Estimating Stochastic Closures Using Sparsity-Promoting Ensemble Kalman Inversion

Jinlong Wu, Tapio Schneider, Andrew Stuart

Joint IS-ENES3/ESiWACE2 Virtual Workshop, March 17th, 2021

Motivation: Climate Change

mm | 10m | 100m | 1km | 10km | 100km | 1000km | 10000km

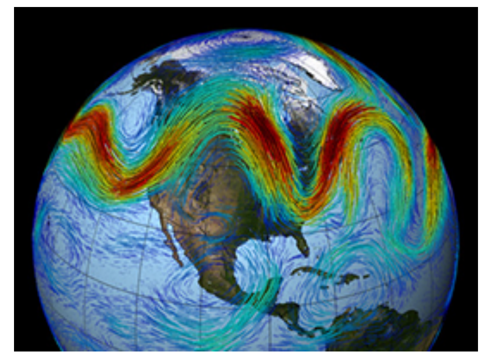
Cloud microphysics

Turbulence

Cumulus clouds

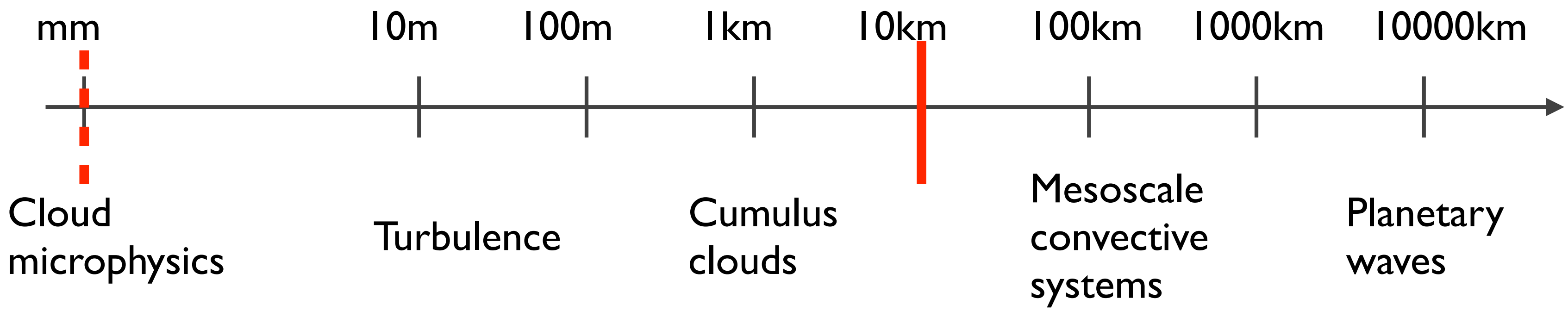
Mesoscale convective systems

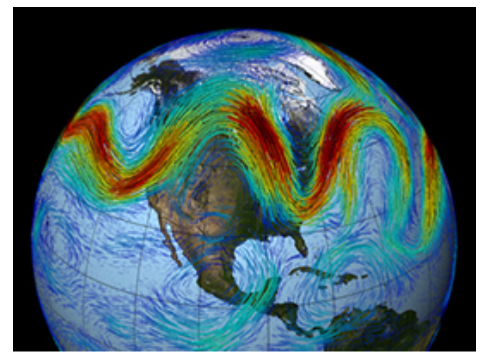
Planetary waves



Motivation: Climate Change

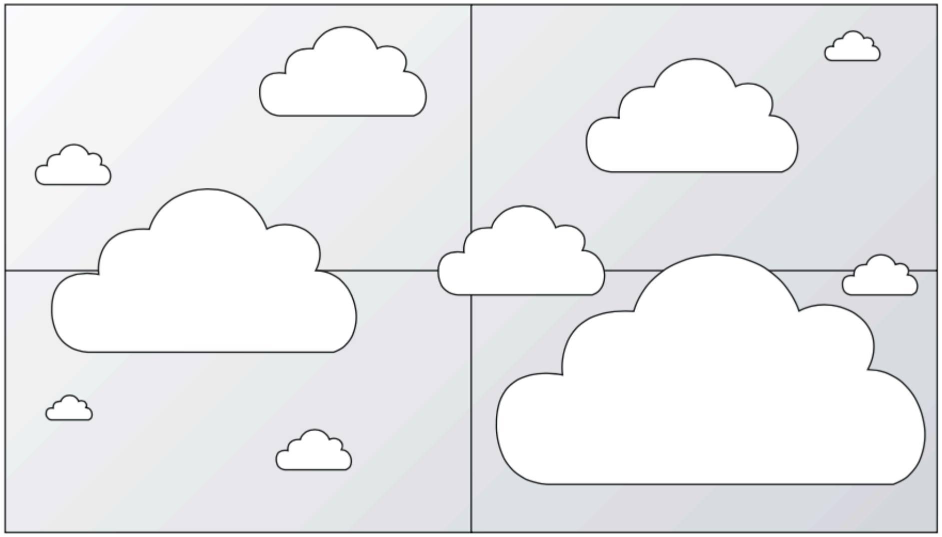
Closure models are needed
for unresolved processes





Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)

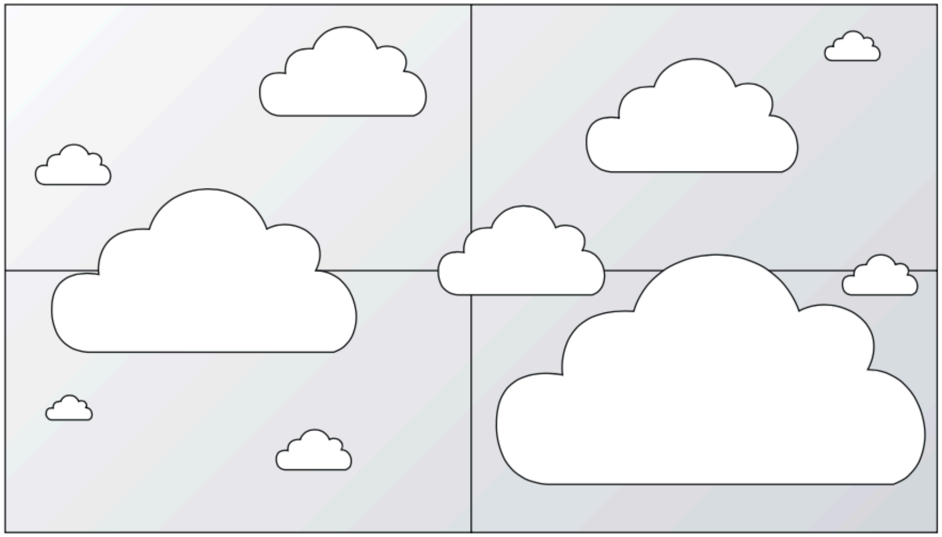


$$\frac{\partial u}{\partial t} = f(u)$$

- In practical simulations, it's usually impossible to resolve all the scales.

