Assessment of the Advanced Very High Resolution Radiometer (AVHRR) for Snowfall Retrieval in High Latitudes Using *CloudSat* and Machine Learning

MOHAMMAD REZA EHSANI,^a ALI BEHRANGI,^{a,b} ABISHEK ADHIKARI,^a YANG SONG,^a GEORGE J. HUFFMAN,^c ROBERT F. ADLER,^d DAVID T. BOLVIN,^{c,e} AND ERIC J. NELKIN^{c,e}

^a *Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, Arizona*

^b *Department of Geosciences, The University of Arizona, Tucson, Arizona*

^c *Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland*

^d *Earth System Science Interdisciplinary Center, Cooperative Institute for Climate and Satellites, University of Maryland, College Park, College Park, Maryland*

^e *Science Systems and Applications, Inc., Lanham, Maryland*

(Manuscript received 14 October 2020, in final form 6 April 2021)

ABSTRACT: Precipitation retrieval is a challenging topic, especially in high latitudes (HL), and current precipitation products face ample challenges over these regions. This study investigates the potential of the Advanced Very High Resolution Radiometer (AVHRR) for snowfall retrieval in HL using *CloudSat* radar information and machine learning (ML). With all the known limitations, AVHRR observations should be considered for HL snowfall retrieval because 1) AVHRR data have been continuously collected for about four decades on multiple platforms with global coverage, and similar observations will likely continue in the future; 2) current passive microwave satellite precipitation products have several issues over snow and ice surfaces; and 3) good coincident observations between AVHRR and *CloudSat* are available for training ML algorithms. Using ML, snowfall rate was retrieved from AVHRR's brightness temperature and cloud probability, as well as auxiliary information provided by numerical reanalysis. The results indicate that the ML-based retrieval algorithm is capable of detection and estimation of snowfall with comparable or better statistical scores than those obtained from the Atmospheric Infrared Sounder (AIRS) and two passive microwave sensors contributing to the Global Precipitation Measurement (GPM) mission constellation. The outcomes also suggest that AVHRR-based snowfall retrievals are spatially and temporally reasonable and can be considered as a quantitatively useful input to the merged precipitation products that require frequent sampling or long-term records.

KEYWORDS: Precipitation; Snowfall; Freezing precipitation; Mixed precipitation; Machine learning

1. Introduction

Quantifying high-latitude precipitation is critical to understanding the current state of Earth's climate, and for water and energy cycle analysis. Precipitation retrieval in high latitudes is also important because snow in the form of the snowpack is a major source of freshwater for many countries including the United States. Furthermore, knowing the snowfall amount on sea ice is essential to derive sea ice thickness from spaceborne altimeters (Kwok and Markus 2018; Song et al. 2020). Therefore, accurate quantification of high-latitude snowfall at regional and global scales is important not only for the present weather observations but also for long-term climate analysis (Adhikari et al. 2018; Brucker and

Markus 2013; Liu and Curry 1997; Webster et al. 2014; Arabzadeh et al. 2020; Skofronick-Jackson et al. 2017).

The number of precipitation gauges in high latitudes has sharply decreased since 1990 and only limited gauges are operating in polar regions and over the ocean. Besides, the traditional gauge measurement techniques for snowfall measurement exhibit high uncertainties and errors; correction factors for wind-induced undercatch can lead to uncertainties as high as 100%, especially in sparsely gauged regions of high latitudes (Behrangi et al. 2019; Fuchs et al. 2001; Goodison et al. 1998; Kidd et al. 2017; Panahi and Behrangi 2019; Yang et al. 2005).

Precipitation retrieval from satellite data is an important topic and has been made for almost four decades. As requested by the user communities, precipitation estimates from individual sensors are often combined to produce gridded products with better temporal sampling. Generally, three types of satellite sensors are used: 1) radar, which allows direct observation of vertical distributions of hydrometeors to be linked to precipitation rate; 2) passive microwave (PMW), which enables retrieval methods to use absorption and scattering properties of hydrometeors to relate the observed radiances at the top of the atmosphere to surface rainfall; and 3) infrared (IR), which provides cloud-top temperature information that can be linked to precipitation rate, generally assuming that colder temperature is related to more intense precipitation rate.

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-20-0240.s1>.

Corresponding author: Ali Behrangi, behrangi@email.arizona.edu

DOI: 10.1175/JHM-D-20-0240.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Radars operating at Ku or Ka bands tend to provide an accurate estimation of precipitation in lower latitudes, as shown by precipitation radars on the Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM) mission core instruments. However, they may not provide sufficient sensitivity to light rain and snowfall that are frequent in high latitudes (Casella et al. 2017; Skofronick-Jackson et al. 2019).

Launched in 2006, *CloudSat* carries a 94-GHz Cloud Profiling Radar (CPR) that enables retrieval of snow and light rainfall with unprecedented signal sensitivity (i.e., the minimum detectable signal of ~ -28 dBZ; Stephens et al. 2002), making it a valuable sensor for high-latitude precipitation analysis. *CloudSat* CPR is the first W-band spaceborne radar, providing near-global retrievals of fine-scale precipitation occurrence and precipitation rate in both frozen and liquid phases. Previous efforts to evaluate the *CloudSat* precipitation products are encouraging, showing great capabilities of *CloudSat* CPR to retrieve falling snow from space (e.g., Ellis et al. 2009; Chen et al. 2016). As a result, *CloudSat* CPR has been widely used for regional and global snowfall analysis and can be considered as a baseline for snowfall estimation (Behrangi et al. 2014a,b, 2016; Kulie et al. 2016; Liu 2008; Palerme et al. 2014). However, *CloudSat* estimates may face two major issues: 1) saturation of signal under intense precipitation events that is reduced rapidly toward higher latitudes and for snowfall (Behrangi et al. 2012), and 2) missing precipitation near surface due to ground clutter (Smalley et al. 2017; Battaglia and Panegrossi 2020; Skofronick-Jackson et al. 2019; Palerme et al. 2019; Bennartz et al. 2019).

While the combination of radars enables precipitation estimation at all phases with unprecedented accuracy, to date, spaceborne radars have not provided sufficient data for global precipitation monitoring because they are 1) fairly new and have limited temporal coverage [i.e., there are only three common ones: TRMM (1997–2015), GPM (2014–present), and *CloudSat* (launched in 2006, providing only daytime observation after 2011, and currently deorbiting)], 2) limited in geographical coverage (i.e., $\sim 36^{\circ}\text{S}$ – 36°N for TRMM and $\sim 65^{\circ}\text{S}$ – 65°N for GPM), and 3) limited in swath coverage (i.e., *CloudSat* provides only nadir observation and radars on TRMM and GPM have narrow swaths).

PMW sensors provide less direct interaction with hydrometeors and generally provide less accurate precipitation rate than radars. However, they have some advantages over radars. For example, there are several PMW sensors in space and they provide longer observational records [e.g., the Special Sensor Microwave Imager (SSM/I) goes back to 1987] and wider swath and geographical coverage (e.g., they often provide pole-to-pole observations) than satellite radars.

However, with respect to high-latitude precipitation estimation, PMW-based retrievals face several challenges such as 1) poor sensitivity of sensors to light rain and snowfall that may lead to large missing or underestimation of precipitation (Behrangi et al. 2012), 2) unknown surface emissivity over snow and ice surfaces (Ferraro et al. 2013), 3) problems in distinguishing light rain and snowfall from cloud (Berg et al. 2006; Lebsack and L'Ecuyer 2011), and 4) the need for prior

knowledge about precipitation phase for retrieval (Liu 2008). Nonetheless, studies have shown that PMW data can offer great potentials for snowfall retrieval (e.g., Levizzani et al. 2011; Liu and Seo 2013; You et al. 2017; Panegrossi et al. 2017; Meng et al. 2017; Tang et al. 2018; Edel et al. 2019; Skofronick-Jackson et al. 2019; Adhikari et al. 2020). For example, Rysman et al. (2018, 2019) have developed a *CloudSat*/CALIPSO-based machine learning approach for snowfall detection and retrieval for the GPM Microwave Imager (GMI); the Snow Retrieval Algorithm for GMI (SLALOM). The authors have demonstrated that the use of *CloudSat* products for training machine learning models is very effective in exploiting GMI capabilities to reproduce snowfall climatology in good agreement with *CloudSat*.

IR-based sensors perhaps provide the least direct information about precipitation, but they are commonly used because they provide 1) high temporal sampling (e.g., from the geostationary platform), 2) high spatial resolution [e.g., from geostationary and low-Earth-orbiting (LEO) sensors], 3) relatively long-term record (e.g., going back to 1979 from geostationary satellites and AVHRR), and 4) global coverage and ample sampling due to their availability on geostationary or several LEO platforms.

Most of the current gridded precipitation products currently use geostationary IR observations either as their main retrieval input or as an alternative estimate to fill the gaps remaining due to insufficient temporal sampling or accuracy of PMW estimates. For example, Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) (Hsu et al. 1997; Sorooshian et al. 2000; Ashouri et al. 2015) retrieves precipitation rate entirely based on IR data with PMW used to link IR observation to precipitation rate. The Integrated Multisatellite Retrievals for GPM (IMERG) also uses IR-based precipitation rate to fill PMW gaps over snow and ice surfaces (where the accuracy of PMW retrieval is questionable) or when PMW observations are too sparse in time to be interpolated with good accuracy through the morphing process used in IMERG (Huffman et al. 2019).

