Deep learning for bias correction of satellite retrievals of orographic precipitation

Haonan Chen, *Senior Member, IEEE*, Luyao Sun, Robert Cifelli, Pingping Xie

***Abstract*—The performance of various composite satellite precipitation products is severely limited by their individual passive microwave (PMW)-based retrieval uncertainties since the PMW sensors have difficulties in resolving heavy rain and/or shallow orographic precipitation systems. Characterizing the error structure of PMW retrievals is crucial to improving precipitation mapping at different space-time scales. To this end, this paper introduces a machine learning framework to quantify the uncertainties associated with satellite precipitation products with an emphasis on orographic precipitation. A deep convolutional neural network (CNN) is designed, which utilizes the ground-based Stage IV precipitation estimates as target labels in the training phase, to reduce biases involved in the precipitation product derived using the NOAA/Climate Prediction Center morphing technique (CMORPH). The products before and after bias correction are evaluated using four independent precipitation events over the coastal mountain region in the western United States, and the impact of topography on satellite-based precipitation retrievals is quantified. Experimental results show that the orographic gradients have a strong impact on precipitation retrievals in complex terrain regions. The accuracy of CMORPH is dramatically enhanced after applying the proposed machine learning-based bias correction technique. Using Stage IV data as references, the overall correlation (CC), normalized mean error (NME), and normalized mean absolute error (NMAE) of CMORPH are improved from 0.55, 32%, 63%, to 0.88, -2%, 39%, respectively, after bias correction for the independent case studies presented in this article. Such a machine learning scheme also has great potential for improved fusion of other or future satellite precipitation retrievals.**

***Index Terms*—Satellite, ground radar, precipitation estimation, complex terrain, deep learning, bias correction**

I. Introduction

SPACE-based precipitation products (SPPs) are of vital importance in regional, continental, and global water, weather, and climate research. The high-resolution SPPs with low latency also serve as critical forcing of hydrological models for streamflow and flood predictions, particularly in regions that lack weather radars and rain gauges. Compared with ground-based weather radars and/or rain gauges that may offer better estimates of precipitation at regional scales, the satellite sensors have great advantages in providing seamless spatial coverage over the globe, especially over the oceans and other remote regions.

A large number of geostationary (GEO) and low earth orbit (LEO) satellites have been launched for hydrometeorological applications. Based on the GEO satellite infrared (IR) data or passive microwave (PMW) measurements from LEO satellite sensors, different precipitation retrieval algorithms have been developed (e.g., [1][2][3]). Through fusion of precipitation estimates from multiple satellites, a number of quasi-global SPPs have been developed in the precipitation research community, including the precipitation estimation from remotely sensed information using artificial neural network (PERSIANN) and its improved versions [4][5], the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) [6], the Global Satellite Mapping of Precipitation (GSMaP) [7], the Climate Prediction Center (CPC) MORPHing technique (CMORPH) [8] [9], and the recent Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) [10]. These SPPs are produced at different temporal and spatial resolutions, and are commonly used for natural disaster monitoring and situational awareness worldwide.

However, the accuracy of SPPs is severely limited by the spatiotemporal sampling of individual PMW and/or IR sensors, as well as the implemented precipitation retrieval algorithms. Although the PMW and IR sensors can do a reasonable job at resolving moderate rain, both struggle with retrieving precipitation during light or heavy rain events. In addition, due to complex land-ocean interactions and orographic enhancement, none of the SPPs performs well in complex terrain regions such as the coastal mountain areas over the western United States [11][12][13].

This article aims to characterize the error structure of SPPs over complex terrain regions using deep learning techniques. In particular, this study uses the CMORPH product for demonstration purposes, which is derived through combining IR data and precipitation retrievals from multiple PMW sensors [8][9]. Therein, the high temporal resolution IR imageries are used to create motion vectors of the cloud systems and the derived cloud motion vectors are then used to interpolate and propagate the PMW-based retrievals to produce complete precipitation estimates at a global scale [8]. In recent years, artificial intelligence and machine learning are increasingly being used in geosciences and remote sensing applications, including precipitation research. For example, Tao et al. presented a deep learning approach to extract useful features from bispectral satellite information, IR, and water vapor channels to identify rain/non-rain regions [14]. Chen et al. developed a

This research was supported by the NOAA JPSS Proving Ground and Risk Reduction (PGRR) program and the NOAA Physical Sciences Laboratory through Grant NA19OAR4320073. (*Corresponding author: Haonan Chen.*)

H. Chen and L. Sun are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA (e-mail: haonan.chen@colostate.edu)

R. Cifelli is with the NOAA Physical Sciences Laboratory, Boulder, CO 80305, USA.

P. Xie is with the NOAA Climate Prediction Center, College Park, MD 20740, USA.

novel two-stage hybrid neural network for rainfall estimation using ground-based and spaceborne precipitation radars [15]. In [16], a multi-layer perceptron model is described for improving precipitation estimates through fusion of IR, PMW, and ground observations in an urban environment. In [17], Lee et al. presented a proof-of-concept using machine learning methods to extract physical properties of convective clouds and automate the detection process using the observations from the advanced baseline imager onboard the Geostationary Operational Environmental Satellite-16 (formerly known as GOES-R). In addition to the precipitation detection and estimation problems, deep learning is often applied for short-term prediction of precipitation (e.g., [18] [19]) and enhancement of weather model forecasts (e.g., [20]).

This paper introduces a deep convolutional neural network (CNN) for bias correction of CMORPH retrievals of coastal orographic precipitation, assuming the ground-based Stage IV multi-sensor quantitative precipitation estimates as references in training the model. Using terrain gradients as auxiliary information, this nonparametric CNN approach can resolve the complex orographic enhancement in precipitation processes that is often missed in many SPPs. The remainder of this article is organized as follows. Section II details the study domain, data sets, and the deep learning technique for bias correction of CMORPH. The case studies and evaluation results are presented in Section III. Section IV provides a thorough discussion on the performance of the proposed bias correction methodology, as well as potential extension of the machine learning model. Section V concludes the study and suggests directions for future research.

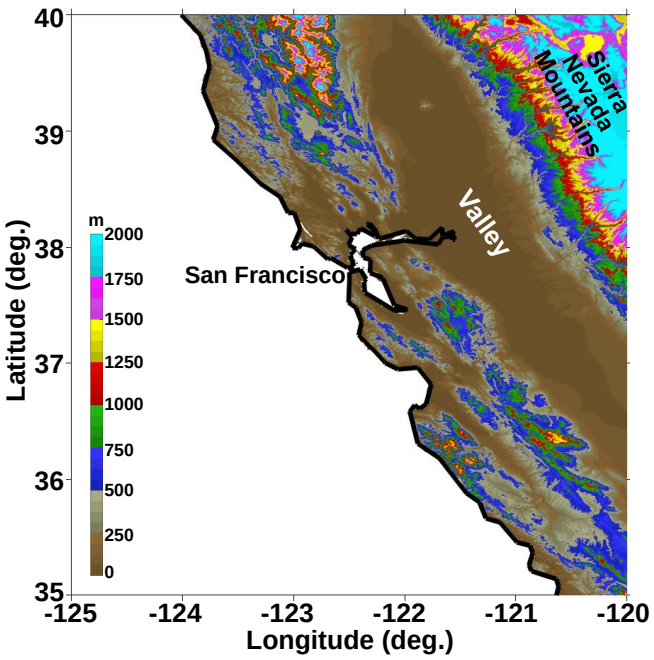

Fig. 1. The digital elevation model (DEM) information of the study domain. The thick black line indicates coastal zone.

II. Study Domain, Datasets, and Methodology

A. Study Domain

In this article, the selected study domain is located in the west coast of the United States, covering 35°N-40°N latitude and 120°W-125°W longitude. The area of this study domain is about 500 km by 500 km, including the whole San Francisco Bay Area and part of the Sierra Nevada mountain range. Figure 1 illustrates the digital elevation model (DEM) information of this region. Here, it should be noted that the current study primarily focuses on the land region (i.e., precipitation over the ocean is not considered).

