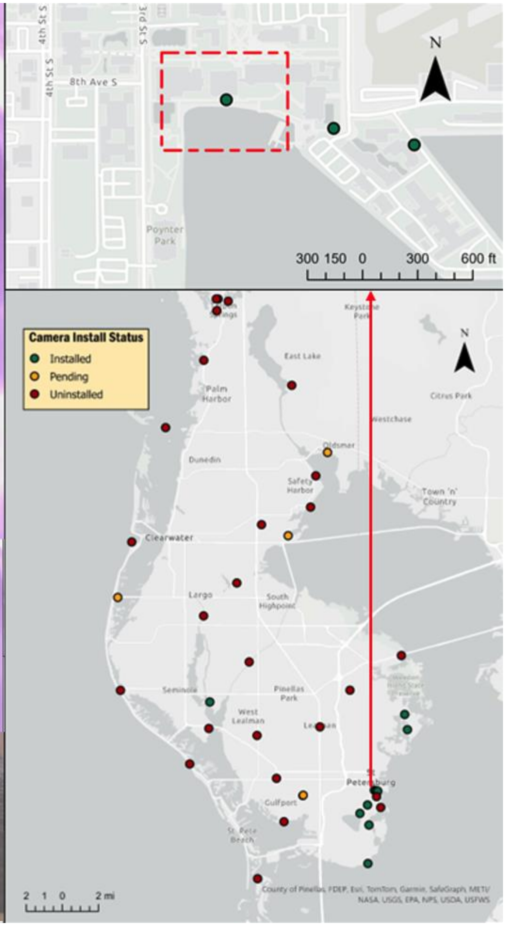
Latent-EnSF for Sparse Data Assimilation of High-Dimensional Dynamical Systems

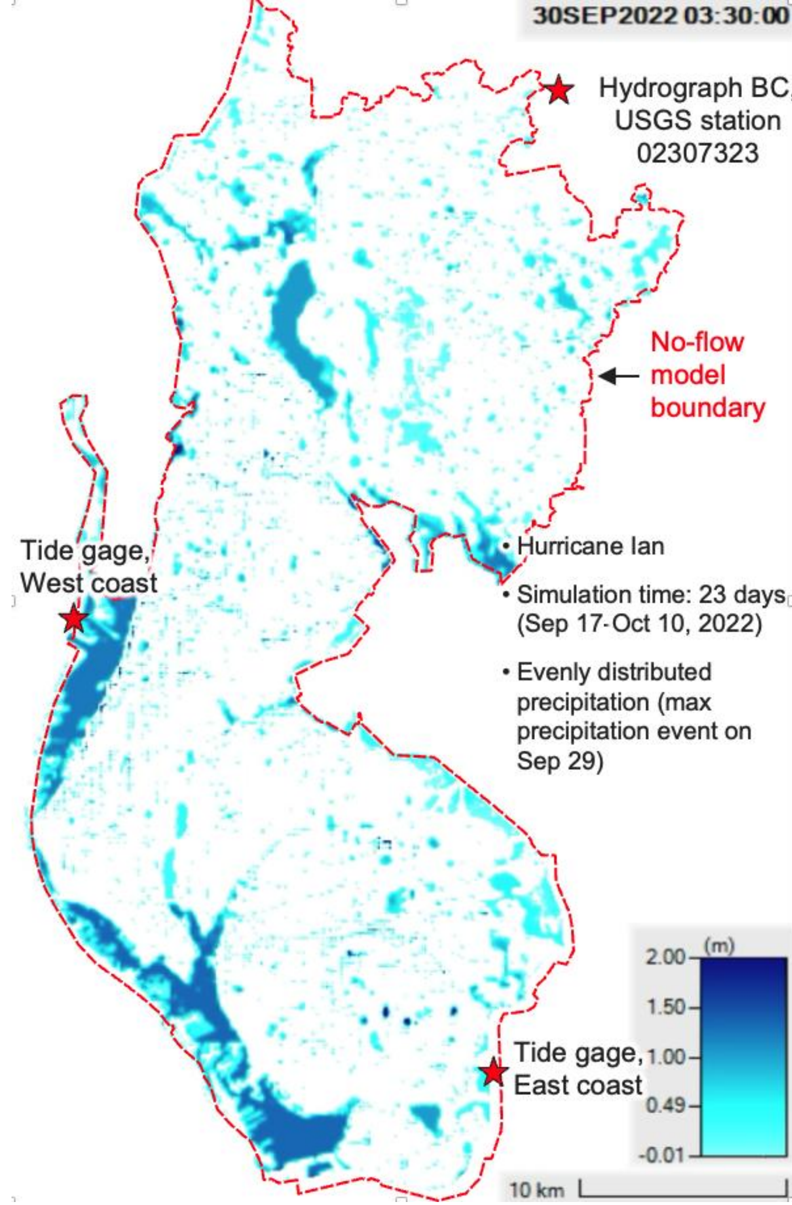
Phillip Si

Georgia Tech

Data Assimilation for Flood Prediction

- Left: Camera image of the water level at 3 different times
- Right: Sparsely located camera sites in Pinellas, Florida

