

Earth and Space Science



RESEARCH ARTICLE

10.1029/2020EA001423

Key Points:

- A proposed machine learning (ML) pipeline to improve precipitation quantification
- The proposed framework applies successfully to precipitation classification and quantification
- Interpretable ML bridges the precipitation physics and the black box mode

Correspondence to:

Y. Wen,
yixinwen@ou.edu; berry.wen@noaa.gov

Citation:

Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y., & Lambriksen, B. (2021). Advancing satellite precipitation retrievals with data driven approaches: Is black box model explainable? *Earth and Space Science*, 8, e2020EA001423. <https://doi.org/10.1029/2020EA001423>

Received 26 AUG 2020

Accepted 1 DEC 2020

Advancing Satellite Precipitation Retrievals With Data Driven Approaches: Is Black Box Model Explainable?

Zhi Li¹ , Yixin Wen^{2,3} , Mathias Schreier⁴, Ali Behrangi⁵ , Yang Hong¹, and Bjorn Lambriksen⁴

¹Hydrometeorology and Remote Sensing Laboratory, School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA, ²NOAA National Severe Storm Laboratory, Norman, OK, USA,

³Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK, USA, ⁴California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, USA, ⁵Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Abstract Satellite-based precipitation retrieval is an essential and long-standing scientific problem. With an increase of observational satellite data, the advances of data-driven approaches such as machine learning (ML)/deep learning (DL) are favored to deal with large data sets and potentially improve the accuracy of precipitation estimates. In this study, we took advantage of new technologies by wrapping up a ML/DL-based model pipeline (LinkNet segmentation + tree ensemble). This approach is applied to the Advanced Microwave Sounding Unit (AMSU) on National Oceanic and Atmospheric Administration 18 and 19 flight, and compared with the MultiRadar MultiSensor. Four simulations were configured to examine the performance gain by incorporating three components: (1) precipitation identification, (2) nonlocal features, and (3) precipitation classification. More importantly, we examined the interpretability of the “black box” model to get a better understanding of the underlying physical connections. First, the results by this model pipeline suggest the advantages of the ML model by reducing the systematic error and instantaneous error to a factor of two. Second, identifying precipitation pixels helps to reduce the systematic error by 130%, and predicting precipitation classification benefits improved correlations by 32%. Last, channels at higher frequencies (beyond 150 GHz) are favored to identify precipitation regions, and also channels at 89 and 150 GHz are ranked as the two most important features to precipitation retrieval. This study explores the potentials of AMSU precipitation estimates with ML algorithms and provides means of interpreting the models to facilitate the better prediction of precipitation.

Plain Language Summary Accurate precipitation estimations could undoubtedly benefit water resources management and decision makings. As the precipitation data explode, especially for remote sensing, there is a growing opportunity to leverage these myriads of data to advance the satellite retrieval algorithms. In this study, we apply this novel machine learning/deep learning pipeline with National Oceanic and Atmospheric Administration (NOAA)-18 and NOAA-19 satellite to compare with an operational retrieval algorithm. Beyond that, the black box model interpretability is explored to discover how the model predicts the outcome. The results prove that this model not only outperforms than the operational product, it but also makes physical sense with respect to its functionality.

1. Introduction

Precipitation is an essential component of global water cycle. Understanding precipitation (including rate, phase, and microphysics) is a long-standing scientific research topic. Heavy precipitation induced water-related natural hazards like flash flood, debris flow, and inundation cause numerous fatalities and economic losses (Hong et al., 2007, 2012; Li et al., 2020). Ground-radar based Quantitative Precipitation Estimation (QPE) like MultiRadar MultiSensor (MRMS) system provides nation-wide seamless radar mosaic and precipitation measurements to issue early warnings of such disasters.

Satellite remote sensing products are one of the novel technologies being applied to measure precipitation during the last 3 decades due to its broad spatial coverage and uninhibited view in complex terrains (Mei et al., 2014; Sarachi et al., 2015). Satellite-based precipitation products (SPPs) utilize the information provided by visible-infrared (VIS-IR) channel from geostationary (GEO) satellites and microwave (MW) channel

© 2020. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

from Low Earth Orbiting satellite (LEO). Despite GEO satellites on board has higher temporal resolution compared to LEO, they inherit some deficiencies such as nonpenetrable in precipitating clouds, and thus GEO-derived precipitation is an indirect estimate, associated with large uncertainties at fine scale. On the other side, MW sensors can penetrate clouds and are sensitive to hydrometeors in the atmosphere, yielding more accurate precipitation estimates (Adler et al., 2001).

The MW precipitation retrieval algorithms have been evolving throughout decades for better SPPs accuracy. Wilheit and Hutchison (1997) related the rainfall rates to the increase of the brightness temperature relative to the cold surface. Weng et al. (2003) utilized the low frequencies to derive rain rates and high frequencies to detect rainy clouds over ocean surface. The National Environmental Satellite Data, Information Service (NESDIS)/Microwave Surface and Precipitation Products System (MSSPS) then derived the IWP with two MW channels (89 and 150 GHz) in the Advanced Microwave Sounding Unit (AMSU) and correlated IWP to surface rain rate to make it operational (Vila et al., 2007). Laviola and Levizzani (2011) proposed a way to retrieve rainfall rate over land and ocean using water vapor absorption lines corresponding to the three 183 GHz channels on AMSU. Even though above methods search for the relationships between hydrometeors and brightness temperature, they are still constrained for three facets (1) large amounts of data to construct the empirical relations, (2) limited adaptivity to local environmental changes, (3) lack of accuracy to accommodate further applications (e.g., early warning of natural hazards).

To take advantage of massive data volumes in recent years, researchers have extensively introduced some data-driven approaches (Wang et al., 2020). Machine learning (ML) or deep learning (DL) due to its powerful capacity to solve highly nonlinear problems, is becoming ubiquitous across research fields. Some ML/DL based approaches in satellite precipitation retrievals, like Precipitation Estimates from Remotely Sensed Information using Artificial Neural Network family (Ashouri et al., 2015; Behrangi et al., 2009; Hong et al., 2004; Hsu et al., 1999; Sadeghi et al., 2019) have demonstrated their potentials. Regarding AMSU platform, Surussavadee and Staelin (2009) first collected multiple channels along with information from Fifth-Generation Penn State/NCAR Mesoscale Model (MM5) cloud resolving model to fit into a separate neural network over different surface types. Nevertheless, it is still challenging to determine the input features to be included in the “black box” model.

Additionally, most of the previous studies of precipitation estimates for AMSU only focus on the collocated pixel point (Surussavadee & Staelin, 2009; Vila et al., 2007; Weng et al., 2003). If a “parallax effect” is included, then it has to be taken into account that the respective interception altitude of different frequency channels and scan strategy result in location shift of the sensed object (Bauer et al., 1998; Guilloteau & Foufoula-Georgiou, 2020). Hence, some studies propose the use of neighboring pixels as auxiliary data (Guilloteau & Foufoula-Georgiou, 2020; Hong et al., 1999). Beyond that, Earth’s surface features, flight dynamics (e.g., azimuthal angle and ascending/descending orbit) contribute to the difference of observed signals and derived rainfall as well (Kummerow et al., 2015; Mo, 2010). In this study, we attempt to answer two scientific questions: (1) Is data-driven based approach applicable in AMSU MW precipitation retrieval? (2) How to interpret the model results with reference to the microphysics of precipitation characteristics? Specifically, the divided specific objectives are:

1. Exploration of the potentials of ML-based model in AMSU precipitation retrieval
2. Evaluation of the performance gain from precipitation identification, composing different input features, and precipitation classification
3. Physical interpretation of the segmentation and ensemble ML/DL model

To achieve that, we inject the level-2 AMSU product onboard National Oceanic and Atmospheric Administration (NOAA) 18 and 19 into the ML/DL models with MRMS data being the reference over the conterminous United States (CONUS). The following chapters are structured as: Section 2 introduces the four designed simulations and modeling details; Section 3 includes the data sets being used in this study; Section 4 presents the results of this study with a case study; and Section 5 concludes this study along with recommendations for future works.