The study domain has a Mediterranean-type climate, with precipitation coming in the winter months and a pronounced dry season extending roughly from April-October [21][22][23]. Every year, this region experiences a few atmospheric river (AR) events during wet season, which are long and narrow regions of intense water vapor transport that can produce floods, mudslides, and debris flows [23][24]. Due to the unique topography and their exposure to AR, many watersheds in this area are prone to flooding. Accurate quantitative precipitation estimation (QPE) is critical for water managers to balance the competing needs for water supply and flood mitigation [25][26].

Unfortunately, it is very difficult to obtain reliable QPE in such a complex terrain region due to the rapid changes of precipitation characteristics as a result of orographic enhancement [22][27][28]. Previous studies have shown that the performance of various SPPs in this region is very poor [12]. The precipitation intensities reported by different SPPs differ sometimes by an order of magnitude [11]. Hence, the objective of this study is to resolve and mitigate the biases associated with SPPs, with an emphasis on the impact of orographic enhancement on precipitation retrievals.

B. Datasets

As mentioned, this study uses CMORPH products to demonstrate the bias correction technique. The CMORPH products are created by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC), which takes advantage of precipitation observations and retrievals from a large number of GEO and LEO satellites (for details, see [8][9]). Briefly speaking, CMORPH uses precipitation estimates that have been derived from LEO satellite PMW measurements exclusively. Cloud motion vectors are generated by computing spatial lag correlations on successive GEO satellite IR images, and are then used to propagate the PMW-based precipitation retrievals. In this way, for regions and time frames that have no PMW retrievals available, CMORPH can still provide complete precipitation mapping through motion vector-based interpolation of available PMW retrievals.

The CMORPH products are produced at global scale with a spatial resolution of 8 km (at the equator) and temporal resolution of 30 min. The products are available from 1998 to the current day. In addition, a bias corrected version (i.e., CMORPH-CRT) is available, which is utilized in this study. The bias correction in producing CMORPH-CRT is based on probability density function (PDF) matching against the

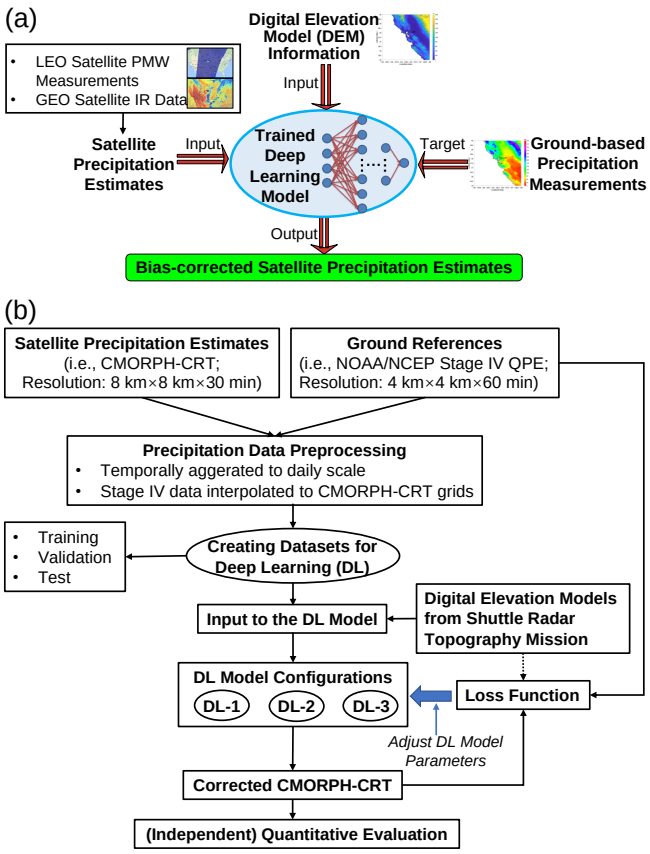

Fig. 2. (a) Conceptual diagram of the machine learning system for bias correction of satellite retrievals of orographic precipitation. Essentially, the system is trained to produce enhanced precipitation estimates using ground-based precipitation measurements as training labels. (b) Flowchart of the bias correction scheme designed for NOAA CMORPH product, which utilizes Stage IV QPE as reference and topography information as a key input feature.

CPC daily gauge analysis over land (0.25° lat/lon grids) and the Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (2.5° lat/lon grids) over ocean [9]. Hereafter, the terms of CMORPH and CMORPH-CRT will be used interchangeably in this article.

In this study, 10 months of CMORPH-CRT data are considered, including all January and February events from 2016 to 2019, as well as March of 2018 and 2019. The data are obtained from ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/8km-30min/. Therein, four precipitation events, namely, 9-12 January 2017, 21-23 March 2018, 13-16 February 2019, and 25-28 February 2019, are excluded in training the deep learning-based bias correction model detailed in Section II-C. These four events are used for independent test of the bias correction model performance.

To train the deep learning model for bias correction of CMORPH-CRT, the National Centers for Environmental Precipitation (NCEP) Stage IV QPE data are used as target labels in the training phase and are assumed as ground truth in the test phase. The Stage IV analysis is based on multi-sensor (radar and gauges) hourly and 6-hr QPE data (i.e., Stage III data) produced by the National Weather Service (NWS) River Forecast Centers (RFCs) [29][30], which are mosaicked into a national product at the NCEP Environmental Modeling Center (EMC) during the Stage IV analysis. The Stage IV QPE, which is often used to verify and evaluate various SPPs, is produced on local 4 km polar-stereographic grids. The products are available at https://data.eol.ucar.edu/cgi-bin/codiac/fgr_form/id=21.093. In this study, the Stage IV QPE at hourly scale is used. In order to match the CMORPH-CRT grid resolution, the Stage IV QPE data are spatially resampled using the nearest-neighbor interpolation method.

It should be mentioned that the gauge data used in generating Stage IV QPE are independent from the CPC gauge analysis in producing CMORPH-CRT. In addition, although the CMORPH-CRT products are produced every 30 min, they essentially represent the precipitation rates for an instantaneous observing time within a half-hour window. In order to eliminate the random error and uncertainty caused by time mismatching, both CMORPH-CRT and Stage IV QPE are aggregated to daily scale in the bias correction analysis.

To quantify the impact of orographic enhancement on satellite precipitation estimation, the digital elevation model (DEM) information is also taken into account as the input feature of the bias correction model. The DEM information is obtained from the shuttle radar topography mission: http://srtm.csi.cgiar.org/download. The high-resolution (30-m) terrain data are spatially averaged at each CMORPH-CRT grid pixel. That is, the uncertainties associated with finer scale (<8 km) terrain features are neglected in this study.

C. Methodology

Figure 2 shows the conceptual diagram of the deep learning (DL) framework for bias correction of CMORPH in complex terrain regions. Roughly speaking, there are four components, including (1) preprocessing of CMORPH, ground-based Stage IV QPE, and corresponding terrain information; (2) construction of the training, validation, and test datasets for the DL models; (3) DL model training and parameter estimation; (4) performance evaluation based on independent test data. In the following, more details of these four components are provided.

As mentioned, both CMORPH, Stage IV QPE, and DEM data are spatially reprocessed to the CMORPH grid pixels. Temporally, both the CMORPH and Stage IV QPE are integrated to daily scale for training and testing the DL model. For the 10-month data selected in this study, the total number of precipitation samples (daily, 8-km grid) is 90,448, excluding the four test events. The precipitation samples are randomly split into training and validation datasets: 90% of the samples are used for training, and the rest (10%) are used as validation data to fine tune the DL model parameters. Three of the four test events, including 9-12 January 2017, 13-16 February 2019, and 25-28 February 2019, have 15,376 precipitation samples for each case. The 21-23 March 2018 test event has 11,532 sample points due to the shorter duration of this event. In addition, the input features are normalized so that the values of the input data are between 0-1 before being sent to the DL model. The normalization will also speed up the learning process. Here, it is worth noting that the corresponding labels (i.e., Stage IV QPE) are not normalized.

