TRIED TPC04 Report

Carl Robinson - 9th Jan 2018

Part 2: Use of time series for prediction of the El Niño phenomenon

Objective, Data and Method

Objective & Method

El Niño is a meteorological phenomenon that occurs periodically in the equatorial Pacific Ocean and affects the global climate. It manifests as a strong variation in the sea surface temperature (SST), and is partly caused by the wind. The aim of this research is to use self-organising maps (SOM) to model SST data taken from four geographical zones during El Niño events, and then try to predict the occurence of the El Niño phenomenon.

An SOM was implemented as a neural network to estimate the probability density function of the data. It uses vector quantisation, so the output is a lower-dimension representation of the input space.

Dataset and Codings

The dataset is a text file containing 436 measurement rows, and 13 columns:

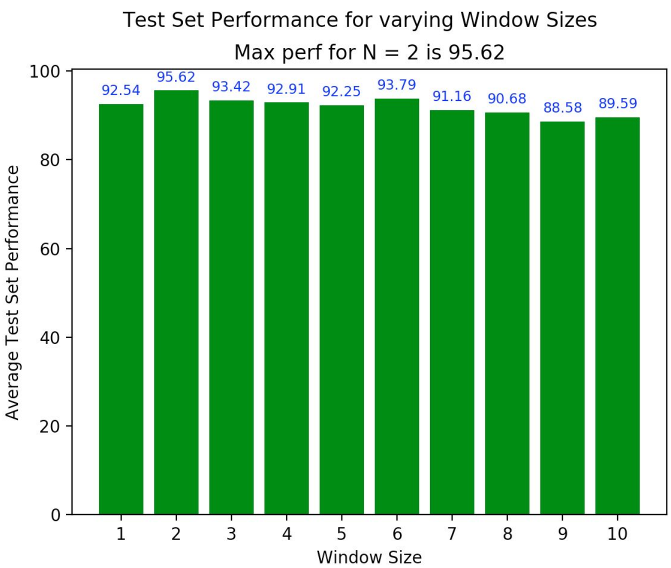
- The date column is in the format YYMM (year and month).
- The 4 SST columns are time series of mean temperatures from the 4 Pacific regions, from which the seasonal component has been removed. This format helps to reveal the inter-annual temperature anomalies. Note that before the year 1970, we have no average temperature values, so the values are set to -9999.00.
- The 8 wind columns are time series of inter-annual wind stress anomalies in the same 4 regions, split into longitude and latitude components. They can be treated as a pseudo wind speed, expressed in $(m/s)^2$.

Results & Conclusions

1) Determining the optimal window size

To find the optimal time-series window size, the test accuracy for window sizes from 1 to 10 was calculated. 75% of the SST measurements were used for training, the remaining 25% used for test. Map dimensions 7*7.

We use two stages of batch SOM training; stage 1 uses a higher initial temperature to consider a wider neighborhood and stage 2 uses smaller temperatures to refine the vector quantization.
Stage 1 parameters: iterations (100); temperature range [5.0 to 1.0];
Stage 2 parameters: iterations (50); temperature range [1.0 to 0.3];

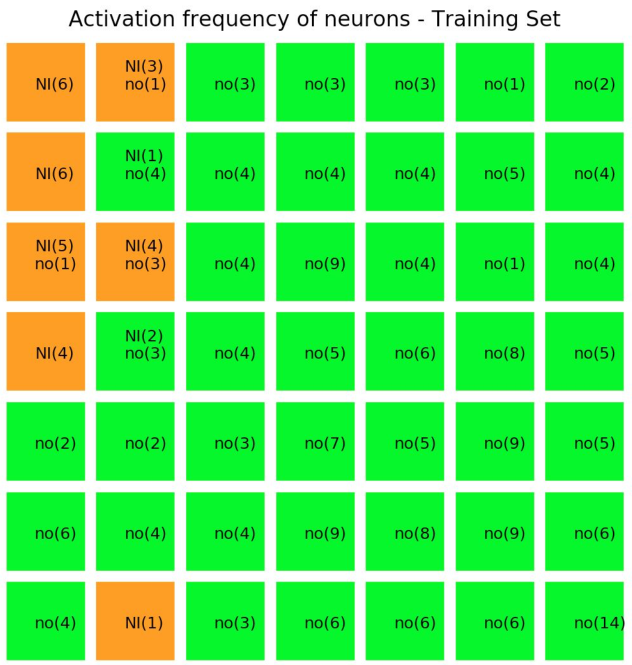


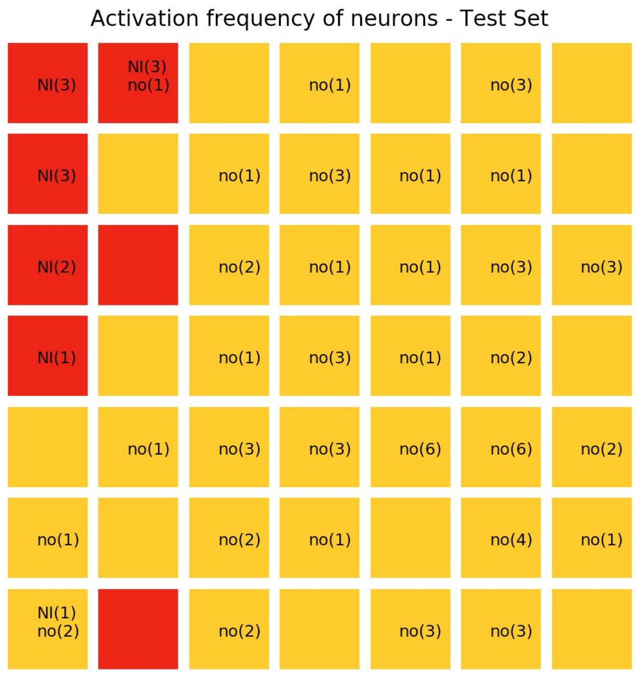
As the SOM referents are initialised with random values, it is necessary to run the training multiple times in order to find a global optimum. Each model was run 10 times, with and the highest accuracy obtained for each window size was retained for comparison.

