Correcting Biased Observation Model Error in Data Assimilation

John Harlim

Department of Mathematics and Department of Meteorology
The Pennsylvania State University

July 11, 2017

Biased observation model error problems

All the Kalman based DA method assumes unbiased observation model error, e.g.,

$$y_i = h(x_i) + \eta_i, \quad \eta_i \sim \mathcal{N}(0, R).$$

Suppose the operator h is un known. Instead, we are only given $\tilde{h}$, then

$$y_i = \tilde{h}(x_i) + b_i,$$

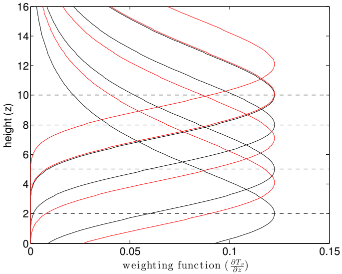
where we introduce a biased model error, $b_i = h(x_i) - \tilde{h}(x_i) + \eta_i$.

Example: Basic radiative transfer model

Consider solutions of the stochastic cloud model[1], $\{T(z), \theta_{eb}, q, f_d, f_s, f_c\}$.
Based on this solutions, define a basic radiative transfer model as follows,

$$h_\nu(x) = \theta_{eb} T_\nu(0) + \int_0^\infty T(z) \frac{\partial T_\nu}{\partial z}(z)\, dz,$$

where T_ν is the transmission between heights z to ∞ that is defined to depend on q.
The weighting function, $\frac{\partial T_\nu}{\partial z}$ are defined as follows:



[1] Khouider, Biello, Majda 2010

Example: Basic radiative transfer model

Suppose the deep and stratiform cloud top height is $z_d = 12$km, while the cumulus cloud top height is $z_c = 3$km. Define $f = \{f_d, f_c, f_s\}$ and $x = \{T(z), \theta_{eb}, q\}$. Then the cloudy RTM is given by,

$$\begin{aligned} h_\nu(x, f) &= (1 - f_d - f_s)\left[\theta_{eb} T_\nu(0) + \int_0^{z_d} T(z) \frac{\partial T_\nu}{\partial z}(z)\, dz\right] \\ &\quad + (f_d + f_s) T(z_t) T_\nu(z_d) + \int_{z_d}^{\infty} T(z) \frac{\partial T_\nu}{\partial z}(z)\, dz \end{aligned}$$

Example: Basic radiative transfer model

Suppose the deep and stratiform cloud top height is $z_d = 12$km, while the cumulus cloud top height is $z_c = 3$km. Define $f = \{f_d, f_c, f_s\}$ and $x = \{T(z), \theta_{eb}, q\}$. Then the cloudy RTM is given by,

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

Example: Basic radiative transfer model

Suppose the deep and stratiform cloud top height is $z_d = 12$km, while the cumulus cloud top height is $z_c = 3$km. Define $f = \{f_d, f_c, f_s\}$ and $x = \{T(z), \theta_{eb}, q\}$. Then the cloudy RTM is given by,

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

One can check that $h_\nu(x,0)$ corresponds to cloud-free RTM.

Observation model error in data assimilation

Suppose the observation is generated with

$$y_\nu = h_\nu(x, f) + \eta, \qquad \eta \sim \mathcal{N}(0, R)$$

The difficulty in estimating the cloud fractions, cloud top heights and (in reality we don't know precisely how many clouds under a column) induces model error.

Observation model error in data assimilation

Suppose the observation is generated with

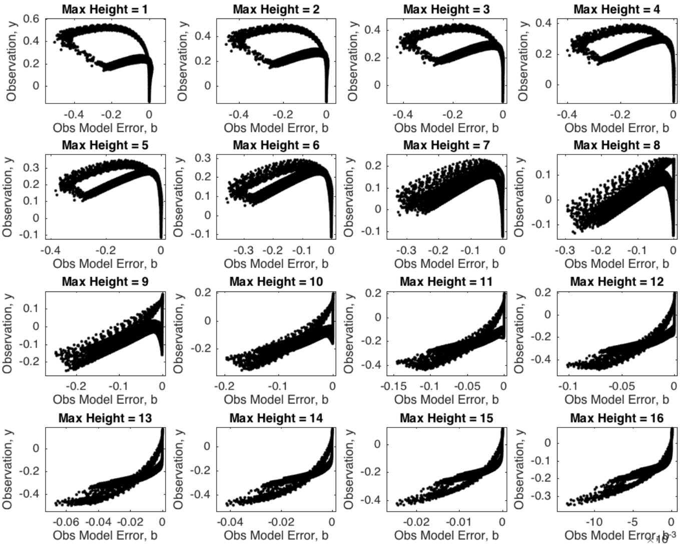
$$y_\nu = h_\nu(x, f) + \eta, \qquad \eta \sim \mathcal{N}(0, R)$$