The core part of the DL model is a convolutional neural network (CNN), which has proven to be effective in precipitation

retrieval and short-term prediction [19][31][32][33]. Figure 3 illustrates the detailed CNN structure designed for CMORPH bias correction, which includes three convolution layers and three pooling layers followed by two fully connected layers. The first layer is the "input layer" for ingesting input data, and the last layer is the "output layer" for producing the learning results that will be compared with the labels. The convolutional and pooling layers will perform feature extraction and transmission, which are the essential components of a CNN [19][34]. It should be noted that this study has tested a large number of hyperparameters including different patch sizes and different numbers of convolutional and pooling layers, using a grid search approach [35]. The structure in Fig. 3 is selected mainly because it renders desirable results without making the network too deep.

In order to quantify the impact of terrain feature on the bias correction performance, the DL model is configured in three different ways, namely, DL-1, DL-2, and DL-3. These three configurations share the same CNN structure, but different input and loss functions. Overall, the requirement on terrain feature is gradually increased in these three configurations, and the relationship between terrain and the corrected CMORPH is gradually enhanced. In particular, in DL-1, only CMORPH precipitation data are used as input and the loss function (Loss1) is solely determined by the precipitation estimates. DL-2 utilizes both CMORPH precipitation data and DEM information as input, and uses the same loss function as DL-1. In contrast, DL-3 uses terrain and precipitation information in both input and the loss function (Loss1+Loss2).

As shown in Eq. 1, the loss function is a customized mean square error (MSE) with different weights, including precipitation (i.e., Loss1) and terrain constraint (i.e., Loss2).

$$\begin{cases} \mathrm{DL}-1 : \mathrm{Loss1} \\ \mathrm{DL}-2 : \mathrm{Loss1} \\ \mathrm{DL}-3 : \mathrm{Loss1}+\mathrm{Loss2} \end{cases} \tag{1a}$$

$$\mathrm{Loss1} = \mathrm{W1} \cdot \frac{1}{N}\sum_{i=1}^{N}\left(R_S - R_G\right)^2 \tag{1b}$$

$$\mathrm{Loss2} = \mathrm{W2} \cdot \frac{1}{N}\sum_{i=1}^{N}\left(R_S - \mathrm{DEM}/25\right)^2 \tag{1c}$$

$$\mathrm{W1} = \begin{cases} 2 & 0 < R_G \le 20 \\ 5 & 20 < R_G \le 40 \\ 10 & 40 < R_G \le 60 \\ 20 & 60 < R_G \le 80 \\ 30 & R_G > 80 \end{cases} \tag{1d}$$

$$\mathrm{W2} = \begin{cases} 2 & 0 < \mathrm{DEM} \le 500 \\ 3 & 500 < \mathrm{DEM} \le 1000 \\ 4 & 1000 < \mathrm{DEM} \le 1500 \\ 5 & 1500 < \mathrm{DEM} \le 2000 \\ 6 & \mathrm{DEM} > 2000 \end{cases} \tag{1e}$$

where N is the sample number; R_S (mm) and R_G (mm) are the corrected CMORPH precipitation from the DL model and corresponding ground-based Stage IV QPE, respectively; W1 and W2 are the weights determined by the daily precipitation amount (mm) and the terrain height (m), respectively. In this study, the statistical distributions of daily precipitation from CMORPH over the study domain are extensively investigated to determine the weights W1 and W2 (larger weight is assigned to heavier precipitation). In addition, DEM/25 is used in Loss2 mainly because the terrain height in this domain is 0-2000 (m). After dividing by 25, the terrain height value is close to the range of daily precipitation amount (i.e., 0-80 mm), although the units are different.

In the training phase, this study selects a 70×70 grid area to cover the domain illustrated in Fig. 1. Considering the spatial correlation of precipitation, a sliding window (patch = 9) is applied when correcting CMORPH at each grid point. To achieve this configuration, 9 grid pixels near the left/right and top/bottom edges are not used in constructing the training pairs (i.e., not corrected), although they are incorporated in correcting relatively interior grid points. The training sample pairs of data and labels are then shuffled, and 1024 samples are fed to the network each time. That is, the batch size is 1024.

For DL-1, the number of input channels is 1 (CMORPH data only), and the dimension of the input data is (batch size = 1024, channel number = 1, patch size = 9×9). For DL-2 and DL-3, there are two channels since the terrain feature is stacked with the precipitation data, and the dimension of the input data is (batch size = 1024, channel number = 2, patch size = 9×9). The Stage IV QPE at the center point of the 9×9 sliding window is used as the target label.

As shown in Fig. 3, the filter/kernel size for the three convolutional layers is 5×5, 3×3, 3×3, respectively. The number of filters in each convolutional layer is 96, 128, and 192, respectively. The number of filters and filter sizes for the (Maxpool) pooling layers are the same with the convolutional layers. The activation function adopted in this study is ReLU, and the learning rate is 0.001. When the last pooling is completed, the feature map dimension becomes 192×1×1, which will be flattened out and connected to the two fully connected layers. Finally, a 1×1 value is obtained, which is the bias-corrected CMORPH for a single grid point of the patch center, and the corrected precipitation data is then compared with corresponding label. After training the model for different configurations, the bias-corrected CMORPH can readily be obtained for any new CMORPH products.

III. Results and Evaluation

A. Application of the Trained DL Model

After training the DL model with different configurations, original CMORPH products for the four independent test events are corrected using the trained models. In addition, the corrected daily precipitation products are aggregated to event totals in order to further highlight the practical performance of the bias correction model. For illustration purposes, Fig. 4 shows the precipitation amounts during the 9-12 January 2017 event, including CMORPH before and after bias correction using DL-1, DL-2, and DL-3. The corresponding Stage IV QPE product is also included in Fig. 4 for comparison.

Comparing the original CMORPH product (Fig. 4a) and Stage IV QPE (Fig. 4b), it can be seen that obvious underestimation of CMORPH occurs in most parts of this region. The

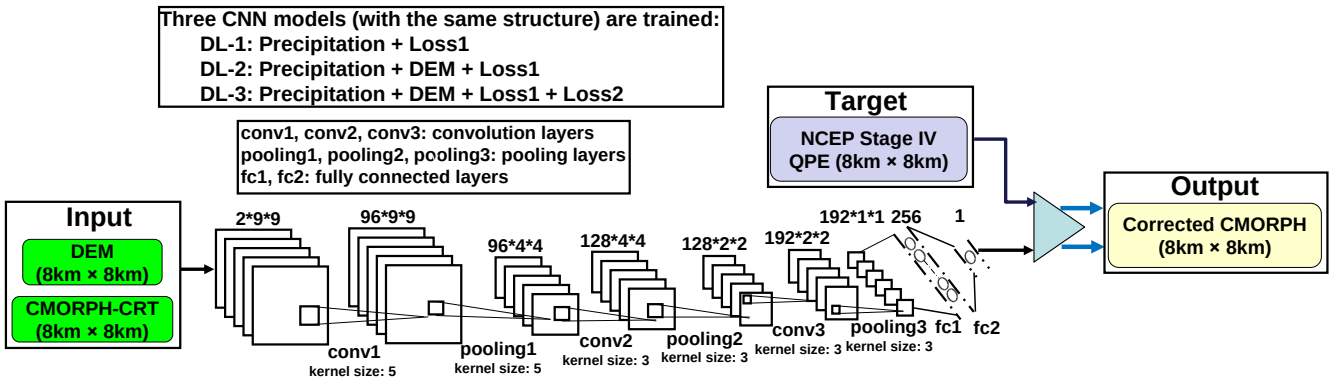

Fig. 3. The detailed structure of the CNN model used for bias correction of satellite precipitation product.

