Hindawi
Advances in Meteorology
Volume 2019, Article ID 6542410, 17 pages
https://doi.org/10.1155/2019/6542410

Research Article

Assessing the Applicability of Random Forest, Stochastic Gradient Boosted Model, and Extreme Learning Machine Methods to the Quantitative Precipitation Estimation of the Radar Data: A Case Study to Gwangdeoksan Radar, South Korea, in 2018

Ju-Young Shin, Yonghun Ro, Joo-Wan Cha, Kyu-Rang Kim, and Jong-Chul Ha

Applied Meteorology Research Division, National Institute of Meteorological Sciences, 33, Seohobuk-ro, Seogwipo-si, Jeju-do 63568, Republic of Korea

Correspondence should be addressed to Yonghun Ro; royh1@korea.kr

Received 12 March 2019; Revised 23 July 2019; Accepted 23 August 2019; Published 7 October 2019

Guest Editor: Jongho Kim

Machine learning algorithms should be tested for use in quantitative precipitation estimation models of rain radar data in South Korea because such an application can provide a more accurate estimate of rainfall than the conventional ZR relationship-based model. The applicability of random forest, stochastic gradient boosted model, and extreme learning machine methods to quantitative precipitation estimation models was investigated using case studies with polarization radar data from Gwangdeoksan radar station. Various combinations of input variable sets were tested, and results showed that machine learning algorithms can be applied to build the quantitative precipitation estimation model of the polarization radar data in South Korea. The machine learning-based quantitative precipitation estimation models led to better performances than ZR relationship-based models, particularly for heavy rainfall events. The extreme learning machine is considered the best of the algorithms used based on evaluation criteria.

1. Introduction

Quantitative precipitation estimation (QPE) using remote sensing data has been widely used to investigate the spatial characteristics of precipitation events [1, 2]. This method can be used to obtain rainfall estimation at ungauged locations, cloud characteristics, and areal rainfall depth [3–6]. The spatial resolution of rain radar data is the finest of all these. While the spatial resolution of satellite images is greater than approximately 10 km, the spatial resolution of rain radar data is approximately 1 km [7–9]. Because of the spatial resolution of rain radar data, it is often applied into rainfall-runoff modeling, particularly in terms of flash flood and urban flood modeling [10, 11]. The accurate forecast of these extreme hydrological events can mitigate damages on the hydraulic infrastructure and prevent the crisis of water-related disaster on human life. The accurate QPE of radar data is the key for the accurate forecast of extreme hydrological events.

Reflectivity and rainfall rate (ZR) relationship-based models have been used broadly for QPE models of rain radar data [12–14]. Because ZR relationship can be changed based on the characteristics of the rainfall event and the radar instrument used, various methodologies are applied to build ZR relationship-based QPE models and correct their estimations [15–18]. However, the ZR relationship-based model still has high uncertainty in a rainfall estimation [19–21].

Machine learning (ML) algorithms have been widely employed to create functional relationships for natural phenomena and data processing. Many ML algorithms were developed and employed to model a function in fields such as meteorology, hydrology, and agriculture. Applications of ML algorithms can provide accurate models of natural phenomena [22–25] and thus can be good candidates for

QPE of rain radar data. Recently, random forest (RF), stochastic gradient boosted model (GBM), and extreme learning machine (ELM) have been actively employed as ML algorithms [26–28]. These advanced ML algorithms, which have been tested recently, would increase our capacity to build QPE model. Chiang et al. [29], proposed a QPE model using a recurrent neural network and three-dimensional radar data. They reported that the ML-based model produced more accurate estimations than the ZR relation-based model. Yu et al. [30] attempted to develop quantitative precipitation forecast (QPF) models of rain radar data using RF and support vector regression. Their proposed methodology focused on QPF models for typhoons in Taiwan and performed well.

To the best of our knowledge, advanced ML algorithms, e.g., RF, GBM, and ELM, have not been employed for QPE of rain radar data in South Korea. This should be resolved because applying ML algorithms may provide more accurate rainfall estimation of rain radar data than the conventional ZR relation-based model. Therefore, this study investigated the applicability of the ML algorithms for QPE using Gwangdeoksan radar station, South Korea, as a case study in order to enhance performance of QPE in radar data. RF, GBM, and ELM are the ML algorithms used; their applicability is investigated using four rainfall events, and their performances for the QPE model are compared. This study can provide fundamental information on the development of QPE model using ML algorithms in South Korea. Particularly, the characteristics of ML algorithm for QPE model of radar data can be briefly investigated in the study. This result can enhance our capacity to understanding ML algorithms in the QPE of radar data. In addition, the most plausible candidate among the employed QPE models will be selected for ML-based QPE model of radar data in South Korea. The selected QPE model can lead to improvements in accuracy of QPE, particularly in extreme rainfall events that cause extreme hydrological events. The improvement in accuracy of QPE may help to mitigate impacts from extreme hydrological events on the destruction of property and human life.

This paper is organized as follows. In Section 2, the characteristics of the radar and ground gauge rainfall data are presented. Section 3 presents a description of the methods employed, e.g., ZR relationship and ML algorithms. The application methodology for the case studies is presented in Section 4. In Section 5, the results of tested QPE models for all events and each event are presented. Finally, the conclusions are presented in Section 6.

2. Data

2.1. Radar and Ground Rainfall Gauge. Gwangdeoksan weather radar station, which has a dual-polarization weather radar with an S-band, is located on the border of Gyeonggi-do and Gangwon-do provinces close to Seoul (latitude $38^{\circ}7'2.5''$, longitude $127^{\circ}26'1''$, and elevation 1064 m), the capital of the Republic of Korea. The observation range of the Gwangdeoksan radar is 240 km, which is enough to cover the northern part of South Korea. Radar data within the effective observation range, 100 km, are applied to QPE. Considering the high elevation of this radar station, the relationship between the radar and ground rainfall gauge data is increased with the application of PPI0. PPI (plan-position indicator) is an intensity-modulated display on which echo signals are shown in plan view with range and azimuth angle displayed in polar coordinates. PPI0 is volume scanned data when the azimuth angle is 0 which represents a condition that can be observed by minimizing blocking in a flat state. Data with a spatial resolution of 1 km × 1 km and stored at 10-minute interval are applied to estimate radar rainfall.

The three main polarization parameters of the radar, i.e., reflectivity, differential reflectivity, and specific differential phases, are applied to QPE in this study. Radar reflectivity refers to the ratio between the transmitted and received energies. The differential reflectivity is the ratio between horizontal and vertical radar reflectivity; it can provide information on the sizes and shapes of raindrops. Specific differential phases are the rate of change of the range in pulse phases, because these are not affected by attenuation, partial beam blockage, or radar miscalibration; they are an attractive parameter to use in QPE [31].

Rainfall rate data from ground gauge stations in Seoul and Gyeonggi-do province within the radar umbrella are analyzed in this study. All stations obtain rainfall data every minute, but the QPE in this study uses rainfall rate data at 10-min intervals. This is to compare the radar data and to minimize the fluctuation of ground gauge data. Figure 1 shows the ground gauges densely distributed across the Korean peninsula; of these, 20 gauges within the radar effective range are selected. The number of 20 stations is located in near Seoul and had severe storm damage in the past. The location of Gwangdeoksan radar and the selected rainfall gauges are described in the zoomed area in Figure 1, and information on each station is given in Table 1. The used data can be downloaded from the data base of Korea Ministration Administration (KMA) at data.kma.go.kr.

2.2. Rainfall Events. Rainfall events for which the depth of the observed daily rainfall exceeded 30 mm from August to November 2018 are used. Four events are selected as case studies. Two events (the first and second) occurred from August 28–29 (event #1) and on September 3 (event #2). The third event occurred from October 5–6 (event #3), and the fourth event happened on November 8 (event #4). Heavy precipitation was observed in Korea, brought on by a rainy front (the Changma front) for event #1 and by low pressure in the northern area for event #2. Event #3 occurred as part of Typhoon KONG-REY, and event #4 was accompanied by the collision of a cold and a warm front.

Table 2 summarizes information on the total rainfall, mean rainfall rate, and standard deviation for each event at the station in Seoul (#108). The amount of rain that fell during event #1 was the largest (larger than 100 mm) in two days. Both events #1 and #4 show that the weather front greatly affected the increase in rainfall or cloud formation. The average rainfall is very large in the rainy season and

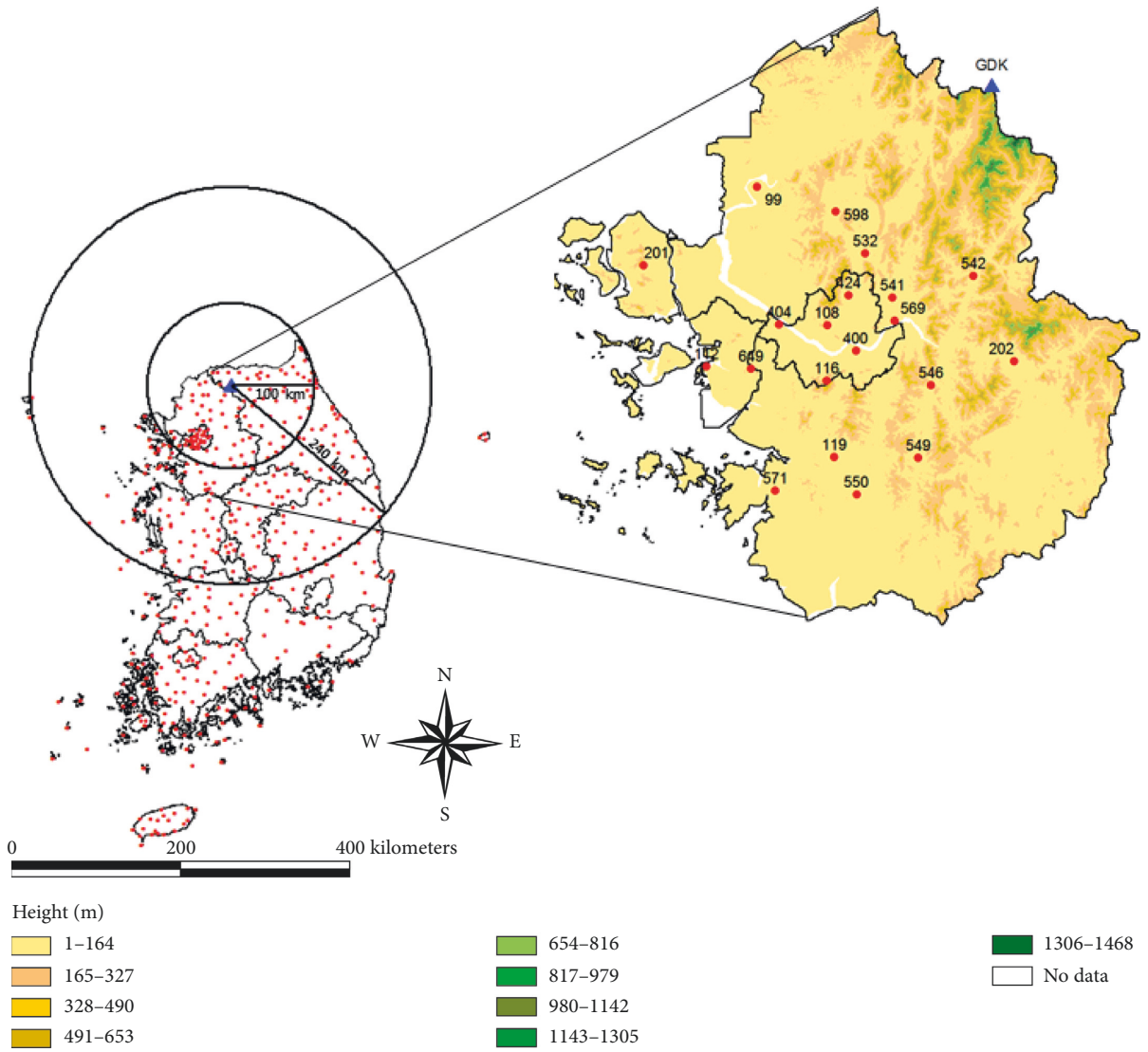

FIGURE 1: Locations of the Korea Meteorological Administration ground gauge stations (red dots) and Gwangdeoksan (GDK) radar (blue triangle) with its radar umbrella (effective radius: 100 km; maximum observation radius: 240 km). The figure in the right-hand panel presents the locations of the ground gauge stations (red dots) used in the current study.

