Modelling Rainfall Prediction using Bayesian Approach with Inter-station Dependencies

(IJCSNT) International Journal of Communication Systems and Network Technologies
ISSN Number: 2053-6283
Volume 13, Issue No. 2, 60–68

https://journals.mirlabs.in/index.php/ijcsnt
DOI-10.18486/ijcsnt/13.2.176
IJCSNT

Ashutosh Sharma[1], Manish Kumar Goyal[2]

Abstract

Rainfall forecasting has been one of the most challenging problem for scientists and engineers around the globe. A rainfall forecast is highly beneficial for efficient management of surface and subsurface water resources in arid regions. Rainfall is a complex hydrological phenomenon which depends on many atmospheric processes. These atmospheric conditions are generally used for rainfall prediction. Two atmospheric variables, Temperature and Cloud Cover, are used in this study for rainfall prediction. Rainfall distribution is not uniform among different stations but stations share a strong correlation, and this property has been used for improve the prediction accuracy of a model. Bayesian network offers framework to incorporate the conditional dependencies between the nodes/variables. K2 algorithm is used to finding out the rainfall dependencies among the stations. Thirty-two stations of Rajasthan, India are used for this study. Rainfall, temperature and cloud cover data of 102 years has been used for this study which is divided into standard 70:30 ratio for training and testing respectively. The prediction results of both cases (considering inter-station rainfall dependencies and without considering inter-station rainfall dependencies) are compared and an improvement in accuracy was found at most of the stations while considering inter-station dependencies.

Keywords

Bayesian Network, Bayesian Classification, Rainfall Forecast, Rainfall Prediction

Received: 16 April 2024; **Revised:** 20 July 2024; **Accepted:** 23 August 2024;
Published: 31 August 2024;

[1,2]Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India.

Corresponding author:
Email: ashutoshsharmaiit@gmail.com

Introduction

Rajasthan is one of the most drought affected states of India. Frequent occurrence of droughts cause huge loss in terms of economy and human lives. Lack of rainfall is the primary reason for all kind of droughts and hence, an accurate prediction of rainfall is required for better management of surface and sub-surface water resources. Rainfall is generally predicted at a station based on the historical records at that station or based on the atmospheric conditions at the station. Although the rainfall distribution over different stations in a region is not uniform but there is strong correlation between rainfalls at different stations. This property of inter-station correlation can be used for improving the rainfall prediction. In this paper, a Bayesian network approach has been used for predictions of rainfall at 32 stations of Rajasthan using Temperature and Cloud cover as predictors. Bayesian network is basically a graphical model which uses nodes and arrows for representing random variables and their conditional dependencies. It combines the principles of probability theory, graph theory, statistics and computer science.[1]

Advantages of Bayesian networks include better data modelling in case of missing data, avoiding overfitting of the data (which generally occurs in complex models), and combining prior knowledge and data.[2] The Bayesian network structure and the conditional probabilities can be found using various techniques available in literature.[3,4] Bayesian network have applications in very wide range of fields of science, engineering and medical sciences. Bayesian network was used as in an integrated model for studying the different processes involved in eutrophication.[5] Bayesian network has found some applications in the field of hydrology and rainfall prediction. The conditional dependency property of Bayesian network makes it suitable for studying the spatial distribution of the rainfall over the large number of stations. Bayesian network was used for weather predictions and the model was combined with numerical atmospheric predictions to make future weather forecasts.[6] Bayesian network can be used for dealing with predictions over a large number of stations. A model based of Bayesian network was used for stochastic weather generation and weather predictions over a network of 100 stations.[7] The rainfall is dependent mainly on the atmospheric conditions over the area and hence, it is beneficial to include some atmospheric variables for rainfall predictions. Rainfall prediction based on a Bayesian network model considering two classes of rainfall and using seven different atmospheric parameters achieved an accuracy above 90 percent.[8] Bayesian network can be used for prediction of rainfall and temperature based on regional and global weather data.[9] A comparison of 13 different combination of five atmospheric variables was done using Bayesian network based model for rainfall prediction at 21 stations of Assam, India.[10]

This study shows a comparison between the rainfall predictions considering the station dependencies and without considering the station dependencies. The inter-station rainfall dependencies are represented in the form of Bayesian network structure. This paper is organized in six sections. Study area and data are described in Section II. A brief overview of Bayesian network is presented in Section III. Methodology and overview of the model is described in Section IV. Section V shows the results and discussion. A comparison between the two techniques for rainfall prediction is presented in this section. Finally, conclusions of the study are presented in Section VI.

