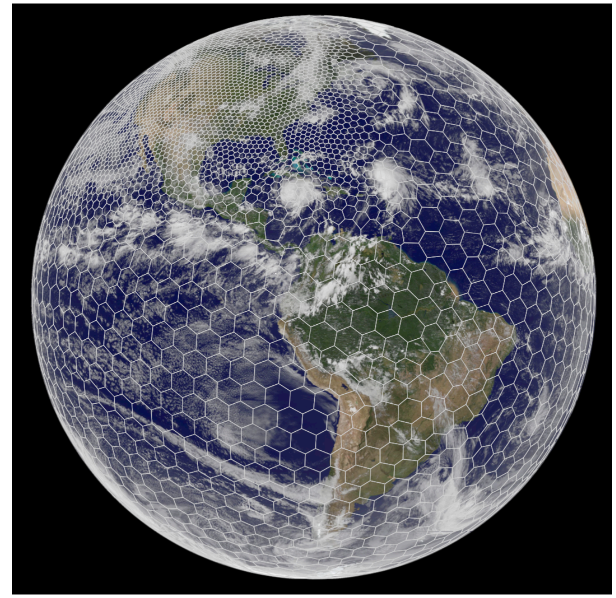
[In general, variables *without* superscripts are random.]

Bayes rule

- compute conditional pdf via

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

MPAS
Model for Prediction Across Scales



Jointly developed, primarily by NCAR and LANL/DOE

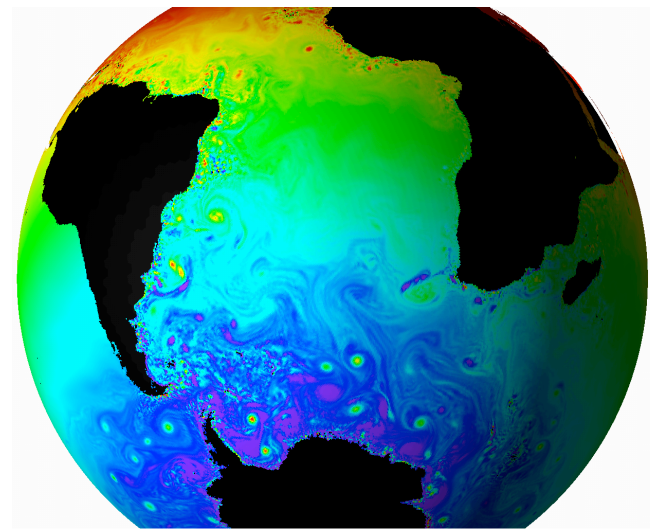
MPAS infrastructure - NCAR, LANL, others.

MPAS - Atmosphere (NCAR)

MPAS - Ocean (LANL)

MPAS - Ice, etc. (LANL and others)

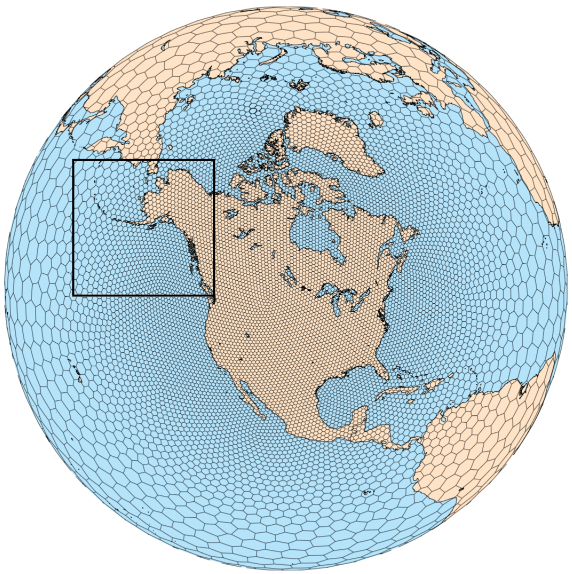
Project leads: Todd Ringler (LANL)
Bill Skamarock (NCAR)

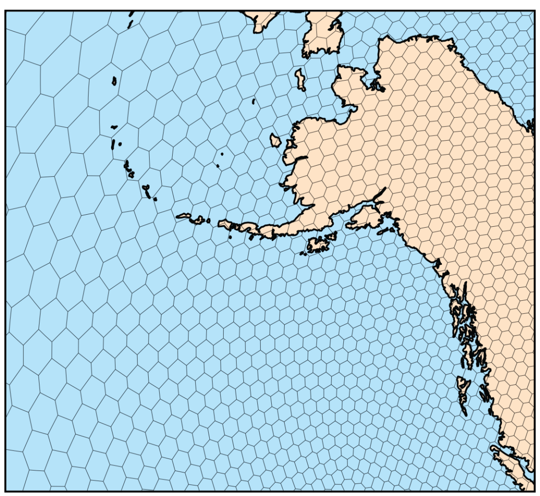


Los Alamos National Laboratory EST. 1943 · U.S. Department of Energy Office of Science · NSF · NESL

MPAS
Model for Prediction Across Scales

MPAS-Atmosphere





Unstructured spherical centroidal Voronoi meshes

Mostly *hexagons*, some pentagons and 7-sided cells.
Cell centers are at cell center-of-mass.
Lines connecting cell centers intersect cell edges at right angles.
Lines connecting cell centers are bisected by cell edge.
Mesh generation uses a density function.
Uniform resolution – traditional icosahedral mesh.

C-grid

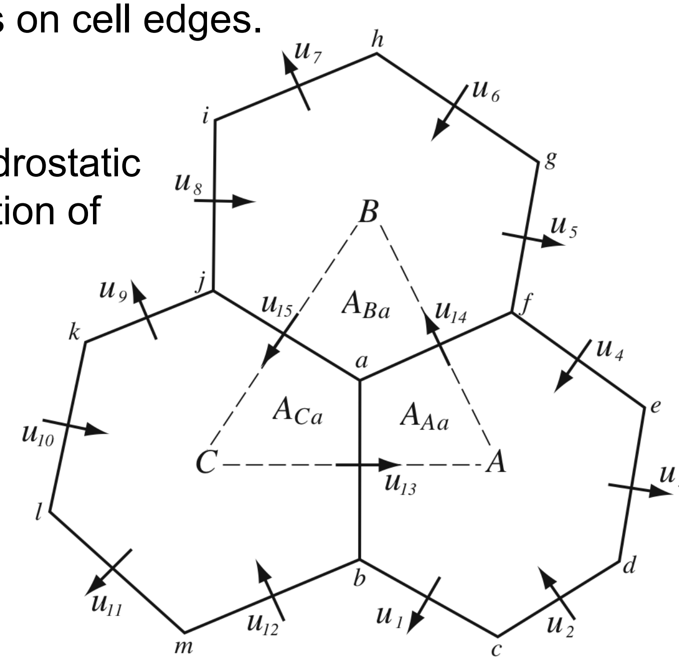
Solve for normal velocities on cell edges.

Solvers

Fully compressible nonhydrostatic equations (explicit simulation of clouds)

Solver Technology

Integration schemes are similar to WRF.

