Extracting 3D Radar Features to Improve Quantitative Precipitation Estimation in Complex Terrain Based on Deep Learning Neural Networks

YUNG-YUN CHENG,[a] CHIA-TUNG CHANG,[a] BUO-FU CHEN,[a] HUNG-CHI KUO,[a,b] AND CHENG-SHANG LEE[a,b]

[a] *Center for Weather Climate and Disaster Research, National Taiwan University, Taipei, Taiwan*
[b] *Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan*

(Manuscript received 14 February 2022, in final form 25 October 2022)

ABSTRACT: This paper proposes a new quantitative precipitation estimation (QPE) technique to provide accurate rainfall estimates in complex terrain, where conventional QPE has limitations. The operational radar QPE in Taiwan is mainly based on the simplified relationship between radar reflectivity Z and rain rate R [$R(Z)$ relation] and only utilizes the single-point lowest available echo to estimate rain rates, leading to low accuracy in complex terrain. Here, we conduct QPE using deep learning that extracts features from 3D radar reflectivities to address the above issues. Convolutional neural networks (CNN) are used to analyze contoured frequency by altitude diagrams (CFADs) to generate the QPE. CNN models are trained on existing rain gauges in northern and eastern Taiwan with the 3-yr data during 2015–17 and validated and tested using 2018 data. The weights of heavy rains ($\geq$10 mm h^{-1}) are increased in the model loss calculation to handle the unbalanced rainfall data and improve accuracy. Results show that the CNN outperforms the $R(Z)$ relation based on the 2018 rain gauge data. Furthermore, this research proposes methods to conduct 2D gridded QPE at every pixel by blending estimates from various trained CNN models. Verification based on independent rain gauges shows that the CNN QPE solves the underestimation of the $R(Z)$ relation in mountainous areas. Case studies are presented to visualize the results, showing that the CNN QPE generates better small-scale rainfall features and more accurate precipitation information. This deep learning QPE technique may be helpful for the disaster prevention of small-scale flash floods in complex terrain.

KEYWORDS: Radars/Radar observations; Weather radar signal processing; Deep learning; Neural networks; Precipitation

1. Introduction

Taiwan suffers from frequent flash floods due to complex and steep terrain interplaying with short-term and small-scale torrential rainfall. Accurate precipitation information is urged to prevent and monitor such meteorological disasters. However, the rain gauges in Taiwan are mainly located on plains, while few are in the mountains, and this uneven rain gauge distribution limits the accuracy of rainfall monitoring (Wu et al. 2016; Chung and Yao 2020).

Central Weather Bureau (CWB) of Taiwan started to construct an S-band weather radar network in 2002 and develop a weather monitoring system–quantitative precipitation estimation and segregation using multiple sensors (QPESUMS) to describe the spatial distribution of rainfall in areas without rain gauges (Gourley et al. 2002; Chung and Yao 2020). The operational quantitative precipitation estimation (QPE) method used in QPESUMS is mainly based on Xin et al. (1997), by which the rain rate is calculated from $Z = 32.5R^{1.65}$, where Z is radar reflectivity (mm^6 m^{-3}), and R is rain rate (mm h^{-1}). This $R(Z)$ formula can be used based on the relationship between rainfall observation and high temporal- and spatial-resolution radar observation to calculate two-dimensional precipitation. Namely, using the $R(Z)$ relation estimates precipitation at each grid point, then combining them to form 2D rain maps. Nevertheless, the disadvantage of the $R(Z)$ relation is that this method only uses the lowest available echo to the surface to estimate the precipitation. Although it is convenient to operate, the quality of precipitation estimates is affected by partial beam blockage, drop size distribution uncertainty, and the distance from the radar station (Sachidananda and Zrnić 1987; Ryzhkov et al. 2014).

Numerous QPE methods using dual-polarized parameters have been proposed in recent years. Adopting the specific differential phase (K_{dp}) parameter makes QPE undisturbed from beam attenuation (Zrnić et al. 2000). Several studies have used differential reflectivity (Z_{DR}), K_{dp}, and Z to calculate raindrop size distribution and obtain new QPE formula (Zhang et al. 2001; Bringi et al. 2002; Brandes et al. 2003). Recent research by Ryzhkov et al. (2014) improved QPE by applying an attenuation formula during heavy rain events. The CWB combined these results and techniques to announce a new generation of QPESUMS system in 2021. The new QPESUMS system considers different radar scanning strategies and detectable parameters, selects the most suitable QPE method, and combines them to obtain 2D precipitation estimates (Chang et al. 2021).

However, there are still limitations to the second-generation QPESUMS system. First, a single formula for individual radar, even using different radar parameters, is insufficient for various weather systems and in different regions and terrains (Ryzhkov et al. 2005a,b; Jou et al. 2015), especially in eastern Taiwan. Second, and more importantly, calculating the accurate precipitation information should consider the vertical structure of the weather system, especially in complex terrain where beam blockage is a serious problem. However, the operational QPE method still uses the single point of radar reflectivity at the lowest available level. Figure 1 shows the correlation coefficient between Z and R at Taipei station. In addition to the low-level Z, the Z at

Corresponding author: Buo-Fu Chen, bfchen@ntu.edu.tw

DOI: 10.1175/WAF-D-22-0034.1

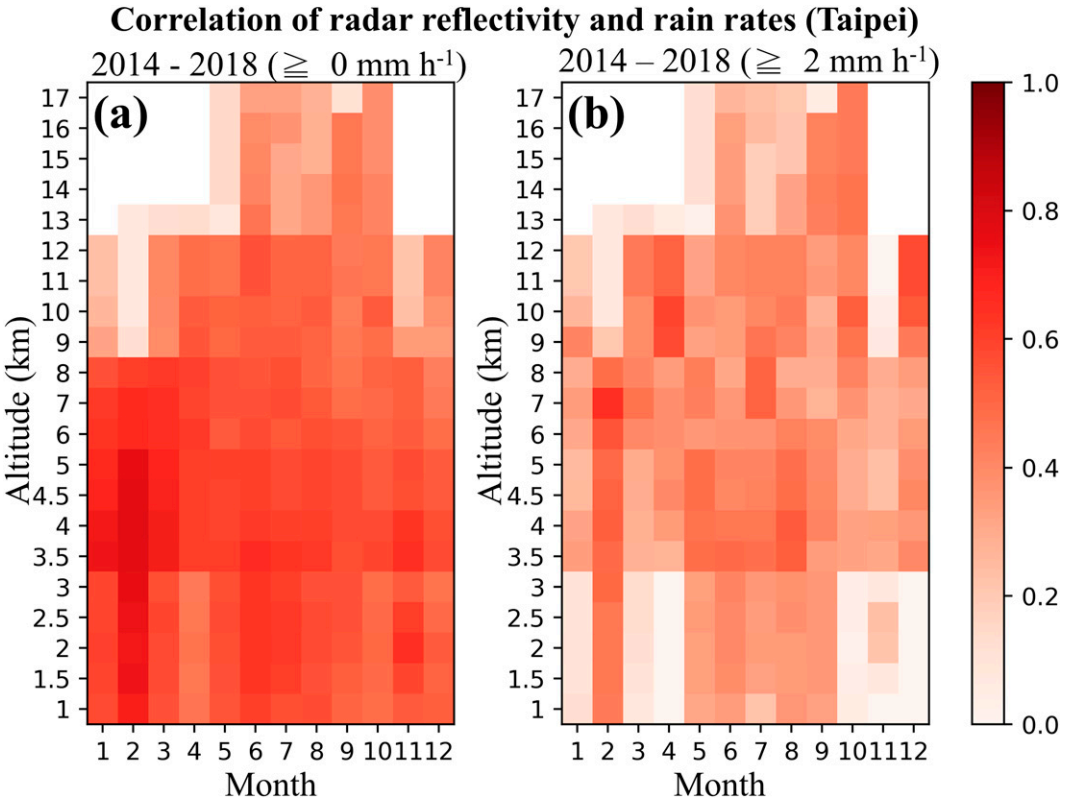

FIG. 1. The correlation coefficient (shading) between radar reflectivity and surface rain rates for (a) all samples and (b) samples with rain rates ≥ 2 mm h^{-1} from 2014 to 2018 at Taipei station. The data are stratified by various months (x axis) and altitudes (y axis).

3–9 km above sea level also well correlates with rain gauge observation, especially in summer and when the rain rate is larger than 2 mm h^{-1} (Fig. 1b). Therefore, it is hypothesized that we could obtain more accurate precipitation information by considering the upper-level reflectivities.

QPE could be regarded as an image recognition task estimating rain rate from 2D or 3D radar reflectivity. Deep learning has become increasingly popular in Earth sciences (Reichstein et al. 2019). Most of the recent studies on the application of artificial intelligence in atmospheric sciences use deep learning networks (Isola et al. 2017; Vandal et al. 2017; Shi et al. 2015, 2017; Sønderby et al. 2020; Becker et al. 2017; Gentine et al. 2018). A convolutional neural network (CNN) is one of the most common models that deal with visual recognition (Krizhevsky et al. 2012). Technically, CNN acquires features from multidimensional data, flattens these extracted features, and conducts nonlinear classification or regression. Lagerquist et al. (2020) used a CNN model to predict tornado occurrence 1 h ahead. Racah et al. (2017) used CNN to automatically search and identify weather systems on satellite cloud images, such as fronts, typhoons, and atmospheric rivers. Chen et al. (2019) used CNN to estimate tropical cyclone intensity based on satellite images. Therefore, this study proposes using CNN to analyze the characteristics of 3D radar data to overcome two major issues of the currently operational QPE in Taiwan. That is, (i) the 3D convective structure is not well considered, limiting the accuracy in the area suffering from beam blockage; and (ii) difficulty of choosing the optimal parameters of the $R(Z)$ formula that is suitable for all terrains.

This study proposes a new deep learning QPE technique that extracts 3D radar echo characteristics through contoured frequency by altitude diagram (CFAD; Yuter and Houze 1995). First, a cluster of CNN models is constructed at existing rain gauges in northern and eastern Taiwan. Subsequently, this study proposes new methods for generating 2D rain maps by the CNN cluster to increase the forecasting practicability. Specifically, we integrate the CNN models trained with the data at neighboring stations to conduct QPE at every grid point by using the observation of that specific grid point. We also compare it with the observation at independent stations to evaluate the performance of this new QPE technique. In addition, this study selects 13 June and 8 October 2018 for case analyses to visualize the results.

