Seasonal Prediction of Rainfall over Cherthala region of Kerala through stepwise Singular Value Decomposition

Prameela S. Bhanu[a] and Archana Nair[b]

[a] Department of Mathematics, Amrita School of Arts and Sciences, KochiAmrita Vishwa Vidyapeetham

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract: In this study, an attempt has been made to predict summer monsoon rainfall over Cherthala region of Alappuzha district of Kerala. For this purpose, rainfall datasets from 1991 to 2014 have been utilized from the Kerala government. The rainfall dataset act as the dependent variable while the sea surface temperature (SST), ocean heat content (OHC) and wind at different locations act as independent variables. A mathematical model based on singular value decomposition multiple linear regression approach has been employed here. Sixty-three models are constructed by taking combinations of different parameters at a time. The accuracy of the forecast has been tested through root mean square error (RMSE). Results suggest that the best model is the prediction given by taking the ocean heat content (OHC) in the 0-50N, 500E-750E with an RMSE of 1.03mm/day. The study implies that OHC can be utilized to predict rainfall over Cherthala region

Keywords: Rainfall, prediction, principal compnent regression, idukki, kerala

1. Introduction

India is an agrarian country because large number of its population depends on agriculture and its related activities. As per 2018, agriculture employed more than 50% of the Indian work force and contributed 17–18% to country's GDP (Dhanabalan, T & Sathish, (2018). The prime source of water for agriculture in India is rainfall which is caused by the south-west monsoon rainfall. It is the seasonal reversal of wind that causes monsoon and brings copious amount of rainfall to India. The monsoon wind first strikes over Kerala, which is the gateway of Indian summer monsoon rainfall. From past few years, Kerala is facing natural calamities in the form of floods and cyclones due to factors which are still unknown to the scientific community.

Many studies are undertaken to predict the amount of rainfall over India using many dynamical and statistical approaches. For example, Multilinear Regression (MLR) is the widely used mathematical technique used for prediction (Navid & Niloy 2018). Artificial Neural Network model was induced in certain predictions (Chattopadhyay S,2007, Cannon, & Kendry, 2002). Data Mining (Kannan et al. 2010, Dutta & Tahbilder, 2014, Sethi et al. 2014) Machine Learning (Mohammed et al. 2020) and Multivariable Polynomial Regression (MPR) (Zaw & Naing 2008) are some other techniques are capable of predicting rainfall. Mostly, the studies are limited in seasonal prediction rather than monthly prediction over India.

Kerala the south most state of India is diverse in climate and geographic nature. The place is known as 'God's own country'. The climate is very favourable to live in, as the place maintains a moderate temperature. Until 2018 everything was expectable. During 2018, the state faced a great flood which caused the death of nearly 500 number of people, many became homeless. The agriculture loss was uncountable. This brings up the relevance of this study. This study could help in predicting rainfall in the following years. Hence, the people as well as the authority can be prepared for the sudden change in climate particularly the rainfall. Although, there are plenty of studies that are carried out in predicting rainfall all over India but no such studies have been done over Kerala especially over Cherthala which is situated in the Alappuzha district of Kerala. Some studies that are based on climate change, monsoon rainfall and trends in seasonal rainfall have taken place. The studies on climate change and wave climate were taken place to know the effects on rice production and to protect the coast from erosion (Swamy, et al. 1979, Anapalli et al. 2000). There were few studies based on the precipitation trends. (Nair et al., 2014, Pal & Al-Tabbaa, 2009). The state is rich in varied land and climate.

Cherthala is a block in Alappuzha district. Cherthala, a coastal town, located at 9°41′13″N, 76°20′10″E, 31 km to south of Kochi, 31km to west of Kottayam and 22 km north of Alappuzha town (see Figure 1a). It is in the 2m elevation (altitude) and covers an area of 16.18 km^2 (6.25 sq mi). The relevance of prediction of rainfall in Cherthala is inevitable and is of prime importance in terms of agriculture and tourism. Cherthala is a place with very good climate and is humid in nature. The south-west monsoon from June to September (known as Edavappathi), affects the climate of the region. On account of its adjacency with ocean, a large part of its climate variability may be controlled by oceans. Monthly and seasonal rainfall prediction of monsoon can help in the field of agricultural and tourism. Here, in this paper we make an effort to fit a mathematical model that would predict the seasonal summer monsoon rainfall over Cherthala based on Stepwise Singular Value Decomposition (SVD), the details of which are presented in section 2 along with the datasets used. The Results of the study are penned down in Section 3 and the conclusions are described in Section 4.

