



AMERICAN METEOROLOGICAL SOCIETY

Journal of Hydrometeorology

EARLY ONLINE RELEASE

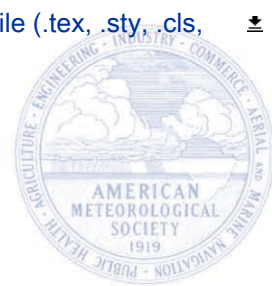
This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/JHM-D-18-0021.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Takbiri, Z., A. Ebtehaj, E. Foufoula-Georgiou, P. Kirstetter, and F. Turk, 2019: A Prognostic Retrieval Approach for Microwave Precipitation Phase Detection over Snow Cover. *J. Hydrometeor.* doi:10.1175/JHM-D-18-0021.1, in press.



A Prognostic Retrieval Approach for Microwave Precipitation Phase

Detection over Snow Cover

Zeinab Takkiri ^{*}, Ardeshir Ebtehaj [†], Efi Foufoula-Georgiou [‡], Pierre-Emmanuel Kirstetter [§], and

F. Joseph Turk [¶]

St. Anthony Falls Laboratory, University of Minnesota, Minneapolis, USA

^{*} *Corresponding author address:* Department of Civil, Environmental and Geo-Engineering and St. Anthony Falls Laboratory, University of Minnesota, Minneapolis, USA.

E-mail: takbi001@umn.edu

[†] Department of Civil, Environmental and Geo-Engineering and St. Anthony Falls Laboratory, University of Minnesota, Minneapolis, USA.

[‡] Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

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

[¶] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

ABSTRACT

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

34 **1. Introduction**

35 More than two billion people rely on glacier and snowmelt for their water supply (Mankin et al.
36 2015). Snowfall accounts for approximately 30 to 90 percent of the global precipitation over mid-
37 to high-latitudes (Levizzani et al. 2011) and is the main input to the accumulation processes of
38 snowpack and glaciers (Radić et al. 2014). In recent decades, snowpack reservoirs have declined
39 and are projected to further decline in the 21st century (Karl et al. 1993; Mote et al. 2005; Pederson
40 et al. 2011). Thus, global monitoring of snowfall from space is key for improved understanding
41 and prediction of ongoing changes in the cryosphere and the implications for sustainable manage-
42 ment of water and food resources — especially in mountainous areas of the world.

43 In the past three decades, significant progress has been made in microwave precipitation re-
44 trieval as part of the Tropical Rainfall Measuring Mission (TRMM) satellite in 1997 (Kummerow
45 et al. 1998). The launch of the Global Precipitation Measurement (GPM) core satellite (Kidd and
46 Huffman 2011; Hou et al. 2014) has provided a unique opportunity for improved understanding
47 of mid-latitude precipitation and its phase change beyond what the TRMM satellite could offer
48 (Skofronick-Jackson et al. 2017).

49 The snowfall microwave scattering signal can be captured at frequencies above 80 GHz as these
50 high frequencies are more sensitive to ice scattering compared to lower frequencies, which largely
51 respond to variations of surface emissivity (Kulie et al. 2010; Skofronick-Jackson and Johnson
52 2011; Gong and Wu 2017; You et al. 2017). Among high-frequency channels, Bennartz and Bauer
53 (2003) found that frequencies around and above 150 GHz provide a strong polarization signal for
54 snowfall detection (Gong and Wu 2017; You et al. 2017; Panegrossi et al. 2017).

55 Remote sensing of snowfall is among the most challenging tasks in precipitation retrieval algo-
56 rithms (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 (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

81 using the AMSU-B data. The study used the NESDIS Microwave Land Surface Emissivity Model
82 (Weng et al. 2001) to provide surface emissivity as an input to the RT model. The largest retrieval
83 error were found to be over snow-covered surfaces.

