Analysis of Hydrologic Time Series Reconstruction Uncertainty Due to Inverse Model Inadequacy

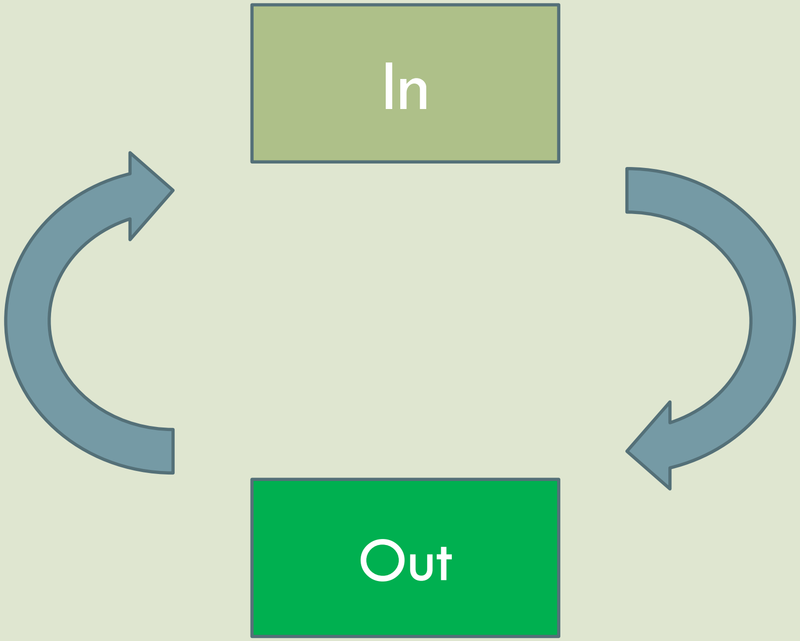
Using the Laguerre Expansion Method

Jiachuan He[1,2], **Scott K. Hansen**[2], and Velimir V. Vesselinov[2]
[1]Institute for Computational Engineering and Sciences, University of Texas at Austin
[2]Computatonal Earth Science Group (EES-16) Los Alamos National Laboratory

LA-UR-17-20066

Talk outline

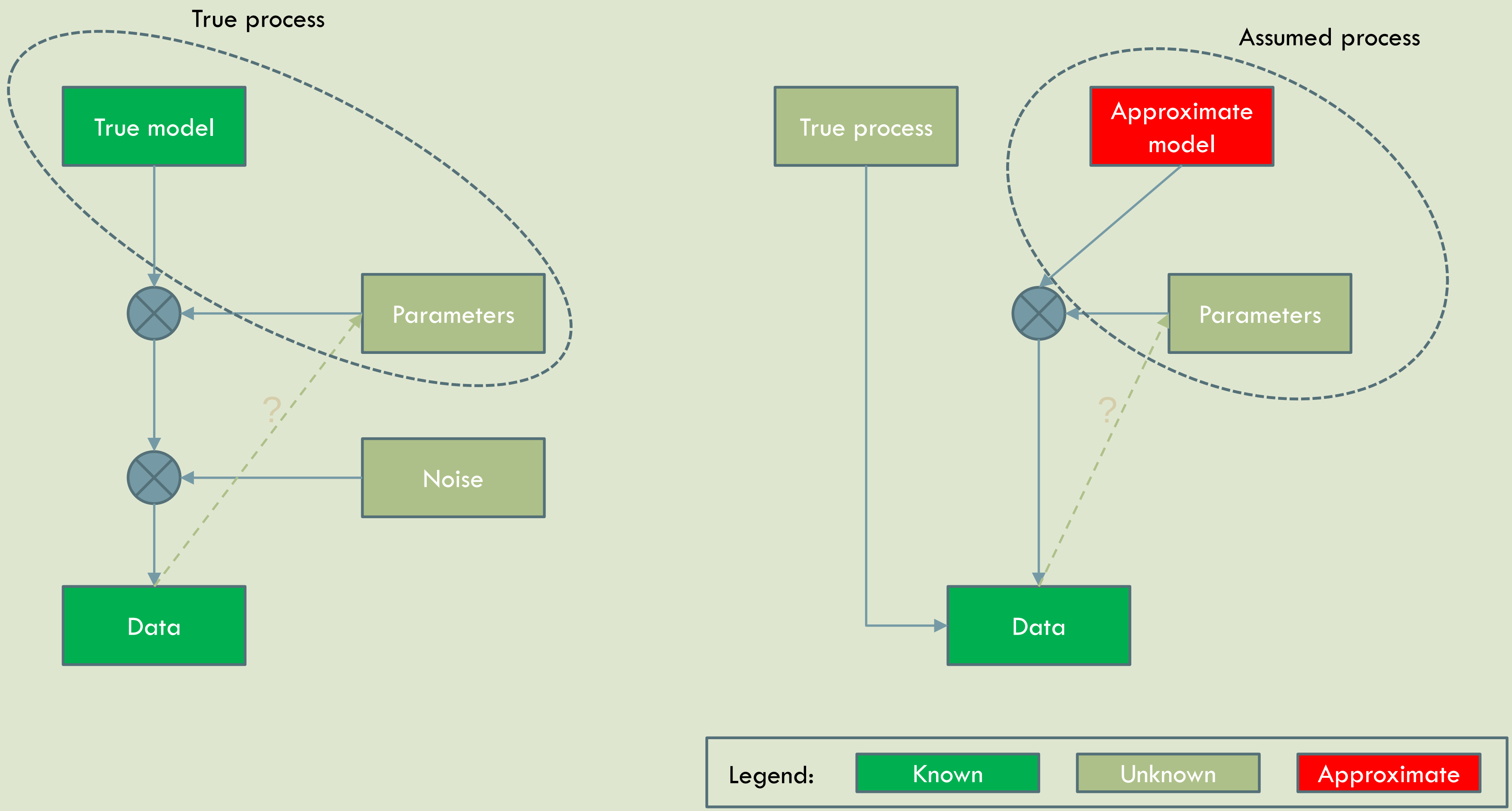
1. What sort of problems are we considering?
2. What is the Laguerre Expansion Method?
3. Derivation of error bounds
4. Numerical study



