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

Gérôme Andry, Sacha Lewin, François Rozet, Omer Rochman, Victor Mangeleer, 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

Operational weather prediction relies on estimating the current atmospheric state using observational data. Once observations are assimilated, key downstream tasks follow such as

$$\textbf{Reanalysis: } p(x^{1:L} \mid y^{1:L})$$
$$\textbf{Filtering: } p(x^{L} \mid y^{1:L})$$
$$\textbf{Forecasting: } p(x^{K+1:L} \mid x^{1:K}).$$

Traditionally, data assimilation and simulation are treated as separate steps. For example, forecasts depend on a prior full-state estimate used as the initial condition.

Pipeline

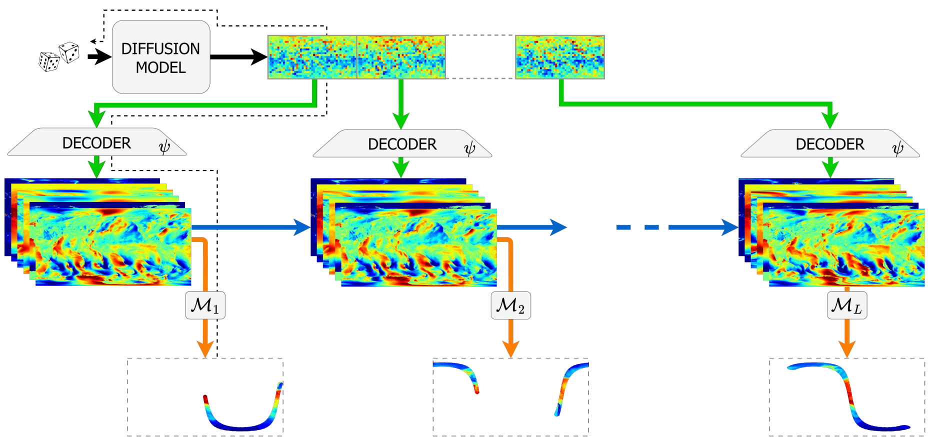

Figure 1. After the training of an autoencoder, we encode the high-resolution dataset into the latent space and train a diffusion model to generate windows of W consecutive states. At inference, the dynamics $p(x^{i+1}|x^{i})$ of the system are approximated by the combination of the diffusion model and the **decoder**. Using modern posterior sampling methods for diffusion, we can incorporate any observation process $p(y^{i}|x^{i})$ without retraining.

Assimilation gallery

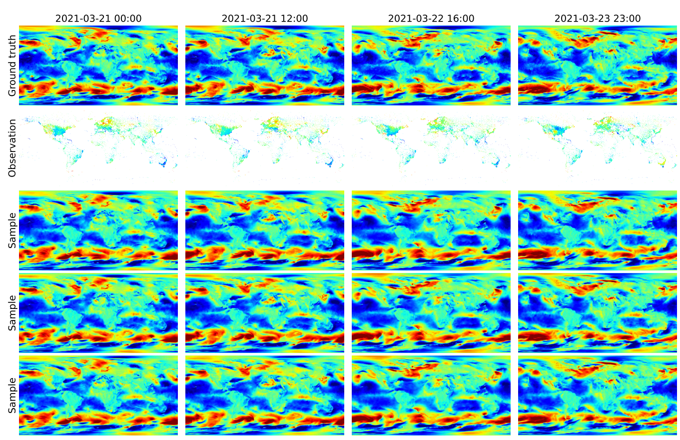

Figure 2. Reanalysis samples of U wind component at 10m along with simulated station observations used for assimilation.

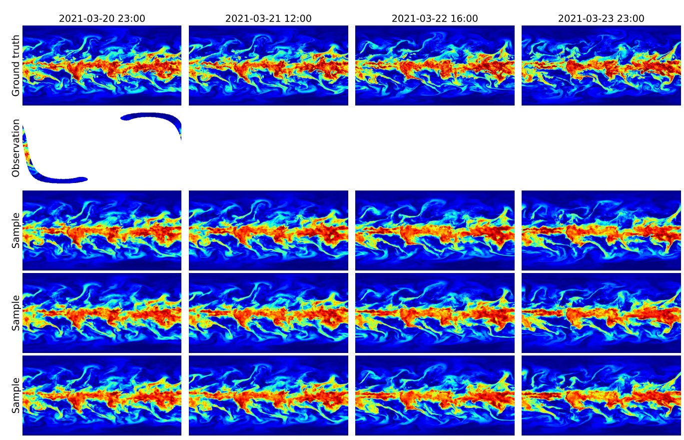

Figure 3. Forecasting samples of specific humidity at 700 hPa along with simulated satellite scan used to initialize the forecast.

Autoencoder results

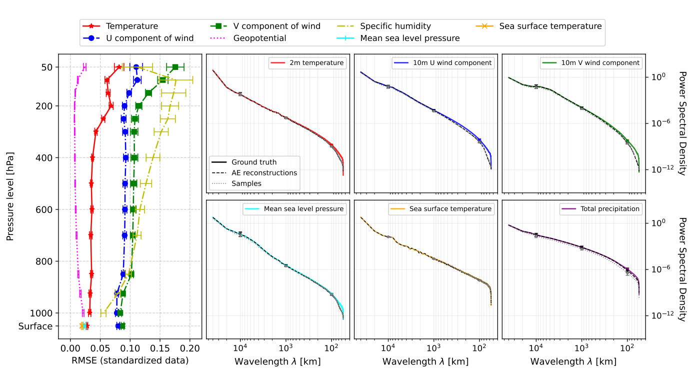

Figure 4. Root mean square error (RMSE) for standardized reconstructions of surface and atmospheric variables across pressure levels (left). Surface variable spectra comparison between ERA5, reconstructions and generated samples (right).

Assimilation metrics

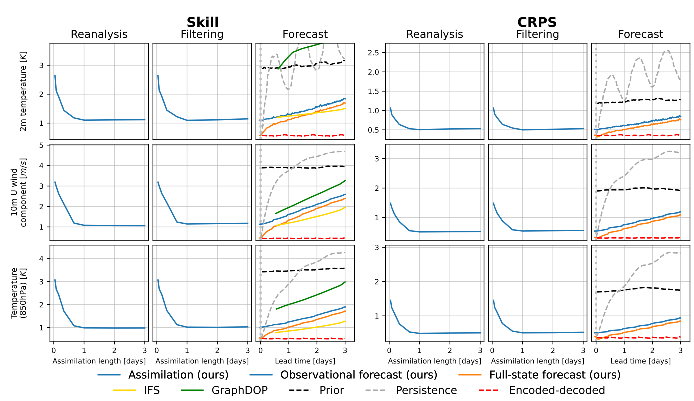

Figure 5. Skill and CRPS for different assimilation tasks in January 2023. Metrics are computed over 10-member ensemble for 8 different dates.