Scaling up simulation-based inference with diffusion models

GWFreeride workshop
Sexten Center for Astrophysics
January 26, 2026

Gilles Louppe
g.louppe@uliege.be
http://glouppe.github.io

LIÈGE université


From a noisy observation y...

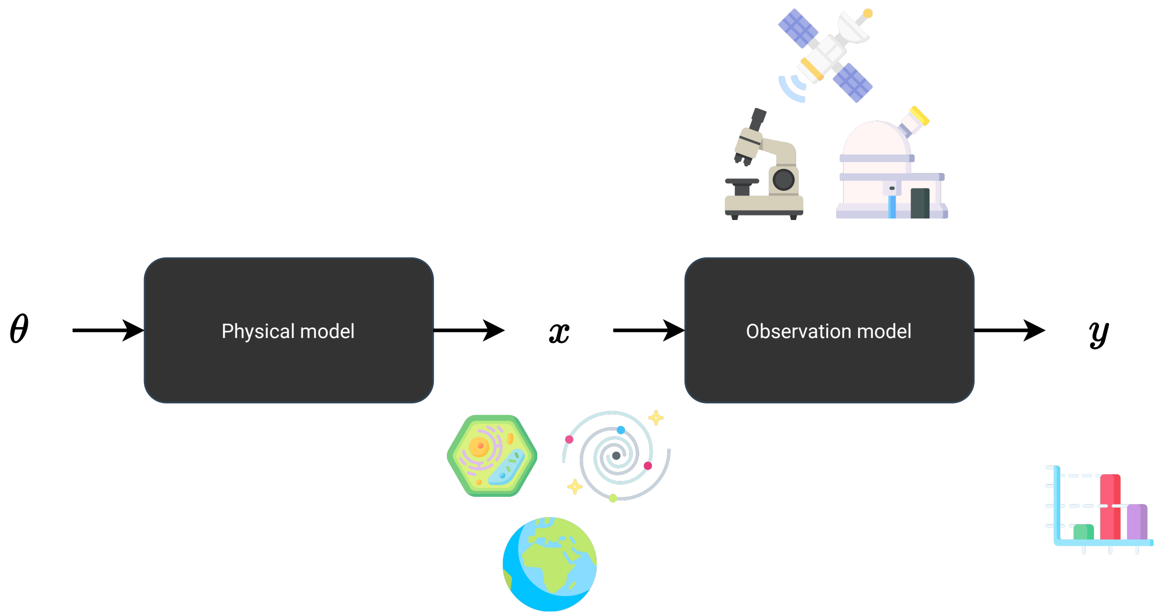

… can we recover
all plausible states x?

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

$$0 = \nabla \cdot u$$

... or model parameters

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



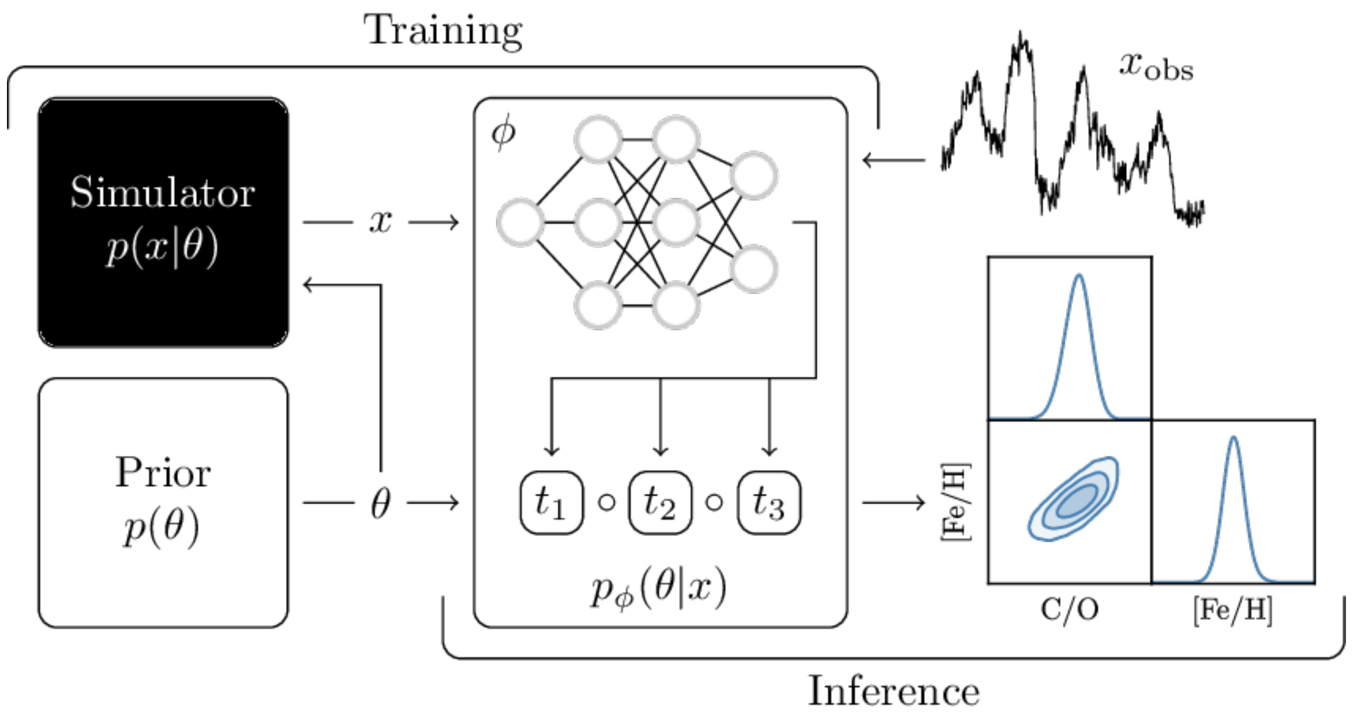
Inverse problems in science

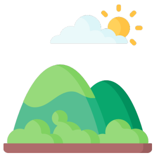
Given noisy observations y, estimate either

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

Neural Posterior Estimation (recap)

Neural posterior estimation (NPE) is a **simulation-based inference** algorithm that learns to approximate the posterior distribution $p(\theta|x)$ (or $p(\theta|y)$) of parameters θ given x (or a noisy observation y).





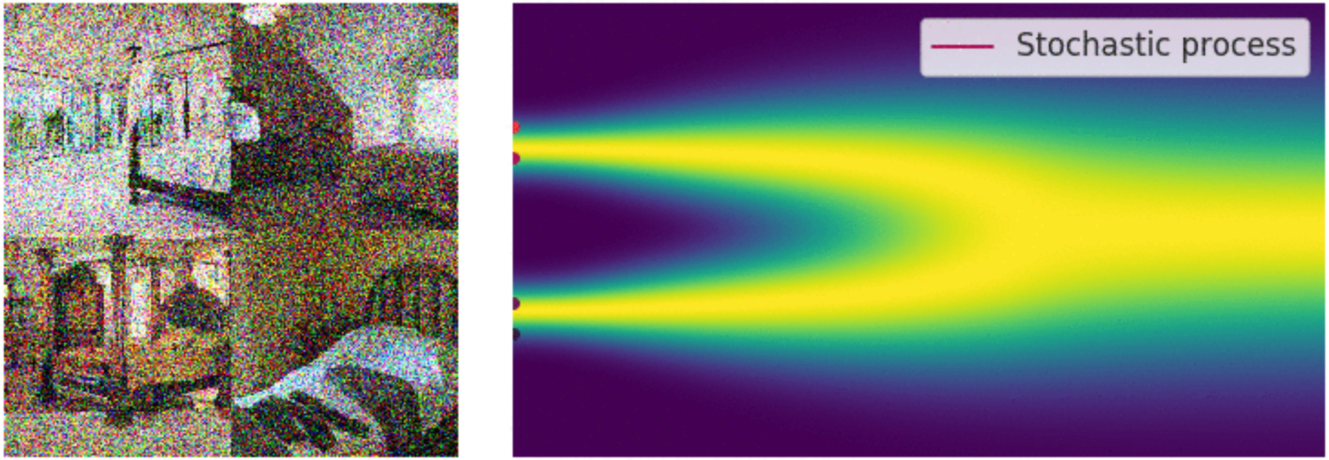
What if the posterior distribution is high-dimensional? ($O(10^3+)$)

Diffusion models 101

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

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

where x_t is the perturbed sample at time t.



