Deep Learning for Precipitation Retrievals Using ABI and GLM Measurements on the GOES-R Series

Yifan Yang, Haonan Chen, Kyle Hilburn, Robert J. Kuligowski, Robert Cifelli

***Abstract*—Satellite sensors have been widely used for precipitation retrieval, and a number of precipitation retrieval algorithms have been developed using observations from various satellite sensors. The current operational rainfall rate quantitative precipitation estimate (RRQPE) product from the geostationary operational environmental satellite (GOES) offers full disk rainfall rate estimates based on the observations from the advanced baseline imager (ABI) aboard the GOES-R series. However, accurate precipitation retrieval using satellite sensors is still challenging due to the limitations on spatio-temporal sampling of the satellite sensors and/or the uncertainty associated with the applied parametric retrieval algorithms. In this article, we propose a deep learning framework for precipitation retrieval using the combined observations from the ABI and geostationary lightning mapper (GLM) on the GOES-R series to improve the current operational RRQPE product. Particularly, the proposed deep learning framework is composed of two deep convolutional neural networks (CNNs) that are designed for precipitation detection and quantification. The cloud-top brightness temperature from multiple ABI channels and the lightning flash rate from the GLM measurement are used as inputs to the deep learning framework. To train the designed CNNs, the precipitation product multi-radar multi-sensor (MRMS) system from the National Oceanic and Atmospheric Administration (NOAA) is used as target labels to optimize the network parameters. The experimental results show that the precipitation retrieval performance of the proposed framework is superior to the currently operational GOES RRQPE product in the selected study domain, and the performance is dramatically enhanced after incorporating the lightning data into the deep learning model. Using the independent MRMS product as a reference, the deep learning model can reduce the retrieval uncertainty in the operational RRQPE product by at least 31% in terms of the mean squared error and normalized mean absolute error, and the improvement is more significant in moderate to heavy rain regions. Therefore, the proposed deep learning framework can potentially serve as an alternative approach for GOES precipitation retrievals.**

***Index Terms*—Precipitation retrieval, GOES-R series, deep learning, convolutional neural networks.**

I. Introduction

PRECIPITATION is a critical component in the regional and global water cycles. Accurate precipitation estimation is important for predicting water availability and understanding the risks of flooding and/or drought, especially in the context of climate changes which intensify the Earth's water cycle and increase the frequency of extreme weather events.

Traditionally, precipitation measurement was accomplished by using rain gauges. However, the existing rain gauge network may not be sufficient to capture the precipitation distributions due to its limited coverage and low density, especially over the oceans and other remote regions [1]. Also, rain gauges require regular maintenance to ensure their proper functioning and accuracy. Issues such as gauge blockage, evaporation, wind effects, and measurement errors can affect the quality and reliability of the collected data. The ground-based multi-radar multi-sensor (MRMS) system was operationalized in 2015 by the National Oceanic and Atmospheric Administration (NOAA) [2]. This system primarily integrates about 180 operational weather radars from the continental United States (CONUS) and Canada with the numerical weather prediction model data and rain gauge observations to generate quantitative precipitation estimates. Nevertheless, the performance of precipitation estimates from MRMS is limited in the western United States and some other areas due to the lack of weather radar coverage over complex terrain regions. In addition, the MRMS products are only available over the contiguous United States [2].

In contrast to rain gauges and ground-based weather radars, satellite sensors have wide coverage over the globe. Therefore, various geostationary Earth orbit (GEO) satellites and low Earth orbit (LEO) satellites have been deployed and used for observing clouds and precipitation. Based on the GEO infrared (IR) data or passive microwave (PMW) measurements from LEO satellite sensors, different precipitation retrieval algorithms have been developed (e.g., [3], [4], [5]). Among the LEO and GEO satellites, the latest generation of geostationary operational environmental satellite (GOES) satellites (i.e., GOES-R series) has unique advantages in continuously monitoring precipitation at large scales and at a high temporal resolution. Measurements from the multi-channel advanced baseline imager (ABI) and the geostationary lightning mapper (GLM) aboard the GOES-R series are widely used to investigate various atmospheric phenomena. Therein, the current operational GOES-16 (formerly known as GOES-R) rainfall rate quantitative precipitation estimate (RRQPE) product is derived to assign each earth-navigated pixel a rainfall rate

This research was supported in part by the NOAA JPSS Proving Ground and Risk Reduction (PGRR) program and in part by the National Science Foundation (NSF) CAREER program. (*Corresponding author: Haonan Chen.*)

Yifan Yang, Haonan Chen, and Kyle Hilburn are with Colorado State University, Fort Collins, CO 80523, USA (e-mail: haonan.chen@colostate.edu).

Robert J. Kuligowski is with the NOAA/NESDIS Center for Satellite Applications and Research (STAR), College Park, MD 20740, USA (e-mail: bob.kuligowski@noaa.gov).

Rob Cifelli is with the NOAA Physical Sciences Laboratory, Boulder, CO 80305, USA (e-mail: rob.cifelli@noaa.gov).

[5]. The algorithm in this product includes eight linear and eight non-linear predictors based on the ABI measurements. The algorithm derives the rainfall rates in two steps. The first step is to identify pixels that are experiencing rainfall. After that, it retrieves rainfall rates for pixels where rainfall has been detected. Although the GOES-16 RRQPE product can complement ground-based observations well in many instances, the performance of this precipitation product still needs to be improved, especially in the western United States where orographic precipitation processes are often undetected by the RRQPE and in mesoscale convective systems (MCSs) that frequent the midwestern United States [6][7][8][9]. The currently operational GOES-16 product is calibrated based on passive microwave-derived rainfall rates, which may not fully capture the variability and complexity of the precipitation within the MCS, resulting in underestimation or overestimation of rainfall rates by RRQPE.

Several machine learning-based frameworks for precipitation retrievals using satellite data have been developed and shown to be effective in recent years. For example, the precipitation estimation from remotely sensed information using artificial neural networks cloud classification system (PERSIANN CCS) was designed in [10], which extracted the local and regional cloud features from satellite IR imagery, and transferred the satellite cloud images into pixel rain rates through artificial neural networks. Chen et al. [11] improved satellite-based precipitation retrievals using a deep learning-based data fusion approach, which incorporated the dual-polarization measurements from a high-resolution ground radar network. Upadhyaya et al. [12] described a quantitative precipitation estimation algorithm based on the GOES-16 ABI observations, as well as low-level environmental information and a numerical weather prediction model. However, these previous algorithms are either based on feedforward neural networks, which may not be sufficient to extract the rich information contained in the satellite data. More importantly, none of these studies have incorporated the GLM data, which is expected to provide critical information about heavy convective precipitation.

To this end, this article proposes a deep learning-based framework for improving the operational GOES-16 precipitation product using both ABI and GLM measurements. The essential components of this precipitation retrieval framework are two convolutional neural networks (CNNs), which are widely used in image classification and computer vision problems since they have the advantage of being invariant to scaling and translation, meaning they can recognize and classify objects regardless of their size or scale within an image. [13], [14]. In CNNs, the shared weights among the neurons reduce the number of parameters, and allow the network to extract the more important high-level features of images, such as topology and spatial information. In addition, the pooling operation in CNNs shrinks the size of the feature maps, which overcomes the distortions and translations. In the proposed model, the convolutional layers in CNNs can capture the spatial features of precipitation from the multi-channel satellite observations, including the lightning data from GLM measurements. This article also performs a detailed feature analysis to quantify the impacts of different input data on precipitation retrieval performance at different precipitation intensities.

This paper is organized as follows. In Section II, the study domain in the southeastern United States, the ABI, GLM, and MRMS datasets, and the proposed deep learning framework are detailed. Section III presents an in-depth analysis on the discrimination of the input features for the deep learning model, and the implementation results using the proposed framework over the selected study domain. A discussion about the performance of the proposed framework is also given in this section. Finally, conclusions based on the experimental results and suggested future work are provided in Section IV.

II. Datasets, Study Domain, and Methodology

A. ABI and GLM Measurements and MRMS Datasets

The GOES-16 (current GOES-East) satellite was jointly developed and operated by NOAA and the National Aeronautics and Space Administration (NASA). It covers North and South America and the Atlantic Ocean to the west coast of Africa. In this article, we utilize the ABI and GLM measurements acquired from the GOES-16 satellite as inputs to the deep learning model. To train the model, the ground-based MRMS quantitative precipitation estimation (QPE) product is used for the target labels. The data from May to September in 2019 are divided into training and validation data. The data from May to September in 2020 are used for the independent test. The period of May to September is selected mainly because it is the rainy season in our study domain. Detailed information on each dataset used in this article will be provided in the following.

1) GOES-16 ABI Measurement: The full-disk ABI measurement is an imaging radiometer with 10 minutes temporal resolution, which images Earth's weather, oceans, and environment using 16 spectral bands associated with different center wavelengths and spatial resolutions, including 2 visible, 4 near-infrared, and 10 infrared bands [15]. For the purpose of precipitation retrieval, we select 5 infrared spectral bands 8, 10, 11, 14, and 15, with 6.2 µm, 7.3 µm, 8.4 µm, 11.2 µm, and 12.3 µm center wavelength, respectively. These wavelengths have been shown to be particularly sensitive to water vapor [15]. On the contrary, the near-infrared and visible bands in ABI measurements are less sensitive to water vapor and precipitations compared to these selected five infrared bands. All these five bands have a spatial resolution of 2 km. The radiance data from these selected bands are converted into brightness temperatures (BTs), which are incorporated into the input features of the deep learning model.

Previous studies [16], [17] have demonstrated that the brightness temperature differences (BTDs) derived from the GOES-16 ABI bands can be utilized to obtain precipitation retrieval products, as well as land surface temperature and cloud-top pressure products. For instance, the BTD between 6.2 µm band 8 and 7.3 µm band 10 approximates the concentration and distribution of water vapor [15]. The BTD between 8.4 µm band 11 and 11.2 µm band 14 can differentiate between thick and thin cirrus, and between clouds made of ice and those made of liquid [15]. Consequently, the methodology described herein incorporates five BTD features in addition to the BT features.

