Adaptive ensemble Kalman filtering of nonlinear systems

Tyrus Berry

September 29, 2016

MiniCV: http://math.gmu.edu/~berry/

Background:

- PhD Mathematics at GMU, Advisor: Tim Sauer
- Postdoc at PSU with John Harlim
- NSF Big Data Postdoc at GMU (current)

Research Interests:

- Geometry of data and nonparametric statistics
- Data-driven and model-free forecasting
- Filtering/forecasting with model error

This is also joint work with Franz Hamilton (postdoc at NC State)

What is the filtering problem?

- Consider a discrete time dynamical system:

$$\begin{aligned} x_k &= f_k(x_{k-1}, \omega_k) \\ y_k &= h_k(x_k, \nu_k) \end{aligned}$$

- Where x_k is the state variable, ω_k is stochastic forcing, and the maps f_k define the dynamics
- The maps h_k are called the observation functions, ν_k is observation noise, and y_k is a noisy observation

What is the filtering problem?

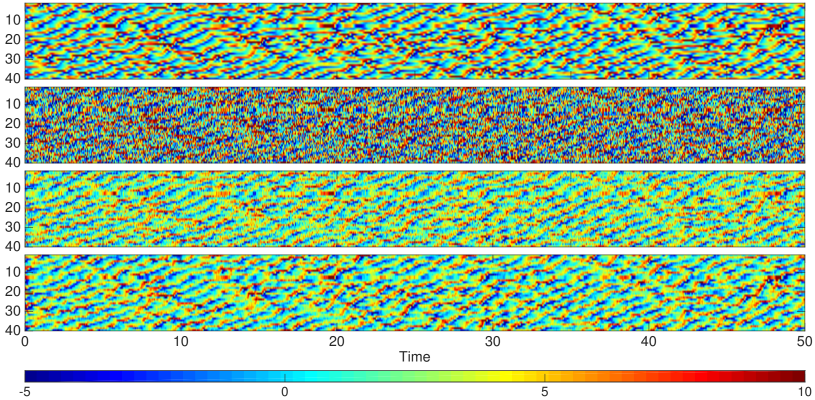
- Consider a discrete time dynamical system:

$$\begin{aligned} x_k &= f_k(x_{k-1}, \omega_k) \\ y_k &= h_k(x_k, \nu_k) \end{aligned}$$

- Given the observations $y_1, ..., y_k$ we define three problems:
- **Filtering:** Estimate the current state $p(x_k \,|\, y_1, ..., y_k)$
- **Forecasting:** Estimate a future state $p(x_{k+\ell} \,|\, y_1, ..., y_k)$
- **Smoothing:** Estimate a past state $p(x_{k-\ell} \,|\, y_1, ..., y_k)$

What is the filtering problem?

$$\frac{dx^i}{dt} = -x^{i-2}x^{i-1} + x^{i-1}x^{i+1} - x^i + F$$



Two Step Filtering to Find $p(x_k \mid y_1, ..., y_k)$

- Assume we have $p(x_{k-1} \mid y_1, ..., y_{k-1})$
- **Forecast Step:** Find $p(x_k \mid y_1, ..., y_{k-1})$
- **Assimilation Step:** Perform a Bayesian update,

$$p(x_k \mid y_1, ..., y_k) \propto p(x_k \mid y_1, ..., y_{k-1}) p(y_k \mid x_k, y_1, ..., y_{k-1})$$

Posterior $\propto$ **Prior** $\times$ **Likelihood**

Kalman Filter

- Assume linear dynamics/obs and additive Gaussian noise

$$
\begin{aligned}
x_k &= F_{k-1}x_{k-1} + \omega_k & \omega_k \sim \mathcal{N}(0, Q) \\
y_k &= H_k x_k + \nu_k & \nu_k \sim \mathcal{N}(0, R)
\end{aligned}
$$

- For linear systems, easy observability condition:

$$
\tilde{H}_k^\ell = \begin{pmatrix} H_k \\ H_{k+1}F_k \\ \vdots \\ H_{k+\ell+1}F_{k+\ell}\cdots F_k \end{pmatrix}
$$