The paper is organized as follows. Section 2 describes data, model, and methods for generating 2D rain maps with the CNN cluster. Section 3 evaluates the overall performance of the CNN QPE model at the rain gauges. Independent verification of the generated 2D rain maps and two case studies are

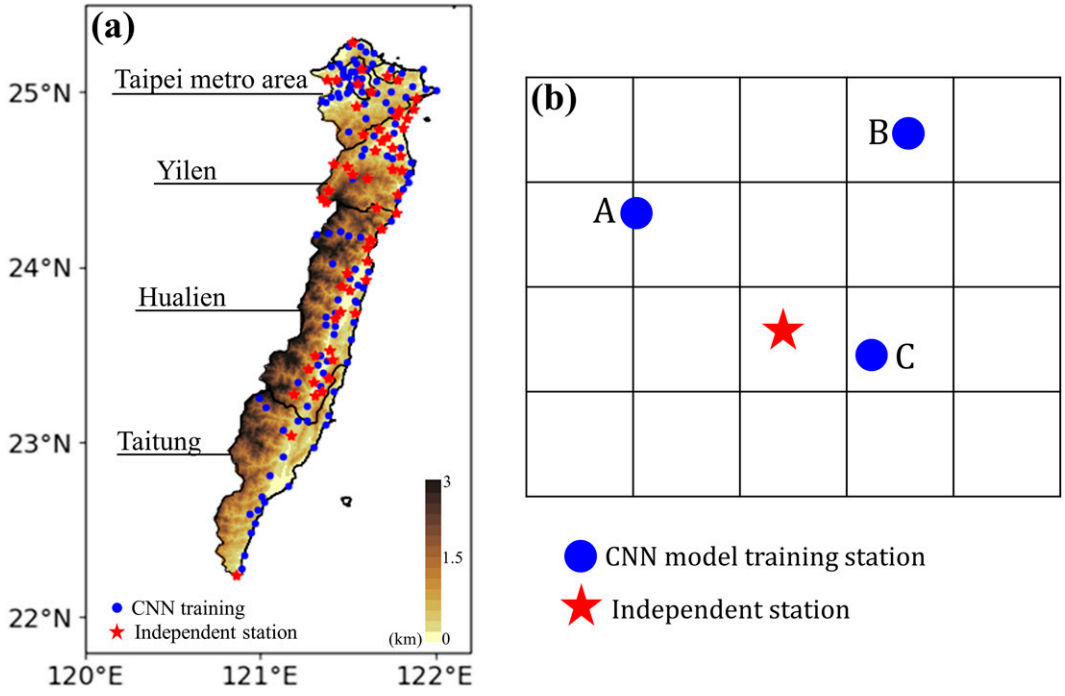

FIG. 2. (a) The terrain (shading) of the study area, stations for training deep learning models, and independent stations for verification. (b) Schema of QPESUMS grid, stations for training, and stations for independent verification.

discussed and visualized in section 4. The summary and future direction are provided in section 5.

2. Data and methodology

a. Rainfall data and the problem setup

This study proposes a deep learning QPE technique that utilizes 3D QPESUMS radar reflectivity as input to estimate 2D precipitation. The target areas include northern and eastern Taiwan (i.e., Taipei Metro Area, Yilan, Hualien, and Taitung). Hourly rainfall data of 166 rain gauges during 2014–18 were used as the ground truth, or the target of the model, including 13 CWB conventional weather stations and 153 automatic rain gauge stations (Fig. 2a). To establish a cluster of CNN models at existing rain gauges, data from 109 of the 166 stations were selected for training CNN models. These 109 rain gauges have almost no data loss from 2014 to 2018 (Fig. 2, blue dots).

On the other hand, this study used the data from the remaining 57 stations (Fig. 2, red stars) to evaluate the proposed QPE technique independently. These 57 gauges have good observations after 2018 but have substantial data loss during 2014–17, so they are unsuitable for training CNN models. Note that our final goal is to generate the 2D rainfall at each grid point, as the example shown in Fig. 2b, consisting of $4 \times 5 = 20$ grid points. Thus, this study tested some methods (see section 2d for more detail) to estimate the rainfall at the 57 independent stations (Fig. 2b, red stars) based on the trained CNN models at the nearby stations (Fig. 2b, blue dots). Suppose the rainfall at these 57 independent stations can be accurately estimated by blending estimates from nearby models, which are trained with data at locations represented by the blue dots in Fig. 2b, with inputs of the specific target location represented by the red star in Fig. 2b. In that case, we argue that rainfall estimates on the grid are useable.

b. Radar data and data preprocessing

This subsection introduces the QPESUMS radar reflectivities used in this study. The QPESUMS interpolated observations from multiple radars in Taiwan to form a high-resolution 3D gridded dataset covering the region of 20°–27°N and 118°–123.5°E. The horizontal and temporal resolutions are 0.0125° latitude–longitude and 10 min on 21 vertical layers.

The preprocessing process assists the CNN in learning the storm's vertical structure and saves computing resources; it transforms the 3D reflectivity into the contoured frequency by altitude diagram (CFAD) as the model input. Previous studies used CFADs to combine horizontal and vertical storm features and to easily display and describe the convective life cycle (Yuter and Houze 1995; Storer et al. 2014).

To obtain a CFAD, the reflectivity frequency distribution of each radar scan was calculated within a radius of 2.5 km to the target station (Fig. 3, left) and radar reflectivity is classified into 15 5-dBZ bins, which are commonly used in a CFAD. For example, the CFAD at Taipei station at 0610 local standard time (LST) 8 September 2018 (Fig. 3, right) is useful for distinguishing that the deep convection cloud developed up to 12 km, and the radar reflectivity value decreased with height. Also, the upper-layer values of the convective cloud are

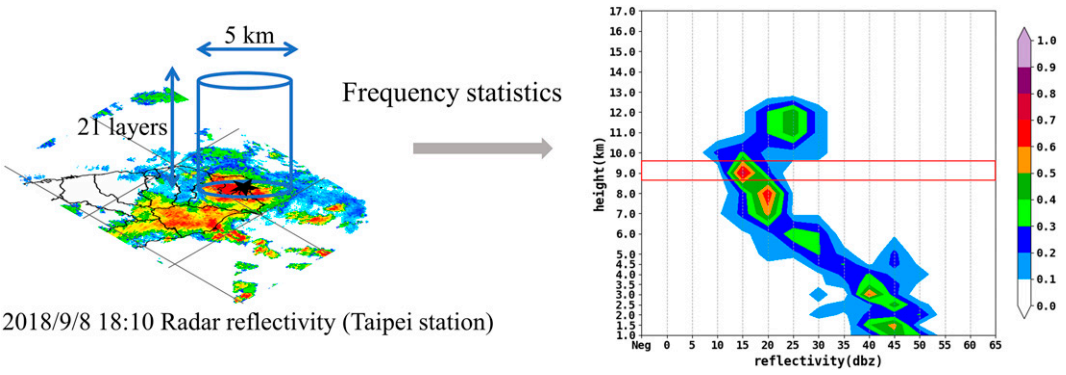

FIG. 3. Transforming (left) 3D radar reflectivity into (right) the CFAD. The shading in the right panel represents the frequency of radar reflectivity on each layer.

around 15–20 dBZ. This information might be helpful in training a CNN QPE model, especially when low-level reflectivity is blocked in complex terrain areas.

c. *CNN and model design*

This study chooses an approach of first training 109 independent per-station CNN models, which extract features from the CFAD and regresses to the hourly rainfall. Training per-station models for individual stations is the most straightforward way to make the model learn local characteristics for QPE in complex terrain, as long as we ensure that the model does not overfit the validation/training data (see later the results regarding Fig. 5).

As the input CFAD array only has dimensions of 21 × 15 × 6, we do not need a very deep CNN that may easily overfit the data. The proposed model consists of three kinds of neural layers (Fig. 4). The convolution layer is the core of a CNN model, utilizing convolution kernels to detect spatial features and update the weight during training via gradient descent with backpropagation, in order to minimize the loss function. Compared to traditional machine learning approaches, which cannot learn the spatial relationships in

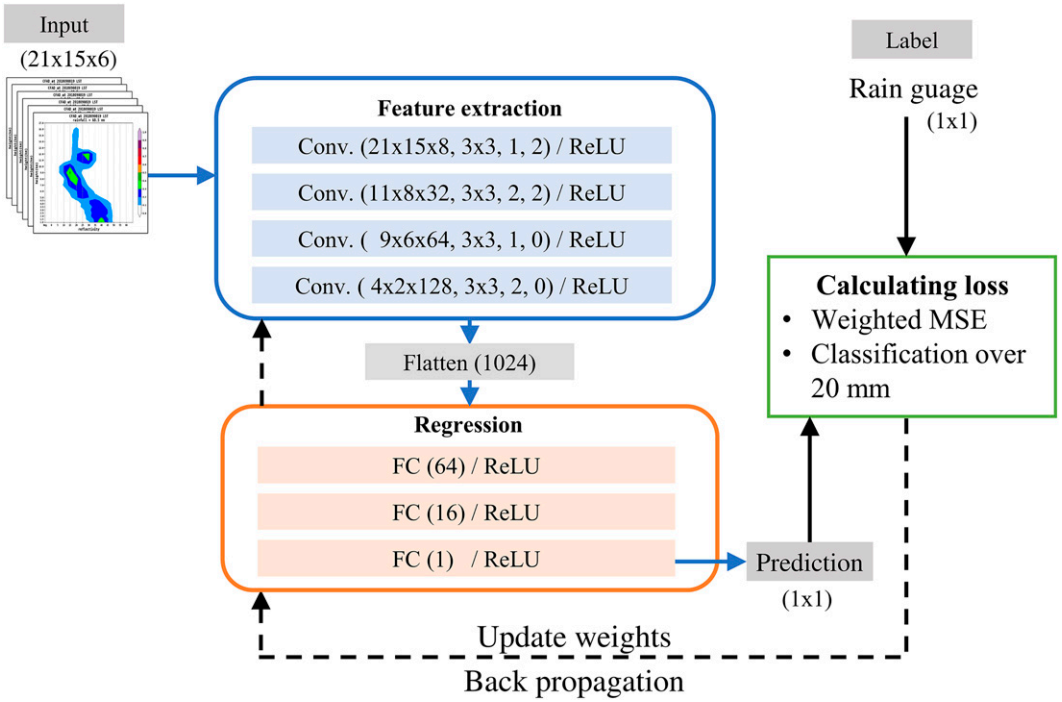

FIG. 4. The architecture of the proposed deep learning model. The blue rectangle inside the "feature extraction" module indicates [convolution layer (output dimension, kernel size, stride, padding)/activation function]. The orange rectangle inside the regression module indicates [fully connected layer (output dimension)/activation function]. The blue arrow means the direction of prediction. The black arrow means the backward propagation for training. The loss function of this model has two terms: weighted MSE and classification over 20 mm.

images, convolution kernels can assist a deep learning model in training directly with images. The activation function layer allowing the model to be nonlinear is an equation with no adjustable parameters—different from the convolution layers, where each convolutional kernel contains weights that can be adjusted by training. Currently, rectified linear unit (RELU) is the most commonly used activation function in CNNs; it is defined as

$$f(x) = \max(x, 0). \tag{1}$$

This equation only allows the result greater than zero from the previous layer to pass to the next layer and can improve the learning efficiency of the model (Nair and Hinton 2010; Glorot et al. 2011). The third kind of layer is the fully connected layer, similar to that in a traditional artificial neural network (ANN; Jain et al. 1996). The purpose of the fully connected layer is to consider the overall characteristics of the data and make judgments, i.e., using the results extracted from convolutional layers for subsequent classification or regression. Finally, according to some ablation experiments and the characteristic of the CFAD that lacks contours, our proposed model omits the pooling layer usually used in deeper CNNs, following Chen et al. (2018).

