LATENT DIFFUSION MODELS AND MULTI-CHANNEL DATA INTEGRATION

A new road to high-resolution statistical downscaling

This presentation participates in OSPP

EGU

Outstanding Student & PhD candidate Presentation contest

Thea Quistgaard[1,2], Peter L. Langen[1], Tanja Denager[2], Raphael J. M. Schneider[2], and Simon Stisen[2]

[1]*iClimate, Department of Environmental Science, Aarhus University*, [2]*Department of Hydrology, Geological Survey of Denmark and Greenland (GEUS)*

AARHUS UNIVERSITY   GEUS

iCLIMATE
AARHUS UNIVERSITY INTERDISCIPLINARY CENTRE FOR CLIMATE CHANGE

OVERVIEW

To understand and mitigate climate change, we require **high-resolution climate projections**, especially when investigating **impacts on a local scale**. This work is focused on developing models for **statistical downscaling** of climate model outputs for **Danish hydrology** by integrating advanced **deep learning** and computer science techniques. We present a **Diffusion Model** [1,2] tailored for generating high-resolution (2.5 km x 2.5 km) climate variables relevant for hydrological models: precipitation, evaporation, and temperature.

Our approach is inspired by advancements made in Computer Vision, specifically in the areas of **latent image generation** and **conditional diffusion models** [3,4]. It further incorporates **multi-channel climate data** and geographical features to improve model accuracy and stability. The model attempts to map between two reanalysis products, **ECMWF Reanalysis v5** (ERA5) [5] and the **DANish atmospheric ReAnalysis** (DANRA) [6], as a proof of concept, and future work will focus on creating a **synchronized training dataset** between DANRA and realisations from the Regional Climate Model ensemble **EURO-CORDEX** [7], to develop a model for actual climate projection downscaling.

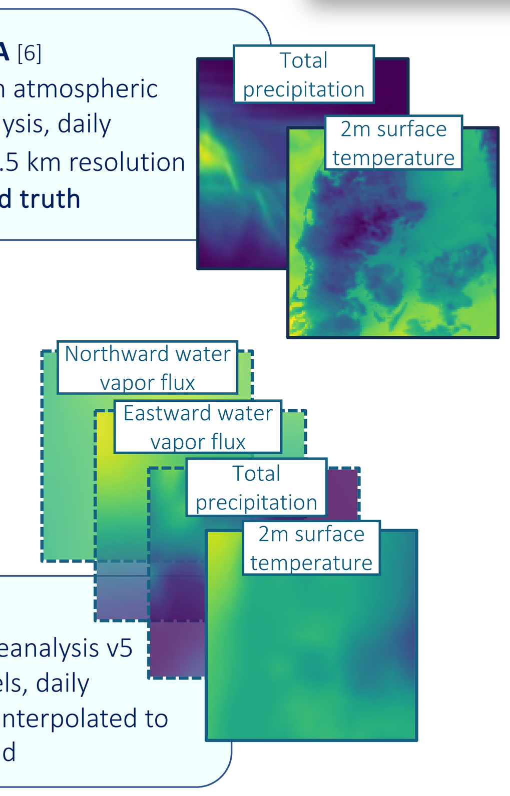
This work is part of a broader effort to **narrow the gap** between **global climate projections** and **local-scale analysis**. It covers the initial phase, where future advancements will focus on integrating this model with **RCM** and possibly **GCM** outputs instead of reanalysis data, to explore ensemble downscaling methods. Our end-goal is to **establish a robust pipeline for translating global climate projections into more actionable local climate statistics**.

WHAT'S NEW?

DATA

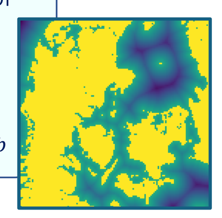
DANRA [6]
DANish atmospheric ReAnalysis, daily
2.5 x 2.5 km resolution
Ground truth

ERA5 [5]
ECMWF Reanalysis v5
Single levels, daily
Bilinearly interpolated to DANRA grid



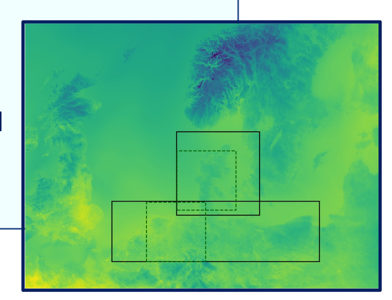
TRAINING

Loss function
- Signed Distance Function for weighted loss
- Hybrid loss [6] to capture multi-modality $L_{hybrud} = L_{simple} + \lambda L_{vlb}$



