Assimilating Irregularly Spaced, Sparse Observations with Hierarchical Bayesian Reduced Stochastic Filters

John Harlim[1]

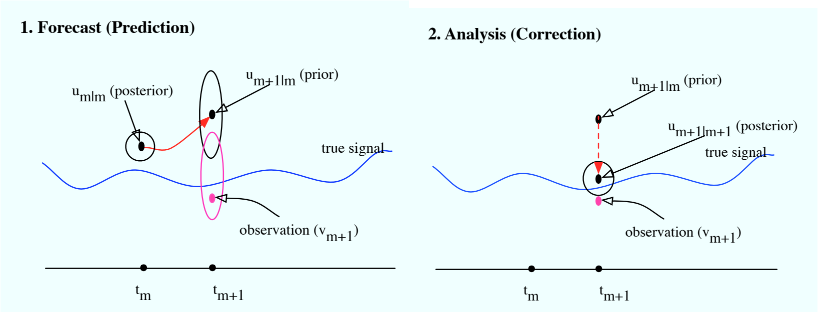
Kristen A. Brown

Department of Mathematics
North Carolina State University

Jan 31, 2013

[1]Supports: ONR N00014-11-1-1310, ONR MURI N00014-12-1-0912, NC State startup and FRPD funds.

Filtering/Data Assimilation



$$
\begin{aligned}
u_{m+1} &= \tilde{f}(u_m) + \epsilon_{m+1}, \quad \epsilon_m \sim \mathcal{N}(0, Q), \\
\tilde{v}_m &= g(u_m) + \tilde{\epsilon}_m, \quad \tilde{\epsilon}_m \sim \mathcal{N}(0, R)
\end{aligned}
$$

Filtering/Data Assimilation

Practically, filtering proceeds as follows:

- **Prediction step:** Given posterior statistical estimates $\bar{u}^+_{m-1} \equiv \mathbb{E}(U_{m-1}), C^+_{m-1} \equiv Cov(U_{m-1})$, compute prior statistical estimates $\bar{u}^-_m, C^-_m$ with

  $$u^-_m = \tilde{f}(u^+_{m-1}) + \epsilon_m, \quad \epsilon_m \sim \mathcal{N}(0, Q).$$

  Probabilistically, this can be generalize to

  $$p(u_{m-1}|\tilde{v}_{m-1}) \xrightarrow{\tilde{f}} p(u_m|\tilde{v}_{m-1}) \equiv p(u_m).$$

Filtering/Data Assimilation

Practically, filtering proceeds as follows:

- **Prediction step:** Given posterior statistical estimates $\bar{u}^+_{m-1} \equiv \mathbb{E}(U_{m-1})$, $C^+_{m-1} \equiv Cov(U_{m-1})$, compute prior statistical estimates $\bar{u}^-_m$, C^-_m with

$$u^-_m = \tilde{f}(u^+_{m-1}) + \epsilon_m, \quad \epsilon_m \sim \mathcal{N}(0, Q).$$

  Probabilistically, this can be generalize to

$$p(u_{m-1}|\tilde{v}_{m-1}) \xrightarrow{\tilde{f}} p(u_m|\tilde{v}_{m-1}) \equiv p(u_m).$$

- **Correction Step:** Given discrete-time data $\tilde{v}_m$, apply Bayes' theorem to obtain a posterior statistical estimates $\bar{u}^+_m$, C^+_m of

$$\begin{aligned} p(u_m|\tilde{v}_m) &\propto p(u_m)p(\tilde{v}_m|u_m) \\ &\propto \exp(-\frac{1}{2}\|u_m - \bar{u}^-_m\|^2_{C^{-1}_m} - \frac{1}{2}\|\tilde{v}_m - g(u_m)\|^2_{R^{-1}}), \end{aligned}$$

  where

$$\tilde{v}_m = g(u_m) + \tilde{\epsilon}_m, \quad \tilde{\epsilon}_m \sim \mathcal{N}(0, R).$$

Motivation

- Observations from nature are often noisy, measured at irregularly spaced and sparse locations, e.g., radiosondes in the ocean. On the other hand, the typical predictor model is resolved on regularly spaced grid points.

Motivation

- Observations from nature are often noisy, measured at irregularly spaced and sparse locations, e.g., radiosondes in the ocean. On the other hand, the typical predictor model is resolved on regularly spaced grid points.
- The typical approach for filtering such a data set is based on observation model,

$$\tilde{v}_m = g(\vec{u}_m) + \tilde{\epsilon}_m, \quad \tilde{\epsilon}_m \sim \mathcal{N}(0, r^o),$$

  where g is an observation operator that includes a change of variables as well as interpolation. Here, g maps the model state $\vec{u}$ to irregularly located observations $\tilde{v}$.

Motivation

- Observations from nature are often noisy, measured at irregularly spaced and sparse locations, e.g., radiosondes in the ocean. On the other hand, the typical predictor model is resolved on regularly spaced grid points.
- The typical approach for filtering such a data set is based on observation model,

$$\tilde{v}_m = g(\vec{u}_m) + \tilde{\epsilon}_m, \quad \tilde{\epsilon}_m \sim \mathcal{N}(0, r^o),$$

  where g is an observation operator that includes a change of variables as well as interpolation. Here, g maps the model state $\vec{u}$ to irregularly located observations $\tilde{v}$.
- This approach assumes that $\tilde{v}$ are raw data.

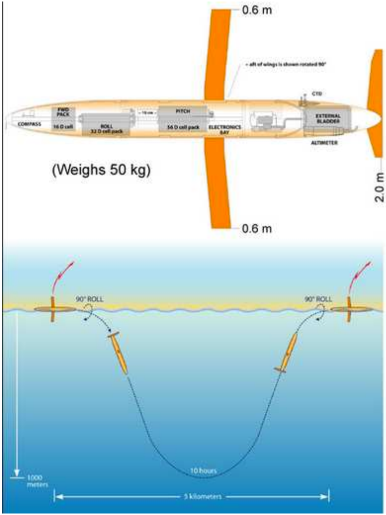
Motivation

- Observations from nature are often noisy, measured at irregularly spaced and sparse locations, e.g., radiosondes in the ocean. On the other hand, the typical predictor model is resolved on regularly spaced grid points.
- The typical approach for filtering such a data set is based on observation model,

$$\tilde{v}_m = g(\vec{u}_m) + \tilde{\epsilon}_m, \quad \tilde{\epsilon}_m \sim \mathcal{N}(0, r^o),$$

  where g is an observation operator that includes a change of variables as well as interpolation. Here, g maps the model state $\vec{u}$ to irregularly located observations $\tilde{v}$.
- This approach assumes that $\tilde{v}$ are raw data.
- We often use the same approach to assimilate processed data, ignoring the uncertainties associated with the processing scheme.

Motivation: Glider instrument



Motivation

- Processing data is often unavoidable since raw data can be very noisy and uninformative. The processing scheme varies from simple interpolations to inversion of complicated nonlinear differential equations. e.g. cloud clearing algorithm in satellite data [Chahine 1973].

Motivation

- Processing data is often unavoidable since raw data can be very noisy and uninformative. The processing scheme varies from simple interpolations to inversion of complicated nonlinear differential equations. e.g. cloud clearing algorithm in satellite data [Chahine 1973].
- The goal of this talk is to understand the effect of filtering processed data (from various interpolation schemes) in the presence of model errors.

