PREDICTION OF RAIN FALL USING MACHINE LEARNING PATTERNS

***Dr.M.Ajay Kumar*[1]., *CH. Bhargavi*[2]., *B.Maneesha*[3]., *T.Uha Vaishnavi*[4]., *M.Sindhu*[5]**

1 Associate Professor, Department Of ECE., Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India (ajayajju126@gmail.com)

2,3,4,5 B.Tech ECE, (19RG1A04D4, 19RG1A04C7, 19RG1A04H1, 20RG5A0414),

Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

Abstract: The efficacy of the models varies depending on the trends and odd forms of the dataset, however they have been used by machine learning algorithms to create precise and smooth prediction models for various time series. The state of Meghalaya was chosen as a case study for this investigation because of its varied climate and strategic location on the globe. Linear Regression (LR), Regression Trees (RT), Gaussian Process Regression (GPR), Support Vector Machines (SVM), and Ensembles of Trees (ET) were the five machine learning methods utilized to construct the ideal ML model for Meghalaya rainfall forecasting. The effectiveness of each model was measured by its RMSE, or root mean squared error. The parameters' optimization as predictors to impart in the training dataset was also inspected in the investigation. Almost every dataset from several districts in Meghalaya identified LR, SVM, and GPR to be highly excellent result-oriented models. However, the most accurate predictions may be made using LR and SVM algorithms. The analysis identified the frequency of wet days as the most sensitive and efficient single parameter as predictor, suggesting that this factor may be used to generate the prediction models needed to carry out the research.

Keywords:

- Machine Algorithms;
- Support VectorMachine;
- Rainfall;
- Regressionmodels;
- Root mean squared error.

1. Introduction

Rainfall forecasting has been striving for an ideal model that can reliably provide numbers for a given region or climate. In order to foresee potential intense rainfall, Yang et al. (2007) used a geographical data mining strategy. Due to the complicated nonlinear data pattern involved in rain prediction, several new methods have been developed to increase forecast precision. The most complex aspect of weather forecasting is the need to develop a model of predictive algorithm for reliable precipitation. One research that used environmental machine learning methods to improve predicting accuracy was conducted in 2008 (Hong). Stochastic models have received a great deal of focus in hydrologic time series modeling. Compared to the wavelet modeling framework (WMF) (Sang, 2013), stochastic techniques such as Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA), and Auto Regressive Integrated Moving Average (ARIMA) are computationally intensive.

Momani (2009) has used ARIMA's seasonal components to make rain forecasts. The proposed model could not accurately predict the heaviest downpours. Stochastic models have been demonstrated to work effectively for a simple model with a small quantity of data, but as the dataset shows a high degree of polynomial, the simulation becomes more challenging to build (Wang et al., 2014). The systems may provide useful guidance to forecasters for operational decision-making during severe rainfall events, as shown by the study of Li and Lai (2004) in the field of operational hydrology. Each of these studies tried to build a prediction model by using past data. The factors required to anticipate rainfall are exceedingly complex and subtle, even over a very short time period. Pham et al. (2017) note that a classification and regression trees (CART) model may be used to predict shallow landslides when GIS is used to produce a map of likely land slide sites.

Both physical and data-driven models may be used to predict precipitation. Several research (Abbott and Marohasy, 2012, 2014) have shown that data-driven models exceed physical-model predictions. Data-driven models make use of a number of data-driven or statistical approaches to make predictions, whereas physical models mimic all of the basic physical processes in accordance with the rules of physics. Nanda et al. (2013) looked at and analyzed different ANN-based prediction models using time series data and found that Functional-link Artificial Neural Network (FLANN) was more successful for prediction than ARIMA (1,1,1).

Support Vector Machines (SVM; Vapnik, 1998) is a machine learning method based on the principle of statistical learning. The Tsengwen river basin was the focus of an investigation by Lin and Jhong (2015), who discovered that an SVM combined with a multi-objective genetic algorithm enhanced hourly typhoon rainfall estimates and might be utilized to help in disaster warning systems.

An assortment of data-driven approaches include artificial neural networks, autoregressive integrated moving averages, K-nearest neighbors, multivariate linear regression, and support vector machines for regression. Numerous research, including those by Abbot and Marohasy (2012, 2014), Deo and Sahin (2015), Farajzadeh et al. (2014), Karamouz et al. (2008), Mekanik et al. (2013), and Ramirez et al. (2005), have employed ANNs to accurately predict precipitation. Precipitation cannot always be accurately predicted by a single global model (Solomatine & Ostfeld, 2008).

Precipitation and other meteorological features are increasingly being predicted using models based on AI. The relationships between input and output variables are the primary focus of artificial intelligence models in order to provide reliable forecasts (Solomatine and Ostfeld, 2008). Deo and Sahin (2015) constructed an ANN model to forecast monthly rainfall and Evapotranspiration

Index for eight candidate stations in eastern Australia using training data from 1915 to 2005 and simulated data from 2006 to 2012. Unnikrishnan and Jothiprakash (2017) conducted a similar study in an effort to foretell future rainfall.

An auto regressive process is one that is reliant on both past information and a random perturbation. Moving average processes are linear, which is why white noises are linear. It is worth noting that ANN is similar to the auto regressive integrated moving average (ARIMA) model, which is also used for precipitation forecasting (Kaushik and Singh 2008; Kim et al. 2011; Narayanan et al. 2013).

According to Valipour et al. Monthly predictions of temperature and precipitation were made using the Box-Jenkins time series seasonal ARIMA method (Kaushik and Singh, 2008). Similarly, a research employed non-parametric Mann-Kendall and Cumulative Sum tests to locate patterns and transitions in the summer rainfall time series (Kim et al., 2011). The inflow of the Dez dam reservoir was predicted using ARMA and ARIMA models, and the results were compared to those obtained using static and dynamic ANN by Valipour et al. (2013). Wu and Lin (2017) conducted a research in which numerical prediction models were utilized to make predictions 1-3 days into the future based on the basic principles of the atmosphere.

Using rainfall predictors from LENS and Radar from 2017 to 2018, as well as machine learning tools LightGBM and XGBoost, a study (Lee et al., 2019) proposes a quantitative precipitation prediction scheme by machine learning to improve the practicality of Numerical Weather Prediction (NWP). An efficient early warning system for extremely short-term heavy rainfall has been suggested by a research (Moon et al., 2019) that makes use of machine learning methods. In 2019, Damavandi and Shah employed five machine learning approaches based on climatological factors such air temperature, geo-potential height, relative humidity, and elevation to estimate monthly rainfall for the Indus Basin. In this research, the Pearson correlation coefficient and the mean absolute error were used to evaluate the accuracy of the predictions.

