LATENT DIFFUSION MODELS AND MULTI-CHANNEL DATA INTEGRATION

A new road to high-resolution statistical downscaling

Thea Quistgaard1,2, Peter L. Langen1, Tanja Denager, Raphael J. M. Schneider3, and Simon Stisen2
1Climate, Department of Environmental Science, Aarhus University, 2Department of Hydrology, Geological Survey of Denmark and Greenland (GEUS)

Contact
Thea Quistgaard, quistgaard@ens.au.dk
Department of Environmental Science, Aarhus University,
Frederiksbergvej 398, 4000 Roskilde, Denmark

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 (DM) tailored for generating high-resolution 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 realizations 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.

DIFFUSION MODEL
Generative diffusion models (e.g.,) 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.

WHAT’S NEW?

TRAINING

- Loss function
  - Signed Distance Function for weighted loss
- Hybrid loss to capture multi-modality
- Augment 𝑘–transform + Mask

DATA

- Data split
  - Training: 5 years (1991–1999)
  - Validation: 1 year (2000)
  - Test: 100 randomly selected dates in 2005–2015
  - Data randomly shifted (geographically) for visibility

ON THIS MODEL

An ongoing effort is currently underway to speed up the training processes through the implementation of a Latent Diffusion Model (DM). Alongside this, the conditioning will be modified to incorporate a larger geographical and temporal domain for capturing large scale atmospheric dynamics.

WHAT’S NEXT?

ON NEXT STEPS

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

PRELIMINARY RESULTS

ONTHE PIPELINE

Creating synchronized and suited datasets for training models to climate projections and not just reanalysis data.

REFERENCES