EH 691-I Impact of solar activity on precipitation in India

Special Topics in Earth Sciences: Data Analysis for Earth System Processes

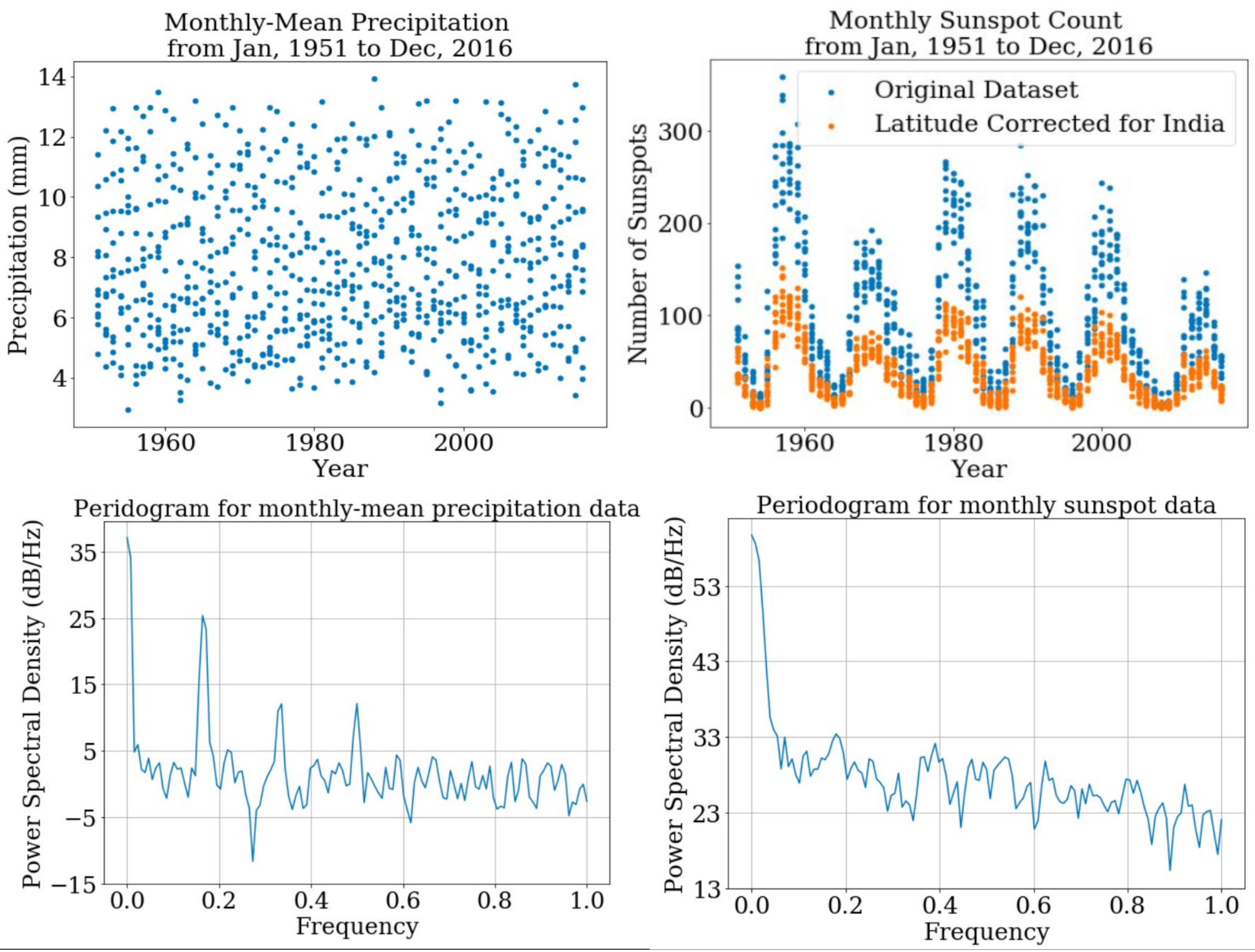
Naman Jain, Pranjali Jain
Discipline of Computer Science and Engineering, IIT Gandhinagar