2) GOES-16 GLM Measurement: The GLM measurement uses a single-channel, near-infrared optical transient detector to identify and indicate the presence of lightning. It measures the in-cloud, cloud-to-cloud, and cloud-to-ground lightning activities continuously over the Americas and adjacent ocean regions with near-uniform 8 km spatial resolution and 20 seconds refresh rate [18]. In the GLM measurement, the lightning data are ingested and assembled into three categories of high-level products: event, group, and flash, which can be easily and conveniently utilized for scientific research and broader operational applications. For instance, the flash product from the GOES-16 GLM measurement has been employed in [19] to improve convective precipitation forecasting. The GLM measurement has also been found to be valuable for estimating locations of strong radar echoes [20].

In the GOES-16 GLM measurement, an event, which is the basic unit of data from the GLM measurement, is defined as the occurrence of a single pixel exceeding the background threshold during a single frame. A group is defined as multiple simultaneous adjacent events in pixels that occur in the same time integration frame. The reason for defining groups is that a lightning discharge is usually present in more than one pixel during a single time integration frame. The location information for the groups is stored in Earth-based latitude and longitude coordinates. A flash is defined as a set of groups sequentially separated in time by 330 ms or less and in space by no more than 16.5 km (nominally two pixels). Thus, a flash may include one group with a single event or it may consist of many groups that contain many events. The spatial information for flashes is in latitude and longitude to be consistent for the groups and flashes. In this article, we apply the high-level flash product as an additional input feature to our deep learning model in order to identify areas of deep convection with associated heavy rain rates.

3) Ground-Based MRMS Product: The ground-based MRMS product was operationalized in 2015 by NOAA, which integrates multiple radars, satellites, surface observations, upper air observations, lightning reports, rain gauges, and numerical weather prediction models in the continental United States and Canada, and generates quantitative precipitation estimates. The integration of radar with multi-sensor data provides more accurate diagnoses of physical processes in the atmosphere than using radar data alone. Furthermore, a bias correction algorithm was later developed by Ware [2], [21], which takes the density of the gauge network and the distance from the gauge into consideration. It properly weights the radar estimates and the gauge observations at each grid point to improve the quantitative precipitation estimates. The bias correction process for MRMS data can be summarized into the following three stages: a) calculation of hourly radar-gauge differences; b) interpolation of differences onto the MRMS grid via an inverse-distance-weighted mean scheme; c) subtraction of interpolated differences from radar QPE. The bias correction is performed separately for different subdomains within the CONUS region. The presented results by Zhang et al. in [2] demonstrate that this bias correction approach consistently improves the radar QPE accuracy throughout the year across the CONUS domain. This bias-corrected MRMS product has a spatial resolution of 1 km and includes the accumulation results ranging from 1 to 72 hours. In this article, we use the hourly bias-corrected MRMS rainfall estimates as the target labels to train the deep learning model, which has been identified as the most precise precipitation rainfall rate product during the period of this study.

B. Study Domain

In this article, a rectangular region in the southeastern United States is selected as the study domain, which spans from 81°W to 97°W in longitude and 31°N to 39°N in latitude, covering an area of approximately 1600 km $\times$ 800 km as shown by the red rectangle in Fig. 1. This region was chosen due to the fact that it primarily consists of relatively flat terrains with similar characteristics as the digital elevation model (DEM) information in Fig. 1 presents. The deployment of radars and rain gauge sites across this area has been extensive, owing to the absence of complex terrain features like high mountains or lakes [2]. As a result, the MRMS precipitation estimates in this region are considered reliable and accurate with the bias-correction algorithm, making them suitable as target labels for our deep learning model. Although the main challenge is the mesoscale convective precipitation in western and the mid-western United States as described above, this article is describing a prototype algorithm and the selected region is relatively "simple" in terms of terrain and has accurate QPE for training.

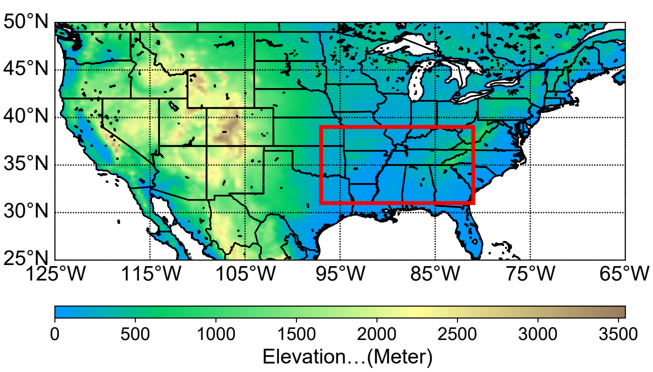


Fig. 1. Location of the study domain (denoted by the red rectangle) in the CONUS. The background color indicates the digital elevation model (DEM) information.

C. Methodology

In this subsection, we present the deep learning-based framework for precipitation, which includes data preprocessing, components and structure of the deep learning model, essential aspects of model training and testing, and a comprehensive evaluation of the retrieval performance using appropriate metrics.

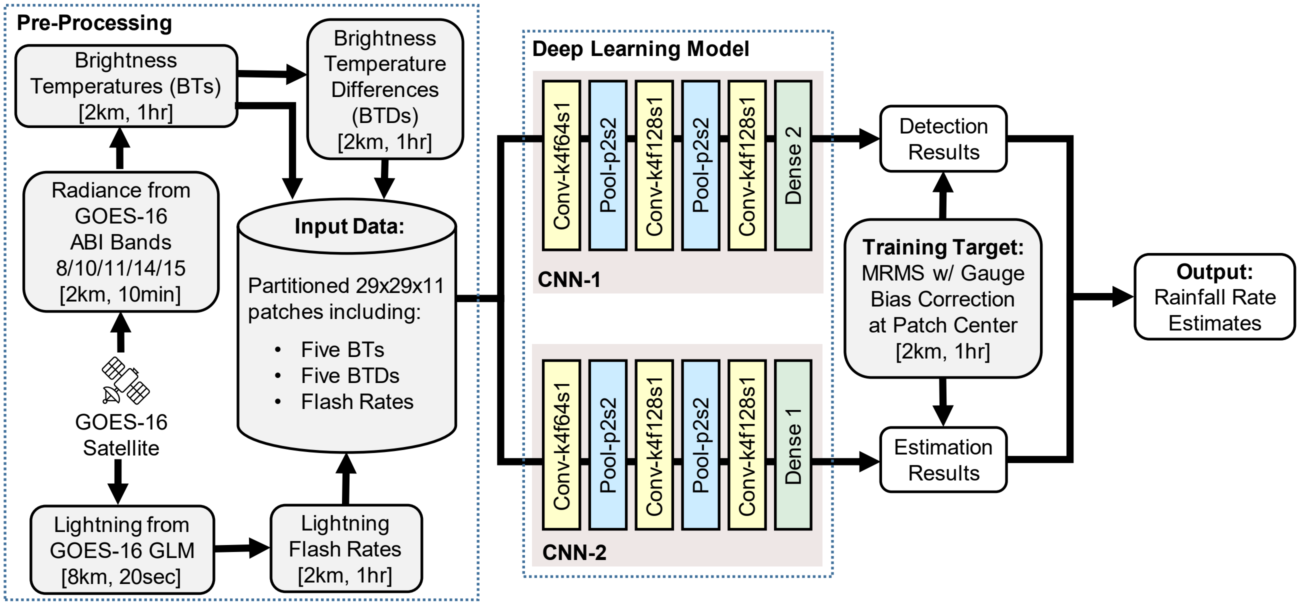


Fig. 2. The deep learning framework for precipitation retrievals using GOES ABI and GLM data. In the convolutional (Conv-) and pooling (Pool-) layers, k, f, s, and p represent kernel size, number of feature maps, stride size, and pooling size, respectively. The spatio-temporal resolution of the data and data pre-processing is also illustrated. The rainfall rate estimates are determined by combining the detection results from CNN-1 and estimation results from CNN-2.

1) Data Pre-Processing: The data pre-processing phase comprises four key parts: a) converting the radiance data from the GOES-16 ABI measurement into BT and BTD features; b) forming the hourly flash rate feature from the GOES-16 GLM measurement; c) partitioning the features into patches; d) labeling the partitioned patches using the MRMS data in the training phase. Each of them will be demonstrated elaborately in the following.

a) BTs and BTDs: To convert the spectral radiance data of the selected bands in the GOES-16 ABI measurement into the BT, we use the transformation equation [22]

$$\mathrm{BT}_i = [fk_2/(\ln((fk_1/L_i)+1)) - bc_1]/bc_2, \tag{1}$$

where $i \in \{8, 10, 11, 14, 15\}$ is the index for the GOES-16 ABI band; BT_i is the brightness temperature converted from the band i spectral radiance; L_i is the spectral radiance data from the band i; bc_1 and bc_2 are the spectral response function offset and the scale correction term, respectively; fk_1 and fk_2 are the Planck function coefficients, which are derived as

$$fk_1 = 2hc\nu, \tag{2}$$

$$fk_2 = 2h\frac{c}{b}\nu, \tag{3}$$

where h is the Planck constant; b is the Boltzmann constant; c is the velocity of light; ν is the band central wavenumber. In this article, the values of bc_1, bc_2, fk_1, and fk_2 are loaded from the GOES-16 ABI measurement data files.

Once the BT features are calculated, we also include the following BTD features:

$$\begin{aligned}
\mathrm{BTD}_{10,\ 8} &= \mathrm{BT}_{10} - \mathrm{BT}_{8},\\
\mathrm{BTD}_{11,14} &= \mathrm{BT}_{11} - \mathrm{BT}_{14},\\
\mathrm{BTD}_{14,10} &= \mathrm{BT}_{14} - \mathrm{BT}_{10},\\
\mathrm{BTD}_{11,14} &= \mathrm{BT}_{11} - \mathrm{BT}_{14},\\
\mathrm{BTD}_{14,15} &= \mathrm{BT}_{14} - \mathrm{BT}_{15},
\end{aligned} \tag{4}$$

where BTD_{i_1,i_2}, $i_1, i_2 \in \{8, 10, 11, 14, 15\}$ is the brightness temperature difference of BT_{i_1} subtracting BT_{i_2}.

