PR-Smoother: Simulator-Preserving Non-Gaussian Smoothing for Data Assimilation

Yuta Tarumi (RIKEN, Japan)

My Webpage

Github

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation: state estimate from observations

Research problem: Estimate the hidden physical state X_t by fusing noisy, incomplete observations with a dynamical-model forecast inside a Bayesian update.

Why it matters: Estimating the true system state from observations is a common, fundamental task across the natural sciences: it powers daily weather prediction and is becoming equally vital in oceanography, hydrology, and other sciences.

Key challenges:

- Observations are noisy and incomplete.
- Non-linear dynamics and observation maps create non-Gaussian posteriors.
- State trajectories are not available for training.
- Million-dimensional states make particle-filter weights collapse.

Our strategy: $p_\theta(X_{1:T} \mid Y_{1:T}) \leftarrow q_{\theta,\phi}(X_{1:T} \mid Y_{1:T})$

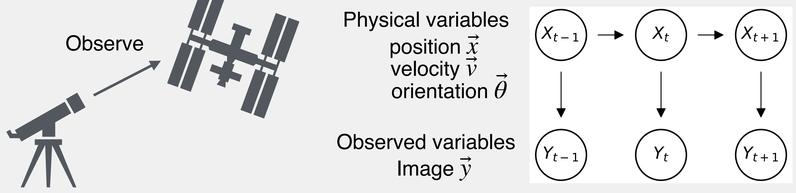


Figure 1: A typical data assimilation problem settings

Problem setting: Data assimilation

What is available?

Data assimilation combines observations $Y_{1:T}$ and prescribed models f_θ, h_θ. We consider a state space model (SSM) based on the models:

$$p(X_{t+1} \mid X_t) = \mathcal{N}[X_{t+1}; f_\theta(X_t), Q]$$
$$p(Y_t \mid X_t) = \mathcal{N}[Y_t; h_\theta(X_t), R]$$
$$p(X_1) = \mathcal{N}[X_1; m_1, \Sigma_1]$$

Observation dataset: $\{Y_{1:T}^{(n)}\}_{n=1}^N$

What do we want?

- Smoother distribution $p_\theta(X_{1:T} \mid Y_{1:T})$
- Calibrated physical parameters θ for f_θ, h_θ

	Non-Gauss	High-D	No state labels	Explicit models
EnKF, ETKF	×	✔	✔	✔
4DVar	×	✔	✔	✔
PF	✔	×	✔	✔
Latent assimilation	✔	✔	△	×
Obs-to-obs	✔	✔	✔	×
PR-smoother	✔	✔	✔	✔

Method: PR-Smoother

Simulator-preserving variational inference

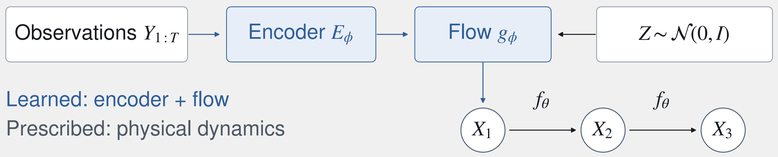
Learn an observation-conditioned posterior around a prescribed simulator, keeping the physical dynamics explicit.

Initial-state posterior

Encode the observations, then transform a Gaussian sample.

$$Z \sim \mathcal{N}(0, I), \qquad X_1 = g_\phi\Big(Z; E_\phi(Y_{1:T})\Big).$$

A conditional normalizing flow gives a flexible, non-Gaussian initial-state density for variational training.



Deterministic rollout shown; noisy dynamics add learned corrections.

Trajectory posterior

$$q_{\theta,\phi}(X_{1:T} \mid Y_{1:T}) = q_\phi(X_1 \mid Y_{1:T}) \prod_{t=2}^{T} q_{\theta,\phi}(X_t \mid X_{t-1}, Y_{t:T}).$$

Strong constraint: deterministic dynamics

$$X_t = f_\theta(X_{t-1}).$$

Initial-state uncertainty induces a distribution over physical trajectories.