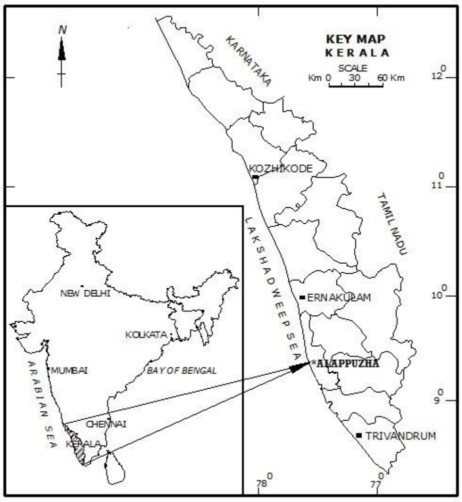

Figure 1: Map showing Kerala and the Alappuzha district.

2. Methodology

The data of monsoon rainfall from 1991-2014 is collected and is studied and plotted to understand the rainfall in Cherthala. The seasonal, monthly and daily rainfall has been observed for the months of June, July, August, and September.

In order to develop a model, parameters that effect rainfall over Kerala has been found. Since, Kerala is the gateway of monsoon and the monsoon dynamics play an important role in contributing rainfall over Kerala. It is also known that the Arabian Sea plays major role in modulating the monsoon, hence, parameters that effect the monsoon dynamics has been taken into consideration for developing the model. The parameters that are taken into consideration are

2.1.Sea Surface Temperature (SST)

Sea surface temperature is the water temperature closest to the ocean surface, i.e., between 1 millimetre (0.04 in) and 20 metres (70 ft) below the sea surface. SST was one of the first oceanographic variables to be measured. However, SST affect climate pattern to some extent (Huang et al. 2014, 2015 Liu et al. 2014). The two regions that are considered here are 0-5^{0}N, 50^{0}E to 75^{0}E (SST1) and 5^{0}N-10^{0}N, 50^{0}E to 75^{0}E (SST2).

2.2.Ocean Heat Content (OHC)

Ocean heat content is the energy absorbed by ocean. The main source of energy is obtained from sunlight. Ocean is the largest absorber of solar energy on earth (Levitus et al. 2017). It has the ability to collect a huge amount of heat without increasing the temperature. Thus, oceans play a vital role in maintaining the climate of earth. Moreover, some heat from clouds, water vapour and greenhouse gases are also been absorbed by oceans. This heat absorbed is never destroyed and is been transferred to cooler latitudes. 30% of the ocean heat is moved from the upper ocean layer to deeper ocean and this impacts marine eco-system. One of the influences of ocean heat content is the rainfall. (Venugopal, et al. 2018) Researches have proven the increase in ocean heat current(Cheng, et al. 2021). The two regions that are considered here are 0-5^{0}N, 50^{0}E to 75^{0}E (OHC1) and 5^{0}N-10^{0}N, 50^{0}E to 75^{0}E (OHC2).

2.3.Wind

The two regions that are considered here 0-5^{0}N, 50^{0}E to 75^{0}E (WND1) and 5^{0}N-10^{0}N, 50^{0}E to 75^{0}E (WND2).

2.4.Correlation

Correlation is a statistic that measures the degree to which two variables move in relation to each other. The correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0. A correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. A correlation of 0.0 shows no linear relationship between the movements of the two variables.