The difficulty in estimating the cloud fractions, cloud top heights and (in reality we don't know precisely how many clouds under a column) induces model error.
In an extreme case, we consider filtering with a cloud-free RTM:

$$y_\nu = h_\nu(x, 0) + b_\nu$$

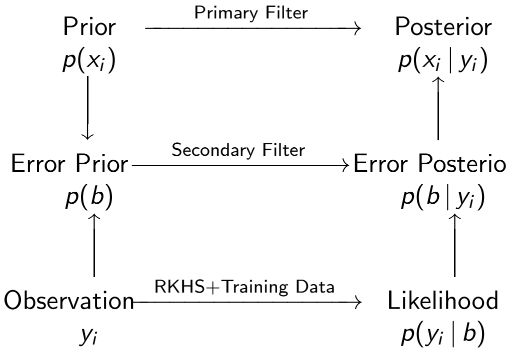
where $b_\nu = h_\nu(x, f) - h_\nu(x, 0) + \eta$ is model error with bias.

Observations (y_ν) v Model error (b_ν)



State estimation of the model error

We propose a secondary filter to estimate the statistics for b_i as follows:

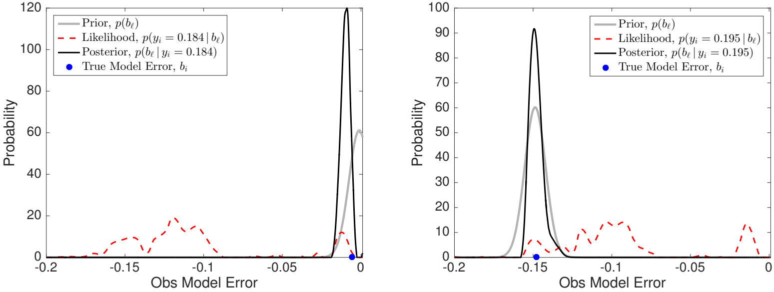


A machine learning technique, kernel embedding of conditional distribution[2], is employed to train a nonparametric likelihood function.

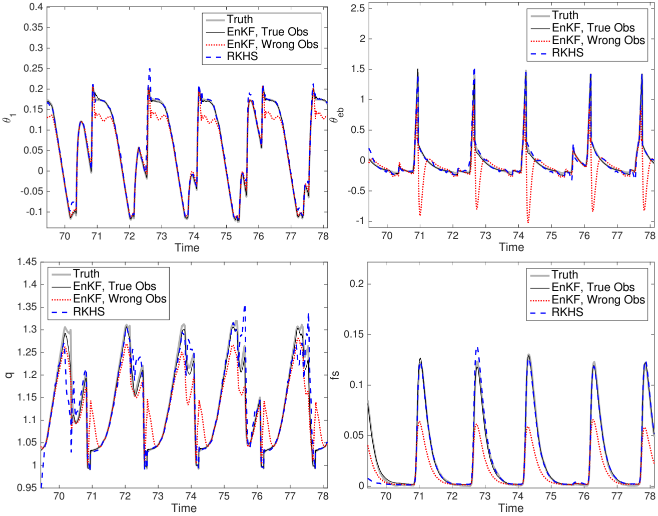
[2] Song, Fukumizu, Gretton, 2013.

