

ARTICLE

Rainfall Estimation using Image Processing and Regression Model on DWR Rainfall Product for Delhi-NCR Region

Kuldeep Srivastava* Ashish Nigam

Regional Meteorological Centre, India Meteorological Department, New Delhi 110003, India

ARTICLE INFO

Article history

Received: 6 May 2020

Accepted: 22 May 2020

Published Online: 30 June 2020

Keywords:

Rainfall estimation

Rainfall analysis

Doppler Weather Radar

Precipitation Accumulation Product

Image processing

Linear regression model

ABSTRACT

Observed rainfall is a very essential parameter for the analysis of rainfall, day to day weather forecast and its validation. The observed rainfall data is only available from five observatories of IMD; while no rainfall data is available at various important locations in and around Delhi-NCR. However, the 24-hour rainfall data observed by Doppler Weather Radar (DWR) for entire Delhi and surrounding region (up to 150 km) is readily available in a pictorial form. In this paper, efforts have been made to derive/estimate the rainfall at desired locations using DWR hydrological products. Firstly, the rainfall at desired locations has been estimated from the precipitation accumulation product (PAC) of the DWR using image processing in Python language. After this, a linear regression model using the least square method has been developed in R language. Estimated and observed rainfall data of year 2018 (July, August and September) was used to train the model. After this, the model was tested on rainfall data of year 2019 (July, August and September) and validated. With the use of linear regression model, the error in mean rainfall estimation reduced by 46.58% and the error in max rainfall estimation reduced by 84.53% for the year 2019. The error in mean rainfall estimation reduced by 81.36% and the error in max rainfall estimation reduced by 33.81% for the year 2018. Thus, the rainfall can be estimated with a fair degree of accuracy at desired locations within the range of the Doppler Weather Radar using the radar rainfall products and the developed linear regression model.

1. Introduction

RADAR is an acronym for Radio Detection and Ranging. It is a device capable of detecting objects at far off distances, measuring the distance or range of the object by using electromagnetic waves. Radars have assisted weather predictions for over fifty years but its operational use in hydrologic applications spans only a decade or so. The hydrological applications of the radar have gained traction with the introduction of

Doppler Weather Radars.

Doppler weather radar makes use of the Doppler Effect to measure velocity of moving targets it detects. It works by detecting the change in the frequency of the transmitted and returned signal arising due to the movement of the target. The velocity component of a target relative to the radar beam is known as the “radial velocity”. The radar used for the purpose of this research is a Dual Polarized Doppler Weather Radar installed at IMD HQ, New Delhi. It is a C-Band Radar (Frequency: 4-8 GHz, Wavelength:

**Corresponding Author:*

Kuldeep Srivastava,

Regional Meteorological Centre, India Meteorological Department, New Delhi 110003, India;

Email: kuldeep.imd@gmail.com

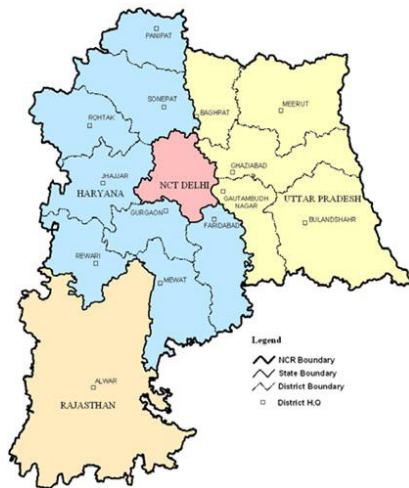


Figure 2. The district wise map of different regions covered

3. Data Sources

Data used for this study are:

(1) Surface Rainfall Intensity (SRI): The SRI generates an image of the rainfall intensity in a user selectable surface layer with constant height above ground. A user definable topographical map is used to find the co-ordinates of this surface layer relative to the position of the radar. This map is also used to check for regions, where the user selected surface layer is not accessible to the radar. These parts of the image will be filled with the NO DATA value. The product provides instantaneous values of rainfall intensity. The estimated values of reflectivity are converted to SRI by using $Z=AR^b$ [8] where R is the rainfall intensity, A and b are constants. The values of A & b vary from season to season and place to place.

