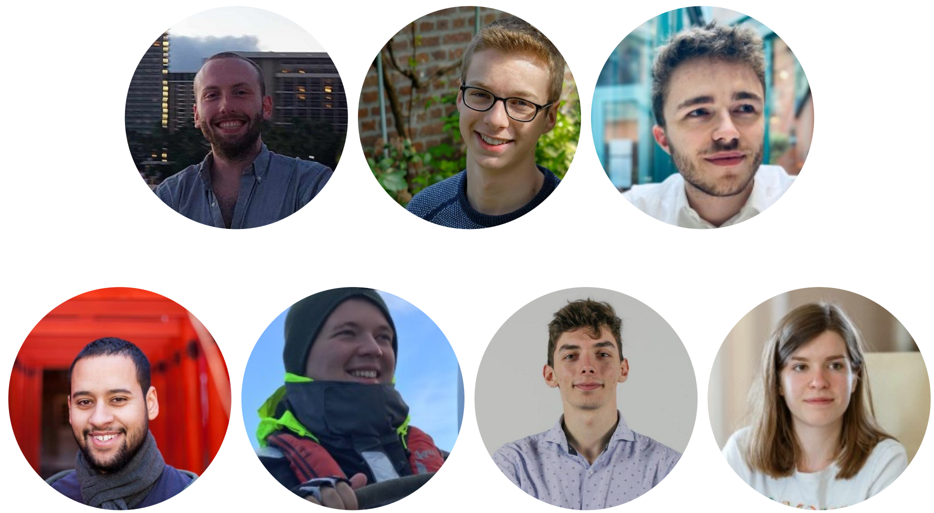
Appa: Bending weather dynamics with latent diffusion models

Séminaire UCLouvain
May 13, 2025

Gilles Louppe
g.louppe@uliege.be



(Gérome, François, Sacha, Omer, Victor, Matthias, Elise)

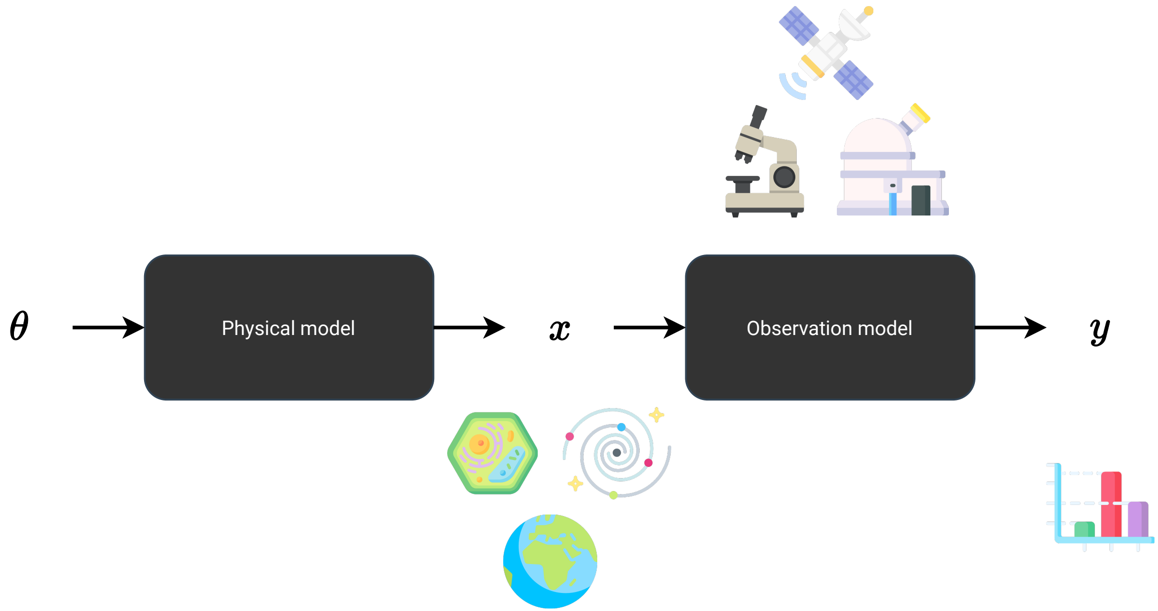
From a noisy observation y...

… can we recover all plausible images x?

