Enhancing Rainfall Forecast Accuracy in Urban Monsoon Systems Using an MSSA-LSTM Hybrid Model

Priya Shejule [1*](spriya@iitg.ac.in), **Sreeja Pekkat** [2],

[1*,2] Department of Civil Engineering, Indian Institute of Technology Guwahati, India

Introduction

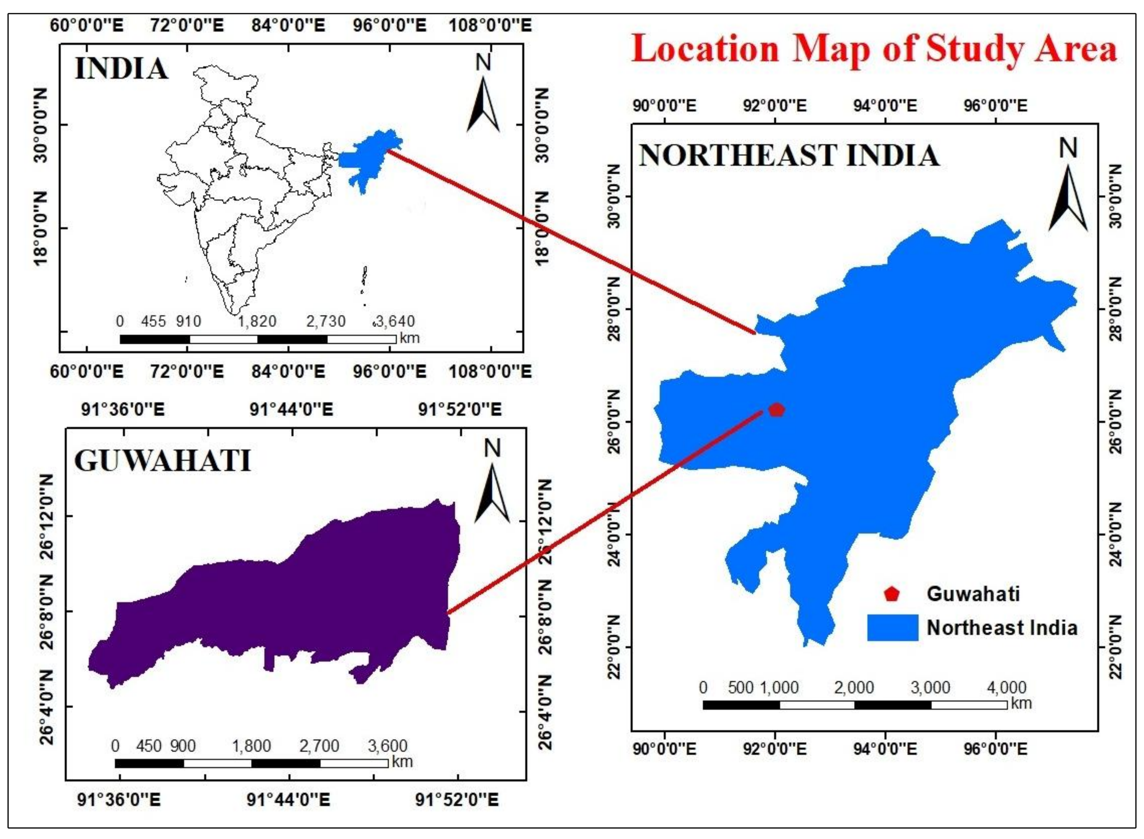
- Accurate and reliable rainfall forecasting is critical for mitigating hydrometeorological disasters such as floods and droughts, particularly in flood-prone regions like Guwahati, Northeast India.
- Complex topography enhancing orographic lifting and humid subtropical climate of the city amplify its vulnerability to extreme monsoon events.
- Rainfall, closely linked to hydrometeorological disasters (Panda and Sahu, 2019).
- Forecasting hourly rainfall at synoptic scales remains challenging due to its non-linear dynamics, non-stationary behavior, and multiscale variability, compounded by noise in meteorological data.
- This study addresses the predictability of synoptic monsoon processes by proposing a hybrid model employing Multivariate Singular Spectrum Analysis (MSSA) and a Long Short-Term Memory (LSTM) network for multistep ahead hourly rainfall forecast

