More data is not always better: Why and how feature-based data assimilation can be useful.

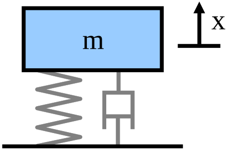
Spence Lunderman

Matthias Morzfeld, Jesse Adams, and Rafael Orozco

University of Arizona

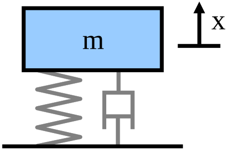
April 18, 2018

A simple data assimilation example



$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

A simple data assimilation example



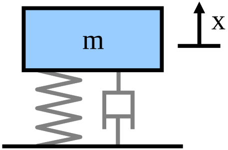
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

Noisy observations every
$\Delta t = 0.5$ seconds

$$z_k = x(\Delta t \cdot k) + \eta_k,$$

for $\eta_k \sim \mathcal{N}\left(0, 1E-3\right)$, i.i.d.

A simple data assimilation example

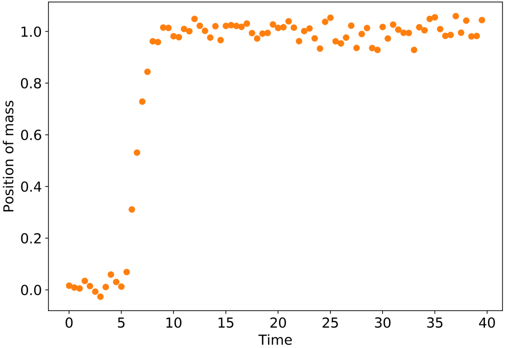


$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

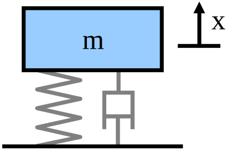
Noisy observations every
$\Delta t = 0.5$ seconds

$$z_k = x(\Delta t \cdot k) + \eta_k,$$

for $\eta_k \sim \mathcal{N}\left(0, 1E-3\right)$, i.i.d.



A simple data assimilation example

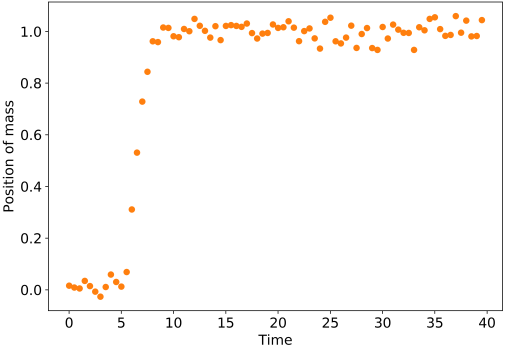


$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

Noisy observations every
$\Delta t = 0.5$ seconds

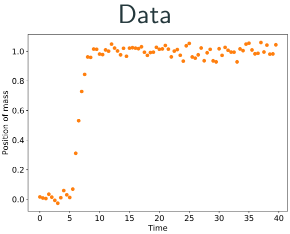
$$z_k = x(\Delta t \cdot k) + \eta_k,$$

for $\eta_k \sim \mathcal{N}\left(0, 1E-3\right)$, i.i.d.



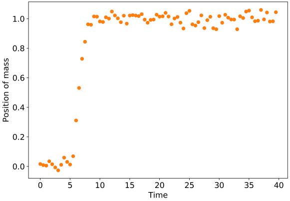
True parameters: $\begin{pmatrix} \zeta \\ \omega \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$

A simple data assimilation example



A simple data assimilation example

Data

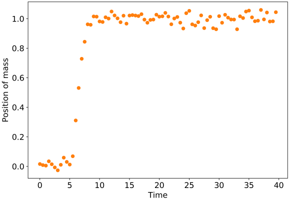


Model

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

A simple data assimilation example

Data



Model

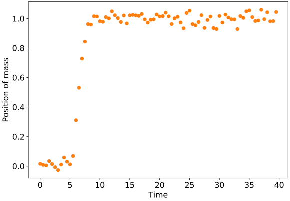
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

Bayes' Rules

$$p\left(\zeta, \omega \mid \text{data}\right) \propto p\left(\text{data} \mid \zeta, \omega\right) p\left(\zeta, \omega\right)$$

A simple data assimilation example

Data



Model

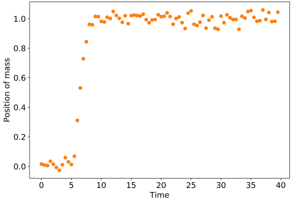
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

Bayes' Rules

$$p\left(\zeta, \omega \mid \text{data}\right) \propto p\left(\text{data} \mid \zeta, \omega\right) p\left(\zeta, \omega\right)$$

A simple data assimilation example

Data



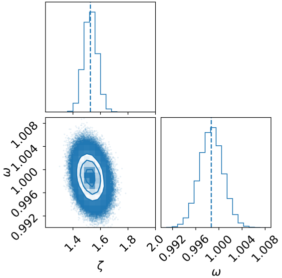
Model

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

Bayes' Rules

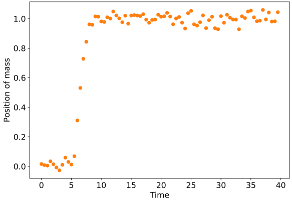
$$p\Big(\zeta, \omega \,\Big|\, \text{data}\Big) \propto p\Big(\text{data} \,\Big|\, \zeta, \omega\Big)\, p\Big(\zeta, \omega\Big)$$

Results:



A simple data assimilation example

Data



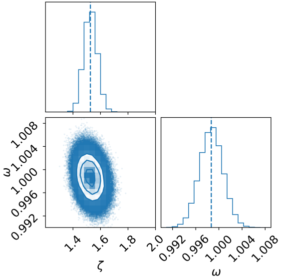
Model

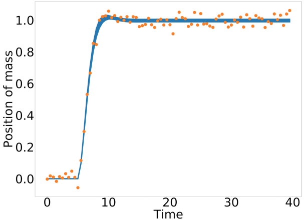
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t-5)$$

Bayes' Rules

$$p\left(\zeta,\omega \mid \text{data}\right) \propto p\left(\text{data} \mid \zeta,\omega\right) p\left(\zeta,\omega\right)$$

Results:

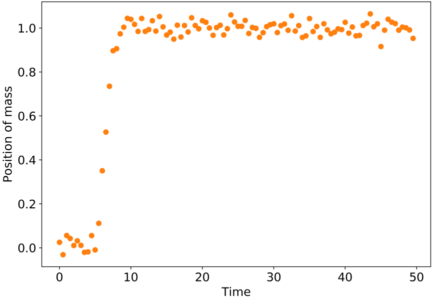




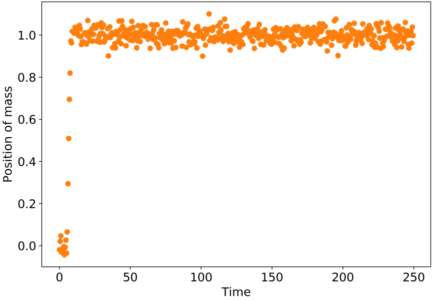
Trajectories of 50 samples from the posterior distribution

Is steady state data "redundant" in parameter estimation?

Recording data for 50 seconds:

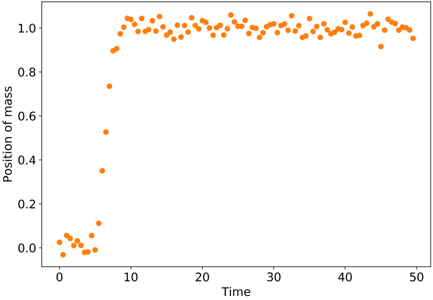


Recording data for 250 seconds:

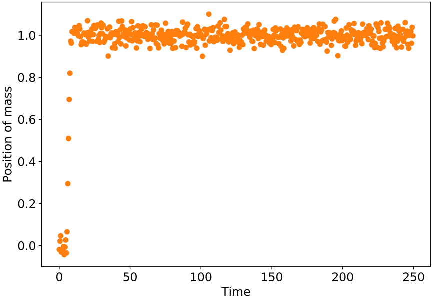


Is steady state data "redundant" in parameter estimation?

Recording data for 50 seconds:



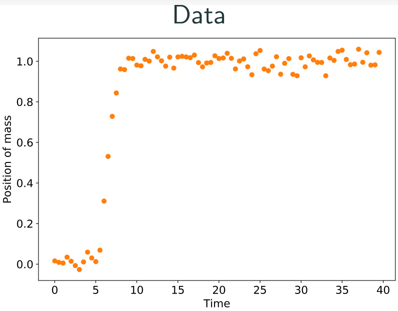
Recording data for 250 seconds:



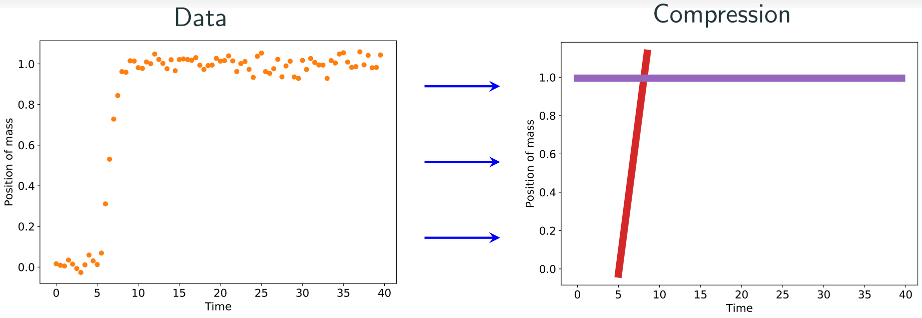
Seconds of data	ζ estimation	ω estimation
50	1.48 ± 0.04	0.99 ± 0.03
100	1.48 ± 0.04	0.99 ± 0.03
150	1.47 ± 0.04	0.99 ± 0.03
200	1.48 ± 0.04	0.99 ± 0.03
250	1.47 ± 0.04	0.99 ± 0.03