To accommodate the 1-hour temporal resolution of the bias-corrected MRMS data, the BTs and BTDs with a resolution of 10 minutes are averaged over the six scans within an hour. By doing so, the resulting averaged BTs and BTDs are treated as the representative BT and BTD features for that specific hour.

b) Lighting Flash Rates: In GOES-16 GLM measurement, flashes are recorded and stored with the associated longitude, latitude, and radiant energy. To form the hourly flash rate feature, we initialize a grid with a 2 km spatial resolution over the selected study domain for each hour. We then find the coordinates of the flashes within this hour and assign the grid pixel value with the occurrence of the flashes at the corresponding location. It should be noted that the flash occurrence is counted regardless of the released flash energy. Also, the spatial resolution formed hourly flash rate feature is 2 km while the spatial resolution of the GOES-16 GLM measurement is 8 km. When the coordinates of the flashes are not perfectly matched to the grid coordinates, we find the closest locations in the feature grid as the corresponding locations for the flashes.

c) Partitioning: In summary, a total of 11 features are stacked together as the input features to the deep learning model, including 5 BT features, 5 BTD features, and the GLM flash rate feature. To apply these features to the deep learning model, we partition them into $11 \times 29 \times 29$ patches with stride size 1 and feed them to the deep learning model next. The selection of 29×29 patch size in each band is based on the following consideration. CNNs excel at capturing spatial information from data. By utilizing nearby information surrounding the patch center, we aim to effectively determine the rainfall rate at that specific location. However, if the patch size is too large, it may encompass numerous precipitation patterns that could potentially confuse the CNN and yield inaccurate results. We have conducted a series of experiments involving different patch sizes to determine the optimal choice. After careful evaluation, we found that a patch size of 29×29 yielded the best performance. Therefore, we made the decision to employ a 29×29 patch size for each band in our framework.

d) Ground-Based MRMS Product: In the training phase, we first down-sample the MRMS data from 1 km to 2 km to match the resolution of the input features and generate target labels for the partitioned patches. Specifically, we retain every other cell on the MRMS grid, both column-wise and row-wise. This sub-sampling approach allows us to reduce the spatial resolution of the MRMS data while still capturing the essential information needed for our analysis. The MRMS estimates at the center of the partitioned patches are then utilized to label the $11 \times 29 \times 29$ patches. In the case of precipitation detection, the partitioned patches are labeled as precipitation if the MRMS rainfall rates exceed 0.1 mm/hr, and as non-precipitation otherwise. For precipitation estimation, the MRMS estimates at the center of the partitioned patches are taken as the target values of the partitioned patches.

2) Deep Learning Framework: Fig. 2 details the deep learning framework. Once the GOES-16 ABI and GLM measurements have been pre-processed, we apply the partitioned patches to the deep learning model. The deep learning model consists of two components: a detection CNN module (CNN-1) and an estimation CNN module (CNN-2), both of which have three convolutional layers, two average pooling layers, and one dense layer. For all the convolutional layers in CNN-1 and CNN-2, the kernel size is set to 4 (i.e., the width and height of kernels are 4) with stride size 1. The numbers of feature maps in three convolutional layers for both CNN-1 and CNN-2 are set to 64, 128, and 128, respectively. All the pooling layers in CNN-1 and CNN-2 use an average pooling mask with size 2 and stride 2. In CNN-1, the feature maps at the third convolutional layer are fully connected to the dense layer, which has 2 nodes to predict the detection results, i.e., precipitation and non-precipitation. In CNN-2, the feature maps at the third convolutional layer are fully connected to a single-node dense layer for precipitation estimation.

In the training phase, the losses between the predicted results and the actual target labels provided by the MRMS data are calculated for both modules, in which the CNN-1 uses the binary cross-entropy loss, and the CNN-2 uses a weighted mean squared error (MSE) loss to address the precipitations according to their target rainfall rates. The weighted MSE loss in CNN-2 is defined as

$$\text{Loss}(\theta) = \frac{1}{N}\sum_{n=1}^{N}(1+y_n)(\hat{y}_n(\theta)-y_n)^2, \tag{5}$$

where θ denotes the model parameters in CNN-2, N is the total number of training patches; y_n and $\hat{y}_n$ are the actual and the predicted label for the nth training patch, respectively. One can notice that the heavier the precipitation in a patch is, the larger weight it gets. Since the actual rainfall rate y_n is always non-negative, $(1 + y_n)$ ensures the weight is always greater than or equal to 1. The losses are then back-propagated to the CNN modules to update the network parameters individually.

In the testing phase, the test data features are pre-processed into $11 \times 29 \times 29$ patches similar to the training data, and then applied to the deep learning model. The final precipitation retrievals are determined by fusing the results from the two modules. Specifically, if CNN-1 detects precipitation for a given patch, the rainfall rate at the center location of that patch is estimated based on the outcome of CNN-2. Conversely, if CNN-1 does not detect precipitation in a patch, the corresponding rainfall rate at the center location for that patch is 0 mm/hr.

3) Performance Evaluation: The detection performance of the proposed deep learning framework is evaluated in terms of Heidke skill score (HSS), critical success index (CSI) [23], probability of detection (POD), and false alarm ratio (FAR), which are given as

$$\text{HSS} = \frac{2(\text{TP}\times\text{TN}-\text{FN}\times\text{FP})}{(\text{TP}+\text{FN})(\text{FN}+\text{TN})+(\text{TP}+\text{FP})(\text{FP}+\text{TN})}, \tag{6}$$

$$\text{CSI} = \frac{\text{TP}}{\text{TP}+\text{FN}+\text{FP}}, \tag{7}$$

$$\text{POD} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{8}$$

$$\text{FAR} = \frac{\text{FP}}{\text{TN}+\text{FP}}, \tag{9}$$

where TP (true positive) is the number of pixels predicted and labeled as precipitation; FP (false positive) is the number of pixels labeled as non-precipitation, but predicted as precipitation; FN (false negative) is the number of pixels labeled as precipitation, but predicted as non-precipitation; TN (true negative) is the number of pixels predicted and labeled as non-precipitation.

The integrated HSS and CSI metrics directly reflect the model skill. Normally, higher POD, HSS, CSI, and lower FAR indicate better model performance. Furthermore, to investigate the precipitation detection performance at different rainfall rates, the evaluation metrics are computed based on a number of rainfall rate thresholds. This is done by transforming the deep learning-based estimates and the ground-based MRMS product into binary arrays. Specifically, if the rainfall rate of the deep learning-based estimate or the ground-based MRMS product at a grid pixel is higher than the threshold, “1” is assigned to this grid pixel, otherwise “0” will be assigned. For each threshold, we compute the confusion matrix between the transformed binary arrays from the proposed deep learning-based rainfall rate estimates and the corresponding ground-based MRMS product to further obtain the four detection evaluation metrics above.

The performance of the precipitation estimation is evaluated in terms of mean squared error (MSE), mean absolute error (MAE), normalized mean error (NME), and the normal mean absolute error (NMAE), which are calculated as

$$\mathrm{MSE} = \frac{1}{M}\sum_{m=1}^{M}(\hat{y}_m - y_m)^2, \tag{10}$$

$$\mathrm{MAE} = \frac{1}{M}\sum_{m=1}^{M}|\hat{y}_m - y_m|, \tag{11}$$

$$\mathrm{NME} = \frac{\sum_{m=1}^{M}\hat{y}_m - y_m}{\sum_{m=1}^{M}y_m}, \tag{12}$$

$$\mathrm{NMAE} = \frac{\sum_{m=1}^{M}|\hat{y}_m - y_m|}{\sum_{m=1}^{M}y_m}, \tag{13}$$

where $m \in [1, M]$ is the testing patch index; y_m is the rainfall rate from MRMS ground truth, $\hat{y}_m$ is the predicted rainfall rate estimates; M is the total number of testing patches. It is important to compute and show all of these metrics to gain a comprehensive understanding of the model's performance. Each metric captures different aspects of the errors, providing valuable insights. In other words, MSE and MAE quantify the squared and the absolute errors, respectively, while NME and NMAE provide a relative perspective by normalizing the errors.

To assess the efficacy of precipitation estimation under varying intensities, we adopt the identical rainfall rate thresholds employed in evaluating the detection performance of precipitation. At each threshold level, we determine the grid pixel indices where the ground-based MRMS product exceeds the threshold value. We subsequently calculate the four categories of estimation evaluation errors between the predicted rainfall rates and the ground-based MRMS product exclusively for these selected pixel indices.

III. Experimental Results and Discussion

In this section, we commence by first conducting a comprehensive examination of the impact of the input features for precipitation retrieval. Then, the experimental results of the proposed deep learning-based precipitation retrieval and the operational product over the selected study domain are presented and evaluated using the metrics in Section II-C3, encompassing both hourly and daily accumulated results. Lastly, an extensive discussion of the precipitation retrieval performance is provided.

A. Features Analysis

To investigate the impact of each input feature on the precipitation retrieval for different levels of precipitation rainfall rates, Fisher discriminant analysis (FDA) [24] was applied to the 11 input features across four distinct rainfall rate thresholds, including 1, 4, 7, and 10 mm/hr. By employing FDA, we were able to quantitatively measure the discrimination of the features based on the statistics of the dataset, where the highest Fisher measure indicates the most informative and discriminating feature. For each threshold, the Fisher measure FDA(i) for the band i feature was calculated as

$$\mathrm{FDA}(i) = \frac{|\mu_0(i) - \mu_1(i)|^2}{\sigma_0^2(i) + \sigma_1^2(i)}, \tag{14}$$

where $\mu_0(i)$ and $\mu_1(i)$ are the pixel mean of the band i feature over all the pixels in patches whose precipitation estimation target labels are above and below the threshold, respectively; $\sigma_0^2(i)$ and $\sigma_1^2(i)$ are the pixel variance of the band i feature over all the pixels in patches whose precipitation estimation target labels are above and below the threshold, respectively. In this analysis, we selected 64,000 patches randomly from the data in 2019 and obtained the Fisher measures of the 11 input features, which are presented in Fig. 3.

