Received: July 8, 2018

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Time Series Analysis of Large Scale Rainfall Data Using Regression Automata Models

Tulasi Sunitha Manepalli [1]* **Chamakuzhi Subramanian [1]**

[1]*Department of Computer Science & Engineering, Jain University, India*
* Corresponding author's Email: tulasi80@gmail.com

Abstract: Rainfall is considered as the most important phenomena of the climate system. Due to the lack of adequate irrigation facilities, agriculture becomes vulnerable, which is the backbone of a country's economy. The rainfall can be able to predict by using selective appropriate predictors. Though several models have been developed for forecasting and predicting in Time Series (TS), there is no ideal model to predict the rainfall. In recent years, Automata is useful for forecasting and prediction of hydrological TS because automata help to predict the rainfall from the uncertainty data. The motivation of this work is to design a reliable tool for predicting daily rainfall in advance using Regression Automata (RA) models. The proposed method uses three different RA models for predicting rainfall from the collected data for four stations in Queensland State. The results clearly show that the all the three RA models can predict the rainfall very efficiently in various terms such as error rate, coefficients and mean square error.

Keywords: Automata, Forecasting, Logistic regression automata model, Rainfall prediction, Time-series, Regression model.

1. Introduction

In several years, operational Numerical Weather Prediction (NWP) and climate prediction models have evolved tremendously the continuous improvement of computing technologies and of the underlying algorithms at the foundations of these models [1]. The prediction models for climate and NWP are transitioning to more sophisticated Earth System Models (ESM) as the major challenges faced by these algorithms [2]. The time integration drives several key aspects of a prediction model or NWP such as solution accuracy, the effectiveness of uncertainty quantification, time-to-solution, energy or money-to-solution and robustness [3]. The highest goals of the climatic prediction model or NWP is the quantification of its uncertainty bounds and the forecast accuracy. In addition, the NWP requires requirement on the timeliness of delivery of the forecast [4]. The weather can be predicted by using Time-Series Data Mining (TSDM), the time-series forecasting problems have attracted wide attention to solving the problems by providing a way to explore past behaviour in the future. Recently, various technologies have been proposed for time-series forecasting [5]. TS is a sequence of data which is associated with time, such as daily temperature measurement. In many application areas such as environmental, economic, finance, and medicine, a stretch of values on the same scale indexed by a time will occur naturally [6].

The aim of the Time Series Analysis (TSA) is to formulate TS data in order to gain knowledge, fit low dimensional models, and make forecasts. The TS is a non-negative and precisely different way in various fields. However, there are many challenges to be faced when it involved application area as an example in manufacturing and weather [7]. Researcher in manufacturing tries to survey any techniques that have been proposed specifically for modelling and processing TS for temporal data. The determination of knowledge from TS can be used in weather precipitation or prediction by the researcher for human being benefits [8]. Rainfall-Runoff (R-R) modeling is an important topic in hydrology research. It aims to capture the R-R association and understand its process. A few examinations have closed the prevalence of time series information over other information driven models, for example, Artificial Neural Network (ANN), Fuzzy strategies, Auto-Regressive Integrative Moving Average (ARIMA), and Auto Regressive with exogenous data sources (ARX) models. The main obstacle of utilizing ANN is that the information investigation model will be unable to be translated. The conventional fuzzy frameworks don't have any learning algorithm to assemble the examination demonstrate [9, 10]. Heavy rainfall has been identified as the most common trigger of shallow landslides and slope instabilities worldwide [11]. For these types of landslides, one of the measures for risk mitigation is the adoption of early warning systems depends upon rainfall thresholds that identify the critical amount of precipitation for landslide triggering [12]. The empirical rainfall thresholds are used in an early warning system of landslides for forecasting the possible occurrence of rainfall-induced landslides [13]. Cellular Automata (CA) model as a self-organizing approach provides a simulation for predicting the rainfall through simple local interaction rules [14].

The aim of the proposed work is to design an effective representation technique for time-series with dimensionality reduction and also design an efficient Automata based framework for Modeling Time series. The method of performing time-series mining task on climate data based on the proposed approach and comparing it with existing techniques to predict the rainfall. Enormous data are present for the forecasting of data and dimension reduction technique is used to produce the required data. Python is used to forecast the rainfall from the acquired data. The remaining paper is discussed as: Section 2 gives the description of the models analyzed by various researchers that relate to this study. Section 3 provides a description of the development of the proposed methodology used for predicting the rainfall. Sections 4 present the results obtained by various experiments and the conclusions are made in Section 5, respectively.

2. Literature review

An earlier research on several automata techniques for predicting the rainfall is described below. In this scenario, brief evaluations of some important contributions and limitations are presented.

J.C. Bennett, D.E. Robertson, P.G. Ward, H.P. Hapuarachchi, and Q.J. Wang [15] established an efficient hourly rainfall-runoff model by testing the hypothesis which is the combination of hourly streamflow data with Simple Disaggregation (SD) of daily rainfall data. The framework is tested on a range of mesoscale catchments (150-3500 km2) with four rainfall-runoff models. When sub-daily rainfall data are unavailable, the establishment of hydrological models for continuous streamflow forecasting systems by testing the SD for the models and small catchments. Though the SD is a very straightforward and effective way to predict, this model will not work in small catchments because information from sub-daily rainfall is likely to play a major role in the identification of model parameters.