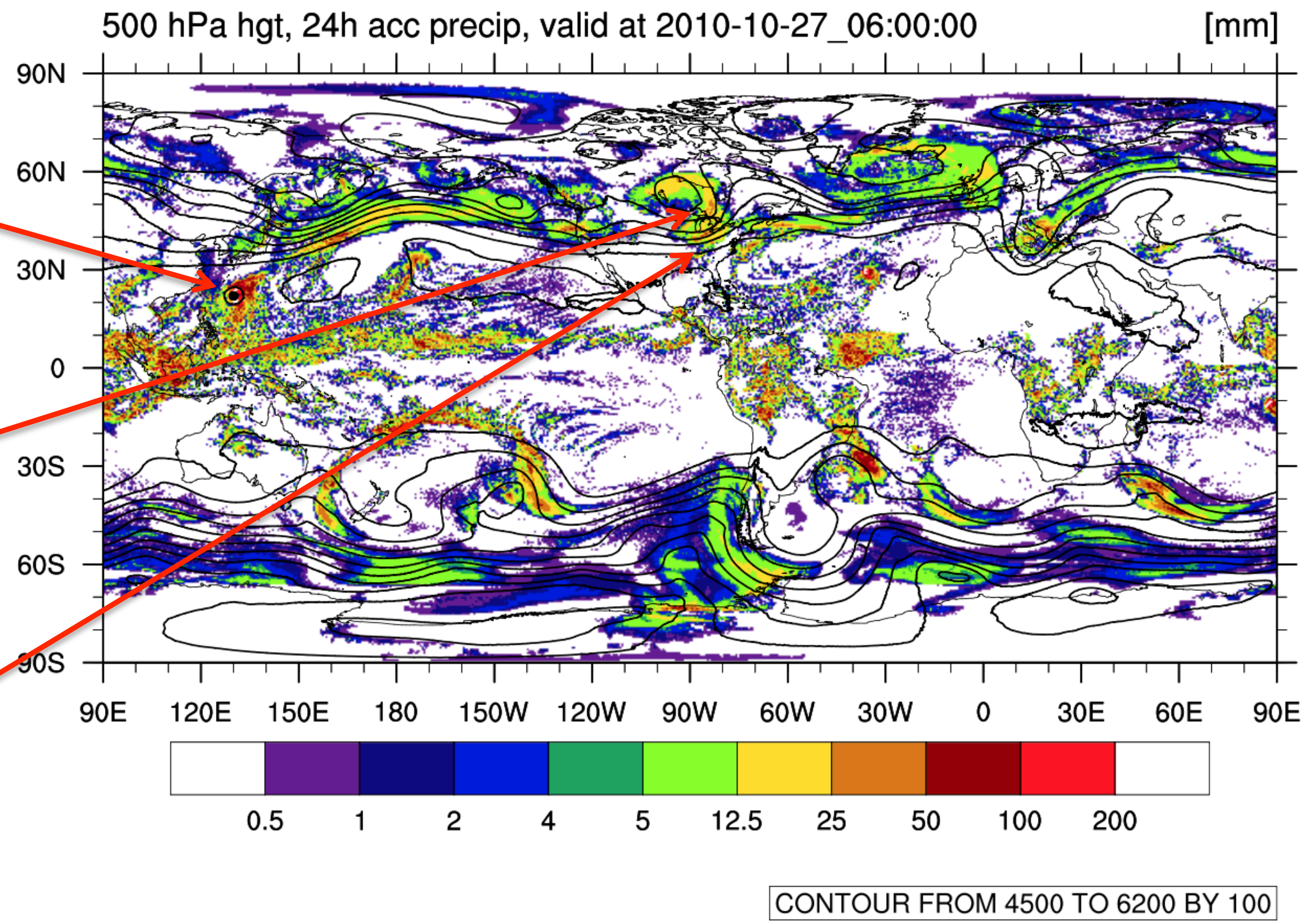


3-km Global MPAS-A Simluation

Tropical cyclone Chaba (cat 3)

Observed extratropical cyclone with the lowest recorded surface pressure in North America

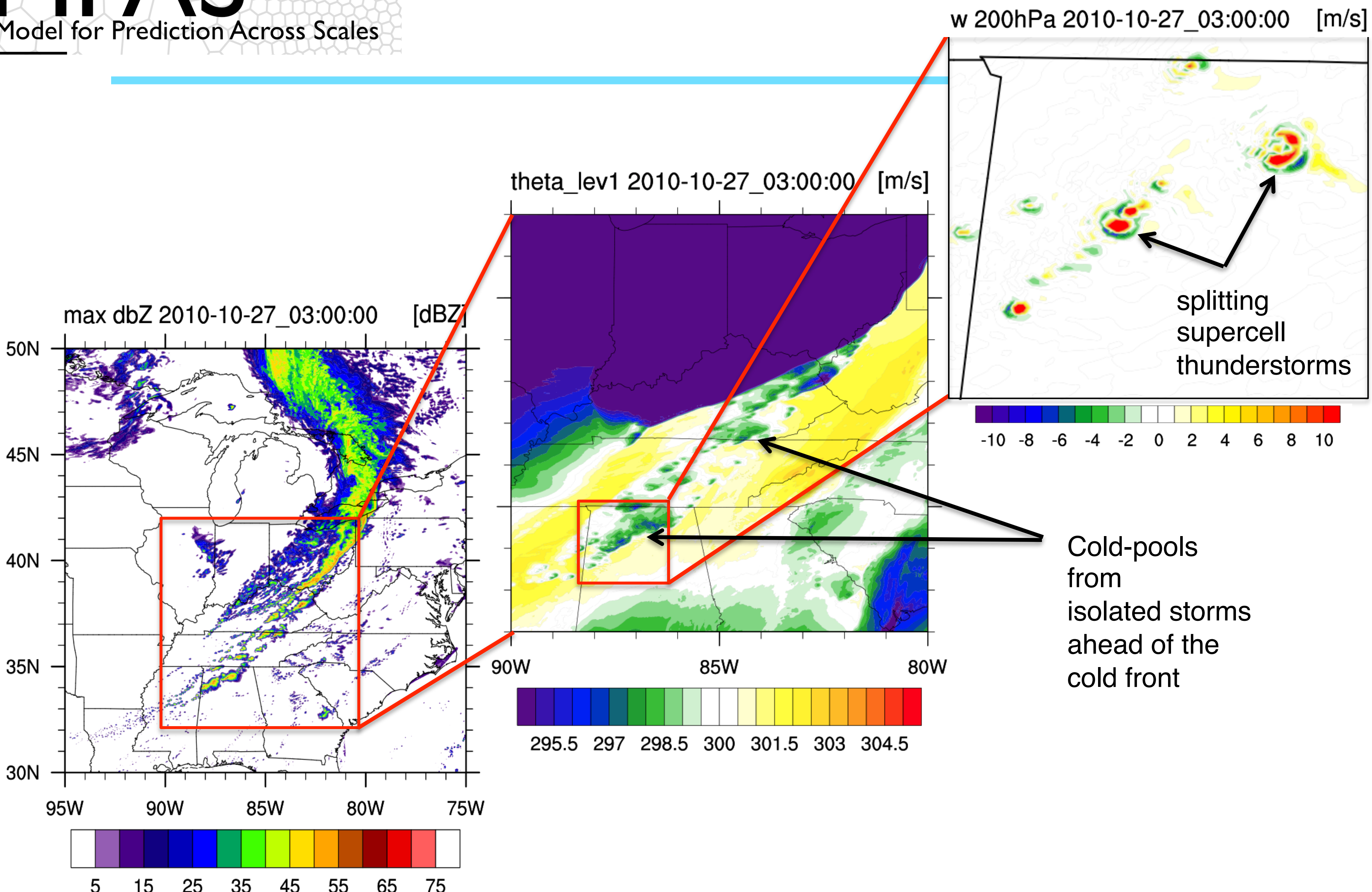
Isolated tornadic storms observed in warm sector ahead of the cold front



Courtesty of Bill Skamarock

MPAS

Model for Prediction Across Scales



Courtesty of Bill Skamarock

MPAS/DART

Data Assimilation Research Testbed (DART)

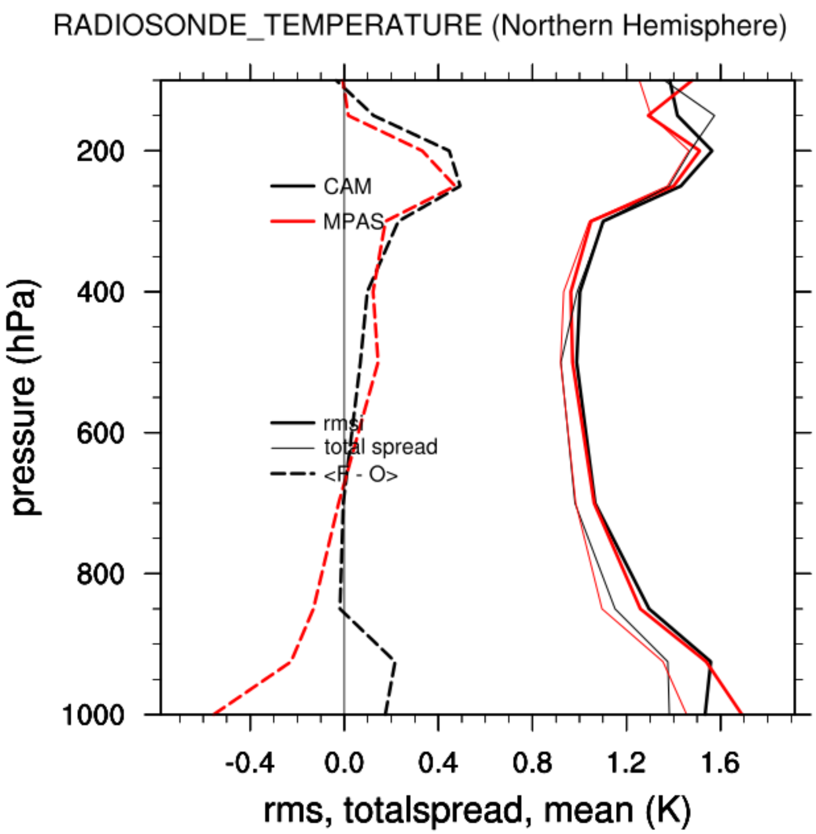
- Provides algorithm(s) for ensemble Kalman filter (EnKF)
- General framework, used for several models
- Parallelizes efficiently to 100's of processors
- Developed by Jeff Anderson and team; see http://www.image.ucar.edu/DAReS/DART/

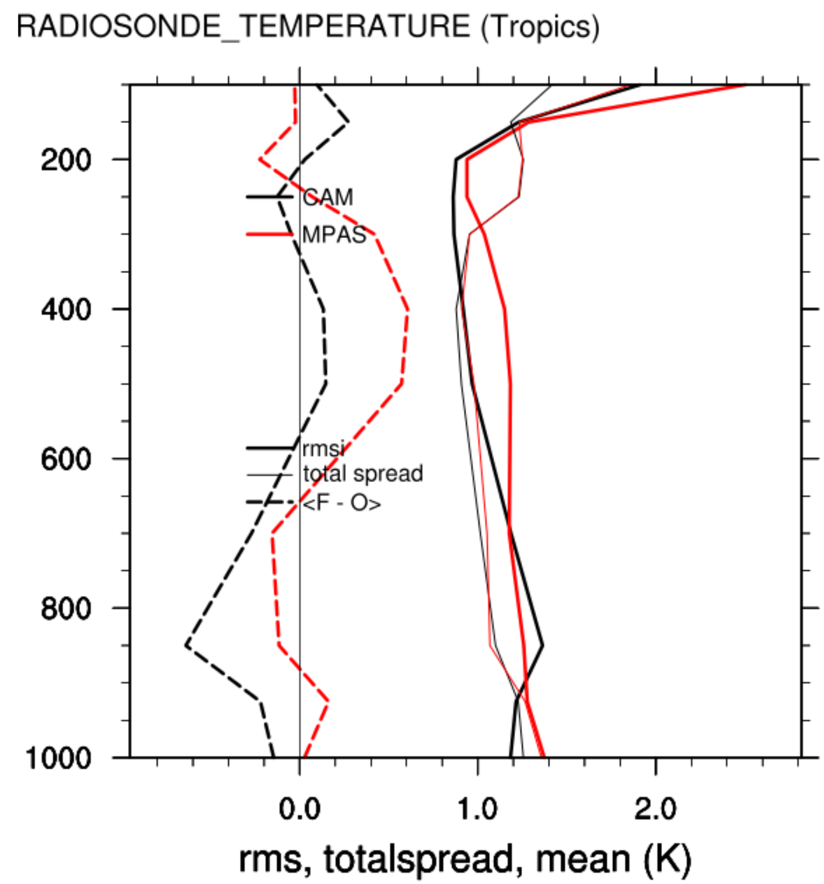
MPAS/DART

- MPAS-specific interfaces + obs operators (conventional, GPS)
- Month-long experiments with 6-hourly cycling are stable, with results comparable to those from Community Atmosphere Model (CAM 4)/DART