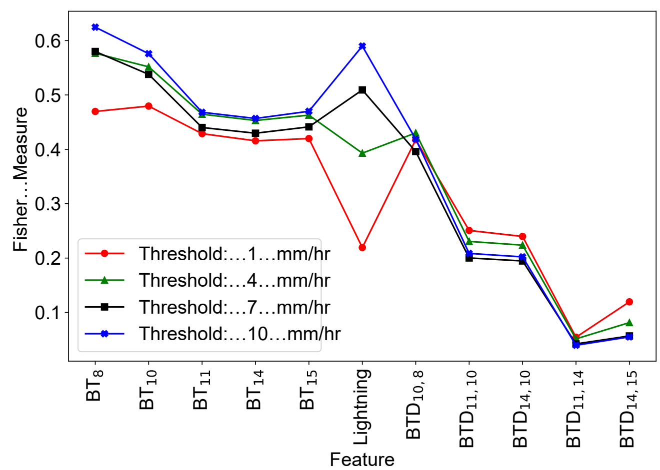


Fig. 3. Fisher measures of the input features at different rainfall intensities determined by the rainfall rate thresholds. The x-axis includes all the input features to the deep learning model, where BT_i and BTD_{i_1,i_2}, $i, i_1, i_2 \in \{8, 10, 11, 14, 15\}$ refer to the BT from the GOES-16 ABI band i, and the BTD from band i_1 subtracting band i_2, respectively. Lightning refers to the hourly flash rate obtained from the GOES-16 GLM measurement.

As depicted in Fig. 3, we note that the Fisher measurements of both BT and BTD features remained relatively constant across varying rainfall rate thresholds. In general, BT features exhibited greater impact than BTD features, with the BT feature from ABI band 8, having a center wavelength of 6.2 µm, being the most informative. Nonetheless, with regards to the hourly flash rate feature (denoted as *Lightning* in Fig. 3), its Fisher measurements exhibited remarkable variations for different rainfall rate thresholds. As the rainfall rate threshold increased from 1 mm/hr to 10 mm/hr, the Fisher measure of the hourly flash rate feature increased from 0.22 to 0.59 approximately. Furthermore, the hourly flash rate feature became the second most informative feature when the rainfall rate threshold was 10 mm/hr. The Fisher Measure results for the Lightning feature are consistent with the fact that higher rain rates are usually associated with deeper and more intense convection compared to lower rain rates. These findings suggest that the lightning data could potentially play a pivotal role in precipitation retrieval, particularly for heavy precipitation.

B. Experimental Setups

Once the GOES-16 ABI and GLM measurements were pre-processed and partitioned into $11 \times 29 \times 29$ patches, a total of 6.4 million patches from the data in 2019 were randomly selected for training the designed deep learning model, for which the numbers of precipitation patches and non-precipitation patches are equal. Then, the deep learning model was validated on 6.4 million validation patches, which were also randomly selected from the data in 2019, and exclusive to the training patches. Finally, the trained deep learning model was tested on the test patches in 2020. For the testing set, we sampled 12.8 million patches from the dataset collected in 2020, for which the numbers of precipitation patches and non-precipitation patches are equal. The reason for the sampling was to ensure a well-balanced representation of both precipitation and non-precipitation patches in the testing dataset. Including all non-precipitation samples from the dataset would have led to an imbalance in the testing set, potentially biasing and inflating certain evaluation metrics such as TN, MSE, and MAE. By using a testing set with a balanced representation of precipitation and non-precipitation samples, we were able to obtain reliable and unbiased evaluation results for the framework's performance. Before being applied to the deep learning model, the pixel values in the patches were normalized between 0 and 1 for each individual feature as

$$\mathrm{BT}_{i,norm} = \frac{\mathrm{BT}_i - \mathrm{BT}_{i,min}}{\mathrm{BT}_{i,max} - \mathrm{BT}_{i,min}}, \tag{15}$$

where $\mathrm{BT}_{i,max}$ and $\mathrm{BT}_{i,min}$ are the maximum and the minimum values of the BT from band i in the training set. The normalization for the BTD and flash rate features were done in a similar manner. It should be noted that the BT features, the BTD features, and the flash rate features had different maximum and minimum values, which needed to be taken into account in normalization. The maximum and minimum values for the input features used in this article are presented in Table I.

TABLE I
MAXIMUM AND MINIMUM VALUES USED FOR NORMALIZING THE INPUT FEATURES.

feature	Maximum	Minimum
BTD_8 (K)	257.44	200.37
BTD_{10} (K)	273.46	199.81
BTD_{11} (K)	310.46	198.95
BTD_{14} (K)	314.75	198.63
BTD_{15} (K)	308.96	198.19
$\mathrm{BTD}_{10,\ 8}$ (K)	29.08	-5.89
$\mathrm{BTD}_{11,14}$ (K)	49.86	-1.37
$\mathrm{BTD}_{14,10}$ (K)	54.40	-1.80
$\mathrm{BTD}_{11,14}$ (K)	10.83	-7.40
$\mathrm{BTD}_{14,15}$ (K)	8.32	-0.43
Lightning (Count/hr)	5	0

Both CNN modules used ReLU as the activation function at each neuron. The deep learning model was optimized using the Adam optimizer with learning rates 2e-6 and 2e-7 for CNN-1 and CNN-2, respectively. The training batch size was set to 100 and the training epoch was set to 100. All experiments were implemented using the Tensorflow platform [25]. In the following figures and tables, the proposed framework is referred to as "CNN w/ GLM".

In order to experimentally quantify the impact of the GLM data, another model with the same framework in Fig. 2 was trained except without using the GLM measurement for comparison, hence the input patches for this model were of size $10 \times 29 \times 29$. Otherwise, the loss functions, activations, optimizer, hyper-parameters, and the network architecture beside the input layer remained the same. This model is referred to as "CNN w/o GLM".

C. Experimental Results

In this section, the experimental results are presented, including the hourly and daily accumulated results. Results for both accumulation periods were evaluated using the metrics in Section II-C3 with different thresholds, including the performance comparison of the proposed deep learning-based precipitation retrieval with and without GLM measurement, as well as the currently operational GOES-16 RRQPE product.

1) Hourly Retrieval Results: Fig. 4 shows a visual example of the hourly precipitation estimates at 23:00 UTC, May 17, 2020, including the CNN-based precipitation retrieval with and without the hourly flash rate feature from the GLM measurement, the MRMS ground truth, and the currently operational GOES-16 RRQPE product. It can be seen from Fig. 4 that the CNN-based precipitation retrieval outperformed the operational product as it was able to identify heavy precipitation areas that the operational product missed. The CNN-based retrieval provided better estimates for both weak and intense precipitation, particularly in areas with intense precipitation when using lightning data compared to the model without lightning data. Despite some false alarmed areas, such as the region near 85°W, 37°N, the estimates from the CNN-based retrieval with the lightning data had lower errors than the estimates from the CNN-based retrieval without the lighting data and the operational product, and the operational product completely failed to indicate the precipitation over this area. These visual and comparative observations between the CNN-based retrieval with and without GLM corroborated our conclusion in Section III-A that the lightning data from the GLM measurement played a crucial role in precipitation retrieval.

The quantitative evaluation results of the CNN-based precipitation retrieval and the operational product on the validation data and the test data were summarized in Table II, where the best scores were indicated in bold. The evaluation metrics for detection focused on distinguishing between precipitation and non-precipitation, whereas the assessment metrics for estimation encompassed all levels of precipitation events. In general, the CNN-based retrieval approach outscored the operational GOES-16 RRQPE product. For the integrated HSS

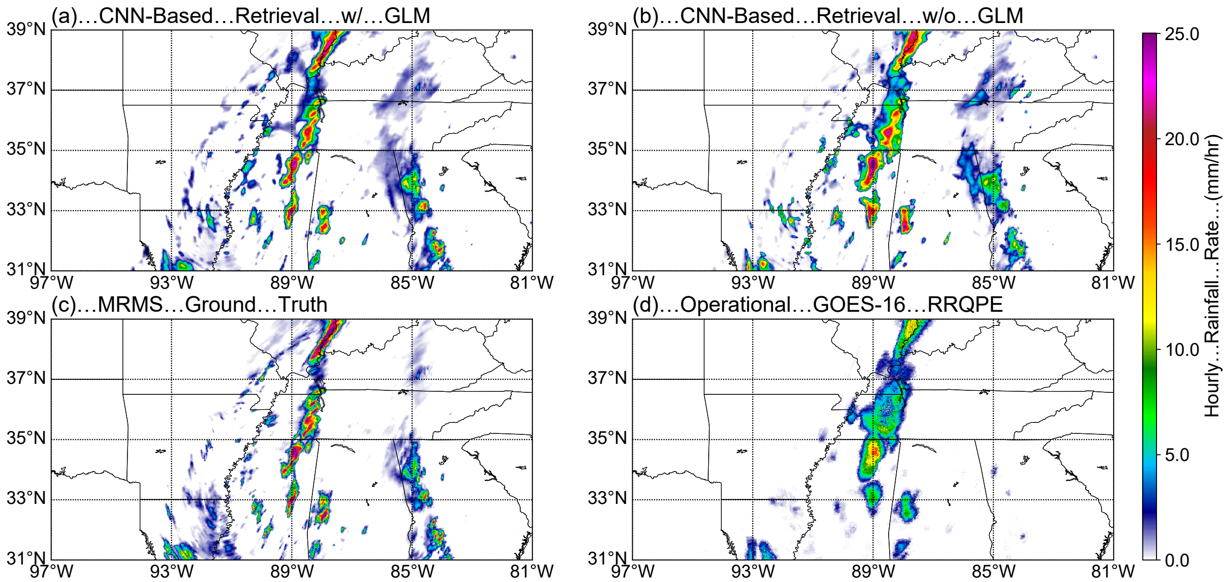


Fig. 4. Example of hourly precipitation estimates at 23:00 UTC, May 17, 2020: (a) CNN-based retrieval with GLM data; (b) CNN-based retrieval without GLM data; (c) ground-based MRMS product; (d) operational GOES-16 RRQPE product.

TABLE II
PERFORMANCE OF HOURLY PRECIPITATION RETRIEVALS FROM THE DESIGNED CNN MODELS AND THE OPERATIONAL GOES-16 RRQPE PRODUCT BASED ON THE VALIDATION AND TEST DATA.

