SPECIAL

Global Precipitation Measurement (GPM)

COLLECTION

A Prognostic Nested *k*-Nearest Approach for Microwave Precipitation Phase Detection over Snow Cover

ZEINAB TAKBIRI AND ARDESHIR EBTEHAJ

Department of Civil, Environmental and Geo-Engineering, and St. Anthony Falls Laboratory,
University of Minnesota, Twin Cities, Minneapolis, Minnesota

EFI FOUFOULA-GEORGIOU

Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, California

PIERRE-EMMANUEL KIRSTETTER

Advanced Radar Research Center, University of Oklahoma, and National Severe Storms Laboratory,
Norman, Oklahoma

F. JOSEPH TURK

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

(Manuscript received 3 February 2018, in final form 8 November 2018)

ABSTRACT

Monitoring changes of precipitation phase from space is important for understanding the mass balance of Earth's cryosphere in a changing climate. This paper examines a Bayesian nearest neighbor approach for prognostic detection of precipitation and its phase using passive microwave observations from the Global Precipitation Measurement (GPM) satellite. The method uses the weighted Euclidean distance metric to search through an a priori database populated with coincident GPM radiometer and radar observations as well as ancillary snow-cover data. The algorithm performance is evaluated using data from GPM official precipitation products, ground-based radars, and high-fidelity simulations from the Weather Research and Forecasting Model. Using the presented approach, we demonstrate that the hit probability of terrestrial precipitation detection can reach to 0.80, while the probability of false alarm remains below 0.11. The algorithm demonstrates higher skill in detecting snowfall than rainfall, on average by 10%. In particular, the probability of precipitation detection and its solid phase increases by 11% and 8%, over dry snow cover, when compared to other surface types. The main reason is found to be related to the ability of the algorithm in capturing the signal of increased liquid water content in snowy clouds over radiometrically cold snow-covered surfaces.

1. Introduction

More than two billion people rely on glacier and snowmelt for their water supply (Mankin et al. 2015). Snowfall accounts for approximately 30%–90% of the global precipitation over mid- to high latitudes (Levizzani et al. 2011) and is the main input to the accumulation processes of snowpack and glaciers (Radić et al. 2014). In recent decades, snowpack reservoirs have declined and are projected to further decline in

the twenty-first century (Karl et al. 1993; Mote et al. 2005; Pederson et al. 2011). Thus, global monitoring of snowfall from space is key for improved understanding and prediction of ongoing changes in the cryosphere and the implications for sustainable management of water and food resources—especially in mountainous areas of the world.

In the past three decades, significant progress has been made in microwave precipitation retrieval as part of the Tropical Rainfall Measuring Mission (TRMM) satellite in 1997 (Kummerow et al. 1998). The launch of the Global Precipitation Measurement (GPM) core

Corresponding author: Zeinab Takbiri, takbi001@umn.edu

satellite (Kidd and Huffman 2011; Hou et al. 2014) has provided a unique opportunity for improved understanding of midlatitude precipitation and its phase change beyond what the TRMM satellite could offer (Skofronick-Jackson et al. 2017).

The snowfall microwave scattering signal can be captured at frequencies above 80 GHz as these high frequencies are more sensitive to ice scattering compared to lower frequencies, which largely respond to variations of surface emissivity (Kulie et al. 2010; Skofronick-Jackson and Johnson 2011; Gong and Wu 2017; You et al. 2017). Among high-frequency channels, Bennartz and Bauer (2003) found that frequencies around and above 150 GHz provide a strong polarization signal for snowfall detection (Gong and Wu 2017; You et al. 2017; Panegrossi et al. 2017).

Remote sensing of snowfall is among the most challenging tasks in precipitation retrieval algorithms (Bennartz and Bauer 2003; Skofronick-Jackson et al. 2004; Noh et al. 2009; Kongoli et al. 2015). Detection of snowfall is challenging because it involves complex and dynamic interactions between the snowfall scattering signal and the surface. First, compared to rainfall, the snowfall backscattering is much weaker (Grody 1991; Kim et al. 2008; Kulie et al. 2010) and depends on more complex microphysical characteristics snowfall such as shape and density of snowflakes (Liu 2008; Petty et al. 2010; Skofronick-Jackson and Johnson 2011). These characteristics are difficult to accurately parameterize as of today. Second, the already weak snowfall scattering signal tends to be masked by the increased atmospheric emissivity and liquid water content in precipitating conditions (Liu and Seo 2013; Wang et al. 2013; Panegrossi et al. 2017). Third, changes in surface emissivity due to snow accumulation on the ground can significantly alter the snowfall microwave signal. Dry snow cover scatters the upwelling surface radiation at frequencies above 20 GHz (Ulaby and Stiles 1980; Hallikainen et al. 1987) similar to the snowfall (Grody 2008). As a result, the snowfall microwave signature gradually weakens as snow accumulates on the ground (Ebtehaj and Kummerow 2017). The snow-cover scattering evolves with time as a function of snow-cover metamorphism. For example, a small amount of liquid water content (e.g., 2%) significantly reduces the snow-cover scattering and increases its absorptivity (Stiles and Ulaby 1980; Hallikainen et al. 1986, 1987). Hence, snow cover has a time-varying effect on snowfall upwelling signal.

Physical and empirical approaches have been developed for microwave retrievals of snowfall. Skofronick-Jackson et al. (2004) presented a physical method to retrieve snowfall during a blizzard over the eastern United States using high-frequency observations from the Advanced Microwave Sounding Unit B (AMSU-B) instrument. Kim et al. (2008) simulated atmospheric profiles of a blizzard storm with the mesoscale MM5 model and a delta-Eddington-type radiative transfer (RT) model to produce a storm-scale database for snowfall retrieval using AMSU-B observations. Noh et al. (2009) used a large number of snowfall profiles from airborne, surface, and satellite radars, as well as an atmospheric RT model (Liu 1998) to generate a regional database for snowfall retrievals using the AMSU-B data. The study used the NESDIS Microwave Land Surface Emissivity Model (Weng et al. 2001) to provide surface emissivity as an input to the RT model. The largest retrieval errors were found to be over snow-covered surfaces.

Empirical passive microwave snowfall retrieval algorithms largely rely on ancillary data of precipitation radar and air temperature profile. A family of these algorithms relies on thresholding the brightness temperature at different channels (e.g., Staelin and Chen 2000; Chen and Staelin 2003; Kongoli et al. 2003). For example, Kongoli et al. (2015) developed a statistical approach that partitions high-frequency brightness temperatures ($\geq$89 GHz) into two distinct warm and cold weather regimes by thresholding the brightness temperature at 53 GHz.

Another class of empirical approaches relies on Bayesian techniques. These techniques use a database or a lookup table that relates brightness temperatures of snowing clouds to the radar snowfall observations along with the atmospheric temperature profile. As an example, Liu and Seo (2013) used matched observations from the CloudSat Profiling Radar (CPR), the AMSU-B, and NOAA's Microwave Humidity Sounder (MHS). More recently, Sims and Liu (2015) used the CloudSat radar and multiple ground-based reanalysis data, including near-surface air temperature, atmospheric moisture, low-level vertical temperature lapse rate, surface skin temperature, surface pressure, and land cover types to diagnose precipitation phase partitioning. This algorithm is deployed in the GPM operational precipitation retrievals (Kummerow et al. 2015). It is worth noting that most of these algorithms use reanalysis wet-bulb temperature that exhibits the strongest correlation with the precipitation phase (Matsuo et al. 1981; Motoyama 1990; Lundquist et al. 2008; Kienzle 2008; Ye et al. 2013). However, the reanalysis data are often available at coarse spatial scales with significant uncertainty, which hamper the applicability for accurate detection of snowfall (Harpold et al. 2017).

In this paper, we examine a prognostic Bayesian k-nearest neighbor (KNN) algorithm that strictly relies on observed passive microwave brightness temperatures and does not use any online reanalysis data of temperature

and moisture profiles. This approach is based on a weighted distance metric applied on an a priori database to detect overland precipitation phase. The a priori database is populated with combined radar–radiometer observations from the GPM satellite. This database is then stratified using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor to account for effects of the background snow-cover emission. We demonstrate that the algorithm shows improved skill in detection of snowfall over snow cover and can predict the likelihood of precipitation phase changes in the atmospheric boundary layer, which is not well observed by the GPM radar.

In summary, the presented algorithm isolates a few physically relevant candidate vectors of brightness temperatures in the database via a weighted Euclidean distance. Using these isolated candidates, the method detects the precipitation and its phase, based on a probabilistic decision rule. To test the performance of the proposed approach, the database is populated by merging the outputs of both passive (Sims and Liu 2015) and active (Iguchi et al. 2017) GPM products using all overland observations from June 2015 to May 2016. We compare the results with the ground-based Multi-Radar Multi-Sensor (MRMS) data over the conterminous United States (CONUS; Zhang et al. 2011, 2016). The outputs of a high-fidelity mesoscale simulation model are also used for further evaluation of the results over high altitudes, during the Olympic Mountains Experiment (OLYMPEX) in 2015 (Houze et al. 2017).

The paper is structured as follows. Section 2 briefly describes the database and the phase detection method used on the operational GPM radar and radiometer products. Section 3 elaborates on the effects of snow cover on passive microwave signal of snowfall at different frequency channels by analyzing a large dataset of GPM observations. Section 4 explains the proposed KNN algorithm followed by the results presented in section 5. Concluding remarks and future directions of the research are discussed in section 6.

2. Database description

The Dual-Frequency Precipitation Radar (DPR) aboard the GPM core satellite detects precipitation reflectivity at Ka (35 GHz) and Ku band (13.6 GHz). The GPM Microwave Imager (GMI) captures the upwelling emission/scattering signals of the surface and the atmosphere at 13 frequency channels ranging from 10 to 183 GHz. On the one hand, observations by the DPR and the GMI high-frequency channels (>80 GHz) provide information about the microwave signature of precipitation and more specifically about snowfall ice scattering. On the

other hand, observations by the low-frequency channels (>80 GHz) add information about the land surface characteristics that leads to improved detection skill by the presented algorithm. This study uses level-II near-surface precipitation phase retrieval from DPR (active) product (2A-DPR-V04, normal scan), GMI (passive) product (2A-GPROF-V04), and the level 1B calibrated GMI brightness temperatures.

In DPR level II, the precipitation phase is determined by the dual-frequency retrieval approach that uses the differential attenuation between the Ku- and the Ka-band reflectivity values (Iguchi et al. 2012, 2017). The differential attenuation method ingests ancillary atmospheric profile data such as air temperature, pressure, and the microphysical parameterization of the snow and rain particle size distribution. The DPR surface retrieval is inferred from the near-surface reflectivity observations in the clutter-free region. Above relatively flat surfaces, the altitude of this region varies from 1 to 2 km from nadir to the edge of the DPR swath. The depth of this region is often increased over complex terrains. As a result, any precipitation within the ground clutter region cannot be detected by the radar. Moreover, DPR has limited capability to detect light precipitation with a rate below ~ 0.2 mm h^{-1} (Hou et al. 2014; Kubota et al. 2014).

Unlike the DPR that provides range-resolved information about the precipitation backscattering, the GMI observes an integration of precipitation scattering in a continuum that extends from the land surface to the top of the atmosphere. As previously explained, the current operational algorithm for passive detection of precipitation phase relies on thresholding of the near-surface wet-bulb temperature (Sims and Liu 2015). The wet-bulb temperature is processed offline from reanalysis of ancillary data, which often suffer from different sources of uncertainty, especially due to its coarse resolution over topographic features and structurally complex land surfaces (Li et al. 2008).

For implementing and testing the proposed algorithm, we create a reference product (REF) for precipitation occurrence and phase change. This REF product is based solely on the occurrence information from the DPR data. For determining the precipitation phase, we use the inner-swath phase information from both GPM active and passive products. None of these products provides direct phase estimation; however, each has unique information based on the atmospheric and surface conditions. Specifically, the REF product determines the phase by applying a logical operator to both active and passive products. The radar phase retrieval is reported as solid, liquid, and mixed, while the phase probability in GPROF is from zero (solid) to one (liquid). We therefore first discretize the GPROF phase probability into solid (phase

probability less than 0.5) and liquid (phase probability greater than 0.5) to match the radar phase. Second, we assign the phase of REF precipitation as solid or liquid when both active and passive phases are solid or liquid. Otherwise, the phase is labeled as mixed. Therefore, the mixed phase in the REF product refers to those cases where the precipitation phases from the passive and active products do not agree and thus should not be literally interpreted. By combining the active and passive phase information through this logical rule, we implicitly address the limitations of DPR in identifying precipitation phase change in the ground clutter region which overlaps with the boundary layer.

It is important to note that the so-called mixed category depends on the threshold (0.5), used for discretization of the passive phase. Understanding the impacts of this threshold on the retrieval requires a thorough investigation outside the scope of this study. It is worth noting that choosing this threshold results in 12% of mixed phase data in the REF product, in which 10% corresponds to liquid passive phase and solid active phase (scenario 1) and 2% otherwise (scenario 2). The first scenario might be related to those conditions where the melting layer is in the clutter region. The second scenario may be related to a temperature inversion near the surface that causes a refreezing of precipitation.

The MODIS daily snow-cover fraction (MOD10A1; Hall et al. 2002) and land surface skin temperature (MOD11A1; Wan 2014) are added to the database. The data are used from the MODIS sensor on board the *Terra* satellite. The MODIS snow cover and surface temperature data are available at a resolution of 500 and 1000 m, respectively. We assume that the total daytime snow-cover fraction does not change significantly between consecutive overpasses of the GPM and *Terra* satellites within one day. Note that this assumption could give rise to some degree of uncertainty when the data are matched with nighttime precipitation. We consider a 5-km DPR pixel as a snow-covered surface when more than 50% of the enclosed high-resolution snow fraction data indicates the presence of snow cover on the ground. It is also assumed that the temperature does not vary significantly within a 5-km DPR pixel and is considered to be the average of the cloud-free MODIS temperature data. As the liquid water content of global snowpack is not available, we define dry (wet) snow when the skin and air temperature are both below (above) 0°C (Baggi and Schweizer 2009).

To account for atmospheric radiometric signals, we also added the integrated liquid and ice water content of the clouds, as well as the integrated water vapor content of the atmospheric column from the second version of the Modern-Era Retrospective Analysis for Research

and Applications (MERRA-2-M2I1NXASM; Gelaro et al. 2017). The MERRA-2 data are hourly single-level diagnostic products at $0.625° \times 0.5°$, which are derived from version 5 of the NASA Goddard Earth Observing System (GEOS-5) and the Atmospheric Data Assimilation System (ADAS).