Averaged over 100 experiments

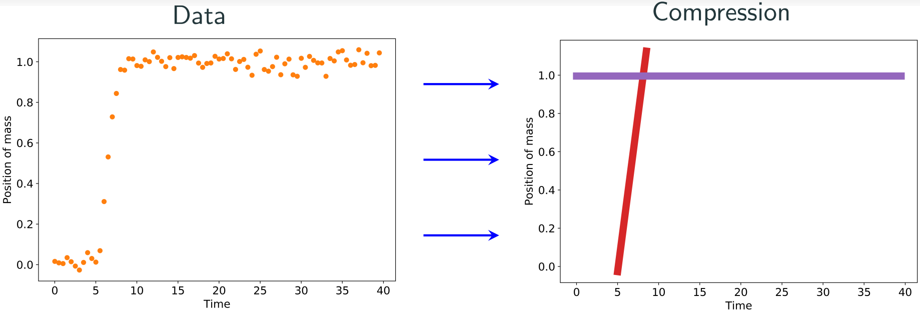
Compressing the data further



Compressing the data further



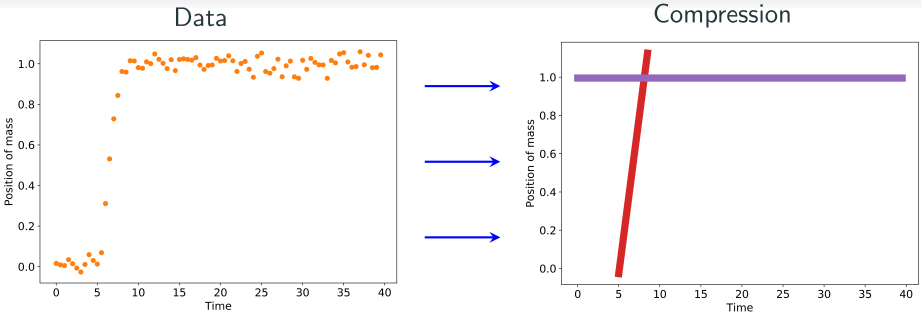
Compressing the data further



From the data, we extract two features

1. The slope of a linear fit to the 7 data points collected after $t = 5$.
2. The average of the last 25 seconds of data.

Compressing the data further



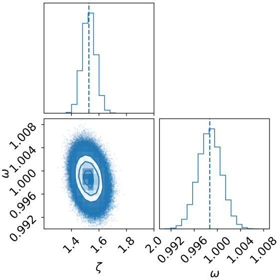
From the data, we extract two features

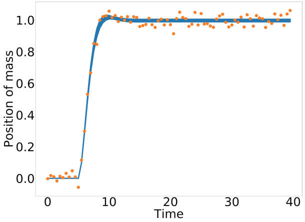
1. The slope of a linear fit to the 7 data points collected after $t = 5$.
2. The average of the last 25 seconds of data.

What is the distribution of parameters ζ, ω given the compressed data?

Comparing Results

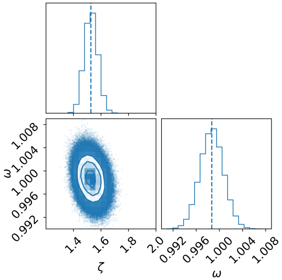
Assimilation without compression

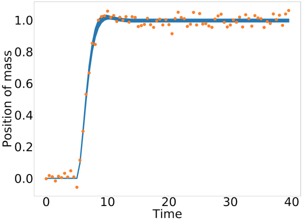




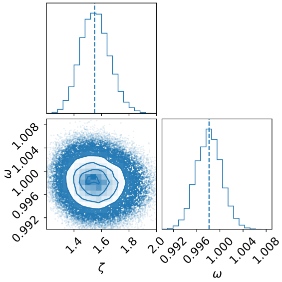
Comparing Results

Assimilation without compression



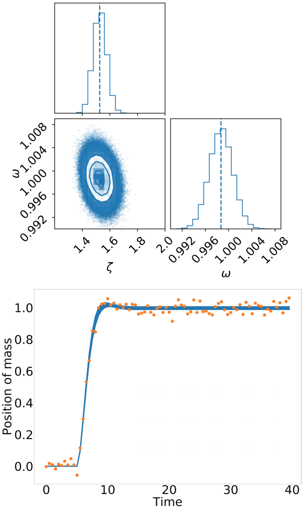


Assimilation with compression

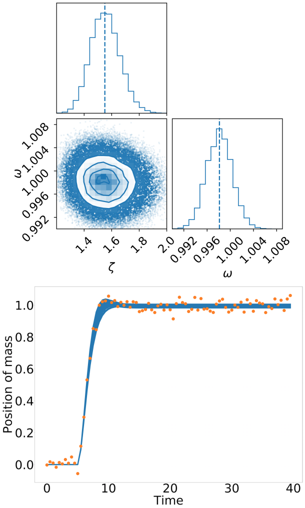


Comparing Results

Assimilation without compression

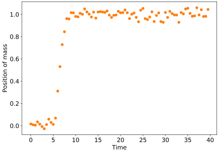


Assimilation with compression



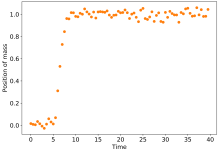
Feature-based data assimilation

Data



Feature-based data assimilation

Data

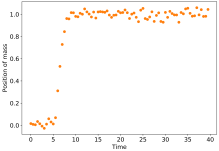


Model

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature-based data assimilation

Data



Model

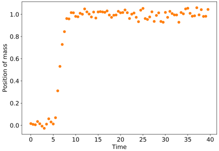
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

$$\mathcal{F}\big(\text{ data }\big) = \text{ features}$$

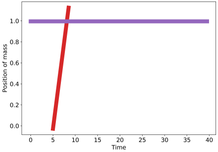
Feature-based data assimilation

Data



$\longrightarrow$
$\longrightarrow$
$\longrightarrow$

Feature selection



Model

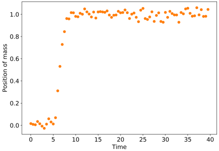
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

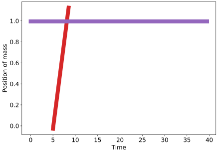
$$\mathcal{F}\Big(\text{ data }\Big) = \text{ features}$$

Feature-based data assimilation

Data



Feature selection



Model

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

$$\mathcal{F}(\text{ data }) = \text{ features}$$

Bayes' Rule: $$p\Big(\zeta, \omega \,|\, \mathcal{F}(\text{ data })\Big) \propto p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta, \omega\Big) p\Big(\zeta, \omega\Big)$$