(2) Precipitation Accumulation (PAC): The PAC product is a second level product. It takes SRI products of the same type as input and accumulates the rainfall rates in a user-definable time period (look back time). Every time a new SRI product is generated, the PAC generation starts again. The display shows the colour coded rainfall amount in [mm] for the defined time period. Precipitation Accumulation (PAC) Products generated by the DWR have been downloaded (in .gif format) from radar section of the official website of India Meteorological Department (https://mausam.imd.gov.in/imd_latest/contents/index_radar.php) that are updated on a daily basis. A bash script was written to automate this process. At a fixed time, each day, the script downloads the PAC product gif images from the official website and saves them in the repository with that day's date.

(3) Those PAC products that were not available on the

website have been generated in the Radar Lab from the RAW data of the concerned date (Figure 3). Raw Radar data of the concerned date was sent to the input folder of the IRIS software. This raw data was then ingested by the IRIS software. After the ingestion, Surface Rainfall Intensity (SRI) product was created for every 10-min interval. Once all the SRI products were created, RAIN-1 product was created for each hour. Thus, a total of 24 RAIN-1 products were created for the whole day. After that, a single RAIN-N (PAC) product was created for the whole 24-hour duration by overlapping all the RAIN-1 products (N can be programmed to have any value from 1 to 24).

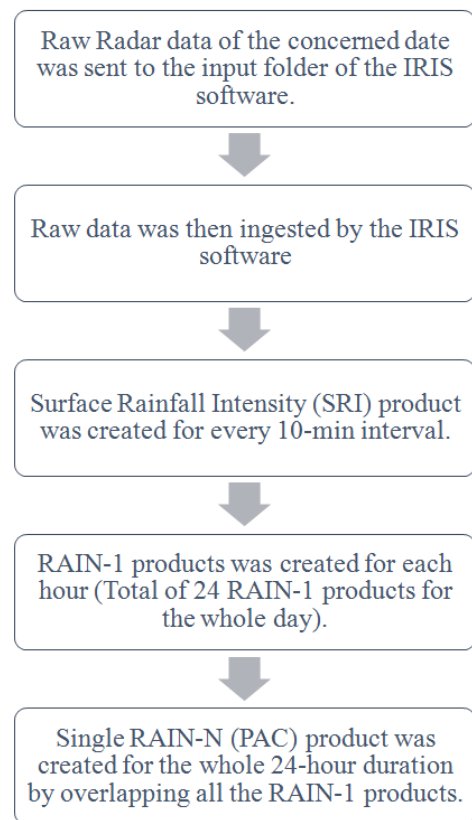


Figure 3. PAC product generation procedure

(4) Realised Rainfall data has been taken from the Month end rainfall reports submitted to Regional Meteorological Centre (RMC) Delhi by Meteorological Centre (MC) Lucknow and Meteorological Centre Chandigarh. These reports are submitted in the form of excel sheets. These sheets show date wise distribution of rainfall as well as cumulative amount of rainfall for each month at various observatories, automatic weather stations and automatic rain gauges.

4. Methodology

Visual Studio Code Editor was used to code the script in

Python 3 which would read the image and carry out image processing. Pixel coordinates corresponding to each concerned station were determined using Microsoft Paint and then coded in the script. The script first reads the PAC image (in .gif format) using PIL/Pillow library and then converts it into OpenCV image format. Image processing using OpenCV library was carried out to determine the amount of rainfall. The algorithm for calculating the rainfall amount takes into account the value of the pixel at the desired coordinate (Figure 4). The RGB combination corresponding to each colour in the Precipitation accumulation scale (present on the right side of the PAC image) and the associated rainfall range were stored in an array. The pixel coordinates for each of concerned station were also stored in an array. For each pixel coordinate, the RGB value was calculated. Based on the RGB value, the rainfall was calculated. This step was repeated for all the pixels lying in a 3*3 grid surrounding the station coordinates. The average rainfall value of all the nine pixels was allotted to the concerned station. This was done to eliminate/reduce the effect of noise.