Study Area and Data

A. Study Area

Rajasthan is largest state of India by area and it has an area of 342,000 km^2 which accounts for around 10.4% area of the whole country. The state is situated between $23°3.5'N$ and $30°14'N$ latitude, and $69°27'E$ and $78°19'E$ longitude. Figure 1 shows location of Rajasthan on India map. The unique climatological characteristics of state which include low rainfall, extremes diurnal and annual temperatures, and high wind velocity, make it unique and different from other states.[11] Around 70% of the state is covered by the 'Thar Desert' which is also known as The Great Indian Desert. The climate over the state varies from arid to semi-arid. On the west side of the Aravali Hill Range the climate is arid, marked by low rainfall, extreme diurnal and annual temperature, low humidity and high wind velocities, whereas on the east side of Aravali Hill Range, the climate is semi-arid to semi-humid marked by higher rainfalls and lower wind velocity. The average annual rainfall of the state is 574 mm but there is high spatial variation of rainfall over the state. The western part of the state receives less rainfall compared to the eastern part of the state. The average annual rainfall in the western part varies between 100 to 400 mm, in the eastern part between 550 mm to 1020 mm and in the southwest part around 1638 mm.

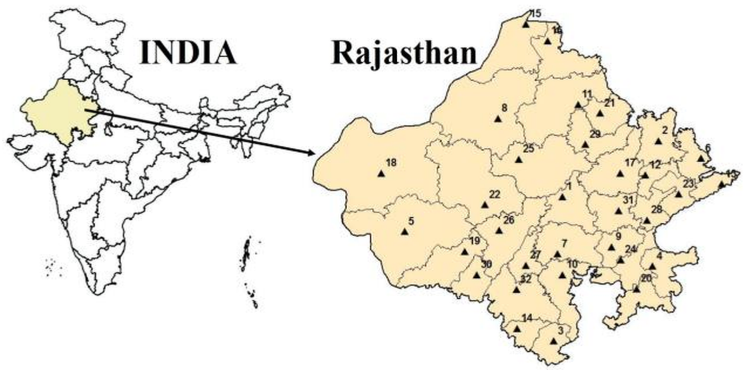

Figure 1. Study area.

B. Data

Mean monthly data of rainfall, temperature and cloud cover has been used for this work. Data was obtained from IMD website India water portal (http://www.indiawaterportal.org/met_data) from 1901 to 2002. The 102 years monthly data was divided into 70:30 ratio for training and testing of the model. Data from Jan, 1901 to Dec, 1970 was used for training and rest of the data was used for testing. Two atmospheric variables used are:

1. Temperature:

Mean monthly temperature in ^{o}C is used for this study. The data was obtained for all 32 stations of Rajasthan from India Water Portal website.

2. Cloud cover:

It is defined as the fraction of the sky obscured by clouds at a particular location and time. The value of cloud cover varies between 0 and 1 where 0 represents the condition when the sky is clear and 1 represents the condition when whole sky is covered with clouds.

Bayesian Network

A Bayesian network is a probabilistic graphical model which consists of nodes and arrows joining the nodes, and conditional probabilities. A node represents a random variable which can either be continuous or discrete, and an arrow shows dependency of one node on other. The graph containing nodes and arrows is called Bayesian network structure or a DAG (Directed Acyclic Graph). The node at the tail of arrow is called the parent node and node at head is called child node. 'Acyclic' in DAG means there is no complete cycle of arrows, which ensures a node can never be its own parent or child. In addition to this graph which is "qualitative" part, a Bayesian network also contains "quantitative" part which are the conditional probabilities at each node. The conditional probabilities are represented in the form of Conditional probability distribution (CPD) if the node is a continuous variable or Conditional probability table (CPT) if the node is discrete variable. It quantifies the effect of node on its immediate node. So the Bayesian network can be described a combination of Bayesian network structure (B_P) and Conditional probabilities (B_S).[3] Bayesian network is a combination of principles from graph theory, probability theory, computer science and statistics.[1]

The Joint Probability at any node can be expressed as the product of conditional probabilities as:

$$P_B\left(x_1, x_2, ..., x_n\right) = \prod_{i=1}^{n} P_B\left(x_i \mid x_{i+1}, ..., x_n\right) \tag{1}$$

where x_i represents the i-th node and n is the total number of nodes. Using conditional dependency property of Bayesian network, a node x_i is independent of all other nodes in the network structure except its parent nodes (π_i), where π_i denotes the parent nodes of i-th node. Hence the Joint Probability distribution is reduced to:

$$P_B\left(x_1, x_2, ..., x_n\right) = \prod_{i=1}^{n} P_B\left(x_i \mid \pi_i\right) \tag{2}$$