typhoons, such as in events #1 and #3. The largest variance was observed in event #1. This heavy rainfall during the rainy season is representative of the summer monsoon climate in South Korea.

3. Methods

3.1. ZR Relationship. Radar rainfall can be defined by the relationship between radar parameters and rainfall gauge data. A variety of synthetic algorithms have been proposed to estimate quantitative radar rainfall based on the polarization parameters applied [32, 33]. The basic form of the equation, which is well-known as ZR relationship, is given as follows:

$$R = \theta_0 x_1^{\theta_1}, \ \ldots, x_d^{\theta_d}, \quad (d = 1, 2, \ \ldots, n), \tag{1}$$

where R is the ground gauge rainfall rate (mm/h), $x_1, \ \ldots, x_d$ are radar polarization parameters such as reflectivity, differential reflectivity, and specific differential phase, and $\theta_0, \ \ldots, \theta_d$ are the parameters of the ZR relationship. The main radar polarization parameters are defined as the following equations:

$$\begin{aligned} Z &= 10 \log (Z_{\mathrm{H}}), \\ \mathrm{DR} &= 10 \log \left(\frac{Z_{\mathrm{H}}}{Z_{\mathrm{V}}} \right), \\ \mathrm{KD} &= \left| \frac{\varnothing_{\mathrm{DP}} (r_2) - \varnothing_{\mathrm{DP}} (r_1)}{r_2 - r_1} \right|, \end{aligned} \tag{2}$$

where Z is the radar reflectivity, (changed from $\mathrm{mm}^6\mathrm{m}^{-3}$ to dBZ); Z_{H} and Z_{V} are horizontal and vertical reflectivity; DR is differential reflectivity (dB); KD is specific differential phase (deg km^{-1}); and $\varnothing_{\mathrm{DP}}$ and r are phases of the radar beam pulse and given range, respectively. Because the ZR relationship stands on the physical phenomena, the results

Table 1: Ground precipitation gauge stations selected for this study.

Name	Code	Latitude	Longitude
Paju	99	37.885	126.766
Seoul	108	37.571	126.965
Incheon	112	37.477	126.624
Suwon	119	37.272	126.985
Ganghwa	201	37.707	126.446
Yangpyeong	202	37.488	127.494
Gwanak	116	37.445	126.964
Gangnam	400	37.513	127.046
Gangseo	404	37.573	126.829
Gangbuk	424	37.639	127.025
Uijeongbu	532	37.734	127.073
Namyangju	541	37.634	127.150
Daeseongri	542	37.684	127.380
Gwangju	546	37.435	127.259
Yongin	549	37.270	127.221
Osan	550	37.187	127.048
Guri	569	37.582	127.156
Hwaseong	571	37.195	126.820
Yangju	598	37.831	126.990
Bupyeong	649	37.472	126.750

Table 2: Precipitation events selected based on observed rainfall data from Seoul station.

No.	Periods of precipitation events	Total (mm)	Mean (mm/h)	Standard deviation (mm/h)
1	2018.08.28. 11 : 50–2018.08.29. 22 : 20	138.5	4.0	11.9
2	2018.09.03. 08 : 50–2018.09.03. 21 : 30	34.5	2.7	5.8
3	2018.10.05. 08 : 20–2018.10.06. 12 : 20	92.0	3.3	3.3
4	2018.11.08. 01 : 30–2018.11.08. 23 : 40	64.0	2.8	3.5

of ZR relationship can be used to interpret characteristics of precipitation events, unlike the ML algorithms. The ML algorithms used in this study are the predictive models. Though they can be used to predict rain rate, extracting physical meaning from the results is difficult. For example, the parameters of ZR relationship can be used to identify type of cloud, type of precipitation, and type of storm events. In the case of the ML algorithms, prediction models for each variable of interest such as type of cloud, type of precipitation, and type of storm events have to be individually built.

3.2. Machine Learning (ML) Algorithms

3.2.1. Random Forest. RF has been widely applied in regression and forecasting problems [34–37]. It was proposed by Breiman [38] and uses bagging (called bootstrapping in statistics) to build a number of decision trees with a controlled variance. Each decision tree in the RF is grown using randomly selected samples. Subsequently, the nodes in each tree use randomly selected features (called input variables). The RF has two major steps: (1) randomness and (2) ensemble learning. The randomness in the RF comes from random sampling of the entire data set, and the selection of features with which every classification and regression tree (CART) is built. The data set is randomly sampled with replacement to create a subset with which to train one CART. At each node, optimal split rule is determined by using the one of the randomly selected features from the employed features.

The ensemble learning method in the RF means that all individual decision trees in a collection of decision trees (called an ensemble) contribute to a final prediction. A training subset is created after the random selection step. The CART without pruning is used to construct a single decision tree. To grow K trees in the ensemble, this process (resampling a subset and training an individual tree) is repeated K times. The final predicted value comes from averaging the results of all the individual trees. The *ranger* library in *r package* is used to construct the RF model in the current study [39].

3.2.2. Stochastic Gradient Boosted Model. GBM is a method widely used in classification and regression problems; it was proposed by Friedman [40]. Decision stumps or regression trees are used widely as weak classifiers in the GBM [40–42]. In the GBM, weak learners are trained to decrease loss functions, e.g., mean square errors. Residuals in the former weak learners are used to train the current weak learners. Therefore, the value of the loss function in the current weak learners decreases. The bagging method is employed to reduce correlation between weak learners, and each weak learner is trained with subsets sampled without replacement from the entire data set. The final prediction is obtained by combining predictions by a set of weak learners.

The GBM and RF adapted ensemble learning with a decision tree model (the weak learner). Both models produce

one prediction based on a combination of predictions from a set of weak learners. Though the methods seem to be similar, they are based on different concepts. The major difference between the GBM and RF is that the tree in the GBM is fit on the residual of a subset of the former trees while the RF trains a set of weak learners using a number of subsets. Therefore, the GBM can reduce bias of prediction while the RF method can reduce variance of prediction. Therefore, the RF can be trained in parallel computing, whereas the GBM cannot. The *gbm* library in *r package* (https://github.com/gbm-developers/gbm) is used to construct a GBM in the current study.

3.2.3. Regularized Extreme Learning Machine. ELM was originally developed and then extended to generalized single-hidden layer feed-forward networks in which the hidden layer need not to be neuron alike. ELM is a single-layer network in which the weights and biases between input and hidden layers are randomly generated [43]. Unlike traditional iterative learning algorithms, the randomly initiated input weights and biases of ELM remain fixed without need to iteratively tuned, and the output weights are determined analytically. Hence, the model can be trained in a single iteration which significantly reduces the training time of ELM and makes ELM efficient for online and real-time applications. The ELM can be formulized using the following equations:

$$\mathbf{Y} = \mathbf{H}\boldsymbol{\beta}, \tag{3}$$

where $\mathbf{Y}$, $\boldsymbol{\beta}$, and $\mathbf{H}$ are the outputs, weight matrix between hidden and output layers, and the output vector of the hidden layer (called nonlinear feature mapping), respectively, and

$$\mathbf{H} = f_{\mathrm{a}}(\mathbf{X}\mathbf{W} + \mathbf{B}), \tag{4}$$

where $f_{\mathrm{a}}(\cdot)$, $\mathbf{W}$, $\mathbf{B}$, and $\mathbf{x}$ are the activation function, weight matrix between the input to hidden layer, bias, and inputs, respectively. In the current study, the sigmoid function ($f_{\mathrm{a}}(x) = 1/(1 + \exp(-x))$) is used as the activation function in the ELM. Since the weights ($\mathbf{W}$) and bias ($\mathbf{B}$) are randomly generated and the activation function ($f_{\mathrm{a}}(\cdot)$) is known in the ELM, $\mathbf{H}$ represents the deterministic variables from a data set. Thus, only $\boldsymbol{\beta}$ needs to be estimated in the ELM.

In the ELM, finding an appropriate weight set is to avoid overfitting. Tuning weights in the ELM can be considered a fitting linear regression model using the ordinary least square method. Ridge regression was employed to attenuate multicollinearity in the data set by adding the norm of the parameters to the parameter estimations in the regression model [44]. The ELM model also adapted this strategy for weight tuning. The ELM attempts to perform better generalization by achieving the smallest training error and the smallest output weight norm. This minimization problem can take the form of ridge regression or regularized least squares as follows [45]:

$$\min \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\|\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}\|^2, \tag{5}$$

where the first term of the objective function is l_2, the norm regularization term that controls the complexity of the model; the second term is the training error associated with the learned model; and $C > 0$ is a tuning parameter. The ELM gradient equation can be solved analytically, and the closed-form solution can be written as follows:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^T\mathbf{H} + \frac{1}{C}\mathbf{I}\right)^{-1}\mathbf{H}^T\mathbf{Y}, \tag{6}$$

where $\mathbf{I}$ is an identity matrix. The ELM models used in this study are the regularized ELM model.

4. Application Methodology

To examine the applicability of the three ML algorithms, their input variables and hyperparameters should be defined. Z, DR, and KD in the polarization radar data have been widely employed as input variables for the QPE model. Therefore, these variables are used as input variable candidates in the ML algorithms. The tested models with input variable combinations are presented in Table 3.