To form the database with a uniform spatial sampling density, the GMI brightness temperatures and the MERRA-2 reanalysis data are mapped onto the DPR grids and scanning time using the spatial nearest neighbor interpolation and temporal linear interpolation techniques. The high-resolution MODIS snow-cover data are mapped onto and then averaged over the nearest DPR grids. We collect and process two consecutive years of data, from June 2014 to May 2016, which lead to a database with more than 5×10^9 matched data pairs. The data from the first year (June 2014–May 2015) are applied to build the database and the data from the second year (June 2015–May 2016) are used to test the proposed algorithm.

3. The effect of snow cover on terrestrial snowfall signal

Precipitation spectral signatures and their dependence on snow-cover scattering are studied by analyzing the entire dataset (June 2014–May 2016) for three surface types (ground without snow cover, wet snow, and dry snow) and for both liquid and solid phases of precipitation. Each land–atmospheric class is further divided into five bins of precipitation intensity r with equal logarithmic width, $\log_2(r_{i+1}/r_i) = 1$, centered at 0.5, 1, 2, 4, and $8\,mm\,h^{-1}$. We first quantify the effects of snow cover on the precipitation signal over each surface type by calculating the mean values of the brightness temperatures for different precipitation phases and intensities at frequency bands 10–19, 36–89, 166, 183 ± 3, and 183 ± 7 GHz (Fig. 1). Then, we demonstrate that the snowfall signal exhibits a weaker scattering signal than rainfall and reveal that there exists a nonunique relationship between the brightness temperatures and snowfall rate over snow-covered surfaces. Last, we highlight why precipitation phase detection could be less uncertain over dry than wet snow cover using the presented approach.

The first three columns in Figs. 1a–i focus on the signatures of rainfall over land surfaces with no snow cover, wet snow cover, and dry snow cover, where both active and passive products indicate liquid phase. The signatures over the ground with no snow cover are mainly affected by the upwelling surface emission, the upwelling emission by cloud liquid water path, as well as scattering by the cloud ice particles and large raindrops. As it is well
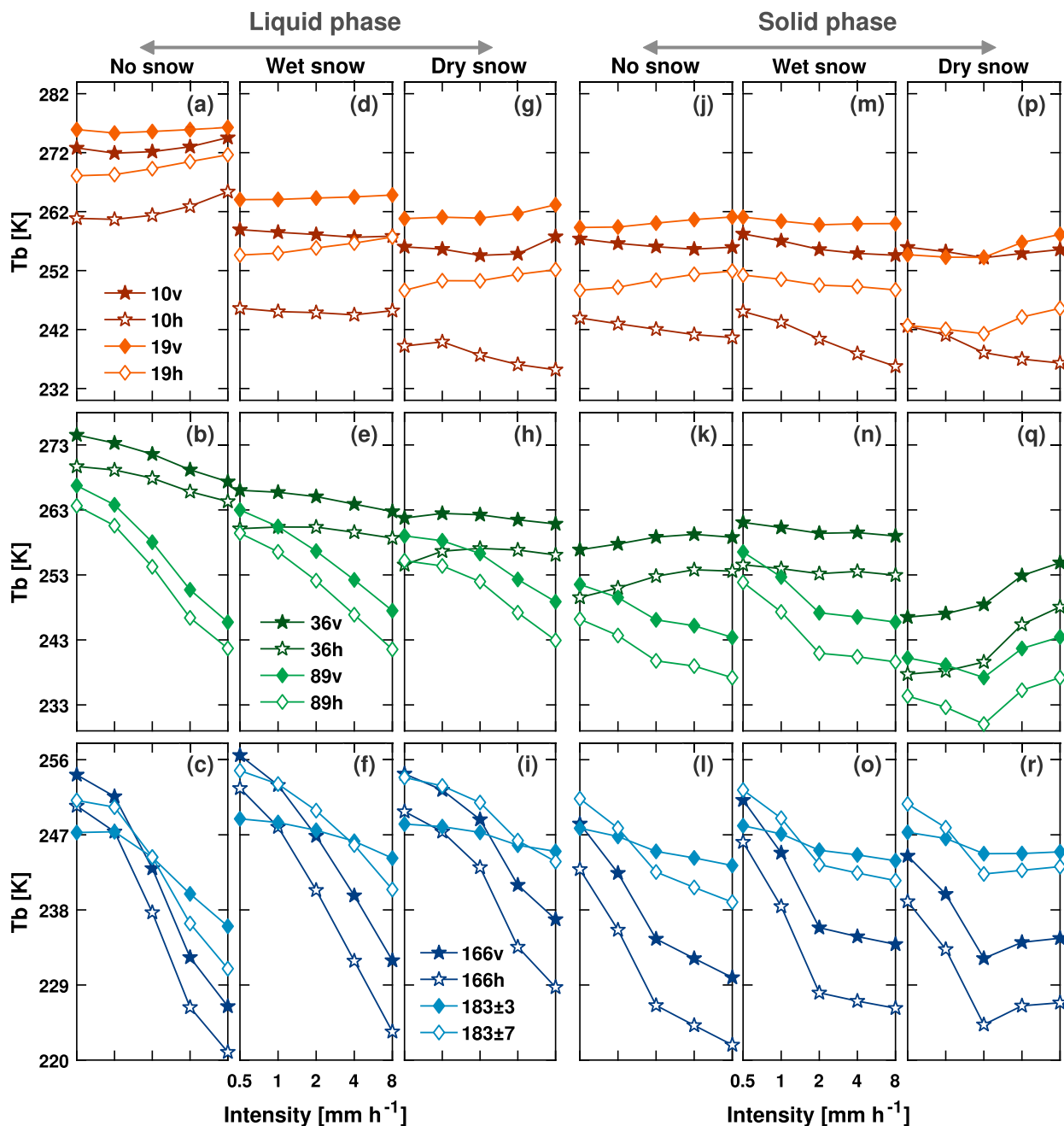
FIG. 1. Variation of mean brightness temperatures (June 2014–May 2016) in response to changes in precipitation intensity for different land–atmosphere classes including the (a)–(i) liquid and (j)–(r) solid precipitation phase over the ground (no snow) and wet and dry snow cover. The bins are five logarithmically (base 2) spaced intervals with the width of $1 \, mm \, h^{-1}$ in the log space.

understood, due to strong background emission at frequencies 10–36 GHz, the overland precipitation microwave signal is difficult to be separated from the surface contributed signal in these channels. For example, due to the rainfall emission, the mean brightness temperature at 10h GHz only increases by less than 5 K as the intensity increases from 0.5 to $8 \, mm \, h^{-1}$ (Fig. 1a).

On average, we observe that over all three land surface types, the brightness temperatures monotonically decrease for frequencies above 80 GHz as the rainfall intensity increases. However, the significance of scattering decreases over snow-covered surfaces (Figs. 1d–i). For example, at 89 and 166 GHz, the average decrease of brightness temperature per $1 \, mm \, h^{-1}$ increase in rainfall

intensity is about 3.0 and 3.6 K (Figs. 1d,g), while these rates are around 1.2 and 2.3 K over the dry snow cover (Figs. 1h,i). As expected, the 183 ± 3 GHz is the least sensitive channel to the changes of rainfall rate. This channel becomes almost insensitive to the rainfall intensity when the snow accumulates on the ground and exhibits less than 0.2 K of cooling effect per unit rainfall (Fig. 1i).

The last three columns in Figs. 1j–r present brightness temperatures of snowfall over the three explained land surface types, where both active and passive products indicate solid phase. Similar to the overland rainfall, the emission and the scattering signals become more significant from low- to high-frequency channels. Over the surfaces with no snow cover, when the snowfall intensity increases from 0.5 to 8 mm h^{-1}, the brightness temperatures at frequencies ≤ 36 GHz increase ~ 6 K (Fig. 1 j). This warming could be due to increased cloud liquid water path (from 75 to 101 g m^{-2}; Figs. 2a,d), water vapor path (from 9.5 to 13.1 kg m^{-2}; Figs. 2c,f), and surface temperature (from 273 to 274.2 K; Figs. 2g,i).

As a result of the snowfall scattering, the average brightness temperature at the 166-GHz frequency channel (Figs. 1l,o,r) decreases about 14–20 K, which corresponds to a cooling rate of 1.75–2.50 K per unit snowfall rate. This observation reaffirms the importance of 166 GHz for snowfall retrieval compared to the 89-GHz channel (see Bennartz and Bauer 2003; Shi et al. 2010; Skofronick-Jackson et al. 2013; You et al. 2017). When the precipitation intensity increases from 0.5 to 8 mm h^{-1}, the average decrease in brightness temperatures at 166 (89) GHz is about 18–30 (10–22) K for rainfall and 10–20 (2–9) K for snowfall over all examined land surface types. Therefore, the scattering signal weakens when the precipitation falls in the solid form; however, this weakening effect is less significant at 166 GHz than 89 GHz. In particular, over the ground with no snow cover, the signal becomes weaker approximately by 30% and 57% at 166 and 89 GHz, respectively, while these rates are 44% and 80% over the dry snow cover.

Observations demonstrate that the snowfall scattering signal decreases at frequencies ≥ 89 GHz when snow begins to accumulate on the ground. An interesting observation is the nonmonotonic response of the observed brightness temperatures to the snowfall rate over snow-covered surfaces. For example, over the dry snow, the brightness temperatures at ≥ 89 GHz increase when the snowfall intensity varies from 2 to 4 mm h^{-1}, showing an irregular transition from a scattering to an emission regime (Figs. 1q,r). Although less pronounced, a similar pattern is observed over the wet snow cover (Figs. 1n,o).

The possible reasons for this anomaly could be related to an emission signal from either the atmosphere or the land surface. The atmospheric-related reasons can be due to the enhanced emission from the cloud liquid water and/or the water vapor path; both of them often increase with increasing snowfall intensity (Liu and Seo 2013; You et al. 2017; Ebtehaj and Kummerow 2017). The land-surface-related causes largely correspond to the increased surface temperature and/or changes in the snow-cover physical properties. To find the most prominent contributing factor, we analyzed the variations of liquid, ice, and vapor water path derived from MERRA-2 data, the surface temperature derived from MODIS, and the mean snowfall intensity at different latitudes (Figs. 2a–f).

Over the ground with no snow cover, as the average precipitation intensity increases, the liquid and ice water path increase during rainfall and even more significantly during snowfall. Specifically, the liquid water path increases from 14% to 26% (Figs. 2a,d) and the ice water path increases about 23% and 37% (Figs. 2b,e) for raining and snowing events, respectively. Over dry snow cover, there is no evidence of any additional changes either in liquid or in ice water path that could cause the observed irregularity. Figure 2f shows that the water vapor path increases about 2.5 kg m^{-2} between snowfall intensities of 2 and 8 mm h^{-1} over the dry snow cover, which cannot be the main reason for the observed anomaly. The reason is that the sensitivity of the 166-GHz channel to variation of water vapor decreases significantly for snowfall intensities > 0.8 mm h^{-1} (You et al. 2017). Therefore, we speculate that the anomaly could be largely due to a surface effect.

The MODIS surface temperature data (Wan 2014) do not show any significant dependency on the rate of snowfall (Fig. 2i). Therefore, we hypothesize that the detected emission could be due to either an unknown retrieval uncertainty or, more likely, to the climatology of the snowfall and snow cover dynamics. The database shows that light but prolonged snowfall intensities (<2 mm h^{-1}) occur at latitudes above $>55°$N over dry and thicker snow cover (Fig. 2j). However, high-intensity but less-frequent snowfall is more likely to occur over lower latitudes with a thinner snow-cover climatology. In other words, the high snowfall rates mostly represent the climatology of lower latitudes with thinner depth of snow cover, less volume scattering, and thus stronger surface emission than the thicker snow cover of higher latitudes.

The above observations from Figs. 1 and 2 lead us to hypothesize that the distance between vectors of brightness temperature encodes a similarity metric that can be used to discriminate the precipitation from the background signal. A larger distance indicates larger
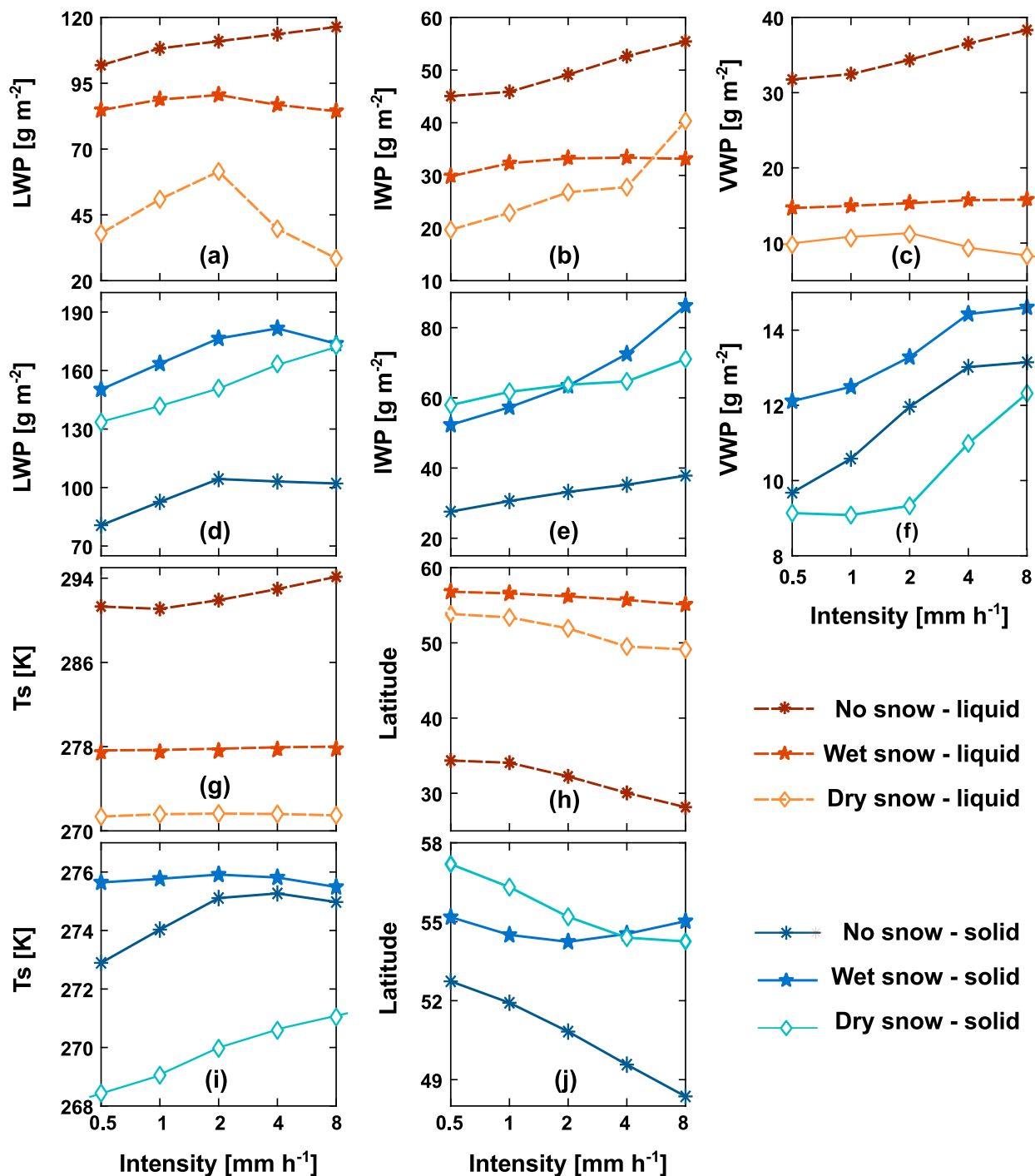
FIG. 2. Average variations of the (a),(d) cloud liquid water path (LWP); (b),(e) cloud ice water path (IWP); (c),(f) water vapor path (WVP); (g),(i) skin temperature (Ts); and (h),(j) latitudes against the precipitation intensity. The ice and liquid water paths are extracted from the MERRA-2 data (M2I1NXASM; Gelaro et al. 2017) and the surface temperature data are from MODIS (MOD11A1; Wan 2014) from June 2014 until May 2016. The intensity bins are the same as Fig. 1.