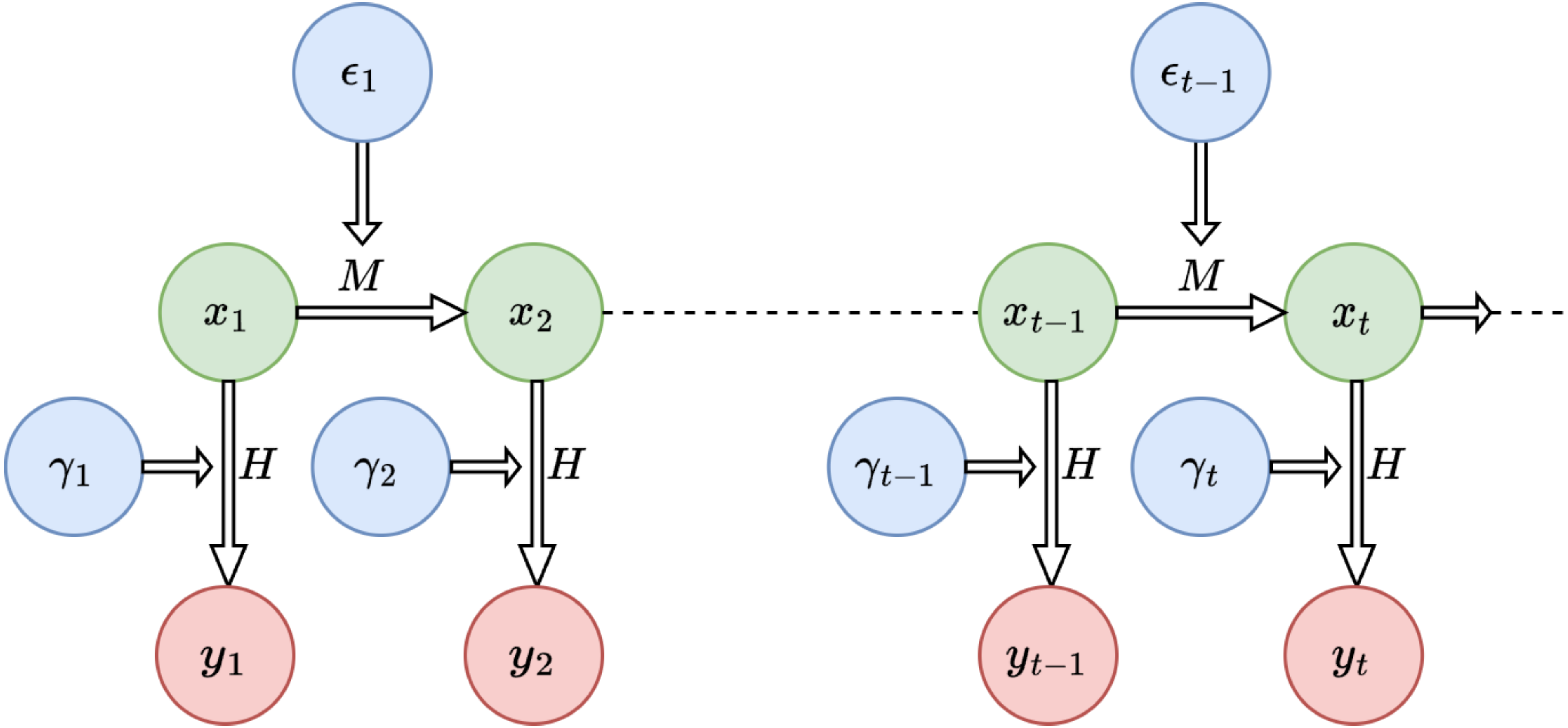




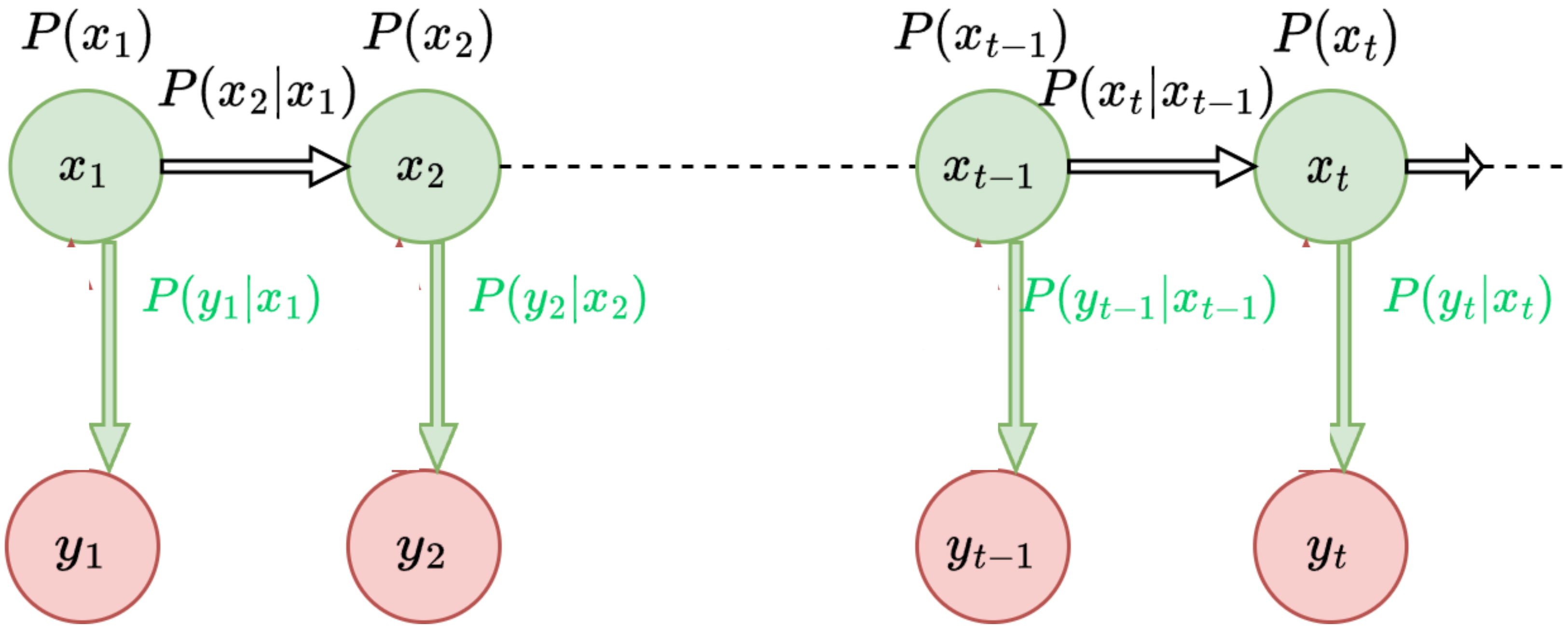
Flood depth from high resolution HEC-RAS Simulation

