Superparameterization and Dynamic Stochastic Superresolution (DSS) for Filtering Sparse Geophysical Flows

Andrew J. Majda

June 2013

Outline

1. Filtering
   - Filtering: obtaining the best statistical estimation of a nature system from the partial observations.
   - Fourier Domain Kalman Filter (FDKF) with regularly spaced sparse observations.
2. Filtering with Superparameterization
   - linear, analytically solvable model,
   - model error coming from finite discrete approximations.
3. Filtering with Dynamic Stochastic Superresolution (DSS)
   - nonlinear model,
   - using cheap stochastic models to forecast the true nonlinear dynamics.

- **Test Models for Filtering with Superparameterization**, John Harlim and A. J. Majda, submitted, SIAM J. Multiscale Modeling and Simulation, September 9, 2012.
- **Dynamic Stochastic Superresolution of sparseley observed turbulent systems**, M. Branicki and A. J. Majda, submitted, Journal of Computational Physics, May 17, 2012.
- **New methods for estimating poleward eddy heat transport using satellite altimetry**, S. Keating, A. J. Majda and K. S. Smith, Monthly Weather Review, February 9, 2012.

Basic Notions of Filtering and Test Models for Filtering with Superparameterization

I. Filtering the Turbulent Signal

1.1. Kalman Filter

- True signal $\vec{u}_{m+1} \in \mathbb{R}^N$, which is generated from

$$\boxed{\vec{u}_{m+1} = F\vec{u}_m + \vec{\sigma}_{m+1}},$$

- Observation $\vec{v}_{m+1} \in \mathbb{R}^M$:

$$\boxed{\vec{v}_{m+1} = G\vec{u}_{m+1} + \vec{\sigma}^o_{m+1}},$$

where matrix $G \in \mathbb{R}^{M\times N}$ and $\vec{\sigma}^o_m = \{\sigma^o_{j,m}\}$ is an M-dimensional Gaussian while noise vector with zero mean and covariance

$$R^o = \langle \vec{\sigma}^o_m \otimes (\vec{\sigma}^o_m)^T \rangle = \{\langle \vec{\sigma}^o_{i,m} (\vec{\sigma}^o_{j,m})^T \rangle\} = \{\delta(i-j) r^o\}$$

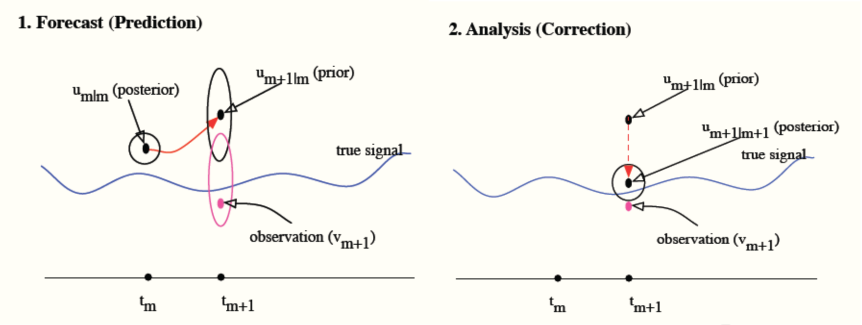
- Forecast model:

$$\boxed{\vec{u}^M_{m+1} = F^M \vec{u}^M_m + \vec{\sigma}^M_{m+1}},$$

where $F^M \in \mathbb{R}^{N\times N}$ and $\vec{\sigma}^M_m$ is an M-dimensional Gaussian while noise vector with zero mean and covariance

$$R^M = \langle \vec{\sigma}^M_m \otimes (\vec{\sigma}^M_m)^T \rangle.$$

Goal: Estimate the true state: $\vec{u}_{m+1} \in \mathbb{R}^N$ from the imperfect prediction model and the observations of the true signal.



