Overcoming model and observation error in data assimilation using manifold learning

Tyrus Berry
Dept. of Mathematical Sciences
George Mason University

Joint Math Meetings
January 17, 2019

Joint work with John Harlim, Franz Hamilton, and Tim Sauer
FILTERING OVERVIEW

▶ Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]
FILTERING OVERVIEW

- Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- Filtering: Given \( y_1, \ldots, y_k \) estimate \( x_k \) or \( P(x_k \mid y_1, \ldots, y_k) \)
FILTERING OVERVIEW

- Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- Filtering: Given \( y_1, \ldots, y_k \) estimate \( x_k \) or \( P(x_k \mid y_1, \ldots, y_k) \)
  - Forecast: From \( P(x_k \mid y_1, \ldots, y_k) \) find prior \( P(x_{k+1} \mid y_1, \ldots, y_k) \)
FILTERING OVERVIEW

- Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- Filtering: Given \( y_1, \ldots, y_k \) estimate \( x_k \) or \( P(x_k \mid y_1, \ldots, y_k) \)
  - Forecast: From \( P(x_k \mid y_1, \ldots, y_k) \) find prior \( P(x_{k+1} \mid y_1, \ldots, y_k) \)
  - Assimilate: Combine prior with likelihood \( P(y_{k+1} \mid x_{k+1}) \)
FILTERING OVERVIEW

- Consider the standard filtering problem,
  \[ x_i = f(x_{i-1}) + \omega_{i-1} \]
  \[ y_i = h(x_i) + \eta_i \]

- Filtering: Given \( y_1, ..., y_k \) estimate \( x_k \) or \( P(x_k \mid y_1, ..., y_k) \)
  - Forecast: From \( P(x_k \mid y_1, ..., y_k) \) find prior \( P(x_{k+1} \mid y_1, ..., y_k) \)
  - Assimilate: Combine prior with likelihood \( P(y_{k+1} \mid x_{k+1}) \)

- Kalman-based Filtering:
FILTERING OVERVIEW

- Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- Filtering: Given \( y_1, \ldots, y_k \) estimate \( x_k \) or \( P(x_k \mid y_1, \ldots, y_k) \)
  - Forecast: From \( P(x_k \mid y_1, \ldots, y_k) \) find prior \( P(x_{k+1} \mid y_1, \ldots, y_k) \)
  - Assimilate: Combine prior with likelihood \( P(y_{k+1} \mid x_{k+1}) \)

- Kalman-based Filtering:
  - Forecast: Local Linear (EKF) or Ensemble (EnKF)
FILTERING OVERVIEW

- Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- Filtering: Given \( y_1, \ldots, y_k \) estimate \( x_k \) or \( P(x_k | y_1, \ldots, y_k) \)
  - Forecast: From \( P(x_k | y_1, \ldots, y_k) \) find prior \( P(x_{k+1} | y_1, \ldots, y_k) \)
  - Assimilate: Combine prior with likelihood \( P(y_{k+1} | x_{k+1}) \)

- Kalman-based Filtering:
  - Forecast: Local Linear (EKF) or Ensemble (EnKF)
  - Assimilate: Gaussian assumption + Bayesian posterior
Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- **Model error**: Specify variables, unknown dynamics
- **Observation error**: Specify dynamics, unknown mapping
- **Both unknown**: Underdetermined
Bias in Observation Models

- Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

- True observation function \( h(x) \) is unknown

- Assume we have a guess \( g(x) \) and

\[ y_i = h(x_i) + \eta_i = g(x_i) + b_i + \eta_i \]

- Bias: \( b_i \equiv h(x_i) - g(x_i) \)
Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i, \eta_i) \]

True observation function \( h \) is unknown

Assume we have a guess \( g \) and

\[ y_i = h(x_i, \eta_i) = g(x_i, \eta_i) + b_i \]

Stochastic Bias: \( b_i \equiv h(x_i, \eta_i) - g(x_i, \eta_i) \)
## Example 1: Lorenz-96

- **40-dimensions:**
  \[
  \dot{x}_j = x_{j-1}(x_{j+1} - x_{j-2}) - x_j + 8
  \]

- Observe 20 variables, 7 are ‘cloudy’

  \[
  h(x_k) = \begin{cases}
  x_k & \xi_i > 0.8 \\
  \beta_k x_k - 8 & \text{else}
  \end{cases}
  \]

  \[
  \beta_k \sim \mathcal{N}(0.5, 1/50).
  \]

  \[
  \xi_i \sim \mathcal{U}(0, 1)
  \]
**Example 1: Lorenz-96**

- The result is a bimodal distribution, “cloudy/clear”
- Obs Model Error = True Obs - \( g(\text{True State}) \)
CORRECTING THE BIAS