Georgia Tech

Data Assimilation: Physical System

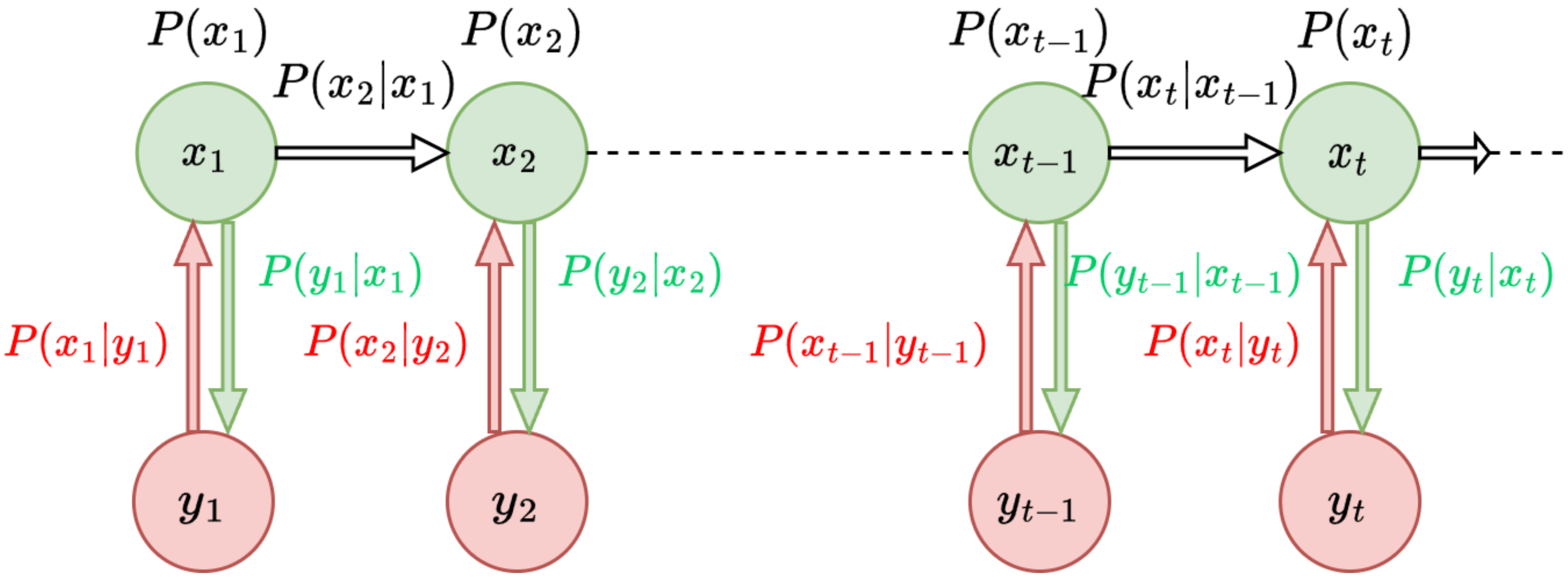


Bayesian Filtering



Prediction: $P(x_t|y_{1:t-1}) = \int P(x_t|x_{t-1})P(x_{t-1}|y_{1:t-1})dx_{t-1}$

Bayesian Filtering



Prediction: $P(x_t|y_{1:t-1}) = \int P(x_t|x_{t-1})P(x_{t-1}|y_{1:t-1})dx_{t-1}$

Update: $P(x_t|y_{1:t}) = \frac{1}{Z}P(y_t|x_t)P(x_t|y_{1:t-1})$

Georgia Tech

Score-Matching, SDEs, and Diffusion Models

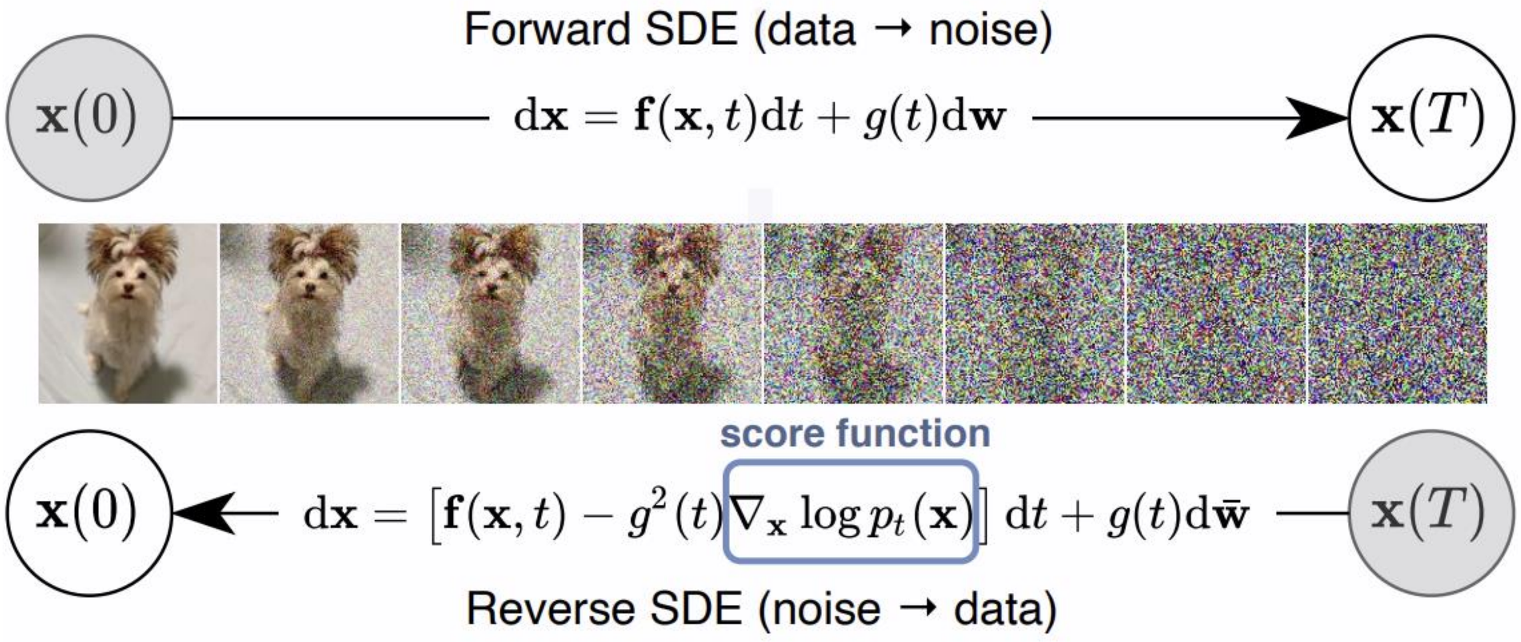
- Score-matching SDEs create a forward noisy SDE

$$dx = f(x,\tau)d\tau + g(\tau)dw$$

- And couple it with a reverse-time SDE for sampling

$$dx = [f(x,\tau) - g^2(\tau)\nabla_x \log P_\tau(x)]dt + g(\tau)d\bar{w}$$

- Image from Song et. al (2021)



Score-Matching, SDEs, and Diffusion Models

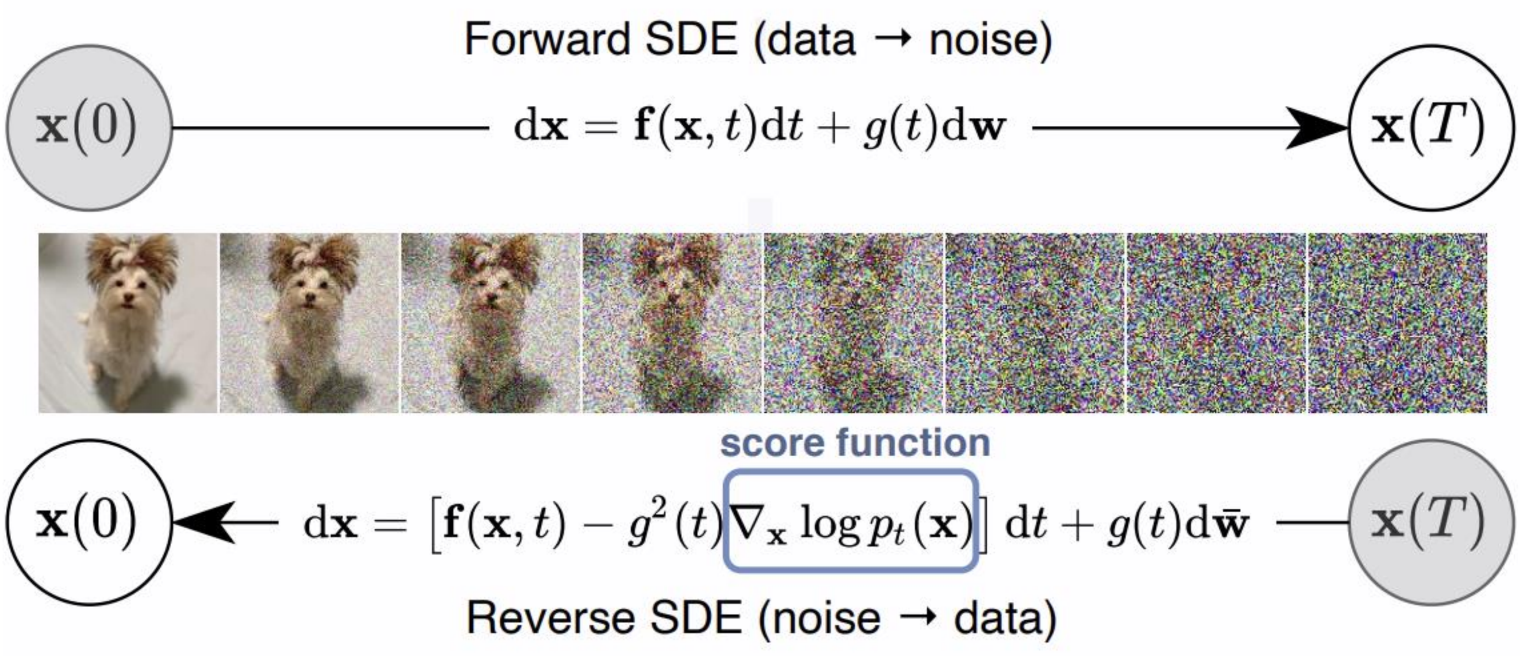
- Score-matching SDEs create a forward noisy SDE

$$dx = f(x,\tau)d\tau + g(\tau)dw$$

- And couple it with a reverse-time SDE for sampling

$$dx = [f(x,\tau) - g^2(\tau)\boxed{\nabla_x \log P_\tau(x)}]dt + g(\tau)d\bar{w}$$

- Image from Song et. al (2021)



Georgia Tech

Ensemble Score Filters (EnSF) (Bao et. al, 2023)

- Models the score function instead of the precise distribution and draw samples by the reverse SDE process

$$dx = [f(x, \tau) - g^2(\tau)\nabla_x \log P_\tau(x)]d\tau + g(\tau)d\bar{w}$$

$$\nabla_x \log P\left(x_{t,\tau} \middle| y_{1:t}\right) = \nabla_x \log P\left(x_{t,\tau} \middle| y_{1:t-1}\right) + h(\tau)\nabla_x \log P\left(y_t \middle| x_{t,\tau}\right).$$

- Shown to be able to handle high dimensional systems like a million-dimensional Lorenz-96 system

EnSF with Sparse Observations

- EnSF relies on full-dimension observations
- Struggles when incorporating sparse observation functions due to the score function being ill-defined at non-observed points