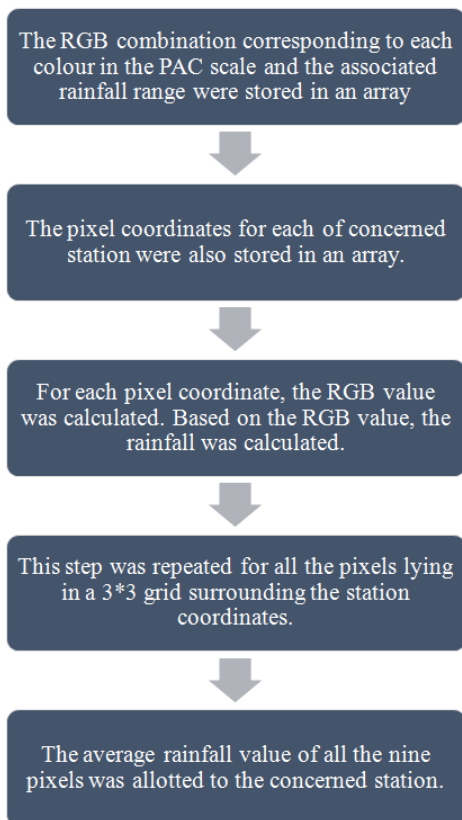


Figure 4. Image processing methodology

Data, thus calculated, was stored in csv format date wise, corresponding to PAC product of each date. We refer to this data as “Estimated Rainfall”. This was done for both 2018 and 2019. The above calculated data and the

data from various ground stations (We refer to this data as “Observed Rainfall”) for the year 2018 was then used to train a linear regression model in R. The model was then tested on the estimated and observed rainfall data of 2019 and the results were studied and analysed (Figure 5). R Studio Integrated Development Environment (IDE) was used for the purpose of the modelling the linear regression model and testing it out.

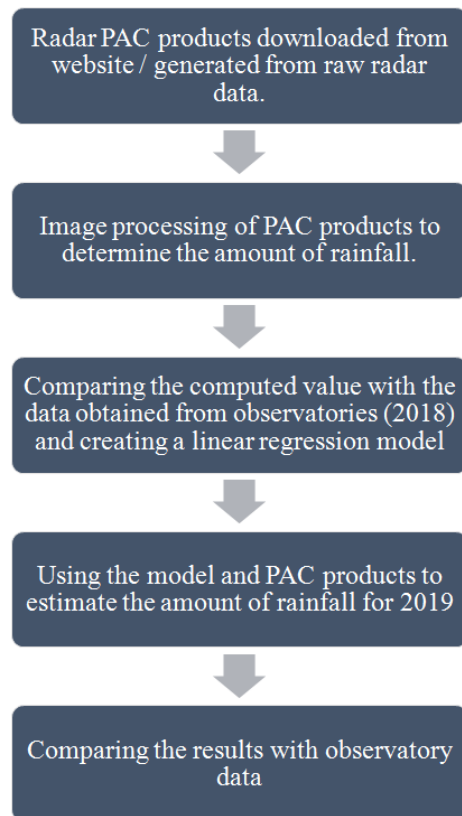


Figure 5. Research methodology

5. Results and Discussion

5.1 Analysis of Observed and DWR Estimated Rainfall for 2018

On comparing the observed and DWR estimated rainfall for 2018, it is found that there is a high positive correlation between the two (Figure 6). This comparison was done without use of the developed linear regression model. The minimum rainfall values shown by both the methods were identical while the mean rainfall value estimated using RADAR is 35.8% less than the mean rainfall observed during the concerned months in 2018. The maximum rainfall value estimated using RADAR is 57.33% less than the maximum rainfall observed during the concerned months in 2018 (Table 1).

