**RAINFALL PREDICTION USING MACHINE LEARNING**

B. Priyanka

*KV SubbaReddy Engineering College, Kurnool, A.P, India*

Ch. SriLakshmi prasanna, G. Jhansi, C. Pavani, K. Lakshmi

*KV SubbaReddy Engineering College, Kurnool, A.P, India*

# 

***Abstract***: India Meteorological Department has implemented state level medium range rainfall forecast system applying multi model ensemble technique, making use of model outputs of state-of-the-art global models from the five leading global NWP centers. The pre-assigned grid point weights on the basis of anomaly correlation coefficients (CC) between the observed values and forecast values are determined for each constituent model utilizing two season datasets and the multi model ensemble forecasts are generated at the same resolution on a real-time basis. The ensemble forecast fields are then used to prepare forecasts for each state, taking the average value of all grid points falling in a particular district. In this paper, we describe the development strategy of the technique and performance skill of the system during 15 years of rain fall at different states in india. The study demonstrates the potential of the system for predicting future rainfall forecasts for upcoming years and scale over Indian region. District wise performance of the ensemble rainfall forecast reveals that the technique, in general, is capable of providing reasonably good forecast skill over most states of the country, particularly over the states where the monsoon systems are more dominant.

**I. INTRODUCTION**

There has been long demand from the user community for district level quantitative weather forecasts in short to medium range time scale. The quantitative rainfall forecast for smaller spatial distribution such as district level over highly complex inhomogeneous region like India is a very challenging task. For the generation of district level quantitative rainfall forecasts, one has to depend on the forecasts from dynamical Numerical Weather Prediction (NWP) models. During the last two decades, NWP methods have acquired greater skills and are playing an increasingly important role in the operational weather forecasting. But rainfall prediction skill of NWP models is still not adequate to satisfactorily address detailed aspects of Indian summer monsoon. This is because of large spatial and temporal variability of rainfall and some inherent limitations of NWP models. There are various factors like topography, prevailing synoptic situation and its interaction with mesoscale systems, lack of observations, etc., are some of the key factors which pose difficulties for numerical weather prediction of any region, and so Indian region is not an exception. Considering the need of farming sector, India Meteorological Department (IMD) has upgraded the Agro-Meteorological Advisory Service from agro climate zone to district level. As a major step, IMD started issuing district level weather forecasts from 1 June 2008 for meteorological parameters such as rainfall, maximum and minimum temperature, relative humidity, surface wind and cloud octa up to 5 days in quantitative terms (Roy Bhowmik et al 2009). These forecasts are generated through multi-model ensemble (MME) system making use of model outputs of state-ofthe-art three global models from the leading global NWP centres. These forecasts are made available on the national website of IMD (www.imd.gov.in). During summer monsoon 2009, the number of ensemble members is increased from three to five. In the present study, we describe the development strategy of the MME technique, used for high resolution rainfall forecasts over Indian region and demonstrate the prediction skill of the technique during summer monsoon 2009. In our previous study (Roy Bhowmik and Durai 2008, 2010), performance skill of MME at 50 km horizontal resolution for district level short range rainfall forecasts during summer monsoon 2007 was demonstrated from the use of four coarser grid models namely (i) IMD limited area model at 75 km horizontal resolution, (ii) IMD MM5 at 45 km horizontal resolution, (iii) National Centre for Medium Range Weather Forecasting (NCMRWF) MM5 at 30 km resolution, and (iv) NCMRWF T-80 (grid space ∼156 over the tropics). At 50 km resolution, MME could cover only 250 districts. The encouraging results of the study motivated authors for further research to increase the forecast period and model resolution using improved rainfall outputs of state-of-the-art high resolution global models from leading NWP centres to meet the operational requirement of farming community.

## **II. LITERATURE SURVEY**

**Improved weather and seasonal climate forecasts from multimodel super ensemble**

India Meteorological Department has implemented district level medium range rainfall forecast system applying multimodel ensemble technique, making use of model outputs of state-of-the-art global models from the five leading global NWP centres. The pre-assigned grid point weights on the basis of anomaly correlation coefficients (CC) between the observed values and forecast values are determined for each constituent model at the resolution of 0.25° ×0.25° utilizing two season datasets (1 June–30 September, 2007 and 2008) and the multimodel ensemble forecasts (day-1 to day-5 forecasts) are generated at the same resolution on a real-time basis. The ensemble forecast fields are then used to prepare forecasts for each district, taking the average value of all grid points falling in a particular district. In this paper, we describe the development strategy of the technique and performance skill of the system during summer monsoon 2009. The study demonstrates the potential of the system for improving rainfall forecasts at five days time scale over Indian region. Districtwise performance of the ensemble rainfall forecast reveals that the technique, in general, is capable of providing reasonably good forecast skill over most states of the country, particularly over the states where the monsoon systems are more dominant.

