




Prediction of heavy rainfall days over a peninsular Indian station using the machine learning algorithms

KANDULA V SUBRAHMANYAM^{1,*} , C RAMSENTHIL², A GIRACH IMRAN¹, ANIKET CHAKRAVORTY³, R SREEDHAR², E EZHILRAJAN⁴, D BALA SUBRAHAMANYAM¹, RADHIKA RAMACHANDRAN¹, KARANAM KISHORE KUMAR¹, M RAJASEKHAR² and C S JHA⁵

¹Space Physics Laboratory, Vikram Sarabhai Space Centre, Indian Space Research Organisation, Thiruvananthapuram, India.

²Satish Dhawan Space Centre-SHAR, Indian Space Research Organisation, Sriharikota, India.

³Department of Space, North Eastern Space Applications Centre, Shillong, India.

⁴ISRO Propulsion Complex, Indian Space Research Organisation, Mahendragiri, India.

⁵National Remote Sensing Centre, ISRO, Hyderabad, India.

*Corresponding author. e-mail: kv_subrahmanyam@vssc.gov.in; kvsm2k@gmail.com

MS received 27 January 2021; revised 1 July 2021; accepted 5 July 2021

Advance prediction of heavy rainfall days over a given location is of paramount importance as heavy rainfall impacts ecosystems, leads to floods, accounts largely for the total rainfall over the region and its prediction is highly desired for the efficient management of weather-dependent activities. Traditionally, Numerical Weather Prediction models serve the purpose of weather predictions, but they have their constraints and limitations. In this regard, artificial intelligence and machine learning tools have gained popularity in recent years. In the present study, we have employed the Gaussian Process Regression (GPR) approach, one of the machine learning methods, on a long time-series rainfall data for the determination of heavy and light rainfall days. Climatological data of daily rainfall for a period of 116 years from 1901 to 2016 over Sriharikota (13.66°N, 80.23°E), a coastal island location in India, is used for training the GPR model for the identification of the heavy and light category of rainy days. The performance of the GPR model is investigated by predicting the heavy and light rainfall days per year over Sriharikota. K-nearest neighbour, random forest, and decision tree models are also used and results are compared. The validation of GPR results shows that the performance of the proposed model is satisfying (root mean square error = 0.161; mean absolute error = 0.126; mean squared error = 0.026), especially for the heavy rainfall days. Furthermore, GPR model is extended to prediction of spatial distribution of monthly rainfall over the Indian region. Results obtained from the present study encourages the utilization of the GPR model as one of the promising machine learning tools for the prediction of heavy rainfall days over a given location.

Keywords. Heavy rainfall; light rainfall; machine learning; Gaussian process regression.

1. Introduction

Rainfall is one of the important elements of the hydrological cycle. Due to the complexities involved in precipitation process (from convection to cloud formation to rainfall), an accurate prediction of the occurrence of rainfall episodes over a given location becomes a difficult task without appropriate tools. Moreover, the regional and seasonal dependence of rainfall over a given location further add up to the complexity of the problem. In the Indian context, the Asian summer monsoonal season accounts for more than 80% of the total rainfall over the country (e.g., Ding and Sikka 2006). Despite having a consistency in the pattern of rainfall over different seasons, researchers working towards the prediction of rainfall, find it difficult to provide a credible tool for the prediction of rainfall episodes of different categories (e.g., French *et al.* 1972). The occurrence of heavy rainfall eventually leads to major floods as seen over Mumbai in July 2005, over Chennai and Bengaluru in October 2005, and more recently over Kerala in August 2018 (Goswami *et al.* 2006; Rajeevan *et al.* 2008). The occurrence of such extreme rainfall events is increasing regionally as well as globally, and there is no imminent pattern in such events as they have large variabilities over different geographical regions (Parthasarthy *et al.* 1995; Dash *et al.* 2009). Some of the heavy rainy events led to severe floods in many parts of the country during recent years and have caused large-scale destruction of life and property. Since 1950s, numerical weather forecasting has advanced remarkably. A wide range of rainfall forecasts is employed in the area of Numerical Weather Prediction (NWP) at regional and global scales. Synoptic weather forecasting is generally dominated by the computer-based NWP models, which are based on the governing equations of atmosphere and parameterization of physical processes. Over the last five to six decades, these models are serving as the backbone of weather forecasting and their credibility has been consistently increasing ever since their inception (e.g., Bauer *et al.* 2015; Subrahmanyam *et al.* 2019).