$$\dot{u} = -u\nabla u + \frac{1}{Re}\nabla^2 u - \frac{1}{\rho}\nabla p + f$$

$$0 = \nabla \cdot u$$

... or parameters $\theta = \{Re, \rho, f\}$**?**

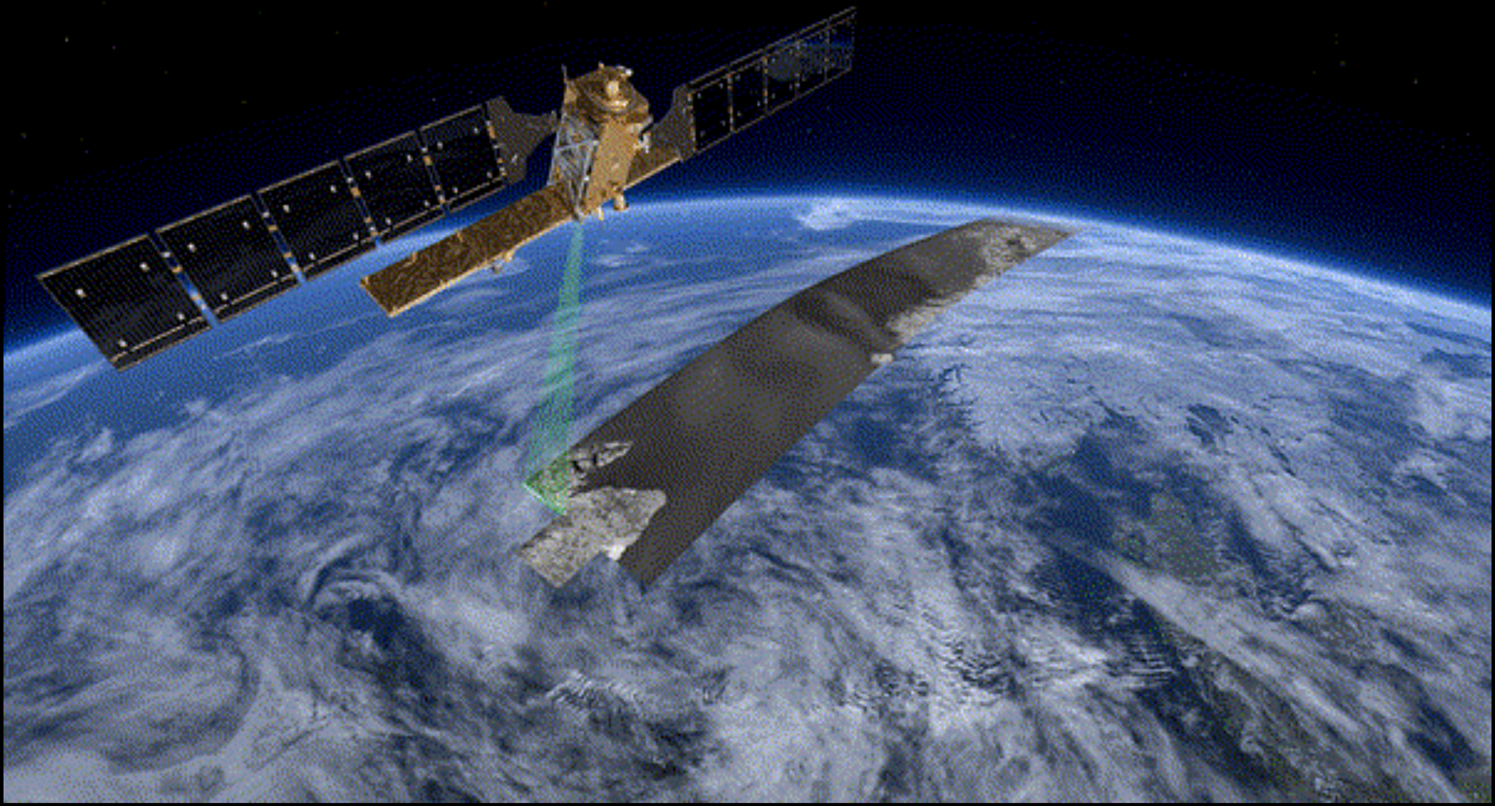


Bayesian inverse problems

Given a noisy observation y, estimate either

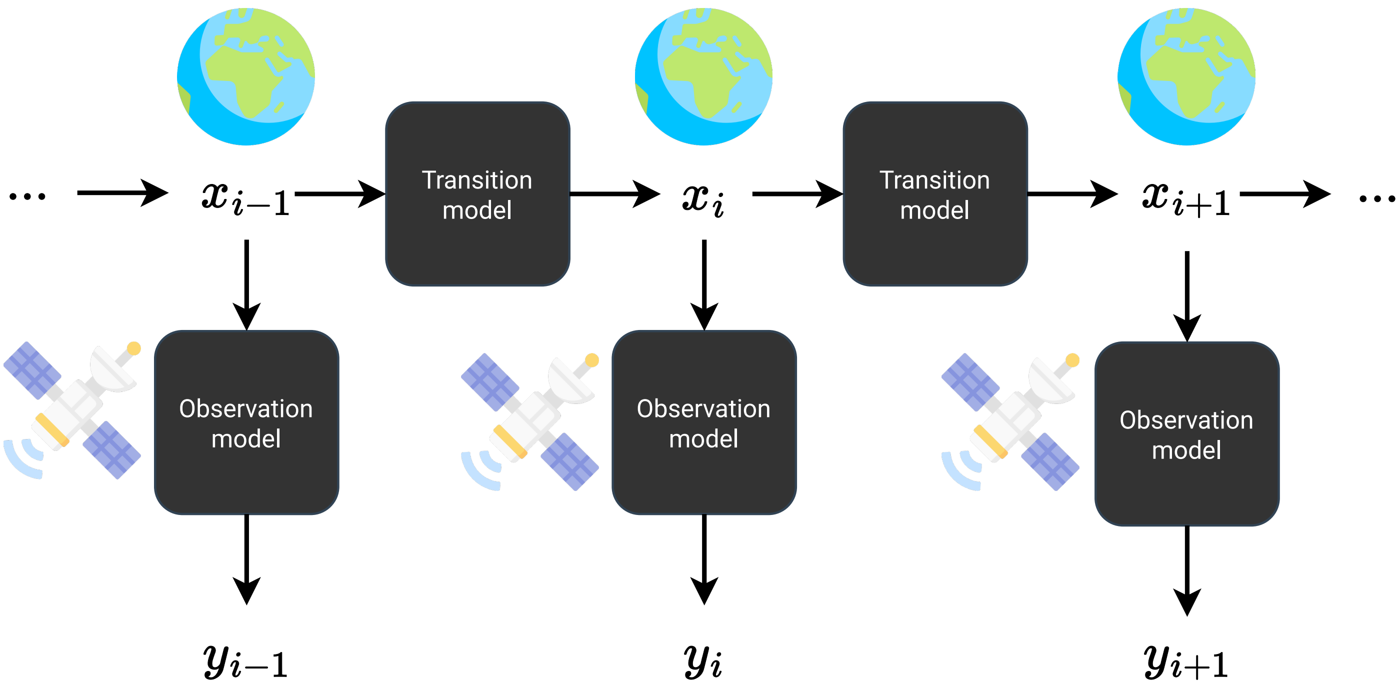
- the posterior distribution $p(x|y) \propto p(x)p(y|x)$ of latent states x, or
- the posterior distribution $p(\theta|y)$ of parameters θ,

assuming a prior $p(\theta)$ and forward models $p(x|\theta)$ and $p(y|x)$.



Today: from a sequence noisy observations $y_{1:L}$ (e.g., from satellites), can we recover past and present atmospheric states $x_{1:L}$?

Data assimilation



The goal of **data assimilation** is to estimate plausible trajectories $x_{1:L}$ given one or more noisy observations y (or $y_{1:L}$) as the posterior

$$p(x_{1:L}|y) = \frac{p(y|x_{1:L})}{p(y)} p(x_0) \prod_{i=1}^{L-1} p(x_{i+1}|x_i).$$

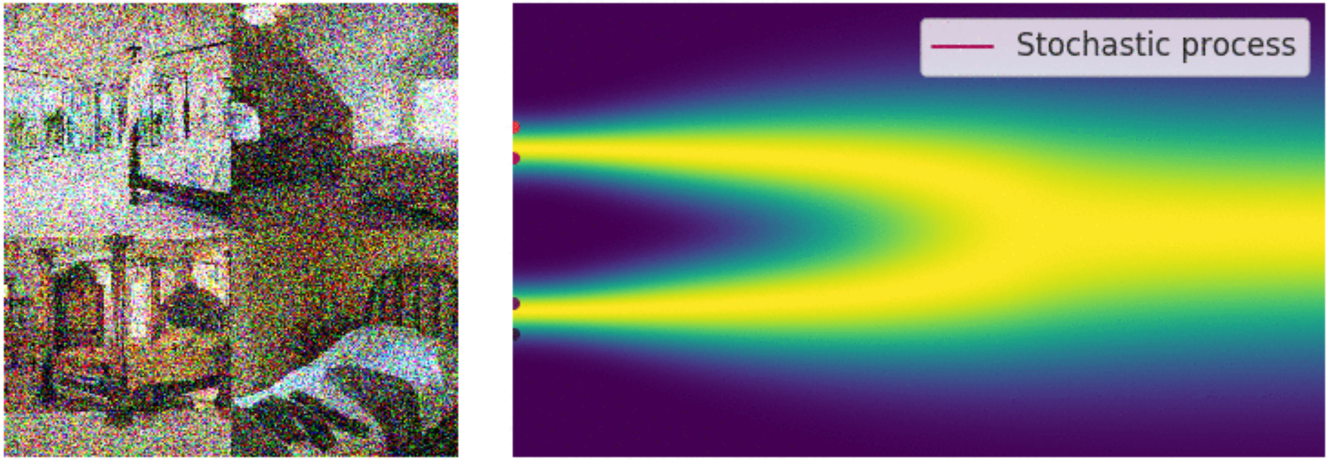
Diffusion models for inverse problems

Diffusion models 101

Samples $x \sim p(x)$ are progressively perturbed through a diffusion process described by the forward SDE

$$\text{d}x_t = f_t x_t \text{d}t + g_t \text{d}w_t,$$

where x_t is the perturbed sample at time t.



