Seasonal prediction using unsupervised feature learning and regression

Mahesh Mohan[1], Cheng Tang[1], Claire Monteleoni[1], Timothy DelSole[2] and Benjamin Cash[2]

[1]George Washington University, [2]Center for Ocean- Land Atmosphere Studies, George Mason University

Abstract

We propose to use machine learning to discover indices from the SST (Sea Surface Temperature) field data and compare their prediction performance to that of the Nino3.4 index on tasks related to ENSO. As a first step in this direction, this work focuses on predicting the time-series of monthly temperature anomalies in Texas, from the time series for the whole ocean SST field, ending 6 months prior.

Why Texas temperature anomalies?

Drought/heat wave in Texas in 2011 raised critical questions about the role of ocean temperatures and the extent to which such events can be predicted in the future [3].

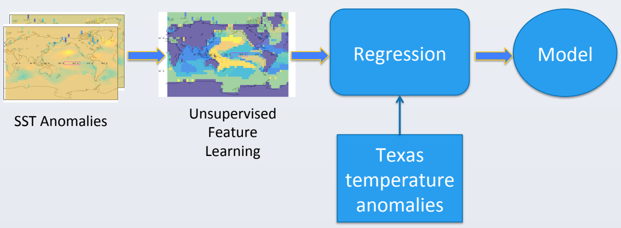
Why new features?

Many of the currently used climate indices, including Nino3.4, were originally defined by human experts. The goal here is to obtain a data driven method to learn the climate indices automatically.

What kind of feature generation methods?

In this project, we explore the use of clustering based approaches (specifically k-means and spectral clustering) to generate features from the SST field.

Proposed framework



Results

All our experiments were performed on data from the MLOST dataset [4]. The data was preprocessed by smoothing using a 3-month moving average filter. The evaluation framework used in this paper, closely follows the "progressive validation" error, analyzed in [6]. Prediction performance was then determined by computing two metrics, the Normalized RMSE ($1-\frac{RMSE}{\sigma}$), and the correlation.

Effect of number of features

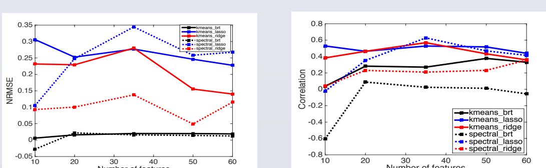


Fig. 1. Absolute value of the Normalized RMSE (left) and Correlation (right) for a 1 month look-ahead task. Both plots show that 35 features suffice for good prediction. This indicates the existence of a strong latent structure, since the original input data had significantly larger number of features.

Predicted Time-series

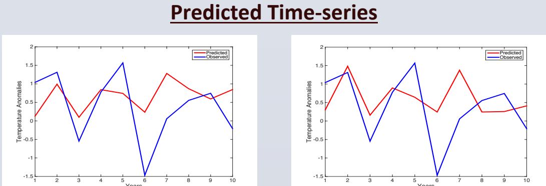


Fig. 2. The predicted time series by using k-means + LASSO (left) and spectral + LASSO (right).

Effect of the lag

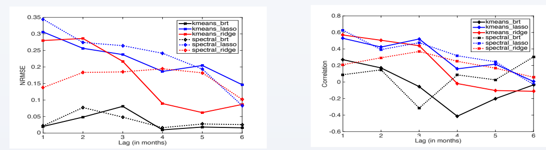


Fig. 3. Above figure shows the performance (correlation on the left and NRMSE on the right) for predicting Texas temperatures at different lags. The correlation of the predicted time series decreases with the lag. However, the results remain statistically significant for lags of up to 4 months.

Examples of learned features

K-means, 10 features

K-means, 30 features

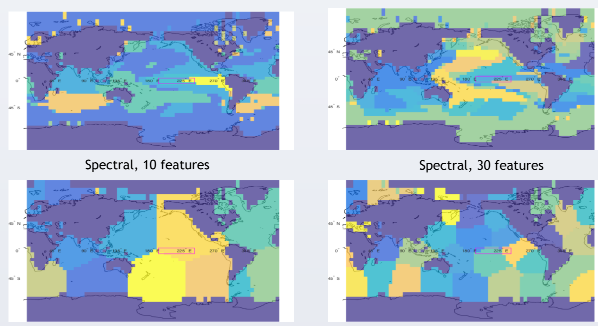


Fig. 4. Figure shows the features generated using k-means (top) and spectral clustering (bottom). The features generated using k-means clustering do not enforce any spatial contiguity constraints on the clusters and manage to capture a significant number of known patterns in the SST field.

- Summary of preliminary results
  - Prediction using the proposed method outperformed the prediction using Nino 3.4 indices.
  - For shorter lags, the predicted time series have a significant correlation with the Texas temperature anomalies.
- Future directions
  - Study other supervised regression tasks related to ENSO, e.g. prediction of temperature/precipitation in other regions.
  - Explore other notions of error, to quantify the prediction of extremes.
  - Identify key regions in the SST field that enable better prediction.

Acknowledgements

We acknowledge the NOAA Merged Air Land and SST Anomalies data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, http://www.esrl.noaa.gov/psd/

References

[1] WCRP, "World Climate Research Programme: Grand Challenges." http://www.wcrp-climate.org/grandcha.shtml, 2013.

[2] S. Liess, A. Kumar, P. K. Snyder, J. Kawale, K. Steinhaeuser, F. H. M. Semazzi, A. R. Ganguly, N. F. Samatova, and V. Kumar, "Different modes of variability over the tasman sea: Implications for regional climate," *Journal of Climate*, vol. 27, pp. 8466–8486, 2015/05/19 2014.

[3] M. Hoerling, A. Kumar, R. Dole, J. W. Nielsen-Gammon, J. Eischeid, J. Perlwitz, X.-W. Quan, T. Zhang, P. Pegion, and M. Chen, "Anatomy of an extreme event," *Journal of Climate*, vol. 26, no. 9, pp. 2811–2832, 2013.

[4] T. M. Smith, R. W. Reynolds, T. C. Peterson, and J. Lawrimore, "Improvements to noaa's historical merged land-ocean surface temperature analysis (1880-2006)," *Journal of Climate*, vol. 21, no. 10, pp. 2283–2296, 2008.

[5] D. Giannakis and A. J. Majda, "Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability," *Proceedings of the National Academy of Sciences*, vol. 109, no. 7, pp. 2222–2227, 2012.

[6] E. Szekely, D. Giannakis, and A. J. Majda, "Extraction and predictability of intraseasonal signals in infrared brightness temperature data," *4th International Workshop on Climate Informatics*, Sept. 2014.