Data driven computing by the morphing fast Fourier transform ensemble Kalman filter in epidemic spread simulations

Loren Cobb
Jan Mandel, Jonathan Beezley, Ashok Krishnamurthy
Department of Mathematical and Statistical Sciences

Research funded by the U.S. National Institutes of Health and the U.S. National Science Foundation

Correspondence: Loren.Cobb@ucdenver.edu
The Epidemic Problem

• If an entirely new and unknown pandemic infectious disease emerges — as HIV, Ebola, and SARS did in the 20th century — then how can we best track the progress of the pandemic?

• If the disease has a very high mortality rate, then we can expect that air, rail, and road transport will be severely restricted. In this case, the disease spreads as a traveling wave, fueled by person-to-person contact.

• The “Morphing FFT EnKF” is a powerful Bayesian data assimilation method for tracking traveling waves.

• The following simulation illustrates...
(an animated epidemic follows the next click)
Infected

Infected (persons)

Data Min = 0, Max = 20060
Infected
Spatial Epidemic Model (S-I-R)

\[ S = \text{density of susceptible population}, \]
\[ I = \text{density of infected population}, \]
\[ R = \text{density of removed population (recovered, dead, quarantined)}. \]

\[ \frac{\partial S(x,y,t)}{\partial t} = -\beta SJ, \]
\[ \frac{\partial I(x,y,t)}{\partial t} = \beta SJ - \gamma I, \]
\[ \frac{\partial R(x,y,t)}{\partial t} = \gamma I, \]
\[ J(x,y,t) = \iint I(x-\phi, y-\psi, t) K(\phi, \psi) d\phi d\psi. \]
Organization of this talk

1. The Kalman Filter (KF) for Bayesian data assimilation.
2. The Ensemble KF (EnKF) for high-dimensional problems.
3. Use of the Fast Fourier Transform (FFT) for covariance stabilization.
4. Image registration by algorithmic morphing.
5. Application of the Morphing FFT EnKF to spatial epidemic tracking.
The Tracking Problem

Given a state vector $u \in \mathbb{R}^n$, with known dynamics $u_{t+1} = f(u_t) + q_t$, where $q_t$ is (possibly random) input, and we can only occasionally observe some linear function of the state, with measurement error:

$$d_t = Hu_t + e_t, \ e_t \sim \text{IN}(0, R),$$

then create a new Bayesian estimate of the state vector $u_t$ every time a new observation $d_t$ arrives.

One solution: the Kalman Filter (KF).
The Kalman Filter (KF) algorithm

1. Let $u$ be the estimated state vector of the system, with superscripts $a$ and $f$ for “analysis” (posterior) and “forecast” (prior) estimates.
   
   $u \in \mathbb{R}^n$.

2. The forecast step simply advances the model:
   
   $u_{t+1}^f = f(u_t^a)$.

3. When new data $d$ arrives, the analysis step performs a Bayesian update of both the state and covariance estimates:
   
   
   $u_t^a = u_t^f + K_t (d_t - Hu_t^f)$, and $C_t^a = (I - K_t H)C_t^f$, where
   
   $K_t = C_t^f H^T (HC_t^f H^T + R)^{-1}$ is the Kalman Gain matrix.

4. Use results at step 3 as initial values and repeat step 2 until the next datum arrives. Then apply step 3 and repeat.
The Ensemble KF algorithm (EnKF)

1. Start with an initial ensemble,
\[
A = (x_1^0, x_2^0, \ldots , x_N^0) \in \mathbb{R}^{n \times N}
\]

   n: dimension of state vector,
   N: number of ensemble members.

2. The forecast step applies to each member of the ensemble:
\[
u_j^f = f(u_j^0), j = 1, \ldots , N
\]

3. Analysis step employs a random perturbation e:
\[
u_j^a = u_j^f + C^f H^T (HC^f H^T + R)^{-1} (d + e_j - Hu_j^f)
\]

4. Use results at step 3 as initial values and repeat step 2 until the next datum arrives. Then apply step 3 and repeat.
The Covariance Problem

• If the dimensionality of the state vector is high (one million elements is not uncommon) then the covariance matrix is too large to be stored.

• The ensemble solves that problem — each covariance can be estimated from the ensemble — but the covariance matrix will not be of full rank (because there are too few ensemble members).

• Our FFT method generates stable covariance estimates even with very small ensembles.
Covariance Estimation with FFT

- Our method exploits the fact that the state vector describes a random field that is approximately stationary, and that the covariance of this field decays with spatial distance.

- Tapering is an effective alternative, but it is computationally much more expensive than the following:
  
  - In the frequency domain, we assume that the off-diagonal elements of the covariance can be neglected, that is, Fourier modes with different frequencies are approximately independent. With this key assumption in place, then:

  - The Kalman filter formulas consist of manipulating diagonal matrices in the frequency domain, and all the computations become simple. If $H = I$ and $R = rI$, then

  $\hat{u}_k^a = \hat{u}_k + \hat{c} \cdot (\hat{c} + r)^{-1} \cdot (\hat{d} + \hat{e}_k - \hat{u}_k^f)$. 
FFT vs Taper (simulated 1D data)
The EnKF artifact problem

• The EnKF posterior estimate of the state vector is locally a linear combination of ensemble state vectors. This creates artifacts (ghosts) that then become incorporated into the model (as erroneous outbreaks of disease).

• Example, from wildfire project, shown below. Morphing eliminates most of the artifacts:
Morphing FFT EnKF

• Given an ensemble of similar epidemics, with differences only in the location and shape of the traveling wave, we construct three vector fields that describe the changes necessary to morph each epidemic into a standard form. Then we apply the EnKF to these three vector fields. The result is called the Morphing FFT EnKF.

• An example of morphing:
Test Case Demonstration

• In the following slides we will show how each of the major filter types performs with a simple epidemic tracking problem:
  – The standard Ensemble Kalman Filter (EnKF)
  – The EnKF with FFT stabilization of covariance
  – The Morphing EnKF
  – The Morphing FFT EnKF
Standard EnKF
FFT EnKF
Morphing EnKF
Morphing FFT EnKF
Conclusions

1. The FFT procedure for stabilizing the covariance in an Ensemble Kalman Filter is computationally more efficient than the tapering procedure, with comparable performance.

2. The morphing variant of the EnKF appears to work well for tracking certain kinds of epidemics — highly infectious air-borne diseases moving through populations that are prevented from long-distance travel.

3. When populations have air, rail, and car transportation, then a different kind of filtering will be required, perhaps some variant of the particle filter.

Thank You!