A Composite Method of Rainfall Rates for a Multi-Parameter Phased Array Weather Radar and XRAIN using Machine Learning

Shota Ochi[1], Noritsugu Shiokawa[1], Yoshimasa Egashira[1], Masakazu Wada[1],
Tetsuya Kobayashi[2], Tadashi Yamada[3]

[1]TOSHIBA Corporation, Tokyo, Japan
[2]TOSHIBA Infrastructure Systems & Solutions Corporation, Kanagawa, Japan
[3]Chuo University, Tokyo, Japan

1. Introduction

As part of its activities in the Strategic Innovation Program, Toshiba has developed X-band multi-parameter phased array weather radar (MP-PAWR) capable of high-speed three-dimensional observations by electrical scanning in the elevation direction using a phased-array technique and mechanical scanning in the azimuth direction. The terms "multi-parameter" and "MP" mean dual-polarization. The MP-PAWR is expected to enable rapid detection of severe weather events such as torrential rains and tornados.

Obtaining accurate rainfall rate data over wide areas requires composite rainfall rate data from multiple radars because single-radar observations have limitation in range, their accuracy deteriorates due to rainfall attenuation, and buildings or topography can create blank areas. There is thus a need for composite methods where rainfall rates are obtained from MP-PAWR alongside other radars. Conventional radars in Japan include the X- and C-band parabolic dual-polarization radars that comprise the Ministry of Land, Infrastructure, Transport and Tourism’s eXtended RAdar Information Network (XRAIN). The objective of this study was to improve the accuracy of rainfall rates by generating composite rainfall rate data from MP-PAWR data and XRAIN composite data. Assuming raw observation data from all radars that comprise XRAIN are unavailable to an MP-PAWR operator, we propose a machine learning-based method to selectively composite MP-PAWR data and XRAIN composite data. MP-PAWR and XRAIN composite data accuracy depends on the observations and weather conditions, so we use 17 features representing the MP-PAWR and XRAIN observation conditions and weather conditions as input to the machine learning model.

We used data from the MP-PAWR installed at Saitama University (Figure 1) and actual XRAIN composite data to generate composite rainfall rates. We evaluated the accuracy of the composite rainfall rate data using true values obtained from rain gauges.

MP-PAWR antenna (left) and the MP-PAWR radome installed at Saitama University (right). (Adapted from Takahashi et al. [1])

2. The Proposed Composite Method

2.1 Concept

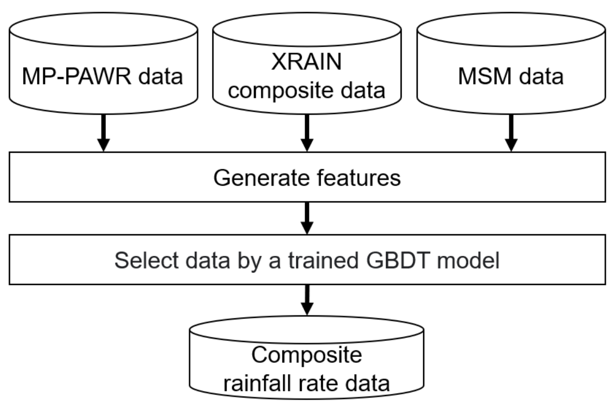

Figure 2. Concept of the proposed composite method.

Figure 2 shows the concept of the proposed composite method, which selectively composites MP-PAWR data and XRAIN composite data on Cartesian axes of latitude and longitude with a horizontal resolution of about 250 m. The selection is made by a gradient boosting decision tree (GBDT) model, which is a machine learning model that uses

17 features to estimate which of MP-PAWR data or XRAIN composite data are closer to the actual ground rainfall. Those features, representing the MP-PAWR and XRAIN observation conditions and weather conditions, are generated from MP-PAWR, XRAIN composite, and MSM data. Section 2.2 describes these source data, and Section 2.3 describes the features.

2.2 Data Description

This section describes the data used in this study.