So far, most of the gridded merged satellite precipitation products provide quasi-global (e.g., 60°S – 60°N) coverage due to the limited coverage (e.g., 60°S – 60°N) of high-quality IR data from geostationary satellites and low-quality precipitation retrievals from PMW sensors in the cold conditions typical of high latitudes. IMERG V06 provides global 90°S – 90°N product but has gaps over snow and ice surfaces poleward of 60° latitude where PMW are low quality and there is no geostationary IR data for precipitation retrieval. In contrast, the Global Precipitation Climatology Project (GPCP; Adler et al. 2003; Huffman et al. 2009; Adler et al. 2018; Huffman et al. 2020) uses IR-based precipitation retrievals from AIRS and the Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) sensors to fill in the high latitudes. Precipitation rates from AIRS and TOVS are calculated using an algorithm developed by Susskind et al. (1997). The use of TOVS/AIRS also provides relatively long-term records going back to 1979.

With growing users' requests for pole-to-pole coverage, the producers are devoting additional efforts to extend their

TABLE 1. Summary of datasets used in the present study. Variables used from each product are also provided.

Product	Variables	Spatial and temporal resolution	Coverage
<i>CloudSat</i> 2C-SNOW-PROFILE P1_R05	Snowfall rate	Orbital, 1.4 km $\times$ 1.7 km	81°S/N
<i>CloudSat</i> 2C-CLDCLASS-lidar P1_R05	Cloud types	Orbital, 1.4 km $\times$ 1.7 km	81°S/N
<i>NOAA-18</i> AVHRR	11- and 12- μ m TBs and cloud probability	Orbital, 1.1 km $\times$ 1.1 km at nadir	90°S/N
GPM_2AGPROFNOAA18MHS CLIM 5	Precipitation rate	Orbital, 17 km $\times$ 17 km at nadir	90°S/N
GPM_2AGPROFF16SSMIS CLIM 5	Precipitation rate	Orbital, 12 km $\times$ 12 km at nadir	90°S/N
Stage IV V01	Precipitation rate	Gridded, hourly, 4 km $\times$ 4 km	CONUS
AIRS V006	Precipitation rate	Orbital, 15 km $\times$ 15 km at nadir	90°S/N
MERRA-2	TWET, TPW, near-surface snowfall rate	Gridded, hourly, 0.5° $\times$ 0.625°	90°S/N
ERA5	Snowfall rate	Gridded, hourly, 30 km $\times$ 30 km	90°S/N
GPCP 1DD V1.3	Precipitation rate	Gridded, daily, 1° $\times$ 1°	90°S/N
GPCC full data V2018	Precipitation rate, correction factors	Gridded, daily, 1° $\times$ 1°	90°S/N
NOAA AutoSnow V001	Surface type	Gridded, daily, 4 km $\times$ 4 km	90°S/N

coverage (with no gaps) to the poles; at least this is the case for IMERG and the second-generation Climate Prediction Center (CPC) morphing technique (CMORPH-2) (Joyce and Xie 2011; Xie et al. 2018). These products also target subdaily (e.g., hourly, or higher) temporal resolution, requiring alternatives for IR-based precipitation estimation from sensors besides AIRS. While the Cross-track Infrared Sounder (CrIS) aboard the U.S. *Suomi National Polar-Orbiting Partnership* (*Suomi NPP*) can provide comparable observations to AIRS, another alternative is the use of AVHRR, as there are several of them on NOAA platforms, going back to 1981. Furthermore, AVHRR-like observations seem likely to continue [e.g., the Visible Infrared Imaging Radiometer Suite (VIIRS) on *NPP*].

In light of quality snowfall retrieval from *CloudSat* over land and ocean and availability of coincident observations between *CloudSat* and AVHRR (e.g., on the *NOAA-18* platform), the present study investigates the potential of using machine learning (ML) to establish a relationship between *CloudSat* snowfall rate and AVHRR observations and products (i.e., brightness temperature and cloud probability) as well as auxiliary information (surface wet-bulb temperature, total precipitable water, and surface type). *CloudSat* potential for snowfall retrieval using ML is investigated for PMW [e.g., Microwave Humidity Sounder (MHS)] in several studies. The results indicate that the *CloudSat*-based PMW retrieval algorithms using ML can detect and estimate both intense and weak snowfall events with high accuracy (Adhikari et al. 2020; Edel et al. 2019; Rysman et al. 2018, 2019).

However, to the best of our knowledge, there is no study focused on snowfall retrieval at high latitude using IR. Upon successful retrieval, the AVHRR can provide snowfall estimates with high temporal sampling and long-term records required by the merged precipitation products, especially those used for climate studies (e.g., GPCP). This study addresses the following scientific questions:

- Can ML help to establish a robust relationship between *CloudSat* snowfall rate and brightness temperature and cloud properties from AVHRR in high latitudes?
- Does the addition of auxiliary variables improve the detection and estimation of the snowfall?

- How well does the ML-based method developed in this study perform in comparison with snowfall estimates from *CloudSat*, ground-based, current PMW, reanalysis, and AIRS products?

2. Datasets

Table 1 summarizes datasets used in this study. A brief description of each dataset is given as follows:

a. AVHRR

The Advanced Very High Resolution Radiometer instrument is a spaceborne sensor that measures the reflectance of Earth in five spectral bands (0.6, 0.9, 3.5, 11, and 12 μ m). AVHRRs have been carried on the NOAA family of polar-orbiting platforms (POES) and European MetOp satellites. AVHRRs offer fairly continuous global coverage since June 1979, with morning and afternoon acquisitions available (Frey et al. 2012). In the present work, AVHRR data are obtained from the AVHRR Pathfinder Atmospheres–Extended (PATMOS-x) product, providing a high-quality Climate Data Record (CDR) of multiple cloud properties along with AVHRR brightness temperatures (Heidinger et al. 2014). Here, the AVHRR sensor on *NOAA-18* is used.

b. *CloudSat*

The CPR is a 94-GHz nadir-looking radar aboard *CloudSat* that measures the power backscattered by clouds (and other objects) as a function of distance from the radar with the cross-track, along-track, and vertical resolution of 1.4 km, 1.7 km, and 500 m, respectively. Two *CloudSat* products were used in this study: (i) 2B-CLDCLASS-lidar that combines *CloudSat* CPR and *CALIPSO* lidar measurements to classify clouds into stratus (St, Sc), cumulus (Cu, including cumulus congestus), nimbostratus (Ns), altocumulus (Ac), altostratus (As), deep convective (cumulonimbus), or high (cirrus and cirrostratus) clouds (Mace and Zhang 2014), and (ii) 2C-SNOW-PROFILE (Wood et al. 2013) providing near-surface and profile of snowfall rate. Good performance has been reported for the 2C-SNOW-PROFILE product over land (Smalley et al. 2014; Chen et al. 2016). However, *CloudSat* may be affected by underestimation

of intense surface snowfall rates mostly due to attenuation issues, and ground clutter (Battaglia and Panegrossi 2020; Skofronick-Jackson et al. 2019; Bennartz et al. 2019; Palerme et al. 2019).

c. MERRA-2

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) provides data beginning in 1980. MERRA-2 was introduced to replace the original MERRA dataset because of the advances made in the assimilation system that enable assimilation of modern hyperspectral radiance and microwave observations, along with GPS radio occultation datasets. It also uses NASA's ozone profile observations that began in late 2004. Additional advances in both the GEOS model and the GSI assimilation system are included in MERRA-2 (Gelaro et al. 2017). This study uses surface wet-bulb temperatures (TWET), total precipitable water (TPW), and surface snowfall rate from MERRA-2.

d. ERA5

ERA5 provides hourly estimates of many atmospheric, land, and oceanic climate variables. The data cover Earth on a 30-km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80 km. ERA5 includes information about uncertainties for all variables at reduced spatial and temporal resolutions. ERA5 combines vast amounts of historical observations into global estimates using advanced modeling and data assimilation systems (Hersbach et al. 2020). This study uses convective and large-scale snowfall rates to compare ERA5 total snowfall with other products.

e. GPCP

GPCP 1DD V1.3 product provides daily precipitation estimates on a 1° grid over the entire globe for the period from October 1996 to the present. The 1DD product is consistent with the version 2.3 monthly product in the sense that the 1DD approximately sums to the monthly satellite–gauge precipitation estimates. The 1DD precipitation product is produced by optimally merging estimates computed from microwave, infrared, and precipitation gauge analyses (Adler et al. 2018).

f. GPCC

The Global Precipitation Climatology Center (GPCC) was established in 1989 at the request of the WMO. It is operated by the German Weather Service [Deutscher Wetterdienst (DWD)] and contributes to the World Climate Research Program (WCRP). GPCC provides gridded maps of monthly and daily precipitation data on Earth's land surface (except over Antarctica) based on in situ rain gauge data for monitoring and research of Earth's climate. Depending on their applications, various GPCC products are available. Here the full data daily (V2018) is used, currently available at 1° × 1° resolution from 1982 through 2016, incorporating all stations with the highest quality-control process, but it is not available in real time and does not correct for gauge undercatch problems (Schneider et al. 2017; Ziese et al. 2018).

g. MHS and SSMIS precipitation rates based on GPROF

Precipitation estimation from two representative GPM constellation sensors, based on the latest version of the GPM

Profiling (GPROF2017) algorithm for GPM V05, were used for comparative analysis. The two sensors are 1) MHS, a five-channel passive microwave radiometer (from 89 to 190 GHz) that flies with AVHRR on NOAA-18 and coincides well with CloudSat observations during the period of study (2007–10), and 2) the SSMIS aboard the Defense Meteorological Satellite Program (DMSP) F16 satellite. SSMIS provides brightness temperature observations at frequencies ranging from 19 to 183 GHz over a swath width of 1707 km (Sun and Weng 2008; Yan and Weng 2008). GMI was not used in this study as its coverage is limited to 65°S–65°N and does not cover the entire high-latitude region studied here.