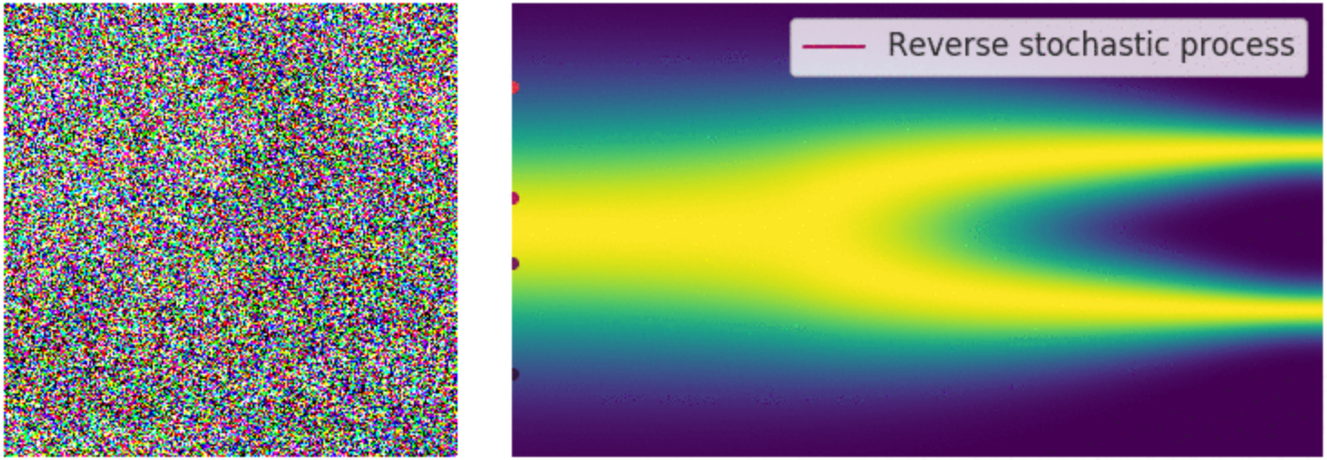
Forward diffusion process.

Credits: Song, 2021.

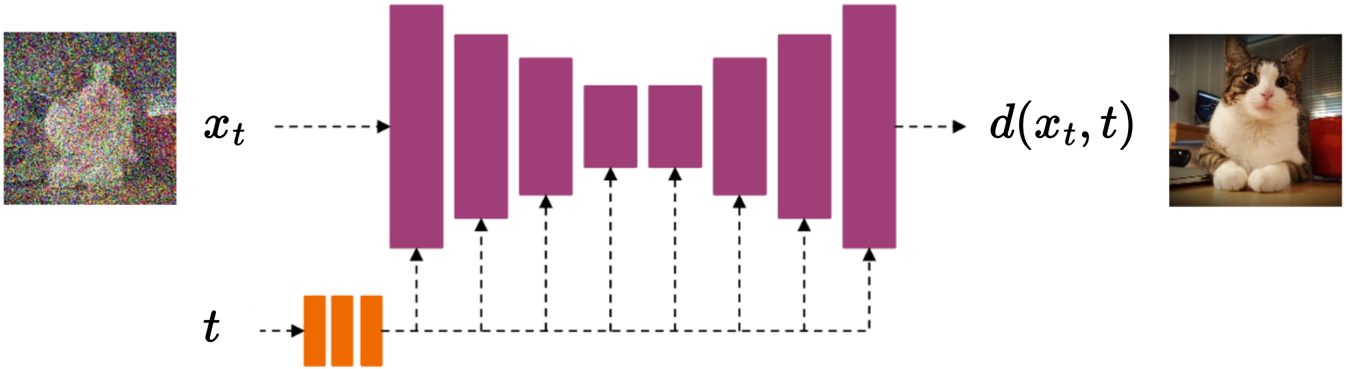
The reverse process satisfies a reverse-time SDE that can be derived analytically from the forward SDE as

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

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, allows to use

$$s_\theta(x_t, t) = \Sigma_t^{-1}(d_\theta(x_t, t) - x_t)$$

as a score estimate of $\nabla_{x_t} \log p(x_t)$ in the reverse SDE.

Inverting single observations

To turn a diffusion model $p_\theta(\mathbf{x})$ into a conditional model $p_\theta(\mathbf{x}|y)$, we can **hard-wire** conditioning information y as an additional input to the denoiser $d_\theta(x_t, t, y)$ and train the model on pairs (x, 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 hard-wire the neural 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

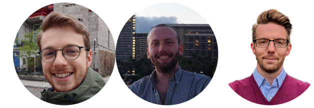
$$p(y|x_t) = \int p(y|x)p(x|x_t)\mathrm{d}x.$$

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

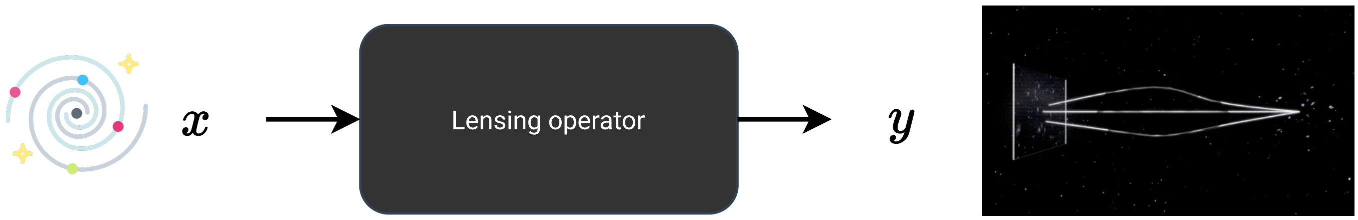
Our approac

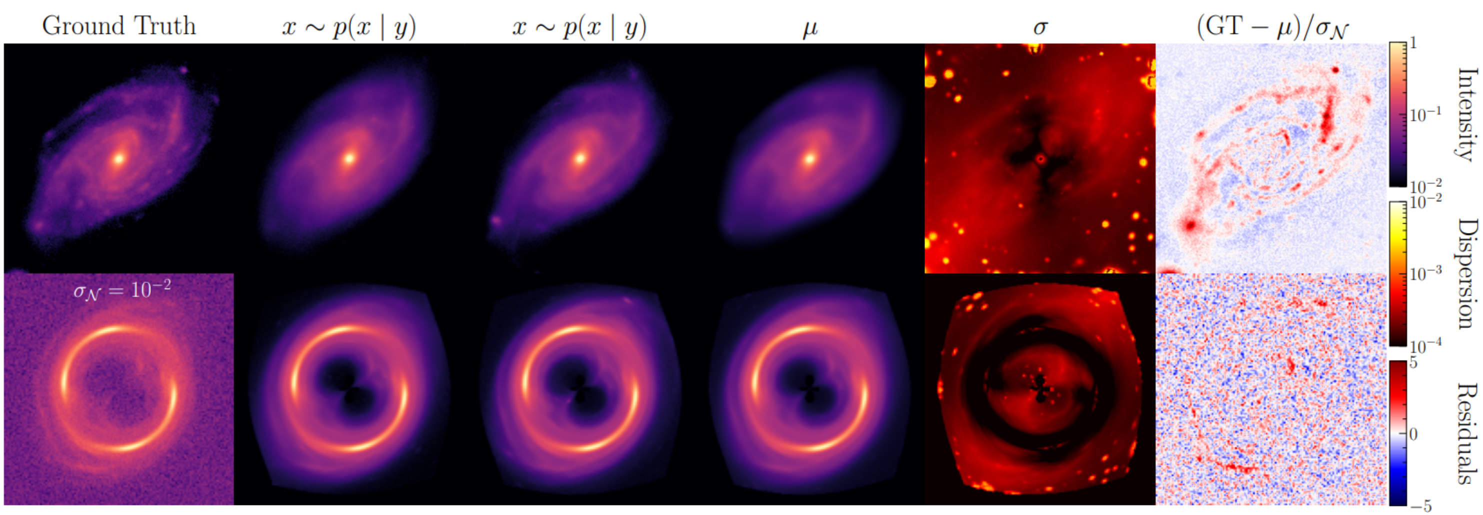
- Assume $, \Sigma_y)$.
- Assume here $\mathbb{E}[x|x_t]$ ng Tweedie
- Then $p($
- The sco