2. Materials and Methods

The AMSU on board the NOAA15 was first launched in July 1998, aiming to accelerate the development of more atmospheric products, for example, Cloud Liquid Water (CLW), Rain Rate (RR), and Ice Water Path (IWP). Afterward, NOAA-16, NOAA-17, NOAA-18, and NOAA-19 were sequentially launched with AMSU sensors to compensate for the spatiotemporal gaps. AMSU contains A and B two modules. The AMSU-A includes 15 channels which are designed to obtain the atmospheric temperature profiling; AMSU-B has 5 channels at high frequencies measuring atmospheric moisture profile. AMSU adopts cross-track scan strategy that is perpendicular to the satellite's orbital track to gain a broader swath. Due to the frequency differences, the AMSU-A has an instantaneous view of 3.3° within $\pm 48^\circ$, providing a nominal spatial resolution at nadir of 48 km. The AMSU-B provides a spatial resolution at nadir of around 15 km because of higher frequency channels and makes three scans of 90 observations with a spacing of 1.1° . Due to the advantage of large swath of 1,650 km, it covers more global areas per scan than other constellation satellites.

2.1. Matching MRMS and AMSU Observations

The collocation of the MRMS data uses two steps: In the first step, we filtered the MRMS data sets that coincide with the overpass of the AMSU instrument. The orbits of NOAA-18 are very similar, and NOAA-18 follows NOAA-19 in orbit—with a time difference of ~ 1 h in 2012 and 2013. In the second step, the MRMS data points got averaged to AMSU resolution. The MRMS spatial resolution (~ 1 km) is much smaller than the footprint size of any AMSU-A (~ 45 km) or AMSU-B channel (~ 15 km) channel, so we average the MRMS data to make the data sets comparable. A collocation algorithm was applied, that looked for the MRMS data point closest to the center of the AMSU-footprint. Based on this center point, a mean value was calculated with the surrounding pixels to get a footprint average. We used all data points within a circular footprint for nadir, assuming a circular footprint of 15/45 km, respectively. We expanded this to elliptic footprints with increasing eccentricity for higher scan angles, based on calculated Earth-surface projection calculations. The multiple-use of MRMS points was allowed because a possible overlap of footprints was possible. We calculated two sets, one for the channels of AMSU-B (15 km at nadir) and the other one for channels of AMSU-A (45 km at nadir), so both options were available.

2.2. Data Split

Total 4,163 orbits are split into training, testing, and validation period with the relative sample sizes of 0.8, 0.1, and 0.1, respectively. For each orbit, we crop the granule into patches composed by 64×64 pixels. These patches are considered only when the areal mean of surface precipitation rate exceeds 0.5 mm/h to eliminate no-rain samples. Consequently, the selected imageries for training, testing, and validation category are 8693, 568, 661, respectively. The training samples can further be increased with data augmentation methods, but in this rainfall identification task, increase in sample sizes only bring marginal benefits.

2.3. Nonlocal Features

Apart from the auxiliary data set used in this study, nonlocal features (i.e., neighboring pixels) are included to account for the aforementioned “parallax effect” and to preserve the continuity of the precipitation structure. Inspired by Hong et al. (1999), four variability indices of the brightness temperature (VI, VC, VM, and VX) are included in this study, and their definitions are in the following. VI is to examine the mean 85 GHz brightness temperature difference; VC is defined as the brightness temperature difference between the center pixel and averaged value of eight surrounding pixels; VX is the maximum temperature enhancement due to emission for the 23.8 GHz; VM is the maximum depression due to scattering at 85 GHz.

$$VI = \frac{1}{n} \sum_{i=1}^n |T_{Bc} - T_{Bi}|$$

$$VC = T_{Bc} - \frac{1}{n} \sum_{i=1}^n T_{Bi}$$

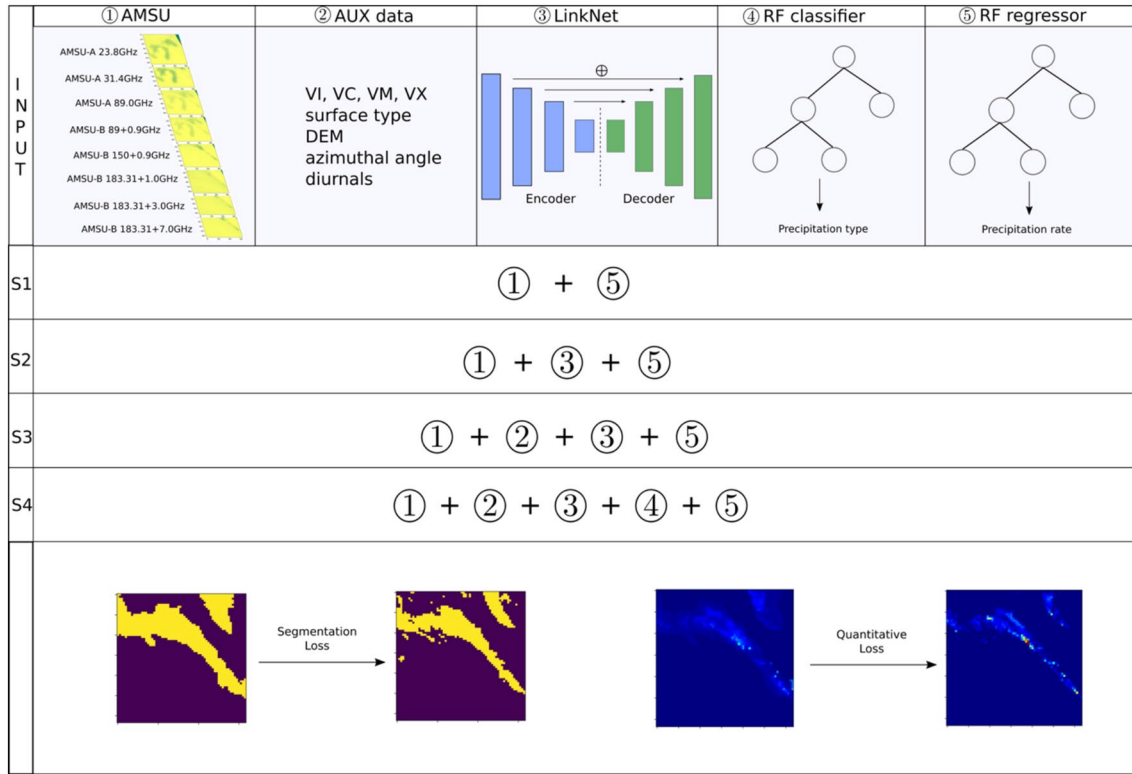


Figure 1. Schematic view of input components and four designed simulations (S1, S2, S3, and S4).

$$VX = \text{Max}(T_{Bc} - T_{Bi})$$

$$VM = \text{Min}(T_{Bc} - T_{Bi})$$

where n is the number of surrounding pixels; T_{Bc} is brightness temperature in the central pixel; T_{Bi} is brightness temperature of the surrounding pixels. The nonlocal features (VI, VC, VM, VX) are proven effectively to detect convective/stratiform precipitation in several studies (Hong et al., 1999; Kummerow, 1996).

2.4. Algorithm Description

The semantic segmentation is a prevalent deep learning model based on Convolutional Neural Networks (CNN) in image segmentation tasks (LinkNet; Chaurasia & Culurciello, 2017). It adopts a LinkNet structure in which input features are first upsampled by four times and then downsampled to its original size. Each block in Figure 1 is a combination of multiple kernels that are used to convolve the preprocessed satellite imagery. The advantages of the LinkNet structure over other frameworks are (1) minimum tunable parameters and (2) increased receptive field. Hereafter, we refer the rainfall identification problem (i.e., rain or no rain) to the segmentation task. The cost function is the weighted sum of Dice score, Intersection over Union (IOU) and Balanced Cross Entropy (BCE) as described in Table 1. With respect to the optimizer, the Rectified Adam (RAdam) is chosen, which is a variant of Adam that rectifies the variance of the adaptive learning rate (Liu et al., 2019). The results are then sigmoid transformed and filtered with thresholds to predict the dichotomy based on the optimized Receiver Operating Characteristic (ROC) curve. After rain/no-rain segmentation, the Random Forest (RF) is used as the ensemble algorithm for precipitation prediction. The numbers of estimators and max depth are set to 100 and 50 for all experiments in order to control variables, and the cost function used is the Mean Square Error (MSE). The schematic flowchart of the processes described is illustrated in Figure 1.