**Improving tropical precipitation forecasts from a multi analysis super ensemble**

This paper utilizes forecasts from a multianalysis system to construct a superensemble of precipitation forecasts. This method partitions the computations into two time lines. The first of those is a control (or a training) period and the second is a forecast period. The multianalysis is derived from a physical initialization–based data assimilation of “observed rainfall rates.” The different members of the reanalysis are produced by using different rain-rate algorithms for physical initialization. The basic rain-rate datasets are derived from satellites’ microwave radiometers, including those from the Tropical Rainfall Measuring Mission (TRMM) satellites and the Special Sensor Microwave Imager (SSM/I) data from three current U.S. Air Force Defense Meteorological Satellite Program (DMSP) satellites. During the training period, 155 experiments were conducted to find the relationship between forecasts from the multianalysis dataset and the best “observed” estimates of daily rainfall totals. This relationship is based on multiple regression and defined by statistical weights (which vary in space.) The forecast phase utilizes the multianalysis forecasts and the statistics from the training period to produce superensemble forecasts of daily rainfall totals. The results for day 1, day 2, and day 3 forecasts are compared to various conventional forecasts with a global model. The superensemble day 3 forecasts of precipitation clearly have the highest skill in such comparisons.

**Experimental realtime multi-model ensemble (MME) prediction of rainfall during monsoon 2008: Large scale medium range aspects**

Realistic simulation/prediction of the Asian summer monsoon rainfall on various space–time scales is a challenging scientific task. Compared to mid-latitudes, a proportional skill improvement in the prediction of monsoon rainfall in the medium range has not happened in recent years. Global models and data assimilation techniques are being improved for monsoon/tropics. However, multi-model ensemble (MME) forecasting is gaining popularity, as it has the potential to provide more information for practical forecasting in terms of making a consensus forecast and handling model uncertainties. As major centers are exchanging model output in near real-time, MME is a viable inexpensive way of enhancing the forecasting skill and information content. During monsoon 2008, on an experimental basis, an MME forecasting of large-scale monsoon precipitation in the medium range was carried out in real-time at National Centre for Medium Range Weather Forecasting (NCMRWF), India. Simple ensemble mean (EMN) giving equal weight to member models, bias-corrected ensemble mean (BCEMn) and MME forecast, where different weights are given to member models, are the products of the algorithm tested here. In general, the aforementioned products from the multi-model ensemble forecast system have a higher skill than individual model forecasts. The skill score for the Indian domain and other sub-regions indicates that the BCEMn produces the best result, compared to EMN and MME. Giving weights to different models to obtain an MME product helps to improve individual member models only marginally. It is noted that for higher rainfall values, the skill of the global model rainfall forecast decreases rapidly beyond day-3, and hence for day-4 and day-5, the MME products could not bring much improvement over member models. However, up to day-3, the MME products were always better than individual member models.

**High resolution daily gridded rainfall data for Indian Region: Analysis of break and active monsoon spells**

In this communication, we discuss the development of a very high resolution (0.5° x 0.5°) daily rainfall data-set for mesoscale meteorological studies over the Indian region. The dataset was developed using quality-controlled rainfall data from more than 3000 rain gauge stations over India. The analysis consists of daily rainfall data for all the seasons for the period 1971-2005. A well-tested interpolation method (Shepard's method) was used to interpolate the station data into regular grids of 0.5° x 0.5° lat. x long. After proper validation, it has been found that the present dataset is better compared to other available datasets. A few case studies have been shown to demonstrate the utility of the dataset for different mesoscale meteorological analyses. However, since the data density is not kept uniform, there is a possibility of temporal inhomogeneity and therefore, the present dataset cannot be used for trend analysis. The dataset is freely available from the India Meteorological Department, Pune.