Motivation

- Processing data is often unavoidable since raw data can be very noisy and uninformative. The processing scheme varies from simple interpolations to inversion of complicated nonlinear differential equations. e.g. cloud clearing algorithm in satellite data [Chahine 1973].
- The goal of this talk is to understand the effect of filtering processed data (from various interpolation schemes) in the presence of model errors.
- First, we will assess various standard deterministic interpolation schemes on a simple 1D test model to mimic turbulence. Second, we will discuss a statistical interpolator scheme, applied on a stiff 2D QG model.

Hierarchical Bayesian framework

Consider v_m to be the interpolated observations at the regular model grid points. Filtering interpolated (processed) data can be interpreted as follows:

$$p(u_m|v_m, \tilde{v}_m) \quad \propto \quad p(u_m)p(\tilde{v}_m|v_m, u_m)p(v_m|u_m)$$

Hierarchical Bayesian framework

Consider v_m to be the interpolated observations at the regular model grid points. Filtering interpolated (processed) data can be interpreted as follows:

$$
\begin{aligned}
p(u_m|v_m, \tilde{v}_m) &\propto p(u_m)p(\tilde{v}_m|v_m, u_m)p(v_m|u_m) \\
&\propto p(u_m)p(v_m|u_m, \tilde{v}_m)
\end{aligned}
$$

Hierarchical Bayesian framework

Consider v_m to be the interpolated observations at the regular model grid points. Filtering interpolated (processed) data can be interpreted as follows:

$$
\begin{aligned}
p(u_m|v_m,\tilde{v}_m) &\propto p(u_m)p(\tilde{v}_m|v_m,u_m)p(v_m|u_m) \\
&\propto p(u_m)p(v_m|u_m,\tilde{v}_m)
\end{aligned}
$$

In the presence of model errors, we filter with a surrogate prior model:

$$
p(u_m) \approx \tilde{p}(u_m|\theta)
$$

Hierarchical Bayesian framework

Consider v_m to be the interpolated observations at the regular model grid points. Filtering interpolated (processed) data can be interpreted as follows:

$$\begin{aligned} p(u_m|v_m,\tilde{v}_m) &\propto p(u_m)p(\tilde{v}_m|v_m,u_m)p(v_m|u_m) \\ &\propto p(u_m)p(v_m|u_m,\tilde{v}_m) \end{aligned}$$

In the presence of model errors, we filter with a surrogate prior model:

$$p(u_m) \approx \tilde{p}(u_m|\theta)$$

There are various approaches for choosing surrogate prior. For this talk, we will consider a simple, "Mean Stochastic Model", whose designed is based on various turbulent closure approximations.

One-dimensional test model

We consider a linear SPDE

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} + f(x,t) - \Gamma u + \sigma(x)\dot{W}(t).$$

where, the damping and adding terms are added to mimic rapid energy transfer due to nonlinearity [Majda & H 2012].

One-dimensional test model

We consider a linear SPDE

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} + f(x,t) - \Gamma u + \sigma(x)\dot{W}(t).$$

where, the damping and adding terms are added to mimic rapid energy transfer due to nonlinearity [Majda & H 2012].

We consider Γ to alternate between damping and anti-damping to simulate intermittent instability. Numerically, we model the eigensolution of Γ for the first 20 resolved modes with a two-state Markov jump process from the following sample spaces:

modes(k)	stable	unstable
1-10	$\gamma_w = 1.3$	$\gamma_s = 1.6$
11-20	$\gamma^+ = 2.27$	$\gamma^- = -0.04$
21-220	$\bar{\gamma} = 1.5$	$\bar{\gamma} = 1.5$

One-dimensional test model

We consider a linear SPDE

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} + f(x,t) - \Gamma u + \sigma(x)\dot{W}(t).$$

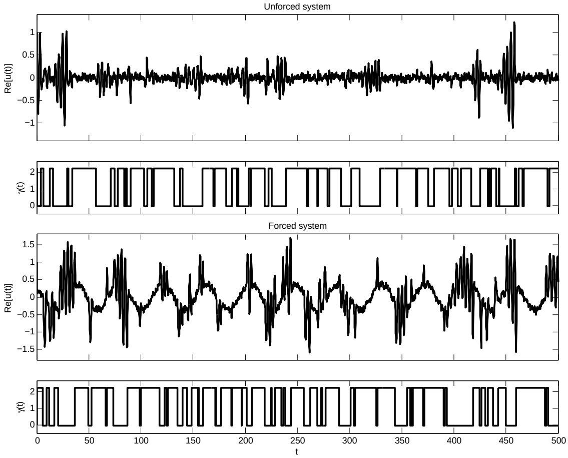
where, the damping and adding terms are added to mimic rapid energy transfer due to nonlinearity [Majda & H 2012].

We consider Γ to alternate between damping and anti-damping to simulate intermittent instability. Numerically, we model the eigensolution of Γ for the first 20 resolved modes with a two-state Markov jump process from the following sample spaces:

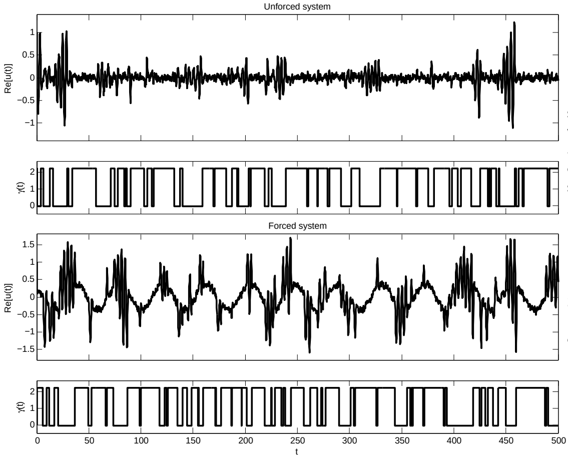
modes(k)	stable	unstable
1-10	$\gamma_w = 1.3$	$\gamma_s = 1.6$
11-20	$\gamma^+ = 2.27$	$\gamma^- = -0.04$
21-220	$\bar{\gamma} = 1.5$	$\bar{\gamma} = 1.5$

The noise amplitude, σ, is chosen such that the system has a statistical steady state for $f = 0$.

Intermittent instability



Intermittent instability



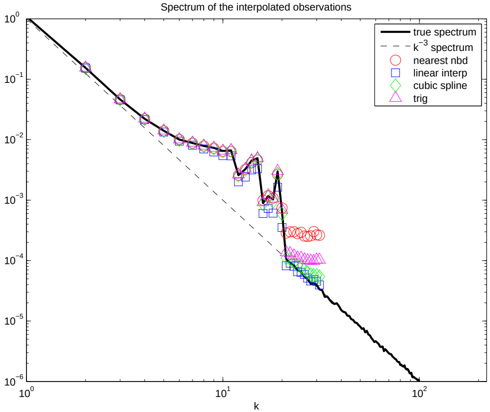
Mean Stochastic Model: A simple prior model for filtering with model errors is to use average damping $\bar{\gamma}$ that can be extracted from the equilibrium statistics. That is,

$$\frac{du_k}{dt} = (-\bar{\gamma}_k + \mathrm{i}\omega)u_k + f_k + \sigma_k \dot{W}_k$$

Such a poor man's model is motivated by the turbulent closure approach [Del Sole 2004, Majda & Timofeyev 2004].