As mentioned above, we used three data types (MP-PAWR, XRAIN composite, and MSM data) to generate the 17 features for input to the GBDT model. MP-PAWR data were obtained from the MP-PAWR installed at Saitama University. XRAIN composite data are composite ground rainfall rates from XRAIN X- and C-band dual-polarization radars. MSM data are numerical weather prediction data using the Japan Meteorological Agency (JMA) meso-scale model.

We also used data obtained from rain gauges of the Automated Meteorological Data Acquisition System (AMeDAS), a JMA meteorological observation system, to train the GBDT model and evaluate the composite rainfall rate data. Specifically, we used data obtained from 41 AMeDAS rain gauges in the MP-PAWR observation area.

Table 1 describes the MP-PAWR, XRAIN composite, and MSM data, along with AMeDAS rain gauge data. Figure 3 shows the locations of the MP-PAWR at Saitama University, the X-band dual-poralization radars (X-MP radars) of XRAIN near the MP-PAWR, and the AMeDAS rain gauges in the MP-PAWR range.

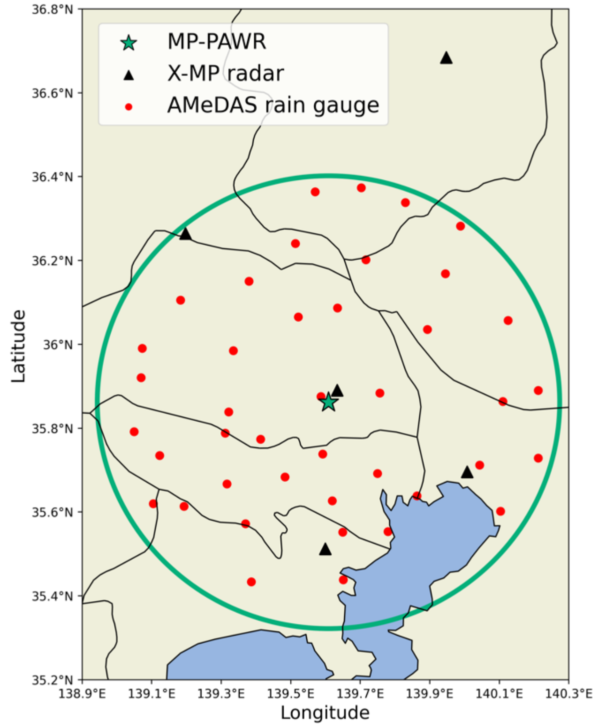

Figure 3. Locations of MP-PAWR, XRAIN X-MP radars, and AMeDAS rain gauges.

Table 1. Description of Data Used in This Study

Data	Description
MP-PAWR Data	Data obtained from the MP-PAWR Installed at Saitama University ● Obs. range: 80km (data within 60 km used in this study) ● Obs. Azimuth: 360 deg. ● Obs. Elevation: 0 – 90 deg. ● Range resolution: 150 m ● Azimuth beamwidth: < 1.2 deg. ● Temporal resolution: 30 seconds
XRAIN Composite Data	Composite rainfall rate data of X-band and C-band dual-polarization radars of XRAIN ● Horizontal resolution: about 250 m ● Temporal resolution: 1 minute
MSM Data	NWP data using the Meso-Scale Model by JMA ● Horizontal resolution: 5 km ● Altitudes: surface and 16 pressure levels (from 1000hPa to 100hPa, including 850hPa and 700hPa) ● Initial time: every 3 hours including 12:00 UTC (6 times a day) ● Delivered time: each initial time + 2.5 hours ● Forecast time interval: 1 hour (surface), 3 hours (pressure level) Note that the information above is about the specifications of the data those forecast time until 39 hours.
AMeDAS Rain Gauge Data	Data obtained from the rain gauges of Automated Meteorological Data Acquisition System (AMeDAS) which is a JMA meteorological observation system ● Rainfall resolution: 0.5 mm ● Temporal resolution: 10 minutes and 1 hour (1 hour data used in this study)

2.3 Features

This section describes the 17 features of the GBDT model inputs. We selected these features through trial and error to obtain a trained GBDT model with high generalization performance for various rainfall cases.

