Modelling the Indian Summer Monsoon

History, Goals and Methodologies

Shivam Patel[1]
Asif Shaikh[2]
Under guidance of Prof S Juneja[3]

[12]Electrical Engineering, IITB
[3]Tata Institute of Fundamental Research

21st May, 2022

Indian Summer Monsoon Rainfall (ISMR)

Summer Monsoon in India is the major rainfall season in India, accounting for over 90% of rainfall in the western and central India, and 50-75% rainfall in the southern and northwestern regions of the country.

India is an agrarian country, with over 50% population being directly or indirectly dependent on the agriculture industry. In such a country, reliable and timely rain forecasts are of utmost importance for a good crop and economy growth.

References

New statistical models for long-range forecasting of southwest monsoon rainfall over India, Rajeevan et al (2006), National Climate Centre, IMD Pune

Introduction

The weighted average of Indian Rainfall over all areas during all parts of the year is almost 89cm. There is a 10% variability in rainfall YoY. This may seem nominal, but even a small fluctuation can affect agriculture and related industries.

A testament to the detrimental effect of such fluctuations can be judged by the economic loss incurred due to the below deficit rainfall in 2002, 04.

The failure of statistical methods for monsoon predictions needed serious reconsiderations, as they were able to forecast accurately in normal rain years, but were unreliable for forecasting deficit or excessive rainfall years.

Methods Used

For the Long Range Forecast of the ISMR, three main methods are used -

1. Statistical Methods use the historic relation between various predictors like global atmosphere-ocean parameters, and the ISMR. This relation is mathematically quantified using models and such models are further used for making predictions.
2. Empirical Methods are based on analysing time series data of the past rainfall. Such methods do not use any predictors, and hence cannot be relied on to give accurate predictions in case of extremities. Thus, these models give probability distribution of rainfall, rather than an assertive value.
3. Dynamical Methods are the most widely used inferential approaches today. They are based on studying the general circulation models of the atmosphere and oceans to simulate the summer monsoon circulation and associated rainfall.