Defining the feature-based likelihood

$$p\Big(\mathcal{F}(\text{ data }) \mid \zeta, \omega\Big) \sim \mathcal{N}\Big(\mathcal{F}(\zeta,\omega), R\Big)$$

Defining the feature-based likelihood

$$p\Big(\mathcal{F}(\text{ data })\,|\,\zeta\,,\omega\Big) \sim \mathcal{N}\Big(\mathcal{F}(\zeta,\omega)\,,\,R\Big)$$

Estimate R:

Defining the feature-based likelihood

$$p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta\,,\omega\Big) \sim \mathcal{N}\Big(\mathcal{F}(\zeta,\omega)\,,\; R\Big)$$

Estimate R:

1. Generate perturbed data, $z^j = \text{ data} + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, Q)$

Defining the feature-based likelihood

$$p\Big(\mathcal{F}(\text{ data })\,|\,\zeta\,,\omega\Big) \sim \mathcal{N}\Big(\mathcal{F}(\zeta,\omega)\,,\,R\Big)$$

Estimate R:

1. Generate perturbed data, $z^j = \text{ data} + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, Q)$
2. Each perturbed data leads to a perturbed feature: $f_j = \mathcal{F}(\,z_j\,)$

Defining the feature-based likelihood

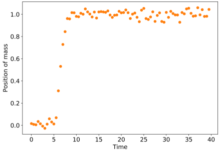
$$p\Big(\mathcal{F}(\text{ data })\,|\,\zeta,\omega\Big) \sim \mathcal{N}\Big(\mathcal{F}(\zeta,\omega),\, R\Big)$$

Estimate R:

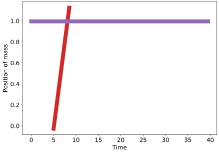
1. Generate perturbed data, $z^j = \text{ data} + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, Q)$
2. Each perturbed data leads to a perturbed feature: $f_j = \mathcal{F}(\, z_j\,)$
3. Define R as the sample covariance of these perturbed features (f_j).

Feature-based data assimilation

Data



Feature selection



Model

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

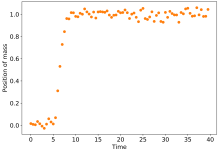
Feature function

$$\mathcal{F}(\text{ data }) = \text{ features}$$

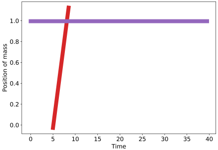
Bayes' Rule: $$p\Big(\zeta,\omega \,|\, \mathcal{F}(\text{ data })\Big) \propto p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta,\omega\Big)p\Big(\zeta,\omega\Big)$$

Feature-based data assimilation

Data



Feature selection



Model

$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

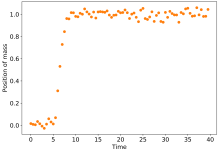
$$\mathcal{F}(\text{ data }) = \text{ features}$$

Bayes' Rule: $p\Big(\zeta, \omega \,|\, \mathcal{F}(\text{ data })\Big) \propto p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta, \omega\Big) p\Big(\zeta, \omega\Big)$

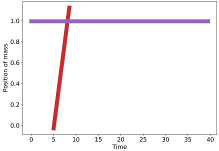
Sampling the posterior result in model trajectories that fit our features.

Feature-based data assimilation

Data



Feature selection



Model

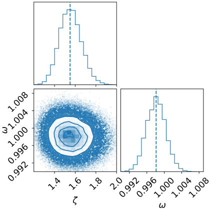
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

$$\mathcal{F}(\text{ data }) = \text{ features}$$

Bayes' Rule: $p\Big(\zeta,\omega \,|\, \mathcal{F}(\text{ data })\Big) \propto p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta,\omega\Big) p\Big(\zeta,\omega\Big)$

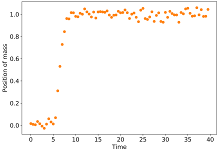
Sampling the posterior result in model trajectories that fit our features.



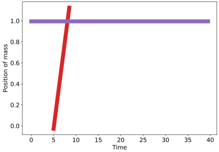
Posterior

Feature-based data assimilation

Data



Feature selection



Model

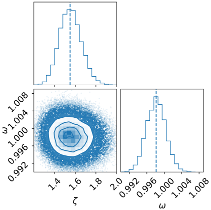
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

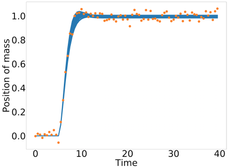
$$\mathcal{F}(\text{ data }) = \text{ features}$$

Bayes' Rule: $p\Big(\zeta, \omega \,|\, \mathcal{F}(\text{ data })\Big) \propto p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta, \omega\Big) p\Big(\zeta, \omega\Big)$

Sampling the posterior result in model trajectories that fit our features.



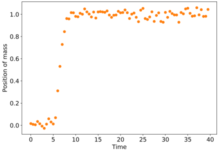
Posterior



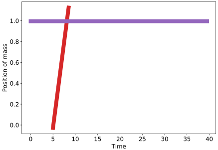
Sampled trajectories

Feature-based data assimilation

Data



Feature selection



Model

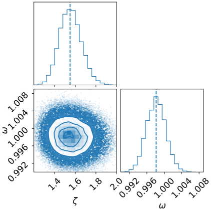
$$\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = H(t)$$

Feature function

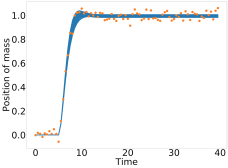
$$\mathcal{F}(\text{ data }) = \text{ features}$$

Bayes' Rule: $p\Big(\zeta,\omega \,|\, \mathcal{F}(\text{ data })\Big) \propto p\Big(\mathcal{F}(\text{ data }) \,|\, \zeta,\omega\Big)p\Big(\zeta,\omega\Big)$

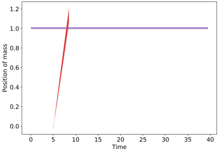
Sampling the posterior result in model trajectories that fit our features.



