NAVAL
POSTGRADUATE
SCHOOL

Data Assimilation for High-Dimensional Systems

— Challenges, algorithms and opportunities

Wei Kang
U.S. Naval Postgraduate School

2018 IMA Workshop

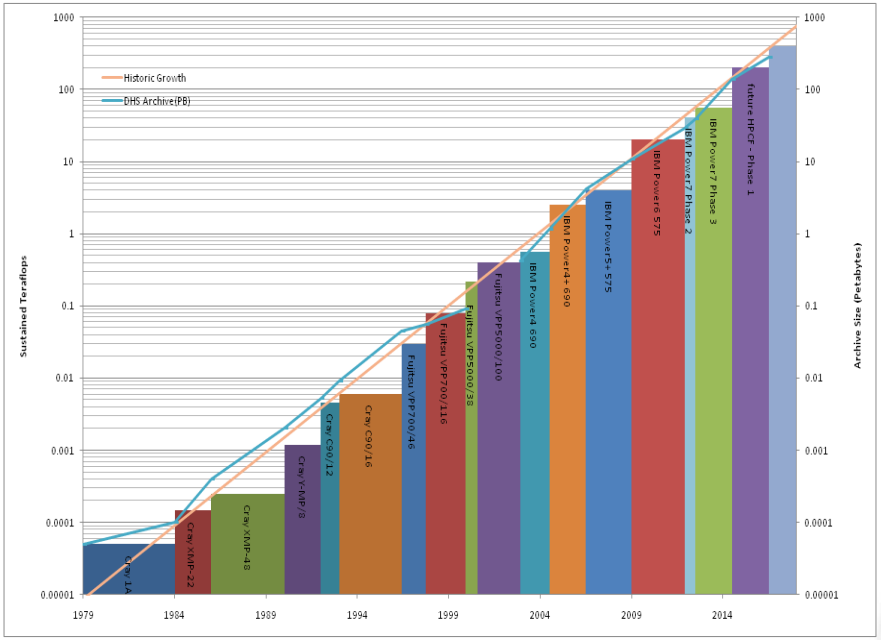
Collaboration

- Liang Xu (NRL)
- Sarah King (NRL)
- Kai Sun (UTK)
- Junjian Qi (UCF)
- Isaac Kaminer (NPS)
- Qi Gong (UCSC)
- Lucas Wilcox (NPS)
- Arthur J. Krener (NPS)
- Randy Boucher (Army)

- Data assimilation
- Numerical weather prediction
- Partial observability
- Power systems
- Swarms of autonomous vehicles
- Numerical algorithms

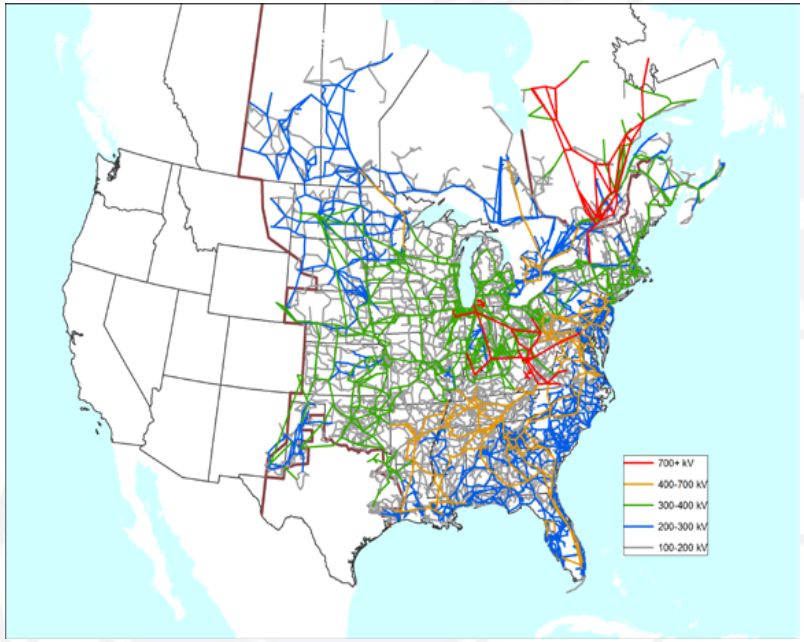
High-Dimensional Systems

Numerical Weather Prediction - ECMWF Global Model



ECMWF Global Model: 8×10^7 model variables are updated every 12 hours using 1.3×10^7 observations

Power System - Eastern Interconnection



EI: A largest electrical grid in North America. A simplified model has more than 25K buses, 28K lines, 8K transformers, $1,000$ generators.