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



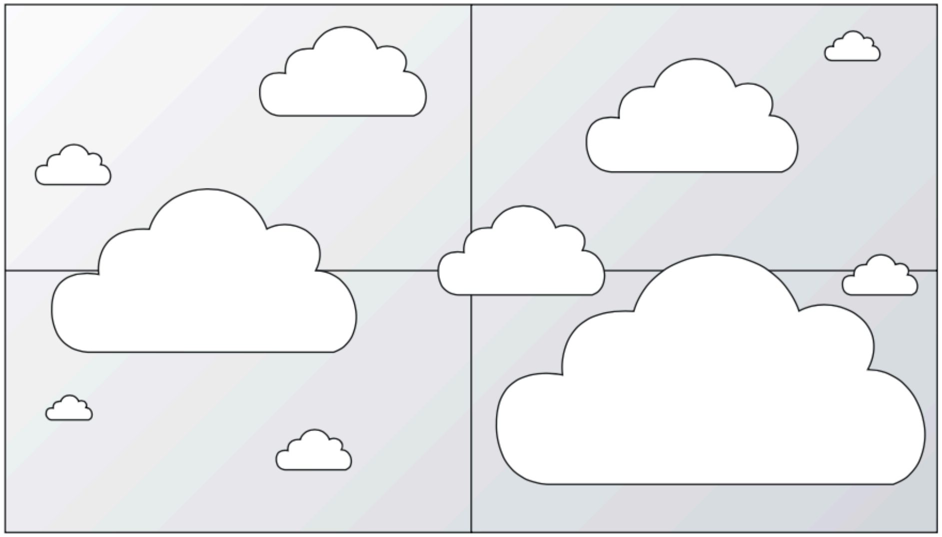
$$\frac{\partial u}{\partial t} = f(u) \qquad U = \mathcal{F}(u)$$

Filter

- In practical simulations, it's usually impossible to resolve all the scales.

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



Filter

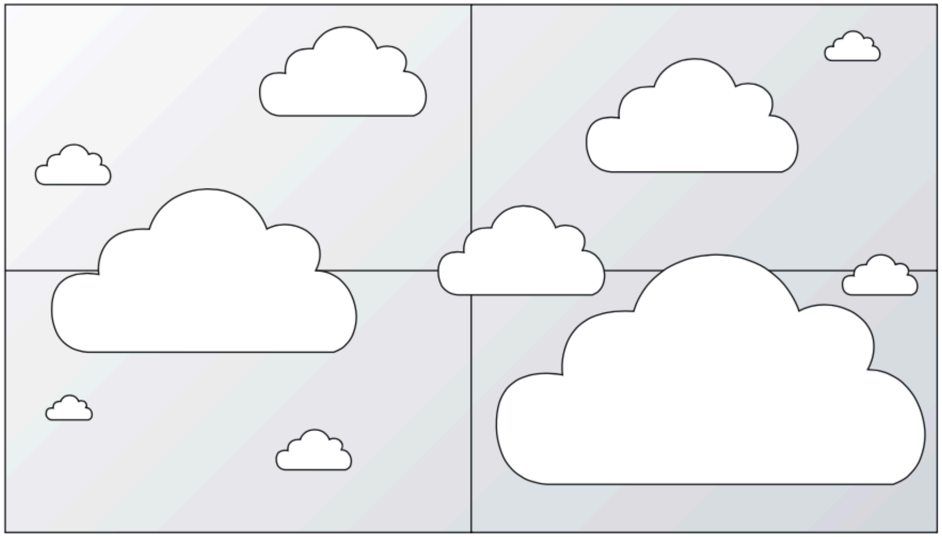
$$\frac{\partial u}{\partial t} = f(u) \qquad U = \mathcal{F}(u)$$

$$\frac{\partial U}{\partial t} = \overline{f}(U; u)$$

- In practical simulations, it's usually impossible to resolve all the scales.
- To describe the impact of unresolved physics, closure models are needed.

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



$$\frac{\partial u}{\partial t} = f(u) \qquad U = \underbrace{\mathcal{F}}_{\text{Filter}}(u)$$

$$\frac{\partial U}{\partial t} = \overline{f}(U; u)$$

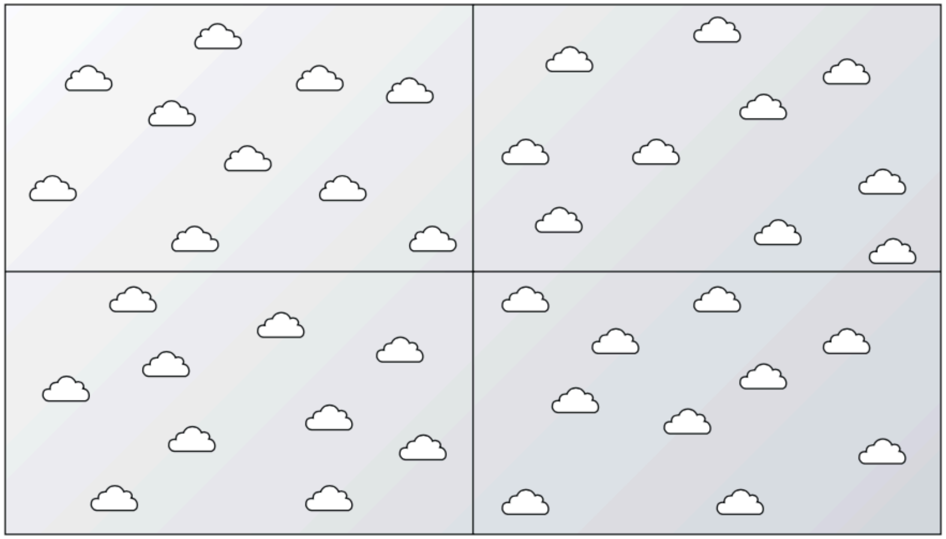
$$\frac{\partial U}{\partial t} = \tilde{f}(U)$$

$$\overline{f}(U; u) \approx \tilde{f}(U)$$

- In practical simulations, it's usually impossible to resolve all the scales.
- To describe the impact of unresolved physics, closure models are needed.

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



Filter

$$\frac{\partial u}{\partial t} = f(u) \qquad U = \mathcal{F}(u)$$

$$\frac{\partial U}{\partial t} = \overline{f}(U;u)$$

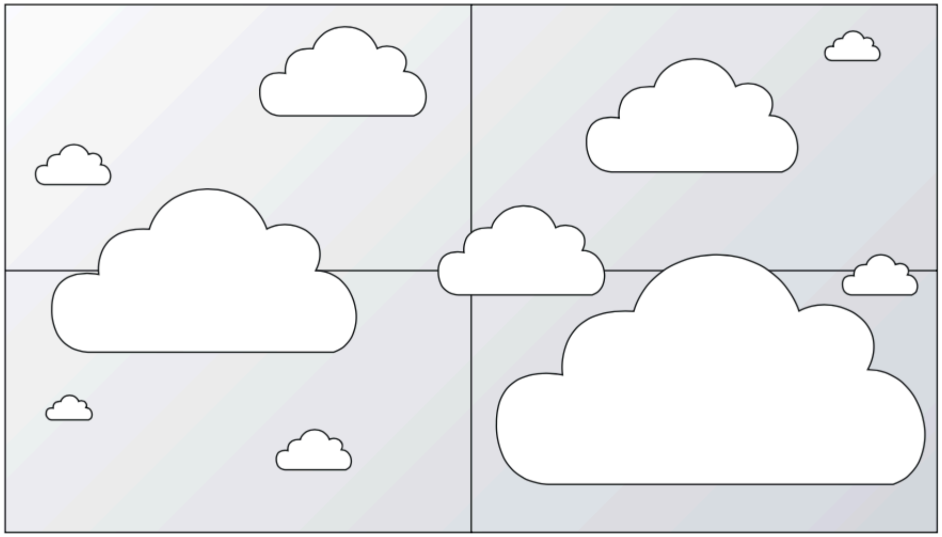
$$\frac{\partial U}{\partial t} = \tilde{f}(U)$$

$$\overline{f}(U;u) \approx \tilde{f}(U)$$

- In practical simulations, it's usually impossible to resolve all the scales.
- To describe the impact of unresolved physics, closure models are needed.