At each CNN model training station (Fig. 2, blue dots), six pieces of 10-min CFADs (6 × 10 min = 1 h) for the previous hour are concatenated as model input, 21 × 15 × 6 in dimensions (Fig. 4, upper left). As shown in Fig. 4, our model contains two parts: the feature extraction module and the regression module.

The feature extraction module includes four convolution layers, each followed by a ReLU. The kernel size is 3 × 3 for all layers, a common setting in CNNs. The filter sizes, which means the number of convolutional filters in a given layer, are set as 8, 32, 64, and 128, respectively. A key reason for selecting the number of convolutional filters is that we hope to transform the 21 × 15 × 6 input CFADs to 1024 extracted features, a common and arbitrarily selected feature number in neural networks. Also, to down sample the data to 1024 features, strides and padding are subsequently determined as [1, 2, 1, 2] and [2, 2, 0, 0] for the four layers. The stride is the parameter to decide how many paces the kernel will move from pixel to pixel. For instance, if the stride equals 2, the kernel will slide two pixels after finishing scanning the former pixels (Albawi et al. 2017). The padding is the strategy in the convolution layer. If the padding equals 2, the model will add extra 2 pixels with 0 in each column and row around the boundary of the image to ensure each pixel in the image is scanned the same number of times (Albawi et al. 2017).

Of note, some other combinations of convolution layer numbers, filter numbers, strides, and padding numbers were tested, such as one with three convolution layers and another with five. Results show that the CNN with three convolution layers has the poorest performance. The five-layer CNN performs similarly to the CNN with four convolution layers, but the four-layer CNN appears to achieve more stability (i.e., less fluctuation in the curve of validation loss). Therefore, we decide to use four convolution layers for the final configuration.

In the wake of feature extraction, we flatten the features from CFADs and propagate them into the second-part regression module. Three fully connected layers are used to regress to the hourly rainfall, and the results also pass through a ReLU after each layer. The neuron sizes of the layers are selected as 64, 16, and 1, following common practice in machine learning. Note that reducing the number of features with each successive fully connected layer enables the model finally estimates the hourly rainfall, a single value.

Regarding the data-use strategy, the CFADs and the paired hourly rainfall labels at each CNN training station (Fig. 2, blue dots) were categorized into three groups: training, validation, and testing. This study divided every 4 days from 2014 to 2017 to form the training and validation datasets. In each 4-day cycle, the first 3 days belong to the training dataset, and the latest day is used for validation. The purpose is to maintain the statistical consistency between the training and validation datasets to avoid model bias due to seasonal and diurnal cycles. Furthermore, we selected the CFADs in 2018 as the testing dataset to evaluate the CNN model's performance. Note that the testing dataset for evaluating the model's performance at the training site and the independent verification dataset for evaluating the model's performance at independent stations (Fig. 2, red stars; refer to section 2a) are both from 2018. In other words, the testing dataset is separated from training/validation by time, whereas the independent verification dataset is separated from training/validation/testing by location. Thus, the evaluation based on the testing dataset can cooperate with the independent station evaluation to gain a comprehensive insight into the model performance.

In machine learning, the loss function helps the model fit optimal weights through gradient descent, which attempts to find the global minimum loss based on the training dataset. The most commonly used loss function for regression is the mean square error (MSE):

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2, \tag{2}$$

where N is the sample number of training data, X is model prediction (i.e., hourly rainfall estimates), and Y is the target from rain gauge data. MSE could magnify the punishment for large-error samples compared to calculating the absolute error. However, using MSE as a loss function has the drawback of making the model ignore the small number of imbalanced samples (i.e., extreme rainfall) in the training dataset.

Unfortunately, in our study, the frequencies of different rainfall levels are highly imbalanced. Taking the Taipei station as an example, there was only 0.5% of data greater than 10 mm h^{-1} and 90% of data on sunny days (rain rate = 0) according to hourly rainfall data from 2014 to 2018. To solve this problem, we used two alternative components in the loss function, instead of the MSE, to improve the model performance for heavy rainfall events: weighted MSE and classification of rainfall over 20 mm.

Weighted MSE is described here. Before gathering the errors together, multiplying the weight based on the target value

emphasizes the accuracy of the heavy rain (Shi et al. 2017) so that Eq. (2) is rewritten as

$$\text{weighted MSE} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{w}(Y_i) \times (X_i - Y_i)^2, \tag{3}$$

where $\mathbf{w}(Y_i)$ is designed as

$$\mathbf{w}(Y_i) = \begin{cases} 1 & \text{if } Y_i < 1 \text{ mm} \\ Y_i & \text{if } Y_i \geq 1 \text{ mm} \end{cases}. \tag{4}$$

When the rain gauge data (Y_i) is greater than 1 mm h^{-1}, the weight **w** would be the rain gauge data itself. On the other hand, when the rain gauge data (Y_i) is less than 1 mm h^{-1}, the **w** would be one.

The other loss term is described: classification of rainfall over 20 mm. The purpose of this term is to set the following goal for the CNN "classify if it is heavy rain or not first, then do the regression for the certain region." We hope the model can learn the classification for heavy rainfall greater than 20 mm h^{-1}. This component of the loss function is defined as

$$\begin{matrix}\text{classification over 20 mm} \\ \text{(only if } Y_i \geq 20)\end{matrix} = \begin{cases} 0 & \text{if } 20 \leq X_i \\ 200 & \text{if } 10 \leq X_i < 20 \\ 400 & \text{if } 0 < X_i < 10 \\ 600 & \text{if } X_i = 0 \end{cases}. \tag{5}$$

This component is only effective when the rainfall (Y_i) is over 20 mm h^{-1}, and additional punishments are given if the model predicts a value that falls in various categories (with a 10 mm h^{-1} interval).

Finally, a cluster of CNN models at the 109 training stations (Fig. 2) is constructed based on the above CNN design. Each CNN model's weights are independent and should not share with other stations because every station has its own features, including geographic location, distances from radars, and terrain height. Each CNN model was trained on the Intel Xeon CPU E5-1650 v2, using the TensorFlow 2 framework. It only took 30 min to complete 200 epochs for training at each station, and most of the CNN models only required 50–150 epochs to reach a minimum of the validation loss for learning this task and getting the optimal weight. The performance of the CNN cluster is described in section 3.

d. *Producing gridded QPE product*

The gridded QPE product is demanded to increase the QPE's practicability of weather forecasting for hydrometeorology use, such as calculating the average rainfall in certain river basins. This study proposes two gridded QPE products based on the CNN clusters, which have been trained at the 109 rain gauge sites, to provide 2D rain maps. As mentioned in section 2a, data from 57 independent stations are prepared to verify the performance of the two gridded QPE products. It is worth noting that what we do here is to conduct QPE based on the radar CFAD at each point rather than simply doing interpolation according to the QPEs at nearby rain gauges. It is hypothesized that each of the trained CNNs is representative of the nearby area, and using the radar observation of each point could better reflect the fine-scale rainfall features. Moreover, if the rainfall at the 57 independent stations can be accurately estimated, we can argue that rainfall estimates at every grid point are useable.

The first method is the nearest-station method. Assuming that the CNN of the nearest station has similar precipitation characteristics to the target location, this method uses the target point's CFAD as input and feeds it into the CNN of the nearest station to get the QPE. As illustrated in Fig. 2b, this method utilizes the CFAD at the independent station (Fig. 2b, red star) and feeds it into the CNN of the nearest training station C. The advantage of this method is that it is easy to operate and effectively utilizes the CFAD at each target grid point.

The second method is the three stations blending method. This method feeds the target point's CFAD into the CNNs of the three nearest stations and integrates the results based on the inverse distance weighted interpolation (IDW; Bartier and Keller 1996). As shown in Fig. 2b, utilizing the CFAD at the independent station (red star) as the input for the three nearest CNNs (station A, B, and C) and combining the three rainfall estimates with the IDW leads to the final QPE.

The IDW interpolates the information into the target point using a weighting function inversely proportional to the square of the distance, defined as

$$A_i = \frac{\sum_{k=1}^{3} \frac{1}{r_{ik}^2} \times O_k}{\sum_{k=1}^{3} \frac{1}{r_{ik}^2}}, \tag{6}$$

where A_i, O_k, and r_{ik} represent the hourly rainfall at the target point, the observation, and the distance between the target point and the observation. Although this method is more comprehensive than only considering the nearest CNN, it would smooth the extreme rainfall.

On the other hand, two additional baseline methods plotting 2D rain maps are used for verifying our gridded QPE products: $R(Z)$ relation and Cressman method (Cressman 1959). Both $R(Z)$ and Cressman methods have been used at CWB. The $R(Z)$ was the basic method for producing radar QPE in the operational QPESUMS. The Cressman method is one of the most common operational objective analyses for interpolating rain gauge observations into a 2D rain map. The equation of the Cressman method is

$$A_i^n = F_i^{n-1} + \frac{\sum_{k=1}^{N} W_{ik} \times (O_k - F_k^{n-1})}{\sum_{k=1}^{N} W_{ik}}, \tag{7}$$

where A, F, and O_k are estimated value of precipitation, guess (background) value (first guess value = 0), and observation, respectively. The subscripts i and k represent the target point and the location of observation, respectively. The term n means the result of the nth adjustment and W is the weighting factor, defined as

$$W_{ik} = \begin{cases} \dfrac{R^2 - r_{ik}^2}{R^2 + r_{ik}^2} & \text{if } r_{ik} < R \\ 0 & \text{if } r_{ik} \geq R \end{cases}, \tag{8}$$

where r is the distance between the target grid point and the observation, and R is the radius of influence. Similar to the CWB setting, the initial guessing field (F_1) is 0, and adjust the value three times given different influence ranges (R): 10, 7, and 5 km, respectively.

3. The verification of CNN models at the training stations

a. *Examination of overfitting, benefits of the per-station approach, and impacts of upper-level Z*

Before evaluating the model performance against the baseline $R(Z)$ relation, several experiments shown in this subsection justify or explore: (i) the trained models do not overfit the validation and testing data; (ii) the advantage of the per-station model approach; and (iii) the contribution of upper-level CFAD information on QPE.

First, analysis is conducted to evaluate the CNN estimation accuracy for the training, validation, and testing datasets. As a deep learning model generally overfits the training data because of its large number of neurons, it is critical to confirm that the trained model does not overfit the validation and testing data. As we examined the hourly rainfall estimated by the CNN models and the corresponding observation (Figs. 5a–c), the 109 CNN models for the 109 training stations show a good performance on the training data according to the linear-regression line slope of 0.91 and a correlation coefficient (C.C.) of 0.72 (Fig. 5a). On the other hand, the CNN models have similar performance on both the validation (Fig. 5b) and testing (Fig. 5c) datasets. The slopes and C.C. in Figs. 5b and 5c are slightly lower than those based on the training data, but the regression line still falls in the interval of 20% error (Fig. 5, gray shaded area). Furthermore, the boxplots in Fig. 5d show that the error distributions for the testing dataset are comparable to that of the validation set and even slightly better for hourly rainfall > 30 mm, implying that the CNN models can successfully capture features in the CFAD and stably estimate rainfall for future applications.

