Project 2: ENSO prediction

Feedforward neural networks have been used extensively in El Niño prediction. In this project, you will perform actual El Niño predictions using a Gaussian Density Neural Network (GDNN) as described in Petersik and Dijkstra, GRL (2020). In a GDNN there are two output neurons, one for the mean and one for the standard deviation of the quantity to be predicted.

(i) Read the Petersik and Dijkstra, GRL (2020) paper and its Supplementary Information and become familiar with the ensemble training which is used. In the project below, you will focus on a so-called 'hindcast' of the most recent large amplitude 2015-2016 El Niño and on near-future predictions.

The code can be found at https://github.com/CCSS-Utrecht/ninolearn and can be run in Spyder (within Anaconda). There is a short manual which explains how you can make predictions of the ONI (explained in Petersik and Dijkstra, GRL (2020)) using a GDNN. Install this code and modify n_segments=2, n_members_segment=1 in the cross_training call in s2_training.py. You also need to change the paths in the file s0_start.py.

(ii) Use this as standard case (given features and hyperparameter settings) with initial conditions at February 1, 2015 to see if it runs and how long it takes on your machine. You make a prediction of the ONI for up to 9 months ahead with the GDNN and compare with observations. During training, you may get warnings from TensorFlow but these can be ignored. Monitor the loss functions during training.

(iii) First you are going to study the quality of the prediction (think of a quality measure when comparing observations and the GDNN results) versus the number of segments (standard: 2, go up to 5) and the number of ensembles (standard: 1, go up to 3). How does the ensemble training improve the quality of the prediction? Note that training can take a while with more ensemble members and more segments.

(iv) For adequate values of the number of segments and number of ensembles (depending in (iii)), study the quality measure of the prediction when using June 1, 2015 as initial conditions. Try to explain the difference in the results for both initial conditions.

(v) For the case of February 1, 2015 initial conditions, study the effect of the different features, in particular by leaving out the warm water volume, on the quality of the predictions.

(vi) *Optional:* Make a prediction for the near-future ONI, using February 1, 2022 as initial conditions (for which observations are available) for lead times up to 9 months.