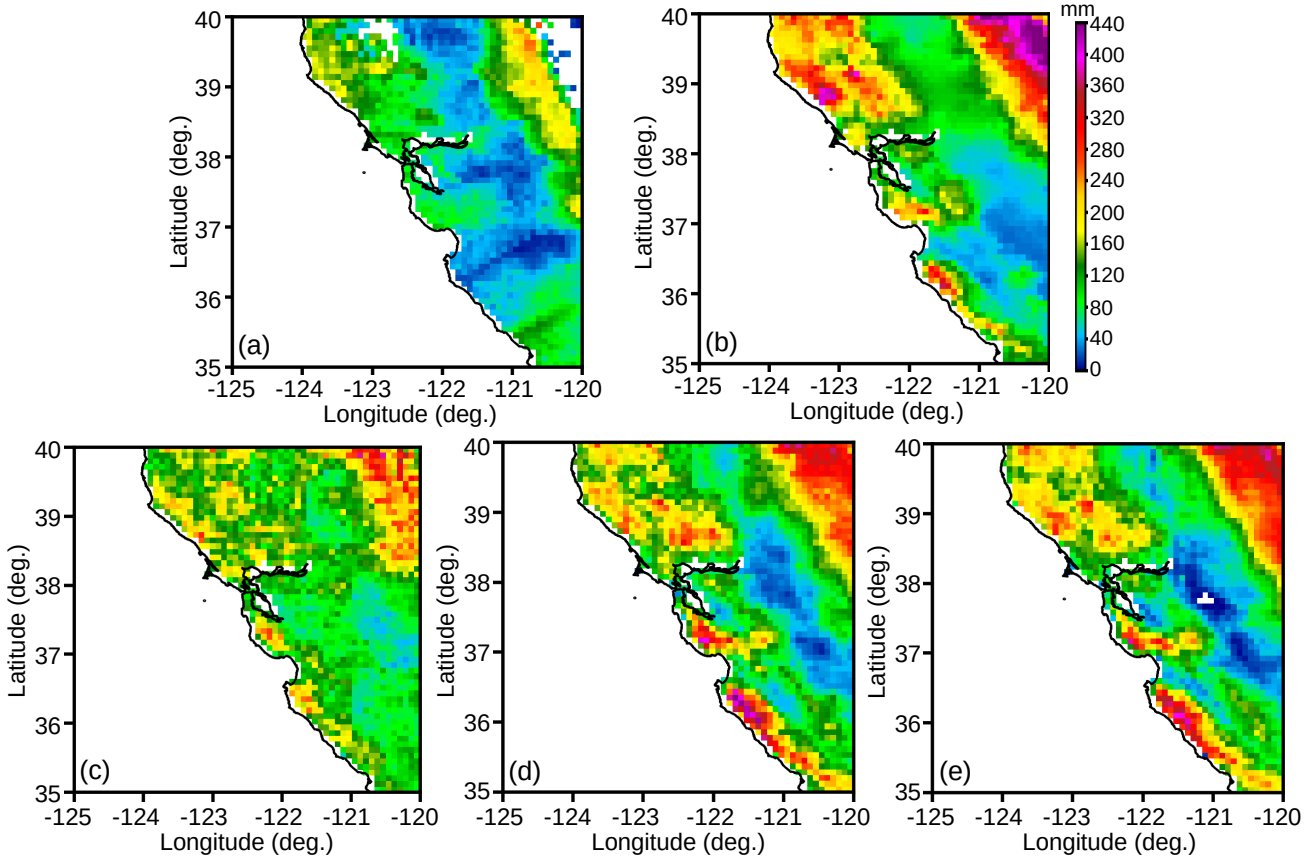

Fig. 4. Precipitation amounts during the 9-12 January 2017 event: (a) original CMORPH-CRT; (b) Stage IV QPE (i.e., reference); (c)-(e) are bias corrected CMORPH-CRT using the trained DL-1, DL-2, and DL-3, respectively.

pattern of CMORPH precipitation estimates does not show a clear orographic gradient as the Stage IV QPE reveals (see also the terrain map in Fig. 1). In addition, CMORPH does not provide any valid estimates in some of the high elevation areas due to snow contamination on the PMW-based retrievals. In contrast, Stage IV QPE shows a significant amount of precipitation in the mountain regions. This is also indicated by the scatter plot of CMORPH versus Stage IV QPE in Fig. 5a, where CMORPH shows zero values when Stage IV QPE reported significant precipitation accumulations.

Scrutinizing the bias corrected CMORPH in Fig. 4c-e, we can conclude that all of the methods are able to recover the missed precipitation over the mountain regions since the CNN-based DL model is incorporating precipitation information from surrounding grid pixels (patch size is 9×9) when correcting the CMORPH grid pixel at the center of the patch. Nevertheless, compared to the Stage IV QPE, the corrected product based on DL-1 shows significant underestimation of precipitation over the mountain regions. It is also apparent that the CMORPH product is over corrected in the valley region, resulting in overestimation of precipitation. This is mainly due to the lack of terrain information in the DL-1 configuration.

After including the terrain information in the bias correction model, the corrected CMORPH products from DL-2 and DL-3

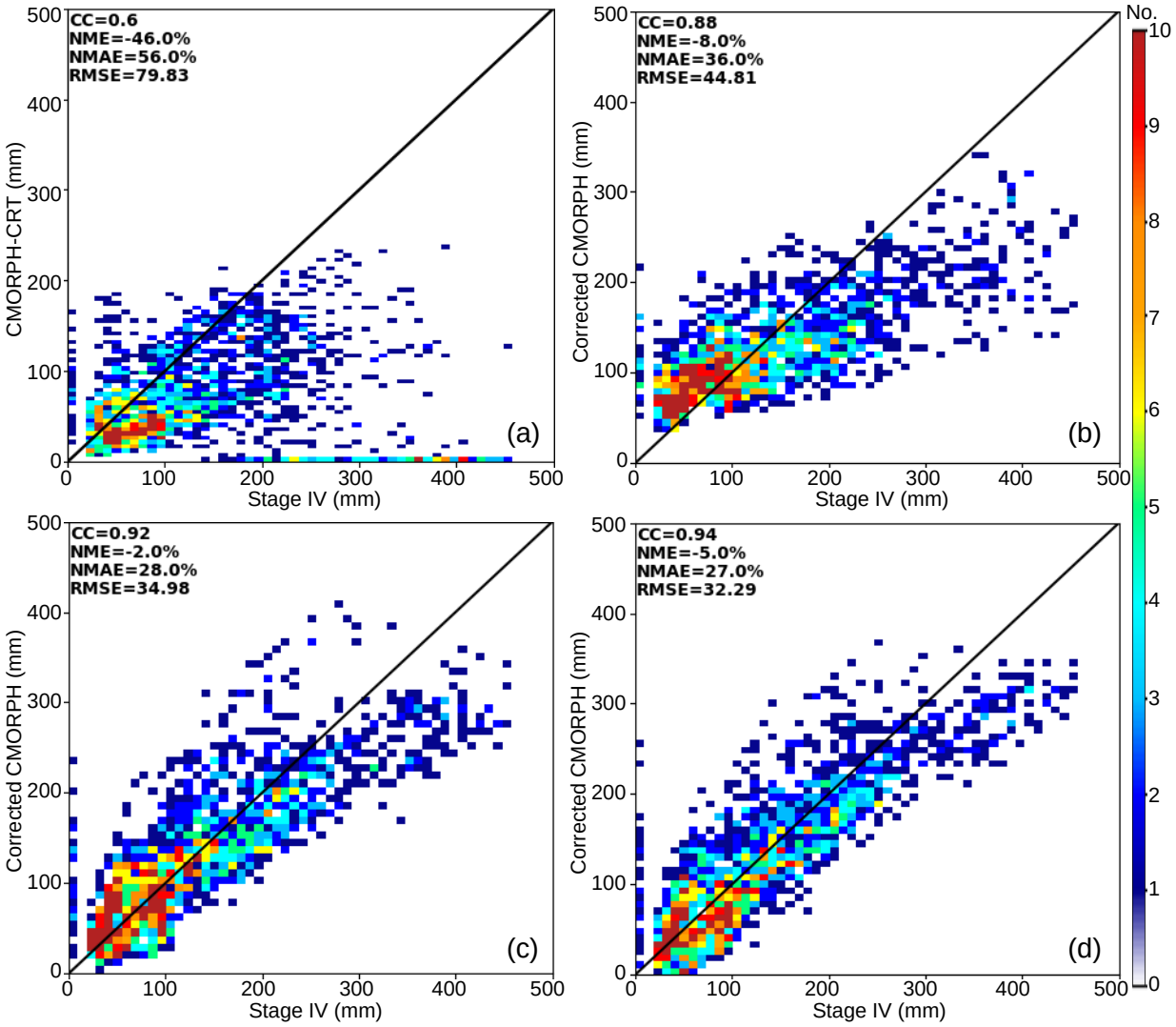

Fig. 5. Scatter plots of CMORPH versus Stage IV QPE during the precipitation event illustrated in Fig. 4 (9-12 January 2017): (a) original CMORPH-CRT versus Stage IV QPE; (b)-(d) are bias corrected CMORPH-CRT using DL-1, DL-2, and DL-3 models, respectively, versus Stage IV QPE.

show much more improvement than that based on DL-1. Both Fig. 4d and Fig. 4e agree with the Stage IV QPE very well in terms of precipitation accumulation and pattern. Not only the underestimation over high elevation regions is alleviated, but also the over-correction in the valley is addressed to a large extent. These findings are further demonstrated by the scatter plots in Fig. 5b-d. In addition, it is observed that the difference between the corrected products based on DL-2 and DL-3 is not significant, which implies that inclusion of additional terrain constraint in the loss function will probably not change the results dramatically as long as the terrain information is included as an input feature. But it is worth noting that DL-3 does have a slightly better performance than DL-2 in terms of the overall agreement with Stage IV QPE as indicated in Fig. 5c-d.