Bayesian Network Learning

Bayesian network learning is the process of finding out the Bayesian network structure (Structure learning) and computation of conditional probabilities (Parameter learning) from the available data for the random variables. Various learning techniques are available for both structure and parameter learning.[1] We have used K2 algorithm for structure learning, which is a greedy search algorithm to for finding the Bayesian network structure (B_S).[3] It starts with assumption that all nodes have no parent and nodes are ordered. For every node (x_i) in the given order, the algorithms adds parents (π_i) which has maximum increment in a score which defines the quality of network. The algorithms stops when no improvement in score is obtained by the further addition of nodes.[3] Maximum likelihood parameter estimation (MLE) techniques is used for parameter estimation.

Methodology

Rainfall is a natural phenomenon which varies both in time and space domain. It is not uniform over a large area and is driven by many atmospheric processes. The local rainfall at a station is dependent on the local atmospheric conditions prevailing over that area. To incorporate the effect of local conditions many atmospheric variable which include temperature, wind speed, humidity, cloud cover etc. have been used for forecasting the rainfall. Apart from using atmospheric variable at every station individually, the relation between the rainfalls at different stations can also be used for prediction. Although the rainfall is not uniform over the stations, but it has been noted that stations share a strong correlation in terms of rainfall. This rainfall correlation between the stations has been incorporated in the model and a comparison has been made with predictions made taking stations independent.

In this paper, we have compared two techniques for rainfall prediction. One in which rainfalls at different stations are assumed to be independent of each other and second in which the rainfalls at stations are assumed to be inter-dependent. In both the cases the rainfall is predicted using two atmospheric variables (Temperature and Cloud cover) individually as well as together.

A. Stations independent of each other: Different Bayesian networks are prepared for every station. Three different Bayesian network models are used by varying the predictor (atmospheric variable). (Figure 2)

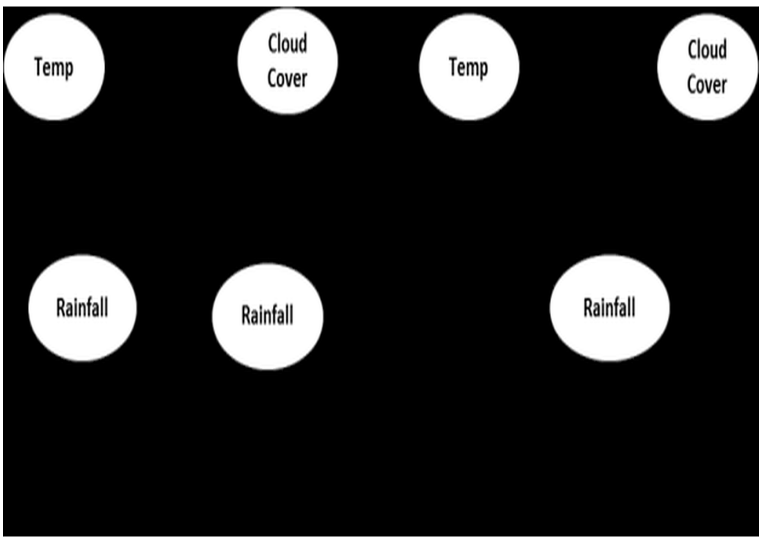

Figure 2. Bayesian model structure for independent station.

B. Stations are inter-dependent: Bayesian network structure learning is used to find out the dependencies between the stations based on rainfall. K2 algorithm is used for BN structure learning. Using the obtained Bayesian network structure, three Bayesian network models were prepared using rainfall, cloud cover, and rainfall and cover as predictors.

Steps:

1. Discretization of data
2. Bayesian network structure construction
3. Model training
4. Model Testing

1. Discretization of data

The Bayesian models were developed considering the discretized nodes. The data for rainfall, temperature and cloud cover was discretized in three classes. Three classes for rainfall represent very less rainfall, average rainfall and good rainfall with ranges 0 mm-10 mm, 10mm-200mm and more than 200mm respectively.

2. Bayesian network structure construction

The Bayesian network structure was obtained using K2 algorithm. The nodes were ordered alphabetically and the maximum number of parents for a node were taken as three.

3. Model Training

Out of 102 years monthly data, first 70 years data was used to train the model. Training includes computation of conditional probabilities at all the nodes.