Comparison with CAM/DART

- August 2008, 6-h cycling, conventional obs + GPS
- 120-km MPAS, 1-deg CAM FV



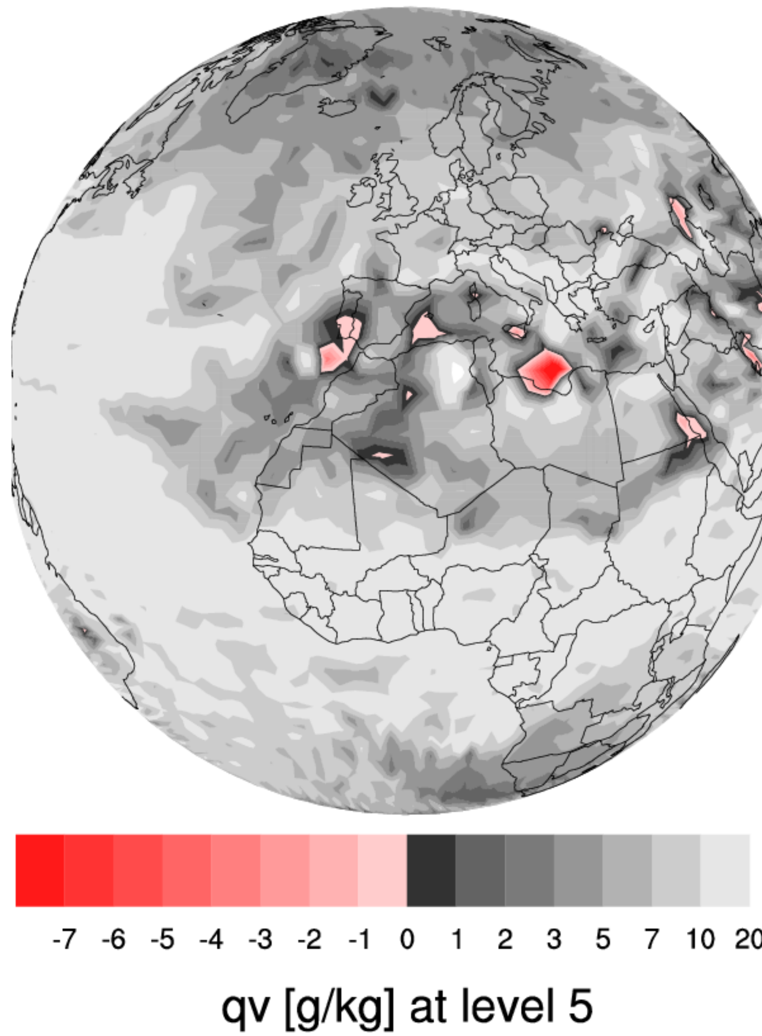


Courtesty of Soyoung Ha

MPAS/DART Moisture Analysis

Specific humidity,
12Z 6 Aug 2008,
member 1

Negative values!



Courtesty of Soyoung Ha

EnKF and Positive-Definite Variables

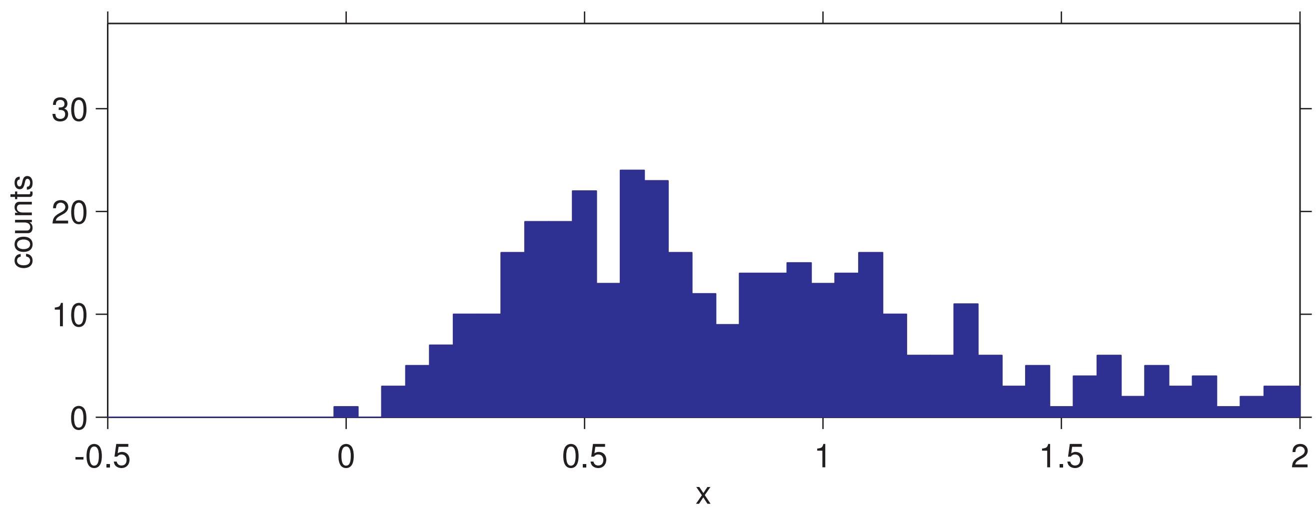
KF (and EnKF) consider only mean and covariance of $\mathbf{x}_k$

- linear updates for $\bar{\mathbf{x}}_k = E(\mathbf{x}_k)$ and $\mathbf{P}_k = \mathrm{cov}(\mathbf{x}_k)$
- implements Bayes rule when $p(\mathbf{x}_k)$ and $p(\mathbf{y}_k^o|\mathbf{x}_k)$ are Gaussian

Positive-definite variables are not Gaussian

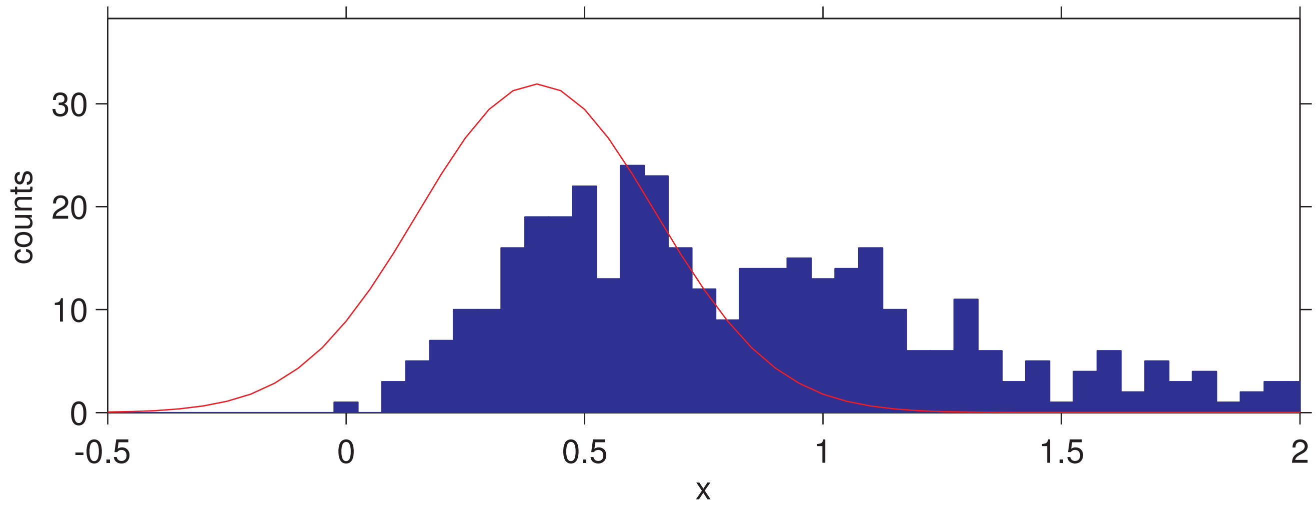
EnKF and +ive Variables (cont.)

