Appa: Bending Weather Dynamics with Latent Diffusion Models for Global Data Assimilation

Gérôme Andry, François Rozet, Sacha Lewin, Victor Mangeleer, Omer Rochman, Matthias Pirlet, Elise Faulx, Marilaure Grégoire and Gilles Louppe

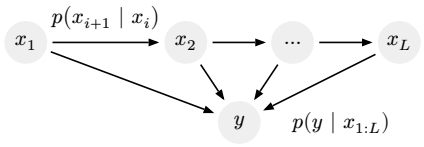
TL;DR Score-based data assimilation models can produce global atmospheric trajectories at 0.25-degree resolution and 1-hour intervals. Using a spatio-temporal latent diffusion model trained on ERA5 reanalysis data, it can be conditioned on any types of observations to infer the posterior distribution of plausible state trajectories, without retraining.

Problem statement

Data assimilation (DA) addresses the problem of inferring the posterior distribution

$$p(x_{1:L} \mid y) = \frac{p(y \mid x_{1:L})}{p(y)} \underbrace{p(x_1) \prod_{i=1}^{L-1} p(x_{i+1} \mid x_i)}_{\text{Markovian prior}}$$

for dynamical systems (atmospheres, oceans, ...) given noisy or incomplete observations.



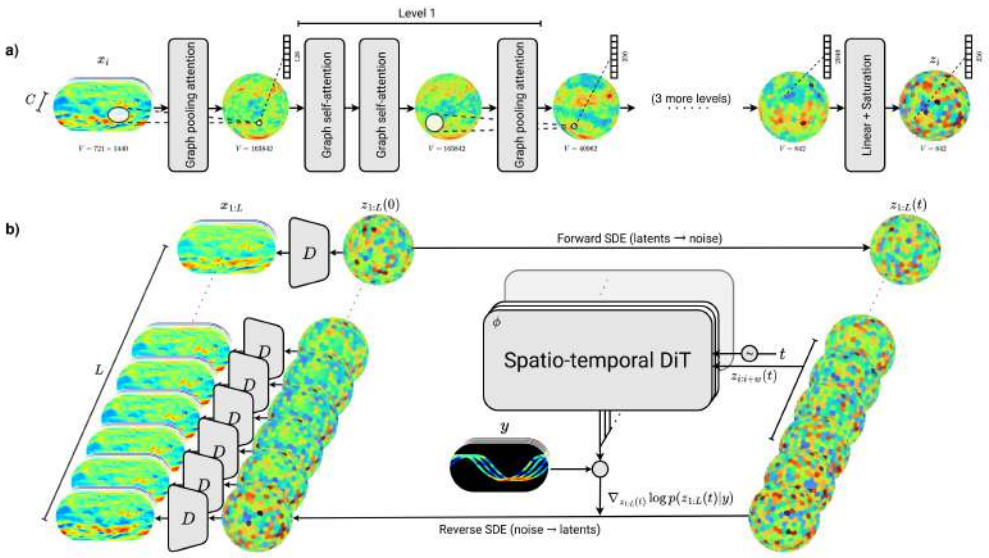


Figure 2. Overview of Appa's architecture. a) The autoencoder maps high-dimensional atmospheric states to a compact latent representation. b) The latent diffusion process operates entirely in the compressed space, where latent trajectories are perturbed via a forward SDE to create noisy trajectories. Any type of observations can be incorporated in the reverse diffusion process to produce a posteriori samples without additional training.

Auto-encoder results

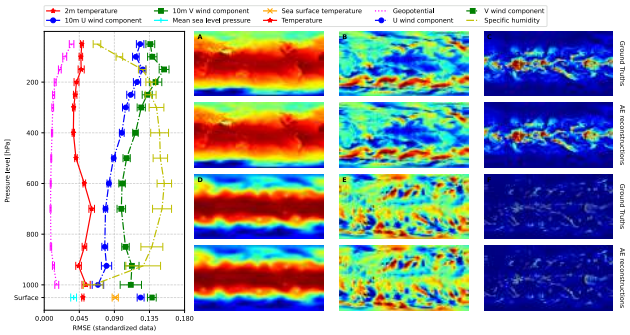


Figure 3. Root mean square error (RMSE) for standardized reconstructions of surface and atmospheric variables across pressure levels. Lower values indicate better reconstruction quality.