Three ML algorithms use variables from both lag-zero (L0) and lag-one (L1) radar data for input variable while lag-zero and lag-one radar data, respectively, are used to construct ZR relationships. Since radar data measure the amount of cloud in the air, there is a short time difference between radar data and ground gauge observation. The time difference depends on the precipitation event conditions such as wind speed, cloud movements, and types of cloud. As the ZR relationship cannot account for time lag in its formula, QPE models based on the ZR relationship are constructed using different time-lag data and their appropriateness are investigated.

The ML algorithms can use both lag-zero and lag-one radar data simultaneously. In addition, the number of variables from the radar data (three) is much smaller than the number of data points (greater than thousands). A larger number of input variables might improve the predictability of the ML algorithms employed. Additionally, since this input variable setting can take the time lag in modeling into account automatically, additional processes such as the ZR relationship are unnecessary in ML-based models.

To evaluate the performances of the models constructed, the data set should be grouped into training and test data sets. The data from stations #112, #201, #400, #546, and #571 are used as randomly selected test data. The data at the other stations are used for the training data set. For the case of all events, the numbers of training and test data are 3652 and 1209, respectively. The numbers of training data for event 1 to 4 are 1079, 319, 1173, and 1081, respectively. The numbers of test data for event 1 to 4 are 318, 107, 441 and 343, respectively. To build a regression model using ML algorithms, their hyperparameters should be tuned. The number of the trees is the most sensitive hyperparameter for the RF and GBM [30]; hence, the number of trees for the RF and GBM are optimized. The tuning parameter and the number of

Table 3: Tested models for quantitative precipitation estimation from radar data.

Model	Input variables				
	Z	Z, DR	Z, KD	DR, KD	Z, DR, KD
ZR-L1	ZR1-L1	ZR2-L1	ZR3-L1	ZR4-L1	ZR5-L1
ZR-L0	ZR1-L0	ZR2-L0	ZR3-L0	ZR4-L0	ZR5-L0
RF	RF1	RF2	RF3	RF4	RF5
GBM	GBM1	GBM2	GBM3	GBM4	GBM5
ELM	ELM1	ELM2	ELM3	ELM4	ELM5

hidden nodes are the hyperparameters of the ELM. The relationship between the tuning parameter and the number of hidden nodes presents a trade-off relationship such as the Pareto frontier. Thus, after one parameter is fixed, another will be optimized. In this study, the tuning parameter is fixed ($C = 0.5$), and the number of hidden nodes is optimized.

In the current study, leave-one-out cross-validation (LOOCV) is employed to optimize the hyperparameters of the three ML algorithms. The root-mean-square error (RMSE) between the estimates and observations is calculated for the ML algorithms trained by the data set that does not include any station among all those in the training set. The expected numbers of train and test data are 3227 (approximate 94%) and 233 (approximate 6%), respectively. The fifteen models were trained, and their performances were evaluated using the test data set. The RMSEs without each station are calculated, and average value of these RMSEs is the criterion for measuring appropriateness of the hyperparameters. The results of the LOOCV are presented in Figure 2. The optimal numbers of the tree for the RF and GBM are 380 and 4200, respectively; any numbers greater than these do not lead to significant improvements in increasing the performance of the RF and GBM. The optimal number of hidden nodes for the ELM is 950. These numbers are used for the hyperparameters of ML algorithms.

QPE models are built for five case studies. The first case study uses all data including the four precipitation events. The other case studies built QPE models for each of the precipitation events. The first case study was carried out to evaluate the overall performances of the QPE models constructed. The results of the other case studies may provide detailed examinations of the performance of the different rainfall events. The RMSE, Pearson correlation, mean absolute error (MAE), mean bias (Mbias), and relative root-mean-square error (RRMSE) are employed as evaluation criteria. Equation (7) gives the equation of the RMSE:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - O_i)^2}, \tag{7}$$

where E_i, O_i, and n are ith radar estimation, ith observed precipitation data point, and the number of data points, respectively. The correlation can be calculated using the following equation:

$$\mathrm{correlation} = \sqrt{\frac{\sum_{i=1}^{n}(E_i - \overline{E})(O_i - \overline{O})}{\sum_{i=1}^{n}(E_i - \overline{E})\sum_{i=1}^{n}(O_i - \overline{O})}}, \tag{8}$$

where $\overline{E}$ and $\overline{O}$ are the means of the radar estimates and observed precipitation data, respectively. MAE, Mbias, and RRMSE equations are given in equations (9)–(11), respectively.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|E_i - O_i\right|, \tag{9}$$

$$\mathrm{MBias} = \frac{1}{n}\sum_{i=1}^{n}(E_i - O_i), \tag{10}$$

$$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\overline{o}} \times 100. \tag{11}$$

5. Results

5.1. Overall Performance of the QPE Models. The overall performances of the constructed QPE models are evaluated using rainfall and radar data for all the events. The training and test data sets are constructed from the data set that included all the rainfall events. Evaluation criteria for the constructed QPE models are applied to the test data set. Results of evaluation criteria are presented in Figure 3. The ML-based models lead to lower RMSEs than ZR relationship-based models.

When the number of input variables increases, the RMSE becomes smaller. For the ZR relationship-based models, RMSEs of ZR-L1-based models are smaller than those of ZR-L0-based models. The result means that usage of lag-data may provide more information onto QPE of the employed radar data. Models that include all the available input variables lead to lowest RMSE values. The second lowest RMSE is observed for models that use Z and DR as input variables. Models using DR and KD lead to the largest RMSE. Based on RMSE, the ELM5 (using Z, DR, and KD) is the best model for QPE of the radar data. Correlation results are similar to the results of the RMSE. The ML-based models give larger correlations than ZR relationship-based models. For correlation, the cases using all input variables provide the largest correlation values. Based on MAE, models using Z and DR as input variables lead to the smallest MAE. The best model based on MAE is the ELM2 (using Z and DR). Based on MBias, ZR-L0-based models are the best models, with an MBias close to zero. Estimations by ZR-L1-based models have positive biases except for the ZR4-L1, while those of ML-based models have negative biases. RRMSEs of all employed QPE models are larger than 100%. Based on the RRMSEs, the ELM is the best model for QPE of radar data. The second best model is the RF.

Estimation-verse observation plots are presented in Figure 4. Rainfall rate estimations are underestimated for large amounts of rainfall rates (larger than 40 mm/h). These underestimations for large amount of rainfall rate are clearly observed in the results of ZR-L0-based models.

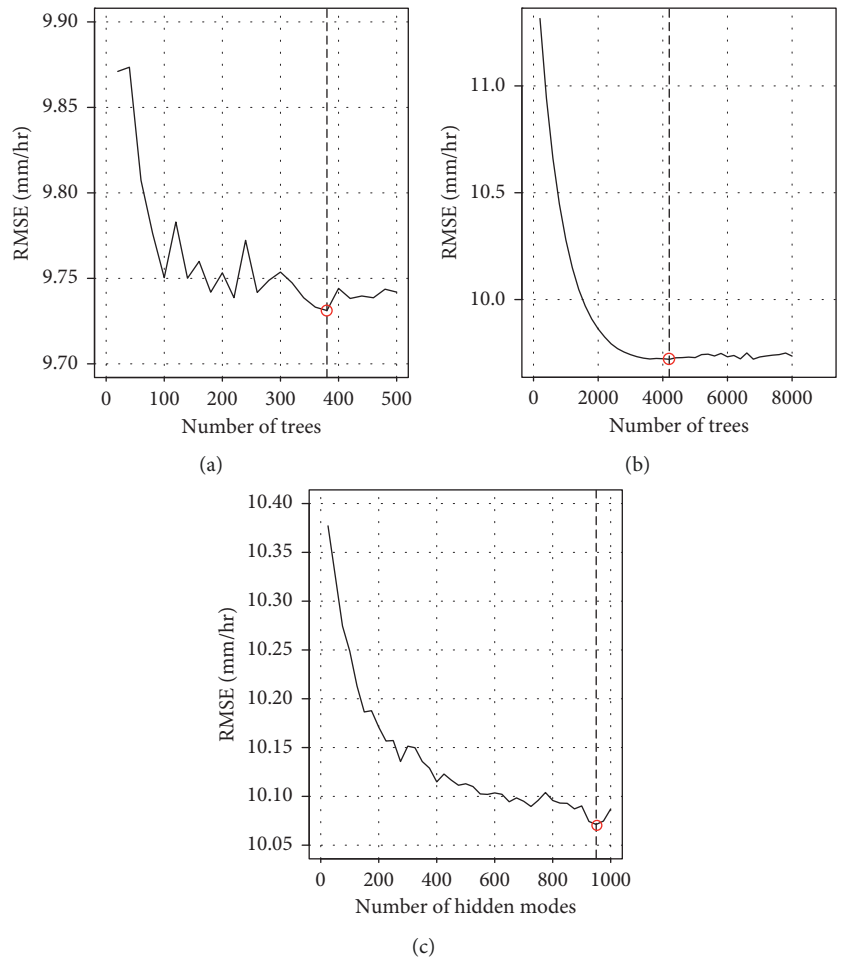

Figure 2: Leave-one-out cross-validation results of hyperparameters for the random forest (the number of trees), stochastic gradient boosted model (the number of trees), and extreme learning machine (the number of hidden nodes) models. The red circles indicate the selected optimal points of the employed hyperparameters based on the root-mean-square error. LOOCV results of (a) RF, (b) GBM, and (c) ELM.

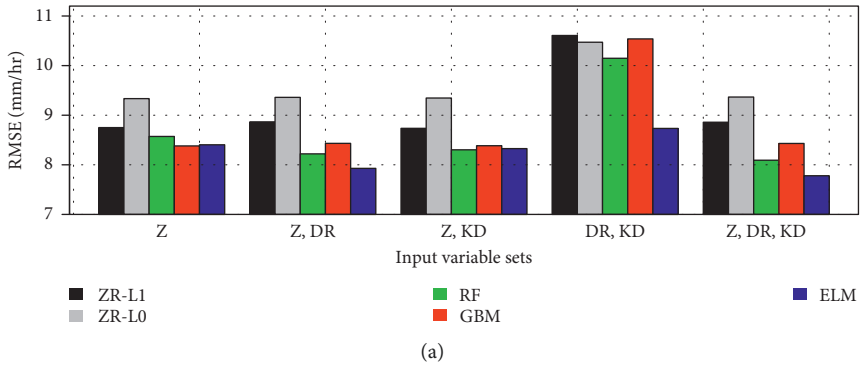

Figure 3: Continued.

