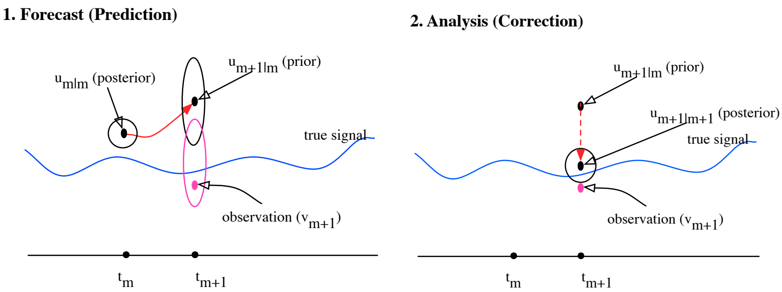
Mathematical Strategies for Filtering Turbulent Systems: Sparse Observations, Model Errors, and Stochastic Parameter Estimation

John Harlim

Courant Institute of Mathematical Sciences, New York University

July 1, 2009

What is filtering?



The correction step is an application of Bayesian update

$$p(u_{m+1|m+1}) \equiv p(u_{m+1|m}|v_{m+1}) \sim p(u_{m+1|m})p(v_{m+1}|u_{m+1|m})$$

Kalman filter formula produces the optimal unbiased posterior mean and covariance by assuming linear model and Gaussian observations and forecasts errors.

The standard Kalman filter algorithm for solving:

$$\begin{aligned} u_{m+1} &= Fu_m + \bar{f}_m + \sigma_{m+1} \\ v_m &= Gu_m + \sigma_m^o \end{aligned}$$

The standard Kalman filter algorithm for solving:

$$\begin{aligned} u_{m+1} &= Fu_m + \bar{f}_m + \sigma_{m+1} \\ v_m &= Gu_m + \sigma_m^o \end{aligned}$$

Forecast (Prediction)

$$\begin{aligned} &\text{A)} \quad \bar{u}_{m+1|m} = F\bar{u}_{m|m} + \bar{f}_m, \\ &\text{B)} \quad R_{m+1|m} = FR_{m|m}F^* + R, \end{aligned}$$

The standard Kalman filter algorithm for solving:

$$
\begin{aligned}
u_{m+1} &= Fu_m + \bar{f}_m + \sigma_{m+1} \\
v_m &= Gu_m + \sigma_m^o
\end{aligned}
$$

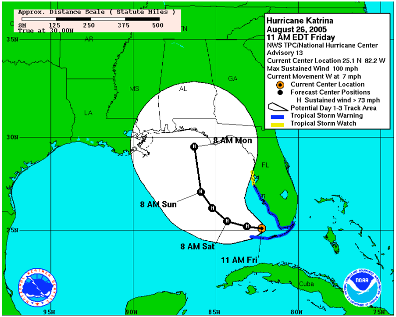
Forecast (Prediction)

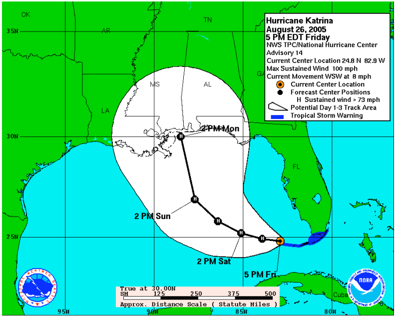
$$
\begin{aligned}
&\text{A)} \quad \bar{u}_{m+1|m} = F\bar{u}_{m|m} + \bar{f}_m, \\
&\text{B)} \quad R_{m+1|m} = FR_{m|m}F^* + R,
\end{aligned}
$$

Analysis (Correction)

$$
\begin{aligned}
&\text{D)} \quad \bar{u}_{m+1|m+1} = (\mathcal{I} - K_{m+1}G)\bar{u}_{m+1|m} + K_{m+1}v_{m+1} \\
&\text{E)} \quad R_{m+1|m+1} = (\mathcal{I} - K_{m+1}G)R_{m+1|m}, \\
&\text{F)} \quad K_{m+1} = R_{m+1|m}G^T(GR_{m+1|m}G^T + R^o)^{-1}.
\end{aligned}
$$

Example of application: predicting path of hurricane





Computational and Theoretical Issues:

- How to handle large system? Perhaps $N = 10^6$ state variables (e.g., 200 km resolved Global Weather Model)

Computational and Theoretical Issues:

- How to handle large system? Perhaps $N = 10^6$ state variables (e.g., 200 km resolved Global Weather Model)
- Where is the computational burden? Propagating covariance matrix of size $N \times N$ (6N minutes = 300,000 hours).

Computational and Theoretical Issues:

- How to handle large system? Perhaps $N = 10^6$ state variables (e.g., 200 km resolved Global Weather Model)
- Where is the computational burden? Propagating covariance matrix of size $N \times N$ (6N minutes = 300,000 hours).
- Handling nonlinearity! Why not particle filter? Convergence requires ensemble size that grows exponentially with respect to the ensemble spread relative to observation errors rather than to the state dimension per se(Bengtsson, Bickel, and Li 2008).