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



$$\frac{\partial u}{\partial t} = f(u) \qquad U = \mathcal{F}(u)$$

Filter

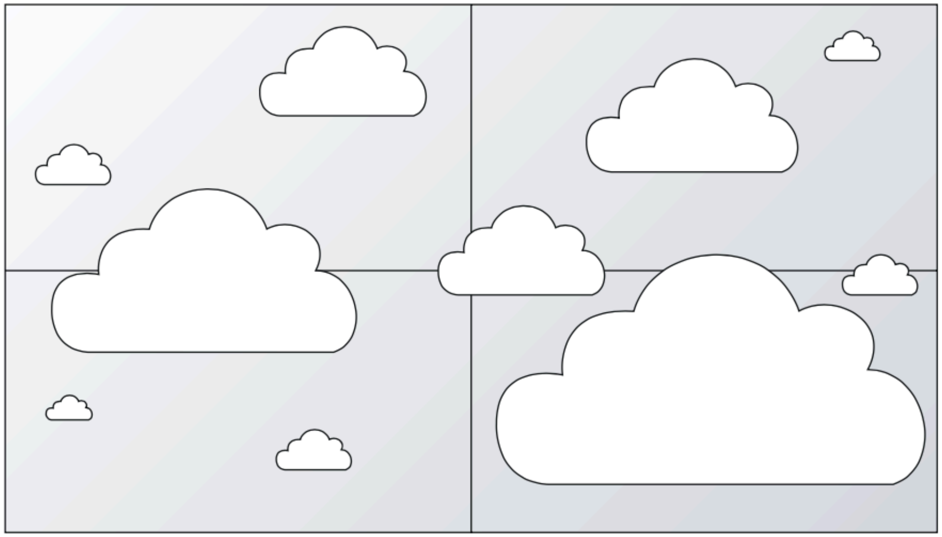
$$\frac{\partial U}{\partial t} = \overline{f}(U; u)$$

$$\frac{\partial U}{\partial t} = \tilde{f}(U)$$

$$\overline{f}(U; u) \approx \tilde{f}(U) + M(U; \alpha) + \sqrt{\sigma(U; \gamma)}\dot{W}$$

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



$$\frac{\partial u}{\partial t} = f(u) \qquad U = \underset{\text{Filter}}{\mathcal{F}}(u)$$

$$\frac{\partial U}{\partial t} = \overline{f}(U;u)$$

$$\frac{\partial U}{\partial t} = \tilde{f}(U)$$

$$\overline{f}(U;u) \approx \tilde{f}(U) + M(U;\alpha) + \sqrt{\sigma(U;\gamma)}\dot{W}$$

Dictionary learning

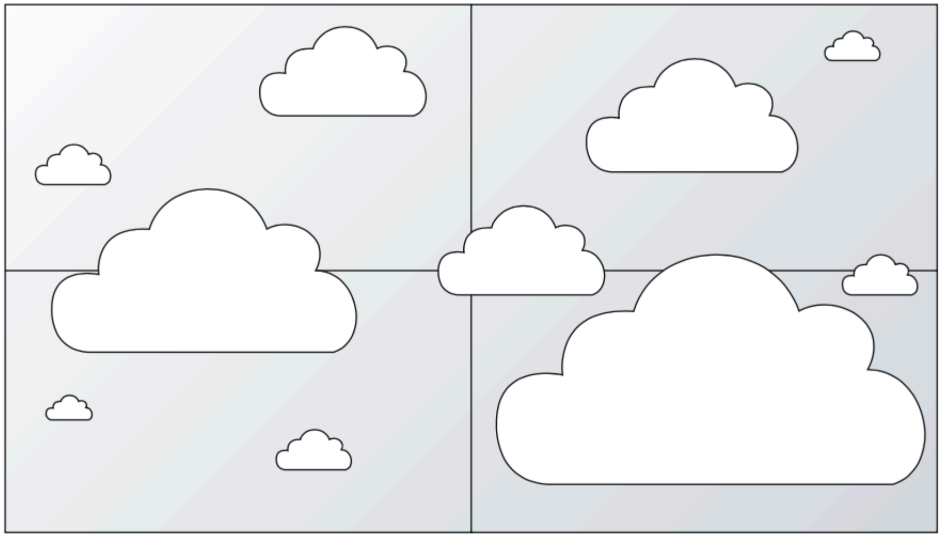
RKHS

Neural network

...

Closure Problem

Figure source: Palmer, Nature Reviews Physics (2019)



$$\frac{\partial u}{\partial t} = f(u) \qquad U = \underset{\text{Filter}}{\mathcal{F}}(u)$$

$$\frac{\partial U}{\partial t} = \overline{f}(U;u)$$

$$\frac{\partial U}{\partial t} = \tilde{f}(U)$$

$$\overline{f}(U;u) \approx \tilde{f}(U) + M(U;\alpha) + \sqrt{\sigma(U;\gamma)}\dot{W}$$

$$\theta = \{\alpha, \gamma\}$$

- We estimate the unknown parameters θ from some data y using ensemble Kalman inversion.

Ensemble Kalman Inversion (EKI)

- Inverse problem to be solved:

$$\text{Data} \quad y = G(\theta) + \eta, \quad \eta \sim N(0, \Gamma)$$

- EKI formula (j: index of ensemble, m: index of EKI step):

$$\theta_{m+1}^{(j)} = \theta_m^{(j)} + C_m^{\theta G} \left(C_m^{GG} + \Gamma \right)^{-1} \left(y_{m+1}^{(j)} - G(\theta_m^{(j)}) \right)$$

- Advantage of EKI:
  - Derivative-free optimization;
  - Robust to noisy evaluations of the forward map G;
  - The core task is equivalent to a convex optimization problem.

Iglesias, Law, Stuart, Inverse Problems **29** 045001 (2013)

Learning Stochastic Closures

Schneider, Stuart, Wu, arXiv:2004.08376 (2020)

Case I: Noisy Lorenz 63 System

True System: $\dot{x} = \tilde{f}(x) + \sqrt{\sigma}\dot{W}$

Where $\tilde{f}(x)$ **is the standard Lorenz 63 system :**

$$\frac{dx_1}{dt} = \alpha(x_2 - x_1)$$
$$\frac{dx_2}{dt} = x_1(\rho - x_3) - x_2$$
$$\frac{dx_3}{dt} = x_1 x_2 - \beta x_3$$