- Our goal is to find $p(b_i \mid y_i)$
- We can then correct our observation function
  \[ \hat{h}(x^f_i) \equiv g(x^f_i) + \hat{b}_i \]
- Where $\hat{b}_i = \mathbb{E}_{p(b_i \mid y_i)}[b_i]$
- Since $\hat{b}_i$ random:
  - Inflate the obs noise covariance
  - Use $\hat{R}_{b_i} = \mathbb{E}_{p(b_i \mid y_i)}[(b_i - \hat{b}_i)(b_i - \hat{b}_i)^\top]$
CORRECTING THE BIAS

- Need to find $p(b_i \mid y_i)$
- From the forecast step we have a prior $p(b_i)$
  - Forecast $x_i^f \Rightarrow$ Bias estimate: $y_i - g(x_i^f)$
  - Prior $p(b_i) = \mathcal{N}(y_i - g(x_i^f), P_i^y)$
- Use Bayes’ $p(b_i \mid y_i) = p(b_i)p(y_i \mid b_i)$
- Need the likelihood $p(y_i \mid b_i)$
- Use kernel estimation of conditional distributions
LEARNING THE CONDITIONAL DISTRIBUTION

- Given training data \((y_i, b_i)\) our goal is to learn \(p(y_i | b_i)\).
- For a kernel \(K(\alpha, \beta) = e^{-\frac{||\alpha - \beta||^2}{\delta^2}}\) we define Hilbert spaces
  \[
  \mathcal{H}_y = \left\{ \sum_{i=1}^{N} a_i K(y_i, \cdot) : \vec{a} \in \mathbb{R}^N \right\}
  \]
  \[
  \mathcal{H}_b = \left\{ \sum_{i=1}^{N} a_i K(b_i, \cdot) : \vec{a} \in \mathbb{R}^N \right\}
  \]
- Eigenvectors \(\phi_\ell\) of \(K_{ij} = K(y_i, y_j)\) are a basis for \(\mathcal{H}_y\).
- Similarly \(\varphi_k\) are a basis for \(\mathcal{H}_b\).
LEARNING THE CONDITIONAL DISTRIBUTION

- We assume that $p(y | b)$ can be approximated in $\mathcal{H}_y \otimes \mathcal{H}_b$

- Let $C_{ij}^{yb} = \langle \phi_i, \varphi_j \rangle$ and $C_{ij}^{bb} = \langle \varphi_i, \varphi_j \rangle$ then define

$$
C^{y | b} = C^{yb}(C^{bb} + \lambda I)^{-1}
$$

- We can then define a consistent estimator of $p(y | b)$ by

$$
\hat{p}(y | b) = \sum_{i,j=1}^{N} C_{i,j}^{y | b} \phi_i(y) \varphi_j(b) \hat{q}(y)
$$
CORRECTING THE BIAS

- Below plots have $y_i \approx -4$
- Left is clear, right is cloudy
- Notice bimodal likelihood
OVERVIEW

- **Learning Phase:** Given training data set \((x_i, y_i)\)
  - Compute the biases \(b_i = y_i - g(x_i)\)
  - Learn the conditional distribution \(p(y | b)\)
- **Filtering:** Forecast \(x_i^f \Rightarrow \text{Bias estimate: } y_i - g(x_i^f)\)
  - Prior \(p(b) = \mathcal{N}(y_i - g(x_i^f), P_i^y)\)
  - Likelihood \(p(y_i | b)\) from learning phase
  - Apply Bayes: \(p(b | y_i) = p(b)p(y_i | b)\)
- Estimate bias \(\hat{b}_i\) and correct the observation
OVERVIEW

Prior $p(x_i)$ $\rightarrow$ Primary Filter $\rightarrow$ Posterior $p(x_i | y_i)$

Error Prior $p(b)$ $\rightarrow$ Secondary Filter $\rightarrow$ Error Posterior $p(b | y_i)$

Observation $y_i$ $\rightarrow$ RKHS+Training Data $\rightarrow$ Likelihood $p(y_i | b)$
LORENZ-96 RESULTS
LORENZ-96 RESULTS

- Works well with small measurement noise
- Observations need to be precise, but not accurate
Bias in Observation Models

- Consider the standard filtering problem,
  \[ x_i = f(x_{i-1}) + \omega_{i-1} \]
  \[ y_i = h(x_i) + \eta_i \]

- True observation function \( h(x) \) is unknown

- Guess \( g(x) \) and bias: \( b_i \equiv h(x_i) - g(x_i) \)

- Previously: Given training data, \( \{(x_i, y_i)\} \)

- Now: Only have observations, \( \{y_i\} \).