$]).$

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

Inverting gravitational lenses

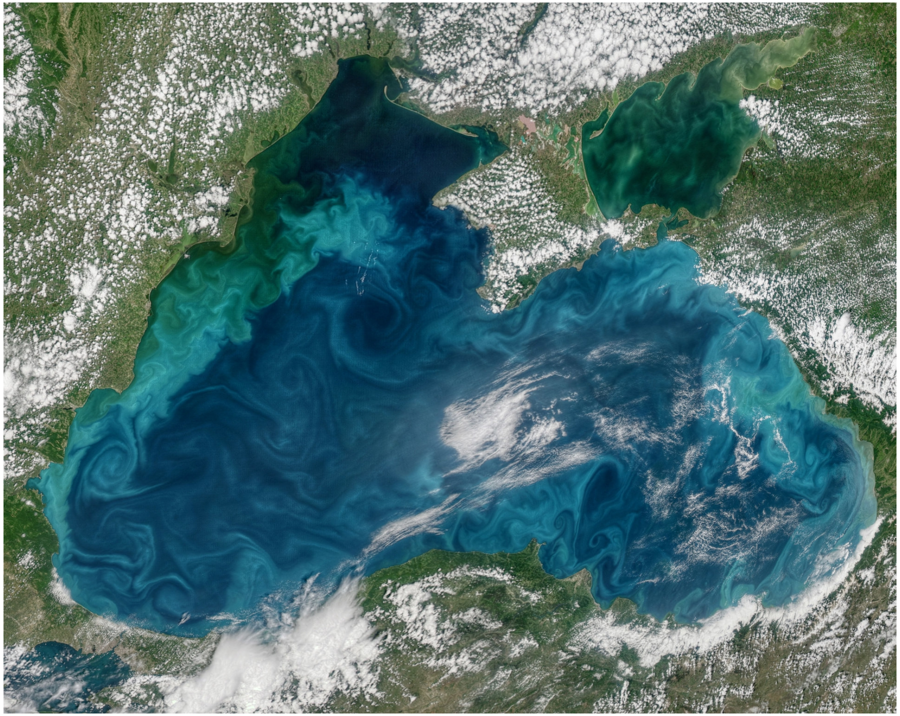


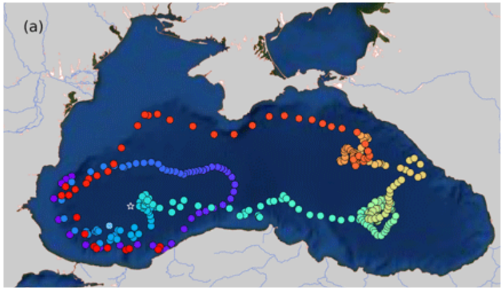


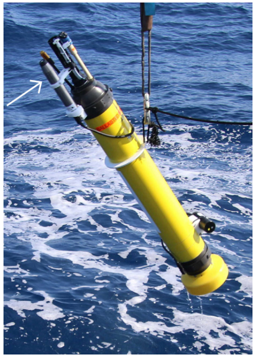
Posterior source galaxies x can be recovered from gravitional lenses y by zero-shot posterior sampling from a diffusion prior $p(x)$ of galaxy images.

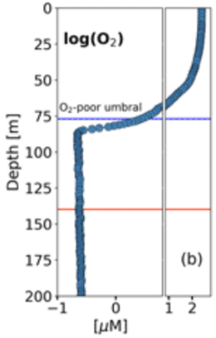
Credits: Adam et al., 2022 (arXiv:2211.03812).

Nowcasting Black Sea hypoxia from satellite observations



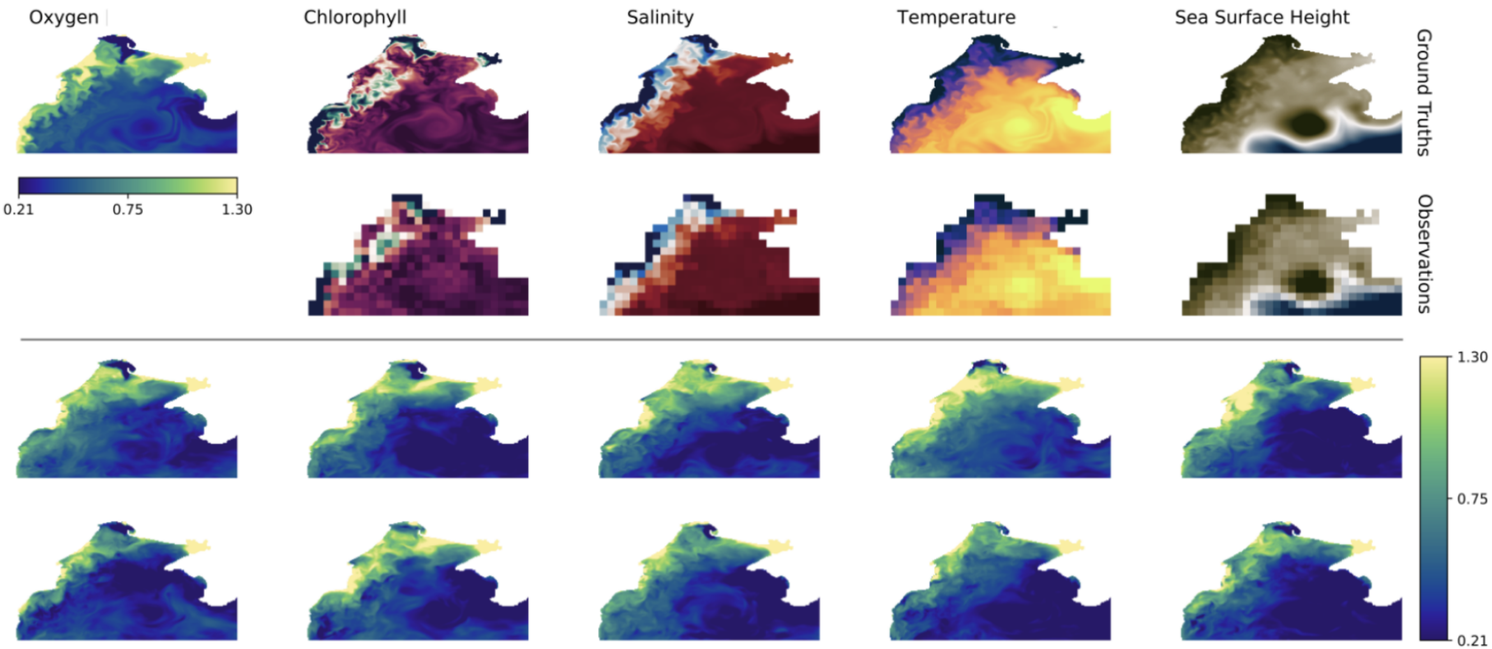






How do hypoxic zones evolve in response to climate change? Can we monitor them from space or with sparse measurements?