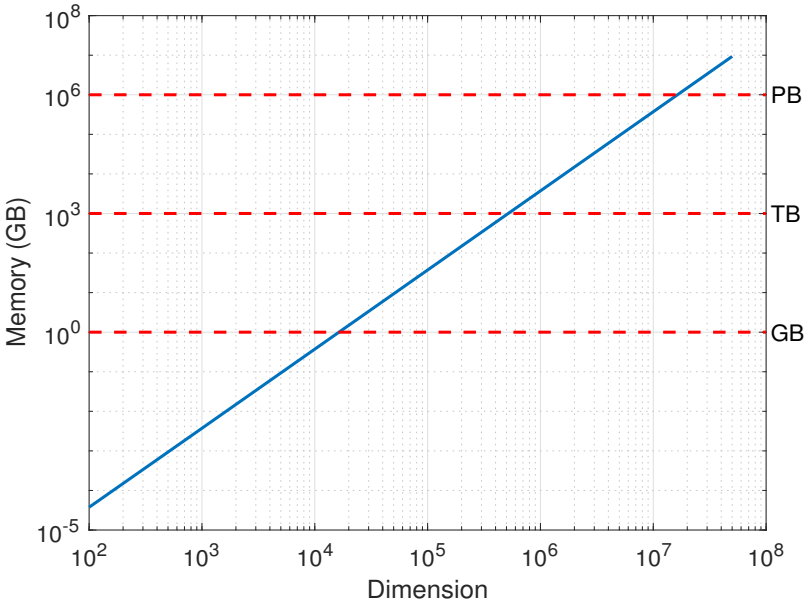
High-Dimensional Systems

Challenges

The scalability of algorithms is limited by several factors: computational load, I/O overhead and required memory size, degree of parallelism, and power consumption.

Covariance Matrix

dimension vs RAM size

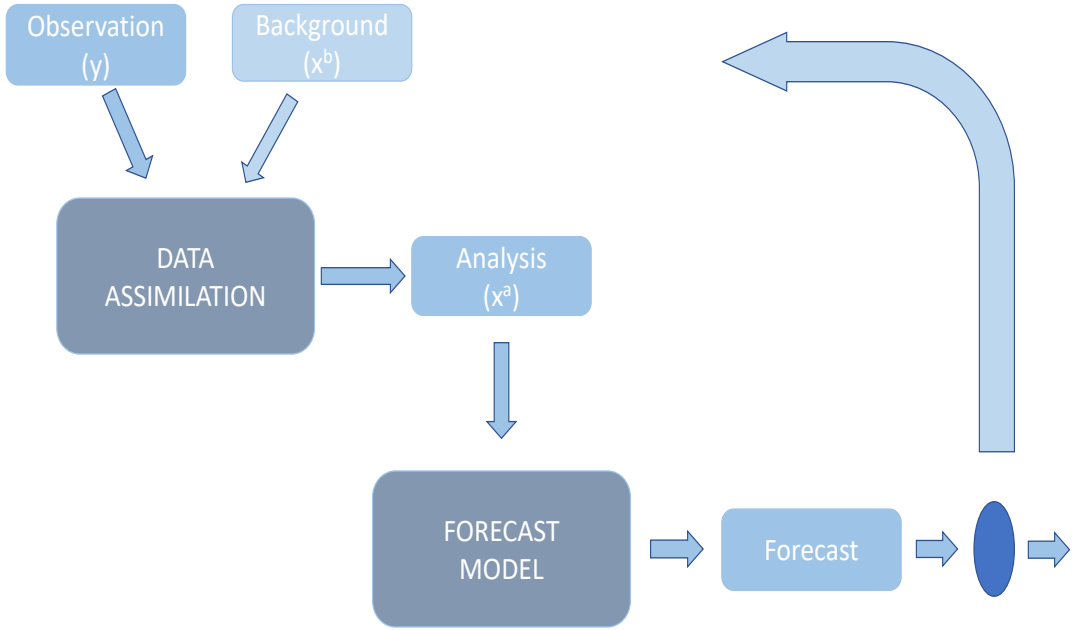


Quantitative Change
Becomes Qualitative

Data Assimilation

Liang Xu - NRL

Numerical Weather Prediction



Data Assimilation

System Model

$$\begin{aligned} x_{k+1} &= \mathcal{M}(x_k) + \eta_k, \quad x_k \in \mathbb{R}^n, \quad \eta_k \sim \text{model uncertainty}, Q \\ y_k &= \mathcal{H}(x_k) + \delta_k, \quad y_k \in \mathbb{R}^m, \quad \delta_k \sim \text{sensor noise}, R \end{aligned}$$

Linearization

$$\begin{aligned} x_{k+1} &= M_k x_k + \cdots \\ y_k &= H_k x_k + \cdots \end{aligned}$$

Both n and m are large. In daily operations, only a small part of sensor data is used.

Cost function

$$\mathcal{J}(x_0) = \frac{1}{2}(x_0 - x^b)^T (P_0^b)^{-1}(x_0 - x^b) + \frac{1}{2}\sum_{i=0}^{N}(y_i - \mathcal{H}(x_i))^T R^{-1}(y_i - \mathcal{H}(x_i))$$

4D-Var estimation

$$\begin{cases} \min\limits_{x_0^a} \mathcal{J}(x_0^a) \\ x_k^a = \mathcal{M}(x_{k-1}^a), \quad k = 1, 2, \cdots, N \end{cases}$$

The trajecotry, $x_0^a, x_1^a, \cdots, x_N^a$, is called an *analysis*; x^b is the initial estimate, or *background*; P_0^b is a positive definite matrix (fixed).