Effect of interpolation on the energy spectrum

Interpolated spectrum from irregularly spaced observations with noise variance $r^o = 2 \times 10^{-5} < M^{-3}$ to the model resolved grid points.



Effect of interpolation on error covariance matrix

- The interpolation **operator** considered here is **linear**, so the interpolated noises are Gaussian. This may not be true in general.

Effect of interpolation on error covariance matrix

- The interpolation **operator** considered here is **linear**, so the interpolated noises are Gaussian. This may not be true in general.
- if one interpolates observations with i.i.d. noises, then the interpolated error covariance matrix is **not** diagonal.

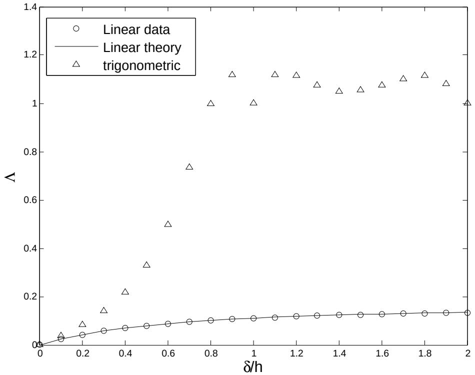
Effect of interpolation on error covariance matrix

- The interpolation **operator** considered here is **linear**, so the interpolated noises are Gaussian. This may not be true in general.
- if one interpolates observations with i.i.d. noises, then the interpolated error covariance matrix is **not** diagonal.
- For piecewise linear interpolated noises, we have:

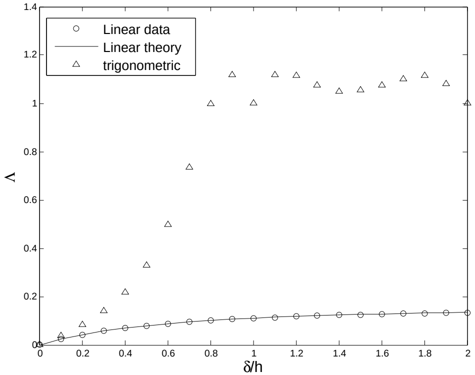
Proposition. Let $\{\sigma_j = \sigma(x_j)\}_{j=0}^{2M}$ be i.i.d. noises with variance r^o at regularly spaced grid points. Let us perturb a single observation site $\tilde{x}_j$ by δ, i.e., $\tilde{x}_j = x_j + \delta$. Then the ratio between the largest off-diagonal term and the smallest diagonal term is,

$$\Lambda \equiv \frac{\max_{k \neq k'} |R^o_{k,k'}|}{\min_k |R^o_{k,k}|} \leq \frac{2(\delta^2 + 2\delta h)}{(2M+1)(\delta + h)^2 - 2(\delta^2 + 2\delta h)}.$$

Effect of interpolation on error covariance matrix



Effect of interpolation on error covariance matrix



For general linear interpolation schemes, we can compute Λ explicitly but the computation is rather involved.

Fourier Domain Kalman Filter (FDKF)

For the second step in the hierachical Bayesian framework, we consider the approximate diagonal filter

$$
\begin{aligned}
du_k &= (-\bar{\gamma}_k + \mathrm{i}\omega)u_k\,dt + f_k\,dt + \sigma_k\,dW_k, \quad |k| \le M, \\
v_{k,m} &= u_k(t_m) + \sigma^o_{k,m}, \quad \sigma^o_{k,m} \sim \mathcal{N}(0, C^o)
\end{aligned}
$$

where $v_{k,m} = \mathbb{E}(V_{k,m}|\tilde{v}_m)$ and $C^o = Cov(V_{k,m}|\tilde{v}_m)$ are the first and second moments of the random variable for the processed data, $V_{k,m}$, given raw data $\tilde{v}_m$.

Fourier Domain Kalman Filter (FDKF)

For the second step in the hierachical Bayesian framework, we consider the approximate diagonal filter

$$\begin{aligned} du_k &= (-\bar{\gamma}_k + \mathrm{i}\omega) u_k \, dt + f_k \, dt + \sigma_k \, dW_k, \quad |k| \le M, \\ v_{k,m} &= u_k(t_m) + \sigma^o_{k,m}, \quad \sigma^o_{k,m} \sim \mathcal{N}(0, C^o) \end{aligned}$$

where $v_{k,m} = \mathbb{E}(V_{k,m} | \tilde{v}_m)$ and $C^o = Cov(V_{k,m} | \tilde{v}_m)$ are the first and second moments of the random variable for the processed data, $V_{k,m}$, given raw data $\tilde{v}_m$.

- If Λ is small, we can ignore the cross-covariance terms in C^o and consider diagonal FDKF.

Fourier Domain Kalman Filter (FDKF)

For the second step in the hierachical Bayesian framework, we consider the approximate diagonal filter

$$\begin{aligned} du_k &= (-\bar{\gamma}_k + \mathrm{i}\omega)u_k\,dt + f_k\,dt + \sigma_k\,dW_k, \quad |k| \leq M, \\ v_{k,m} &= u_k(t_m) + \sigma^o_{k,m}, \quad \sigma^o_{k,m} \sim \mathcal{N}(0, C^o) \end{aligned}$$

where $v_{k,m} = \mathbb{E}(V_{k,m}|\tilde{v}_m)$ and $C^o = Cov(V_{k,m}|\tilde{v}_m)$ are the first and second moments of the random variable for the processed data, $V_{k,m}$, given raw data $\tilde{v}_m$.

- If Λ is small, we can ignore the cross-covariance terms in C^o and consider diagonal FDKF.
- Alternatively, one can consider observation model,

$$\tilde{v}_m = g(u_k(t_m)) + \tilde{\varepsilon}_m \quad \tilde{\varepsilon}_m \in \mathcal{N}(0, r^o),$$

  and perform the coupled filtering problem in Fourier domain with an appropriate g.

Effect of interpolation on the filtered solutions

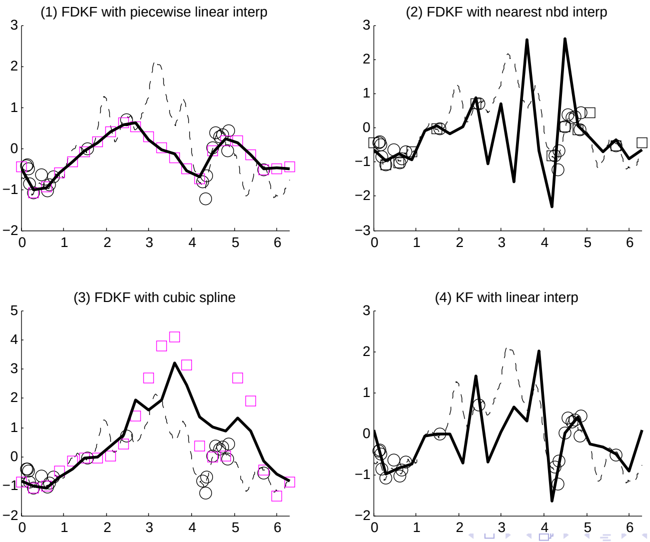
Table: Weakly irregularly spaced observations: Average RMS errors and spatial correlation for numerical experiments with sparse $2M+1=21$ observations and observation noise error 0.4583 with variance spectrum $\hat{r}^o = 0.01 > E_5 = 5^{-3}$.