A. Douinot, H. Roux, and D. Dartus [16] implemented an objective function called Discharge Envelop Catching (DEC) for RR model calibration. The DEC met the two major objectives such as enabling the end-user to define an acceptable uncertainty for each part of the simulated hydro-graph and considering the uncertainty of discharge observations. The approach DEC was used to flash the floods which were explained on MARINE, an existing hydrological model. The demonstration made by the existing approach highlighted the objective function of DEC by identifying the strength and weakness of the DEC framework.

M. Espínola, J. A. Piedra-Fernández, R. Ayala, L. Iribarne, S. Leguizamón, and J. Z. Wang [17] developed a new algorithm called RAinfall with Cellular Automata (RACA) for simulating the water evaporation, groundwater flow in 3D satellite images and weather phenomena of rainfall. The RACA allowed the users to make decisions by obtaining from the 3D results, simulation and numerical related to the cumulative flow and maximum level of water. The approach RACA concentrated on major issues such as how the climatic conditions affect the water level in a particular area, estimating the future water supply of a population, establishing future construction projects and urban planning away from locations with the high probability of flooding.

M. Alvioli, and R. L. Baum, [18] implemented the Transient Rainfall Infiltration and Grid-based Regional Slope-stability model (TRIGRS) for the distribution of rainfall-induced shallow landslides. Within the message passing interface framework, the TRIGRS four time-demanding execution modes were parallelized, namely both the unsaturated and saturated model with infinite and finite soil depth options. The results were obtained both on a high-performance multi-node machine and on commercial hardware showing different limits of applicability of the new code. The number of runs required to collect one statistically significant set of results was certainly a limitation in the TRIGRS model, and the speed of the code represented a significant variable.

M.R. Bendre, R.R. Manthalkar, and V.R. Thool [19] proposed an approach was used for predicting monthly and daily weather data in scarcity zone. The linear and polynomial regression methods were modeled with an iterative approach for the purpose of predictions. The experimental results demonstrated that the proper choice of input parameter with the approach produced good accuracy and season-wise temperature, humidity and rainfall conditions. The effectiveness and performance of the method were analyzed by statistical measure and testing strategies. The regression method was lacking by a better performance of general predictions due to the extension of the data size limit.

M. Herrera, A.P. Ramallo-González, M. Eames, A.A. Ferreira, and D.A. Coley [20] introduced a new Location-Specific Mortality Risk (LSMR) focused definition of heat waves. The new mathematical model was also proposed for the creation of TS that represent the location-specific mortality. When the temperature was high during the day and night, the LMSR focused on identifying periods which was strongly linked to mortality. The representation of extremes was provided by the LMSR method than the common reference weather files. The drawback of the LSMR framework selected the month in such a way that the Design Summer Year (DSY) represented a slightly warmer than average summer, but contained no heat waves and thus selected weather caused excess mortality.

To overcome the above issues, the proposed method introduced three regression models such as linear regression, support vector and logistic regression model. The database is huge and there may some data missing in the dataset. So, the Pre-processing is used to normalize the missing data in the database. Once the preprocessing is done, then the data is sent to predict the rainfall.

3. Proposed methodology

Prediction of rainfall is still a huge challenge to the climatologists. Most of the burning issues of our time like global warming, floods, drought, heat waves, soil erosion and many other climatic issues are directly related to rainfall. Agriculture is still the major source of economic activities in the most of the countries of the world. Therefore, predicting the rainfall correctly is very important, nowadays, most of the linear models and the findings were inconclusive. Fig. 1 shows the proposed methodology of our work. In this paper, the approach employs the three different RA techniques to predict rainfall. The prediction of outcome (e.g. presence or absence of rainfall) is obtained by using the RA models based on values of a collection of predictor variables.

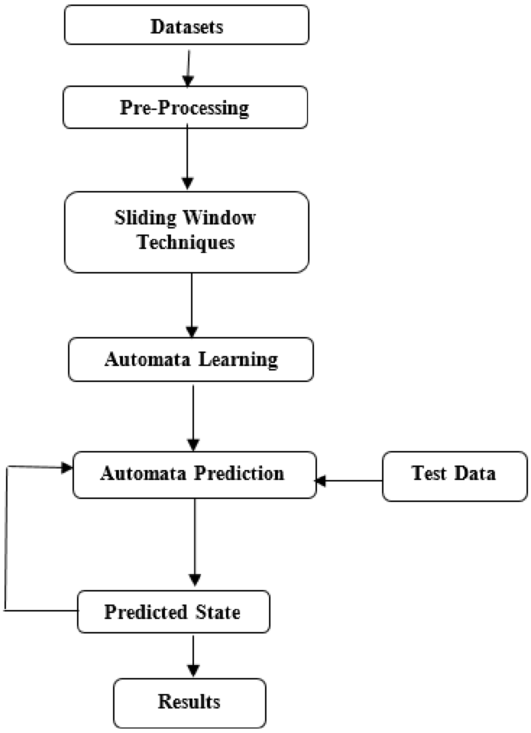

Figure.1 Basic structure of the proposed methodology

3.1 Preprocessing

The unwanted columns from the dataset are filtered and the only day, year, month and rainfall are taken. Day, month, year is combined into a singular column of data ranging from 1954 January 1 to 2015 December 31 and its corresponding rainfall are tabulated. Then, the rainy days and non-rainy days are labelled as 0 and 1 using a threshold of rainfall above zero.