84 Empirical passive microwave snowfall retrieval algorithms largely rely on ancillary data of pre-
85 cipitation radar and air temperature profile. A family of these algorithms relies on thresholding
86 the brightness temperature at different channels (e.g., Staelin and Chen 2000; Chen and Staelin
87 2003; Kongoli et al. 2003). For example, Kongoli et al. (2015) developed a statistical approach
88 that partitions high-frequency brightness temperatures (≥ 89 GHz) into two distinct warm and
89 cold weather regimes by thresholding the brightness temperature at 53 GHz.

90 Another class of empirical approaches relies on Bayesian techniques. These techniques use
91 a database or a look-up table that relates brightness temperatures of snowing clouds to the radar
92 snowfall observations along with the atmospheric temperature profile. As an example, Liu and Seo
93 (2013) used matched observations from the CloudSat Profiling Radar (CPR), the AMSU-B, and
94 the NOAA'S Microwave Humidity Sounder (MHS). More recently, Sims and Liu (2015) used the
95 CloudSat radar and multiple ground-based reanalysis data, including near-surface air temperature,
96 atmospheric moisture, low-level vertical temperature lapse rate, surface skin temperature, surface
97 pressure, and land cover types to diagnose precipitation phase partitioning. This algorithm is
98 deployed in the GPM operational precipitation retrievals (Kummerow et al. 2015). It is worth
99 noting that most of these algorithms use reanalysis wet-bulb temperature that exhibits the strongest
100 correlation with the precipitation phase (Matsuo et al. 1981; Motoyama 1990; Lundquist et al.
101 2008; Kienzle 2008; Ye et al. 2013). However, the reanalysis data are often available at coarse
102 spatial scales with significant uncertainty, which hamper the applicability for accurate detection of
103 snowfall (Harpold et al. 2017).

104 In this paper, we examine a prognostic Bayesian k -nearest neighbor (KNN) algorithm that
105 strictly relies on observed passive microwave brightness temperatures and does not use any on-
106 line reanalysis data of temperature and moisture profiles. This approach is based on a weighted
107 distance metric applied on an a priori database to detect overland precipitation phase. The a priori
108 database is populated with combined radar-radiometer observations from the GPM satellite. This
109 database is then stratified using data from the Moderate Resolution Imaging Spectroradiometer
110 (MODIS) sensor to account for effects of the background snow-cover emission. We demonstrate
111 that the algorithm shows improved skill in detection of snowfall over snow cover and can predict
112 the likelihood of precipitation phase changes in the atmospheric boundary layer, which is not well
113 observed by the GPM radar.

114 In summary, the presented algorithm isolates a few physically relevant candidate vectors of
115 brightness temperatures in the database via a weighted Euclidean distance. Using these isolated
116 candidates, the method detects the precipitation and its phase, based on a probabilistic decision
117 rule. To test the performance of the proposed approach, the database is populated by merging the
118 outputs of both passive (Sims and Liu 2015) and active (Iguchi et al. 2010) GPM products using all
119 overland observations from June 2015 to May 2016. We compare the results with the ground-based
120 Multi-Radar Multi-Sensor (MRMS) data over the Conterminous United States (CONUS) (Zhang
121 et al. 2011, 2016). The outputs of a high-fidelity mesoscale simulation model are also used for
122 further evaluation of the results over high altitudes, during the Olympic Mountains Experiment in
123 2015 (OLYMPEX, Houze et al. 2017).

124 The paper is structured as follows: Section 2 briefly describes the database and the phase detec-
125 tion method used on the operational GPM radar and radiometer products. Section 3 elaborates on
126 the effects of snow cover on passive microwave signal of snowfall at different frequency channels
127 by analyzing a large dataset of GPM observations. Section 4 explains the proposed KNN algo-

128 rithm followed by the results presented in Section 5. Concluding remarks and future directions of
129 the research are discussed in Section 6.