A linear solution

$$x_k^a = x_k^b + g_k, \quad 0 \le k \le N$$
$$g = PH^T \left(HPH^T + R\right)^{-1} (y - Hx^b)$$
$$g = \begin{bmatrix} g_0 \\ \vdots \\ g_N \end{bmatrix}, P = (P_{ij})_{i,j=0}^N, H = diag(H_0, \cdots, H_N), R, y, x^b, \cdots$$

Or define

$$z = \left(HPH^T + R\right)^{-1} (y - Hx^b)$$

Then

$$x_k^a = x_k^b + g_k, \quad 0 \le k \le N$$
$$g = PH^T z$$
$$HPH^T z + Rz = (y - Hx^b)$$

A 4D-Var algorithm

$$
\begin{aligned}
& f_{N+1} = 0 \\
& \text{for } k = N : 0 \\
& \quad f_k = M_k^T f_{k+1} + H^T z_k, \\
& g_0 = P_0^b f_0 \\
& \text{for } k = 0 : N \\
& \quad g_{k+1} = M_k g_k,
\end{aligned}
$$

- $M_k g_k$ and $M_k^T f_k$ are computed using a tangent linear model and an adjoint model.
- The equation $HPH^T + Rz_k = (y - Hx^b)$ is solved using conjugate gradient algorithm.

4D-Var

- 4D-Var has been widely used in today's numerical weather prediction (NWP). It was adopted shortly after the introduction of 3D-Var.
- Historically 3D-Var and 4D-Var are inspired by optimal control in which a cost function is minimized (Sasaki (1958), Lorenc (1986)).
- It is an effective method to provide estimation results with an affordable computational load.
- The method does not provide information about error covariance.
- It requires the development and maintenance of tangent linear models and adjoint models.

Ensemble KF

Ensemble $x^i_{k|k}$, $1 \leq i \leq N_{ens}$ $\qquad N_{ens} << n$

Forecast
$$x^i_{k+1|k} = \mathcal{M}(x^i_{k|k}) + \eta_k$$
$$y^i_{k+1|k} = \mathcal{H}(x^i_{k|k})$$
$$\bar{x}_{k+1} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} x^i_{k+1|k}$$
$$\bar{y}_{k+1} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} y^i_{k+1|k}$$

Ensemble KF (Cont.)

Analysis

$$\Delta X = \frac{1}{\sqrt{N_{ens}-1}}\left(\begin{bmatrix} x^1_{k+1|k} & \cdots & x^{N_{ens}}_{k+1|k} \end{bmatrix} - \bar{x}_{k+1}\mathbb{1}\right) \qquad n \times N_{ens}$$

$$\Delta Y = \frac{1}{\sqrt{N_{ens}-1}}\left(\begin{bmatrix} y^1_{k+1|k} & \cdots & y^{N_{ens}}_{k+1|k} \end{bmatrix} - \bar{y}_{k+1}\mathbb{1}\right) \qquad m \times N_{ens}$$

$$K = \Delta X \Delta Y^T (\Delta Y \Delta Y^T + R)^{-1} \qquad n \times m$$

$$x_{k+1|k+1} = \bar{x}_{k+1} + K(y_{k+1} - \bar{y}_{k+1})$$

Update

$$D = I - \Delta Y^T (\Delta Y \Delta Y^T + R)^{-1} \Delta Y \qquad N_{ens} \times N_{ens}$$

$$P_{k+1|k+1} = \Delta X D \Delta X^T$$

$$\begin{bmatrix} x^1_{k+1|k+1} & x^2_{k+1|k+1} & \cdots x^{N_{ens}}_{k+1|k+1} \end{bmatrix}$$

$$= x_{k+1|k+1} + \Delta X \left(\sqrt{(N_{ens}-1)D}\right)^T$$

EnKF

- EnKF does not require tangent linear model and adjoint model
- It contains partial information about error statistics
- Undersampling and rank deficiency
- Filter divergence
- Inbreeding - systematically underestimate the analysis error covariance
- Spurious correlations - covariance between state components that are not physically related

Localization - components physically far a way are uncorrelated.

Set $P_{ij} = 0$ for i and j far away

Suppose y is a part of state variables at $i_1, \cdots, i_m$. Let ρ be a sparse matrix that defines the sparsity of P. Define $\rho^{xy} \in \mathbb{R}^{n \times m}$ and $\rho^{yy} \in \mathbb{R}^{m \times m}$

$$P^{xy}_{st} = \rho_{i_s i_t}, P^{yy}_{st} = \rho_{i_s i_t}$$

Then

$$K_{loc} = \rho^{xy} .* (\Delta X \Delta Y^T) \left(\rho^{yy} .* (\Delta Y \Delta Y^T) + R \right)^{-1}$$

Inflation - rescale ΔX and ΔY by a small factor.

Sparsity-based Filters

Sparsity-based filters: The goal is to avoid rank deficiency, provide more error covariance information, and achieve granularity control for optimal parallelism.

A variety of parallel computing architectures are available; and new technologies are being developed rapidly.

- Multi-core CPU
- General-purpose GPU
- Clusters or massively parallel computing
- Grid computing
- Application-specific integrated curcuits
- .....