Table 1
Computational Metrics

Metrics	Formula	Best value
Dice (f1 score)	$\frac{2TP}{2TP + FP + FN}$	1
BCE	$-\beta p \log(\hat{p}) + (1 - \beta)(1 - p) \log(1 - \hat{p})$	0
IOU	$\frac{TP}{TP + FP + FN}$	1
POD/recall	$\frac{TP}{TP + FN}$	1
FAR	$\frac{FP}{TP + FP}$	0
CSI	$\frac{TP}{TP + FN + FP}$	1
BIAS	$\frac{\sum_{i=1}^n (x_i - y_i)}{\sum_{i=1}^n y_i}$	0
MAE	$\frac{\sum_{i=1}^n (x_i - y_i)}{n}$	0
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}$	0
CC	$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$	1

Notes. True positives, true negatives, false positives, and false negatives are denoted as TP, TN, FP, and FN, respectively in the table; x and y correspond to estimates and references. The sample sizes are referred as n .

Abbreviations: BCE, Balanced Cross Entropy; CC, correlation coefficient; CSI, critical success index; FAR, false alarm rate; IOU, Intersection over Union; MAE, Mean Absolute Error; POD, probability of detection; RMSE, Root Mean Square Error.

2.5. Training Strategy

Based on the separated samples discussed earlier, the training and validation samples are first injected into the algorithm to tune the model parameters. The validation samples are evaluated at the end of each epoch to gain the out-of-bag estimates. Afterward, the best parameters were determined based on the validation results instead of training results to improve the robustness of the algorithm. The number of epochs is set to 100, and the best parameter set is obtained at the 86th epoch with the lowest cost value. We acknowledge that having more training epochs may improve the training scores, but the final decision is made based on the best validation score.

2.6. Experiment Design

Four simulations are configured for the precipitation rate prediction to discover the underlying relations: (1) Simulation 1 incorporates AMSU-A 3 channels and AMSU-B 5 channels directly into ensemble trees without rain segmentation; (2) Simulation 2 includes rain segmentation before predicting rainfall rate; (3) Simulation 3 in contrast to Simulation 2, incorporates not only local features (8 channels), but also nonlocal features (VI, VC, VM, VX), surface types, diurnals, azimuthal angles, and DEM; (4) Simulation 4 adds ensemble tree classification to predict precipitation types prior to prediction in addition to the same process to Simulation 3. In summary, Simulation 2 is designed to investigate the importance of rain segmentation, and Simulation 3 uncovers the feature importance while Simulation 4 is to explore the performance gain by including precipitation types.

2.7. Evaluation Indicators

All the metrics containing categorical comparisons and continuous comparisons are summarized in Table 1. Among these metrics, Dice, BCE, IoU are used for the cost function in the training process and the remaining are used in the evaluation stage including probability of detection (POD), false alarm rate (FAR), critical success index (CSI), BIAS, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and correlation coefficient (CC).

2.8. Interpretable ML/DL

The ML models are oftentimes conceptualized as “black box” because of the perceived inability to understand how ML makes its predictions.

However, the importance of uncovering the “black box” is obvious: (1) Knowing how the model works; (2) Understanding the limitations of the model; (3) Innovating the model for better performance; (4) Identifying novel structures and relationships in data aided by ML to guide future research. For both segmentation and ensemble algorithms, we attempt to interpret the model intermediate products. For the CNN-based segmentation task, the Gradient-weighted Class Activation Mapping (Grad-CAM) approach proposed by Selvaraju et al. (2019) is adopted to perceive the relative importance of channels to rainfall identification problem. The basic idea is to use the gradients of the weights of the network propagating to the final convolutional layer to produce a localized heatmap which highlights the importance of each input features. With respect to the ensemble trees-based algorithm, the respective feature importance is calculated based on the averaged decrease of impurity from each feature. It is worth noting that the inferred importance is only relative to each feature, which is also regarded the limitation of this mean-impurity based approach. That is to say, the intercorrelated features may be less significant because the algorithm only needs to utilize one while the others are recognized as redundant information.

Table 2
Channels From AMUS-A and -B Used in This Study

Name	Frequency (GHz)	Polarization	NEDT ^a (K)
c1_ AMSU-A	23.8	Vertical	0.30
c2_ AMSU-A	31.4	Vertical	0.30
c15_ AMSU-A	89.0	Vertical	0.50
c1_ AMSU-B	89.9 ± 0.9	Vertical	0.37
c2_ AMSU-B	150 ± 0.9	Vertical	0.84
c3_ AMSU-B	183.31 ± 1.0	Vertical	1.06
c4_ AMSU-B	183.31 ± 3.0	Vertical	0.70
c5_ AMSU-B	183.31 ± 7.0	Vertical	0.60

^aNEDT stands for noise-equivalent temperature which is used to quantify the sensitivity of the detector of thermal radiation. The channel number is followed by “c.”

3. Data Sets

3.1. Advanced Microwave Sounding Unit

AMSU inferred rainfall rate has been made operational in the Microwave integrated Retrieval System (MiRS) at (<https://www.star.nesdis.noaa.gov/mirs/index.php>). The current algorithm applies 1D Variational Retrieve Scheme (1DVRs) to provide robust hydrometeorological variables including the CLW and IWP at different channels, followed by a vertical integration to obtain the vertical profile. Thereafter, the rainfall rate variable is calculated by a multilinear regression model with parameters trained offline over different surface types (Boukabara et al., 2013; Iturbide-Sanchez et al., 2011). In this study, we obtained this operational product as benchmark to compare with our data-driven approach.

There are 4,163 orbits in total overlaid with the CONUS, starting from March 1, 2012 to July 30, 2013. In this study, three window channels for AMSU-A on board NOAA18 and NOAA19 are selected with the frequency at 23.8 GHz, 31.4 GHz, and 89.0 GHz because they are linked to atmospheric moistures and are sensitive to intense and moderate rain; the other

five scattering channels for AMSU-B are all included because they are more sensitive to snow and light rains over land. The selected channels are summarized in Table 2.

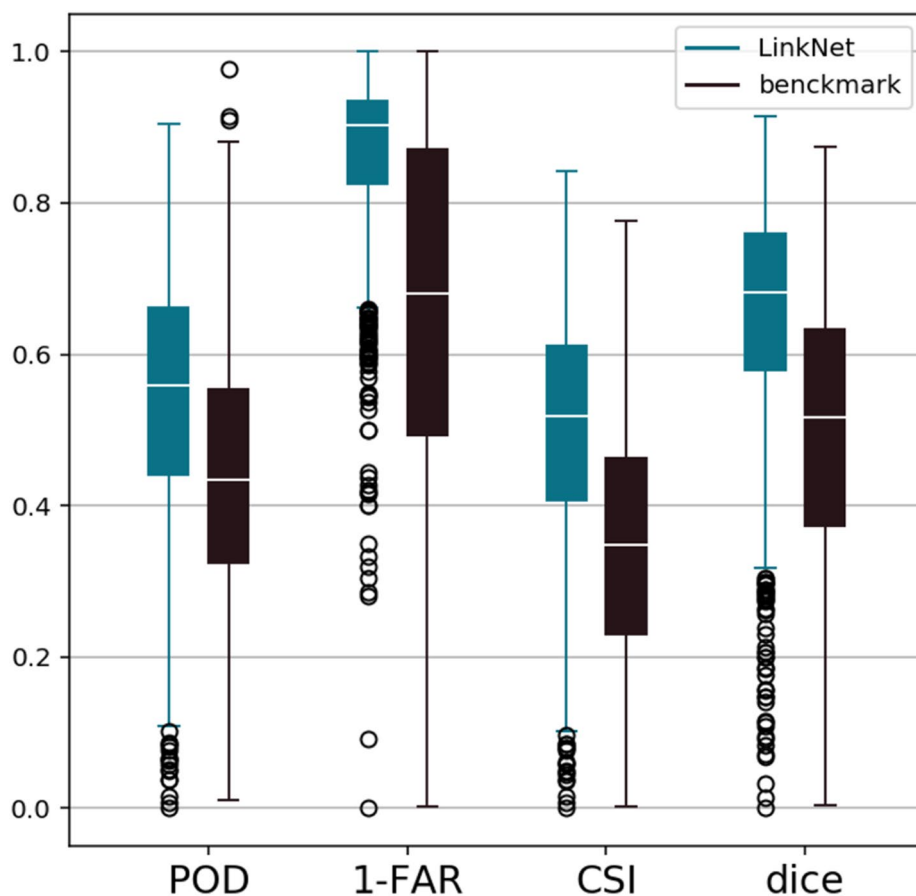


Figure 2. The boxplot of segmentation results for four categories. Each sample is the 64 by 64 patch, comparing LinkNet model (dark green) and the benchmark (black) against MRMS. CSI, critical success index; FAR, false alarm rate; MRMS, MultiRadar MultiSensor; POD, probability of detection.