130 **2. Database Description**

131 The dual-frequency precipitation radar (DPR) aboard the GPM core satellite detects precipita-
132 tion reflectivity at Ka- (35 GHz) and Ku-band (13.6 GHz). The GPM Microwave Imager (GMI)
133 captures the upwelling emission/scattering signals of the surface and the atmosphere at 13 fre-
134 quency channels ranging from 10 to 183 GHz. On the one hand, observations by the DPR and the
135 GMI high-frequency channels (> 80 GHz) provide information about the microwave signature of
136 precipitation and more specifically about snowfall ice scattering. On the other hand, observations
137 by the low-frequency channels (> 80 GHz) add information about the land surface characteristics
138 that leads to improved detection skill by the presented algorithm. This study uses level-II near-
139 surface precipitation phase retrieval from DPR (active) product (2A-DPR-V04, Normal Scan),
140 GMI (passive) product (2A-GPROF-V04) and the level 1B calibrated GMI brightness tempera-
141 tures.

142 In DPR level-II, the precipitation phase is determined by the dual-frequency retrieval approach
143 that uses the differential attenuation between the Ku- and the Ka-band reflectivity values (Iguchi
144 et al. 2010, 2012). The differential attenuation method ingests ancillary atmospheric profile data
145 such as air temperature, pressure, and the microphysical parameterization of the snow and rain
146 particle size distribution. The DPR surface retrieval is inferred from the near-surface reflectivity
147 observations in the clutter-free region. Above relatively flat surfaces, the altitude of this region
148 varies from 1 to 2 km from nadir to the edge of the DPR swath. The depth of this region is often
149 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 $\sim 0.2 \text{ mm h}^{-1}$ (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

174 logical rule, we implicitly address the limitations of DPR in identifying precipitation phase change
175 in the ground clutter region which overlaps with the boundary layer.

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

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

197 To account for atmospheric radiometric signals, we also added the integrated liquid and ice wa-
 198 ter content of the clouds, as well as the integrated water vapor content of the atmospheric column
 199 from the second version of the Modern-Era Retrospective analysis for Research and Applications
 200 (MERRA-2-M2I1NXASM, Gelaro et al. 2017). The MERRA-2 data are hourly single-level diag-
 201 nostic products at $0.625^\circ \times 0.5^\circ$, which are derived from the version 5 of the NASA Goddard Earth
 202 Observing System (GEOS-5) and the Atmospheric Data assimilation system (ADAS).

203 To form the database with a uniform spatial sampling density, the GMI brightness temperatures
 204 and the MERRA-2 reanalysis data are mapped onto the DPR grids and scanning time using the
 205 spatial nearest neighbor interpolation and temporal linear interpolation techniques. The high-
 206 resolution MODIS snow-cover data are mapped onto and then averaged over the nearest DPR
 207 grids. We collect and process two consecutive years of data, from June 2014 to May 2016, which
 208 lead to a database with more than 5×10^9 matched data pairs. The data from the first year (June
 209 2014 to May 2015) are applied to build the database and the data from the second year (June 2015
 210 to May 2016) are used to test the proposed algorithm.

211 **3. The effect of snow cover on terrestrial snowfall signal**

212 Precipitation spectral signatures and their dependence on snow-cover scattering are studied by
 213 analyzing the entire dataset (June 2014 to May 2016) for three surface types (ground without snow
 214 cover, wet snow, and dry snow) and for both liquid and solid phases of precipitation. Each land-
 215 atmospheric class is further divided into 5 bins of precipitation intensity r with equal logarithmic
 216 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
 217 snow cover on the precipitation signal over each surface type by calculating the mean values of
 218 the brightness temperatures for different precipitation phases and intensities at frequency bands
 219 10–19, 36–89, 166, 183 ± 3 , and 183 ± 7 GHz (Fig. 1). Then, we demonstrate that the snowfall

220 signal exhibits a weaker scattering signal than rainfall and reveal that there exists a non-unique
221 relationship between the brightness temperatures and snowfall rate over snow-covered surfaces.
222 Lastly, we highlight why precipitation phase detection could be less uncertain over dry than wet
223 snow cover using the presented approach.