Credits: Work in progress with Victor Mangeleer and Marilaure Grégoire.



Posterior oxygen maps $p(x|y)$ can be recovered from satellite observations y of the surface, by zero-shot posterior sampling from a diffusion prior $p(x)$ of the the Black Sea dynamics.

Credits: Work in progress with Victor Mangeleer and Marilaure Grégoire.

Combining multiple observations

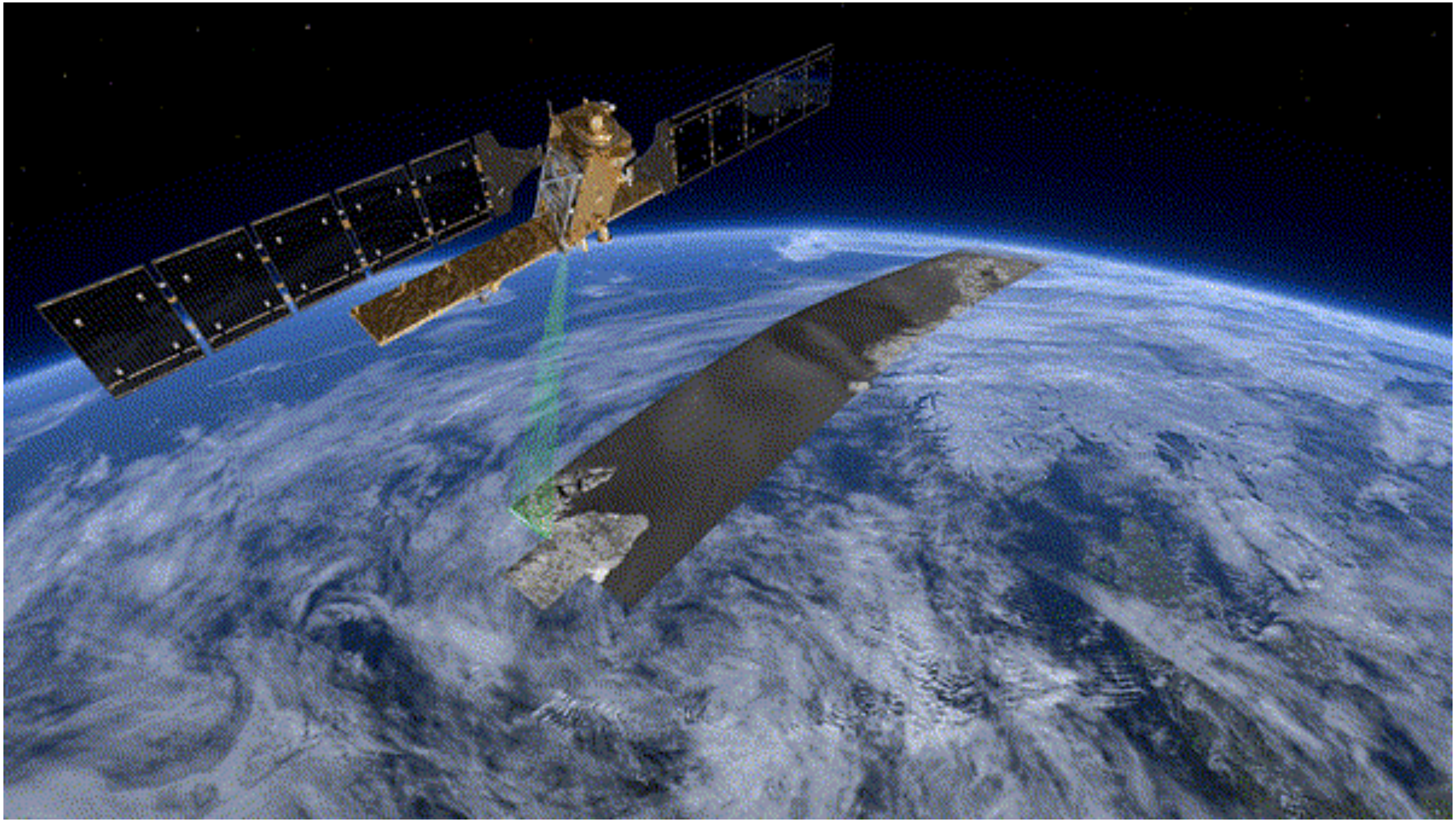
The additivity of scores extends naturally to **multiple observations**:

$$\nabla_{x_t} \log p(x_t | y_1, y_2, ...) = \nabla_{x_t} \log p(x_t) + \sum_i \nabla_{x_t} \log p(y_i | x_t).$$

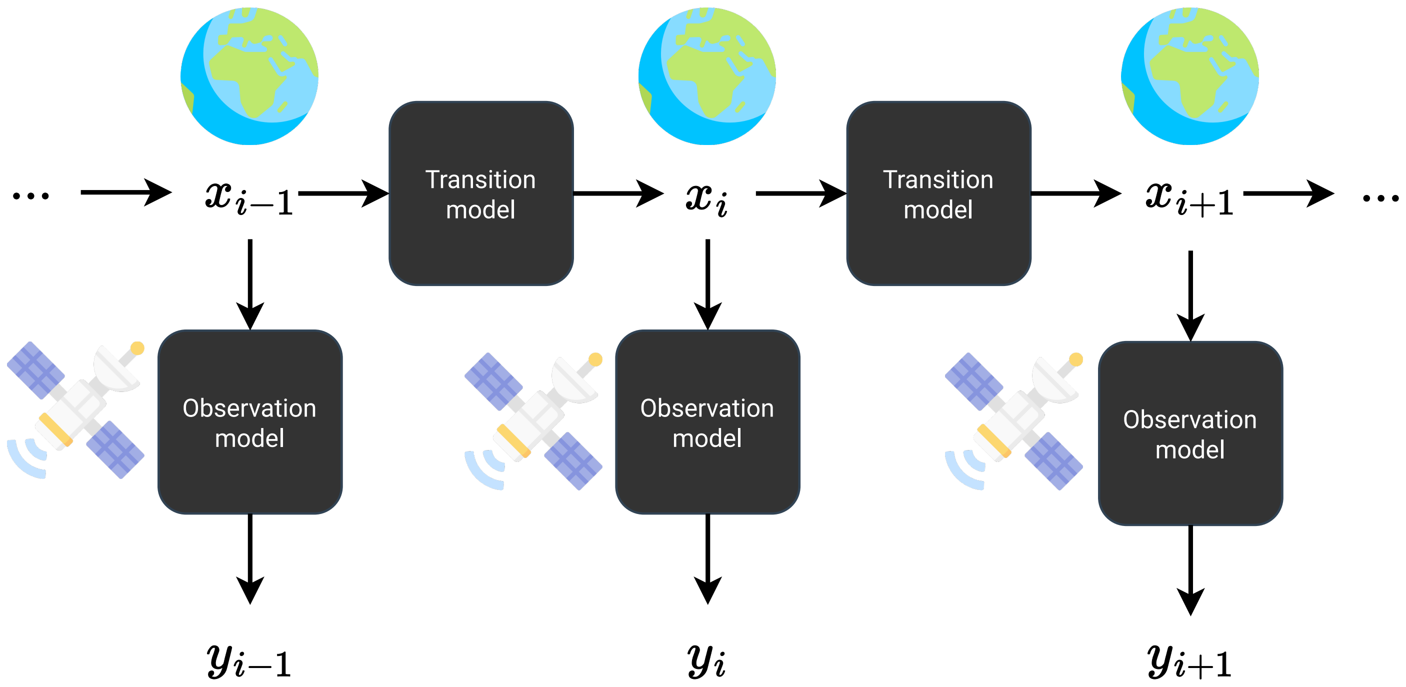
This allows combining:

- Multiple observation modalities (e.g., satellite + in-situ)
- Multiple independent measurements
- Observations at different times
- Members of a population model

What if the posterior distribution is even larger? ($O(10^6+)$)

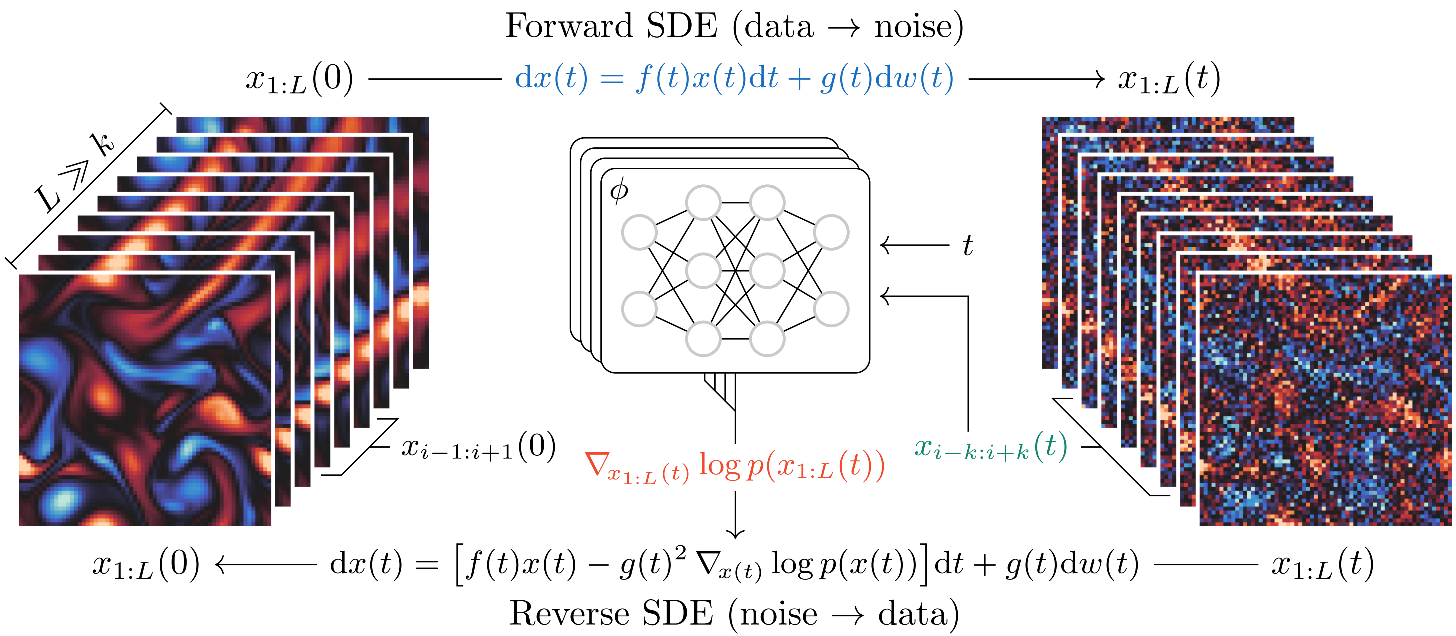


How can we create a comprehensive record of Earth's atmospheric evolution to understand climate change and improve weather prediction?



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

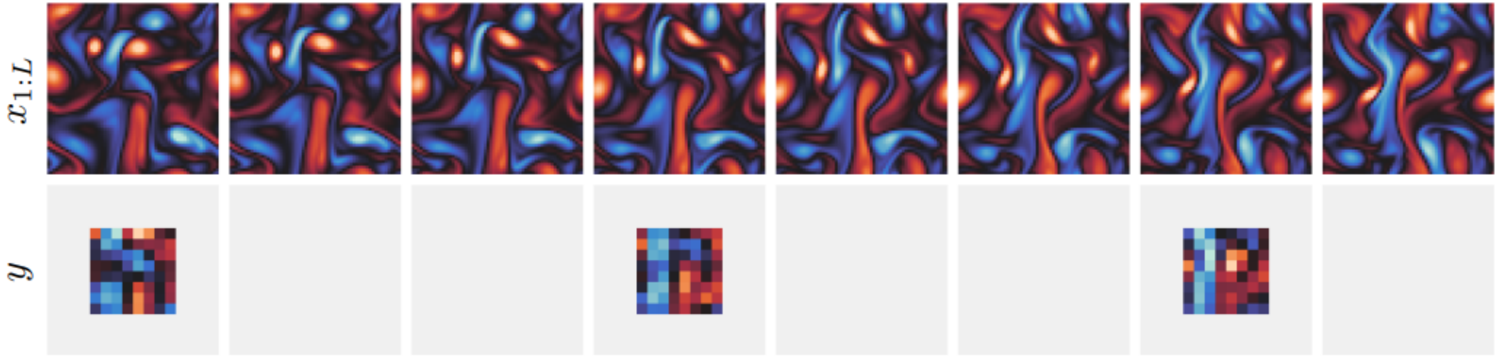
$$p(x_{1:L}|y) = \frac{p(y|x_{1:L})}{p(y)} p(x_0) \prod_{i=1}^{L-1} p(x_{i+1}|x_i).$$



Score-based data assimilation

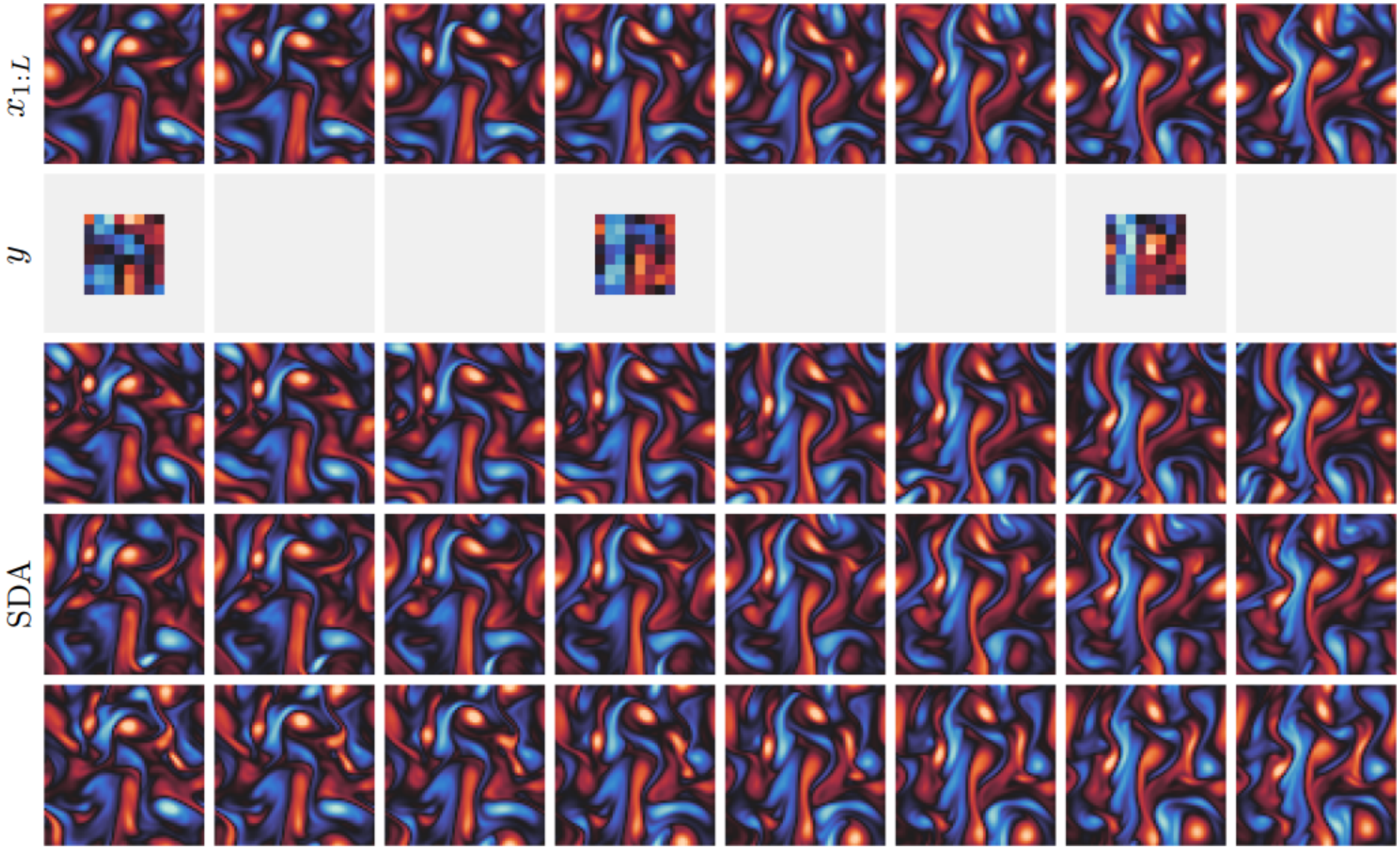
- Build a score-based generative model $p(x_{1:L})$ of arbitrary-length trajectories*.
- Use zero-shot posterior sampling to generate plausible trajectories from noisy observations y.

