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 $y_1, ..., y_k$ estimate x_k or $P(x_k \mid y_1, ..., y_k)$

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 $y_1, ..., y_k$ estimate x_k or $P(x_k \,|\, y_1, ..., y_k)$
  - Forecast: From $P(x_k \,|\, y_1, ..., y_k)$ find prior $P(x_{k+1} \,|\, y_1, ..., y_k)$

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 $y_1, ..., y_k$ 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 $P(y_{k+1} \mid x_{k+1})$

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 $y_1, ..., y_k$ 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:
  - 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 $y_1, ..., y_k$ 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:
  - Forecast: Local Linear (EKF) or Ensemble (EnKF)
  - Assimilate: Gaussian assumption + Bayesian posterior

Bias in Observation Models

- Consider the standard filtering problem,

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

- 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: $b_i \equiv h(x_i) - g(x_i)$

Bias in Observation Models

- Consider the standard filtering problem,

$$\begin{aligned} x_i &= f(x_{i-1}) + \omega_{i-1} \\ y_i &= h(x_i, \eta_i) \end{aligned}$$

- 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:

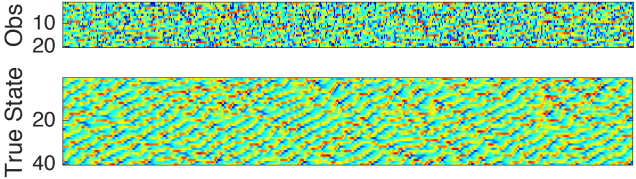
$$\dot{x}_j = x_{j-1}(x_{j+1} - x_{j-2}) - x_j + 8$$

- Observe 20 variables, 7 are 'cloudy'

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

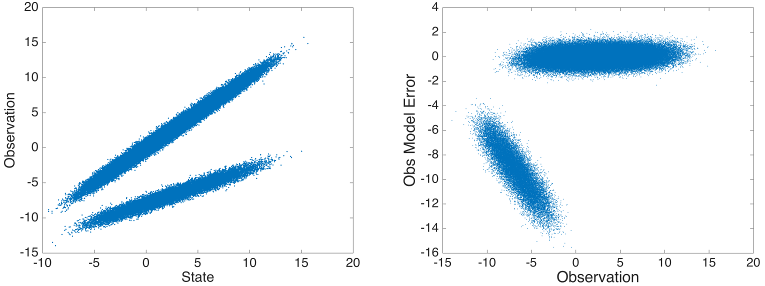
$$\beta_k \sim \mathcal{N}(0.5, 1/50).$$

$$\xi_i \sim \mathcal{U}(0, 1)$$



EXAMPLE 1: LORENZ-96

- The result is a bimodal distribution, “cloudy/clear”
- Obs Model Error = True Obs - g(True State)



CORRECTING THE BIAS

- Our goal is to find $p(b_i \mid y_i)$
- We can then correct our observation function

$$\hat{h}(x_i^f) \equiv g(x_i^f) + \hat{b}_i$$

- 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 $p(y_i \mid b_i)$
- 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 ϕ_ℓ of $K_{ij} = K(y_i, y_j)$ are a basis for $\mathcal{H}_y$.
- Similarly φ_k are a basis for $\mathcal{H}_b$.

Learning the conditional distribution

- We assume that $p(y \mid b)$ can be approximated in $\mathcal{H}_y \otimes \mathcal{H}_b$
- Let $C^{yb}_{ij} = \langle \phi_i, \varphi_j \rangle$ and $C^{bb}_{ij} = \langle \varphi_i, \varphi_j \rangle$ then define

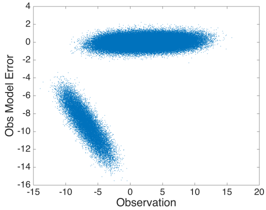
$$C^{y|b} = C^{yb} \left(C^{bb} + \lambda I \right)^{-1}$$