Four Features Associated with MP-PAWR Data

We generated four features associated with MP-PAWR data: rainfall rate, KDP ratio, path-integrated attenuation (PIA), and distance between MP-PAWR and each grid point of the composite rainfall rate data. These features represent the MP-PAWR observation conditions.

Figure 4 shows the steps for generating the rainfall rate, KDP ratio, and PIA.

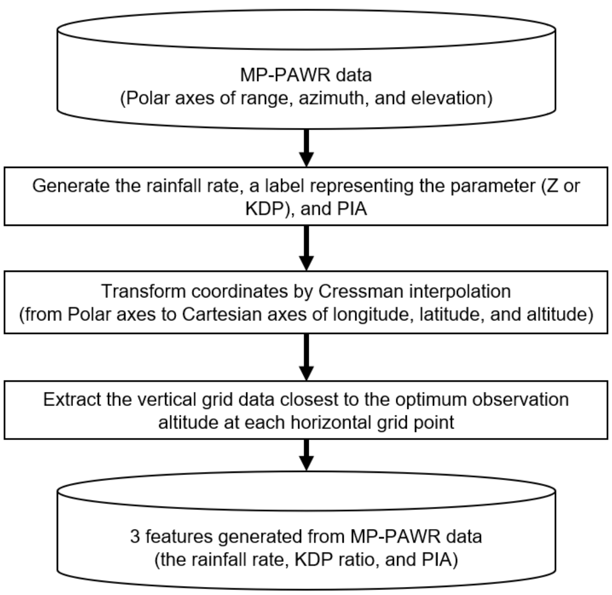

Figure 4. Method to generate three features from MP-PAWR data.

First, we generate the rainfall rate, a label representing the parameter used for rainfall rate estimation, and PIA in polar coordinates. Here, we calculate the MP-PAWR rainfall rate using the same rainfall rate estimation method as XRAIN [2]. This method calculates the rainfall rate from either radar reflectivity (Z [dBZ]) or the rate of change in distance in phase difference between polarizations [◦/km], called KDP depending on the observation conditions. KDP rainfall rates are generally more accurate than Z-calculated rates. Therefore, the label representing the parameter used for rainfall rate estimation (0 representing Z, 1 representing KDP) is a feature that represents the rainfall rate quality. PIA is calculated from KDP using the XRAIN method [3].

Next, we transform the rainfall rate, the label representing the rainfall rate estimation parameter, and PIA from polar to Cartesian longitude, latitude, and altitude coordinates by Cressman interpolation. The horizontal grid resolution is approximately 250 m, the same as XRAIN composite data and the generated composite data, while the vertical grid is set at 100 m intervals for altitudes between 500 and 1000 m and at 250 m intervals for altitudes between 1250 and 5250 m. The KDP ratio is obtained by transforming the label representing the rainfall rate estimation parameters to Cartesian grid points. The KDP ratio, a value between 0 and 1, indicates how much of the rainfall rate at each Cartesian grid point is attributed to the rainfall rate calculated by KDP. The closer the KDP ratio is to 1, the greater the attribution by KDP, and the closer it is to 0, the greater the attribution by Z.

Finally, we extract vertical grid data closest to the observation altitude with the highest correlation with ground rainfall (the optimum observation altitude; identified in a previous analysis) at each horizontal grid point.

Four Features Associated with XRAIN Composite Data

We generated four features associated with XRAIN composite data: the rainfall rate, the average distance between five X-MP radars near the MP-PAWR (KANTOU, SHINYOKO, FUNABASHI, YATTAJIMA, and UJIIE) and the grid point of the composite rainfall rate data, and the average of the rainfall rate and the percentage of grid points with observed rainfall in a 120 × 120 km area centered on the MP-PAWR at Saitama University. These features represent the XRAIN observation conditions and area rainfall distribution.