In order to further compare the improvement brought by DL-1, DL-2, and DL-3, Fig. 6 shows the probability density functions (PDFs) of precipitation during this verification event. The PDFs indicate that the original CMORPH is overestimating light rain but underestimating heavy rain; the intersection of PDFs of CMORPH and Stage IV QPE is around 100 mm (event total) for this particular event. After applying the DL model, PDFs of the three corrected CMOPRH products are closer to that of Stage IV QPE, especially the corrected products based on DL-2 and DL-3.

Bias corrected CMORPH for other three test events are not shown here since they essentially exhibit the same trends in performance. The PDF structures are similar to those shown in Fig. 6. This is not a surprise given that precipitation characteristics such as the distribution patterns are rather similar during different (AR) events [23] [36][37]. For the sake of illustration, Fig. 7 shows the scattergrams of CMORPH versus Stage IV QPE for all the four validation events combined. Overall, the terrain information has a strong impact on the bias correction performance, and the orographic enhancement of precipitation can be learned very well by DL-2 and DL-3. The distribution of corrected values using DL-2 and DL-3 agree with Stage IV QPE much better than that based on DL-1, which does not include terrain information. Compared to DL-2, the results from DL-3 are more concentrated along the 1:1 line, which

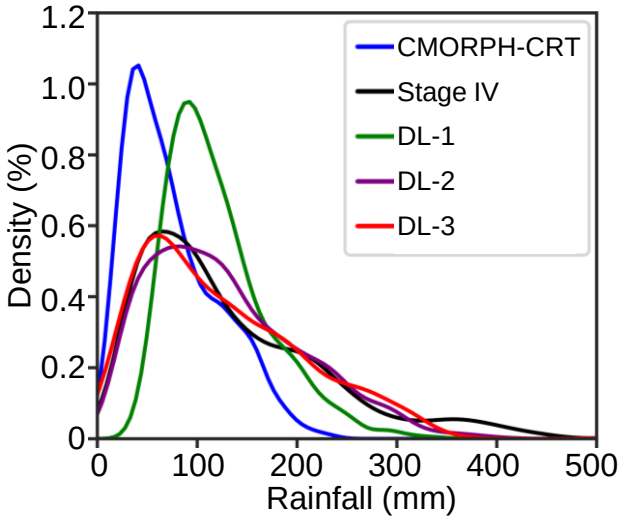

Fig. 6. Probability density functions (PDFs) of precipitation estimates during the 9-12 January 2017 event (event totals, see also Fig. 4). Different colors indicate different products, including Stage IV QPE and CMORPH before and after bias correction.

further demonstrates the positive influence of adding a terrain-based constraint in the loss function.

Nevertheless, Figs. 5 and 7 also show that all the CMORPH products, even after bias correction, are underestimating the precipitation peaks that occur in the Sierra Nevada. Since this high elevation region is often characterized by mixed-phase precipitation, which is ignored in the PMW-based retrievals ingested by CMORPH, the DL model uses terrain and surrounding precipitation information to predict actual precipitation in this case. Although the terrain feature is considered in DL-2 and DL-3, the convolution and pooling process is likely smoothing out the precipitation peaks since surrounding precipitation intensity is noticeably lower.

B. Quantitative Evaluation of the DL-based Bias Correction Performance

In order to quantify the bias correction performance, the following metrics are computed, including the Pearson correlation coefficient (CC), normalized mean error (NME), normalized mean absolute error (NMAE), and root-mean-squared error (RMSE), which are respectively defined as:

$$\mathrm{CC} = \frac{\sum[(R_S - < R_S >)(R_G - < R_G >)]}{\sqrt{\sum(R_S - < R_S >)^2}\sqrt{\sum(R_G - < R_G >)^2}} \quad (2a)$$

$$\mathrm{NME} = \frac{< R_S - R_G >}{< R_G >} \quad (2b)$$

$$\mathrm{NMAE} = \frac{< |R_S - R_G| >}{< R_G >} \quad (2c)$$

$$\mathrm{RMSE} = \sqrt{< (R_S - R_G)^2 >} \quad (2d)$$

where R_S is satellite precipitation estimate from CMORPH; R_G is ground-based Stage IV QPE; and the angle brackets stand for sample average. In addition, a threshold of 0 mm (at daily scale) is applied in computing these scores. That is, only precipitating grid pixels (both CMORPH and Stage IV products) are considered. In this way, the blank regions with snow contamination in the original CMORPH product are ignored for fair comparison.

The quantitative evaluation scores for the 9-12 January 2017 event and the four test events combined are indicated in Fig. 5 and Fig. 7, respectively. For clarity, Fig. 8 summarizes the evaluation results for all the individual events, as well as the overall scores combining all the samples from the four test events. Obviously, the CC scores of CMORPH for all the four events are improved after bias correction. Compared to the original CMORPH, the improvement is more and more significant from DL-1 to DL-2 and then to DL-3. In particular, the overall CC score of original CMORPH combining all the test events is 0.55. After applying the DL-1, Dl-2, and DL-3 models, the CC score is improved to 0.77, 0.85, and 0.88, respectively.

In terms of NME, the original CMORPH products are underestimating precipitation during all the four test events (i.e., NME scores are negative). After applying the three trained models, the bias corrected products during the 9-12 January 2017 event still show some underestimation. In contrast, the corrected products generally exhibit overestimation during other three test events except the DL-2 based product for 21-23 March 2018 and the DL-3 based product for 25-28 February 2019. The underestimation of the DL-3 based product for 25-28 February 2019 is negligible (NME = -2%). It is also noticed that in general the magnitude of NME scores is getting lower after bias correction, indicating the effectiveness of the DL-based approach. The product based on DL-1 during the 21-23 March 2018 event has the highest NME score, likely due to the over correction of CMORPH in the valley regions. In addition, the NME scores for the corrected products based on DL-2 and DL-3 models are lower than those based on DL-1 model. Nevertheless, it should be noted that although NME is a good overall indicator of underestimation or overestimation, there is a potential cancellation of overestimation and underestimation in calculating NME for a large amount of samples.

Figure 8c shows the NMAE scores, which do not involve cancellation of overestimation and underestimation. Clearly, the bias corrected products based on DL-2 and DL-3 have much better performance than the original CMORPH or the corrected products based on DL-1. The DL-3 model has slightly better performance than the DL-2 model in terms of the NMAE. For the four events combined, the overall NMAE score of original CMORPH is 63%. After applying the DL-1, Dl-2, and DL-3 models, the NMAE score decreases to 61%, 44% 39%, respectively. This dramatic improvement is further demonstrated by the RMSE scores illustrated in Fig. 8d. After bias correction using DL-3, the RMSE of event totals for the four events combined reduced from 62.1 mm (RMSE of original CMORPH) to 33.5 mm.