Must be full rank for some ℓ

Kalman Filter

- Assume linear dynamics/obs and additive Gaussian noise

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + \omega_k & \qquad \omega_k \sim \mathcal{N}(0, Q) \\ y_k &= H_k x_k + \nu_k & \qquad \nu_k \sim \mathcal{N}(0, R) \end{aligned}$$

- Assume current estimate is Gaussian:

$$p(x_{k-1} \mid y_1, ..., y_{k-1}) = \mathcal{N}(\hat{x}^a_{k-1}, P^a_{k-1})$$

Kalman Filter: Forecast Step

- At time $k-1$ we have mean $\hat{x}^a_{k-1}$ and covariance P^a_{k-1}
- Linear combinations of Gaussians are still Gaussian so:
  - $p(F_{k-1}x_{k-1} \,|\, y_1, ..., y_{k-1}) = \mathcal{N}(F_{k-1}\hat{x}^a_{k-1}, F_{k-1}P_{k-1}F^\top_{k-1})$
  - $p(x_k \,|\, y_1, ..., y_{k-1}) = \mathcal{N}(F_{k-1}\hat{x}^a_{k-1}, F_{k-1}P_{k-1}F^\top_{k-1} + Q)$
- Define the *Forecast mean:* $\hat{x}^f_k \equiv F_{k-1}\hat{x}^a_{k-1}$
- Define the *Forecast covariance:* $P^f_k \equiv F_{k-1}P^a_{k-1}F^\top_{k-1} + Q$

Kalman Filter: Defining the Likelihood function

- Recall that $y_k = H_k x_k + \nu_k$ where $\nu_k \sim \mathcal{N}(0, R)$ is Gaussian
- The forecast distribution: $p(x_k \mid y_1, ..., y_{k-1}) = \mathcal{N}(\hat{x}_k^f, P_k^f)$
- **Likelihood:** $p(y_k \mid x_k, y_1, ..., y_{k-1}) = \mathcal{N}(H_k \hat{x}_k^f, H_k P_k^f H_k^\top + R)$
- Define the *Observation mean:* $y_k^f = H_k \hat{x}_k^f$
- Define the *Observation covariance:* $P_k^y = H_k P_k^f H_k^\top + R$

Kalman Filter: Assimilation Step

- Gaussian prior × Gaussian likelihood ⇒ Gaussian posterior

$$
\begin{aligned}
p(y|x)p(x) \propto \exp\Big\{ & -\frac{1}{2}(y - Hx)^\top (P^y)^{-1}(y - Hx) \\
& - \frac{1}{2}(x - \hat{x}^f)^\top (P^f)^{-1}(x - \hat{x}^f)\Big\} \\
\propto \exp\Big\{ & -\frac{1}{2}x^\top((P^y)^{-1} + H(P^f)^{-1}H^\top)x \\
& + x^\top(H^\top(P^y)^{-1}y - (P^f)^{-1}\hat{x}^f)\Big\}
\end{aligned}
$$

- Posterior Covariance: $P^a = \left((P^f)^{-1} + H^\top(P^y)^{-1}H\right)^{-1}$
- Posterior Mean: $x^a = P^a\left(H^\top(P^y)^{-1}y - (P^f)^{-1}\hat{x}^f\right)$

Kalman Filter: Assimilation Step

- **Kalman Equations:** (after some linear algebra...)
  - Kalman Gain: $K_k = P_k^f H_k^\top (P_k^y)^{-1}$
  - Innovation: $\epsilon_k = y_k - y_k^f$
  - Posterior Mean: $\hat{x}_k^a = \hat{x}_k^f + K_k \epsilon_k$
  - Posterior Covariance: $P_k^a = (I - K_k H_k) P_k^f$
- $\hat{x}_k^a$ is the least squares/minimum variance estimator of x_k

Kalman Filter Summary

$$x_k^f = F_{k-1}x_{k-1}^a$$

$$\begin{aligned} P_k^f &= F_{k-1}P_{k-1}^a F_{k-1}^T + Q_{k-1} \\ P_k^y &= H_k P_k^f H_k^T + R_{k-1} \end{aligned}$$

$$\begin{aligned} K_k &= P_k^f H_k^T (P_k^y)^{-1} \\ P_k^a &= (I - K_k H_k)P_k^f \end{aligned}$$

$$\begin{aligned} \epsilon_k &= y_k - y_k^f = y_k - H_k x_k^f \\ x_k^a &= x_k^f + K_k \epsilon_k \end{aligned}$$