Posterior



Sampled trajectories



Sampled features

General framework for feature-based data assimilation

Given:

- Mathematical or computational model $\mathcal{M}$ with parameters θ.
- Data z.
- A feature function $\mathcal{F}$ that maps data to features f.

General framework for feature-based data assimilation

Given:

- Mathematical or computational model $\mathcal{M}$ with parameters θ.
- Data z.
- A feature function $\mathcal{F}$ that maps data to features f.

Set up:

- Define the prior distribution $p(\theta)$.

General framework for feature-based data assimilation

Given:

- Mathematical or computational model $\mathcal{M}$ with parameters θ.
- Data z.
- A feature function $\mathcal{F}$ that maps data to features f.

Set up:

- Define the prior distribution $p(\theta)$.
- Define the likelihood distribution

$$p\Big(\mathcal{F}(z)\,|\,\theta\Big) \sim \mathcal{N}\Big(\mathcal{F}(\mathcal{M}(\theta))\,,\ R\Big)$$

where R is the sample covariance of perturbed features.

General framework for feature-based data assimilation

Given:

- Mathematical or computational model $\mathcal{M}$ with parameters θ.
- Data z.
- A feature function $\mathcal{F}$ that maps data to features f.

Set up:

- Define the prior distribution $p(\theta)$.
- Define the likelihood distribution

$$p\Big(\mathcal{F}(z)\,|\,\theta\Big) \sim \mathcal{N}\Big(\mathcal{F}(\mathcal{M}(\theta))\,,\,R\Big)$$

 where R is the sample covariance of perturbed features.

- Define the posterior distribution $p\Big(\theta\,|\,\mathcal{F}(\,z\,)\Big)$ via Bayes' rule.

General framework for feature-based data assimilation

Given:

- Mathematical or computational model $\mathcal{M}$ with parameters θ.
- Data z.
- A feature function $\mathcal{F}$ that maps data to features f.

Set up:

- Define the prior distribution $p(\theta)$.
- Define the likelihood distribution

$$p\Big(\mathcal{F}(z)\,|\,\theta\Big) \sim \mathcal{N}\Big(\mathcal{F}(\mathcal{M}(\theta)),\, R\Big)$$

  where R is the sample covariance of perturbed features.
- Define the posterior distribution $p\Big(\theta\,|\,\mathcal{F}(\,z\,)\Big)$ via Bayes' rule.

What results (hopefully!) is a distribution of the parameters θ that yield features similar to $\mathcal{F}(\,z\,)$.

When may this be useful?

When may this be useful?

Data compression without information loss.

- If we can compress our data in such a way that we do not lose too much information.

When may this be useful?

Data compression without information loss.

- If we can compress our data in such a way that we do not lose too much information.

Complex systems and low dimensional models.

- A low dimensional model may describe some qualitative aspect of your complex system and you need to define a feature function to compare the data from the complex system and your model.

When may this be useful?

Data compression without information loss.

- If we can compress our data in such a way that we do not lose too much information.

Complex systems and low dimensional models.

- A low dimensional model may describe some qualitative aspect of your complex system and you need to define a feature function to compare the data from the complex system and your model.

Models and data at different scales.

- When your model and data are characterized by different scales (spatial, temporal, or both), features can filter out the differences.

Another example: Lynx and hare populations

Lotka-Volterra Equations:

$$\dot{x} = \alpha x - \beta xy$$
$$\dot{y} = -\gamma y + \delta xy$$

for $\alpha, \beta, \gamma, \delta > 0$

We want to estimate the parameters $\alpha, \beta, \gamma, \delta$ and the initial conditions $x(0), y(0)$.

Another example: Lynx and hare populations

Lotka-Volterra Equations:

$$\dot{x} = \alpha x - \beta xy$$

$$\dot{y} = -\gamma y + \delta xy$$

for $\alpha, \beta, \gamma, \delta > 0$

We want to estimate the parameters $\alpha, \beta, \gamma, \delta$ and the initial conditions $x(0), y(0)$.

Data:

We use the lynx and hare fur data from the Hudson's Bay Company between the years 1917 to 1927; our data D has size 2×11.

Another example: Lynx and hare populations

Lotka-Volterra Equations:

$$\dot{x} = \alpha x - \beta xy$$

$$\dot{y} = -\gamma y + \delta xy$$

for $\alpha, \beta, \gamma, \delta > 0$

We want to estimate the parameters $\alpha, \beta, \gamma, \delta$ and the initial conditions $x(0), y(0)$.

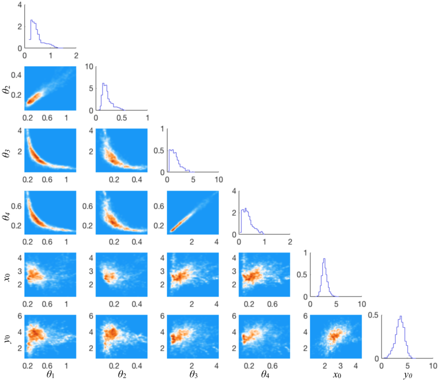
Data:

We use the lynx and hare fur data from the Hudson's Bay Company between the years 1917 to 1927; our data D has size 2×11.