Conventionally, NWP models are used as the primary tools for the prediction of irregular heavy rainfall events. Though the NWP models provide fairly reliable prediction of extreme weather events, occasionally these traditional approach yields inconsistent results about the occurrence of heavy rainfall events. In such a scenario, the

Machine Learning (ML) algorithm of Artificial Intelligence (AI) offers a promising alternative approach for the prediction of rainfall events. Recently, there has been a growing interest in the usage of AI in the weather as well as the climatic forecasts. ML algorithms have attained wider acceptance in several research domains (Schneider *et al.* 2017; O’Gorman and Dwyer 2018; Reichstein *et al.* 2019). ML is a sub-area of AI, where computers can be trained to learn a probable relationship between different variables from a large training dataset, and it provides easier evaluation with high performance compared to the NWP models (Mosavi *et al.* 2012; Mekanik *et al.* 2013). There are some studies on the prediction of rainfall and other parameters in regional as well as the global scales (Schneider *et al.* 2017; Scher and Messori 2018). Goswami and Srividya (1996) designed a neural network to predict rainfall over the Indian region. Venkatesan *et al.* (1997) have used Error Back Propagation approach for the prediction of Indian summer monsoon rainfall. Sahai *et al.* (2000) have also used Artificial Neural Network (ANN) methods for prediction of all Indian summer monsoon rainfall. Chaudhuri and Chattopadhyay (2005) have developed a feed-forward neural network for the estimation of maximum surface temperature and relative humidity. Chattopadhyay (2007) has carried out a comparative study by using different ANNs for prediction of the rainfall. Vathsala and Koolagudi (2017) have predicted the peninsular Indian summer monsoon rainfall using multi-layered perceptron (MLP) model. Karmakar *et al.* (2009) have developed an Artificial Neural network (ANN) model for estimation of the rainfall over the Indian region. Wang and Sheng (2010) have proposed a Generalization Regression Neural Network (GRNN) for forecasting the yearly rainfall over China. Hung *et al.* (2009) have forecasted the hourly rainfall in Bangkok using Generalized Feed Forward Neural Network. Nagahamulla *et al.* (2011) have used ANN to predict the seasonal monsoon rainfall in Sri Lanka and Kashiwao *et al.* (2017) have developed a prediction model for hourly rainfall prediction in Japan using Multilayer Perceptron (MLP) and Radial Basis Function Neural Network (RBFNN). Yan *et al.* (2020) developed a deep learning approach for precipitation forecast.

Although AI-based meteorological prediction has its advantages, it comes with its own set of challenges and limitations. Dueben and Bauer (2018) documented the challenges and various

designing choices for forecasting based on ANN. Most of the studies mentioned above were limited to the prediction of hourly, daily and monthly or average rainfall over a region using different AI- and ANN-based algorithms. Performances of these methods were found to be largely dependent on their training method and the historical datasets used in the study. Whenever the training data volume was not sufficient and did not cover large dynamical values, the ANN and ML algorithms failed to provide promising results. Despite good progress in the area of AI and ML techniques, there are very limited studies on the prediction of heavy rainfall events over a given location.

The aforementioned studies were for the prediction of seasonal, daily and monthly rainfall, but there is no study seen in the literature on the prediction of heavy rainfall days, which impact the livelihood of many people and trigger havoc in the country. Therefore, in the present study, we propose an AI-based approach for the prediction of heavy and light rainfall events for Sriharikota, a coastal station on the eastern coastline of the Indian peninsula. Currently, there is a whole gamut of AI algorithms to choose from, however, this work only focuses on the efficacy of the Gaussian Process Regression (GPR) which is a nonparametric, Bayesian approach to regression. We have used rainfall data for a period of 80 years spanning from 1901 to 1980 for the training of AI model, and subsequently allowed the model to make prediction of light and heavy rainy days from 1981 to 2016. This article is organized into four sections. Section 2 provides details about the model, data and method of analysis. Section 3 describes important results in the context of AI/ML approach. Section 4 highlights the summary and implications of the present results on futuristic prediction techniques.

2. Data pre-processing and description of GPR algorithm

In the present work, we have focused on AI/ML technique for prediction of light to heavy rainy days for Sriharikota (13.82°N, 80.22°E), which is a coastal station located on the eastern coastline of the Indian peninsula (figure 1). The climate of Sriharikota is mainly wet (higher humidity) and it encounters both southwest and northeast Indian monsoon. The climatological average annual rainfall over Sriharikota is about 1311.9 mm. More