radiometric dissimilarity that could be used for improved detection of the precipitation from the background signal. Using the database, we calculate the average distance between the vectors of brightness temperatures for the

clear sky (no precipitation) and precipitating atmosphere over the three land surface types (Fig. 3). In this figure, the shaded areas in light blue (orange) represent the detected emission (scattering) signal. The key observation
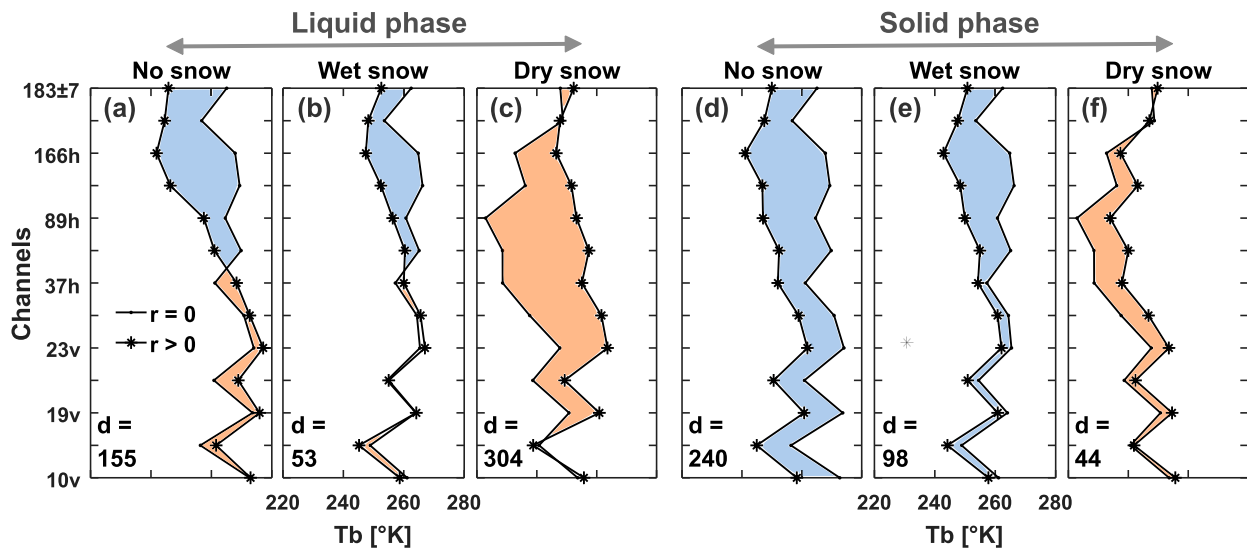
FIG. 3. Average distance between vectors of mean brightness temperatures in the database from June 2014 to May 2016 for a clear sky ($r = 0$) and a near-surface precipitating atmosphere ($r > 0$) with (a)–(c) liquid and (d)–(f) solid phase overland precipitation with no snow cover, wet snow cover, and dry snow cover. The blue and orange shaded areas indicate the cooling (scattering) and warming (emission) signals of precipitation. The mean root squared distance between the no precipitating (clear sky) and precipitating atmosphere is also presented for each land–atmosphere class.

is that when the snow-cover scattering increases, the precipitation signal transitions from a scattering to an emission regime. The wet snow cover weakens the precipitation scattering as it is less emissive than the ground with no snow cover. However, the less emissive dry snow reveals the precipitation emission signal.

For the liquid phase, we can see that the rainfall scattering over the ground with no snow cover is manifested by a large distance between the high-frequency channels ≥ 89 GHz, while the distance over lower-frequency channels is insignificant (Fig. 3a). This distance shrinks over the wet snow cover (Fig. 3b), where the dominant precipitation signal is still due to its scattering over high-frequency channels. This shrinkage is largely explained because wet snow is not a strong scatterer and thus reduces slightly the surface emission and the high-frequency scattering of rainfall. However, when the surface emission is significantly reduced over the dry snow (Fig. 3c), the emission of rainfall can be detected as a warming signal across almost all frequency channels.

For the solid phase, the distance is relatively large between the background and precipitation signals when there is no snow on the ground (Fig. 3d). This distance captures a shift across all frequency channels and a reduced polarization signal above 37 GHz. The shift is largely due to the differences between the surface temperature of clear sky versus a snowing atmosphere, while the reduced polarization is chiefly due to diffused scattering of the snowflakes. Similar to the liquid precipitation, this

distance shrinks when the ground is covered with wet snow, where the shift between the background temperatures almost vanishes as the surface temperature increases. We can see that when the snowfall is occurring over dry snow, an emission signal is observed, chiefly in response to the increased liquid and water vapor paths (see Liu and Seo 2013; You et al. 2015, 2016; Ebtehaj and Kummerow 2017). The MERRA-2 data indicate increases of $\sim 58\,\mathrm{g\,m^{-2}}$ and $4.8\,\mathrm{kg\,m^{-2}}$ in liquid and vapor water paths, respectively, when snowfall occurs. This emission signal indirectly indicates the likelihood of precipitation by increasing the brightness temperatures rather than a direct physical signature of precipitation. Because of this emission signal, the vector of snowfall brightness temperatures becomes dissimilar to the surface emission, which could lead to improved snowfall retrievals over dry snow cover—if a proper distance metric is used to quantify the dissimilarity.

4. A nested nearest neighbor algorithm for precipitation phase detection

The nearest neighbor matching has been successfully utilized for passive microwave retrieval of rainfall using the TRMM data (Ebtehaj et al. 2015, 2016) and for microwave mapping of flood inundation using the Special Sensor Microwave Imager/Sounder observations (Takbiri et al. 2017). In this section, we introduce a prognostic algorithm that relies on a nested k-nearest neighbor

matching that finds the best representation of a query brightness temperature in the database to detect precipitation occurrence and phase. The criterion for matching relies on the hypothesis that similar vectors of brightness temperatures represent similar atmospheric profiles. In other words, an observed pixel-level vector of brightness temperature for a precipitating atmosphere is more similar to a collection of precipitating brightness temperatures in the database than those that refer to a nonprecipitating atmosphere. Here, we define the similarity metric by a weighted Euclidean distance between the query vector of observed brightness temperatures and those stored in the a priori database, described in section 2.

To set the notation, hereafter, the vector of brightness temperatures is denoted by $\mathbf{Tb}$ and the ancillary data containing information on the precipitation occurrence, phase, and snow cover are represented by the vector $\mathbf{r}$. The database is pruned to contain balanced information over two different land surface types $\{\mathcal{L}\}_{s=1}^{2}$ and four independent atmospheric conditions $\{\mathcal{A}\}_{a=1}^{4}$. The land surface types are defined only based on the presence ($s = 1$) and absence ($s = 2$) of snow cover, while the atmospheric conditions refer to the clear sky ($a = 1$), liquid ($a = 2$), solid ($a = 3$), and mixed ($a = 4$) precipitating atmosphere.

To reduce the algorithmic complexity, we do not differentiate between the dry and wet snow cover in the database. Each land class consists of pairs of $\{(\mathbf{Tb}_m, \mathbf{r}_m)\}_{m=1}^{M}$, where $M = 2 \times 10^{7}$ are evenly distributed between clear and precipitating sky. Those pairs in the precipitating sky are also evenly distributed between a raining $\mathcal{A}_2$, mixed $\mathcal{A}_3$, and snowing $\mathcal{A}_4$ atmosphere. It is important to note that the pairs are randomly selected from their parent class to avoid any bias toward a specific land or atmospheric class.

In summary, for a given land surface type and a query vector of brightness temperatures $\mathbf{y}$, the algorithm relies on a 3-tuple $\{(k_a, \mathbf{W}_a, p_a)\}_{a=1}^{3}$, where k_a is the number of nearest neighbors, $\mathbf{W}_a$ is the atmospheric weight matrix over each land surface type used in the weighted Euclidean distance $d_m = (\mathbf{y} - \mathbf{Tb}_m)^{\mathrm{T}} \mathbf{W}_a (\mathbf{y} - \mathbf{Tb}_m)$, and p_a denotes a detection probability measure. The weight matrix accounts for the relative importance of the channel combinations for detection of precipitation and its phases (Ebtehaj and Kummerow 2017). Specifically, given the land surface types $\mathcal{L}_s$, after finding the k-nearest neighbors $\{(\mathbf{Tb}_i, \mathbf{r}_i)\}_{i=1}^{k}$ of each query vector $\mathbf{y}$, a nested decision-making process is made to detect precipitation and its phase based on the probability measure p_a.

In the first step, the algorithm uses $(k_1, \mathbf{W}_1, p_1)$ to search for the k_1-nearest neighbors of $\{\mathbf{Tb}_i\}_{i=1}^{k_1}$ and the

corresponding ancillary information in the database. Then, a binary decision is made to label the vector $\mathbf{y}$ as a precipitating observation, when the number of precipitating neighbors n_p is greater than $p_1 \times k_1$. For precipitating $\mathbf{y}$, the algorithm identifies the precipitation phase by running a new k-nearest neighbor search using $(k_2, \mathbf{W}_2, p_2)$ through those precipitating neighbors $\{(\mathbf{Tb}_j, \mathbf{r}_j)\}_{j=1}^{n_p}$ that are found in the first step, where $k_2 < p_1 \times k_1$. Then, as explained before, a binary decision is made to label $\mathbf{y}$ as liquid precipitation, if the number of raining neighbors $n_l = \max(n_l, n_s, n_m)$ is greater than $p_2 \times k_2$, where n_m and n_s are the number of mixed and solid precipitation elements among the k_2-nearest brightness temperatures $\{\mathbf{Tb}_i\}_{i=1}^{k_2}$. If those conditions are not satisfied, the algorithm continues similarly to find if the phase of $\mathbf{y}$ is solid or mixed. An algorithmic flowchart is presented in Fig. 4.

To determine the optimal values of the input parameters k_a and p_a, we compute the receiver operating characteristic (ROC) curves (Fig. 5), which characterize the trade-off between the hit and false alarm rates. The probability of hit is defined as the fraction of occurred events that were correctly detected, while the false alarm rate is a fraction of events that did not occur but were incorrectly detected by the algorithm. Let a represent the number of correctly detected events, c the number of missed events, b the number of false detection, and d the number of correct rejection. Then, the probabilities of hit and false alarm are defined as $a/(a + c)$ and $b/(b + d)$, respectively. The optimal value of k_a is chosen based on the maximum area under the ROC curves (Hanley and McNeil 1982), while the best detection probability p_a is chosen where the curvature of the ROC is maximum.