Good data, bad models?

Data error conception

Model error conception

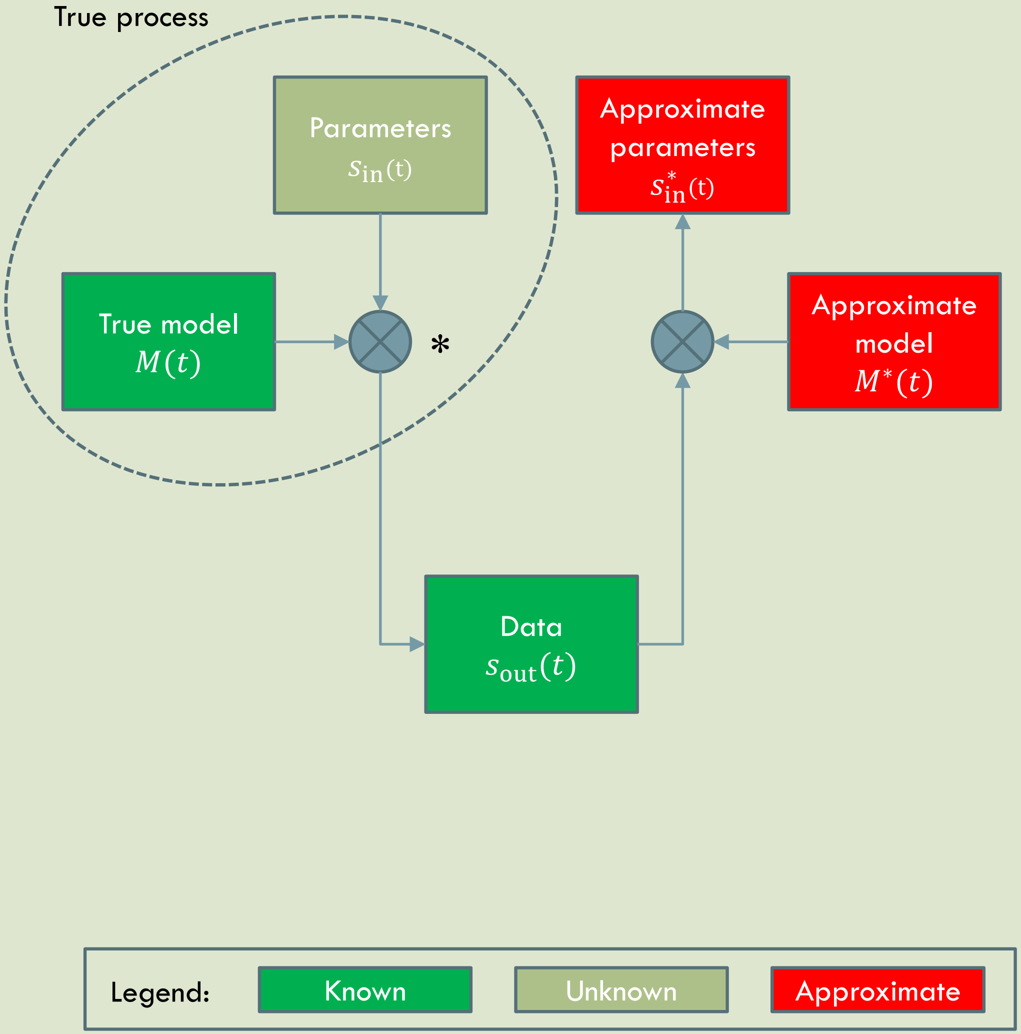


Model error

Important, but difficult to quantify: need to consider the **"space of all plausible models"**