Step 1. Forecast:
Run the forecast model from step m to $m+1$,

$$\vec{u}^M_{m+1|m} = F\vec{u}^M_{m|m} + \vec{\sigma}^M_{m+1}.$$

Compute the **prior** mean and covariance

$$\vec{u}^M_{m+1|m} = F^M \vec{u}^M_{m|m},$$
$$R^M_{m+1|m} = F^M R^M_{m|m} (F^M)^T + R^M.$$

Step 2. Analysis:
Compute **posterior** mean and variance

$$\vec{u}^M_{m+1|m+1} = \vec{u}^M_{m+1|m} + K_{m+1}(\vec{v}_{m+1} - G\vec{u}^M_{m+1|m}),$$
$$R^M_{m+1|m+1} = (\mathcal{I} - K_{m+1}G)R^M_{m+1|m},$$

where K_{m+1} is the Kalman gain matrix

$$K_{m+1} = \frac{R^M_{m+1|m} G^T}{G R^M_{m+1|m} G^T + R^o}.$$

1.2. Fourier Domain Kalman Filter (FDKF).

Canonical Filtering Problem: Plentiful Observations

$$\frac{\partial}{\partial t}\vec{u}(x,t) = \mathcal{L}\left(\frac{\partial}{\partial x}\right)\vec{u}(x,t) + \sigma(x)\dot{\vec{W}}(t), \qquad \vec{u} \in \mathbb{R}^s,$$
$$\vec{v}(x_j,t_m) = G\vec{u}(x_j,t_m) + \sigma^o_{j,m}.$$

The dynamics is realized at $2N+1$ discrete points $\{x_j = jh, j = 0,1,\ldots,2N\}$ such that $(2N+1)h = 2\pi$. The observations are attainable at all the $2N+1$ grid points. The observation noise $\sigma^o_m = \{\sigma^o_{j,m}\}$ are assumed to be zero mean Gaussian variables and are spatial and temporal independent.

Finite Discrete Fourier expansion of $\vec{u}(x,t)$:

$$\vec{u}(x_j,t_m) = \sum_{|k|\le N} \vec{\hat{u}}(t_m)e^{ikx_j}, \qquad \hat{u}_{-k} = \hat{u}_k^*,$$

$$\vec{\hat{u}}(t_m) = \frac{h}{2\pi}\sum_{j=0}^{2N} \vec{u}(x_j,t_m)e^{-ikx_j}.$$

Fourier Analogue of the Canonical Filtering Problem:

$$\vec{\hat{u}}_k(t_{m+1}) = F_k\vec{\hat{u}}_k(t_m) + \vec{\sigma}_{k,m+1},$$
$$\vec{\hat{v}}_k(t_m) = G\vec{\hat{u}}_k(t_m) + \vec{\sigma}^o_{k,m}.$$

Then the original $(2N+1)s \times (2N+1)s$ filtering problem reduces to study $2N+1$ independent $s \times s$ matrix Kalman filtering problems.

1.3. FDKF with regularly spaced sparse observations.

Assume there are N model grid points. We consider the observations at every p model grid points such that the total number of observation is M with $M = N/p$.

Sparse Regularly Spaced Observations in Fourier Space is expressed as follows:

$$\vec{\hat{u}}_k(t_{m+1}) = F_k \vec{\hat{u}}_k(t_m) + \vec{\sigma}_{k,m+1}, \qquad |k| \leq N/2,$$

$$\vec{\hat{v}}_l(t_m) = G \sum_{k \in \mathcal{A}(l)} \vec{\hat{u}}_k(t_m) + \vec{\sigma}^o_{l,m}, \qquad |l| \leq M/2,$$

where the aliasing set of wavenumber l is defined as

$$\mathcal{A}(l) = \{k : k = l + Mq | q \in \mathbb{Z}, |k| \leq N/2\}$$

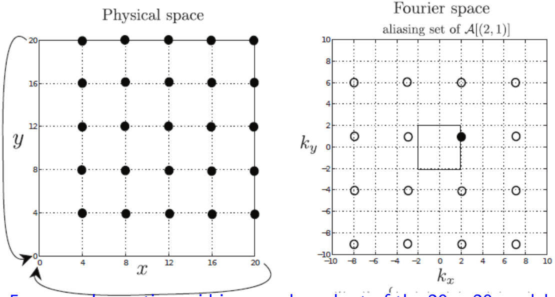


Figure 1: 5×5 sparse observation grid is a regular subset of the 20×20 model mesh so that every $P = 4$ model mesh node is observed. Here $N = 20$ and $M = 5$. There are 25 aliasing sets in all: $A(i,j)$ with $i, j \in \mathbb{Z}$ and $-2 \leq i, j \leq 2$. All primary modes lie inside the region $-2 \leq k_x, k_y \leq 2$.