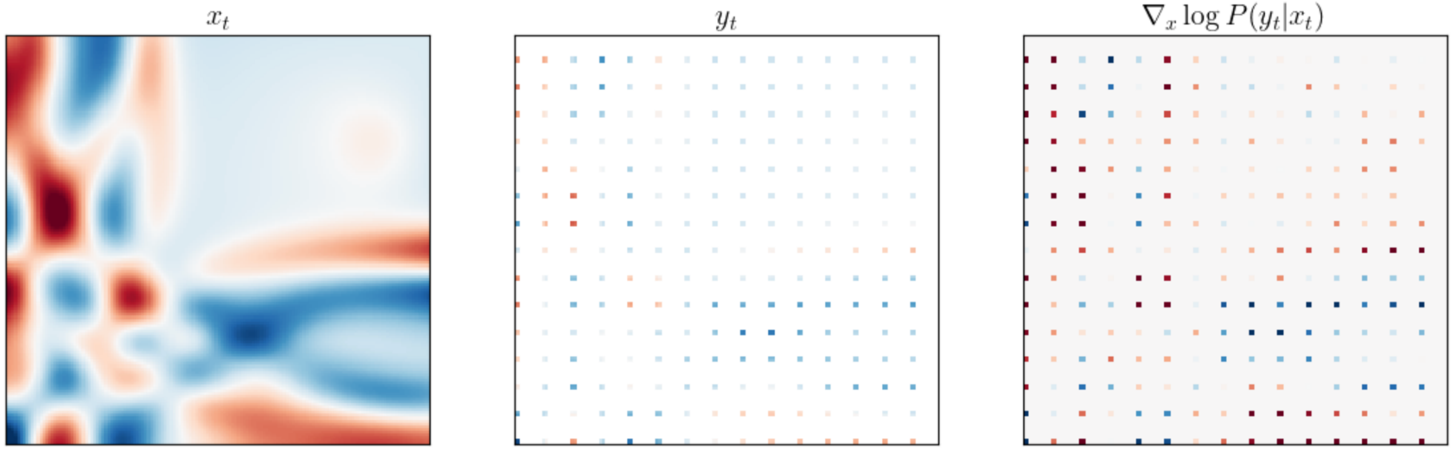


Figure 2: The gradient of the log-likelihood function $\nabla_x \log P(y_t|x_t)$ (right) vanishes at the points where the sparse observation data y_t (middle) do not have any information of the state x_t (left).

Sparsity (continued)

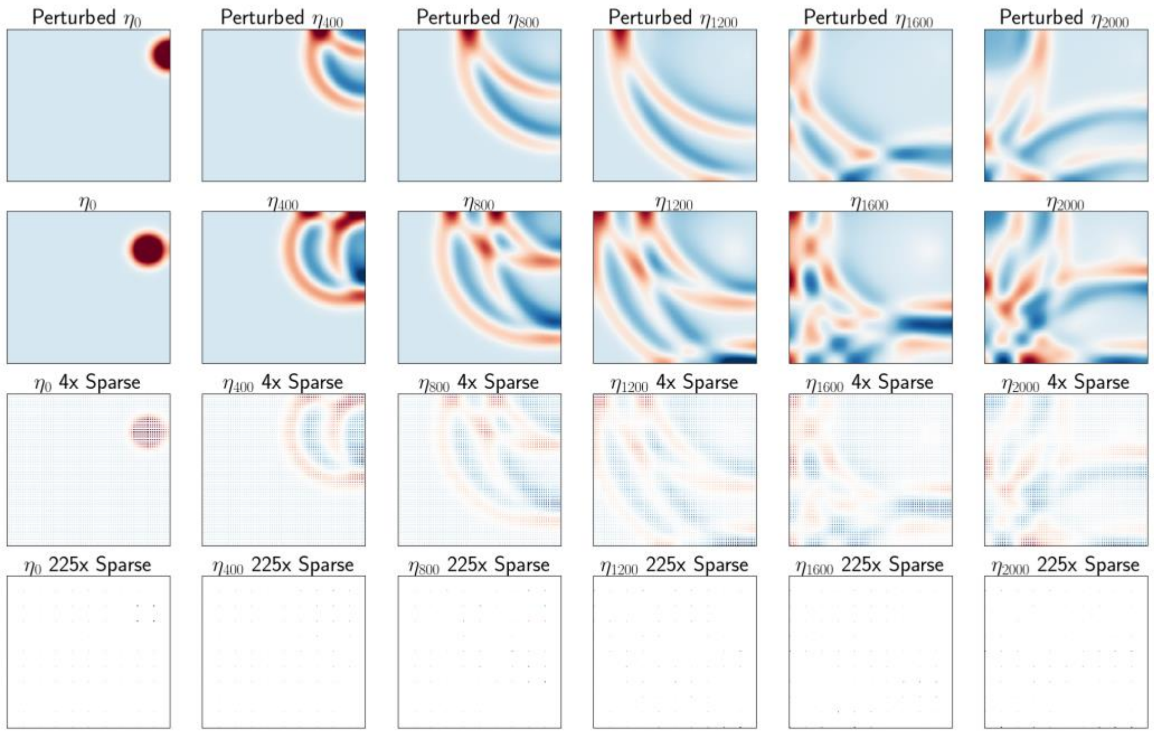


Figure 4: Evolution of states and sparse observations.