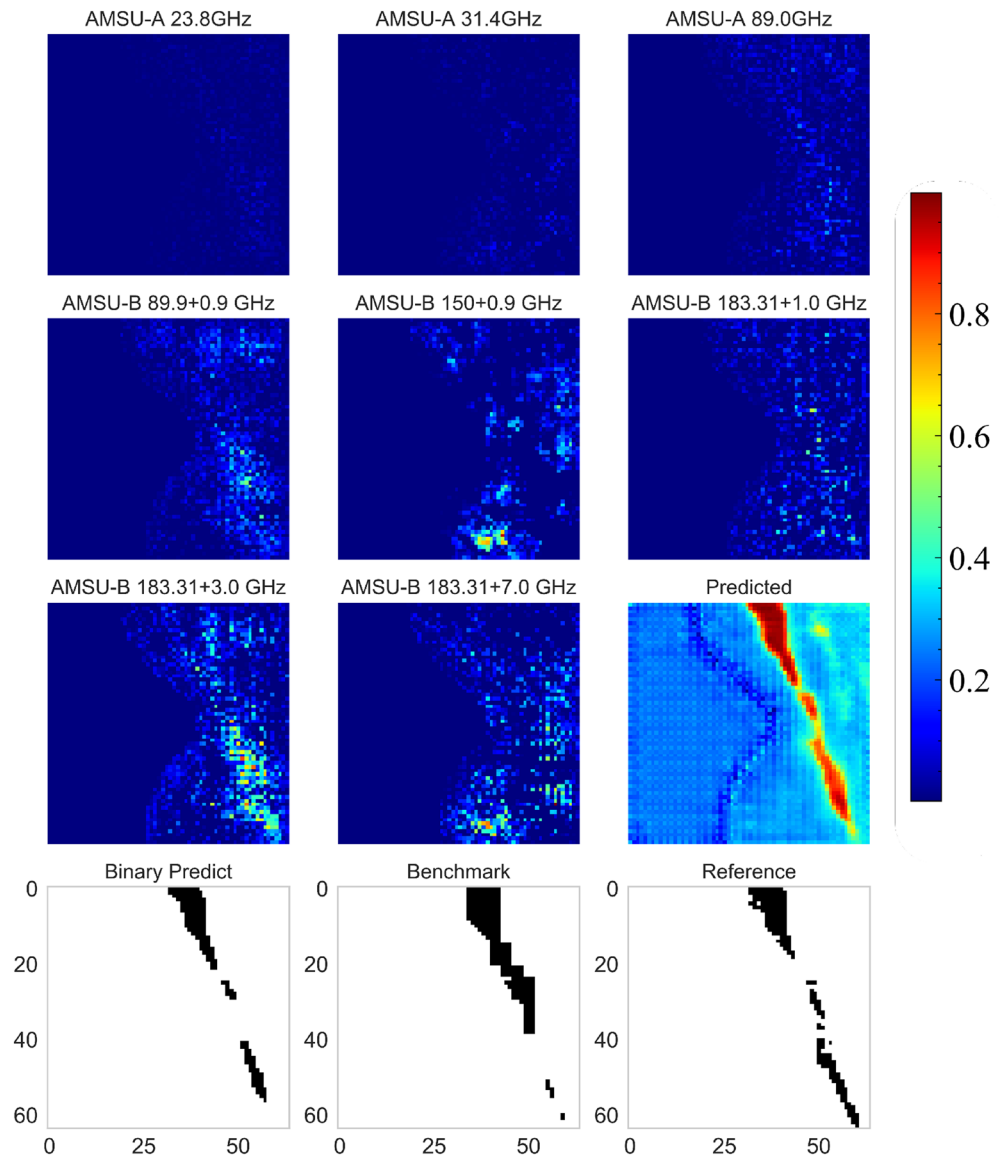


Figure 3. The weights of each channel contributed to the final prediction. The binary predict is obtained from a single threshold approach.

3.2. MultiRadar MultiSensor

The MRMS product provides a radar-gauge merged precipitation estimates over the CONUS (Zhang et al., 2016). It integrates 180 operational radars, including 146 S-band and 30 C-band radars, creating a seamless 3D radar mosaic across the CONUS and Southern Canada. The rigorous quality control steps have made this product been one of the most accurate QPEs in the CONUS (Chen et al., 2020; Li et al., 2020) and also made it intensively applied to flash flood monitoring. The precipitation flags in the MRMS provides eight distinct features in total: missing, no precipitation, cool stratiform, warm stratiform, snow, overshooting, convective, hail, and warm rain.

3.3. Auxiliary Data

Beyond precipitation measurements, the surface type data set and the Digital Elevation Model (DEM) data are included from https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MCD12C1_T1 and https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SRTM_RAMP2_TOPO. The surface type data contain two main categories:

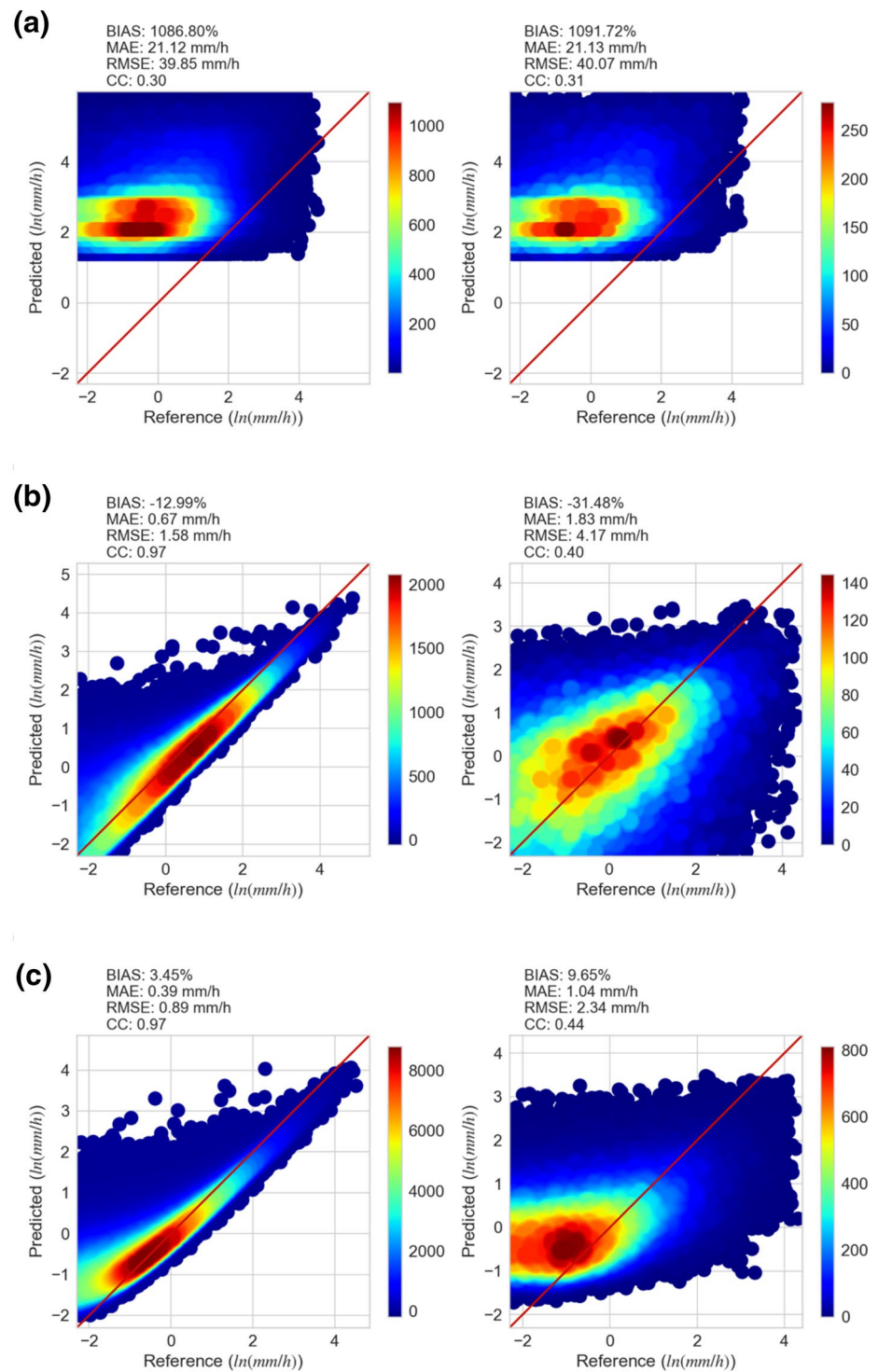


Figure 4. Scatter plot of benchmark and four simulation results. In each subgraph, the left panel is training period and the right is testing period: (a) Benchmark, (b) Simulation 1, (c) Simulation 2, (d) Simulation 3, and (e) Simulation 4. CC, correlation coefficient; MAE, Mean Absolute Error; RMSE, Root Mean Square Error.