Plotting these on a bar chart shows that the maximum performance was obtained when the window size N = 2.

2) Studying the Optimal Case of Window Size N=2

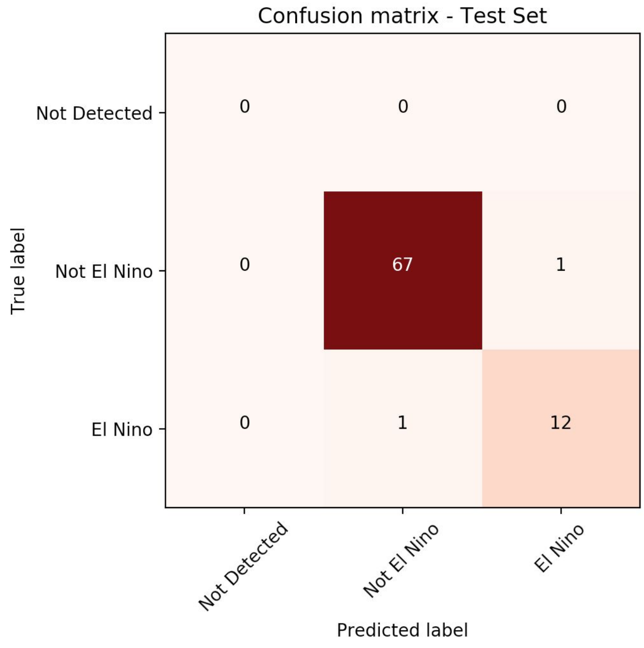
We can visualise the classification results on both the training and the test set, using an SOM trained on the dataset of optimal window size N = 2. Before training, the criterion Norm ($SST1_{t+1}$)>1 was used to classify windowed time segments as El Nino (NI) / not El Nino (no). After training, each referent is classified using a majority vote based on the observations assigned to it. The class and the number of training data per class (frequency hit) are displayed below. We can immediately see that the same pattern of referents classification occurs across both sets.





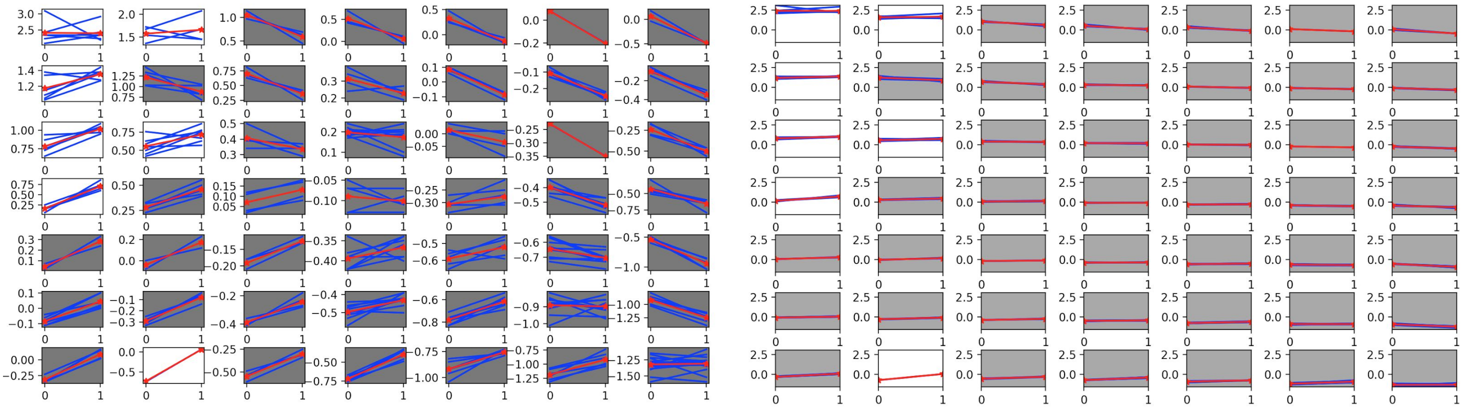
Plotting the confusion matrices highlights the fact that a small number of 'not El Nino' observations were incorrectly classified as 'El Nino' in both training and test sets, and vice-versa also. All other observations were correctly classified. The classification accuracies are very high in both cases.

Training set accuracy: 96.73%



Test set accuracy: 97.53%

Next we plot the profiles of each of the referents, and the data of the training set that they have captured. On the left the axes are scaled to the referent, and on the right the axes are set to the same scale for comparison.



Lastly we can visualise the map components (variables). There is a clear diagonal order to the values, with the highest values at the top left, and the lowest values in the bottom right. This corresponds with the fact that the referents classified as ‘El Nino’ are found in the top left corner of the map, showing that a high activation indicates the presence of El Nino.