Schemes	RMS error	PCorr	Λ
1. FDKF with piecewise linear interp	0.3835	0.91	0.16
2. FDKF with nearest nbd	0.4417	0.89	
3. FDKF with cubic spline	0.4184	0.88	0.64
4. Physical space KF with linear interp	0.5136	0.87	
5. Coupled FDKF with linear interp	0.4843	0.88	
6. Decoupled FDKF with linear interp	0.5089	0.87	
7. Coupled FDKF with trig interp	0.4618	0.89	1.08
8. Decoupled FDKF with trig interp	0.5010	0.85	1.08

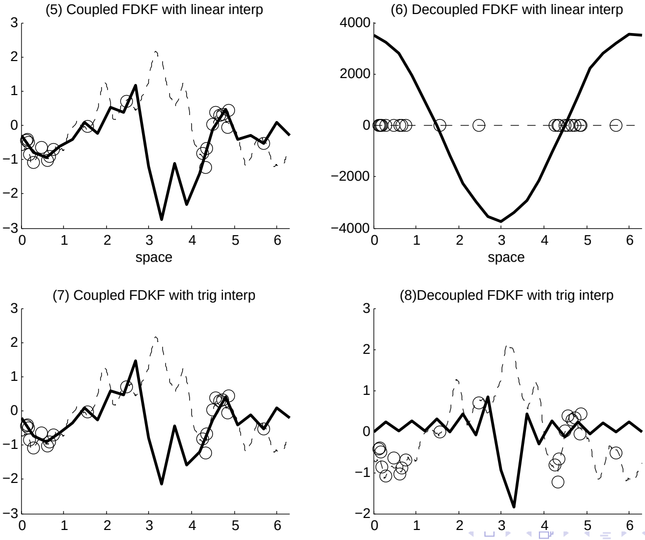
Table: Extremely irregularly spaced and sparse observations: Average RMS errors and spatial correlation for numerical experiments with sparse $2M+1=21$ observations and observation noise error $\hat{r}^o = 0.01$.

Schemes	RMS error	SCorr	Λ
1. FDKF with piecewise linear interp	0.6774	0.83	1.55
2. FDKF with nearest nbd	1.4507	0.61	
3. FDKF with cubic spline	1.0161	0.47	72.58
4. Physical space KF with linear interp	1.5488	0.57	
5. Coupled FDKF with linear interp	0.9160	0.78	
6. Decoupled FDKF with linear interp	3507.9	0	
7. Coupled FDKF with trig interp	0.9198	0.77	$\mathcal{O}(10^5)$
8. Decoupled FDKF with trig interp	1.7558	0	$\mathcal{O}(10^5)$

Effect of interpolation on filtered solutions: extremely irregularly spaced observations



Effect of interpolation on filtered solutions: extremely irregularly spaced observations



Summary from the 1D test model

- Processing data with interpolation will induce small scale oscillatory artifact with higher order interpolation schemes.

Summary from the 1D test model

- Processing data with interpolation will induce small scale oscillatory artifact with higher order interpolation schemes.
- We find that the FDKF with linear interpolation is more accurate than the higher order interpolation schemes or even the standard physical space Kalman filter. This result is quite surprising, considering that the approximate filter ignores the non-diagonal terms in the processed error covariance matrix.

Summary from the 1D test model

- Processing data with interpolation will induce small scale oscillatory artifact with higher order interpolation schemes.
- We find that the FDKF with linear interpolation is more accurate than the higher order interpolation schemes or even the standard physical space Kalman filter. This result is quite surprising, considering that the approximate filter ignores the non-diagonal terms in the processed error covariance matrix.
- How robust is the linear interpolation in 2D?

Summary from the 1D test model

- Processing data with interpolation will induce small scale oscillatory artifact with higher order interpolation schemes.
- We find that the FDKF with linear interpolation is more accurate than the higher order interpolation schemes or even the standard physical space Kalman filter. This result is quite surprising, considering that the approximate filter ignores the non-diagonal terms in the processed error covariance matrix.
- How robust is the linear interpolation in 2D?
- Next, we'll consider a popular spatial statistical interpolator, kriging, on a more realistic problem in 2D setup.

Two Layer Quasi-Geostrophic Model

The dynamical equations for the perturbed variables about uniform shear with stream function $\Psi_1 = -Uy, \Psi_2 = Uy$:

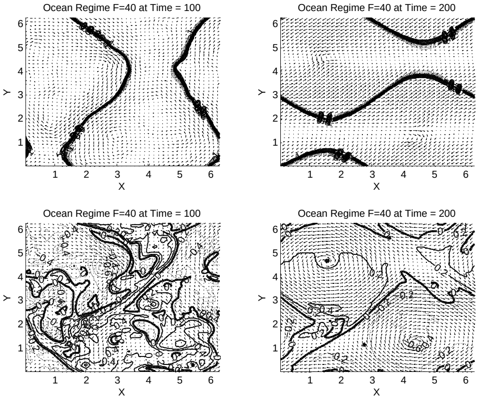
$$\begin{aligned} \frac{\partial q_1}{\partial t} + J(\psi_1, q_1) + U\frac{\partial q_1}{\partial x} + (\beta + k_d^2 U)\frac{\partial \psi_1}{\partial x} + \nu\nabla^8 q_1 &= 0 \\ \frac{\partial q_2}{\partial t} + J(\psi_2, q_2) - U\frac{\partial q_2}{\partial x} + (\beta - k_d^2 U)\frac{\partial \psi_2}{\partial x} + \nu\nabla^8 q_2 + \kappa\nabla^2\psi_2 &= 0 \end{aligned}$$

q_j is the quasi-geostrophic potential vorticity given as

$$q_j = \nabla^2\psi_j + \frac{k_d^2}{2}(\psi_{3-j} - \psi_j), \quad j = 1, 2,$$

with $\vec{u} = \nabla^{\perp}\psi$, $k_d = \sqrt{8}/L_d$ [see Smith et al. 2002].
We'll consider numerically stiff regime $F = 1/L_d = 40$ [Kleeman & Majda 2005].

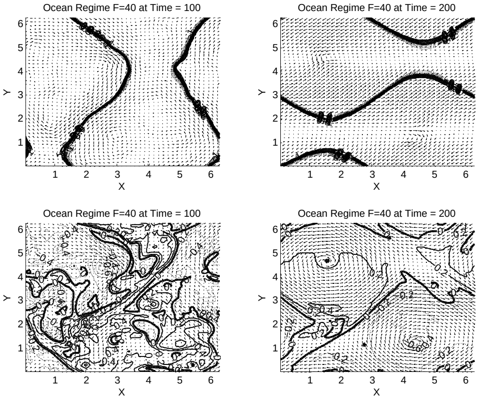
The two-layer QG model with baroclinic instability



Barotropic mode $\psi_b = \frac{\psi_1+\psi_2}{2}$ (top), baroclinic mode $\psi_c = \frac{\psi_1-\psi_2}{2}$ (bottom)

Resolve on 128×128 grid points

The two-layer QG model with baroclinic instability

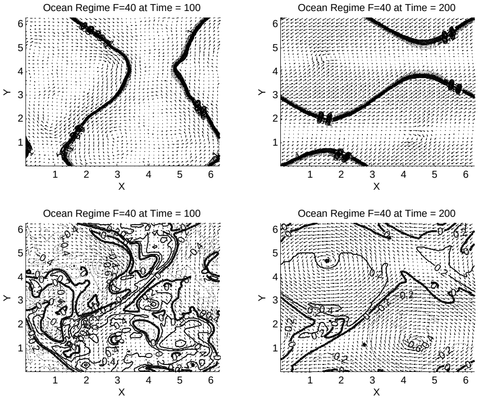


Barotropic mode $\psi_b = \frac{\psi_1+\psi_2}{2}$ (top), baroclinic mode $\psi_c = \frac{\psi_1-\psi_2}{2}$ (bottom)

Resolve on 128×128 grid points

In this regime, we'll have two competing behavior (zonal jet mode vs Rossby mode blocking).