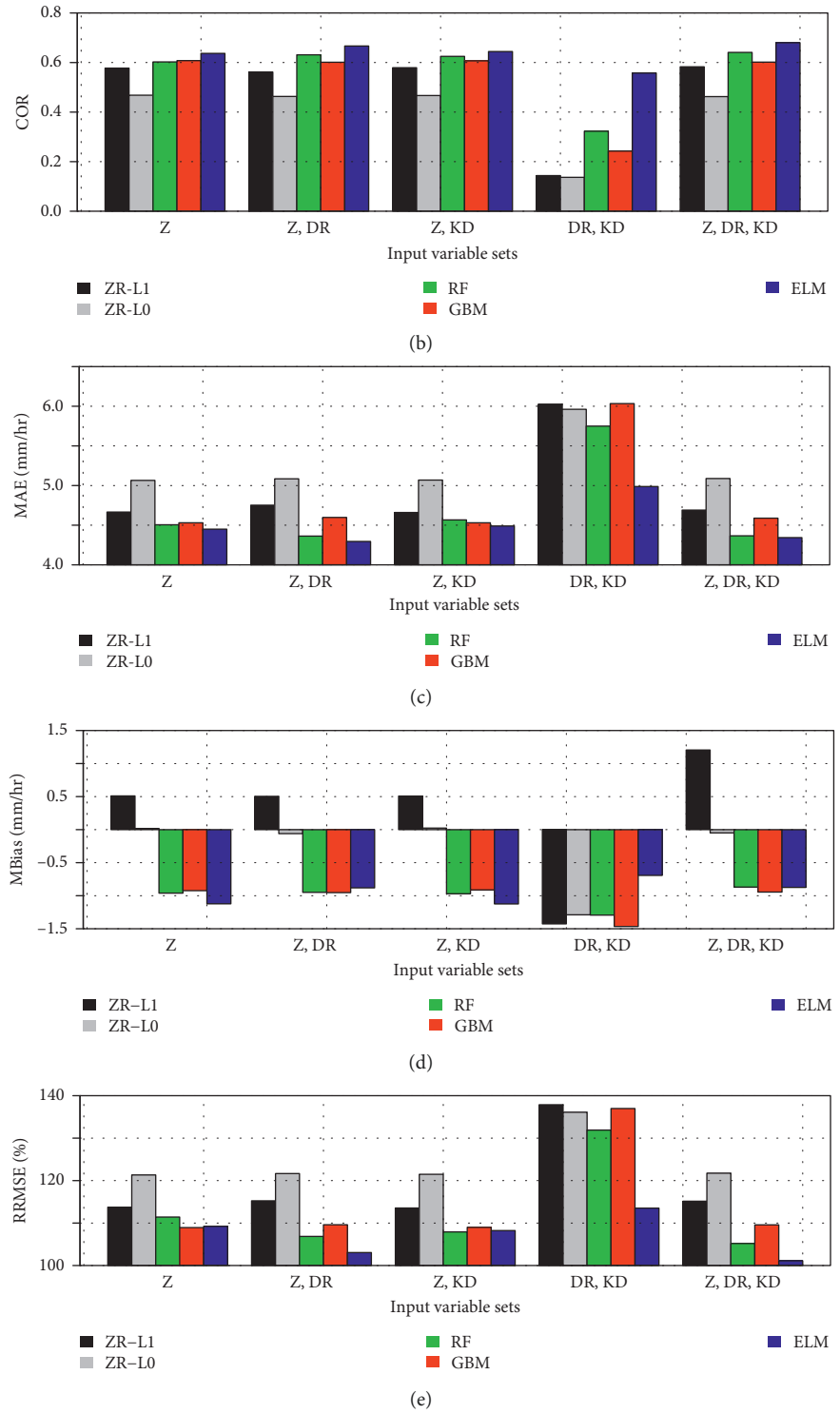

Figure 3: (a) Root-mean-square error (RMSE), (b) correlation (COR), (c) mean absolute error (MAE), (d) mean bias (Mbias), and (e) relative RMSE (RRMSE) of rainfall rate estimations by the tested quantitative precipitation estimation models for all rainfall events studied.

Mbias of ZR-L0-based models is close to zero. To meet the value of Mbias estimate, rainfall rate estimations for small and medium amounts of rainfall rates are overestimated. Mbias of ZR-L1-based models have positive values, and estimations are underestimated for large amount of precipitation, which indicates that a large overestimation occurs for a small amount of precipitation. These overestimations also are observed in Figure 4. The ELM5 leads to the best estimation performance in Figure 4. Circles by the ELM5 are located closer to the diagonal line than other models, while ZR5-L1

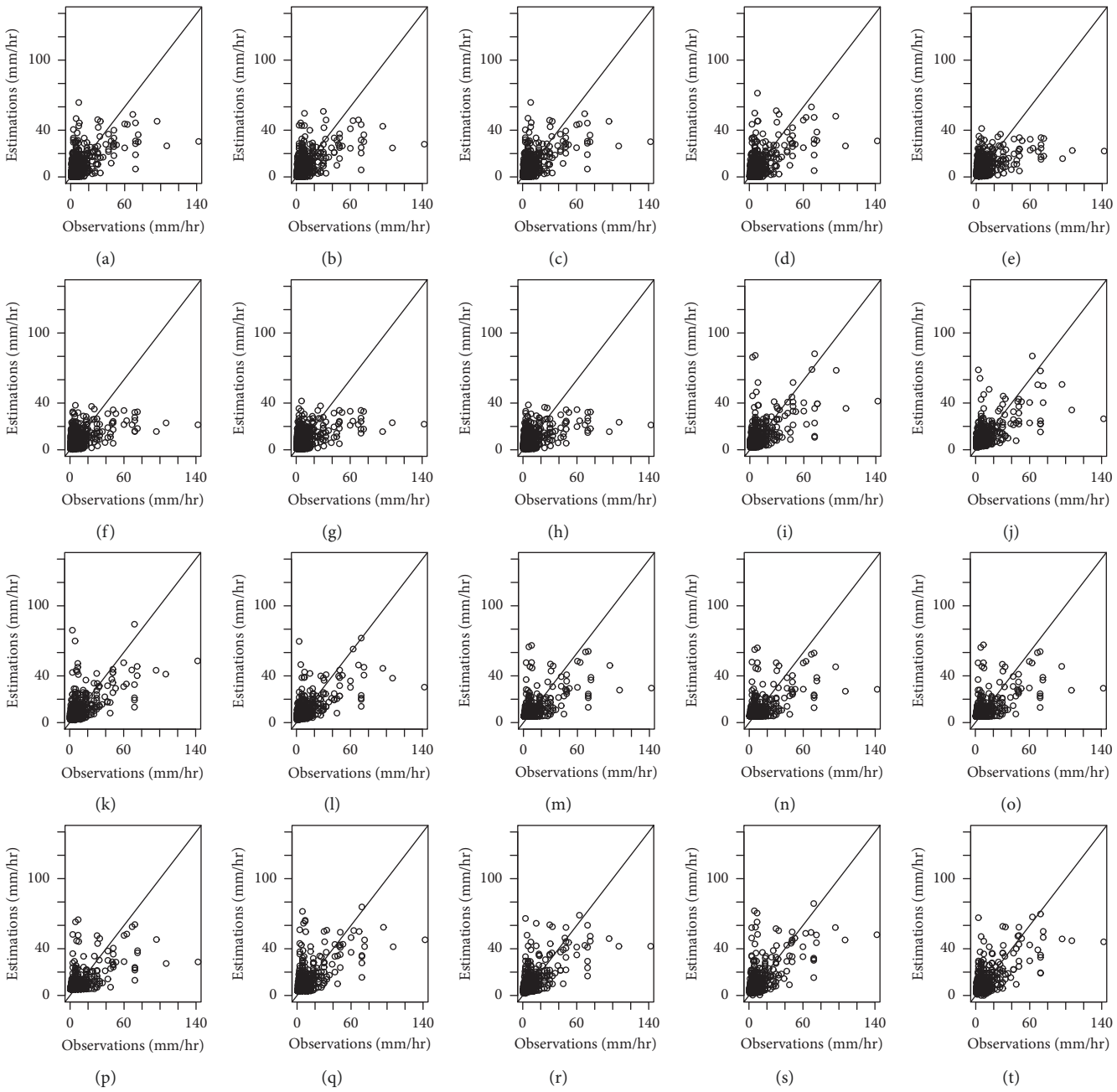

Figure 4: Plots of rainfall rate estimation versus observations for the models tested for all precipitation events studied. (a) ZR1-L1 (RMSE: 8.75, *R*: 0.58). (b) ZR2-L1 (RMSE: 8.87, *R*: 0.56). (c) ZR3-L1 (RMSE: 8.74, *R*: 0.58). (d) ZR5-L1 (RMSE: 8.86, *R*: 0.58). (e) ZR1-L0 (RMSE: 9.33, *R*: 0.47). (f) ZR2-L0 (RMSE: 9.36, *R*: 0.46). (g) ZR3-L0 (RMSE: 9.35, *R*: 0.47). (h) ZR5-L0 (RMSE: 9.37, *R*: 0.46). (i) RF1 (RMSE: 8.56, *R*: 0.6). (j) RF2 (RMSE: 8.26, *R*: 0.63). (k) RF3 (RMSE: 8.36, *R*: 0.62). (l) RF5 (RMSE: 8.18, *R*: 0.63). (m) GBM1 (RMSE: 8.38, *R*: 0.61). (n) GBM2 (RMSE: 8.43, *R*: 0.61). (o) GBM3 (RMSE: 8.38, *R*: 0.24). (p) GBM5 (RMSE: 8.43, *R*: 0.6). (q) ELM1 (RMSE: 8.37, *R*: 0.64). (r) ELM2 (RMSE: 7.99, *R*: 0.66). (s) ELM3 (RMSE: 8.33, *R*: 0.64). (t) ELM5 (RMSE: 7.91, *R*: 0.67).

and GBM5 models seem to provide poor performances. The circle distribution for these models is L-shaped (orthogonal shape) in Figure 4.

5.2. Performances of the Constructed QPE Models for Single-Rainfall Events. Parameters of ZR relationship differ depending on rainfall events characteristics. The performances of the QPE model differ from the rainfall events. Hence, to obtain an accurate QPE, the QPE model should be built for every rainfall event. To investigate the applicability and performance of QPE models, all the tested QPE models are built using data from each precipitation event. RMSEs of the tested QPE models for single-rainfall events are presented in Table 4. ML-based models are selected for the best models based on RMSEs. For events #1 and #2, the ELM5 and ELM3 lead to the lowest RMSEs, respectively. Based on RMSEs, RF2 and RF5 lead to the best performance for events

Table 4: Root-mean-square errors (RMSEs) of rainfall rates estimated by quantitative precipitation estimation models for selected rainfall events.

Event no.	Model	Input variables				
		Z	Z, DR	Z, KD	DR, KD	Z, DR, KD
1	ZR-L1	11.97	11.91	11.98	18.12	11.92
	ZR-L0	14.59	14.62	14.62	17.86	14.66
	RF	11.29	11.30	11.01	15.52	11.20
	GBM	11.05	11.05	11.03	16.33	11.06
	ELM	10.30	10.11	10.21	13.12	*10.06*
2	ZR-L1	5.39	5.26	5.40	7.00	5.30
	ZR-L0	6.13	5.99	6.15	7.51	6.00
	RF	5.25	5.16	5.14	6.34	5.10
	GBM	5.08	5.16	5.07	6.53	5.15
	ELM	4.93	5.17	*4.71*	5.99	5.05
3	ZR-L1	3.33	3.33	3.33	3.57	3.33
	ZR-L0	3.34	3.33	3.34	3.57	3.33
	RF	3.47	*3.17*	3.43	3.54	3.20
	GBM	3.35	3.31	3.35	3.48	3.32
	ELM	3.64	3.91	3.75	3.85	3.99
4	ZR-L1	2.88	2.86	2.88	3.24	2.86
	ZR-L0	3.05	3.02	3.05	3.26	3.02
	RF	2.93	2.97	2.89	3.39	*2.85*
	GBM	2.89	2.86	2.88	3.22	2.86
	ELM	2.91	2.90	2.90	3.08	2.89

Italicized numbers indicate the smallest RMSEs among those calculated during the same rainfall events.