h. Stage IV

For the evaluation purpose, besides independent CloudSat snowfall rates, hourly accumulated precipitation data from the National Centers for Environmental Prediction (NCEP) Stage IV was used over the United States. NCEP Stage IV data assimilate both radar and gauge observations, which are further adjusted based on 12 River Forecast Centers. The spatial resolution of Stage IV data is 4 km and is available in polar stereographic grid (Lin 2011). With quality control, the product agrees closely with gauge measurement (Zhang et al. 2018), so this study uses Stage IV product for validation of the case study.

i. AIRS

The Atmospheric Infrared Sounder aboard the Earth Observing System (EOS) Aqua satellite contains 2378 infrared and four visible/near-infrared channels. AIRS's main objective is to provide highly accurate temperature profiles within the atmosphere plus additional atmospheric products in combination with the Advanced Microwave Sounding Unit (AMSU) and the Humidity Sounder for Brazil (HSB). Precipitation from AIRS is obtained using an algorithm described in Susskind et al. (1997). The algorithm estimates a cloud volume proxy computed from sounding estimates of cloud-top pressure, fractional cloud cover, and relative humidity profiles. The cloud volume is converted to precipitation using regression against the First GARP Global Experiment (FGGE) rain stations. Since the loss of AMSU-A2 in 2016, a modified version of the algorithm was developed that uses only IR data. This version that is now produced operationally is used in the present study.

j. NOAA AutoSnow

Another important auxiliary dataset used in this study is the surface snow/ice index (AutoSnow), which is derived from the global multisensor automated snow/ice cover maps (Romanov 2017). Surface indices are represented by binary values, where 0 stands for open water, 1 for land, 2 for snow over land, and 4 for ice over water. This is a daily global gridded product with a spatial resolution of 4 km. The surface indices are generated from combined snow/ice retrievals with various satellite sensors operating in visible, infrared, and microwave spectral bands.

3. Methodology

In the first step, each CloudSat pixel was matched up with corresponding AVHRR, MERRA-2, and AutoSnow pixels for

the period 2007–10. Then pixels meeting certain spatial and temporal distance conditions (i.e., 5 km and 5 min, respectively) that have good quality were considered for allocation to the training and testing data pools; 80% of the matched pixels are used for training and validation, and 20% for testing the ML algorithm. Note that samples from different orbits were used for training and testing to guarantee the independence of the training and test sets. Stratified sampling was used to ensure that the distributions of variables are the same in all three datasets and to enhance the statistical robustness of the training, testing, and validation sets. In stratification, members of the population are divided into homogeneous subgroups before sampling, which guarantees that the ratio of snow/no-snow instances in all sets (training, validation, and test) is the same.

Among several ML algorithms available, random forest (RF) was selected, based on the initial evaluations and literature review (e.g., Adhikari et al. 2020; Edel et al. 2019; Ehsani et al. 2020). Random forest is an ensemble of decision trees (DTs), generally trained via the bagging method (or sometimes pasting). DT builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated DT is incrementally developed. At each step, the DT selects the best variable and threshold that allow the dataset to be split between two classes (i.e., snow and no-snow for detection). This process is reproduced for each subsequent subset until either all cases in a node fall into the same class or the maximum tree development is reached. The result is a tree with decision nodes and leaf nodes.

The random forest algorithm introduces extra randomness when growing trees; instead of searching for the very best feature when splitting a node, it searches for the best feature among a random subset of features. The algorithm results in greater tree diversity, which trades a higher bias for a lower variance, generally yielding an overall better model. For a classification problem, the RF returns a statistical probability of snowfall occurrence based on casted votes by all developed trees (Ehsani et al. 2020; Adhikari et al. 2020; Gislason et al. 2004; Ham et al. 2005).

The RF can also be used in a regression problem, where it predicts snowfall rate of a given footprint (or pixel) based on the predictions by all decision trees. RF relies on several DTs and is based on a random process; therefore, RF is less prone to overfitting (Hastie et al. 2009). RF runs efficiently on large datasets, is computationally lighter than other tree assembly methods, and can capture nonlinear relationships between input features and labels, so it has been widely used in satellite remote sensing research (Kühnlein et al. 2014).

The proposed ML-based retrieval algorithm is based on two steps. In the first step and based on the features, the classification/detection RF model determines whether each pixel is a snow pixel or not. Then, the RF regression model estimates the snowfall rate for the snow pixels. A k -fold cross validation (here $k = 10$) was performed to check the robustness of the RF models with respect to the training samples. Three ML models were developed in this study. The first model (i.e., AVHRR 0), uses AVHRR's brightness temperature at 11 and 12 μm (TB 11, TB 12), AVHRR's cloud probability (CP), and

AutoSnow's surface type (hereafter ST) as inputs. Only IR bands were used, so the AVHRR-based precipitation retrievals can be used during both day and night. The combination of the TB 11 and TB 12 are also known to be effective in detecting thin cirrus clouds that often produce no precipitation (Inoue et al. 1987a,b; Kurino 1997). The second model, hereafter referred to as the AVHRR I, uses TB 11, TB 12, CP, MERRA-2's TWET, and ST as inputs. The third model, hereafter referred to as AVHRR II, uses MERRA-2's TPW in addition to the second model's features. TWET is the only reanalysis variable chosen for use in both AVHRR I and AVHRR II due to the significance of the wet-bulb temperature for snowfall detection/estimation. TWET, which combines the effects of temperature and moisture, is a key parameter for separating solid and liquid precipitation. TWET also yields the highest skill score for determining precipitation phase and can reduce uncertainties due to regional differences, especially compared to the commonly used near-surface air temperature (Sims and Liu 2015; Behrangi et al. 2019). TPW was also added to AVHRR II since it is shown to be important variable for precipitation retrieval (e.g., Gutman et al. 2004; Hou et al. 2000). Comparison of AVHRR I and AVHRR II models enables assessment of how TPW affects the retrieval results. The value of adding near-surface temperature and humidity information to infrared observations for precipitation detection has also been shown in Behrangi et al. (2015). GPROF also uses TPW and near-surface air temperature from reanalysis to determine precipitation regime.

AVHRR-ML products were compared with AIRS and MHS precipitation products at the footprint level using statistical metrics for snowfall detection and estimation. *CloudSat* was matched up separately with AIRS and MHS for the period 2007–10 by considering differences in their pixel sizes (Behrangi et al. 2012), and TWET was used to separate rain from snow according to Sims and Liu (2015). This method was also used to separate rain and snow for other precipitation products used in this study. AVHRR-ML products and AIRS were also compared with other precipitation products listed in Table 1 through various plots including time series, geographical maps, and Taylor diagrams. Results are presented for both Northern and Southern Hemispheres.

4. Results

Because AVHRR CP is an option to be included as a feature (i.e., input) of the ML-based retrieval method, it was first compared with the 2B-CLDCLASS-lidar for an overall assessment and consistency check. Figure 1 shows the distribution of AVHRR CP (i.e., shown with boxplots) for 2B-CLDCLASS-lidar clear sky and each cloud type, separately over land and ocean and for the Northern Hemisphere (NH) and Southern Hemisphere (SH) in high latitudes. This was obtained by identifying coincident observations between AVHRR and 2B-CLDCLASS-lidar with the temporal distance no more than 5 min, and spatial distance less than 5 km. The boxplots show the median values of AVHRR CP as well as the first and third quartiles (solid horizontal lines). The results indicate

FIG. 1. Distribution of AVHRR CP for 2B-CLDCLASS-lidar clear sky and each cloud type, separately over land and ocean in NH and SH high latitudes (poleward of 60° latitude) for the years 2007–10. Solid horizontal lines show the median values of AVHRR CP, as well as the first and third quartiles. The number of samples is displayed for each category.

that AVHRR CP provides valuable insights about the presence of the clouds as it has much higher cloud probability for all 2B-CLDCLASS-lidar cloud types than clear sky. The lower CP of AVHRR for high clouds (i.e., Cirrus) compared to 2B-CLDCLASS-lidar is also helpful for precipitation retrieval as no precipitation is expected from Cirrus clouds. However, the AVHRR CP corresponding to 2B-CLDCLASS-lidar clear sky is slightly high, especially in SH, suggesting that identifying clear cases with AVHRR CP may not be accurate, which can consequently lead to higher false alarm ratios. If CP of 50% be selected as a threshold for determining cloudy scenes, AVHRR CP seems to perform well in NH. Nonetheless, some of the differences are due to differences in observing systems, retrieval algorithms (Wang et al. 2016), and temporal and spatial distance between *CloudSat* and AVHRR pixels. Because in the present work ML is used for retrieval, it was decided to use AVHRR CP as an input with no need for identifying any specific threshold for CP values. The importance of CP for snowfall retrieval will be discussed later.

Statistical metrics at footprint level for AIRS, AVHRR 0, AVHRR I, AVHRR II, and GPM-MHS with reference to *CloudSat* surface snowfall rate are summarized in Table 2. A brief description of these metrics is also provided in the appendix. The results indicate that among AVHRR products,

AVHRR 0 (i.e., the model using only AVHRR features) has weaker performance compared to AVHRR I and AVHRR II. AVHRR I (the model without TPW) has a slightly better performance than other products in terms of probability of detection (POD). AVHRR II (the model with TPW) has comparable results with AIRS in terms of false alarm ratio (FAR), Heidke skill score (HSS), and root-mean-squared error (RMSE) while having lower bias, and higher correlation coefficient. GPM-MHS has the weakest scores in comparison with other products. Considering the weaker performance of AVHRR 0 relative to other AVHRR variations, in what follows only performance of AVHRR I and AVHRR II are compared to other products. In Edel et al. (2019), for a RF detection scheme for MHS, the authors find POD of 0.62, FAR of 0.34, and HSS of 0.51. In a similar study, Rysman et al. (2018), for a RF detection scheme for GMI (lower latitudes) reported POD, FAR, and HSS equal to 0.83, 0.12, and 0.84, respectively. In Rysman et al. (2019) a *CloudSat*-based machine learning scheme for surface snowfall rate retrieval is evaluated and the scores show relative bias of −13%, RMSE of 0.08 mm h^{-1} , and a correlation coefficient of 0.7.

