PROGRAMME OF THE EUROPEAN UNION

IMPLEMENTED BY ECMWF

Uncertainty Quantification Using Machine Learning Diffusion Model for CARRA2 Reanalysis

Swapan Mallick [a], Jelena Bojarova [a], Kasper Tølløse [b], Xiaohua Yang [b], Per Dahlgren [c], Eivind Støylen [c], & Harald Schyberg [c]

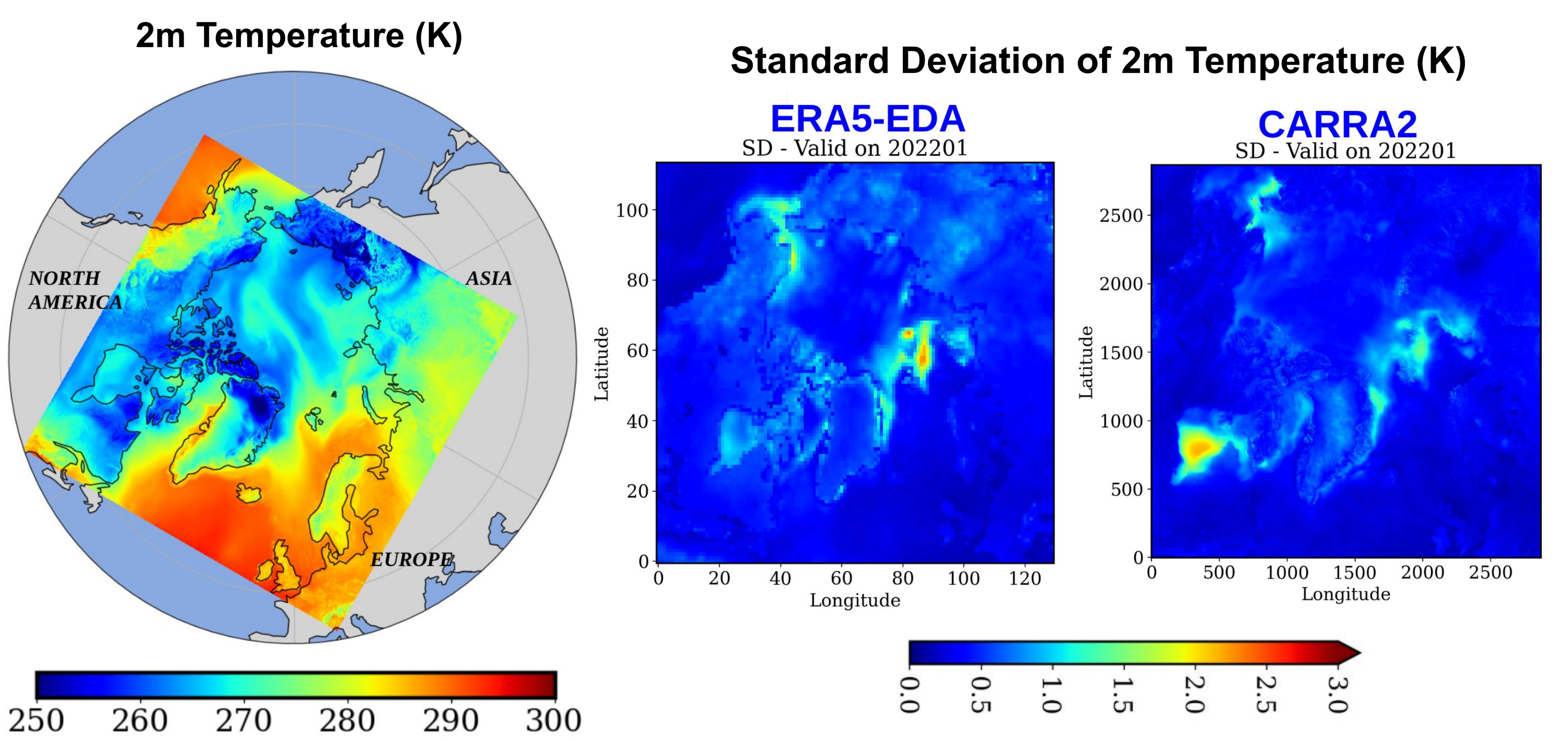
[a] Swedish Meteorology and Hydrology (SMHI), [b] Danish Meteorological Institute (DMI), [c] Norwegian Meteorological Institute (Met-Nor)