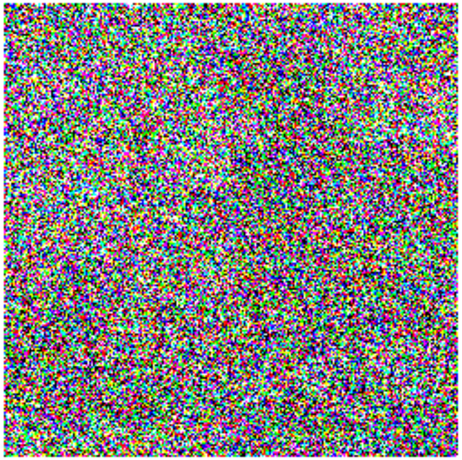
Forward diffusion process.

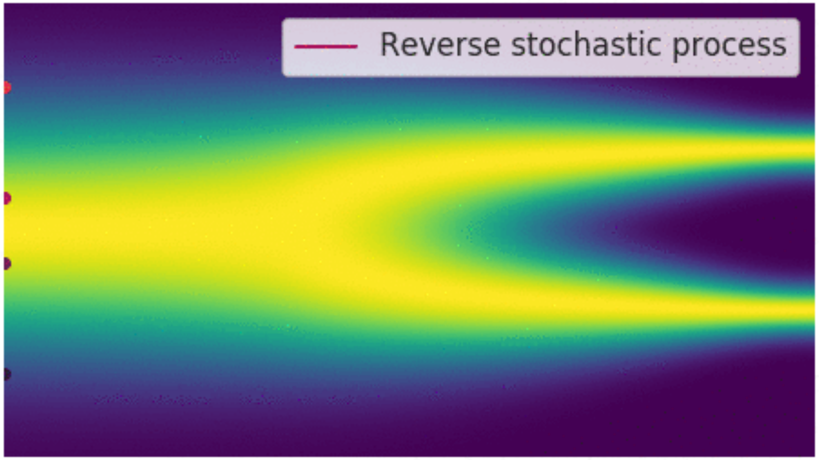
—
Credits: Song, 2021.

The reverse process satisfies a reverse-time SDE that can be derived analytically from the forward SDE as

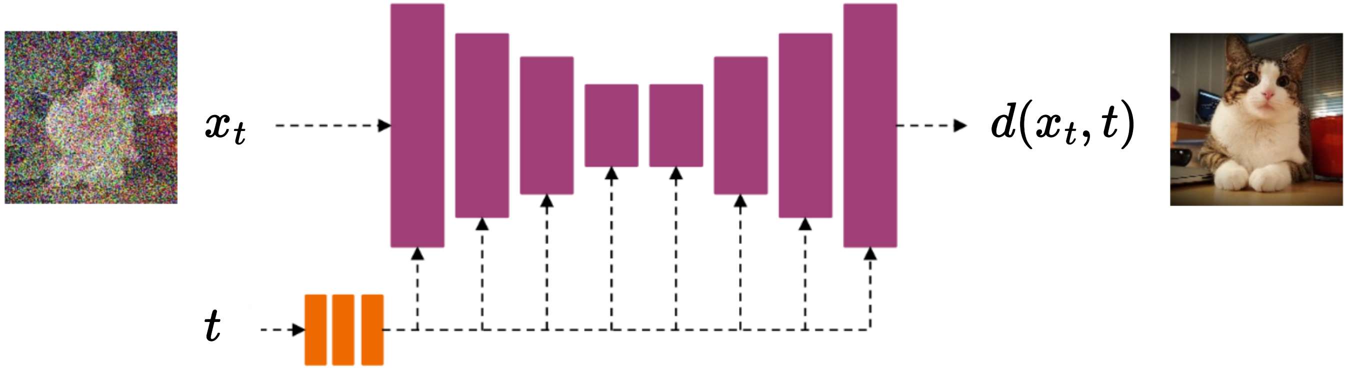
$$\mathrm{d}x_t = \left[f_t x_t - g_t^2 \nabla_{x_t} \log p(x_t)\right] \mathrm{d}t + g_t \mathrm{d}w_t.$$

Therefore, to generate data samples $x_0 \sim p(x_0) \approx p(x)$, we can draw noise samples $x_1 \sim p(x_1) \approx \mathcal{N}(0, \Sigma_1)$ and gradually remove the noise therein by simulating the reverse SDE from $t = 1$ to 0.





Reverse denoising process.



The score function $\nabla_{x_t} \log p(x_t)$ is unknown, but can be approximated by a neural network $d_\theta(x_t, t)$ by minimizing the denoising score matching objective

$$\mathbb{E}_{p(x)p(t)p(x_t|x)} \left[||d_\theta(x_t, t) - x||_2^2 \right] .$$

The optimal denoiser d_θ is the mean $\mathbb{E}[x|x_t]$ which, via Tweedie's formula, relates to the score function as

$$\nabla_{x_t} \log p(x_t) \approx \Sigma_t^{-1}(d_\theta(x_t, t) - x_t).$$

Inverting an observation

Let's assume an observation $y \sim p(y|x)$ for some unknown latent state x.

To turn a diffusion model $p_\theta(x)$ into a conditional model $p_\theta(x|y)$, we can add conditioning information y at each step of the reverse process by adding y as an additional input to the denoiser $d_\theta(x_t, t, y)$.

Using the Bayes' rule, the posterior score $\nabla_{x_t} \log p(x_t|y)$ to inject in the reverse SDE can be decomposed as

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(y|x_t) - \cancel{\nabla_{x_t} \log p(y)}.$$

This enables **zero-shot posterior sampling** from a diffusion prior $p(x_0)$ without having to pre-wire the denoiser to the observation model $p(y|x)$.

Approximating $\nabla_{x_t} \log p(y|x_t)$

We want to estimate the score $\nabla_{x_t} \log p(y|x_t)$ of the noise-perturbed likelihood

$$p(y|x_t) = \int p(y|x)p(x|x_t)\mathrm{d}x.$$

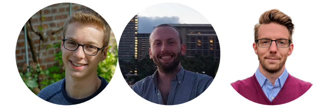
Our approach:

- Assume a linear Gaussian observation model $p(y|x) = \mathcal{N}(y|Ax, \Sigma_y)$.
- Assume the approximation $p(x|x_t) \approx \mathcal{N}(x|\mathbb{E}[x|x_t], \mathbb{V}[x|x_t])$, where $\mathbb{E}[x|x_t]$ is estimated by the denoiser and $\mathbb{V}[x|x_t]$ is estimated using Tweedie's covariance formula.
- Then $p(y|x_t) \approx \mathcal{N}(y|A\mathbb{E}[x|x_t], \Sigma_y + A\mathbb{V}[x|x_t]A^T)$.
- The score $\nabla_{x_t} \log p(y|x_t)$ then approximates to

$$\nabla_{x_t}\mathbb{E}[x|x_t]^T A^T (\Sigma_y + A\mathbb{V}[x|x_t]A^T)^{-1}(y - A\mathbb{E}[x|x_t]).$$

—

See also Daras et al (2024) for a survey on diffusion models for inverse problems.