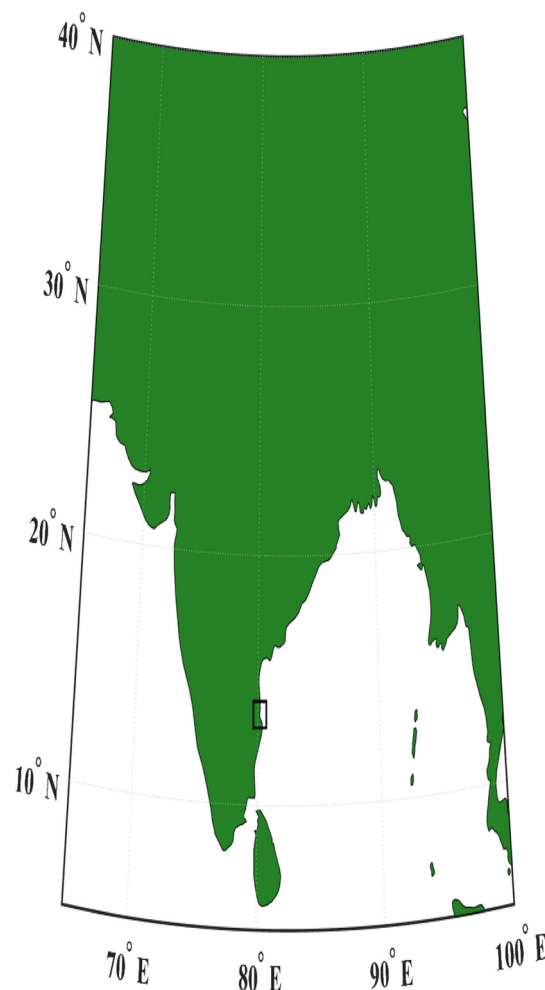


Figure 1. Box denotes the selected location.

than 50% of annual rainfall occurs between October and December over Sriharikota. About 14% of average annual rainfall days are due to moderate to heavy rain event days. The study of heavy rainfall rates is essential for launch missions like Polar Satellite Launch Vehicle (PSLV), Geo-Synchronous Satellite Launch Vehicle (GSLV) and pre-launch day-to-day activities at Satish Dhawan Space Centre (SDSC), Sriharikota. Here, we have used high-resolution spatially gridded (0.25° × 0.25°) daily rainfall data of India Meteorological Department (IMD) extracted for Sriharikota for a period of 116 years from 1901 to 2016 forms the main database (Rajeevan *et al.* 2006). The IMD data provides 24-h accumulated rainfall amount on a daily basis. We have classified the rainfall data into a total of nine distinct and standard categories from ‘No Rain’ to ‘Exceptionally very high rain’ (IMD). Table 1 provides standard threshold of rainfall amounts used by IMD for categorization to

Table 1. Classification of various categories of rainfall (in table, we can convert percentage to equivalent day. This will show, how many days in a year can we expect one particular category of rain. You simply need to multiply % with 365).

Classification	Rainfall (RF) per day	Number of days $116 \times 366 = 42,456$	Percentage of occurrence (%)
No rain	0 mm	29,672	69.88
Trace	$0.01 \leq \text{RF} \leq 0.04$ mm	0	0
Very light rain	$0.1 \leq \text{RF} \leq 2.4$ mm	3269	7.69
Light rain	$2.5 \leq \text{RF} \leq 7.5$ mm	3026	7.12
Moderate rain	$7.6 \leq \text{RF} \leq 35.5$ mm	4359	10.26
Rather heavy rain	$35.6 \leq \text{RF} \leq 64.4$ mm	1069	2.51
Heavy rain	$64.5 \leq \text{RF} \leq 124.4$ mm	684	1.61
Very heavy rain	$124.5 \leq \text{RF} \leq 244.5$ mm	226	0.53
Exceptionally heavy rain	$\text{RF} \geq 244.5$ mm	19	0.0044

different classes of rainy events. Here, we have chosen ‘heavy rain’ and ‘light rain’ cases to demonstrate the potential for ML-based model.

2.1 Gaussian processes regression (GPR)

In the present study, we have used GPR model for the prediction of light and heavy rainy days. This is a supervised learning model and it can be used for classification and regression problems. The outputs are discrete class labels in a classification problem, while others are predictions of continuous quantities in regression. Gaussian process (GP) is a collection of random variables and it is described by its mean and covariance functions. Covariance function in GPR also called as kernel plays a crucial role in determining the shape of prior and posterior of the GPR. Rasmussen and Williams (2006) have described the details about Gaussian processes and their covariance functions in neural network algorithms. Kernel functions in the GPR model characterize the correlation among different input points in the process. Various kernel functions can be used for the GP model, such as radial basis function (RBF), Matern kernel, rational quadratic kernel, ex-sine-squared kernel and dot-product-kernel. The inverse of the length-scale determines the relevance of input data. In the present study, we have used the RBF kernel as optimization in the GPR model and White kernel as the noise level. Liu *et al.* (2001) also used an adaptive RBF neural network for estimation of rainfall. Equation (1) represents the general form of GPR model (Rasmussen and Williams 2006) and figure 2 depicts the flowchart of GPR processes performed in the present study.

Let us assume a given training dataset $\mathcal{D} = \{(x_i, y_i) \mid i = 1, \dots, n\}$, where y_i is the daily accumulated rainfall

corresponding to a given day x_i . In the finite set of given data \mathcal{D} , $f(x(1)), f(x(2)), \dots, f(x(n))$ is the random variable and obeys the Gaussian distribution. The statistical form of the general GP can be represented by the mean function $m(x)$ along with its covariance function $k(x, x^*)$ and is given in equation (1)

$$m(x) = E[f(x)] \quad (1)$$

$$Y = F \times X + \varepsilon \quad (2)$$

where X is the input vector, F is the weight of the model and Y is the observed target. The error associated with the GP is given by ε , i.e., $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, where σ_n^2 stands for the variance of the noise.