**Rainfall analysis for Indian monsoon region using the merged rain gauge observations and satellite estimates**

Objective analysis of daily rainfall at the resolution of 1° grid for the Indian monsoon region has been carried out merging dense land rainfall observations and INSAT derived precipitation estimates. This daily analysis, being based on high dense rain gauge observations was found to be very realistic and able to reproduce detailed features of Indian summer monsoon. The inter-comparison with the observations suggests that the new analysis could distinctly capture characteristic features of the summer monsoon such as north-south oriented belt of heavy rainfall along the Western Ghats with sharp gradient of rainfall between the west coast heavy rain region and the rain shadow region to the east, pockets of heavy rainfall along the location of monsoon trough/low, over the east central parts of the country, over north-east India, along the foothills of Himalayas and over the north Bay of Bengal. When this product was used to assess the quality of other available standard climate products (CMAP and ECMWF reanalysis) at the gird resolution of 2.5°, it was found that the orographic heavy rainfall along Western Ghats of India was poorly identified by them. However, the GPCC analysis (gauge only) at the resolution of 1° grid closely discerns the new analysis. This suggests that there is a need for a higher resolution analysis with adequate rain gauge observations to retain important aspects of the summer monsoon over India. The case studies illustrated show that the daily analysis is able to capture large-scale as well as mesoscale features of monsoon precipitation systems. This study with data of two seasons (2001 and 2003) has shown sufficiently promising results for operational application, particularly for the validation of NWP models.

**III.EXISTING SYSTEM**

In our previous studies conducted by Roy Bhowmik and Durai (2008, 2010), an attempt was made to evaluate the performance of a Multi-Model Ensemble (MME) forecasting system designed for district-level short-range rainfall prediction during the southwest monsoon season of 2007. The MME was configured at a horizontal resolution of 50 km and aimed to improve the accuracy and reliability of rainfall forecasts over the Indian region. This system utilized the combined outputs of four numerical weather prediction (NWP) models with varying horizontal resolutions and dynamical cores. These included: (i) the India Meteorological Department (IMD) Limited Area Model with a spatial resolution of 75 km, (ii) the IMD implementation of the MM5 (Mesoscale Model version 5) operating at a finer resolution of 45 km, (iii) the NCMRWF (National Centre for Medium Range Weather Forecasting) MM5 model at a resolution of 30 km, and (iv) the global T-80 spectral model developed by NCMRWF with a grid spacing of approximately 156 km over tropical regions.

The ensemble forecasting approach aimed to utilize the strengths of each model to generate improved rainfall forecasts by minimizing individual model biases and uncertainties. However, despite operating at an enhanced resolution of 50 km, the MME system could only cover approximately 250 districts across the country. Given that India has over 700 districts with diverse geographical and climatological characteristics, this limited coverage significantly constrained the broader application and operational impact of the model. Furthermore, the limited spatial resolution and regional coverage meant that fine-scale rainfall variations and local weather phenomena could not be captured accurately, which is especially critical for regions prone to localized heavy rainfall, flash floods, and agricultural dependence on timely precipitation.

### **Disadvantages of the Existing System**

While the earlier approaches made significant contributions to operational weather forecasting in India, they suffer from several inherent limitations that affect their efficacy in the current data-driven and machine-intelligent era. One of the foremost drawbacks of the existing system is its **limited geographical coverage**. The MME system, even at a resolution of 50 km, was only capable of providing rainfall predictions for around 250 districts. This excluded a large portion of the country, particularly rural, hilly, and underserved regions that are equally vulnerable to monsoonal variations and extreme weather events. The insufficient district-level granularity reduces the practical usefulness of the forecasts for state and local-level planning in sectors such as agriculture, water resource management, and disaster preparedness.

Secondly, the existing models were **entirely dependent on deterministic and numerical techniques**, lacking the integration of modern **machine learning or artificial intelligence-based approaches**. Machine learning techniques, with their ability to learn from vast amounts of historical weather data and detect complex, non-linear relationships, have demonstrated superior performance in a variety of predictive tasks, including weather forecasting. The absence of such methodologies in the existing system limits its adaptability, scalability, and responsiveness to rapidly changing atmospheric conditions.

Another critical limitation is the **restricted and insufficient dataset** utilized in the earlier system. The availability of historical weather data was constrained to only a limited number of districts, which posed a significant barrier to comprehensive data analysis and training of more robust models. This data scarcity not only affected the accuracy of rainfall forecasts but also reduced the ability to validate and calibrate the models across different geographical and meteorological contexts.