3.2 Sliding window techniques

One of the key problems in the learning task is to determine the length of the sliding window, i.e., how many historical data points the prediction would rely on. E_{in} and E_{out} are the fitting mean square error in the training data and testing data respectively to process the data in parallel over different states. One can infer that by increasing the parallel processing complexity (sliding window length), E_{in} decreases sharply, while E_{out} becomes increasingly worse, which is typically the result of overfitting. In practice, it favours simpler models in order to minimize the risk of overfitting. The models, of which window length is less than 5, have relatively small E_{out}.

3.3 Automata learning

The subjects of CA deal with large collections (usually infinite in order to avoid boundary problems) of interconnected finite automata, each finite automaton being thought of as a cell. The approach uses the Wolfram's classification of CA for predicting the rainfall in three stations of QLD state. The Wolfram's classification [21] is predicting the detailed properties of a particular CA, it is often enough just to know what class the cellular automaton was in. The second problem is that Turing universal computation and the possible relation between the complexity of CA are tackled by the Wolfram's classification.

The analysis of Wolfram's includes a one-dimensional (1D) study, order ($k = 2; r = 2$), where $r \epsilon Z$ the number of neighbours and $k \epsilon Z$ is the cardinality of the finite alphabet and find the behavior of the same classes in other CA rule spaces.

In a 1D array, a finite automaton called an Elementary Cellular Automaton (ECA) is well defined. The automaton updates two states and also the closest neighbors' state in discrete time depends on its own state and synchronously, all cells updating their states.

Wolfram's classes can be described as:

- Class I. CA evolving chaotically.
- Class II. Includes all previous cases, known as a class of complex rules.
- Class III. CA evolving periodically
- Class IV. CA evolving to a homogeneous state.

Otherwise explained, in the case of a given CA,

1. The evolution is dominated by sets of cells without any defined pattern for any random and longtime initial condition, then it belongs to Class I.
2. The non-trivial structures dominated the evolution of emerging and traveling along the evolution space. These spaces are periodic, chaotic or uniform can coexist, then it belongs to Class II. This class is frequently tagged such as simply complex, complex behavior, complexity or dynamics.
3. The blocks of cells dominated the periodically repeated evolution for any random initial condition, then it belongs to Class III.
4. Class IV contains the unique state of its alphabet dominated the evolution for any random initial condition.

3.4 Regression automata model

The estimation of relationships among variables is evaluated by the RA which is a set of statistical processes. The focus of the RA is on the relationship between an independent variable or dependent variable and the analysis includes various techniques for evaluating and modeling several variables. The use of RA has a substantial overlap with the field of machine learning that is used for forecasting and prediction. RA explores the forms of relationship between independent and dependent variables. In this work, the regression analysis model is used to predict the daily rainfall prediction for the four stations. There are three regression models are used to predict rainfall such as Linear Regression Automata model (LRA), Support Vector Regression Automata model (SVRA), and Logistic Regression Automata model (LORA).

3.4.1. Linear regression automata model

LRA is a method used for defining the relationship between one or more independent variables or explanatory variables, denoted by (X) and a dependent variable (Y). For multiple explanatory variables, the process is defined as Multiple Linear Regression (MLR).

The general equation for an LRA is given as in Eq. (1),

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

Where y denotes the dependent variable (rainfall) and x_i where i= 1,2,...,n, denotes independent variables and $\beta$is called the intercept.

3.4.2. Support vector regression automata model

The SVRA model is a learning approach firstly used in the pattern recognition problems. Later on, it was modified and used in the regression problems. The SVRA is considered as a thriving algorithm in the learning problems. In the regression problems, training procedure includes obtaining the correlation or non-linear mapping function f(x) between both learners (i.e. input and output of the learner). The SVR [22] aims to provide a non-linear mapping