In addition, to realize the benefit of training CNN models in the per-station approach, we conducted experiments of training a general CNN model. That is, rainfall estimates of the 109 stations are from a single CNN model with fixed weights for various stations. Note that, as the training dataset will be huge in this manner, we randomly select 8000 rainfall events for each of the 109 stations in 2014–17 to form the training and validation datasets for training the general model.

The performance comparison between the general model and 109 per-station models for the testing dataset (Fig. 6) reveals that the per-station CNN models outperform the general model with significantly smaller biases. Furthermore, to understand the model performance in complex terrain, we categorized the rain gauges above and under 200-m altitude into the mountain and plain areas, respectively. The difference in the performance is more obvious for the rainfall samples in the 52 mountain stations (Fig. 6a) than that of the 57 plain stations (Fig. 6b), and the general model has worse performance in the mountains presumably because of the difficulty to learn the local-specialized features. Thus, these results imply that, as each station has its unique features, utilizing a general CNN model cannot obtain accurate estimates at each station.

Last but not least, this study also performs a sensitivity experiment on CFAD heights to consolidate the claim that 2D CFAD images have more predictive power than point reflectivity because upper-level information also correlates with hourly rainfall. Upper-level information is especially critical in the mountains where low-level radar beams are blocked.

Three kinds of CFADs are tested as the model input data: 0–2, 0–8, and 0–17 km (Fig. 7). The 0–17-km CFAD is the standard CFAD used in this study, and a 0–2-km CFAD is the corresponding 0–17-km CFAD in that values at >2-km levels are replaced with zero. Although three CFAD experiments have similar performance in plain areas (Fig. 7d), the CNN models using 0–2-km CFADs have worse performance in mountain areas (Fig. 7b). Moreover, adding upper-level radar information can efficiently help the CNN model to obtain more accurate precipitation for heavy rainfall events. It is worth noting that 0–8-km CFADs lead to comparable performance with 0–17 km CFADs, consistent with an observation regarding Fig. 1 that most of the high correlation parts are below 8 km. Nevertheless, we select to use 0–17-km CFADs as the final model input.

b. *Evaluating CNN model performance against the baseline R(Z) relation*

This section aims to evaluate the performance of the CNN QPE models (section 2c) based on the rainfall data at training stations (i.e., the testing dataset) and compare it to the baseline $R(Z)$ relation results. It is worth noting that, although numerous QPE methods using dual-polarized parameters have been proposed, the $R(Z)$ relation used for CWB operational radar QPE is still the best baseline for this study because dual-polarized radar observation has not yet been available in our research area.

As mentioned in section 2a, the rainfall data of 2018 are used for model verification, and the performance evaluation is based on mean absolute error (MAE), one of the most common indexes to verify the model error, and mean absolute percentage error (MAPE). MAPE is defined as

$$\text{MAPE} = \sum_{i=1}^{n} \left| \frac{X_i - Y_i}{Y_i} \right| \times \frac{100\%}{n}, \tag{9}$$

where X_i, Y_i, and n represent the CNN rainfall estimates, the observation, and the sample number of the testing dataset. The difference between the MAPE and MAE is that although the value of $|X_i - Y_i|$ is identical, the MAPE better reflects the improvement for different rainfall levels. Again, we categorized the rain gauges above and under 200-m altitude into the mountain and plain areas.

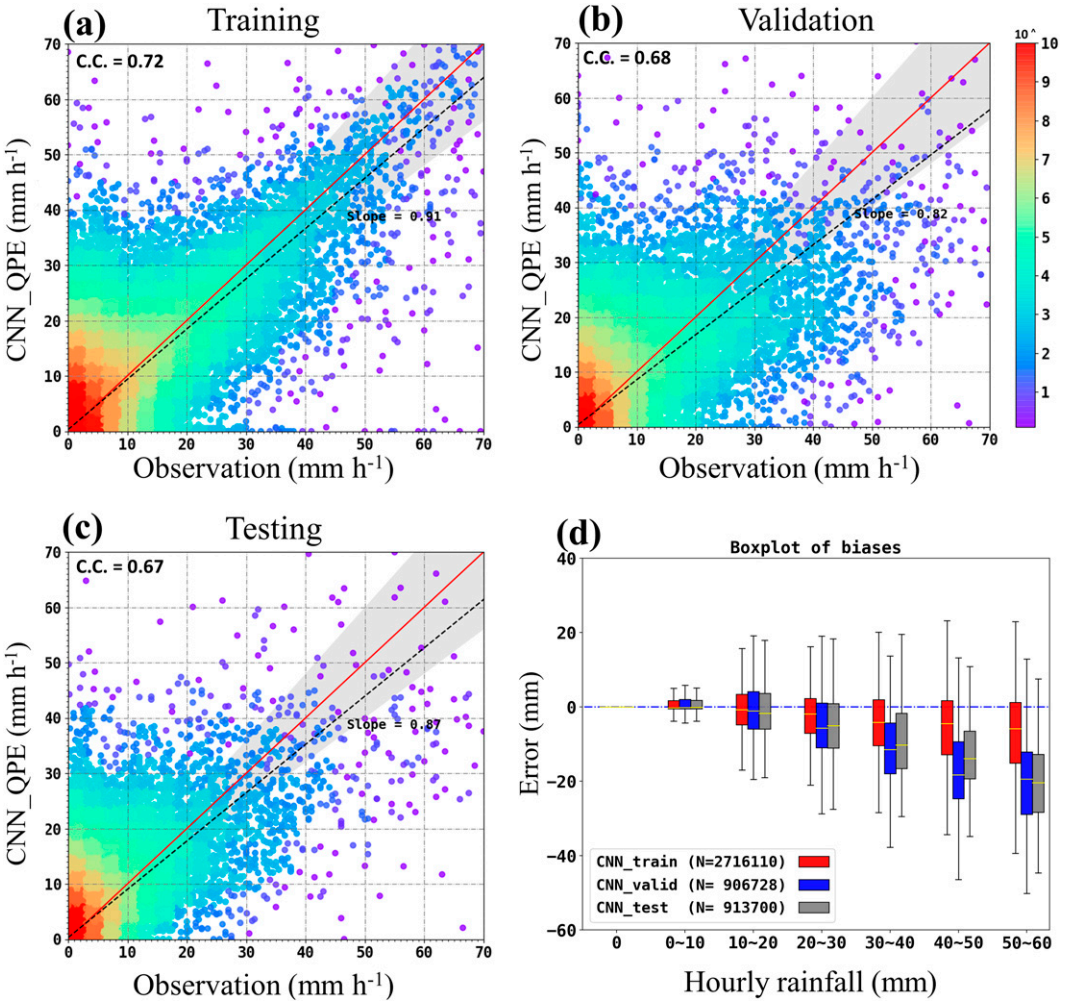

FIG. 5. (a)–(c) Heat scatterplots of hourly rainfall estimated by CNN models (y axis) and gauge observation (x axis) for the 109 stations for (a) training, (b) validation, and (c) testing datasets. Color shading represents the sample density. The red line and gray shading indicate the perfect regression line and interval of 20% error, respectively. The black dashed line is the regression line with the slope indicated. The upper-left corner shows the correlation coefficient (C.C.). (d) Boxplots of error (y axis) in rainfall estimates for various rain-rate bins (x axis) for training (red), validation (blue), and testing (gray) datasets. Numbers N represent the sample numbers. The boxplot's center line, two box edges, and upper and lower whiskers represent the median, 25th and 75th percentiles, and 75th/25th percentile ±1.5 interquartile range, respectively.

Furthermore, the bootstrapping test is used for examining the statistical significance of the comparisons between CNN models and the $R(Z)$ relation; it can be done with the following steps. Let N be the number of data samples in the evaluation set, $K = 1000$ be the number of bootstrap replicates, u be the evaluation metric, and c be the confidence level of the test (95%),

1) Do the following N times:
   (i) Draw N data samples from the evaluation set.
   (ii) Compute u for both models—e.g., the CNN and $Z(R)$ relation.
   (iii) Compute the difference between the two models: $\Delta u = u_{Z(R)} - u_{\mathrm{CNN}}$.
2) The result is a distribution of N differences between the two models. In this distribution of N differences, find the percentile level corresponding to a $\Delta u = 0$. This percentile level is the p value of the test.
3) Compare the p value to 1 minus the confidence level. If $p < 1 - c$, we can reject the null hypothesis [the CNN and $Z(R)$ relation have the same performance] and conclude that the CNN is significantly better.

The first quantitative analysis compares the performance between the CNN QPE models and the $R(Z)$ relation. As shown in Fig. 8, the 109 CNN training stations are categorized into 52 mountain stations (Figs. 8a,c) and 57 plain stations (Figs. 8b,d). Generally, although both the pattern of the CNN

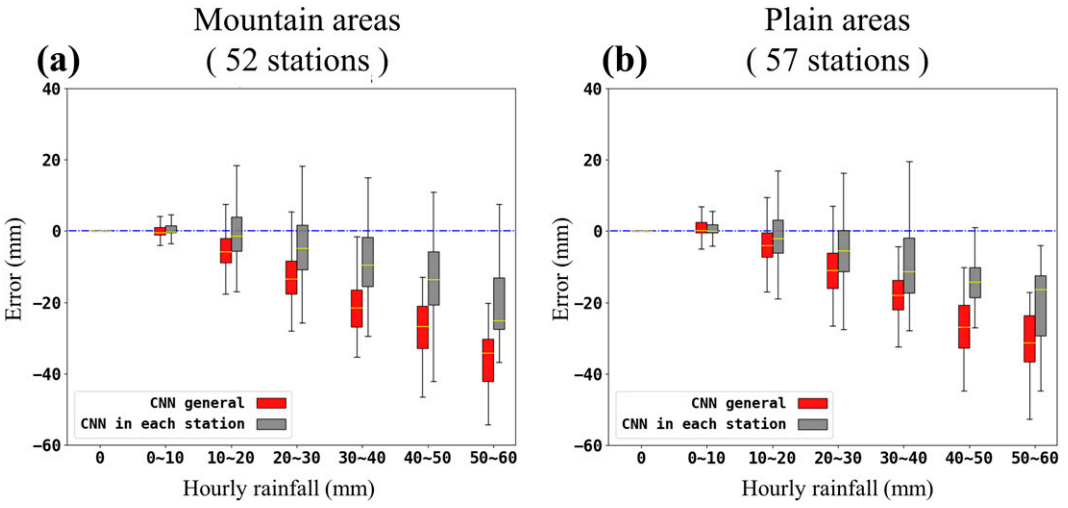

FIG. 6. As in Fig. 5d, but for the rainfall estimates by the general model (red) and the pre-station models (gray) for stations in (a) mountain areas and (b) plain areas. Note that this figure is based on the testing dataset.

and $R(Z)$ reveal underestimation of heavy rain (>20 mm h^{-1}), the CNN model has better performance than the $R(Z)$ relation according to the slopes (Fig. 8, black dashed line and texts) and samples that fall in the interval of 20% error, especially in the mountain area (Figs. 8a,c) and for heavy rainfall events. For the plain area, although the improvement of the CNN cluster is not comparable to the improvement in mountain areas, the CNN cluster still reduces the underestimation for heavy rainfall events (Figs. 8b,d). The results suggest that CNN models analyzing upper-layer radar information outperform the baseline $R(Z)$ relation.