Only a small number of research projects have combined many statistical and machine methods for a single dataset. In this paper, we combine LR, SVM, RT, GPR, and ERT regression methods in MATLAB to create a rainfall forecast model for five districts in the Indian state of Meghalaya by reducing root mean squared error (RMSE). An additional goal of this research is to determine the optimum number of predictors necessary to develop a reliable prediction model for the climatic time series.

2. Materials and Methods

2.1 Dataset

Locations between 250°00′N and 900°00′E, and 260°00′N and 920°00′E will be analyzed. The water portal of India was scoured for various time series datasets from the five districts of Meghalaya: i) Jaintia Hills (JH), ii) East Garo Hills (EGH), iii) West Garo Hills (WGH), and iv) East Khashi Hills (EKH), and v) West Khashi Hills (WKH). In all, eight measurements were made for this analysis. Rainfall (Rn), cloud cover (CC), maximum temperature (MaxT), lowest temperature (MinT), ground frost frequency (GFF), and the frequency of wet days (WDF). Rainfall was used as the goal variable while the other 7 characteristics were used as predictors in the training model developed using this data set. The information is structured according to the four seasons that Meghalaya experiences: spring, summer, autumn, and winter. All other monthly data characteristics were averaged out, but the rainfall totals were added together for each season. For the training set, we utilized data collected between 1901 and 1980, and for the testing set, we used data collected between 1981 and 2000.

2.2 Machine Learning Models

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to provide plausible forecasts of how to react to new information. Predictive models are built with the use of classification and regression in supervised learning. Continuous reactions may be predicted using regression methods. There are many different supervised and unsupervised machine learning algorithms available, and it might be difficult to decide which one to use. There isn't a silver bullet or universal solution. Sometimes you just have to try different things to find the appropriate algorithm. To further understand how various regression models play out in practice, you may utilize the Regression Learner App (RLA). A variety of regression models may be combined into one using a supervised machine learning technique. Fox (1997) introduced Linear Regression models; later, Solomatine and Dulal (2003) introduced Regression Trees; Rasmussen (2003) introduced Gaussian Process Regression models; Müller et al. (2005) introduced Support Vector Machines; and Vapnik (1995, 2000) introduced Ensembles of Regression Trees.

2.2.1 Linear Regression Models

Using one or more predictor variables, linear regression provides a description of a continuous response variable. It's useful for analyzing experimental, financial, and biological data, as well as understanding and predicting the behavior of complex systems. To develop a linear model, statisticians utilize linear regression. A dependent variable y (also known as the response) is described in terms of its connection to one or more independent variables Xi (also known as the predictors) in this model. The standard formula for linear regression analysis is:

$$y = \beta_0 + \sum \beta_i X_i + C_i \quad (1)$$

where β represents linear parameter estimates to be computed and ϵ represents the error terms.

All 4 linear regression models (linear, interactions linear, robust linear and stepwise linear) have easy interpretability where linear and robust linear models have very low flexibility. Regression Learner uses the *fitlm* function to train Linear, Interactions Linear, and Robust Linear models. The app uses the *stepwiselm* function to train Stepwise Linear models.

2.2.2 Regression Trees

Binary recursive partitioning is used to construct a regression tree; this is an iterative technique that first divides the data into partitions or branches, and then further divides each partition into smaller groups as the algorithm progresses up each branch. At the outset, each record from the Training Set (the already-classified data used to build the tree) is assigned to a single division. The program then starts dividing the data into the first two branches, splitting each field into binary if at all feasible. The algorithm chooses the division that results in the smaller squared sum of outliers relative to the mean. Each of the new branches is then subjected to the aforementioned splitting rule. This procedure is repeated until each node meets the minimum node size set by the user, at which point it becomes a terminal node. Even if a node is smaller than the required minimum size, it is still regarded a terminal node if its sum of squared deviations from the mean is zero..

2.2.3 Support Vector Machines

Vapnik (1995) proposed the supervised learning model known as the support vector machine (SVM). Using the Vapnik-Kuhn-Tucker (VKT) concept of structural risk minimization (SRM), SVM seeks a hyperplane in a high-dimensional space to divide data.

Vapnik's (1995, 2000) Chervonenkis (VC) dimension. SVM is a binary classification model used for a classification problem. The binary classifier presupposes the presence of two classes, each of which can be reliably distinguished by the decision surface. A multi-class problem may be solved by using a series of binary classifiers. There are two types of flags employed in this investigation. One represents an incident and zero represents the backdrop.

This is a high-level explanation of how SVM works. Take, for example, an n-dimensional vector x, with dimensions Rn, and a one-dimensional vector y, with dimensions 1,+1, as the sample data. By calculating the following equation, the SVM determines a hyperplane with a maximum margin.

$$\min \frac{1}{2}(w)^T . w + C \sum_{i=1}^{k} £_i \quad (2)$$

$$\text{s. t.} \quad y_i(w^T . \varphi(w_i) + b) \geq 1 - £_i, £_i \geq 0, i = 1, \dots, k$$

where $\emptyset(x_i)$ maps the input space to the feature space. C > 0 is a penalty factor that controls the trade-off between minimization of the classification error and maximization of the margin. w, b, and £ are optimized during the training phase.
The optimal decision surface can be determined by introducing the Lagrange multipliers and finally the classification function is represented as,

$$f(x) = \text{sgk}\,(y_i a_i k(x_i, x) + b^*)$$

where a_i is the support vector and b^* is the bias, $k\ (x_i, x) = \langle \emptyset(x_i), \emptyset(x) \rangle$ is the kernel function.

2.2.4 Gaussian Process Regression

Nonparametric kernel-based probabilistic models known as Gaussian process regression (GPR) models. Using the fitrgp command, a GPR model may be trained. Take the sample data (xi,yi); i=1,2,...,n, where xi is taken from the distribution Rd and yi is drawn from the distribution R. Predicting the value of a response variable ynew, given a new input vector xnew and the training data, is the problem that a GPR model seeks to solve. Formally, a linear regression model looks like this:

$$y = x^T \beta + \varepsilon \quad (3)$$

where $\varepsilon \sim N\ (0, \sigma^2)$. The error variance σ^2 and the coefficients β are estimated from the data. A GPR model explains the response by introducing latent variables, $f\ (x_i), i=1, 2, \dots, n$, from a Gaussian process (GP), and explicit basis functions, h. The covariance function of the latent variables captures the smoothness of the response and basis functions project the inputs x into a p-dimensional feature space.