2.5.Multicollinearity

Collinearity is a linear association between two independent variables. Two variables are perfectly collinear if there is an exact linear relationship between them. For example, X_1 and X_2 are perfectly collinear if there exist parameters λ_0 and λ_1 such that, for all observations i, we have

$$X_{2i}=\lambda_0+\lambda_1 X_{1i}.$$

Multicollinearity refers to a situation in which more than two independent variables in a multiple regression model are highly linearly related. We have perfect multicollinearity if, for example as in the equation above, the correlation between two independent variables is equal to 1 or −1. In practice, we rarely face perfect multicollinearity in a data set. More commonly, the issue of multicollinearity arises when there is an approximate linear relationship among two or more independent variables. Mathematically, a set of variables is perfectly multicollinear if there exist one or more exact linear relationships among some of the variables. For example, we may have

$$\lambda_0+\lambda_1 X_{1i}+\lambda_2 X_{2i}+...+\lambda_k X_{ki}=0$$

holding for all observations i, where λ_k are constants and X_{ki} is the i^{th} observation on the k^{th} explanatory variable.

We can explore one issue caused by multicollinearity by examining the process of attempting to obtain estimates for the parameters of the multiple regression equation

$$Y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+...+\beta_k X_{ki}+\varepsilon_i.$$

The ordinary least squares estimates involve inverting the matrix X^TX where

$$X=\begin{pmatrix} 1 & X_{11} & ... & X_{k1} \\ . & . & & . \\ . & . & ... & . \\ . & . & & . \\ 1 & X_{1N} & ..... & X_{kN} \end{pmatrix}$$

is an $N \times (k+1)$ matrix, where N is the number of observations and k is the number of independent variables (with N required to be greater than or equal to $k+1$). If there is an exact linear relationship (perfect multicollinearity) among the independent variables, at least one of the columns of X is a linear combination of the others, and so the rank of X (and therefore of X^TX) is less than $k+1$, and the matrix X^TX will not be invertible. If there is no exact linear relationship among the variables, the matrix X^TX has an inverse.

2.6.Regression

Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. The most common form of regression analysis is linear regression, in which one finds the line (or a more complex linear combination) that most closely fits the data according to a specific mathematical criterion. Regression analysis is primarily used for two conceptually distinct purposes. First, regression analysis is widely used for prediction and forecasting. Second, in some situations regression analysis can be used to infer casual relationships between the independent and dependent variables. Importantly, regressions by themselves only reveal relationships between a dependent variable and a collection of independent variables in a fixed dataset.

$$Y_i=f(X_i,\beta)+\varepsilon_i.$$

2.7.Multiple linear Regression

Multiple linear regression (MLR), also is simply known as multiple regression, is an extension of linear regression. The purpose of MLR is to model the linear relationship between independent variables and dependent variable.

$$Y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+...+\beta_k X_{ki}+\varepsilon_i.$$

where,

Y_i =dependent variable

X_{ki} = independent variables

β_0=y-intercept (constant term)

β_k =slope coefficients for each explanatory variable

ε_i=the model's error term (also known as the residuals)

2.8.Singular Value Decomposition

Singular Value Decomposition (SVD) is a method in linear algebra. It has numerous applications in statistics, computer science and machine learning. The method helps in decomposing a matrix to a reduced form for easier calculations.

$X=U*\Sigma*V$

where U is a Unitary Matrix, Σ is a Diagonal Matrix and V is a Conjugate Transpose of Unitary Matrix.

Algorithm: Rainfall Prediction in Cherthala

Input: Rainfall data set from 1991 to 2014 at different stations

Output: Best model for predicting the rainfall

Step 1: Collecting parameters and choosing them through correlation

Step 2: Import the data to excel file

Step 3: Fill the missing data

Step 4: Sum data and Standard Deviation is analysed using matlab

Step 5: Reduction of data is done using SVD

Step 6: Using MLR prediction for each year is calculated

Step 7: Error is calculated using matlab

Step 8: Best Model is selected

3. Results and Discussions

3.1 Inter-annual variation of seasonal and monthly rainfall

The inter-annual variation of seasonal (Figure 2a) and monthly (Figure 3) rainfall over Cherthala region of Kerala from 1991 to 2014 has been studied in this section. The trend is also depicted as dark black dotted line with the timeseries plot of rainfall. The figure suggests fluctuating behaviour rainfall with many anomalies in the seasonal scale. The mean seasonal rainfall is found to be 1758.4mm and a standard deviation is 250mm, indicating a variability of 14%. The highest rainfall happened in the year 2007 with 2270mm of rainfall and the least in 2012 with 1291 mm of rainfall. The standardised rainfall suggests five years (1999, 2002, 2003,2004,2012) of below normal rainfall whereas 3 instances of above normal rainfall (2006, 2007,2013) are found. The other years are recorded as normal. It can be observed that for a decade 1991-2002, the rainfall was normal, except in 1992. The trend analysis suggests an increase of 9.77mm/year of rainfall in the district.