The two-layer QG model with baroclinic instability



Barotropic mode $\psi_b = \frac{\psi_1 + \psi_2}{2}$ (top), baroclinic mode $\psi_c = \frac{\psi_1 - \psi_2}{2}$ (bottom)

Resolve on 128×128 grid points

In this regime, we'll have two competing behavior (zonal jet mode vs Rossby mode blocking).

We'll consider sparse observations of only the barotropic streamfunction.

Stochastic Models for Filtering the barotropic mode

Recall that

$$\frac{\partial q_b}{\partial t} + J(\psi_b, q_b) + \beta \frac{\partial \psi_b}{\partial x} + \kappa \nabla^2 \psi_b + \nu \nabla^8 q_b + s(\psi_c, q_c) = 0$$

where $q_b = \nabla^2 \psi_b$.

Stochastic Models for Filtering the barotropic mode

Recall that

$$\frac{\partial q_b}{\partial t} + J(\psi_b, q_b) + \beta \frac{\partial \psi_b}{\partial x} + \kappa \nabla^2 \psi_b + \nu \nabla^8 q_b + s(\psi_c, q_c) = 0$$

where $q_b = \nabla^2 \psi_b$.
Define $\psi(x, y, t) = \sum_{k,\ell} \hat{\psi}_{k,\ell}(t) e^{\mathrm{i}(kx+\ell y)}$ such that each horizontal mode has the following form

$$d\hat{\psi}(t) = (-d + \mathrm{i}\omega)\hat{\psi}(t)dt + \hat{f}(t)dt + \text{NL terms}$$

Stochastic Models for Filtering the barotropic mode

Recall that

$$\frac{\partial q_b}{\partial t} + J(\psi_b, q_b) + \beta\frac{\partial \psi_b}{\partial x} + \kappa\nabla^2\psi_b + \nu\nabla^8 q_b + s(\psi_c, q_c) = 0$$

where $q_b = \nabla^2\psi_b$.
Define $\psi(x,y,t) = \sum_{k,\ell}\hat{\psi}_{k,\ell}(t)e^{\mathrm{i}(kx+\ell y)}$ such that each horizontal mode has the following form

$$d\hat{\psi}(t) = (-d + \mathrm{i}\omega)\hat{\psi}(t)dt + \hat{f}(t)dt + \text{NL terms}$$

Replace the nonlinear terms and all of the baroclinic components by Ornstein-Uhlenbeck processes. That is,

$$d\hat{\psi}(t) = (-d + \mathrm{i}\omega)\hat{\psi}(t)dt + \hat{f}(t)dt + \sigma dW(t)$$

and our task is to parameterize d, ω, σ.

Stochastic Models for Filtering the barotropic mode

Recall that

$$\frac{\partial q_b}{\partial t} + J(\psi_b, q_b) + \beta \frac{\partial \psi_b}{\partial x} + \kappa \nabla^2 \psi_b + \nu \nabla^8 q_b + s(\psi_c, q_c) = 0$$

where $q_b = \nabla^2 \psi_b$.
Define $\psi(x, y, t) = \sum_{k,\ell} \hat{\psi}_{k,\ell}(t) e^{\mathrm{i}(kx+\ell y)}$ such that each horizontal mode has the following form

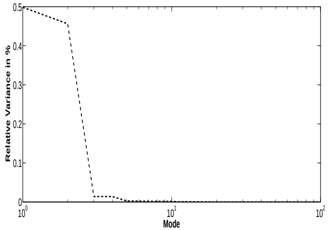
$$d\hat{\psi}(t) = (-d + \mathrm{i}\omega)\hat{\psi}(t)dt + \hat{f}(t)dt + \text{NL terms}$$

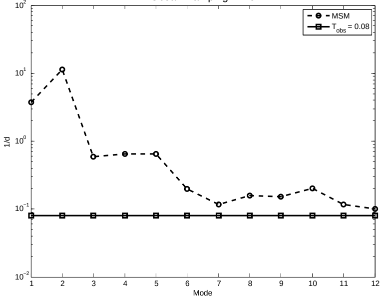
Replace the nonlinear terms and all of the baroclinic components by Ornstein-Uhlenbeck processes. That is,

$$d\hat{\psi}(t) = (-d + \mathrm{i}\omega)\hat{\psi}(t)dt + \hat{f}(t)dt + \sigma dW(t)$$

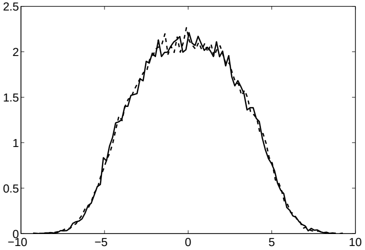
and our task is to parameterize d, ω, σ.
Offline Parameterization Strategy: Regression to empirical statistics from a long time series (Mean Stochastic Models).

Equilibrium Statistics

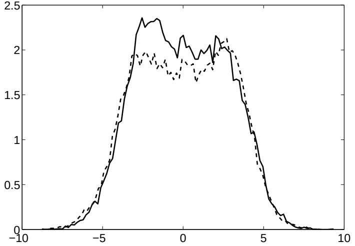




OCEAN MODE 1: (1,0)

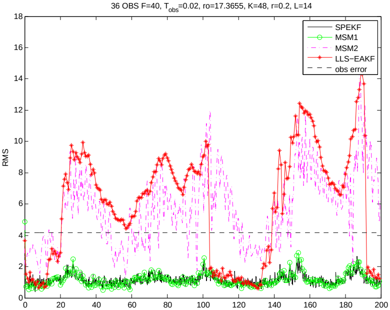


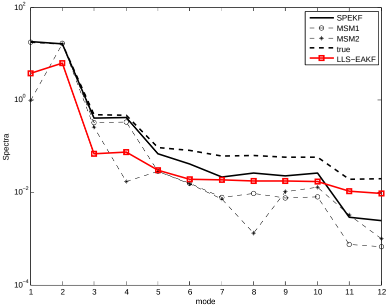
OCEAN MODE 2: (0,1)



Regularly Spaced Sparse Observations [H & Majda 2009]

RMS errors and energy spectrum recovery on regularly spaced 36 observations. Here, we also consider different nonlinear filter SPEKF [Gershgorin, H & Majda 2010].

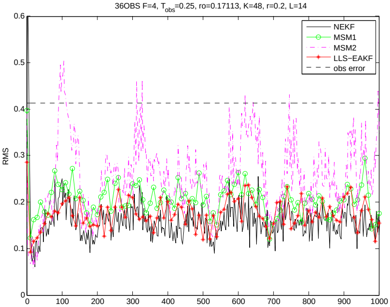


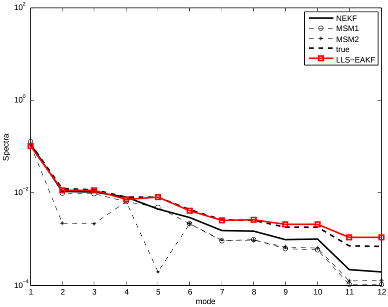


The MSM is better than Ensemble Kalman Filter with perfect model!