Covariance matrix in physical domain — Covariance matrix in Fourier domain

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} & R_{16} & R_{17} & R_{18} & R_{19} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} & R_{26} & R_{27} & R_{28} & R_{29} \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} & R_{36} & R_{37} & R_{38} & R_{39} \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} & R_{46} & R_{47} & R_{48} & R_{49} \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} & R_{56} & R_{57} & R_{58} & R_{59} \\ R_{61} & R_{62} & R_{63} & R_{64} & R_{65} & R_{66} & R_{67} & R_{68} & R_{69} \\ R_{71} & R_{72} & R_{73} & R_{74} & R_{75} & R_{76} & R_{77} & R_{78} & R_{79} \\ R_{81} & R_{82} & R_{83} & R_{84} & R_{85} & R_{86} & R_{87} & R_{88} & R_{89} \\ R_{91} & R_{92} & R_{93} & R_{94} & R_{95} & R_{96} & R_{97} & R_{98} & R_{99} \end{pmatrix} \longrightarrow \begin{pmatrix} \begin{matrix} R^a_{11} & R^a_{12} & R^a_{13} \\ R^a_{21} & R^a_{22} & R^a_{23} \\ R^a_{31} & R^a_{32} & R^a_{33} \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} R^b_{11} & R^b_{12} & R^b_{13} \\ R^b_{21} & R^b_{22} & R^b_{23} \\ R^b_{31} & R^b_{32} & R^b_{33} \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} R^c_{11} & R^c_{12} & R^c_{13} \\ R^c_{21} & R^c_{22} & R^c_{23} \\ R^c_{31} & R^c_{32} & R^c_{33} \end{matrix} \end{pmatrix}$$

The aliased Fourier modes in geophysical systems with quadratic, advection-type nonlinearity are expected to be relatively weakly correlated.

In such systems the quadratic nonlinearities *do not directly couple the Fourier modes contained in the same aliasing set*; that is, if mode $\mathbf{k}$ is in the aliasing set $\mathcal{A}$, the quadratic couplings in the dynamics of $u_{\mathbf{k}}$ have the form

$$\frac{du_{\mathbf{k}}}{dt} \propto \sum u_{\mathbf{l}} u_{\mathbf{m}}, \qquad \mathbf{k} \in \mathcal{A}, \mathbf{l}, \mathbf{m} \notin \mathcal{A}$$

II. Test Models for Superparameterization

Features of superparameterization algorithm:

- intermittent strongly unstable fluctuations, and
- moderate scale separation without statistical equilibration ($\epsilon = 1/6$ to $1/10$).

2.1. Test model.

Decompose a turbulent field

$$U = \bar{u}(X,t) + u'(X,x,t,\tau),$$
$$X = \epsilon x, \qquad \tau = t\epsilon^{-1}.$$

The scalar multiscale test model:

$$(1) \qquad \frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \underbrace{\langle cov(u') \rangle (X,t)}_{nonlinear\ \ covariance\ \ eddy\ \ flux} + F_{ext}(X,t),$$

$$(2) \qquad \frac{\partial u'}{\partial \tau} + P'(\bar{u}, \partial_x)u' = -(-\Gamma(\partial_x)u' + \sigma(x)\dot{W}(\tau)),$$

where

$$P(\partial_X) = A\partial_X^3 - \nu\partial_X^2 + c\partial_X + d, \qquad F_{ext} = \bar{F} + \Lambda(X)\cdot \dot{W}(t),$$

$$\langle cov(u')(X,t) \rangle = \epsilon \int_0^{\epsilon^{-1}} cov(u')(X,t,\tau)d\tau = \epsilon \int_0^{\epsilon^{-1}} \overline{u'u'}(X,t,\tau)d\tau,$$

and P' is a constant coefficient differential operator that depends explicitly on the mean variable $\bar{u}$.