The rainfall distribution on a monthly scale is shown in Figure 3 for June, July, August and September. The average rainfall in the month of June is observed to be 618mm and the standard deviation is 165mm. The variability in the month of June is found to be 26%. The maximum rainfall in the year is recorded in the year 2013 (1061 mm) whereas the least is observed in 2012 (288.7mm). It is noted that, out of 24 years, five years show below normal rainfall (1992, 1996,2008,2012 and 2014) whereas two instances of above average rainfall is found (1991 and 2013). Also, from 1999 to 2007, the rainfall in the district follows a same structure. The trend analysis of rainfall suggests an insignificant but small increase of 1mm/year of rainfall.

In the month of July (figure 2b), an average rainfall of 503 mm occurs on a monthly scale. The standard deviation is found to be 145 mm which is much less than the June rainfall. The rainfall in this month ranges from 200mm to 800mm. The highest rainfall is observed in the year 2007 (804mm) and the least in 2012 (237.1 mm). The variability in the month of July is 2% more than the July month. It can be observed that 1998, 2000, 2002, 2004 and 2012 are below normal years while 1993,1996,2007 and 2013 above normal rainfall had occurred. The trend depicts a decrease of 0.46mm/year of rainfall.

In the month of August, the mean monthly rainfall is recorded as 366mm with standard deviation of 149mm. The variability in this region is found to be increased to 40% from June and July rainfall. The rainfall range is noted to be between 100 to 800mm with the maximum in the year 201 and least in 1993. There are three instances of below normal (1993,2005, 2008) and two instances (2000,2014) of above normal. In all the other years, the rainfall is said to be normal. The trend analysis suggests an increase of 1.58mm/year of rainfall. The maximum variability in monthly rainfall is observed in the month of September where 48% of variance is found. The standard deviation is 130 mm and the mean is 270mm.

The rainfall in September month is much less than the other months, yet the variability is more in the September rainfall. The maximum rainfall is 531 mm which is observed in the year 1998 and minimum is 55.8mm (1991). Four instances of below normal (1991,1993,2002,2003) and 3 instances (1998,2007 and 2008) of above normal rainfall years are found. The analysis of trend suggests a significant increasing pattern of rainfall. The increase is found to be 7.64 mm/year.