Objective

- To identify nature of noise in the meteorological data under consideration
- To investigate the interdependency between rainfall and meteorological variables to understand the factors influencing rainfall dynamics using hybrid approach of forecast integrating multivariate SSA and LSTM, i.e., MSSA–LSTM

Study Area and Data Collection

- Guwahati city, North-east of India
- Data type: Indian Monsoon Data Assimilation and Analysis (IMDAA) gridded hourly data available at https://rds.ncmrwf.gov.in Duration: 2015-2019



- Longitude- 91º34'25" and 91º52'00" E;Latitude- 26º4'45" and 26º13'25" N
- Mean annual rainfall- 1698 mm (Shejule and Pekkat 2024)
- Climate zone- Humid subtropical climate

Figure 1: Study area

Methodology Flowchart

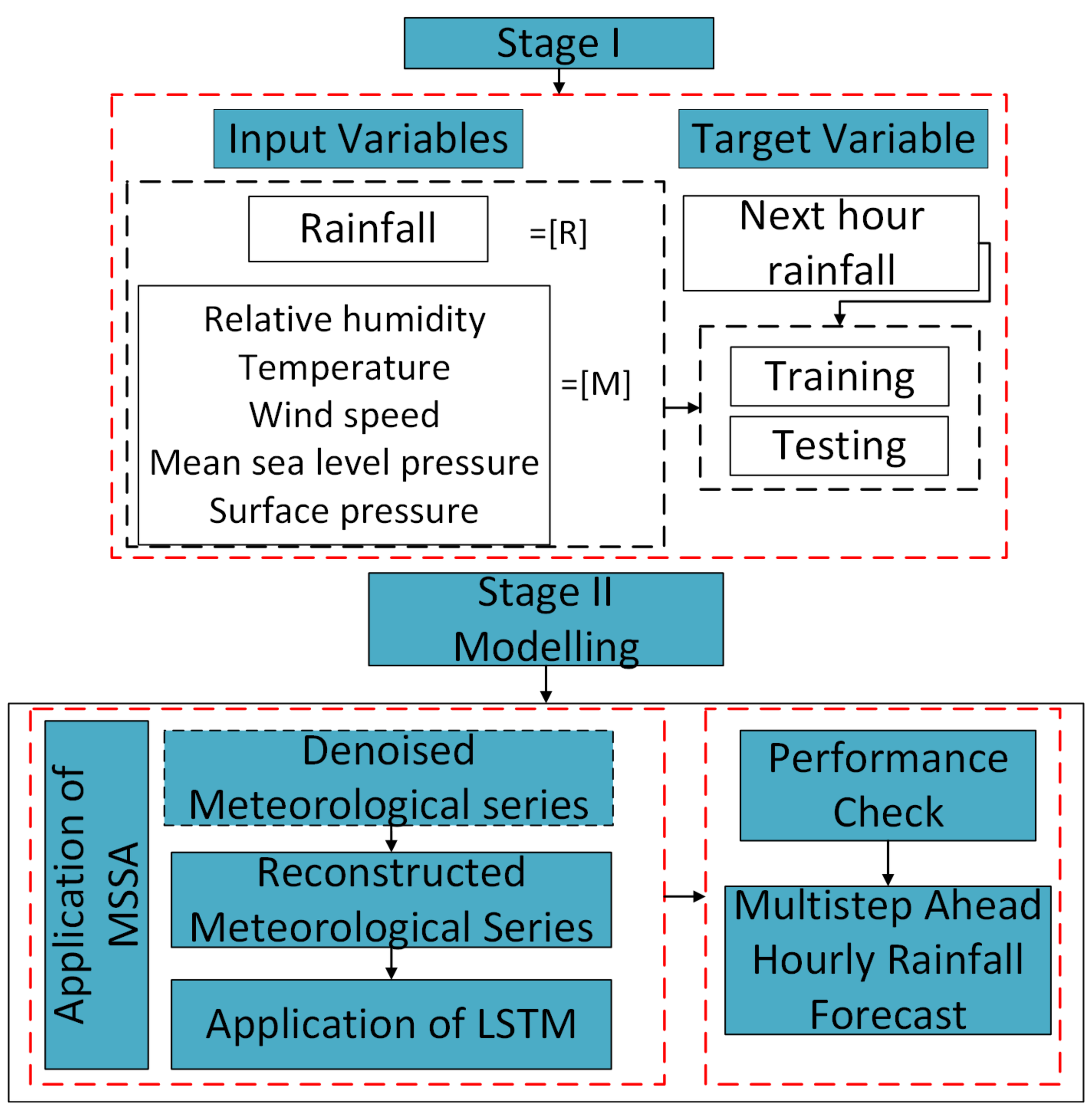


Figure 2: Flowchart of methodology

Results

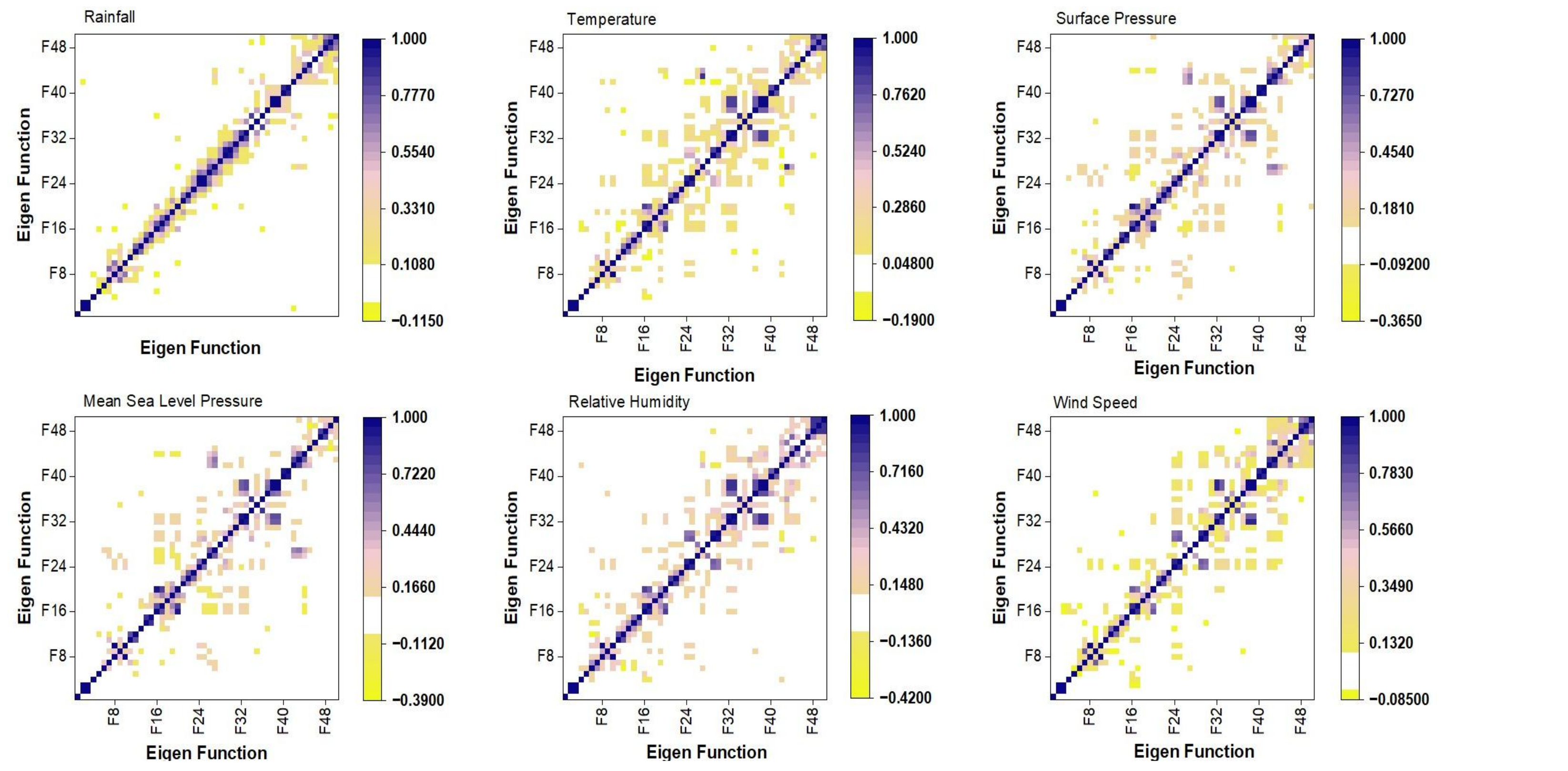


Figure 3: W-correlation for the first 50 reconstructed components