Regularly Spaced Sparse Obs.

Numerically less stiff with $F = 4$ (larger radius of deformation)!





Ordinary Kriging [Cressie 1993]

- Kriging is a maximum likelihood estimator of a random field modeled by a stationary Gaussian process,

$$Z(s) = \mu(s) + \delta(s), \quad \delta(s) \sim \mathcal{N}(0, C(s, s)).$$

Ordinary Kriging [Cressie 1993]

- Kriging is a maximum likelihood estimator of a random field modeled by a stationary Gaussian process,

$$Z(s) = \mu(s) + \delta(s), \quad \delta(s) \sim \mathcal{N}(0, C(s,s)).$$

- The main idea is to fit observations

$$\vec{Z} = [Z(\vec{x}_1), \ldots, Z(\vec{x}_M)]^T$$

to an empirically chosen isotropic parametric function $C(\vec{x}, \vec{y}) = C(\|\vec{x} - \vec{y}\|; \theta)$. Computationally, it involves solving low-dimensional nonlinear optimization problem, "on-the-fly", so, kriging is data-driven.

Ordinary Kriging [Cressie 1993]

- Kriging is a maximum likelihood estimator of a random field modeled by a stationary Gaussian process,

$$Z(s) = \mu(s) + \delta(s), \quad \delta(s) \sim \mathcal{N}(0, C(s,s)).$$

- The main idea is to fit observations

$$\vec{Z} = [Z(\vec{x}_1), \ldots, Z(\vec{x}_M)]^T$$

  to an empirically chosen isotropic parametric function $C(\vec{x}, \vec{y}) = C(\|\vec{x} - \vec{y}\|; \theta)$. Computationally, it involves solving low-dimensional nonlinear optimization problem, "on-the-fly", so, kriging is data-driven.

- Ordinary kriging assumes $\mu(s)$ is locally constant. The estimator for Z at grid point $\vec{x}$ is given by $\mathbb{E}(Z(\vec{x})|\vec{Z})$ with uncertainties characterized by $Cov(Z(\vec{x}), Z(\vec{x}')|\vec{Z})$.

Implementation

- Use the variogram, $2\gamma(\vec{x}-\vec{y}) \equiv Var(Z(\vec{x})-Z(\vec{y}))$ to obtain

$$C(\vec{x},\vec{y}) = C(\|\vec{x}-\vec{y}\|) = r^o - \gamma(\vec{x}-\vec{y}).$$

Implementation

- Use the variogram, $2\gamma(\vec{x}-\vec{y}) \equiv Var(Z(\vec{x})-Z(\vec{y}))$ to obtain

$$C(\vec{x},\vec{y}) = C(\|\vec{x}-\vec{y}\|) = r^o - \gamma(\vec{x}-\vec{y}).$$

- Use the available observations to construct

$$2\hat{\gamma}(r) = \frac{1}{|N(r)|}\sum_{i,j\in N(r)}\left(\hat{\delta}(\vec{x}_i)-\hat{\delta}(\vec{x}_j)\right)^2,$$

  where $N(r) \equiv \{i,j : \|\vec{x}_i - \vec{x}_j\| \le r\}$ and $\hat{\delta}(\vec{x}) \equiv Z(\vec{x}) - \hat{\mu}(\vec{x})$ is the residual with estimator $\hat{\mu}(\vec{x})$ obtain via median polishing (this estimator is a local average value over the same rows/columns.).

Implementation

- Use the variogram, $2\gamma(\vec{x}-\vec{y}) \equiv Var(Z(\vec{x})-Z(\vec{y}))$ to obtain

$$C(\vec{x},\vec{y}) = C(\|\vec{x}-\vec{y}\|) = r^o - \gamma(\vec{x}-\vec{y}).$$

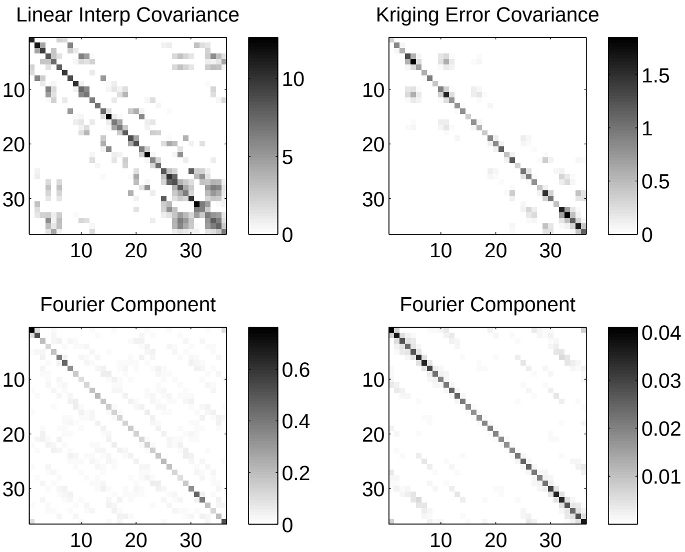
- Use the available observations to construct

$$2\hat{\gamma}(r) = \frac{1}{|N(r)|}\sum_{i,j\in N(r)} \left(\hat{\delta}(\vec{x}_i) - \hat{\delta}(\vec{x}_j)\right)^2,$$

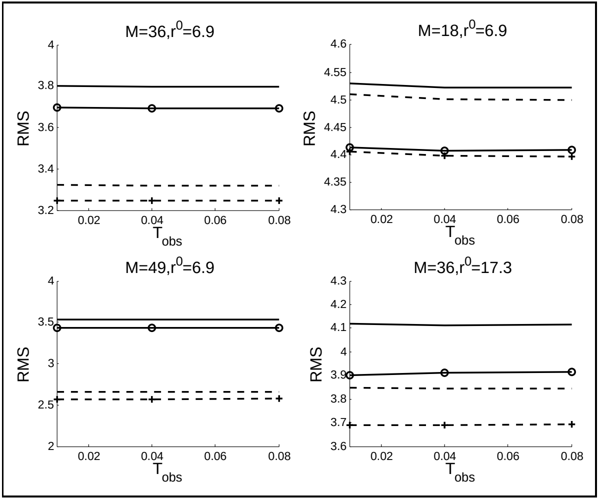
  where $N(r) \equiv \{i,j : \|\vec{x}_i - \vec{x}_j\| \leq r\}$ and $\hat{\delta}(\vec{x}) \equiv Z(\vec{x}) - \hat{\mu}(\vec{x})$ is the residual with estimator $\hat{\mu}(\vec{x})$ obtain via median polishing (this estimator is a local average value over the same rows/columns.).
- Fit $\hat{\gamma}$ to an appropriate parametric function. We choose

$$\hat{\gamma}^*(r) = \sigma^2 \exp(-\rho r), \quad r \geq 0.$$

Covariance Estimators



Filtering Skill (averaged RMS errors from 50 random choices of irregularly spaced observation networks)

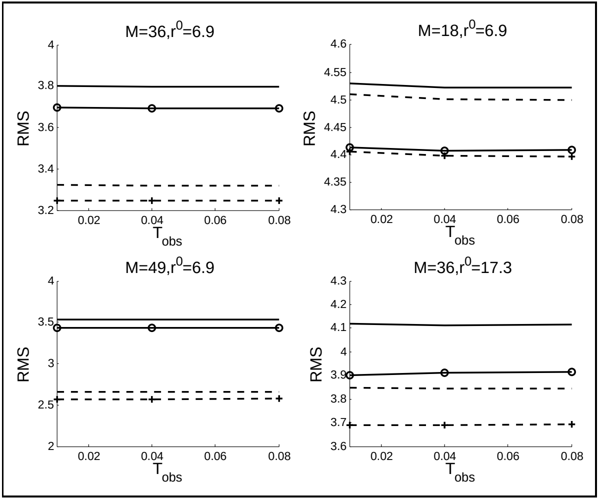


Interpolate to 6 $\times$ 6 grid points.