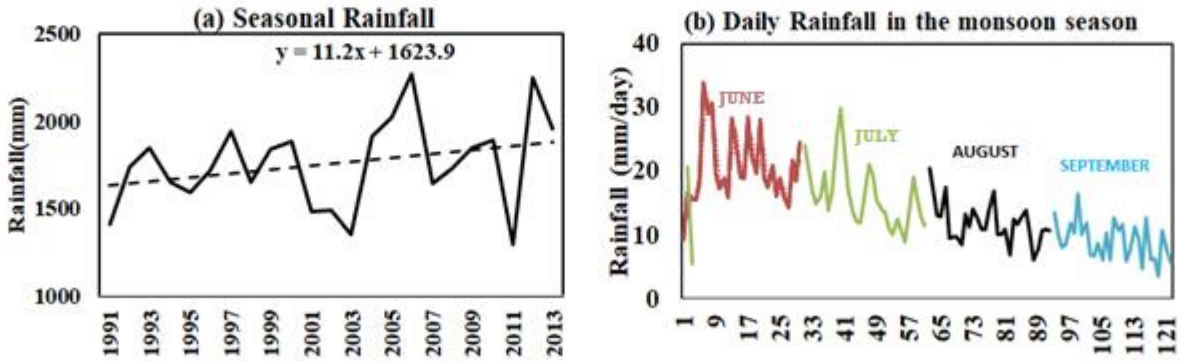

Figure 2 The rainfall over Cherthala region on (a) Seasonal (b) Daily scale

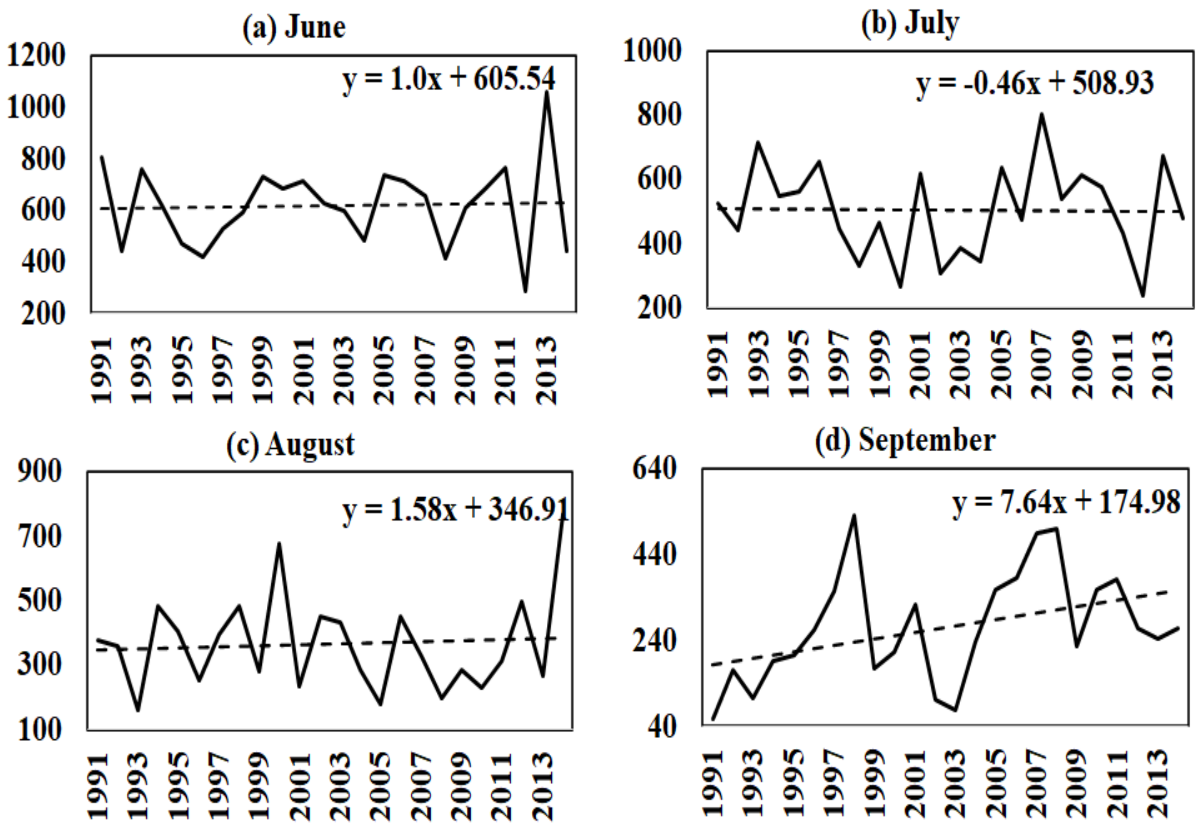

Figure 3 Monthly rainfall distribution over Cherthala region of Alappuzha, Kerala.

3.2 Daily Rainfall Scenario

The daily average rainfall scenario over Cherthala region is shown in Figure 1b for the 122 days of summer monsoon rainfall. The red color line shows the variation in the month of June, the green for July, black for August and blue for September. The average rainfall in the month of June is 20.6 mm/day with a standard deviation of 5mm/day. A maximum of 33 mm/day of rainfall has occurred in the month of June. A small increase of 0.006mm/day has been observed for June. In the month of July, the average rainfall is 16mm/day with a standard deviation of 4.95mm/day. The maximum rainfall that has occurred in the month of July is 19.84 mm/day and the minimum is 8.98mm/day. A decrease of 0.29 mm/day has been noted in the month of July. Similar to July, August and September month also shows a decrease of 0.16 mm/day and 0.14 mm/day. The mean rainfall in the month of August and September is 11mm/day and 9 mm/day. The standard deviation for August and September are 3.1mm/day and 2.9 mm/day. The maximum one-day rainfall realized in a single day is 20mm/day and 16mm/day for August and September respectively.