Case I: Noisy Lorenz 63 System

Modeled System: $\dot{x} = \tilde{f}(x) + \boxed{\sqrt{\sigma}}\dot{W}$

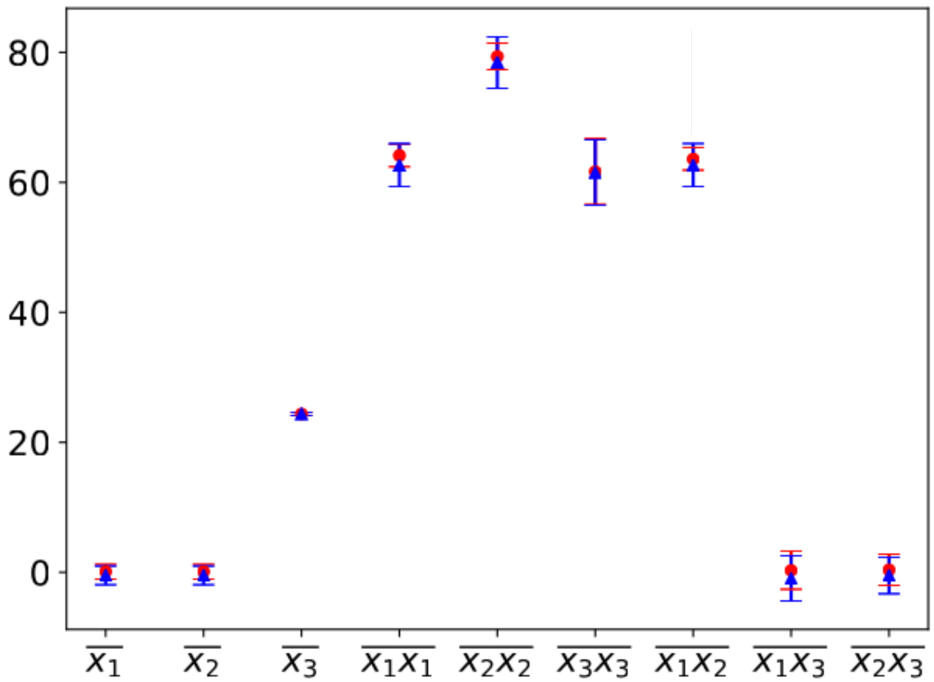
Where $\tilde{f}(x)$ **is the standard Lorenz 63 system :**

$$\begin{aligned}
\frac{dx_1}{dt} &= \alpha(x_2 - x_1) \\
\frac{dx_2}{dt} &= x_1(\rho - x_3) - \boxed{M(x_2)} \\
\frac{dx_3}{dt} &= x_1 x_2 - \beta x_3
\end{aligned}$$

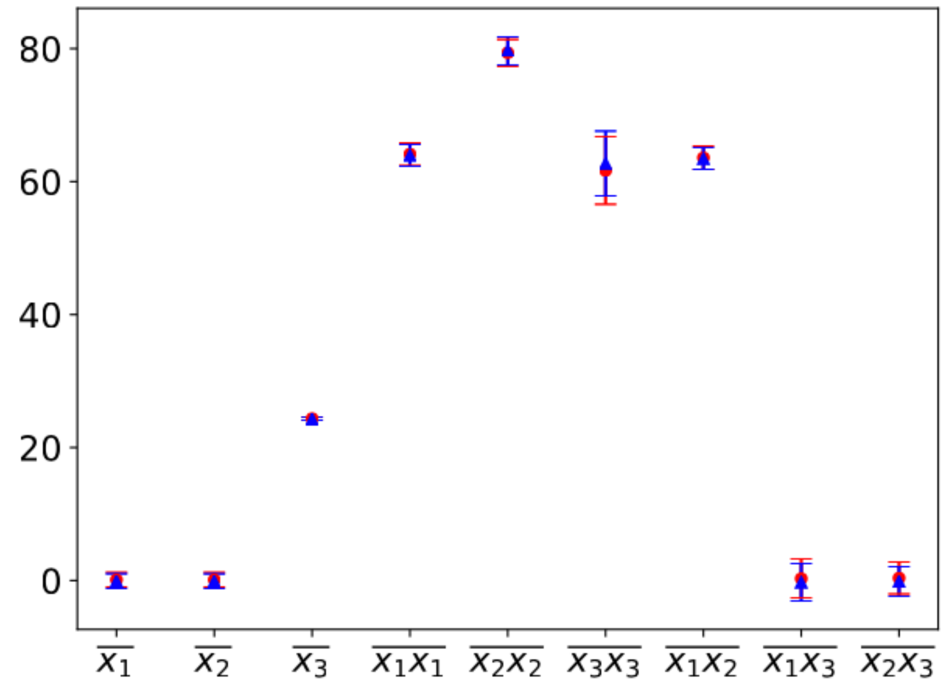
Data: $\{\overline{x_1}, \overline{x_2}, \overline{x_3}, \overline{x_1 x_1}, \overline{x_2 x_2}, \overline{x_3 x_3}, \overline{x_1 x_2}, \overline{x_1 x_3}, \overline{x_2 x_3}\}$

