

NINOLEARN PREDICTIONS MANUAL

Miriam Sterl

June 9, 2021

Introduction

NinoLearn is a research framework for the application of machine learning (ML) methods for the prediction of the El Niño-Southern Oscillation (ENSO). It contains methods for downloading relevant data from their sources, reading raw data, postprocessing it and then access the postprocessed data in an easy way. Moreover, it contains models for the ENSO forecasting. For code, information and documentation on the NinoLearn framework, see <https://ppetersik.github.io/ninolearn/intro.html> and <https://github.com/CCSS-Utrecht/ninolearn>.

This manual outlines how to use the code to make forecasts of the Oceanic Niño Index (ONI) together with an uncertainty in the predictions, using Gaussian Density Neural Network (GDNN) models. For more information, see Petersik and Dijkstra (2020).

1 Installation and set-up

You only have to go through these steps if you don't already have the NinoLearn files on your local machine. In order to be able to run all the code, you will need to have the following Python packages installed:

- `numpy`
- `matplotlib`
- `pandas`
- `xarray`
- `dask`
- `scikit-learn`
- `netcdf4`
- `tensorflow`

Fork the following repository from GitHub and create a copy on your local machine:
<https://github.com/CCSS-Utrecht/ninolearn>

2 Running the code

In your local `ninolearn` repository, go to the folder `predictions`. Here you find six numbered files.

In the first file, `s0.start.py`, you are required to fill in the first date (month and year) for which you want the prediction to start, as well as the paths to your root directory and data directory.

After that, you can run all six files in order of their numbers, without having to change anything else in the code. You do not have to do this all at once; the only important thing is that you stick to the order. Below, it is outlined what happens in each step.

- Step 0: Start (s0_start.py).** First, fill in the full path to the directory where you have your local `ninolearn` folder, and the full path to the directory where you want to store data. Second, fill in the month and year where you want the prediction to start. Then, run the code.
- Step 1: Data downloading and preprocessing (s1_data.py).** Run the code. Some new directories are created and the data used for training the models is downloaded and preprocessed. Also, the data from the IRI/CPC ENSO Predictions Plume (https://iri.columbia.edu/our-expertise/climate/forecasts/enso/current/?enso_tab=enso-sst_table) is downloaded. Furthermore, the latest common date of all data sets is saved; this will be the final date of the data sets to be used for the training.
- Step 2: Training the models (s2_training.py).** Run the code. The GDN models are trained and saved. This takes a while!
- Step 3: Making the predictions (s3_predict.py).** Run the code. The trained models are loaded and used to make predictions of the mean and standard deviation of the ONI for the nine coming three-month seasons. This also takes a while, though not as long as step 2. These predictions will be saved as a `.csv` file in the subfolder `forecasts` within your data directory. You can load it into Python later as a Pandas DataFrame using the `pandas.read_csv` function.
- Step 4: Plotting the predictions (s4_predict.py).** This step is optional. Run the code to generate a plot of the predictions together with their uncertainty (1 standard deviation). If available, the last observation before the start of the prediction period is also plotted, as well as the predictions from IRI/CPC. Also, if available, the observed values of the ONI for the predicted period are plotted. The plot is saved as a `.png` as well as a `.pdf` file in the subfolder `forecasts` within your data directory.
- Step 5: Removing data (s5_finish.py).** Run the code. The directories and files with (raw and processed) data, trained models, and information saved in between (such as lead times) are removed. Next month you can have a fresh start! Note: the folder with the forecast data is NOT automatically removed. You can keep it to save the predictions every month. This folder also contains all the information required to run step 4 (plotting), so the plot can still be adjusted.

References

- Petersik, P. J. and Dijkstra, H. A. (2020). Probabilistic Forecasting of El Niño Using Neural Network Models. *Geophysical Research Letters*, 47(6):e2019GL086423.