- One-dimensional example: sample from $p(x_k)$



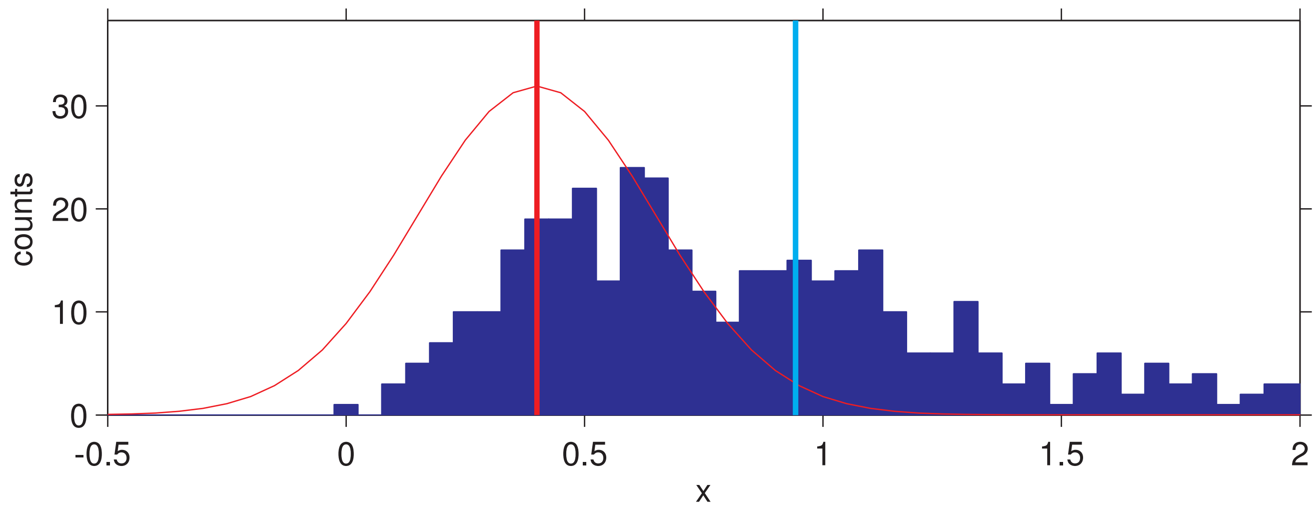
EnKF and +ive Variables (cont.)

- $p(y^o|x_k)$ and Gaussian obs error



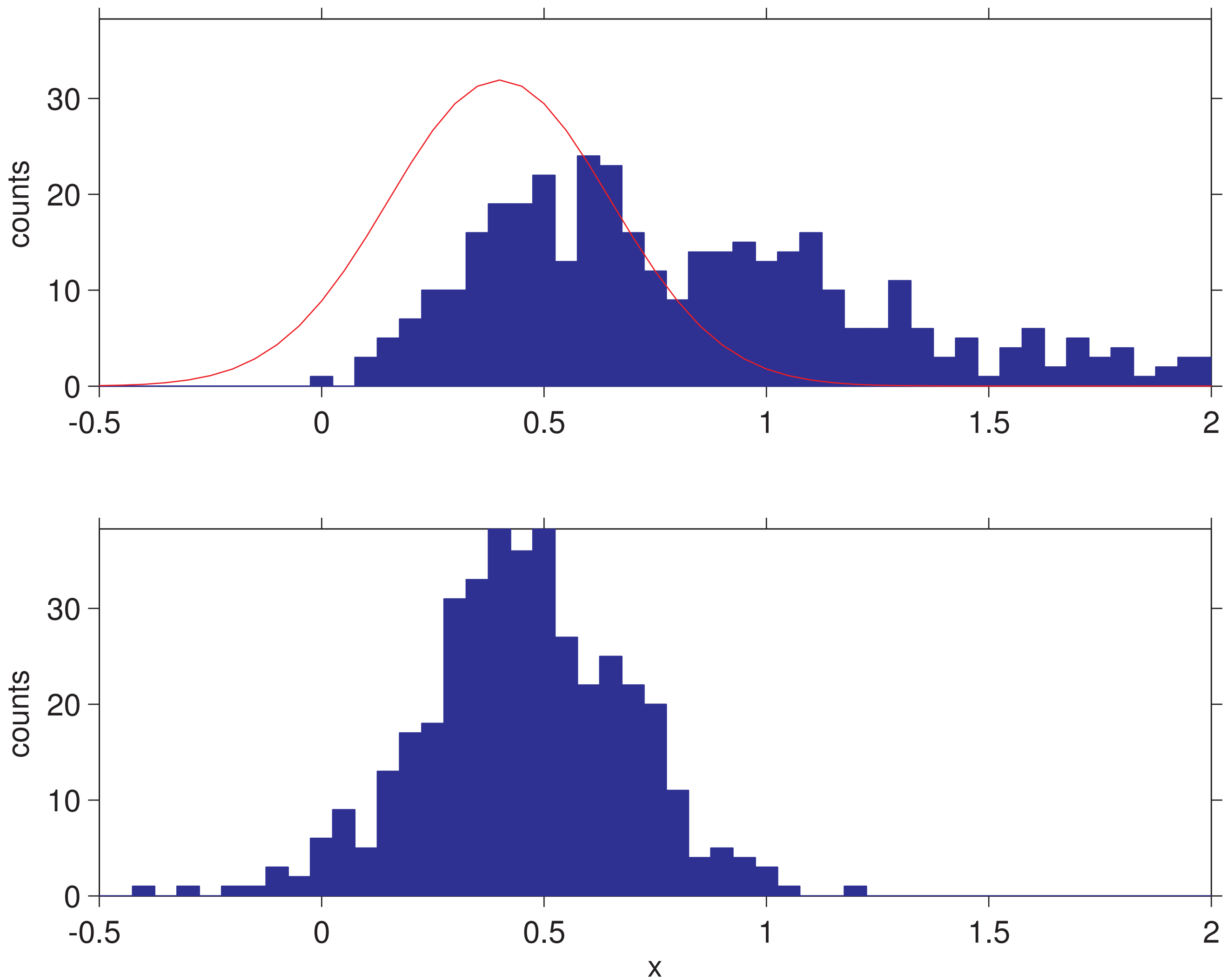
EnKF and +ive Variables (cont.)

- ▷ prior mean and obs value ($y^o = 0.4$)



EnKF and +ive Variables (cont.)

- updated sample produced by EnKF includes some $x^i < 0$



Part II

Particle filters offer potential solution for non-Gaussian DA

Part II: Overview

- ▷ Simplest particle filter requires very large ensemble size, growing exponentially with the problem size.
- ▷ Can the use of the optimal proposal density fix this?
- ▷ What exactly is the "problem size?"

Background I: Particle Filters (PFs)

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

- works with samples from desired pdf, rather than pdf itself
- 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)

PFs (cont.)

Elementary particle filter:

- begin with members $\mathbf{x}^i_{k-1}$ drawn from $p(\mathbf{x}_{k-1}|\mathbf{y}^o_{k-1})$
- begin with members $\mathbf{x}^i_{k-1}$ and weights w^i_{k-1} that "represent" $p(\mathbf{x}_{k-1}|\mathbf{y}^o_{k-1})$
- compute $\mathbf{x}^i_k$ by evolving each member to t_k under the system dynamics
- re-weight, given new obs $\mathbf{y}^o_k$: $w^i_k \propto w^i_{k-1} p(\mathbf{y}^o_k|\mathbf{x}^i_k)$
- resample

Background II: Importance Sampling

Basic idea

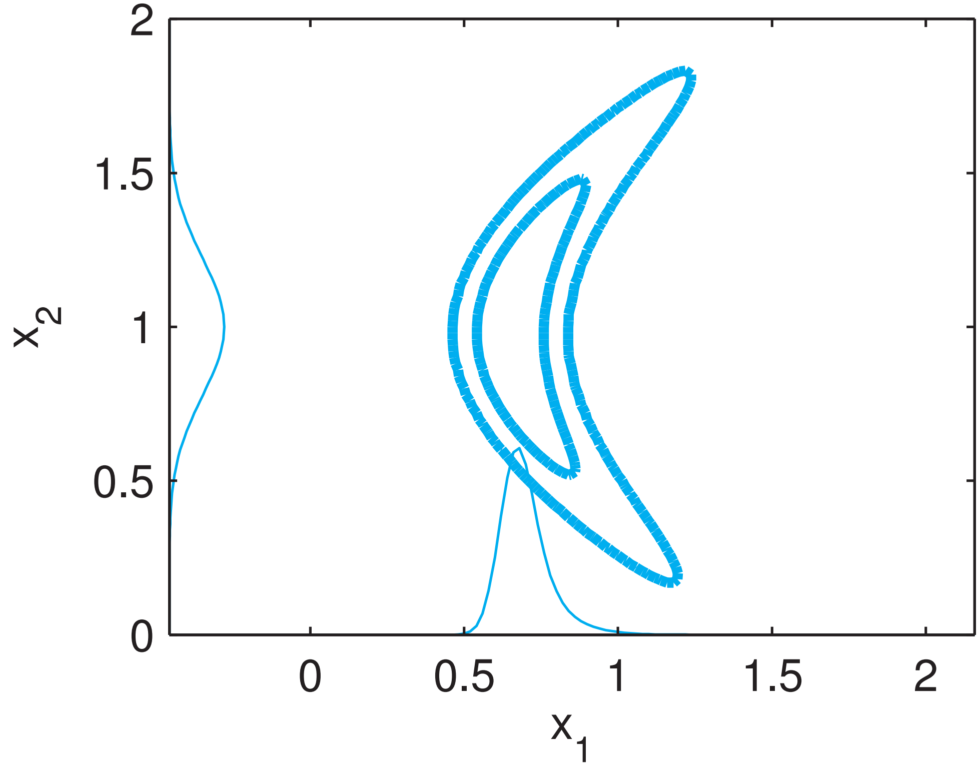
- suppose $p(\mathbf{x})$ is hard to sample from, but $\pi(\mathbf{x})$ is not.
- draw $\{\mathbf{x}^i\}$ from $\pi(\mathbf{x})$ and approximate

$$p(\mathbf{x}) \approx \sum_{i=1}^{N_e} w^i \delta(\mathbf{x} - \mathbf{x}^i), \quad \text{where } w^i \propto p(\mathbf{x}^i)/\pi(\mathbf{x}^i)$$