Case I: Moments

Red: True system   Blue: Modeled system



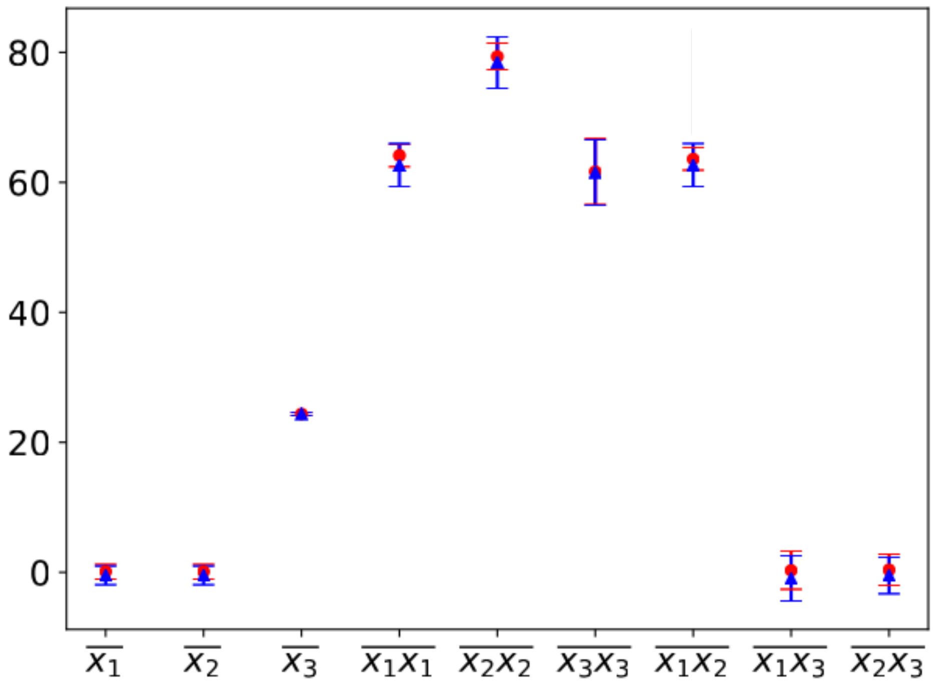
Without stochastic term



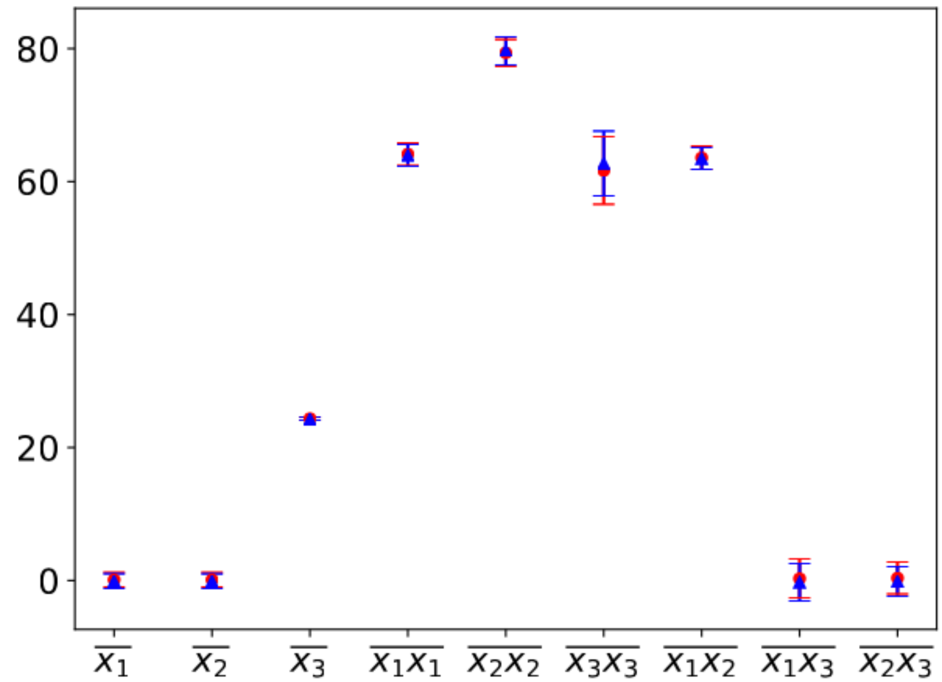
With stochastic term

Case I: Moments

Red: True system Blue: Modeled system



Without stochastic term

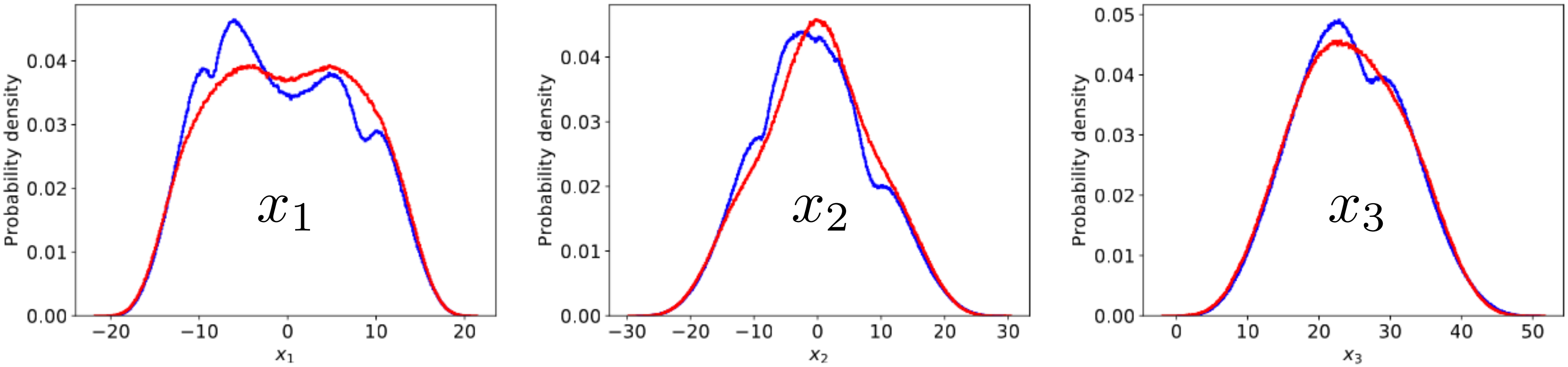


With stochastic term

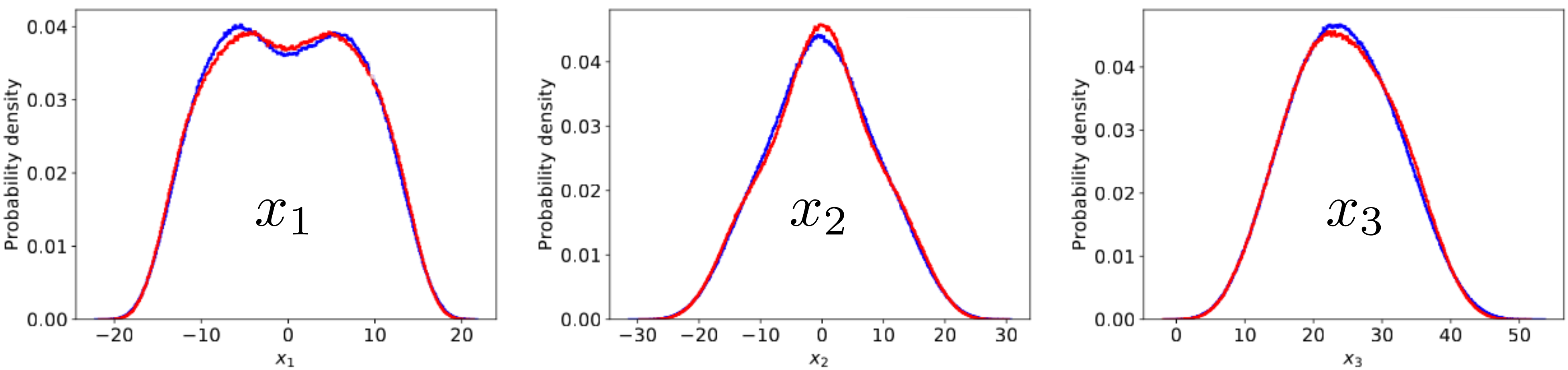
- Does it mean the two models are equally good?

Case I: Invariant Measures

Red: True system Blue: Modeled system



Without stochastic term



With stochastic term

Case II: Reduced Lorenz 63 System

- Lorenz 63 system in transformed coordinates $\dot{a} = g(a)$:

$$g_1(a) = 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3,$$
$$g_2(a) = -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3,$$
$$g_3(a) = -0.63a_1 - 13a_3 + 0.43a_1a_2 + 0.49a_2a_3.$$

- Reduced-order model:

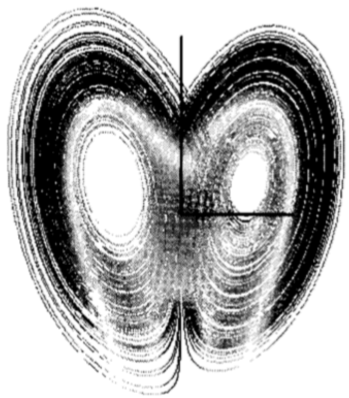
$$\dot{a}_1 = 2.3a_1 - 0.49a_1a_2$$
$$\dot{a}_2 = -62 - 2.7a_2 + 0.49a_1^2$$

Palmer, Quarterly Journal of the Royal Meteorological Society, 127(572), 279-304 (2001)

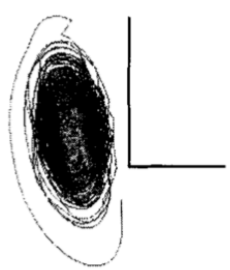
Case II: Reduced Lorenz 63 System

- Comparison of trajectories:

True system



Reduced-order system



Reduced-order system with simple stochastic term (white noises)

Palmer, Quarterly Journal of the Royal Meteorological Society, 127(572), 279-304 (2001)

Case II: Reduced Lorenz 63 System

- Lorenz 63 system in transformed coordinates $\dot{a} = g(a)$:

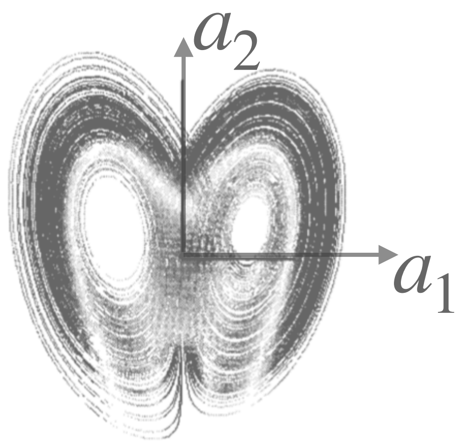
$$
\begin{aligned}
g_1(a) &= 2.3a_1 - 6.2a_3 - 0.49a_1a_2 - 0.57a_2a_3, \\
g_2(a) &= -62 - 2.7a_2 + 0.49a_1^2 - 0.49a_3^2 + 0.14a_1a_3, \\
g_3(a) &= -0.63a_1 - 13a_3 + 0.43a_1a_2 + 0.49a_2a_3.
\end{aligned}
$$

- Reduced-order model:

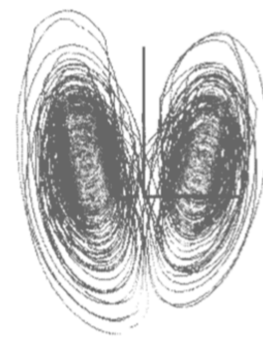
$$
\begin{aligned}
\dot{a}_1 &= 2.3a_1 - 0.49a_1a_2 + M_1(a_1, a_2) + \sqrt{\sigma_1(a_1, a_2)}\dot{W} \\
\dot{a}_2 &= -62 - 2.7a_2 + 0.49a_1^2 + M_2(a_1, a_2) + \sqrt{\sigma_2(a_1, a_2)}\dot{W}
\end{aligned}
$$

Case II: Reduced Lorenz 63 System

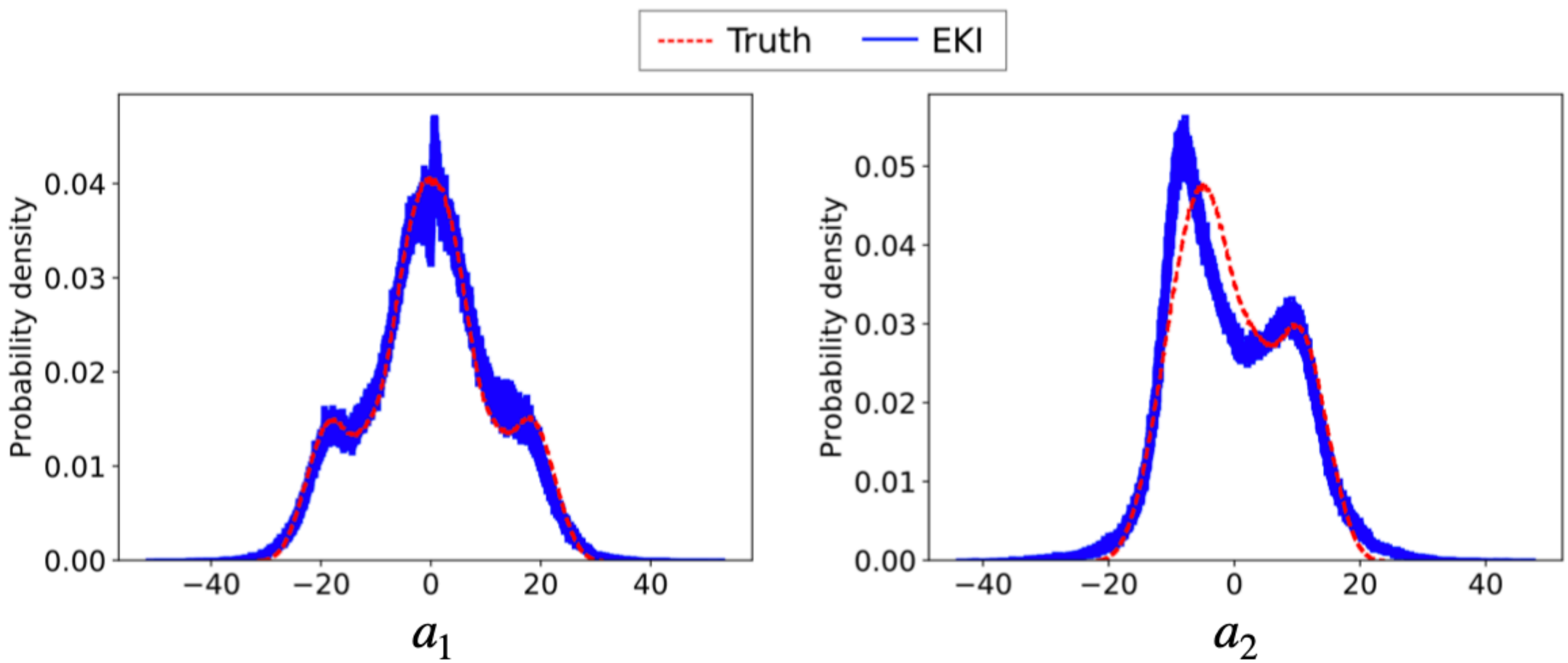
True system



Reduced-order system with simple stochastic term (white noises)



Palmer, Quarterly Journal of the Royal Meteorological Society, 127(572), 279-304 (2001)



Sparsity-Promoting Ensemble Kalman Inversion

Schneider, Stuart, Wu, arXiv:2007.06175 (2020)

Ensemble Kalman Inversion (EKI)

- Inverse problem to be solved:

$$y = G(\theta) + \eta, \quad \eta \sim N(0, \Gamma)$$

- EKI formula (j: index of ensemble, m: index of EKI step):

$$\theta_{m+1}^{(j)} = \theta_m^{(j)} + C_m^{\theta G} \left(C_m^{GG} + \Gamma \right)^{-1} \left(y_{m+1}^{(j)} - G(\theta_m^{(j)}) \right)$$

- Advantage of EKI:
  - Derivative-free optimization;
  - Robust to noisy evaluations of the forward map G;
  - The core task is equivalent to a convex optimization problem.

Iglesias, Law, Stuart, Inverse Problems **29** 045001 (2013)

Ensemble Kalman Inversion (EKI)

- Introduce a new variable $w = G(\theta)$:

$$v = (\theta, w)^\top,$$
$$\Psi(v) = (\theta, G(\theta))^\top.$$

- Formulating a noisy observed dynamical system ($Hv = w, H^{\perp}v = \theta$):

$$v_{m+1} = \Psi(v_m)$$
$$y_{m+1} = Hv_{m+1} + \eta_{m+1}.$$

Ensemble Kalman Inversion (EKI)

- Introduce a new variable $w = G(\theta)$:

$$v = (\theta, w)^\top,$$
$$\Psi(v) = (\theta, G(\theta))^\top .$$

- Formulating a noisy observed dynamical system ($Hv = w, H^{\perp}v = \theta$):

$$v_{m+1} = \Psi(v_m)$$
$$y_{m+1} = Hv_{m+1} + \eta_{m+1}.$$

- Solving the optimization problem:

$$J_m^{(j)}(v) := \frac{1}{2}\left|y_{m+1}^{(j)} - Hv\right|_\Gamma^2 + \frac{1}{2}\left|v - \Psi\left(v_m^{(j)}\right)\right|_{C_m^{\Psi\Psi}}^2,$$
$$v_{m+1}^{(j)} = \arg\min_v J_m^{(j)}(v).$$

Albers et al, Inverse Problems **35** 095007 (2019)

Sparse EKI

- Objective function of standard EKI:

$$\frac{1}{2}v^\top \left(H^\top\Gamma^{-1}H + C_m^{-1}\right) v - \left(C_m^{-1}\Psi(v_m^{(j)}) + H^\top\Gamma^{-1}y^{(j)}\right)^\top v.$$