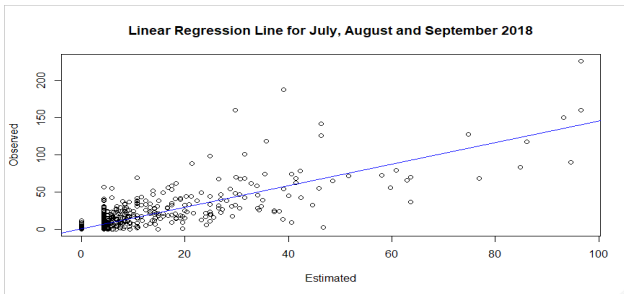


Figure 6. Scatter plot between observed and estimated rainfall values for July, August and September 2018

Table 1. Estimated and observed rainfall comparison for 2018

Estimated Rainfall	Observed Rainfall
Min. : 0.000	Min. : 0.000
Mean : 4.419	Mean : 6.888
Max. : 96.500	Max. : 226.200

5.2 Analysis of Observed and DWR Estimated Rainfall for 2019

On comparing the observed and DWR estimated rainfall for 2019, it is found that there is a high positive correlation between the two (Figure 7). This comparison was done without the use of the developed linear regression model. The minimum rainfall values shown by both were identical while the mean rainfall value estimated using RADAR is 22.85% less than the mean rainfall observed during the concerned months in 2019. The maximum rainfall value estimated using RADAR is 28.24% less than the maximum rainfall observed during the concerned months in 2019 (Table 2).

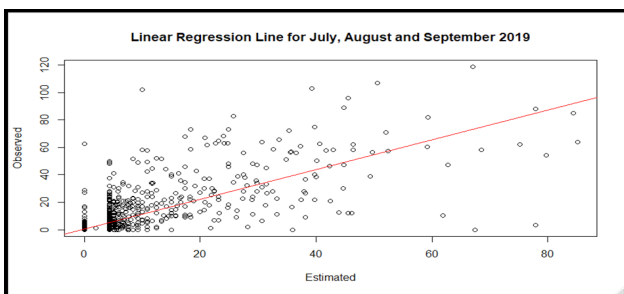


Figure 7. Scatter plot showing positive correlation between observed and estimated rainfall values for July, August and September 2019

Table 2. Estimated and observed rainfall comparison for 2019

Estimated Rainfall	Observed Rainfall
Min.: 0.000	Min.: 0.000
Mean: 3.285	Mean: 4.258
Max.: 85.100	Max.: 118.600

5.3 Development of Model using Least Square Method

The model is found using the least square method using the rainfall data from 2018. The minimum and maximum residuals associated with the model are -66.388 and 130.167 respectively (Table 3). The residuals, which are vertical distance between data points and the regression line, are both positive and negative, in this case which shows that data points are evenly distributed on both sides of the regression line.

Table 3. Model residuals

Residuals:	Min	1Q	Median	3Q	Max
	-66.388	-0.461	-0.461	-0.461	130.167

The equation derived from the model is:

$$\text{Actual Rainfall} = 1.45454 * (\text{Rainfall estimated from PAC product}) + 0.4607 \tag{1}$$

The standard error associated the coefficient of the independent variable (i.e. rainfall estimated using PAC product) is 0.02557 (Table 4). Such a low value indicates that there is only a small error in the coefficient. The ratio of coefficient value and its associated standard error gives t-value and its high value (in our case: 56.879) indicates high degree of accuracy of the coefficient of the independent variable.

Table 4. Statistical parameters of the model

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.46071	0.29202	1.578	0.115
Estimated	1.45454	0.02557	56.879	<2e-16

The value of R-squared is 0.6895 which indicates that 68.95% of the total variation in the dependent variable is being explained by our equation/model. The p-value is less than 2.2e-16 and thus it is less than 10%, 5% and even 1% level of significance. Therefore, we can safely reject the null hypothesis and conclude that our independent variable is significant.