Sparsity-based Filters

Sparsity based methods

- Approximately sparse error covariance

$$N_{sp} = \text{maximum number of nonzero entries in columns}$$
$$\mathcal{I}_i(P) = \text{indices of nonzero entries in the } i\text{th-column}$$

- Component-based numerical model

$$\mathcal{M}(x_k^{sp}; \mathcal{I}) \text{ or } \mathcal{M}^{comp}$$
$$\mathcal{I} = \text{ indices of entries to be evaluated}$$

A progressive approach

Assume

$$M_k P_k M_k^T = P_k + \Delta P_{k+1}$$

To estimate ΔP_{k+1}, assume

$$M_{k+1} = I + \Delta M_k$$
$$x_{k+1} = \mathcal{M}(x_k) = x_k + \Delta(x_k)$$

Then

$$\begin{aligned} M_k P_k M_k^T &= (I + \Delta M_k) P_k (I + \Delta M_k^T) \\ &= P_k + \Delta M_k P_k + (\Delta M_k P_k)^T + \cdots \\ &\approx (\mathcal{M}(x_k + \delta P_k) - \mathcal{M}(x_k))/\delta \\ &+ (\mathcal{M}(x_k + \delta P_k) - \mathcal{M}(x_k))^T/\delta - P_k \end{aligned}$$

Prograssive KF

Background $x_{k|k}$ and $P^{sp}_{k|k}$ (sparse covariance approximation)

Forecast

$$x_{k+1|k} = \mathcal{M}(x_{k|k})$$
$$y_{k+1|k} = \mathcal{H}(x_{k|k})$$
$$P^{sp}_{k+1|k} = \left(\mathcal{M}^{comp}(x^{sp}_{k|k} + \delta P^{sp}_{k|k}) - \mathcal{M}^{comp}(x^{sp}_{k|k})\right)/\delta + \left(\mathcal{M}^{comp}(x^{sp}_{k|k} + \delta P^{sp}_{k|k}) - \mathcal{M}^{comp}(x^{sp}_{k|k})\right)^T/\delta - P^{sp}_{k|k} + Q$$

Analysis

$$K = P^{sp}_{k+1|k} H^T_{k+1} (H_{k+1} P^{sp}_{k+1|k} H^T_{k+1} + R)^{-1}$$
$$P^{sp}_{k+1|k+1} = (I - K H_{k+1}) P^{sp}_{k+1|k}$$
$$x_{k+1|k+1} = x_{k+1|k} + K(y_{k+1} - y_{k+1|k})$$

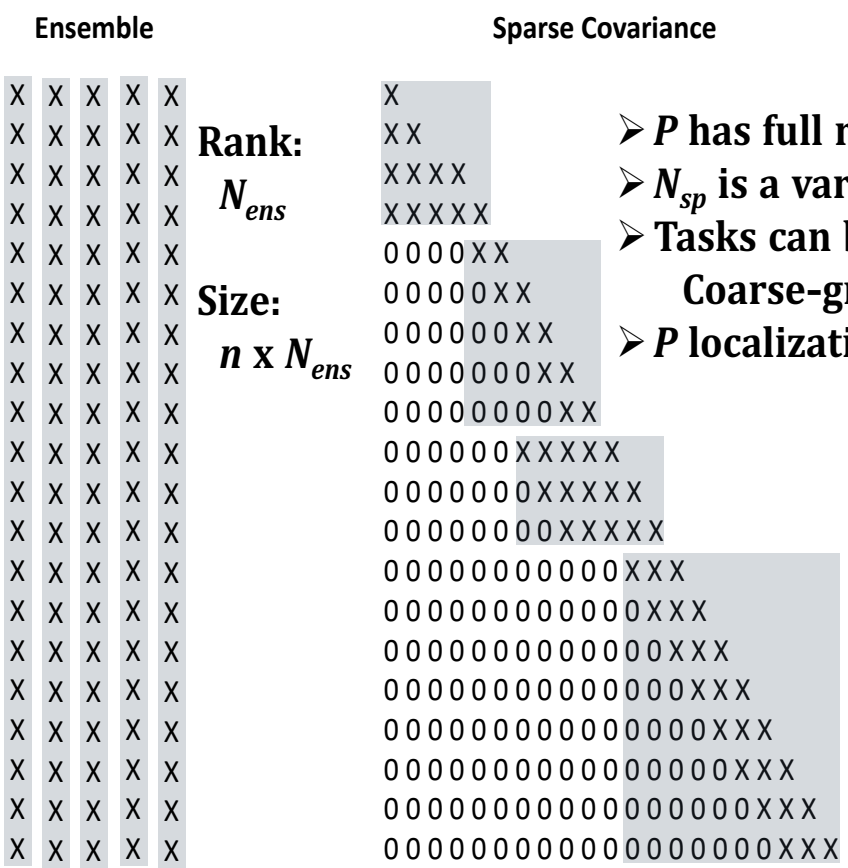
Sparsity-based Filters

Computational load

Progressive KF Full model	number of model components evaluation
$\mathcal{M}(x_k)$ $\mathcal{M}\left(x_k + \delta P_k(:,i)\right)$ $i = 1, 2, \cdots, n$	$(n+1)nN_p$ N_p- progressive steps
Progressive KF Component-based model	
$\mathcal{M}(x_k)$ $\mathcal{M}\left(x_k + \delta P_k(:,i), \mathcal{I}_i(P)\right)$ $i = 1, 2, \cdots, n$	$(N_{sp}+1)nN_p$
Ensemble KF	
$\mathcal{M}(x_k^i)$ $i = 1, 2, \cdots, N_{ens}$	$N_{ens}n$