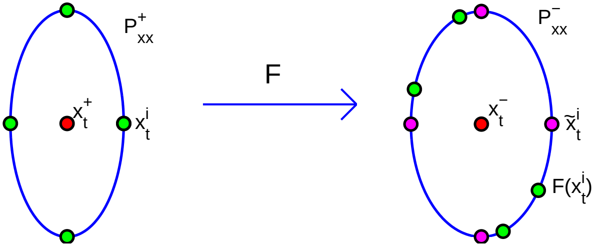
What about nonlinear systems?

- Consider a system of the form:

$$
\begin{aligned}
x_{k+1} &= f(x_k) + \omega_{k+1} & \omega_{k+1} &\sim \mathcal{N}(0, Q) \\
y_{k+1} &= h(x_{k+1}) + \nu_{k+1} & \nu_{k+1} &\sim \mathcal{N}(0, R)
\end{aligned}
$$

- More complicated observability condition (Lie derivatives)
- **Extended Kalman Filter (EKF)**:
  - Linearize $F_k = Df(\hat{x}_k^a)$ and $H_k = Dh(\hat{x}_k^f)$
- Problem: State estimate $\hat{x}_k^a$ may not be well localized
- Solution: Ensemble Kalman Filter (EnKF)

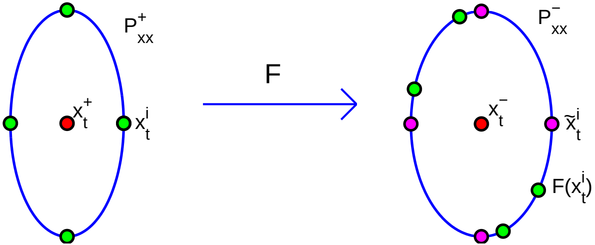
Ensemble Kalman Filter (EnKF)



Generate an ensemble with the current statistics (use matrix square root):

$$
\begin{aligned}
x_t^i &= \text{"sigma points" on semimajor axes} \\
x_t^f &= \frac{1}{2n}\sum F(x_t^i) \\
P_{xx}^f &= \frac{1}{2n-1}\sum (F(x_t^i) - x_t^f)(F(x_t^i) - x_t^f)^T + Q
\end{aligned}
$$

Ensemble Kalman Filter (EnKF)



Calculate $y_t^i = H(F(x_t^i))$. Set $y_t^f = \frac{1}{2n}\sum_i y_t^i$.

$$\begin{aligned} P_{yy} &= (2n-1)^{-1}\sum (y_t^i - y_t^f)(y_t^i - y_t^f)^T + R \\ P_{xy} &= (2n-1)^{-1}\sum (F(x_t^i) - x_t^f)(y_t^i - y_t^f)^T \\ K &= P_{xy}P_{yy}^{-1} \text{ and } P_{xx}^a = P_{xx}^f - KP_{yy}K^T \\ x_{t+1}^a &= x_t^f + K(y_t - y_t^f) \end{aligned}$$

Parameter Estimation

- When the model has parameters p,

$$x_{k+1} = f(x_k, p) + \omega_{k+1}$$

- Can *augment* the state $\tilde{x}_k = [x_k, p_k]$
- Introduce trivial dynamics for p

$$\begin{aligned} x_{k+1} &= f(x_k, p_k) + \omega_{k+1} \\ p_{k+1} &= p_k + \omega^p_{k+1} \end{aligned}$$