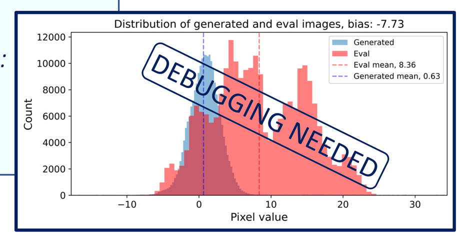
Data split
- *Training*: 9 years (1991-1999)
- *Validation*: 1 year (2000)
- *Test*: 100 randomly selected dates in 2001-2005

Data randomly shifted (geographically) for stability

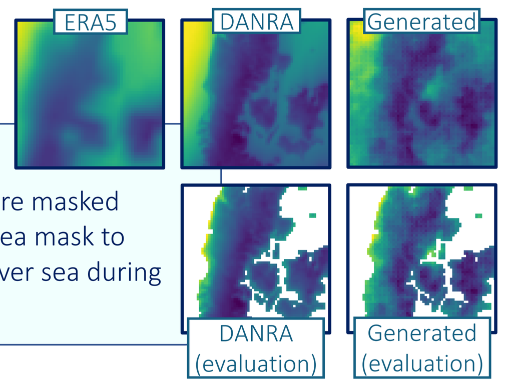


PRELIMINARY RESULTS

Quantitative
Distributions of generated and evaluation images (masked to land only)
Not capturing the mean:
May be an issue with standardization in code.



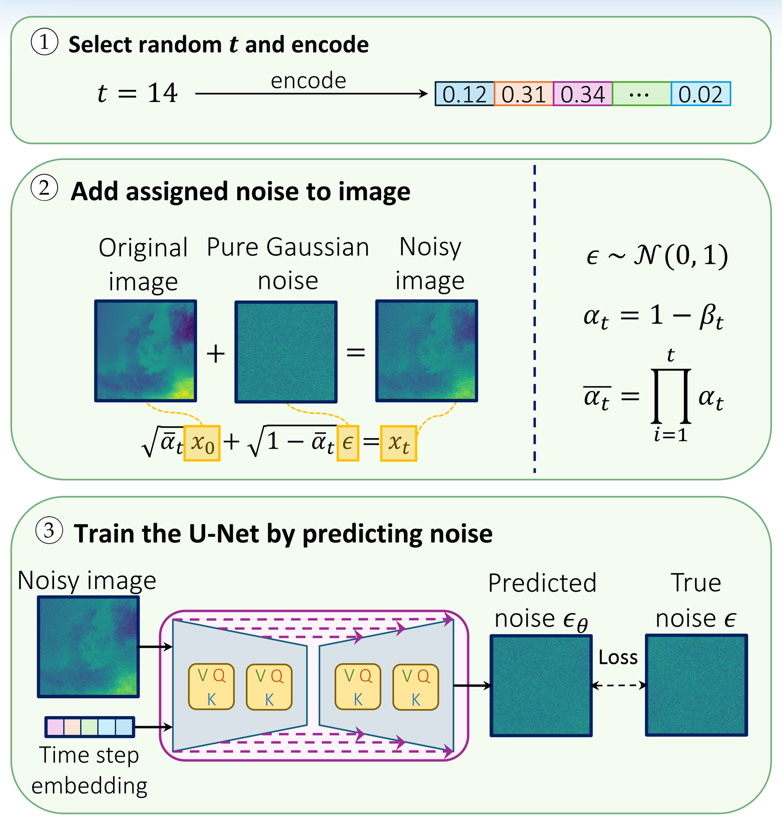
Qualitative
Predictions are masked with a land-sea mask to omit pixels over sea during evaluation.