Introduction

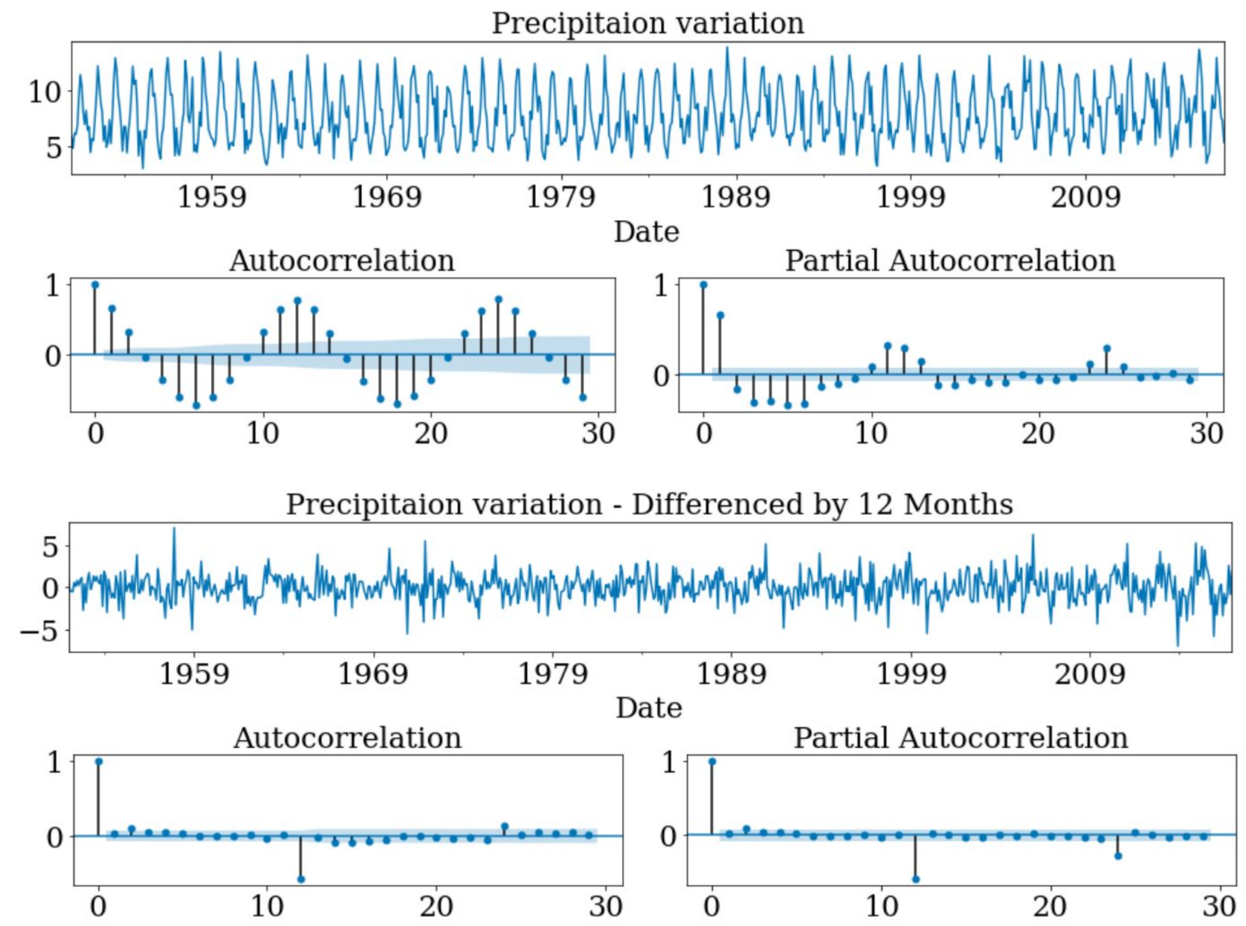
We analyze the relationship between monthly sunspot numbers and average monthly precipitation measured at meteorological stations in India. We check the forecasting abilities of popular time series models on the precipitation data with and without additional information of sun activity.