In addition to the footprint level statistics, it is important to assess products' snowfall pattern and geographical distribution. Figure 2 shows seasonal maps of snowfall rate constructed

TABLE 2. Summary of statistics at footprint level for AIRS, AVHRR 0, AVHRR I, AVHRR II, and GPM-MHS with reference to *CloudSat*. Statistics for AVHRR I and AVHRR II are presented for the test set that is independent of the data used for training. GPM-MHS is based on GPROF, version 5, and AIRS precipitation is obtained from AIRS V6.

Metrics	AIRS	AVHRR 0	AVHRR I	AVHRR II
F1 score	0.516	0.420	0.490	0.470
Areal bias	1.294	1.734	1.547	1.282
Probability of detection	0.592	0.578	0.667	0.583
False alarm ratio	0.543	0.670	0.590	0.580
HSS score	0.408	0.268	0.382	0.384
Correlation coefficient	0.333	0.250	0.301	0.353
Bias	1.416	1.680	1.451	1.206
RMSE	0.096	0.125	0.114	0.102

from AVHRR I, AVHRR II, AIRS, and GPM-MHS over NH poleward of 60° latitude compared to the corresponding ones from *CloudSat*, ERA5, MERRA-2, GPM-SSMIS, GPCP, and GPCC (over land only). The horizontal line at 81°N/S

represents the highest latitude that *CloudSat* can observe. The results over land indicate that AVHRR products can produce the general snowfall pattern, but both tend to overestimate compared to *CloudSat*, especially AVHRR I in DJF [e.g., for AVHRR I and for JJA, RMSE is 0.09, and correlation coefficient (CC) is 0.78, and for AVHRR II RMSE is 0.05 and CC is 0.77]. Compared to GPCP and GPCC, both AVHRR products have relatively good performance, especially AVHRR II (e.g., CC with respect to GPCC for MAM is 0.706 and 0.681 for AVHRR I and AVHRR II, respectively). GPM-MHS underestimates snowfall compared to all products except for GPM-SSMIS, probably because both GPM products use the GPROF algorithm. AIRS underestimates compared to AVHRR products and is less similar to GPCC and GPCP, as GPCP is largely adjusted by GPCC over land (e.g., CC with respect to GPCC for MAM is 0.602).

Over the ocean in the NH, AVHRR I tends to show larger values compared to all other products especially above 81°N, where no *CloudSat* reference is available for training. AVHRR II has lower seasonal snowfall rates than AVHRR I, which might be due to the additional information made available by TPW for snowfall retrieval near the pole. AIRS also overestimates considerably during JJA, which might be related to

FIG. 2. Comparison of surface snowfall rate from AVHRR I, AVHRR II, AIRS, and MHS products with that of *CloudSat*, ERA5, MERRA-2, SSMIS, GPCP, and GPCC products over NH above 60°N for the years 2007–10. The black horizontal lines show the highest latitude where *CloudSat* observations are available (81°N).

FIG. 3. As in Fig. 2, but over the SH.

frequent false alarms due to sea ice extent. AVHRR II agrees well with *CloudSat* and reanalysis products in terms of pattern and magnitude, especially at latitudes lower than 81°N (e.g., CC is 0.679 with *CloudSat* in JJA, 0.751 with MERRA-2 in MAM, and 0.773 with ERA5 in DJF). GPM-MHS and GPM-SSMIS both tend to largely underestimate compared to other products but can capture the overall snowfall pattern.

Similar maps are also produced in SH high latitudes, over land (i.e., mainly Antarctica) and ocean and are shown in Fig. 3. Over Antarctica, AVHRR I tends to overestimate compared to other products in DJF, but in other seasons it produces comparable estimates to reanalysis (e.g., CC with MERRA-2 is 0.750, 0.770, 0.816, and 0.790 for different seasons). AVHRR II agrees better with other products, except GPM-MHS and GPM-SSMIS as they underestimate, similar to what was observed over land–ocean in NH. AIRS has a good agreement with GPCP (e.g., CC is 0.861, 0.939, 0.909, and 0.897 for different seasons), probably because AIRS is the main satellite input to GPCP in high latitudes (Adler et al. 2003, 2018), especially over the Antarctic where in situ observations are very limited. GPCP shows different snowfall patterns near the pole in JJA and SON when compared to reanalysis and *CloudSat* (e.g., in JJA CC is 0.581, 0.621, and 0.643 with *CloudSat*, ERA5, and MERRA-2, respectively).

Over the ocean in SH, AVHRR I, AVHRR II, AIRS, and GPCP overestimate compared to the reanalysis and *CloudSat*;

however, GPROF-based products underestimate. While in MAM, JJA, and SON, AVHRR products are in better agreement with *CloudSat* and reanalysis than GPCP and AIRS (except for the Weddell Sea), AVHRR products show different behavior (both pattern and magnitude) in DJF, where they seem to agree more with GPCP and AIRS.

Scrutinizing average monthly snowfall rate over high-latitude land in NH (Fig. 4) indicates that there is a relatively good agreement among all products in terms of the monthly mean pattern. GPM-MHS and GPM-SSMIS capture the overall pattern but significantly underestimate the mean snowfall rate. Both AVHRR I and AVHRR II show the best agreement with GPCC among all the products (CC is 0.968 and 0.986 for AVHRR I and AVHRR II, respectively). Compared to AVHRR I, AVHRR II has lower values in most months and generally is in a better agreement with GPCC. The largest difference between AVHRR II and GPCC occurs in March when AVHRR I almost matches GPCC in terms of mean monthly precipitation rate. AVHRR I is especially larger than both AVHRR II and GPCC in April and May, but it is almost identical to the two products in January and February. This suggests that the gauge undercatch issue in GPCC (see Behrangi et al. 2018) may not justify the lower value of GPCC than AVHRR I in April and May, as the gauge undercatch issue is often largest in winter. AIRS underestimates compared to other products (except GPM products) in March,

FIG. 4. Annual cycle of average monthly snowfall rate from different products for the latitude band 60°–81°N for the years 2007–10.

April, and May but shows a very good agreement with ERA5 in summer and fall months (CC is 0.965).

Over the ocean in NH, AVHRR I captures the overall pattern (CC is 0.961, 0.943, and 0.942 with *CloudSat*, ERA5, and MERRA-2, respectively) but overestimates compared to all other products in all months except for JJA where AIRS overestimates significantly. AVHRR II has a very good agreement with reanalysis and *CloudSat* products in most months (i.e., CCs greater than 0.960 for different seasons). The improvement in AVHRR II compared to AVHRR I can be attributed to the information made available by TPW. AIRS tends to underestimate in February, March, and April, and GPCP's performance is probably affected by AIRS since it uses AIRS as the main input. While over land GPM products tend to capture the overall monthly mean snowfall pattern, over the ocean they show different patterns from all other products. Besides, they continue to underestimate compared to other products.

Over Antarctica, AVHRR I overestimates from October to February while AVHRR II produces mean snowfall rates closer to *CloudSat*. The addition of TPW seems effective in reducing large biases of AVHRR I from October to February. GPCP overestimates in June, July, August, and September. AIRS and GPCP are highly correlated (i.e., CC is 0.987), perhaps because, over high-latitude ocean, AIRS is the main source of input to the GPCP. In general, there is a good agreement between *CloudSat*, and reanalysis compared to other products. GPM-SSMIS and GPM-MHS show patterns similar to AVHRR I but produce lower snowfall rates compared to all other

products (e.g., CC between MHS and AVHRR I is 0.969). The large underestimation of the GPM products is consistent with the mass budget analysis of snowfall accumulation over Antarctica using the Gravity Recovery and Climate Experiment (GRACE) and ice discharge observations (Behrangi et al. 2020). GPM-SSMIS seems to underestimate less than GPM-MHS, perhaps due to the extra channels it carries. Note that GPCC has no coverage over Antarctica, so GPCC was not included.

Over the ocean in SH, 1) AVHRR I overestimates all other products across all months except February and March; 2) GPCP and AIRS are highly correlated and show different monthly mean snowfall patterns compared to *CloudSat* and reanalysis products; 3) *CloudSat* and reanalysis products show comparable monthly distributions, but *CloudSat* produces slightly lower mean snowfall rates than reanalysis products (i.e., CC is 0.985 and 0.983 with ERA5 and MERRA-2, respectively); 4) AVHRR II has closer values to reanalysis and *CloudSat* than AVHRR I but both AVHRR products estimates seem to miss the monthly distribution patterns suggested by *CloudSat* and reanalysis products; and 5) GPROF-based products underestimate all other products but show patterns similar to GPCP and AIRS (e.g., CC for MHS is 0.903 and 0.944 with AIRS and GPCP, respectively). Overall, and by comparing these figures, one may conclude that AVHRR estimates have better agreement with *CloudSat* and reanalysis products over NH than SH.