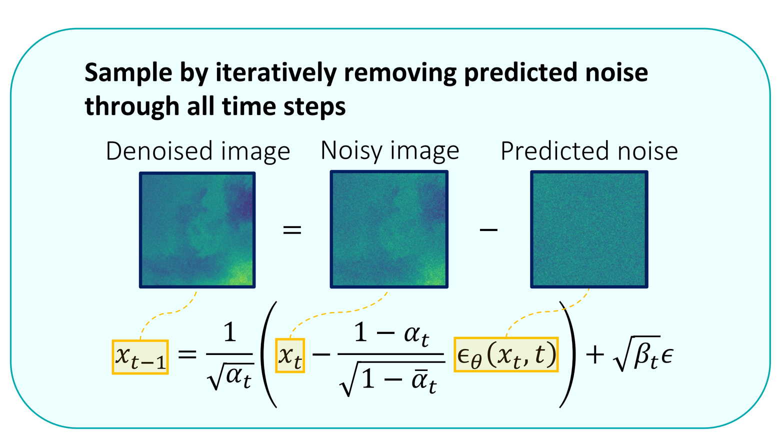
DIFFUSION MODEL

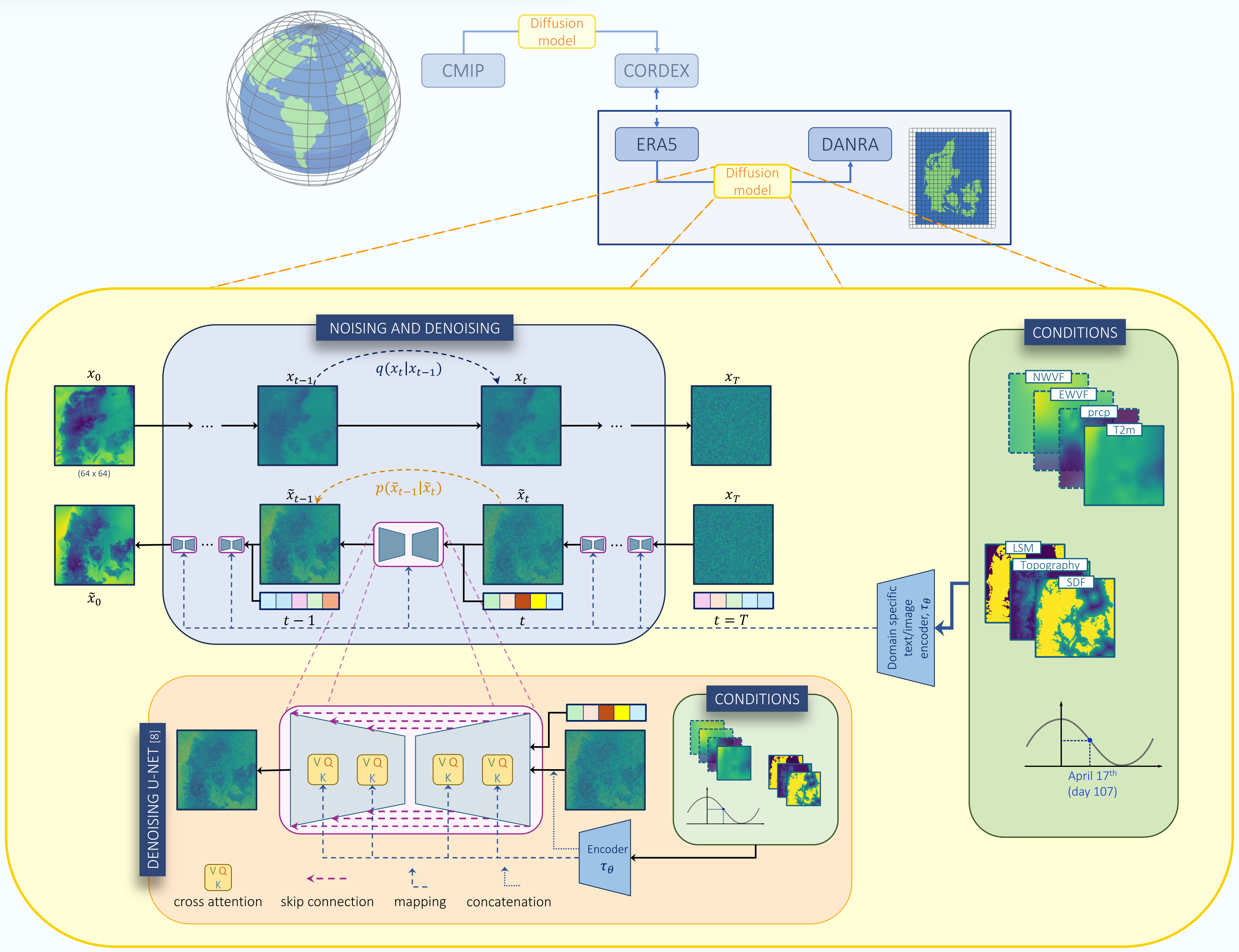
Generative diffusion models [1] [2] work by **computing a forward diffusion** process by iteratively adding gaussian noise with cosine sampling, then **learning the reverse diffusion** process with a U-Net, and finally **sampling through iterative denoising** of random noise.