Using Ornstein-Uhlenbeck process with damping operator Γ and white noise forcing $\sigma(x)\dot{W}(x,\tau)$ for the eddies, it becomes a linear stochastic differential equation in Fourier space,

$$\frac{d\hat{u}'_k}{d\tau} + \tilde{P}'(\bar{u}, ik)\hat{u}'_k = -\gamma_k \hat{u}'_k + \sigma_k \dot{W}_k,$$

with the Fourier coefficient defined by the spectral integral

$$u'(X,x,t,\tau) = \int_{\mathbb{R}} \hat{u}'_k(X,t,\tau) e^{ikx} dW_k. \tag{3}$$

Therefore,

$$cov(u')(X,t,\tau) = \int_{\mathbb{R}} C_k(X,t,\tau)dk, \quad \text{where} \quad C_k \equiv \overline{\hat{u}'_k(\hat{u}'_k)^*}. \tag{4}$$

The linear deterministic ODE with coefficients depending on $\bar{u}$ for the covariance C_k:

$$\begin{aligned} \frac{dC_k}{d\tau} &= -(\tilde{P}'_k + (\tilde{P}'_k)^* + \gamma_k + \gamma_k^*)C_k + \sigma_k\sigma_k^*, \\ C_k(\tau = 0) &= C_{k,0}. \end{aligned} \tag{5}$$

2.2. Numerical implementation.

1. Compute C_k as a solution of IVP in (5) for fixed $\bar{u}$ for various modes

$$C_k(\tau) = e^{-2\lambda_k \tau} C_{k,0} + \frac{\sigma_k^2}{2\lambda_k}(1 - e^{-2\lambda_k \tau}),$$

where

$$\lambda_k = \frac{(\tilde{P}'_k + (\tilde{P}'_k)^*) + (\gamma_k + \gamma_k^*)}{2}$$

2. Compute the turbulent fluctuation $\langle cov(u') \rangle$ using the spectral integral (4) and empirical time average with constant ϵ

$$\begin{aligned}\langle cov(u') \rangle &= \epsilon \int_0^{\epsilon^{-1}} \int_{\mathbb{R}^n} C_k(\tau) dk d\tau \\ &= \int_{\mathbb{R}^n} \left[\frac{\sigma_k^2}{2\lambda_k} + \frac{\epsilon}{2\lambda_k} \left(1 - e^{-2\lambda_k \epsilon^{-1}} \left(C_{k,0} - \frac{\sigma_k^2}{2\lambda_k} \right) \right) \right] dk\end{aligned}$$

3. Integrate the large-scale PDE in (1) with large time step Δt on a coarsely resolved period domain by assuming that the turbulent fluctuation $\langle cov(u') \rangle (X, t)$ is constant over the time interval $(t, t + \Delta t)$.

Remark: pairwise small scale solutions $\langle cov(u') \rangle (X, t)$ obtained by freezing $\bar{u}$ at two distinct locations $X_j \neq X_l$ do not interact directly.

2.3. Small-scale intermittency.

To model intermittency, we choose

$$\frac{\tilde{P}'_k + (\tilde{P}'_k)^*}{2} = -f(\bar{u})A_k,$$

where

$$A_k = \bar{A}e^{-\delta|k|}|k|^2.$$

In this paper, we choose quadratic

$$f(\bar{u}) = \gamma_k + \alpha - \bar{u}^2.$$

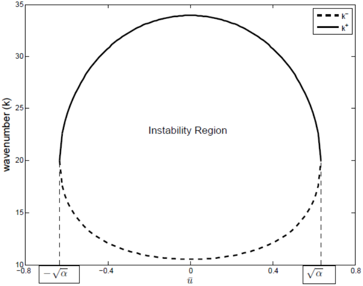


Figure 2: Boundary of unstable modes as a function of $\bar{u}$. Recall $\lambda_k = \frac{(\tilde{P}'_k + (\tilde{P}'_k)^*) + (\gamma_k + \gamma_k^*)}{2}$.