Methods Used

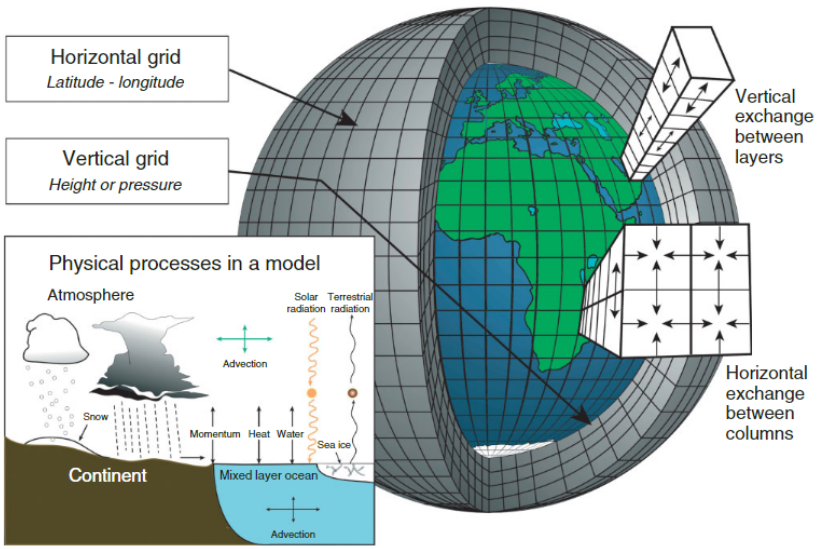

Figure: General Circulation Models

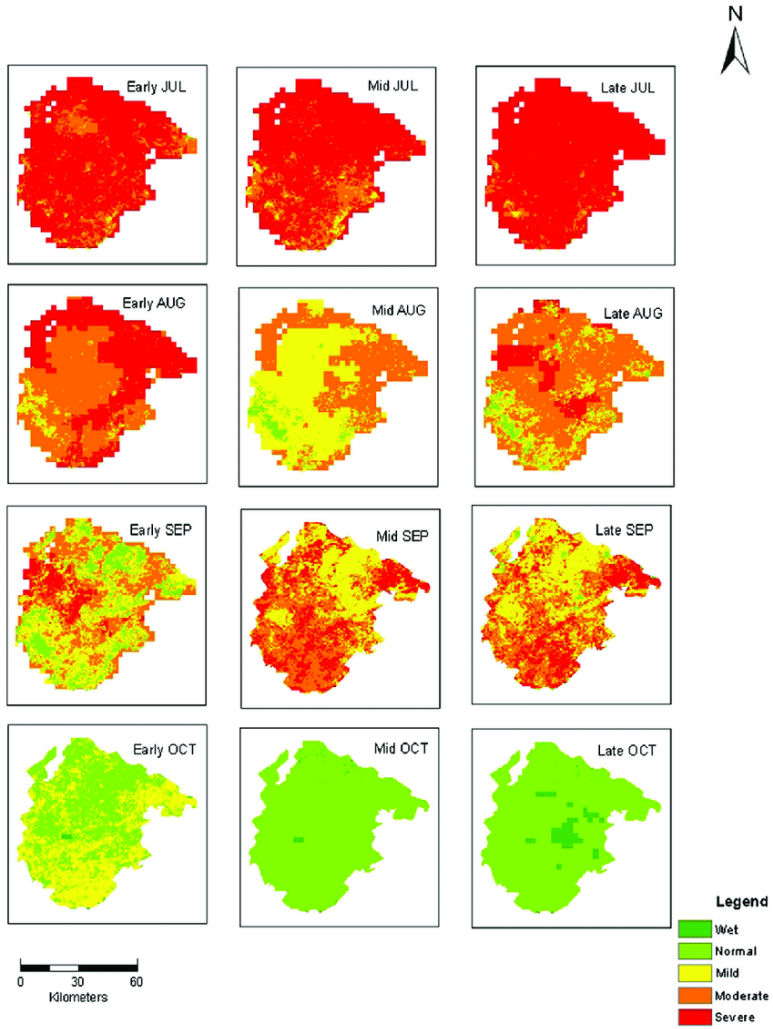

Figure: 2002 Drought, Rajasthan

Methods Used

Statistical models were predominantly used from 1988 - 2002. These were based on a 16-predictor power regression and parametric models. However, the mean forecast error was more than that due to climatological analysis. This model also failed to predict the severe drought of 2002. 21.5% deficit in rainfall was observed, with 57% fall in July rain.

Such shortcomings led to the development of a 2 stage forecasting system, where the forecasts were released twice a year, once in April and once in June.

Development of New Models

The major operational failure of power regressor models in 2002, 04 forecasts led to the critical examination of the following model parameters:

- Judging the relation between the the parameters and the ENSO (El Nino Southern Oscillation).
- A critical look at the number of optimal parameters of the model, and the model development period. Newer statistical methods were also explored in hope of a better performance.