In view of above, it can be inferred that the daily rainfall in almost all the months are decreasing.

3.3 Mathematical Model

To find the best model to predict rainfall over Cherthala, we have evaluated 63 models, by taking different parameters into consideration.

(a) One parameter fitting

Six type of model is fitted by taking one parameter at a time as described in methodology. Figure 4a shows the RMSE by fitting the model. It can be seen that the RMSE ranges from 1.02mm/year to 1.08mm/year. The highest error is found for the parameter SST2 and the least is found in the OHC1.

(b) Two parameter model

Fifteen combinations were made by selecting two parameters. At a time. The RMSE ranges from 1.04 to 1.15 mm/day. The least RMSE is found in the combination of OHC1 and OHC2 (Figure 4b).

(c) Three parameter model

Twenty combinations were made by selecting three parameters at a time. The RMSE ranges from 1 to 1.25 mm/year where the least is 1.09 mm/year and highest is found as 1.22 mm/year. The least RMSE is noticed in the combination of SST1, OHC1 and OHC2 (Figure 4c).

(d) Four parameter model

By combination of four parameters, fifteen models were created. The least RMSE is recorded as 1.14 mm/year which is resulted as the combination of SST1, OHC1, OHC2 and WND1. The range of error in this case is from 1.1 mm/day to 1.27 mm/day (Figure 4d).

(e) Five parameter model

Six models were constructed by taking five combinations of parameters. The range of values is from 1.19 to 1.28mm/year. The least is obtained in the combination of SST1, SST2, OHC1, OHC2 and WND1.

(f) Six parameters model

By taking six combinations at a time, the RMSE is found to be 1.26 mm/day.

From the above results, it is found that, from all the 63 models, the best model is considered as the model with only OHC1 as the parameter i.e. the ocean heat content in the region 0-5, 50 -75E. Nevertheless, in all the models that has been constructed, the relation of Cherthala rainfall and ocean heat content can be found.