Dataset

Precipitation data is obtained after taking average of all the locations in India over each month from Jan, 1951 to Dec, 2016. Sunspot data was publicly available in monthly units from Jan, 1749 and Mar, 2017. There are a total of 792 instances of both the variables from Jan, 1951 to Dec, 2016. We split the data into 90:10 for training:validation purposes. The figures below describe the dataset and its periods.



ACF and PACF for Precipitation



Since we plan on fitting ARMA and ARMAX model, we plot the ACF and PACF for precipitation data to obtain model parameters and time steps.

We can clearly see in the graph "Precipitation Variation" that there exists seasonality in data.

To fit a stationary time series model to the precipitation data, we have to remove seasonal trends from the data.

After differencing by 12 months that is the yearly monsoon cycle, we obtain the required parameters.

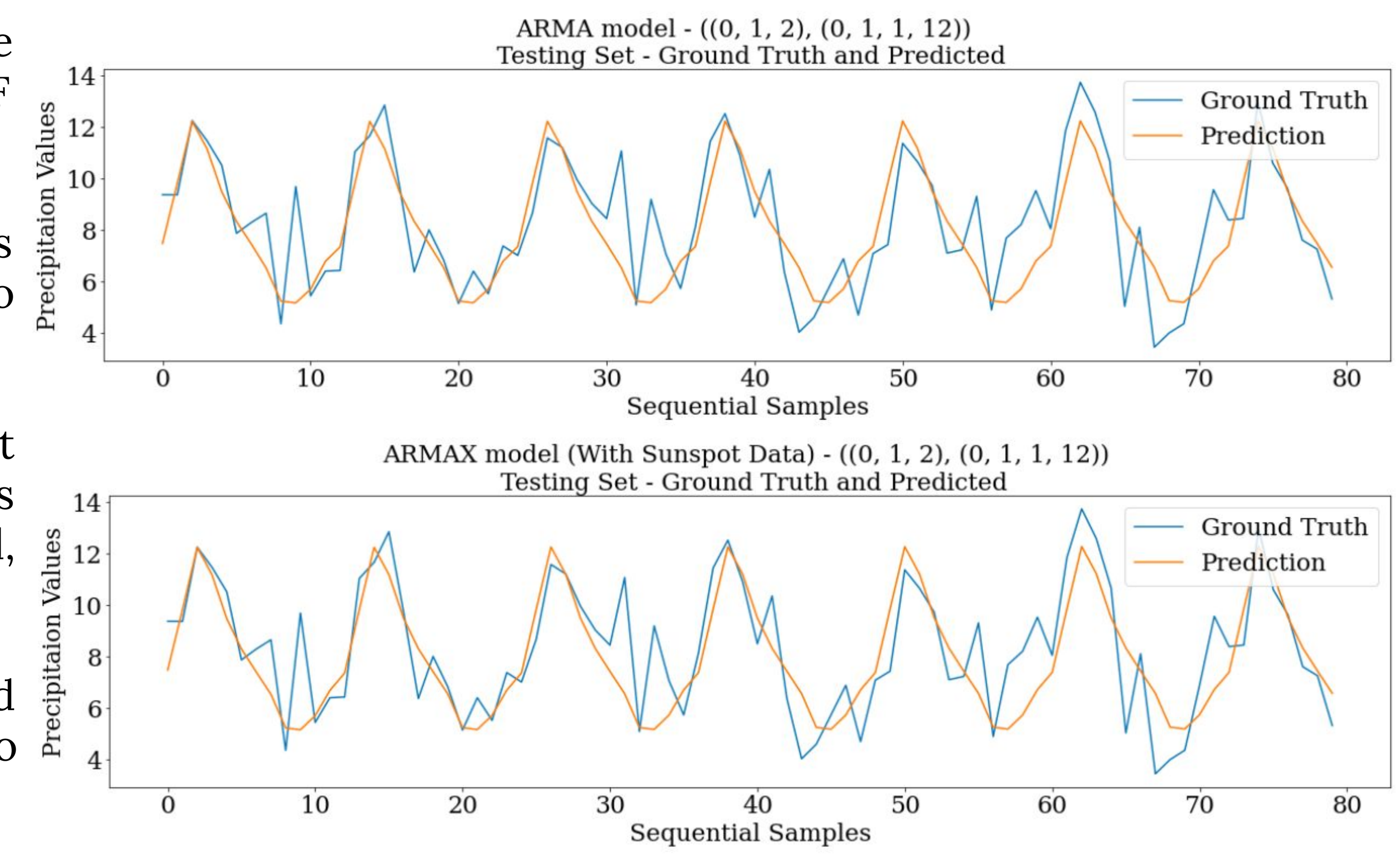
ARMA and ARMAX models

We ran a grid search over the range values near the ACF/PACF values on ARMA model.

We obtained ((2,1,2),(2,1,2,12)) as the set of parameters leading to lowest RMSE score.

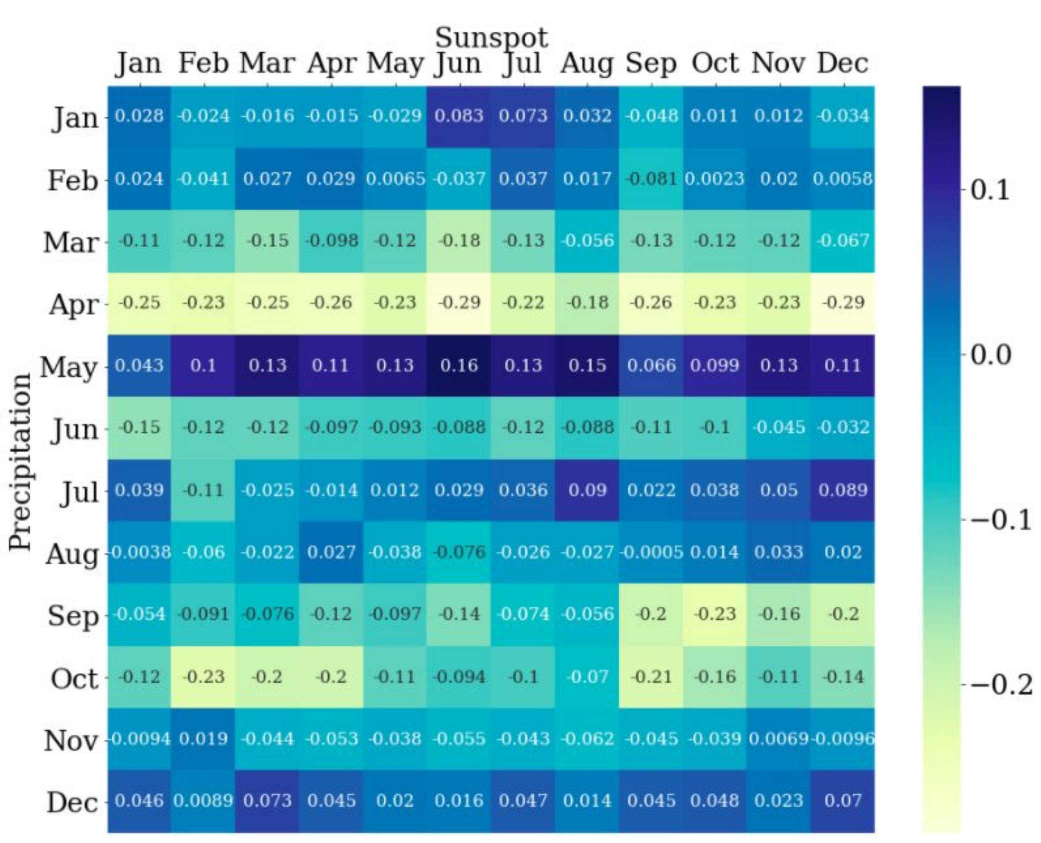
By utilising the sunspot data at various delays as the exogenous variable in the ARMA model, ARMAX model is trained.