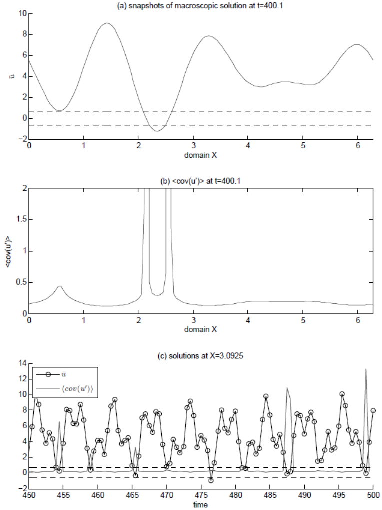


Figure 3: Deterministic forcing solutions. $\epsilon = 1/10$.

2.4. Superparameterization.
Retain large-scale dynamics (1), but make various space-time discrete approximations in solving the small-scale dynamics (2) to reduce the computational cost.

Test model

$$(1)\qquad \frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \langle cov(u')\rangle(X,t) + F_{ext}(X,t),$$

$$(2)\qquad \frac{\partial u'}{\partial \tau} + P'(\bar{u},\partial_x)u' = -(-\Gamma(\partial_x)u' + \sigma(x)\dot{W}(\tau)),$$

Traditional superparameterization introduces an artificial **scale gap** L and solves the small-scale dynamical equations in (2) locally on a periodic domain, which introduces two types of models:

1. model errors due to finite spatial and temporal discrete approximations, and
2. model errors due to truncation of the direct interaction between nonlocal fluxes.

In our test model, superparameterization only introduces model errors of the first type. We approximate the covariance integral over the lattice with wavenumbers $k_j = j/L$ such that

$$\begin{aligned}\langle cov(u')\rangle &= \int_{\mathbb{R}^n}\left[\frac{\sigma_k^2}{2\lambda_k} + \frac{\epsilon}{2\lambda_k}(1 - e^{-2\lambda_k\epsilon^{-1}}\left(C_{k,0} - \frac{\sigma_k^2}{2\lambda_k}\right))\right]dk \\ &= \frac{1}{L^n}\sum_j\left[\frac{\sigma_{k_j}^2}{2\lambda_{k_j}} + \frac{\epsilon}{2\lambda_{k_j}}(1 - e^{-2\lambda_{k_j}\epsilon^{-1}})\left(C_{k_j,0} - \frac{\sigma_{k_j}^2}{2\lambda_{k_j}}\right)\right].\end{aligned}$$

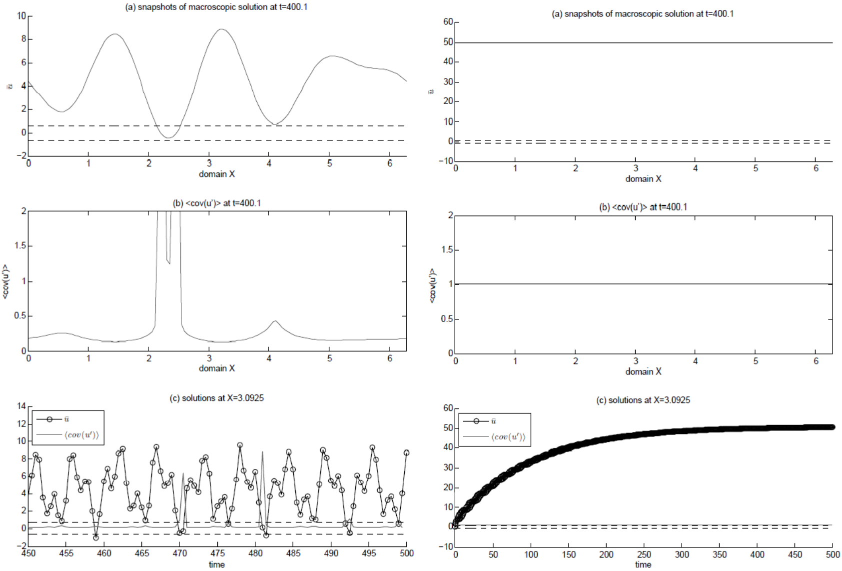


Figure 4: Deterministic forced superparameterized approximate solutions with $L = 2$ (left) and $L = 0.1$ (right). $\epsilon = 1/10$. Recall $k_j = j/L$.