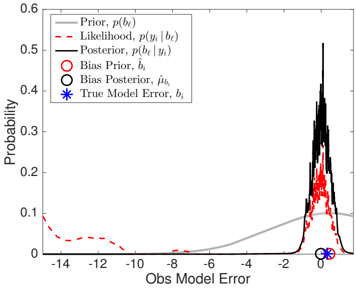
- We can then define a consistent estimator of $p(y \mid b)$ by

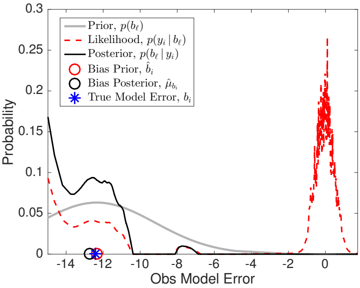
$$\hat{p}(y \mid b) = \sum_{i,j=1}^{N} C^{y|b}_{i,j} \phi_i(y) \varphi_j(b) \hat{q}(y)$$

Correcting the bias

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



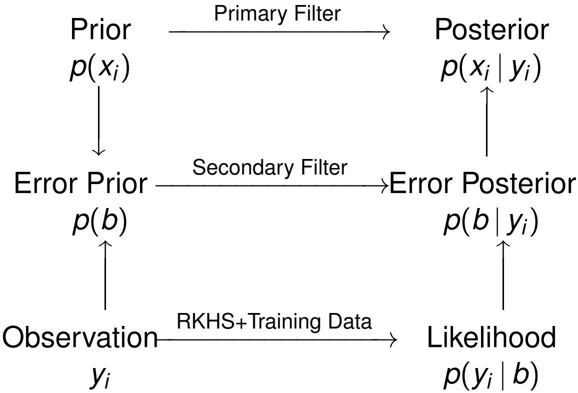




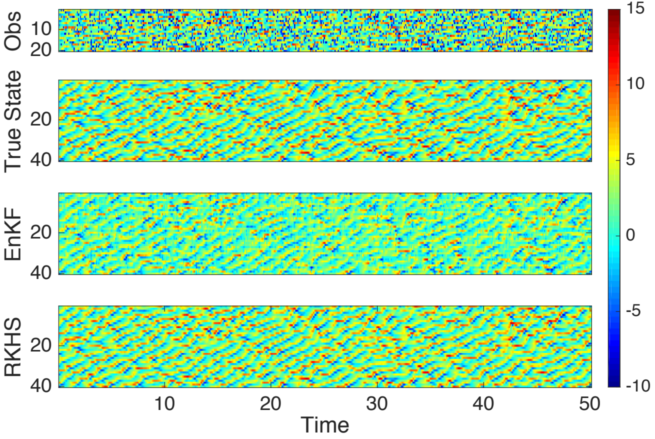
OVERVIEW

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

Overview

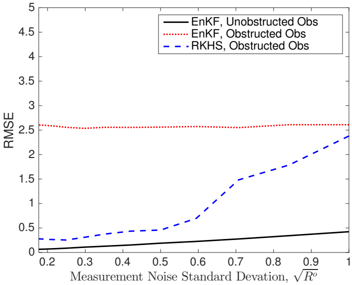

Prior
$p(x_i)$
Primary Filter
Posterior
$p(x_i \mid y_i)$
Error Prior
$p(b)$
Secondary Filter
Error Posterior
$p(b \mid y_i)$
Observation
y_i
RKHS+Training Data
Likelihood
$p(y_i \mid b)$

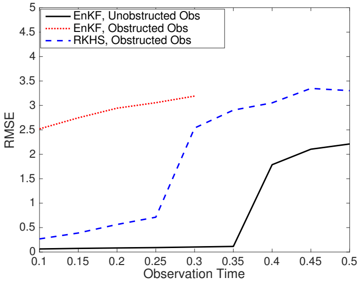
Lorenz-96 Results



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)

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

- Takens' embedding to identify similar states:

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

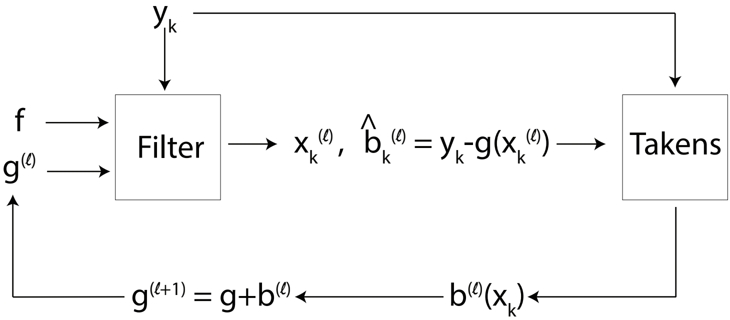
- Smooth the bias with local linear interpolation:

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

- Update the observation function:

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

Iterative Bias Estimation



Example 2: Lorenz-63

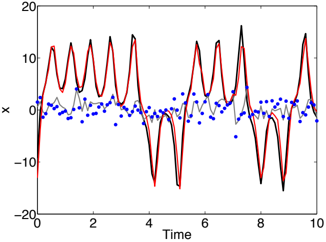
- 3-dimensional chaotic ODE
- True Obs:

$$h(\vec{x}) = h\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} \sin(x_1) \\ x_2 - 6 \\ \cos(x_3) \end{bmatrix}$$

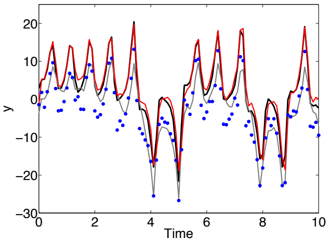
- Guess:

$$g(\vec{x}) = g\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

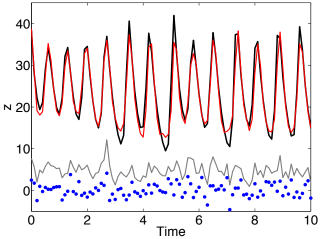
EXAMPLE 2: LORENZ-63



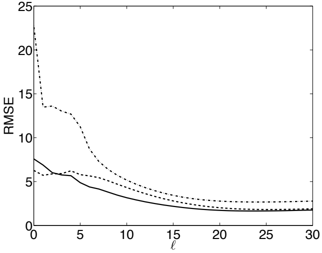
(a)



(b)



(c)

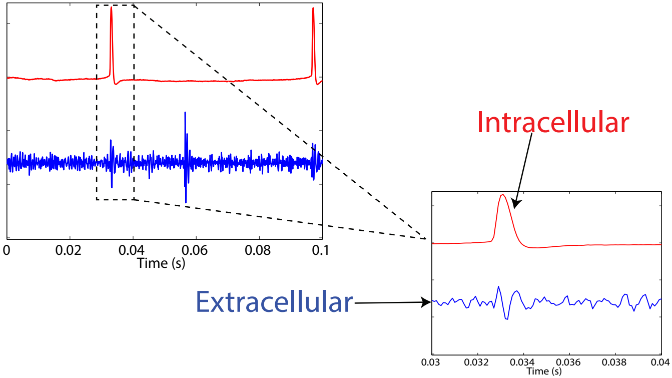


(d)

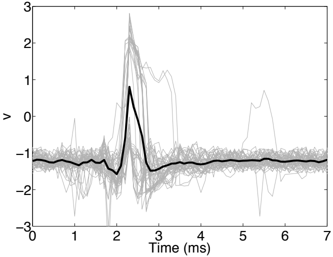
Putting the Two Methods Together

- Step 1: Iterative Estimation
  - Using historical observations, offline
- Step 2: Conditional Estimation
  - Use data from step 1 to train RKHS, online

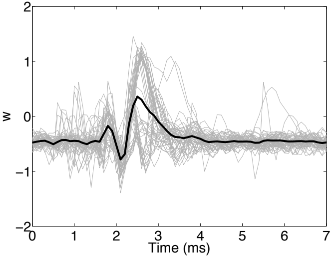
Example 3: Intracellular from Extracellular