We consider an important **special class** of problems with convolution structure:

$$s_{\text{out}}(t) = \int_0^t M(t-\tau)s_{\text{in}}(\tau)\,d\tau$$



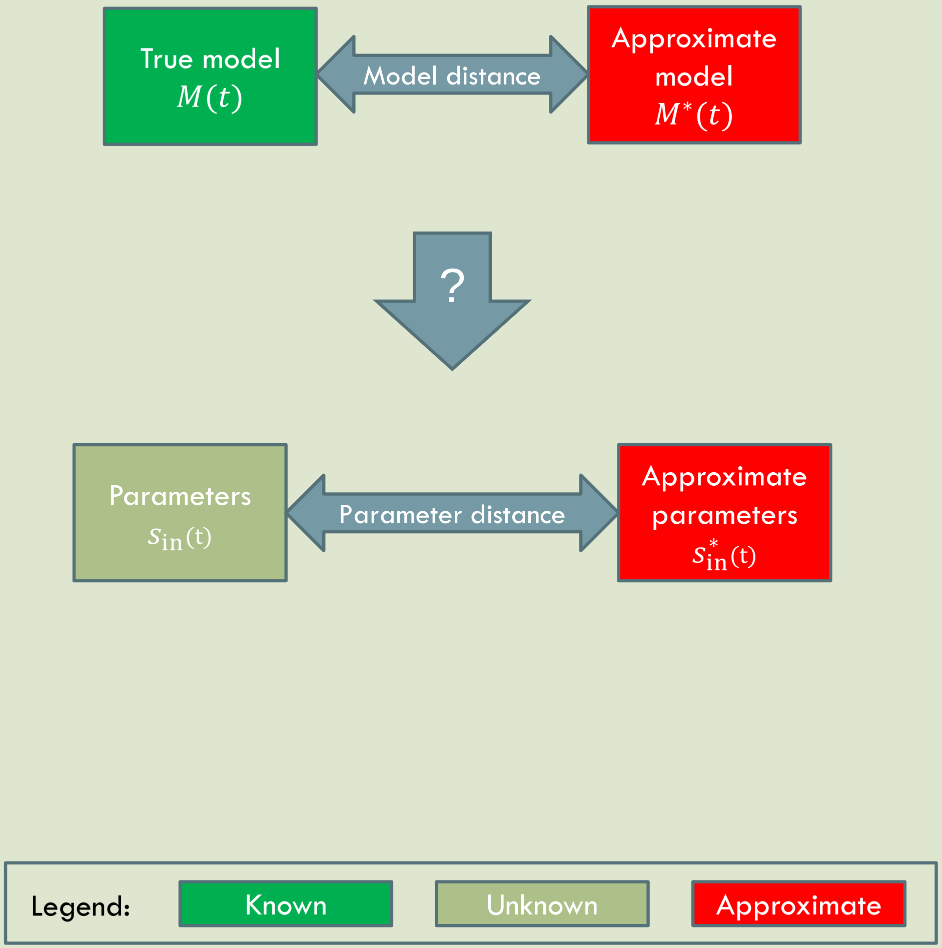
Model error

Here, a perfect model would generate perfect reconstruction.