Sparsity-based Filters

Avoid rank deficiency and achieve granularity control



- P has full rank
- N_{sp} is a variable
- Tasks can be grouped in different size:
  Coarse-grained, medium-grained, fine-grained
- P localization is straightforward

Sparsity-based Filters

Lorenz-96 model

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, 2, \cdots, m$$

$$x_{m+1} = x_1$$

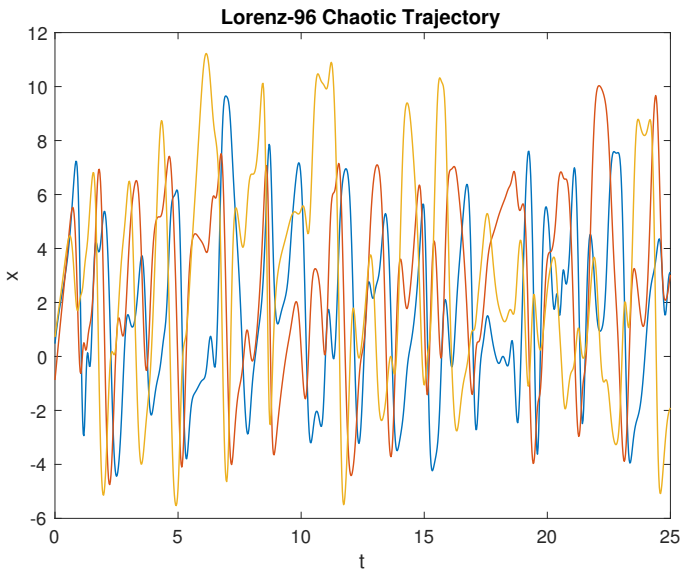
Discretization - 4th-order RK

$$x_k = \mathcal{M}(x_{k-1})$$

$$\Delta t = 0.025$$

$$F = 8$$

$$m = 40$$



Sparsity-based Filters

A comparison

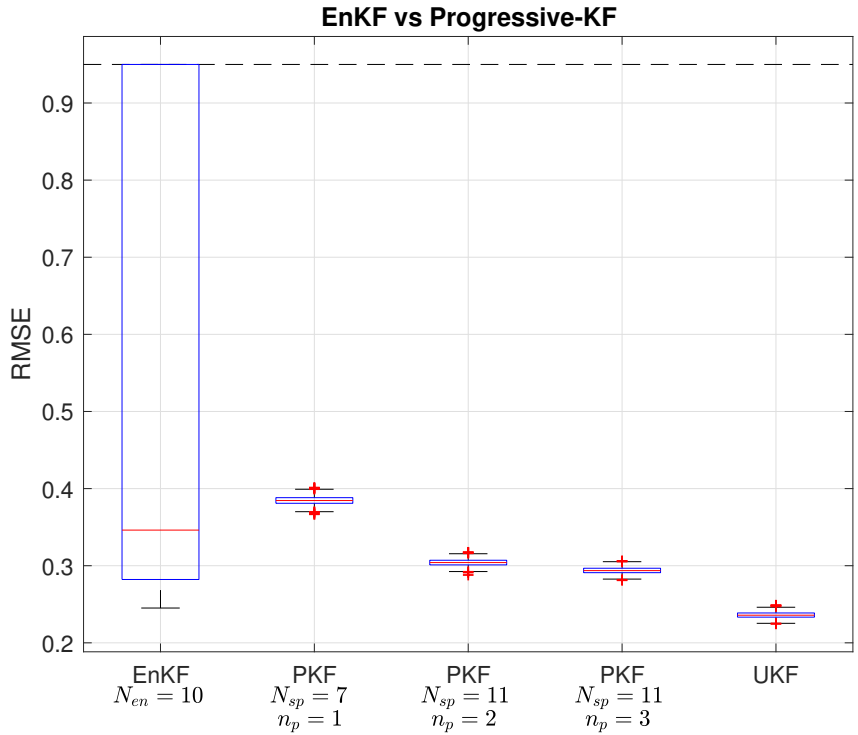
$N = 1000$ initial states in $[-1\ 1]$ - uniform distribution.