Approxim

We want to e likelihood

Our approac

- Assume , $\Sigma_y)$.
- Assume here $\mathbb{E}[x|x_t]$ ng Tweedie
- Then $p($
- The sco

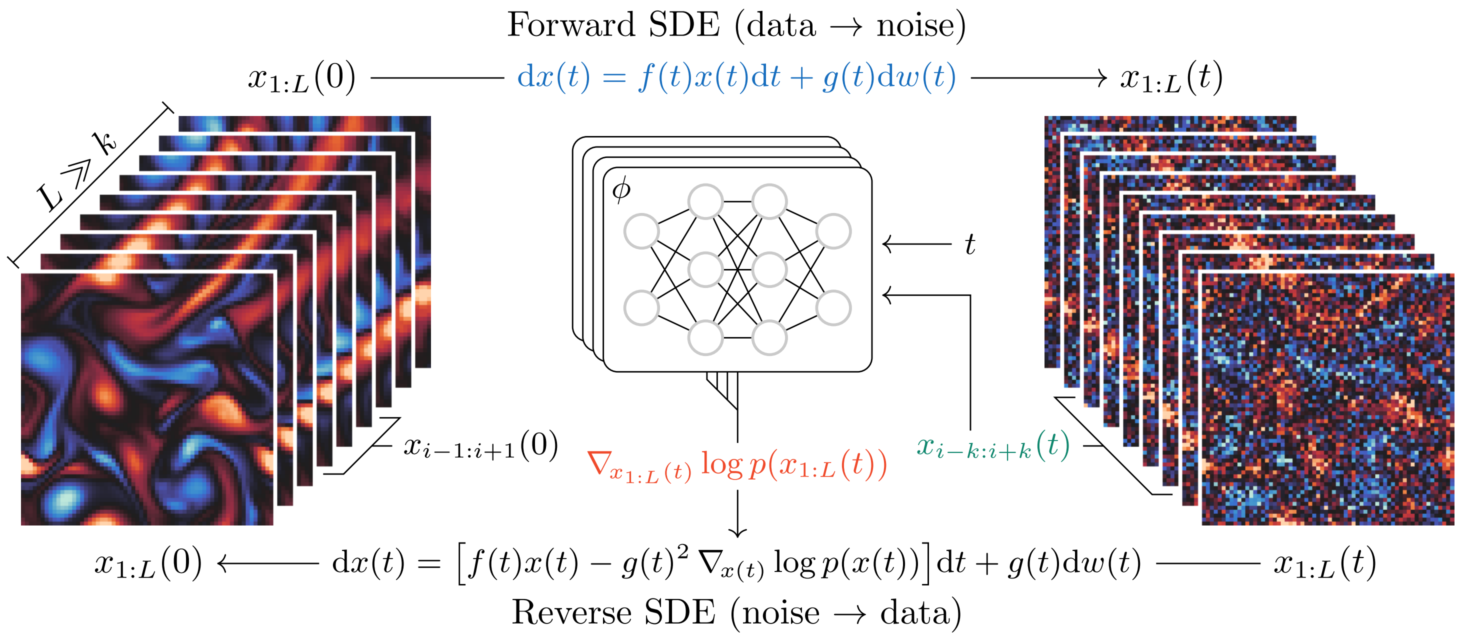
]).

—

See also Daras et al (2024) for a survey on diffusion models for inverse problems.

Data assimilation with diffusion models

Score-based data assimilation (SDA)

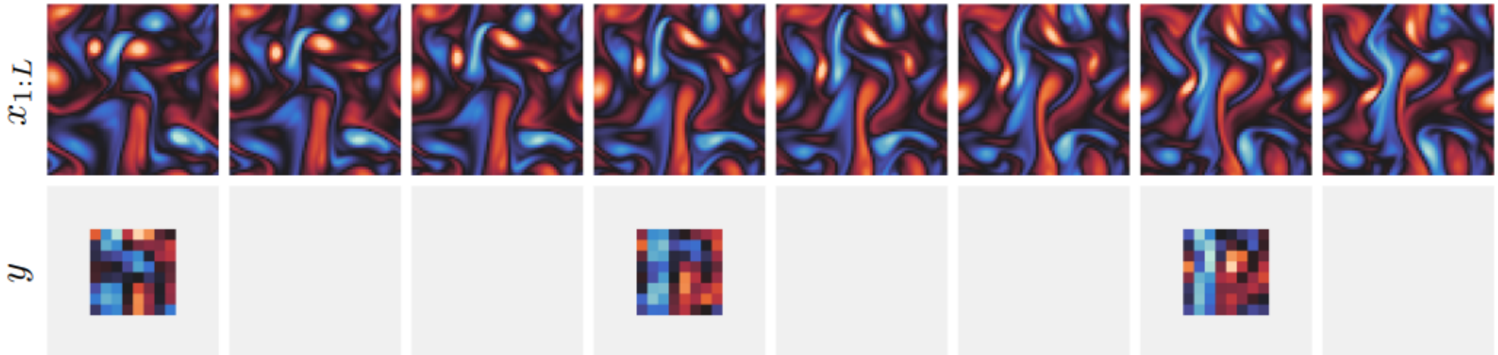


Our approach:

- Build a score-based generative model $p(x_{1:L})$ of arbitrary-length trajectories.
- Use zero-shot posterior sampling to generate plausible trajectories from noisy observations y.