We wish to bound the **parameter distance** based on the **model distance**.

We *define* the parameter distance based on the L2 distance of of s_{in} and s_{in}^*:

$$\int_0^\infty (s_{\mathrm{in}}(\mathrm{t}) - s_{\mathrm{in}}^*(\mathrm{t}))^2 dt$$



Analytic strategy in one slide

1. Expand functions as Fourier series in basis of scaled (by T) Laguerre functions, $\phi_n(t/T)$, work with their coefficients:

Function	Vector
$s_{\text{in}}(\text{t}) = \sum a_n \, \phi_n(t/T)$	$\mathbf{a}$
$s_{\text{in}}^*(\text{t}) = \sum a_n^* \phi_n(t/T)$	$\mathbf{a}^*$
$M(t) = \sum b_n^* \phi_n(t/T)$	$\mathbf{b}$
$M^*(t) = \sum b_n^* \phi_n(t/T)$	$\mathbf{b}^*$
$s_{\text{out}}(\text{t}) = \sum c_n \, \phi_n(t/T)$	$\mathbf{c}$
$s_{\text{out}}^*(\text{t}) = \sum c_n^* \phi_n(t/T)$	$\mathbf{c}^*$

2. Our forward problem,

$$s_{\text{out}}(t) = \int_0^t M(t-\tau) s_{\text{in}}(\tau)\, d\tau,$$

translates to

$$\mathbf{c} = \mathbf{B}\mathbf{a}$$

where

$$\mathbf{B} = T \begin{bmatrix} \mathbf{b}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{b}_1 - \mathbf{b}_0 & \mathbf{b}_0 & \mathbf{0} & & \mathbf{0} \\ \mathbf{b}_2 - \mathbf{b}_1 & \mathbf{b}_1 - \mathbf{b}_0 & \mathbf{b}_0 & & \mathbf{0} \\ \vdots & & & \ddots & \vdots \\ \mathbf{b}_N - \mathbf{b}_{N-1} & \mathbf{b}_{N-1} - \mathbf{b}_{N-2} & \mathbf{b}_{N-2} - \mathbf{b}_{N-3} & \cdots & \mathbf{b}_0 \end{bmatrix}$$

3. By Parseval's theorem:

$$\int_0^\infty (s_{\text{in}}(\text{t}) - s_{\text{in}}^*(\text{t}))^2 dt = \|\mathbf{a} - \mathbf{a}^*\|_2^2$$

The Laguerre Functions

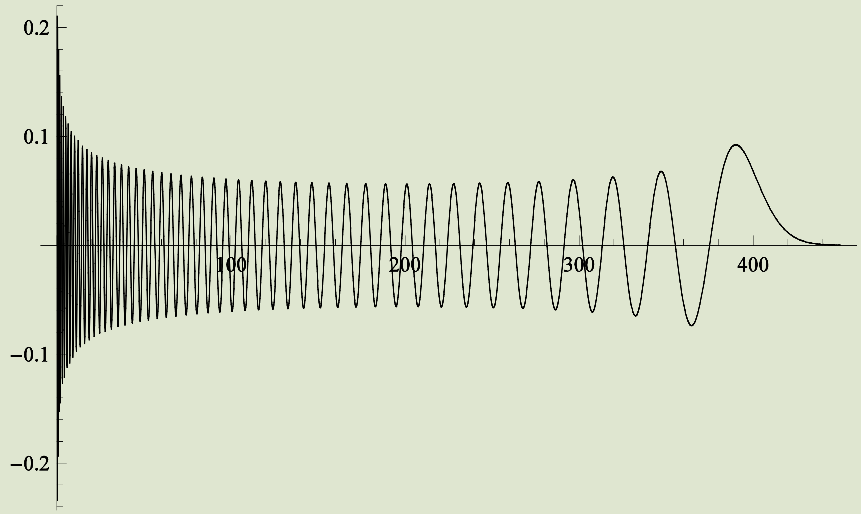
- Laguerre functions are defined:

$$\phi_n(t) = \frac{e^{t/2}}{n!}\frac{d^n}{dt^n}\{e^{-t}t^n\}, \qquad n = 0,1,2,\ldots$$