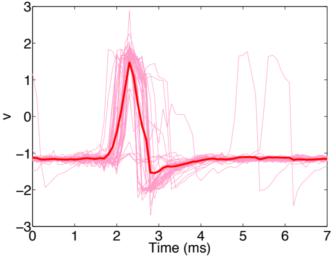
Example 3: Intracellular from Extracellular



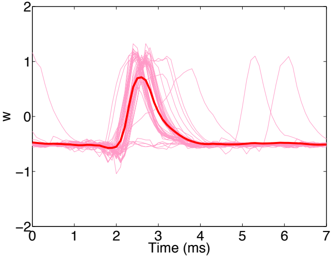
(a)



(b)



(c)



(d)

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

- Consider a 7-dim'l model for a column of atmosphere
  - Baroclinic anomaly potential temperatures, θ_1 and θ_2
  - Boundary layer anomaly potential temperature, θ_{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

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

Khouider, B., J. Biello, and A. J. Majda, 2010: A stochastic multicloud model for tropical convection.

Example 4: Multi-cloud "Satellite-like" obs

- Extrapolate anomaly potential temperature at height z

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

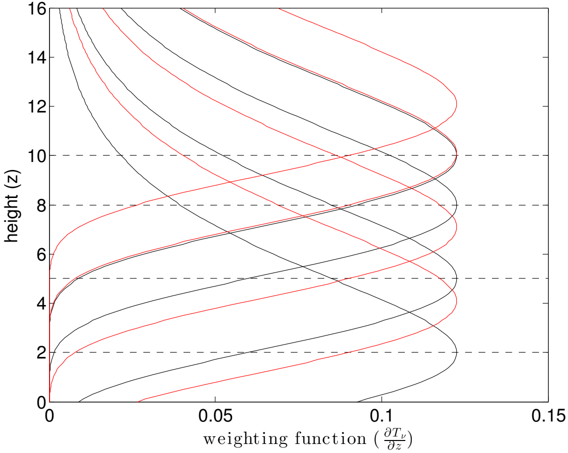
- Brightness temperature-like quantity at wavenumber-ν

$$\begin{aligned} h_\nu(x, f) = (1 - f_d - f_s)\Big[(1 - f_c)\big(\theta_{eb} T_\nu(0) + \int_0^{z_c} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\big) \\ + f_c T(z_c) T_\nu(z_c) + \int_{z_c}^{z_d} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz\Big] \\ + (f_d + f_s) T(z_d) T_\nu(z_d) + \int_{z_d}^{\infty} T(z)\frac{\partial T_\nu}{\partial z}(z)\,dz, \end{aligned} \tag{1}$$