Feature selection:

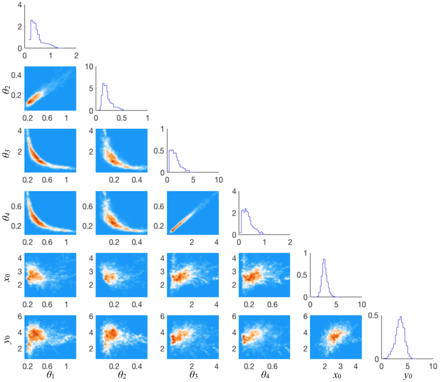
In the singular-value decomposition of our data, $D = USV^*$, we take our feature vector to be the first singular values and the associated left and right singular vectors.

Lynx and hare populations

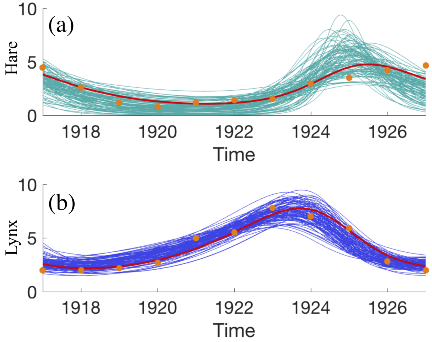


2D and 1D marginal distributions of the parameters $\alpha, \beta, \gamma, \delta, x(0), y(0)$.

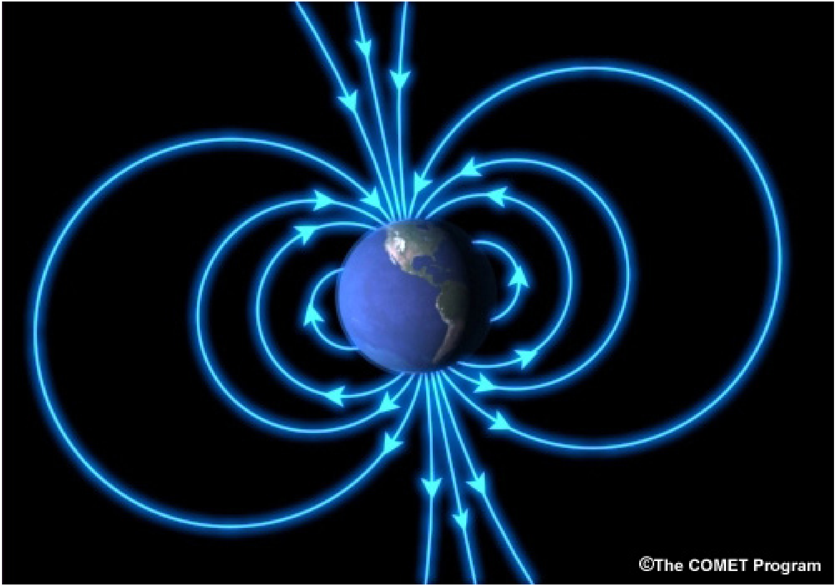
Lynx and hare populations



2D and 1D marginal distributions of the parameters $\alpha, \beta, \gamma, \delta, x(0), y(0)$.



The true data is plotted in orange and the trajectories of 100 samples of the posterior distribution are shown.


©The COMET Program

Example: Dipole reversal rates.

- We consider the Earth's magnetic dipole field over time scales of tens of millions of years.

Example: Dipole reversal rates.

- We consider the Earth's magnetic dipole field over time scales of tens of millions of years.
- At these time scales, the geomagnetic dipoles exhibit reversals, i.e., the north and south pole switch.

Example: Dipole reversal rates.

- We consider the Earth's magnetic dipole field over time scales of tens of millions of years.
- At these time scales, the geomagnetic dipoles exhibit reversals, i.e., the north and south pole switch.

Data:

The occurrence of dipole reversals is documents over the last 150 Myr by the "geomagnetic polarity time scale" (Cande and Kent, 1995; Lowrie and Kent, 2004).

Example: Dipole reversal rates.

- We consider the Earth's magnetic dipole field over time scales of tens of millions of years.
- At these time scales, the geomagnetic dipoles exhibit reversals, i.e., the north and south pole switch.

Data:

The occurrence of dipole reversals is documents over the last 150 Myr by the "geomagnetic polarity time scale" (Cande and Kent, 1995; Lowrie and Kent, 2004).

Model:

The model of Buffett et al. (2013) which is an SDE of the form

$$dx = f(x)dt + g(x)dW.$$

Example: Dipole reversal rates.

- We consider the Earth's magnetic dipole field over time scales of tens of millions of years.
- At these time scales, the geomagnetic dipoles exhibit reversals, i.e., the north and south pole switch.

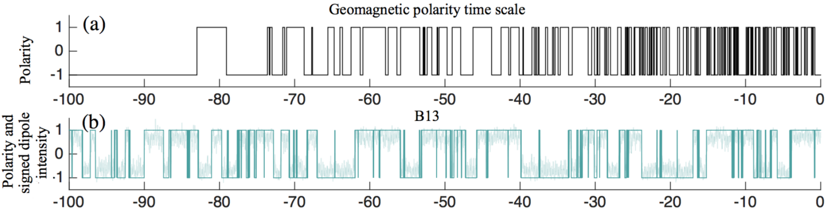
Data:

The occurrence of dipole reversals is documents over the last 150 Myr by the "geomagnetic polarity time scale" (Cande and Kent, 1995; Lowrie and Kent, 2004).