2.5. Test model for superparameterization and other closure approximations.

Test models for superparameterization

$$\frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \langle cov(u')\rangle_L(X,t) + F_{ext}(X,t),$$

$$\langle cov(u')\rangle_L(X,t) = \epsilon \int_0^{\epsilon^{-1}} \int_{\mathbb{R}^n} C_k(X,t,\tau)dkd\tau.$$

Bare-truncation model

$$\frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \qquad F_{ext}(X,t)$$

Statistical equilibrium closure model

$$\frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \langle cov(u')\rangle_\infty(X,t) + F_{ext}(X,t),$$

$$\langle cov(u')\rangle_\infty(X,t) \equiv \lim_{\epsilon\to 0} \epsilon \int_0^{\epsilon^{-1}} \int_{\mathbb{R}^n} C_k(X,t,\tau)dkd\tau = \int_{\mathbb{R}^n} \frac{\sigma_k^2}{2\lambda_k}dk.$$

III. Filter Performance on Test Models

- Parameters:
  - Scale separation parameter $\epsilon = 1/10$.
  - Total grid points of large scale mean dynamics: $N = 128$.
  - Regularly sparse observations at every p model grid points, with $p = 4, 8, 16$ and 32.
  - Observation noise: $r^o = 1.41$, about 23% − 25% of the covariance of $\bar{u}$.
  - Observation time: $t_{obs} = 0.5$, much shorter than the temporal correlation.
- Measurements of filtering skill:

$$\mathrm{RMS} = \frac{1}{T - T_0} \sum_{m=T_0+1}^{T} \sqrt{\langle(\vec{u}_m^+ - \vec{u}_m)^2\rangle_N},$$

$$\mathrm{SC} = \frac{1}{T - T_0} \sum_{m=T_0+1}^{T} \frac{\left\langle(\vec{u}_m^+ - \langle\vec{u}_m^+\rangle_N)(\vec{u}_m - \langle\vec{u}_m\rangle_N)\right\rangle_N}{\sqrt{\langle(\vec{u}_m^+ - \langle\vec{u}_m^+\rangle_N)^2\rangle_N \langle(\vec{u}_m - \langle\vec{u}_m\rangle_N)^2\rangle_N}}$$

3.1. Stochastically forced prior models.

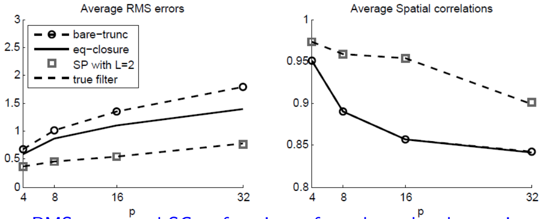


Figure 5: Average RMS errors and SC as functions of p, where the observation error $\sqrt{r^o} = 1.18$.