Subsequently, the rainfall events are divided into seven groups with 10 mm h^{-1} intervals (Fig. 9). We quantify the improvement of the CNN model compared with the $R(Z)$ performance for different rainfall intervals and areas. Their error distributions, MAE, and MAPE show consistent results in the preliminary quantitative analysis (Fig. 8) with more details. Specifically, the CNN outperforms $R(Z)$ quantitatively for 10–60-mm intervals in the mountain areas (Fig. 9a) and has similar performance in the plain areas for <40-mm rainfall (Fig. 9b). Moreover, the bootstrapping test results show that the performance of the CNN models is significantly better than the $R(Z)$ relation for 20–40-mm rainfall in the mountain area. Figure 9 also shows that, for 10–60-mm intervals, the CNN model performance can improve about 21% in MAPE and MAE compared to the $R(Z)$ relation. However, the CNN performs poorly in MAPE for rainfall less than 10 mm (Figs. 9c,d). The reason is that rainfall values in this range are too small, so that the MAPE would be high. Nevertheless, if we change the verification metrics to MAE, the CNN model has comparable performance with the $R(Z)$ relation in the 0–10-mm interval. Moreover, mountain stations benefit more from the CNN model, especially in 20–40-mm ranges. For 30–40-mm rainfall in the mountain area, CNN's MAE improves up to 50% compared with the $R(Z)$ relation (Figs. 9e,f).

Additionally, we analyze the results of some particular regions in our study area (Fig. 10). All regions performed similarly to that we have mentioned above. We select the Hualien area and the Taipei metro area as examples. The Hualien area, with the most complex terrain in Taiwan, is where the CNN model gets the most significant improvement in this study (Figs. 10b,d,f). In contrast, the Taipei metro area is the research area with the least mountains in this study. As shown in Fig. 10, CNNs could improve the underestimation of $R(Z)$ in both the Taipei metro area and the Hualien area. But note that, from the boxplot in the Hualien area (Fig. 10f), it is clearly shown that CNNs have a smaller bias than $R(Z)$ in heavy rainfall events (about 35% improvement). In summary, it is indicated that the upper-level information is important to help the CNN model estimate the precipitation in complex terrain, where radar beam blockage is a serious problem, i.e., in Hualien.

4. Evaluation of the 2D QPE products

a. Independent station verification

Recall that the final goal is to produce 2D rain maps; it is thus important to evaluate the performance of the gridded QPE products (section 2d) at every grid point. As it is impossible to have ground-truth observation everywhere, an improvising way to verify the 2D QPE products is an evaluation based on the 57 independent stations. This subsection evaluates the performance of the nearest-station method, the three-stations-blending method, and the $R(Z)$ relation for the 57 independent stations (Fig. 2, red stars), of which the data are not used for training and validating the CNN cluster. If the rainfall at the independent stations can be accurately estimated, we can fairly argue that rainfall estimates at each grid point are useable.

First, we compare the performance of these two gridded QPE products (the nearest-station method and the three-stations-blending method) for the independent stations with

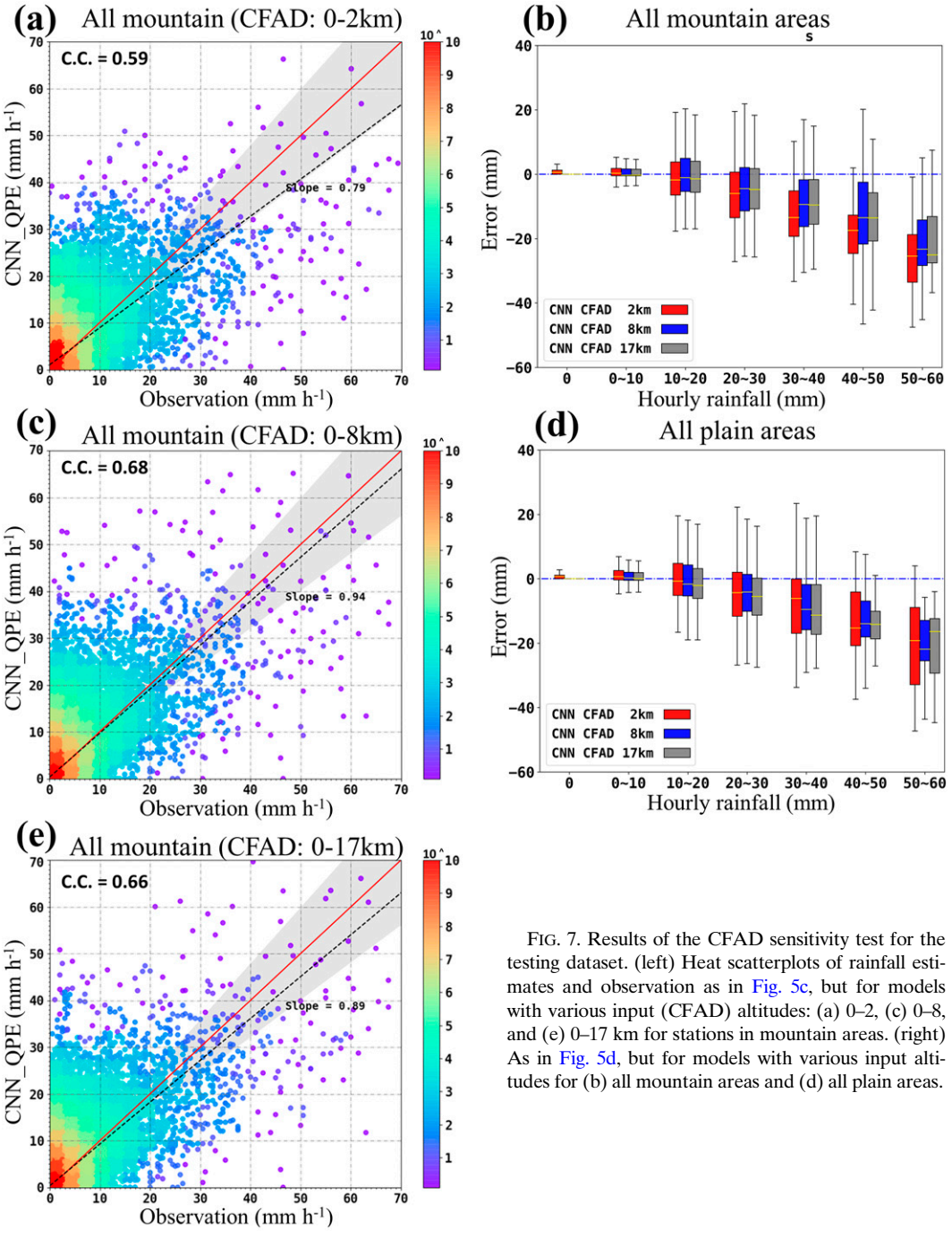

FIG. 7. Results of the CFAD sensitivity test for the testing dataset. (left) Heat scatterplots of rainfall estimates and observation as in Fig. 5c, but for models with various input (CFAD) altitudes: (a) 0–2, (c) 0–8, and (e) 0–17 km for stations in mountain areas. (right) As in Fig. 5d, but for models with various input altitudes for (b) all mountain areas and (d) all plain areas.

the CNN clusters' performance for the 109 CNN training stations (Fig. 2, blue dots). Note that this is not a homogeneous comparison, but it is important to realize if the qualities of these two gridded QPE products and the CNN clusters are comparable. If they have similar performance, it could be concluded that the QPE is stable when the CNN model is applied to the nearby area around the training station and increases confidence when we produce the 2D rain maps.

As shown in Fig. 11, both the nearest-station and the three-stations-blending methods have shown consistent results with the CNN cluster conducting QPE at the CNN training stations (Figs. 11a,b). Although the three stations blending method has the best performance for rainfall less than 50 mm h^{-1} from the MAE and MAPE (Figs. 11c,d), it might underestimate the hourly rainfall when the observation is greater than 50 mm h^{-1}, especially in plain areas. Note again that, compared to the $R(Z)$ relation with uneven performance in the

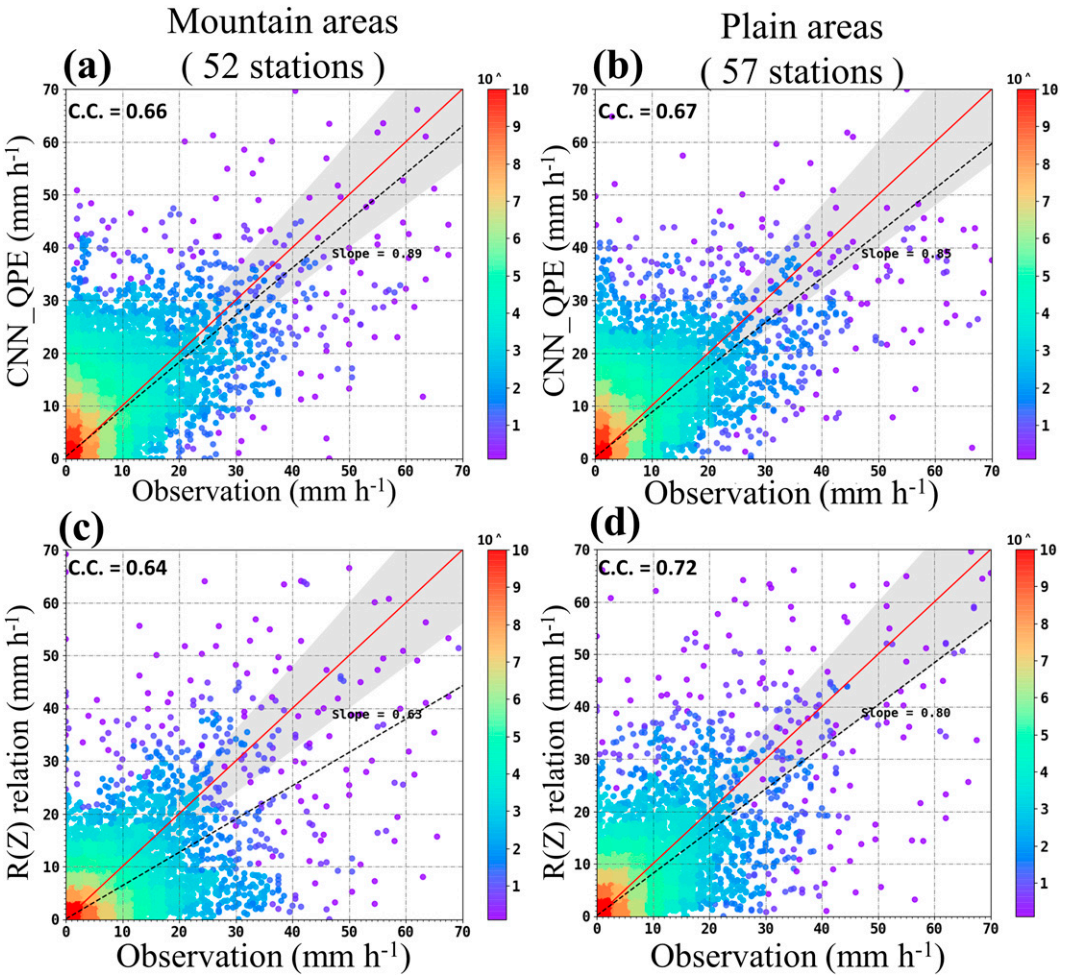

FIG. 8. For the testing dataset, the scatterplots of precipitation estimates and observation as in Fig. 5c, but for stations in (a) mountain areas and (b) plain areas. (c),(d) As in (a) and (b), but for precipitation estimates by the $R(Z)$ relation.

mountain and plain areas (Fig. 10f), the CNNs can consider the upper-level reflectivities to reduce the problem of radar beam blockage and improve the QPE performance in the mountains.