For the scope of this paper, the intercept value in Eq. (1) has been ignored, since it is very small (i.e. less than 1mm) and on the assumption that when the radar indicates that there was no rainfall, there was no observed rainfall.

5.4 Verification of Model Estimated and Observed Rainfall

After training the linear regression linear model with 2018 rainfall data, the model was tested on 2019 rainfall data

and the results have been shown in Table 5.

Table 5. Image estimated rainfall, Model estimated rainfall and Observed rainfall comparison for 2019

Rainfall estimated directly from image	Rainfall estimated by model	Rainfall Observed
Min.: 0.000	Min.: 0.000	Min.: 0.000
Mean: 3.285	Mean: 4.7778	Mean: 4.258
Max.: 85.100	Max.: 123.7811	Max.: 118.600

With the use of the linear regression model, it can be seen that the mean rainfall value estimated using model is 12.2% more than the mean rainfall observed during the concerned months in 2019 and the maximum rainfall value estimated using model is 4.3% more than the maximum rainfall observed.

The linear regression linear model was re-run on the training data set (2018 rainfall data) and the results have been shown in Table 6.

Table 6. Image estimated rainfall, Model estimated rainfall and Observed rainfall comparison for 2018

Rainfall estimated directly from image	Rainfall estimated by model	Rainfall Observed
Min.: 0.000	Min.: 0.000	Min.: 0.000
Mean: 4.419	Mean: 6.4278	Mean: 6.888
Max.: 96.500	Max.: 140.3628	Max.: 226.200

With the use of the linear regression model, it can be seen that the mean rainfall value estimated using model is 6.68% less than the mean rainfall observed during the concerned months in 2018 and the maximum rainfall value estimated using model is 37.94% less than the maximum rainfall observed.

5.5 Discussion

With the use of linear regression model, the error in mean rainfall estimation reduced by 46.58% and the error in max rainfall estimation reduced by 84.53% for the year 2019. The error in mean rainfall estimation reduced by 81.36% and the error in max rainfall estimation reduced by 33.81% for the year 2018.

The root mean square error between the model estimated and observed rainfall value was found to be as low as 0.9807644 mm (i.e. less than 1mm) which is evident in Figure 8, Figure 9 and Figure 10

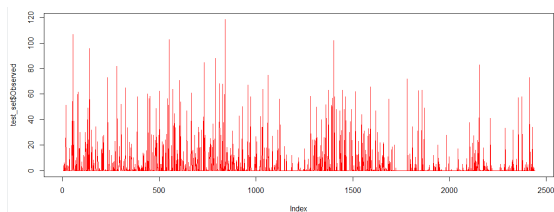


Figure 8. Plot of Observed value corresponding to different index values

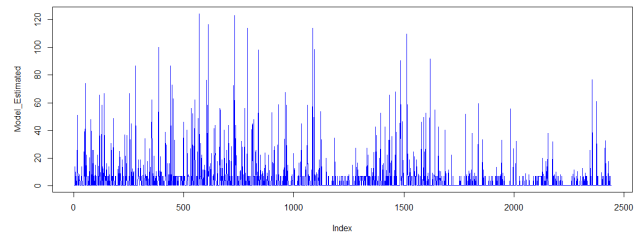


Figure 9. Plot of Model Estimated values corresponding to different index values

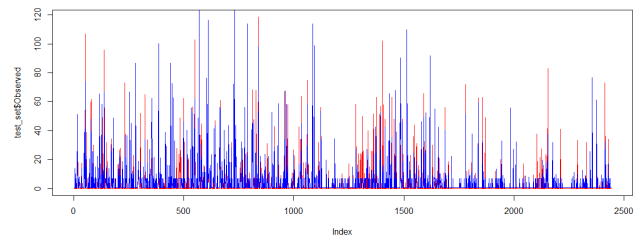


Figure 10. Comparison of Model Estimated and Observed values

6. Conclusions

Rainfall data was successfully extracted and actual rainfall was estimated from the hydrological products of the RADAR using image processing and regression for the monsoon period for the years 2018 and 2019. From the above results and discussion, it can be concluded that the rainfall at any desired location can be estimated using hydrological products generated by the Doppler Weather Radar.