- Need to tune the covariance of ω^p_{k+1}

Example of Parameter Estimation

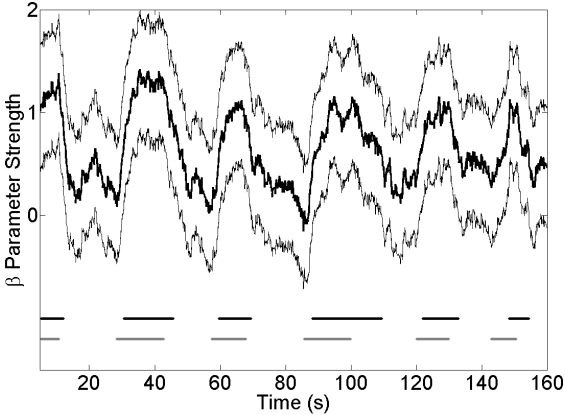
Consider the Hodgkin-Huxley neuron model, expanded to a network of n equations

$$\begin{aligned}
\dot{V}_i &= -g_{Na} m^3 h (V_i - E_{Na}) - g_K n^4 (V_i - E_K) - g_L (V_i - E_L) \\
&\quad + I + \sum_{j \neq i}^{n} \Gamma_{\mathrm{HH}}(V_j) V_j \\
\dot{m}_i &= a_m(V_i)(1 - m_i) - b_m(V_i) m_i \\
\dot{h}_i &= a_h(V_i)(1 - h_i) - b_h(V_i) h_i \\
\dot{n}_i &= a_n(V_i)(1 - n_i) - b_n(V_i) n_i \\
\Gamma_{\mathrm{HH}}(V_j) &= \beta_{ij} / (1 + e^{-10(V_j + 40)})
\end{aligned}$$

Only observe the voltages V_i, recover the hidden variables and the connection parameters β

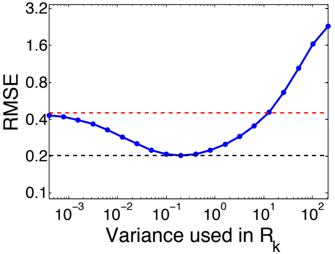
Example of Parameter Estimation

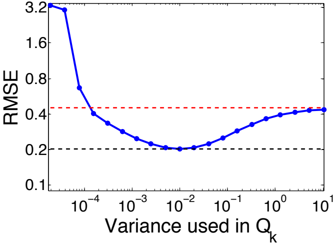
Can even turn connections on and off (grey dashes)
Variance estimate $\Rightarrow$ statistical test (black dashes)



Nonlinear Kalman-type Filter: Influence of Q and R

- Simple example with full observation and diagonal noise covariances
- Red indicates RMSE of unfiltered observations
- Black is RMSE of ‘optimal’ filter (true covariances known)





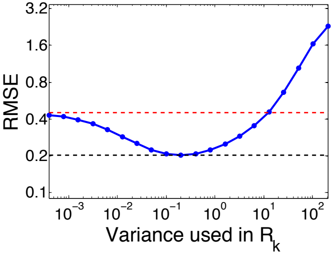
Nonlinear Kalman-type Filter: Influence of Q and R

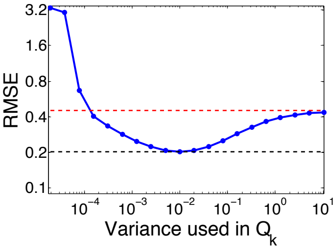
Standard Kalman Update:

$$P_k^f = F_{k-1} P_{k-1}^a F_{k-1}^T + Q_{k-1}$$
$$P_k^y = H_k P_k^f H_k^T + R_{k-1}$$

$$K_k = P_k^f H_k^T (P_k^y)^{-1}$$
$$P_k^a = (I - K_k H_k) P_k^f$$

$$\epsilon_k = y_k - y_k^f = y_k - H_k x_k^f$$
$$x_k^a = x_k^f + K_k \epsilon_k$$





Adaptive Filter: Estimating Q and R

- Innovations contain information about Q and R

$$\begin{aligned}\epsilon_k &= y_k - y_k^f \\ &= h(x_k) + \nu_k - h(x_k^f) \\ &= h(f(x_{k-1}) + \omega_k) - h(f(x_{k-1}^a)) + \nu_k \\ &\approx H_k F_{k-1}(x_{k-1} - x_{k-1}^a) + H_k \omega_k + \nu_k\end{aligned}$$

- IDEA: Use innovations to produce samples of Q and R :

$$\begin{aligned}\mathbb{E}[\epsilon_k \epsilon_k^T] &\approx HP^f H^T + R \\ \mathbb{E}[\epsilon_{k+1} \epsilon_k^T] &\approx HFP^e H^T - HFK\mathbb{E}[\epsilon_k \epsilon_k^T] \\ P^e &\approx FP^a F^T + Q\end{aligned}$$