Email: Swapan.Mallick@smhi.se

Climate Change Service
climate.copernicus.eu

Introduction

The Copernicus Arctic Regional Reanalysis Second Generation (**CARRA2**) is a high-resolution climate dataset that spans the entire **pan-Arctic** area. Accurately quantifying uncertainty is crucial for enhancing the precision and dependability of weather and climate forecasting systems. This research investigates advanced uncertainty quantification methods using a machine learning-based Denoising Diffusion Probabilistic Model (**DDPM**). A key goal of the **CARRA2** project is to assess **ensemble uncertainties** in important climate variables. To accomplish this, DDPM is applied to estimate uncertainties between the high-resolution **CARRA2** data and the coarser **ERA5-EDA** datasets. The **CARRA2** simulations are characterized by the following: Model – **Harmonie-AROME** Ensemble, **CY46**; Data Assimilation – **3D-VAR**; Horizontal Resolution – **2.5 km**; Grid Points – **2880 x 2880**; Vertical Levels – **65**; Initial and Boundary Conditions – **ERA5-EDA**.



Diffusion Model (DL ⊂ ML ⊂ AI)

The Denoising Diffusion Probabilistic Model (DDPM), introduced in 2020, is a fundamental diffusion model that frames both the forward (adding noise) and reverse (removing noise) processes as Markov chains. It uses a fixed, linear noise schedule and requires a large number of steps to generate high-quality outputs. In the forward process, Gaussian noise is gradually added over T steps. During the reverse process, a U-Net is trained to predict and eliminate noise step by step. In our experiments, all T set to 4000 steps for a single image. The benefits of these methods include their simplicity and solid theoretical basis. They produce high-quality samples that can often outperform Generative Adversarial Networks (GANs) in some scenarios.

Mathematical expressions are as follows:

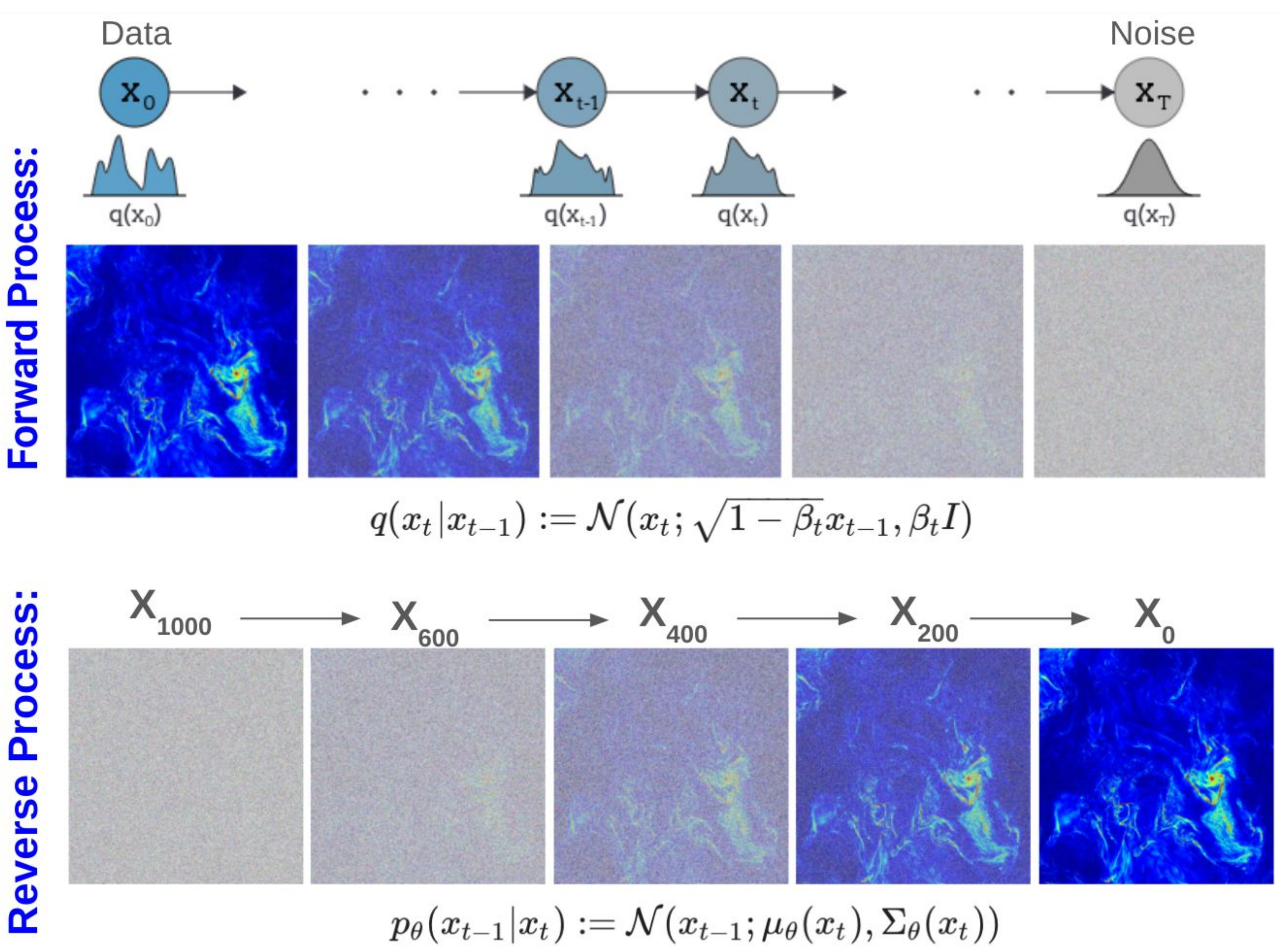
1. Forward Process:

$$q(x_t \mid x_{t-1}) = N\left(x_t ;\ \sqrt{1-\beta_t}\, x_{t-1},\ \beta_t I\right)$$