224 The first three columns in Fig. 1 a–i focus on the signatures of rainfall over land surfaces with
225 no snow cover, wet snow cover, and dry snow cover, where both active and passive products
226 indicate liquid phase. The signatures over the ground with no snow cover are mainly affected
227 by the upwelling surface emission, the upwelling emission by cloud liquid water path, as well as
228 scattering by the cloud ice particles and large raindrops. As it is well understood, due to strong
229 background emission at frequencies 10–36 GHz, the overland precipitation microwave signal is
230 difficult to be separated from the surface contributed signal in these channels. For example, due to
231 the rainfall emission, the mean brightness temperature at 10h GHz only increases by less than 5 K
232 as the intensity increases from 0.5 to 8 mmh⁻¹ (Fig. 1 a).

233 On average, we observe that over all three land surface types, the brightness temperatures mono-
234 tonically decrease for frequencies above 80 GHz as the rainfall intensity increases. However, the
235 significance of scattering decreases over snow-covered surfaces (Fig. 1 d–i). For example, at 89
236 and 166 GHz, the average decrease of brightness temperature per 1 mmh⁻¹ increase in rainfall
237 intensity is about 3.0 and 3.6 K (Fig. 1 d, g), while these rates are around 1.2 and 2.3 K over the
238 dry snow cover (Fig. 1 h, i). As expected, the 183 ± 3 GHz is the least sensitive channel to the
239 changes of rainfall rate. This channel becomes almost insensitive to the rainfall intensity when
240 the snow accumulates on the ground and exhibits less than 0.2 K of cooling effect per unit rainfall
241 (Fig. 1 i).

242 The last three columns in Fig. 1 j–r present brightness temperatures of snowfall over the three
243 explained land surface types, where both active and passive products indicate solid phase. Similar

244 to the overland rainfall, the emission and the scattering signals become more significant from low
245 to high-frequency channels. Over the surfaces with no snow cover, when the snowfall intensity in-
246 creases from 0.5 to 8 mmh⁻¹, the brightness temperatures at frequencies ≤ 36 GHz increase ~ 6 K
247 (Fig. 1 j). This warming could be due to increased cloud liquid water path (from 75 to 101 g m⁻²,
248 Fig. 2 a, d), water vapor path (from 9.5 to 13.1 kg m⁻², Fig. 2 c, f), and surface temperature (from
249 273 to 274.2 K, Fig. 2 g, i).

250 Due to the snowfall scattering, the average brightness temperature at 166 GHz frequency channel
251 (Fig. 1 l, o, r) decreases about 14 to 20 K, which corresponds to a cooling rate of 1.75 to 2.50 K
252 per unit snowfall rate. This observation reaffirms the importance of 166 GHz for snowfall retrieval
253 compared to the 89 GHz channel (see Bennartz and Bauer 2003; Shi et al. 2010; Skofronick-
254 Jackson et al. 2013; You et al. 2017). When the precipitation intensity increases from 0.5 to
255 8 mmh⁻¹, the average decrease in brightness temperatures at 166 (89) GHz is about 18 to 30 (10
256 to 22) K for rainfall and 10 to 20 (2 to 9) K for snowfall — over all examined land surface types.
257 Therefore, the scattering signal weakens when the precipitation falls in the solid form; however,
258 this weakening effect is less significant at 166 GHz than 89 GHz. In particular, over the ground
259 with no snow cover, the signal becomes weaker approximately by 30% and 57% at 166 and 89
260 GHz, respectively, while these rates are 44 and 80% over the dry snow cover.

261 Observations demonstrate that the snowfall scattering signal decreases at frequencies ≥ 89 GHz
262 when snow begins to accumulate on the ground. An interesting observation is the non-monotonic
263 response of the observed brightness temperatures to the snowfall rate over snow-covered surfaces.
264 For example, over the dry snow, the brightness temperatures at ≥ 89 GHz increase when the
265 snowfall intensity varies from 2 to 4 mmh⁻¹, showing an irregular transition from a scattering to
266 an emission regime (Fig. 1 q, r). Although less pronounced, a similar pattern is observed over the
267 wet snow cover (Fig. 1 n, o).