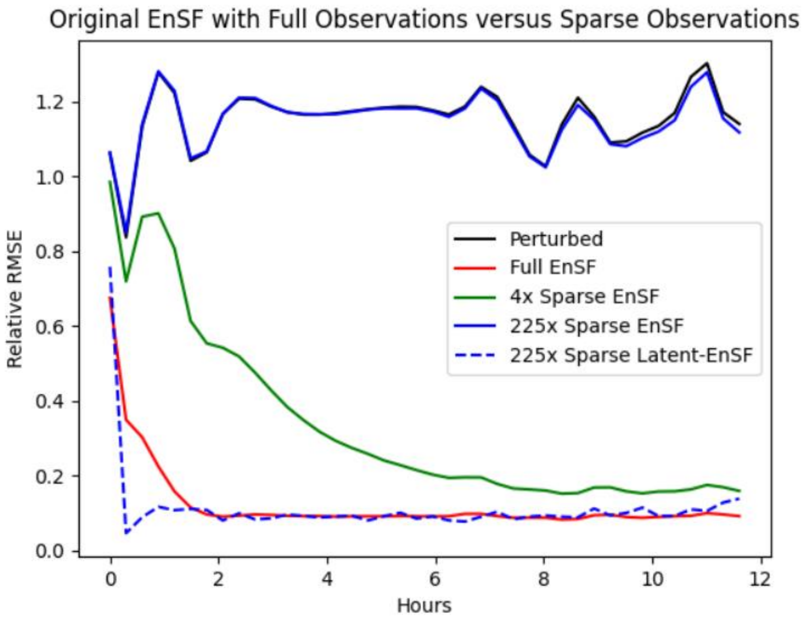


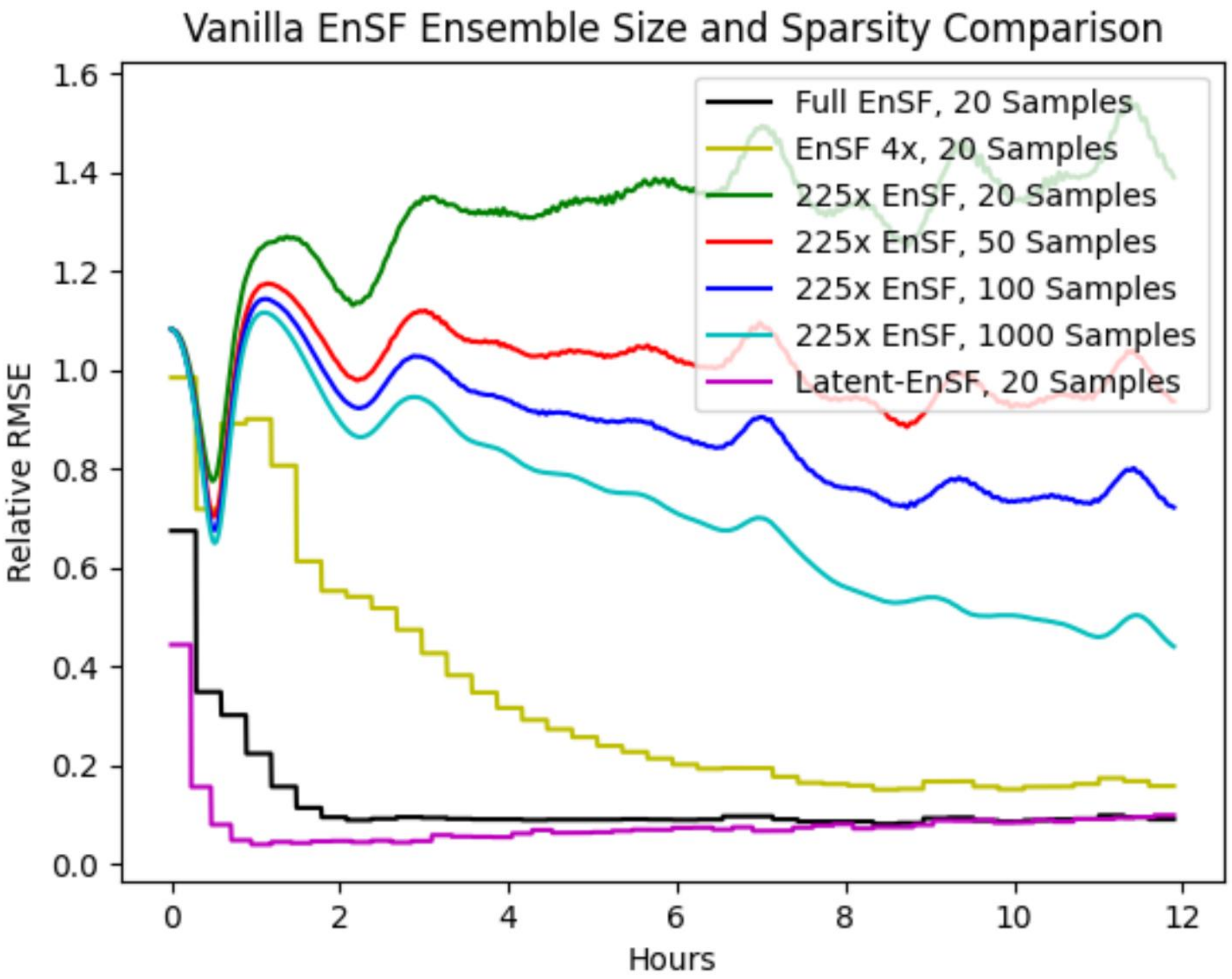
Figure 5: Relative RMSE of EnSF and Latent-EnSF for sparse observations.

$$\nabla_x \log P(y_t|x_{t,\tau}) = (y_t - H(x_{t,\tau}))^T \Gamma_t^{-1} \nabla_x H(x_{t,\tau})$$

Georgia Tech

Sparsity and the EnSF

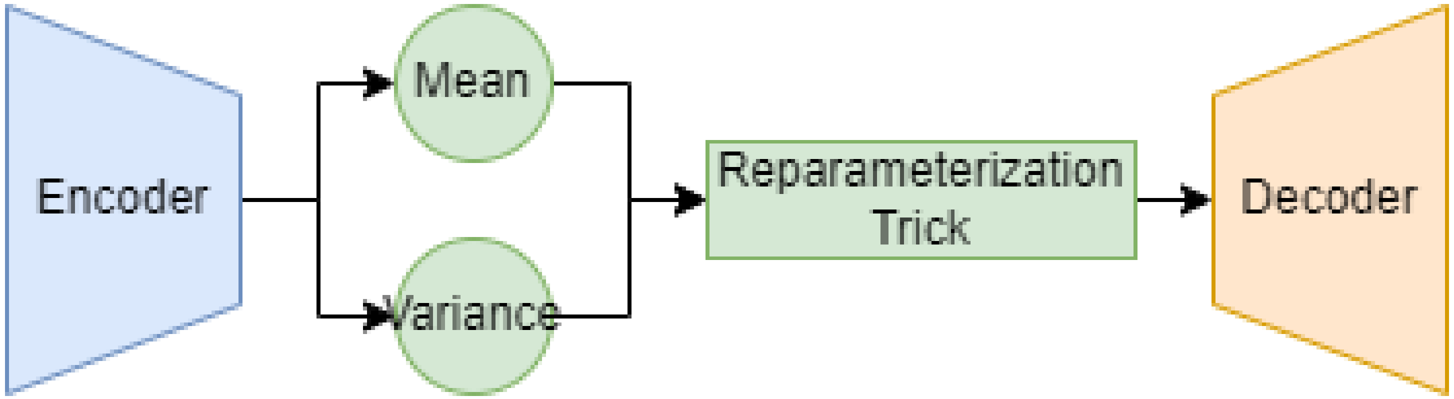
- EnSF gets better prior estimates by having more ensemble members



Variational Autoencoders (Kingma and Welling, 2022)

- Variational Autoencoders approximate a low-dimensional manifold of the original data by bottlenecking the latent dimension
- Allows for low-dimensional nonlinear representations
- Trained by minimizing for the ELBO loss, with a MSE loss in the state space and a KLD term (with respect to a N(0, 1) gaussian) in the latent space

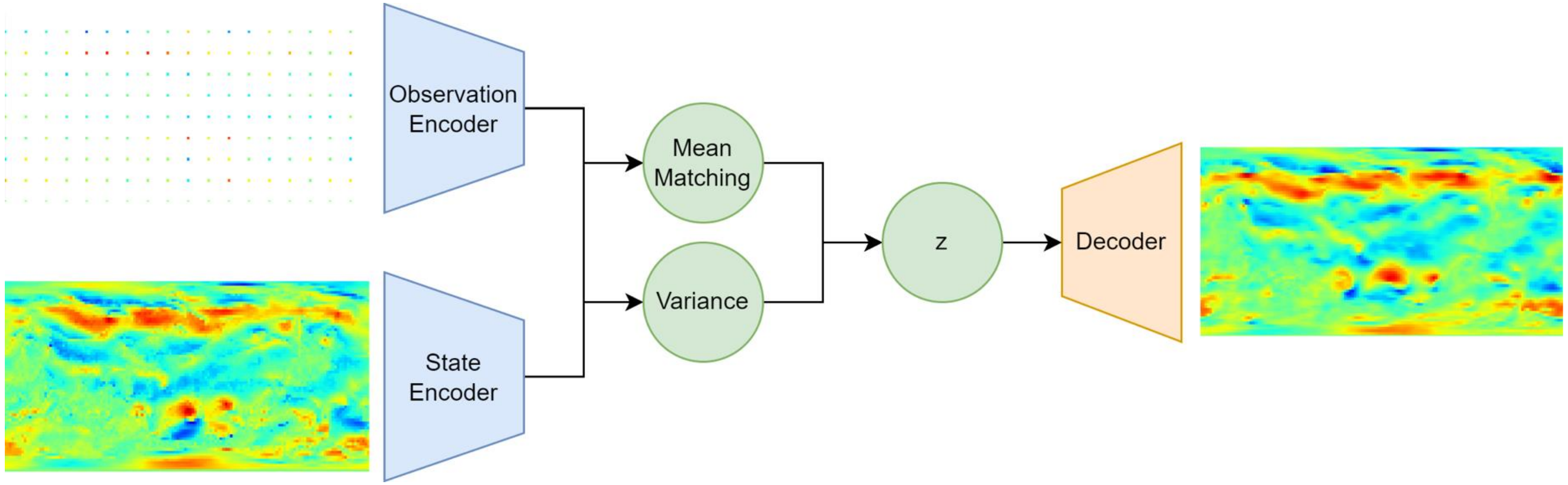
$$\mathrm{KLD}(\mu, \Sigma) = \sum_{i=1}^{r} -\log(\sigma_i) + \frac{\sigma_i^2 + \mu_i^2}{2} - \frac{1}{2}$$