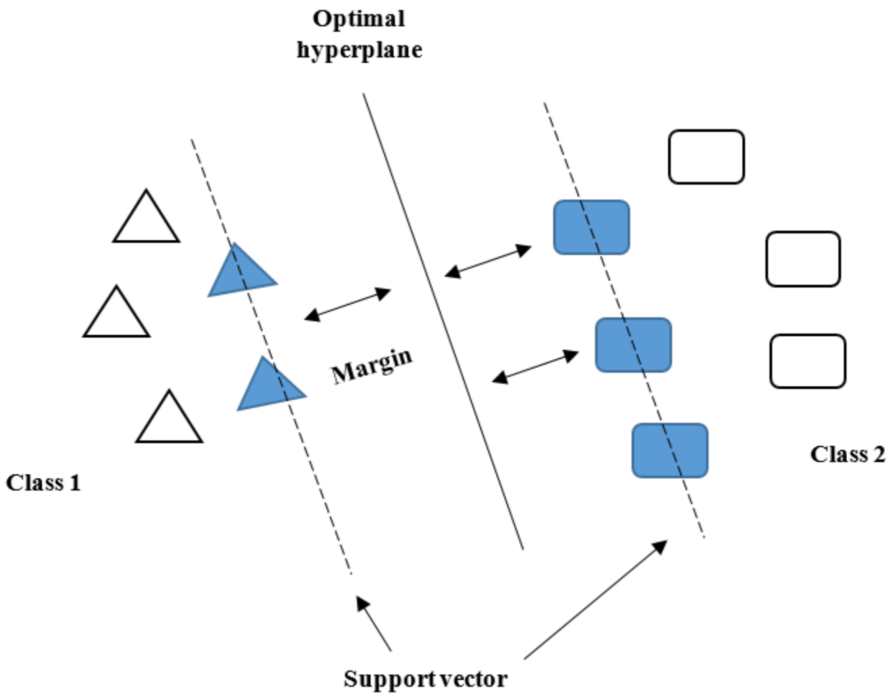

Figure.2 The conceptual illustration of SVR model

function to map the training data $xi, yi; i = 1, \ldots, n$to a high dimensional feature space. Then, the non-linear relationship between both learners can be described by a regression function as follows in the Eq. (2).

$$f(x) = W^T \varphi(x) + b \quad (2)$$

Where $w$and $b$are the coefficients to be adjusted, f(x) is a mapping function.

In fact, SVR is an optimizing problem in which objective function is given in Eq. (3).

$$Min_{w,b,\xi^*,\xi} R_\varepsilon(W, \xi^*, \xi) = 0.5 w^T w + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \quad (3)$$

Where C is the trade-off parameter between the first and second terms of the equation, ξ^*, ξ is a features of the space, R is a risk factor.

By solving the above-described optimization problem the coefficient of Eq. (2) can be found as follows in Eq. (4).

$$w = \sum_{i=1}^{n} (\beta_i - \beta_i^*) \varphi(x_i) \quad (4)$$

Where β_i is the Lagrangian coefficients. The following Eq. (5) describes the SVR regression function.

$$f(x) = \sum_{i=1}^{n} (\beta_i - \beta_i^*) k(x_i, x_j) + b \quad (5)$$

Where the Kernel function is denoted by $K(x_i, x_j)$, b is a constant.

Among the family of Kernel functions, the most commonly used ones are the Gaussian Radial Basis Functions (RBF) and the polynomial. There are no specific guidelines for determining the proper Kernel type for specific data patterns. The concept of SVR concept is visually illustrated in Fig. 2.

As mentioned, the SVR parameters affect the accuracy of the prediction. Hence, it is essential to select appropriate parameters. The parameter C takes care of the trade-off between the degree of the training error and the model flatness; large values of C result in only minimizing the empirical risk.

3.4.3. Logistic regression automata model

Logistic regression [23] allows one to predict a discrete outcome, such as whether it will rain today or not, from many types of variables that may be dichotomous, continuous, discrete, or a mix of any of these. Generally, the response or dependent variable is dichotomous, such as success/failure or presence/absence, i.e., the dependent variable can take the value 0 or 1 with a probability of failure or success. Then, this kind of variable is called a Binary (or Bernoulli) variable.

Consider a simple k variable regression model in Eq. (6).

$$E(Y \vee X) = \beta_0 + \beta_1 X_1 + \ldots. + \beta_p X_p \quad (6)$$

where $k = p + 1$. We would logically let
y_i=
0 if the i-th unit does not have the characteristic.
1, if the i-th unit does possess that characteristic.

Generally, where the response variable is binary, the shape of the response function is indicated by considerable empirical evidence that should be non-linear (in a variable). A monotonically increasing (or decreasing) S-shaped (or reversed S-shaped) curve could be a better choice. This kind of curve is obtained, if the regression chooses the specific form of the function as defined in Eq. (7).

$$\pi(X) = \frac{exp(Z)}{1+exp(Z)} \tag{7}$$

Where $Z = X\beta$. This function called as the logistic response function. Here Z is called the linear predictor defined by the Eq. (8).

$$Z = ln\left(\frac{\pi}{1-\pi}\right) \tag{8}$$

The model in terms of Y would be written as in the Eq. (9).

$$E(Y \vee X) = \pi(X) \tag{9}$$

It is a well-known problem that the binary response model violates a number of Ordinary Least Squares (OLS) assumptions. Hence it is a common practice to use the Maximum Likelihood (ML) method based on Iterative Re-weighted Least Squares (IRLS) algorithm.

