Linear Theory for Filtering Nonlinear Multiscale Systems with Model Error[1]

John Harlim[2]

Department of Mathematics and Department of Meteorology
The Pennsylvania State University

Collaborator: Tyrus Berry
Department of Mathematics
The Pennsylvania State University

May 21, 2014

[1]Proceedings of the Royal Society A 2014, 470 20140168.

[2]Supports: Office of Naval Research and the National Science Foundation

Outline of the talk:

- What is the linear theory? How do you account for model error such that it simultaneously produces optimal filtering and accurate climatological statistical estimates?

Outline of the talk:

- What is the linear theory? How do you account for model error such that it simultaneously produces optimal filtering and accurate climatological statistical estimates? Answer: It is possible with an appropriate stochastic parameterization.

Outline of the talk:

- What is the linear theory? How do you account for model error such that it simultaneously produces optimal filtering and accurate climatological statistical estimates? Answer: It is possible with an appropriate stochastic parameterization.
- What could happen if the appropriate stochastic parameterization is not available to us?

Outline of the talk:

- What is the linear theory? How do you account for model error such that it simultaneously produces optimal filtering and accurate climatological statistical estimates? Answer: It is possible with an appropriate stochastic parameterization.
- What could happen if the appropriate stochastic parameterization is not available to us? Answer: One easily ends up with a model that gives accurate filtering but bad climatological estimates, or vice versa.

Outline of the talk:

- What is the linear theory? How do you account for model error such that it simultaneously produces optimal filtering and accurate climatological statistical estimates? Answer: It is possible with an appropriate stochastic parameterization.
- What could happen if the appropriate stochastic parameterization is not available to us? Answer: One easily ends up with a model that gives accurate filtering but bad climatological estimates, or vice versa.
- If we have an appropriate stochastic parameterization, how should we fit the parameters?

Outline of the talk:

- What is the linear theory? How do you account for model error such that it simultaneously produces optimal filtering and accurate climatological statistical estimates? Answer: It is possible with an appropriate stochastic parameterization.
- What could happen if the appropriate stochastic parameterization is not available to us? Answer: One easily ends up with a model that gives accurate filtering but bad climatological estimates, or vice versa.
- If we have an appropriate stochastic parameterization, how should we fit the parameters? Answer: For nonlinear problems, it is more natural to obtain the parameters "online" as part of the filtering procedure.

Linear theory:

Consider a two-dimensional linear filtering problem,

$$\begin{aligned} dx &= (a_{11}x + a_{12}y)\,dt + \sigma_x dW_x, \\ dy &= \frac{1}{\epsilon}(a_{21}x + a_{22}y)\,dt + \frac{\sigma_y}{\sqrt{\epsilon}} dW_y, \\ dz &= x\,dt + \sqrt{R}dV \end{aligned}$$

for a slow variable $x \in \mathbb{R}$ and fast variable $y \in \mathbb{R}$.

Linear theory:

Consider a two-dimensional linear filtering problem,

$$\begin{aligned} dx &= (a_{11}x + a_{12}y)\,dt + \sigma_x dW_x, \\ dy &= \frac{1}{\epsilon}(a_{21}x + a_{22}y)\,dt + \frac{\sigma_y}{\sqrt{\epsilon}} dW_y, \\ dz &= x\,dt + \sqrt{R} dV \end{aligned}$$

for a slow variable $x \in \mathbb{R}$ and fast variable $y \in \mathbb{R}$.
Under some "technical" assumption, we ask the following question:
Find α and σ such that the filtered estimates of the problem above can be approximated by the filter estimates of the following reduced model:

$$dX = \alpha X\,dt + \sigma\,dW.$$

Linear theory:

There exists a unique optimal reduced filter given by the one-dimensional linear model with parameters:

$$\begin{aligned}
\alpha &= \tilde{a}(1-\epsilon\hat{a}), \\
\sigma^2 &= \sigma_x^2 + \epsilon(-2\hat{a}\sigma_x^2 + \sigma_y^2 \frac{a_{12}^2}{a_{22}^2}),
\end{aligned}$$

where $\tilde{a} = a_{11} - a_{12}a_{21}a_{22}^{-1} < 0$ and $\hat{a} = a_{12}a_{21}a_{22}^{-2}$. The optimality is in the sense of both the mean and covariance statistics are close to estimates from the true filter.