$N_{filter} = 4000$ filter steps

$m = 20$ measurement locations

$R = I$

Filter	Size	CMPT EVAL
EnKF	$N_{ens} = 10$	400
P-KF	$N_{sp} = 7$ $N_p = 1$	320
P-KF	$N_{sp} = 11$ $N_p = 2$	480x2
P-KF	$N_{sp} = 11$ $N_p = 3$	480x3



Sparsity-based Filters

Unscented KF (UKF)

σ-points $\quad x^i_{k|k}, \qquad 0 \le i \le 2n$

$$x^0_{k|k} = x_{k|k}$$

Forecast $\quad x^i_{k+1|k} = \mathcal{M}(x^i_{k|k}), \qquad y^i_{k+1|k} = \mathcal{H}(x^i_{k|k}), \qquad 0 \le i \le 2n$

$$\bar{x}_{k+1} = \sum_{i=0}^{2n} w_i x^i_{k+1|k}, \qquad \bar{y}_{k+1} = \sum_{i=0}^{2n} w_i y^i_{k+1|k}$$

$$P_{k+1|k} = \sum_{i=0}^{2n} w_i \Delta x^i_{k+1} (\Delta x^i_{k+1})^T + Q$$

$$\Delta x^i_{k+1} = x^i_{k+1|k} - \bar{x}_{k+1}$$

$$w_0 = \frac{\kappa}{n+\kappa}, \; w_i = \frac{\kappa}{2(n+\kappa)}$$

UKF (Cont.)

Analysis

$$P^{yy} = \sum_{i=0}^{2n} w_i \Delta y_{k+1}^i (\Delta y_{k+1}^i)^T + R, \quad \Delta y_{k+1} = y_{k+1|k}^i - \bar{y}_{k+1}$$

$$P^{xy} = \sum_{i=0}^{2n} w_i \Delta x_{k+1}^i (\Delta y_{k+1}^i)^T$$

$$K P^{yy} = P^{xy}$$

$$x_{k+1|k+1} = \bar{x}_{k+1} + K(y_{k+1} - \bar{y}_{k+1})$$

Update

$$P_{k+1|k+1} = P_{k+1|k} - K(P^{xy})^T$$

$$x_{k+1|k+1}^i = x_{k+1|k+1} + \sqrt{(n+\kappa)P_{k+1|k+1}}, \; i = 1, 2, \cdots, n$$

$$x_{k+1|k+1}^i = x_{k+1|k+1} - \sqrt{(n+\kappa)P_{k+1|k+1}}, \; i = n+1, \cdots, 2n$$

Sparsity-based Filters

Sparsity of square root matrix

Theorem (S. Toledo). If P is a symmetric positive definite matrix. The amount of storage for a Cholesky decomposition of P is $O(n + 2\eta(P))$, where $\eta(P)$ is the number of nonzero entries in P.

Assumption: The sparsity patterns of P and $\sqrt{P}$ are known - $I(P)$, $I(\sqrt{P})$.

Sparse UKF

Sparse $x^0_{k|k} = x_{k|k}$

σ-points σ^i, $\mathcal{I}_i$ (sparsity index) $\quad 1 \le i \le n$

Forecast

$$x^0_{k+1|k} = \mathcal{M}(x^0_{k|k}),$$

$$x^i_{k+1|k} = \mathcal{M}^{comp}(x^0_{k|k} + \sigma^i), \qquad x^{i+n}_{k+1|k} = \mathcal{M}^{comp}(x^0_{k|k} - \sigma^i)$$

$$y^i_{k+1|k} = \mathcal{H}(x^i_{k+1|k} \triangleright_{\mathcal{I}_i} x^0_{k+1|k}), \qquad 1 \le i \le 2n$$

$$\bar{x}_{k+1} = \sum_{i=0}^{2n} w_i (x^i_{k+1|k} \triangleright_{\mathcal{I}_i} x^0_{k+1|k}), \qquad \bar{y}_{k+1} = \sum_{i=0}^{2n} w_i y^i_{k+1|k}$$

$$P^{sp}_{k+1|k} = \sum_{i=0}^{2n} w_i \left(\Delta x^i_{k+1} (\Delta x^i_{k+1})^T \right)^{sp} + Q$$

$$w_0 = \frac{\kappa}{n+\kappa}, \; w_i = \frac{\kappa}{2(n+\kappa)}, \; \Delta x^i_{k+1} = x^i_{k+1|k} \triangleright_{\mathcal{I}_i} x^0_{k+1|k} - \bar{x}_{k+1}$$

$x_1^{sp} \triangleright_{\mathcal{I}} x_2$ - merging operation.

Sparse UKF (Cont.)

Analysis

$$P^{yy} = \sum_{i=0}^{2n} w_i \Delta y_{k+1}^i (\Delta y_{k+1}^i)^T, \quad \Delta y_{k+1}^i = y_{k+1|k}^i - \bar{y}_{k+1}$$

$$P^{xy} = \sum_{i=0}^{2n} w_i \Delta x_{k+1}^i (\Delta y_{k+1}^i)^T$$

$$K P^{yy} = P^{xy}$$

$$x_{k+1|k+1} = \bar{x}_{k+1} + K(y_{k+1} - \bar{y}_{k+1})$$

Update

$$P_{k+1|k+1}^{sp} = P_{k+1|k}^{sp} - \left(K(P^{xy})^T\right)^{sp}$$

$$\sigma^i, \mathcal{I}_i \sim \sqrt{(n+\kappa) P_{k+1|k+1}^{sp}}, \qquad i = 1, 2, \cdots, n$$