- In the linear case this is rigorous and was first solved by Mehra in 1970

Adaptive Filter: Estimating Q and R

- To find Q and R we estimate H_k and F_{k-1} from the ensemble and invert the equations:

$$\begin{aligned} \mathbb{E}[\epsilon_k \epsilon_k^T] &\approx HP^f H^T + R \\ \mathbb{E}[\epsilon_{k+1} \epsilon_k^T] &\approx HFP^e H^T - HFK\mathbb{E}[\epsilon_k \epsilon_k^T] \end{aligned}$$

- This gives the following *empirical* estimates of Q_k and R_k:

$$\begin{aligned} P_k^e &= (H_{k+1}F_k)^{-1}(\epsilon_{k+1}\epsilon_k^T + H_{k+1}F_k K_k \epsilon_k \epsilon_k^T)H_k^{-T} \\ Q_k^e &= P_k^e - F_{k-1}P_{k-1}^a F_{k-1}^T \\ R_k^e &= \epsilon_k \epsilon_k^T - H_k P_k^f H_k^T \end{aligned}$$

- Note: P_k^e is an empirical estimate of the background covariance

An Adaptive Kalman-Type Filter for Nonlinear Problems

We combine the estimates of Q and R with a moving average

Original Kalman Eqs.

$$
\begin{aligned}
P_k^f &= F_{k-1} P_{k-1}^a F_{k-1}^T + Q_{k-1} \\
P_k^y &= H_k P_k^f H_k^T + R_{k-1}
\end{aligned}
$$

$$
\begin{aligned}
K_k &= P_k^f H_k^T (P_k^y)^{-1} \\
P_k^a &= (I - K_k H_k) P_k^f
\end{aligned}
$$

$$
\begin{aligned}
\epsilon_k &= y_k - y_k^f \\
x_k^a &= x_k^f + K_k \epsilon_k
\end{aligned}
$$

Our Additional Update

$$
\begin{aligned}
P_{k-1}^e &= F_{k-1}^{-1} H_k^{-1} \epsilon_k \epsilon_{k-1}^T H_{k-1}^{-T} \\
&+ K_{k-1} \epsilon_{k-1} \epsilon_{k-1}^T H_{k-1}^{-T}
\end{aligned}
$$

$$
\begin{aligned}
Q_{k-1}^e &= P_{k-1}^e - F_{k-2} P_{k-2}^a F_{k-2}^T \\
R_{k-1}^e &= \epsilon_{k-1} \epsilon_{k-1}^T - H_{k-1} P_{k-1}^f H_{k-1}^T
\end{aligned}
$$

$$
\begin{aligned}
Q_k &= Q_{k-1} + (Q_{k-1}^e - Q_{k-1})/\tau \\
R_k &= R_{k-1} + (R_{k-1}^e - R_{k-1})/\tau
\end{aligned}
$$

How does this compare to inflation?

- We extend Kalman's equations to estimate Q and R
- Estimates converge for linear models with Gaussian noise
- When applied to nonlinear, non-Gaussian problems
  - We interpret Q as an additive inflation
  - Q can have complex structure, possibly more effective than multiplicative inflation?
  - Downside: many more parameters than multiplicative inflation
- Somewhat less ad hoc than other inflation techniques?

Observability and Parameterization of Q

Recall:

$$\begin{aligned} P^e_{k-1} &= F^{-1}_{k-1} H^{-1}_k \epsilon_k \epsilon^T_{k-1} H^{-T}_{k-1} + K_{k-1} \epsilon_{k-1} \epsilon^T_{k-1} H^{-T}_{k-1} \\ Q^e_{k-1} &= P^e_{k-1} - F_{k-2} P^a_{k-2} F^T_{k-2} \end{aligned}$$

Together these equations imply that:

$$\begin{aligned} H_k F_{k-1} Q^e_k H^T_{k-1} &= \epsilon_k \epsilon^T_{k-1} + H_k F_{k-1} K_{k-1} \epsilon_{k-1} \epsilon^T_{k-1} \\ &\quad - H_k F_{k-1} P^a_{k-1} F^T_{k-1} H^T_{k-1} \end{aligned}$$

Set C_k equal to the right hand side (we simply compute C_k). Parameterize $Q^e_k = \sum_{i=1}^{s} q_i \hat{Q}_i$ where q_i are scalar parameters and $\hat{Q}_i$ are 'shape' matrices.