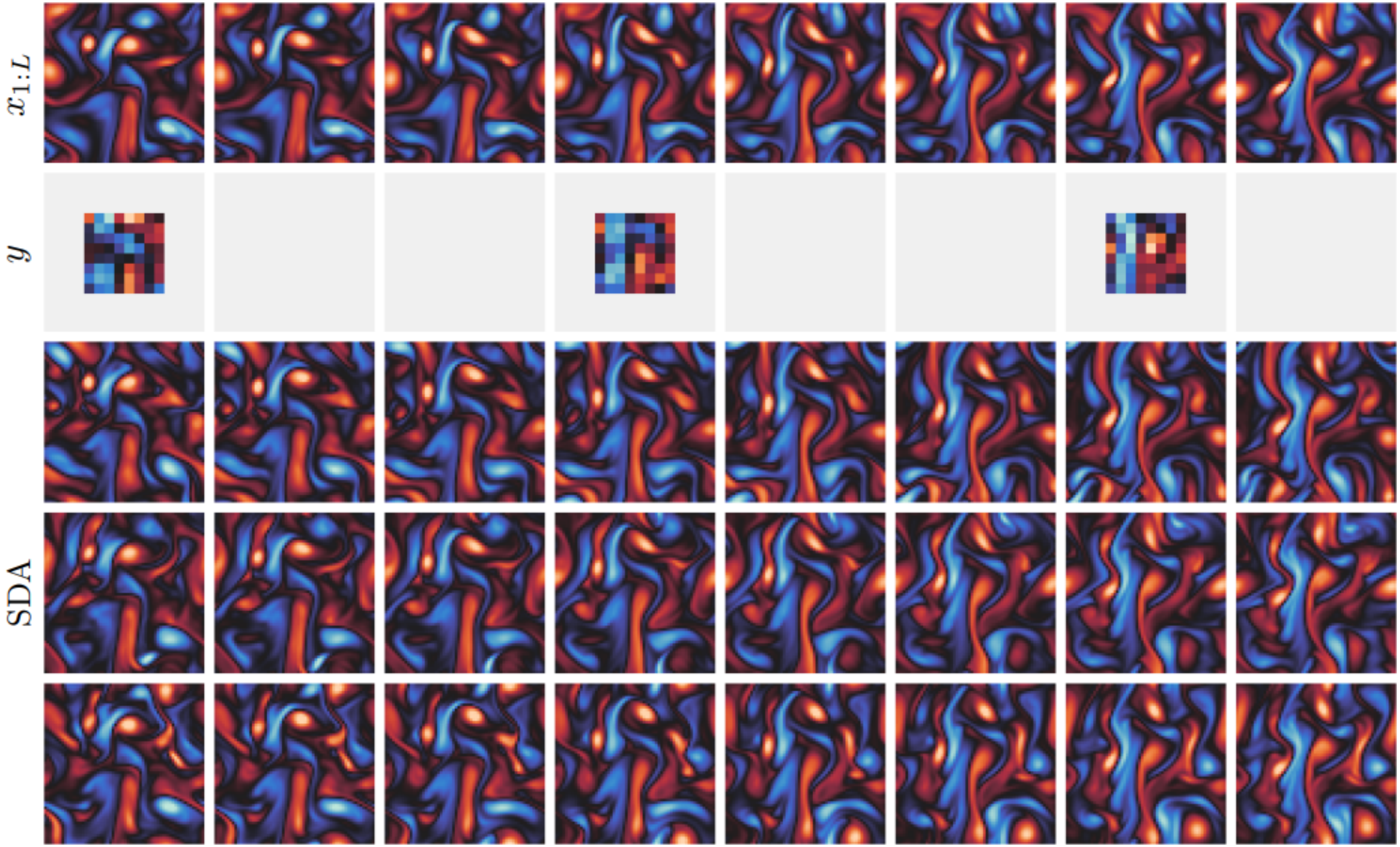
—

Credits: Rozet and Louppe, 2023 (arXiv:2306.10574).



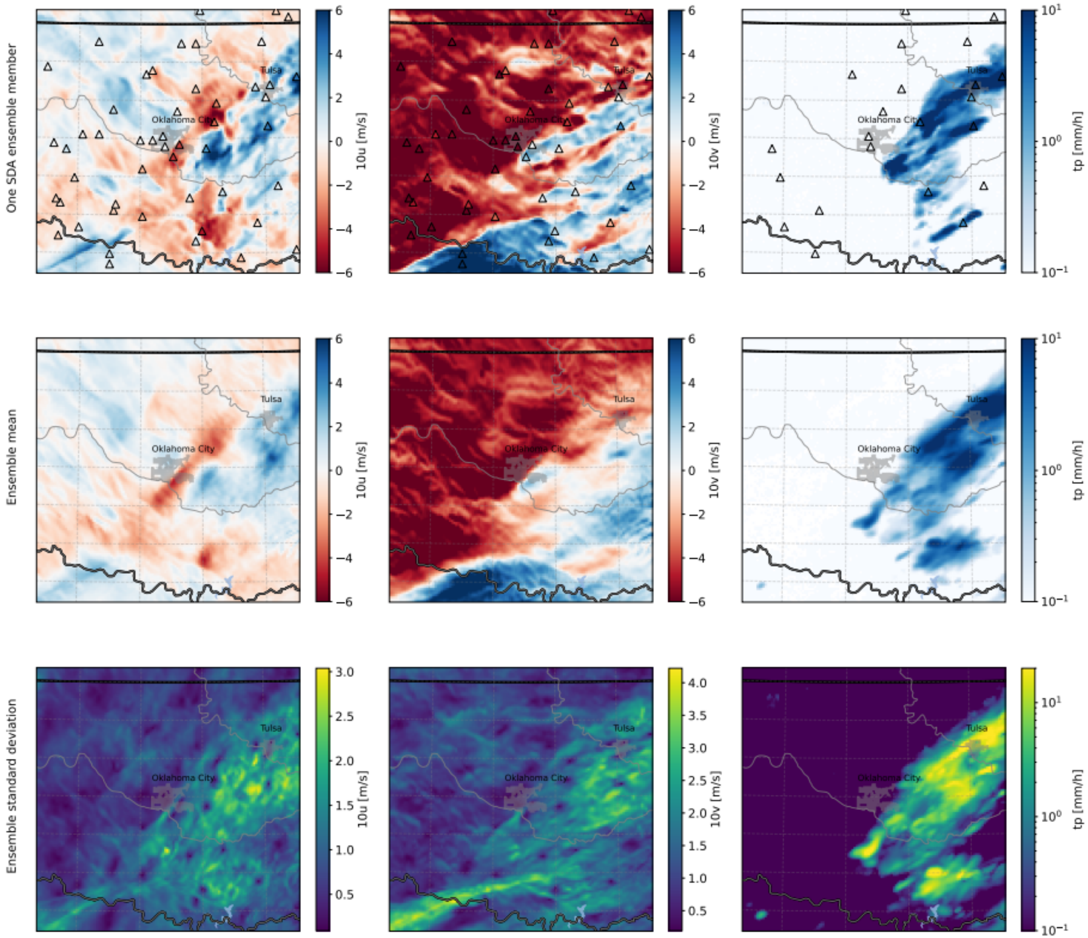
Sampling trajectories $x_{1:L}$ from
noisy, incomplete and coarse-grained observations y.

—
Credits: Rozet and Louppe, 2023 (arXiv:2306.10574).



Sampling trajectories $x_{1:L}$ from
noisy, incomplete and coarse-grained observations y.

—
Credits: Rozet and Louppe, 2023 (arXiv:2306.10574).



SDA can assimilate noisy weather observations
to produce stochastic ensembles at the regional scale.

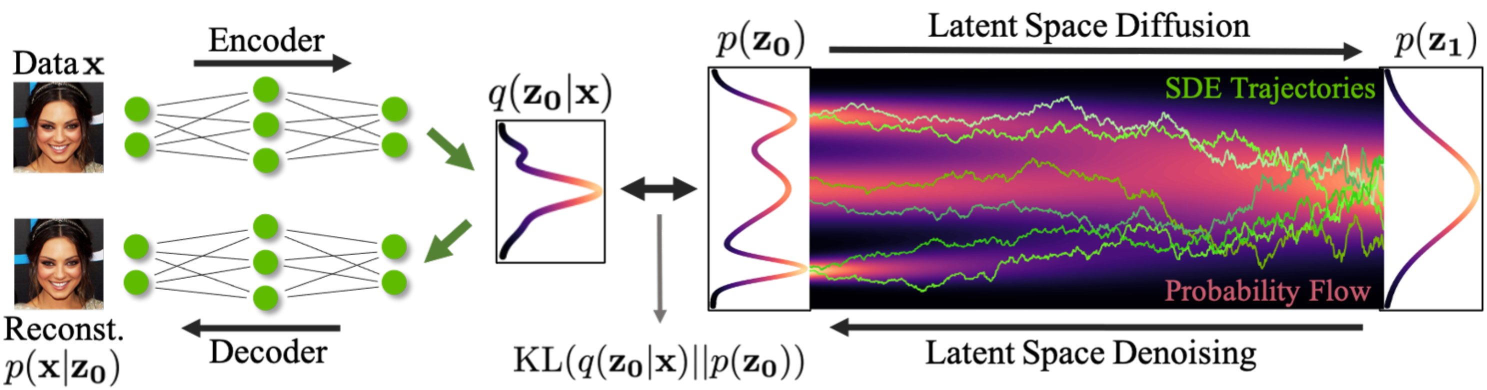
—
Credits: Manshausen et al, 2024 (arXiv:2406.16947).

Scaling SDA to a high-resolution Earth model?

Latent diffusion models (LDMs)

Diffusion models scale to high-dimensional data by generating in the latent space of an autoencoder $x \approx D(E(x))$.

If x is the data and $z = E(x)$ is the latent representation produced by an encoder E, then the diffusion model generates $z \sim p_\theta(z)$ which is then decoded into $x = D(z)$ by a decoder D.



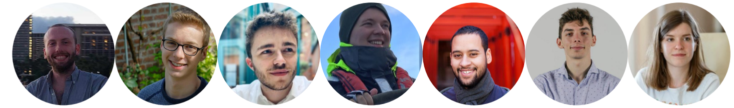
—
Credits: Vahdat et al, 2021.