- Setting $f = 0$ is the clear sky model

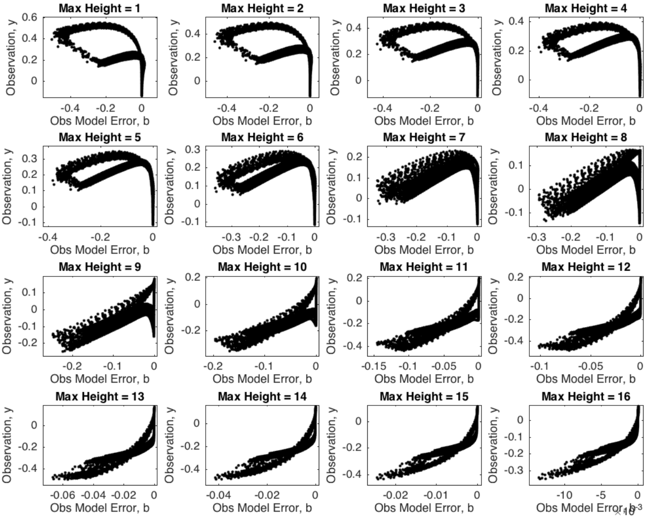
EXAMPLE 4: MULTI-CLOUD "SATELLITE-LIKE" OBS

- Weighting functions define RTM at different wavenumbers



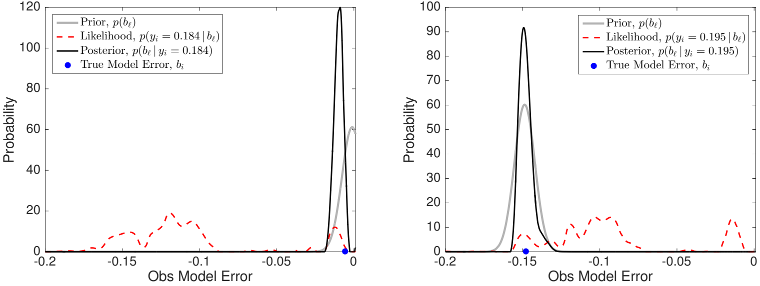
Example 4: Multi-cloud "Satellite-like" obs

- Biases at the 16 observed wavenumbers

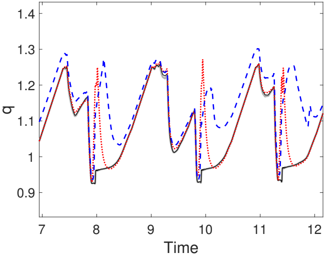


Example 4: Multi-cloud "Satellite-like" obs

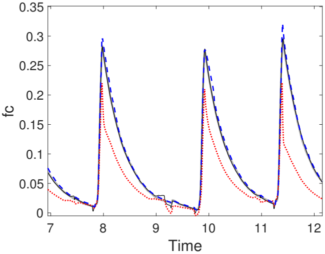
- Multimodal likelihood functions



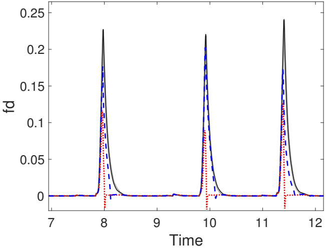
Example 4: "Satellite-like" obs (Iterative)

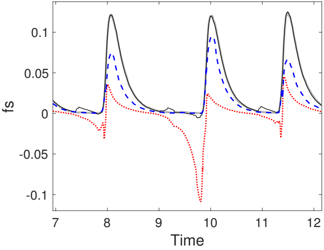


(e)

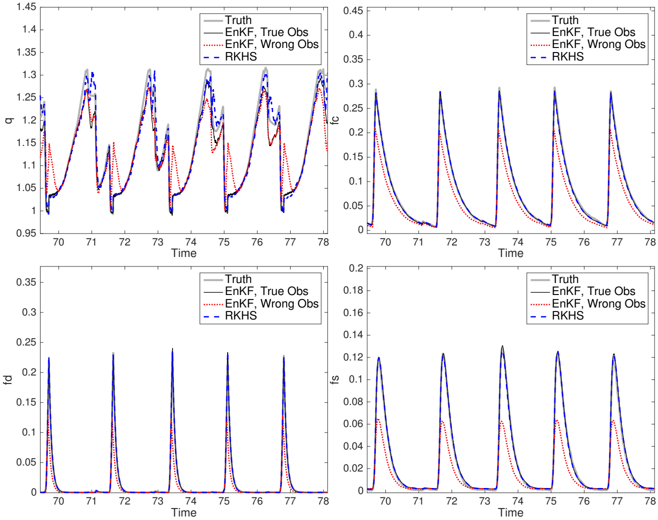


(f)





Example 4: "Satellite-like" obs (RKHS)



References

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

- J. Harlim, T. Berry, Correcting biased observation model error in data assimilation. Monthly Weather Review (2017).
- F. Hamilton, T. Berry, T. Sauer, Correcting Observation Model Error in Data Assimilation (preprint).
- F. Hamilton, T. Berry, T. Sauer, Tracking intracellular dynamics through extracellular measurements. PloS One (2018).