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.

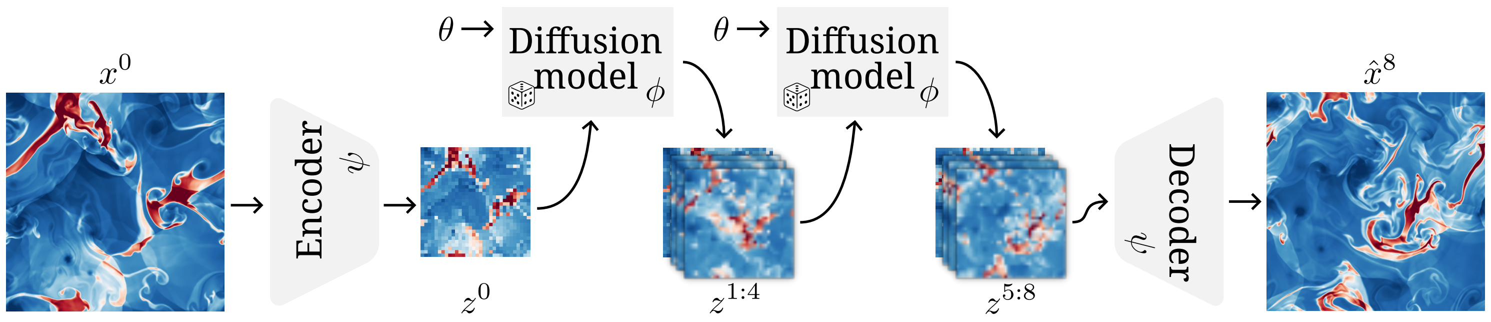
Credits: Rozet and Louppe, 2023 (arXiv:2306.10574).

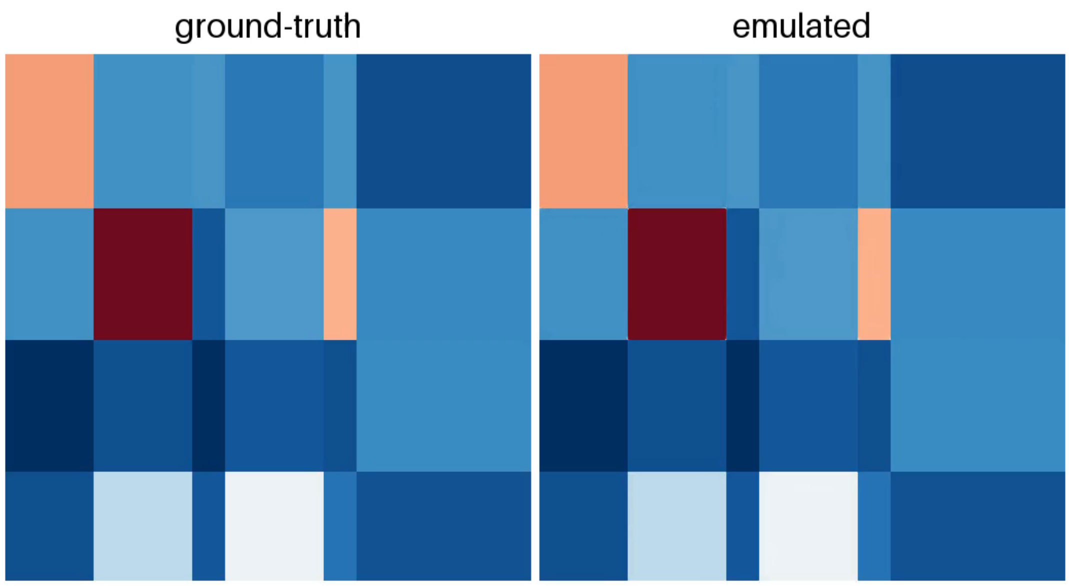
... but does it scale to a whole Earth model?

At 0.25° resolution, for 6 atmospheric variables, 13 pressure levels, hourly time steps, and 14 days of simulation, a trajectory $x_{1:L}$ contains $721 \times 1440 \times 6 \times 13 \times 24 \times 14 = 27 \times 10^9$ variables.

$O(10^9)$ **variables (or more) is needed**
to capture the complexity of the atmosphere.

Latent diffusion models for physics emulation





LDMs trained on compressed latent states $z = E(x)$ remain accurate even at high compression rates.

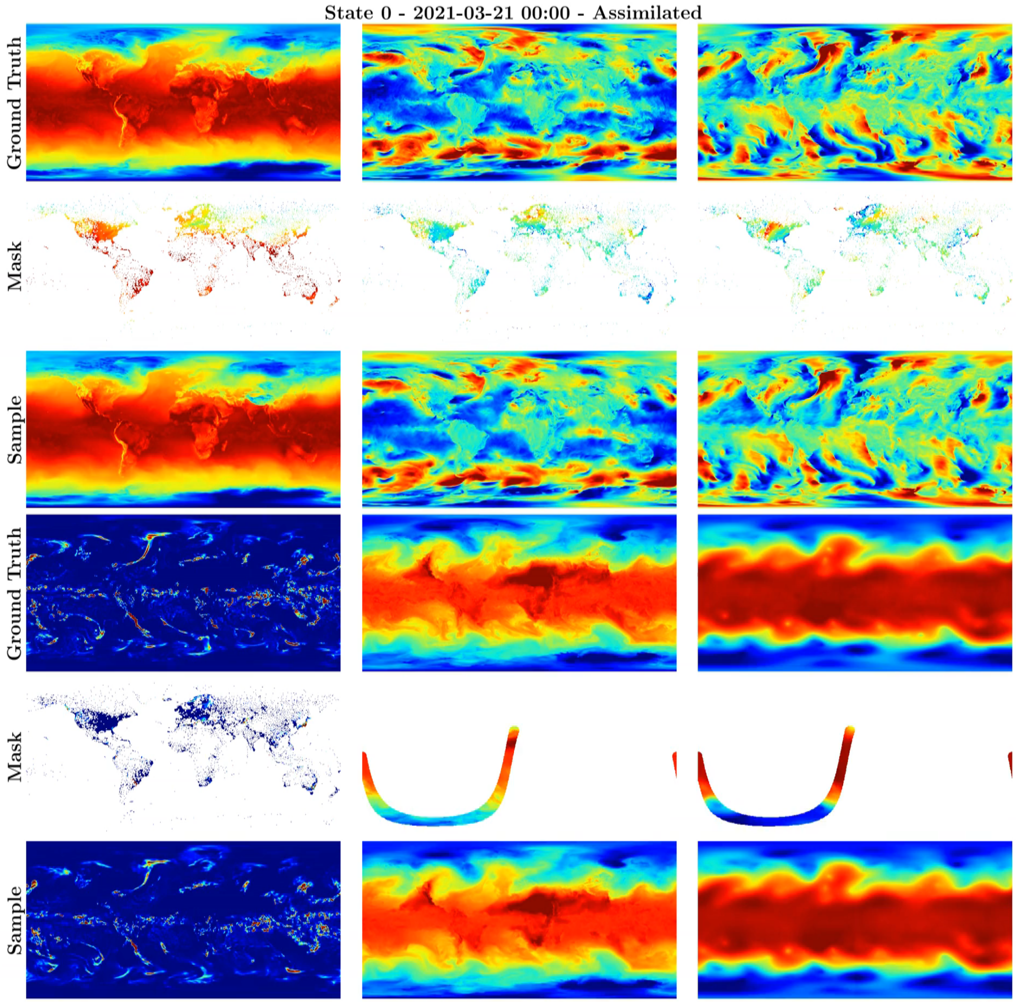
Credits: Rozet et al, 2025 (arXiv:2507.02608).