#3 and #4, respectively. Overall, RMSEs of the models using Z and KD data are lower than those in the models using other input variable sets for event #2. For event #3, RMSEs of the models using Z and DR data are lower than the models using other input variable sets. Differences between RMSEs of ML and ZR relationship-based models are very small in event #4. Although RF5 is selected as the best model in the event #4 based on RMSEs, the difference between RMSEs of RF5 and ZR5-L1 is 0.01. Practically, the performances of ZR-L1-based models are the best for event #4 based on RMSEs.

Table 5 presents the correlations of the tested QPE models for single-rainfall events. The correlations of the QPE models for events #1 and #2 are much larger than those for events #3 and #4. While the RMSEs of events #1 and #2 are larger than events #3 and #4, their correlations are higher than those of events #3 and #4. Results indicate that the QPE models lead to good estimation performance for heavy rainfall events. The largest correlation values are observed in ML-based models. ELM2 and ELM3 lead to the largest correlations for events #1 and #2, respectively. RF2 and RF5 provide the largest correlations for events #3 and #4, respectively. Results of MAE are similar to the results of RMSE. Based on Mbias, ZR-L1-, ZR-L0-, GBM-, and ELM-based models lead to the best performance for events #1 to #4, respectively. Detailed MAE and MBias results are not contained in the current manuscript.

Table 6 presents the RRMSEs of the tested QPE models for single-rainfall events. The correlations of the QPE models for events #3 and #4 are smaller than those for events #1 and #2, unlike the results of RMSE and correlation. The smallest

Table 5: Correlations of rain rate estimations by quantitative precipitation estimation models for each selected rainfall event.

Event no.	Model	Input variables				
		Z	Z, DR	Z, KD	DR, KD	Z, DR, KD
1	ZR-L1	0.750	0.753	0.749	0.154	0.752
	ZR-L0	0.603	0.600	0.600	0.201	0.597
	RF	0.785	0.785	0.800	0.513	0.796
	GBM	0.799	0.799	0.800	0.425	0.799
	ELM	0.829	*0.837*	0.830	0.687	0.836
2	ZR-L1	0.721	0.736	0.721	0.455	0.732
	ZR-L0	0.615	0.639	0.611	0.332	0.637
	RF	0.745	0.750	0.749	0.593	0.756
	GBM	0.758	0.749	0.759	0.559	0.750
	ELM	0.793	0.785	*0.805*	0.649	0.788
3	ZR-L1	0.308	0.309	0.308	0.007	0.308
	ZR-L0	0.287	0.292	0.286	−0.056	0.292
	RF	0.228	*0.423*	0.230	0.208	0.404
	GBM	0.298	0.362	0.298	0.207	0.358
	ELM	0.364	0.417	0.355	0.371	0.389
4	ZR-L1	0.418	0.431	0.419	0.047	0.429
	ZR-L0	0.275	0.299	0.279	−0.065	0.298
	RF	0.433	0.418	0.449	0.018	*0.459*
	GBM	0.422	0.435	0.423	−0.004	0.433
	ELM	0.395	0.409	0.400	0.258	0.415

Italicized numbers indicate the largest correlations among those calculated during the same rainfall events.

Table 6: Relative root-mean-square errors (RRMSEs) of rain rate estimations by quantitative precipitation estimation models for each selected rainfall event.

Event no.	Model	Input variables				
		Z	Z, DR	Z, KD	DR, KD	Z, DR, KD
1	ZR-L1	85.5	85.1	85.6	129.5	85.2
	ZR-L0	104.3	104.5	104.5	127.6	104.7
	RF	81.1	80.2	78.5	112.2	79.8
	GBM	78.8	79.0	79.1	116.8	79.2
	ELM	73.5	72.1	72.9	93.5	*71.8*
2	ZR-L1	77.8	75.9	77.9	101.0	76.4
	ZR-L0	88.5	86.4	88.8	108.4	86.6
	RF	76.4	74.0	74.5	91.6	73.6
	GBM	73.1	74.6	73.2	94.1	74.5
	ELM	71.2	75.0	*68.1*	86.5	72.6
3	ZR-L1	63.6	63.7	63.6	68.3	63.7
	ZR-L0	63.8	63.8	63.8	68.3	63.8
	RF	66.3	*60.6*	65.6	67.4	60.8
	GBM	64.1	63.4	64.1	66.6	63.4
	ELM	69.6	75.8	71.9	73.3	76.9
4	ZR-L1	63.5	63.0	63.5	71.4	63.0
	ZR-L0	67.4	66.7	67.3	72.0	66.7
	RF	64.7	65.3	64.1	75.0	*63.0*
	GBM	63.6	63.0	63.5	71.4	63.0
	ELM	64.2	63.8	63.8	68.0	63.4

Italicized numbers indicate the smallest RRMSEs among those calculated during the same rainfall events.

RRMSEs for events #1 and #2 (heavy rainfall events) are 71.8% and 68.1%, respectively. For events #3 and #4 (light rainfall events), the smallest RRMSEs are 60.6% and 63.0%,

respectively. Overall difference between the smallest RRMSEs of heavy and light rainfall events are approximately 10%. The difference between the smallest RMSE of heavy and light rainfall events is approximately 4.3 mm/hr. Because 4.3 mm/hr is larger than the smallest RMSE of event #3, the RRMSE difference is relatively smaller than RMSE difference. The result indicates that the QPE models provide similar performances for heavy and light rainfall events based on RRMSE measures.

To evaluate the tested QPE models for single-rainfall events, rainfall rate estimation versus observation plots for event #1 and #4 is presented in Figures 5 and 6, respectively. The tested models excluding ZR-L0-based models lead to good estimation performances for event #1. Of the tested models, ELM5 gives the best estimation performance. Some circles are aligned to approximately 80 mm/h based on observations in Figure 5. This is a recurrent issue in QPE of rain radar data. When the observed rainfall rates are the same but the observed parameters of radar data are different, this phenomenon occurs. This result indicates that all tested QPE models cannot solve this issue. For event #4 in Figure 6, all the tested QPE models lead to poor performances of rainfall rate estimation. In the observed small magnitude of rainfall rates, the QPE models tend to overestimate rainfall rates. On the other hand, the QPE models provide an underestimation for the observed large magnitude of rainfall rates. Five lines are observed at approximately 3 mm/h, 6 mm/h, 9 mm/h, 12 mm/h, and 15 mm/h based on observations in all the subfigures presented in Figure 6. The observed rainfall depths for these small rainfall rates are 0.5 mm, 1 mm, 1.5 mm, 2 mm, and 2.5 mm, and their duration is 10 minutes. Because event #4 is light, a large number of small rainfall rates are observed. The phenomena wherein parameters of radar are different for the same amount of observed rainfall rate occur frequently. Due to this phenomenon, the tested QPE models show poor performances in event #4.

Radar rainfall rate fields for events #1 and #4 are illustrated to investigate the difference between ZR relation- and ML-based models in Figures 7 and 8. Figure 7 presents radar rainfall rate fields of event #1 at 20 : 10, 28th August 2018. The range of rainfall rates is from 0 to 100 mm/h for event #1. The ML-based QPE models provide larger magnitudes of rainfall rates for very small magnitudes of rainfall rates based on estimates by ZR2-L1. For heavy magnitudes of rainfall rates, estimates of all QPE models are similar. The GBM leads to the largest magnitude of rainfall rate estimation in Figure 7. Rainfall rate estimates on ground gauge stations for the ZR2-L1 are larger than those of ML-based models. Due to the high magnitude of rainfall rate at these points, the ZR2-L1 overestimates rainfall rate in Figure 3. In areas where there are no ground gauge stations, the ZR2-L1 estimates smaller rainfall rates than other models. Figure 8 presents radar rainfall rate fields of event #4 at 12 : 40, 8th November 2018. Rainfall rates range from 0 to 15 mm/h for event #4. Overall results of rainfall rate estimates by the tested models are similar to the results shown in Figure 7. The ELM leads to the largest magnitude of rainfall rate estimation.

6. Discussion

The comparison results of the ZR relationship- and ML-based models show that the application of ML algorithms can lead to an improvement in the QPE of radar data in the tested rainfall events. This result supports the notion that the ML algorithm could be used in the development of QPE models of radar data in South Korea. Increasing the number of variables for the input variables of the ZR relationship-based models results in very small improvements. In some events, this increment does not improve performances of QPE models. It can be inferred that Z is the most critical variable for the ZR relationship-based model. Additionally, the application of other variables is often an inefficient way to build the ZR relationship-based model.

The performances of the ML-based models improve when Z and additional variables such as DR and KD are applied as input variables. In particular, a combination of Z and DR for input variables of the ML-based models leads to a good QPE performance. Studies have reported that this combination leads to the best performance among combinations of Z, DR, and KD for the ZR relationship-based model [46, 47]. In many cases, application of DR, except for combinations of DR and KD, lead to a large improvement in the QPE using the ML-based model, unlike the ZR relationship-based model. The results imply that the ML-based models could consider other variables in QPE. Because the ML-based model can extract a large amount of information from the input variables and use this information in QPE of rain radar data, performances of the ML-based models may be better than those of ZR relationship-based models. Based on results of RMSE for individual events, the RF model with three variables provided the smallest RMSEs in events #2 and #4. Otherwise, RMSEs of other RF models were smaller than those of the RF model with three variables. In addition, there is a very small difference (0.06 mm/hr) between RSMEs of RF models with three variables and with Z and KD. The RF model with Z and KD is the best model when taking into consideration of parsimony for event #2. Hence, the RF model with three variables can be considered for suboptimal in events #1, #2, and #3.

Computation times to build QPE models differ depending on the ML algorithms employed. RF has the shortest computation time, and its computation time with the data sets of all events is approximately 1 minute. The computation times of the GBM and ELM are approximately 7 minutes and 3 minutes, respectively. As the measuring interval of rainfall data is 10 minutes, the computation time should be shorter than 5 minutes. The RF- and ELM-based models proposed in the current study can be applied for QPE, but the GBM has to be modified before application.