Linear theory:

There exists a unique optimal reduced filter given by the one-dimensional linear model with parameters:

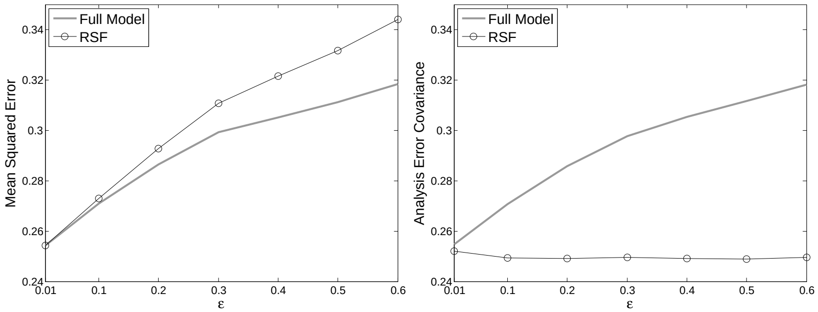
$$
\begin{aligned}
\alpha &= \tilde{a}(1 - \epsilon \hat{a}), \\
\sigma^2 &= \sigma_x^2 + \epsilon(-2\hat{a}\sigma_x^2 + \sigma_y^2 \frac{a_{12}^2}{a_{22}^2}),
\end{aligned}
$$

where $\tilde{a} = a_{11} - a_{12}a_{21}a_{22}^{-1} < 0$ and $\hat{a} = a_{12}a_{21}a_{22}^{-2}$. The optimality is in the sense of both the mean and covariance statistics are close to estimates from the true filter.

Notations:

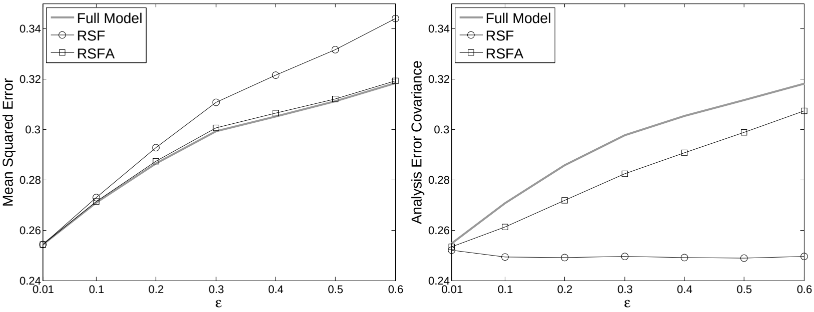
- RSF: $\alpha = \tilde{a}$, $\sigma^2 = \sigma_x^2$ (Classical averaging theory).
- RSFA: $\alpha = \tilde{a}$, $\sigma^2 = \sigma_x^2 + \epsilon\sigma_y^2 \frac{a_{12}^2}{a_{22}^2}$ [Gottwald & Harlim 2013].

Numerical Results: True filter vs RSF



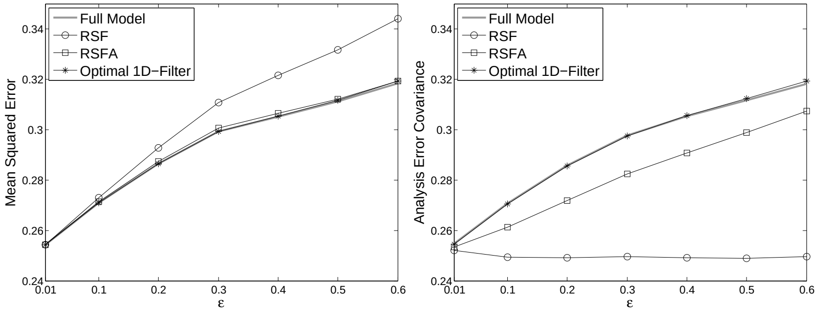
$$\begin{aligned}\text{Actual error covariance estimate} &= \frac{1}{T}\sum_{m=1}^{T}(x_m - \hat{x}_m)^2 \\ \text{Steady state filter error covariance estimate} &= \mathbb{E}[(x-\hat{x})^2].\end{aligned}$$

Numerical Results: True filter vs RSFA



Improved mean estimates, but the covariance estimates are still underestimated for large ϵ!