The GP predictive distribution at a test input x^* is given by

$$P(F(x^*) | \mathcal{D}, x^*) = \mathcal{N}(\mu(m(x^*)), \sigma_n^2(m(x^*))) \quad (3)$$

$$\mu(m(x^*)) = k^{*\text{T}}(K + \sigma_n^2 I)^{-1} Y \quad (4)$$

$$\sigma_n^2(m(x^*)) = k^{**} k^{*\text{T}} (K + \sigma_n^2 I)^{-1} k^* \quad (5)$$

where K is the kernel matrix constructed as $K_{ij} = k(x_i, x_j)$, $k^{**} = k(x^*, x^*)$, $k^* = k(X, x^*)$ and k is the squared exponential covariance function. Equation (6) represents the form of the squared exponential function, which has been successfully applied to different applications.

$$k_{\text{SE}}(x_i, x_j) = \sigma_f^2 \exp \left[-(1/2) (x_i - x_j)^{\text{T}} M^{-2} (x_i - x_j) \right] \quad (6)$$

where $\{M, \sigma_f^2, \sigma_n^2\}$ is the set of hyperparameters, denoted by θ and σ_f^2 defines the variance of the kernel function. The optimal hyperparameter θ can be

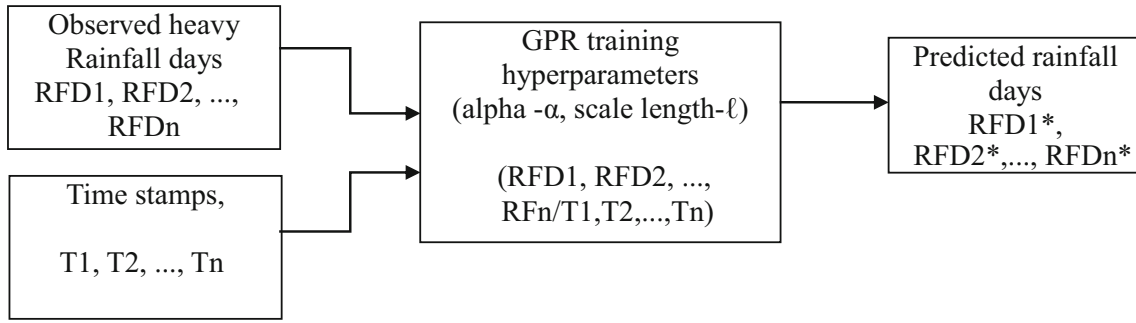


Figure 2. The flowchart diagram for implementation of the GPR model in the present study.