One Feature Generated from MP-PAWR Data and XRAIN Composite Data

We generated the difference between the rainfall rates of MP-PAWR and XRAIN composite data as a feature.

Eight Features Generated from MSM Data

We selected or generated eight features from MSM data. From surface data, we selected low cloud cover, horizontal wind velocity, and pressure. We selected vertical velocity from 700-hPa level data. We also generated difference values for 850 and 700-hPa levels of relative humidity, temperature, vertical velocity, and horizontal wind velocity. These features represent weather conditions that might be useful for generating a model that considers rainfall types (convective rainfall, typhoons, stratiform rainfall, etc.). We considered the following points when generating these features:

- Applicability to real-time processing
  As MSM data are delivered at the initial time + 2.5 h, we used predicted data after the initial time + 3 h with a 30-min margin. We generate features for each time point by linear interpolation of the predictions before and after each time point in the time direction.

- Consideration of misalignment risk
  We used averaged data over a 55 × 55 km area centered on the MSM data grid point closest to each composite rainfall rate grid point at the center.

3. Training and Evaluation Method

3.1 Dataset

We used data comprising 461 h between July and October each year from 2019 to 2022, including various rainfall cases. This dataset contained only data where the AMeDAS rain gauges observed hourly rainfall of at least 0.5 mm. To increase the number of evaluation cases while reducing the data used for training as little as possible, we divided training and evaluation datasets into two groups. Table 2 shows the evaluation cases for each group. Each group has three evaluation cases: convective rainfall, typhoon, and stratiform rainfall. For each group, the trained GBDT model was generated from data excluding the evaluation cases of each group. The training data consisted of 415 h for Group 1 and 413 h for Group 2.

Table 2. Evaluation Cases.

Group	Type	Period	Number of data
1	Convective	2020/08/12 11:00 - 20:00	62
	Typhoon	2019/10/12 00:00 - 24:00	815
	Stratiform	2019/10/29 06:00 - 19:00	314
2	Convective	2020/08/22 16:00 - 24:00	27
	Typhoon	2019/09/08 02:00 - 2019/09/09 09:00	435
	Stratiform	2020/10/17 10:00 - 23:00	487

3.2 Training Method

We used LightGBM [4] as the GBDT model implementation. The training label represents MP-PAWR data or XRAIN composite data, whichever has the smaller absolute error relative to hourly rainfall as obtained from the rain gauges. We used hourly average rainfall rates in the MP-PAWR data and XRAIN composite data to match the sampling frequency of the rain gauges. We also used hourly averages of each feature except KDP ratios in the training process. For KDP ratios, we used the value $r_{\mathrm{KDP\_hour}}$ calculated by Equation (1), with hourly features representing the contribution of the KDP-calculated rainfall rate to hourly rainfall. In Equation (1), n is the number of hourly samples, Rr is the rainfall rate in each data sample, and r_{KDP} is the KDP ratio in each data sample.

$$r_{\mathrm{KDP\_hour}} = \frac{\sum_{k=0}^{n} Rr(k) * r_{\mathrm{KDP}}(k)}{\sum_{k=0}^{n} Rr(k)} \qquad (1)$$

3.3 Evaluation Method

We performed evaluations under some operational assumptions. Specifically, we generated the composite rainfall rate for every minute using by-minute features, then evaluated hourly averages of composite rainfall rate data using the root mean square error (RMSE) with hourly rainfall obtained from the AMeDAS rain gauges as true values. We calculated the RMSE using Equation (2), where A [mm/h] is the hourly rainfall obtained from the AMeDAS rain gauges, R [mm/h] is the hourly average in the evaluated rainfall rate data, and N is the number of data.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_i - A_i)^2} \qquad (2)$$

4. Evaluation Results

4.1 Accuracy

Table 3 shows the RMSE of rainfall rates in the XRAIN composite data, MP-PAWR data, the composite data generated by the maximum method (Max.), which selects the maximum value from MP-PAWR data and XRAIN composite data (the “Max method”), and the composite data generated by the proposed method. Here, we evaluated the Max method as an example of a simple composite method. Figure 5 shows the RMSE ratio for each rainfall rate data to the RMSE of XRAIN composite data shown in Table 3. Note that the RMSE of MP-PAWR data is inferior to that of XRAIN composite data because MP-PAWR data are single radar observations, while XRAIN composite data are composites from many X- and C-band dual-polarization radars at least five radars near the MP-PAWR shown in Figure 2.

