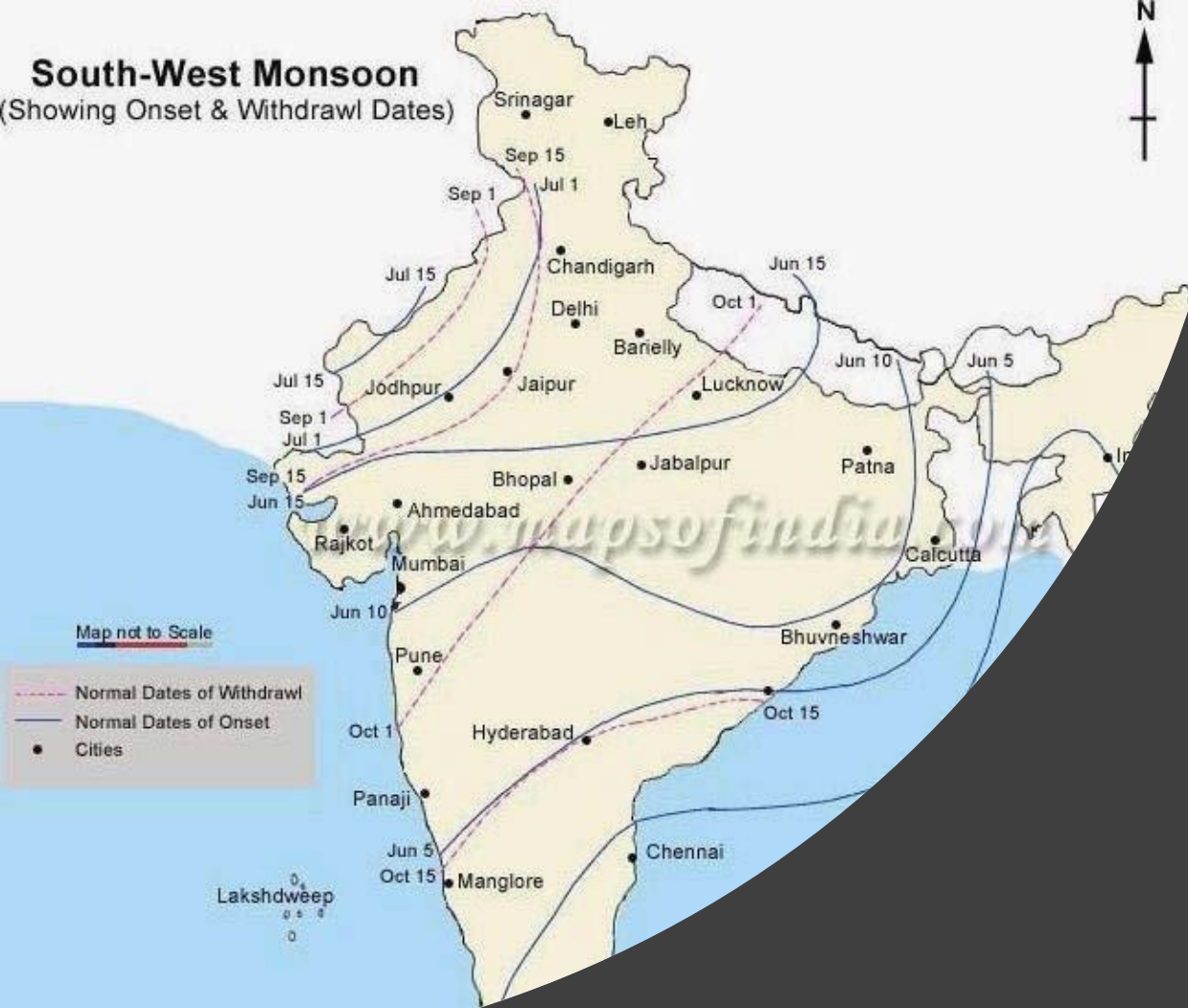# Data Preparation

For Kerala, in the paper - Concept Drift Detection in Data Stream Clustering and its Application on Weather Data

South-West Monsoon
(Showing Onset & Withdrawl Dates)

Experiments are conducted to study the clustering behavior of weather data and its relationship to the physical phenomena occurring in the environment.

Onset and withdrawal of south-west Monsoon and thunderstorms occurrences are the weather phenomena selected for the study.

Data Used in the Paper…

In **Cochin, Kerala**, normally the onset of southwest monsoon happens during *June first week*, and it withdraws during the *beginning of October*.

Also, unpredictable rain and thunderstorm are common during the *summer season*.

Data collected in the *years 2016 and 2017* are used for this study.

WORK DONE ON DATA

- The Min-Max procedure is performed to normalize the data values to the range [0,1].

- The first model will be created in an offline way using the initial 5000 samples of the stream. K-means clustering is the algorithm used in this stage.

- Thereafter, the model is updated online, along with the stream.