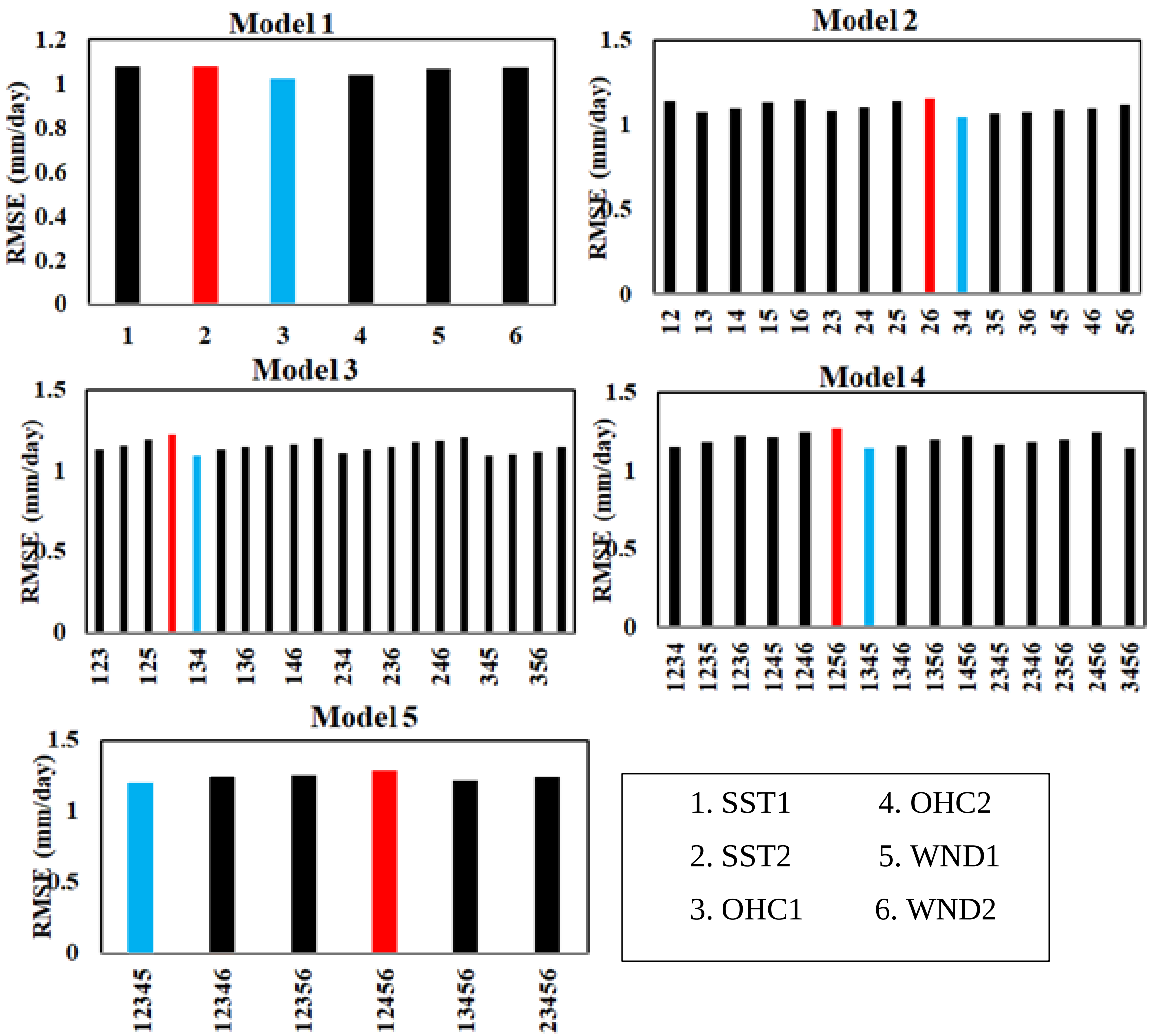

Figure 4: The RMSE (mm/day) of the developed model with different parameters. The blue bars show the least RMSE and the red the highest

4. Conclusions

It is seen that; from past few years many abnormalities are seen in the occurrence of rainfall. Hence, it is needed to predict the rainfall with certain accuracy. In this study, a mathematical model is built to predict the seasonal rainfall over Cherthala region of Alappuzha district of Kerala. Different parameters such as sea surface temperature (SST), ocean heat content (OHC) and wind at different regions are utilized for the development of model. The model is based on singular value decomposition and stepwise multiple linear regression is employed to test the accuracy of parameters and model.

The main results from the study are as follows:

1) The analysis of seasonal rainfall suggests the frequency of drought years are more than wet years. For every 10 years, three years are drought and two are wet.

2) Out of four months, September month shows lot of variability and a significant increasing trend of 7.64 mm/year is recorded.

3) Though the daily rainfall trend is decreasing but the seasonal rainfall is increasing.

4) Out of six variables, ocean heat content shows the best accuracy while predicting Cherthala rainfall. An RMSE of 1.03mm/day is recorded during the development.

So, in view of above, it can be inferred that the mathematical model is able to predict the rainfall over Cherthala with sufficient accuracy.