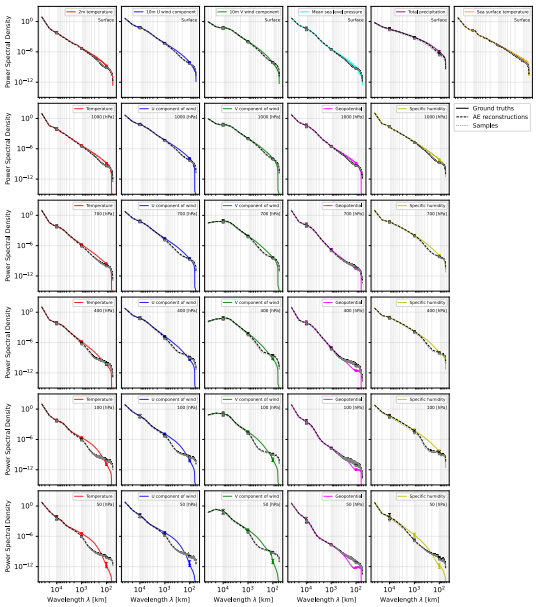


Figure 4. Power spectral density of ground truth, autoencoder reconstructions, and unconditional diffusion samples across wavelengths for surface variables (top row) and atmospheric variables at selected pressure levels (lower rows).

Assimilation results

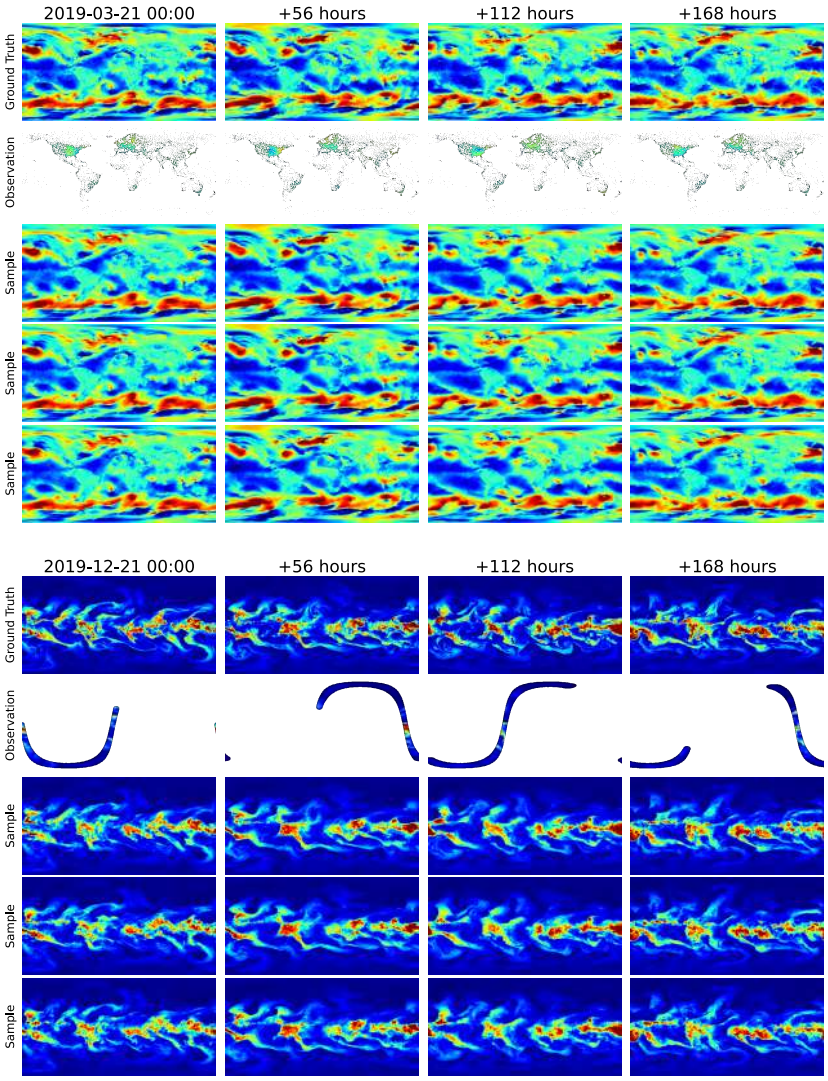


Figure 6. Reanalysis samples for 10m U component of wind (top) and specific humidity at 300 hPa (bottom) with an assimilation window of 1 week at a 1-hour resolution.