Sparsity-based Filters

A comparison

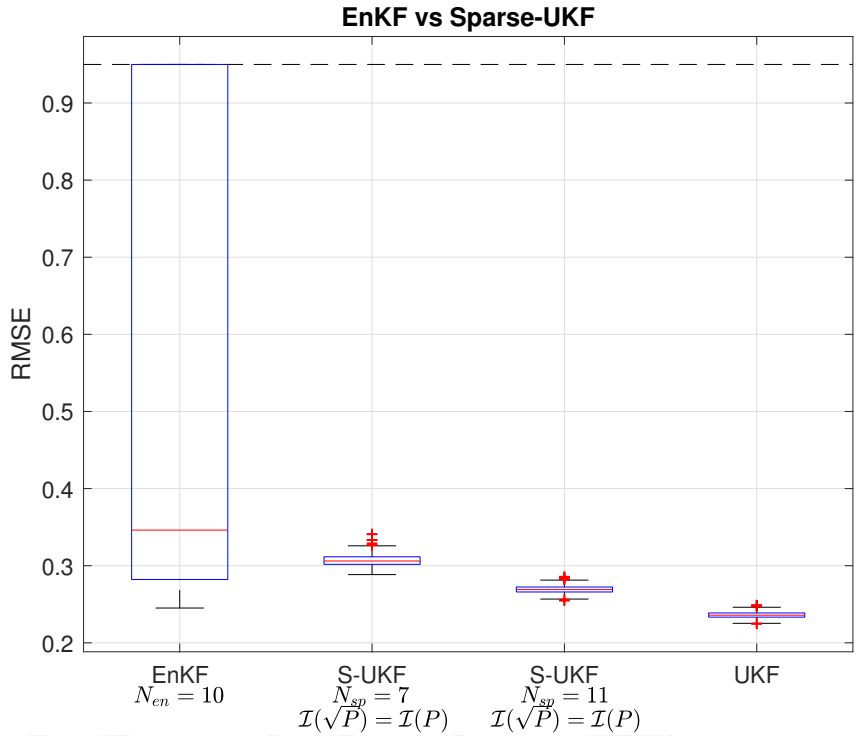
$N = 1000$ initial states in $[-1\ 1]$ - uniform distribution.

$N_{filter} = 4000$ filter steps

$m = 20$ measurement locations

$R = I$

Filter	Size	CMPT EVAL
EnKF	$N_{ens} = 10$	400
S-UKF	$N_{sp} = 7$	640
S-UKF	$N_{sp} = 11$	960



Summary

Progressive KF

- Full rank covariance
- Sparse covariance matrix reduces I/O load and memory size.
- Component-based model reduces computational load.
- Requirement: $P_{k+1} \approx P_k + MP_k + P_k M^T$

Sparse UKF

- Full rank covariance
- Sparse covariance matrix reduces I/O load and memory size.
- Component-based model reduces computational load.
- Requirement: Cholesky decomposition (additional computation and memory).

Partial Observability

Partial Observability - A quantitative definition

- x - space variable, λ - sensor location, $u(x,t)$ - state trajectory
- $y(t) = h(u(x,t),\lambda)$ - system output (sensor data)
- Background variation (data assimilation cost function)

$$J(u,\delta u,\lambda) = \delta u^T P_1 \delta u + ||y(\cdot,\lambda;u+\delta u) - y(\cdot,\lambda;u)||_{P_2}$$

Definition. Let $\epsilon > 0$ be a positive number. The observability ambiguity, ρ, is defined as

$$\rho^2 = \max_{\delta u} \; ||\mathcal{P}(\delta u)||_W$$

subject to: $||J(u^B,\delta u,\lambda)|| \leq \epsilon$

system model of $u(x,t)$

$\mathcal{P}(\delta u) \in W$ (subspace for estimation)

The ratio, ρ/ϵ, is called the unobservability index.

Partial Observability

Empirical Gramian

- Let $\{w_i\}$ be an orthonormal basis in W. Let $\Delta y(t)$ be the variation of the output. The Gramian is defined as follows

$$G = \langle w_i, w_j \rangle_{P_1} + \langle \Delta y_i(t), \Delta y_j(t) \rangle_{P_2}.$$

- Let $\sigma_{\min}$ be the smallest eigenvalue of G then $1/\sqrt{\sigma_{\min}}$ is a first order approximation of the unobservability index ρ/ϵ.

- The large scale and high dimensions in NWP make it less desirable or even impossible to make the entire state space observable. A finite number of modes is enough to provide an accurate approximation.

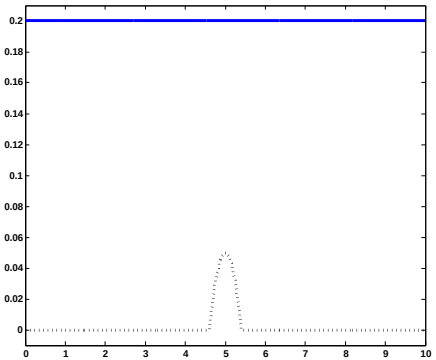
Shallow Water Equations (SWEs)

Simplest atmospheric flow model

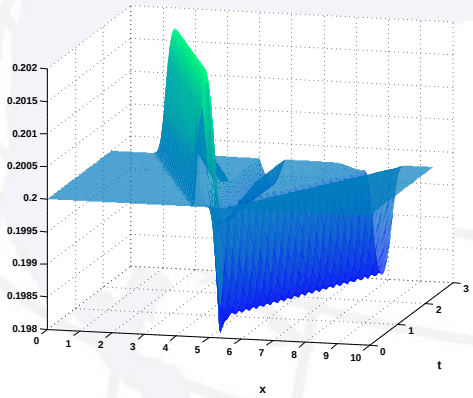
$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} + g\frac{\partial P}{\partial x} = 0$$

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} + \phi\frac{\partial u}{\partial x} = 0$$

- x- horizontal distance, u-horizontal velocity, p-fluid depth
- g-gravitational constant, $P(x)$-height of the obstacle
- $\phi = gp$-geopotential.



