# Predicting Indian Monsoonal Rainfall Using a Sequential Neural Network on Principal Components of Atmospheric Data


Science: Earth Science

Thomas Favata

Mentors: Hoot Thompson and Thomas Maxwell

NASA Goddard Space Flight Center, Greenbelt, MD 20771

NASA Center for Climate Simulation, Code 606.2


## Introduction

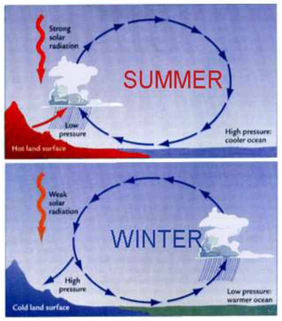


Figure 1) Pictured above is the background of a monsoon and what meteorological factors contribute to its creation. http://carolinasweatherplus.blogspot.com/

Monsoonal variation effects many people across India, and a better way to predict the extent of this phenomena would greatly benefit the Indian people. Surface temperature has been used in most literature to create statistically significant forecasts of the Indian Monsoon. Empirical Orthogonal Functions (EOFs) and their Principal Components (PCs) were calculated to find the spatial and time series components respectively. This research uses a sequential neural network to incorporate the PCs of the heights of pressure layers in addition with the temperature data to create a more comprehensive prediction method of Indian monsoons.

## Data

•**MERRA-2** – The Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) is an atmospheric reanalysis with a resolution of 0.625ºx0.5º that goes back to 1980. The two variables used in this research were skin temperature and the height of a pressure level.

•**IITM**– This is a data set from the Indian Institute of Tropical Meteorology called the Longest Instrumental Rainfall of the Indian Regions. The rainfall data for All-India was included in this research as the training data for the neural network. June, July, August and September were used to find data as these are the months of the Indian Monsoon. The IITM All-India data set has total rainfall data over the period from 1813-2006.

## Methods

The PCs calculated from the 20CRV2c reanalysis were used as input to a neural network implemented using the Keras package. Backpropagation was used to train the network to forecast IITM AISMR timeseries with a one year lag time. The training was halted when the validation error reached a minimum

### EOFs and PCs

**EOFs** - Empirical Orthogonal Functions are used to find both the time series and spatial patterns within a specific matrix of data. When calculated, the eigenvalues provide the percent variance for each mode. This establishes the level of overall variance accounted for with the mode.

**PCs** – Principal Components are the temporal pattern within empirical orthogonal functions. Component scores are calculated and are compared through time with the rainfall data from the training data set. The component scores have to be scaled to the data and are by themselves meaningless. These values produce a prediction by the Sequential neural network. When the prediction is displayed alongside the training data they determine the extent of machine learning from the neural network.

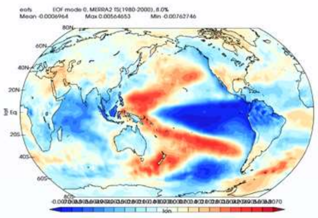

Figure 2) The spatial distribution of the first MERRA2 surface temperature EOF, which represents the dynamics of El Nino

**Principal Components of El Nino**

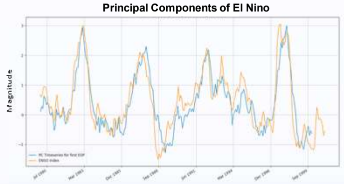

Figure 3) A comparison of the dynamics of the first (rescaled) MERRA2 surface temperature principal component and the ENSO index, shown here for 1980-2015

### Loss and Validation Loss

**Loss** – Loss is the error function used to classify how close the prediction is to the validation data. The loss function used in this sequential model is a mean squared error. The function should be minimized for the best results.

**Validation Loss** – Validation loss takes into account the strength of learning as well as overprediction. Overprediction occurs when the learning taking place is only focusing on the patterns within the training data and not the overall patterns the training data is showing. In a result where the loss and validation loss are low there is learning taking place.

## Results

### Skin Temperature

Skin temperature is the temperature at the surface of the Earth. This is an important variable used within models to predict monsoonal rainfall. It has been found to be extremely important in prediction by Gadgil et al (2007), Cannon and McKendry (1999), and Sahai et al (2003).