Based on the results of this study, a comparison of the performances of the employed algorithms can be carried out. The ELM leads to the best performance for the case that includes all the events. For single events, the best algorithms are different. The ELM provides a good performance for heavy rainfall events, while the RF is considered a good algorithm for light rainfall events. The difference in performance between the RF and ELM is small in the light

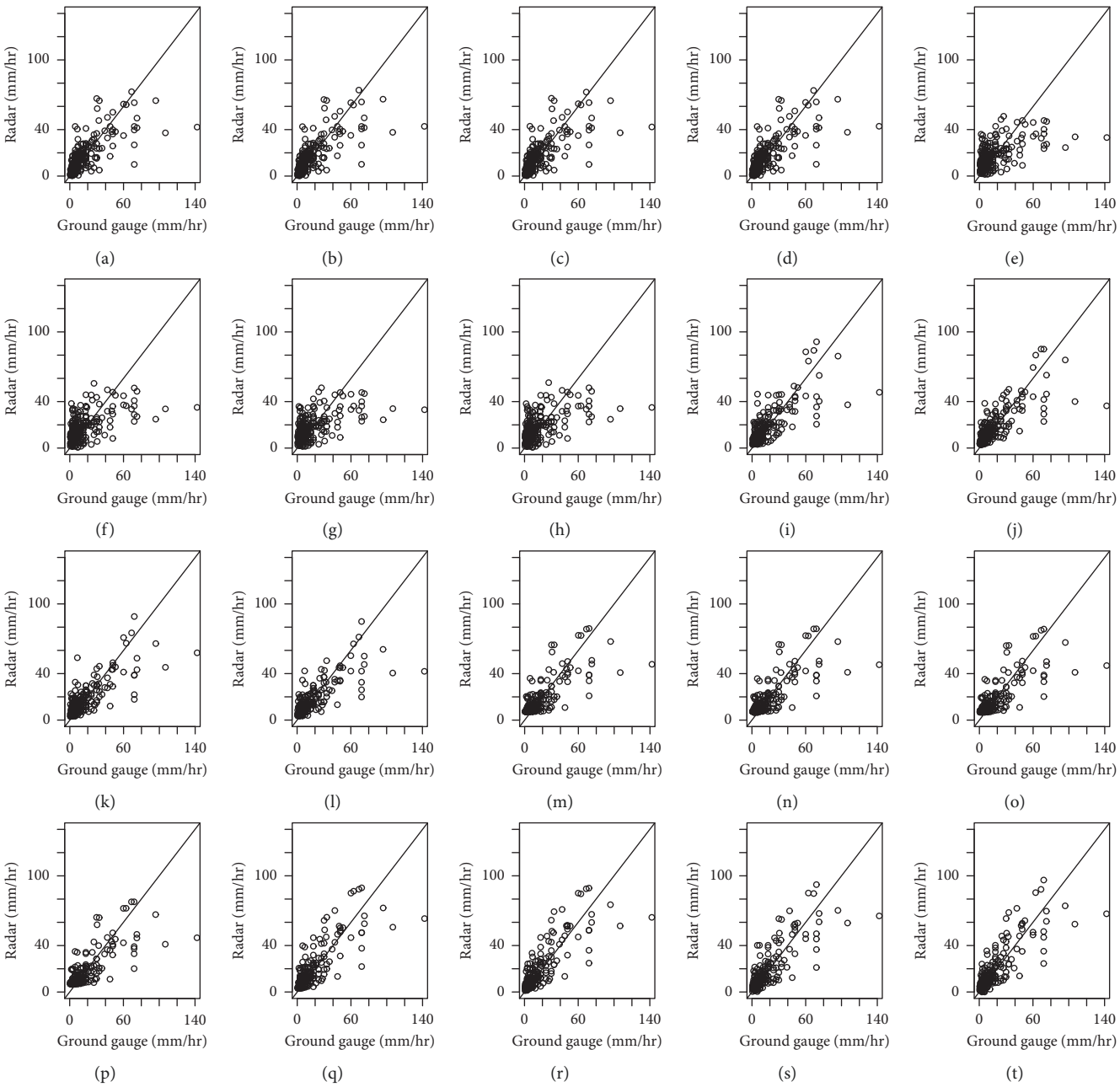

Figure 5: Plots of rainfall rate estimations versus observations of tested models for precipitation event #1. (a) ZR1-L1 (RMSE: 11.97, *R*: 0.75). (b) ZR2-L1 (RMSE: 11.91, *R*: 0.75). (c) ZR3-L1 (RMSE: 11.98, *R*: 0.75). (d) ZR5-L1 (RMSE: 11.92, *R*: 0.75). (e) ZR1-L0 (RMSE: 14.59, *R*: 0.6). (f) ZR2-L0 (RMSE: 14.62, *R*: 0.6). (g) ZR3-L0 (RMSE: 14.62, *R*: 0.6). (h) ZR5-L0 (RMSE: 14.66, *R*: 0.6). (i) RF1 (RMSE: 11.29, *R*: 0.79). (j) RF2 (RMSE: 11.3, *R*: 0.79). (k) RF3 (RMSE: 11.01, *R*: 0.8). (l) RF5 (RMSE: 11.2, *R*: 0.8). (m) GBM1 (RMSE: 11.05, *R*: 0.8). (n) GBM2 (RMSE: 11.05, *R*: 0.8). (o) GBM3 (RMSE: 11.03, *R*: 0.42). (p) GBM5 (RMSE: 11.06, *R*: 0.8). (q) ELM1 (RMSE: 10.3, *R*: 0.83). (r) ELM2 (RMSE: 10.11, *R*: 0.84). (s) ELM3 (RMSE: 10.21, *R*: 0.83). (t) ELM5 (RMSE: 10.06, *R*: 0.84).

rainfall events. Hence, the best ML algorithm for case studies performed in the current study is the ELM. Each ML algorithm tested in this study uses popular setting. The comparison results of the ML algorithm for QPE models can be altered by adopted setting and used data. For example, in this study, the CART is used for the decision tree in the RF. Other decision tree models can be used in the RF such as inference dichotomiser 3 and chi-squared automatic iteration detection. RF with other decision tree models can outperform to the ELM model for QPE in South Korea. Thus, the results in the current study should be restricted to these data sets and the ML algorithms with adopted setting.

Variation of neural network models like artificial neural, recurrent neural, and deep neural networks may have a high applicability building QPE models of radar data in South Korea, because the ELM is developed based on a neural network. Additionally, enhancing the precision of rainfall gauges may lead to improvements in the performance of

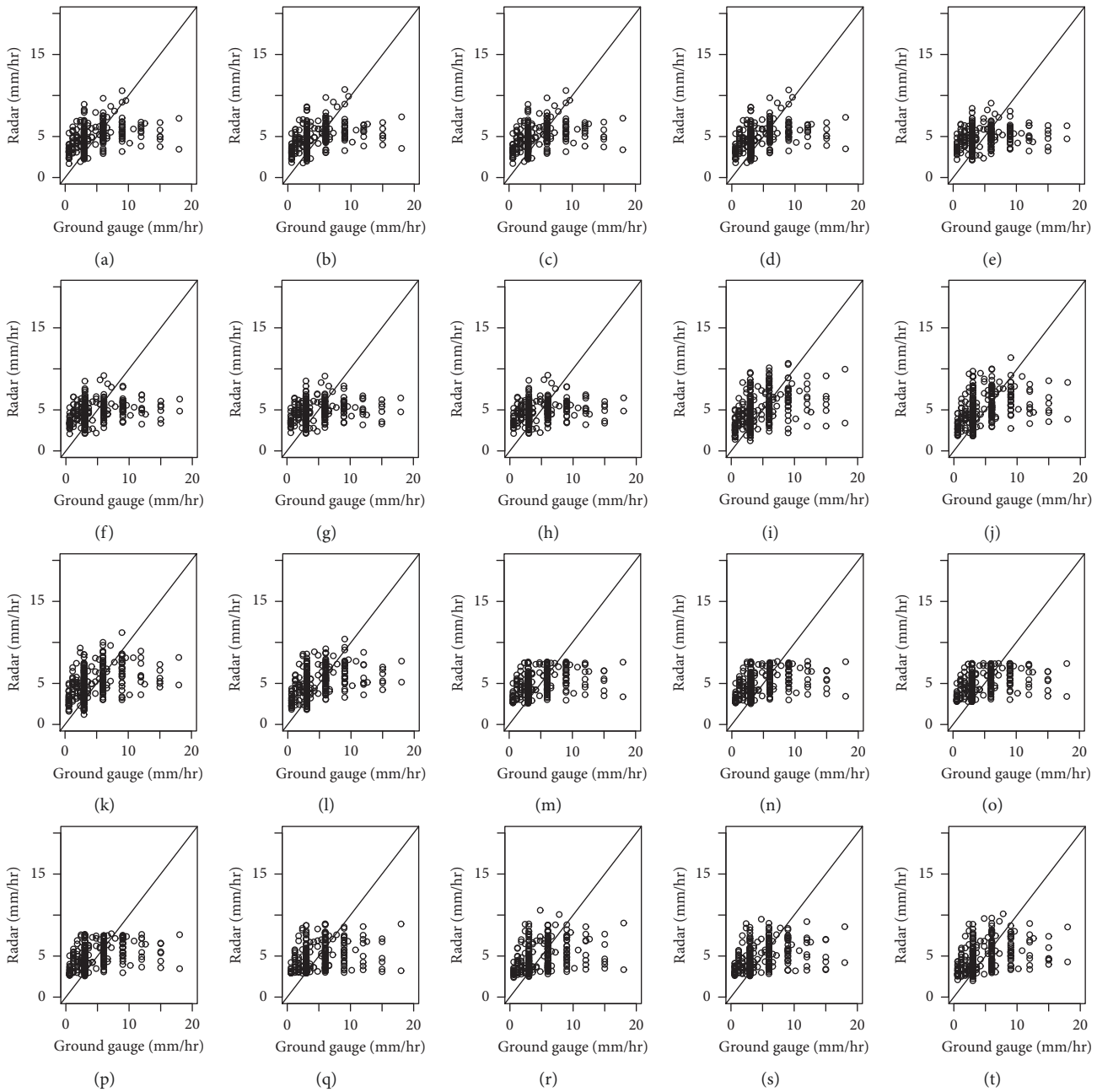

Figure 6: Plots of rainfall rate estimations versus observations of tested models for precipitation event #4. (a) ZR1-L1 (RMSE: 2.88, *R*: 0.42). (b) ZR2-L1 (RMSE: 2.86, *R*: 0.43). (c) ZR3-L1 (RMSE: 2.88, *R*: 0.42). (d) ZR5-L1 (RMSE: 2.86, *R*: 0.43). (e) ZR1-L0 (RMSE: 3.05, *R*: 0.27). (f) ZR2-L0 (RMSE: 3.02, *R*: 0.3). (g) ZR3-L0 (RMSE: 3.05, *R*: 0.28). (h) ZR5-L0 (RMSE: 3.02, *R*: 0.3). (i) RF1 (RMSE: 2.93, *R*: 0.43). (j) RF2 (RMSE: 2.97, *R*: 0.42). (k) RF3 (RMSE: 2.89, *R*: 0.45). (l) RF5 (RMSE: 2.85, *R*: 0.46). (m) GBM1 (RMSE: 2.89, *R*: 0.42). (n) GBM2 (RMSE: 2.86, *R*: 0.42). (o) GBM3 (RMSE: 2.88, *R*: 0.42). (p) GBM5 (RMSE: 2.86, *R*: 0.43). (q) ELM1 (RMSE: 2.91, *R*: 0.39). (r) ELM2 (RMSE: 2.9, *R*: 0.41). (s) ELM3 (RMSE: 2.9, *R*: 0.4). (t) ELM5 (RMSE: 2.89, *R*: 0.41).