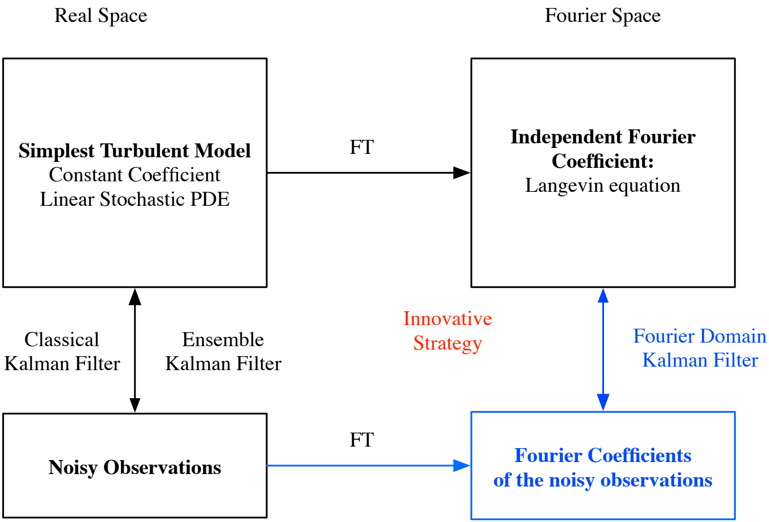
Computational and Theoretical Issues:

- How to handle large system? Perhaps $N = 10^6$ state variables (e.g., 200 km resolved Global Weather Model)
- Where is the computational burden? Propagating covariance matrix of size $N \times N$ (6N minutes = 300,000 hours).
- Handling nonlinearity! Why not particle filter? Convergence requires ensemble size that grows exponentially with respect to the ensemble spread relative to observation errors rather than to the state dimension per se(Bengtsson, Bickel, and Li 2008).
- Some successful strategies: Ensemble Kalman filters (ETKF of Bishop et al. 2001, EAKF of Anderson 2001). Each involves computing singular value decomposition (SVD).

Computational and Theoretical Issues:

- How to handle large system? Perhaps $N = 10^6$ state variables (e.g., 200 km resolved Global Weather Model)
- Where is the computational burden? Propagating covariance matrix of size $N \times N$ (6N minutes $=$ 300,000 hours).
- Handling nonlinearity! Why not particle filter? Convergence requires ensemble size that grows exponentially with respect to the ensemble spread relative to observation errors rather than to the state dimension per se(Bengtsson, Bickel, and Li 2008).
- Some successful strategies: Ensemble Kalman filters (ETKF of Bishop et al. 2001, EAKF of Anderson 2001). Each involves computing singular value decomposition (SVD).
- However, these accurate filters are not immune from "catastrophic filter divergence" (diverge beyond machine infinity) when observations are sparse, even when the true signal is a dissipative system with "absorbing ball property".

Filtering in Frequency space



Filtering Stochastically forced advection-diffusion equation

$$\frac{\partial u(x,t)}{\partial t} = -\frac{\partial}{\partial x}u(x,t) + \bar{F}(x,t)$$

Filtering Stochastically forced advection-diffusion equation

$$\frac{\partial u(x,t)}{\partial t} = -\frac{\partial}{\partial x}u(x,t) + \bar{F}(x,t) + \mu\frac{\partial^2}{\partial x^2}u(x,t) + \sigma(x)\dot{W}(t)$$

Filtering Stochastically forced advection-diffusion equation

$$\begin{aligned}\frac{\partial u(x,t)}{\partial t} &= -\frac{\partial}{\partial x}u(x,t) + \bar{F}(x,t) + \mu\frac{\partial^2}{\partial x^2}u(x,t) + \sigma(x)\dot{W}(t)\\ v(\tilde{x}_j, t_m) &= u(\tilde{x}_j, t_m) + \sigma_m^o, \quad \tilde{x}_j = j\tilde{h}, (2N+1)\tilde{h} = 2\pi.\end{aligned}$$

where $\sigma_m^o \sim \mathcal{N}(0, r^o)$.

Filtering Stochastically forced advection-diffusion equation

$$\begin{aligned}\frac{\partial u(x,t)}{\partial t} &= -\frac{\partial}{\partial x}u(x,t) + \bar{F}(x,t) + \mu\frac{\partial^2}{\partial x^2}u(x,t) + \sigma(x)\dot{W}(t)\\ v(\tilde{x}_j,t_m) &= u(\tilde{x}_j,t_m) + \sigma_m^o, \quad \tilde{x}_j = j\tilde{h}, (2N+1)\tilde{h} = 2\pi.\end{aligned}$$

where $\sigma_m^o \sim \mathcal{N}(0, r^o)$.

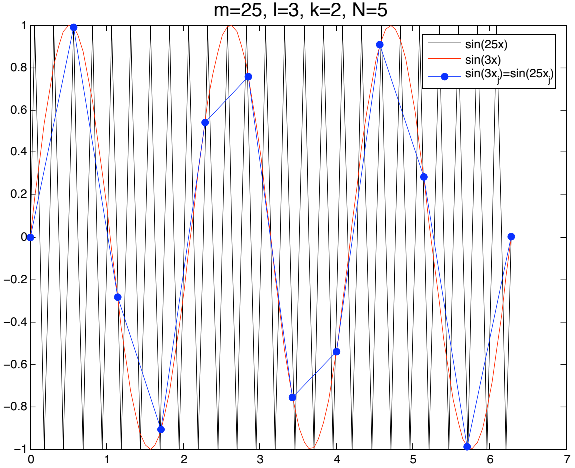
In Fourier Domain, we reduce filtering (2N+1) dimensional problem to filtering decoupled scalar stochastic Langevin equations:

$$\begin{aligned}d\hat{u}_k(t) &= [(-\mu k^2 - \mathrm{i}k)\hat{u}_k(t) + \hat{F}_k(t)]dt + \sigma_k dW_k(t)\\ \hat{v}_{k,m} &= \hat{u}_{k,m} + \hat{\sigma}^o_{k,m}\end{aligned}$$

where $\hat{\sigma}^o_{k,m} \sim \mathcal{N}(0, r^o/(2N+1))$.