Secondary Bayesian filter

$$p(b|y_i) \propto p(b)p(y_i|b)$$

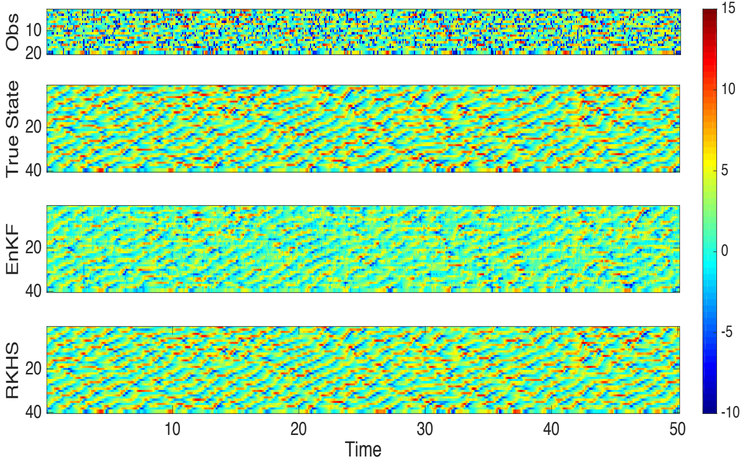


Filter estimates (with adaptive tuning of R and Q).



Example: Lorenz-96

Biased occurs random in space and times.



Nonparametric likelihood functions

We will use the kernel embedding of conditional distribution.[3]

Let X be a r.v on $\mathcal{M}$ and distribution $P(X)$. Given a kernel $K : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$, the Moore-Aronszajn theorem states that there exists a Reproducing Kernel Hilbert Space (RKHS) $L^2(\mathcal{M}, q)$.

This means that we can evaluate any function $f \in L^2(\mathcal{M}, q)$ as follows:

$$f(x) = \langle f, K(x, \cdot) \rangle_q.$$

[3] Song, Fukumizu, Gretton, 2013.

Nonparametric likelihood functions

The kernel embedding of conditional distribution $P(Y|B)$ is defined as,

$$\mu_{Y|b} = \mathbb{E}_{Y|b}[\tilde{K}(Y,\cdot)] = \int_{\mathcal{N}} \tilde{K}(y,\cdot)dP(y|b).$$

Given $g \in L^2(\mathcal{N}, \tilde{q})$,

$$\begin{aligned} \mathbb{E}_{Y|b}[g(Y)] &= \int_{\mathcal{N}} g(y)dP(y|b) = \int_{\mathcal{N}} \langle g, \tilde{K}(y,\cdot)\rangle_{\tilde{q}} dP(y|b) \\ &= \Big\langle g, \int_{\mathcal{N}} \tilde{K}(y,\cdot)dP(y|b)\Big\rangle_{\tilde{q}} = \langle g, \mu_{Y|b}\rangle_{\tilde{q}}. \end{aligned}$$

One can verify that

$$\mu_{Y|b} = q\mathcal{C}_{YB}\mathcal{C}_{BB}^{-1}K(b,\cdot),$$

where

$$\mathcal{C}_{BY} = \int_{\mathcal{M}\times\mathcal{N}} K(b,\cdot)\otimes\tilde{K}(y,\cdot)\, dP(b,y)$$

is the kernel embedding of $P(B,Y)$ on appropriate Hilbert spaces.

Data-driven nonparametric likelihood functions

Given $\{b_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$, apply the *diffusion maps*[4] to learn the data-driven orthonormal basis functions $\varphi_j(b) \in L^2(\mathcal{M}, q)$ and $\tilde{\varphi}_k(y) \in L^2(\mathcal{M}, \tilde{q})$. Let

$$p(y|b) = \sum_k \mu_{Y|b,k} \tilde{\varphi}_k(y) \tilde{q}(y)$$

where

$$\begin{aligned} \mu_{Y|b,k} &= \langle p(\cdot|b), \tilde{\varphi}_k \rangle = \mathbb{E}_{Y|b}[\tilde{\varphi}_k] = \langle \mu_{Y|b}, \tilde{\varphi}_k \rangle_{\tilde{q}} \\ &= \langle q \mathcal{C}_{YB} \mathcal{C}_{BB}^{-1} K(b, \cdot), \tilde{\varphi}_k \rangle_{\tilde{q}} \\ &= \ldots \\ &= \sum_j \varphi_j(x) [C_{YB} C_{BB}^{-1}]_{kj} \end{aligned}$$

where

$$[C_{YB}]_{jk} = \langle \mathcal{C}_{YB}, \tilde{\varphi}_j \otimes \varphi_k \rangle_{\tilde{q} \otimes q} \approx \frac{1}{N} \sum_{i=1}^N \tilde{\varphi}_j(y_i) \varphi_k(b_i),$$

$$[C_{BB}]_{jk} = \langle \mathcal{C}_{BB}, \varphi_j \otimes \varphi_k \rangle_q \approx \frac{1}{N} \sum_{i=1}^N \varphi_j(b_i) \varphi_k(b_i),$$

[4] Coifman & Lafon 2006, Berry & H, 2016.