		HSS	CSI	POD	FAR	MSE (mm^2)	MAE (mm)	NME	NMAE
Validation Data	CNN w/ GLM	**0.73**	**0.74**	0.78	0.06	**27.62**	**1.98**	**-0.04**	**0.48**
	CNN w/o GLM	0.70	**0.74**	**0.91**	0.23	48.32	2.76	-0.10	0.66
	Operational Product	0.52	0.54	0.55	**0.03**	73.97	3.55	-0.52	0.75
Test Data	CNN w/ GLM	**0.71**	0.73	0.77	0.06	**31.79**	**2.49**	**-0.09**	**0.52**
	CNN w/o GLM	0.67	**0.74**	**0.91**	0.24	50.59	3.56	-0.11	0.74
	Operational Product	0.53	0.55	0.57	**0.03**	71.49	3.65	-0.53	0.76

and CSI metrics, the CNN-based retrieval with the lightning data had at least 32% improvement on the testing data. For the estimation errors, it reduced the MSE, MAE, and NMAE by at least 31%. Although the CNN-based model without GLM had higher POD, it also had a relatively high FAR.

To highlight the precipitation performance for different rainfall rates, seven thresholds were applied in calculating the evaluation metrics, including 1, 2, 4, 7, 10, 15, and 20 mm/hr. These metrics versus thresholds were plotted in Fig. 5. It can be seen that, in general, the CNN-based retrieval approach produced higher skill scores and lower errors compared to the operational GOES-16 RRQPE product. From Fig. 5 (a)-(c) and (e)-(f), we can observe that when the rainfall rate thresholds increased from 10 to 20 mm/hr, the detection performance of all models dropped and the estimation error raised, which implied that the precipitation retrieval for heavy rain was more challenging than the retrieval for weak and moderate rain. Similarly, Fig. 5 (e)-(f) reveals that an increase in the rainfall rate thresholds from 1 to 20 mm/hr led to a decline in estimation performance for all models. This was expected as higher rain rates are typically associated with more intense convective features that are often transient and more challenging to retrieve from a satellite remote sensing perspective. Notice that the normalized errors in Fig. 5 (g)-(h) dropped as the thresholds increased because of the normalization in the calculation. It is remarkable that the CNN-based retrieval model with GLM exhibited a more reliable performance than both the CNN-based model without GLM and the operational product, particularly for heavy precipitation events where the hourly rainfall rate exceeded 10 mm/hr. This conclusion is consistent with the visual example illustrated in Fig. 4 and the FDA investigation in Fig. 3.

2) Daily Retrieval Results: The daily precipitation retrieval results were obtained by aggregating the precipitation estimates over 24 hours of a day. An example of the precipitation estimates on Aug 20, 2020, is presented in Fig. 6, including the CNN-based precipitation retrieval with and without the hourly flash rate feature from GLM measurement, the MRMS ground truth, and the currently operational GOES-16 RRQPE product. In Fig. 6 (c), an area of precipitation with a number of embedded convective cells can be observed in the southeastern part of the study domain. Most of these rainfall patterns were revealed by the CNN-based retrieval with the lightning data in Fig. 6 (a), albeit with some underestimation. Conversely, these heavy precipitation areas were largely missed by the CNN-based retrieval without the lightning data and the operational

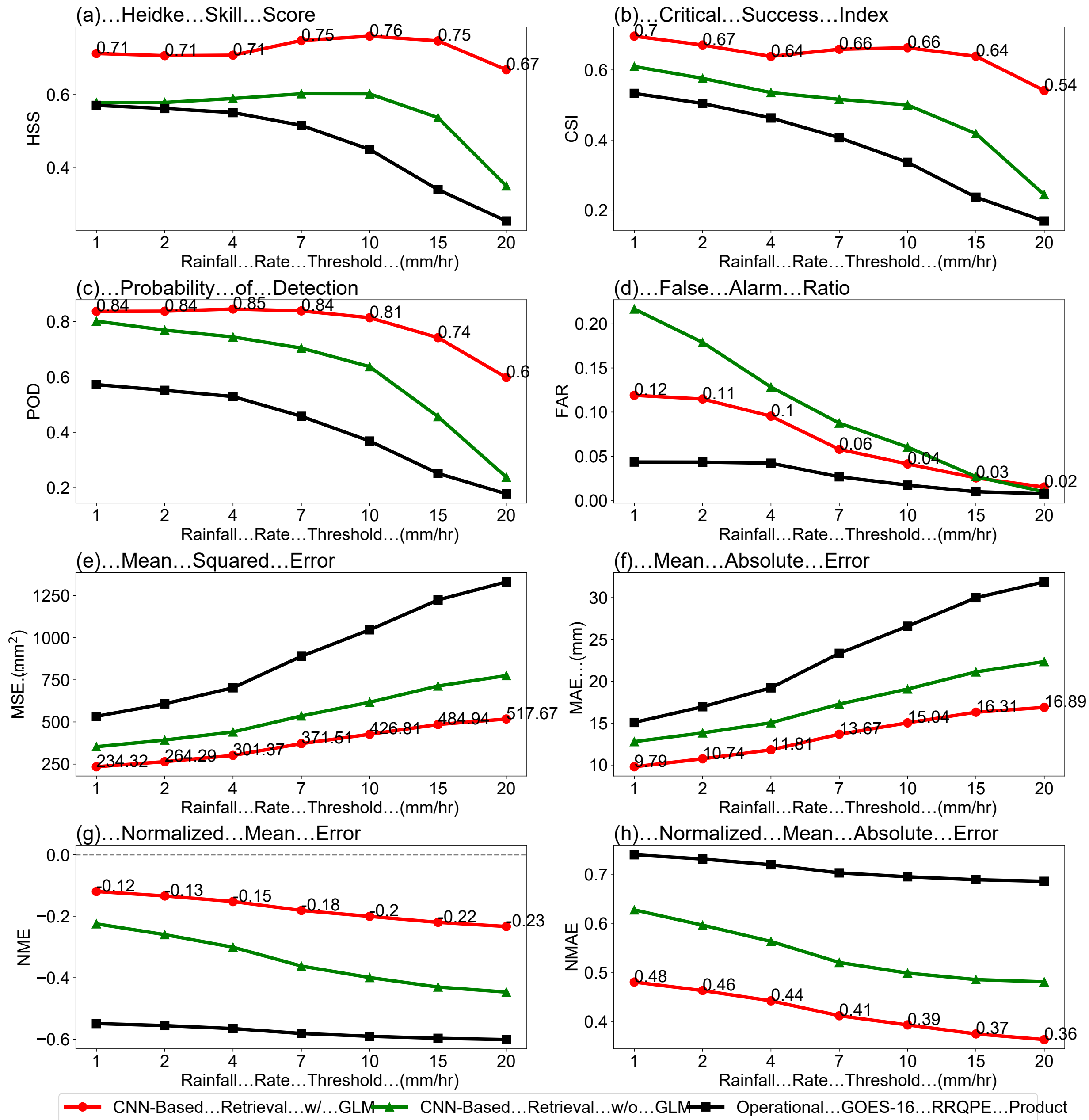


Fig. 5. The quantitative evaluation scores of the operational GOES-16 RRQPE product and deep learning-based retrievals with and without using lightning information. The evaluation scores are calculated for hourly rainfall accumulations based on all the test data, using different thresholds: (a) HSS, (b) CSI, (c) POD, (d) FAR, (e) MSE, (f) MAE, (g) NME, (h) NMAE. The evaluation scores of the CNN-based retrievals with GLM data are marked along the red curves.

product, leading to substantial inaccuracy and underestimation. In contrast, for the moderate precipitation in the northeast part of the study domain, both of the CNN-based retrievals were able to more accurately delineate the precipitation areas and their intensities compared to the operational product. In Fig. 6 (b), the estimates without using GLM measurement seemed to have a coarser resolution than the estimates in Fig. 6 (a). Since the CNN-based retrieval model without GLM exhibits a relatively high false alarm rate for weak to moderate precipitation, after aggregating the estimates to a daily scale, the delineation of areas with weak to moderate precipitation appeared less distinct compared to the estimates obtained from the CNN-based retrieval model without GLM model, making the estimates look smoother.

The daily precipitation evaluation with thresholds was carried out as follows. We first obtained the daily precipitation estimates of the days in the testing months in 2020. Note that we completely skipped the days with abnormal/missing data in ABI or GLM measurements, as well as the days without recorded precipitation events within the chosen study domain. We then calculated the daily precipitation evaluation metrics for each remaining day with thresholds 1, 5, 10, 20, 30, 40, and 50 mm/day. Finally, for each threshold, the daily precipitation evaluation metrics were averaged over the remaining days. The averaged daily precipitation evaluation metrics were plotted in

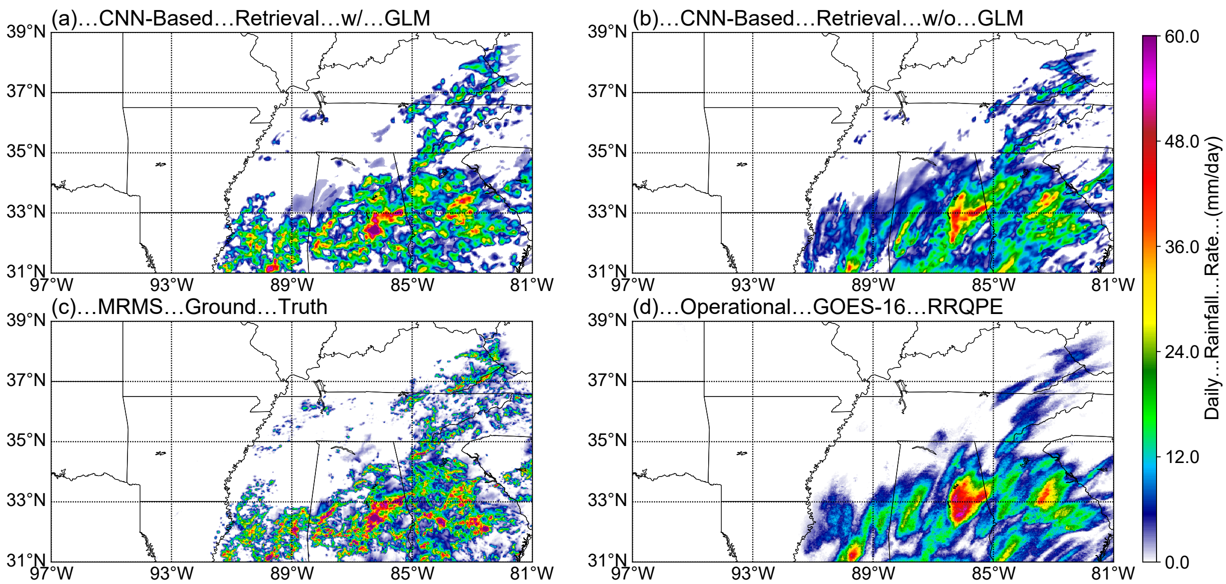


Fig. 6. Example of daily precipitation accumulations on August 20, 2020: (a) CNN-based retrieval with GLM data; (b) CNN-based retrieval without GLM data; (c) ground-based MRMS product; (d) operational GOES-16 RRQPE product.