Direct Observation Encoder Matching

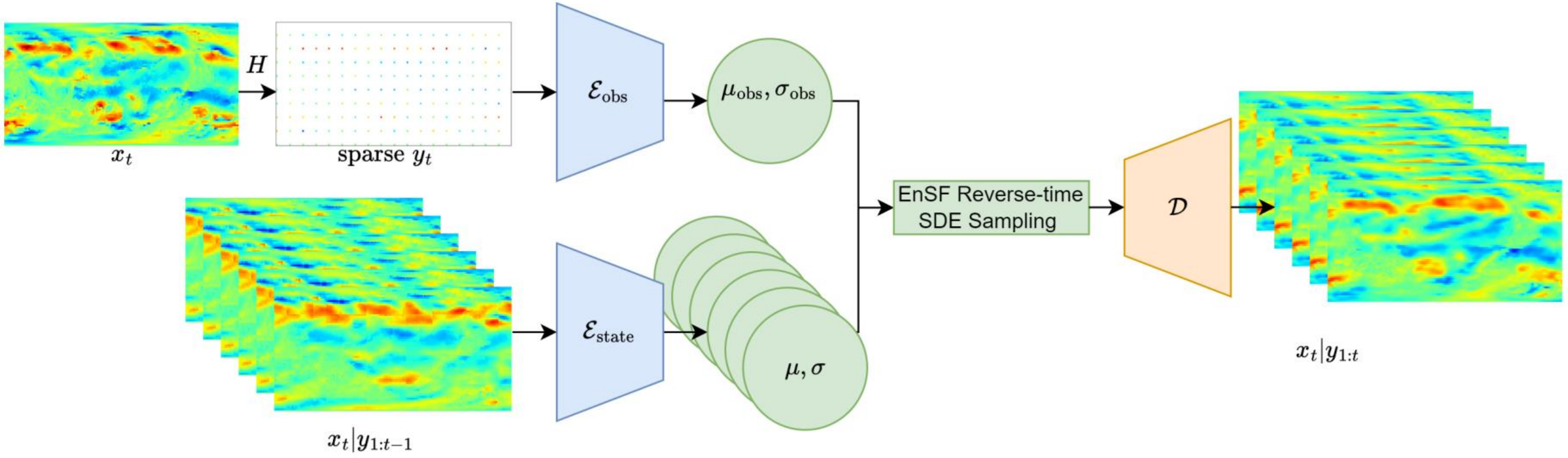
- When training the model, we match the means of the observation-space and state-space encoders and directly model the encoding without interpolation
- Loss can be written as

$$
\begin{aligned}
\ell_t(\theta) =& \|x_t - \mathcal{D}(z_t)\|_2^2 + \|x_t - \mathcal{D}(z_t^{\text{obs}})\|_2^2 && \text{(Reconstruction Term)} \\
&+ \|\mathcal{E}(x_t) - \mathcal{E}_{\text{obs}}(y_t)\|_2^2 && \text{(Latent Matching Term)} \\
&+ \lambda D_{\text{KL}}(\nu_t|\nu) + \lambda D_{\text{KL}}(\nu_t^{\text{obs}}|\nu) && \text{(Regularization Term)}
\end{aligned}
$$



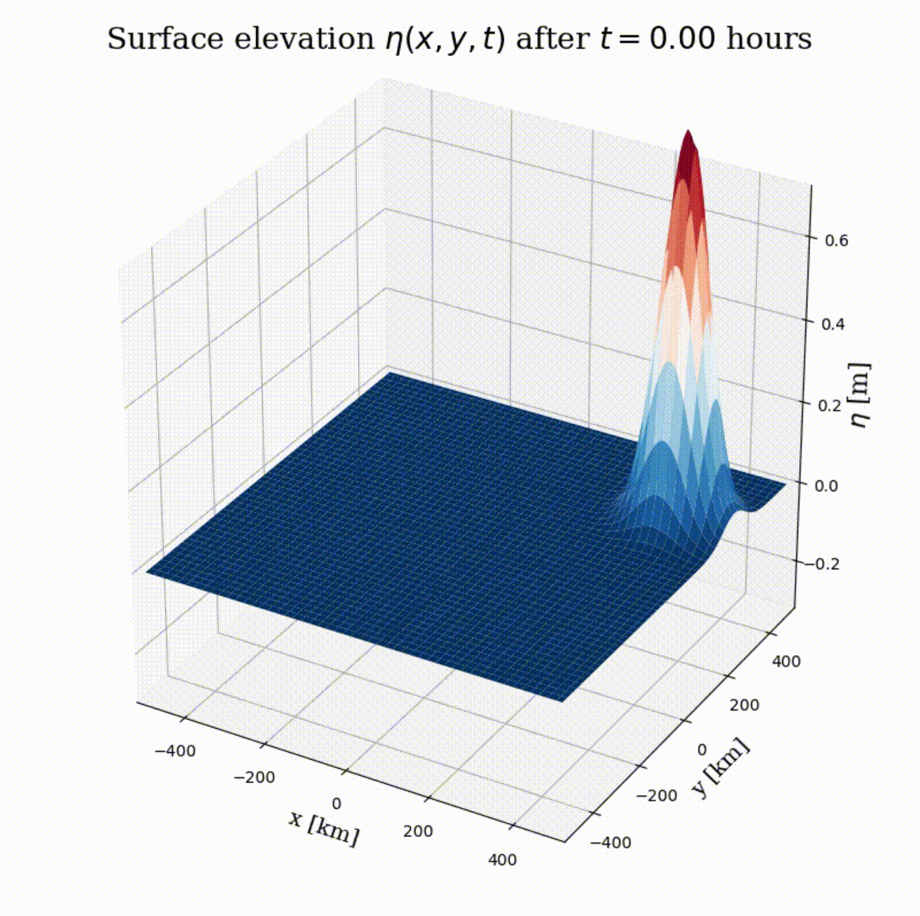
Relaxing Sparse Data Assimilation with EnSF

- Casting the sparse observations to the latent space results in the observation function being incorporated into the VAE versus explicitly defined
  - Allows for assimilation when the observation function is not explicitly known
- Additionally, since the VAEs regularize the latent space, the estimated covariance is easier to tune compared to standard autoencoders



Shallow-Water Equations and Example Simulation

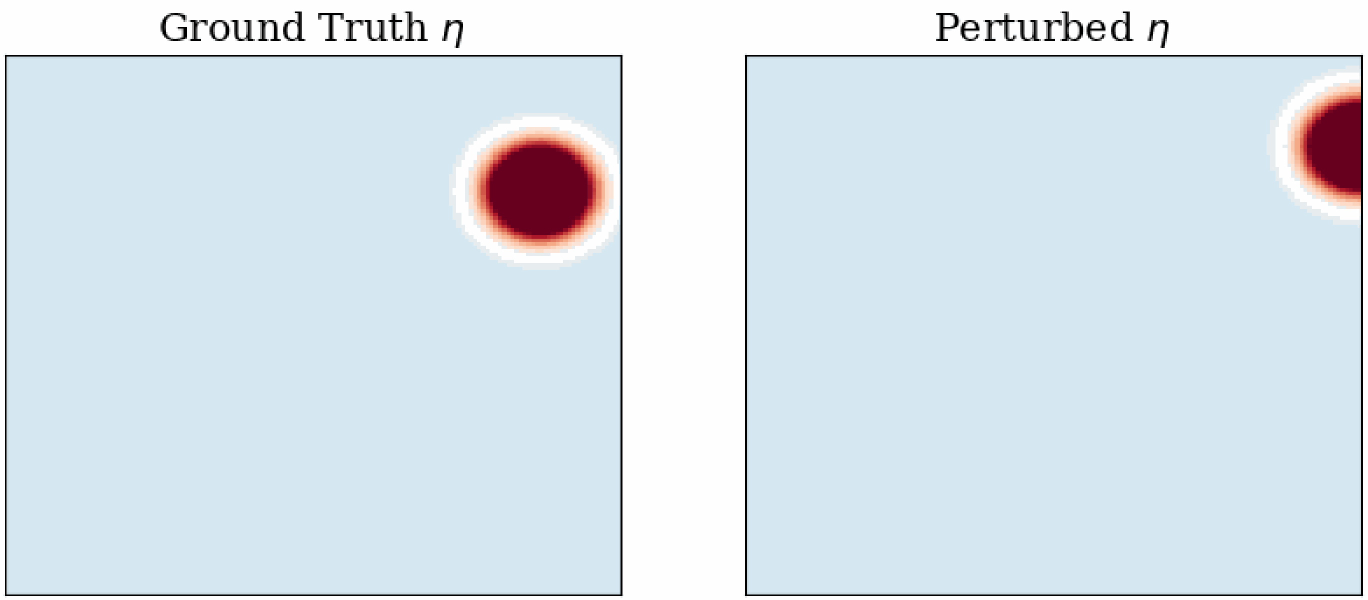
- Shallow-water equations originate from Navier-Stokes equations and model the flow
- Starts off with big gaussian perturbation
- Our simulations are then run for 2000 time steps using the upwind method, keeping track of height and velocity in the u and v directions
- Grid size of 150x150