QPE models for light rainfall events. Precision for some of the employed rainfall gauges is 3 mm/hr. Although this precision is good enough to measure rainfall rates for long duration or heavy rainfall events, it should be higher for estimating rainfall rates of light rainfall events. For example, when parameters of radar data for two points are different but their observed rainfall rates are the same, the QPE model has to fail estimations of rainfall rates at two points. If the precision of the rainfall gauge increases, the observed rainfall rates may be different and could result in a more accurate constructed QPE model. As shown in Figure 6, three lines can be observed in all the subfigures. Values of ground gauge for first, second, and third lines are 3, 6, and 9 mm/hr. These three lines indicate that the observed rainfall rates at ground gauge station are the same when the parameters of radar data are different. If precisions of these gauge stations become

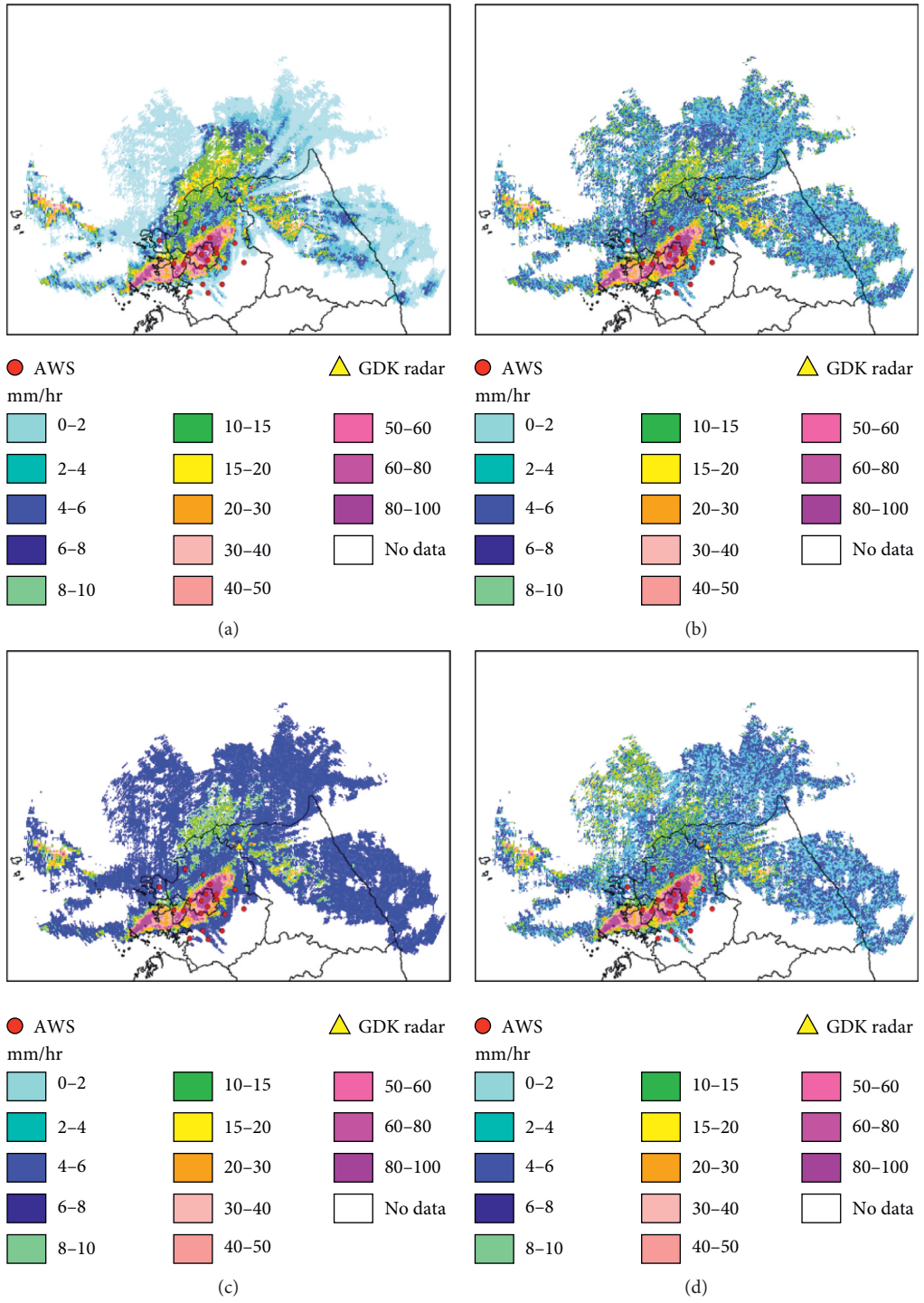

Figure 7: Radar rainfall rate fields for four selected quantitative precipitation estimation models (ZR2–L1, RF3, GBM3, and ELM5) for event #1 (August 28, 2018; 20 : 10). (a) ZR. (b) RF. (c) GB. (d) EL.

better, these lines may be disappeared and the data points in the lines are dissipated. This dispersion of data points, caused by the high precision of measuring instrument, may lead to improvement of the performance of QPE models for light rainfall event.

The tested QPE models lead to good performances for heavy rainfall events but not for light rainfall events. This characteristic of QPE with rain radar data is also observed in this study. ML-based QPE models outperform ZR relationship-based models for events #3 and #4, albeit insignificantly. As mentioned above, the ML-based QPE models show good performances by efficiently extracting information from given radar data. If additional variables can be applied in QPE models, the performance of the QPE

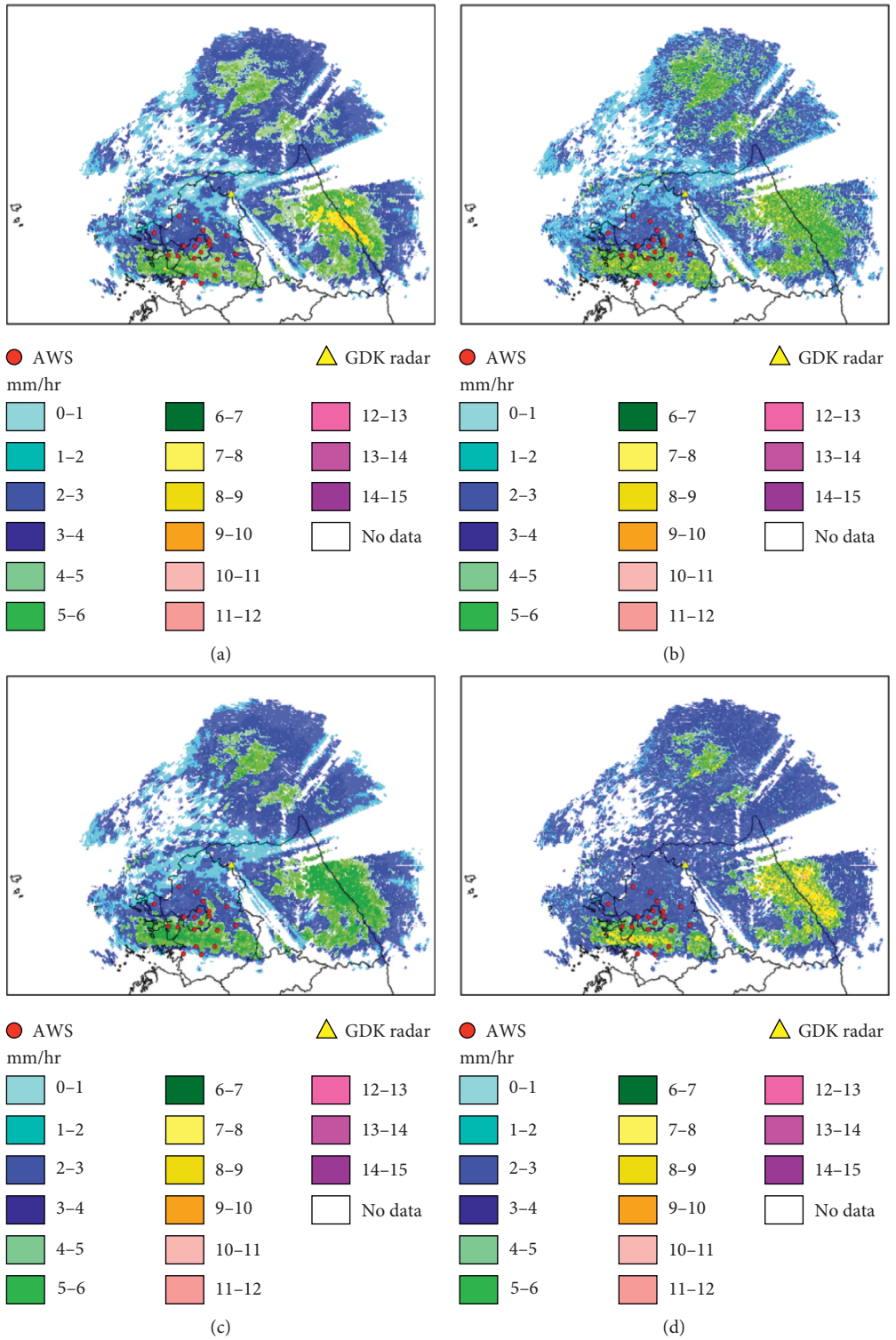

Figure 8: Radar rainfall rate fields of four selected quantitative precipitation estimation models (ZR2–L1, RF5, GBM2, and ELM5) for event #4 (November 8, 2018; 12 : 40). (a) ZR2-L1. (b) RF5. (c) GBM2. (d) ELM5.

model may improve, particularly for light rainfall events. Thus, various sets of input variables that are frequently used in conventional QPE algorithms should be tested for ML-based QPE models to improve the performance of QPE models for light rainfall events.

7. Conclusions

The applicability of three ML algorithms in QPE models is investigated using case studies of polarization radar data of four rainfall events from Gwangdeoksan radar station, Gyeonggi-do,

South Korea. Various combinations of input variable sets are also tested for QPE models. Conventional ZR relation-based models are also constructed and compared to ML-based models. In the current study, we reach the following conclusions:

(1) ML algorithms can be applied to build a QPE model of polarization radar data. Overall, the ML-based QPE models outperform or are equal to ZR relationship-based models. ML algorithms can extract information from radar data more efficiently than the ZR relationship, which leads to an improvement in QPE of the radar data.