How to deal with Sparse Regularly Spaced Observations?

ALIASING !!



Recall Aliasing Formula:

- Fine mesh: $f(x_j) = \sum_{|k|\le N} \hat{f}_{fine}(k)\mathrm{e}^{\mathrm{i}kx_j}$ where $x_j = jh$ and $(2N+1)h = 2\pi$.

Recall Aliasing Formula:

- Fine mesh: $f(x_j) = \sum_{|k|\leq N} \hat{f}_{fine}(k)\mathrm{e}^{\mathrm{i}kx_j}$ where $x_j = jh$ and $(2N+1)h = 2\pi$.
- Coarse mesh: $f(\tilde{x}_j) = \sum_{|\ell|\leq M} \hat{f}_{coarse}(\ell)\mathrm{e}^{\mathrm{i}\ell\tilde{x}_j}$ where $\tilde{x}_j = j\tilde{h}$ and $(2M+1)\tilde{h} = 2\pi$.

Recall Aliasing Formula:

- Fine mesh: $f(x_j) = \sum_{|k|\le N} \hat{f}_{fine}(k)\mathrm{e}^{\mathrm{i}kx_j}$ where $x_j = jh$ and $(2N+1)h = 2\pi$.
- Coarse mesh: $f(\tilde{x}_j) = \sum_{|\ell|\le M} \hat{f}_{coarse}(\ell)\mathrm{e}^{\mathrm{i}\ell\tilde{x}_j}$ where $\tilde{x}_j = j\tilde{h}$ and $(2M+1)\tilde{h} = 2\pi$.
- Suppose the coarse grid points $\tilde{x}_j$ coincide with the fine mesh grid points x_j at every $P = (2N+1)/(2M+1)$ fine grid points.

Recall Aliasing Formula:

- Fine mesh: $f(x_j) = \sum_{|k|\leq N} \hat{f}_{fine}(k)\mathrm{e}^{\mathrm{i}kx_j}$ where $x_j = jh$ and $(2N+1)h = 2\pi$.
- Coarse mesh: $f(\tilde{x}_j) = \sum_{|\ell|\leq M} \hat{f}_{coarse}(\ell)\mathrm{e}^{\mathrm{i}\ell\tilde{x}_j}$ where $\tilde{x}_j = j\tilde{h}$ and $(2M+1)\tilde{h} = 2\pi$.
- Suppose the coarse grid points $\tilde{x}_j$ coincide with the fine mesh grid points x_j at every $P = (2N+1)/(2M+1)$ fine grid points.
- Since $\mathrm{e}^{\mathrm{i}k\tilde{x}_j} = \mathrm{e}^{\mathrm{i}(\ell+q(2M+1))\tilde{x}_j} = \mathrm{e}^{\mathrm{i}\ell\tilde{x}_j}$,

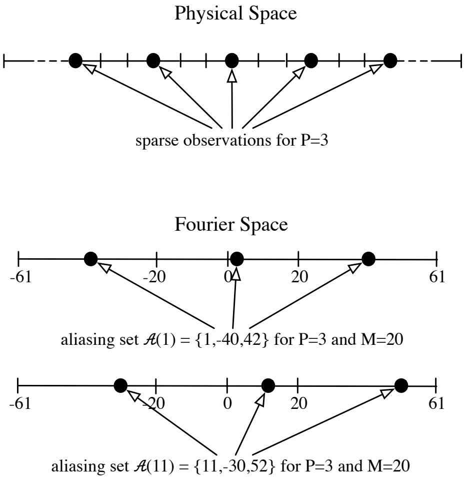
Recall Aliasing Formula:

- Fine mesh: $f(x_j) = \sum_{|k|\le N} \hat{f}_{fine}(k)\mathrm{e}^{\mathrm{i}kx_j}$ where $x_j = jh$ and $(2N+1)h = 2\pi$.
- Coarse mesh: $f(\tilde{x}_j) = \sum_{|\ell|\le M} \hat{f}_{coarse}(\ell)\mathrm{e}^{\mathrm{i}\ell\tilde{x}_j}$ where $\tilde{x}_j = j\tilde{h}$ and $(2M+1)\tilde{h} = 2\pi$.
- Suppose the coarse grid points $\tilde{x}_j$ coincide with the fine mesh grid points x_j at every $P = (2N+1)/(2M+1)$ fine grid points.
- Since $\mathrm{e}^{\mathrm{i}k\tilde{x}_j} = \mathrm{e}^{\mathrm{i}(\ell+q(2M+1))\tilde{x}_j} = \mathrm{e}^{\mathrm{i}\ell\tilde{x}_j}$,
- We deduce

$$\hat{f}_{coarse}(\ell) \quad = \quad \sum_{k_j\in\mathcal{A}(\ell)} \hat{f}_{fine}(k_j), \quad |\ell| \le M,$$

where $\mathcal{A}(\ell) = \{k : |k| \le N, k = \ell + q(2M+1), q \in \mathbb{Z}\}$

Consider the following sparse observations: 123 grid pts (61 modes) but only 41 observations (20 modes) available



Aliasing Formula:

Observation at time t_m becomes:

$$\hat{v}_{\ell,m} = \sum_{k_j \in \mathcal{A}(\ell)} \hat{u}_{k_j,m} + \hat{\sigma}^o_{\ell,m}, = G\vec{\hat{u}}_{\ell,m} + \hat{\sigma}^o_{\ell,m}$$

where $G = [1, 1, \ldots, 1]$ and $\hat{\sigma}^o_{\ell,m} \sim \mathcal{N}(0, r^o/(2M+1))$.