4. Model Testing

Model performances were tested for remaining 32 years data. The data for evidence variables (Temperature and Cloud cover) was given and the rainfall was predicted at 32 stations. To evaluate the performance of the model, prediction accuracy was computed based on the percentage of correct prediction.

We have used BNT MATLAB Toolbox for this work.[12]

Results and Discussion

The Bayesian network structure was obtained using K2 algorithm as shown in Figure 3. The numbers represent the station numbers arrange alphabetically. Table I shows the model accuracy for all the stations for the case when the inter-station dependencies were considered. The station numbers are in accordance with the Figure 3. 1, 2, and 3 represent the cases when Temperature, Cloud cover, and both (Temp and Cloud Cover) are used a predictor. Cloud cover is better predictor than the temperature at all the stations. The combination of both predictors outperformed both the predictors used individually.

The overall accuracy for case 1, 2 and 3 are 67.31%, 77.84% and 78.35 %. Out of 32 stations, 10 stations have accuracy above 80%.These stations lies in the northern, western and southern parts of the state, whereas station in the center and eastern regions have less accuracy. Only 7 stations have accuracy below 75 %, these stations are located nearby in the north-eastern part of the station. This part of the state have high variations rainfall compared to other parts.

Figure 4 shows the comparison between the accuracy obtained with considering inter-station dependencies (using BN Structure) and without considering inter-station dependencies (Independent stations). 1, 2 and 3 represent the cases when Temperature, Cloud cover, and both (Temp and Cloud Cover) are used a predictor. It is clear from the results the taking into account the dependencies among the stations has improved the results at almost all the station, except a few stations which performed better independently. These stations are located in the north-eastern part of the stations, and have less accuracy (Table I)

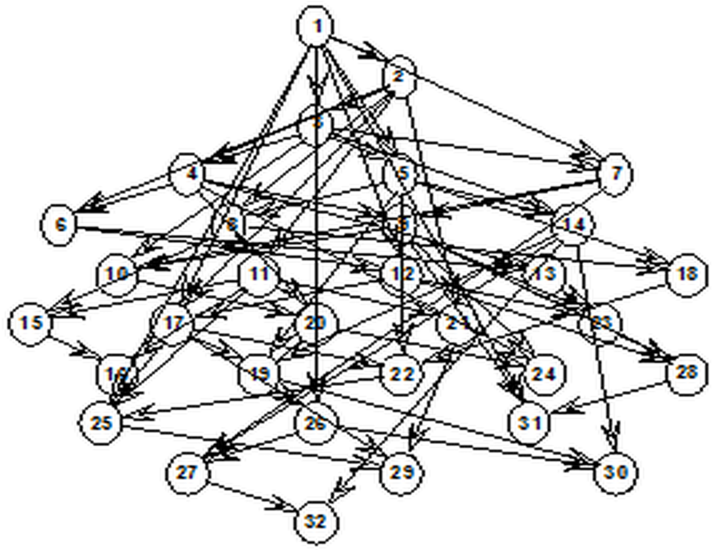

Figure 3. Bayesian network structure obtained using K2 algorithm.

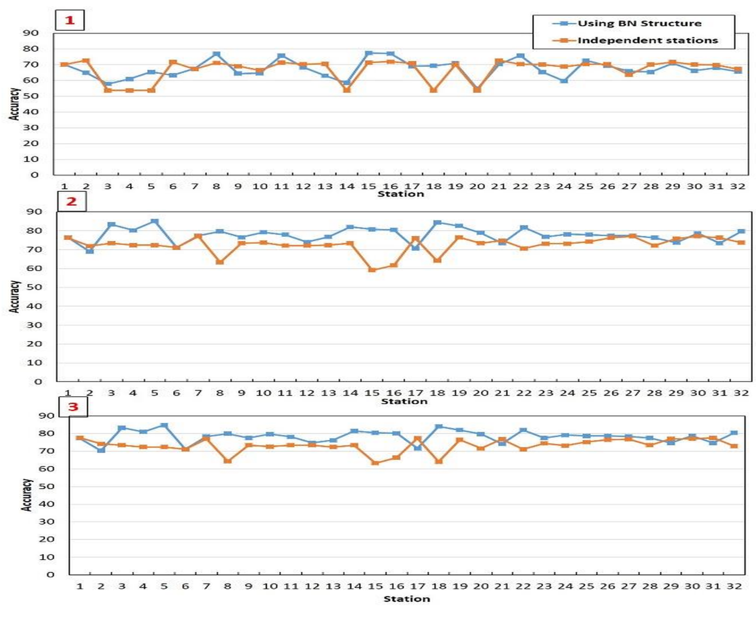

Figure 4. Comparison of prediction accuracy with and without considering station dependencies.

