Project 2: Downscaling temperature and wind data for Lake Garda
Deadline: Friday June 30, 17:00

Starting Notebook: Project_2_pre.ipynb

In this project, you will downscale two-dimensional atmospheric model quantities from a relatively low resolution (18 km) to a medium (9 km) and (optionally, this is a bonus project) to high (3 km) resolution using a GAN as in Stengel et al., (2020) (paper to be discussed on June 16, is already on BB) over a 288km × 288km region as shown in Fig. 1. You can either use the explicit codes from previous notebooks, Tensorflow or PyTorch (which you can also run in Google Colabs), or the ones provided by Stengel et al., (2020), which are also provided.

The data (size about ∼9 GB) can be downloaded from

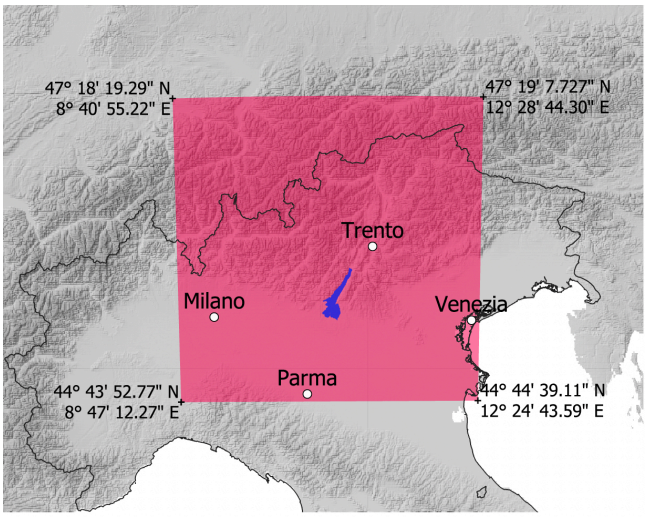

Figure 1: *Region of interest in Northern Italy.*

https://webspace.science.uu.nl/∼dijks101/ATCD-Project2 and the notebook will guide you to read in the data. The NCEP data are originally on a 25 km grid and are interpolated to an 18 km grid.

The low resolution (LR) data are from NCEP GFS over the period Jan 1, 2017 to March 31, 2018 (which will be used as training data) and over the whole year 2021 (which will be used as test data). The higher resolution training data are based on dynamic downscaling using WRF, both at a 9 km (MR) and a 3 km (HR) resolution.

a.
Make plots of the mean wind fields (U_{10}, V_{10}) and the surface temperature T_{2m}

for the training data.

b. whats that?
Compute the probability density distribution of the RMSE between the NCEP data and both the WRF 9km and 3 km resolutions for all three variables (U_{10}, V_{10}, T_{2m}. Explain shortly the differences in these (so-called baseline) distributions.

c.
Use a GAN first for T_{2m} variable on the LR-MR downscaling problem. In the GAN, there are three hyper parameters: the learning rate, the batch size and a parameter α, which measures the importance of the adversarial loss in the total loss. See Stengel et al. (2020) for the training method. What are 'best' values of these hyper parameters?

d.
Plot the RMSE for the test data. Can the GAN improve on the baseline result? Plot several 'best' downscaled temperature patterns in 2021 and explain the regions with the largest errors.

e.
Repeat the same as c) and d) but now for the wind vector veld (U_{10}, V_{10}), using a different GAN than in c.

f. *(Bonus)* Next perform the MR-HR downscaling only for the T_{2m} variable as in c) and d) and compare the results with the WRF (3 km) data. Make plots of the 'best' and 'worst' case temperature profiles (compare NCEP, downscaled HR and WRF-HR).