Aliasing Formula:

Observation at time t_m becomes:

$$\hat{v}_{\ell,m} = \sum_{k_j \in \mathcal{A}(\ell)} \hat{u}_{k_j,m} + \hat{\sigma}^o_{\ell,m}, = G\vec{\hat{u}}_{\ell,m} + \hat{\sigma}^o_{\ell,m}$$

where $G = [1, 1, \ldots, 1]$ and $\hat{\sigma}^o_{\ell,m} \sim \mathcal{N}(0, r^o/(2M+1))$.

Reduced Filters

- With the aliasing formula above, we reduce filtering $(2N+1)$ dimensional system with $(2M+1)$ observations, where $M < N$, to decoupled $P = (2N+1)/(2M+1)$ dimensional problem with scalar observations (FDKF).

Aliasing Formula:

Observation at time t_m becomes:

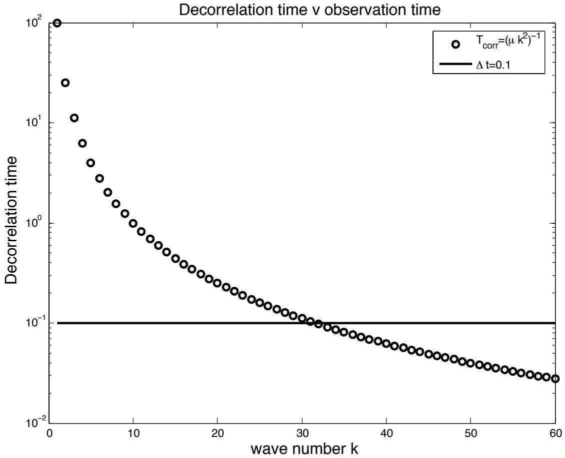
$$\hat{v}_{\ell,m} = \sum_{k_j \in \mathcal{A}(\ell)} \hat{u}_{k_j,m} + \hat{\sigma}^o_{\ell,m}, = G\vec{\hat{u}}_{\ell,m} + \hat{\sigma}^o_{\ell,m}$$

where $G = [1, 1, \ldots, 1]$ and $\hat{\sigma}^o_{\ell,m} \sim \mathcal{N}(0, r^o/(2M+1))$.

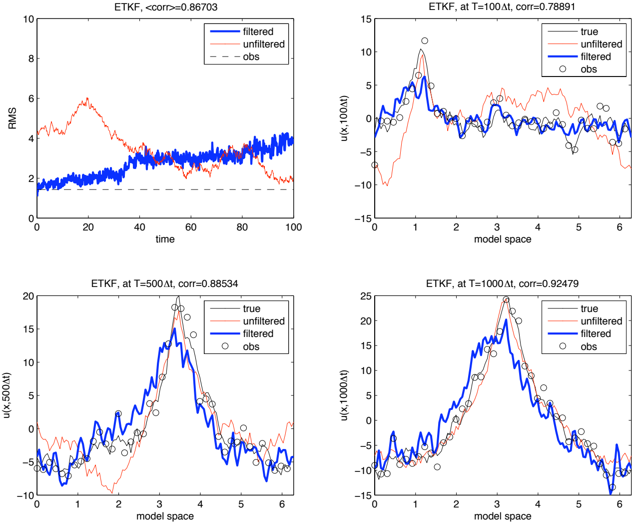
Reduced Filters

- With the aliasing formula above, we reduce filtering $(2N+1)$ dimensional system with $(2M+1)$ observations, where $M < N$, to decoupled $P = (2N+1)/(2M+1)$ dimensional problem with scalar observations (FDKF).
- When the energy spectrum of the system decays as a function of wavenumbers, we can ignore the high wavenumbers (e.g., RFDKF, SDAF).

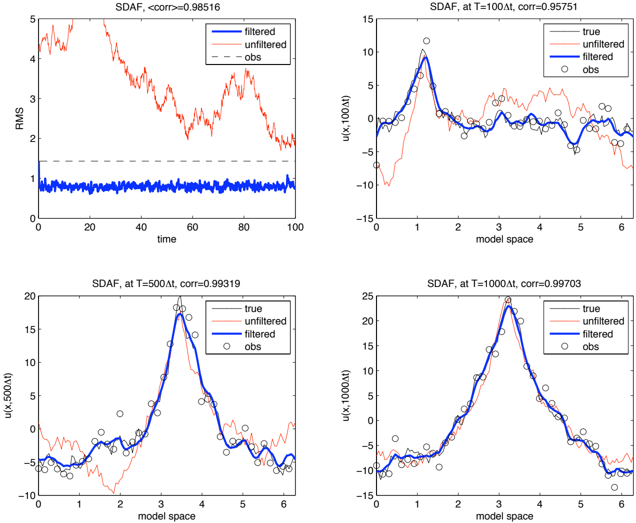
Decorrelation time vs observation time:



Ensemble Kalman Filter diverges with ensemble size $150 > N = 123$. Extreme event, $\Delta t_2 = 0.1, E_k = k^{-5/3}, P = 3, r^o = 2.05$



Reduced Filter produces high skill Spontaneous development of extreme event for $\Delta t_2 = 0.1$ and $E_k = k^{-5/3}$, $P = 3$, $r^o = 2.05$



Summary of Part I:

Summary of Part I:

- In our assessment, we find that filtering sparsely observed linear problem with ensemble Kalman filter with ensemble size larger than the model dimensionality does not guaranteed convergence solution.