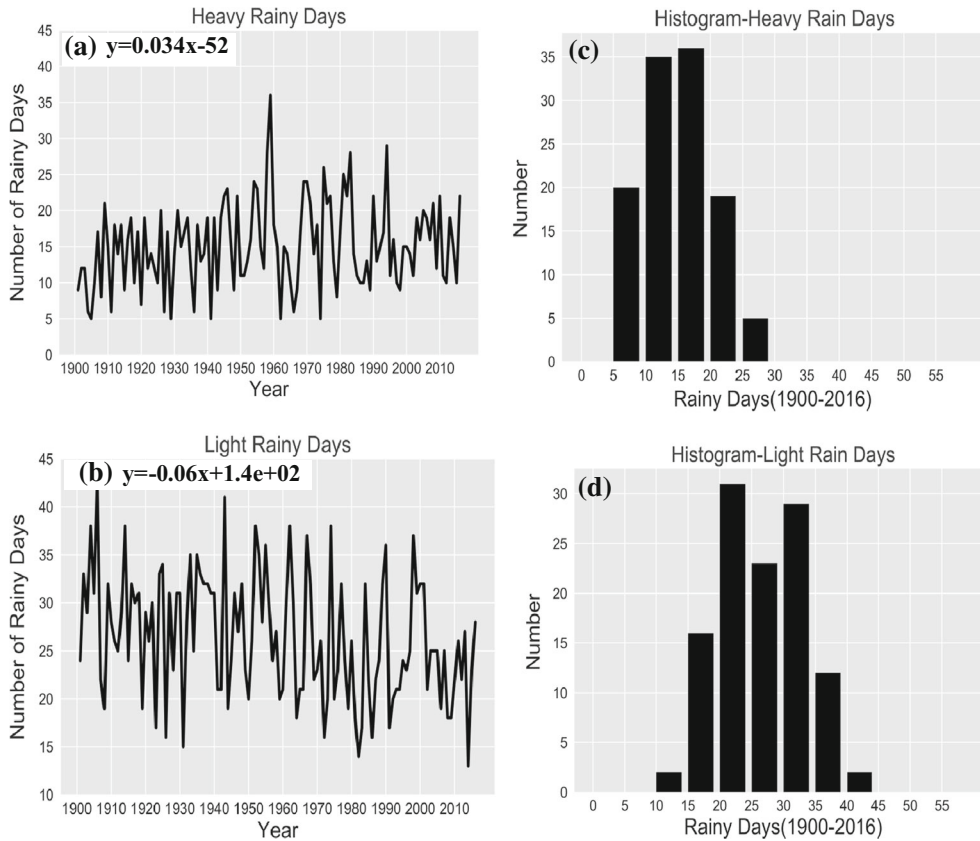


Figure 3. Time-series of yearly occurrence of heavy and light rainy days during 1901–2016 (a and c) and histograms of heavy and light rainy days (b and d).

solved by the maximum likelihood estimate function $L(\theta)$, expressed as:

$$L(\theta) = -\log(P(Y|X, \theta)). \tag{7}$$

To get the prediction accuracy using GPR model, the time-series of heavy rainfall data have been divided into two sets, such as training (80%) and testing (20%) data. In this experiment, yearly observed 116 years (1901–2016) of heavy and light rainfall days along with their corresponding time stamps are given as input to the GPR model. After

appropriate training and cross-validation, the GPR model could provide a better prediction of heavy and light rainfall days as discussed in section 3.

3. Experiment results and discussion

Figure 3(a and c) shows the yearly occurrence of heavy and light rainy days and 3(b and d) shows the histograms of heavy and light rainy days over Sriharikota during 1901–2016. It is observed from the figure that there is an increasing trend in heavy

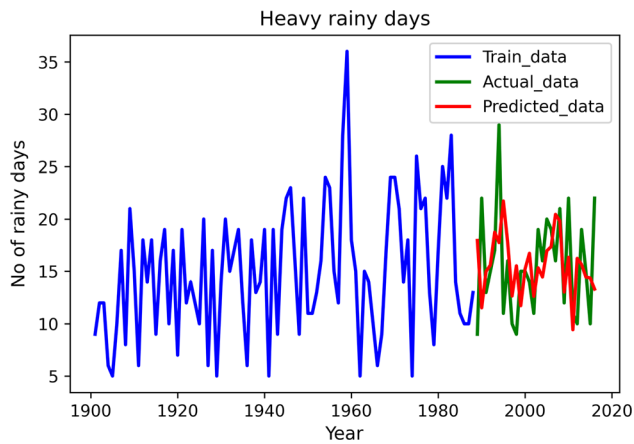


Figure 4. The prediction of heavy rainfall days using the GPR model.

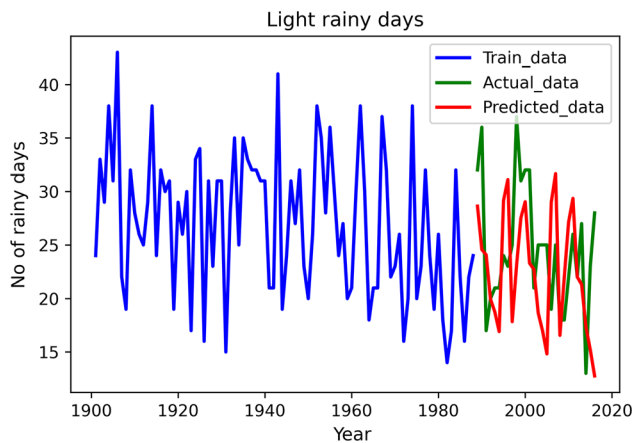


Figure 5. Same as figure 4, but for light rainfall days.

rainfall days (0.342 per decade with 95% confidence), while a decreasing trend is observed in light rainy days (−0.602 per decade with 95% confidence). The present observations are corroborating with the existing results. It has been observed that frequency of occurrence of heavy rain events is increased significantly over India during 1950–2015 (Goswami *et al.* 2006; Kulkarni *et al.* 2020) and a decreasing trend observed in the moderate rain events (Goswami *et al.* 2006). The rise in heavy rainfall events is seen over several geographical locations and sometimes it exceeds the seasonal mean rainfall (Westra *et al.* 2013). Histograms of heavy and light rainy days are shown in figure 3(b and d), which showed a Gaussian distribution with skewness of 0.47 and 0.19, respectively.