Weak constraint: noisy dynamics

$$q_{\theta,\phi}(X_t \mid X_{t-1}, Y_{t:T}) = \mathcal{N}\left(X_t; f_\theta(X_{t-1}) + m_t, S_t\right).$$

Corrections use the previous state and remaining observations.
Gaussian transitions can still yield non-Gaussian trajectory distributions.

Loss function

$$\mathcal{L}(\theta, \phi; Y_{1:T}) = \mathbb{E}_{q_{\theta,\phi}}\left[\sum_{t=1}^{T} \log p_\theta(Y_t \mid X_t)\right] - \mathrm{KL}\Big(q_{\theta,\phi}(X_{1:T} \mid Y_{1:T}) \,\|\, p_\theta(X_{1:T})\Big)$$

Fit the observations while respecting the physical trajectory prior.
Maximize the mean ELBO over observation windows.

ϕ: encoder, flow, and transition corrections.
θ: unknown physical parameters and sensor biases.

Amortized inference
Reuse the trained network on new observation windows.

Exact limiting cases

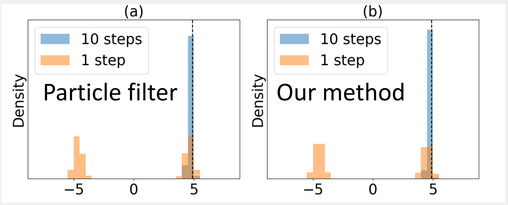
With sufficient flow capacity and unrestricted Gaussian corrections, the family contains the exact smoother for deterministic dynamics and linear-Gaussian state-space models.

Experiments

Lorenz96, 4D

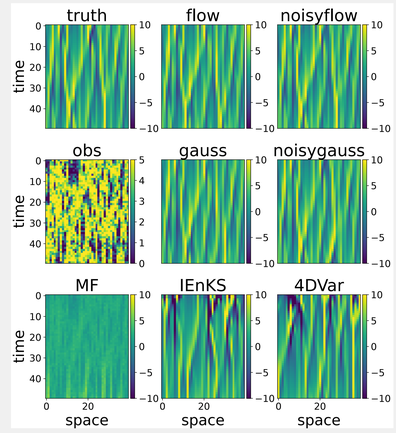
- $h(x) = x^2$
- "10 steps": $p(X_{10} \mid Y_{1:10})$:
- "1 step": $p(X_{10} \mid Y_{10})$

Our method recovers the results of the particle filter with 5 million particles.



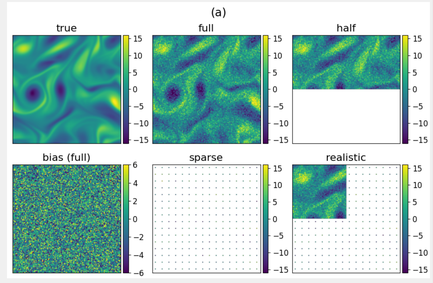
Lorenz96, 40D

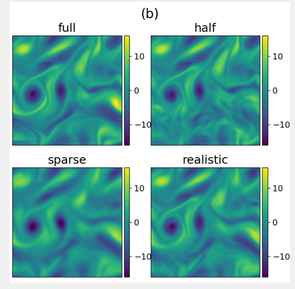
- $h(x) = \min(5, x^2)$
- 50-step window: model $p(X_{1:50} \mid Y_{1:50})$
- Our methods ("flow": strong-constraint and "noisyflow": weak-constraint) recover the state variables well.



Kolmogorov flow, 16,384D

- $h(x) = x$ (subsampled)
- Four observation networks: "full", "half", "sparse", and "realistic"
- 10-step window: model $p(X_{1:10} \mid Y_{1:10})$
- The forcing term F, Reynolds number Re, and the bias field are learned from observations





	Full	Half	Sparse	Realistic
θ recovery	✔	✔	×	✔
State Estimate	✔	×	✔	✔

Takeaway:

PR-Smoother learns non-Gaussian posteriors without state supervision. It scales to 16,384 state variables while keeping prescribed physical models explicit for parameter and sensor bias calibration.

NeurIPS

arXiv