A GP is a set of random variables, such that any finite number of them have a joint Gaussian distribution. If $\{f(x), x \in \mathbb{R}^d\}$ is a GP, then given n observations $x_1, x_2, \dots, x_n$, the joint distribution of the random variables $f(x_1), f(x_2), \dots, f(x_n)$ is Gaussian. A GP is defined by its mean

function $m(x)$ and covariance function, $k(x, x')$. That is, if $\{f(x), x \in \mathbb{R}^d\}$ is a Gaussian process,
then $E\ (f(x))=m(x)$ and $Cov[f(x), f(x')]=E[\{f(x)-m(x)\}\{f(x')-m(x')\}]=k(x, x')$. There are four types of GPR models: Rational Quadratic, Squared Exponential, Matern 5/2 and Exponential. Each model has hard interpretability and automatic flexibility to fit the dataset.

2.2.5 Ensembles of Trees

A regression tree ensemble is a predictive model composed of a weighted combination of (4) multiple regression trees. In general, combining multiple regression trees increases predictive performance. To boost regression trees using *LSBoost*, one can use *fitrensemble*. To bag regression trees or to grow a random forest (Breiman, 2001), *fitrensemble* or *TreeBagger* can beused.

3. Results and discussion

3.1 Seven parameters as predictors

A regression model was constructed using a total of seven factors to forecast precipitation. In addition, we use machine learning methods in the form of Regression (LR) models, Regression Ensembles of Regression Trees (ET), Regression Trees (RT), Gaussian Process Regression (GPR) models, and Support Vector Machines (SVM). Rainfall (Rn) was chosen as the response, and the input parameters (predictors) were the frequency of wet days (WDF), potential evaporation (PE), vapor pressure (VP), cloud cover (CC), maximum temperature (MaxT), minimum temperature (MinT), and ground frost frequency (GFF). Table 1 displays the RMSE values achieved by several machine learning techniques.

Table 1 RMSEs at different Machine Learning Regression models.

Spring		**Summer**	
District/Model	RMSE	District/Model	RMSE
Jaintia Hills		**Jaintia Hills**	
Stepwise LR	109.02	Linear SVM	392.91
Linear SVM	114.98	LR	432.42
Exponential GPR	116.83	Robust LR	427.28
East Garo Hills		**East Garo Hills**	
Linear SVM	46.52	Linear SVM	254.82
Quadratic SVM	47.29	Fine Tree	257.71
Exponential GPR	48.89	Robust LR	269.11
West Garo Hills		**West Garo Hills**	
Rational Quadratic GPR	34.72	LR	188.67
Squared Exponential GPR	34.40	Linear SVM	188.65
Matern 5/2 GPR	33.99	Medium Gaussian SVM	213.08
East Khashi Hills		**East Khashi Hills**	
Exponential GPR	127.59	Linear SVM	469.06
Rational Quadratic GPR	130.28	Exponential GPR	483.07
Medium Gaussian SVM	130.40	Matern 5/2 GPR	486.52
West Khashi Hills		**West Khashi Hills**	
Exponential GPR	87.87	Linear SVM	403.50
Matern 5/2 GPR	89.00	Boosted Trees	414.26
Rational Quadratic GPR	90.07	Bagged Trees	419.97

Three models with the smallest root-mean-squared errors (RMSEs) are chosen and listed in Table 1 to demonstrate the range of regression models. While LR and SVM yield minimal RMSE, GPR models produce the best answers in three districts (WGH, EKH, and WGH) throughout the spring. However, SVM results in the lowest RMSE across all jurisdictions during the summer months. Therefore, creating a prediction model using SVM and GPR was successful in these two seasons. Using these models, Fig. 1 displays the expected trends across all five regions during a twenty-year time span (1981-2000).

All anticipated district patterns are quite close to the actual district shapes. When compared to the RMSE of the training models, the RMSE of the predicted values is often within 5%. This indicates that the regression models have good predictive ability. Additionally, support vector machine was deemed most valuable in all districts to show its constancy throughout the summer. The dataset has extremely excellent correlations (Table 4), which is why the linear SVM method works uniformly well across all jurisdictions and can provide more accurate predictions.