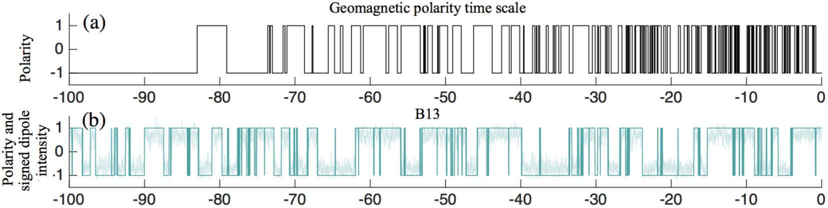
Model:

The model of Buffett et al. (2013) which is an SDE of the form

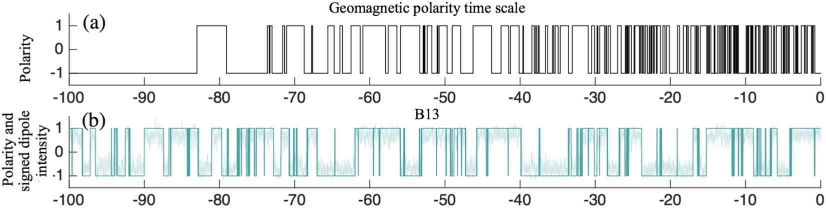
$$dx = f(x)dt + g(x)dW.$$



Example: Variations in Earth's dipole reversal rates.

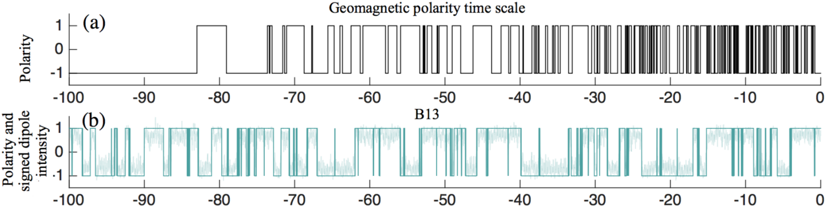


Example: Variations in Earth's dipole reversal rates.



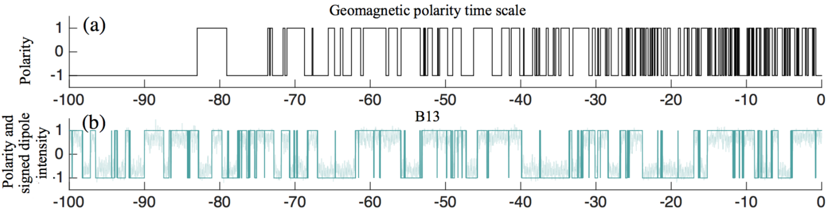
- Data assimilation techniques struggle because the data only tells us the sign of the SDE.

Example: Variations in Earth's dipole reversal rates.



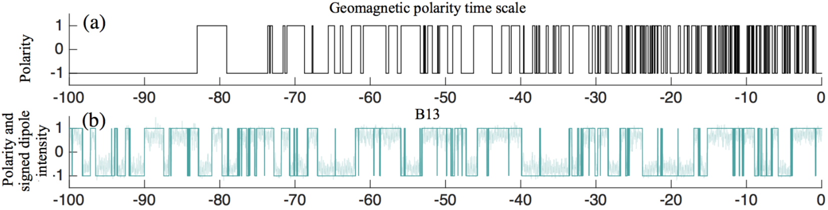
- Data assimilation techniques struggle because the data only tells us the sign of the SDE.
- A "chron" is period during which the dipole polarity is constant.

Example: Variations in Earth's dipole reversal rates.



- Data assimilation techniques struggle because the data only tells us the sign of the SDE.
- A "chron" is period during which the dipole polarity is constant.
- The B13 model has a constant mean chron duration (MCD).

Example: Variations in Earth's dipole reversal rates.



- Data assimilation techniques struggle because the data only tells us the sign of the SDE.
- A "chron" is period during which the dipole polarity is constant.
- The B13 model has a constant mean chron duration (MCD).
- The recorded geomagnetic polarity does **not** have a constant MCD.

Example: Dipole reversal rates.

To allow the model's MCD to vary over time, we modify the model with a time-varying, piecewise constant parameter $\theta(t)$,

$$dx = f(x)dt + \theta(t)g(x)dW.$$

Example: Dipole reversal rates.

To allow the model's MCD to vary over time, we modify the model with a time-varying, piecewise constant parameter $\theta(t)$,

$$dx = f(x)dt + \theta(t)g(x)dW.$$

Goal:
Estimate the parameter $\theta(t)$.

Example: Dipole reversal rates.

To allow the model's MCD to vary over time, we modify the model with a time-varying, piecewise constant parameter $\theta(t)$,

$$dx = f(x)dt + \theta(t)g(x)dW.$$

Goal:
Estimate the parameter $\theta(t)$.

Feature Selection:
For every 1 Myr, we calculate the average MCD using a sliding window averaged over 10 years. Our " feature data" is a vector of 149 MCD values.

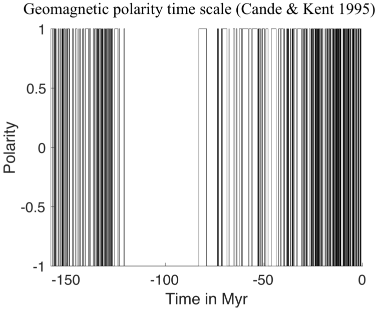
Example: Dipole reversal rates.