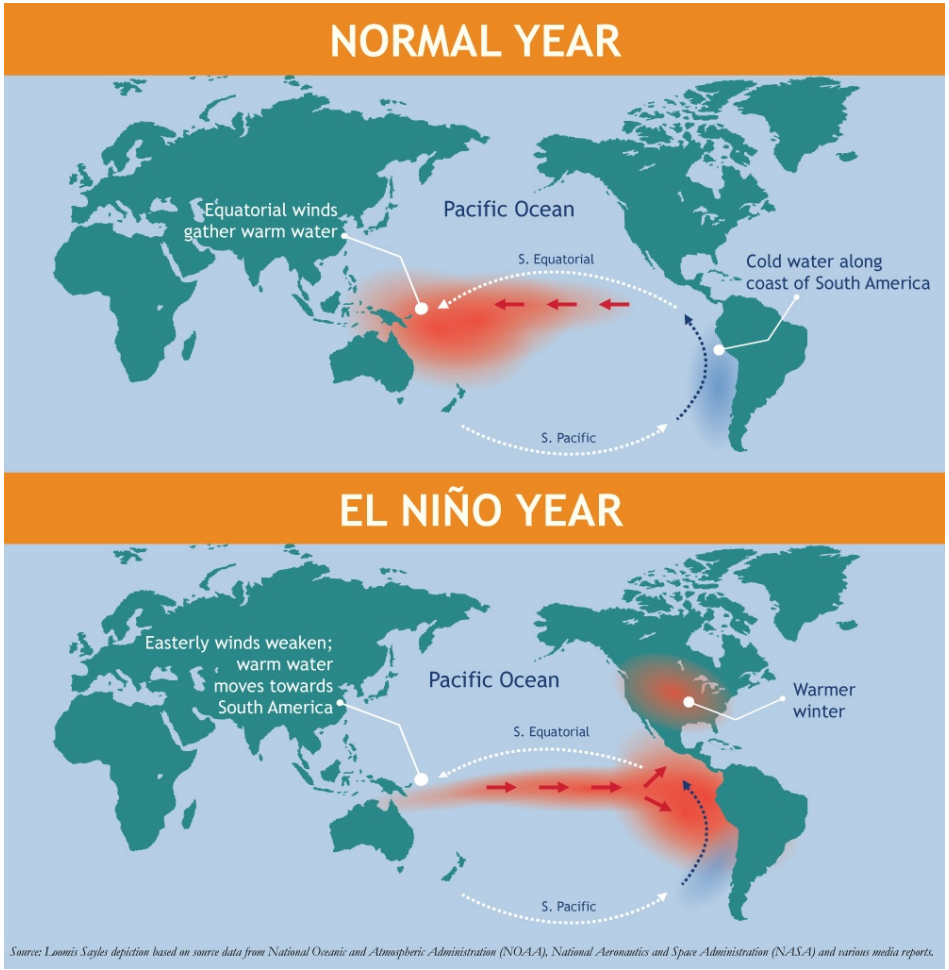

Figure: El Nino Anomaly

Parameters Used

IMD used 6 different parameters for making predictions of the ISMR. These were -

1. North Atlantic Sea Surface Temperature (SST)
2. Equatorial South Eastern Indian Ocean SST anomaly
3. East Asia Surface Pressure Anomaly
4. Europe land surface air temperature anomaly
5. Northwest Europe surface pressure anomaly tendency
6. Warm Water Volume (WWV) Anomaly over the Pacific Ocean

These parameters were included in the model from climatological considerations and judging the correlation between them and the ISMR.