- Imposing sparsity within EKI (recall that $H^\perp v = \theta$):

$$\begin{aligned} \min_v \quad & \frac{1}{2}v^\top Qv + q^\top v \\ \text{s.t.} \quad & \left|H^\perp v\right|_{\ell_1} \leq \gamma. \end{aligned}$$

Sparse EKI

- Objective function of standard EKI:

$$\frac{1}{2}v^\top\left(H^\top\Gamma^{-1}H + C_m^{-1}\right)v - \left(C_m^{-1}\Psi(v_m^{(j)}) + H^\top\Gamma^{-1}y^{(j)}\right)^\top v.$$

- Imposing sparsity within EKI (recall that $H^\perp v = \theta$):

$$\begin{aligned} \min_{v} \quad & \frac{1}{2}v^\top Q v + q^\top v \\ \text{s.t.} \quad & \left|H^\perp v\right|_{\ell_1} \leq \gamma. \end{aligned}$$

- Introduce variables $v_i = v_i^+ - v_i^-$ and $|v_i| = v_i^+ + v_i^-$:

$$\begin{aligned} \min_{v^+,v^-} \quad & \frac{1}{2}\left(v^+ - v^-\right)^\top Q\left(v^+ - v^-\right) + q^\top\left(v^+ - v^-\right) \\ \text{s.t.} \quad & H^\perp\left(v^+ + v^-\right) \leq \gamma,\ v^+ \geq 0,\ v^- \geq 0. \end{aligned}$$

Tibshirani, Journal of the Royal Statistical Society: Series B **58** (1996)

Case III: Noisy Lorenz 63 System

True System: $\dot{x} = \tilde{f}(x) + \sqrt{\sigma}\dot{W}$

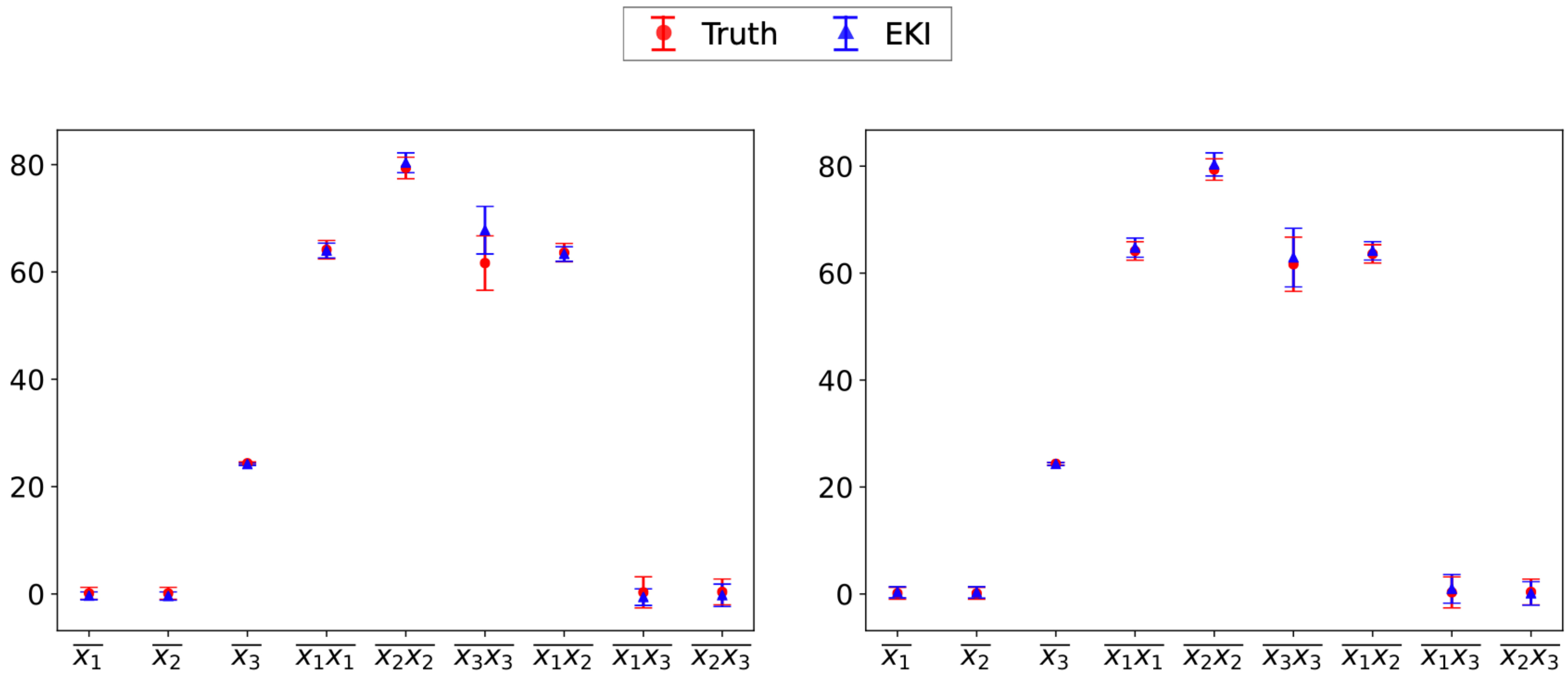
Where $\tilde{f}(x)$ **is the standard Lorenz 63 system :**

$$\frac{dx_1}{dt} = \alpha(x_2 - x_1)$$
$$\frac{dx_2}{dt} = x_1(\rho - x_3) - x_2$$
$$\frac{dx_3}{dt} = x_1 x_2 - \beta x_3$$

Modeled system: $\dot{X}_k = \sum_{i=1}^{9} \theta_{ki}\phi_i(X) + \sqrt{\sigma}\dot{W}_k, \quad k = 1, 2, 3$