Simulation code obtained from https://github.com/jostbr/shallow-water

Perturbed State

- We perturb the initial position of the gaussian bump which leads to different dynamics later down the line
- For following assimilation experiments, we will start out with the perturbed state and use observations from the true state for assimilation



Results

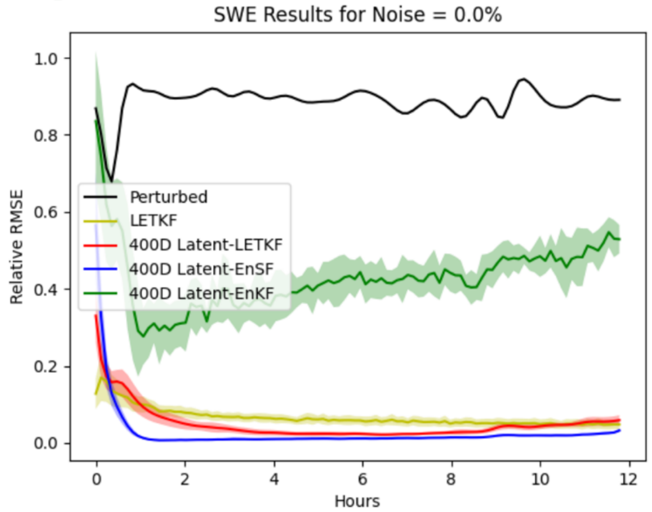


Figure 13: Relative RMSE of Latent-EnSF compared to baselines with no observation noise.

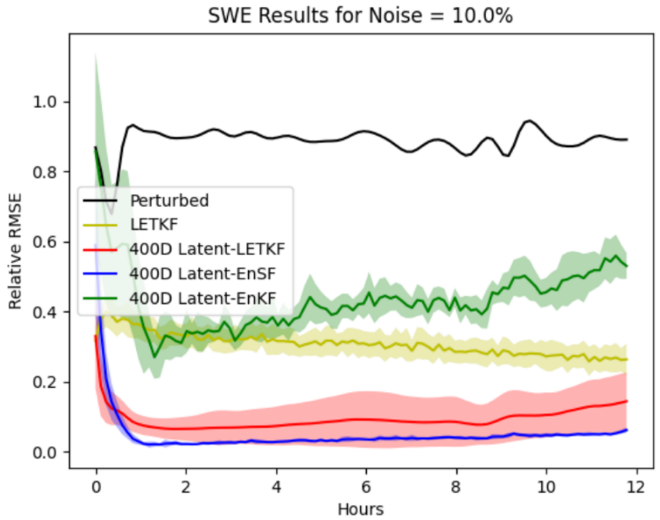


Figure 6: Relative RMSE of Latent-EnSF compared to baselines with 10% observation noise.

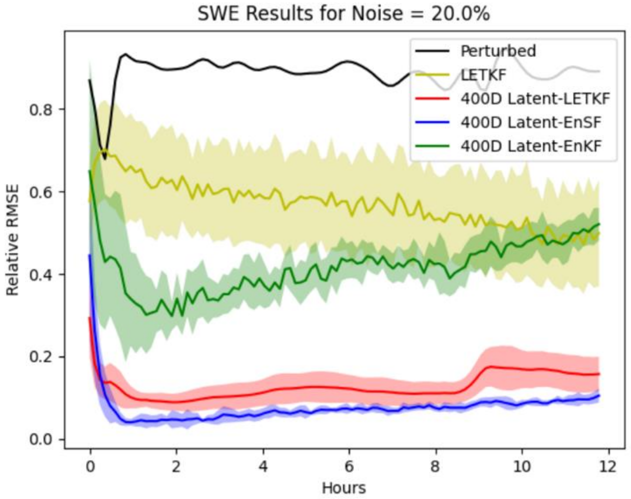
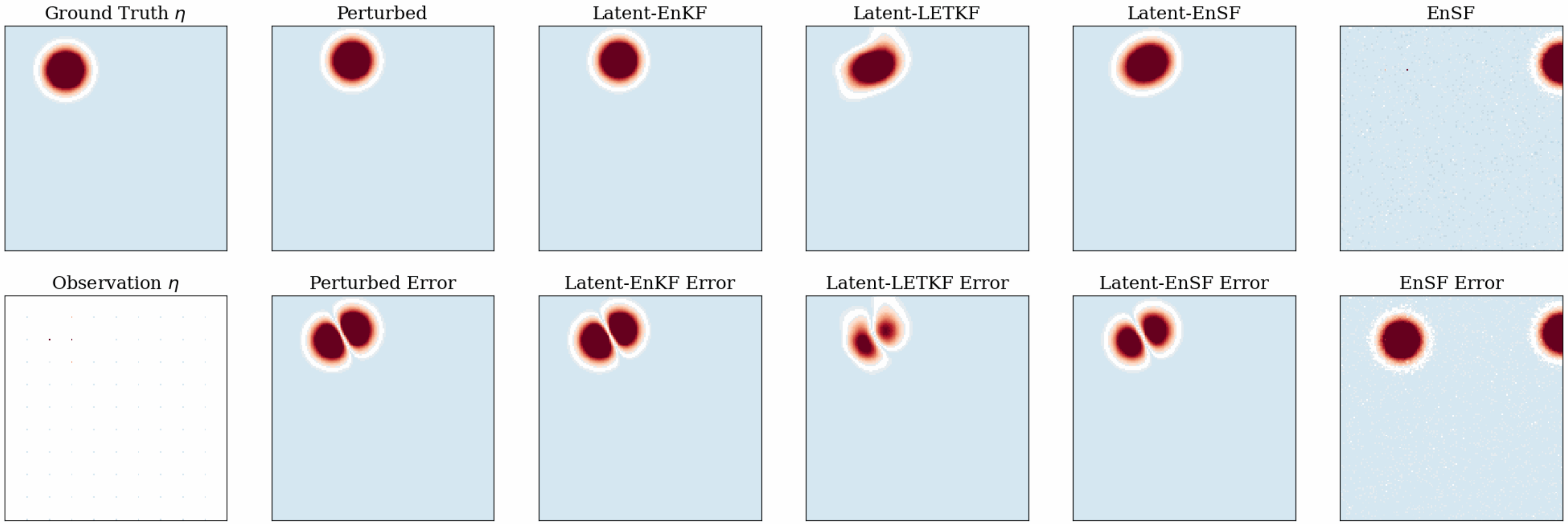


Figure 14: Relative RMSE of Latent-EnSF compared to baselines with 20% observation noise.