Fig. 7. It can be seen from Fig. 7 that the CNN-based retrieval approach performed better than the operational GOES-16 RRQPE product with lower errors. For the more challenging heavy precipitation retrieval, the CNN-based retrieval with the lightning data exhibited large improvements when compared to the CNN-based retrieval without the lightning data and the operational product. Although the operational product had the lowest FAR as shown in Fig. 7 (d), the FAR of the CNN-based retrieval with GLM still maintained reasonable values. In Fig. 7 (g), the negative values of NMEs imply that all three models underestimated the precipitation overall, with the CNN-based retrieval with lightning data demonstrating the lowest NME. These observations are consistent with the underestimation of precipitations in Fig. 6 using the CNN-based retrieval. Furthermore, our comparison of the CNN-based retrieval with and without GLM highlights the significance of the lightning data for precipitation retrieval, particularly as the rainfall rates increase.

D. Discussion

1) Deep Learning for Precipitation Retrieval: The experimental results have demonstrated the efficacy and superiority of the deep learning model for precipitation retrieval using the ABI and GLM measurements on the GOES-R series satellite, particularly for moderate and heavy precipitation. From these results, one can conclude that the deep learning-based precipitation retrieval can significantly enhance the currently operational product to improve the precipitation retrieval accuracy, with brightness temperatures from the selected ABI bands and the flash rates from the GLM measurement as the input features. Thus, the proposed deep learning framework can potentially serve as an alternative approach for GOES precipitation retrievals. In comparison to other deep learning models with wide usages, e.g., VGG-16 with 13 convolutional layers and 3 dense layers [26], [27], our proposed deep learning model only has 3 convolutional layers and 1 dense layer for each CNN module. Consequently, the proposed deep learning model is lightweight and efficient, providing precipitation estimates without consuming excessive computational time. However, it is important to note that further investigation is necessary to determine the generalization capability of deep learning models for precipitation retrieval.

From another perspective, the deep learning model learns a mapping from the BTs, BTDs, and flash rates to the precipitation rainfall rate. These BT and BTD features are derived from the observations of the GOES-16 ABI measurement and are on the top or near the clouds with a 2 km spatial resolution. The flashes in the GOES-16 GLM measurement account for all in-cloud, cloud-to-cloud, and cloud-to-ground lightning activities with 8 km resolution. On the other hand, the target labels used for training the deep learning model are the ground-based MRMS dataset and have a 1 km spatial resolution. In the training phase, the MRMS data are resampled to 2 km spatial resolution to provide the target labels for the training patches. Hence, errors may inevitably arise during sampling and providing the target labels of the training patches. In spite of the possible errors raised from the different resolutions of datasets, the environments between the clouds and the grounds are variable. Any atmospheric changes may affect the learned mapping, and cause unstable or inaccurate precipitation estimates.

Moreover, it is anticipated that the operational RRQPE product on the GOES-R series satellite has a global or at least hemispheric coverage, while the study domain in this article is only a region in the southeastern United States. When we expand the study domain to the whole globe, further studies are required to justify whether multiple deep learning models or only one deep learning model is adequate to offer accurate hemispheric coverage of rainfall rate estimates. An additional

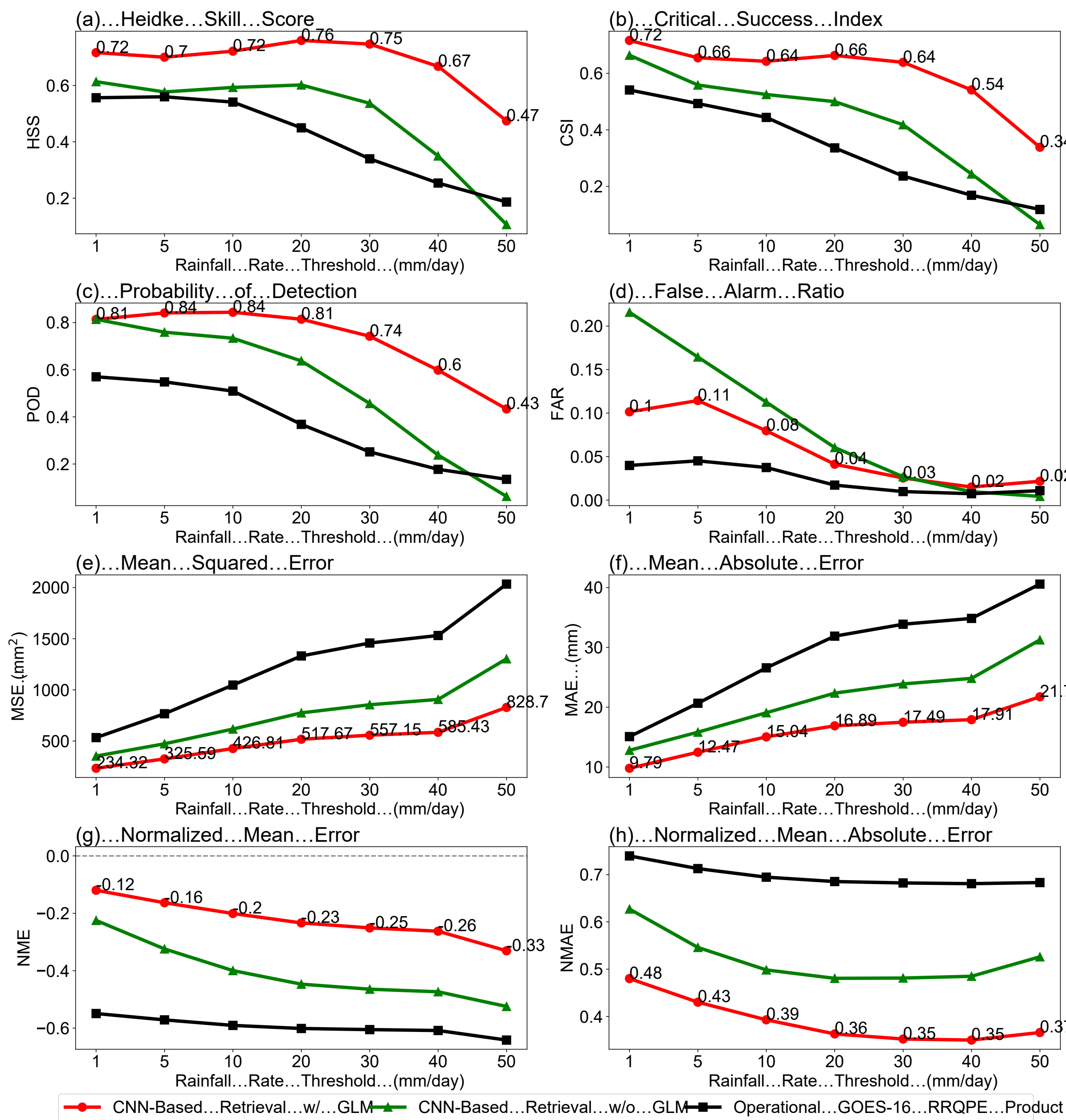


Fig. 7. The quantitative evaluation scores of the operational GOES-16 RRQPE product and deep learning-based retrievals with and without using lightning information. The evaluation scores are calculated for daily rainfall accumulations based on all the test data, using different thresholds: (a) HSS, (b) CSI, (c) POD, (d) FAR, (e) MSE, (f) MAE, (g) NME, (h) NMAE. The evaluation scores of the CNN-based retrievals with GLM data are marked along the red curves.

concern regarding the expansion of the proposed framework pertains to the parallax shift in ABI measurements [28]. That is, the cloud-top features observed by the GOES-R satellite sensors appear to be displaced away from the satellite sub-point. The parallax shift becomes more noticeable as the study domain moves closer toward the edges of the ABI's field of view. For the study domain in this article, the parallax effects were less significant compared to other areas, such as the western United States, due to its proximity to the satellite sub-point. Furthermore, it should be acknowledged that one of the advantages of CNNs lies in their ability to be shift-invariant, which prompted us to overlook the parallax effect in our study. Nevertheless, it is important to note that this parallax shift will pose a more substantial challenge when extending our work to encompass the western United States or the entire western hemisphere. Despite the missing or invalid data in GOES-16 ABI and GLM measurements, the reliability and accuracy of the targets of the training data, i.e., ground-based MRMS product, are crucial when training a deep learning model for precipitation retrieval. However, as presented in [2], the coverage of the MRMS dataset is mostly limited to the continental United States. Even so, the deployment of the ground radars over some complex terrains may not be as sufficient as those in plain areas which have relatively flat terrain. These complex areas include the regions where the deployment of radars is challenging due to various environmental obstacles such as mountains, oceans, and dense vegetation. For offering a global precipitation retrieval product, these complex terrains are unavoidable and further study on deep learning-based precipitation over complex terrains needs to be further investigated. One benefit of the deep learning-based precipitation retrieval is to leverage the transfer learning technique for the precipitation retrieval over different regions [29], [30]. By employing transfer learning, the deep learning model trained on the present study domain can be adapted for other regions by retaining most of the previously learned knowledge and making minor modifications based on the characteristics of the new region.