5. Results and validation

To test the performance of the proposed approach, the terrestrial precipitation and its phase are retrieved over the inner swath of the GMI overpasses from June 2015 to May 2016. As the phase outputs of the algorithm are discrete values for solid (0), mixed (0.5), and liquid (1), the temporal mean values associated with these phases could reveal the overall sensitivity of the algorithm to the seasonal variations of surface temperature and emissivity. To that end, the phase indices are averaged at orbital levels over the summer and winter for the nested KNN algorithm and the standard active and passive GPM products (Fig. 6). To quantify statistical agreements between the results of the algorithm and those of the REF maps, we calculate the annual probability of detection, false alarm, and the Heidke skill score (Doswell et al. 1990) for the presented results
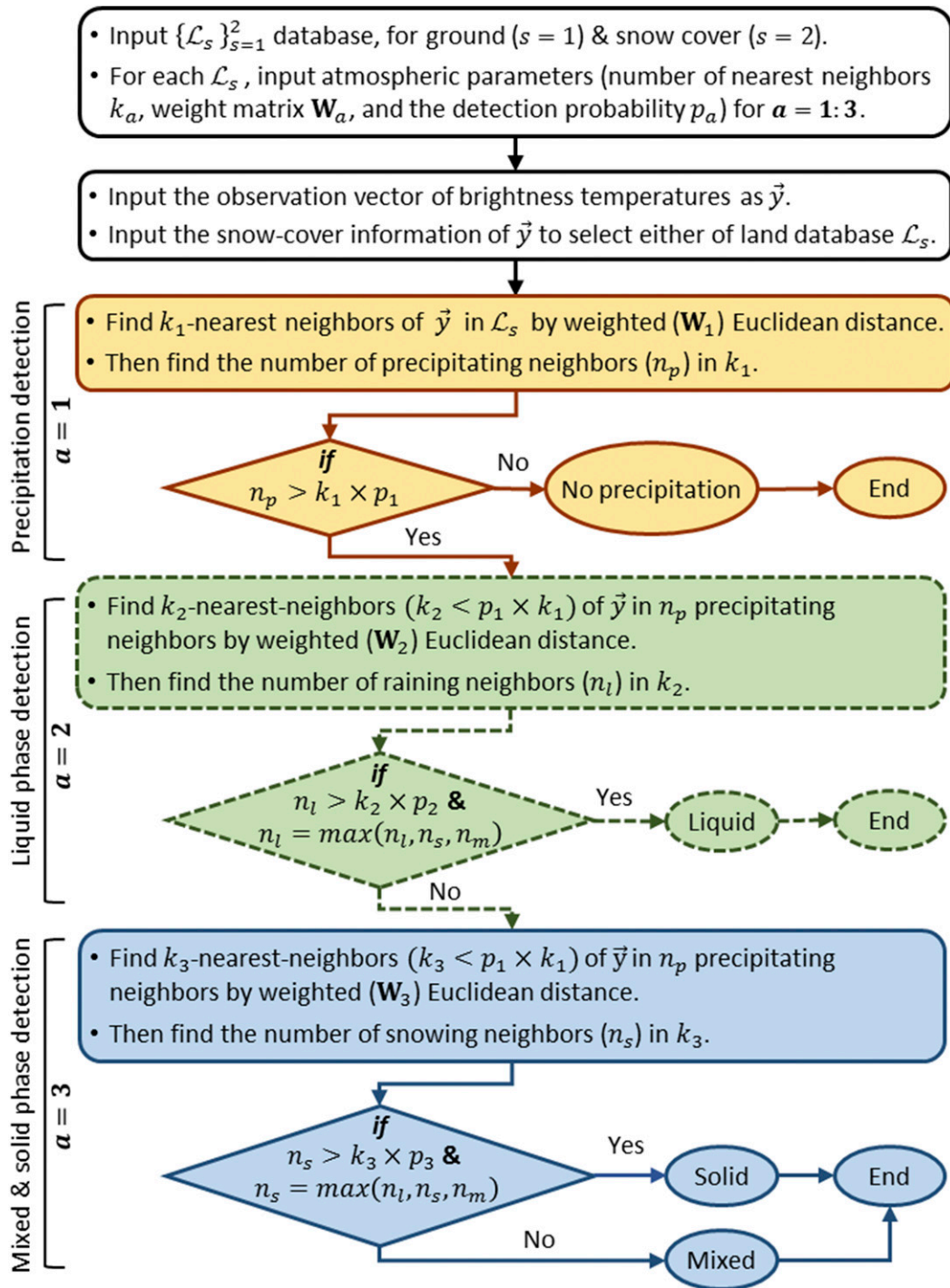
FIG. 4. Algorithm flowchart of the proposed weighted KNN algorithm for detection of precipitation occurrence and phase.

in Figs. 7–9. We also compare the algorithm outputs with the precipitation phase products of the MRMS on a seasonal basis (Figs. 10, 11). Finally, some results are presented at a storm scale to demonstrate the detection capabilities of the algorithm for a few precipitation events that are coincidentally captured by the DPR and high-resolution ground-based radars (Figs. 12, 13) and simulated by the Weather Research and Forecasting (WRF) Model (Fig. 14) during OLYMPEX (Houze et al. 2017).
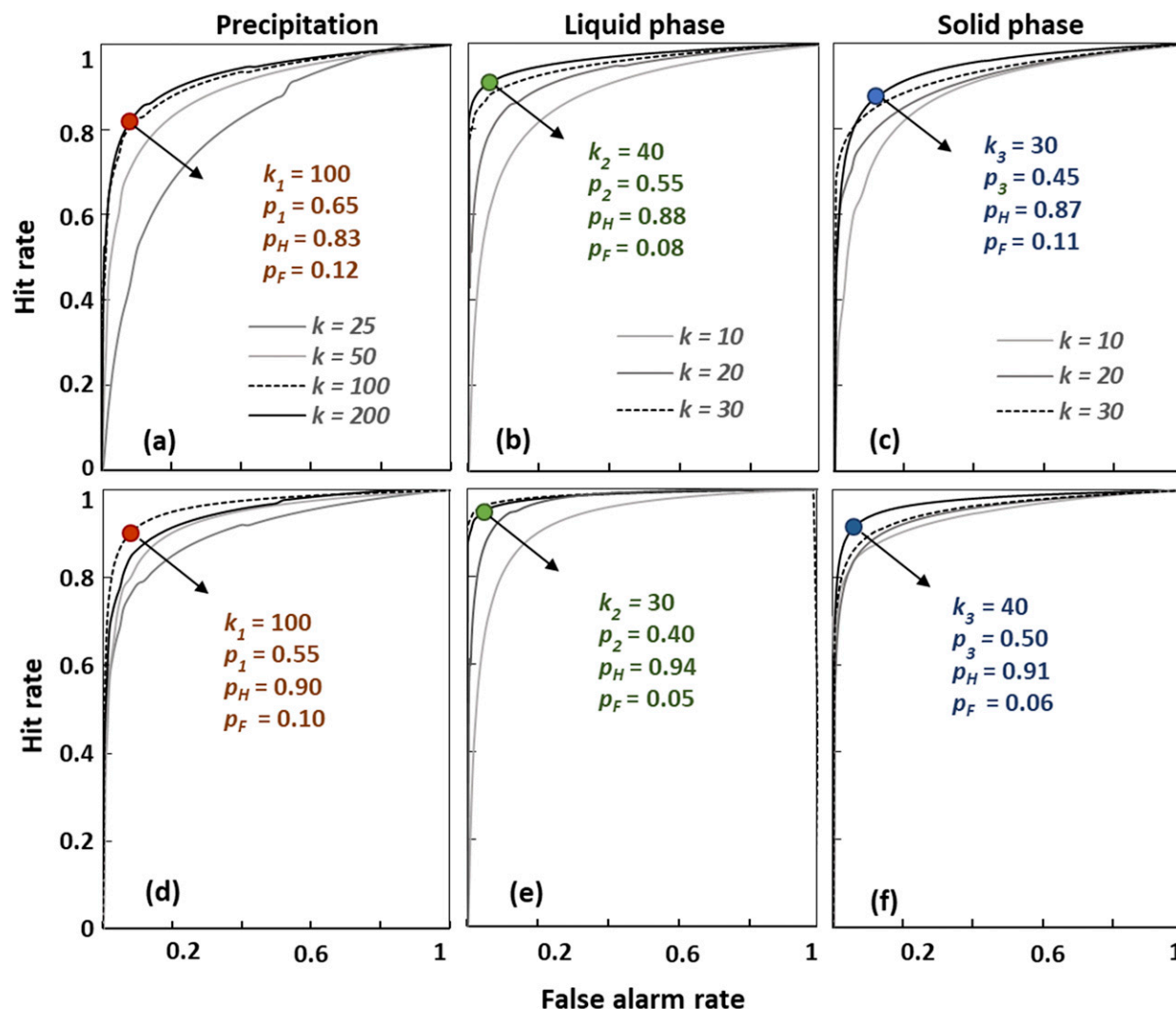
FIG. 5. Trade-off curves between the probability of hit p_H and false alarm p_F calculated with different numbers of k-nearest neighbors for detection of the precipitation occurrence and phase over (a)–(c) non-snow-covered surfaces and (d)–(f) snow-covered surfaces. The optimal values of k and the detection probabilities p are given in each subplot.

a. Global retrievals

The seasonal average of the quasi-global maps of precipitation phase are presented in Fig. 6, for the inner-swath data products by the DPR, GPROF, KNN, and REF. The results are shown as a probability continuum of phase transition from the liquid (0) to solid (1), at the grid resolution 0.1°. These results are mapped where the precipitation is detected only by the DPR for two seasons. The seasons are defined as summer (June–October 2015 and May 2016) and winter (November 2015–April 2016) of the Northern Hemisphere.

Overall, since the phase of the passive product is dictated by the reanalysis data, the results mostly follow the climatology patterns of near-surface wet-bulb temperature and are smoother than those of the active

product (Figs. 6a–d). The smoothness of the GPROF retrievals could also be due to its ability in retrieving the light precipitation with intensities below the minimum detectable rate by the DPR ($<0.2\,\mathrm{mm\,h}^{-1}$), as the GPROF also uses precipitation data from MRMS ground-based radar in its a priori database to increase the retrieval sensitivity to snowfall. Comparison of the official passive and active products remains outside the scope of this research; however, there seem to be notable differences in the spatial patterns of precipitation phases in these products. The difference in the source of ancillary data could be a reason for the observed discrepancies, which largely exist over mountainous terrains such as the Andes, Tibetan highlands, Rockies, Scandinavian Mountains, Alps, and Zagros
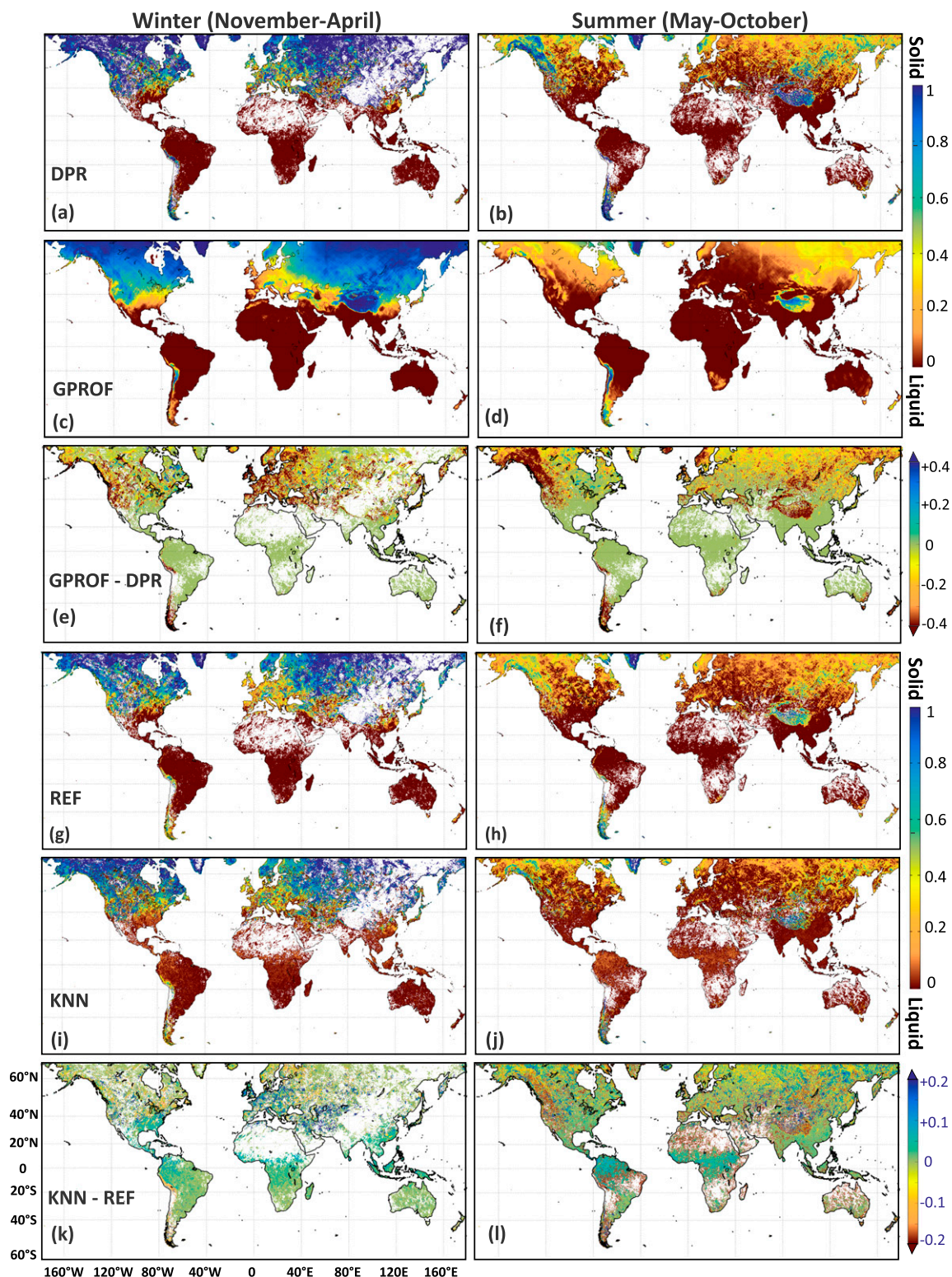
FIG. 6. Seasonal probability of the precipitation phase change. The average phase of the (a),(b) DPR; (c),(d) GPROF; (g),(h) REF (merged); and (i),(j) KNN algorithm, as well as the differences between (e),(f) the DPR and GPROF products and (k),(l) the REF and KNN products. The differences are shown where both products detect the precipitation occurrence.
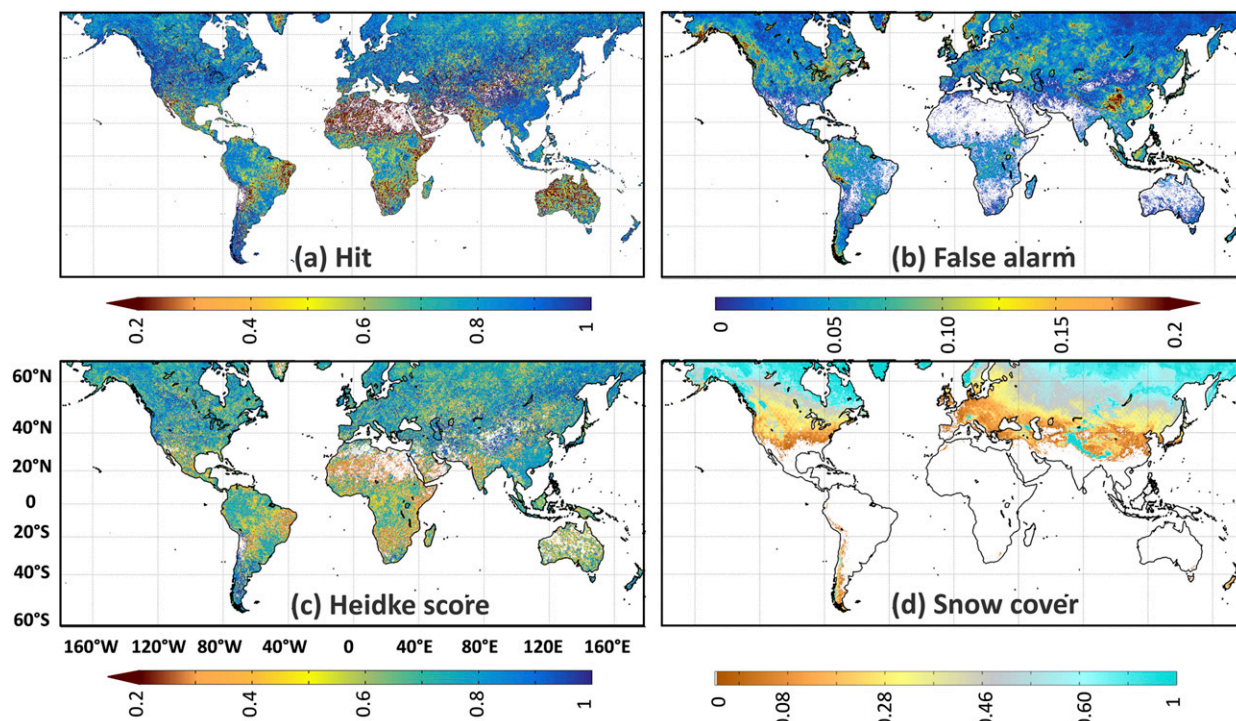
FIG. 7. The mean annual map of the (a) probability of hit, (b) probability of false alarm, and (c) HSS obtained by comparing the pixel-level results of the KNN algorithm with the REF product for the detection of precipitation occurrence. (d) The map of snow-cover fraction is also obtained from the MODIS data (MOD10A1; Hall et al. 2002) coincident with GPM inner-swath overpasses from June 2015 to May 2016.