In Group 1 in Figure 5(a), the RMSE of the composite rainfall rate generated by the Max method is worse than the XRAIN composite data for the convective and stratiform rainfall cases. By contrast, the proposed method generated composite rainfall rate data more accurate than the XRAIN composite data in all cases. Specifically, the composite rainfall rate data generated by the proposed method improves RMSE by 13% for convective rainfall, 3% for typhoons, and 18% for stratiform rainfall, as compared to XRAIN composite data.

Similarly, in Group 2 in Figure 5(b), the RMSE of the composite rainfall rate generated by the Max method was worse than XRAIN composite data for all three cases, and the RMSE of the composite rainfall rate generated by the proposed method improves over the XRAIN composite data in all three cases. The improvement ratios are 13% for convective rainfall, 2% for typhoons, and 16% for stratiform rainfall.

The above confirms that the proposed method can generate composite rainfall rates from MP-PAWR data and XRAIN composite data that are more accurate than either alone.

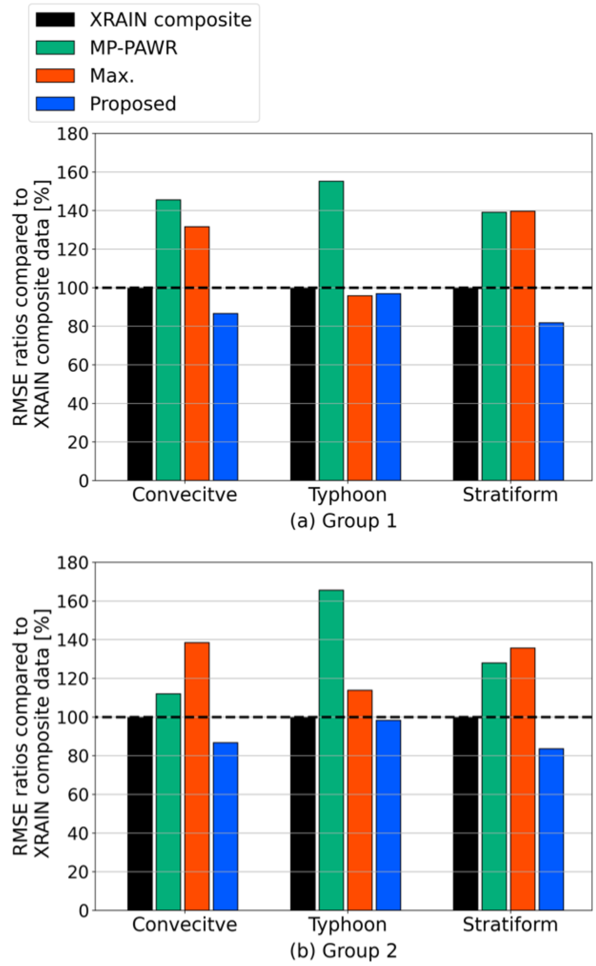

Figure 5. RMSE ratios compared to XRAIN composite data.

Table 3. RMSE of XRAIN composite data, MP-PAWR data, and composite rainfall rate data, as generated by the Max method and the proposed method.