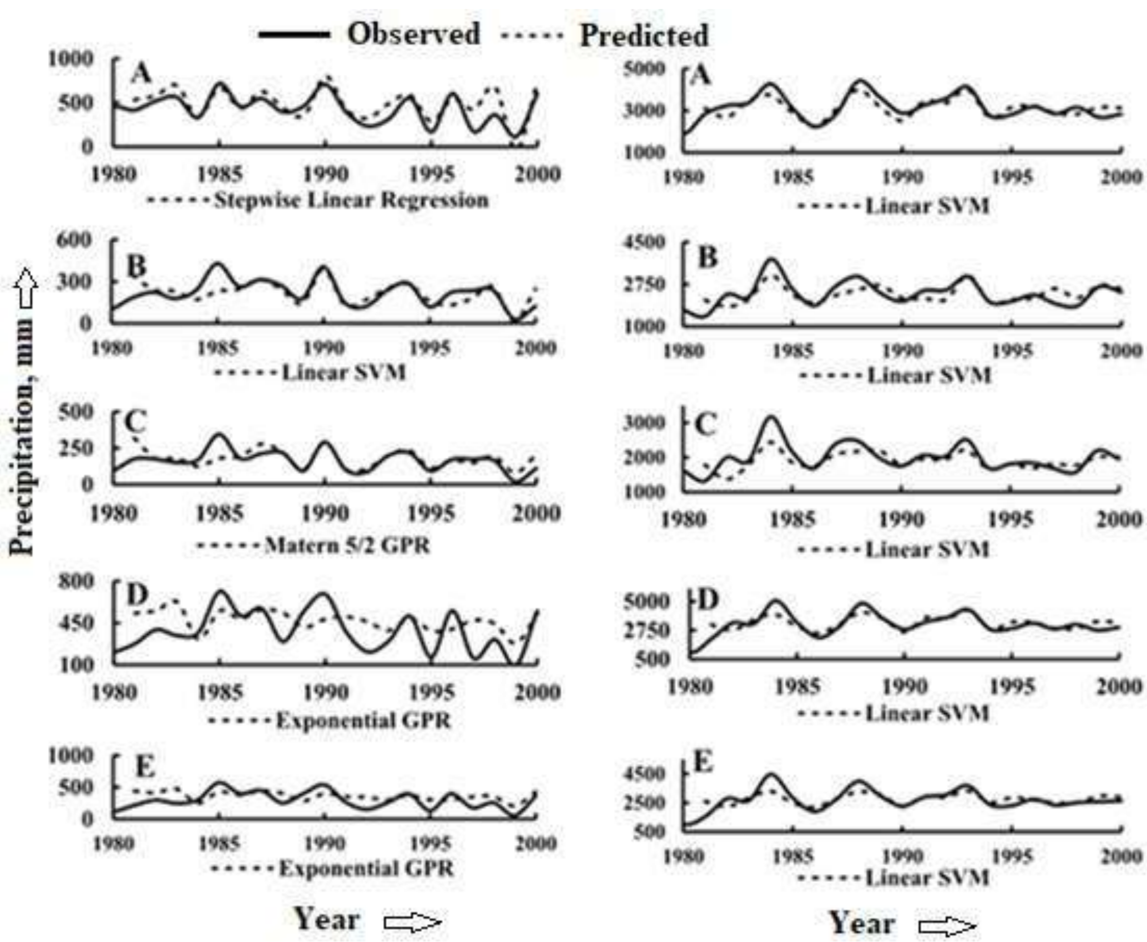


Fig. 1 Predicted rainfall for spring season (left) and for summer season (right) in different districts of

Meghalaya. *A= Jaintia Hills; B=East Garo Hills; C=West Garo Hills; D=East Khashi Hills; E= West Khashi Hills.*

Autumn and winter provide circumstances that differ from spring and summer. The RMSE values for each regression model in the fall and winter are shown in Table 2. One cannot point to a single model to observe preeminence in carrying out top-tier fall model. This occurred because rainfall patterns vary considerably between regions. In comparison to the other five districts, WGH received the least amount of rain while EKH received the most. Prediction accuracy is maximized by selecting models with the lowest root-mean-squared errors (RMSEs), and observable trends may be shown in Fig. 2. The situation during the winter months is essentially the same, with no universal model being suitable for use across all regions. So, instead of focusing on seasonal analysis for fall and winter, we tried some new things with our models. Different districts were successful in using SVM, RT, and GPR to forecast rainfall.

Table 2 Variations of RMSE using different predictive Machine Learning algorithms

Autumn		Winter	
District/Model	RMSE	District/Model	RMSE
Jaintia Hills		**Jaintia Hills**	
Linear SVM	110.00	Medium Tree	20.56
Exponential GPR	123.37	Bagged Trees	24.44
Bagged Trees	126.77	Boosted Trees	23.20
East Garo Hills		**East Garo Hills**	
Boosted Trees	73.60	Medium Tree	12.56
Fine Tree	79.80	Linear SVM	12.66
Bagged Trees	79.80	Robust LR	12.36
West Garo Hills		**West Garo Hills**	
Linear SVM	66.07	Squared Exponential GPR	11.47
Boosted Trees	66.19	Rational Quadratic GPR	11.47
Bagged Trees	68.40	Matern 5/2 GPR	11.53
East Khashi Hills		**East Khashi Hills**	
Boosted Trees	129.27	Medium Tree	24.25
Exponential GPR	135.47	Robust LR	24.65
Fine Tree	135.73	Linear SVM	24.78
West Khashi Hills		**West Khashi Hills**	
Robust LR	114.58	Stepwise LR	17.51
Exponential GPR	111.40	Squared Exponential GPR	19.35
Linear SVM	113.44	Matern 5/2 GPR	19.36

3.2 Efficiency of parameters

Seven indicators were used as predictors in this analysis. It is possible that a more sensitive parameter or set of parameters might emerge through developing several models with varying values for those parameters (Bobby et al., 1999). The sensitivity analysis in this research was conducted by varying the values of the predictors. This process was performed many times in order to create the best model possible with the fewest possible predictors.

Parameter effectiveness was evaluated by selecting data from Table 3. When using all 6 parameters as predictors, there were 7 possible permutations, and the RMSE for each machine learning model was unique. Take, for example, the effectiveness of We did WDF with the following variables selected as predictors: PE, VP, CC, MaxT, MinT, and GFF. RMSE in each region is largest when WDF is not included in the set of predictors for that season, as summarized in Fig. 3. Therefore, WDF is the most sensitive of this bunch.

.