- $\pi(\mathbf{x})$ is the *proposal density*

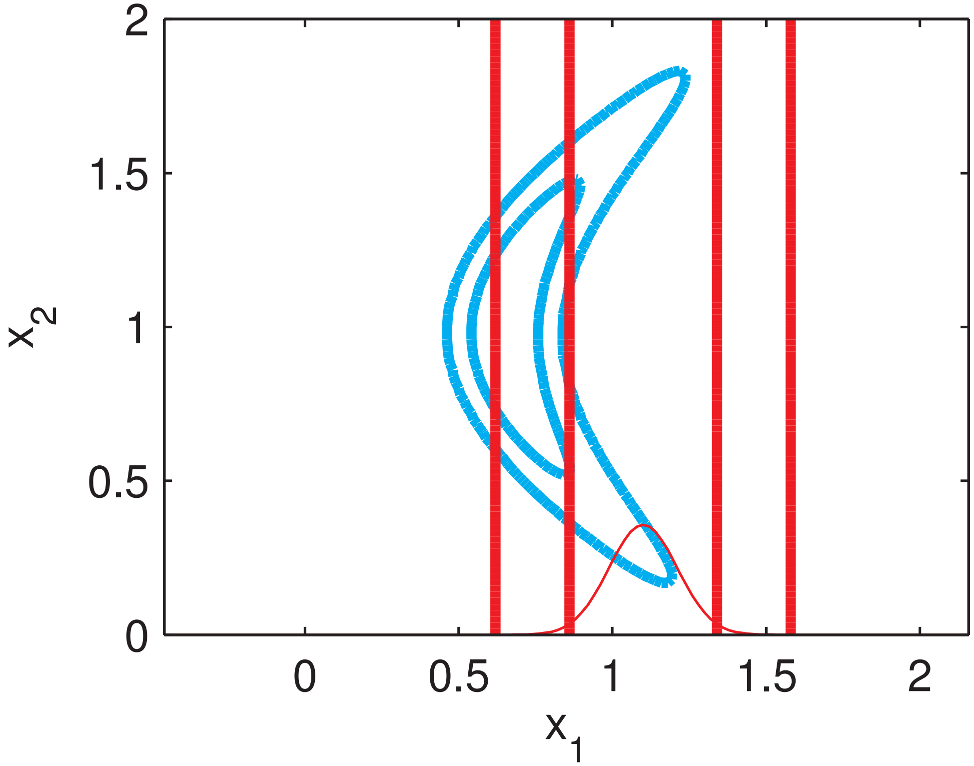
IS Example

- ▷ $p(x_1, x_2)$ for 2D state (x_1, x_2); thin lines indicate marginal pdfs



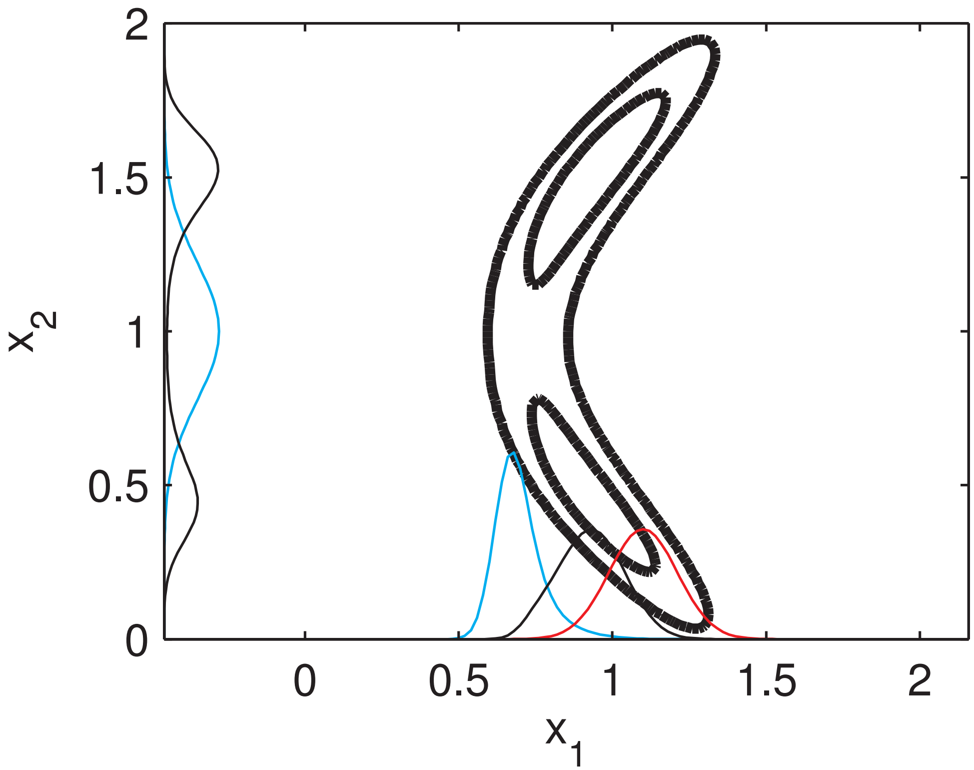
IS Example (cont.)

- observation $y = x_1 + \epsilon$, with realization $y^o = 1.1$
- $p(y^o|x_1, x_2)$ does not depend on x_2



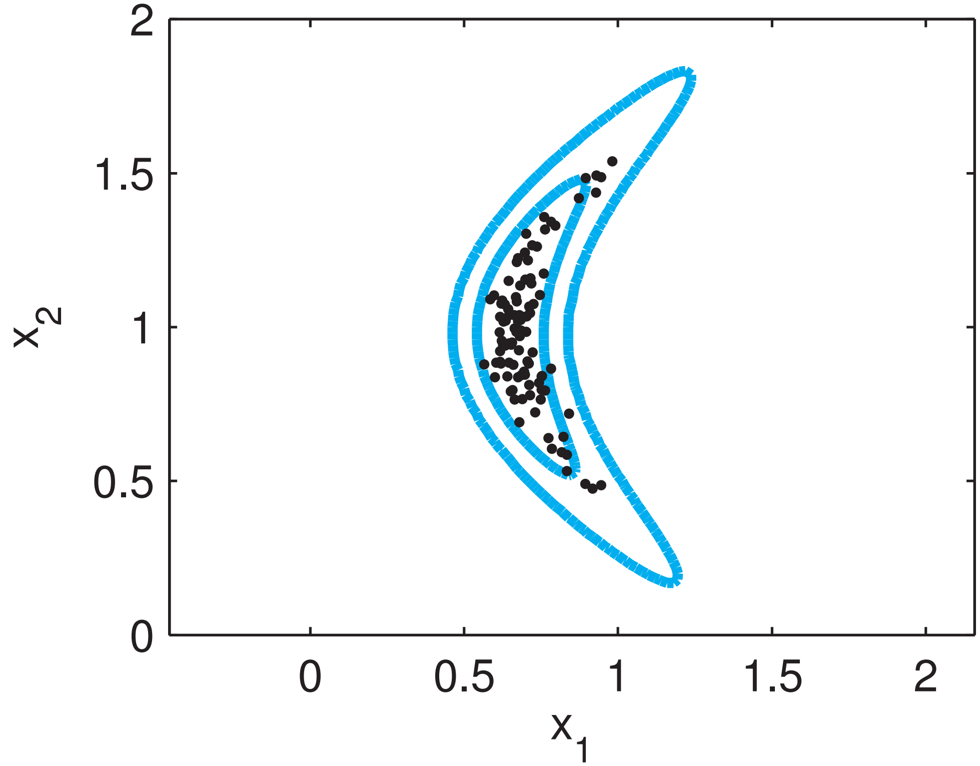
IS Example (cont.)

- $p(x_1, x_2 | \, y^o)$



IS Example (cont.)