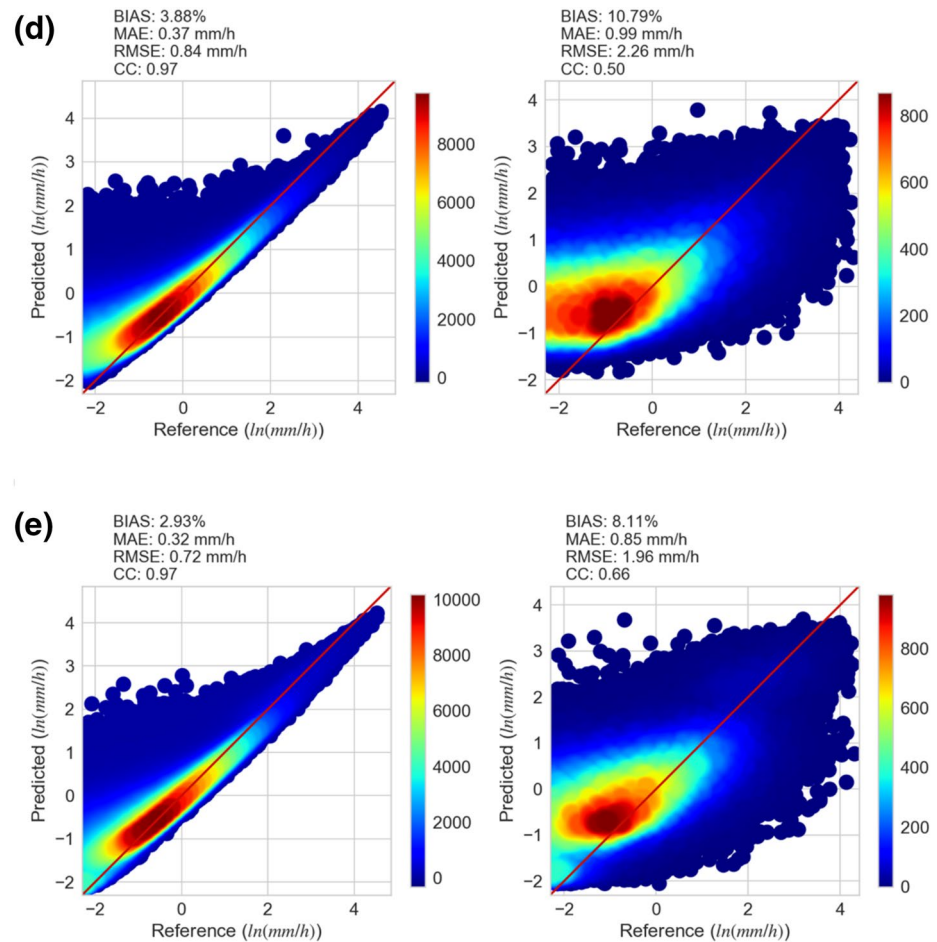


Figure 4. continued

ries (land and water). Over land, nine classes of natural vegetation, three classes of developed lands, two classes of mosaic lands, and three classes of nonvegetated lands are further specified to compensate the distinct emissivity from different land covers. Both the surface type and the DEM are retrieved from orbit locations with nearest neighbors.

4. Results and Discussions

4.1. Rainfall Segmentation

Figure 2 illustrates the performance (four categorical indices) of the DL-based approach (LinkNet) and benchmark. For the PODs, the indicator of the detectability, the median value is around 0.56%, 30% higher than benchmark; the 25th and 75th percentile of the values for the LinkNet are 0.42 and 0.65 while 0.35 and 0.56 for the benchmark, which indicates the advancements of the LinkNet in rainfall detection. Additionally, the LinkNet also has prominent improvement of 1-FAR over the benchmark as the median value rises from 0.7 to 0.9; the other two interquartile are 0.82 (25th) and 0.92 (75th) for the LinkNet while 0.5 (25th) and 0.85 (75th) for benchmark. Notably, the interquartile range for the LinkNet (0.1) is 71% lower than benchmark, which indicates the variance is rather small. The CSI, as a combination of POD and FAR, is 56% higher than the benchmark for LinkNet with respect to the median. The Dice score is another variant of diagnostic testing that weights twice on the true positives (TP), and the median Dice score for the LinkNet is still about 40% higher than benchmark. In summary, the deep learning-based approach is significantly improved across all the indicators than the benchmark.

Table 3
Performance Gains (%) for Four Simulations

Item	BIAS (%)	MAE (%)	RMSE (%)	CC (%)	Baseline
Simulation 1	−102.9	−91.4	−89.6	29.0	Benchmark
Simulation 2	−130.7	−42.9	−43.9	10	Simulation 1
Simulation 3	11.8	−4.81	−3.42	13.6	Simulation 2
Simulation 4	−24.8	−14.1	−13.7	32.0	Simulation 3

Abbreviations: CC, correlation coefficient; MAE, Mean Absolute Error; RMSE, Root Mean Square Error.

With the improved rain segmentation performance, it is still doubtful about the logics behind: How does the deep learning model make decisions? By further applying the Grad-CAM (Selvaraju et al., 2019), one can interpret the black box and perceive what is behind the scene. In Figure 3, the weights of each input channels are depicted from which the higher frequency channels (i.e., beyond 89 GHz) contribute significantly more to the final end-product than the low frequencies (i.e., 23.8 and 31.4 GHz). Especially at 183.31 ± 3.0 GHz, it shows the highest similarity with the final product in the southeast. Despite the scattering signal is even stronger at 183.31 ± 7.0 GHz with further penetration compared to other two channels at 183.31 GHz, it is possibly saturated by surface emission (Edel et al., 2019; G. Hong et al., 2005). This result aligns well with previous studies that mainly relied on the scattering signals to identify rain or delineate precipitating clouds (Ferraro et al., 2000; Grody, 1991; G. Hong et al., 2005; Weng et al., 2003).

4.2. Precipitation Predictions

The scatter plot from Figure 4 depicts the performance of the benchmark and four simulations for both the training period and testing period; Table 3 complements the illustration of performance gain with cascading improvement. Since the testing period is more representative to attest the robustness of the algorithm, we thus mainly focus on the testing results. First, the benchmark offsets from the reference the most in Figure 4a, with the systematic bias over 1,000% and the instantaneous errors (MAE and RMSE) are 21.13 and 40.07 mm/h; the CC is about 0.30, which is not considered successful. Figure 4b shows the testing results and validation results from Simulation 1, where the eight channels from the AMSU are injected into ensemble algorithm without identifying rain region. All the indicators have demonstrated profound improvements contrasting benchmark (BIAS = −30%, MAE/RMSE = 1.83/4.14 mm/h, CC = 0.4). However, because of the imbalanced sample sizes of rainy and nonrainy pixels, the algorithm has been trained mostly without rains and thus results in an underestimation. After the implementation of rain segmentation in