- They form an orthogonal basis on $[0,\infty)$:

$$\int_0^\infty \phi_m(t)\phi_n(t)\,dt = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

- They allow “Fourier” analysis on time series



The 100th Laguerre function

Error bounds

$\mathbf{B}$ and $\mathbf{B}^*$ have full rank, and so are invertible:

$$\|\mathbf{a} - \mathbf{a}^*\|_2 = \left\|\left(\mathbf{I} - \mathbf{B}^{*-\mathbf{1}}\mathbf{B}\right)\mathbf{a}\right\|_2$$

We can prove (this is not obvious) that $\mathbf{I} - \mathbf{B}^{*-\mathbf{1}}\mathbf{B}$ is itself a lower triangular Toeplitz matrix.

Using the identity $\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}$, we can directly compute the main and sub-diagonal elements of $\mathbf{I} - \mathbf{B}^{*-\mathbf{1}}\mathbf{B}$.

Lower error bounds

Using this, we arrive at reconstruction error lower bounds which are based on dominant components of model error:

$$|\mathbf{a}_0| \left| 1 - \frac{\mathbf{b}_0}{\mathbf{b}^*{}_0} \right| \leq \|\mathbf{a} - \mathbf{a}^*\|_2$$

$$\left| \mathbf{a}_0 \left(1 - \frac{\mathbf{b}_0}{\mathbf{b}^*{}_0} \right) \right|^2 + \left| \frac{\mathbf{a}_0}{\mathbf{b}^*{}_0} \left(\mathbf{b}_1 - \mathbf{b}^*{}_1 \left(\frac{\mathbf{b}_0}{\mathbf{b}^*{}_0} \right) \right) + \mathbf{a}_1 \left(1 - \frac{\mathbf{b}_0}{\mathbf{b}^*{}_0} \right) \right|^2 \leq \|\mathbf{a} - \mathbf{a}^*\|_2^2$$

In the special case of $s_{\text{in}}(\text{t}) = e^{-t/2T}$ (and we may select T arbitrarily):

$$\left| 1 - \frac{\mathbf{b}_0}{\mathbf{b}^*{}_0} \right| \leq \frac{\|\mathbf{a} - \mathbf{a}^*\|_2}{\|\mathbf{a}\|_2}$$

Multiple collection locations

If we have a single source and m data collection locations, the problem is over-determined. *Define* the solution to minimize

$$\sum_{i=1}^{m} \|\mathbf{c} - \mathbf{c}^{\,*}\|_2^2 \,.$$

We have multiple system matrices, $\mathbf{B}_1$, $\mathbf{B}_2$, …, $\mathbf{B}_m$, and model matrices, $\mathbf{B}_1^*$, $\mathbf{B}_2^*$, …, $\mathbf{B}_m^*$. Define

$$\mathbf{B}_\otimes = \begin{bmatrix} \mathbf{B}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{B}_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}_m \end{bmatrix}, \qquad \mathbf{B}_\otimes^* = \begin{bmatrix} \mathbf{B}_1^* & 0 & \cdots & 0 \\ 0 & \mathbf{B}_2^* & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}_m^* \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{I}_N \\ \vdots \\ \mathbf{I}_N \end{bmatrix}$$

Multiple collection locations

By straightforward computation, we can show that the solution is

$$\|\mathbf{a} - \mathbf{a}^*\|_2 = \frac{1}{m}\left\|\left(\mathbf{I}_{mN} - \mathbf{B}_{\otimes}^{*-1}\mathbf{B}_{\otimes}\right)\mathbf{D}\mathbf{a}\right\|_2$$

Which can be rewritten as

$$\|\mathbf{a} - \mathbf{a}^*\|_2 = \frac{1}{m}\sum_{i=1}^{m}\left\|\left(\mathbf{I}_N - \mathbf{B}_i^{*-1}\mathbf{B}_i\right)\mathbf{a}\right\|_2$$

We see no expected utility in additional monitoring locations.