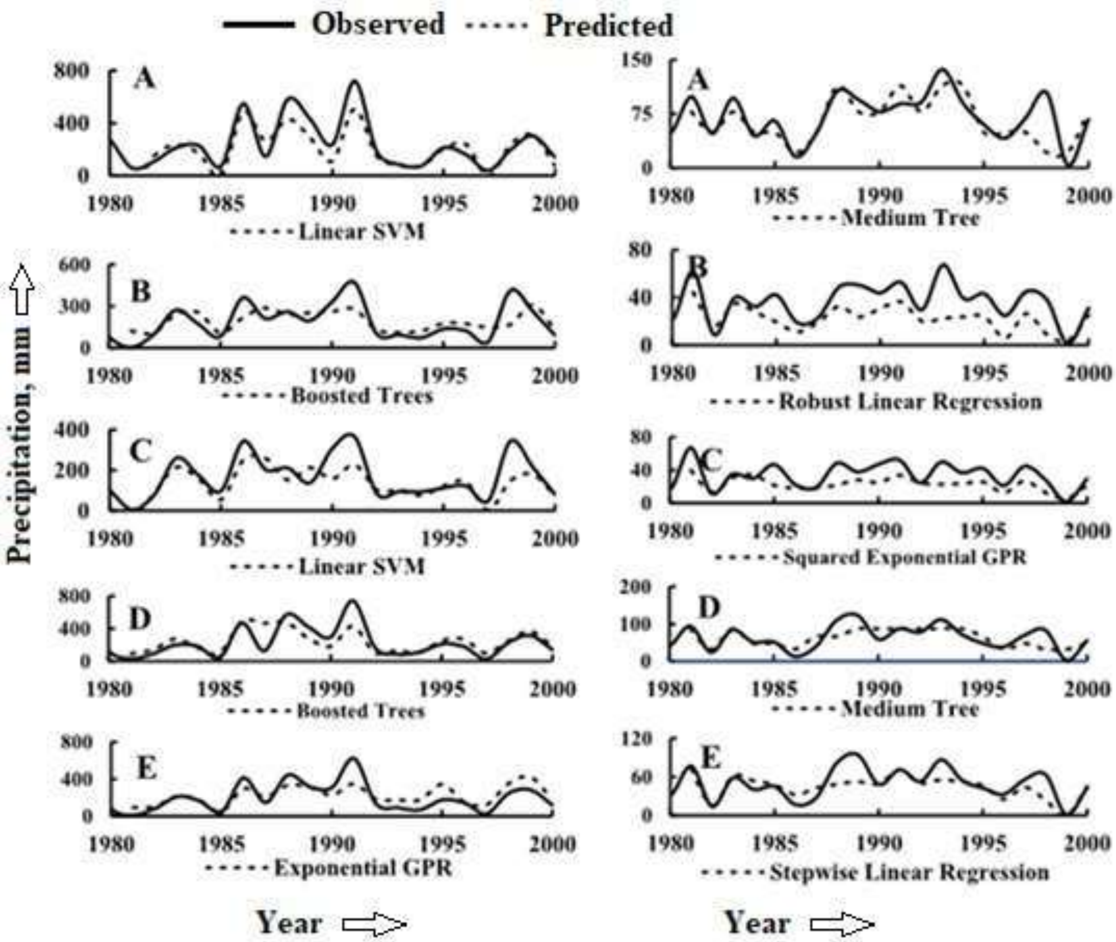


Fig. 2 Predicted rainfall patterns for autumn (left) and for winter (right) season in different districts of Meghalaya. *A= Jaintia Hills; B=East Garo Hills; C=West Garo Hills; D=East Khashi Hills; E= West Khashi Hills.*

Table 3 Range of iterations for different numbers of predictors.

No. of predictors	Response	No. of Combinations
06	Precipitation	07

05	Precipitation	21
04	Precipitation	35
03	Precipitation	35
02	Precipitation	21
01	Precipitation	07

The same procedure was repeated for 5, 4, 3, 2 and 1 predictors. The number of iterations is different for each group of predictors (Table 3). The details are checked thoroughly for each generation so that the optimum number of predictors can be achieved. Machine learning models are verified by considering lowest RMSE for every single iteration. Entire simulation comes to the decision that wet day frequency (WDF) itself as a single predictor gives the best model considering minimum RMSE. For all the other cases and groups of predictors the response in respect to RMSE was not satisfactory. So, WDF is the most efficient and single predictor to develop the machine learning models for prediction of rainfall.

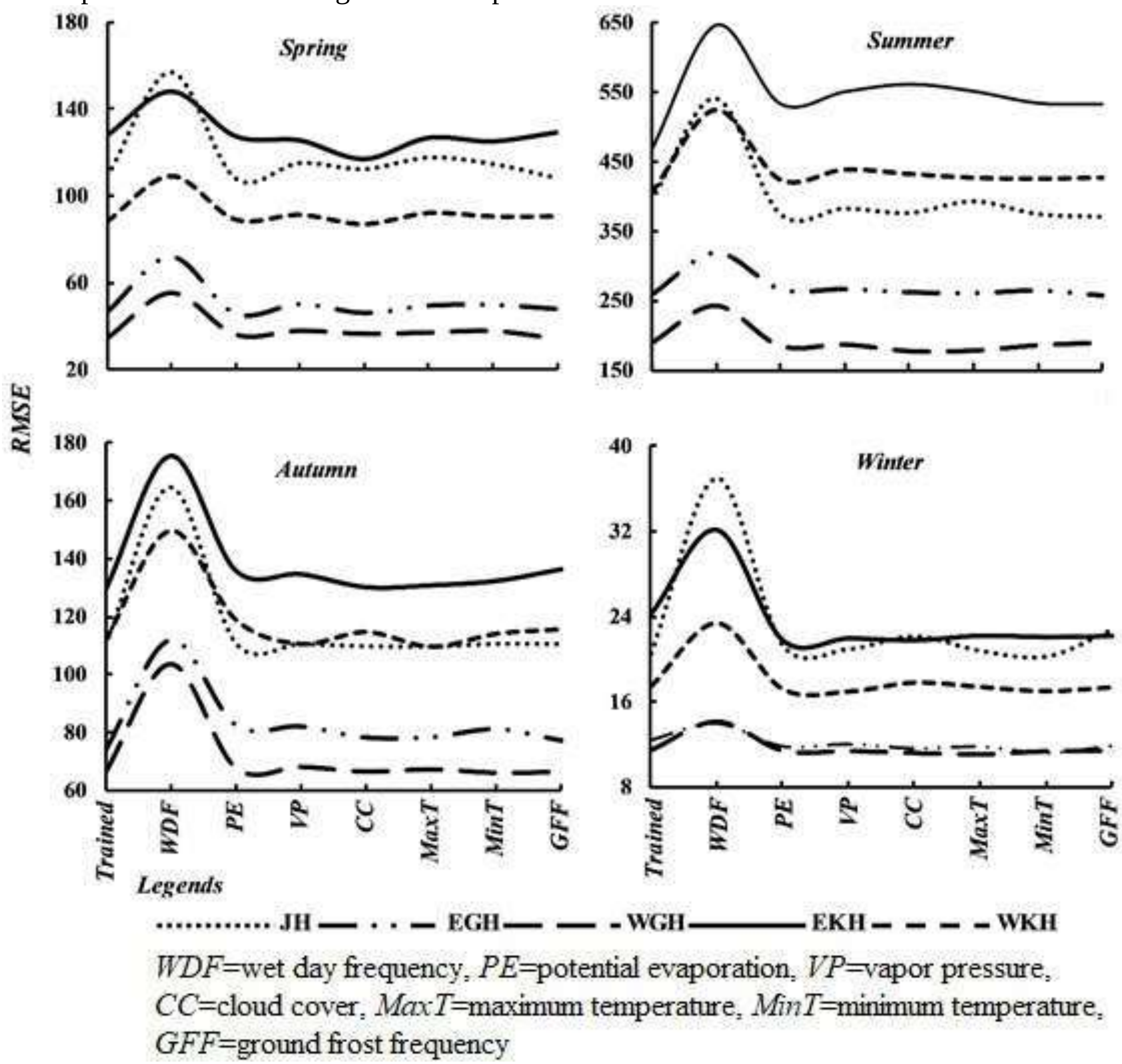


Fig. 3 Seasonal variations of RMSE from district to district.