As AVHRR products show overestimation in most months compared to other products, one may use seasonal *CloudSat*

FIG. 5. Comparison of surface snowfall rate from AVHRR products after zonal adjustment with *CloudSat* climatology (AVHRR I C and AVHRR II C) with AVHRR products before adjustment and *CloudSat* over NH land above 60°N for the years 2007–10. The white horizontal lines show the highest latitude where *CloudSat* observations are available (81°N).

climatology to bias adjust AVHRR estimates, so the patterns can be better compared with other products. Figure 5 shows the AVHRR products after zonal adjustment with *CloudSat* climatology over land in NH. Similar climatology bias adjustment has been employed in the recent version of GPCP (V3.1) (Huffman et al. 2020) in which climatology of precipitation estimates from advanced sensors are used to bias adjust other products with longer length of data record. The outcome of AVHRR bias adjustment shows that AVHRR captures patterns that are generally comparable to *CloudSat*. Similar plots over ocean in NH and land and ocean in SH are presented in supplemental (Figs. S1–S3 in the online supplemental material).

Also, relatively large biases can be observed between AVHRR products themselves. Since the only difference between the models is TPW, it is insightful to look at how TPW impacts the snowfall prediction. One way to see what the impact of TPW on snowfall detection is, and to assess the AVHRR snowfall detection capabilities of the different RF modules in the different regions, is to compare the global maps of snowfall occurrence (normalized by the total number of occurrences in each grid point) obtained for AVHRR I and AVHRR II. This would also be useful to understand if the features described in the maps in geographic maps and zonal plots are related to the detection or to the estimation component of each module. Based on this plot (Fig. 6), it can easily be seen that adding TPW changes the detection phase considerably. In NH and over land, detection increases in DJF and SON by adding TPW while over the ocean detection decreases in DJF and increases in SON (increase in SON is over 81 where no *CloudSat* observation is available). In SH, addition of TPW generally decreases the detection especially in DJF, MAM, and SON. These observations imply that TPW at least has an important role in detection phase.

Exploring Taylor diagrams over land and ocean in the NH for snowfall (Fig. 7) suggests that AIRS often has the highest correlation with GPCP (i.e., 0.665, 0.690, 0.895, and 0.507 over

land and 0.917, 0.871, 0.867, and 0.758 over ocean), which is not surprising because GPCP uses AIRS in high latitudes. AVHRR estimates are close to each other as well as AIRS (except in MAM over the ocean), suggesting that AVHRR can produce estimates with patterns comparable to AIRS. This means that there is a potential for using both AIRS and AVHRR estimates to improve temporal sampling for combined products by filling current gaps in PMW estimates due to snow and ice surfaces. A single sensor (e.g., AIRS) does not provide the high temporal sampling required by IMERG.

In terms of standard deviation, AVHRR I, AVHRR II, ERA5, and MERRA-2 have the closest values to GPCP over the ocean in NH, but *CloudSat* shows a relatively lower agreement with GPCP than other products, especially over the ocean (e.g., CC is 0.599, 0.529, 0.654, and 0.489 for different seasons, respectively). As AVHRR products are trained by *CloudSat*, the difference between *CloudSat* and AVHRR suggests that the type of observation (i.e., IR vs radar) and sampling differences (i.e., *CloudSat* has much less sampling than AVHRR) can result in a fairly large deviation in the snowfall estimates between AVHRR and *CloudSat*. This can also be related to the ability of the retrieval method (i.e., RF) to learn from such datasets.

Over Antarctica, AIRS, AVHRR I, and AVHRR II have the highest correlation with GPCP (i.e., 0.939, 0.897, and 0.917 in DJF; 0.939, 0.904, and 0.940 in MAM; 0.909, 0.559, and 0.614 in JJA; and 0.897, 0.777, and 0.786 in SON, respectively), while *CloudSat*, GPM-MHS, and GPM-SSMIS are relatively weakly correlated with GPCP and show larger standard deviation (e.g., CC for *CloudSat* is 0.705, 0.773, 0.581, and 0.708). The large standard deviation of *CloudSat* is probably due to *CloudSat* sampling as can be inferred from the noisier maps shown in Figs. 2 and 3, and its high sensitivity to detect light snowfall. GPROF-based products have very similar performance in most Taylor plots and often show relatively good correlation with GPCP, suggesting that their pattern is not too different from GPCP, but they have large biases as shown

FIG. 6. Comparison of the maps of snowfall occurrence (% , normalized by the total number of occurrences in each grid point) obtained for AVHRR I and AVHRR II in NH and SH. The differences is due to addition of TPW as a feature to the AVHRR II.

earlier. Similar to the NH, AIRS, AVHRR I, and AVHRR II are often close to each other and GPCP over Antarctica, but they show larger differences over the Southern Ocean. Over the Southern Ocean, the products have a larger spread among themselves (and GPCP) than that observed over Antarctica and NH. The observed large differences among the products over the Southern Ocean was also noticed in Behrangi et al. (2014a,b) and Behrangi et al. (2016).

The performance of AVHRR I and AVHRR II was also investigated for a case study over the United States. This event was observed by AVHRR at 1940 UTC 20 December 2010. Inputs used by the ML algorithm to retrieve snowfall for this event are presented in Fig. S4 in supplemental. These inputs are used to detect snowfall occurrence and estimate snowfall rate for this event using AVHRR I (Fig. 8a) and AVHRR II (Fig. 8b) products. The results are then compared with GPM-SSMIS (Fig. 8c), GPM-MHS (Fig. 8d), MERRA-2 (Fig. 8e), and Stage IV (Fig. 8f) snowfall rates.

It can be seen that AVHRR products can detect the event but AVHRR I shows smaller snowfall rates than other products. This observation along with better performance of AVHRR II on footprint level, its higher capacity to detect the overall pattern in geographic maps, zonal plots, and Taylor diagrams, suggests that TPW plays an important role for snowfall detection and estimation. Similar to MERRA-2,

AVHRR II shows a wider snowfall event than other products, but AVHRR II seems to produce a smooth field with generally lower extreme rates compared to the other products. In contrast, GPM-SSMIS and GPM-MHS suggest that the snowfall event is narrower than that observed by Stage IV and other products.

Nonetheless, it is important to recognize that Stage IV may also contain errors in detecting and retrieving snowfall rates (e.g., Prat and Nelson 2015; Wen et al. 2016). It is also noteworthy to mention that the selected case study is particularly intense, while *CloudSat* is known to be affected by underestimation of intense surface snowfall rates mostly due to attenuation issues (see Battaglia and Panegrossi 2020; Skofronick-Jackson et al. 2019) and ground clutter (especially over land, see Bennartz et al. 2019). Such underestimation may also affect the AVHRR-based results shown here because AVHRR products use *CloudSat* as a reference for training phase.

A valuable feature of the RF algorithm is that it provides insights into the importance of inputs used in detection and estimation tasks. Figure 9 demonstrates the importance of RF inputs for both AVHRR models, separately for detection and estimation. In the absence of TPW, TWET is the most important feature for both detection and estimation followed by AVHRR TB 11 and AVHRR CP. The large value of TWET is mainly due to its important role in separating precipitation

FIG. 7. Taylor diagram for the latitude band 60° – 81° N for the years 2007–10, averaged by season. The reference is GPCP. The radial axis shows the standard deviation, and the angular coordinate shows the correlation to the reference.

phase, as it is also used in *CloudSat*. TPW is also important as it can determine precipitation regime. TPW has also been used in the GPROF algorithm. High importance of TPW for detection is consistent with what was implied by the detection plots (Fig. 6). TB 11 and TB 12 are highly correlated, reducing the importance of the individual TB feature. Training a separate model using only TB 11 showed that TB 11 can exceed other individual features in terms of importance for both detection and regression. Using the features importance analysis, the value of adding surface type (water, land, snow, and ice) as an input feature to RF was also investigated, but it appears to have the lowest importance among all other features.

In the interpretation of the features importance, it is important to note the following: 1) Feature importance measures are highly dependent upon the accuracy of the model and the results are model dependent. RF and other classical machine learning algorithms have limitations, and their results should be interpreted sagaciously. In other words, since the response surface is highly rough and bumpy, RF may find a local optimum in the response surface; more

robust algorithm like deep neural networks (DNNs) may find better solutions. 2) Brightness temperatures and cloud probability fields of AVHRR used for training the machine learning algorithm are matched up with corresponding *CloudSat* snowfall rate meeting certain spatial and temporal conditions. The matchup process introduces noise to the features and decreases the capacity of the machine learning algorithms to fully exploit the information within these features.

5. Concluding remarks

Despite its importance, accurate precipitation estimation has remained challenging in high latitudes and over frozen surfaces. This is partly due to the large uncertainties in retrieving precipitation from PMW sensors and the lack of sufficient coverage from spaceborne radars with high sensitivity to light rain and snowfall (e.g., *CloudSat*). As a result, merged precipitation products that aim to produce pole-to-pole estimates have used or considered precipitation estimates from

FIG. 8. A case study representing a snow event over the U.S. northern plains observed by AVHRR at 1940 UTC 20 Dec 2010. Snowfall rates estimates are shown for (a) AVHRR I, (b) AVHRR II, (c) SSMIS-F17, (d) MHS-N18, (e) MERRA-2, and (f) Stage IV. Local time differences with the comparison reference (i.e., Stage IV) are shown at the top-right corner of each panel.

infrared sensors as an alternative. For example, in high latitudes, GPCP uses AIRS precipitation estimates, CMORPH-2 is planning to use AVHRR in its future generation, and IMERG is already investigating options to fill in the remaining gaps in the PMW estimates over frozen surfaces in high latitudes.