Observability and Parameterization of Q

We now need to solve:

$$C_k = \sum_{i=1}^{s} q_i H_k F_{k-1} \hat{Q}_i H_{k-1}^T$$

We vectorize the equation as

$$\text{vec}(C_k) = \sum_{i=1}^{s} q_i \text{vec}(H_k F_{k-1} \hat{Q}_i H_{k-1}^T) = A_k [q_1, ..., q_s]^T$$

where A_k is an m^2-by-l matrix where the i-th row is given by $\text{vec}(H_k F_{k-1} \hat{Q}_i H_{k-1}^T)$.
We can the solve for the parameters $[q_1, ..., q_s]^T$ by least squares.

Adaptive Filter: Application to Lorenz-96

- We will apply the adaptive EnKF to the 40-dimensional Lorenz96 model integrated over a time step $\Delta t = 0.05$

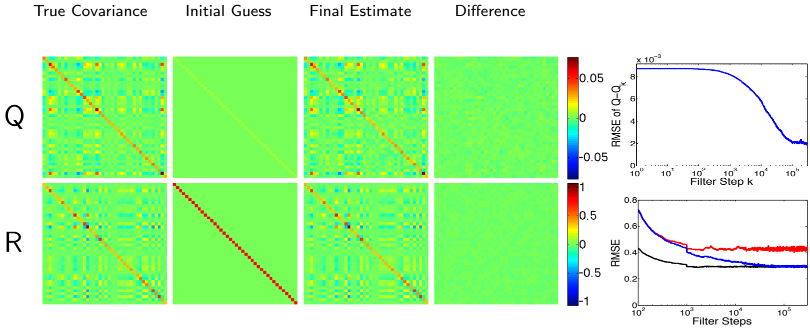
$$\frac{dx^i}{dt} = -x^{i-2}x^{i-1} + x^{i-1}x^{i+1} - x^i + F$$

- We augment the model with Gaussian white noise

$$\begin{aligned} x_k &= f(x_{k-1}) + \omega_k & \omega_k = \mathcal{N}(0, Q) \\ y_k &= h(x_k) + \nu_k & \nu_k = \mathcal{N}(0, R) \end{aligned}$$

- We will consider full and sparse observations
- The Adaptive EnKF uses $F = 8$
- We will consider model error where the true $F^i = \mathcal{N}(8, 16)$

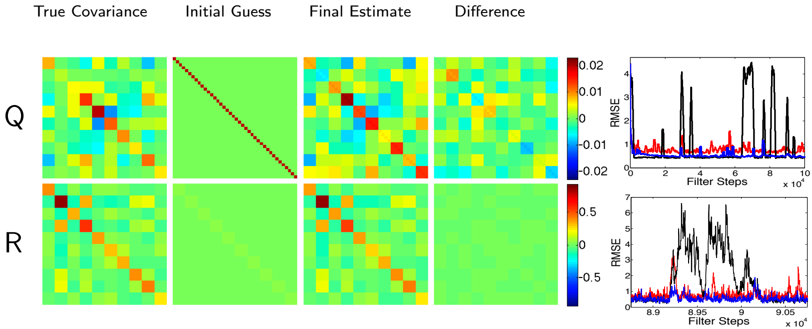
Recovering Q and R, Full Observability



RMSE shown for the initial guess covariances (red) the true Q and R (black) and the adaptive filter (blue)

Recovering Q and R, Sparse Observability

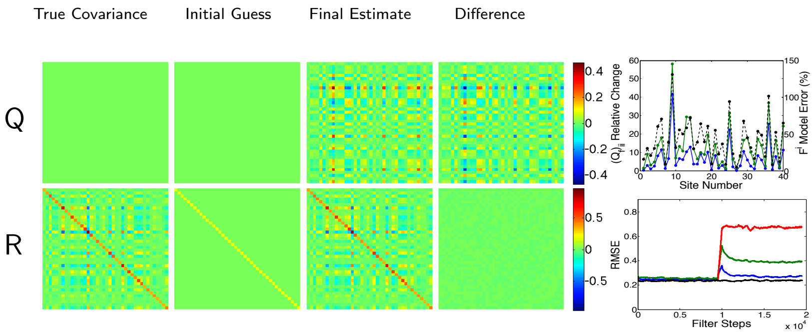
Observing 10 sites results in divergence with the true Q and R