Mountains (Figs. 6e,f)—where precipitation is mostly triggered by topographic features.

The observed differences are not surprising because of complications in both active and passive retrievals due to reduced ice scattering in shallow orographic lifting, heterogeneity of surface roughness, and radiometric complexity of high-elevation snow and ice cover (Tian and Peters-Lidard 2010). The phase discrepancies also seem to be larger when it comes to identifying precipitation phase in the summer. For example, over the Tibetan highlands, the active products classify most of the summer precipitation as snowfall while the passive product results in more liquid precipitation, especially over the Hengduan Mountains in southeast China.

Figures 6i and 6j show the results of the KNN algorithm in summer and winter and compare them with the REF map (Figs. 6k,l). Overall, we observe a good agreement between the KNN outputs and the REF target precipitation product. The differences are more pronounced in the summer than the winter and mostly accumulated over the mountainous and dense vegetation regions (Figs. 6k,l). For example, we observe that, in the summer, the detection probability of solid and mixed phases are negatively biased ($\sim-12\%$) over the Rockies and the Andes. However, in winter, this probability is positively biased over small parts of the Scandinavian Mountains in northern Europe ($\sim+15\%$). Some of these mountainous biases are mainly attributed to the false detection of precipitation occurrence rather than its phase (Fig. 7b). Additionally, over the tropical forests, the algorithm falsely detects some mixed precipitation phases. Over dense vegetative surfaces the microwave polarization signal becomes very weak (Prigent et al. 1997) due to incoherent vegetation scattering. The lack of a pronounced polarization signal could be the main reason for the reduced discriminatory power of the KNN approach that relies on the Euclidean distance as a similarity metric.

Visual inspection of the global maps shows a good spatial and seasonal agreement between the KNN and REF. The proximity of these two products at the global scale is quantified by three measures including the Spearman's correlation ρ, the root-mean-square error (RMSE), and the Kullback–Leibler (KL) divergence in Table 1. The KL divergence $\mathrm{KL}\|(P\|Q) = \sum_{i=1}^{n} P(i)/Q(i)$ is a nonsymmetric and nonnegative measure that captures the proximity of two probability distributions P and Q and is zero when they are identical. To compute the KL divergence, between the probability histograms of the REF (P) and KNN outputs (Q), we discretize P and Q
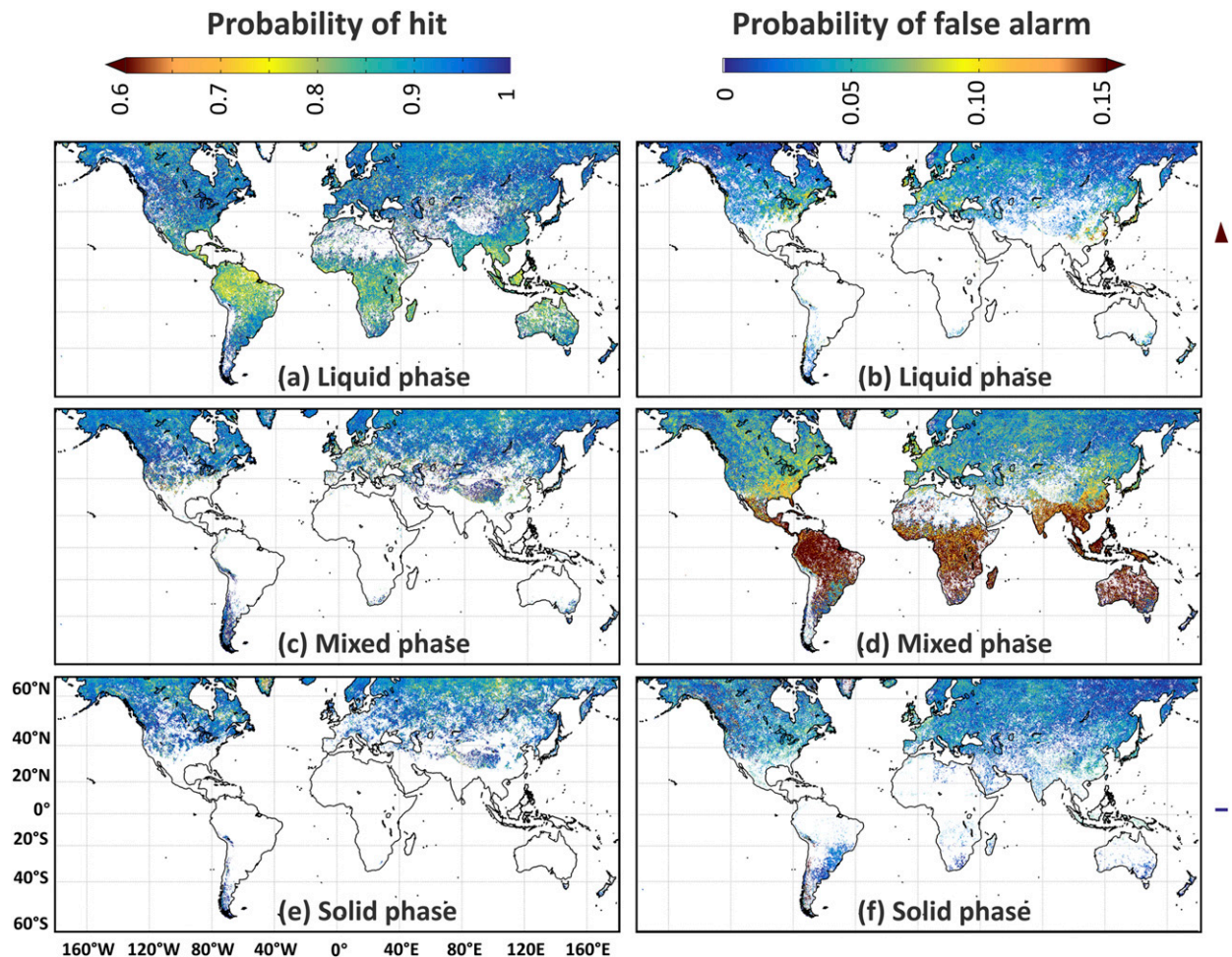
FIG. 8. The mean annual map of the probability of hit and false alarm by the KNN algorithm for the detection of the (a),(b) liquid phase; (c),(d) mixed phase; and (e),(f) solid phase. The results are obtained for all GPM inner-swath overpasses from June 2015 to May 2016.

with $n = 20$ probability intervals. The RMSE and KL values are normalized between 0 and 1 for interpretation convenience. As is evident, the correlation between the KNN and REF products is around 0.89–0.91 in winter and summer, indicating that the algorithm is not excessively sensitive to the seasonal changes in land surface radiometric properties. The normalized RMSE also remains below 14% in both seasons. We see that the KL values slightly increase from winter (0.06) to summer (0.10), which indicates that, on average, the KNN method may exhibit improved detection skills when the extent of the global snow cover is larger in winter than summer.

To further reveal the error structure of the instantaneous pixel-level retrievals, we used three statistical measures including the probability of hit, probability of false alarm, and the Heidke skill score HSS $= 2(ad - bc)/[(a + c)(c + d) + (a + b)(b + d)]$ (Doswell et al. 1990), which ranges from a no skill ($-\infty$) to a perfect skill (1). Recall that a is the number of correctly detected

events, c is the number of missed events, b is the number of false detection, and d is the number of correct rejection. To have an adequate number of samples, these quality measures are calculated using the entire validation period from June 2015 to May 2016 (Table 2, Figs. 7, 8).

The annual maps of the probability of hit, false alarm, and HSS score are used to evaluate the detection skill of the KNN approach against the DPR as a reference (Fig. 7). The probability of hit over the snow-covered regions is relatively high. The reason is that the presence of snow on the ground reduces the surface emission, which could lead to better detection of the precipitation emission signal (Fig. 3)—similar to radiometrically cold ocean surfaces. The low detection rates are mostly over the areas where the DPR has a low sampling rate. Thus, lack of skills in these regions could be partly due to lack of samples in the database. A high probability of false alarm (~ 0.2) is seen over some mountainous regions such as the Tibetan highlands and the western Rockies.
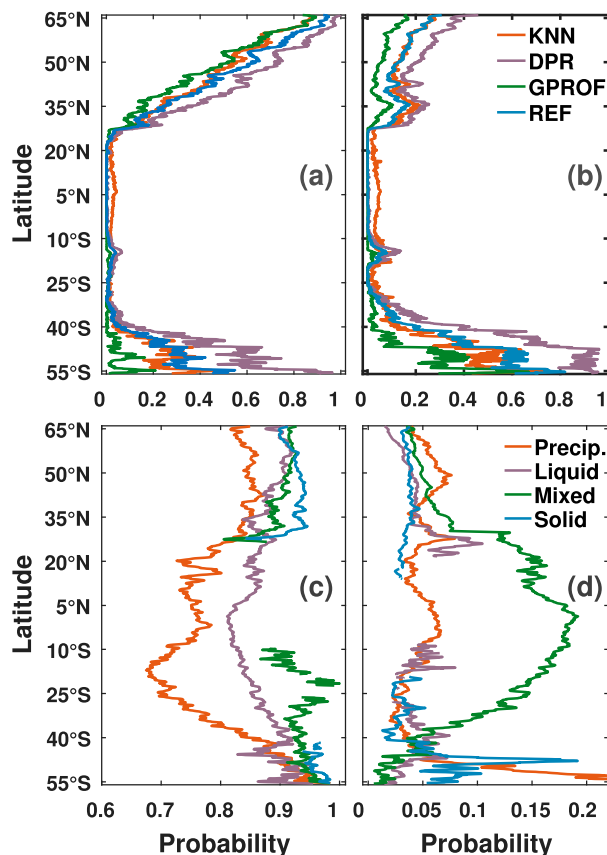
FIG. 9. Zonal mean values of the probability of precipitation phase change from liquid ($p = 0$) to solid ($p = 1$) by the KNN, DPR, GPROF, and REF products in (a) winter (November–April) and (b) summer (May–October). Zonal mean values of (c) probability of hit and (d) false alarm for the detection of the precipitation occurrence and its phase change by comparing the KNN results with the REF product.

The false detection, mostly in liquid phase, gives rise to negative biases in detecting frozen and mixed precipitation (Fig. 6l). High (~0.80), medium (~0.66), and low values (~0.50) of HSS are observed over the snow cover, tropical forests, and undersampled deserts such as the Sahara, respectively (Fig. 7c).

The conditional probability of hit and false alarm are calculated for liquid, mixed, and solid phases (Fig. 8), with respect to the REF product. For separating the errors of the precipitation and phase detection, the probabilities are obtained assuming that the precipitation is correctly detected by the KNN algorithm. Similar to the precipitation detection, the algorithm displays improved phase detection capabilities over snow-covered surfaces (Fig. 8). The probability of hit for the liquid, mixed, and the solid phase is mostly greater than 0.85 and reaches 0.95 over the high altitudes of North America. However, we observe a relatively lower detection rate of around 0.74 for liquid precipitation over the tropical and

subtropical regions such as the rain forest of Amazonian and central Africa. The results show that the low probability of detection for the liquid phase is mostly because the algorithm detects some false mixed phase precipitation (Fig. 8d). We speculate that this error could be partly attributed to the reduced skill of the algorithm over vegetated surfaces. The reduced detection skill of the algorithm could also be partly due to warm rain occurrences over the heterogeneous land surface of tropical and subtropical regions where cloud ice scattering is not significant.