LDMs for inverse problems

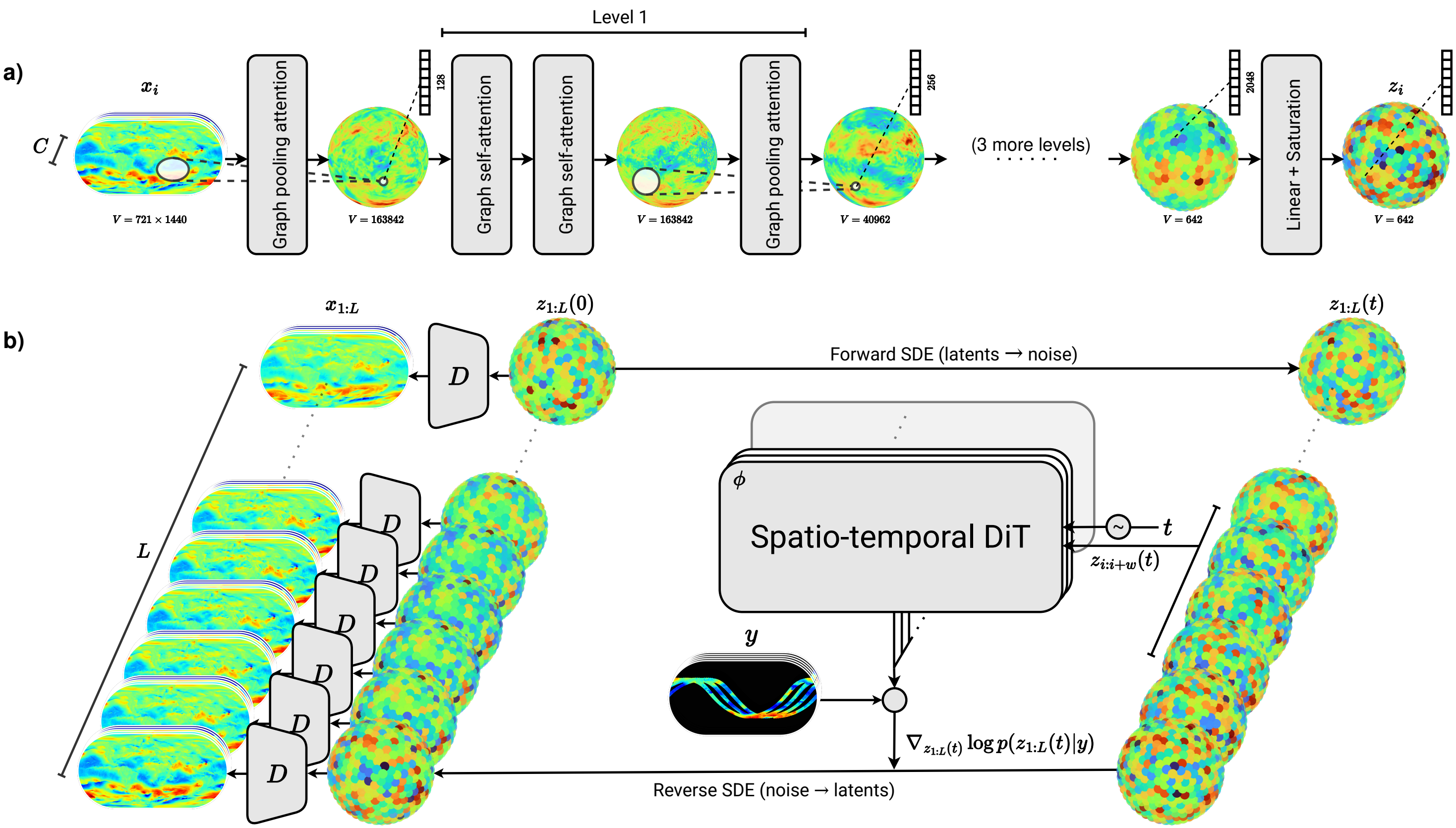
$$x \sim p_\theta(x|y) \Leftrightarrow z \sim p_\theta(z|y), x = D(z)$$

But how to sample from $p_\theta(z|y)$?

- If the observation model in data space is $p(y|x) = \mathcal{N}(y|M(x), \Sigma_y)$, then an observation model in the latent space can be defined as $p(y|z) = \mathcal{N}(y|M(D(z)), \Sigma_y)$.
- Regular posterior sampling algorithms can then be used to sample from the posterior distribution $p_\theta(z|y)$.



Appa: Bending weather dynamics with LDMs



—

Credits: Andry et al, 2025 (arXiv:2504.18720).

Appa is made of three components:

- a 500M-parameter **autoencoder** that compresses the data space x into a latent space z with a 450x compression factor;
- a 1B-parameter **diffusion model** that generates latent trajectories $z_{1:L}$;
- a zero-shot **posterior sampling algorithm** adapted from MMPS (Rozet et al, 2024) that samples from the posterior distribution $p(z_{1:L}|y)$.

—

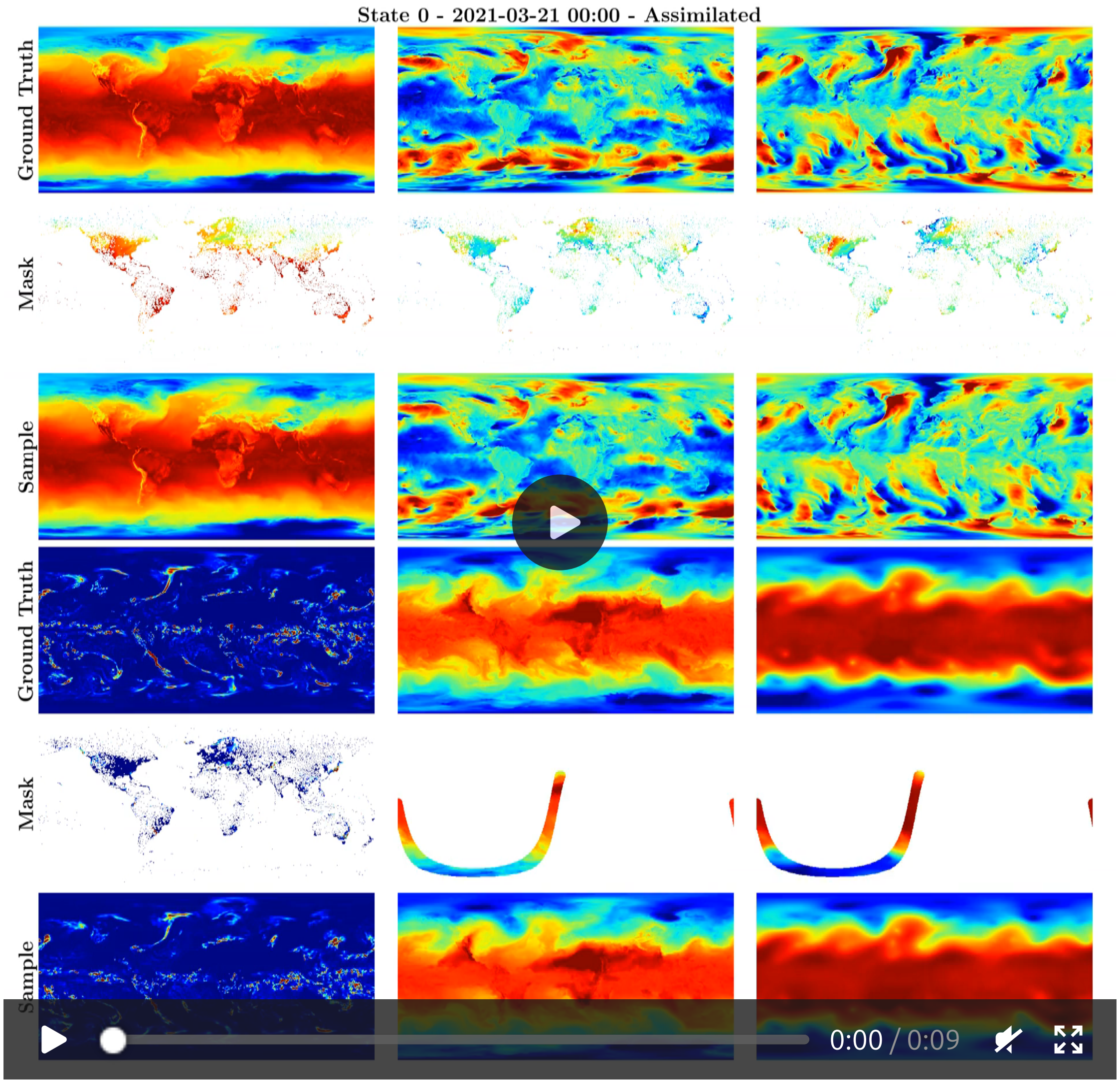
Credits: Andry et al, 2025 (arXiv:2504.18720).