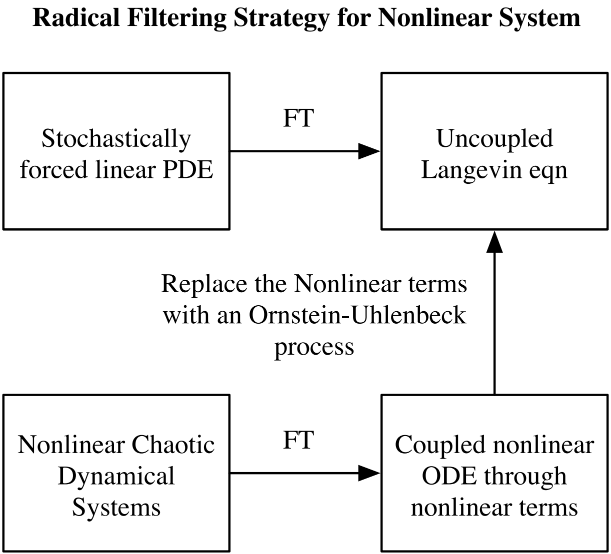
Summary of Part I:

- In our assessment, we find that filtering sparsely observed linear problem with ensemble Kalman filter with ensemble size larger than the model dimensionality does not guaranteed convergence solution.
- FDKF suggests that ignoring the cross covariance between different aliasing sets is not only computationally advantageous but it also produces more accurate solutions.

Summary of Part I:

- In our assessment, we find that filtering sparsely observed linear problem with ensemble Kalman filter with ensemble size larger than the model dimensionality does not guaranteed convergence solution.
- FDKF suggests that ignoring the cross covariance between different aliasing sets is not only computationally advantageous but it also produces more accurate solutions.
- Intuitively, this works because the reduced filter avoids the spurious correlations between different wave numbers.

Nonlinearity

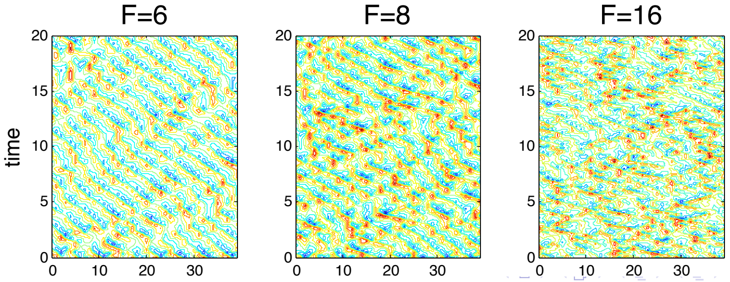


Filtering turbulent nonlinear dynamical systems

L-96 model (Lorenz 1996), 40-dim, "absorbing ball property".

$$\frac{du_j}{dt} = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F, \quad j = 0, \ldots, J-1$$

	F	λ_1	N^+	KS	T_{corr}
Weakly chaotic	6	1.02	12	5.547	8.23
Strongly chaotic	8	1.74	13	10.94	6.704
Fully turbulent	16	3.945	16	27.94	5.594



The “poorman’s” Climatological Stochastic Model (CSM):

- Fourier coefficients of normalized L-96 [MAG05]:

$$d\hat{u}_k(t) = [(-d_k + \mathrm{i}\omega_k)\hat{u}_k(t) + E_p^{-1}(F - \bar{u})\delta_{k,0}]dt + NL$$

The “poorman’s” Climatological Stochastic Model (CSM):

- Fourier coefficients of normalized L-96 [MAG05]:

$$d\hat{u}_k(t) = [(-d_k + \mathrm{i}\omega_k)\hat{u}_k(t) + E_p^{-1}(F - \bar{u})\delta_{k,0}]dt + NL$$

- Replace the nonlinearity with Ornstein-Uhlenbeck process:

$$d\hat{u}_k(t) = [(-\gamma_k + \mathrm{i}\omega_k)\hat{u}_k(t) + E_p^{-1}(F - \bar{u})\delta_{k,0}]dt + \sigma_k dW_k(t).$$

The "poorman's" Climatological Stochastic Model (CSM):

- Fourier coefficients of normalized L-96 [MAG05]:

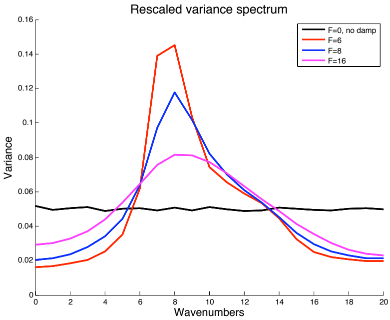
$$d\hat{u}_k(t) = [(-d_k + \mathrm{i}\omega_k)\hat{u}_k(t) + E_p^{-1}(F - \bar{u})\delta_{k,0}]dt + NL$$