Parameters Used

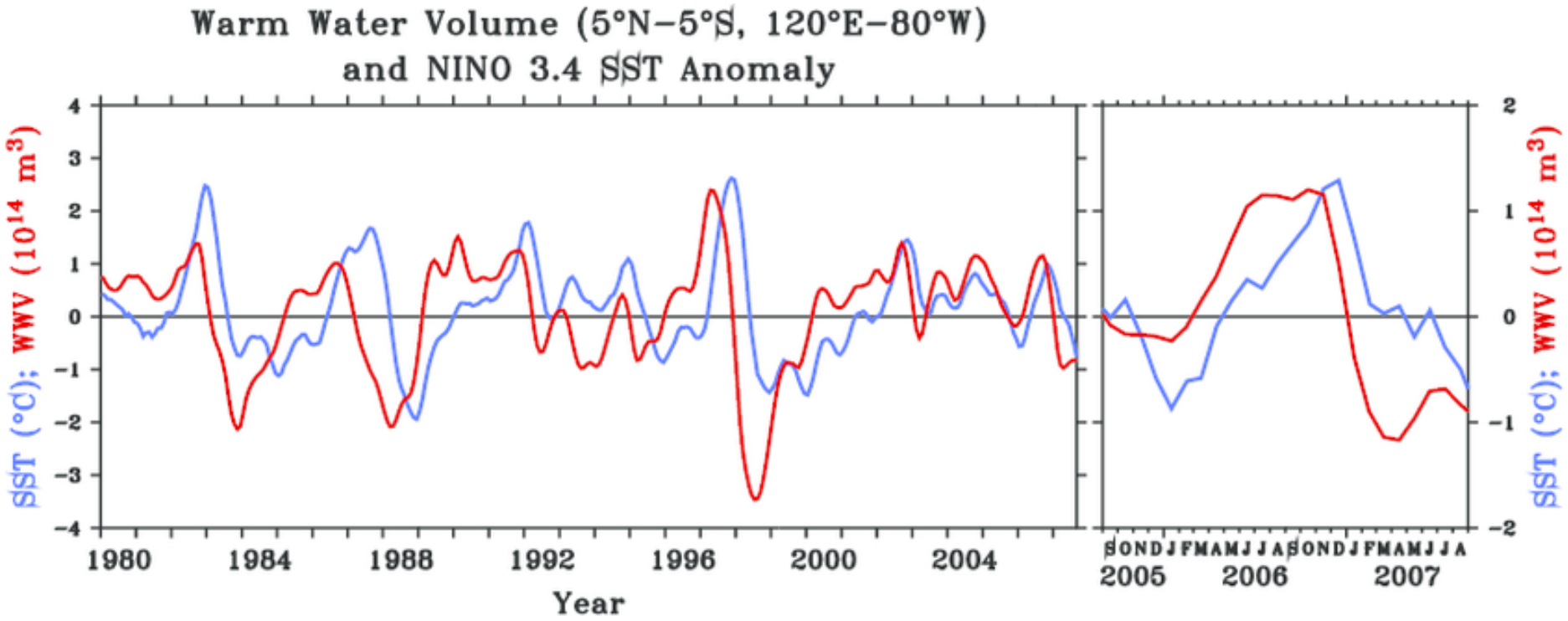

Figure: Correlation between El Nino SST and WWV Anomaly

Indian Ocean SST Anomaly

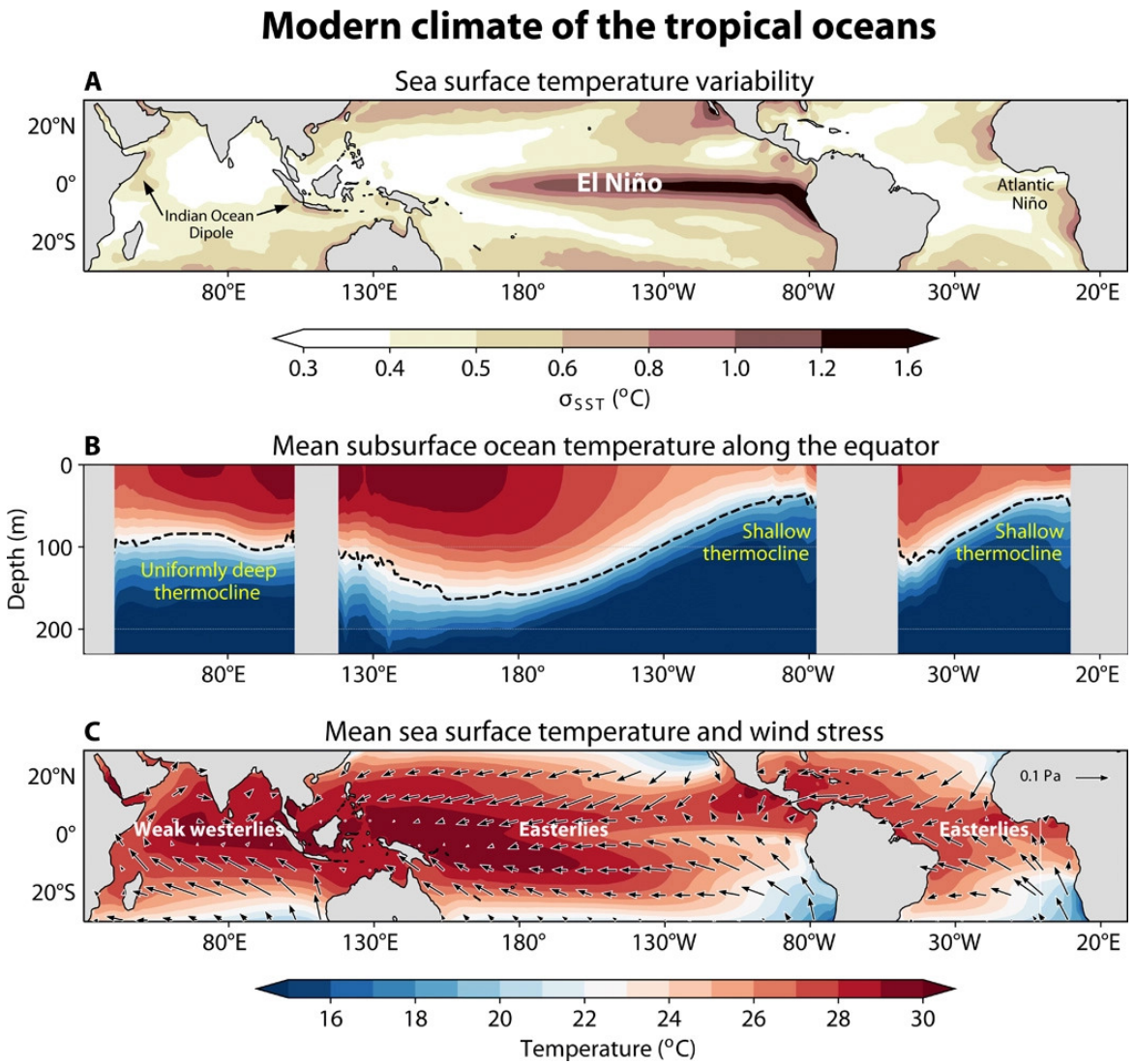

Figure: Indian Ocean SST Anomaly due to the El Nino

Preprocessing and Training

The six various parameters used can give $2^6 - 1$ possible combinations of models, where we can choose any non zero number of parameters for training our model on.
All of the data used was normalised in the 1971-2000 time period.
Training data was used as a 23yr sliding window data, where the model for each year's prediction was trained on previous 23yrs data. This number was experimentally found from backtesting the algorithm.

Conclusions and Inferences

- There are some inherent problems in the statistical models such as epochal variation in the predictand predictor relationship, intercorrelation between the predictors etc.
- During the extreme ISMR years, the predictions were relatively accurate than the previous models.
- Using both parametric (EMR) and non parametric (PPR) models gives us varying results, with some predictions being accurate for a particular model.
- Non parametric nature of PPR is better posed to account for the non linearity that may exist between the features and the target variables.