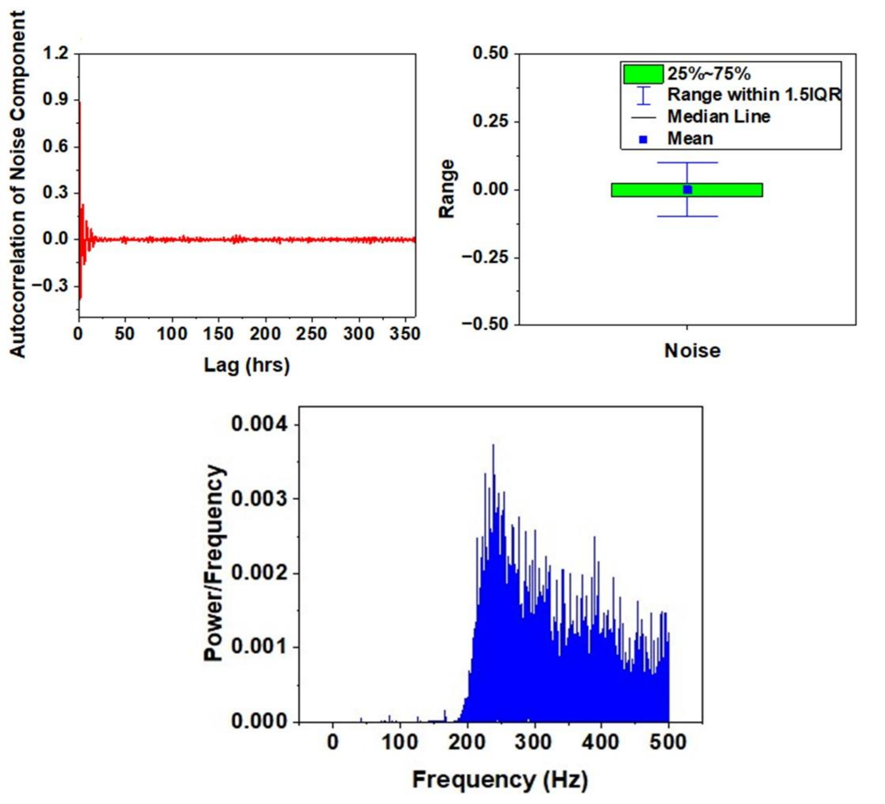


Figure 4: Correlogram plot, box plot, power spectrum of a noise component

- The correlogram plot of red noise exhibits a slow decline as the lag increases, while that of white noise rapidly diminishes (Elsner and Tsonis 1996).
- The power spectrum plot, Fig 4., shows that the power of a signal gradually decreases at higher frequencies.
- The reconstructed meteorological series are fed to the LSTM model for training purposes.