(2) Application of the ML algorithms for QPE models improves rainfall rate estimations for heavy events in South Korea by far. The performances of the ML-based QPE model are significantly improved based on performances of ZR relationship-based models for heavy rainfall events. This improvement will be helpful in modeling floods and forecasting flash floods.

(3) ELM algorithm may be the best ML algorithm among the tested ML algorithm with the adopted setting for QPE models of radar data in South Korea. Overall, the ELM outperforms other tested QPE models in QPE of radar data employed in the current study. Based on evaluation results of single-rainfall events, the ELM also leads to the best performance in two heavy rainfall events. Although the ELM is not the best QPE model for the two light rainfall events, the performances of QPE models using ELM are comparable to other QPE models.

In the current study, four rainfall events in 2018 were employed to evaluate the applicability of ML algorithms for the QPE model of polarization radar data as the radar instrument in Gwangdeoksan radar was updated. Future rainfall events should be included in data sets to further investigate the applicability and characteristics of ML algorithms in the QPE of polarization radar data in South Korea. In addition, the applicability of ML algorithms for QPF should be examined. Because ML algorithms show high applicability in QPE, they make good candidates for modeling functions between radar data and QPF.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the Korea Meteorological Administration Research and Development Program "Development of Application Technology on Atmospheric Research Aircraft" under Grant (1365003069).

References

[1] B. Martens, P. Cabus, I. De Jongh, and N. E. C. Verhoest, "Merging weather radar observations with ground-based measurements of rainfall using an adaptive multiquadric surface fitting algorithm," *Journal of Hydrology*, vol. 500, pp. 84–96, 2013.

[2] S. Yang and S. W. Nesbitt, "Statistical properties of precipitation as observed by the TRMM precipitation radar," *Geophysical Research Letters*, vol. 41, no. 15, pp. 5636–5643, 2014.

[3] W. F. Krajewski and J. A. Smith, "Radar hydrology: rainfall estimation," *Advances in Water Resources*, vol. 25, no. 8–12, pp. 1387–1394, 2002.

[4] J. Kim, J. Lee, D. Kim, and B. Kang, "The role of rainfall spatial variability in estimating areal reduction factors," *Journal of Hydrology*, vol. 568, pp. 416–426, 2019.

[5] W. Cho, J. Lee, J. Park, and D. Kim, "Radar polygon method: an areal rainfall estimation based on radar rainfall imageries," *Stochastic Environmental Research and Risk Assessment*, vol. 31, no. 1, pp. 275–289, 2017.

[6] T. Park, T. Lee, D. Ko, J. Shin, and D. Lee, "Assessing spatially dependent errors in radar rainfall estimates for rainfall-runoff simulation," *Stochastic Environmental Research and Risk Assessment*, vol. 31, no. 7, pp. 1823–1838, 2016.

[7] J. Simpson, R. F. Adler, and G. R. North, "A proposed tropical rainfall measuring mission (TRMM) satellite," *Bulletin of the American Meteorological Society*, vol. 69, no. 3, pp. 278–295, 1988.

[8] R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie, "CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution," *Journal of Hydrometeorology*, vol. 5, no. 3, pp. 487–503, 2004.

[9] Y. Wang and V. Chandrasekar, "Quantitative precipitation estimation in the CASA X-band dual-polarization radar network," *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 10, pp. 1665–1676, 2010.

[10] G. Delrieu, I. Braud, A. Berne et al., "Weather radar and hydrology," *Advances in Water Resources*, vol. 32, no. 7, pp. 969–974, 2009.

[11] T. Lee, J. Shin, T. Park, and D. Lee, "Basin rotation method for analyzing the directional influence of moving storms on basin response," *Stochastic Environmental Research and Risk Assessment*, vol. 29, no. 1, pp. 251–263, 2015.

[12] N. Balakrishnan, D. S. Zrnić, J. Goldhirsh, and J. Rowland, "Comparison of simulated rain rates from disdrometer data employing polarimetric radar algorithms," *Journal of Atmospheric and Oceanic Technology*, vol. 6, no. 3, pp. 476–486, 1989.

[13] A. V. Ryzhkov and D. S. Zrnić, "Comparison of dual-polarization radar estimators of rain," *Journal of Atmospheric and Oceanic Technology*, vol. 12, no. 2, pp. 249–256, 1995.

[14] S. Verrier, L. Barthès, and C. Mallet, "Theoretical and empirical scale dependency of Z-R relationships: evidence, impacts, and correction," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 14, pp. 7435–7449, 2013.

[15] J. A. Smith and W. F. Krajewski, "Estimation of the mean field bias of radar rainfall estimates," *Journal of Applied Meteorology*, vol. 30, no. 4, pp. 397–412, 1991.

[16] M.-K. Suk, K.-H. Chang, J.-W. Cha, and K.-E. Kim, "Operational real-time adjustment of radar rainfall estimation over the South Korea region," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 91, no. 4, pp. 545–554, 2013.

[17] C. Yoo, C. Park, J. Yoon, and J. Kim, "Interpretation of mean-field bias correction of radar rain rate using the concept of linear regression," *Hydrological Processes*, vol. 28, no. 19, pp. 5081–5092, 2014.

[18] L. Yang, Y. Yang, P. Liu, and L. Wang, "Radar-Derived quantitative precipitation estimation based on precipitation classification," *Advances in Meteorology*, vol. 2016, Article ID 2457489, 16 pages, 2016.

[19] D.-J. Seo and J. A. Smith, "Rainfall estimation using raingages and radar—A Bayesian approach: 1. Derivation of estimators," *Stochastic Hydrology and Hydraulics*, vol. 5, no. 1, pp. 17–29, 1991.

[20] D.-J. Seo and J. A. Smith, "Rainfall estimation using raingages and radar—A Bayesian approach: 2. An application," *Stochastic Hydrology and Hydraulics*, vol. 5, no. 1, pp. 31–44, 1991.

[21] Q. Dai, M. A. Rico-Ramirez, D. Han, T. Islam, and S. Liguori, "Probabilistic radar rainfall nowcasts using empirical and theoretical uncertainty models," *Hydrological Processes*, vol. 29, no. 1, pp. 66–79, 2013.

[22] A. Mosavi, P. Ozturk, and K.-w. Chau, "Flood prediction using machine learning models: literature review," *Water*, vol. 10, no. 11, p. 1536, 2018.

[23] P. Ganguli and M. J. Reddy, "Ensemble prediction of regional droughts using climate inputs and SVM-copula approach," *Hydrological Processes*, vol. 28, no. 19, pp. 4989–5009, 2013.

[24] D. J. Gagne II, A. McGovern, and M. Xue, "Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts," *Weather and Forecasting*, vol. 29, no. 4, pp. 1024–1043, 2014.

[25] Y. Tao, X. Gao, A. Ihler, S. Sorooshian, and K. Hsu, "Precipitation identification with bispectral satellite information using deep learning approaches," *Journal of Hydrometeorology*, vol. 18, no. 5, pp. 1271–1283, 2017.

[26] X. Lu, Y. Ju, L. Wu, J. Fan, F. Zhang, and Z. Li, "Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models," *Journal of Hydrology*, vol. 566, pp. 668–684, 2018.

[27] B. T. Nolan, C. T. Green, P. F. Juckem, L. Liao, and J. E. Reddy, "Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA," *Journal of Hydrology*, vol. 559, pp. 428–441, 2018.

[28] H. I. Erdal and O. Karakurt, "Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms," *Journal of Hydrology*, vol. 477, pp. 119–128, 2013.

[29] Y.-M. Chiang, F.-J. Chang, B. J.-D. Jou, and P.-F. Lin, "Dynamic ANN for precipitation estimation and forecasting from radar observations," *Journal of Hydrology*, vol. 334, no. 1-2, pp. 250–261, 2007.

[30] P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, "Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting," *Journal of Hydrology*, vol. 552, pp. 92–104, 2017.

[31] M. R. Kumjian, "Principles and applications of dual-polarization weather radar. Part I: description of the polarimetric radar variables," *Journal of Operational Meteorology*, vol. 1, no. 19, pp. 226–242, 2013.

[32] A. V. Ryzhkov, S. E. Giangrande, and T. J. Schuur, "Rainfall estimation with a polarimetric prototype of WSR-88d," *Journal of Applied Meteorology*, vol. 44, no. 4, pp. 502–515, 2005.

[33] R. Cifelli, V. Chandrasekar, S. Lim, P. C. Kennedy, Y. Wang, and S. A. Rutledge, "A new dual-polarization radar rainfall algorithm: application in Colorado precipitation events," *Journal of Atmospheric and Oceanic Technology*, vol. 28, no. 3, pp. 352–364, 2011.

[34] T. Yang, A. A. Asanjan, E. Welles, X. Gao, S. Sorooshian, and X. Liu, "Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information," *Water Resources Research*, vol. 53, no. 4, pp. 2786–2812, 2017.

[35] J. Fan, W. Yue, L. Wu et al., "Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China," *Agricultural and Forest Meteorology*, vol. 263, pp. 225–241, 2018.

[36] B. Pang, J. Yue, G. Zhao, and Z. Xu, "Statistical downscaling of temperature with the random forest model," *Advances in Meteorology*, vol. 2017, Article ID 7265178, 11 pages, 2017.

[37] C. Choi, J. Kim, J. Kim, D. Kim, Y. Bae, and H. S. Kim, "Development of heavy rain damage prediction model using machine learning based on big data," *Advances in Meteorology*, vol. 2018, Article ID 5024930, 11 pages, 2018.

[38] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[39] M. N. Wright and A. Ziegler, "Ranger: a fast implementation of random forests for high dimensional data in C++ and R," *Survival Analysis*, vol. 77, no. 1, p. 17, 2017.

[40] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[41] A. Torres-Barrán, Á. Alonso, and J. R. Dorronsoro, "Regression tree ensembles for wind energy and solar radiation prediction," *Neurocomputing*, vol. 326-327, pp. 151–160, 2019.

[42] S. A. Naghibi, H. R. Pourghasemi, and B. Dixon, "GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran," *Environmental Monitoring and Assessment*, vol. 188, no. 1, p. 44, 2016.

[43] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[44] M. Peña and H. van den Dool, "Consolidation of multimodel forecasts by ridge regression: application to pacific sea surface temperature," *Journal of Climate*, vol. 21, no. 24, pp. 6521–6538, 2008.

[45] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2012.

[46] M. J. Simpson and N. I. Fox, "Dual-polarized quantitative precipitation estimation as a function of range," *Hydrology and Earth System Sciences*, vol. 22, no. 6, pp. 3375–3389, 2018.

[47] B.-C. Seo, B. Dolan, W. F. Krajewski, S. A. Rutledge, and W. Petersen, "Comparison of single- and dual-polarization-based rainfall estimates using NEXRAD data for the NASA Iowa flood studies project," *Journal of Hydrometeorology*, vol. 16, no. 4, pp. 1658–1675, 2015.