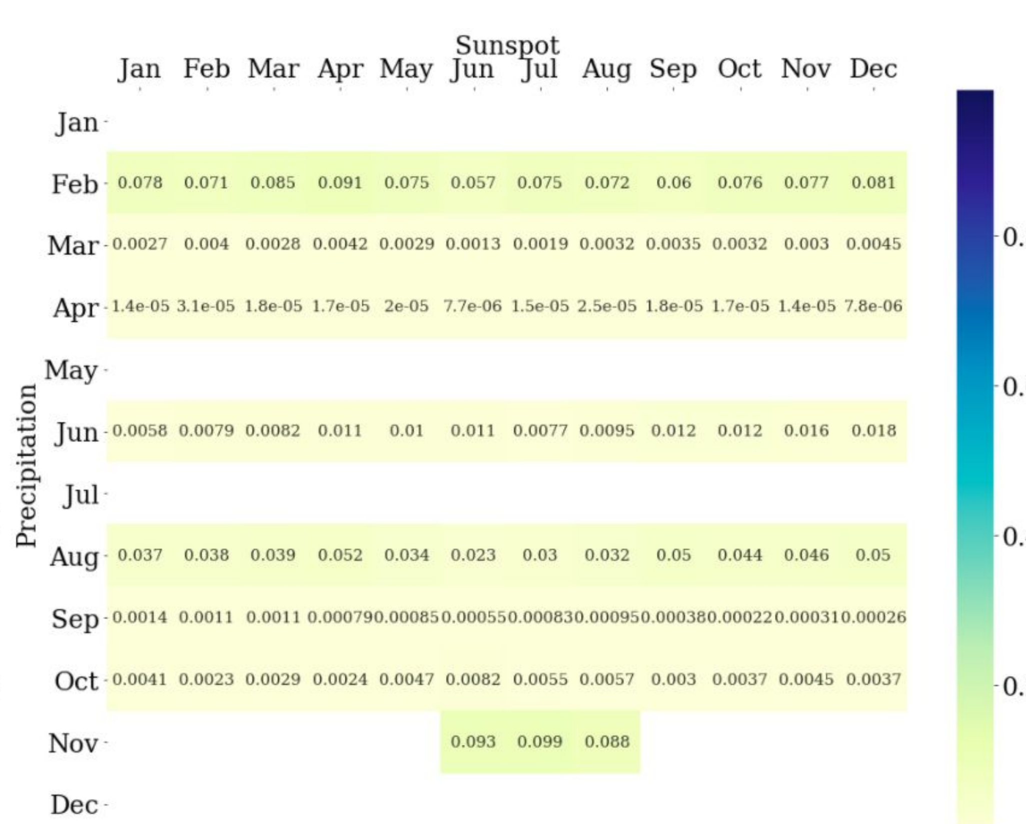
Results for both ARMA and ARMAX models are very close to each other.



Correlation Analysis

The precipitation for some particular months might be significantly correlated with the behavior of the Sun, whereas for other months, it may almost entirely depend on local weather conditions.