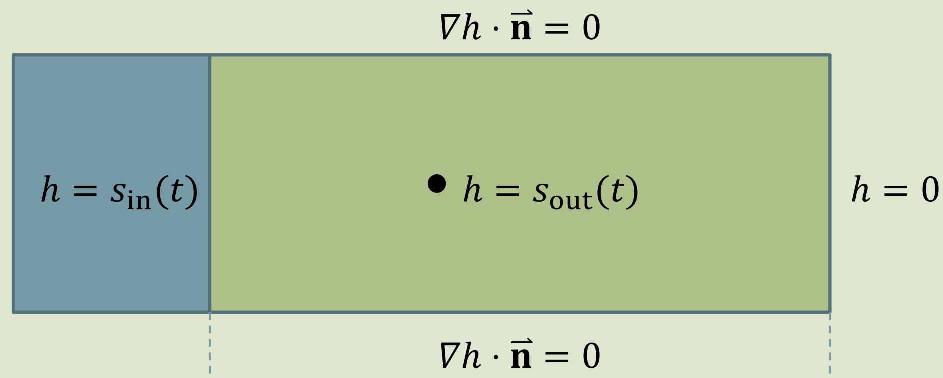
Monte Carlo numerical setup

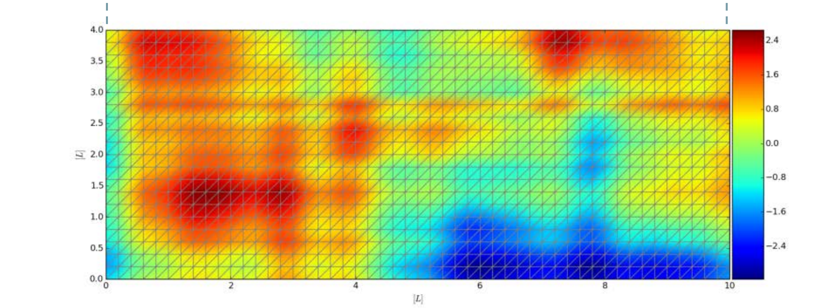
500 aquifers defined by **multi-Gaussian K-fields** (known mean) in hydraulic connection with a river, with a single well in the midst of each.

River level transient is always $s_{\mathrm{in}}(t) = \exp(-t/T)$, for fixed T.

Well level for each is $s_{\mathrm{out}}(t)$.

Interpretive model has identical geometry, but **homogeneous K-field** with correct mean.