Group	Type	Before Composite		Composite data	
		XRAIN composite	MP-PAWR	Max. method	Proposed method
1	Convective	6.17	8.98	8.12	5.34
	Typhoon	5.75	8.92	5.51	5.57
	Stratiform	1.92	2.67	2.68	1.57
2	Convective	8.38	9.38	11.6	7.26
	Typhoon	3.48	5.76	3.96	3.42
	Stratiform	2.61	3.34	3.54	2.18

4.2 Feature Importance

To analyze the behavior of the trained GBDT models, we confirmed the feature importance of each trained GBDT model. Table 4 shows the feature importance of the trained GBDT models for each group. In both groups, the sum of feature importance representing the MP-PAWR and XRAIN observation conditions (Nos. 1–6 and 9) was about 47%, while the sum of feature importance representing weather conditions (Nos. 7, 8, and 10–17) was about 53%. These results indicate that the trained GBDT models select MP-PAWR or XRAIN with consideration of both radar observation and weather conditions.

4.3 Composite Rainfall Rate Map

Figure 6 shows the rainfall rate map of XRAIN composite data, MP-PAWR data, and the composite rainfall rate data generated by the proposed method in a typhoon case (2019/10/12, 15:00) in Group 1 as an example. Here, missing areas and areas outside the MP-PAWR observation range were simply filled in with XRAIN composite data in Figure 6(c).

Figure 6(c) shows that MP-PAWR data are selected more in the strong rainfall area near the MP-PAWR, while XRAIN data are selected in other areas, including accuracy deteriorated areas due to rainfall attenuation of the MP-PAWR. We can also see that the MP-PAWR observations are partly reflected in the strong rainfall area in the upper right area. In cases where strong rainfall occurs over a wide area, such as during typhoons, the rainfall rate tends to be underestimated due to the large rainfall attenuation, and we assume that the trained GBDT models select larger rainfall rates in such cases.

Table 4. Feature Importance

No.	Category	Feature	Importance [%]	
			Group1	Group2
1	MP-PAWR Data	Rainfall rate [mm/h]	7.2	6.9
2		KDP ratio	5.4	4.1
3		Path Integrated Attenuation [dB]	5.4	4.3
4		Distance between radar site and grid-point [km]	3.1	2.9
5	XRAIN Composite Data	Rainfall rate [mm/h]	7.2	7.8
6		Average distance between each X-band radar site and grid-point [km]	4.5	4.0
7		Average of rainfall rate [mm/h]	4.2	3.4
8		Ratio of grid points where rainfall was observed	3.7	4.7
9	MP-PAWR Data and XRAIN Composite Data	Difference value between rainfall rate of MP-PAWR data and XRAIN composite data [mm/h]	14.4	17.1
10	MSM Data (Surface)	Low cloud cover [%]	7.2	7.8
11		Horizontal wind velocity [m/s]	3.3	4.1
12		Pressure [Pa]	5.4	5.7
13	MSM Data (700hPa level)	Vertical velocity [Pa/s]	6.4	5.7
14	MSM Data (Difference value of 850hPa level and 700hPa level)	Relative humidity [%]	5.8	4.9
15		Temperature [K]	5.7	6.5
16		Vertical velocity [Pa/s]	5.1	5.0
17		Horizontal wind velocity [m/s]	5.7	5.1