Appa: Bending weather dynamics with LDMs

Appa is made of three components:

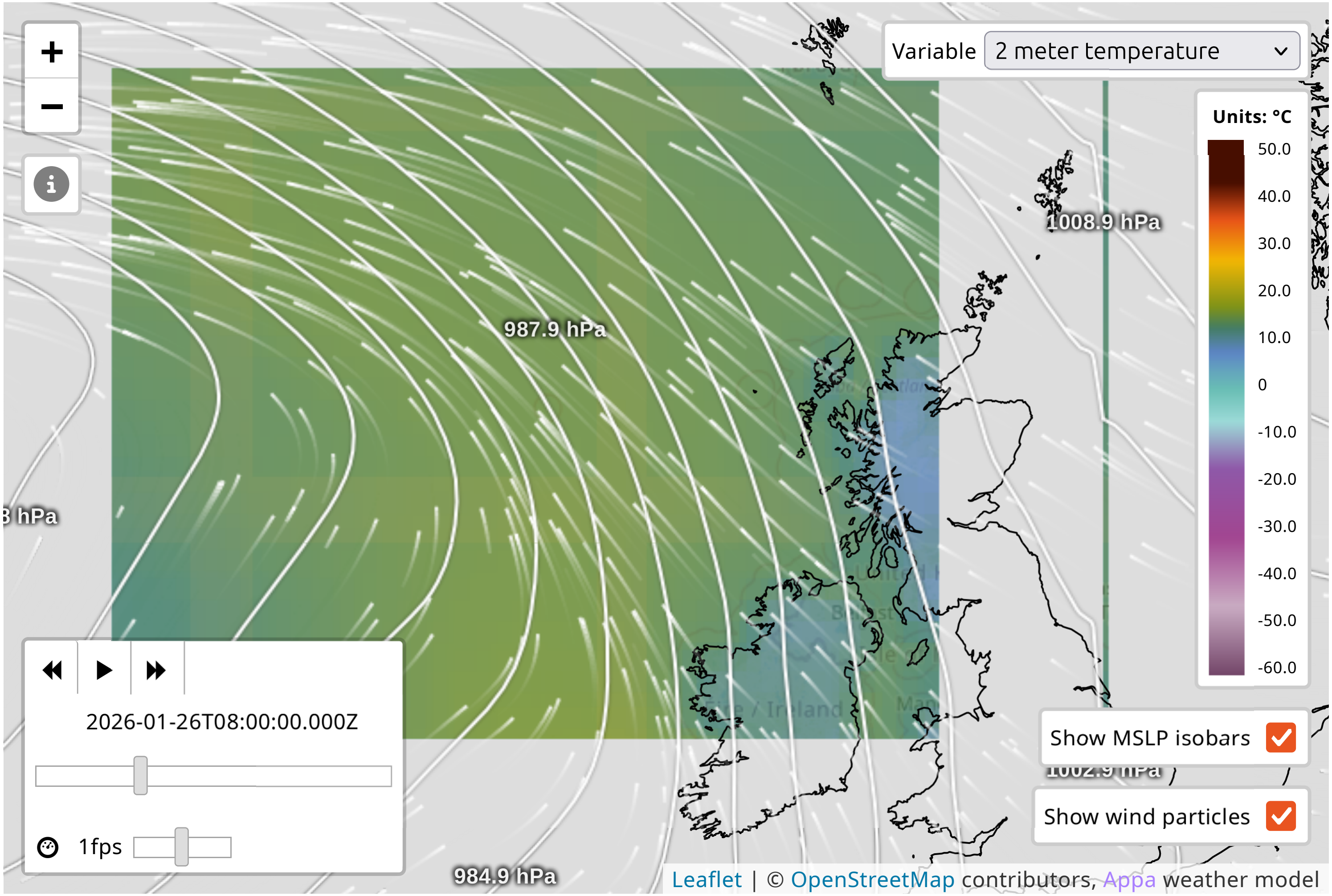
- a 500M-parameter **autoencoder** that compresses the data space x into a latent space z with a 450x compression factor;
- a 1B-parameter **latent diffusion model** that generates latent trajectories $z_{1:L}$;
- a **posterior sampling algorithm** adapted from MMPS (Rozet et al, 2024) that samples from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

Credits: Andry et al, 2025 (arXiv:2504.18720). New version soon on arXiv!



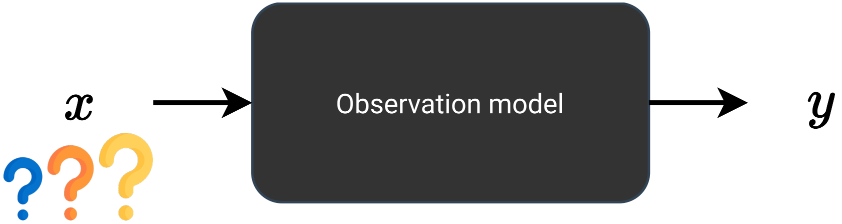
Reanalysis of past data $p(x_{1:L} | y_{1:L})$.

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



Live demo at appa.montefiore.uliege.be.

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



Learning priors from noisy observations

Assume only observations $y \sim p(y)$ and a known observation model $p(y|x)$.

The objective of **Empirical Bayes** is find a prior model $q_\theta(x)$ such that