IV. Discussion

The bias corrected products and quantitative evaluation results have demonstrated that machine learning is very effective in capturing the error structure of satellite retrievals of orographic precipitation. In particular, after incorporating the

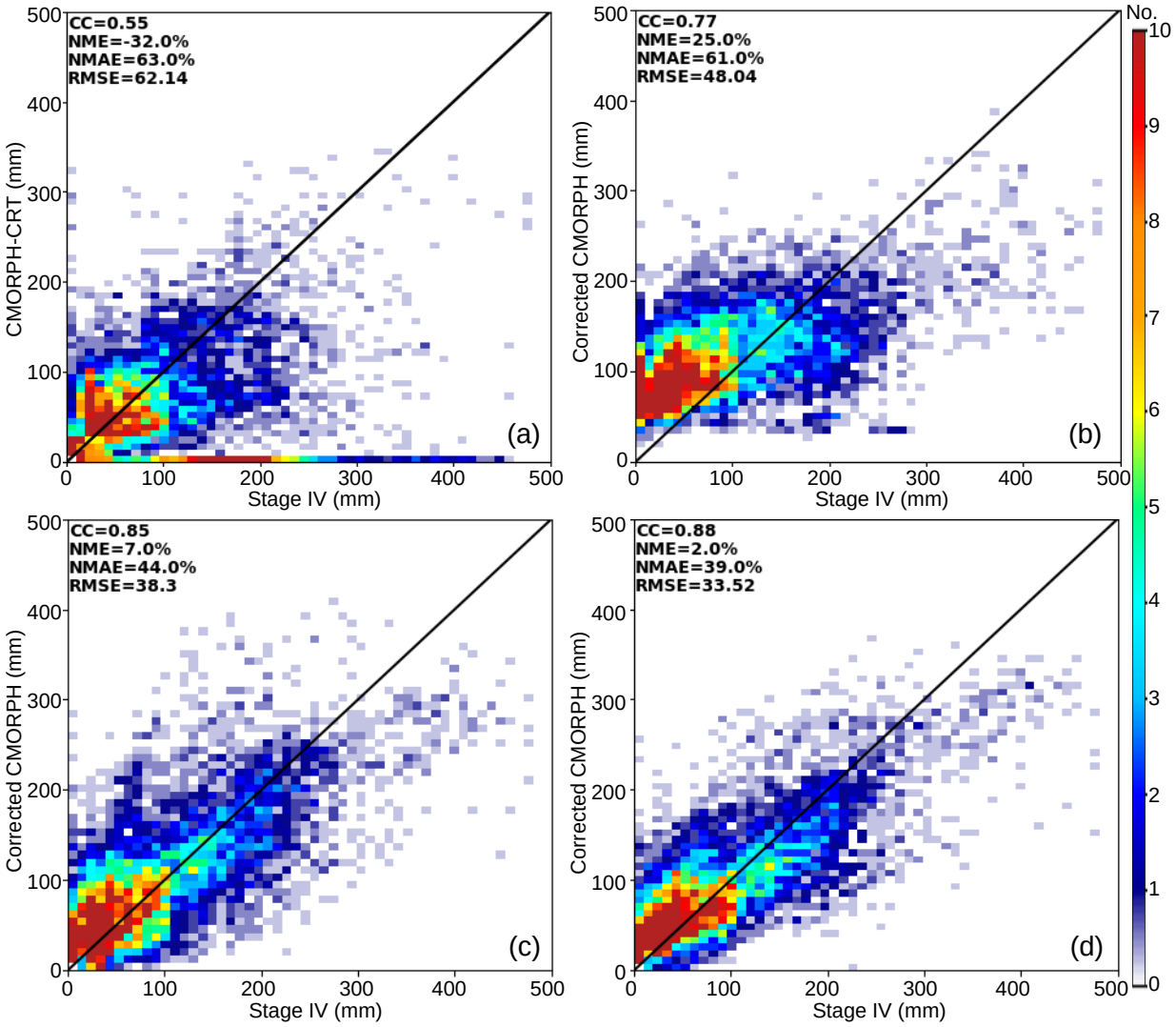

Fig. 7. Scatter plots of CMORPH versus Stage IV QPE during the four independent validation events: (a) original CMORPH-CRT versus Stage IV QPE; (b)-(d) are bias corrected CMORPH-CRT using DL-1, DL-2, and DL-3 models, respectively, versus Stage IV QPE.

terrain feature as an additional input in the DL model, the bias correction performance is dramatically improved. This implies that the terrain elevation information plays a critical role in learning the orographic enhancement of precipitation. The performance can be further enhanced if a terrain constraint is imposed in the loss function of the machine learning model. Such a constraint (i.e., physics-guided loss) also makes the bias correction model fall under the concept of physics-guided artificial intelligence.

Although significant progress has been made in improving CMORPH products over the western U.S., a few relevant issues should be considered in general applications of the proposed bias correction technique. First of all, the precipitation peaks are underestimated, even after bias correction. This is mainly due to the implementation of convolutional and pooling processes across a large area in order to predict the missed precipitation over the mountains (e.g., blank regions in Fig. 4a). Unfortunately, that is where the precipitation peaks occur, and there is no efficient or better way to handle this issue in the current bias correction model framework. A possible solution is to improve individual PMW-based retrievals before merging them into a composite product. Therein, a precipitation classification scheme could be incorporated to support mixed phase precipitation retrievals. In this way, both the earlier stage PMW-based retrievals (hence the composite product such as CMORPH) and the bias corrected versions should be enhanced, especially in capturing precipitation currently missed in CMORPH.

In addition, this article only focuses on the complex terrain regions in Northern California over the Western U.S. However, CMOPRH is a global product. Although numerous studies have been performed to quantify uncertainty of CMORPH at global scale, almost all the studies are conducted at coarse resolution (0.25-degree lat/lon grids or even lower resolution). A major concern in resolving the error structure of CMORPH at global scale at its native 8-km resolution is the lack of high-fidelity ground precipitation references. As such, extra attention should be paid when applying the bias correction model in other domains, especially where orographic enhancement is a main driver of precipitation. Our recommendation