In addition, the performance of deep learning models is highly dependent on the quality and the quantity of the dataset. It is crucial that the training data encompasses a vast number of precipitation patterns to achieve desirable results. Consequently, we aim to include as many precipitation patterns as possible during the training phase. In case the testing data differ considerably from the training data and are unfamiliar to the deep learning model, the testing performance may be adversely affected. However, it is not feasible to incorporate all possible precipitation patterns into the training data. Another consideration with a large amount of data is that, although the training can be done offline, the elapsed time for training a deep learning model will markedly increase. One remedy is to implement the continual lifelong learning technique for deep learning models [31], [32]. Once the deep learning models have been well-trained but some unseen precipitation patterns arise, the continual lifelong learning technique, e.g., the elastic weight consolidation (EWC) in [32], can adjust the deep learning model parameters when encountering previous unseen precipitation patterns without forgetting the previous knowledge catastrophically, for which the deep learning model can maintain the performance on the previous precipitation patterns but also be adaptive to the newly unseen patterns. Also, it is no longer a restrictive requirement that the training data must contain all intensities of precipitation data.

Finally, in this article, we only included BT, BTD, and lightning as input features for precipitation retrieval. Other environmental factors, such as terrain information, water vapor, temperature, and wind magnitude/direction, also impact the precipitation and therefore can be incorporated with the BT, BTD, and lightning features for a potentially more accurate precipitation retrieval. One way to include these additional features without starting the training process over is the network expansion for deep learning neural networks [33], [34]. For example, the terrain and elevation information has been shown to be important factors for the satellite-based precipitation retrieval in [35], [36], [37]. The deep learning model can incorporate the elevation data by adding an extra input layer to the deep learning model and continue learning a mapping function from BT, BTD, flash rate, and elevation to the precipitation rainfall rates.

2) Lightning Data in Precipitation Retrieval: From the experimental results, it can be seen that after including the hourly flash rate feature from the GOES-16 GLM measurement as an additional input to the deep learning model, the performance of the deep learning-based precipitation retrieval was outstandingly improved. This finding supports the inference on the lightning data using FDA in Section III-A, highlighting the crucial role of lightning information in precipitation retrievals.

However, in this article, only the flash rate from the GLM measurement was used, regardless of the energy released from the flashes. One main reason we did not use the released energy directly as the feature is that it would not provide clear information about the absence of lightning. If we used the hourly released energy from the flashes as a feature, we could not justify whether a small number of flashes occurred within a short period of time but released massive of energy intensely, or a large number of flashes occurred mildly within an hour. In contrast, the hourly flash rate can explicitly reflect the occurrence of the flashes, which usually take place with the precipitations. This also avoids possible errors in estimating the released energy from the flashes. Other than that, the GLM measurement provides three products, event, group, and flash, while in this article, we selected the most integrated flash product. A supplemental study on how each of them impacts the deep learning-based precipitation retrieval could be conducted thoroughly to further improve the deep learning-based retrieval performance. Lastly, in the data pre-processing for the GLM measurement, the grids with a 2 km spatial resolution were formed, while the spatial resolution for the GLM measurement is 8 km. This spatial resolution mismatch can cause inaccurate flash rate features. The resolution of the GLM measurement can be potentially enhanced by using the super-resolution generative adversarial networks [38], [39], so that a more precise flash rate feature can be formed to match the 2 km spatial resolution of the selected infrared bands from the ABI measurements.

IV. Conclusions

Although ground-based radars can provide accurate quantitative precipitation estimates, their coverage is limited because of the difficulties in deploying ground radars over complex terrains, such as mountains and oceans. The satellite sensors can overcome this issue as they have wide coverage over the globe. The quantitative precipitation retrieval products are derived from the observations from these satellite sensors, e.g., the currently operational RRQPE product on the GOES-R series satellite. In different precipitation regimes, including orographic precipitation in the western United States and in mesoscale convective systems, the precipitation estimates from the operational RRQPE product on the GOES-R series satellite are affected and result in inaccuracy estimates [40], [6].

In light of this issue, we propose a deep learning framework for precipitation retrieval using the ABI and GLM measurements on the GOES-R series satellite. This framework leverages the BT and BTD features derived from the infrared bands on the ABI measurement and the hourly flash rate features based on the GLM measurement. These features are partitioned into patches and applied to a deep learning model as the inputs. Then, the deep learning model, comprising two CNN modules, is trained with the ground-based MRMS dataset as target labels, and tested over the selected study domain in the southeastern United States. The experimental results demonstrate that the proposed framework is capable of providing promising precipitation estimates with high accuracy, and has the potential to enhance the currently operational RRQPE product. Based on these experimental results, two main conclusions are drawn as follows.

1) The test results indicate that the deep learning-based framework for precipitation retrieval using ABI and GLM measurements can enhance the currently operational GOES-16 RRQPE product. Visually, the CNN-based retrieval model with lightning data yields more encouraging estimates and captures more precipitation patterns, especially heavy precipitations, than the operational RRQPE product on the GOES-R series satellite. Quantitatively, compared to the operational product, the CNN-based retrieval model has better detectability by achieving at least 32% improvement in terms of the integrated HSS and CSI metrics, and can reduce at least 31% estimation error in terms of MSE, MAE, and NMAE.
2) To investigate the impact of the lighting data on deep learning-based precipitation retrieval, we conducted FDA on 11 input features and controlled experiments, in which the deep learning model was trained and tested with and without using the hourly flash rate feature. The experimental results show that in comparison, the CNN-based retrieval with lightning data significantly outscored the CNN-based retrieval without lightning data, highlighting the importance of lightning in deep learning-based precipitation retrieval. The flash rate feature can significantly help the CNN-based retrieval model improve the accuracy of the rainfall rate estimates, particularly for moderate and heavy precipitations.

However, it is worth noting that although the CNN-based retrieval with lightning data has exhibited encouraging results, several avenues for improvement remain unexplored. First, it is important to acknowledge that the study domain in this article is limited to a rectangular region of approximately 1600 km $\times$ 800 km in the southeastern United States. To develop an operational algorithm for precipitation retrieval, the coverage must be expanded to encompass the globe or at least the hemisphere. Therefore, future work will focus on developing a deep learning-based precipitation retrieval framework to extend the study domain and test the precipitation retrieval performance over a larger domain. Second, as the study domain expands, it is inevitable to consider the regions with complex terrain, such as mountains, lakes, and oceans. Incorporating terrain or elevation information as input features to the deep learning model could potentially enhance the accuracy of the precipitation retrieval framework. In addition to terrain information, other factors typically associated with precipitation, including water vapor, ground temperature, cloud height, and wind magnitude/direction, can also be integrated as input features to improve the deep learning-based precipitation retrieval framework. Finally, as the study domain expands and more data become available, transfer learning and continual lifelong learning techniques could be employed to enhance the efficiency and accuracy of the deep learning-based precipitation retrieval framework.

Acknowledgment

The authors would like to thank Dr. Nachiketa Acharya (NOAA/PSL and CU/CIRES) for his careful internal review on this article. We would also like to thank the anonymous reviewers for providing careful comments on this article. GOES-16 data used in this research are accessible from the NOAA National Centers for Environmental Information (NCEI) at https://www.ncdc.noaa.gov/airs-web/search. The intermediate features in the deep learning models as well as the model parameters are available upon request.

References

[1] C. Kidd, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, "So, how much of the earth's surface is covered by rain gauges?" *Bulletin of the American Meteorological Society*, vol. 98, no. 1, pp. 69–78, 2017.

[2] J. Zhang, K. Howard, C. Langston, B. Kaney, Y. Qi, L. Tang, H. Grams, Y. Wang, S. Cocks, S. Martinaitis *et al.*, "Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities," *Bulletin of the American Meteorological Society*, vol. 97, no. 4, pp. 621–638, 2016.

[3] C. Kummerow, Y. Hong, W. Olson, S. Yang, R. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, and 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.

[4] S.-A. Boukabara, K. Garrett, W. Chen, F. Iturbide-Sanchez, C. Grassotti, C. Kongoli, R. Chen, Q. Liu, B. Yan, F. Weng, R. Ferraro, T. J. Kleespies, and H. Meng, "Mirs: An all-weather 1dvar satellite data assimilation and retrieval system," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, pp. 3249–3272, 2011.

[5] R. J. Kuligowski, "A self-calibrating real-time goes rainfall algorithm for short-term rainfall estimates," *Journal of Hydrometeorology*, vol. 3, no. 2, pp. 112–130, 2002.

[6] L. Sun, H. Chen, Z. Li, and L. Han, "Cross validation of goes-16 and noaa multi-radar multi-sensor (mrms) qpe over the continental united states," *Remote Sensing*, vol. 13, no. 20, 2021.

[7] H. Chen, R. Cifelli, P. Xie, and E. J. Thompson, "Improving GOES-R/ABI rainfall estimates with ground-based radar observations using machine learning," in *AGU Fall Meeting Abstracts*, vol. 2020, Dec. 2020, pp. A142–0012.

[8] H. Chen, R. Cifelli, and P. Xie, "Precipitation identification and quantification using ABI and GLM observations aboard the GOES-R series: A machine learning perspective," in *AGU Fall Meeting Abstracts*, vol. 2021, Dec. 2021, pp. A35F–1700.

[9] H. Chen, Y. Yang, K. Hilburn, and M. Azimi-Sadjadi, "Impacts of the GLM measurements aboard the GOES-R series satellites on severe convective precipitation detection and retrievals," in *AGU Fall Meeting Abstracts*, vol. 2022, Dec. 2022, pp. GC32G–0688.

[10] Y. Hong, K.-L. Hsu, S. Sorooshian, and X. Gao, "Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system," *Journal of Applied Meteorology*, vol. 43, no. 12, pp. 1834–1853, 2004.

[11] 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, 2019.

[12] S. A. Upadhyaya, P.-E. Kirstetter, R. J. Kuligowski, and M. Searls, "Towards improved precipitation estimation with the goes-16 advanced baseline imager: Algorithm and evaluation," *Quarterly Journal of the Royal Meteorological Society*, vol. 148, no. 748, pp. 3406–3427, 2022.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[14] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[15] T. J. Schmit, S. S. Lindstrom, J. J. Gerth, and M. M. Gunshor, "Applications of the 16 spectral bands on the advanced baseline imager (abi)," *Journal of Operational Meteorology*, vol. 6, no. 4, pp. 33–46, 2018.