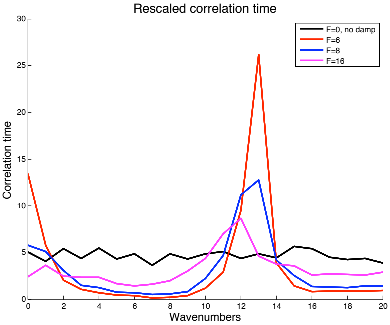
- Replace the nonlinearity with Ornstein-Uhlenbeck process:

$$d\hat{u}_k(t) = [(-\gamma_k + \mathrm{i}\omega_k)\hat{u}_k(t) + E_p^{-1}(F - \bar{u})\delta_{k,0}]dt + \sigma_k dW_k(t).$$

- Fit the damping coefficient γ_k and stochastic noise strength σ_k to the equilibrium variance and correlation time.

Equilibrium Variance and Correlation Time



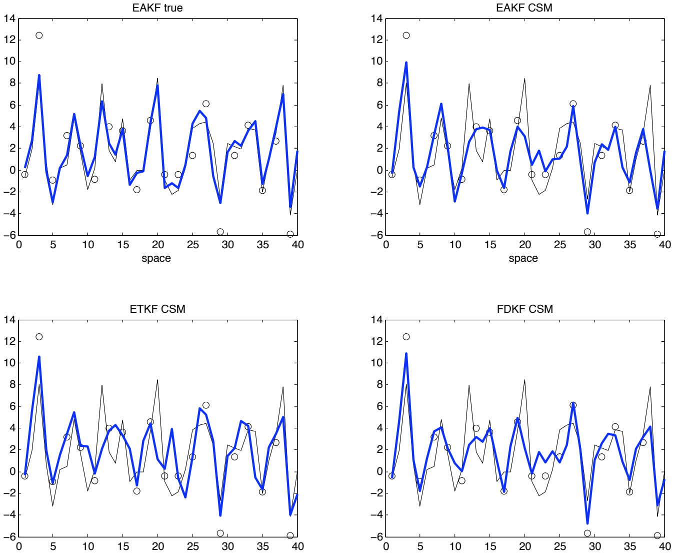


Regularly spaced sparse observations: weakly chaotic regime
$F = 6, P = 2, r^o = 1.96, \Delta t = 0.234$.

This is a regime where EAKF true is superior.

perfect model	RMS	corr.
EAKF true	0.82	0.95
ETKF true	∞	-
No Filter	2.8	-
model error	**RMS**	**corr.**
EAKF CSM	2.20	0.64
ETKF CSM	2.50	0.55
FDKF CSM	2.07	0.69

Regularly spaced sparse observations: weakly chaotic regime
$F = 6, P = 2, r^o = 1.96, \Delta t = 0.234$



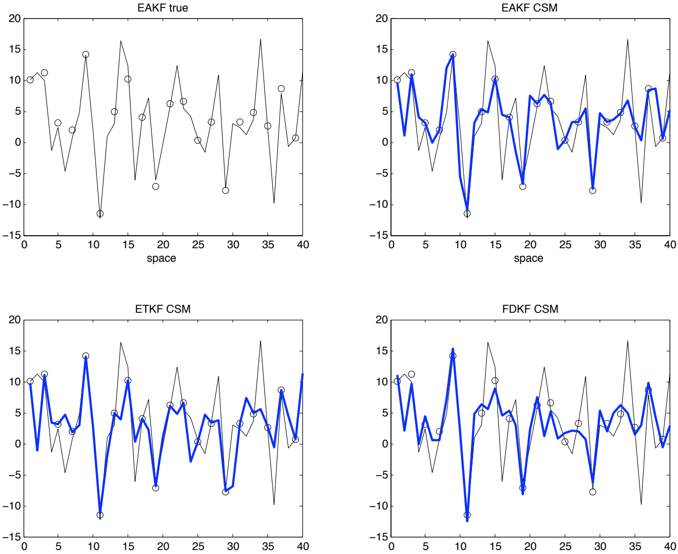
Regularly spaced sparse observations: fully turbulent regime
$F = 16, P = 2, r^o = 0.81, \Delta t = 0.078.$

This is a regime where FDKF is superior.

Scheme	RMS	corr.
EAKF true	∞	-
ETKF true	∞	-
No Filter	6.3	0
model error	RMS	corr.
EAKF CSM	5.15	0.61
ETKF CSM	5.80	0.54
FDKF CSM	4.80	0.66

Regularly spaced sparse observations: fully turbulent regime

$F = 16, P = 2, r^o = 0.81, \Delta t = 0.078$



Summary of Part II:

Summary of Part II:

- We demonstrate that in the fully turbulent regime, perfect model is not necessarily needed for filtering. In our example, we encounter an ensemble collapse which yields filter divergence beyond machine infinity.

Summary of Part II:

- We demonstrate that in the fully turbulent regime, perfect model is not necessarily needed for filtering. In our example, we encounter an ensemble collapse which yields filter divergence beyond machine infinity.
- In the presence of model errors through CSM, our reduced filtering strategy produces better solutions.

Summary of Part II:

- We demonstrate that in the fully turbulent regime, perfect model is not necessarily needed for filtering. In our example, we encounter an ensemble collapse which yields filter divergence beyond machine infinity.
- In the presence of model errors through CSM, our reduced filtering strategy produces better solutions.
- Practically, our radical strategy is independent of tunable parameters, SVD, and ensemble size.