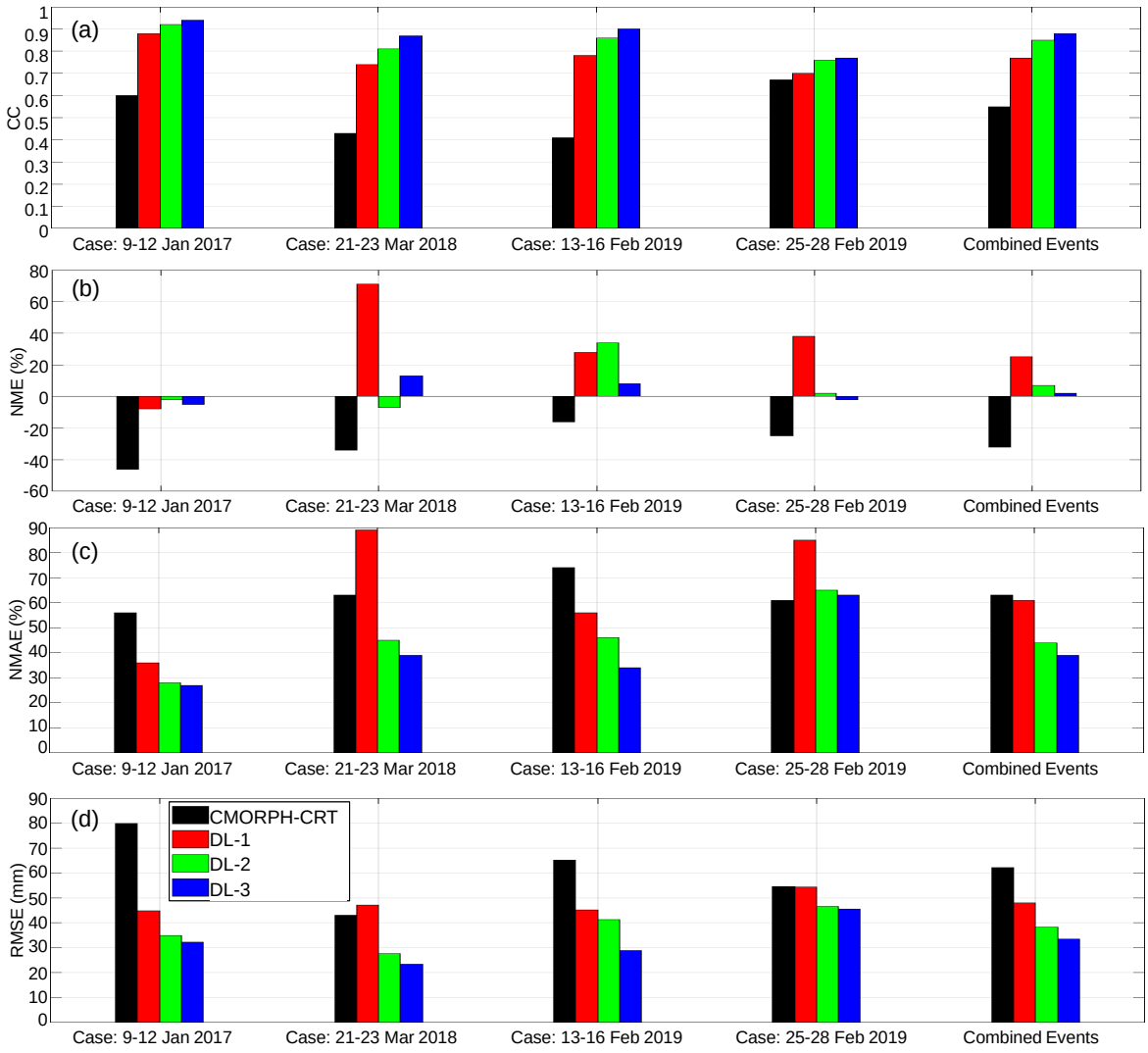

Fig. 8. Quantitative evaluation results of CMORPH before and after bias correction for the four independent validation events and four-event combined product: (a) CC; (b) NME; (c) NMAE; (d) RMSE.

is that if reliable ground-based precipitation products such as the Stage IV QPE used in this study are available, the DL model should be retrained with local terrain and ground precipitation references. At the very least, the fully-connected layers in the CNN model should be fine tuned with a small amount of local data (also refer to the transfer learning concept in [19][38]). However, if there is no ground precipitation reference available, it is suggested that one can either retrain the model with data from a similar climate regime or directly use the model trained in this article. Future work will quantify the general applicability of this DL model and extend it to different domains or global scale.

We also want to note again that different precipitation events have fairly similar characteristics in the selected study domain. Overall, this region is dominated by shallow stratiform rain in the wet seasons [39]. Although the instantaneous precipitation rate is not high, a precipitation event may last a few days, resulting in significant amount of rain, especially over the mountain areas. Hence, this article only uses a number of typical precipitation cases in training the DL models, and the 10-month training data can represent local precipitation features well. When applying the proposed technique in other regions, it is entirely possible that more data will be required in training the DL models. In addition, adding other features such as temperature and water vapor imagery may further enhance the DL-based bias correction model, especially for correcting CMORPH products for small scale, rapidly changing precipitation events.

V. SUMMARY

Precipitation estimation using satellite measurements over complex terrain regions remains a formidable challenge due to the orographic enhancement of precipitation, especially during shallow precipitation events. As part of our continuing effort to improve QPE over the mountain ranges of the western United States, this study has developed a deep learning framework to characterize and correct the biases associated with satellite retrievals of orographic precipitation. The essential component of the framework is a convolutional neural network (CNN), which has been trained with three different configurations (i.e.,

DL-1, DL-2, and DL-3) in order to quantify the impact of terrain features on precipitation retrievals. In short, DL-1 only uses the satellite precipitation products as input, and the loss is defined purely based on precipitation estimates. In DL-2 and DL-3, both the satellite precipitation products and terrain elevations are used as input features. DL-2 uses the same loss as DL-1, whereas DL-3 has an additional terrain-based constraint in the defined loss function.

A demonstration study has been performed using the NOAA CMORPH product in Northern California as an example. The ground-based Stage IV QPE products are utilized as references in training and evaluating the deep learning models. In particular, 10-month data from 2016 to 2019 were used in this analysis, among which four precipitation events were excluded from the total dataset and were used for independent model verification. Primary conclusions are summarized as follows:

1) All the three trained DL-based bias correction schemes can improve the performance of CMORPH product. Not only the precipitating areas detected by the original CMORPH were enhanced, but also the missed precipitating regions over high elevation regions were recovered since the CNN was able to learn the spatial correlation of precipitation and provide effective estimates for the missing grid points using surrounding precipitation information.

2) Before incorporating the terrain feature in the DL model, the overall CC score of the event totals combining all the samples from the four test cases was improved from 0.55 (original CMORPH) to 0.77 (corrected CMORPH based on DL-1). The NMAE score was improved from 63% to 61%, and the RMSE score was improved from 62.1 mm to 48.0 mm.

3) After including the terrain feature in the DL model, the improvement is dramatic, indicating the significant impact of orographic enhancement on precipitation formation and subsequent retrievals. The overall CC score of the bias corrected CMORPH using DL-2 and DL-3 was improved to 0.85 and 0.88, respectively. The NMAE of the corrected products using DL-2 and DL-3 decreased to 44% and 39%, respectively. The RMSE dropped to 38.3 mm and 33.5 mm, respectively.

4) The only difference between DL-2 and DL-3 was the inclusion of a terrain based precipitation constraint in the loss function of DL-3. Although this would not dramatically change the results, the corrected product based on DL-3 indeed showed improved performance over that based on DL-2. To some degree, this implies the importance of physics-guided machine learning rather than purely data-driven models.

5) Since the terrain information is currently not used in the operational PMW- and/or IR-based precipitation retrieval algorithms (e.g., [1] [2][40][41]), it is suggested that the orographic precipitation gradient should be considered an additional input or a constraint in the development of future algorithms. Also, machine learning has a great potential to enhance the current parametric retrieval algorithms since machine learning is more effective in extracting information from multi-source multi-dimensional data, which should be taken into account in future algorithm development.

In addition, the DL-based bias correction framework is designed with high flexibility. It can easily be adapted to process other SPPs such as IMERG product. Besides the SPPs and terrain features, the model can be extended to incorporate other features such as wind, temperature, and water vapor imagery. The bias correction model can also be applied to other regions and at different time scales. Nevertheless, detailed demonstration is yet to be done. Future work will focus on further improving the satellite precipitation retrievals from all these aspects. Generalization capability of this DL-based approach, especially in regions where the training labels are not available, should also be investigated.

ACKNOWLEDGMENT

The authors would like to thank Janice L. Bytheway (NOAA/PSL and CU/CIRES) and the anonymous external reviewers for providing careful review and comments on this article. The extensive discussion with Ying Lin (NOAA/NWS/EMC) on the Stage IV QPE data is also acknowledged.

REFERENCES

[1] G. Vicente, R. Scofield, and W. Menzel, "The operational GOES infrared rainfall estimation technique," *Bull. Amer. Meteor. Soc.*, vol. 79, pp. 1883–1898, 1998.

[2] C. Kummerow, Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, and T. T. Wilheit, "The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors," *Journal of Applied Meteorology*, vol. 40, no. 11, pp. 1801–1820, 2001.

[3] R. R. Ferraro, F. Weng, N. C. Grody, L. Zhao, H. Meng, C. Kongoli, P. Pellegrino, S. Qiu, and C. Dean, "NOAA operational hydrological products derived from the advanced microwave sounding unit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 5, pp. 1036–1049, 2005.

[4] K. lin Hsu, X. Gao, S. Sorooshian, and H. V. Gupta, "Precipitation estimation from remotely sensed information using artificial neural networks," *Journal of Applied Meteorology*, vol. 36, no. 9, pp. 1176 – 1190, 1997.

[5] P. Nguyen, M. Ombadi, V. A. Gorooh, E. J. Shearer, M. Sadeghi, S. Sorooshian, K. Hsu, D. Bolvin, and M. F. Ralph, "Persiann dynamic infraredâ€"rain rate (pdir-now): A near-real-time, quasi-global satellite precipitation dataset," *Journal of Hydrometeorology*, vol. 21, no. 12, pp. 2893 – 2906, 2020.

[6] G. Huffman, D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, and Coauthors, "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *J. Hydrometeor.*, vol. 8, pp. 38–55, 2007.

[7] T. Ushio, K. Sasashige, T. Kubota, and Coauthors, "A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data," *J. Meteor. Soc. Japan*, vol. 87A, pp. 137–151, 2009.

[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. 287–296, 2004.