$$q_\theta(y) = \int p(y|x) q_\theta(x) \mathrm{d}x$$

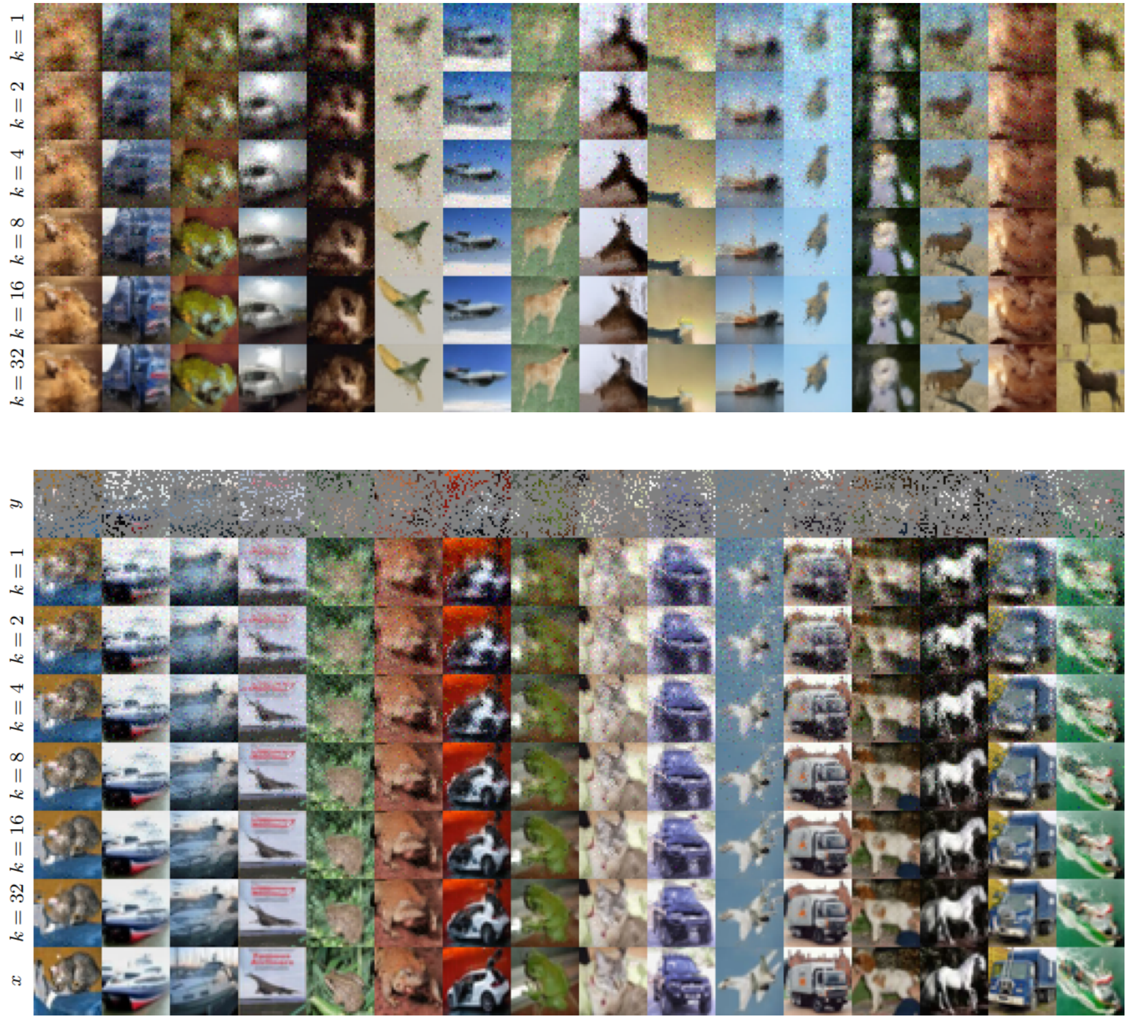
is closest to $p(y)$.

Our approach:

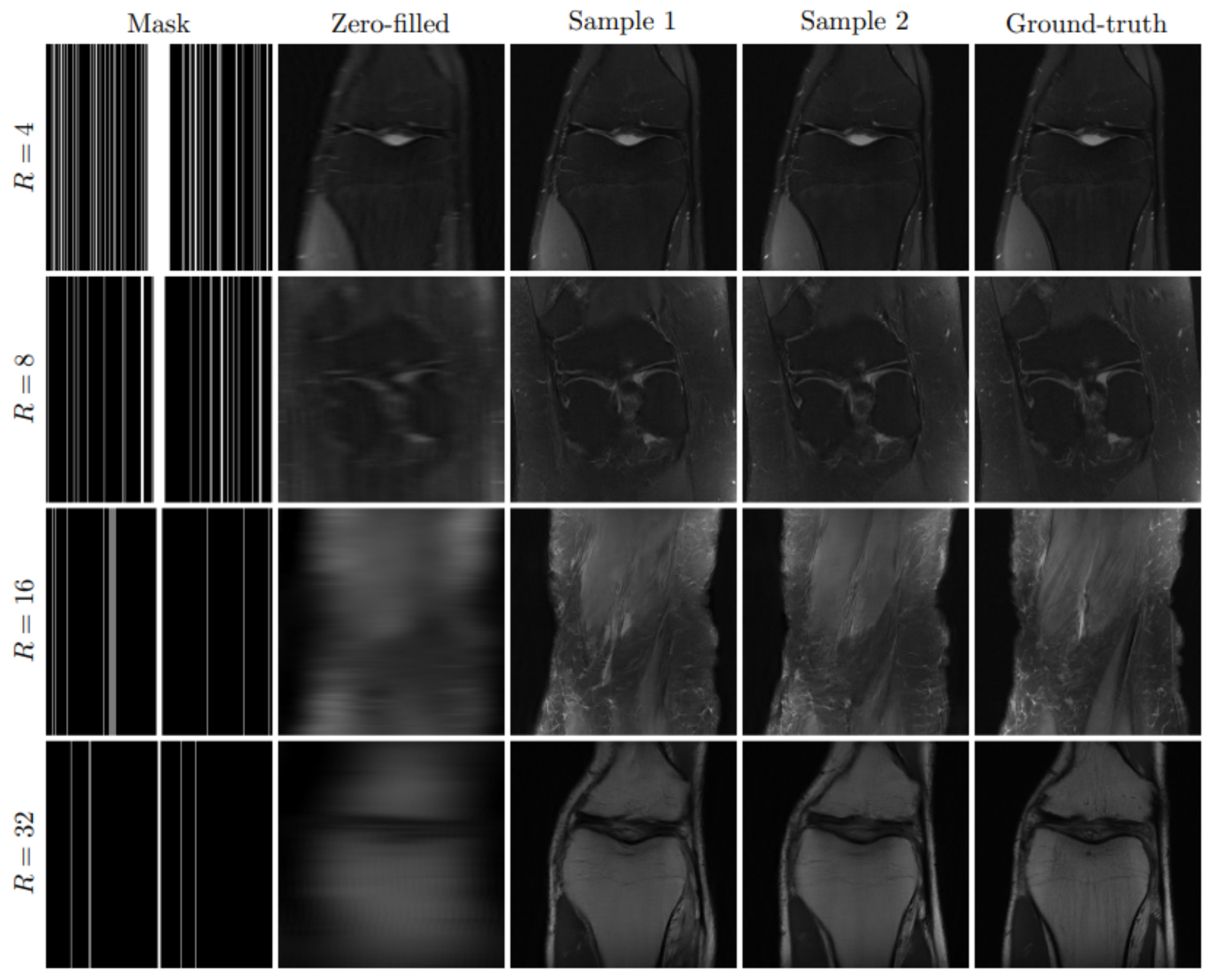
- If we parameterize the latent state x with a diffusion prior $q_\theta(x)$, then Expectation-Maximization can be used to maximize $q_\theta(y)$.
- It can be shown that the EM update

$$\theta_{k+1} = \arg\max_\theta \mathbb{E}_{p(y)} \mathbb{E}_{q_{\theta_k}(x|y)} \left[\log q_\theta(x) \right],$$

where $q_{\theta_k}(x|y)$ is obtained by posterior sampling from $q_{\theta_k}(x)$, leads to a sequence of parameters θ_k such that $\mathbb{E}_{p(y)} \left[\log q_{\theta_k}(y) \right]$ is monotonically increasing and converges to a local optimum.



Samples from the prior $q_{\theta_k}(x)$ (top) and the posterior $q_{\theta_k}(x|y)$ (bottom) along the EM iterations when training from corrupted CIFAR-10 images.



Posterior samples for accelerated MRI using a diffusion prior trained only from observations with subsampled frequencies.

Conclusions

Deep generative models unlock high-dimensional Bayesian inference in complex physical models: **new scientific questions become accessible**.

Next challenges:

- Rigorous validation: when and why these methods work (or not).
- Resolution (in space and time): can we go higher?
- Misspecification: what if the prior, the physical, or the observation models are wrong?

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

The end.