Initial conditions



Solution

Algorithm for Optimal Sensor Placement

Optimal sensor location - maximizing the unobservability index

$$\max \sigma_{min}(\lambda)$$

$$\text{subject to: } \lambda_{\text{lower}} < \lambda < \lambda_{\text{upper}}$$

where σ_{min} is the smallest eigenvalue of G.

1. Initialize λ_0, P_1, P_2
2. Approximate the current solution to the maximization problem by computing the smallest eigenvalue σ of the gramian matrix using the TLM.
3. Update λ using BFGS step

$$m_k(\lambda) = \sigma(\lambda_k) + \left(\frac{d\sigma}{d\lambda_k}\right)^T (\lambda - \lambda_k) + \frac{1}{2}(\lambda - \lambda_k)^T B_k (\lambda - \lambda_k)$$

   where B is the pseudo-Hessian.

4. Repeat until converges

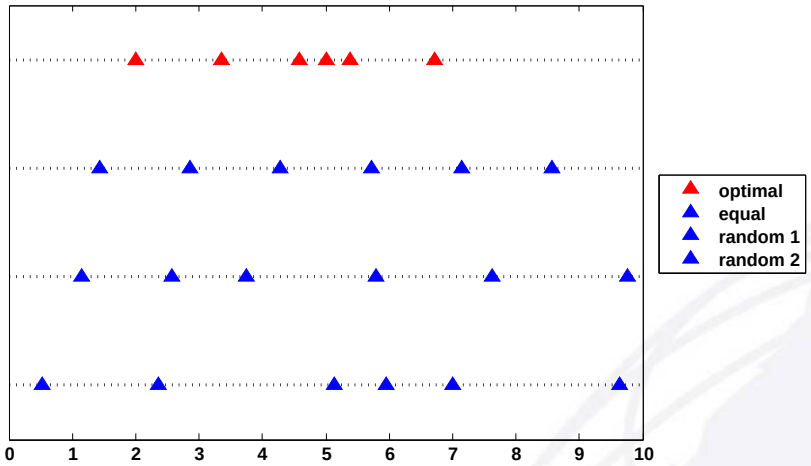
Parameters

N_λ	$=$	6	six sensors
N_{coef}	$=$	5	number of coefficients for each u, ϕ
L	$=$	10	length of x interval
T	$=$	2.3	time interval
N_t	$=$	250	
ρ	$=$	.01	
R	$=$	$1.6e-3$	variance in sensor data
σ_u^b	$=$	$5e-5$	used to compute L_c
σ_ϕ^b	$=$	.02	used to compute L_c
L_c^{-1}	$=$	$\begin{pmatrix} \sigma_u^b \hat{L} & 0 \\ 0 & \sigma_\phi^b \hat{L} \end{pmatrix}$	weight matrix P_1

where $\hat{L} = I + \gamma^{-1}\left(\frac{c^4}{2\Delta x^4}(L_{xx})^2\right)$, $L_{xx} = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & & & & \\ 0 & & & & & 1 \\ 1 & 0 & \cdots & & 1 & -2 \end{pmatrix}$

NAVAL POSTGRADUATE SCHOOL

Monte Carlo Numerical Experiments



	ρ/ϵ	RMSE $u^a(0)$	RSME of u^a	RMSE $\phi^a(0)$	RSME of ϕ^a
Optimal	9.29	0.3577	0.1174	0.5186	0.1591
Equal	13.05	0.4419	0.1484	0.5787	0.1974
Rand 1	19.08	0.4323	0.1459	0.5734	0.1951
Rand 2	14.17	0.4245	0.1431	0.5671	0.1914
Equal Imp	28.8%	19.1 %	20.9 %	10.4 %	19.4 %
Rand 1 Imp	51.3 %	17.3 %	19.5 %	9.6%	18.45 %
Rand 2 Imp	34.4 %	15.7 %	17.96 %	8.6 %	16.88 %

The optimal sensor locations improved the estimation accuracy in the 4D-Var data assimilation for the observed variable ϕ and the unobserved variable u.

Some Remarks

- Partial observability and estimation for large scale systems have also been applied to power systems and networked swarms.
- Unobservability index is a worst-case measure of observability. Other quantitative measure exists for various types of systems and sensor networks.
- User-knowledge and observability?
- Estimation of variables observable in a zero measure set?

Some Remarks

- Partial observability and estimation for large scale systems have also been applied to power systems and networked swarms.
- Unobservability index is a worst-case measure of observability. Other quantitative measure exists for various types of systems and sensor networks.
- User-knowledge and observability?
- Estimation of variables observable in a zero measure set?

Thank you!