After the GPR model is trained, the regression model results can be displayed by the response plot, i.e., the predicted response *vs.* record number as shown in figure 4. Figure 4 shows the prediction of heavy rainfall days using the GPR model. Blue

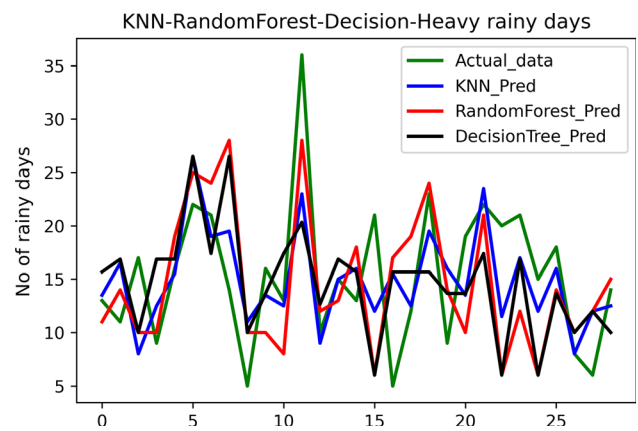


Figure 6. The prediction of heavy rainfall days using the KNN, random forest and decision tree models.

Table 2. Comparison of accuracy of various models.

Model used	MAE	MSE	RMSE
GPR	0.126	0.026	0.161
KNN	0.185	0.047	0.218
Random forest	0.216	0.069	0.262
Decision tree	0.197	0.047	0.217

curve shows the data (1901–1988) used to train the model. Green and red colour line indicates the actual and predicted heavy rainfall days from 1989 to 2016. Overall, the predictions are good and able to reproduce the variability of heavy rainfall days as shown in figure 4. Interestingly, it could capture the peak years of heavy rainfall days. Figure 5 is the same as figure 4, for light rainfall days, which also showed a good prediction compared to observations. Further, we also used various other models such as k-nearest neighbour (KNN), random forest and decision tree models to predict the heavy rainfall days as shown in figure 6. All the models were able to reproduce the variability in heavy rainfall days as seen in figure 6. We further evaluate the model performance by using the validation parameters, such as Root Mean Square Error (0.161), Mean Square Error (0.026), and Mean Absolute Error (0.126), for the model predicted results. The comprehensive comparison is shown in table 2. The performance of difference GPR model is compared with other models using the following model statistics.

We extended the model for prediction of daily average rainfall over the Indian region (60°–100°E; 0°–35°N) using 36 years (1979–2014) of data from Global Precipitation Climatology Project (GPCP). The GPCP daily rainfall product is an integration of various global precipitation satellite datasets of land and oceans and a gauge analysis overland

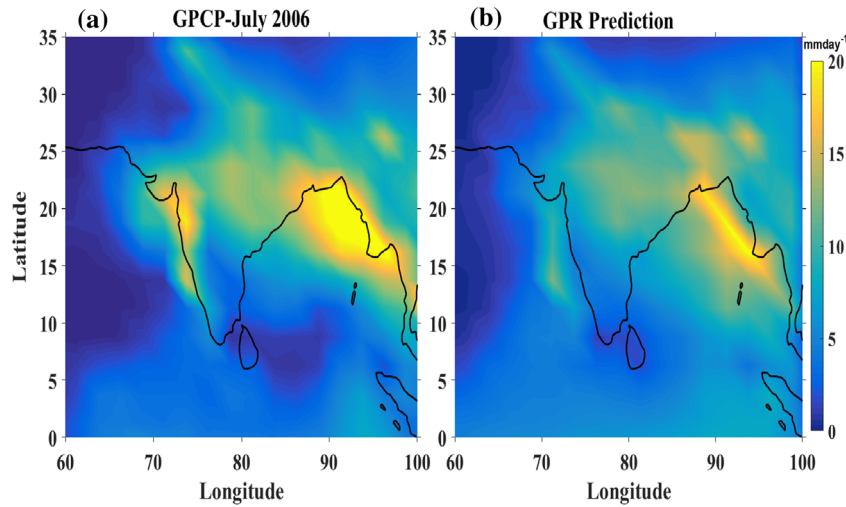


Figure 7. (a) GPCP daily rainfall and (b) predicted GPR daily rainfall for July 2006.

(Huffman *et al.* 1997). This data has been widely used for various regional and global precipitation studies. The GPR model is trained from 1979 to 2005 (27 years) for each grid ($1^\circ \times 1^\circ$) point over the entire Indian region and validated for 2006–2014. Figure 7 shows the GPR-predicted rainfall along with the GPCP rainfall for July 2006. Also the Gujarat and Maharashtra high rainfall was not captured. Overall, the GPR prediction shows a good agreement with GPCP observations ($R^2 = 0.84$). During July, which is the peak month of SW-monsoon season, a large amount of rainfall is observed over the head Bay of Bengal and the Western Ghats regions as seen in figure 7(a). The GPR model was able to reproduce the locations with heavy rainfall, however, magnitudes of rainfall predicted by the GPR model were underestimated as against the GPCP observations. Thus, the present study highlights the potential use of GPR in prediction of drastic weather episode such as heavy and light rainy days. However, there are still some limitations for the magnitude. The improvement of amplitude of predicted rainfall could be considered. Further, continuous evaluation or modification of the AI-based model is required, particularly for heavy rainfall events. Also, if ML-based models are blended with conventional NWP output, more accurate forecasts can be achieved, which needs further study.