To understand the reasons for improved retrievals over snow-covered surfaces, the averaged values of the probability of hit and false alarm are stratified based on precipitation occurrence $\mathcal{D}_1$ at liquid $\mathcal{D}_2$, mixed $\mathcal{D}_3$, and the solid $\mathcal{D}_4$ phase over different land surface types $\{\mathcal{L}\}_{s=1}^3$, where $s = 1$–3 denotes the ground, wet, and dry snow cover (Table 2). The probability of precipitation detection increases by almost 11% from the ground to the dry snow cover, and 3% from wet to dry snow. An increase of 8%–11% is also observed in the probability of hit in detection of solid and liquid phase over dry snow, where the largest detection rate of 94% is obtained for the snowfall. The results show that the probability of false alarm also increases in detection of precipitation occurrence over snow cover, whereas it decreases when it comes to the detection of its phase. Because, once precipitation is detected, due to significant differences between the signatures of rain and snowfall, the probability of false alarm is markedly reduced. Table 2 quantifies the dependency of the probability of hit and false alarm on the annual percentage of the dry snow cover. For precipitation detection, the probability of hit increases by about 10% as the annual percentage of dry snow increases from zero to more than 70%, while the probability of false alarm increases between 2% and 4%. As is evident, for precipitation phase detection, both probabilities show improvements of around 4%.

b. Comparison with the ground-based radar

1) COMPARISON WITH MRMS SYSTEM

To further evaluate the performance of the KNN algorithm, we compare its outputs against a precipitation product derived from the MRMS system (Zhang et al. 2011, 2016). MRMS mosaics three-dimensional volume scan observations from 146 S-band dual-polarization Weather Surveillance Radar-1988 Doppler (WSR-88D) and 31 C-band single polarization Canadian radars. The product optimally integrates the radar observations with simulations of atmospheric models as well as hourly gauge data to produce seamless precipitation rate and
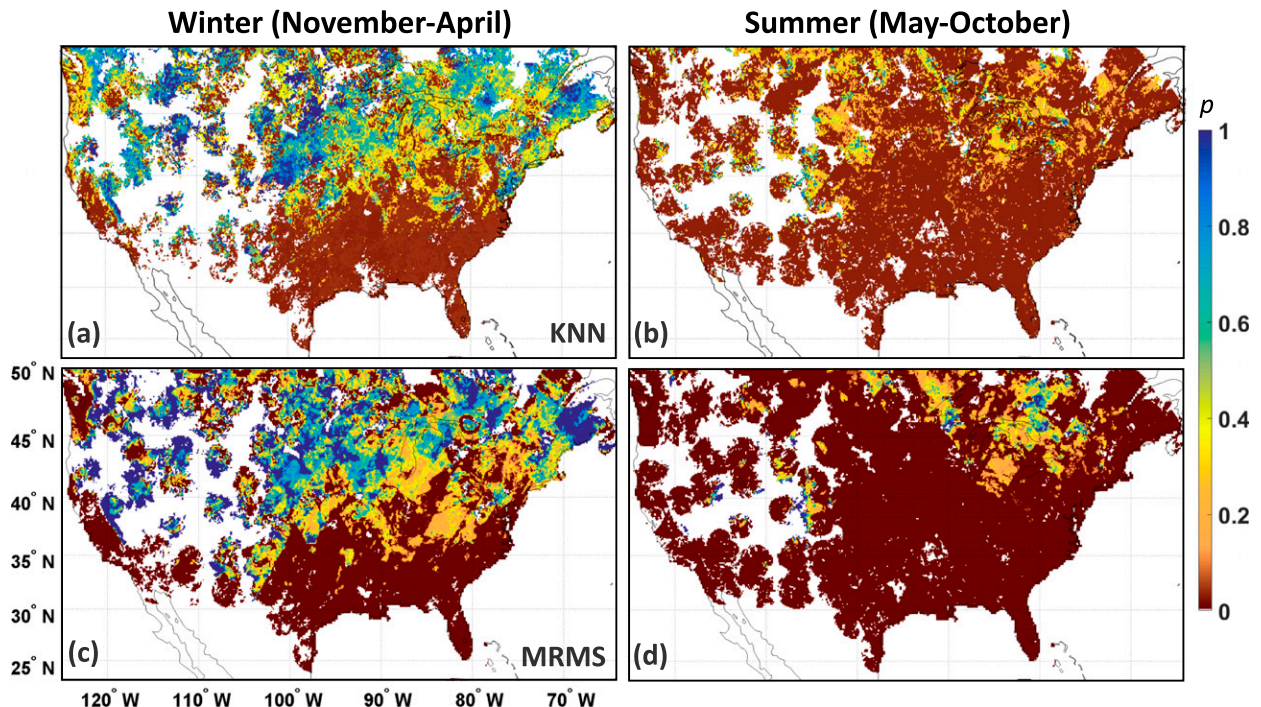
FIG. 10. Mean seasonal maps of the probability of precipitation phase change from liquid ($p = 0$) to solid ($p = 1$) for KNN in (a) winter and (b) summer, and for the MRMS in (c) winter and (d) summer, from June 2015 to May 2016.

phase estimates over the CONUS, at spatial resolution of 1 km at every 2 min. The MRMS products are further quality controlled and gauge adjusted at finescale following the procedure described in Kirstetter et al. (2012) to derive a consistent and high-quality surface precipitation.

To determine the precipitation phase, MRMS uses thresholds on the wet- and dry-bulb temperatures. Specifically, the precipitation is labeled as snowfall when the radar reflectivity exceeds 5 dBZ, the surface temperature is below 2°C, and the surface wet-bulb temperature is below 0°C (Zhang et al. 2016). Thus, the MRMS rain–snow delineation is subject to similar uncertainties as in the passive GPM data (Chen et al. 2016). However, the uncertainties in detecting precipitation are significantly lower than the satellite data because of the higher sensitivity and resolution of the ground-based radar observations, especially over landscapes with no significant orographic features (Kirstetter et al. 2012). To compare with the outputs of the KNN algorithm, a reference surface precipitation is derived by mapping the high-resolution MRMS data onto, and then averaging over, the nearest DPR grids (see Kirstetter et al. 2012, 2014).

Figure 10 shows that the spatial variations of the probability of phase change in the KNN and MRMS are consistent in the winter and summer seasons. The calculated values of KL divergence between KNN and

MRMS are 0.27 and 0.15 in winter and summer, respectively. The values of other calculated similarity metrics (i.e., ρ and RMSE) are also deteriorated from summer to winter (Table 3). These results indicate that even though the KNN shows improved wintertime detection of precipitation compared to those in summertime when
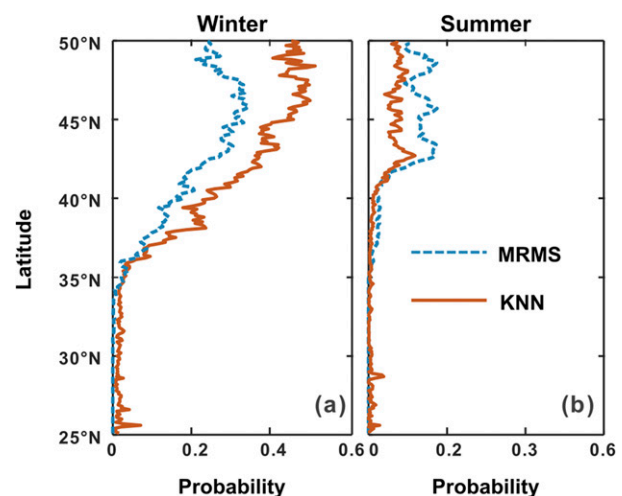

FIG. 11. The zonal mean of the probability of precipitation phase change from liquid ($p = 0$) to solid ($p = 1$) by the KNN and MRMS products in (a) winter (November–April) and (b) summer (May–October), from June 2015 to May 2016.
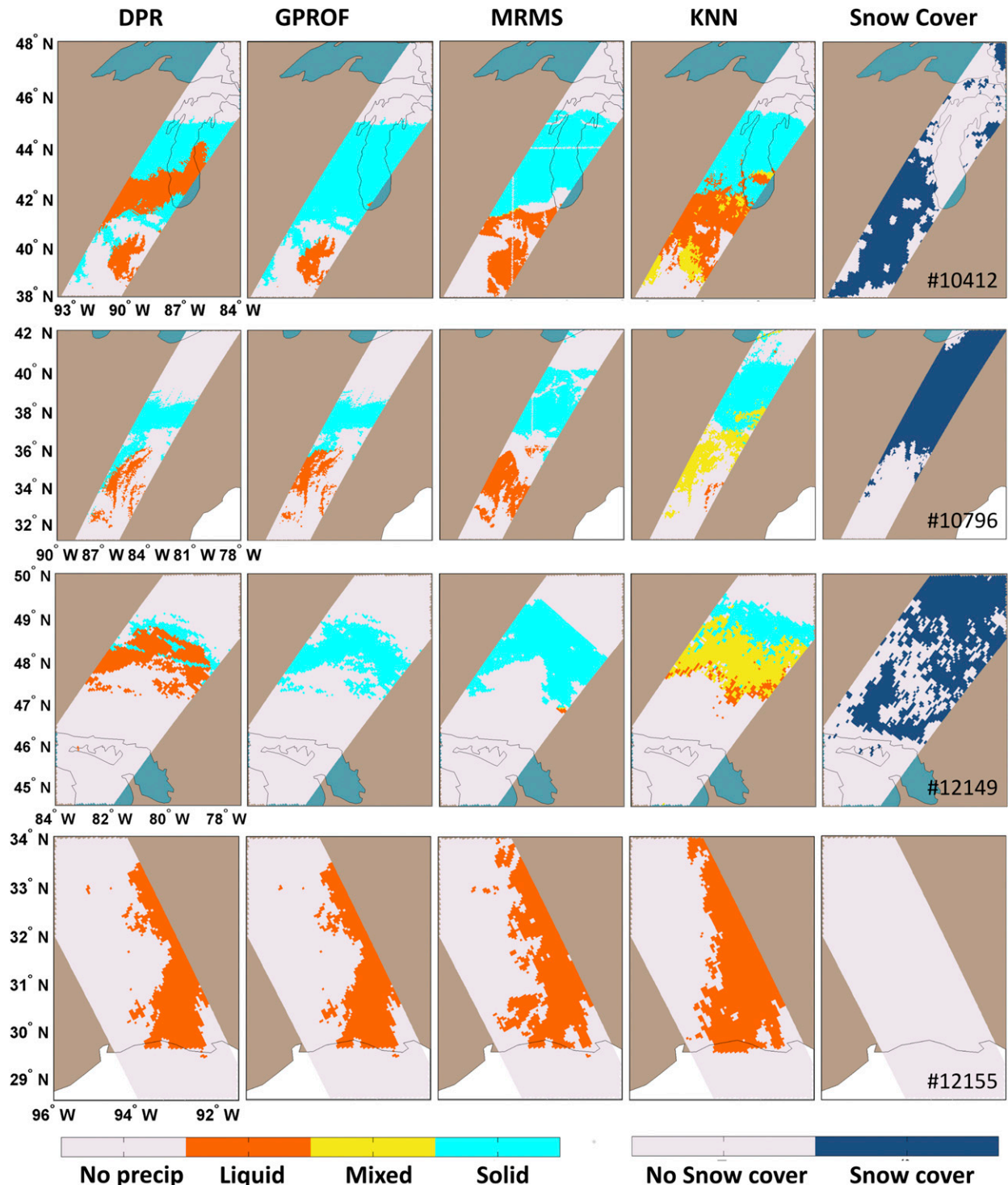
FIG. 12. Orbital-level precipitation phase detection from the KNN, DPR, GPROF, and MRMS for a few GPM overpasses including (top) 10412 on 28 Dec 2015, (top middle) 10796 on 22 Jan 2016, (bottom middle) 12149 on 18 Apr 2016, and (bottom) 12155 on 18 Apr 2016.
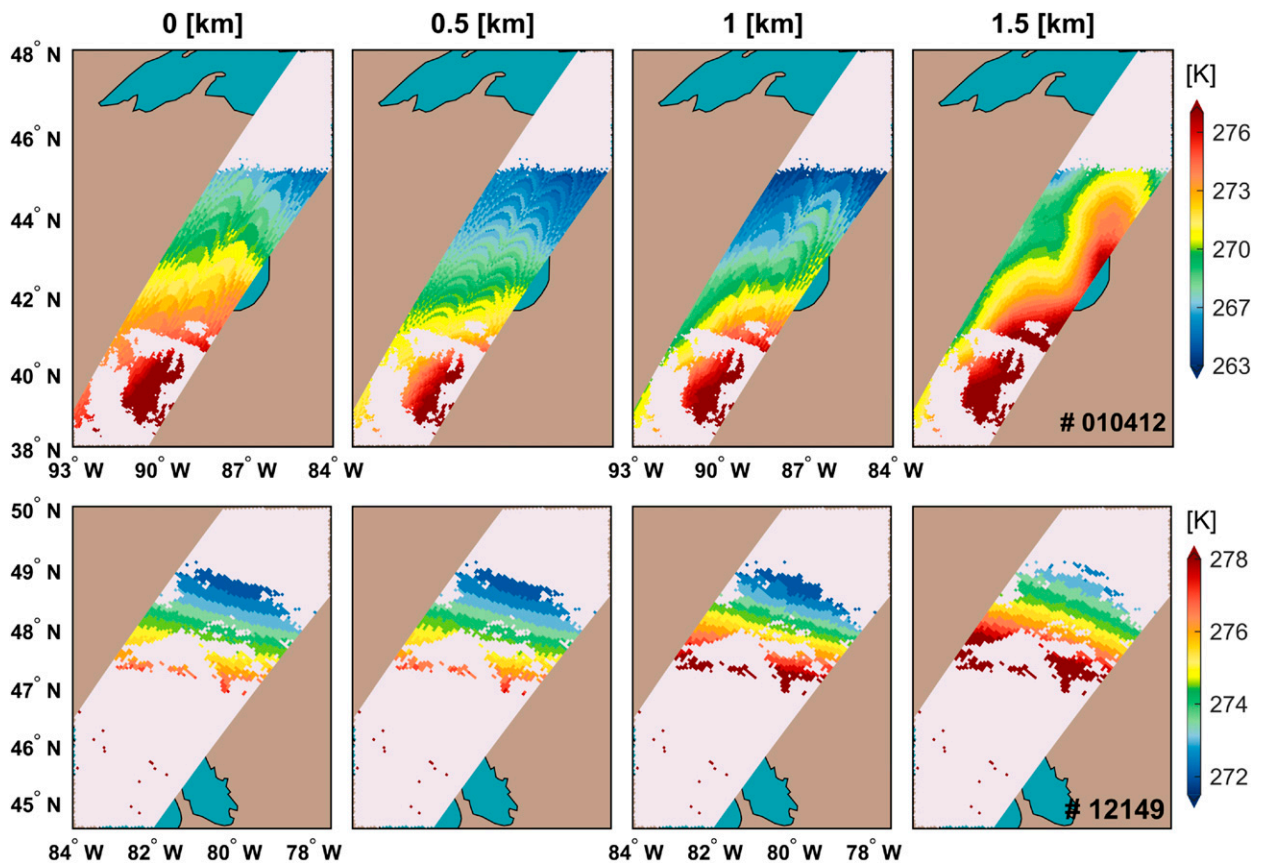
FIG. 13. Inversion of the air temperature at (top) orbit 10412 on 28 Dec 2015 and (bottom) orbit 12149 on 18 Apr 2016. The data (2A-DPRENV) are presented at four ranges from 0 (surface) to 1.5 km.

compared with the REF product (Table 3), the intrinsic error between the satellite and ground-based data is still much larger than the satellite retrieval error, especially in the winter. The zonal mean of the phase transition probabilities (Fig. 11) indicates more similarities at lower latitudes (<40°N), where the uncertainty of precipitation phase change is lower or remains close to zero. At higher latitudes, KNN generates a higher (lower) probability of snowfall occurrence relative to the MRMS in winter (summer). In particular, larger departures occur at latitudes higher than 37°N in winter and 43°N in summer, where the ground is usually covered with snow.