4. Experimental outcome

In this section, the proposed method presents an evaluation of the proposed rainfall prediction algorithm discussed. First, the experimental settings are described in this section, the series of experiments are conducted for evaluating the effectiveness of rainfall prediction algorithms and then the results are presented and discussed.

4.1 Dataset description

The country's economy is based on agriculture and its agricultural products, and crop yield is heavily dependent on the summer monsoon (June-September) rainfall. Therefore, any decrease or increase in annual rainfall will always have a severe impact on the agricultural sector in India. Hence, the prior knowledge of monsoon behaviour will help the Government and farmers to take the advantage of the monsoon season. This knowledge can be very useful in minimizing the crop's damage during the less rainfall in the monsoon season. Forecasting is an important scientific issue in the field of monsoon meteorology. In this study, daily rainfall series were collected from four rainfall stations in the Australian Government Bureau of Meteorology http://www.bom.gov.au/climate/data (Belmont Agforce, Glenlands, Broadmeadows, Gracemere-Lucas stations in QLD state). The changeover dates vary from State to State and year to year. More information can be found at http://www.bom.gov.au/climate/averages/tables/daysavtm.shtml. Apart from some early historical observations, daily rainfall observations are made at 9 am local time and record the precipitation which has fallen in the previous 24 hours.

4.2 Sample output

The method is used for predicting the daily rainfall in parallel TSA. By evaluating various existing methods such as WA-SVM, SVR, this paper implemented the three regression automata model for predicting the daily basis rainfall. The below figure represent the sample output of the proposed method. The sample output describes the prediction of rainfall for daily basis data for January month.

The sample TSA graph for January 2015 month is illustrated in the above figure 3 which shows the nature of LRA, SVRA and LORA prediction rates and its sequence of lag in time. It is inferred that LRA is showing prominent results in terms of accuracy and reduced error rate. The vector nature of SVRA and LORA shows that it is not suitable for time synchronized regression automata models and its lag in the variance of rainfall values.

4.3 Performance Criteria

The performances of the models developed in this study were assessed using standard statistical performance evaluation criteria which included the error ratio metric, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Nash-Sutcliffe Efficiency coefficient (NSE).

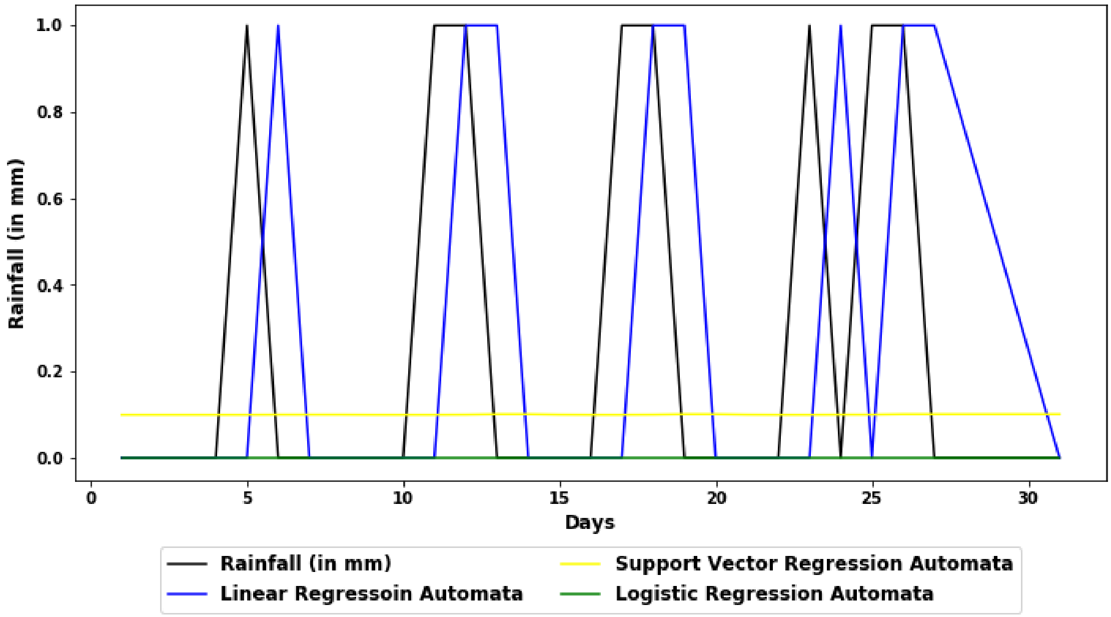

Figure.3 January 2015 Month Prediction

4.3.1. Root mean squared error

The predictive capabilities of the model can be provided by different types of information in RMSE. The RMSE measures the goodness-of-fit relevant to high rainfall values. RMSE is defined in Eq. (10),

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}\left(Q_i^p - Q_i^o\right)^2} \quad (10)$$

Where, n is the number of input samples, Q_i^p and Q_i^o are the observed and predicted rainfall at time t.