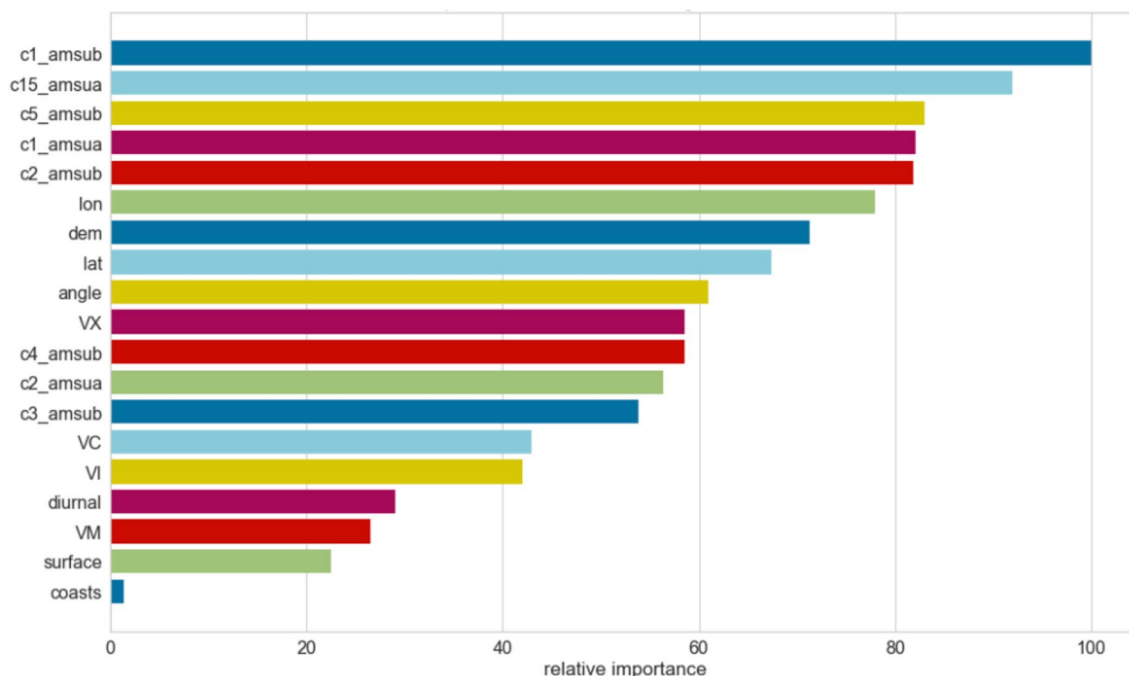


Figure 5. The relative importance (percentage) for each feature for precipitation prediction.

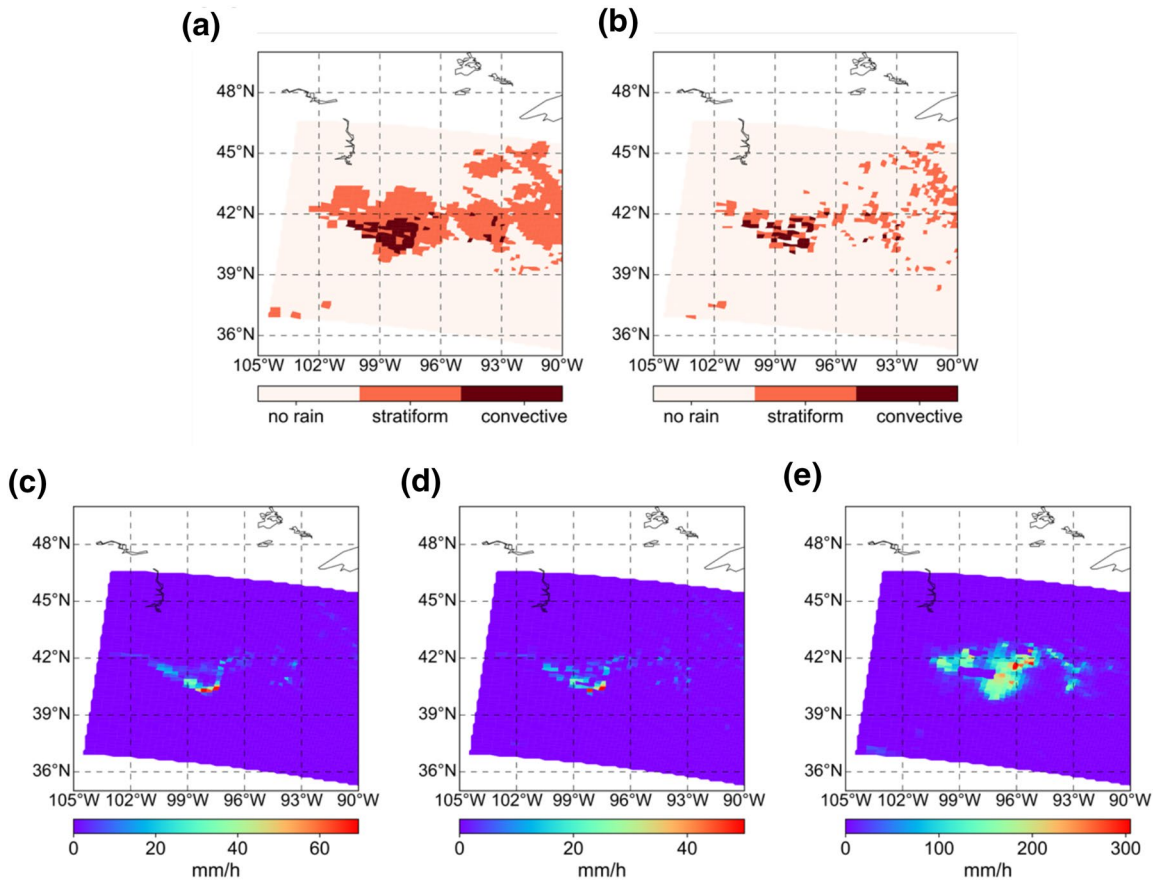


Figure 6. A case study of a summer storm (May 28, 2013). (a) MRMS observed precipitation type, (b) predicted precipitation type, (c) MRMS observed precipitation rate, (d) predicted precipitation rate by the Simulation 4, and (e) precipitation rate by the benchmark. MRMS, MultiRadar MultiSensor.

Simulation 2, the underestimation issue has been addressed as shown in Figure 4c with the BIAS down to 9%; the MAE, RMSE, and CC for Simulation 2 have been further improved to 1.04 mm/h, 2.34 mm/h, and 0.44 in contrast to Simulation 1. With the inclusion of nonlocal features in Simulation 3, the bias slightly increases (BIAS = 10.79%) with others indices (MAE, RMSE, CC) are all improved (0.99 mm/h, 2.26 mm/h, 0.5) as opposed to Simulation 2. Lastly, the feature of precipitation type dramatically improves the performance compared to all other simulations. In contrast to Simulation 3 standalone, the bias and MAE (RMSE) have reduced by 24.8% and 14.1% (13.2%), accompanied with CC rising from 0.50 to 0.66.

Table 3 shows the relative percent changes in each simulation referenced on the corresponding baselines. Among these changes, the most significant one for instantaneous errors (i.e., MAE and RMSE) comes from the ensemble tree algorithm in replacement of the original retrieval which only considered two channels; the largest improvement for the BIAS (130.7%) is by including the segmentation algorithm which reduces zero-rain samples to avoid overweighting on no-rain; for the CC, the greatest change occurs from Simulation 3 to Simulation 4 by adding the prediction of precipitation types with which the predicted precipitation rate can line up with respective precipitation mechanisms (i.e., higher rates for rainfall and lower for snowfall). Notably, nonlocal features only bring marginal gains (below 5% for the MAE and RMSE and even worse BIAS) but it still improves the CC by 13.6%.

In summary, the progressive improvements are from the cascading simulations with more features. First, identifying rainfall regions with segmentation algorithm benefits addressing underestimation problems and controls the bias to be around 10%. Second, the nonlocal features slightly help to reduce the instantaneous errors but increase the correlation. Last, the predicted precipitation types help to largely improve the metrics by above 10%, which illustrates the importance of precipitation classification.