Numerical Results: True filter vs Linear Theory



Linear Theory: Climatological Statistics

On the other hand, the climatological statistics of the two-dimensional true model are given by:

$$
\begin{aligned}
\mathrm{Var(x)} &= \frac{\sigma_x^2 + \epsilon(-2\hat{a}\sigma_x^2 + \sigma_y^2 \frac{a_{12}^2}{a_{22}^2})}{2\tilde{a}(1-\epsilon\hat{a})} + \mathcal{O}(\epsilon^2) \\
T_x &= \frac{1}{\tilde{a}(1-\epsilon\hat{a})} + \mathcal{O}(\epsilon^2)
\end{aligned}
$$

Linear Theory: Climatological Statistics

On the other hand, the climatological statistics of the two-dimensional true model are given by:

$$\begin{aligned} \mathrm{Var}(x) &= \frac{\sigma_x^2 + \epsilon(-2\hat{a}\sigma_x^2 + \sigma_y^2 \frac{a_{12}^2}{a_{22}^2})}{2\tilde{a}(1-\epsilon\hat{a})} + \mathcal{O}(\epsilon^2) \\ T_x &= \frac{1}{\tilde{a}(1-\epsilon\hat{a})} + \mathcal{O}(\epsilon^2) \end{aligned}$$

For the 1D model $dX = \alpha\, dt + \sigma\, dW$, the corresponding statistics are

$$\mathrm{Var}(X) = \frac{\sigma^2}{2\alpha} \quad T_X = \frac{1}{\alpha}$$

Linear Theory: Climatological Statistics

On the other hand, the climatological statistics of the two-dimensional true model are given by:

$$\begin{aligned} \mathrm{Var}(x) &= \frac{\sigma_x^2 + \epsilon(-2\hat{a}\sigma_x^2 + \sigma_y^2 \frac{a_{12}^2}{a_{22}^2})}{2\tilde{a}(1-\epsilon\hat{a})} + \mathcal{O}(\epsilon^2) \\ T_x &= \frac{1}{\tilde{a}(1-\epsilon\hat{a})} + \mathcal{O}(\epsilon^2) \end{aligned}$$

For the 1D model $dX = \alpha\, dt + \sigma\, dW$, the corresponding statistics are

$$\mathrm{Var}(X) = \frac{\sigma^2}{2\alpha} \quad T_X = \frac{1}{\alpha}$$

So by matching these climatological statistics, we obtain the optimal parameters. This is an offline parameterization method, known as "Mean Stochastic Model" [Majda & Harlim 2012].

Nonlinear Test model

Consider [Gershgorin, Harlim, Majda 2010]:

$$
\begin{aligned}
\frac{du}{dt} &= -\lambda_u u + \tilde{b} + \tilde{\gamma} u + \sigma_u \dot{W}_u, \\
\frac{d\tilde{b}}{dt} &= -\lambda_b \tilde{b} + \sigma_b \dot{W}_b, \\
\frac{d\tilde{\gamma}}{dt} &= -\lambda_\gamma \tilde{\gamma} + \sigma_\gamma \dot{W}_\gamma,
\end{aligned}
$$

Nonlinear Test model

Consider [Gershgorin, Harlim, Majda 2010]:

$$
\begin{aligned}
\frac{du}{dt} &= -\lambda_u u + \tilde{b} + \tilde{\gamma} u + \sigma_u \dot{W}_u, \\
\frac{d\tilde{b}}{dt} &= -\lambda_b \tilde{b} + \sigma_b \dot{W}_b, \\
\frac{d\tilde{\gamma}}{dt} &= -\lambda_\gamma \tilde{\gamma} + \sigma_\gamma \dot{W}_\gamma,
\end{aligned}
$$

Consider observations $dz = u\, dt + \sqrt{R} dV$.