### Pressure Heights

The height of a pressure level is the elevation of a specific pressure surface. Due to the compression and expansion of the Earth's atmosphere this process is dynamic. This variable is included in this model for monsoon rainfall prediction and will improve on the models which only contain temperature data.

### Model Generation

The IITM monsoonal rainfall data was used to train the model. The months found to give the best predictions for Indian monsoonal rainfall were August and July. The yearly data for these months were incorporated into the model and ran to see how well the prediction of these data would validate the output data set. The best results included a low training loss accompanied by a validation loss that was low, but not lower than the training loss. This stipulation increases the probability that the results were not an anomaly . This followed by a reasonable prediction of the training data shows improved learning.

**Prediction of Monsoon Using Skin Temperature**

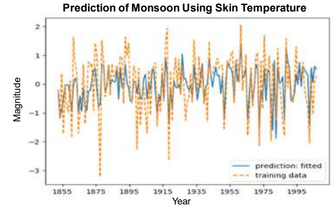


Figure 4) Pictured above is the model based prediction using just the skin temperature data. There is not as strong a prediction when compared to figure 5 where the height data is included.

**Prediction of Monsoon Using Skin Temperature and 500 mb Pressure Level**

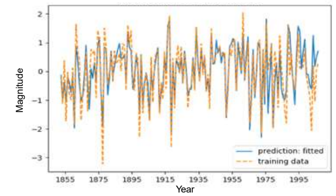


Figure 5) Pictured above is the model based prediction using the skin temperature data and the height of the 500 mb pressure layer. It is a better prediction than figure 4.

**Loss Using Skin Temperature**

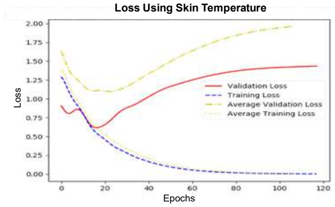


Figure 6) Pictured above is the validation loss and training loss of the prediction model using just the skin temperature. Both the training loss and validation loss are larger than in figure 7. The average values are also larger than in figure 7.

**Loss Using Skin Temperature and 500 mb Pressure Level**

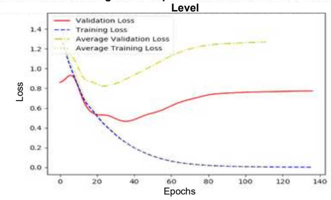


Figure 7) Pictured above is the validation loss and training loss of the prediction model using the skin temperature and the height data for the 500 mb level. Both the training loss and validation loss are smaller when compared to figure 6.

## Conclusions

There is a relationship between total Indian monsoonal rainfall and previous atmospheric conditions. Using both skin temperature and the height of pressure levels as variables, a more accurate prediction can be constructed. This research is still ongoing and the variables that determine the model are changed the performance of the prediction can be improved. The most difficult part of the prediction method is over-prediction. The chance for this is greater with every epoch run, so the challenge is to decrease the training loss as much as possible while still decreasing the validation loss. Finding a balance with prediction is the goal of this research, and if an improved model is found, a connection between the heights of pressure levels and Indian monsoons can be shown. A better prediction of the Indian monsoon will help many people be more prepared for this extreme meteorological event.

## References and Acknowledgements


Sulochana Gadgil, M. Rajeevan, and P. A. Francis, 2007, "Monsoon Variability: Links to Major Oscillations Over the Equatorial Pacific and Indian Oceans." Current Science, 93, 182-194

Alex J. Cannon, and I.G. McKendry, 1999, "Forecasting All-India Summer Monsoon Rainfall Using Regional Circulation Principal Components: A Comparison Between Neural Network and Multiple Regression Models." Int. J. Climatol., 19, 1561-1578

A. K. Sahai, D. R. Pattanaik, V. Satyan, and A. M. Grimm, 2003, "Teleconnections in Recent Time and Prediction of Indian Summer Monsoon Rainfall." Meterol. Atmos. Phys., 84, 217-227 143–156., doi:10.1257/jep.15.4.143.



Thomas Maxwell

NASA Center for Climate Simulation