[16] A. Heidinger, M. Pavolonis, R. Holz, B. A. Baum, and S. Berthier, "Using calipso to explore the sensitivity to cirrus height in the infrared observations from npoess/viirs and goes-r/abi," *Journal of Geophysical Research: Atmospheres*, vol. 115, no. D4, 2010.

[17] Y. Yu, D. Tarpley, J. L. Privette, M. D. Goldberg, M. R. V. Raja, K. Y. Vinnikov, and H. Xu, "Developing algorithm for operational goes-r land surface temperature product," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 936–951, 2008.

[18] S. J. Goodman, R. J. Blakeslee, W. J. Koshak, D. Mach, J. Bailey, D. Buechler, L. Carey, C. Schultz, M. Bateman, E. McCaul Jr *et al.*, "The goes-r geostationary lightning mapper (glm)," *Atmospheric research*, vol. 125, pp. 34–49, 2013.

[19] A. Papadopoulos, T. G. Chronis, and E. N. Anagnostou, "Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale model," *Monthly Weather Review*, vol. 133, no. 7, pp. 1961–1977, 2005.

[20] K. A. Hilburn, I. Ebert-Uphoff, and S. D. Miller, "Development and interpretation of a neural-network-based synthetic radar reflectivity estimator using goes-r satellite observations," *Journal of Applied Meteorology and Climatology*, vol. 60, no. 1, pp. 3–21, 2021.

[21] 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.

[22] T. J. Schmit and M. M. Gunshor, "Chapter 4 - abi imagery from the goes-r series," in *The GOES-R Series*, S. J. Goodman, T. J. Schmit, J. Daniels, and R. J. Redmon, Eds. Elsevier, 2020, pp. 23–34. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128143278000044

[23] J. T. Schaefer, "The critical success index as an indicator of warning skill," *Weather and Forecasting*, vol. 5, no. 4, pp. 570 – 575, 1990.

[24] G. J. McLachlan, *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, 2005.

[25] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: a system for large-scale machine learning." in *Osdi*, vol. 16, no. 2016. Savannah, GA, USA, 2016, pp. 265–283.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[27] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint arXiv:1605.07678*, 2016.

[28] A. C. B. AYALA, J. J. Gerth, T. J. Schmit, S. S. Lindstrom, and J. P. NELSON III, "Parallax shift in goes abi data." *Journal of Operational Meteorology*, vol. 11, no. 2, 2023.

[29] K. Liu, J. He, and H. Chen, "Precipitation retrieval from fengyun-3d mwhts and mwri data using deep learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7619–7630, 2022.

[30] Z. Liu, Q. Yang, J. Shao, G. Wang, H. Liu, X. Tang, Y. Xue, and L. Bai, "Improving daily precipitation estimation in the data scarce area by merging rain gauge and trmm data with a transfer learning framework," *Journal of Hydrology*, vol. 613, p. 128455, 2022.

[31] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural networks*, vol. 113, pp. 54–71, 2019.

[32] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[33] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Speeding up convolutional neural networks with low rank expansions," *arXiv preprint arXiv:1405.3866*, 2014.

[34] R. Liu, J. Lehman, P. Molino, F. Petroski Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," *Advances in neural information processing systems*, vol. 31, 2018.

[35] H. Chen, L. Sun, R. Cifelli, and P. Xie, "Deep learning for bias correction of satellite retrievals of orographic precipitation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[36] B. Henn, A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, "An assessment of differences in gridded precipitation datasets in complex terrain," *Journal of hydrology*, vol. 556, pp. 1205–1219, 2018.

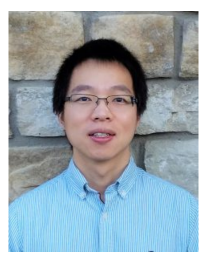
[37] C. Barthlott and N. Kalthoff, "A numerical sensitivity study on the impact of soil moisture on convection-related parameters and convective precipitation over complex terrain," *Journal of the atmospheric sciences*, vol. 68, no. 12, pp. 2971–2987, 2011.

[38] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[39] Z. Zeng, H. Chen, Q. Shi, and J. Li, "Spatial downscaling of imerg considering vegetation index based on adaptive lag phase," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[40] R. S. Schumacher and K. L. Rasmussen, "The formation, character and changing nature of mesoscale convective systems," *Nature Reviews Earth & Environment*, vol. 1, no. 6, pp. 300–314, 2020.

Yifan Yang received the B.S. degree in automation from Beijing University of Posts and Telecommunications, Beijing, China, in 2015, and the M.S. degree in computer engineering from Colorado State University (CSU), Fort Collins, CO, USA, in 2021. He is currently working toward the Ph.D. degree in Computer Engineering at CSU. His research interests include incremental learning, continual lifelong learning, and satellite remote sensing.



Haonan Chen (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Colorado State University (CSU), Fort Collins, CO, USA, in 2017.

He has been an Assistant Professor in the Department of Electrical and Computer Engineering at CSU since August 2020. He is also an Affiliate Faculty with the Data Science Research Institute (DSRI) at CSU. Before joining the CSU faculty, he worked with the Cooperative Institute for Research in the Atmosphere (CIRA) and the National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory, Boulder, CO, USA, from 2012 to 2020, first as a Research Student, then a National Research Council Research Associate and a Radar, Satellite, and Precipitation Research Scientist. His research interests span a broad range of remote sensing and multidisciplinary data science, including radar and satellite remote sensing of natural disasters, polarimetric radar systems and networking, clouds and precipitation observations and processes, big data analytics, and deep learning. He is the author/coauthor of over 80 peer-reviewed journal articles in these areas.

Dr. Chen is a member of the American Geophysical Union (AGU) Precipitation Technical Committee and American Meteorological Society (AMS) Committee on Hydrology. He was a recipient of the National Science Foundation (NSF) Faculty Early Career Development Program (CAREER) Award, Ralph E. Powe Junior Faculty Enhancement Award, IEEE Geoscience and Remote Sensing Society (GRSS) Early Career Award, American Meteorological Society (AMS) Editor's Award, CIRA Research and Service Initiative Award, and the Santimay Basu Prize from the International Union of Radio Science (URSI). He serves as an Associate Editor for the Journal of Atmospheric and Oceanic Technology, URSI Radio Science Bulletin, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING.

Kyle A. Hilburn received the B.S. degree in atmospheric sciences from the University of North Dakota, Grand Forks, ND, USA, in 2000, the M.S. degree in meteorology from Florida State University, Tallahassee, FL, USA, in 2002, and a Ph.D. degree in atmospheric science from Colorado State University, Fort Collins, CO, USA, in 2023.

Dr. Hilburn worked as a Scientist and Lead Software Developer with Remote Sensing Systems, Santa Rosa, CA, USA, during 2002 – 2016. His overall goal is to use passive and active microwave remote sensing to better understand the ocean–atmosphere system and its variability over time. He has worked on calibration, geophysical retrieval algorithms, and validation for a number of sensors including: AMSR-E, AMSR2, Aquarius, GMI, QuikSCAT, SeaWinds, SSM/I, SSMIS, TMI, and WindSat. He also studied the use of microwave satellite observations to better constrain the global water cycle.

He joined the Cooperative Institute for Research in the Atmosphere (CIRA) at Colorado State University as a Research Associate in 2016. His main focus has been on using GOES to better initialize convection in high-resolution weather models. Most recently, Dr. Hilburn has become fascinated in the power of artificial intelligence / machine learning to extract spatio-temporal patterns in satellite imagery. His recent research has used convolutional neural networks to extract precipitation latent heating rates from GOES ABI and GLM to inform numerical weather prediction models. He has developed approaches for visualizing and interpreting what the machine has learned. He is also involved in satellite data applications for monitoring and modeling wildfires.

Dr. Hilburn is a member of the American Geophysical Union and the American Meteorological Society.

Robert J. Kuligowski received a B.S. degree in Meteorology from Penn State University, State College, PA, USA, in 1991. Following three years as an operational weather forecaster at Accu-Weather, Inc., he returned to Penn State for graduate work, receiving his M.S. in Meteorology in 1996. To enhance his background in hydrology, he then switched to the Department of Civil and Environmental Engineering at Penn State for his Ph.D., which was completed in 2000.

Dr. Kuligowski has been a Meteorologist at NOAA/NESDIS/STAR since November 1999 and performs research and development on satellite-based rainfall estimation and nowcasting tools. His primary research interest is in estimating and predicting precipitation, as evidenced by his Master's work on using artificial neural networks to predict short-term precipitation from recent observations, and his Ph.D. work on assimilating satellite-based sounding estimates into a mesoscale numerical weather prediction model to improve fine-scale precipitation forecasts.

Robert Cifelli received the bachelor's degree in geology from the University of Colorado Boulder, Boulder, CO, USA, in 1983, the M.S. degree in hydrogeology from West Virginia University, Morgantown, WV, USA, in 1986, and the Ph.D. degree in atmospheric science from Colorado State University, Fort Collins, CO, in 1996.

He is a Radar Meteorologist with more than 25 years of experience in precipitation research and more than 80 publications in peer-reviewed journals. Since 2009, he has been leading NOAA scientists dedicated to improving precipitation and hydrologic prediction in complex terrain and other geographic regions. He currently leads the Hydrology Applications Division in the NOAA Physical Sciences Laboratory, Boulder, CO, USA. He completed a detail with the Bureau of Reclamation in 2016 through the President's Management Council Interagency Rotation Program. He continues to work closely with the Reclamation Research and Development Office on developing improved weather, climate, and water forecasts of extreme events to better meet water management needs. Dr. Cifelli also works closely with the NOAA's Office of Water Prediction (OWP) on the evaluation of forcings for hydrologic prediction, physical process representation in the National Water Model, and coordination of quantitative precipitation estimation and forecast efforts across NOAA to address internal OWP and external stakeholder needs.

Dr. Cifelli is a member of the American Geophysical Union (AGU), the American Meteorological Society (AMS), and the AMS Water Resources Committee. He is a fellow of the Cooperative Institute for Research in the Atmosphere (CIRA), and recipient of UCAR Award for Education Outreach, Department of Commerce Bronze Medal, and NOAA Administrators Award.