References

1. Cannon, A.J. & McKendry, I.G., 2002: A graphical sensitivity analysis for statistical climate models: application to Indian monsoon rainfall prediction by artificial neural networks and multiple linear regression models. International Journal of Climatology, 22(13), pp.1687-1708, http://dx.doi.org/10.1002/joc.811
2. Chattopadhyay S, 2007: Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India, Acta Geophys. 55, 369–382. https://doi.org/10.2478/s11600-007-0020-8
3. Cheng, L. J., John Abraham, Kevin E. Trenberth, John Fasullo, Tim Boyer, Ricardo Locarnini, Bin Zhang, Fujiang Yu, Liying Wan, Xingrong Chen, Xiangzhou Song, Yulong Liu, Michael E. Mann, Franco Reseghetti, Simona Simoncelli, Viktor Gouretski, Gengxin Chen, Alexey Mishonov, Jim Reagan, And Jiang Zhu, 2021: Upper ocean temperatures hit record high in 2020. Adv. Atmos. Sci., 38(4), 523−530, https://doi.org/10.1007/s00376-021-0447-x
4. Dhanabalan, T., & Sathish, A. (2018). Transforming Indian industries through artificial intelligence and robotics in industry 4.0. International Journal of Mechanical Engineering and Technology, 9(10), 835-845.
5. G.Narayana Swamy, P.Udaya Varma, Abraham Pylee, V.S.Rama Raju and P. Chandramohan, 1979: Wave Climate Off Trivandrum (Kerala), Mahasagar-Bulletin of the National Institute of Oceanography,12(3), 127-133
6. Huang, B., P. Thorne, T. Smith, W. Liu, J. Lawrimore, V. Banzon, H. Zhang, T. Peterson, and M. Menne, 2015: Further Exploring and Quantifying Uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) Version 4 (v4). Journal of Climate, 29, 3119–3142, doi:10.1175/JCLI-D-15-0430.1 (link is external).
7. Huang, B., V.F. Banzon, E. Freeman, J. Lawrimore, W. Liu, T.C. Peterson, T.M. Smith, P.W. Thorne, S.D. Woodruff, and H.-M. Zhang, 2014: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part I. Upgrades and intercomparisons. Journal of Climate, 28, 911–930, doi:10.1175/JCLI-D-14-00006.1
8. Pal, A. Al-Tabbaa, 2009: Trends in seasonal precipitation extremes – An indicator of 'climate change' in Kerala, India , Journal of Hydrology 367(1), 62–69
9. Kannan, M., Prabhakaran, S., & Ramachandran, P., 2010: Rainfall forecasting using data mining technique. International Journal of Engineering and Technology Vol.2 (6), 397-401
10. Levitus, S.; Antonov, J.; Boyer, T.; Baranova, O.; Garcia, H.; Locarnini, R.; Mishonov, A.; Reagan, J.; Seidov, D.; Yarosh, E.; Zweng, M. 2017: NCEI ocean heat content, temperature anomalies, salinity anomalies, thermosteric sea level anomalies, halosteric sea level anomalies, and total steric sea level anomalies from 1955 to present calculated from in situ oceanographic subsurface profile data NOAA National Centers for Environmental Information. doi:10.7289/V53F4MVP
11. Liu, W., B. Huang, P.W. Thorne, V.F. Banzon, H.-M. Zhang, E. Freeman, J. Lawrimore, T.C. Peterson, T.M. Smith, and S.D. Woodruff, 2014: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations. Journal of Climate, 28, 931–951, doi:10.1175/JCLI-D-14-00007.1
12. MAI Navid, NH Niloy, 2018: Multiple Linear Regressions for Predicting Rainfall for Bangladesh, Communications. Vol. 6, No. 1, pp. 1-4. doi: 10.11648/j.com.20180601.11
13. Moulana Mohammed and Roshitha Kolapalli, N.V.V. Golla, Siva Sai Maturi, 2020: Prediction Of Rainfall Using Machine Learning Techniques. International Journal Of Scientific & Technology, Vol. 9 (01), 3236-3240
14. Nair, Archana, K. Ajith Joseph, and K. S. Nair., 2014: Spatio-temporal analysis of rainfall trends over a maritime state (Kerala) of India during the last 100 years, Atmospheric Environment, vol. 88, pp. 123–132. doi:10.1016/j.atmosenv.2014.01.061
15. PS Dutta, H Tahbilder, 2014: Prediction of rainfall using Datamining technique over Assam, Indian Journal of Computer Science and Engineering (IJCSE) 5 (2), 85-90
16. SA Saseendran, KK Singh, LS Rathore, SV Singh, and SK Sinha, 2020: Effects of Climate Change on Rice Production in the Tropical Humid Climate of Kerala, India: Climatic Change [Clim. Change], vol. 44, no. 4, pp. 495-514

17. Sethi, Nikhil and Dr. Kanwal Garg. 2014: Exploiting Data Mining Technique for Rainfall Prediction, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 3982-3984
18. Venugopal, T., Ali, M.M., Bourassa, M.A., Zheng, Y., Goni. G.J., Foltz, G.R., Rajeevan, M., 2018: Statistical Evidence for the Role of Southwestern Indian Ocean Heat Content in the Indian Summer Monsoon Rainfall. Sci Rep 8, 12092. https://doi.org/10.1038/s41598-018-30552-0
19. Zaw, W.T., & Naing, T. 2008: Empirical Statistical Modeling of Rainfall Prediction over Myanmar. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2, 3418-3421