3.3 Efficiency of individual machine learning algorithm

Several different machine learning algorithms were compared in this research to see which ones performed the best. Figure 4 summarizes the results of an evaluation of the accuracy of simulations using root-mean-squared error (RMSE) values generated by Linear Regression (LR) models, Regression Trees (RT), Gaussian Process Regression (GPR) models, Support Vector Machines (SVM), and Ensembles of Regression trees (ET). When the ET algorithm failed to create a useful model for any of the EGH, WGH, or WKH districts, the Linear Regression

(LR) method produced the most models (17 times). The WGH district had the maximum number of models produced (18) by SVM, whereas the WGH, EKH, and WKH districts saw no models built by RT. Integrating the four

LR and SVM were the 53rd and 52nd best models for predicting rainfall during the corresponding seasons (Fig. 4).

Figure 5 provides an alternative justification for narrowing down a machine learning method based on the magnitude of its RMSEs. In particular, SVM built 17 models across all seasons in the WGH area, and produced 23 top models alone for all districts in the summer. Figure 5 displays the scatter of RMSEs for the best models. The RMSE is low in the winter and spring, but it varies widely (from 200 to 500) throughout the summer. Experiments show that predictions work best in the winter and spring.

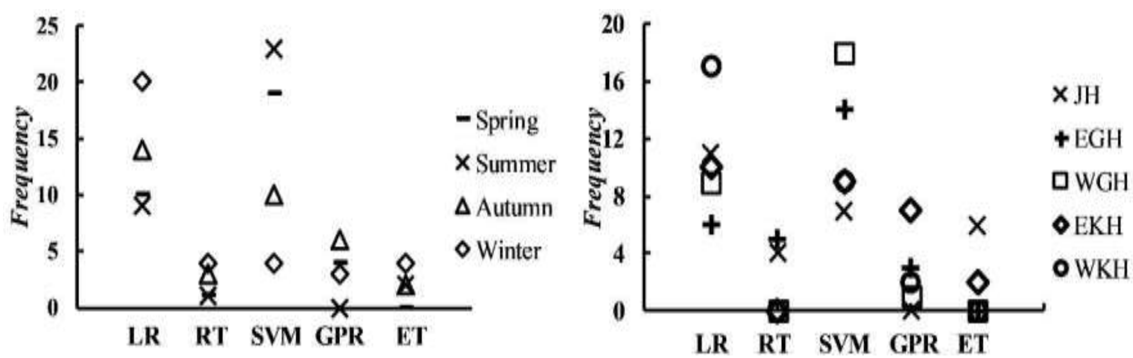


Fig. 4 Rate of execution by all machine learning algorithms.

3.4 Prediction using the efficient predictors

The above analyses show that WDF is the best predictor to build the prediction models. Consequently, WDF based best models are used to predict the rainfall from 1981 to 2000. The patterns are very finely matching with the observed values. Fig. 6 and Fig. 7 present those trends of predictions. However, the study points out the LR and SVM algorithms again worked very well to generate prediction models of perfect regressions. The patterns of prediction are matching with the actual data.