Appa can solve multiple inverse problems **without retraining**:

- reanalysis of past atmospheric states $p(x_{1:L} | y_{1:L})$;
- filtering of the present atmospheric state $p(x_L | y_{1:L})$;
- forecasting of future atmospheric states $p(x_{L+1:L+M} | y_{1:L})$ or $p(x_{L+1:L+M} | x_L)$.

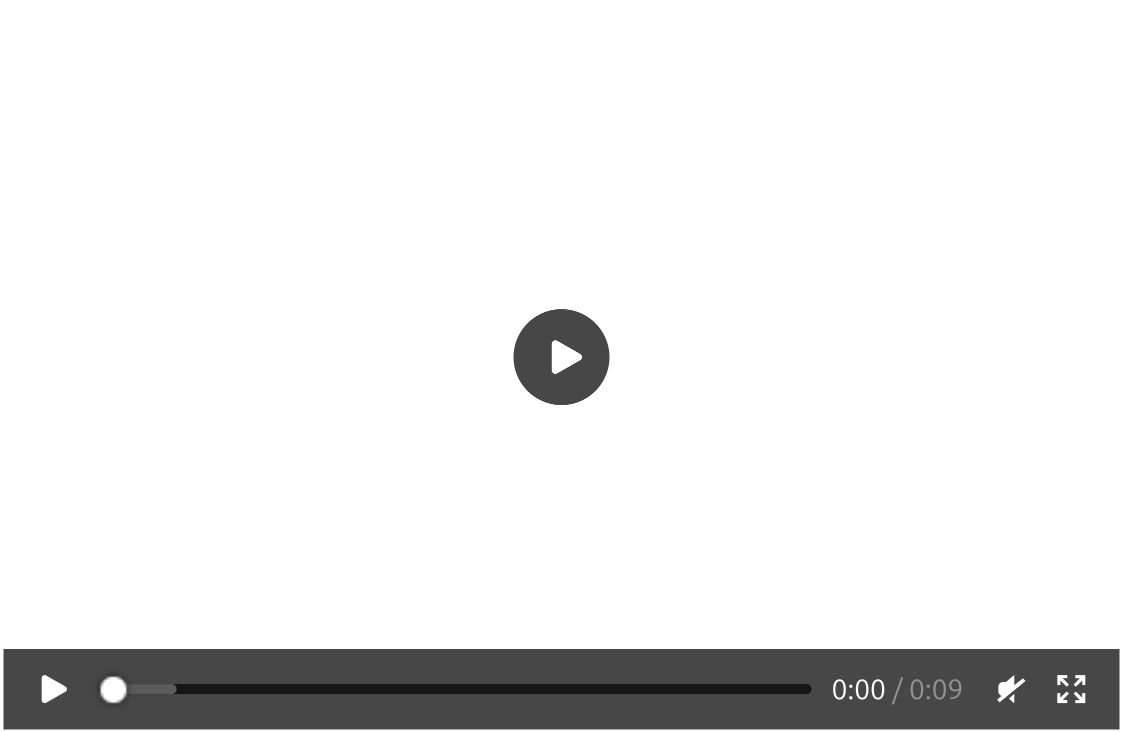
—

Credits: Andry et al, 2025 (arXiv:2504.18720).



Reanalysis of past data $p(x_{1:L} | y_{1:L})$.

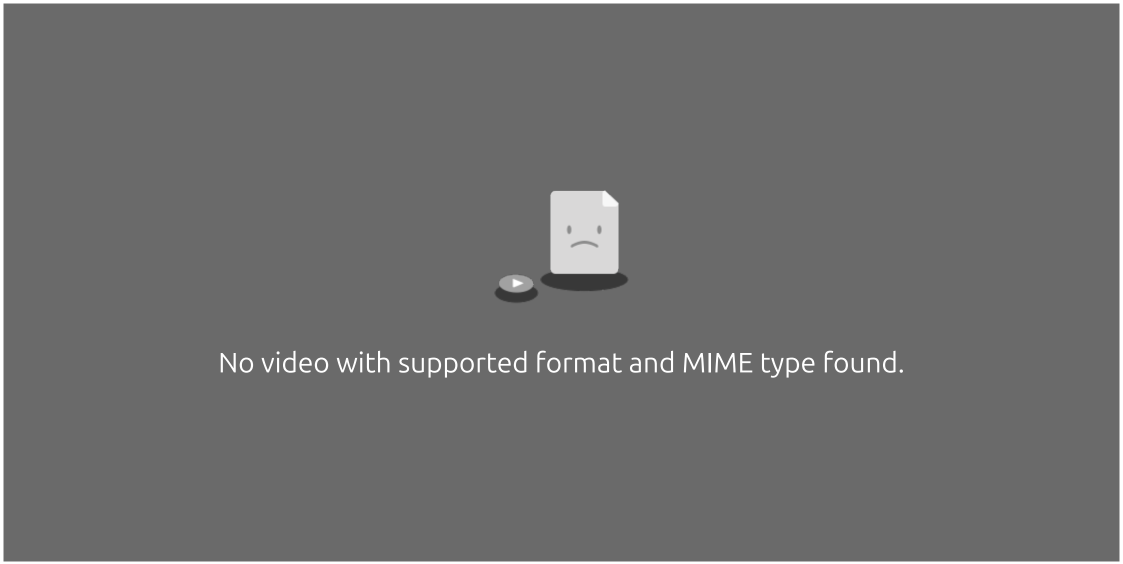
Credits: Andry et al, 2025 (arXiv:2504.18720).



Forecasting of future data $p(x_{L+1:L+M} | y_{1:L})$.