4. Conclusion and future scope

In the present communication, the GPR is applied to historical time-series rainfall data to predict heavy and light rainfall days over Sriharikota

located over east coast of India. The model predicted for heavy and light rainfall days with reasonably good accuracy (RMSE = 0.161). The GPR could reproduce the variability and capture the peak years of heavy rainfall days. The performance of the GPR model was found to be better when compared with the results from other three ML-based models. Further, the GPR model was trained using GPCP daily rainfall data and prediction made for the spatial distribution of monthly rainfall ($1^\circ \times 1^\circ$) over the Indian region which also showed good agreement with the observations ($R^2 = 0.84$ and RMSE = 3.324). The results obtained from the present ML-model encourages the utilization of the GPR model as one of the promising machine learning tools for the prediction of heavy rainfall days over a given location, which impact average rainfall over the region, ecosystems, landscape through erosion processes and leads to major floods. It is recommended that in future the GPR model could be coupled with other ML-based models, so that the prediction accuracy can be improved in terms of magnitude as well as the occurrence of heavy rainfall days. In the world of changing climate where anthropogenic emissions of aerosols influence cloud properties and hence rainfall, the model prediction made in the present study needs to be evaluated continuously for the improvement in the model.

Acknowledgements

The authors are greatly thankful to Dr Tirtha Pratim Das, Shri N Raghu Meetei and Shri R Umamaheswaran, Directorate for Technology

Development and Innovation (DTDI), Indian Space Research Organisation (ISRO), Bangalore for initiating the AI/ML-based training programme. The observational dataset used in the study is the daily gridded rainfall data based on rain gauge observations from India Meteorological Department (IMD) for the period 1901–2016 are also used (Rajeevan *et al.* 2006, https://www.imdpune.gov.in/Clim_Pred_LRF_New/Gridded_Data_Download.html). Authors are also thankful to Director, Space Physics Laboratory, VSSC for constant support and encouragement.

Author statement

Kandula V Subrahmanyam: Conceptualization, modelling, analysis and manuscript. Girach Imran: Modelling and manuscript. C Ramsenthil, Aniket Chakravorty, R Sreedhar and E Ezhilraj: Modelling and visualization. D Bala Subrahmanyam, Radhika Ramachandran, Karanam Kishore Kumar, M Rajasekhar and C S Jha: Supervision and review of the manuscript.