The developed linear regression model can be used to estimate the rainfall amount to a great degree of accuracy over a range of rainfall values. The equation obtained by regression is:

$$\text{Actual Rainfall} = 1.45454 * (\text{Rainfall estimated from PAC product})$$

However, it was also observed that when the rainfall is very heavy (115.6 – 204.4 mm) to extremely heavy (>= 204.5 mm) the difference between the estimated and observed rainfall is large as compared to when the rainfall is light to moderate.

This usage of hydrological products generated by weather radars can be extended to radars located at other locations across India and used to estimate rainfall where no observatories/AWS/ARGs are available.

Acknowledgments

The authors are thankful to Director General of Meteorology (DGM), India Meteorological Department (IMD), New Delhi and Deputy Director General of Meteorology (DDGM), RMC New Delhi for rendering all the facilities. The authors also gratefully acknowledge MC Chandigarh,

MC Lucknow and RMC Delhi for providing the rainfall data of Delhi-NCR and adjoining areas, DWR Division and Radar Lab, I.M.D, New Delhi for providing all the necessary PAC product images to carry out this study.

References

- [1] Anagnostou E.N., Krajewski W.F. Real-Time Radar Rainfall Estimation. Part I: Algorithm Formulation. *Journal of Atmospheric and Oceanic Technology*, 1999a, 16: 189-197.
- [2] Anagnostou E.N., Krajewski W.F. Real-Time Radar Rainfall Estimation. Part II: Case Study. *Journal of Atmospheric and Oceanic Technology*, 1999b, 16: 198-205.
- [3] Anagnostou E.N., Anagnostou M.N., Krajewski W.F., Kruger A., Miriovsky B.J. High-Resolution Rainfall Estimation from X-Band Polarimetric Radar Measurements. *Journal of Hydrometeorology*, 2004, 5: 110-128.
- [4] Brandes E.A. Optimizing Rainfall Estimates with the Aid of Radar. *Journal of Applied Meteorology*. 1975, 14: 1339-1345.
- [5] Brandes E.A., Zhang G., Vivekanandan J. Experiments in Rainfall Estimation with a Polarimetric Radar in a Subtropical Environment. *Journal of Applied Meteorology*, 2012, 41: 674-685.
- [6] Dutta D., Sharma S., Sen G.K., Kannan B.A.M, Venkateswarlu S., Gairola R.M., Das J., Viswanathan G. An Artificial Neural Network based approach for estimation of rain intensity from spectral moments of a Doppler Weather Radar. *Advances in Space Research*, 2011, 47: 1949-1957.
- [7] Krajewski W. F., Smith J. A. Radar hydrology: rainfall estimation. *Advances in Water Resources*. 2002, 25: 1387-1394.
- [8] Marshall J.S., Langille R.C., Palmer W. McK. Measurement of Rainfall by Radar. *Journal of Meteorology*. 1947, 4: 186-192.
- [9] Smith J.A., Baeck M.L., Meierdiercks K.L., Miller A.J., Krajewski W.F. Radar rainfall estimation for flash flood forecasting in small urban watersheds. *Advances in Water Resources*. 2007, 30: 2087-2097.
- [10] Sun X., Mein R.G., Keenan T.D., Elliott J.F. Flood estimation using radar and rain gauge data. *Journal of Hydrology*. 2000, 239: 4-18.
- [11] Wilson J.W., Brandes E.A. Radar Measurement of Rainfall - A Summary. *Bulletin American Meteorological Society*, 1979, 60: 1048-1058.
- [12] Xiao R., Chandrasekar V. Development of a neural network-based algorithm for rainfall estimation from radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, 1997, 35: 160-171.