A manifold learning algorithm: Diffusion Maps[6]

Given $\{x_i\} \in \mathcal{M} \subset \mathbb{R}^n$ with a sampling measure q, the diffusion maps algorithm is a kernel based method that produces orthonormal basis functions on the manifold, $\varphi_k \in L^2(\mathcal{M}, q)$.

[5] Berry & H, 2016

[6] Coifman & Lafon 2006

A manifold learning algorithm: Diffusion Maps[6]

Given $\{x_i\} \in \mathcal{M} \subset \mathbb{R}^n$ with a sampling measure q, the diffusion maps algorithm is a kernel based method that produces orthonormal basis functions on the manifold, $\varphi_k \in L^2(\mathcal{M}, q)$.

These basis functions are solutions of an eigenvalue problem,

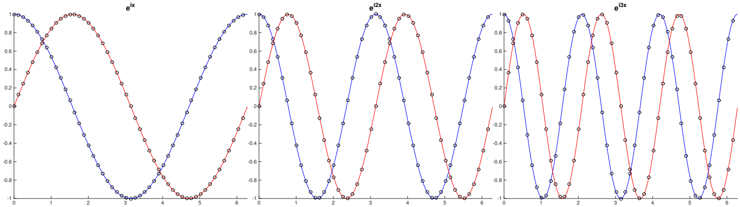
$$q^{-1}\text{div}\Big(q\nabla\varphi_k(x)\Big) = \lambda_k\varphi_k(x),$$

where the weighted Laplacian operator is approximated with an integral operator using a variable bandwidth kernel[5]
with appropriate normalizations.

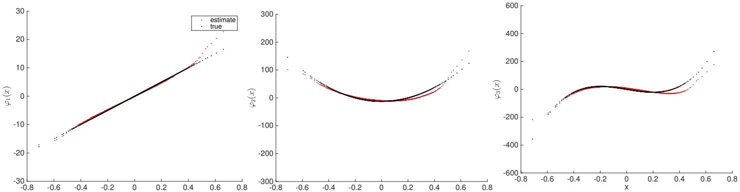
[5] Berry & H, 2016
[6] Coifman & Lafon 2006

Examples:

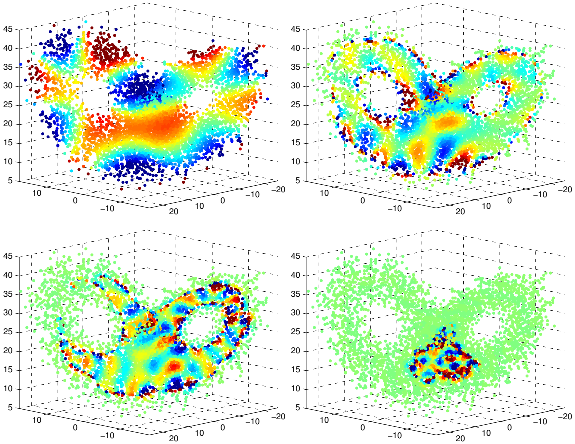
Example: Uniformly distributed data on a circle, we obtain the Fourier basis.



Example: Gaussian distributed data on a real line, we obtain the Hermite polynomials.



Example: Nonparametric basis functions estimated on nontrivial manifold



Remark: Essentially, one can view the DM as a method to learn generalized Fourier basis on the manifold.

References:

1. T. Berry & H, "Correcting biased observation model error in data assimilation", Mon. Wea. Rev. (in press).
2. T. Berry & H, "Variable bandwidth diffusion kernels", Appl. Comput. Harmon. Anal. 40, 68-96, 2016.
3. H, "An introduction to data-driven methods for stochastic modeling of dynamical systems", Springer (submitted for a review).