In summary, the statistical properties of errors for the two gridded QPE products are comparable to each other and similar to that of the evaluations of the CNN models at the 109 training stations. Moreover, the nearest-station method is easier to operate and shows more possibilities to capture the rainfall extremums for rainfall greater than 50 mm h^{-1}. The results imply the feasibility of substituting the local CFAD into a neighboring CNN model to estimate the local precipitation. Therefore, we promote the nearest-station method for producing the 2D rain maps.

b. Two case studies for visualization

This subsection selects two cases of daily rainfall for visualizing the QPE results (Fig. 12): 13 June and 8 October 2018. In the first case, a stationary front was located in the southeast of Taiwan, leading to the rainfall in eastern Taiwan. In the second case, Taiwan was influenced by the northeast monsoon on 8 October 2018. These two cases are visualized by four gridded QPE techniques: The Cressman interpolation method based on 109 rain gauges (Figs. 12a,e) and 109 CNN models (Figs. 12b,f), the $R(Z)$ relation (Figs. 12c,h), and the nearest-station method (Figs. 12d,g).

The Cressman method (section 2d) is used to interpolate the 2D rain map based on the observation of the 109 training stations (Figs. 12a,e), similar to the CWB official rain map. Furthermore, we utilized the same Cressman setting to integrate the QPEs based on the CNN cluster of these 109 stations (Figs. 12b,f). If the 109 sites CNN QPE (Figs. 12b,f) is identical to the 109 sites observation (Figs. 12a,e), it can be considered a perfect QPE.

As shown in the left four panels of Fig. 12, the 109 sites observation and the 109 sites CNN QPE visually exhibit comparable rainfall distribution (Figs. 12a,b,e,f), except the local maximum rainfalls of the 109 sites CNN QPE are higher in southeastern Taiwan. Note that these four rain maps (Figs. 12a,b,e,f) show circle patterns near the rain gauges,

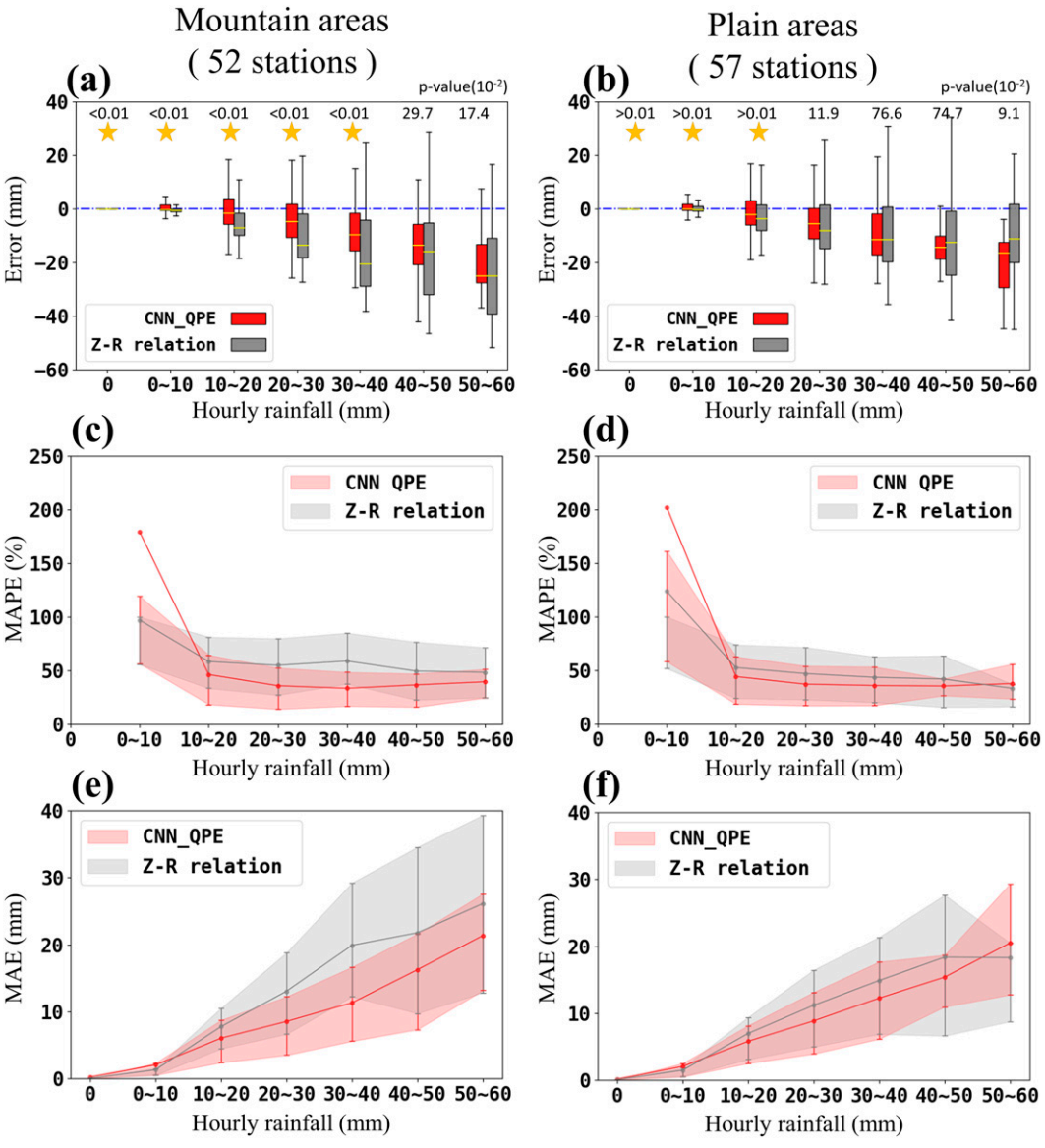

FIG. 9. (top) As in Fig. 5d, but for the error distributions of CNN estimates (red) and $R(Z)$ relation estimates (gray) for testing samples in (a) mountain areas and (b) plain areas. The number and star above the box represent the p value of the bootstrapping test and if the difference is significant. (middle) The MAPE for the CNN estimates and $R(Z)$ estimates in (c) mountain areas and (d) plain areas. The dots and the shading represent each rainfall bin's mean and interquartile range. (e),(f) As in (c) and (d), but for the MAE.

especially in areas where the rain gauge distribution is sparse because the Cressman method uses an unreasonable influence radius. Thus, we suggest that it is important to utilize radar reflectivities at all grids efficiently, and this study thus proposes the nearest-station method to generate the 2D rain maps.

Comparing the nearest-station method (Figs. 12d,h) and the two Cressman methods (Figs. 12a,b,e,f), the nearest-station method generates more small-scale weather features, especially in complex terrain (Fig. 12, "A" and "B"). Note again that using the neighboring CNN model, the nearest-station method can reserve the characteristics of each grid point's information from CFADs and reflect these features on the grid point's precipitation. Therefore, the nearest-station method significantly improves the overestimation of the $R(Z)$ relation (Figs. 12c,g), especially in complex terrain with serious radar beam blockage problems. Furthermore, as shown in Fig. 12 ("C"), there is an obvious boundary close to the coastline of Hualien, and as seen in Fig. 2a, there are about 1.3-km mountains in this region, so the $R(Z)$ relation method cannot accurately estimate the precipitation in the west side of the mountains due to the problem of radar beam blockage.

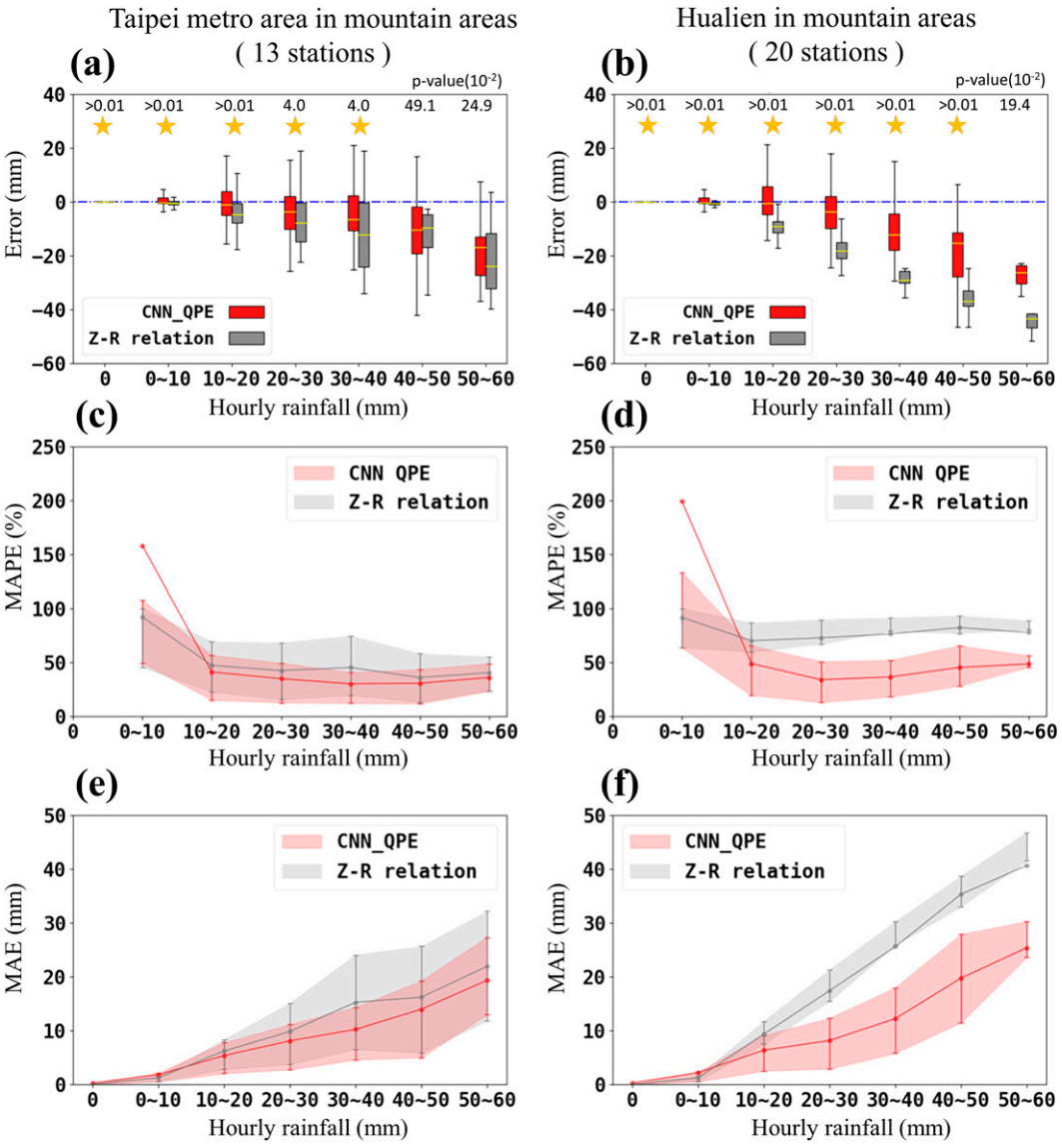

FIG. 10. Corresponding plots of Fig. 9 for mountain stations in (a),(c),(e) Taipei metro area and (b),(d),(f) Hualien area.