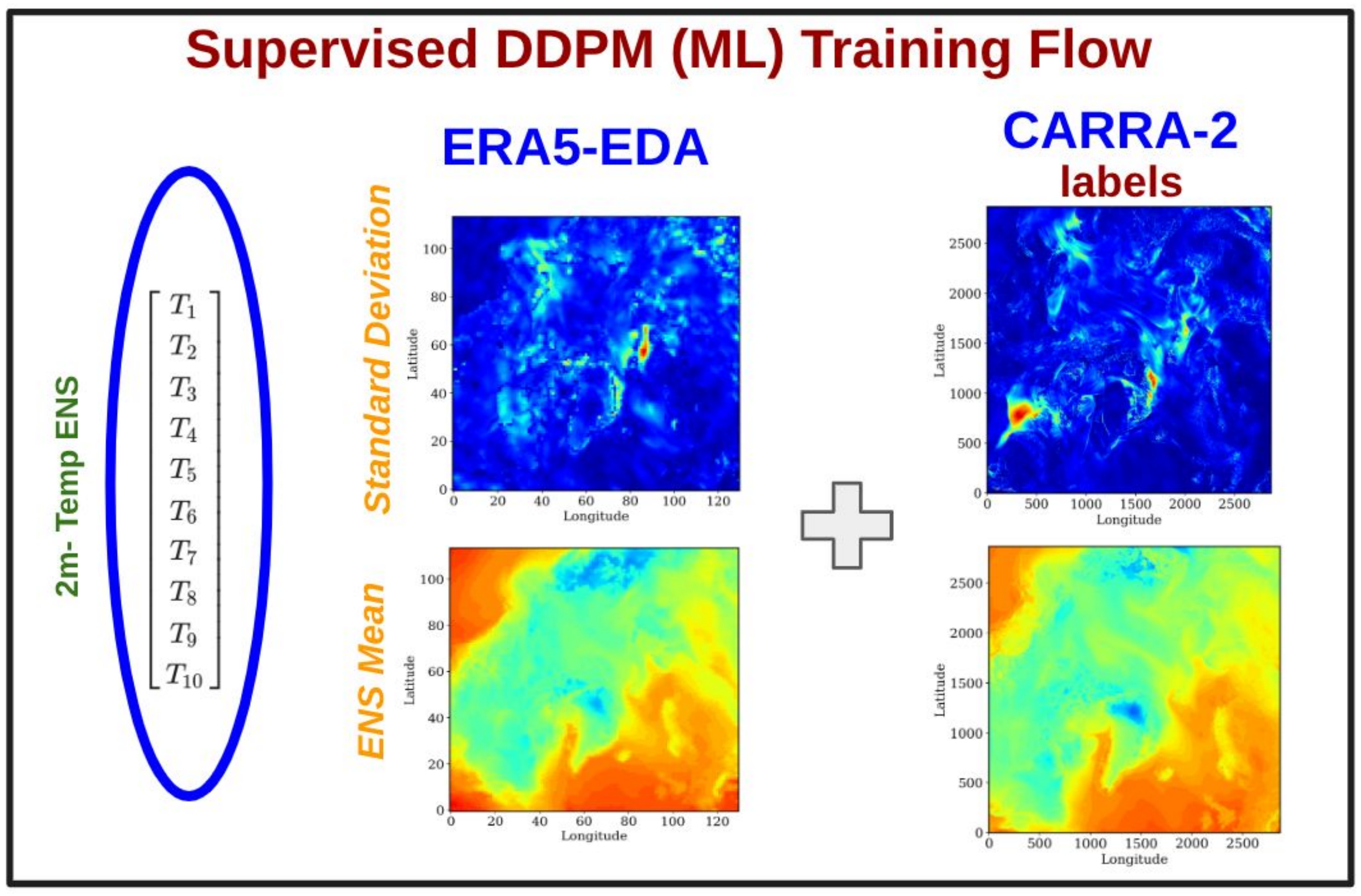
2. Reverse process:

$$p_\theta(x_{t-1} \mid x_t) = N\left(x_{t-1} ;\ \mu_\theta(x_t, t),\ \Sigma_\theta(x_t, t)\ \sqrt{1-\beta_t}\, x_{t-1}\right)$$

The neural network θ learns to predict the mean μ_θ and Σ_θ to reverse the noise.