Moreover, the existing system did not support **real-time or near-real-time data processing**, which is crucial for issuing timely warnings and alerts in the face of severe weather conditions. Without the capability to process and analyze streaming weather data, the system lacks the agility needed to respond to rapidly evolving weather events such as cloudbursts, cyclones, and localized storms.

In conclusion, while the earlier MME-based system laid the foundation for short-range rainfall forecasting at the district level, it now faces considerable challenges in meeting the demands of modern weather prediction. The lack of machine learning integration, limited spatial coverage, inadequate dataset availability, and absence of real-time processing capabilities highlight the urgent need for a more advanced, data-intensive, and intelligent system for accurate and localized rainfall forecasting in India.

**IV. PROPOSED SYSTEM**

In the present study, we outline the strategic development and implementation of an enhanced **Multi-Model Ensemble (MME) technique** tailored for generating high-resolution rainfall forecasts over the Indian region. This system represents a significant improvement over earlier approaches by integrating advanced data processing methods, high-resolution datasets, and modern computational models. The focus of the study is not only on developing a refined forecasting framework but also on demonstrating its **predictive skill and operational efficiency** under varying climatic conditions, particularly during the crucial summer monsoon period. By incorporating higher spatial granularity and leveraging both conventional and machine learning-based analytical methods, the system aims to deliver **more localized, accurate, and actionable rainfall predictions** that can support planning and preparedness across sectors like agriculture, disaster management, and water resource planning.

### **Advantages of the Proposed System**

The proposed system introduces several **key advantages** that address the limitations of earlier forecasting models and enhance the overall performance and usability of rainfall predictions.

One of the major advantages is the **integration of advanced data visualizations**, which allow for dynamic observation of fluctuations, trends, and anomalies in rainfall data over time and space. Through visual tools such as heat maps, time-series plots, and interactive dashboards, users—including researchers, meteorologists, and decision-makers—can easily identify meaningful patterns in the data. These visual insights play a crucial role in guiding the design of predictive models and refining forecasting strategies, thereby improving the overall understanding of rainfall behavior across different regions.

Another significant benefit of the proposed system is its ability to **predict the amount of rainfall across diverse datasets**, including both structured historical meteorological data and real-time streaming data collected from satellite imagery, ground-based weather stations, and radar sources. This dual compatibility makes the system highly versatile and adaptable for both retrospective analysis and forward-looking forecasting. It ensures that predictions are not limited to one specific format or data source, thereby increasing the robustness and accuracy of the model outputs.

Additionally, the proposed system leverages **machine learning algorithms** that can automatically learn from large volumes of past data to detect complex relationships between atmospheric variables and rainfall occurrences. This leads to more accurate and reliable predictions, even in regions with highly variable or unpredictable rainfall patterns. The use of machine learning also enables continuous improvement, as the models can be retrained with new data to enhance their performance over time.

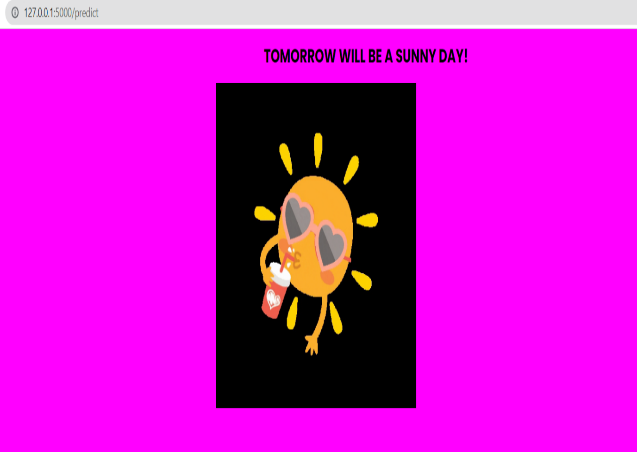
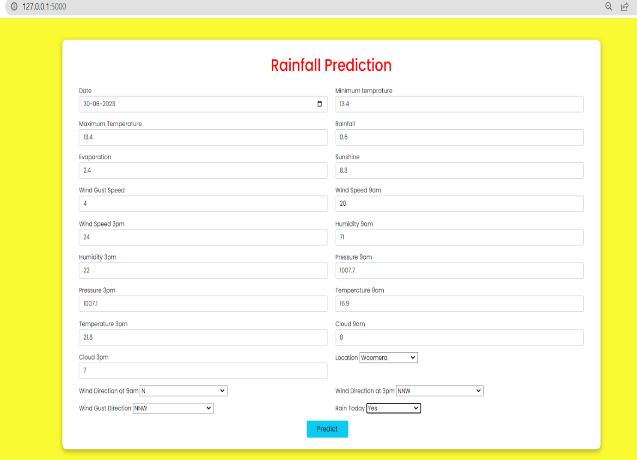
Furthermore, the system offers **wider geographical coverage** by utilizing high-resolution gridded data, enabling district-level or even sub-district-level forecasting. This granularity is especially beneficial for agricultural and disaster response agencies that require localized information to make timely and effective decisions.