- Online maintenance of micro-clusters in CluStream algorithm makes sure that the total number of micro-clusters in main memory does not exceed a predetermined limit at any point of time.

- In these experiments, this limit is fixed to 200.

Automatic Weather Station in ACARR collects most of the weather parameters at three different height levels: 2 meters, 20 meters, and 30 meters. Hence, there is a redundancy in information as far as prediction is concerned.

All these parameters might not be relevant for prediction. A dimensionality reduction technique was applied in the beginning to find out the most relevant dimensions or parameters.

Parameters helpful for the -

Rain related studies	Thunderstorm related studies
• Temperature • Cloud Radiation • Solar Radiation • Net Radiation • Wind Speed • Wind Direction	• Temperature • Humidity • Wind speed

Onset of South-West Monsoon
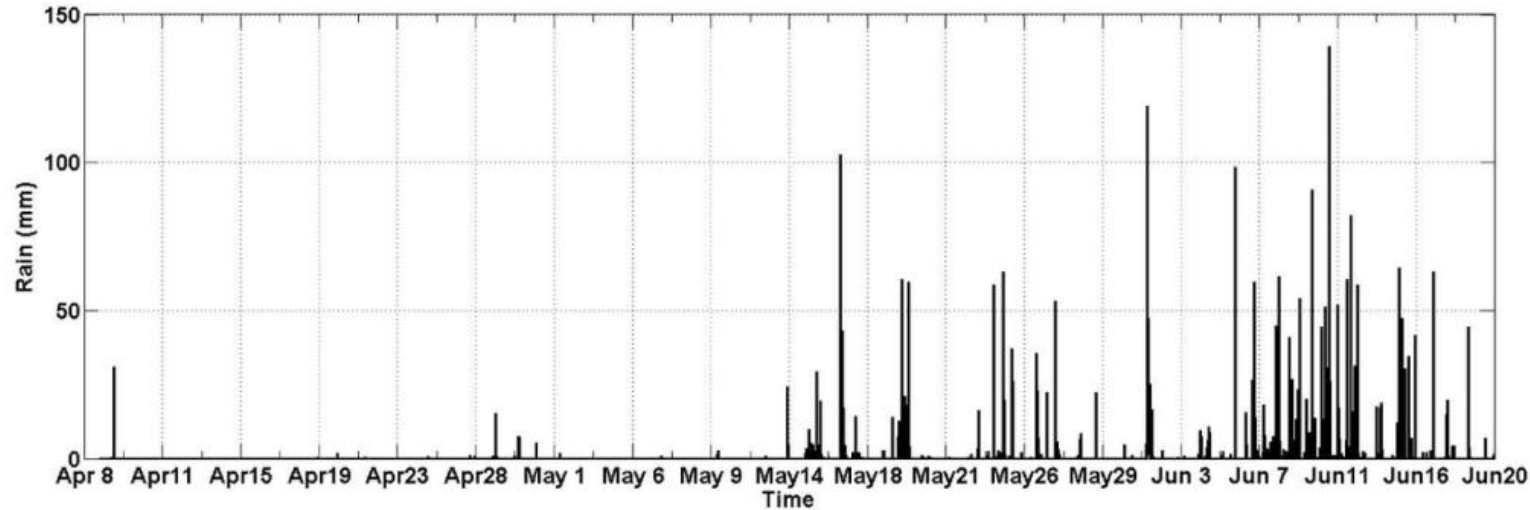
As it is evident from this figure, there is a pre-monsoon shower starting from May 14th onwards. Temperature, Solar radiation and most of the weather parameters show a significant difference in distribution during this period.

A change in all the weather parameters can be observed in the week of 16th to 20th May. During the same time, Pre-Monsoon shower started and it prolonged till the monsoon onset during June first week.