Nonlinear Test model

Consider [Gershgorin, Harlim, Majda 2010]:

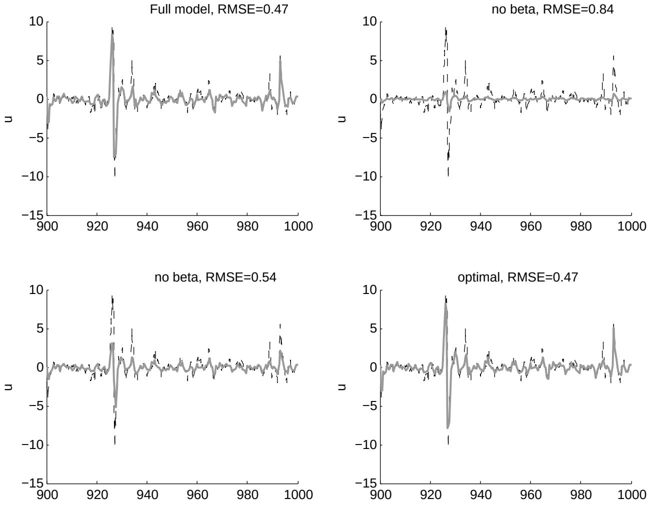
$$\begin{aligned}\frac{du}{dt} &= -\lambda_u u + \tilde{b} + \tilde{\gamma} u + \sigma_u \dot{W}_u,\\ \frac{d\tilde{b}}{dt} &= -\lambda_b \tilde{b} + \sigma_b \dot{W}_b,\\ \frac{d\tilde{\gamma}}{dt} &= -\lambda_\gamma \tilde{\gamma} + \sigma_\gamma \dot{W}_\gamma,\end{aligned}$$

Consider observations $dz = u\,dt + \sqrt{R}dV$.
We find that the optimal reduced model is given by:

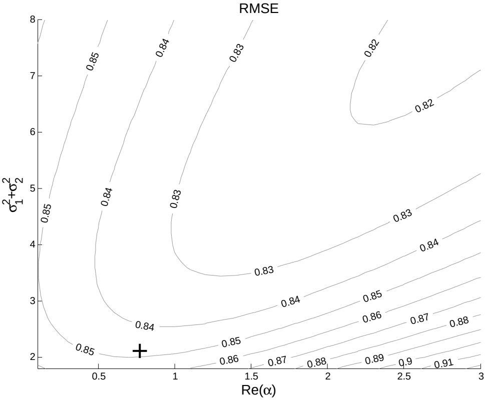
$$dU = -\alpha U dt + \beta U \circ dW_\gamma + \sigma_1 dW_u + \sigma_2 dW_b,$$

for an appropriate choices of parameters $\alpha, \beta, \sigma_1, \sigma_2$.

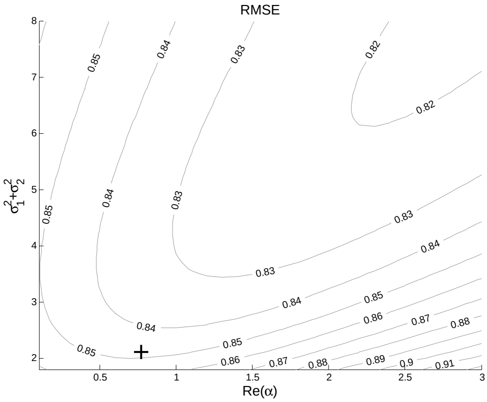
Numerical solutions



Inappropriate Stochastic Parameterization ($\beta = 0$)



Inappropriate Stochastic Parameterization ($\beta = 0$)



On the other hand, the theoretically found model without β (which RMSE=0.54) underestimates the the variance by 83% and correlation time by 52%.

Strategies for high-dimensional problems with model errors

Consider the two-layer Lorenz-96 model as the truth,

$$\begin{aligned}\frac{dx_i}{dt} &= x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F + \frac{h_x}{M}\sum_{j=1}^{M} y_{i,j},\\ \epsilon\frac{dy_{i,j}}{dt} &= y_{i,j+1}(y_{i,j-1} - y_{i,j+2}) - y_{i,j} + h_y x_i,\end{aligned}$$

where $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively and the subscript i is taken modulo N and j is taken modulo M.