RMSE shown for the initial guess covariances (red) the true Q and R (black) and the adaptive filter (blue)

Compensating for Model Error

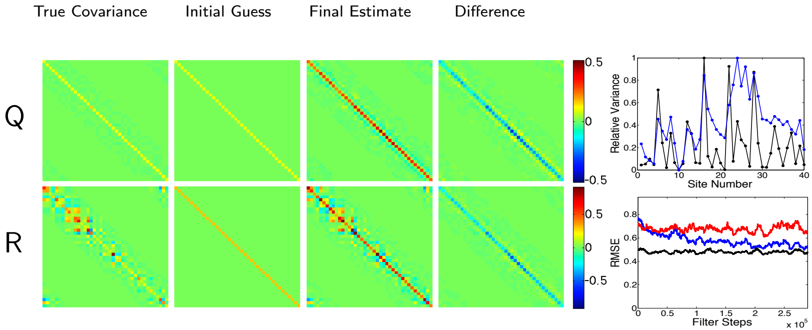
The adaptive filter compensates for errors in the forcing F^i



RMSE shown for the initial guess covariances (red) an Oracle EnKF (black) and the adaptive filter (blue)

Integration with the LETKF

Simply find a local Q and R for each region

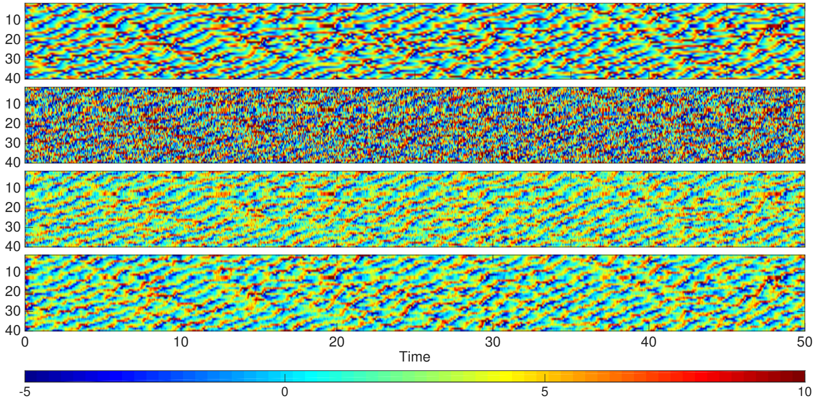


RMSE shown for the initial guess covariances (red) the true Q and R (black) and the adaptive filter (blue)

Kalman-Takens Filter: Throwing out the model...

- Starting with historical observations $\{y_0, ..., y_n\}$
- Form Takens delay-embedding state vectors $x_i = (y_i, y_{i-1}, ..., y_{i-d})^\top$
- Build an EnKF:
  - Apply analog forecast to each ensemble member
  - Use the observation function $Hx_i = y_i$
  - Crucial to estimate Q and R

Kalman-Takens applied to L96



Papers with Franz Hamilton and Tim Sauer

http://math.gmu.edu/~berry/

- Ensemble Kalman filtering without a model. Phys. Rev. X (2016).
- Adaptive ensemble Kalman filtering of nonlinear systems. Tellus A (2013).
- Real-time tracking of neuronal network structure using data assimilation. Phys. Rev. E (2013).

Related/Background Material

- R. Mehra, 1970: On the identification of variances and adaptive Kalman filtering.
- P. R. Bélanger, 1974: Estimation of noise covariance matrices for a linear time-varying stochastic process.
- J. Anderson, 2007: An adaptive covariance inflation error correction algorithm for ensemble filters.
- H. Li, E. Kalnay, T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter.
- B. Hunt, E. Kostelich, I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter.
- E. Ott, et al. 2004: A local ensemble Kalman filter for atmospheric data assimilation.