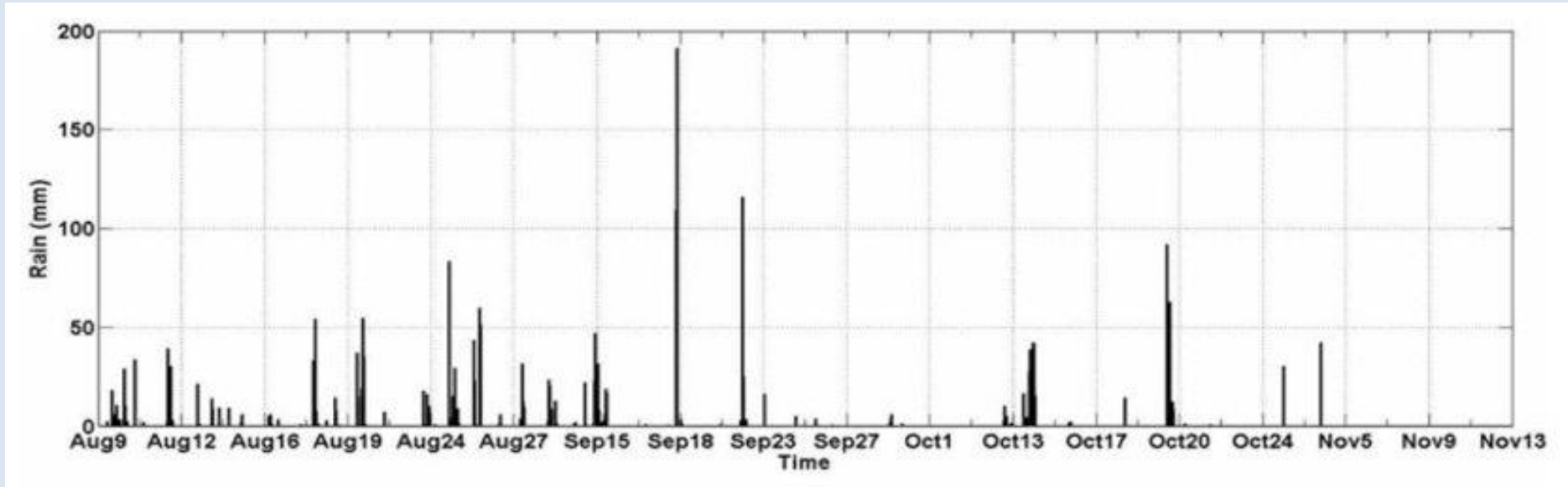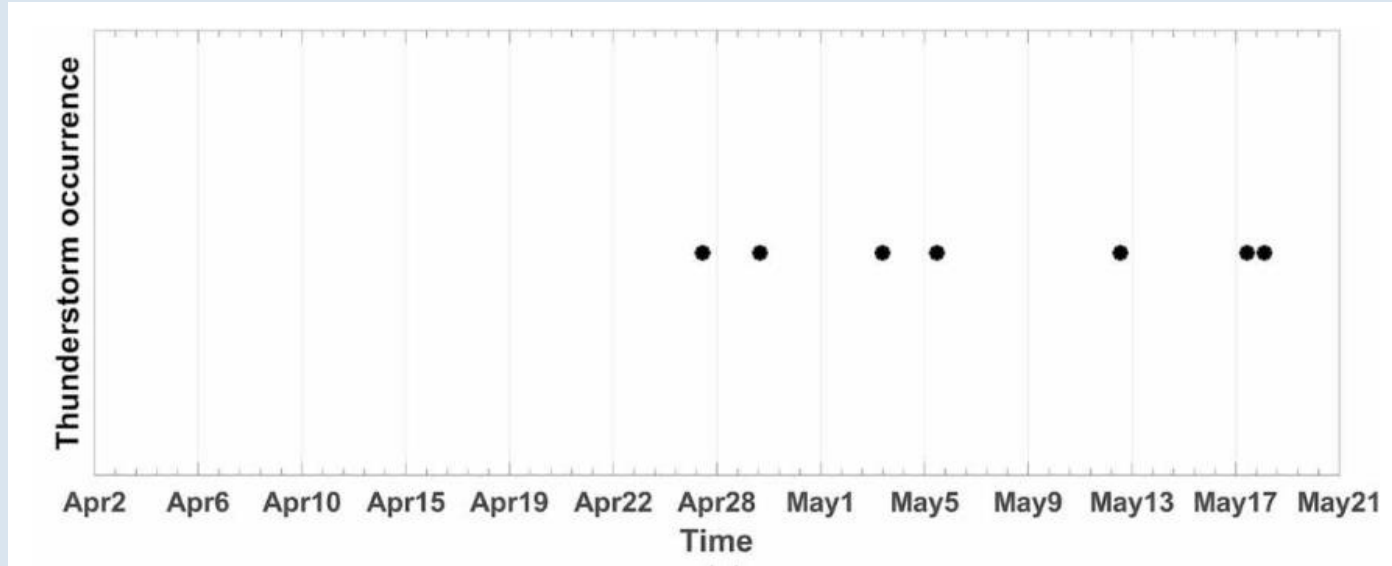# Withdrawal of South-West Monsoon

Figure shows the rainfall received during the southwest monsoon withdrawal period. Concept changes are found relatively frequent compared to the monsoon onset period. These results also support the fact that a significant change in clustering is an indication of a significant weather event.

Thunderstorm

In Kerala, during the summer season, it is common to get unpredicted rainfall with thunderstorm and lightning. Weather parameters show considerable changes during thunderstorms as well. Figure shows the occurrence of thunderstorms during the summer season. Black dots in the figure denote thunderstorm observations.

Temperature, humidity and wind speed are identified to be the most prominent weather parameters that show variations during a thunderstorm. Datastream containing these three parameters is taken to check the clustering structure behavior and its relation to the thunderstorm occurrence.

Thank you

Fahed Shaikh

Shaikh.Fahed@iiitb.ac.in