Strategies for high-dimensional problems with model errors

Consider the two-layer Lorenz-96 model as the truth,

$$\begin{aligned} \frac{dx_i}{dt} &= x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F + \frac{h_x}{M}\sum_{j=1}^{M} y_{i,j}, \\ \epsilon\frac{dy_{i,j}}{dt} &= y_{i,j+1}(y_{i,j-1} - y_{i,j+2}) - y_{i,j} + h_y x_i, \end{aligned}$$

where $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively and the subscript i is taken modulo N and j is taken modulo M.

Given Model:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F$$

Strategies for high-dimensional problems with model errors

Consider the two-layer Lorenz-96 model as the truth,

$$\begin{aligned}\frac{dx_i}{dt} &= x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F + \frac{h_x}{M}\sum_{j=1}^{M} y_{i,j},\\ \epsilon\frac{dy_{i,j}}{dt} &= y_{i,j+1}(y_{i,j-1} - y_{i,j+2}) - y_{i,j} + h_y x_i,\end{aligned}$$

where $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively and the subscript i is taken modulo N and j is taken modulo M.

Given Model:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F$$

Proposed Reduced Filter Model:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F - \alpha x_i + \sum_{j=1}^{N} \sigma_{ij}\dot{W}_j + \sum_{j=1}^{N} \beta_{ij} \circ x_j \dot{V}_j$$

Offline Regression-based Method [Wilks 2005, Arnold, Moroz, and Palmer 2012]

Given training set $x_i(t)$,

- Compute $U(x_i, t) \equiv x_{i-1}(t)(x_{i+1}(t) - x_{i-2}(t)) - ax_i(t) + F - \frac{x_i(t+\delta t) - x_i(t)}{\delta t}$.
- Linear regression fitting to a cubic polynomial: $U(x_i, t) \approx b_0 + b_1 x_i(t) + b_2 x_i(t)^2 + b_3 x_i(t)^3$.
- Take the residual of this fit and apply a second linear regression fitting to an AR(1) model,

We consider this fitting strategy to obtain α and σ on

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - ax_i + F - \alpha x_i + \sigma \dot{W}_i$$

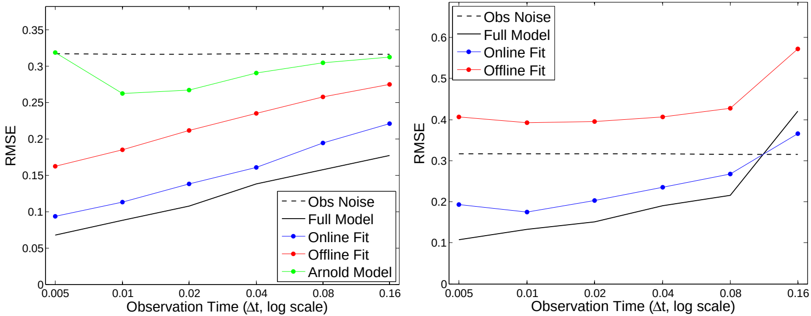
that is, apply regression to a linear model $U(x_i, t) \approx -\alpha x_i(t)$ and the residual to white noise, $\sigma \dot{W}_i$.

Online Parameterization Method to obtain α and σ

- We concatenate the dynamics with $d\alpha/dt = 0$ and apply EnKF.
- We use method developed by [Berry and Sauer 2013] to get σ, which is a reincarnation of [Mehra 1970] on EnKF framework. One can also use the noise estimation method in [Harlim, Mahdi, Majda 2014] which is a reincarnation of [Belanger 1974] on EnKF framework.
- Idea: Recursively, use the innovation vector at many lags to estimate σ.
- Note: We do not use any training data set and we also estimate R simultaneously.
- For multiplicative noise, we don't know how to do it properly, so we set $\beta = 0$.