This study investigates if AVHRR data trained by *CloudSat* snowfall estimates can provide a reasonable estimate for snowfall retrieval in high latitudes (here poleward of 60°S/N) and how the outcomes are compared with snowfall retrievals from the current PMW sensors, AIRS, GPCP, GPCC, and

FIG. 9. Input dataset (i.e., feature) importance acquired from the RF used for snowfall detection and estimation for AVHRR I and AVHRR II products. TB 11 and TB 12 are brightness temperatures at 11 and 12 GHz, CP is AVHRR's cloud probability, TWET and TPW are wet-bulb temperature and total precipitable water from MERRA-2, and ST is surface type from AutoSnow product.

reanalysis products. AVHRR is investigated because 1) AVHRR data have been continuously collected since 1979 on multiple NOAA and recently on MetOp satellites, thus can provide great temporal sampling and long data record; 2) AVHRR on *NOAA-18* coincides well with *CloudSat* when *CloudSat* was operational during both daytime and nighttime, providing sufficient sampling for training and testing the AVHRR snowfall product; and 3) observations similar to AVHRR are expected to continue through VIIRS on NPP platforms.

Machine learning (here random forest) was used to establish relationships between *CloudSat* snowfall rate and AVHRR brightness temperatures (i.e., at 11- and 12- μm bands), AVHRR cloud probability, and auxiliary TWET and TPW information from MERRA-2. Three variations of the AVHRR product were studied, one with only inputs from AVHRR, one with TWET included (AVHRR I) and one with TWET and TPW included (AVHRR II). Initial analysis at footprint level showed that AVHRR I and AVHRR II perform better, so the two products were used for more detailed analysis. Comparisons were made using seasonal and geographical maps, monthly mean snowfall rates, Taylor diagrams, and an event-based case study. The results indicate that the AVHRR products, especially AVHRR II, are capable of detection and estimation of snowfall with accuracies comparable to or better than the current AIRS and GPM microwave (i.e., GPM-MHS *NOAA-18* and GPM-SSMIS *FI6*) products over land and ocean poleward of 60° latitude. The addition of TPW as a feature to the ML algorithm was found to be effective for both snowfall detection and estimation. It was also found that, over the Southern Ocean, the products have larger spread among themselves than over Antarctica and NH, which is consistent with previous studies. Analysis of GPM PMW products (MHS and SSMIS) suggests that while these products can generally capture snowfall patterns, they tend to consistently underestimate and underdetect snowfall events in high-latitude regions studied here. This calls for the need for more rigorous efforts to enhance snowfall detection and estimation from PMW sensors.

The outcomes of this study suggest that, pending major improvements in snowfall retrieval from PMW sensors, AVHRR snowfall retrievals can be considered as a valuable addition to the suite of satellite-based precipitation estimates in high latitudes. This is especially important in the development of the merged precipitation products (e.g., IMERG and GPCP), which require high temporal sampling and long-term records, both offered by the collection of AVHRR and AVHRR-like (e.g., VIIRS) observations.

Nonetheless, our study also showed that AVHRR snowfall retrievals based on RF can result in large regional biases. An important future work is to explore machine learning algorithms with higher capacities such as DNNs to deal with complexity of precipitation retrievals, especially in higher latitudes.

Acknowledgments. Financial support was made available from NASA MEaSUREs (NNH17ZDA001N-MEASURES) and NASA Weather and Atmospheric Dynamics (NNH19ZDA001N-ATDM) grants.

FIG. A1. Schematic representation of misses, hits, false alarms, and correct negatives used in the appendix to calculate the statistical metrics.

Data availability statement. *CloudSat* precipitation and cloud type products were downloaded from *CloudSat* data processing center <http://www.CloudSat.cira.colostate.edu>. MERRA-2, SSMIS, MHS, and AIRS data were obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC) at <https://disc.gsfc.nasa.gov>. ERA-5 data were downloaded from Copernicus: <https://cds.climate.copernicus.eu>. GPCP data were downloaded from <http://gpcp.umd.edu>. GPCC data were obtained from <https://www.dwd.de/EN/ourservices/gpcc>. Stage IV data were obtained from <https://data.eol.ucar.edu/dataset>. AVHRR orbital data were provided through PATMOS-x <https://www.ncdc.noaa.gov/cdr/atmospheric/avhrr-cloud-properties-patmos-x>.

APPENDIX

Definition of the Statistical Measures Used in This Study

Figure A1 is a schematic representation of misses, hits, false alarms, and correct negatives used to calculate the statistical metrics.

Quantitative statistics are obtained using estimated (est) and observed (obs) quantities. Categorical statistics are obtained using the contingency table. The quantitative and categorical statistics used in the present work are calculated as below, where N is the total number of observed and estimated snow pairs, and H , F , M , and Z are hits, false alarm, misses, and correct negative, respectively:

$$\text{F1 score} = \frac{H}{H + \frac{1}{2}(H + F)},$$

$$\text{areal bias} = \frac{H + F}{H + M},$$

$$\text{POD} = \frac{H}{H + M},$$

$$\text{FAR} = \frac{F}{H + F},$$

$$\text{HSS} = \frac{2(HZ - FM)}{(H + M)(M + Z) + (H + F)(F + Z)},$$

correlation coefficient

$$= \frac{\sum_{i=1}^N (\text{obs}_i)(\text{est}_i) - N(\overline{\text{obs}})(\overline{\text{est}})}{\sqrt{\sum_{i=1}^N [(\text{obs}_i)^2 - N(\overline{\text{obs}})^2] \sum_{i=1}^N [(\text{est}_i)^2 - N(\overline{\text{est}})^2]}},$$

$$\text{bias} = \frac{\sum_{i=1}^N (\text{obs}_i - \text{est}_i)}{\sum_{i=1}^N \text{obs}_i}, \quad \text{and}$$

$$\text{RMSE} = \frac{\sum_{i=1}^N (\text{obs}_i - \text{est}_i)^2}{N}.$$

REFERENCES

- Adhikari, A., C. Liu, and M. S. Kulie, 2018: Global distribution of snow precipitation features and their properties from 3 years of GPM observations. *J. Climate*, **31**, 3731–3754, <https://doi.org/10.1175/JCLI-D-17-0012.1>.
- , M. R. Ehsani, Y. Song, and A. Behrangi, 2020: Comparative assessment of snowfall retrieval from microwave humidity sounders using machine learning methods. *Earth Space Sci.*, **7**, e2020EA001357, <https://doi.org/10.1029/2020EA001357>.
- Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- , and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, <https://doi.org/10.3390/atmos9040138>.
- Arabzadeh, A., M. R. Ehsani, B. Guan, S. Hefflin, and A. Behrangi, 2020: Global intercomparison of atmospheric rivers precipitation in remote sensing and reanalysis products. *J. Geophys. Res. Atmos.*, **125**, e2020JD033021, <https://doi.org/10.1029/2020JD033021>.
- Ashouri, H., K. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, <https://doi.org/10.1175/BAMS-D-13-00068.1>.
- Battaglia, A., and G. Panegrossi, 2020: What can we learn from the CloudSat radiometric mode observations of snowfall over the ice-free ocean? *Remote Sens.*, **12**, 3285, <https://doi.org/10.3390/rs12203285>.
- Behrangi, A., M. Lebsock, S. Wong, and B. Lambriksen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *J. Geophys. Res.*, **117**, D20105, <https://doi.org/10.1029/2012JD017979>.
- , K. Andreadis, J. B. Fisher, F. J. Turk, S. Granger, T. Painter, and N. Das, 2014a: Satellite-based precipitation estimation and its application for streamflow prediction over mountainous western U.S. basins. *J. Appl. Meteor. Climatol.*, **53**, 2823–2842, <https://doi.org/10.1175/JAMC-D-14-0056.1>.
- , G. Stephens, R. F. Adler, G. J. Huffman, B. Lambriksen, and M. Lebsock, 2014b: An update on the oceanic precipitation rate and its zonal distribution in light of advanced observations from space. *J. Climate*, **27**, 3957–3965, <https://doi.org/10.1175/JCLI-D-13-00679.1>.
- , H. Nguyen, B. Lambriksen, M. Schreier, and V. Dang, 2015: Investigating the role of multi-spectral and near surface temperature and humidity data to improve precipitation detection at high latitudes. *Atmos. Res.*, **163**, 2–12, <https://doi.org/10.1016/j.atmosres.2014.10.019>.
- , and Coauthors, 2016: Status of high-latitude precipitation estimates from observations and reanalyses. *J. Geophys. Res. Atmos.*, **121**, 4468–4486, <https://doi.org/10.1002/2015JD024546>.
- , A. Gardner, J. T. Reager, J. B. Fisher, D. Yang, G. J. Huffman, and R. F. Adler, 2018: Using GRACE to estimate snowfall accumulation and assess gauge undercatch corrections in high latitudes. *J. Climate*, **31**, 8689–8704, <https://doi.org/10.1175/JCLI-D-18-0163.1>.
- , A. Singh, Y. Song, and M. Panahi, 2019: Assessing gauge undercatch correction in Arctic basins in light of GRACE observations. *Geophys. Res. Lett.*, **46**, 11 358–11 366, <https://doi.org/10.1029/2019GL084221>.
- , A. S. Gardner, and D. N. Wiese, 2020: Comparative analysis of snowfall accumulation over Antarctica in light of ice discharge and gravity observations from space. *Environ. Res. Lett.*, **15**, 104010, <https://doi.org/10.1088/1748-9326/ab9926>.
- Bennartz, R., F. Fell, C. Pettersen, M. D. Shupe, and D. Schuettemeyer, 2019: Spatial and temporal variability of snowfall over Greenland from CloudSat observations. *Atmos. Chem. Phys.*, **19**, 8101–8121, <https://doi.org/10.5194/acp-19-8101-2019>.
- Berg, W., T. L'Ecuyer, and C. Kummerow, 2006: Rainfall climate regimes: The relationship of regional TRMM rainfall biases to the environment. *J. Appl. Meteor. Climatol.*, **45**, 434–454, <https://doi.org/10.1175/JAM2331.1>.
- Bolvin, D. T., R. F. Adler, G. J. Huffman, E. J. Nelkin, and J. P. Poutiainen, 2009: Comparison of GPCP monthly and daily precipitation estimates with high-latitude gauge observations. *J. Appl. Meteor. Climatol.*, **48**, 1843–1857, <https://doi.org/10.1175/2009JAMC2147.1>.
- Brucker, L., and T. Markus, 2013: Arctic-scale assessment of satellite passive microwave-derived snow depth on sea ice using Operation IceBridge airborne data. *J. Geophys. Res. Oceans*, **118**, 2892–2905, <https://doi.org/10.1002/jgrc.20228>.
- Casella, D., G. Panegrossi, S. Dietrich, A. C. Marra, P. Sand, M. S. Kulie, and B. T. Johnson, 2017: Evaluation of the GPM-DPR snowfall detection capability: Comparison with CloudSat. *Atmos. Res.*, **197**, 64–75, <https://doi.org/10.1016/j.atmosres.2017.06.018>.
- Chen, S., and Coauthors, 2016: Comparison of snowfall estimates from the NASA CloudSat cloud profiling radar and NOAA/NSSL Multi-Radar Multi-Sensor system. *J. Hydrol.*, **541**, 862–872, <https://doi.org/10.1016/j.jhydrol.2016.07.047>.
- Edel, L., J. F. Rysman, C. Claud, C. Palermé, and C. Genthon, 2019: Potential of passive microwave around 183 GHz for snowfall detection in the Arctic. *Remote Sens.*, **11**, 2200, <https://doi.org/10.3390/rs11192200>.
- Ehsani, M. R., and Coauthors, 2020: 2019–2020 Australia fire and its relationship to hydroclimatological and vegetation variabilities. *Water*, **12**, 3067, <https://doi.org/10.3390/w12113067>.
- Ellis, T. D., T. L'Ecuyer, J. M. Haynes, and G. L. Stephens, 2009: How often does it rain over the global oceans? The perspective from CloudSat. *Geophys. Res. Lett.*, **36**, L03815, <https://doi.org/10.1029/2008GL036728>.