5. Summary and future work

a. *Summary*

In Taiwan, accurate QPE is urged for preventing and monitoring meteorological disasters. Many dual-polarized meteorological radars have been deployed recently, and numerous QPE methods using dual-polarized parameters have been proposed. However, the operational QPE techniques in Taiwan (e.g., QPESUMS) can only provide accurate rainfall estimates in plain areas and suffer from two critical issues. First, a single formula for individual radar is insufficient for different regions and terrains. Second, in the complex terrain area where beam blockage is a serious problem, using the single-point lowest available radar reflectivity for QPE is risky as the reflectivity may not reflect the near-ground raindrop size.

Therefore, this paper proposes to use a cluster of CNN models to construct 2D rain maps, hoping to establish better relationships between "imperfect" radar observations and "local" rainfall for various regions and terrains to address the first issue. Moreover, rather than using the single-point lowest available echo information, our deep learning QPE extracts features from CFADs to account for the 3D structure of the convective systems, addressing the second issue.

CNN models at existing rain gauges (i.e., CNN training stations) are trained in northern and eastern Taiwan based on the 3-yr data during 2015–17. To ensure the statistical consistency of datasets for establishing CNN models, we divided 4 days as a cycle to construct the training (days 1–3) and validation (day 4) datasets and utilized the 2018 data as the testing dataset. Furthermore, to solve the unbalanced-data problem for extreme rainfall, there are two components (weighted MSE and

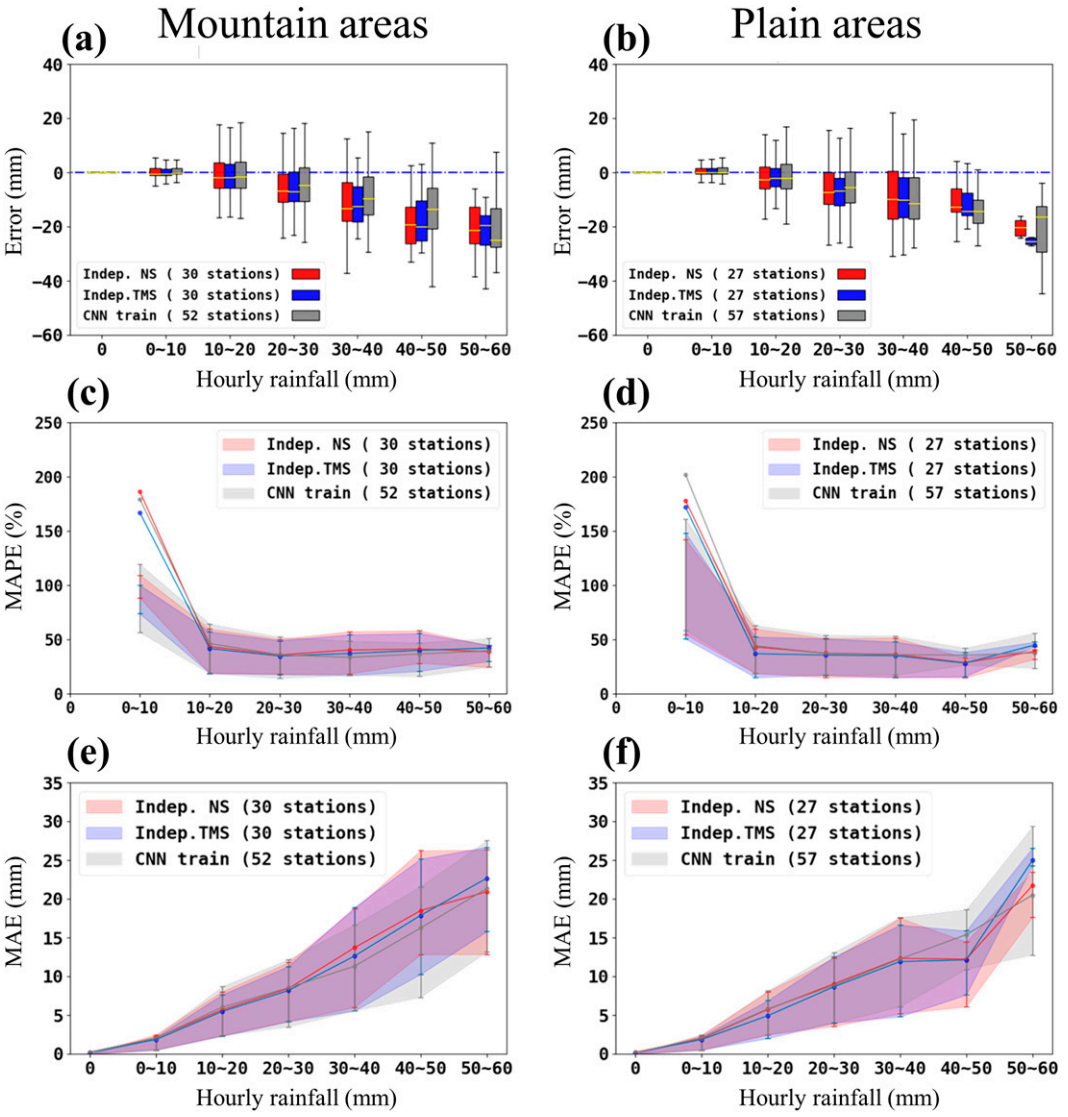

FIG. 11. As in Fig. 9, the performance of rainfall estimation based on the testing and independent verification datasets are compared. The nearest-station method (red), the three-stations-blending method (blue), and original CNN estimates based on the testing dataset (gray) are displayed.

classification over 20 mm) in the calculation of the CNN loss function to increase the weights of heavy rainfall amount (≥20 mm h^{-1}). Verification results based on the testing data from CNN training stations (109 sites) show that the CNN QPE outperforms the $R(Z)$ relation, especially in the mountain area. For example, our proposed QPE shows obvious improvement in the Hualien area, where radar beam blockage is a serious problem.

Moreover, this study proposes two methods (i.e., the nearest-station method and the three-stations blending method) to incorporate the CNN cluster (109 CNN models) for generating 2D rain maps. It is worth noting that an independent verification dataset (i.e., 2018 data from 57 independent stations) is used to confirm the accuracy of these gridded QPE products. It is suggested that the 2D QPE is practical for operational use, provides rainfall in the no-gauge area, and solves the underestimation of the $R(Z)$ relation in the mountains. In addition, two case studies of 13 June and 8 October 2018 are presented to visualize the results. The CNN model generates better features of small-scale weather systems and more accurate precipitation information and overcomes the problem of traditional rainfall interpolation, which cannot provide the flow-dependent influence radius. In summary, the deep learning QPE proposed in this study effectively deals with 3D radar data and learns the storm's vertical structure from CFADs, thus stably estimating the precipitation in complex terrain. Although transitioning deep learning models into operations is a tricky business, and better predictions do not always lead to better decisions, we expect the proposed method may be helpful for disaster prevention of small-scale floods in the future.

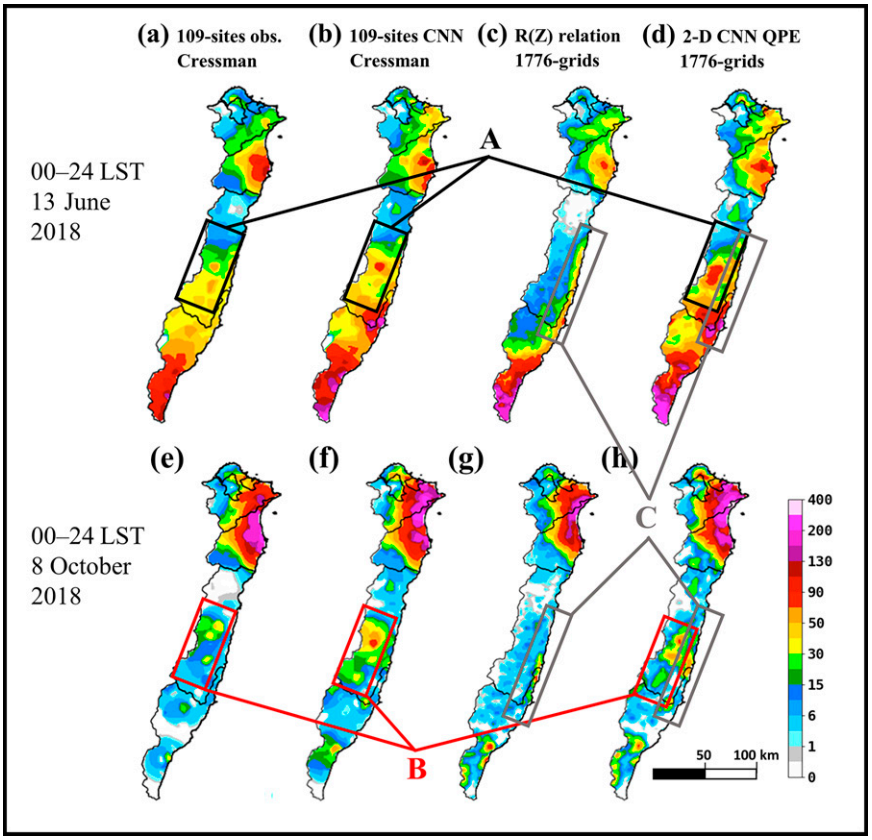

FIG. 12. Two case studies of the daily rainfall (mm) over the study area: (top) 13 Jun 2018 and (bottom) 8 Oct 2018. (a),(e) the Cressman interpolation of the 109 rain gauge observation at training stations ("109-sites obs. Cressman"); (b),(f) the Cressman interpolation of the QPEs by the CNN cluster of the 109 sites ("109-sites CNN Cressman"); (c),(g) the $R(Z)$ relation by using a total of 1776 grid points' radar reflectivities ["$R(Z)$ relation 1776-grids"]; and (d),(h) the nearest-station method by using a total of 1776 grid points' radar reflectivities ("2D CNN QPE 1776-grids").

b. *Limitations of the proposed method and future direction*

Although the CNN QPE improves the underestimation of the current operational QPE in complex terrain, the proposed CNN-QPE technique still has the limitation that negative biases of extreme rainfall over 30 mm h^{-1} exist (Figs. 9a,b, negative bias with MAPE $\approx$ 40%). Future works of using deep learning for QPE should address this shortcoming. The reason for underestimation is presumably because CFADs only reflect the overall structure of the convective system and somehow smooth out the extreme Z values. In addition, our proposed method may have a similar problem to conventional QPE of difficulty adjusting the weights to be suitable for various weather systems.