Figure 12 shows four different satellite overpasses that capture large storms with distinguishable spatial phase change. Overall, the KNN approach is skillful in capturing the occurrence and phase of the near-surface precipitation. As is evident, in case of a single-phase precipitation event (e.g., orbit 12155), the KNN can accurately detect the extent of the storm, especially when a large part of the storm is in liquid form. However, when several phases coexist within the storm (e.g., orbits 10412, 10796, and 12149), discrepancies arise between

the satellite active/passive products and the MRMS data. The produced mixed phase by the KNN retrieval reflects the uncertainty between the satellite active/passive retrievals where a freezing point is likely to occur in the DPR ground clutter zone. For example, the storm on the northern shores of Lake Huron (orbit 12149) is well detected in terms of its spatial extent. The phase detection in the GPM passive product (GPROF) and the MRMS products is consistent since both products rely significantly on the wet-bulb temperature data. However, the DPR product differs significantly from other products and produces more liquid phase over the southern edge of the storm. As is evident, the KNN retrievals capture this discrepancy through a mixed phase detection.

It is surprising that in orbits 10412 and 12149 (Fig. 12), the DPR reports the phase as liquid where the GPROF classifies the phase largely as solid since the discrepancy is often in the other direction. Based on the atmospheric temperature profile derived from environmental ancillary data (2A-DPRENV) used in the active retrieval algorithm, we conclude that the there is a
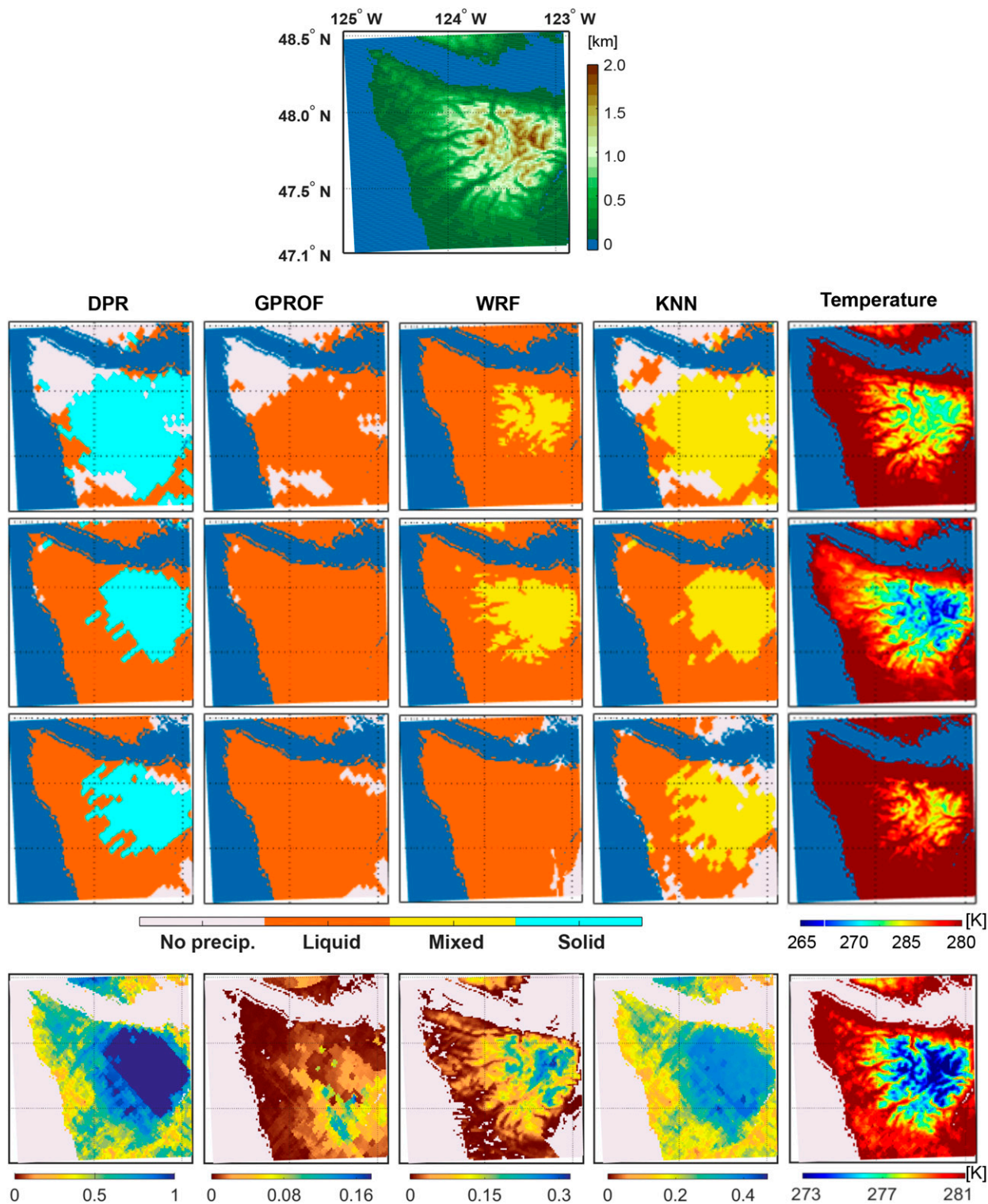
FIG. 14. (top) The digital elevation map (DEM) of the Olympic Mountains and (middle) the precipitation phase by the DPR, GPROF, WRF, and KNN for orbits 9722 (14 Nov 2015, first row), 9773 (17 Nov 2015, second row), and 10019 (3 Dec 2015, third row). (bottom) Average probability of phase for 117 GPM inner-swath overpasses from 1 Nov to 23 Dec 2015. The last column shows the 2-m air temperature from the WRF simulations.

TABLE 1. Quality metrics obtained by comparing the annual probability of phase transition between the KNN results and the reference product (REF). Shown statistics are the normalized root-mean-square difference (RMSD), Spearman's correlation ρ, and the KL divergence.

Metrics	ρ	RMSD	KL ($\Delta p = 0.05$)
Winter (November–April)	0.91	0.12	0.06
Summer (May–October)	0.89	0.14	0.10

temperature inversion when the storm is happening (see Fig. 13). In this case, liquid precipitation can refreeze near the surface and may not be captured by the DPR.

2) COMPARISON WITH THE WRF SIMULATIONS DURING OLYMPEX

The MRMS data lack coverage over mountainous regions, thus we need a venue with rich ground-based observations for further evaluation of the presented approach. There is a wealth of orographic precipitation data during the GPM OLYMPEX (Houze et al. 2017) from 1 November to 23 December 2015. The Olympic Mountains are located in the northwestern corner of Washington State, United States (Fig. 14), with a dominant orographic precipitation regime. This regime is a result of the abrupt uplift of moisture-laden southwest airflow coming from the midlatitude baroclinic storm systems. A few high-elevation snow and precipitation gauges were used during the OLYMPEX field campaign. However, the coarse temporal resolution of the DPR (i.e., 117 partial overpasses), relative to the 56-day duration of OLYMPEX, hamper their use for our purpose. Therefore, we choose the outputs of a high-resolution (1.33 km) hourly WRF simulation by the Northwest

Modeling Consortium over the Olympic Mountains (Mass et al. 2003).

Currier et al. (2017) used the microphysical scheme of WRF to estimate precipitation phase and showed that the results are relatively unbiased when compared with the OLYMPEX ground-based observations. The data are available from November 2015 to May 2016 and contain almost 117 full or partial overlaps with DPR overpasses. First, the DPR retrievals are spatially resampled to match the 1.33-km WRF outputs. Then, the hourly outputs of the WRF are interpolated to match the scanning time of the DPR. To convert the interpolated WRF outputs to discrete precipitation phase, we follow a simple rule. If the ratio of reported snowfall to rainfall intensity is higher (lower) than 0.66 (0.33), then the precipitation is considered as solid (liquid) phase; otherwise, it is labeled as mixed.

Figure 14 illustrates the precipitation phase for the DPR, GPROF, KNN, and the WRF for three GPM orbits (9722, 9773, 10019). We observe that at high-elevation regions, the KNN detects mixed phase over the areas that exhibit phase discrepancies between the GPROF and DPR. We see that these KNN results are in a good agreement with the WRF simulations. However, it is important to note that the precipitation phase partitioning in the WRF outputs is based on cloud microphysical parameters in the atmospheric boundary layer, and thus its mixed-phase precipitation is physically different than the defined mixed-phase category in KNN retrievals.

We calculate and compare the average phase outputs of the DPR, GPROF, KNN, and WRF data for all 117 coincident DPR overpasses. We found that compared with the average phase probability of WRF, the KNN precipitation phase is positively biased by about 28%

TABLE 2. The annual probability of hit and false alarm for the KNN results over different land surface types $\{\mathcal{L}\}_{s=1}^3$ and detection classes $\{\mathcal{D}\}_{i=1}^4$. Here, $s = 1$–3 represent the ground and wet and dry snow-covered surfaces; $i = 1$ denotes the detection of precipitation occurrence; and $i = 2$–4 represent the detection of liquid, mixed, and solid phase, respectively. The results over the dry snow cover $\mathcal{L}_3$ are further stratified based on the annual percentage of the snow.

	Land surface			Annual percentage of dry snow cover $\mathcal{L}_3$				
	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	0–0.10	0.10–0.25	0.25–0.45	0.45–0.70	0.70–1.00
				Probability of hit				
$\mathcal{D}_1$	0.75	0.78	0.86	0.77	0.84	0.86	0.84	0.87
$\mathcal{D}_2$	0.82	0.83	0.90	0.85	0.89	0.90	0.91	0.92
$\mathcal{D}_3$	0.86	0.89	0.92	0.88	0.90	0.93	0.92	0.92
$\mathcal{D}_4$	0.86	0.86	0.94	0.88	0.93	0.94	0.96	0.95
				Probability of false alarm				
$\mathcal{D}_1$	0.06	0.09	0.11	0.09	0.08	0.13	0.09	0.11
$\mathcal{D}_2$	0.10	0.05	0.04	0.07	0.05	0.06	0.04	0.04
$\mathcal{D}_3$	0.09	0.08	0.05	0.08	0.05	0.07	0.06	0.05
$\mathcal{D}_4$	0.08	0.05	0.04	0.08	0.05	0.05	0.04	0.04

TABLE 3. Quality metrics obtained by comparing the annual probability of phase transition between the KNN retrievals and MRMS observations. Shown statistics are the normalized RMSD, Spearman's correlation ρ, and the KL divergence.

Metrics	ρ	RMSD	KL($\Delta p = 0.05$)
Winter (November–April)	0.72	0.29	0.27
Summer (May–October)	0.78	0.21	0.15

(i.e., KNN captures more solid phase than WRF; Fig. 14). However, this bias is about 31% at elevations above 800 m, while reduced to about 24% for lower elevations. Additionally, the results show that over areas with elevations higher than 800 m, the KNN phase bias is significantly smaller compared to both DPR (positive bias ∼48%) and GPROF (negative bias ∼56%). At elevations below 800 m, the KNN is less biased than the positively biased DPR (∼41%), but about 9% more biased than GPROF with a negative bias of ∼19%. Overall, these results indicate that even though the KNN phase detection is consistent with the satellite products, there are notable discrepancies with the WRF simulations over the mountainous regions, which need further investigation.

6. Summary and discussion

We proposed a Bayesian algorithm for detection of precipitation occurrence and phase from satellite observations, with particular emphasis on snowfall detection over snow cover. The algorithm relies on a nested k-nearest neighbor (KNN) search and probabilistic vote rules for detection of precipitation occurrence and its phase. The a priori database in the algorithm contains collocated GMI brightness temperatures (10.65–183 GHz) and DPR precipitation data that were stratified based on snow-cover retrievals from the MODIS sensor on board the *Terra* satellite. The precipitation phase data from the GPM passive and active products were combined to provide a reference database for testing the skill of the algorithm.

The results demonstrated that the weighted Euclidean distance can be used as a similarity metric for precipitation phase detection in a Bayesian setting, with improved results over snow-covered surfaces. We demonstrated that the KNN is able to identify precipitation phase with minimal dependency on ancillary data, such as the near-surface air temperature and moisture. The results showed that the global probability of hit for detection of solid precipitation over dry snow cover could reach up to ∼94%. However, the detection skill of the algorithm is decreased over regions with dense vegetation due to reduced polarization signal. A larger phase discrepancy

was found when the KNN results were compared with the ground-based precipitation phase, which remains to be addressed in future research.

It is important to emphasize that we have used V04 GPM official products. We expect to see fewer discrepancies between the GPM retrievals and the ground-based phase products in the following versions, because the latest version of the GPROF phase detection algorithm benefits from the longer GPM radar/radiometer joint records and the new DPR algorithm relies on an improved parameterization of ice microphysics.

Linking the algorithm with physical or observational databases that contain additional information on snow-cover physical properties (e.g., snow thickness, density, and liquid water content) and vegetation density can be a promising line of research. Furthermore, exploring the ways to constrain the output of the algorithm to the snowfall retrievals by the *CloudSat* radar may also help to improve the accuracy of snowfall detection. A physically realistic definition of mixed-phase precipitation based on cloud microphysics may reduce the uncertainties in phase retrievals. Finally, future research is also required to expand and evaluate the proposed algorithm with direct comparison of its results with ground-based gauge observations.

REFERENCES

Baggi, S., and J. Schweizer, 2009: Characteristics of wet-snow avalanche activity: 20 years of observations from a high alpine valley (Dischma, Switzerland). *Nat. Hazards*, **50**, 97–108, https://doi.org/10.1007/s11069-008-9322-7.

Bennartz, R., and P. Bauer, 2003: Sensitivity of microwave radiances at 85–183 GHz to precipitating ice particles. *Radio Sci.*, **38**, 8075, https://doi.org/10.1029/2002RS002626.

Chen, F. W., and D. H. Staelin, 2003: AIRS/AMSU/HSB precipitation estimates. *IEEE Trans. Geosci. Remote Sens.*, **41**, 410–417, https://doi.org/10.1109/TGRS.2002.808322.