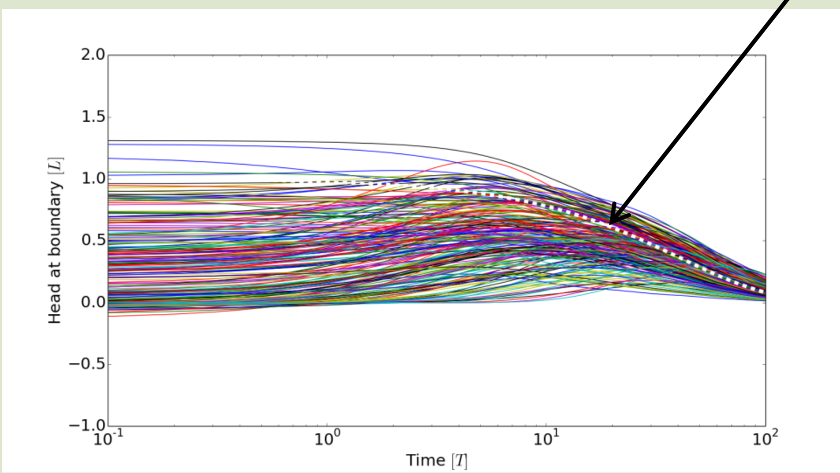




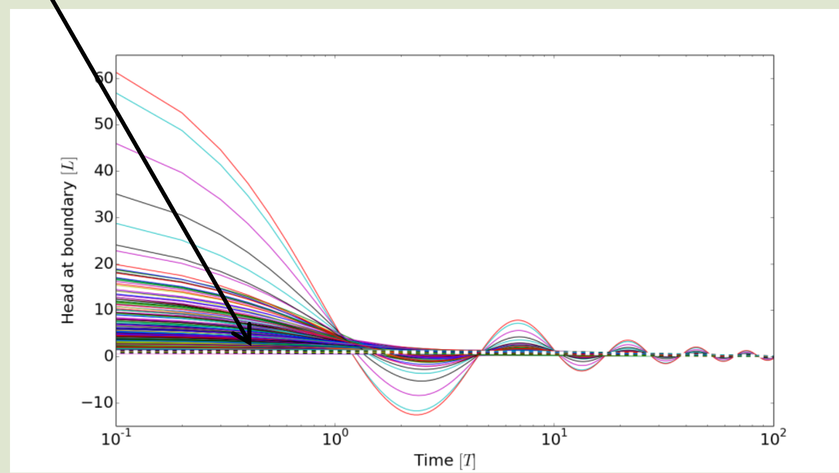
Example K-field realization

Numerical reconstructions: $s_{\text{in}}^{*}(t)$

True $s_{\text{in}}(t)$

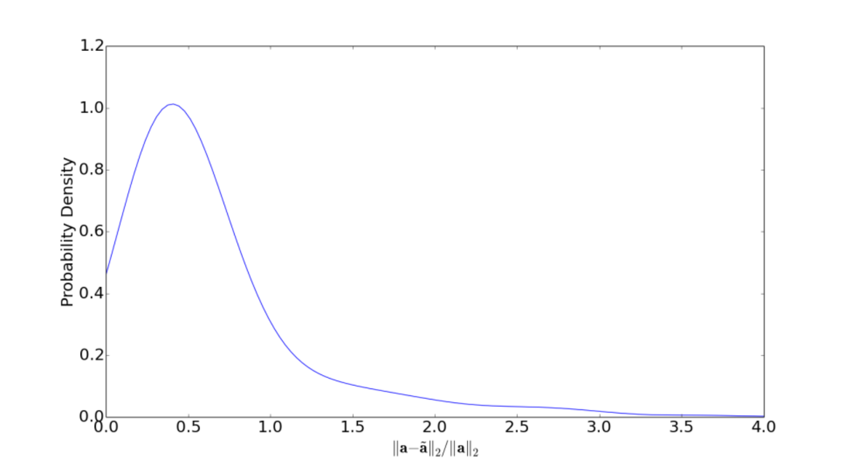


Interpretive model (M^*) peak
before true model (M) peak

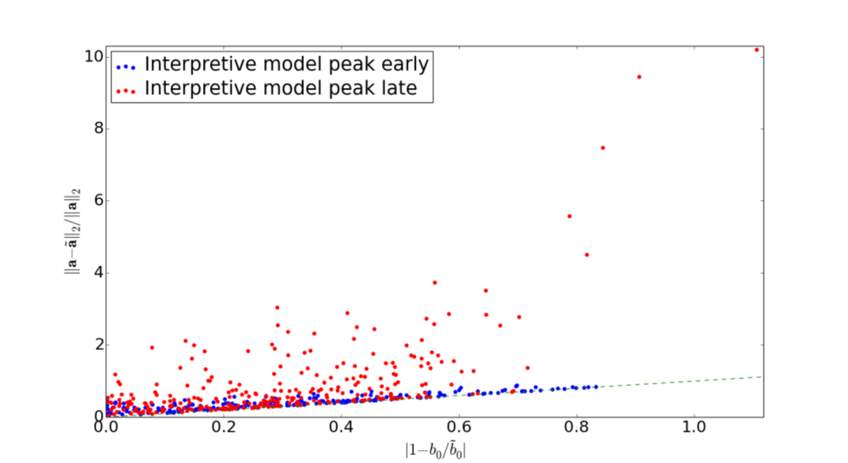


Interpretive model (M^*) peak
after true model (M) peak

L2 reconstruction error



Empirical pdf of normalized reconstruction error



Normalized reconstruction error versus lower error bound

Key points

1. Model error is an often-overlooked factor in geophysical inverse problems.
2. Laguerre expansion allows conversion of convolution inverse problems into matrix inverse problems with a triangular Toeplitz matrix.
3. This structure allows us to compute error bounds using only the **most significant components** of model error.
4. Additional data collection is not expected to improve estimation reliability.
5. We identified a criterion for blind model identification.