Online Model Error Estimation Strategy

The simplest contemporary strategy to cope with model errors for filtering with an imperfect model nonlinear dynamical system depending on parameters, λ,

$$\frac{du}{dt} = F(u, \lambda)$$

is to augment the state variable u, by the parameters λ, and adjoin an approximate dynamical equation for the parameters

$$\frac{d\lambda}{dt} = g(\lambda).$$

Climatological Stochastic Model

$$du(t) = \Big[(-\bar{\gamma} + \mathrm{i}\omega)u(t) + F(t)\Big]dt + \sigma dW(t)$$

Climatological Stochastic Model

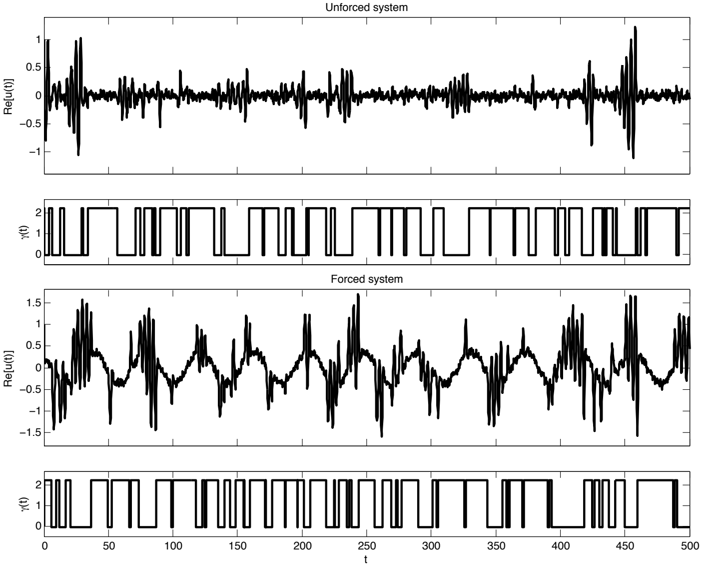
$$du(t) = \Big[(-\bar{\gamma} + \mathrm{i}\omega)u(t) + F(t)\Big]dt + \sigma dW(t)$$

Nonlinear Extended Kalman Filter:

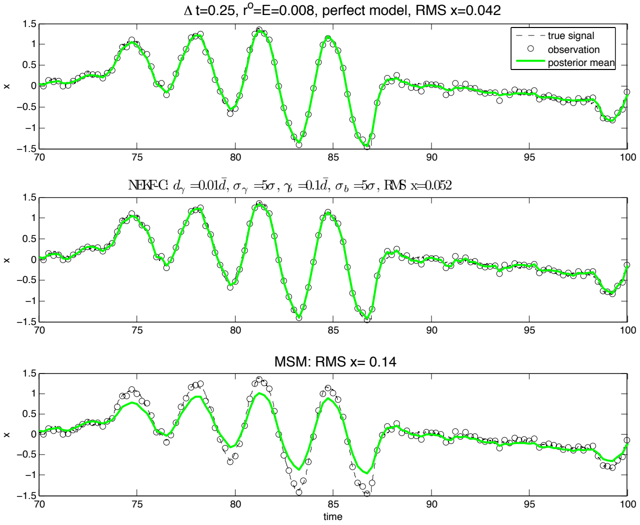
$$\begin{aligned} du(t) &= \Big[(-\gamma(t) + \mathrm{i}\omega)u(t) + F(t) + b(t)\Big]dt + \sigma dW(t) \\ db(t) &= (-\gamma_b + \mathrm{i}\omega_b)b(t)dt + \sigma_b dW_b(t) \\ d\gamma(t) &= -d_\gamma(\gamma(t) - \hat{\gamma})dt + \sigma_\gamma dW_\gamma(t) \end{aligned}$$

We find stochastic parameters $\{\gamma_b, \omega_b, \sigma_b, d_\gamma, \sigma_\gamma\}$ that are robust for high filter skill beyond CSM and in many occasions comparable to the perfect model.

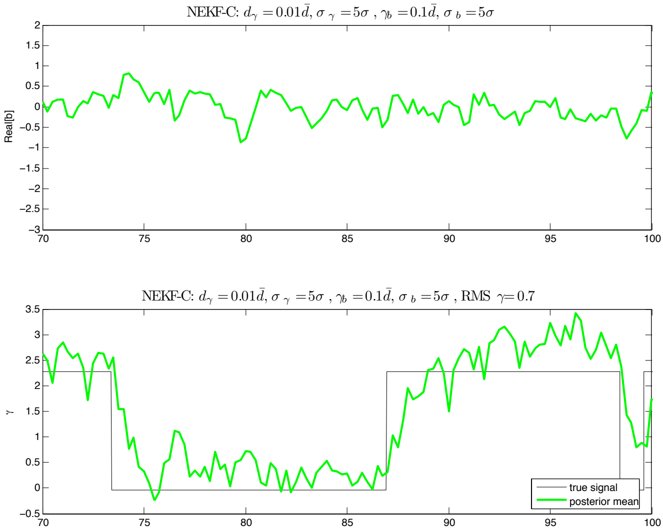
Nature Signals for Unforced and Forced cases