TRAINING



SAMPLING

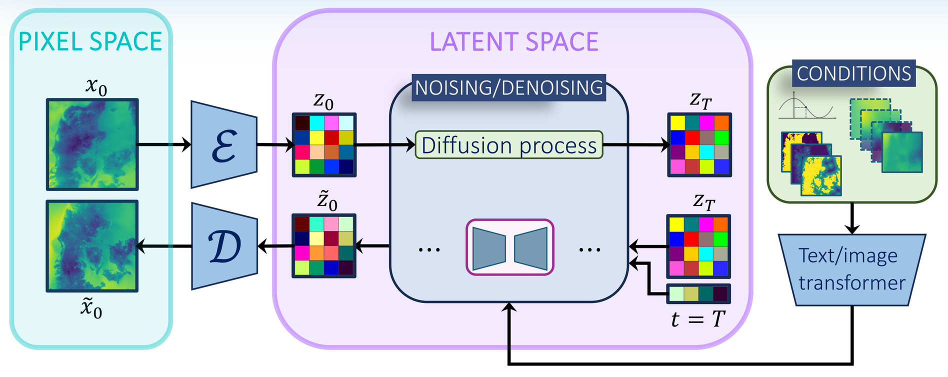




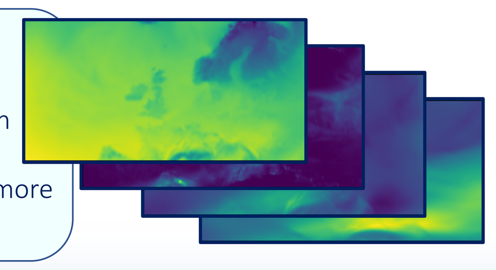
WHAT'S NEXT?

ON THIS MODEL

An ongoing effort is currently underway to speed up the training process through the implementation of a **Latent Diffusion Model** [4]. Alongside this, the conditioning will be modified to incorporate a **larger geographical and temporal domain** for capturing **large-scale atmospheric dynamics**.



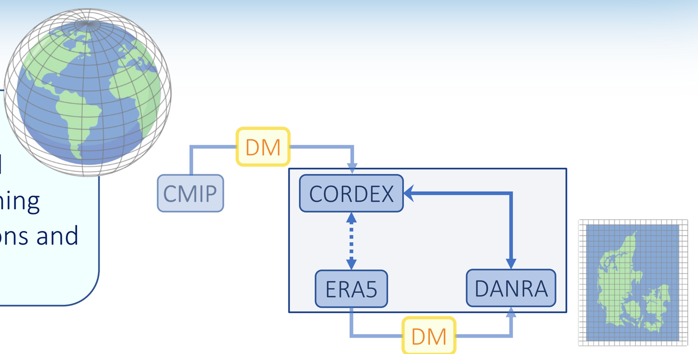
Extended domains
To capture larger atmospheric dynamics, especially in the North Atlantic. Adding a temporal component (synoptic scale) for more accurate projections



ON NEXT STEPS

The immediate next work will focus on generating a synthetic synchronised dataset between select EURO-CORDEX realisations and DANRA, to create a training dataset useful for making climate projection downscaling – for single models and full ensembles.

Next in the pipeline
Creating synchronised and futurized datasets for training model to climate projections and not just reanalysis data.



Contact
Thea Quistgaard, *tquistgaard@envs.au.dk*
Department of Environmental Science, Aarhus University
Frederiksborgvej 399, 4000 Roskilde, Denmark

Acknowledgement
This study is part of the *PEATlands and Climate-driven variability in groundwater depth – Impacts on greenhouse gas Emissions* (PEACE) project.

References

[1] J. SOHL-DICKSTEIN et al. (2015), "Deep Unsupervised Learning using Nonequilibrium Thermodynamics", in *Proceedings of the 32nd ICML, Volume 37, Pages 2256-2265*
[2] J. HO et al. (2020) "Denoising Diffusion Probabilistic Models", in *NeurIPS Proceedings 2020, Volume 33, Pages 6840-6851*
[3] A. NICHOL & P. DHARIWAL (2021) "Improved Denoising Diffusion Probabilistic Models", in *Proceedings of the 38th ICML, Volume 139, Pages 8162-8171*
[4] R. ROMBACH et al. (2022), "High-Resolution Image Synthesis with Latent Diffusion Models", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2022, Pages 10674-10685*
[5] H. HERSBACH et al. (2023), ERA5 hourly data on single levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, (on 07-APR-2023)
[6] X. YANG et al. (2022), "Danish regional reanalysis : Final scientific report of the 2021 national center for climate research work package 1.1.2 danish regional reanalysis"
[7] D. JACOB, J. PETERSEN, B. EGGERT et al. (2014), "Euro-cordex: New high-resolution climate change projections for European impact research," in *Regional Environmental Change, vol. 14, pp. 563–578*
[8] O. RONNEBERGER, P. FISCHER & T. BROX (2015), "U-net: Convolutional networks for biomedical image segmentation" in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*