4.3.2. Mean absolute error

MAE given by Eq. (11) was used to measure the accuracy of forecasting. Smaller values of these parameters indicate higher model accuracy. The balanced perspective of the goodness-of-fit can be yielded by MAE at moderate value distribution of the estimation errors. The MAE is predicted in Eq. (11)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Q_i^p - Q_i^o\right| \quad (11)$$

4.3.3. Nash-Sutcliffe efficiency coefficient

The NSE was used to evaluate the goodness of fit between the observed and the forecasted values. In addition, NSE provides higher values of this coefficient indicate better model performance. The following Eq. (12) represents the NSE coefficient,

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_i^o - Q_i^p\right)^2}{\sum_{i=1}^{n}\left(Q_i^o - \grave{Q}^o\right)^2} \quad (12)$$

where, $\grave{Q}^o$ is the mean of the observed rainfall value.

4.3.4. Error ratio metric

By this way, the impact of predicted algorithm compared to the observed algorithm can be computed as follows in Eq. (13),

$$ErrorRatio = \frac{RMSD_{predictedalgorithm}}{RMSD_{observedalgorithm}} \quad (13)$$

4.4 Experimental analyses

In this section, the performance of the regression models is evaluated by the parameters like RMSE, MAE, NSE and error ratio for the four datasets.

4.4.1. Evaluation of RMSE

The values of the RMSE are obtained from the experimental results for four databases namely ID39043, ID39049, ID39242, ID33229. The results are shown by a graphical representation in figure 4 and the values are presented in table 1. The LRA model has the highest RMSE values (0.452) for ID 39242 station in QLD state. Though all the three models performed well in database, the SVRA model leads better RMSE performance when compared with the other two models.

Table 1. RMSE evaluation for four databases

RMSE	ID39043	ID39049	ID39242	ID33229
LRA	0.392	0.401	0.452	0.410
SVRA	0.336	0.356	0.415	0.356
LORA	0.357	0.379	0.448	0.380

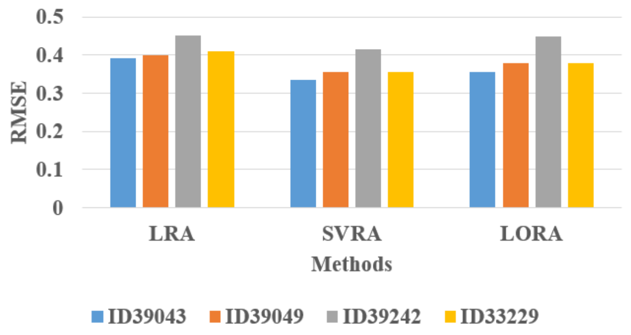

Figure.4 Evaluation of RMSE

Table 2. MAE evaluation for four databases

MAE	ID39043	ID39049	ID39242	ID33229
LRA	0.154	0.161	0.205	0.166
SVRA	0.202	0.216	0.262	0.216
LORA	0.126	0.144	0.201	0.144

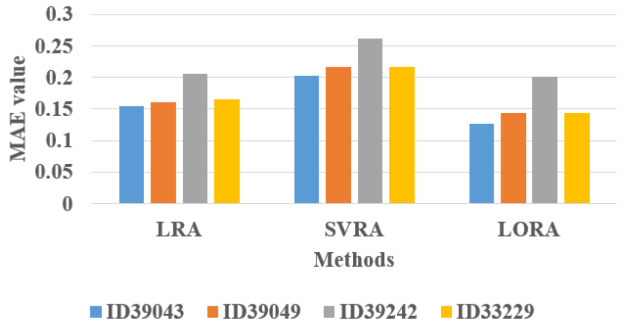

Figure.5 MAE evaluation for datasets

4.4.2. Evaluation of MAE

The rainfall is predicted by the values obtained for MAE parameters which are tabulated in Table 2. The graphical representation for MAE can be obtained in figure 5. Even though the SVRA model yields better performance in RMSE values, the SVRA model provides the poor MAE values in all the four stations. The SVRA model achieved 0.262 MAE values in ID39242 station.

4.4.3. Evaluation of NSE:

The values of the NSE are obtained by the experimental results for four databases namely ID39043, ID39049, ID39242, ID33229. The results are shown by a graphical representation in Fig. 6 and Table 3 contains the NSE values. Here, compared to the other database, the ID39242 performed well in all the three regression methods. The LRA provides 1.481 NSE values, 1.093 values are provided by SVRA, whereas LORA provides 1.437 NSE values.