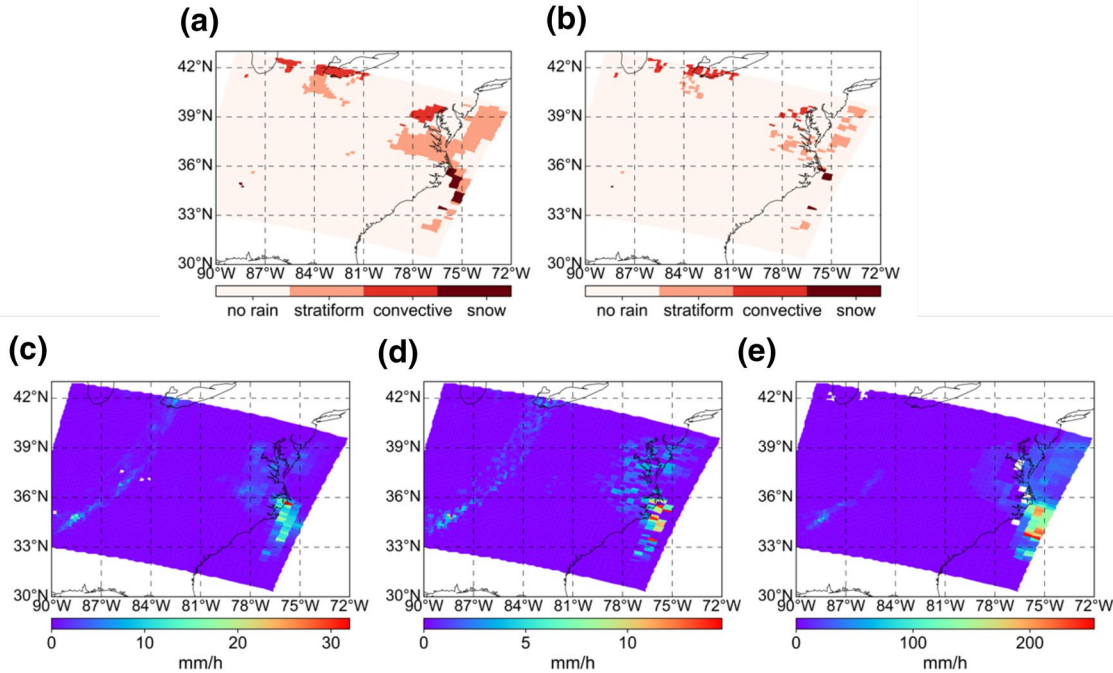


Figure 7. A case study in 2013 North America Blizzard (February 8, 2013). (a) MRMS observed precipitation type, (b) predicted precipitation type, (c) MRMS observed precipitation rate, (d) predicted precipitation rate by the Simulation 4, and (e) precipitation rate by the benchmark. MRMS, MultiRadar MultiSensor.

Figure 5 shows the relative importance of each feature based on the impurity node method. Among these features, the AMSU-B channel 1 (89.9 ± 0.9 GHz) ranks as the most important feature, followed by the AMSU-A channel 15 at frequency 89 GHz. Moreover, the AMSU-B channel 2 (150 ± 0.9 GHz), 5 (183.31 ± 7.0 GHz), and the AMSU-A channel 1 (23.8 GHz) all exceed the relative importance of 80%. Notably, the 89 GHz and 150 GHz channels are currently used to derive the rain rate in the benchmark, which is a physically based approach. Consequently, the data-driven approach can reveal the physics behind. Some localized features such as longitude, latitude, and the DEM have the relative importance beyond 60%, which suggests the location/terrain is a unique feature to distinguish. The nonlocalized features, that is, the VX, VC, VI, and VM are relatively less significant, possibly due to the large grid size (i.e., 40 km) in which the parallel effect is not prominent. Notably, even though surface types are ranked as the least important feature, it does not imply surface types are not correlated with rainfall rates. This is because surface type information is already reflected in the emission map (brightness temperature), and therefore the node-based importance ranking gives less weights on the redundant information.

4.3. Case Study

To further affirm the advances of Simulation 4 in extreme events, we apply the same pipeline with the testing data in NOAA19. Two representative events are selected to investigate the performance, one for a convective rainfall event, and the other one for a snowstorm.

4.3.1. Convective Precipitation

In May 2013, a significant tornado outbreak swept the Midwest CONUS and Great Plains. This system produced a convective rainfall that resulted in a record-breaking rainfall in North Dakota and Nebraska. Figure 6 depicts the performance of our proposed system. It is obvious that the benchmark dramatically overestimates the rainfall rates, which is on par with the previous investigations; however, our model successfully predicts the spatial pattern and the magnitude of the rainfall rates. With respect to the precipitation types, the convective structure in the center (99W, 40N) is well captured by our model. However, the surrounding regions with stratiform rainfall have been underestimated. Overall, this case demonstrates the

capacity of our model to predict the convective precipitation type, preserve the spatial patterns, and reduce the systematic bias compared to the benchmark.

4.3.2. 2013 North American Blizzard

In the early February of 2013, the heavy snowstorm and hurricane-force winds in the northeastern United States caused tremendous damages to the local economy. The snowfall rates in Connecticut and Long Island reached up to 7.5–10 cm/h and the radar reflectivity exceeded 55dBz (Ganetis & Colle, 2015). The NOAA19 flew over these regions at 9 a.m. in the morning that, is able to record the early stage of this event. Figure 7 shows the prediction of precipitation type and rates. First, our model captures the major precipitation systems spatially, such as the dominating convective precipitation in the north and the snowstorm in the coast. However, similar to the convective case, this model still underestimates the surrounding pixels in a precipitation system, which could be improved by applying spatial filter or morphological detections. Second, in terms of the rates, our model certainly exhibits less error compared to the benchmark. Again, the spatial features are well preserved. In this case, our model could predict the snowfall comparably with the MRMS system; while future improvement is also needed to address the underestimations.

This study serves as a broad assessment on this new developed system, with comparison to the existing system. It is obvious that our model outperforms original precipitation estimates across all metrics. Beyond that, this system can successfully predict precipitation types, which is proven to be essential when mapping different precipitation types into rates. However, it might be challenging for a ML/DL model to adapt to local environmental changes. To realize this point and understand the limitations of this model, future studies will focus on the transferability and assessments on different surface types.

5. Conclusions

In this study, we explored the applicability of the ML/DL based approach in satellite precipitation retrieval and interpretability of the “black box” model with connection to physical understanding. The main findings are summarized as follows:

- (1) This data-driven approach (the LinkNet segmentation + ensemble tree) can efficiently predict the surface rain rate with reduced bias, MAE, and RMSE by 99.2%, 96.0%, and 95.0% and with increased CC by 120% compared to benchmark
- (2) Identifying rainfall sufficiently reduces the underestimation (130% reduction on BIAS) caused by unbalanced training samples; Including precipitation types further improves the correlation by 32% between predictions and observations. However, the nonlocal features only bring marginal benefit compared to other gains
- (3) For rain/no rain segmentation, channels at higher frequencies are more efficient to identify rainfall because on land, scattering signals are more distinguishable from the background surface emissions
- (4) For precipitation prediction, the 89 and 150 GHz channels are ranked as the most important features that are also used in the existing retrieval algorithm. Beyond that, the combination of AMSU-A and AMSU-B channels helps to distinguish precipitation signals. Surface types even though are an important indicator, can be learned from the brightness temperature and hence are ranked as the least important feature

This study explores the applicability and efficiency of machine learning in the satellite precipitation retrievals and also how we can interpret the ML “black box” to connect with the physics. We acknowledge that the performance could be further improved by tuning the hyperparameters throughout grid search or evolutionary algorithms. For the future tasks, it is worth developing the percentile-based ensembles that output a range of rainfall rates instead of a deterministic value. Such approach can meet different user demands in operation. For instance, for extreme weather prediction, one may mainly focus on high percentiles to reduce miss-capturing; for drought forecasts, one will rely on low percentiles to safeguard future water management. To other existing SPPs, this work provides a framework that could be utilized to enhance the quantification of precipitation in a more comprehensive way.

Data Availability Statement

The AMSU brightness temperature data on board NOAA-18 and NOAA-19 are accessed from the website (<https://www.ncdc.noaa.gov/cdr/fundamental/amsu-brightness-temperature>) (Ferraro et al., 2016), and the MRMS gauge-corrected hourly rainfall rate is accessed from <https://mtarchive.geol.iastate.edu> (Zhang et al., 2016). Data are retrieved in the continental United States and from 2012 to 2013.

Acknowledgments

We acknowledge NASA Oklahoma EPSCoR Research Initiation Grant and Multidisciplinary Online Training Program in Spring 2020 on Big Data, High Performance Computing and Atmospheric Sciences Information. The authors would like to appreciate the efforts done by NOAA NSSL team for making the MRMS data accessible. The first author is sponsored by University of Oklahoma Hydrology and Water Security (HWS) program (<https://www.ouhydrologyonline.com/>) and Graduate College Hoving Fellowship.