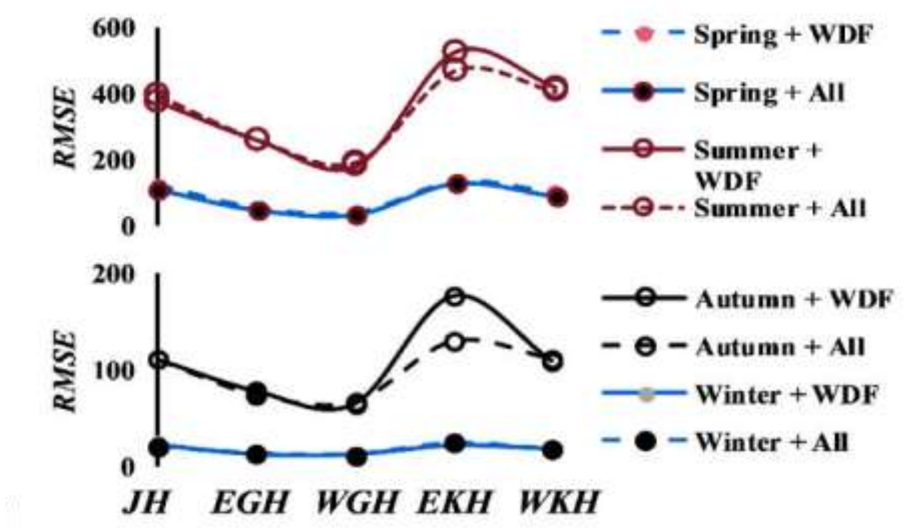
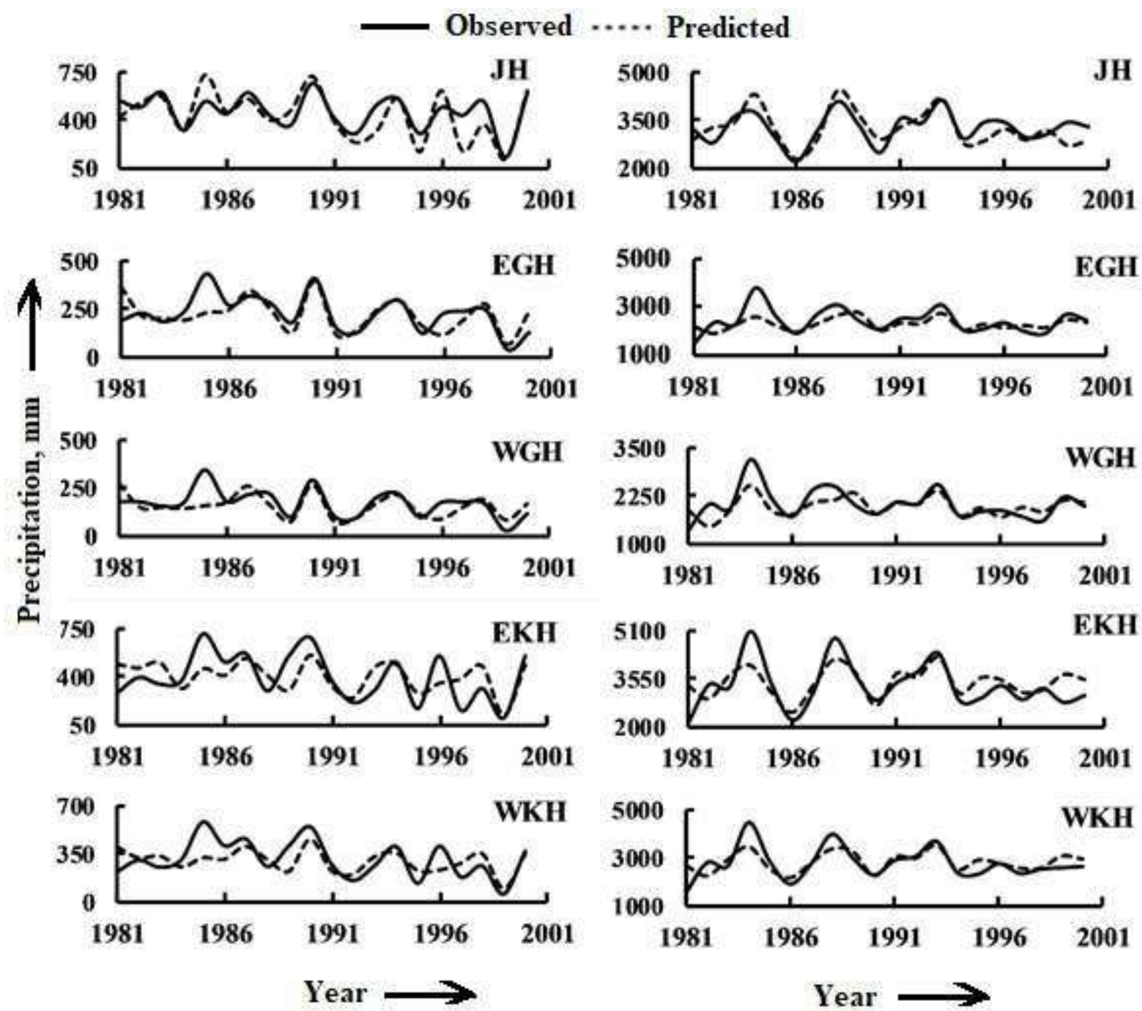
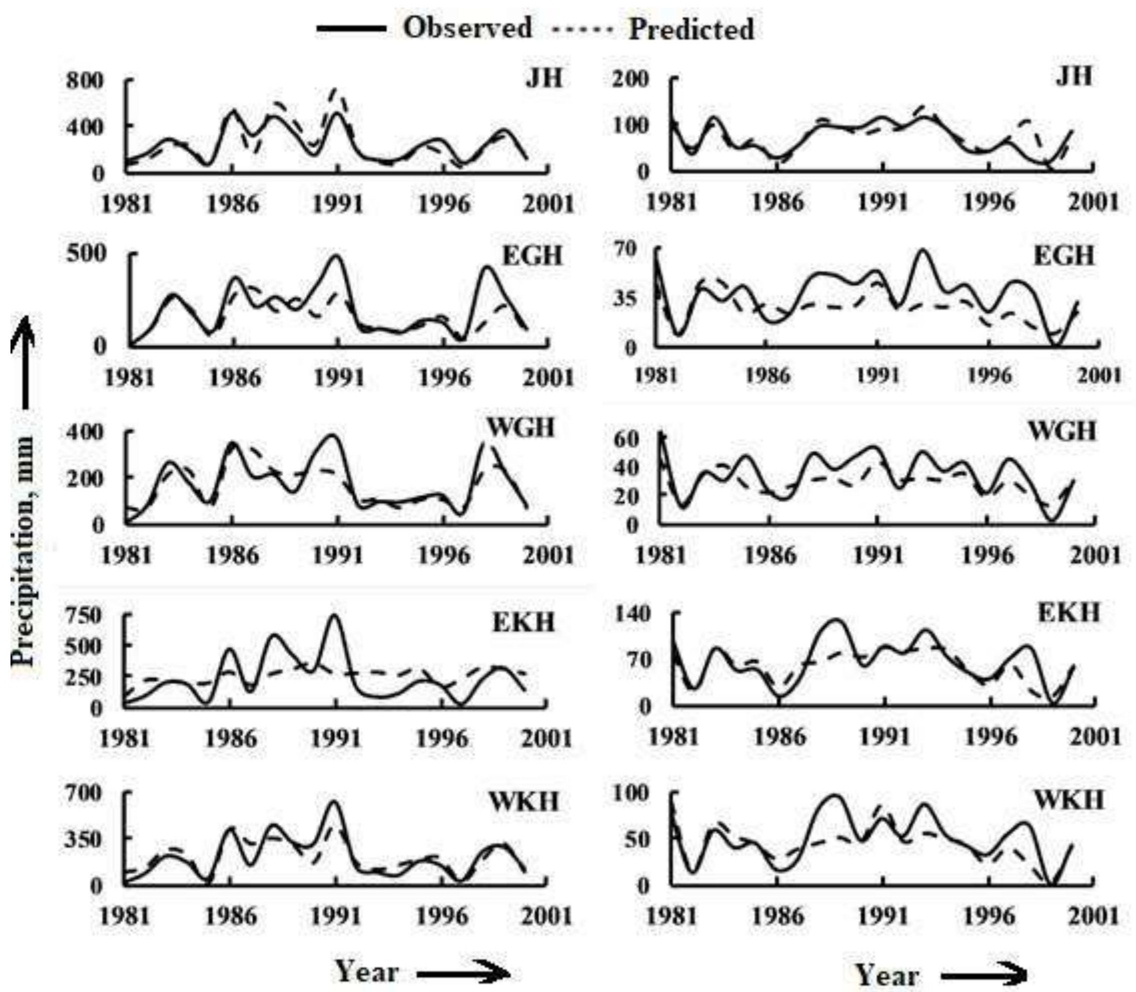


Fig. 5 RMSE values compared within five districts



Legend: *JH=Jaintia Hills; EGH=East Garo Hills; WGH=West Garo Hills; EKH=East Khashi Hills; WKH= West Khashi Hills.*

Fig. 6 Prediction of precipitation using *WDF* as a single predictor for Spring (left) and for Summer (right).



Legend: *JH=Jaintia Hills; EGH=East Garo Hills; WGH=West Garo Hills; EKH=East Khashi Hills; WKH= West Khashi Hills.*

Fig. 7 Prediction of precipitation using the models generated from WDF predictor for Autumn (left) and for Winter (right).

The whole analyses showed that there are suitable machine learning algorithms to establish precise models for prediction of rainfall in each district. However, the accuracy of the model depends on the trend of the training data where RMSE should be minimum for a better prediction. Support vector machine, Gaussian process regression techniques and various Tree models have been used to the dataset and the results are quite satisfactory.