To allow the model's MCD to vary over time, we modify the model with a time-varying, piecewise constant parameter $\theta(t)$,

$$dx = f(x)dt + \theta(t)g(x)dW.$$

Goal:
Estimate the parameter $\theta(t)$.

Feature Selection:
For every 1 Myr, we calculate the average MCD using a sliding window averaged over 10 years. Our " feature data" is a vector of 149 MCD values.



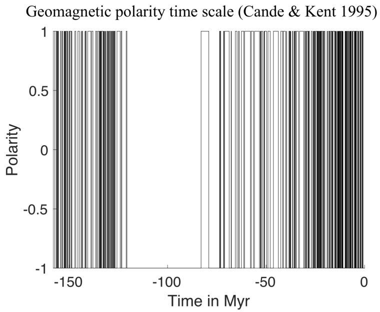
Example: Dipole reversal rates.

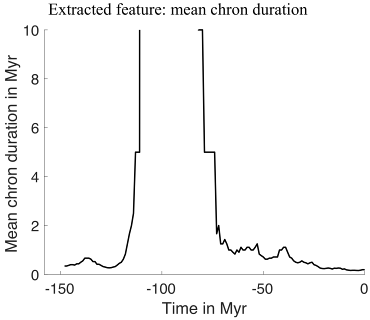
To allow the model's MCD to vary over time, we modify the model with a time-varying, piecewise constant parameter $\theta(t)$,

$$dx = f(x)dt + \theta(t)g(x)dW.$$

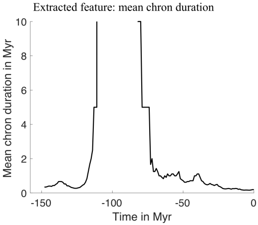
Goal:
Estimate the parameter $\theta(t)$.

Feature Selection:
For every 1 Myr, we calculate the average MCD using a sliding window averaged over 10 years. Our " feature data" is a vector of 149 MCD values.

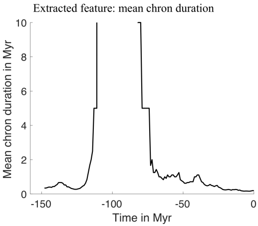


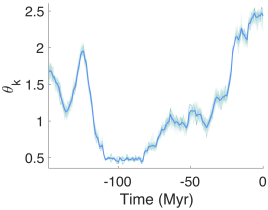


Example: Dipole reversal rates.

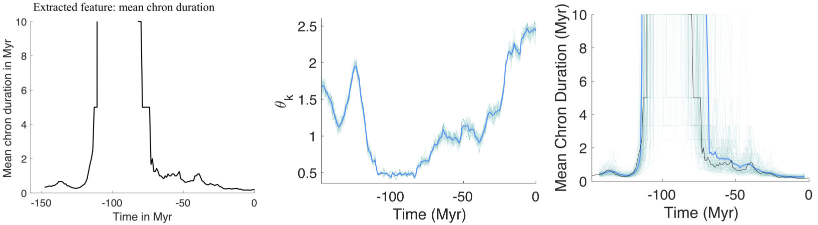


Example: Dipole reversal rates.

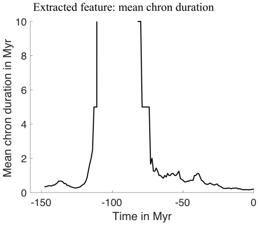


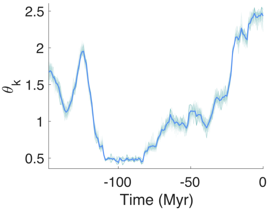


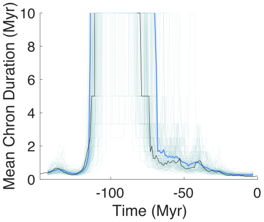
Example: Dipole reversal rates.

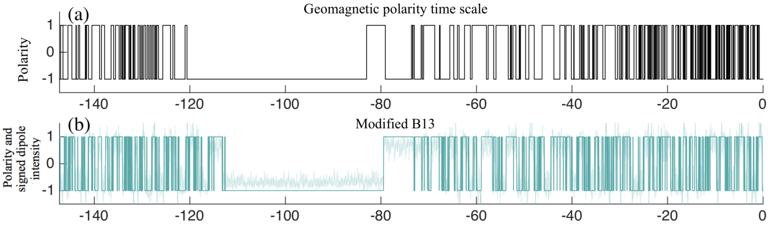


Example: Dipole reversal rates.



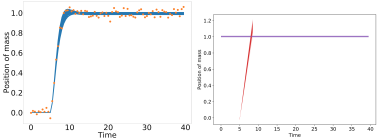




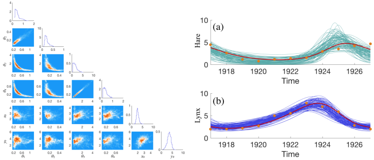


Summary

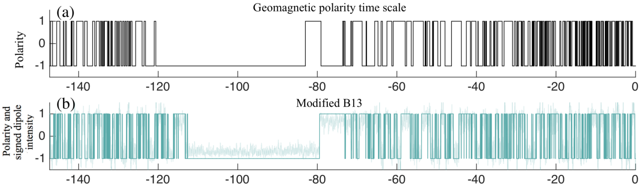
Mass-Spring-Damper System:



Lotka-Volterra:



Earth's dipole reversals:



For more information: Morzfeld M., Adams J., Lunderman S., and Orozco R.: **Feature-based data assimilation in geophysics**, Nonlin. Processes Geophysics (in review)

This work is supported in part by the National Science Foundation under grant DMS- 1619630.