References

- Adler, R. F., Kidd, P., C. G., Morrissey, M., & Goodman, M. H. (2001). Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bulletin of the American Meteorological Society*, 82, 1377–1396.
- Ashouri, H., Hsu, K., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., et al. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96, 69–83. <https://doi.org/10.1175/BAMS-D-13-00068.1>
- Bauer, P., Schanz, L., & Roberti, L. (1998). Correction of three-dimensional effects for passive 679 microwave remote sensing of convective clouds. *Journal of Applied Meteorology*, 37(12), 1619–1632. [https://doi.org/10.1175/1520-0450\(1998\)037<1619:COTDEF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1998)037<1619:COTDEF>2.0.CO;2)
- Behrangi, A., Hsu, K.-I., Imam, B., Sorooshian, S., Huffman, G. J., & Kuligowski, R. J. (2009). PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis. *Journal of Hydrometeorology*, 10(6), 1414–1429. <https://doi.org/10.1175/2009jhm1139.1>
- Boukabara, S. A., Garrett, K., Grassotti, C., Iturbide-Sanchez, F., Chen, W., Jiang, Z., et al. (2013). A physical approach for a simultaneous retrieval of sounding, surface, hydrometeor, and cryospheric parameters from SNPP/ATMS. *Journal of Geophysical Research: Atmospheres*, 118(22), 600–612. <https://doi.org/10.1002/2013jd020448>
- Chaurasia, A., & Culurciello, E. (2017). *LinkNet: Exploiting encoder representations for efficient semantic segmentation*. Paper presented at IEEE Visual Communications and Image Processing. <https://doi.org/10.1109/vcip.2017.8305148>
- Chen, M., Nabih, S., Brauer, N. S., Gao, S., Gourley, J. J., Hong, Z., et al. (2020). Can remote sensing technologies capture the extreme precipitation event and its cascading hydrological response? A case study of hurricane harvey using EF5 modeling framework. *Remote Sensing*, 12, 445. <https://doi.org/10.3390/rs12030445>
- Edel, L., Rysman, J.-F., Claud, C., Palerm, C., & Genthon, C. (2019). Potential of passive microwave around 183 GHz for snowfall detection in the Arctic. *Remote Sensing*, 11, 2200. <https://doi.org/10.3390/rs11192200>
- Ferraro, R. R., Meng, H., Yang, W., & Moradi, I., & NOAA CDR Program. (2016). *NOAA climate data record (CDR) of AMSU-A brightness temperature, version 1*. Silver Spring, MD: NOAA National Centers for Environmental Information (NCEI). <https://doi.org/10.7289/V53RQXD>
- Ferraro, R. R., Weng, F., Grody, N. C., & Zhao, L. (2000). Precipitation characteristics over land from the NOAA-15 AMSU sensor. *Geophysical Research Letters*, 27(17), 2669–2672. <https://doi.org/10.1029/2000gl011665>
- Ganetis, S. A., & Colle, B. A. (2015). The thermodynamic and microphysical evolution of an intense snowband during the Northeast U.S. Blizzard of 8–9 February 2013. *Monthly Weather Review*, 143, 4104–4125. <https://doi.org/10.1175/MWR-D-14-00407.1>
- Grody, N. C. (1991). Classification of snow cover and precipitation using the special sensor microwave imager. *Journal of Geophysical Research*, 96, 7423–7435. <https://doi.org/10.1029/91JD00045>
- Guilloteau, C., & Foufoula-Georgiou, E. (2020). Beyond the pixel: Using patterns and multiscale spatial information to improve the retrieval of precipitation from space-borne passive microwave imagers. *Journal of Atmospheric and Oceanic Technology*, 37(9), 1571–1591. <https://doi.org/10.1175/JTECH-D-19-0067>
- Hong, Y., Adler, R., Negri, A., & Huffman, G. (2007). Flood and landslide applications of near real-time satellite rainfall products. *Natural Hazards*, 43, 285–294. <https://doi.org/10.1007/s11069-006-9106-x>
- Hong, Y., Chen, S., Xue, X., & Hodges, G. (2012). Global precipitation estimation and applications. In N.-B. Chang, & Y. Hong (Eds.), *Multiscale hydrologic remote sensing* (pp. 371–386). Boca Raton: CRC Press.
- Hong, G., Heygster, G., Miao, J., & Kunzi, K. (2005). Detection of tropical deep convective clouds from AMSU-B water vapor channels measurements. *Journal of Geophysical Research*, 110, D05205. <https://doi.org/10.1029/2004JD004949>
- Hong, Y., Hsu, K., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology and Climatology*, 43, 1834–1853. <https://doi.org/10.1175/JAM2173.1>
- Hong, Y., Kummerow, C. D., & Olson, W. S. (1999). Separation of convective and stratiform precipitation using microwave brightness temperature. *Journal of Applied Meteorology and Climatology*, 38, 1195–1213. [https://doi.org/10.1175/1520-0450\(1999\)038<1195:SOCASP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<1195:SOCASP>2.0.CO;2)
- Hsu, K., Gupta, H., Gao, X., & Sorooshian, S. (1999). Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resources Research*, 35, 1605–1618. <https://doi.org/10.1029/1999WR900032>
- Iturbide-Sanchez, F., Boukabara, S.-A., Chen, R., Garrett, K., Grassotti, C., Chen, W., & Weng, F. (2011). Assessment of a variational inversion system for rainfall rate over land and water surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9), 3311–3333. <https://doi.org/10.1109/tgrs.2011.2119375>
- Kummerow, C., Olson, W. S., & Giglio, L. (1996). A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 34, 1213–1232.
- Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N., Ferraro, R., Joseph Munchak, S., & Petkovic, V. (2015). The evolution of the Goddard profiling algorithm to a fully parametric scheme. *Journal of Atmospheric and Oceanic Technology*, 32, 2265–2280. <https://doi.org/10.1175/JTECH-D-15-0039.1>
- Levizzani, V., Laviola, S., & Cattani, E. (2011). Detection and measurement of snowfall from space. *Remote Sensing*, 3, 145–166. <https://doi.org/10.3390/rs3010145>
- Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., et al. (2020). Cross-examination of similarity, difference and deficiency of gauge, radar and satellite precipitation measuring uncertainties for extreme events using conventional metrics and multiplicative triple collocation. *Remote Sensing*, 12, 1258. <https://doi.org/10.3390/rs12081258>
- Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2019). On the variance of the adaptive learning rate and beyond. *Archive preprint*. arXiv: 1908.03265.

- Mei, Y., Anagnostou, E. N., Nikolopoulos, E. I., & Borga, M. (2014). Error analysis of satellite precipitation products in mountainous basins. *Journal of Hydrometeorology*, 15, 1778–1793. <https://doi.org/10.1175/JHM-D-13-0194.1>
- Mo, T. (2010). A study of the NOAA near-nadir AMSU-A brightness temperatures over Antarctica. *Journal of Atmospheric and Oceanic Technology*, 27, 995–1004. <https://doi.org/10.1175/2010JTECHA1417.1>
- Sadeghi, M., Asanjan, A. A., Faridzad, M., Nguyen, P., Hsu, K., Sorooshian, S., & Braithwaite, D. (2019). PERSIANN-CNN: Precipitation estimation from remotely sensed information using artificial neural networks–convolutional neural networks. *Journal of Hydrometeorology*, 20, 2273–2289. <https://doi.org/10.1175/JHM-D-19-0110.1>
- Sarachi, S., Hsu, K., & Sorooshian, S. (2015). A statistical model for the uncertainty analysis of satellite precipitation products. *Journal of Hydrometeorology*, 16, 2101–2117. <https://doi.org/10.1175/JHM-D-15-0028.1>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2019). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128, 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Surussavadee, C., & Staelin, D. H. (2009). Satellite retrievals of arctic and equatorial rain and snowfall rates using millimeter wavelengths. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11), 3697–3707. <https://doi.org/10.1109/TGRS.2009.2029093>
- Vila, D., Ferraro, R., & Joyce, R. (2007). Evaluation and improvement of AMSU precipitation retrievals. *Journal of Geophysical Research*, 112, D20119. <https://doi.org/10.1029/2007JD008617>
- Wang, C., Xu, J., Tang, G., Yang, Y., & Hong, Y. (2020). Infrared precipitation estimation using convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59, 8612–8625. <https://doi.org/10.1109/TGRS.2020.2989183>
- Weng, F., Zhao, L., Ferraro, R. R., Poe, G., Li, X., & Grody, N. C. (2003). Advanced microwave sounding unit cloud and precipitation algorithms. *Radio Science*, 38, 8068. <https://doi.org/10.1029/2002RS002679>
- Wilheit, T. T., & Hutchison, K. D. (1997). Water vapour profile retrievals from SSM/T-2 data constrained by infrared-based cloud parameters. *International Journal of Remote Sensing*, 18(15), 3263–3277. <https://doi.org/10.1080/014311697217071>
- Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., et al. (2016). Multi-radar multi-sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, 97(4), 621–638. <https://doi.org/10.1175/bams-d-14-00174.1>