4. Conclusion

The time series of five districts in the Indian state of Meghalaya was analyzed to determine the best possible prediction model. The prediction models were field-tested over all four seasons. Despite the diversity of the five regions, the analysis identified a subset of machine learning models that performed very well when used to execute prediction models, and the research also revealed that the aforementioned supervised machine learning algorithms are highly pattern-consistent.

excellent reaction, hence the recognition serves the aim of prediction as well. The most successful methods for creating predictive time series models across all areas of research are linear regression (LR), support vector machine (SVM), and Gaussian process regression (GPR). Wet day frequency alone is the best of seven criteria used to create a reliable regression model for forecasting.

References

John R. Abbot & Joseph M. Marohasy 2012. Rainfall predictions in Queensland, Australia using artificial neural networks. Advanced Atmospheric Science, 29(4), 717–730.

According to Abbot and Marohasy (2014). Using artificial neural networks, we optimize the selection of inputs for monthly rainfall forecasts in the Australian state of Queensland. 138, 166-178 (Atmos. Res.).

In Breiman (L.) and 2001. Forests, Random, Machine Learning, 4, pp.

R.C. Deo & M. Sahin 2015. Predicting monthly standardized rainfall and evapotranspiration index in Eastern Australia using artificial neural network model, hydrometeorological parameters, and climatic indicators. 161-162, 65-81. Atmos. Res.

A Learning Framework for Accurate Prediction of Rainfall Rates, Damavandi, H.G., and Shah, R. (2019). a preprint available at arXiv:1901.05885.

Fox, J., 1997. Practical Uses of Regression, Linear Models, and Related Techniques. Specifically, Sage Publications, Inc., of Thousand Oaks, California.

Hong, W., 2008. Machine learning-based rain forecasting technology. The Journal of Computational and Applied Mathematics, 200(1), 41-57.

2014 Farajzadeh, A. Fard, and S. Lotfi. Using a "feed-forward neural network" and a "time series analysis" model, we predict the monthly precipitation and runoff of the Urmia lake basin. 7(B):38-48 Water Resour. Ind.

Case study of long-term seasonal rainfall forecasting using time-delay recurrent neural networks, by M. Karamouz, S. Razavi, and S. Araghinejad, 2008. 22:2(229)-241 Hydrol. Process.

Singh, S., and I. Kaushik 2008. Using a seasonal ARIMA model, we can anticipate monthly precipitation and temperature changes. Journal of Environmental Research and Development Volume 3 Number 2.

Evaluation of temporal-spatial rainfall variability and prediction using a seasonal ARIMA model in Mongolia. Kim B, Hossein S, Choi G. 2011. 15(5), 917–925, KSCE Journal of Civil Engineering.

Y.-M. Lee; C.-M. Ko; S.-C. Shin; and B.-S. Kim. (2019). The Application of Machine Learning to the Correction of Rainfall for Hydrological Purposes. 28:1 (January 2015), pp.125–135 in the Journal of International Environmental Science.

P.W. Li and E.S.T. Lai 2004. Quantitative forecasting of Hong Kong's precipitation over the next several days,
288(1-2), pp. 189-209 in Development.

Lin, Guang-Fu, and Boon-Chi Jhong. 2015. Typhoon rainfall distribution forecasting model that operates in real time. Hydrology Research 521:302–313.

Elmahdi, A.; Mekanik, F.; Imteaz, M.; Gato-Trinidad, S.; 2013. Predicting future rainfall utilizing large-scale climatic patterns using a multiple regression and artificial neural network. Hydrology, 503:11-21 (J.).

P. E. N. M. Momani (2009). Time series analysis model applied to precipitation records in Jordan. Science 5 (American Edition), 599-604.

Moon, S.-H., Kim, Y.-H., Lee, Y.H., & Moon, B.-R. (2019) Applying machine learning to a flash flood early warning system. Hydrology, 568, 1042–1054 (Journal).

2013 S. K. Nanda, D. P. Tripathy, S. K. Nayak, and S. Mohapatra. Artificial neural network (ANN) models for predicting precipitation in India. Journal of Intelligent Systems and Applications, Volume 5 Issue 2:1 (International).

To cite this article: Narayanan, P.; Basistha, A.; Sarkar, S.; Kamna, S.; 2013. We use trend analysis and ARIMA modeling to examine western India's pre-monsoon rainfall data. 345(1), 22-27, of Compt. Rendus Geosci.

In 2005, Ramirez, M., de Campos, V., Haroldo, F., and

Ferreira, N. Rainfall predictions for the Sao Paulo area using an artificial neural network. 301(1), 146-162 in J. Hydrol.

D. P. Solomatine & K. N. Dulal (2003). Tree models outperform neural networks for estimating runoff from precipitation, Biological Systems, 48(3), 399-411.

Ostfeld, A., & Solomatine, D. P. (2008). Data-driven modeling: lessons learned and future directions. 10(1-22)Journal of Hydroinformatics.

Y.-F. Sang. 2013. Foreseeing hydrologic time series using an enhanced wavelet modeling framework. 27(2807-2821)Water Resour. Manage.

2017;Unnikrishnan, P., & Jothiprakash, V. Predicting daily rainfall using data that has been preprocessed using singular spectrum analysis. Hydroinformatics: a Journal.

Published in 2013: Valipour, M., M. Banihabib, and S. Behbahani. Using the Dez dam reservoir as an example, we evaluate the accuracy of monthly inflow forecasts using three different artificial neural network models: autoregressive, moving average, and ARMA. Water Research 476:433–441.

Miche, Y. van Heeswijk, T. Lindh-Knuutila, P. A. J. Hilbers, T. Honkela, E. Oja, and A. Lendasse 2009. Extreme learning machine ensemble models with adaptive parameters for time series forecasting. Comp. Sci. Lecture Notes 5769:305-314

V. N. Vapnik (1995). Introduction to the Theory of Statistical Learning. New York: Springer. In 1998, Vapnik, V.N. The Theory of Statistical Learning. Published by John Wiley & Sons in New York.

(2000). Vapnik, V., and Chapelle, O. Neural Computation, 2(9), 2013-2036, Bounds on Error Expectation for Support Vector Machines

Zhang, P.-P., Pang, S.-S., & Wang, J.-Y., 2014. Analyzing the single-spectrum rainfall in the vicinity of Poyang Lake. 955-959, 3006-3010. Adv. Mater. Res.

(2017). Wu, M.C., and Lin, G.F. The use of an ensemble numerical weather prediction system in Taiwan to provide short-term predictions of precipitation across a hilly watershed. 546, 60–70, J. Hydrol.

Citation: Yang Y., Lin H., Guo Z., Jiang J. Analysis of satellite picture sequences as a data mining technique for predicting severe rainfall. Computers and Geosciences 33 (1), 20-30.