Results

	((2, 1, 2), (2, 1, 2, 12))			(((1, 0, 1), (0, 1, 1, 12))			((0, 1, 2), (0, 1, 1, 12))		
Delay (years)	RMSE	AIC	BIC	RMSE	AIC	BIC	RMSE	AIC	BIC
Without Sunspot	1.555927	2204.092869	2245.039726	1.578595	2191.162924	2209.367246	1.557940	2202.677056	2220.875659
1	1.560709	2206.752370	2252.248877	**1.578519**	2193.447501	2216.202902	1.561144	2204.632278	2227.380531
2	1.556083	2206.147934	2251.644441	**1.571512**	2192.326811	2215.082213	**1.556172**	2203.990053	2226.738307
3	**1.555618**	2206.157254	2251.653761	**1.571800**	2192.473747	2215.229149	**1.555774**	2204.091819	2226.840073
4	1.558328	2205.955197	2251.451704	1.581490	2193.104916	2215.860318	1.560295	2204.843223	2227.591477
5	1.570863	2205.415339	2250.911846	1.597997	2191.508016	2214.263418	1.571670	2203.250822	2225.999076
6	1.563448	2206.733322	2252.229830	1.584106	2193.306448	2216.061850	1.564433	2204.774775	2227.523028
7	1.577077	**2203.657547**	2249.154055	1.600798	**2190.299145**	2213.054546	1.578313	**2201.565827**	2224.314081
8	1.561995	2206.543221	2252.039728	1.582670	2192.866225	2215.621627	1.562652	2204.454803	2227.203057
9	**1.555764**	2206.075585	2251.572093	**1.578467**	2193.697675	2216.453076	1.558589	2204.860825	2227.609079
10	1.560790	2206.832527	2252.329034	1.580051	2193.012870	2215.768272	1.560885	2204.687900	2227.436154
11	1.576966	**2203.991911**	2249.488418	1.593826	**2190.367666**	2213.123068	1.576906	**2201.960011**	2224.708265

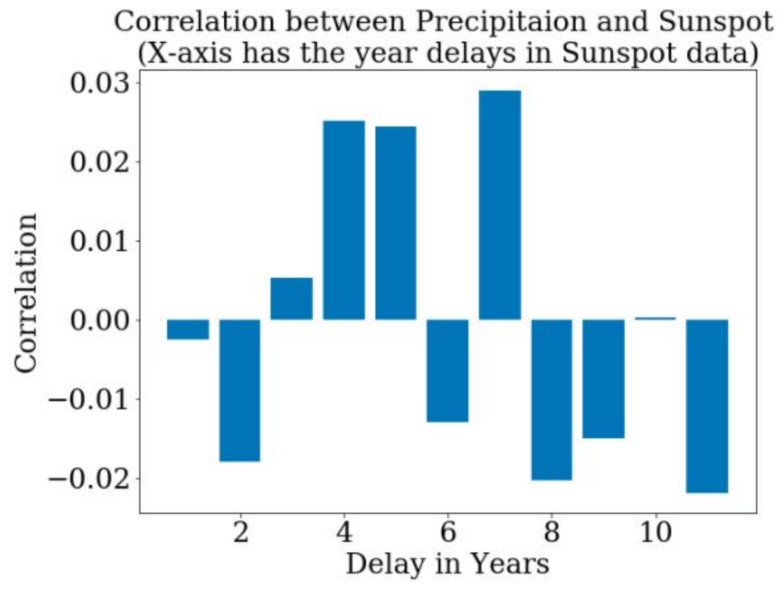
The above table consists of results of the top three parameter sets in terms of RMSE score when we applied gridsearch over the model parameters.

Interestingly, Sunspot data helps in better predictions with a delay of **2, 3** and **9**, in terms of RMSE and with a delay of **7** and **11** in terms of AIC scores.