- Idea: Iteratively estimate the bias
ITERATIVE BIAS ESTIMATION

- Get the filter running with the bad obs $g$ (inflate $R$)

$$\hat{b}_k^{(0)} = y_k - g(x_k^{(0)})$$

- Takens’ embedding to identify similar states:

$$z_k = [y_k, y_{k-1}, \ldots, y_{k-d}]$$

- Smooth the bias with local linear interpolation:

$$b^{(0)}(x_k) = \sum_i e^{-\frac{||z_k - z_i||^2}{\epsilon^2}} \hat{b}_i^{(0)}$$

- Update the observation function:

$$g^{(1)} = g + b^{(0)}$$
Iterative Bias Estimation

\[ f \rightarrow y_k \rightarrow \text{Filter} \rightarrow x_k^{(\ell)} , \hat{b}_k^{(\ell)} = y_k - g(x_k^{(\ell)}) \rightarrow \text{Takens} \]

\[ g^{(\ell+1)} = g + b^{(\ell)} \]

\[ b^{(\ell)}(x_k) \]
EXAMPLE 2: LORENZ-63

- 3-dimensional chaotic ODE
- True Obs:
  \[ h(\vec{x}) = h \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \sin(x_1) \\ x_2 - 6 \\ \cos(x_3) \end{bmatrix} \]
- Guess:
  \[ g(\vec{x}) = g \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \]
EXAMPLE 2: LORENZ-63

Figure 2. Results of filtering noisy Lorenz-63 (a) $x_1$ (b) $x_2$ and (c) $x_3$ time series when true observation function, $h$, is unknown and $R = 2I_{3 \times 3}$. Notice the large difference between the true observations $h(\tilde{x}_k) + n_k$ (blue circles) to the true state variables (solid black curve). We compare the EnKF estimate using the wrong observation function, $g$, without observation model error correction (solid gray lines) and the EnKF estimate with correction (solid red lines) shown. (d) Plot of RMSE vs. iteration of the observation model error correction method, where $\ell = 0$ corresponds to the standard EnKF without correction.
PUTTING THE TWO METHODS TOGETHER

- Step 1: Iterative Estimation
  - Using historical observations, offline

- Step 2: Conditional Estimation
  - Use data from step 1 to train RKHS, online
EXAMPLE 3: INTRACELLULAR FROM EXTRACELLULAR
EXAMPLE 3: INTRACELLULAR FROM EXTRACELLULAR

Fig 7. Results from assimilating the extracellular data to the Fitzhugh-Nagumo model. Thin light traces indicate individual events and the thick dark lines denote the mean waveforms averaged over individual events. Without bias correction (Fig 7 a-b), the filter is unable to compensate for the error resulting in an inaccurate estimate of the (a) intracellular potential and (b) recovery variable dynamics. When we estimate the bias and correct the observation model error (Fig. 7c-d), we are able to learn the mapping from intracellular to extracellular state, and thus get an improved reconstruction of the intracellular potential and recovery dynamics, (c) and (d) respectively.
**EXAMPLE 4: MULTI-CLOUD “SATELLITE-LIKE” OBS**

- Consider a 7-dim’l model for a column of atmosphere
  - Baroclinic anomaly potential temperatures, $\theta_1$ and $\theta_2$
  - Boundary layer anomaly potential temperature, $\theta_{eb}$
  - Vertically averaged water vapor content, $q$
  - Cloud fractions: congestus $f_c$, deep $f_d$, and stratiform $f_s$

- Extrapolate anomaly potential temperature at height $z$

\[
T(z) = \theta_1 \sin \left( \frac{z\pi}{Z_T} \right) + 2\theta_2 \sin \left( \frac{2z\pi}{Z_T} \right), \quad z \in [0, 16]
\]

EXAMPLE 4: MULTI-CLOUD "SATELLITE-LIKE" OBS

- Extrapolate anomaly potential temperature at height $z$

\[
T(z) = \theta_1 \sin\left(\frac{Z\pi}{Z_T}\right) + 2\theta_2 \sin\left(\frac{2Z\pi}{Z_T}\right), \quad z \in [0, 16]
\]

- Brightness temperature-like quantity at wavenumber-$\nu$

\[
h_{\nu}(x, f) = (1 - f_d - f_s) \left[ (1 - f_c)(\theta_{eb} T_{\nu}(0) + \int_0^{z_c} T(z) \frac{\partial T_{\nu}}{\partial z}(z) \, dz) 
+ f_c T(z_c) T_{\nu}(z_c) + \int_{z_c}^{z_d} T(z) \frac{\partial T_{\nu}}{\partial z}(z) \, dz \right] 
+ (f_d + f_s) T(z_d) T_{\nu}(z_d) + \int_{z_d}^{\infty} T(z) \frac{\partial T_{\nu}}{\partial z}(z) \, dz,
\]

- Setting $f = 0$ is the clear sky model
EXAMPLE 4: MULTI-CLOUD “SATELLITE-LIKE” OBS

Weighting functions define RTM at different wavenumbers
EXAMPLE 4: MULTI-CLOUD “SATELLITE-LIKE” OBS

- Biases at the 16 observed wavenumbers
EXAMPLE 4: MULTI-CLOUD “SATELLITE-LIKE” OBS

- Multimodal likelihood functions

![Graphs showing multimodal likelihood functions with and without training data.](image-url)
EXAMPLE 4: “SATELLITE-LIKE” OBS (ITERATIVE)
EXAMPLE 4: “SATELLITE-LIKE” OBS (RKHS)
REFERENCES

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