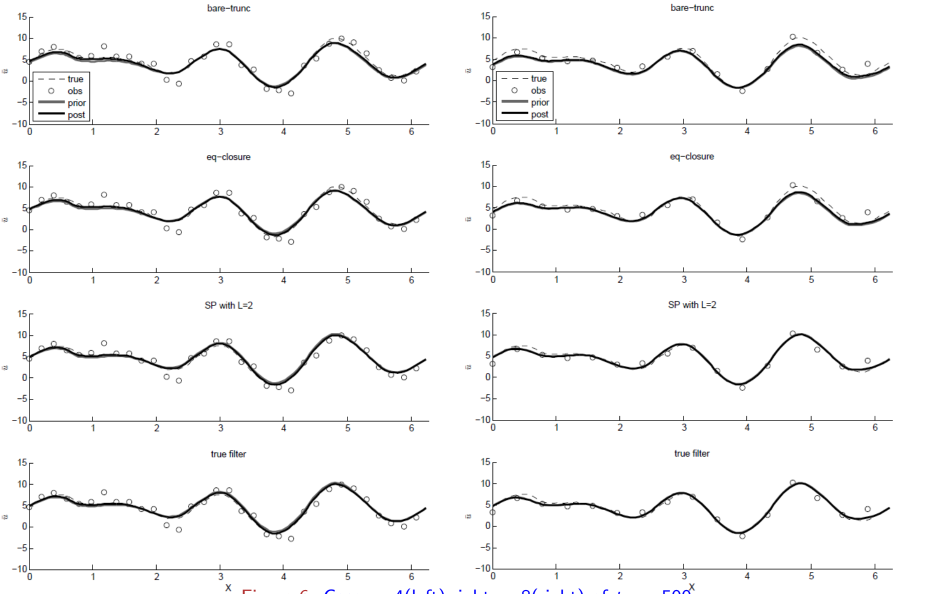


Figure 6: Case p=4(left),right:p=8(right) of $t_m = 500$.

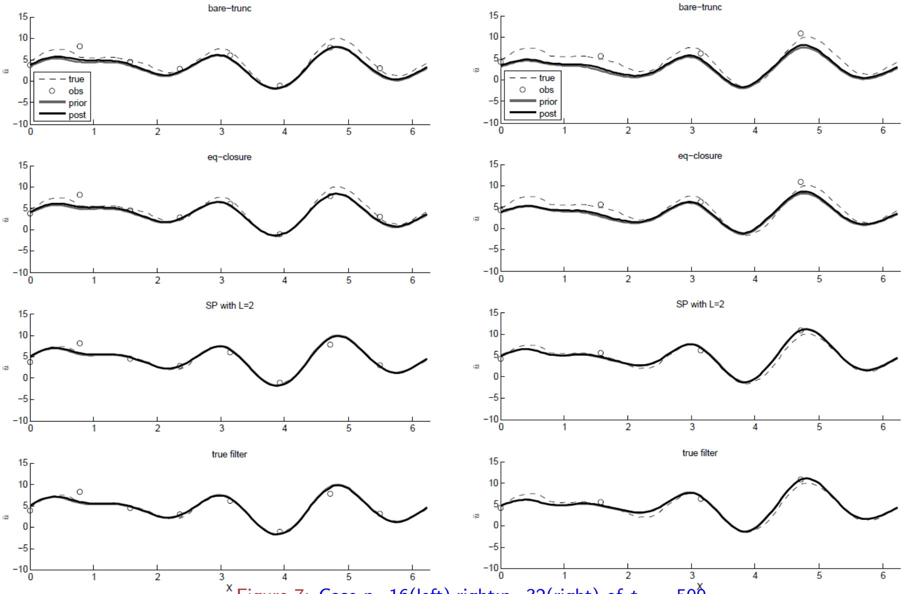


Figure 7: Case p=16(left),right:p=32(right) of $t_m = 500$.

Filtering skill with different scale gap L.

$$\langle cov(u')\rangle_L = \int_{\mathbb{R}} A(k)dk = \frac{1}{L}\sum_j A(k).$$

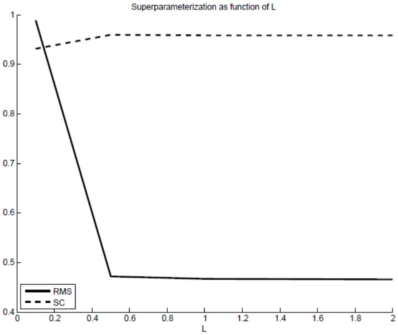


Figure 8: Case $p = 8$: Average RMSE and SC as functions of L. Observation error: $\sqrt{r^o} = 1.18$.

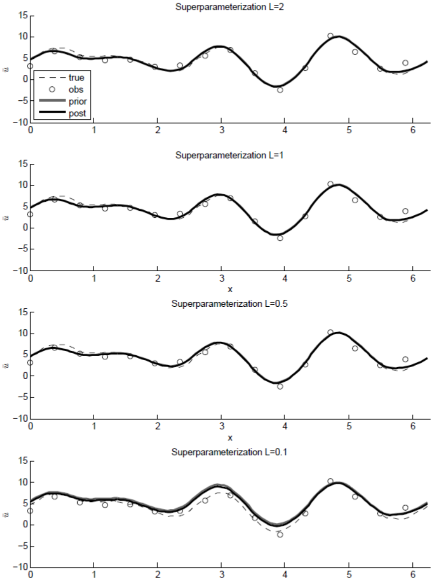


Figure 9: Snapshots at $t_m = 500$ for $p = 8$ for superparameterization with various L.