Three S-band meteorological radars in Taiwan will be upgraded to dual-polarized radars in 2022, and other C-band dual-polarized radars are under construction. Future works may apply this study's idea for improving the CNN cluster's performance by using more input information, such as K_{dp} and Z_{DR}. Furthermore, other meteorological information could be considered in the deep learning framework, including model analysis of subsynoptic weather conditions, the vertical profile of temperature and moisture, and the flow-lifting index on the terrain slope. These approaches could be made with the aid of specific deep learning techniques to integrate heterogeneous data and are worth trying to improve the accuracy and practicability of deep learning–based QPE techniques.

Acknowledgments. Computing resources for this study were mainly provided by the Center for Weather Climate and Disaster Research, National Taiwan University. This work was funded by Grants 109-2625-M-002-021, 111-2111-M-002-016, and 111-2123-M-002-014 of the Ministry of Science and Technology, Taiwan; Project 1102056E of Central Weather Bureau.

Data availability statement. The QPESUMS raw data and rainfall observation in this study were provided by CWB Weather Forecast Center. Due to confidentiality agreements, supporting data can only be made available to bona fide researchers subject to a nondisclosure agreement. Details of the

data and how to request access are available from Dr. Treng-Shi Huang at CWB.

REFERENCES

Albawi, S., T. A. Mohammed, and S. Al-Zawi, 2017: Understanding of a convolutional neural network. *2017 Int. Conf. on Engineering and Technology (ICET)*, Antalya, Turkey, Institute of Electrical and Electronics Engineers, 1–6, https://doi.org/10.1109/ICEngTechnol.2017.8308186.

Bartier, P. M., and C. P. Keller, 1996: Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Comput. Geosci.*, **22**, 795–799, https://doi.org/10.1016/0098-3004(96)00021-0.

Becker, T., B. Stevens, and C. Hohenegger, 2017: Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation. *J. Adv. Model. Earth Syst.*, **9**, 1488–1505, https://doi.org/10.1002/2016MS000865.

Brandes, E. A., G. Zhang, and J. Vivekanandan, 2003: An evaluation of a drop distribution-based polarimetric radar rainfall estimator. *J. Appl. Meteor. Climatol.*, **42**, 652–660, https://doi.org/10.1175/1520-0450(2003)042<0652:AEOADD>2.0.CO;2.

Bringi, V. N., G.-J. Huang, V. Chandrasekar, and E. Gorgucci, 2002: A methodology for estimating the parameters of a gamma raindrop size distribution model from polarimetric radar data: Application to a squall-line event from the TRMM/Brazil campaign. *J. Atmos. Oceanic Technol.*, **19**, 633–645, https://doi.org/10.1175/1520-0426(2002)019<0633:AMFETP>2.0.CO;2.

Chang, P.-L., and Coauthors, 2021: An operational multi-radar multi-sensor QPE system in Taiwan. *Bull. Amer. Meteor. Soc.*, **102**, E555–E577, https://doi.org/10.1175/BAMS-D-20-0043.1.

Chen, B., B.-F. Chen, and H.-T. Lin, 2018: Rotation-blended CNNs on a new open dataset for tropical cyclone image-to-intensity regression. *Proc. 24th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, New York, NY, Association for Computing Machinery, 90–99, https://doi.org/10.1145/3219819.3219926.

Chen, B.-F., B. Chen, H.-T. Lin, and R. L. Elsberry, 2019: Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks. *Wea. Forecasting*, **34**, 447–465, https://doi.org/10.1175/WAF-D-18-0136.1.

Chung, K.-S., and I.-A. Yao, 2020: Improving radar echo Lagrangian extrapolation nowcasting by blending numerical model wind information: Statistical performance of 16 typhoon cases. *Mon. Wea. Rev.*, **148**, 1099–1120, https://doi.org/10.1175/MWR-D-19-0193.1.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374, https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**, 5742–5751, https://doi.org/10.1029/2018GL078202.

Glorot, X., A. Bordes, and Y. Bengio, 2011: Deep sparse rectifier neural networks. *Proc. 14th Int. Conf. on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, PMLR, 315–323, https://proceedings.mlr.press/v15/glorot11a.html.

Gourley, J. J., R. A. Maddox, K. W. Howard, and D. W. Burgess, 2002: An exploratory multisensor technique for quantitative estimation of stratiform rainfall. *J. Hydrometeor.*, **3**, 166–180, https://doi.org/10.1175/1525-7541(2002)003<0166:AEMTFQ>2.0.CO;2.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros, 2017: Image-to-image translation with conditional adversarial networks. *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Institute of Electrical and Electronics Engineers, 5967–5976, https://ieeexplore.ieee.org/document/8100115.

Jain, A. K., J. Mao, and K. M. Mohiuddin, 1996: Artificial neural networks: A tutorial. *Computer*, **29**, 31–44, https://doi.org/10.1109/2.485891.

Jou, J. D., C. J. Jung, and R.G. Hsiu, 2015: Quantitative precipitation estimation using S-band polarimetric radars in Taiwan Meiyu Season (in Chinese with English abstract). *Atmos. Sci.*, **43**, 91–113, http://mopl.as.ntu.edu.tw/web/ASJ/43/43-2-1.pdf.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012: Imagenet classification with deep convolutional neural networks. *Proc. 25th Int. Conf. on Neural Information Processing Systems (NIPS'12)*, Lake Tahoe, NV, Association for Computing Machinery, 1097–1105.

Lagerquist, R., A. McGovern, C. R. Homeyer, D. J. Gagne, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, https://doi.org/10.1175/MWR-D-19-0372.1.

Nair, V., and G. E. Hinton, 2010: Rectified linear units improve restricted Boltzmann machines. *Proc. 27th Int. Conf. on Machine Learning (ICML)*, Haifa, Israel, Association for Computing Machinery, 807–814, https://dl.acm.org/doi/10.5555/3104322.3104425.

Racah, E., C. Beckham, T. Maharaj, S. E. Kahou, Prabhat, and C. Pal, 2017: Extreme weather: A large-scale climate dataset for semisupervised detection, localization, and understanding of extreme weather events. *Proc. 31st Int. Conf. on Neural Information Processing Systems (NIPS'17)*, Vol. 30, Long Beach, CA, Association for Computing Machinery, 3405–3416, https://dl.acm.org/doi/10.5555/3294996.3295099.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven earth system science. *Nature*, **566**, 195–204, https://doi.org/10.1038/s41586-019-0912-1.

Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, P. L. Heinselman, S. E. Giangrande, and D. S. Zrnic, 2005a: The joint polarization experiment: Polarimetric rainfall measurements and hydrometeor classification. *Bull. Amer. Meteor. Soc.*, **86**, 809–824, https://doi.org/10.1175/BAMS-86-6-809.

——, S. E. Giangrande, and T. J. Schuur, 2005b: Rainfall estimation with a polarimetric prototype of WSR-88D. *J. Appl. Meteor. Climatol.*, **44**, 502–515, https://doi.org/10.1175/JAM2213.1.

——, M. Diederich, P. Zhang, and C. Simmer, 2014: Potential utilization of specific attenuation for rainfall estimation, mitigation of partial beam blockage, and radar networking. *J. Atmos. Oceanic Technol.*, **31**, 599–619, https://doi.org/10.1175/JTECH-D-13-00038.1.

Sachidananda, M., and D. S. Zrnić, 1987: Rain rate estimates from differential polarization measurements. *J. Atmos. Oceanic Technol.*, **4**, 588–598, https://doi.org/10.1175/1520-0426(1987)004<0588:RREFDP>2.0.CO;2.

Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Proc. 28th Int. Conf. on Advances in Neural Information Processing Systems (NIPS'15)*, Vol. 28, Montreal, Canada, Association for

Computing Machinery, 802–810, https://dl.acm.org/doi/10.5555/2969239.2969329.

——, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, 2017: Deep learning for precipitation nowcasting: A benchmark and a new model. *Proc. 31st Int. Conf. on Neural Information Processing Systems (NIPS'17)*, Vol. 30, Long Beach, California, Association for Computing Machinery, 5622–5632, https://dl.acm.org/doi/10.5555/3295222.3295313.

Sønderby, C. K., and Coauthors, 2020: MetNet: A neural weather model for precipitation forecasting. arXiv, 2003.12140v2, https://doi.org/10.48550/ARXIV.2003.12140.

Storer, R. L., S. C. van den Heever, and T. S. L'Ecuyer, 2014: Observations of aerosol-induced convective invigoration in the tropical East Atlantic. *J. Geophys. Res. Atmos.*, **119**, 3963–3975, https://doi.org/10.1002/2013JD020272.

Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2017: DeepSD: Generating high resolution climate change projections through single image super-resolution. *Proc. 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'17)*, Halifax, Nova Scotia, Association for Computing Machinery, 1663–1672, https://dl.acm.org/doi/10.1145/3097983.3098004.

Wu, C.-C., T.-H. Yen, Y.-H. Huang, C.-K. Yu, and S.-G. Chen, 2016: Statistical characteristic of heavy rainfall associated with typhoons near Taiwan based on high-density automatic rain gauge data. *Bull. Amer. Meteor. Soc.*, **97**, 1363–1375, https://doi.org/10.1175/BAMS-D-15-00076.1.

Xin, L., G. Recuter, and B. Larochelle, 1997: Reflectivity-rain rate relationship for convective rainshowers in Edmonton: Research note. *Atmos.–Ocean*, **35**, 513–521, https://doi.org/10.1080/07055900.1997.9649602.

Yuter, S. E., and R.-A. Houze Jr., 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and differential reflectivity. *Mon. Wea. Rev.*, **123**, 1941–1963, https://doi.org/10.1175/1520-0493(1995)123<1941:TDKAME>2.0.CO;2.

Zhang, G., J. Vivekanandan, and E. Brandes, 2001: A method for estimating rain rate and drop size distribution from polarimetric radar measurements. *IEEE Trans. Geosci. Remote Sens.*, **39**, 830–841, https://doi.org/10.1109/36.917906.

Zrnić, D. S., T. D. Keenan, L. D. Carey, and P. May, 2000: Sensitivity analysis of polarimetric variables at a 5-cm wavelength in rain. *J. Appl. Meteor. Climatol.*, **39**, 1514–1526, https://doi.org/10.1175/1520-0450(2000)039<1514:SAOPVA>2.0.CO;2.