Table 3. NSE evaluation for four datasets

NSE	ID39043	ID39049	ID39242	ID33229
LRA	0.482	0.409	1.481	0.451
SVRA	0.079	0.108	1.093	0.093
LORA	0.214	0.256	1.437	0.246

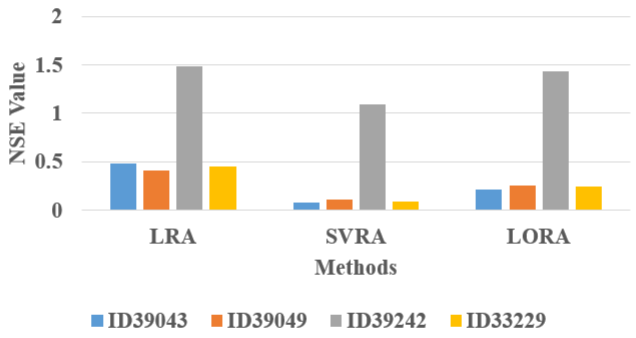

Figure.6 Performance Evaluation for NSE

Table 4. Error Ratio evaluation for four datasets

Error Ratio	ID39043	ID39049	ID39242	ID33229
LRA	2.64×10^{-5}	-3.487×10^{-6}	2.64×10^{-5}	-5.172×10^{-5}
SVRA	0.029	0.047	-0.053	0.046
LORA	0.104	0.114	0.082	0.116

4.4.4. Evaluation of error ratio

The error ratio of the proposed algorithm is calculated for predicting the performance of the three regression models. The values are given in Table 4 shows the best regression model for predicting the daily rainfall for four stations in QLD state. The rainfall can be predicted by the several models, but the best appropriate models are selected by considering their error ratio. Overall, the LRA method performed well to predict the rainfall correctly.

4.5 Comparative analysis

Forecast of daily monsoon rainfall for the year 1954-2015 was done using the developed models.

Table 5. Comparison of proposed method with existing methods

Authors	Methodology Used	RMSE	MAE	NSE
Q. Feng, et al., [24]	WA-SVM	12.689	7.828	0.892
Q. Ouyang and W. Lu, [25]	SVR	5.917	4.593	0.984
	MGGP	7.024	5.480	0.977
	ESN	3.341	1.890	0.992
Geetha, A. and Nasira, G.M., [26]	ARIMA Model	0.464	0.217	-
Proposed Method	LRA	0.410	0.168	0.451
	SVRA	0.357	0.216	0.093
	LORA	0.380	0.144	0.246

The forecasted values of rainfall obtained in this study can be used in agricultural and water resource planning, hydrological model study and climate change study. Table 5 shows the performance criteria of different forecasting methods for daily rainfall forecasting. Note that the results were obtained using different RA techniques coupled with LOR. The proposed methodologies are compared with the existing methods such as Wavelet Analysis-Support Vector Machine coupled model (WA-SVM) [24] in which RMSE values were higher because of using vector method, Echo State Networks (ESN), SVR, and Multi-Gene Genetic Programming (MGGP) [25] were evaluated for 1, 3 and 6 months ahead rainfall forecasting. But, the SVRA method achieved 0.357% RMSE by using regression method.

The existing method (i.e. Q. Ouyang and W. Lu, [25]) provided poor performance in MAE and NSE because the method crossed the lead time. In the proposed method, LRA presents the best MAE and NSE values because linear constant values are used. As an overall result, it is inferred that the proposed regression automata models with LRA and LORA can perform accurate prediction and reduced error rate with parallel processing using a sliding window technique for effective TSA on rainfall forecasting.

5. Conclusion

Long historical daily rainfall depth TS (Jan. 1954–Dec. 2015) at four selected stations (ID39043, ID39049, ID39242, ID33229) have been analyzed. The total daily rainfall depths were calculated and the climatological normal of the total daily rainfall was obtained at all four stations.

The maximum rainfall occurs during March and April. For TS analysis purposes, it was decided to use the series of the Total Quarterly Rainfall (TQR) depth. Analyses of these TS clearly illustrate very strong temporal and spatial variability. The experimental results prove that the LRA method obtained 0.410% of RMSE, 0.357% RMSE of SVR and the last LORA method acquired 0.380% of RMSE when compared with the ARIMA method which obtained 0.464% RMSE. More thorough statistical comparison of the outcome of the three RA models indicates that the third model (Logistic RA model) is the optimal model for forecasting rainfall. In future, the RA model can improve the parallel processing methodology to process many years simultaneously with a synchronized framework with an effective regression model.

References

[1] G. Mengaldo, A. Wyszogrodzki, M. Diamantakis, S. J. Lock, F. X. Giraldo, and N. P. Wedi, "Current and Emerging Time-Integration Strategies in Global Numerical Weather and Climate Prediction", *Archives of Computational Methods in Engineering*, pp.1-22, 2018.

[2] L. Guodong, X. Wenxia, Y. Bing, B. Awudong, and C. Xiaojuan, "A method analysis for hail cloudy prediction based on CNN", *Cluster Computing*, Vol.19, No.4, pp.2015-2026, 2016.

[3] W. Chih-Chiang, "Conceptual weather environmental forecasting system for identifying potential failure of under-construction structures during typhoons", *Journal of Wind Engineering and Industrial Aerodynamics*, Vol.168, pp.48-59, 2017.

[4] C. Lesk, P. Rowhani, and N. Ramankutty, "Influence of extreme weather disasters on global crop production", *Nature*, Vol.529, No.7584, pp.84, 2016.

[5] Y. Chen, Z. Wu, Z. Li, and Y. Zhang, "Research on time series forecasting model based on moore automata", In: *Proc. of International Conf. On Advanced Data Mining and Applications*, pp.98-105, Springer, Berlin, Heidelberg, 2010.