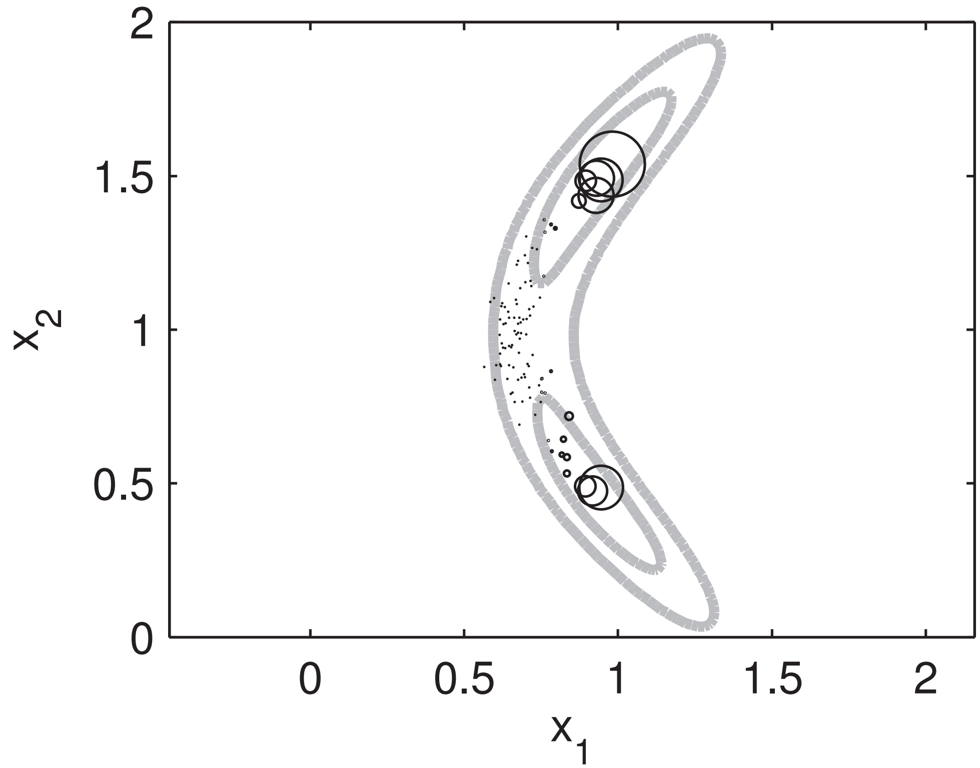
- $\pi(\mathbf{x}) = p(\mathbf{x})$,and sample from $\pi(\mathbf{x})$



- Want to sample from $p(\mathbf{x}|\mathbf{y})$
- IS says we should weight sample from $\pi(\mathbf{x}) = p(\mathbf{x})$ by $p(\mathbf{x}|\mathbf{y})/\pi(\mathbf{x}) = p(\mathbf{y}|\mathbf{x})$

IS Example (cont.)

▷ $p(\mathbf{x}|\mathbf{y})$ and "weighted" ensemble (size $\propto$ weight)



Sequential Importance Sampling

Perform importance sampling sequentially in time

- Given $\{\mathbf{x}_{k-1}^i\}$ from $\pi(\mathbf{x}_{k-1})$, wish to sample from $p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_k^o)$
- choose proposal of the form

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

- Using $p(\mathbf{x}_k^i, \mathbf{x}_{k-1}^i|\mathbf{y}_k^o) \propto p(\mathbf{y}_k^o|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)p(\mathbf{x}_{k-1}^i)$, new weights are

$$w_k^i \propto \frac{p(\mathbf{x}_k^i, \mathbf{x}_{k-1}^i|\mathbf{y}_k^o)}{\pi(\mathbf{x}_k^i, \mathbf{x}_{k-1}^i|\mathbf{y}_k^o)} = \frac{p(\mathbf{y}_k^o|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_k^o)}\, w_{k-1}^i$$

Sequential IS (cont.)

PF literature shows that choice of proposal is crucial

Standard proposal: transition density from dynamics

- $\pi(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_k^o) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$
- $w_k^i \propto p(\mathbf{y}_k^o|\mathbf{x}_k^i)\, w_{k-1}^i$
- members at t_k generated by evolution under system dynamics, as in ensemble forecasting

Sequential IS (cont.)

"Optimal" proposal: Also condition on most recent obs

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

  $$w_k^i \propto p(\mathbf{y}_k^o|\mathbf{x}_{k-1}^i)\, w_{k-1}^i$$
- optimal in sense that it minimizes variance of weights over $\mathbf{x}_k^i$
- several recent PF studies use proposals that either reduce to or are related to the optimal proposal (van Leeuwen 2010, Morzfeld et al. 2011, Papadakis et al. 2010)
- *not* an ensemble forecast; generating members at t_k resembles DA

Degeneracy of PF Weights

- degeneracy $\equiv \max_i w_k^i \to 1$
- common problem, well known in PF literature
- for standard proposal, Bengtsson et al. (2008) and Snyder et al. (2008) show N_e must increase exponentially as problem size increases in order to avoid degeneracy
- What happens with optimal proposal?

A Simple Test Problem

Consider the system

$$\mathbf{x}_k = a\mathbf{x}_{k-1} + \eta_{k-1}, \quad \mathbf{y}_k = \mathbf{x}_k + \epsilon_k$$

where $\mathbf{x}_{k-1} \sim N(0, \mathbf{I})$, $\eta_{k-1} \sim N(0, q^2\mathbf{I})$ and $\epsilon_k \sim N(0, \mathbf{I})$.

A Simple Test Problem

Consider the system

$$\mathbf{x}_k = a\mathbf{x}_{k-1} + \eta_{k-1}, \quad \mathbf{y}_k = \mathbf{x}_k + \epsilon_k$$

where $\mathbf{x}_{k-1} \sim N(0, \mathbf{I})$, $\eta_{k-1} \sim N(0, q^2\mathbf{I})$ and $\epsilon_k \sim N(0, \mathbf{I})$.

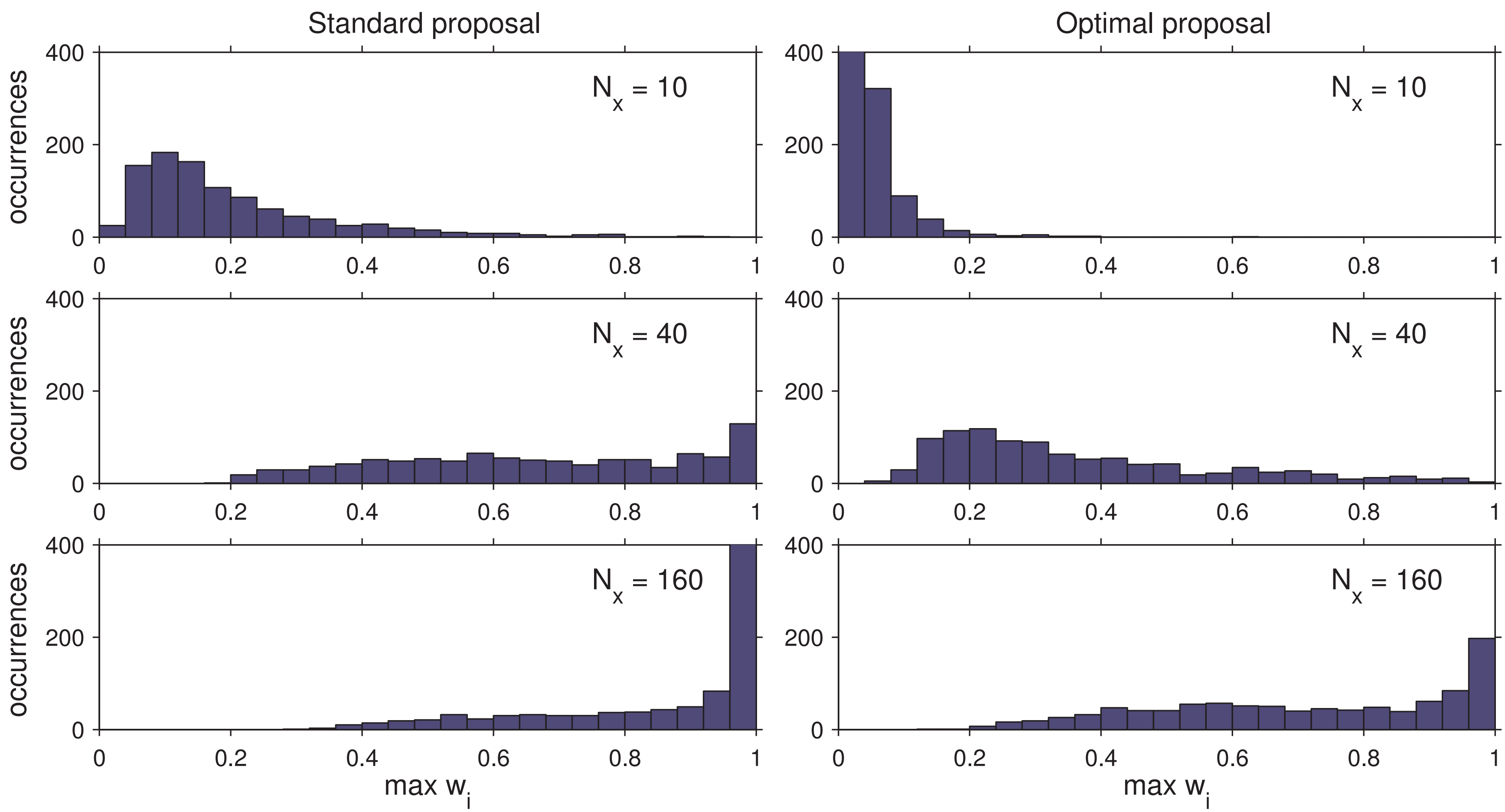
Then

$$\mathbf{y}_k|\mathbf{x}_k^i \sim N(\mathbf{x}_k^i, \mathbf{I}), \quad \mathbf{y}_k|\mathbf{x}_{k-1}^i \sim N\left(a\mathbf{x}_{k-1}^i, \left(1+q^2\right)\mathbf{I}\right).$$

Easy to calculate $w_k^i \propto p(\mathbf{y}_k^o|\mathbf{x}_k^i)$ (standard proposal) or $w_k^i \propto p(\mathbf{y}_k^o|\mathbf{x}_{k-1}^i)$ (optimal proposal).

A Simple Test Problem (cont.)

- histograms of $\max_i w_k^i$ for $N_e = 10^3$, $a = q = 1/2$. 10^3 simulations.
- degeneracy occurs, but optimal proposal clearly reduces it at any N_x



Behavior of Weights

Define

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k^o) = -\log(w_k/w_{k-1}) = \begin{cases} -\log p(\mathbf{y}_k^o|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k^o|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

and let $\tau^2 = \text{var}(V)$.