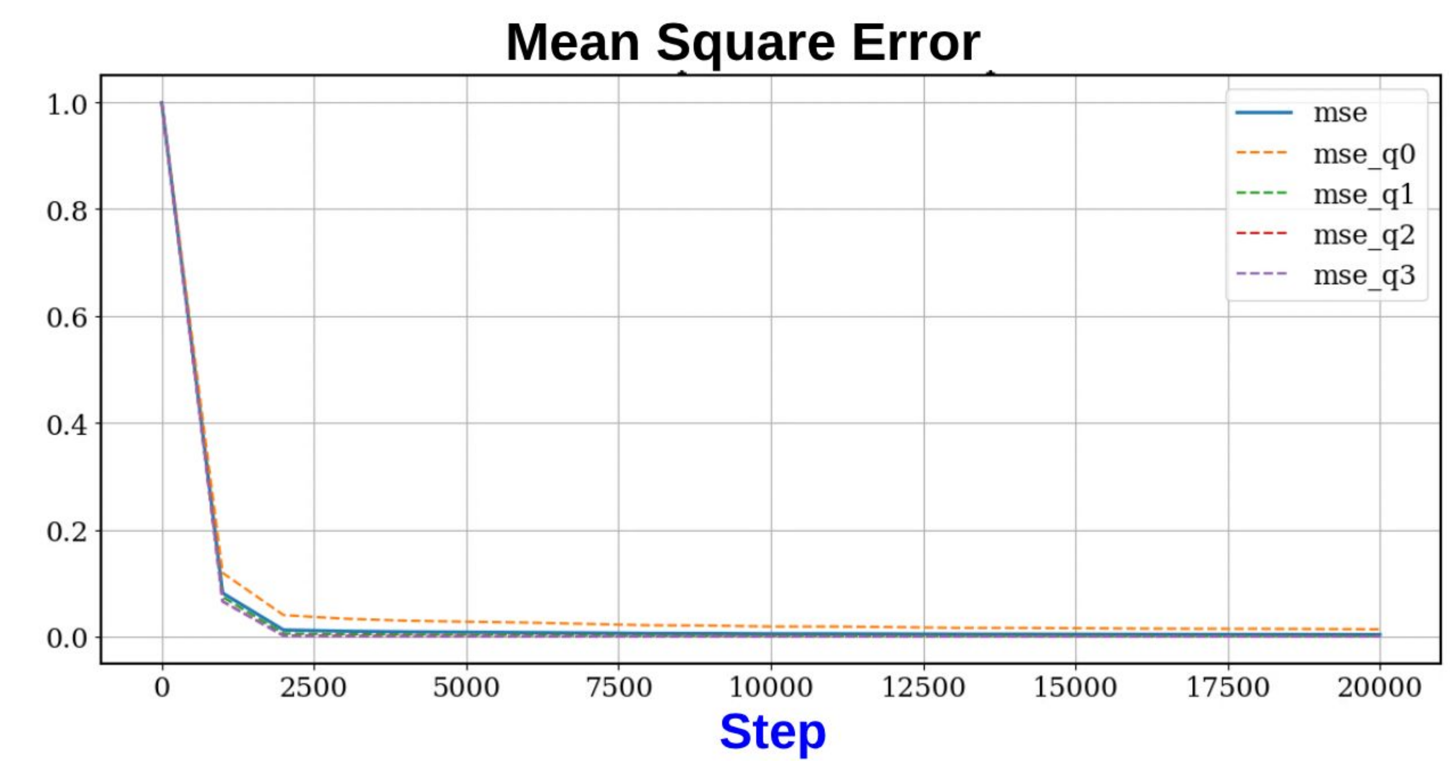


DDPM Training

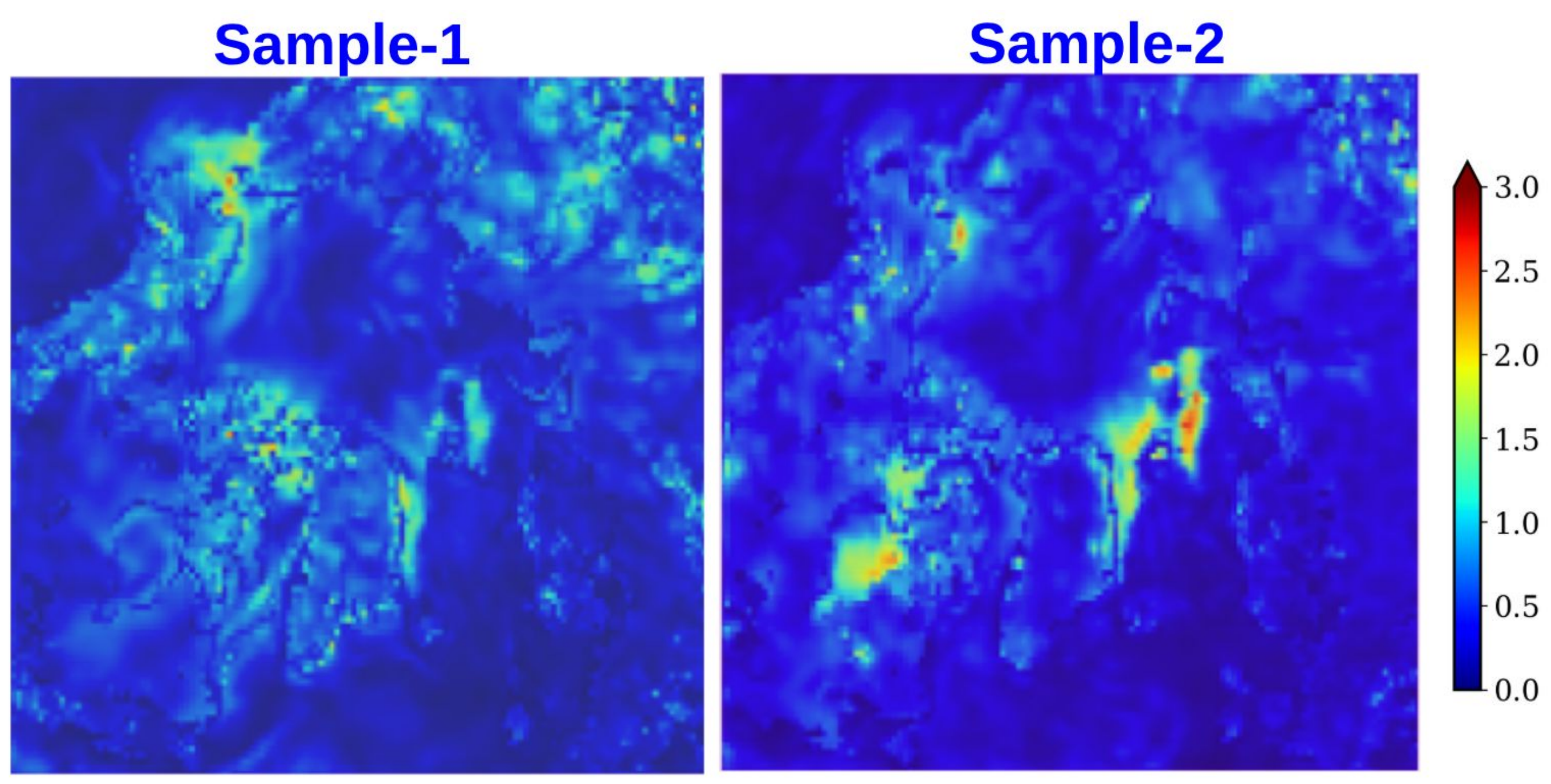


A total of twenty variables were individually selected for model training. These comprised four surface or near-surface variables: 2-meter temperature (in kelvin), meridional and zonal wind components (in meters per second), and surface pressure (in hectopascals). Furthermore, temperature (in kelvin), meridional and zonal wind components (in meters per second), and specific humidity (g/kg) were included at multiple pressure levels, namely 500 hPa, 700 hPa, 900 hPa, and 950 hPa.

Results



Uncertainty Quantification (UQ)
Machine Learning Generated Samples for 2m Temperature (SD)



The figure clearly shows that the diffusion model can capture the uncertainty patterns in the CARRA2 area. It is also important to note that the diffusion model effectively captures the orographic features across various land areas, with particularly clear representation over the Greenland region.

Next, the process of hyperparameter tuning, particularly when applied to extensive datasets in conjunction with score-based data assimilation (SDA), is crucial for improving the efficacy of machine learning (ML) models.

References

- "Copernicus Arctic Regional Reanalysis (CARRA)." https://climate.copernicus.eu/copernicus-arctic-regional-reanalysis-service
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. https://arxiv.org/abs/2006.11239
- Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. https://arxiv.org/abs/2010.02502

Acknowledgements: Copernicus is the Earth observation component of the European Union's space programme. The European Centre for Medium-Range Weather Forecasts (**ECMWF**) has been appointed by the European Commission with funding from the **EU** to operate the Copernicus Climate Change Service on its behalf.