3.2. Controllability. For some initial state x_0, if the system is controllable, then for any state x_1, there exists some time t_1 and some observation v, such that the state at t_1 is x_1.

We use the perfect deterministically forced prior filter model for sparse observations with $p = 4$.

The prior model noise covariance is zero $\Longrightarrow$ The system is uncontrollable $\Longrightarrow$ The Kalman gain matrix is zero $\Longrightarrow$ The filter trusts the prior mean estimates completely.

	Not controllable		Controllable	
Scheme	RMS	SC	RMS	SC
bare truncation	43.1004	0.2314	0.6331	0.9677
eq-closure	28.7007	0.2784	0.5362	0.9677
SP with $L = 2$	1.1180	0.8364	0.2775	0.9859
true filter	1.1320	0.8325	0.2777	0.9859

Table 1: Average RMS errors and SC for filtering deterministic truth with and without controllability. $p = 4$, $\sqrt{r^o} = 1.18$.

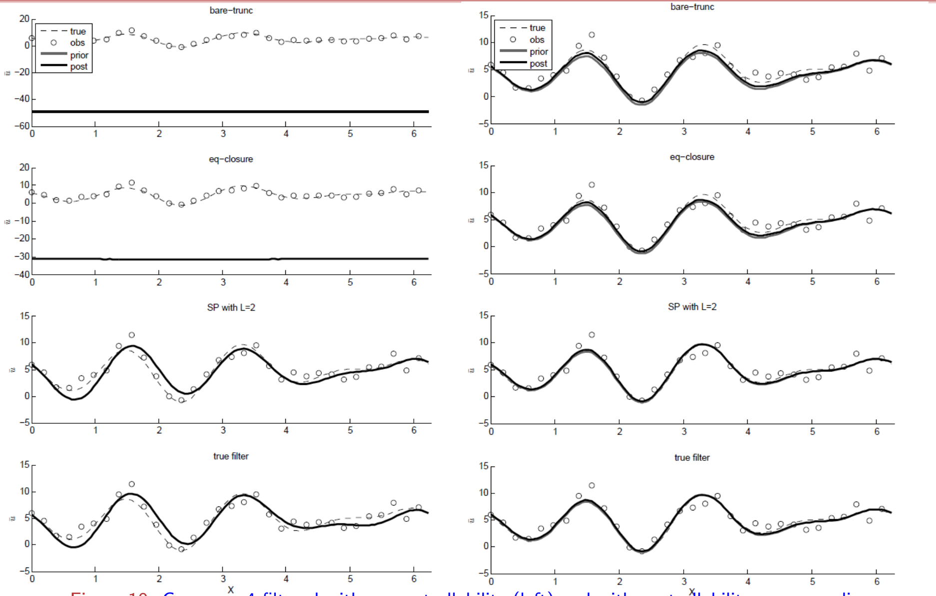


Figure 10: Case $p = 4$ filtered with no controllability (left) and with controllability corresponding to l^{-6} large scale energy spectrum (right) at $t_m = 500$.

Remarks.

- The choice of prior model is very important for sparse observations.
- The small scale dynamics is very important even if the true signal has a very steep spectrum with l^{-6}.

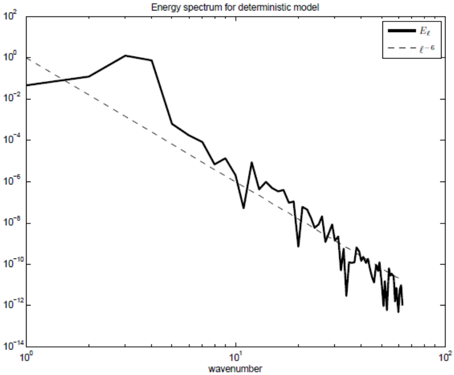


Figure 11: Empiricaly estimated large-scale energy spectrum of the deterministically forced system compared to the l^{-6} spectrum.