Behavior of Weights

Define

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k^o) = -\log(w_k/w_{k-1}) = \begin{cases} -\log p(\mathbf{y}_k^o|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k^o|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

and let $\tau^2 = \text{var}(V)$.

Then for large N_e and large τ,

$$E(1/\max w_k^i) \sim 1 + \frac{\sqrt{2\log N_e}}{\tau}$$

(Bengtsson et al. 2008, Snyder et al. 2008)

Behavior of Weights

Define

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k^o) = -\log(w_k/w_{k-1}) = \begin{cases} -\log p(\mathbf{y}_k^o|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k^o|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

and let $\tau^2 = \text{var}(V)$.

Then for large N_e and large τ,

$$E(1/\max w_k^i) \sim 1 + \frac{\sqrt{2\log N_e}}{\tau}$$

(Bengtsson et al. 2008, Snyder et al. 2008)

As τ^2 increases, N_e must increase as $\exp(2\tau^2)$ to keep $E(1/\max w^i)$ fixed.

The Linear, Gaussian Case

Analytic results possible for linear, Gaussian case with general $\mathbf{R} = \mathrm{cov}(\epsilon_k)$, $\mathbf{Q} = \mathrm{cov}(\eta_k)$ and $\mathbf{P}_k = \mathrm{cov}(\mathbf{x}_k)$.

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

where λ_j^2 are eigenvalues of

$$\mathbf{A} = \begin{cases} \mathbf{R}^{-1/2}\mathbf{H}(\mathbf{M}\mathbf{P}_{k-1}\mathbf{M}^T + \mathbf{Q})\mathbf{H}^T\mathbf{R}^{-1/2}, & \text{std. proposal} \\ (\mathbf{HQH}^T + \mathbf{R})^{-1/2}\mathbf{HMP}_{k-1}(\mathbf{HM})^T(\mathbf{HQH}^T + \mathbf{R})^{-1/2}, & \text{opt. proposal.} \end{cases}$$

- τ^2(opt. proposal) always less than or equal to τ^2(std. proposal), with equality only when $\mathbf{Q} = 0$.

Simple Test Problem, Revisited

Recall

$$\mathbf{x}_k = a\mathbf{x}_{k-1} + \eta_{k-1}, \quad \mathbf{y}_k = \mathbf{x}_k + \epsilon_k$$

where $\mathbf{x}_{k-1} \sim N(0, \mathbf{I})$, $\eta_{k-1} \sim N(0, q^2\mathbf{I})$ and $\epsilon_k \sim N(0, \mathbf{I})$.

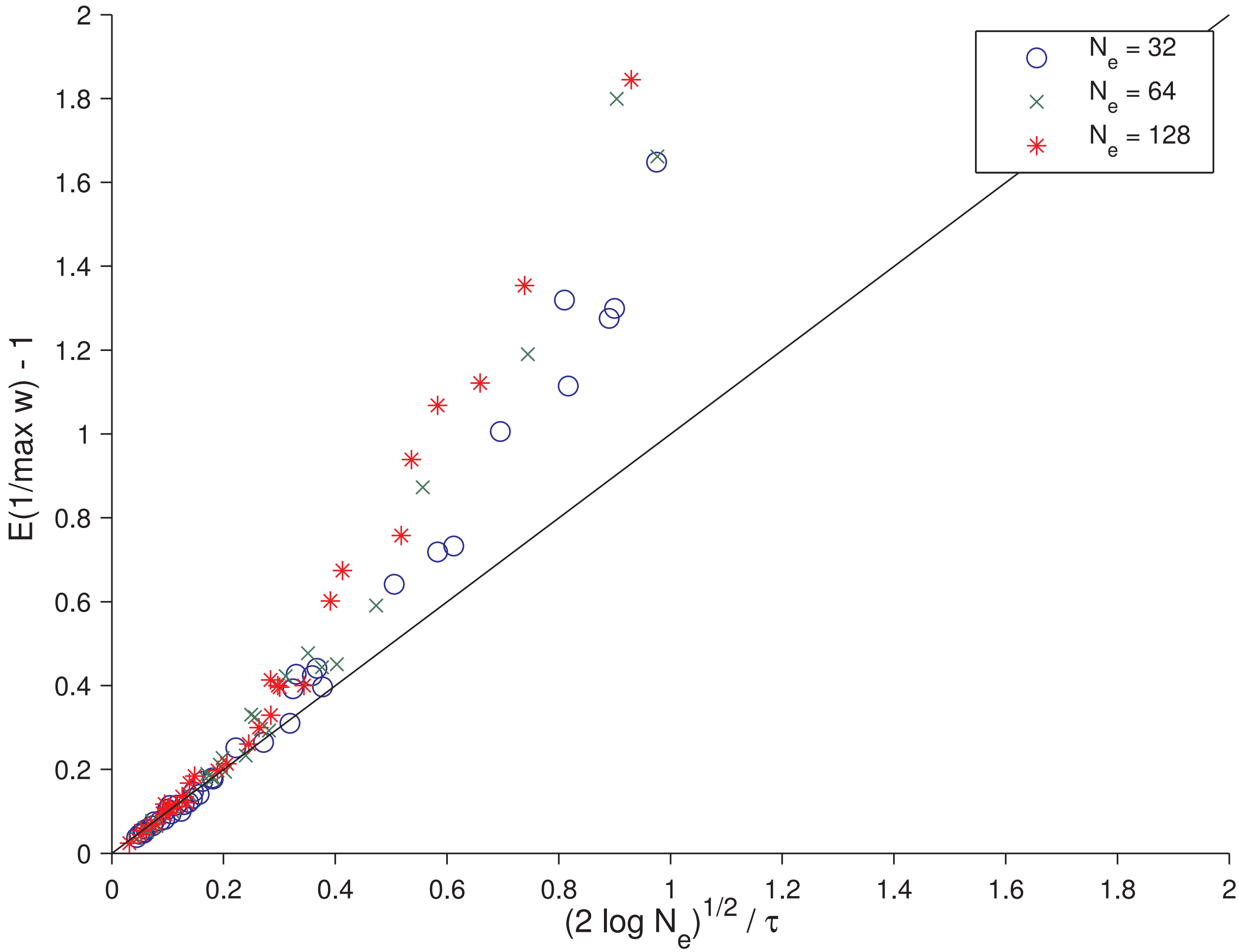
Then

$$\tau^2 = \text{var}(V) = \begin{cases} N_y(a^2 + q^2)\left(\frac{3}{2}a^2 + \frac{3}{2}q^2 + 1\right), & \text{std. proposal} \\ N_y a^2 \left(\frac{3}{2}a^2 + q^2 + 1\right)/(q^2+1)^2, & \text{opt. proposal} \end{cases}$$

- ▷ opt. proposal reduces τ^2 by an $O(1)$ factor for reasonable values of a and q; $a^2 = q^2 = 1/2$ implies a factor of 5 reduction in τ^2.

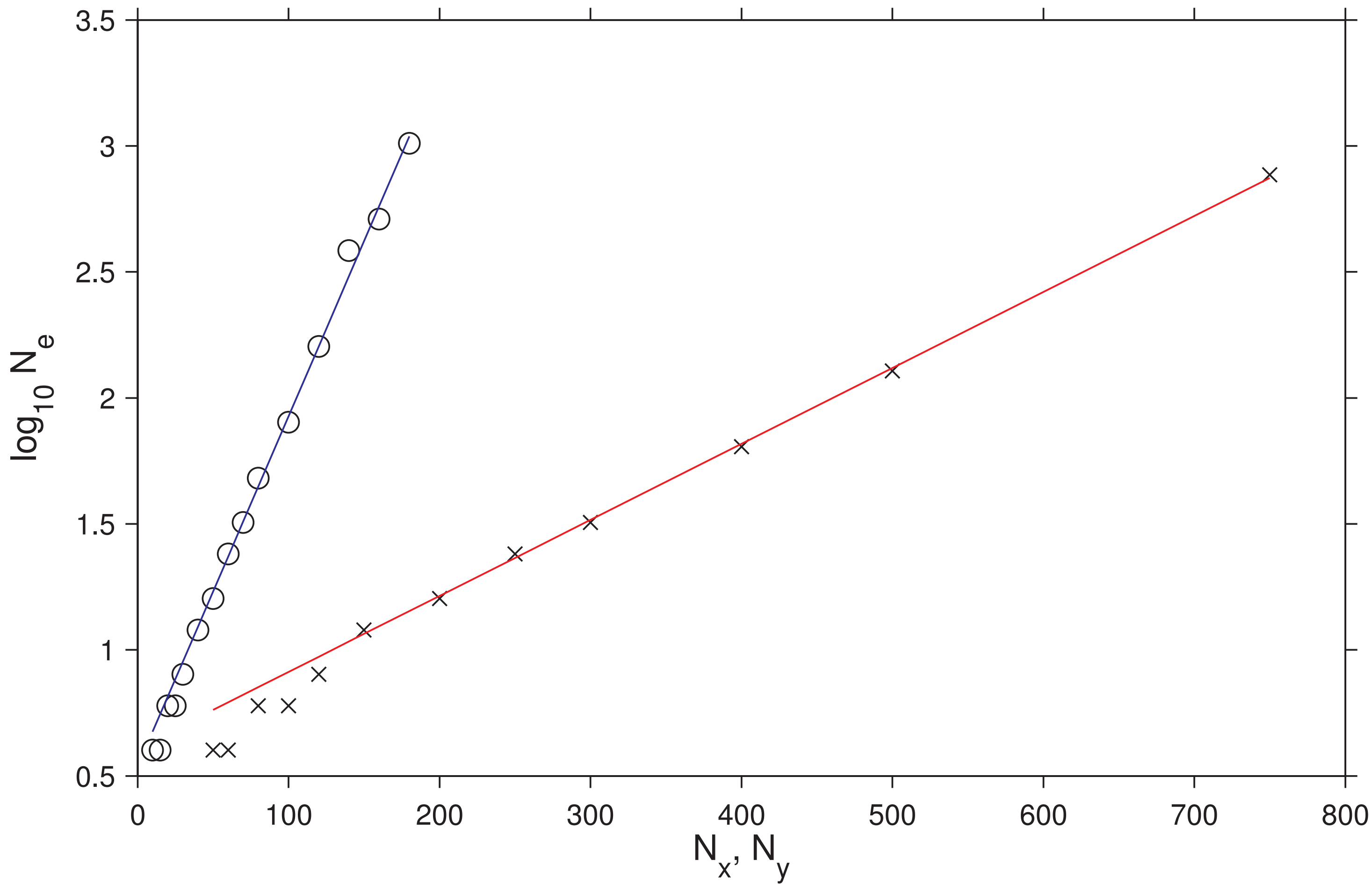
Simple Test Problem, Revisited (cont.)