Animation – 400 Dimensional Latent Space



ERA5 Z500 and U500 Plot

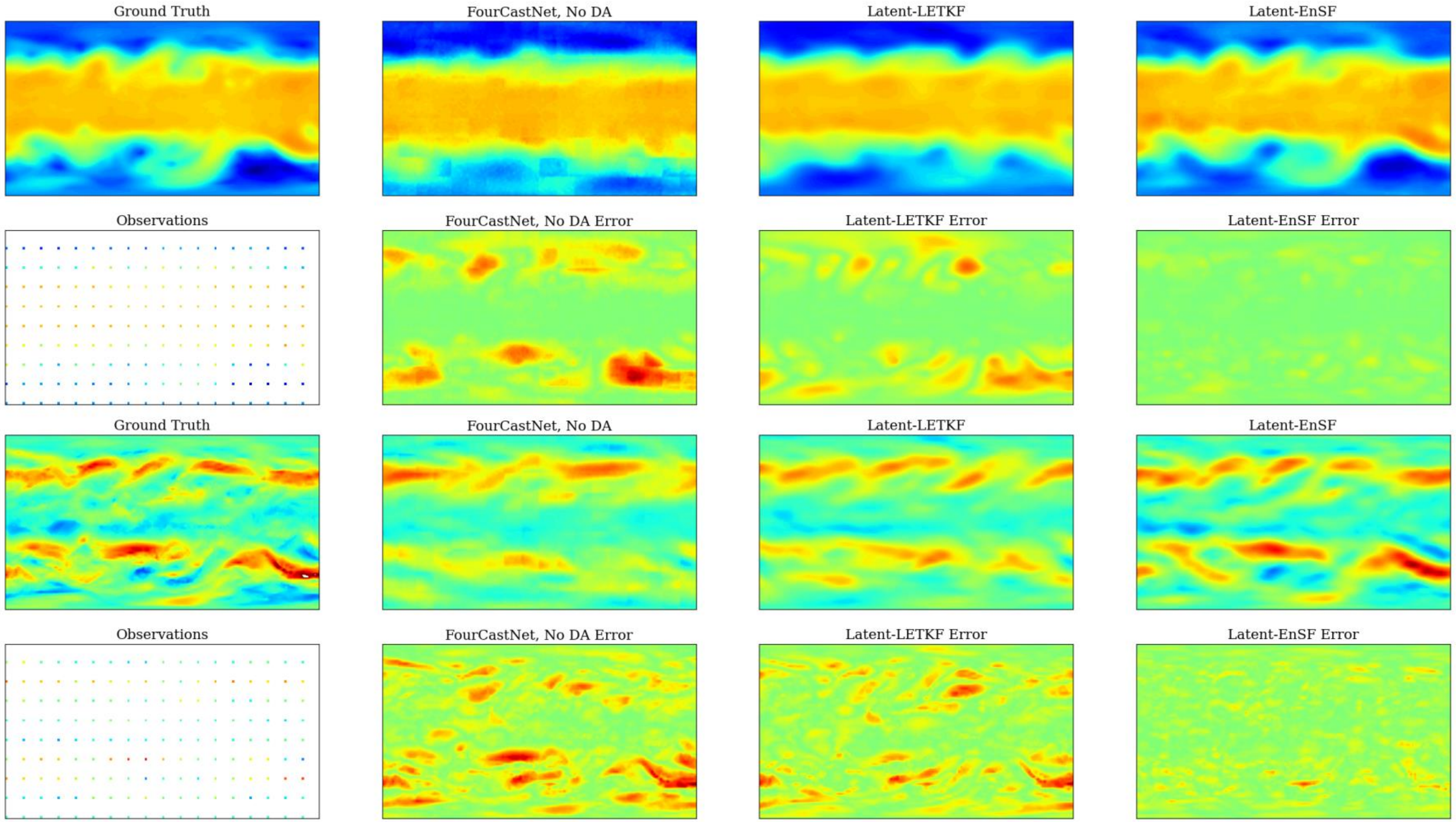
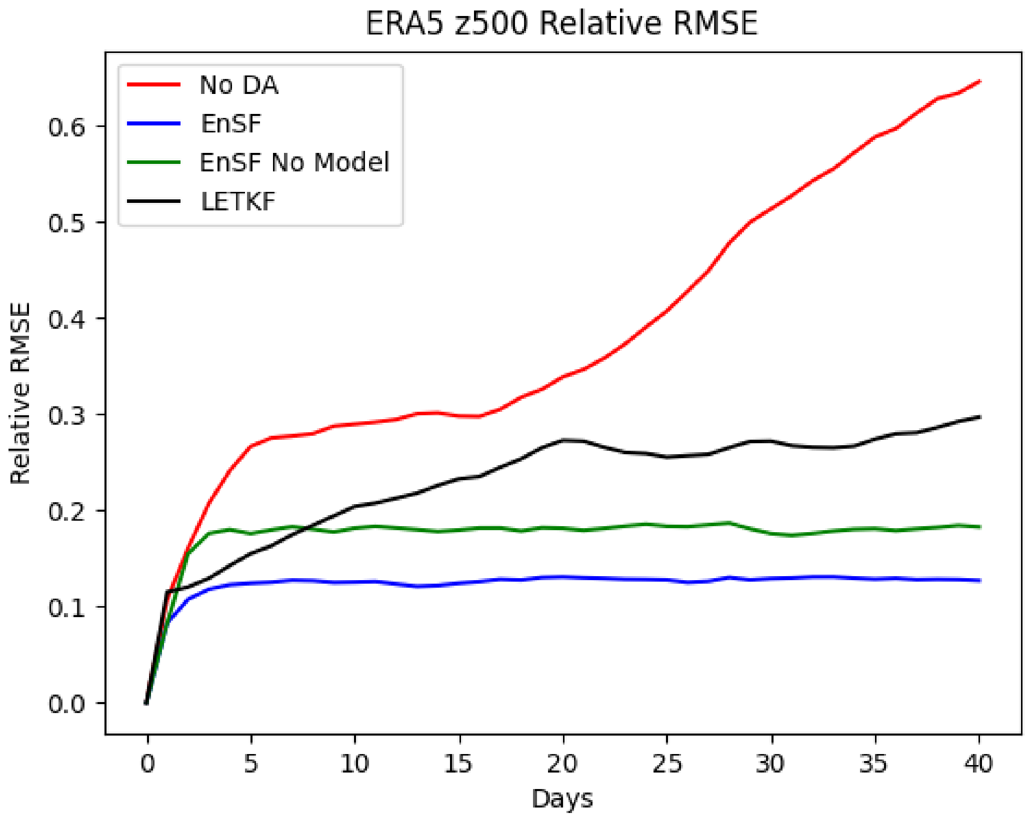


Figure 10: ERA5 Z500 (top two rows) and U500 (bottom two rows) medium-ranged weather forecasting samples after 41 days, along with the errors. Data assimilation is conducted once a day with 64x sparse observations. We compare against a baseline FourCastNet model and Latent-LETKF.

ERA5 Z500 Relative RMSE



Georgia Tech

References

- Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. 1960.
- Bao, F.; Zhang, Z.; Zhang, G. An Ensemble Score Filter for Tracking High-Dimensional Nonlinear Dynamical Systems. 2023.
- Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. 2022.
- Amendola, M.; Arcucci, R.; Mottet, L.; Casas, C. Q.; Fan, S.; Pain, C.; Linden, P.; Guo, Y.-K. Data Assimilation in the Latent Space of a Neural Network. 2020.
- C. Liu, R. Fu, D. Xiao, R. Stefanescu, P. Sharma, C. Zhu, S. Sun, C. Wang, EnKF data-driven reduced order assimilation system. Engineering Analysis with Boundary Elements. 2022.
- Sibo Cheng, I. Colin Prentice, Yuhan Huang, Yufang Jin, Yi-Ke Guo, Rossella Arcucci. Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. 2022.
- Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. 2021.
- Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay,Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, PedramHassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. 2022.
- Si, Phillip, and Peng Chen. "Latent-EnSF: A latent ensemble score filter for high-dimensional data assimilation with sparse observation data." *ICLR 2025.*

Georgia Tech

Ensemble Score Filters (EnSF) (Bao et. al, 2023)

- Ensemble Score Filters take a score-based approach which uses a diffusion-model inspired SDE approach to sample from $P(x_{t+1}|y_{0:t+1})$
- Utilize a reverse time SDE in pseudo-time step τ

$$dx = f(x,\tau)d\tau + g(\tau)dw$$

- w is a d-dimensional Wiener process, and f and g are defined as the following:

$$f(x_{t,\tau},\tau) = \frac{d\log\alpha_\tau}{d\tau}x_{t,\tau} \qquad g^2(\tau) = \frac{d\beta_\tau^2}{d\tau} - 2\frac{d\log\alpha_\tau}{d\tau}\beta_\tau^2$$

with $\alpha_\tau = 1-\tau(1-\epsilon_\alpha)$ and $\beta_\tau^2 = \epsilon_\beta + \tau(1-\epsilon_\beta)$

- This gives a conditional Gaussian distribution

$$x_{t,\tau}|x_{t,0} \sim N(\alpha_\tau x_{t,0}, \beta_\tau^2 I)$$

ERA5 Dataset

- ERA5 is a climate reanalysis dataset with data that goes back to 1940
- For the prediction model, we adapt the FourCastNet model (Pathak et. al, 2022) and assimilate the data once per day with our Latent-EnSF. Both prediction and VAE models are trained on 10 years of data