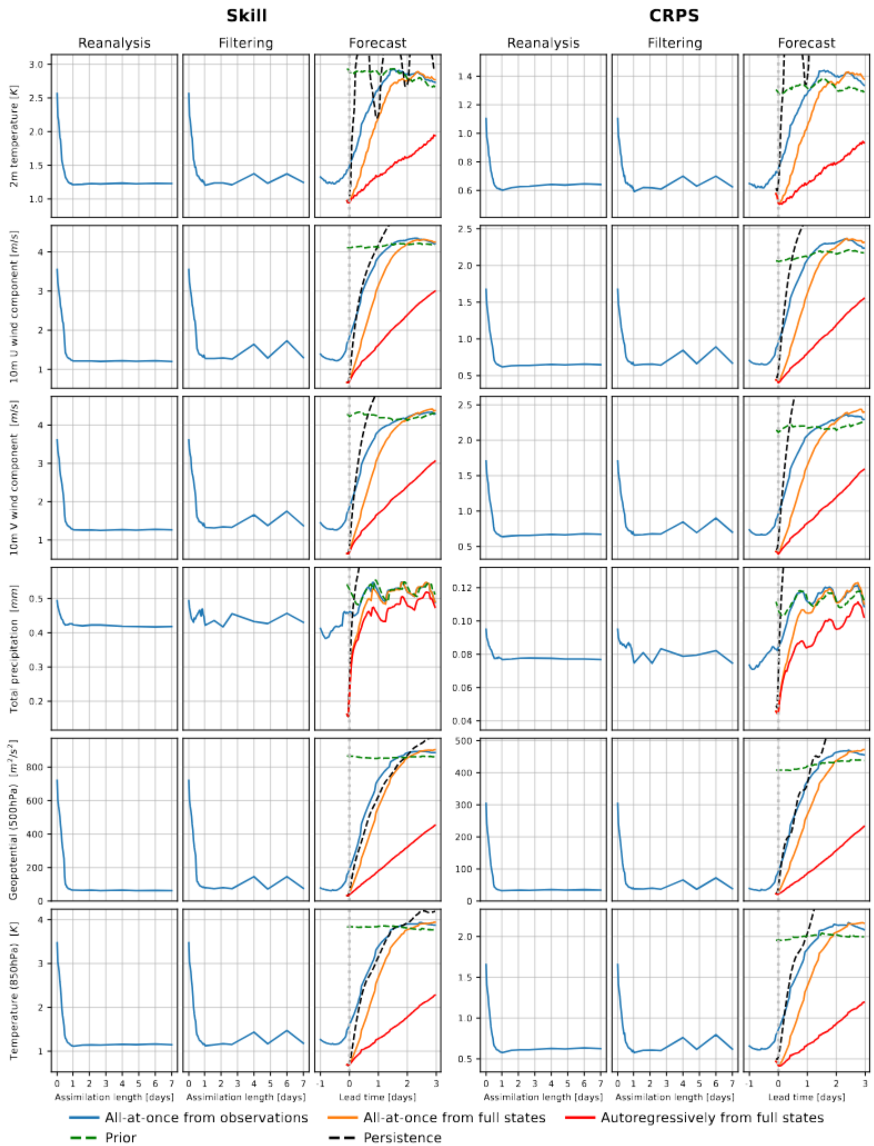
—

Credits: Andry et al, 2025 (arXiv:2504.18720).



Unconditional sampling of (long) trajectories $p(x_{1:L})$.

—
Credits: Andry et al, 2025 (arXiv:2504.18720).



—

Credits: Andry et al, 2025 (arXiv:2504.18720).

Conclusions

Score-based generative models...

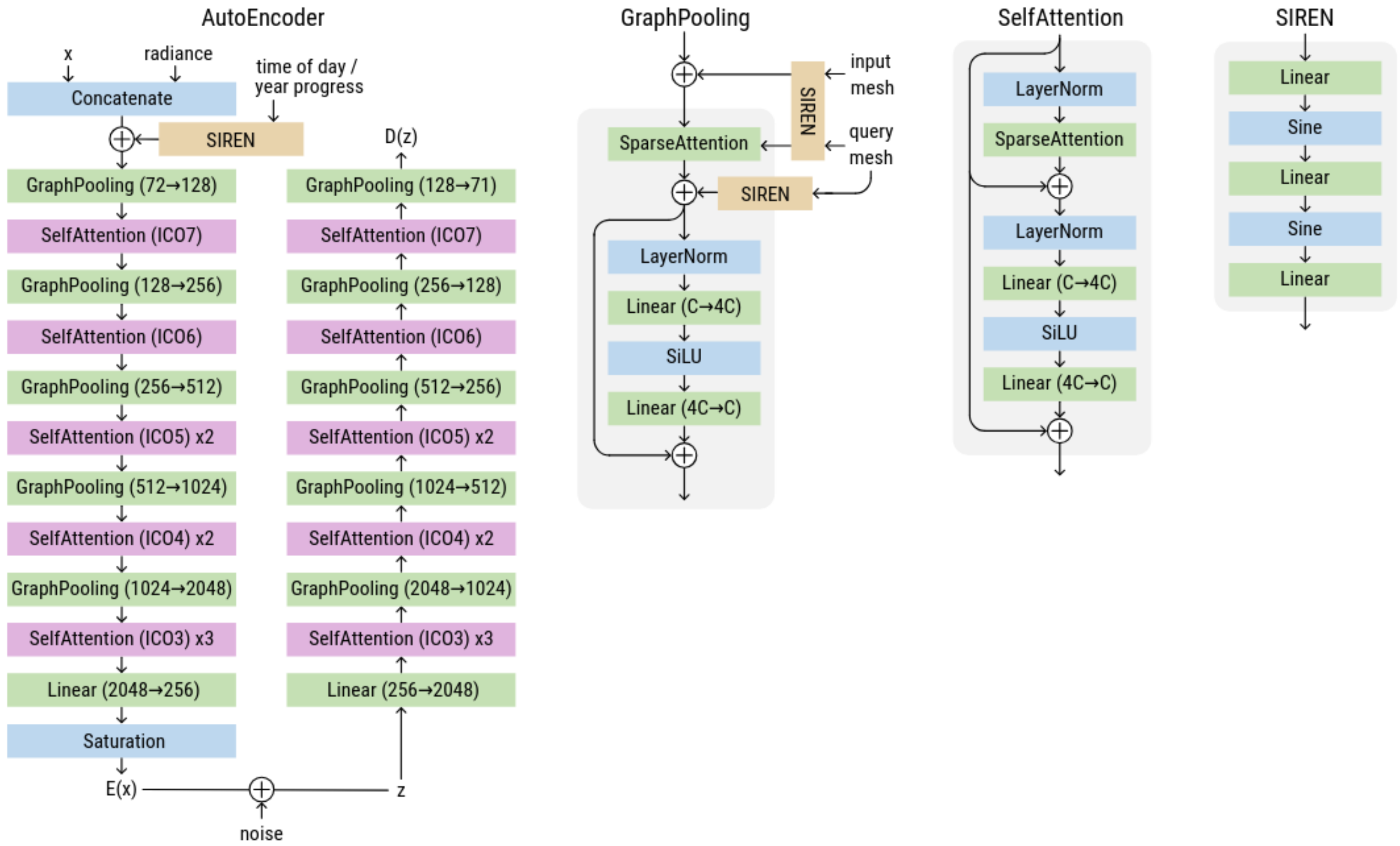
- can be used for high-dimensional inverse problems;
- enable zero-shot posterior sampling, without pre-wiring the network to observations.

Appa...

- is a large weather model that offers a new perspective on data assimilation;
- can be used for reanalysis, filtering and forecasting;
- shows good results for reanalysis and filtering, moderate forecasting skills (for now!).

The end.

Appa: Autoencoder



Appa: Latent denoiser