- Theoretical prediction for $E(1/\max w^i)$ vs. simulations. Expectation is based on 10^3 realizations.



Simple Test Problem, Revisited (cont.)

- minimum N_e such that $E(1/\max w^i) \geq 1/0.8$ for standard proposal (circles) and optimal proposal (crosses) for $a^2 = q^2 = 1/2$.
- ratio of slopes of best-fit lines is 4.6, vs. asymptotic prediction of 5



N_y, N_x and Problem Size

$\tau^2 = \text{var(log likelihood)}$ measures "problem size" for PF

- as τ^2 increases, N_e must increase as $\exp(2\tau^2)$ if $E(1/\max w^i)$ fixed.

Related to obs-space dimension

- in simple example, $\tau^2 \propto N_y$
- given by sum over e-values of obs-space covariance in general linear, Gaussian case—like an effective dimension

Analogy of τ^2 to dimension is incomplete

- τ^2 depends on obs-error statistics, increasing as **R** decreases
- τ^2 depends on proposal

N_y, N_x and Problem Size (cont.)

τ^2 depends explicitly *only* on obs-space quantities

How does N_x affect weight degeneracy?

- asymptotic relation of τ^2 and $E(1/\max w^i)$ requires $V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k)$ to be $\sim$ Gaussian over $\mathbf{x}_k$
- $\sim$ Gaussianity of $V(\mathbf{x}_k)$ only if $N_x = \dim(\mathbf{x})$ is large and components of $\mathbf{x}$ are sufficiently independent

Summary

- As was the case for the standard proposal, the optimal proposal requires N_e to increase exponentially with the "problem size" to avoid degeneracy.
- Exponential rate of increase is quantitatively smaller for the optimal proposal; necessary ensemble size may therefore be *much* smaller in a given problem. Using optimal proposal, PF feasible for problems with τ^2 as large as a few hundred.
- No free lunch: Benefits of optimal proposal dependent on magnitude and form of system noise.

Other Potential Tricks

- Equivalent-weights particle filter (van Leeuwen 2010)
- Use proposals that consider state and obs over a window $[t_{k-L+1}, t_{k-L+2}, \ldots, t_k]$ (Doucet, Briers and Sénécal 2006)
- Consider sequences of proposals, where consecutive pdfs in the sequence are similar/close (Beskos, Crisan and Jasra 2012)
- Spatial localization, in which individual observations influence update only locally (Bengtsson et al. 2003, Lei and Bickel 2011)

Recommendation

New PF algorithms intended for high-dimensional systems should be evaluated first on the simple test problem given here.

References

Bengtsson, T., P. Bickel and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Collections*, **2**, 316–334. doi: 10.1214/193940307000000518.

Morzfeld, M., X. Tu, E. Atkins and A. J. Chorin, 2011: A random map implementation of implicit filters. *J. Comput. Phys.* **231**, 2049–2066.

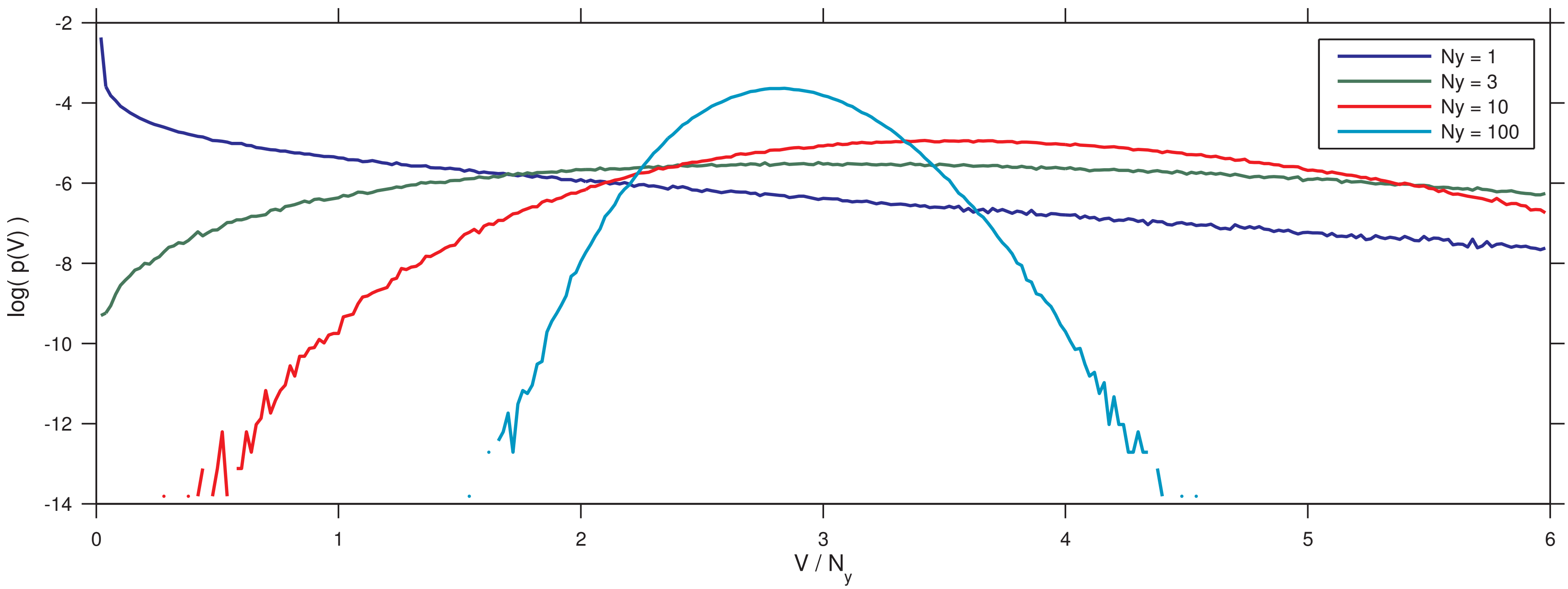
van Leeuwen, P. J., 2010: Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quart. J. Roy. Meteor. Soc.* **136**, 1991–1999.

Papadakis, N., E. Mémin, A. Cuzol and N. Gengembre, 2010: Data assimilation with the weighted ensemble Kalman filter. *Tellus* **62A**, 673–697.

Snyder, C., T. Bengtsson, P. Bickel and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Monthly Wea. Rev.*, **136**, 4629–4640.

N_y, N_x and Problem Size (cont.)

- $\log(p(V))$: $N_y = 1, 3, 10, 100$; $\mathbf{x} \sim N(0, \mathbf{I})$, $\mathbf{H} = \mathbf{I}$, $\epsilon \sim N(0, \mathbf{I})$ (standard proposal, $a^2 + q^2 = 1$)
- recall that max weight depends on left-hand tail of $p(V)$
- as N_y (and N_x) increase, left-hand tail changes and $V \to$ Gaussian (i.e. $\log(p(V))$ approaches a parabola)



Behavior of Weights

Define

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k^o) = -\log(w_k/w_{k-1}) = \begin{cases} -\log p(\mathbf{y}_k^o|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k^o|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

Interested in V as random variable with $\mathbf{y}_k^o$ known and $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ distributed according to the proposal distribution at t_k and t_{k-1}, respectively.

Behavior of Weights

Define

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k^o) = -\log(w_k/w_{k-1}) = \begin{cases} -\log p(\mathbf{y}_k^o|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k^o|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

Interested in V as random variable with $\mathbf{y}_k^o$ known and $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ distributed according to the proposal distribution at t_k and t_{k-1}, respectively.

Suppose each component of obs error is independent.

- $p(\mathbf{y}_k^o|\mathbf{x}_k)$, $p(\mathbf{y}_k^o|\mathbf{x}_{k-1})$ can be written as products over likelihoods for each component $y_{j,k}^o$ of $\mathbf{y}_k^o$
- V becomes a sum over log likelihoods for each component
- if terms in sum are nearly independent, $V \to$ Gaussian as $N_y \to \infty$
- infer asymptotic behavior of $\max w_k^i$ from known asymptotics for sample min of Gaussian