References

- Bauer P, Thorpe A and Brunet G 2015 The quiet revolution of numerical weather prediction; *Nature* **525** 47–55, <https://doi.org/10.1038/nature14956>.
- Chattopadhyay S 2007 Feed forward artificial neural network model to predict the average summer-monsoon rainfall in India; *Acta Geophys.* **55(3)** 369–382.
- Chaudhuri S and Chattopadhyay S 2005 Neuro-computing based short range prediction of some meteorological parameters during the pre-monsoon season; *Soft Comput.* **9(5)** 349–354.
- Dash S K, Kulkarni M A, Mohanty U C and Prasad K 2009 Changes in the characteristics of rain events in India; *J. Geophys. Res.* **114** D10109, <https://doi.org/10.1029/2008JD010572>.
- Ding Y and Sikka D R 2006 Synoptic systems and weather; In: *The Asian Monsoon* (ed.) Wang B, Springer, Berlin, Heidelberg, 135p.
- Dueben P D and Bauer P 2018 Challenges and design choices for global weather and climate models based on machine learning; *Geosci. Model Dev.* **11** 3999–4009, <https://doi.org/10.5194/gmd-11-3999-2018>.
- French M N, Krajewski W F and Cuykendall R R 1972 Rainfall forecasting in space and time using a neural network; *J. Hydrol.* **137(1–4)** 1–31, [https://doi.org/10.1016/0022-1694\(92\)90046-x](https://doi.org/10.1016/0022-1694(92)90046-x).
- Goswami B N, Venugopal V, Sengupta D, Madhusoodanan M S and Xavier P K 2006 Increasing trend of extreme rain events over India in a warming environment; *Science* **314** 1442–1445, <https://doi.org/10.1126/science.1132027>.
- Goswami P and Srividya 1996 A novel neural network design for long range prediction of rainfall pattern; *Curr. Sci.* **70(6)** 447–457.
- Huffman G J, Adler R F, Arkin P, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B and Schneider U 1997 The global precipitation climatology project (GPCP) combined precipitation dataset; *Bull. Am. Meteorol. Soc.* **78** 5–20.
- Hung N Q, Babel M S, Weesakul S and Tripathi N K 2009 An artificial neural network model for rainfall forecasting in Bangkok, Thailand; *Hydrol. Earth Syst. Sci.* **13(8)** 1413–1425.
- Karmakar S, Kowar M K and Guhathakurta P 2009 Long-range monsoon rainfall pattern recognition and prediction for the subdivision ‘EPMB’ Chhattisgarh using deterministic and probabilistic neural network; *Proc. 7th Int. Conf. Adv. Pattern Recognit.*, pp. 367–370.
- Kashiwao T, Nakayama K, Ando S, Ikeda K, Lee M and Bahadori A 2017 A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency; *Appl. Soft Comput.* **56** 317–330.
- Kulkarni A, Sabin T P, Chowdary J S, Koteswara Rao K, Priya P, Gandhi N, Bhaskar P, Bui V K and Sabade S S 2020 Precipitation changes in India; In: *Assessment of climate change over the Indian region* (eds) Krishnan R *et al.*, 48p, https://doi.org/10.1007/978-981-15-4327-2_3.
- Liu H, Chandrasekar V and Xu G 2001 An adaptive neural network scheme for radar rainfall estimation from WSR-88D observations; *J. Appl. Meteor.* **40(11)** 2038–2050.
- Mekanik F, Imteaz M, Gato-Trinidad S and Elmahdi A 2013 Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes; *J. Hydrol.* **503** 11–21.
- Mosavi A, Rabczuk T and Varkonyi-Koczy A R 2012 Reviewing the novel machine learning tools for missing data by novel artificial neural networks; *Expert Syst. Appl.* **39** 10,456–10,464.
- Nagahamulla H R K, Ratnayake U R and Ratnaweera A 2011 Monsoon rainfall forecasting in Sri Lanka using artificial neural networks; *Proc. 6th Int. Conf. Ind. Inf. Syst.*, pp. 305–309.
- O’Gorman P A and Dwyer J G 2018 Using machine learning to parameterize moist convection: Potential for modelling of climate, climate change, and extreme events; *J. Adv. Model. Earth Syst.* **10** 2548–2563, <https://doi.org/10.1029/2018MS001351>.
- Parthasarthy B, Munot A A and Kothawale D R 1995 Monthly and seasonal rainfall series for all India homogeneous regions and meteorological sub-divisions: 1871–1994; Res. Rep. RR-065, Indian Inst. of Trop. Meteorol. Pune, 113p.
- Rajeevan M, Bhate J, Kale J D and Lal B 2006 High resolution daily gridded rainfall data for Indian region: Analysis of break and active monsoon spells; *Curr. Sci.* **9(3)** 296–306.
- Rajeevan M, Bhate J and Jaswal A K 2008 Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data; *Geophys. Res. Lett.* **35** L18707, <https://doi.org/10.1029/2008GL035143>.
- Rasmussen C E and Williams C K I 2006 *Gaussian processes for machine learning*; Massachusetts Institute of Technology, The MIT Press, www.GaussianProcess.org/gpml.

- Reichstein *et al.* 2019 Deep learning and process understanding for data-driven Earth System Science; *Nature* **566** 195–204.
- Sahai A K, Soman M K and Satyan V 2000 All India summer monsoon rainfall prediction using an Artificial Neural Network; *Clim. Dyn.* **16**(4) 291–302.
- Scher S and Messori G 2018 Predicting weather forecast uncertainty with machine learning; *Quart. J. Roy. Meteor. Soc.* **144** 2830–2841, <https://doi.org/10.1002/qj.3410>.
- Schneider T, Lan S W, Stuart A and Teixeira J 2017 Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations; *Geophys. Res. Lett.* **44** 12,396–12,417, <https://doi.org/10.1002/2017GL076101>.
- Subrahmanyam D B, Radhika Ramachandran, Nalini K, Freddy P Paul and Roshny S 2019 Performance evaluation of COSMO numerical weather prediction model in prediction of OCKHI-one of the rarest very severe cyclonic storms: A case study; *Nat. Hazards* **96** 431–459, <https://doi.org/10.1007/s11069-018-3550-2>.
- Vathsala H and Koolagudi S G 2017 Prediction model for peninsular Indian summer monsoon rainfall using data mining and statistical approaches; *Comput. Geosci.* **98** 55–63.
- Venkatesan C, Raskar S D, Tambe S S, Kulkarni B D and Keshavamurthy R N 1997 Prediction of all Indian summer monsoon rainfall using error-back-propagation neural networks; *Meteorol. Atmos. Phys.* **62**(3–4) 225–240.
- Wang Z L and Sheng H H 2010 Rainfall prediction using generalized regression neural network: Case study Zhengzhou; In: *Proc. Int. Conf. Comput. Inf. Sci. (ICIS)*, pp. 1265–1268.
- Westra S, Alexander L V and Zwiers F W 2013 Global increasing trends in annual maximum daily precipitation; *J. Clim.* **26** 3904–3918.
- Yan Q, Ji F, Miao K, Wu Q, Xia Y and Li T 2020 Convolutional residual-attention: A deep learning approach for prediction of nowcasting; *Adv. Meteorol.*, <https://doi.org/10.1155/2020/6484812>.

Corresponding editor: N V CHALAPATHI RAO