[9] P. Xie, R. Joyce, S. Wu, S. H. Yoo, Y. Yarosh, F. Sun, and R. Lin, "Reprocessed, bias-corrected cmorph global high-resolution precipitation estimates from 1998," *Journal of Hydrometeorology*, pp. JHM–D–16–0168.1, 2017.

[10] G. J. Huffman, D. T. Bolvin, D. Braithwaite, K. Hsu, and Coauthors, "NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG)," *Algorithm Theoretical Basis Document (ATBD), Version 5.2*, pp. Greenbelt, MD, 35 pp., 2018.

[11] J. L. Bytheway, M. Hughes, K. Mahoney, and R. Cifelli, "On the uncertainty of high-resolution hourly quantitative precipitation estimates in california," *Journal of Hydrometeorology*, vol. 21, no. 5, pp. 865 – 879, 2020.

[12] Y. Wen, A. Behrangi, H. Chen, and B. Lambrigtsen, "How well were the early 2017 california atmospheric river precipitation events captured by satellite products and ground-based radars?" *Quarterly Journal of the Royal Meteorological Society*, vol. 144, no. S1, pp. 344–359, 2018.

[13] Y. Derin, E. Anagnostou, A. Berne, M. Borga, B. Boudevillain, W. Buytaert, C.-H. Chang, H. Chen, G. Delrieu, Y. Hsu, and et al., "Evaluation of gpm-era global satellite precipitation products over multiple complex terrain regions," *Remote Sensing*, vol. 11, no. 24, p. 2936, Dec 2019.

[14] 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.

[15] H. Chen, V. Chandrasekar, H. Tan, and R. Cifelli, "Rainfall estimation from ground radar and trmm precip-itation radar using hybrid deep neural networks," *Geophysical Research Letters*, vol. 46, p. 10669–10678, 2019.

[16] H. Chen, V. Chandrasekar, R. Cifelli, and P. Xie, "A machine learning system for precipitation estimation using satellite and ground radar network observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 982–994, 2020.

[17] Y. Lee, C. D. Kummerow, and I. Ebert-Uphoff, "Applying machine learning methods to detect convection using goes-16 abi data," *Atmospheric Measurement Techniques Discussions*, vol. 2020, pp. 1–28, 2020.

[18] X. Shi, Z. Chen, H. Wang, and et al., "Convolutional lstm network: A machine learning approach for precip-itation nowcasting," *in Proc. Adv. Neural Inf. Process. Syst.*, p. 802–810, 2015.

[19] L. Han, Y. Zhao, H. Chen, and V. Chandrasekar, "Advancing radar nowcasting through deep transfer learning," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–9, 2021.

[20] L. Han, M. Chen, K. Chen, H. Chen, Y. Zhang, B. Lu, L. Song, and R. Qin, "A deep learning method for bias correction of ecmwf 24-240h forecasts," *Adv. Atmos. Sci.*, vol. 38, no. 9, pp. 1444–1459, 2021.

[21] A. P. Williams, J. T. Abatzoglou, A. Gershunov, J. Guzman-Morales, D. A. Bishop, J. K. Balch, and D. P. Lettenmaier, "Observed impacts of anthropogenic climate change on wildfire in california," *Earth's Future*, vol. 7, no. 8, pp. 892–910, 2019.

[22] H. Chen, R. Cifelli, and A. White, "Improving operational radar rainfall estimates using profiler observations over complex terrain in northern california," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1821–1832, 2020.

[23] A. B. White, B. J. Moore, D. J. Gottas, and P. J. Neiman, "Winter storm conditions leading to excessive runoff above california's oroville dam during january and february 2017," *Bull. Amer. Meteor. Soc.*, vol. 100, pp. 55–70, 2019.

[24] F. M. Ralph, F. Cannon, V. Tallapragada, C. A. Davis, J. D. Doyle, F. Pappenberger, A. Subramanian, A. M. Wilson, D. A. Lavers, C. A. Reynolds, J. S. Haase, L. Centurioni, B. Ingleby, J. J. Rutz, J. M. Cordeira, M. Zheng, C. Hecht, B. Kawzenuk, and L. D. Monache, "West coast forecast challenges and development of atmospheric river reconnaissance," *Bulletin of the American Meteorological Society*, vol. 101, no. 8, pp. E1357 – E1377, 2020.

[25] Y. Ma, V. Chandrasekar, H. Chen, and R. Cifelli, "Quantifying the potential of aqpi gap-filling radar network for streamflow simulation through a wrf-hydro experiment," *Journal of Hydrometeorology*, 2021.

[26] R. Cifelli, L. E. Johnson, J. Kim, T. Coleman, G. Pratt, L. Herdman, R. Martyr-Koller, J. A. FinziHart, L. Erikson, P. Barnard, and M. Anderson, "Assessment of flood forecast products for a coupled tributary-coastal model," *Water*, vol. 13, no. 3, 2021.

[27] R. Cifelli, V. Chandrasekar, H. Chen, and L. E. Johnson, "High resolution radar quantitative precipitation estimation in the san francisco bay area: Rainfall monitoring for the urban environment," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 96A, pp. 141–155, 2018.

[28] H. Chen, R. Cifelli, V. Chandrasekar, and Y. Ma, "A flexible bayesian approach to bias correction of radar-derived precipitation estimates over complex terrain: Model design and initial verification," *Journal of Hydrometeorology*, vol. 20, no. 12, pp. 2367 – 2382, 2019.

[29] D.-J. Seo and J. P. Breidenbach, "Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements," *Journal of Hydrometeorology*, vol. 3, no. 2, pp. 93 – 111, 2002.

[30] B. R. Nelson, O. P. Prat, D.-J. Seo, and E. Habib, "Assessment and implications of ncep stage iv quantitative precipitation estimates for product intercomparisons," *Weather and Forecasting*, vol. 31, no. 2, pp. 371 – 394, 2016.

[31] M. Sadeghi, A. A. Asanjan, M. Faridzad, P. Nguyen, K. Hsu, S. Sorooshian, and D. Braithwaite, "Persiann-cnn: Precipitation estimation from remotely sensed information using artificial neural networks?convolutional neural networks," *Journal of Hydrometeorology*, vol. 20, no. 12, pp. 2273 – 2289, 2019.

[32] C. Wang, J. Xu, G. Tang, Y. Yang, and Y. Hong, "Infrared precipitation estimation using convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8612–8625, 2020.

[33] L. Han, H. Liang, H. Chen, W. Zhang, and Y. Ge, "Convective precipitation nowcasting using u-net model," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–8, 2022.

[34] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *CoRR*, vol. abs/1411.1792, 2014. [Online]. Available: http://arxiv.org/abs/1411.1792

[35] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization," in *25th Annual Conference on Neural Information Processing Systems (NIPS 2011)*, Granada, Spain, 2011.

[36] H. Chen, R. Cifelli, and V. Chandrasekar, "Resolving the precipitation microphysical variability induced by orographic enhancement in complex terrain over the san francisco bay area," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 5415–5418.

[37] E. Sumargo, A. M. Wilson, F. M. Ralph, R. Weihs, A. White, J. Jasperse, M. Asgari-Lamjiri, S. Turnbull, C. Downer, and L. D. Monache, "The hydrometeorological observation network in california's russian river watershed: Development, characteristics, and key findings from 1997 to 2019," *Bulletin of the American Meteorological Society*, vol. 101, no. 10, pp. E1781 – E1800, 2020.

[38] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[39] A. B. White, P. J. Neiman, F. M. Ralph, D. E. Kingsmill, and P. O. G. Persson, "Coastal orographic rainfall processes observed by radar during the california land-falling jets experiment," *Journal of Hydrometeorology*, vol. 4, no. 2, pp. 264 – 282, 2003.

[40] C. Grassotti, S. Liu, Q. Liu, S.-A. Boukabara, K. Garrett, F. Iturbide-Sanchez, and R. Honeyager, *Precipitation Estimation from the Microwave Integrated Retrieval System (MiRS)*. Cham: Springer International Publishing, 2020, pp. 153–168.

[41] R. A. Scofield and R. J. Kuligowski, "Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events," *Weather and Forecasting*, vol. 18, no. 6, pp. 1037 – 1051, 2003.