Table 1. Prediction accuracy considering interstation dependencies.

St No	Name	1	2	3
1	Ajmer	70.05	76.56	77.34
2	Alwar	64.84	69.01	70.31
3	Banswara	57.81	83.33	83.33
4	Baran	60.94	80.21	80.99
5	Barmer	65.36	85.16	84.90
6	Bharatpur	63.28	71.09	71.09
7	Bhilwara	67.45	77.34	78.39
8	Bikaner	76.82	79.69	79.95
9	Bundi	64.32	76.56	77.60
10	Chittorgarh	64.58	79.17	79.69
11	Churu	75.78	77.86	78.13
12	Dausa	68.23	73.96	74.74
13	Dholpur	63.02	76.82	76.30
14	Dungarpur	58.33	82.03	81.51
15	Ganganagar	77.34	80.73	80.47
16	Hanumangarh	77.08	80.47	80.21
17	Jaipur	69.01	70.57	71.61
18	Jaisalmer	69.27	84.38	84.11
19	Jalor	70.83	82.55	82.03
20	Jhalawar	54.69	78.91	79.69
21	Jhunjhunu	70.31	73.44	74.22
22	Jodhpur	75.78	81.77	82.03
23	Karauli	65.36	76.82	77.60
24	Kota	59.64	78.13	79.17
25	Nagaur	72.66	77.86	78.65
26	Pali	69.27	77.34	78.65
27	Rajsamand	65.89	77.34	78.39
28	Sawai Madhopur	65.36	76.30	77.60
29	Sikar	70.83	73.70	74.74
30	Sirohi	66.15	78.65	78.65
31	Tonk	67.97	73.44	74.74
32	Udaipur	65.63	79.69	80.47

Conclusion

Bayesian network model for rainfall prediction has been proposed in the paper which includes the inter-station dependencies. The study is performed over 32 stations of Rajasthan using monthly rainfall data of 102 years. The rainfall dependencies among the stations is represented in the form of Bayesian network structure which was obtained using K2 algorithm. Two atmospheric variable, Temperature and Cloud cover, are used for which mean monthly data was used. Rainfall prediction results obtained considering the inter-station dependencies were compared with prediction results obtained without considering inter-station dependencies. The comparison shows an improvement in prediction accuracy by using inter-station dependencies at most of the stations. Cloud cover was found to be better predictor than temperature. This study can be used for better management of water resources.

References

[1] Ben-Gal I. (2007) Bayesian Networks in Encyclopedia of Statistics in Quality & Reliability. Wiley & Sons.

[2] Heckerman David (1997) Bayesian Networks for Data Mining. Data Mining and Knowledge Discovery 1, 79–119 (1997)

[3] Cooper Gregory E, Herskovits Edwards (1992) A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning, 9, 309-347

[4] Cheng Jie, Bell David A. ,Liu Weiru (1997) An Algorithm for Bayesian Belief Network Construction from Data in the proceeding of AI & STAT'97

[5] Borsuk Mark E., Stow Craig A., Reckhow Kenneth H. (2004) A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling 173 (2004) 219–239

[6] Cofino Antonio S., Cano Rafael, Sordo Carmen, Gutierrez Jose M. (2002) Bayesian Networks for Probabilistic Weather Prediction in ECAI 2002. Proceedings of the 15th European Conference on Artificial Intelligence, IOS Press, 695 - 700 (2002).

[7] Cano Rafael, Sordo Carmen, Gutierrez Jose M. (2004) Applications of Bayesian Networks in Meteorology. in Advances in Bayesian Networks, Gámez et al. eds., 309-327, Springer, 2004.

[8] Nikam Valmik B, Meshram B.B. (2013) Modeling Rainfall Prediction Using Data Mining Method in 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation.

[9] Nandar, A. (2009, December). Bayesian network probability model for weather prediction. In Current Trends in Information Technology (CTIT), 2009 International Conference on the (pp. 1-5). IEEE.

[10] Sharma Ashutosh, Goyal Manish Kumar (2015) Bayesian network model for monthly rainfall forecast. Proceedings of the 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India. Nov, 2015

[11] Rathore, M. S. (2004). State level analysis of drought policies and impacts in Rajasthan, India (Vol. 93). IWMI.

[12] Murphy Kevin P. (2001) The Bayes Net Toolbox for Matlab. Computing Science and Statistics, vol 33, 2001. Code available at https://code.google.com/p/bnt/