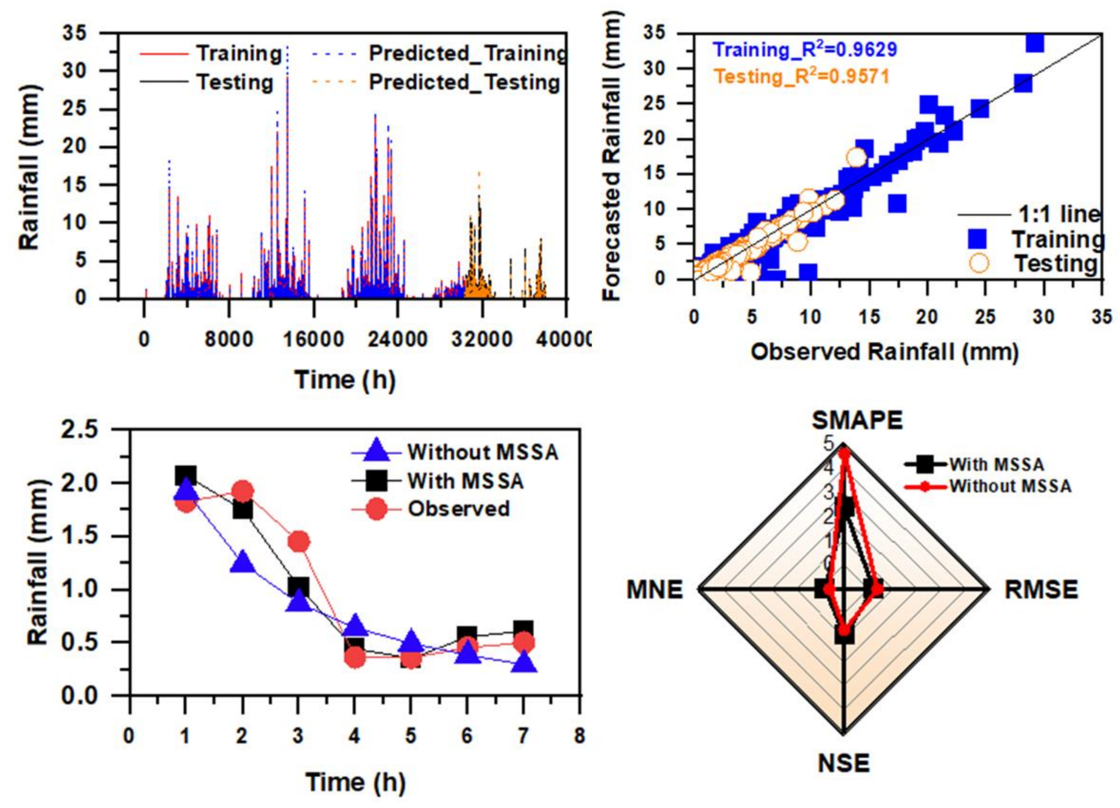


Figure 5: Hourly observed and forecasted rainfall by hybrid MSSA-LSTM model for the period (a) Jan 03, 2015- Apr 29, 2019(b) scatterplot of observed and model forecast during training and testing period (b) Observed and model forecast for next one hour for the period Apr 29, 2019, 17:00:00 to Apr 29, 2019, 23:00:00 (d) Radar chart showing the performance skill of LSTM model in hourly rainfall forecast with and without pre-processing by MSSA

Summary

- The performance of the LSTM model significantly improves as the zero values in the observed data are eliminated after reconstruction by MSSA. In general, the developed hybrid systems outperformed the single model.
- Presence of red noise is detected in the given time series.
- The hybrid systems employing the MSSA-LSTM achieved, on average, a percentage gain compared to its respective single LSTM model of 47.99%, 43.88% for SMAPE, RMSE in the case of the Guwahati city.

References

Elsner, J. B., and A. A. Tsonis, 1996: Singular spectrum analysis: a new tool in time series analysis. *Springer Science & Business Media.*

Panda, A., and N. Sahu, 2019: Trend analysis of seasonal rainfall and temperature pattern in Kalahandi, Bolangir and Koraput districts of Odisha, *India. Atmos. Sci. Lett.*, 20(10), https://doi.org/10.1002/asl.932

Shejule, P. A., & Pekkat, S. (2024). Hybrid Model for Multistep-Ahead Rainfall Forecast in Northeast India: A Comparative Study. *Journal of Hydrometeorology*, 25(8), 1221-1236.