# Project 1: Classification of ENSO extreme phases through SST data

Deadline Friday May 31, 9:30

**Starting Notebook: Project_1_pre.ipynb**

In this project we will use global sea surface temperature anomaly patterns to determine whether there are El Niño conditions or La Niña conditions in the equatorial Pacific. You can either use the explicit codes from previous notebooks, Tensorflow or PyTorch (which you can also run in Google Colabs).

a.
Load in the data from the netcdf file. This is ERSSTv5 SST (sea surface temperature) data set which has been de-seasoned, linearly detrended, and has a 5 month look back running mean applied. See notebook.

b.
Include only samples where an El Niño or La Niña events is occurring (i.e. NINO3.4 > 0.5 or NINO3.4 < −0.5). Make two classes, where 1 is the occurrence of an El Niño and 0 is the occurrence of a La Niña.

Then, divide the SST fields and NINO3.4 output into training (1950 − 2000), validation (e.g. years 2000 − 2010) and testing (> 2010) sets. Make a plot of the NINO3.4 data of the El Niño and La Niña events and show the different date sets (training, validation and testing) in different colours.

Standardise the SST anomaly values of all selected data in b. using the standard deviation of each grid cell within the training set (and set all NaNs, i.e. land points, to zero).

c.
Define a feedforward neural network (FNN) architecture with $h$ hidden layers with $n_h$ neurons, with a learning rate $\eta$, with an activation function $A$, a loss function $J$, ridge regularisation with parameter $\beta$, the number of epochs $n$ and a batch size $b$. The number of output nodes is $n_O = 2$ (two classes). Use initially: $h = 2$, $n_h = 12$, $\eta = 10^{-3}$, $A$ = 'Relu', $J$ = 'Categorical Cross Entropy', $\beta = 10^{-5}$, $n = 10$ and $b = 32$.

d.
Train the FNN on the training data set (using e.g. the Adam optimizer) and study the behaviour of the loss function for both training and validation data. For the standard values of the hyper-parameters, how many epochs are needed to obtain an accuracy score on the training data of 5%? What is validation accuracy in this case?

e.
What is the effect of the learning rate on the training error?

f.
Study the effect of the activation function (apart from 'Relu', choose 'Sigmoid' and 'Softmax') on the training and validation results. Which one gives more ac- curate results for the same number of epochs?

g. Now evaluate the model for the SST test data set and determine model accuracy in correctly classifying El Niño and La Niña. Plot also the so-called confusion matrix, showing the number of the predicted events (false/true) versus the true events.

h. Finally, identify the El Niño and La Niña events that have been misclassified by the FNN and make composite plots of them, separately for El Niño and La Niña events.