- Ferraro, R. R., and Coauthors, 2013: An evaluation of microwave land surface emissivities over the continental United States to benefit GPM-era precipitation algorithms. *IEEE Trans. Geosci. Remote Sens.*, **51**, 378–398, <https://doi.org/10.1109/TGRS.2012.2199121>.
- Frey, C. M., C. Kuenzer, and S. Dech, 2012: Quantitative comparison of the operational NOAA-AVHRR LST product of DLR and the MODIS LST product V005. *Int. J. Remote Sens.*, **33**, 7165–7183, <https://doi.org/10.1080/01431161.2012.699693>.
- Fuchs, T., J. Rapp, F. Rubel, and B. Rudolf, 2001: Correction of synoptic precipitation observations due to systematic measuring errors with special regard to precipitation phases. *Phys. Chem. Earth*, **26B**, 689–693, [https://doi.org/10.1016/S1464-1909\(01\)00070-3](https://doi.org/10.1016/S1464-1909(01)00070-3).
- Gelaro, R., W. McCarty, M. J. Suárez, R. Todling, and A. Molod, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson, 2004: Random forest classification of multisource remote sensing and geographic data. *IGARSS 2004: 2004 Int. Geoscience and Remote Sensing Symp. Proc.*, Vol. 2, 1049–1052, <https://doi.org/10.1109/IGARSS.2004.1368591>.
- Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison. World Meteorological Organization Rep. 67, 212 pp.
- Gutman, S. I., S. R. Sahm, S. G. Benjamin, B. E. Schwartz, K. L. Holub, J. Q. Stewart, and T. L. Smith, 2004: Rapid retrieval and assimilation of ground based GPS precipitable water observations at the NOAA Forecast Systems Laboratory: Impact on weather forecasts. *J. Meteor. Soc. Japan*, **82**, 351–360, <https://doi.org/10.2151/jmsj.2004.351>.
- Ham, J. S., Y. Chen, M. M. Crawford, and J. Ghosh, 2005: Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, **43**, 492–501, <https://doi.org/10.1109/TGRS.2004.842481>.
- Hastie, T., R. Tibshirani, and J. Friedman, 2009: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 745 pp., <https://doi.org/10.1007/978-0-387-84858-7>.
- Heidinger, A. K., M. J. Foster, A. Walther, and X. Zhao, 2014: The Pathfinder Atmospheres—Extended AVHRR climate dataset. *Bull. Amer. Meteor. Soc.*, **95**, 909–922, <https://doi.org/10.1175/BAMS-D-12-00246.1>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hou, A. Y., D. V. Ledvina, A. M. Da Silva, S. Q. Zhang, J. Joiner, R. M. Atlas, G. J. Huffman, and C. D. Kummerow, 2000: Assimilation of SSM/I-derived surface rainfall and total precipitable water for improving the GEOS analysis for climate studies. *Mon. Wea. Rev.*, **128**, 509–537, [https://doi.org/10.1175/1520-0493\(2000\)128<0509:AOSIDS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0509:AOSIDS>2.0.CO;2).
- Hsu, K. L., X. G. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteor.*, **36**, 1176–1190, [https://doi.org/10.1175/1520-0450\(1997\)036<1176:PEFRSI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2).
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, <https://doi.org/10.1029/2009GL040000>.
- , E. F. Stocker, D. T. Bolvin, and E. J. Nelkin, J. Tan, 2019: GPM IMERG Final Precipitation L3 1 day 0.1 degree $\times$ 0.1 degree V06. Goddard Earth Sciences Data and Information Services Center, accessed 23 August 2020, <https://doi.org/10.5067/GPM/IMERGDF/DAY/06>.
- , R. F. Adler, A. Behrangi, D. T. Bolvin, E. J. Nelkin, and Y. Song, 2020: Algorithm Theoretical Basis Document (ATBD) for global precipitation climatology project version 3.1 precipitation data. MEaSUREs project, 32 pp., https://docserver.gesdisc.eosdis.nasa.gov/public/project/MEaSUREs/GPCP/GPCP_ATBD_V3.1.pdf.
- Inoue, T., 1987a: A cloud type classification with NOAA 7 split-window measurements. *J. Geophys. Res.*, **92**, 3991–4000, <https://doi.org/10.1029/JD092iD04p03991>.
- , 1987b: An instantaneous delineation of convective rainfall areas using split window data of NOAA-7 AVHRR. *J. Meteor. Soc. Japan*, **65**, 469–481, https://doi.org/10.2151/jmsj1965.65.3_469.
- Joyce, R. J., and P. Xie, 2011: Kalman filter-based CMORPH. *J. Hydrometeorol.*, **12**, 1547–1563, <https://doi.org/10.1175/JHM-D-11-022.1>.
- Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, <https://doi.org/10.1175/BAMS-D-14-00283.1>.
- Kühnlein, M., T. Appelhans, B. Thies, and T. Nauss, 2014: Improving the accuracy of rainfall rates from optical satellite sensors with machine learning—A random forests-based approach applied to MSG SEVIRI. *Remote Sens. Environ.*, **141**, 129–143, <https://doi.org/10.1016/j.rse.2013.10.026>.
- Kulie, M. S., L. Milani, N. B. Wood, S. A. Tushaus, R. Bennartz, and T. S. L'Ecuyer, 2016: A shallow cumulusiform snowfall census using spaceborne radar. *J. Hydrometeorol.*, **17**, 1261–1279, <https://doi.org/10.1175/JHM-D-15-0123.1>.
- Kurino, T., 1997: A satellite infrared technique for estimating “deep/shallow” precipitation. *Adv. Space Res.*, **19**, 511–514, [https://doi.org/10.1016/S0273-1177\(97\)00063-X](https://doi.org/10.1016/S0273-1177(97)00063-X).
- Kwok, R., and T. Markus, 2018: Potential basin-scale estimates of Arctic snow depth with sea ice freeboards from CryoSat-2 and ICESat-2: An exploratory analysis. *Adv. Space Res.*, **62**, 1243–1250, <https://doi.org/10.1016/J.ASR.2017.09.007>.
- Lebsock, M. D., and T. S. L'Ecuyer, 2011: The retrieval of warm rain from CloudSat. *J. Geophys. Res.*, **116**, D20209, <https://doi.org/10.1029/2011JD016076>.
- Levizzani, V., S. Laviola, and E. Cattani, 2011: Detection and measurement of snowfall from space. *Remote Sens.*, **3**, 145–166, <https://doi.org/10.3390/rs3010145>.
- Lin, Y., 2011: GCIPEOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV data, version 1.0. UCAR/NCAR—Earth Observing Laboratory, accessed 15 June 2020, <https://doi.org/10.5065/D6PG1QDD>.
- Liu, G., 2008: Deriving snow cloud characteristics from CloudSat observations. *J. Geophys. Res.*, **113**, D00A09, <https://doi.org/10.1029/2007JD009766>.
- , and J. A. Curry, 1997: Precipitation characteristics in Greenland-Iceland-Norwegian Seas determined by using satellite microwave data. *J. Geophys. Res.*, **102**, 13 987–13 997, <https://doi.org/10.1029/96JD03090>.
- , and E.-K. Seo, 2013: Detecting snowfall over land by satellite high-frequency microwave observations: The lack of scattering signature and a statistical approach. *J. Geophys. Res. Atmos.*, **118**, 1376–1387, <https://doi.org/10.1002/jgrd.50172>.
- Mace, G. G., and Q. Zhang, 2014: The CloudSat radar-lidar geometrical profile product (RL-GeoProf): Updates, improvements,