$\{\phi_i\}$ contains all the first and second-order polynomial basis functions

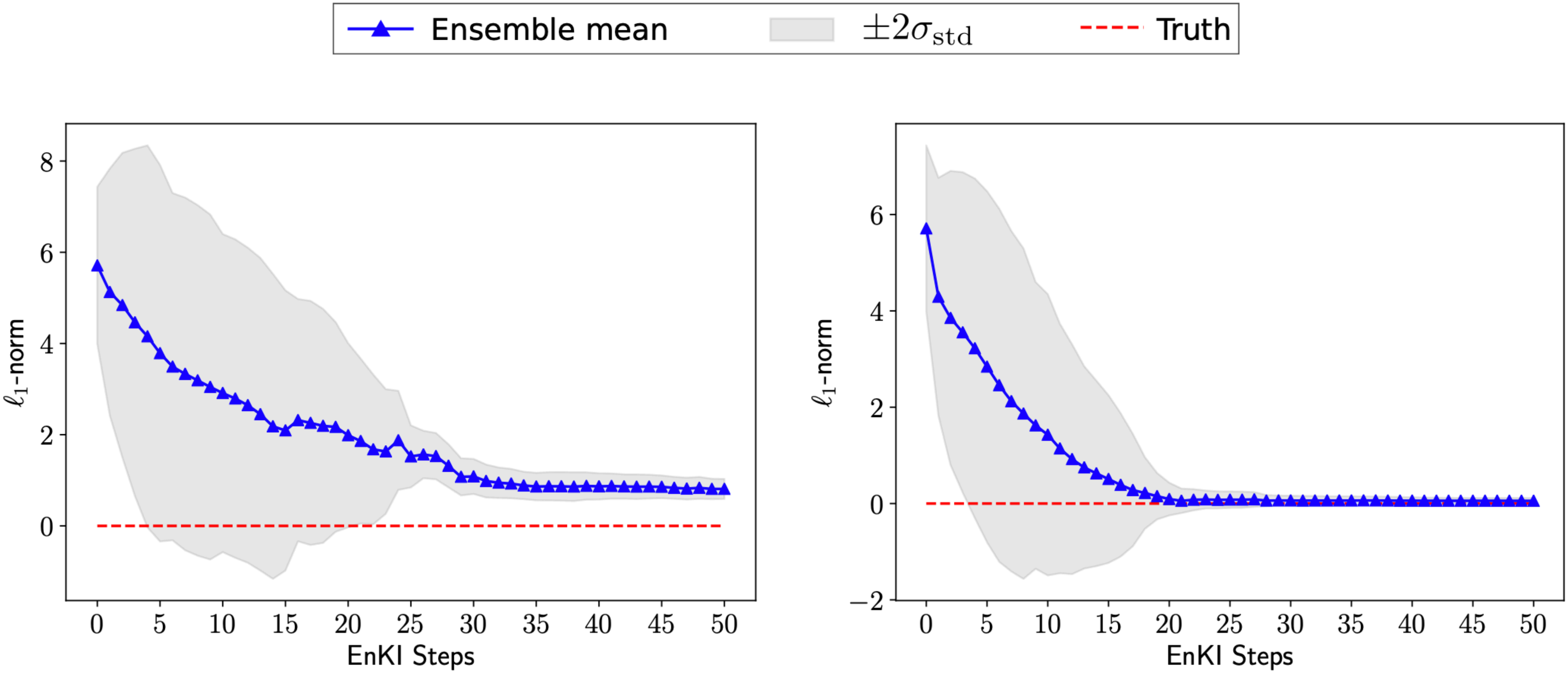
Case III: Moments



(a) Standard EKI

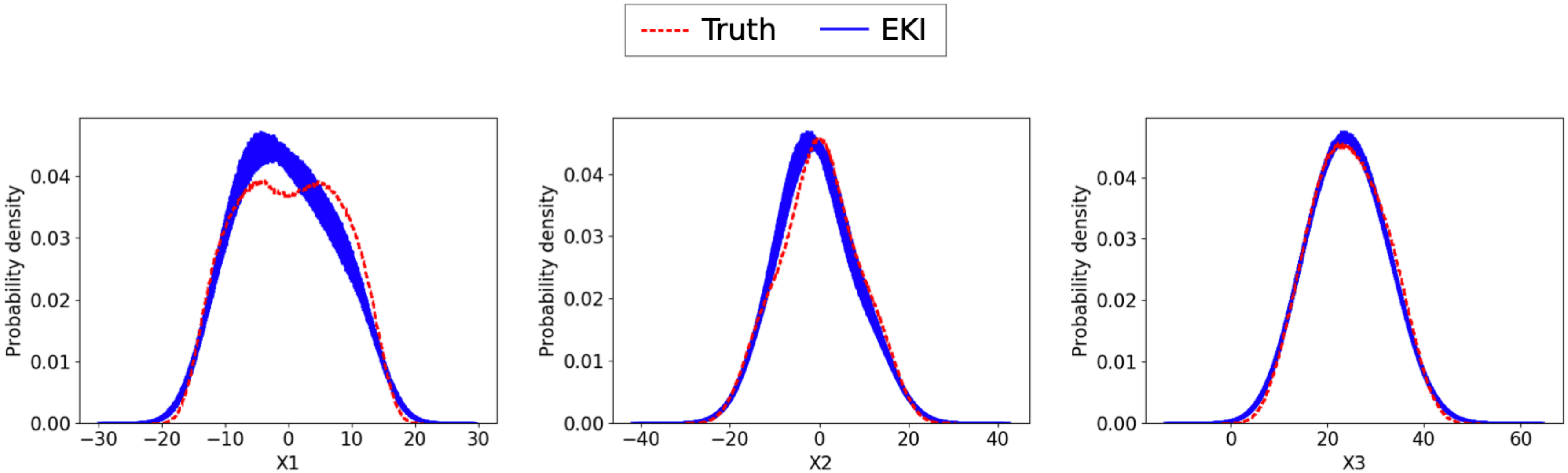
(b) Sparse EKI

Case III: Redundant Coefficients

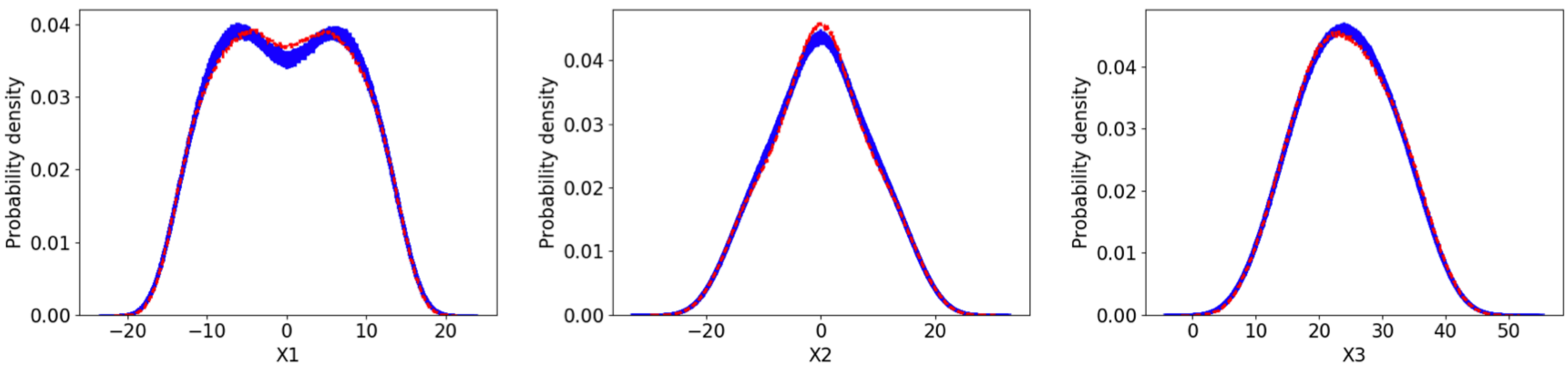


(a) Standard EKI (b) Sparse EKI

Case III: Invariant Measures



(a) X_1 (Standard EKI) (b) X_2 (Standard EKI) (c) X_3 (Standard EKI)



(d) X_1 (Sparse EKI) (e) X_2 (Sparse EKI) (f) X_3 (Sparse EKI)

Summary

- Many complex systems (e.g., earth system) involve **multi-scale dynamics**, and there is usually **no clear scale separation** between resolved and unresolved scales, which justifies the use of **stochastic closures**.
- We demonstrated the learning of stochastic closures based on **time-averaged statistics** using ensemble Kalman inversion (EKI).
- With the assumption that simpler explanation is often better (Occam's razor), we imposed **sparsity constraint** into EKI and showed improved performance of learning stochastic closures.

More details can be found at:

Schneider, Stuart, Wu. Learning Stochastic Closures Using Ensemble Kalman Inversion. arXiv:2004.08376 (2020)

Schneider, Stuart, Wu. Ensemble Kalman Inversion for Sparse Learning of Dynamical Systems from Time-Averaged Data. arXiv:2007.06175 (2020)

Thanks!