Another important advantage is the **real-time processing and prediction capability,** which allows stakeholders to access updated forecasts with minimal delay. This is made possible through the integration of Big Data tools and streaming analytics, ensuring that the system is both fast and scalable, capable of handling high volumes of data with low latency.

In summary, the proposed system not only improves the technical accuracy of rainfall forecasting but also enhances user interaction, data accessibility, and operational efficiency. These features make it a comprehensive and forward-looking solution that is well-suited for addressing the complex challenges associated with weather prediction in the Indian subcontinent.

**V. RESULTS**

The Output Screens are as follows:



**VI. CONCLUSION**

Various visualizations of data are observed which helps in implementing the approaches for prediction.

• Prediction of amount of rainfall for both the types of dataset.

• Observations indicate machine learning models won’t work well for prediction of rainfall due to Fluctuations in rainfall.

Results of this study showed that all these models, in general, have the capability to capture large scale rainfall features of summer monsoon, such as heavy rainfall belt along the west coast, over the domain of monsoon trough and along the foothills of the Himalayas. It has clearly emerged from the results of the skill score that MME is superior to each member model. For the district level forecast, the procedure has showed appreciable skill to predict occurrence and non-occurrence of rainfall, as well as for the rainfall category of moderate rainfall. But it fails to capture heavy rainfall events. Otherwise, performance of the district level forecast for most of the districts has been fairly good, particularly over the monsoon affected states.

**FUTURE WORK:**

Though some significant improvement in accuracy and reliability of NWP product has been driven by adopting MME approach, however, limitations remain, particularly in the prediction of intensity and mesoscale rainfall features causing inland flooding. During recent years, Ensemble Prediction System (EPS) has emerged as a powerful tool for improving medium range weather forecasts. In the EPS, single model is used with multiple sets of initial conditions (Brooks and Doswell 1999) to obtain the final forecast. While Singular Vector and Bred Vector (BV) methods are still widely used in generating initial perturbations, Ensemble Transform of BV, Ensemble Transform Kalman Filter and Ensemble Data Assimilation are also implemented in various centres. Currently, 10 global centres operate EPS for medium range forecasts and they exchange model outputs at the native resolution among themselves. Very recently, with the commissioning of High Performance

**VIII. REFERENCES**

**[**1] Thirumalai, Chandrasegar, et al. "Heuristic prediction of rainfall using machine learning techniques." 2017 International Conference on Trends in Electronics and Informatics (ICEI). IEEE, 2017.

[2] Geetha, A., and G. M. Nasira. "Data mining for meteorological applications: Decision trees for modeling rainfall prediction." 2014 IEEE International Conference on Computational Intelligence and Computing Research. IEEE, 2014

[3] Parmar, Aakash, Kinjal Mistree, and Mithila Sompura. "Machine learning techniques for rainfall prediction: A review." 2017 International Conference on Innovations in information Embedded and Communication Systems. 2017.

[4] Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi. "Rainfall prediction for the Kerala state of India using artificial intelligence approaches." Computers & Electrical Engineering 70 (2018): 66-73.

[5] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April ’2017.

[6] Y. Nieto, V. García-Díaz, C. Montenegro, and R. G. Crespo, ‘‘Supporting academic decision making at higher educational institutions using machine learning-based algorithms,’’ Soft Comput., vol. 23, no. 12, pp. 4145–4153, 2018.

[7] K. V. Pincus, D. E. Stout, J. E. Sorensen, K. D. Stocks, and R. A. Lawson, ‘‘Forces for change in higher education and implications for the accounting academy,’’ J. Accounting Educ., vol. 40, pp. 1–18, Sep. 2016.

[8] A.-P. Pavel, A. Fruth, and M.-N. Neacsu, ‘‘ICT and e-learning—Catalysts for innovation and quality in higher education,’’ Procedia Econ. Finance, vol. 23, pp. 704–711, Jan. 2015.