- and selected results. *J. Geophys. Res. Atmos.*, **119**, 9441–9462, <https://doi.org/10.1002/2013JD021374>.
- Meng, H., J. Dong, R. Ferraro, B. Yan, L. Zhao, C. Kongoli, N.-Y. Wang, and B. Zavadsky, 2017: A 1DVAR-based snowfall rate retrieval algorithm for passive microwave radiometers. *J. Geophys. Res. Atmos.*, **122**, 6520–6540, <https://doi.org/10.1002/2016JD026325>.
- Palmer, C., J. E. Kay, C. Genthon, T. L'Ecuyer, N. B. Wood, and C. Claud, 2014: How much snow falls on the Antarctic ice sheet? *Cryosphere*, **8**, 1577–1587, <https://doi.org/10.5194/tc-8-1577-2014>.
- , C. Claud, N. B. Wood, T. L. Ecuyer, and C. Genthon, 2019: How does ground clutter affect CloudSat snowfall retrievals over ice sheets? *IEEE Geosci. Remote Sens. Lett.*, **16**, 342–346, <https://doi.org/10.1109/LGRS.2018.2875007>.
- Panahi, M., and A. Behrangi, 2019: Comparative analysis of snowfall accumulation and gauge undercatch correction factors from diverse data sets: In situ, satellite, and reanalysis. *Asia-Pac. J. Atmos. Sci.*, **56**, 615–628, <https://doi.org/10.1007/s13143-019-00161-6>.
- Panegrossi, G., J. F. Rysman, D. Casella, A. C. Marra, P. Sanò, and M. S. Kulie, 2017: CloudSat-based assessment of GPM microwave imager snowfall observation capabilities. *Remote Sens.*, **9**, 1263, <https://doi.org/10.3390/rs9121263>.
- Prat, O. P., and B. R. Nelson, 2015: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012). *Hydrol. Earth Syst. Sci.*, **19**, 2037–2056, <https://doi.org/10.5194/hess-19-2037-2015>.
- Romanov, P., 2017: Global Multisensor Automated satellite-based Snow and Ice Mapping System (GMAI) for cryosphere monitoring. *Remote Sens. Environ.*, **196**, 42–55, <https://doi.org/10.1016/j.rse.2017.04.023>.
- Rysman, J. F., G. Panegrossi, P. Sanò, A. Marra, S. Dietrich, L. Milani, and M. Kulie, 2018: SLALOM: An all surface snow water path retrieval algorithm for the GPM Microwave Imager. *Remote Sens.*, **10**, 1278, <https://doi.org/10.3390/rs10081278>.
- , and Coauthors, 2019: Retrieving surface snowfall with the GPM microwave imager: A new module for the SLALOM algorithm. *Geophys. Res. Lett.*, **46**, 13 593–13 601, <https://doi.org/10.1029/2019GL084576>.
- Schneider, U., P. Finger, A. Meyer-Christoffer, E. Rustemeier, M. Ziese, and A. Becker, 2017: Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere*, **8**, 52, <https://doi.org/10.3390/atmos8030052>.
- Sims, E. M., and G. Liu, 2015: A parameterization of the probability of snow–rain transition. *J. Hydrometeor.*, **16**, 1466–1477, <https://doi.org/10.1175/JHM-D-14-0211.1>.
- Skofronick-Jackson, G., and Coauthors, 2017: The Global Precipitation Measurement (GPM) mission for science and society. *Bull. Amer. Meteor. Soc.*, **98**, 1679–1695, <https://doi.org/10.1175/BAMS-D-15-00306.1>.
- , M. Kulie, L. Milani, S. J. Munchak, N. B. Wood, and V. Levizzani, 2019: Satellite estimation of falling snow: A Global Precipitation Measurement (GPM) Core Observatory perspective. *J. Appl. Meteor. Climatol.*, **58**, 1429–1448, <https://doi.org/10.1175/JAMC-D-18-0124.1>.
- Smalley, M., T. L'Ecuyer, M. Lebsock, and J. Haynes, 2014: A comparison of precipitation occurrence from the NCEP Stage IV OPE product and the CloudSat cloud profiling radar. *J. Hydrometeor.*, **15**, 444–458, <https://doi.org/10.1175/JHM-D-13-048.1>.
- , P. Kirstetter, and T. L'Ecuyer, 2017: How frequent is precipitation over the contiguous United States? Perspectives from ground-based and spaceborne radars. *J. Hydrometeor.*, **18**, 1657–1672, <https://doi.org/10.1175/JHM-D-16-0242.1>.
- Song, Y., A. Behrangi, and E. Blanchard-Wrigglesworth, 2020: Assessment of satellite and reanalysis cold season snowfall estimates over Arctic sea ice. *Geophys. Res. Lett.*, **47**, e2020GL088970, <https://doi.org/10.1029/2020GL088970>.
- Sorooshian, S., K. L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSSE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2).
- Stephens, G. L., and Coauthors, 2002: The CloudSat mission and the A-Train. *Bull. Amer. Meteor. Soc.*, **83**, 1771–1790, <https://doi.org/10.1175/BAMS-83-12-1771>.
- Sun, N., and F. Weng, 2008: Evaluation of Special Sensor Microwave Imager/Sounder (SSMIS) environmental data records. *IEEE Trans. Geosci. Remote Sens.*, **46**, 1006–1016, <https://doi.org/10.1109/TGRS.2008.917368>.
- Susskind, J., P. Piraino, L. Rokke, L. Iredell, and A. Mehta, 1997: Characteristics of the TOVS Pathfinder Path A dataset. *Bull. Amer. Meteor. Soc.*, **78**, 1449–1472, [https://doi.org/10.1175/1520-0477\(1997\)078<1449:COTTPP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<1449:COTTPP>2.0.CO;2).
- Tang, G., D. Long, A. Behrangi, C. Wang, and Y. Hong, 2018: Exploring deep neural networks to retrieve rain and snow in high latitudes using multisensor and reanalysis data. *Water Resour. Res.*, **54**, 8253–8278, <https://doi.org/10.1029/2018WR023830>.
- Wang, T., E. J. Fetzer, S. Wong, B. H. Kahn, and Q. Yue, 2016: Validation of MODIS cloud mask and multilayer flag using CloudSat-CALIPSO cloud profiles and a cross-reference of their cloud classifications. *J. Geophys. Res. Atmos.*, **121**, 11 620–11 635, <https://doi.org/10.1002/2016JD025239>.
- Webster, M. A., I. G. Rigor, S. V. Nghiem, N. T. Kurtz, S. L. Farrell, D. K. Perovich, and M. Sturm, 2014: Interdecadal changes in snow depth on Arctic sea ice. *J. Geophys. Res. Oceans*, **119**, 5395–5406, <https://doi.org/10.1002/2014JC009985>.
- Wen, Y., A. Behrangi, B. Lambrigtsen, and P.-E. Kirstetter, 2016: Evaluation and uncertainty estimation of the latest radar and satellite snowfall products using SNOTEL measurements over mountainous regions in western United States. *Remote Sens.*, **8**, 904, <https://doi.org/10.3390/rs8110904>.
- Wood, N. B., T. S. L'Ecuyer, D. G. Vane, G. L. Stephens, and P. Partain, 2013: Level 2C snow profile process description and interface control document. JPL Doc., 21 pp., http://www.CloudSat.cira.colostate.edu/sites/default/files/products/files/2C-SNOW-PROFILE_PDICD.P_R04.20130210.pdf.
- Xie, P., R. Joyce, and S. Wu, 2018: Enhancing the monitoring of global precipitation with the second generation CMORPH integrated satellite estimates. *43rd Climate Diagnostics and Prediction Workshop*, Santa Barbara, CA, NOAA, <https://www.cpc.ncep.noaa.gov/products/outreach/CDPW/43/oral-sessions/presentations/thurs/PXIE-CDPW2018.pdf>.
- Yan, B., and F. Weng, 2008: Intercalibration between special sensor microwave imager/sounder and special sensor microwave imager. *IEEE Trans. Geosci. Remote Sens.*, **46**, 984–995, <https://doi.org/10.1109/TGRS.2008.915752>.
- Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison, 2005: Bias corrections of long-term (1973–2004) daily precipitation

- data over the northern regions. *Geophys. Res. Lett.*, **32**, L19501, <https://doi.org/10.1029/2005GL024057>.
- You, Y., N.-Y. Wang, R. Ferraro, and S. Rudlosky, 2017: Quantifying the snowfall detection performance of the GPM Microwave Imager channels over land. *J. Hydrometeor.*, **18**, 729–751, <https://doi.org/10.1175/JHM-D-16-0190.1>.
- Zhang, J., L. Lin, and R. L. Bras, 2018: Evaluation of the quality of precipitation products: A case study using WRF and IMERG data over the central United States. *J. Hydrometeor.*, **19**, 2007–2020, <https://doi.org/10.1175/JHM-D-18-0153.1>.
- Ziese, M., A. Rauthe-Schöch, A. Becker, P. Finger, A. Meyer-Christoffer, and U. Schneider, 2018: GPCC full data daily version 2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and historic data. Deutscher Wetterdienst, accessed 8 September 2020, https://doi.org/10.5676/DWD_GPCC/FD_D_V2018_100.