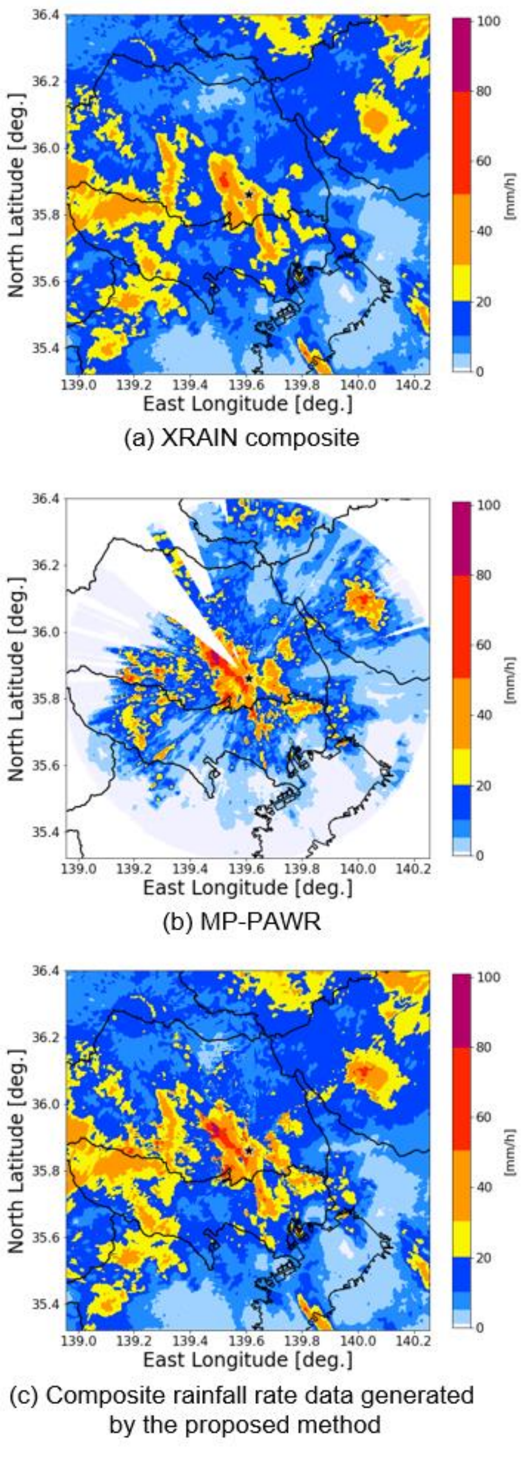

Figure 6. Example maps of XRAIN composite data, MP-PAWR data, and composite rainfall rate data generated by the proposed method.

5. Conclusion

We proposed a new rainfall composite method that aims to generate composite rainfall data from MP-PAWR data and XRAIN composite data that are more accurate than both. The proposed method selectively composites MP-PAWR data and XRAIN composite data on Cartesian latitude and longitude axes using the GBDT model, which estimates whether MP-PAWR data or XRAIN composite data are closer to actual ground rainfall. The method uses 17 features representing the MP-PAWR and XRAIN observation conditions and weather conditions, as input to the machine learning model.

We found composite rainfall rates generated by the proposed method to be more accurate than XRAIN composite data and MP-PAWR data in all six evaluated cases, with RMSE accuracy improvements of 2–18% over XRAIN composite data. A feature importance analysis of the trained GBDT models indicated that the trained GBDT models select MP-PAWR data or XRAIN composite data considering both radar observation and weather conditions.

Acknowledgment

The MP-PAWR data used in this study were provided by the National Institute of Information and Communications Technology (NICT). The operation of the MP-PAWR was supported by the Cross-ministerial Strategic Innovation Promotion Program (SIP) 2nd Phase, “Research on the sophistication and utilization of information about the early onset and forecasting development of Senjo-Kousuitai”.

The XRAIN composite data used in this study were provided by the Ministry of Land, Infrastructure, Transport and Tourism. This dataset was collected and provided under the Data Integration and Analysis System (DIAS), which was developed and operated under a project subsidized by the Ministry of Education, Culture, Sports, Science and Technology.

References

[1] Nobuhiro Takahashi, et al, “Development of Multi-Parameter Phased Array Weather Radar (MP-PAWR) and Early Detection of Torrential Rainfall andTornado Risk,” Journal of Disaster Research Vol.14 No.2, 2019.

[2] S. Tsuchiya, M. Kawasaki, and H. Godo, “Improvement of the Radar Rainfall Accuracy of Xrain by Modifying of Rainfall Attenuation Correction and Compositing Radar Rainfall,” J. of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), Vol.71, No.4, pp. I 457-I 462, 2015 (in Japanese).

[3] H. Godo, M. Naito, and S. Tsuchiya, “Improvement of the observation accuracy of X-band dual polarimetric radar by expansion of the condition to use KDP-R relationship,” J. of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), Vol.70, No.4, pp. I 505-I 510, 2014 (in Japanese).

[4] Guolin Ke, Qi Meng, Thomas Finley, TaifengWang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree, 31st Conference on Neural Information Processing Systems,” 2017.