Chen, S., J. J. Gourley, Y. Hong, Q. Cao, N. Carr, P.-E. Kirstetter, J. Zhang, and Z. Flamig, 2016: Using citizen science reports to evaluate estimates of surface precipitation type. *Bull. Amer. Meteor. Soc.*, **97**, 187–193, https://doi.org/10.1175/BAMS-D-13-00247.1.

Currier, W. R., T. Thorson, and J. D. Lundquist, 2017: Independent evaluation of frozen precipitation from WRF and PRISM in the Olympic Mountains. *J. Hydrometeor.*, **18**, 2681–2703, https://doi.org/10.1175/JHM-D-17-0026.1.

Doswell, C. A., R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585, https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.

Ebtehaj, A. M., and C. D. Kummerow, 2017: Microwave retrievals of terrestrial precipitation over snow-covered surfaces: A lesson from the GPM satellite. *Geophys. Res. Lett.*, **44**, 6154–6162, https://doi.org/10.1002/2017GL073451.

——, R. L. Bras, and E. Foufoula-Georgiou, 2015: Shrunken locally linear embedding for passive microwave retrieval of precipitation. *IEEE Trans. Geosci. Remote Sens.*, **53**, 3720–3736, https://doi.org/10.1109/TGRS.2014.2382436.

——, ——, and ——, 2016: Evaluation of sharp passive rainfall retrievals over snow-covered land surfaces and coastal zones. *J. Hydrometeor.*, **17**, 1013–1029, https://doi.org/10.1175/JHM-D-15-0164.1.

Gelaro, R., and Coauthors, 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.

Gong, J., and D. L. Wu, 2017: Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements. *Atmos. Chem. Phys.*, **17**, 2741–2757, https://doi.org/10.5194/acp-17-2741-2017.

Grody, N. C., 1991: Classification of snow cover and precipitation using the special sensor microwave imager. *J. Geophys. Res.*, **96**, 7423–7435, https://doi.org/10.1029/91JD00045.

——, 2008: Relationship between snow parameters and microwave satellite measurements: Theory compared with Advanced Microwave Sounding Unit observations from 23 to 150 GHz. *J. Geophys. Res.*, **113**, D22108, https://doi.org/10.1029/2007JD009685.

Hall, D. K., G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo, and K. J. Bayr, 2002: MODIS snow-cover products. *Remote Sens. Environ.*, **83**, 181–194, https://doi.org/10.1016/S0034-4257(02)00095-0.

Hallikainen, M. T., F. T. Ulaby, and M. Abdelrazik, 1986: Dielectric properties of snow in the 3 to 37 GHz range. *IEEE Trans. Antennas Propag.*, **34**, 1329–1340, https://doi.org/10.1109/TAP.1986.1143757.

——, ——, and T. E. Van Deventer, 1987: Extinction behavior of dry snow in the 18- to 90-GHz range. *IEEE Trans. Geosci. Remote Sens.*, **GE-25**, 737–745, https://doi.org/10.1109/TGRS.1987.289743.

Hanley, J. A., and B. J. McNeil, 1982: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36, https://doi.org/10.1148/radiology.143.1.7063747.

Harpold, A. A., M. L. Kaplan, P. Z. Klos, T. Link, J. P. McNamara, S. Rajagopal, R. Schumer, and C. M. Steele, 2017: Rain or snow: Hydrologic processes, observations, prediction, and research needs. *Hydrol. Earth Syst. Sci.*, **21**, 1–22, https://doi.org/10.5194/hess-21-1-2017.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1.

Houze, R. A., Jr., and Coauthors, 2017: The Olympic Mountains Experiment (OLYMPEX). *Bull. Amer. Meteor. Soc.*, **98**, 2167–2188, https://doi.org/10.1175/BAMS-D-16-0182.1.

Iguchi, T., and Coauthors, 2012: An overview of the precipitation retrieval algorithm for the dual-frequency precipitation radar (dpr) on the global precipitation measurement (gpm) mission's core satellite. *International Society for Optics and Photonics*, **8528** (9), 85281C.

——, S. Seto, R. Meneghini, N. Yoshida, J. Awaka, M. Le, V. Chandrasekar, and T. Kubota, 2017: GPM/DPR level-2. Algorithm Theoretical Basis Doc., 81 pp., https://pmm.nasa.gov/sites/default/files/document_files/ATBD_DPR_201708_whole_1.pdf.

Karl, T. R., P. Ya. Groisman, R. W. Knight, and R. R. Heim Jr., 1993: Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations. *J. Climate*, **6**, 1327–1344, https://doi.org/10.1175/1520-0442(1993)006<1327:RVOSCA>2.0.CO;2.

Kidd, C., and G. Huffman, 2011: Global precipitation measurement. *Meteor. Appl.*, **18**, 334–353, https://doi.org/10.1002/met.284.

Kienzle, S. W., 2008: A new temperature based method to separate rain and snow. *Hydrol. Processes*, **22**, 5067–5085, https://doi.org/10.1002/hyp.7131.

Kim, M.-J., J. A. Weinman, W. S. Olson, D.-E. Chang, G. Skofronick-Jackson, and J. R. Wang, 2008: A physical model to estimate snowfall over land using AMSU-B observations. *J. Geophys. Res.*, **113**, D09201, https://doi.org/10.1029/2007JD008589.

Kirstetter, P.-E., and Coauthors, 2012: Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar–based National Mosaic QPE. *J. Hydrometeor.*, **13**, 1285–1300, https://doi.org/10.1175/JHM-D-11-0139.1.

——, Y. Hong, J. J. Gourley, Q. Cao, M. Schwaller, and W. Petersen, 2014: Research framework to bridge from the Global Precipitation Measurement mission core satellite to the constellation sensors using ground-radar-based National Mosaic QPE. *Remote Sensing of the Terrestrial Water Cycle*, V. Lakshmi et al., Eds., John Wiley and Sons, 61–79, https://doi.org/10.1002/9781118872086.ch4.

Kongoli, C., P. Pellegrino, R. R. Ferraro, N. C. Grody, and H. Meng, 2003: A new snowfall detection algorithm over land using measurements from the Advanced Microwave Sounding Unit (AMSU). *Geophys. Res. Lett.*, **30**, 1756, https://doi.org/10.1029/2003GL017177.

——, H. Meng, J. Dong, and R. Ferraro, 2015: A snowfall detection algorithm over land utilizing high-frequency passive microwave measurements—Application to ATMS. *J. Geophys. Res. Atmos.*, **120**, 1918–1932, https://doi.org/10.1002/2014JD022427.

Kubota, T., and Coauthors, 2014: Evaluation of precipitation estimates by at-launch codes of GPM/DPR algorithms using synthetic data from TRMM/PR observations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **7**, 3931–3944, https://doi.org/10.1109/JSTARS.2014.2320960.

Kulie, M. S., R. Bennartz, T. J. Greenwald, Y. Chen, and F. Weng, 2010: Uncertainties in microwave properties of frozen precipitation: Implications for remote sensing and data assimilation. *J. Atmos. Sci.*, **67**, 3471–3487, https://doi.org/10.1175/2010JAS3520.1.

Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Oceanic Technol.*, **15**, 809–817, https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2.

——, D. L. Randel, M. Kulie, N. Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015: The evolution of the Goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280, https://doi.org/10.1175/JTECH-D-15-0039.1.

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.

Li, J.-L. F., and Coauthors, 2008: Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations. *Geophys. Res. Lett.*, **35**, L19710, https://doi.org/10.1029/2008GL035427.

Liu, G., 1998: A fast and accurate model for microwave radiance calculations. *J. Meteor. Soc. Japan*, **76**, 335–343, https://doi.org/10.2151/jmsj1965.76.2_335.

——, 2008: A database of microwave single-scattering properties for nonspherical ice particles. *Bull. Amer. Meteor. Soc.*, **89**, 1563–1570, https://doi.org/10.1175/2008BAMS2486.1.

——, 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.

Lundquist, J. D., P. J. Neiman, B. Martner, A. B. White, D. J. Gottas, and F. M. Ralph, 2008: Rain versus snow in the Sierra Nevada, California: Comparing Doppler profiling radar and surface observations of melting level. *J. Hydrometeor.*, **9**, 194–211, https://doi.org/10.1175/2007JHM853.1.

Mankin, J. S., D. Viviroli, D. Singh, A. Y. Hoekstra, and N. S. Diffenbaugh, 2015: The potential for snow to supply human water demand in the present and future. *Environ. Res. Lett.*, **10**, 114016, https://doi.org/10.1088/1748-9326/10/11/114016.

Mass, C. F., and Coauthors, 2003: Regional environmental prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.*, **84**, 1353–1366, https://doi.org/10.1175/BAMS-84-10-1353.

Matsuo, T., Y. Sasyo, and Y. Sato, 1981: Relationship between types of precipitation on the ground and surface meteorological elements. *J. Meteor. Soc. Japan*, **59**, 462–476, https://doi.org/10.2151/jmsj1965.59.4_462.

Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier, 2005: Declining mountain snowpack in western North America. *Bull. Amer. Meteor. Soc.*, **86**, 39–49, https://doi.org/10.1175/BAMS-86-1-39.

Motoyama, H., 1990: Simulation of seasonal snowcover based on air temperature and precipitation. *J. Appl. Meteor.*, **29**, 1104–1110, https://doi.org/10.1175/1520-0450(1990)029<1104:SOSSBO>2.0.CO;2.

Noh, Y.-J. J., G. Liu, A. S. Jones, T. H. V. Haar, and T. H. Vonder Haar, 2009: Toward snowfall retrieval over land by combining satellite and in situ measurements. *J. Geophys. Res.*, **114**, D24205, https://doi.org/10.1029/2009JD012307.

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.

Pederson, G. T., and Coauthors, 2011: The unusual nature of recent snowpack declines in the North American Cordillera. *Science*, **333**, 332–335, https://doi.org/10.1126/science.1201570.

Petty, G. W., W. Huang, G. W. Petty, and W. Huang, 2010: Microwave backscatter and extinction by soft ice spheres and complex snow aggregates. *J. Atmos. Sci.*, **67**, 769–787, https://doi.org/10.1175/2009JAS3146.1.

Prigent, C., W. B. Rossow, and E. Matthews, 1997: Microwave land surface emissivities estimated from SSM/I observations. *J. Geophys. Res.*, **102**, 21 867–21 890, https://doi.org/10.1029/97JD01360.

Radić, V., A. Bliss, A. C. Beedlow, R. Hock, E. Miles, and J. G. Cogley, 2014: Regional and global projections of twenty-first century glacier mass changes in response to climate scenarios from global climate models. *Climate Dyn.*, **42**, 37–58, https://doi.org/10.1007/s00382-013-1719-7.

Shi, J. J., and Coauthors, 2010: WRF simulations of the 20–22 January 2007 snow events over eastern Canada: Comparison with in situ and satellite observations. *J. Appl. Meteor. Climatol.*, **49**, 2246–2266, https://doi.org/10.1175/2010JAMC2282.1.

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 B. T. Johnson, 2011: Surface and atmospheric contributions to passive microwave brightness temperatures for falling snow events. *J. Geophys. Res.*, **116**, D02213, https://doi.org/10.1029/2010JD014438.

——, M.-J. Kim, J. Weinman, and D.-E. Chang, 2004: A physical model to determine snowfall over land by microwave radiometry. *IEEE Trans. Geosci. Remote Sens.*, **42**, 1047–1058, https://doi.org/10.1109/TGRS.2004.825585.

——, B. T. Johnson, and S. J. Munchak, 2013: Detection thresholds of falling snow from satellite-borne active and passive sensors. *IEEE Trans. Geosci. Remote Sens.*, **51**, 4177–4189, https://doi.org/10.1109/TGRS.2012.2227763.

——, 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.

Staelin, D. H., and F. W. Chen, 2000: Precipitation observations near 54 and 183 GHz using the NOAA-15 satellite. *IEEE Trans. Geosci. Remote Sens.*, **38**, 2322–2332, https://doi.org/10.1109/36.868889.

Stiles, W. H., and F. T. Ulaby, 1980: The active and passive microwave response to snow parameters: 1. Wetness. *J. Geophys. Res.*, **85**, 1037, https://doi.org/10.1029/JC085iC02p01037.

Takbiri, Z., A. M. Ebtehaj, and E. Foufoula-Georgiou, 2017: A multi-sensor data-driven methodology for all-sky passive microwave inundation retrieval. *Hydrol. Earth Syst. Sci.*, **21**, 2685–2700, https://doi.org/10.5194/hess-21-2685-2017.

Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, https://doi.org/10.1029/2010GL046008.

Ulaby, F. T., and W. H. Stiles, 1980: The active and passive microwave response to snow parameters: 2. Water equivalent of dry snow. *J. Geophys. Res.*, **85**, 1045–1049, https://doi.org/10.1029/JC085iC02p01045.

Wan, Z., 2014: New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sens. Environ.*, **140**, 36–45, https://doi.org/10.1016/j.rse.2013.08.027.

Wang, Y., G. Liu, E. K. Seo, and Y. Fu, 2013: Liquid water in snowing clouds: Implications for satellite remote sensing of snowfall. *Atmos. Res.*, **131**, 60–72, https://doi.org/10.1016/j.atmosres.2012.06.008.

Weng, F., B. Yan, and N. C. Grody, 2001: A microwave land emissivity model. *J. Geophys. Res.*, **106**, 20 115–20 123, https://doi.org/10.1029/2001JD900019.

Ye, H., J. Cohen, M. Rawlins, H. Ye, J. Cohen, and M. Rawlins, 2013: Discrimination of solid from liquid precipitation over northern Eurasia using surface atmospheric conditions. *J. Hydrometeor.*, **14**, 1345–1355, https://doi.org/10.1175/JHM-D-12-0164.1.

You, Y., N.-Y. Wang, and R. Ferraro, 2015: A prototype precipitation retrieval algorithm over land using passive microwave observations stratified by surface condition and precipitation vertical structure. *J. Geophys. Res. Atmos.*, **120**, 5295–5315, https://doi.org/10.1002/2014JD022534.

——, ——, ——, and P. Meyers, 2016: A prototype precipitation retrieval algorithm over land for ATMS. *J. Hydrometeor.*, **17**, 1601–1621, https://doi.org/10.1175/JHM-D-15-0163.1.

——, ——, ——, 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., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future Plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, https://doi.org/10.1175/2011BAMS-D-11-00047.1.

——, and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.