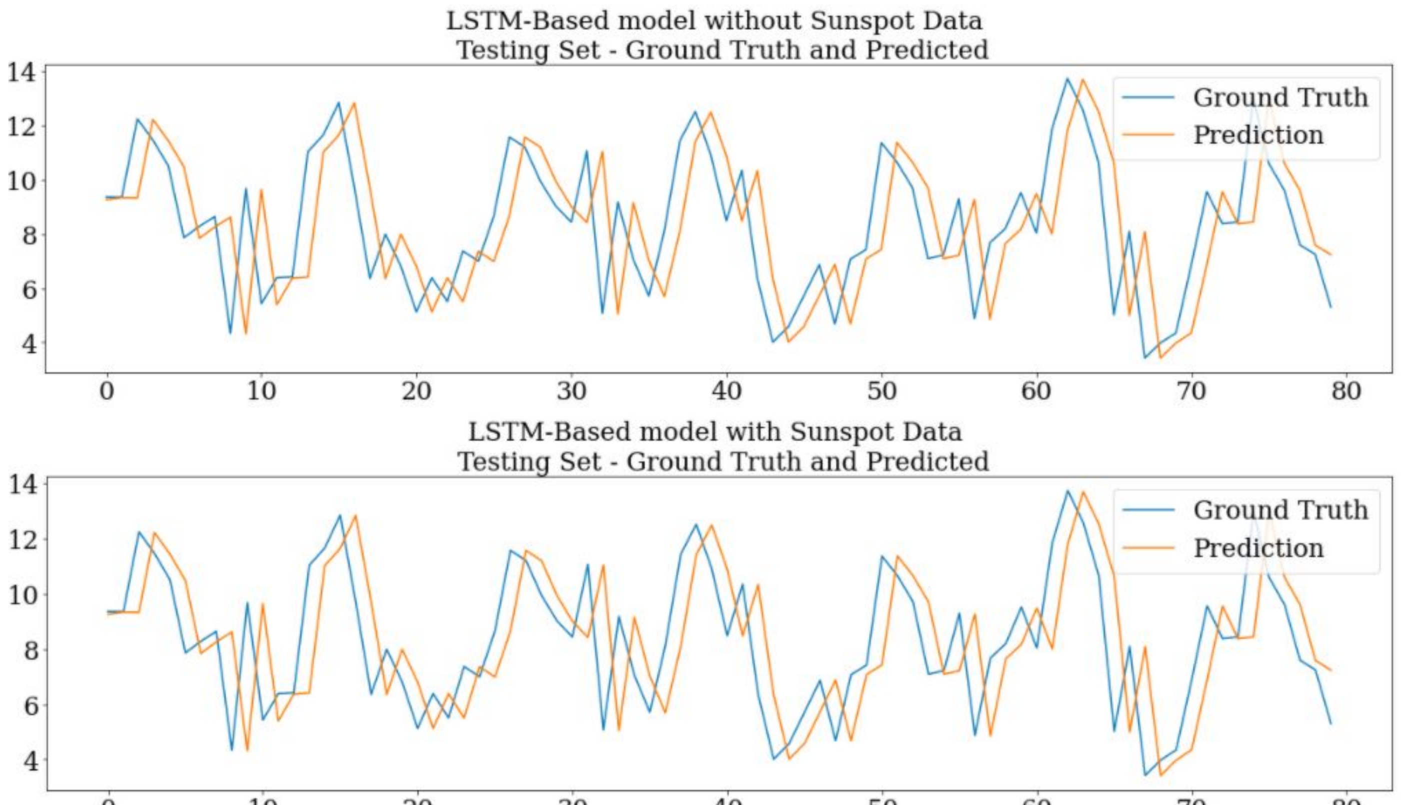
The maximum decrease of **0.13%** in RMSE value is caused by including sunspot data as the exogenous variable in the model.

Optimal time delay

The optimal delay came out to be 7 years in Sunspot data when correlated with the precipitation data of India.



LSTM model



ARMA and ARMAX are linear models who are able to learn the distributions correctly.

We tried adding non-linearity using LSTMs, widely known for their application in sequences and time-series.

We used Multivariate LSTM to account one-step prediction with sunspot data.

The RMSE of both models is very near ~2.45. In LSTM based model, sunspot data is worsening the prediction.

Conclusion

In linear models, sunspot or sun activity is influencing precipitation at various time delays.

LSTMs do not work mainly because of lack of context, lack of data or because it considers sunspot data as noise.

In future, we plan to explore existing models on daily data and other prevalent models.

Reference: Nitka, W. and Burnecki, K., 2019. Impact of solar activity on precipitation in the United States. Physica A: Statistical Mechanics and its Applications, 527, p.121387.