Numerical results ($x \in \mathbb{R}^8, y \in \mathbb{R}^{256}$)

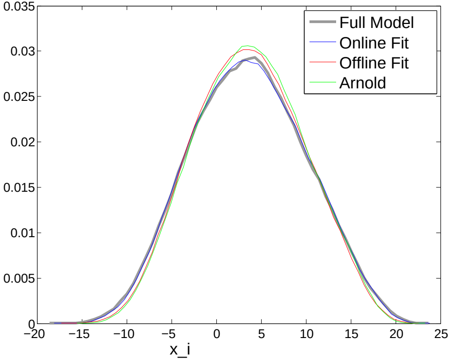
Observe all $x \in \mathbb{R}^8$ (left) and every other grid point (right)

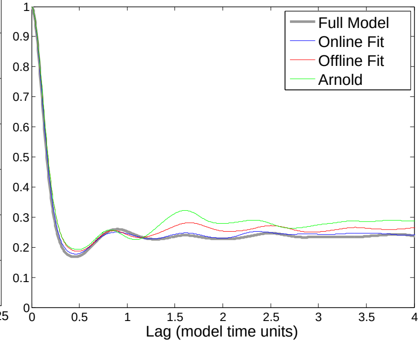


Ensemble size doubles the total state.

Online method does not use any training data set and also estimate R simultaneously.

Climatological Statistics





Summary:

- Linear Theory: There exists a stochastic parameterization for optimal filter and accurate climatological statistics (they are in the form of linear damping and additive noise forcing).

Summary:

- Linear Theory: There exists a stochastic parameterization for optimal filter and accurate climatological statistics (they are in the form of linear damping and additive noise forcing).
- For nonlinear, non-Gaussian configuration, our example suggests an additional multiplicative noise forcing.

Summary:

- Linear Theory: There exists a stochastic parameterization for optimal filter and accurate climatological statistics (they are in the form of linear damping and additive noise forcing).
- For nonlinear, non-Gaussian configuration, our example suggests an additional multiplicative noise forcing.
- When the appropriate parametric form is unknown, one can get good filtering but bad climate model or vice versa.

Summary:

- Linear Theory: There exists a stochastic parameterization for optimal filter and accurate climatological statistics (they are in the form of linear damping and additive noise forcing).
- For nonlinear, non-Gaussian configuration, our example suggests an additional multiplicative noise forcing.
- When the appropriate parametric form is unknown, one can get good filtering but bad climate model or vice versa.
- When the appropriate parametric form is known, a natural parameter estimation method is online (as part of the filtering procedure).

Summary:

- Linear Theory: There exists a stochastic parameterization for optimal filter and accurate climatological statistics (they are in the form of linear damping and additive noise forcing).
- For nonlinear, non-Gaussian configuration, our example suggests an additional multiplicative noise forcing.
- When the appropriate parametric form is unknown, one can get good filtering but bad climate model or vice versa.
- When the appropriate parametric form is known, a natural parameter estimation method is online (as part of the filtering procedure).
- There is nothing inherently wrong with offline parameterization method. For e.g., MSM works in linear and Gaussian setting since the statistics are stationary in this situation.

Summary:

- The offline method we considered fits the bias and random parts of model errors separately, while the online method we considered simultaneously account for the nonlinear feedback of each perturbed parameters.

Summary:

- The offline method we considered fits the bias and random parts of model errors separately, while the online method we considered simultaneously account for the nonlinear feedback of each perturbed parameters.
- Many practical approaches considered estimating either the mean model error (forecast bias), assuming certain structure on the covariance. On the other hand, there are also methods assume model error is completely unbiased random variables and design various covariance inflation methods.

Summary:

- The offline method we considered fits the bias and random parts of model errors separately, while the online method we considered simultaneously account for the nonlinear feedback of each perturbed parameters.
- Many practical approaches considered estimating either the mean model error (forecast bias), assuming certain structure on the covariance. On the other hand, there are also methods assume model error is completely unbiased random variables and design various covariance inflation methods.
- What we found from this academic exercise is that we need to estimate both (or even higher order) statistics simultaneously especially in nonlinear problems, they are simply the first two-moments of a stochastic process that describe model error estimator.