[6] S. Mehrmolaei, and M.R. Keyvanpour, "Time series forecasting using improved ARIMA", In: *Proc. of International Conf. On Artificial Intelligence and Robotics*, pp.92-97 2016.

[7] R.M. Nabilah, Z. Othman, and B. A. Azuraliza, "Approaches of Handling Uncertain Time Series Data towards Prediction", *International Journal of Future Computer and Communication*, Vol.5, No.6, pp.233, 2016.

[8] N.F.M. Radzuan, Z. Othman, and A.A. Bakar, “Uncertain time series in weather prediction”, *Procedia Technology*, Vol.11, pp.557-564, 2013.

[9] T.K. Chang, A. Talei, S. Alaghmand, and M.P. L. Ooi, “Choice of rainfall inputs for event-based rainfall-runoff modeling in a catchment with multiple rainfall stations using data-driven techniques”, *Journal of Hydrology*, Vol.545, pp.100-108, 2017.

[10] D. Lopez, M. Gunasekaran, B.S. Murugan, H. Kaur, and K.M. Abbas, “Spatial big data analytics of influenza epidemic in Vellore, India”, In: *Proc. of International Conf. On Big Data*, pp.19-24, 2014.

[11] G. Lee, H. An, and M. Kim, “Comparing the performance of TRIGRS and TiVaSS in spatial and temporal prediction of rainfall-induced shallow landslides”, *Environmental Earth Sciences*, Vol.76, No.8, pp.315, 2017.

[12] C. Iadanza, A. Trigila, and F. Napolitano, “Identification and characterization of rainfall events responsible for triggering of debris flows and shallow landslides”, *Journal of Hydrology*, Vol.541, pp.230-245, 2016.

[13] R. Giannecchini, Y. Galanti, G. D. A. Avanzi, and M. Barsanti, “Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape”, *Geomorphology*, Vol.257, pp.94-107, 2016.

[14] M.A. Ting, Z.H.O.U. Cheng-Hu, and C.A.I. Qiang-Guo, “Modeling of hillslope runoff and soil erosion at rainfall events using cellular automata approach”, *Pedosphere*, Vol.19, No.6, pp.711-718, 2009.

[15] J.C. Bennett, D.E. Robertson, P.G. Ward, H.P. Hapuarachchi, and Q.J. Wang, “Calibrating hourly rainfall-runoff models with daily forcings for streamflow forecasting applications in meso-scale catchments”, *Environmental Modelling & Software*, Vol.76, pp.20-36, 2016.

[16] A. Douinot, H. Roux, and D. Dartus, “Modelling errors calculation adapted to rainfall–Runoff model user expectations and discharge data uncertainties”, *Environmental Modelling & Software*, Vol.90, pp.157-166, 2017.

[17] M. Espínola, J. A. Piedra-Fernández, R. Ayala, L. Iribarne, S. Leguizamón, and J. Z. Wang, “Simulating rainfall, water evaporation and groundwater flow in three-dimensional satellite images with cellular automata,” *Simulation Modelling Practice and Theory*, Vol.67, pp.89-99, 2016.

[18] M. Alvioli, and R.L. Baum, “Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface”, *Environmental Modelling & Software*, Vol.81, pp.122-135, 2016.

[19] M.R. Bendre, R.R. Manthalkar, and V.R. Thool, “Modeling and predicting weather in agro-climatic scarcity zone using iterative approach”, *Decision*, Vol.44, No.1, pp.51-67, 2017.

[20] M. Herrera, A.P. Ramallo-González, M. Eames, A.A. Ferreira, and D.A. Coley, “Creating extreme weather time series through a quantile regression ensemble”, *Environmental Modelling & Software*, In Press, 2018.

[21] G.J. Martínez, J.C. Seck-Tuoh-Mora, and H. Zenil, “Wolfram’s classification and computation in cellular automata Classes III and IV”, In: *Proc. of International Conf. on Irreducibility and Computational Equivalence*, Berlin, Heidelberg, pp.237-259, 2013.

[22] A. Kavousi-Fard, H. Samet, and F. Marzbani, “A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting”, *Expert systems with applications*, Vol.41, No.13, pp.6047-6056, 2014.

[23] A.R. Imon, M.C. Roy, and S.K. Bhattacharjee, “Prediction of rainfall using logistic regression”, *Pakistan Journal of Statistics and Operation Research*, Vol.8, No.3, pp.655-667, 2012.

[24] Q. Feng, X. Wen, and J. Li, “Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions”, *Water resources management*, Vol.29, No.4, pp.1049-1065, 2015.

[25] Q. Ouyang, and W. Lu, “Monthly Rainfall Forecasting Using Echo State Networks Coupled with Data Preprocessing Methods”, *Water Resources Management*, Vol.32, No.2, pp.659-674, 2018.

[26] A. Geetha and G. M. Nasira, “Time-series modelling and forecasting: modelling of rainfall prediction using ARIMA model”, *International Journal of Society Systems Science*, Vol. 8, No. 4, pp.361-372, 2016.