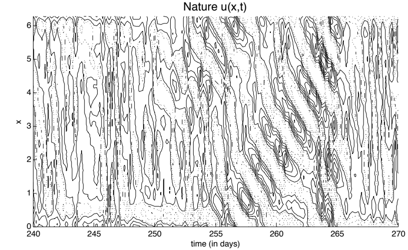
One mode demonstration of the filtered solution

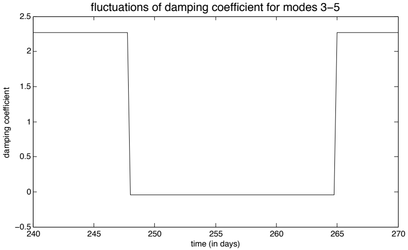


One mode demonstration of the filtered solution

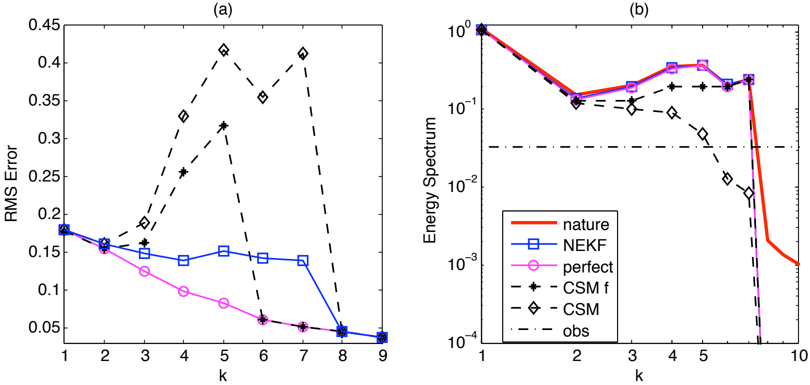


Turbulent system of externally forced barotropic Rossby wave equation with instability through intermittent of negative damping.





Incorrectly specified forcings, observed only 15 observations of 105 grid points

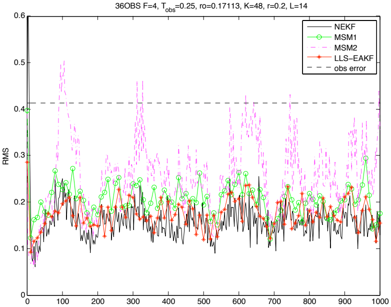


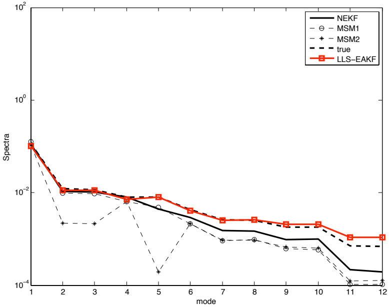
References:

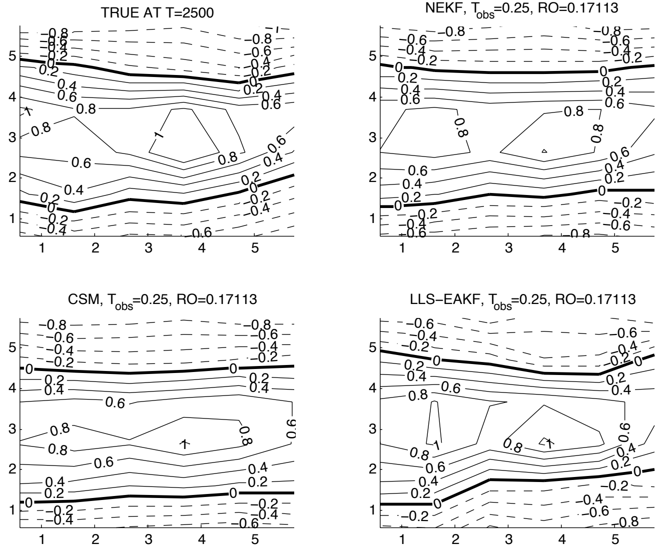
1. Castronovo, Harlim, and Majda, "Mathematical test criteria for filtering complex systems: Plentiful observations", J. Comp. Phys., 227(7), 3678-3714, 2008.
2. Harlim and Majda, "Mathematical test criteria for filtering complex systems: Regularly sparse observations", J. Comp. Phys., 227(10), 5304-5341, 2008.
3. Harlim and Majda, "Filtering nonlinear dynamical systems with linear stochastic models", Nonlinearity 21(6), 1281-1306, 2008.
4. Harlim and Majda, "Catastrophic filter divergence in filtering nonlinear dissipative systems", to appear in Comm. Math. Sci., 2009.
5. Gershgorin, Harlim, and Majda, "Test models for improving filtering with model errors through stochastic parameter estimation", submitted to J. Comp Phys, 2009.
6. Gershgorin, Harlim, and Majda, "Improving Filtering and Prediction of Spatially Extended Turbulent Systems with Model Errors through Stochastic Parameter Estimation", submitted to J. Comp Phys, 2009.
7. Majda and Harlim, "Systematic Strategies for Real Time Filtering of Turbulent Signals in Complex Systems", in preparation (2010).

Latest results on a two-layer QG model (observe 36 uniformly distributed grid points)

Harlim and Majda, "Filtering Turbulent Sparsely Observed Geophysical Flows", submitted to Monthly Weather Review, 2009.






TRUE AT T=2500
NEKF, T_{obs}=0.25, RO=0.17113
CSM, T_{obs}=0.25, RO=0.17113
LLS−EAKF, T_{obs}=0.25, RO=0.17113