$r^o = 6.9 = 10\%E$
$r^o = 17.3 = 25\%E$

kriging (dashes),
filtered kriging (dashes with '+'),
linear interp (solid),
filtered linear interp (solid with 'o').

Filtering Skill (averaged RMS errors from 50 random choices of irregularly spaced observation networks)



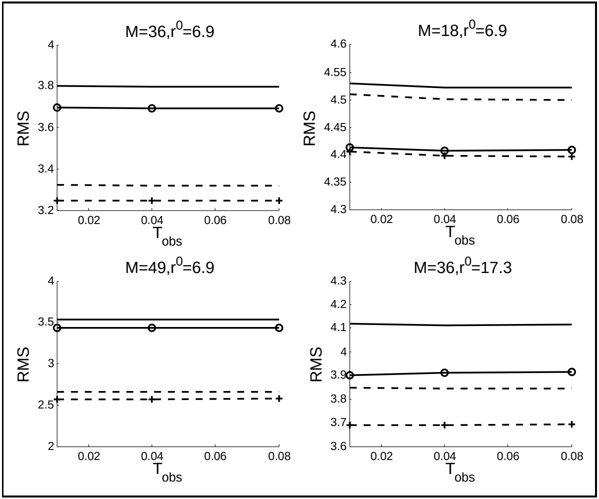
Interpolate to 6 $\times$ 6 grid points.

$r^o = 6.9 = 10\%E$
$r^o = 17.3 = 25\%E$

kriging (dashes),
filtered kriging (dashes with '+'),
linear interp (solid),
filtered linear interp (solid with 'o').

The filtering skill improved as observations are spatially denser.

Filtering Skill (averaged RMS errors from 50 random choices of irregularly spaced observation networks)



Interpolate to 6 $\times$ 6 grid points.

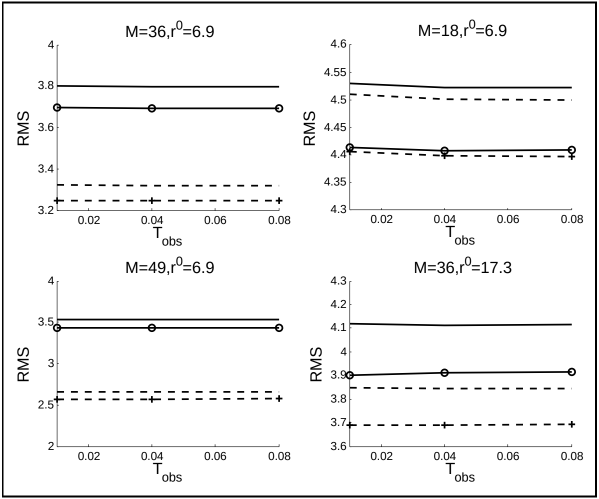
$r^o = 6.9 = 10\%E$
$r^o = 17.3 = 25\%E$

kriging (dashes),
filtered kriging (dashes with '$+$'),
linear interp (solid),
filtered linear interp (solid with 'o').

The filtering skill improved as observations are spatially denser.

Kriging and linear interpolation are comparable when observations are sparse.

Filtering Skill (averaged RMS errors from 50 random choices of irregularly spaced observation networks)



Interpolate to 6 $\times$ 6 grid points.

$r^o = 6.9 = 10\%E$
$r^o = 17.3 = 25\%E$

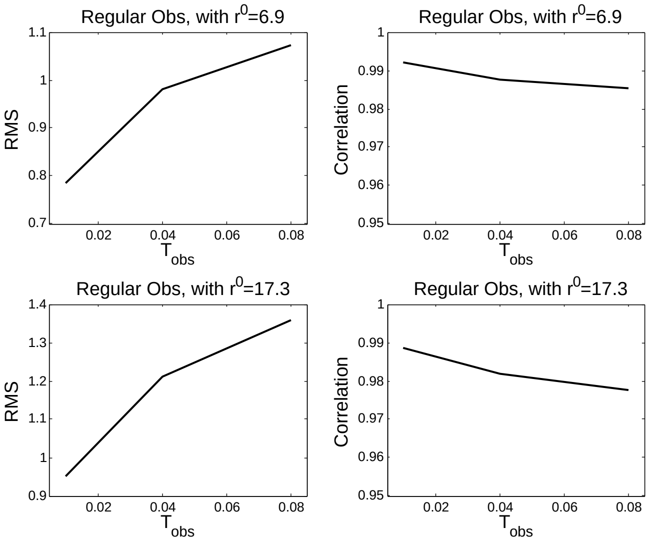
kriging (dashes),
filtered kriging (dashes with '+'),
linear interp (solid),
filtered linear interp (solid with 'o').

The filtering skill improved as observations are spatially denser.

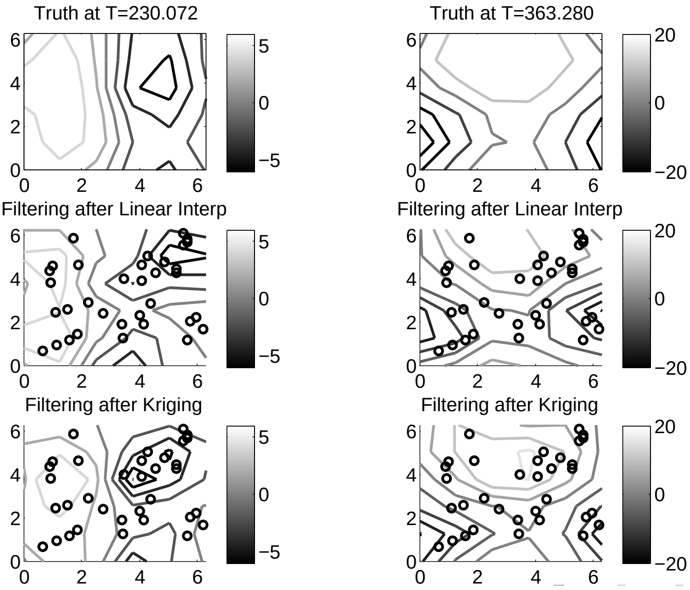
Kriging and linear interpolation are comparable when observations are sparse.

The dynamical constraints through the MSM become important when observation noise is large.

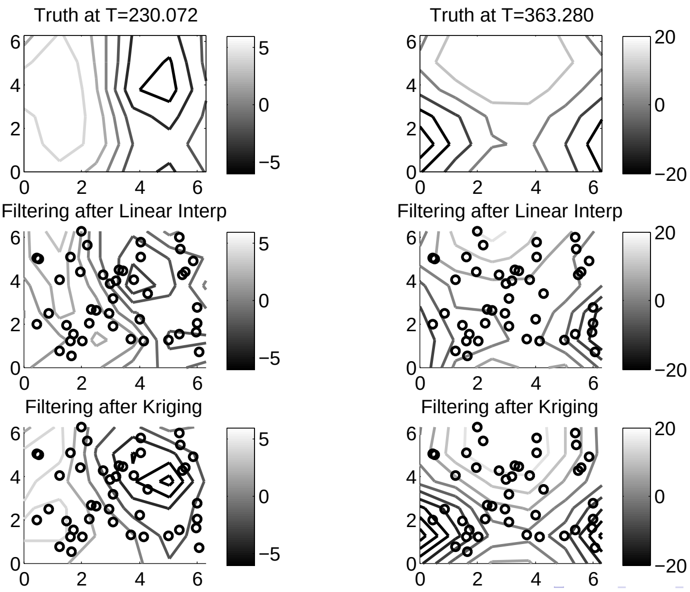
Filtering skill at regularly spaced 6×6 grid points



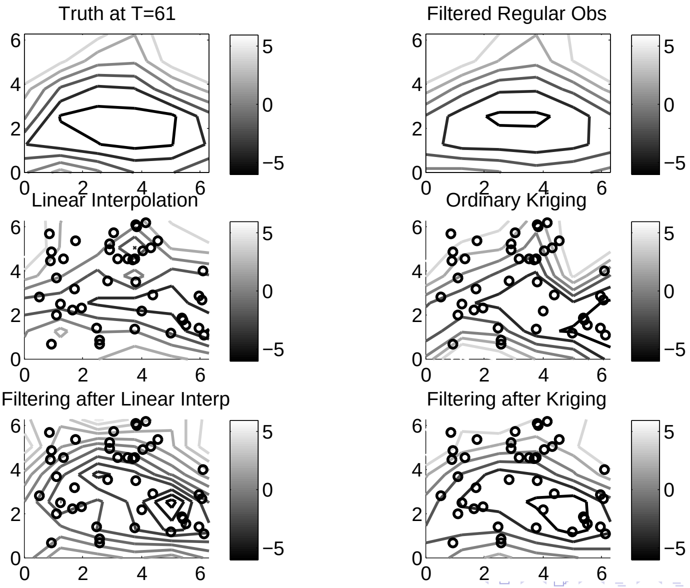
Case M=36, $r^o = 6.9 = 10\%E$



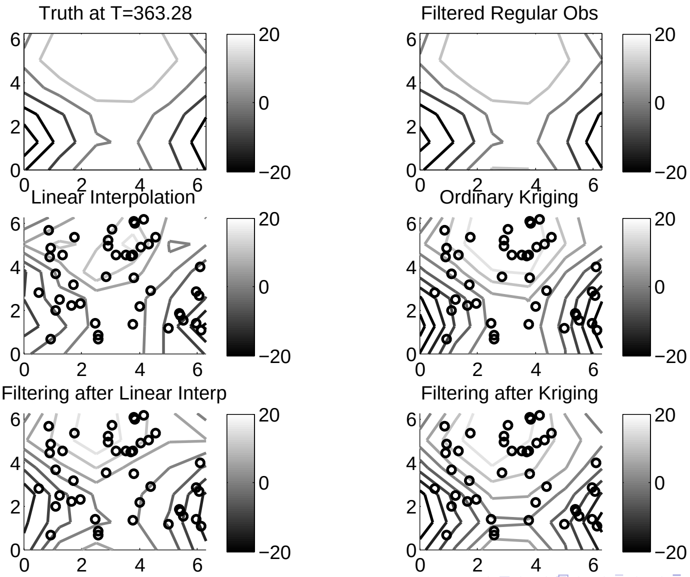
Case M=49, $r^o = 6.9 = 10\%E$



Case M=36, $r^o = 17.1 = 25\%E$



Case M=36, $r^o = 17.1 = 25\%E$



Summary from the 2D test model

- We consider filtering sparse observations of a numerically stiff two-layer QG model with baroclinic instability with a "poor man's" linear stochastic prior model, MSM.

Summary from the 2D test model

- We consider filtering sparse observations of a numerically stiff two-layer QG model with baroclinic instability with a "poor man's" linear stochastic prior model, MSM.
- For regularly spaced observations, MSM beats LLS-EAKF.

Summary from the 2D test model

- We consider filtering sparse observations of a numerically stiff two-layer QG model with baroclinic instability with a "poor man's" linear stochastic prior model, MSM.
- For regularly spaced observations, MSM beats LLS-EAKF.
- For irregularly spaced observations, MSM filtering processed data from kriging performs the best. However, the filtered accuracy is rather poor compared to that with regularly spaced observations.

Summary from the 2D test model

- We consider filtering sparse observations of a numerically stiff two-layer QG model with baroclinic instability with a "poor man's" linear stochastic prior model, MSM.
- For regularly spaced observations, MSM beats LLS-EAKF.
- For irregularly spaced observations, MSM filtering processed data from kriging performs the best. However, the filtered accuracy is rather poor compared to that with regularly spaced observations.
- The filtering skill of MSM can be improved with denser observation network. When observations are really sparse, the filtering skill of MSM on the processed data from linear interpolation and kriging are comparable.

Summary from the 2D test model

- We consider filtering sparse observations of a numerically stiff two-layer QG model with baroclinic instability with a "poor man's" linear stochastic prior model, MSM.
- For regularly spaced observations, MSM beats LLS-EAKF.
- For irregularly spaced observations, MSM filtering processed data from kriging performs the best. However, the filtered accuracy is rather poor compared to that with regularly spaced observations.
- The filtering skill of MSM can be improved with denser observation network. When observations are really sparse, the filtering skill of MSM on the processed data from linear interpolation and kriging are comparable.
- When observations are very noisy, the dynamical constraints through the MSM become more important.

Summary from the 2D test model

- We consider filtering sparse observations of a numerically stiff two-layer QG model with baroclinic instability with a "poor man's" linear stochastic prior model, MSM.
- For regularly spaced observations, MSM beats LLS-EAKF.
- For irregularly spaced observations, MSM filtering processed data from kriging performs the best. However, the filtered accuracy is rather poor compared to that with regularly spaced observations.
- The filtering skill of MSM can be improved with denser observation network. When observations are really sparse, the filtering skill of MSM on the processed data from linear interpolation and kriging are comparable.
- When observations are very noisy, the dynamical constraints through the MSM become more important.
- Obviously, there are many issues that can be improved in the future: Gaussianity, kriging type, resolutions, models, etc

References

Our papers are downloadable from
`http://www4.ncsu.edu/~jharlim/publications.htm`

- J. Harlim, "Interpolating irregularly spaced observations for filtering complex turbulent systems", SIAM J. Sci. Comput. 33(5), 2620-2640, 2011.
- K. Brown and J.Harlim, "Assimilating irregularly spaced sparsely observed turbulent signals with hierarchical Bayesian reduced stochastic filters", J. Comput. Phys. 235, 143-160, 2013.
- A.J. Majda and J. Harlim, "Filtering Complex Turbulent Systems", Cambridge University Press, 2012.
- N. Cressie, Statistics for "Spatial Data, Wiley-Interscience", 1993.