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

277 Over the ground with no snow cover, as the average precipitation intensity increases, the liquid
278 and ice water path increase during rainfall and even more significantly during snowfall. Specifi-
279 cally, the liquid water path increases from 14 to 26% (Fig. 2 a, d) and the ice water path increases
280 about 23 and 37% (Fig. 2 b, e) for raining and snowing events, respectively. Over dry snow cover,
281 there is no evidence of any additional changes neither in liquid nor in ice water path that could
282 cause the observed irregularity. Fig. 2 f shows that the water vapor path increases about 2.5 kg m^{-2}
283 between snowfall intensities 2 and 8 mm h^{-1} over the dry snow cover, which cannot be the main
284 reason for the observed anomaly. The reason is that the sensitivity of the 166 GHz channel to
285 variation of water vapor decreases significantly for snowfall intensities $> 0.8 \text{ mm h}^{-1}$ (You et al.
286 2017). Therefore, we speculate that the anomaly could be largely due to a surface effect.

287 The MODIS surface temperature data (Wan 2014) do not show any significant dependency on
288 the rate of snowfall (Fig. 2 i). Therefore, we hypothesize that the detected emission could be due to
289 either an unknown retrieval uncertainty or more likely, to the climatology of the snowfall and snow
290 cover dynamics. The database shows that light but prolonged snowfall intensities ($< 2 \text{ mm h}^{-1}$)
291 occur at latitudes above $> 55^\circ \text{ N}$ over dry and thicker snow cover (Fig. 2 j). However, high in-

292 tensity but less frequent snowfall is more likely to occur over lower latitudes with a thinner snow
293 cover climatology. In other words, the high snowfall rates mostly represent the climatology of
294 lower latitudes with thinner depth of snow cover, less volume scattering, and thus stronger surface
295 emission than the thicker snow cover of higher latitudes.

296 The above observations from Figs 1 and 2 lead us to hypothesize that the distance between
297 vectors of brightness temperature encodes a similarity metric that can be used to discriminate the
298 precipitation from the background signal. A larger distance indicates larger radiometric dissimi-
299 larity that could be used for improved detection of the precipitation from the background signal.
300 Using the database, we calculate the average distance between the vectors of brightness tempera-
301 tures for the clear-sky (no precipitation) and precipitating atmosphere over the three land surface
302 types (Fig. 3). In this figure, the shaded areas in light blue (orange) represent the detected emission
303 (scattering) signal. The key observation is that when the snow-cover scattering increases, the pre-
304 cipitation signal transitions from a scattering to an emission regime. The wet snow cover weakens
305 the precipitation scattering as it is less emissive than the ground with no snow cover. However, the
306 less emissive dry snow reveals the precipitation emission signal.

307 For the liquid phase, we can see that the rainfall scattering over the ground with no snow cover
308 is manifested by a large distance between the high-frequency channels ≥ 89 GHz, while the dis-
309 tance over lower frequency channels is insignificant (Fig. 3 a). This distance shrinks over the wet
310 snow cover (Fig. 3 b), where the dominant precipitation signal is still due to its scattering over
311 high-frequency channels. This shrinkage is largely explained because wet snow is not a strong
312 scatterer and thus reduces slightly the surface emission and the high-frequency scattering of rain-
313 fall. However, when the surface emission is significantly reduced over the dry snow (Fig. 3 c), the
314 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. 3 d). 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 temperature 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 $\sim 58 \text{ g m}^{-2}$ and 4.8 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 non-precipitating 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 $\vec{T}b$ and the ancillary data containing information on the precipitation occurrence, phase, and snow cover, are represented by the vector \vec{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 $\left\{ \left(\vec{T}b_m, \vec{r}_m \right) \right\}_{m=1}^M$, 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 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 \vec{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 $d_m = \left(\vec{y} - \vec{T}b_m \right)^T \mathbf{W}_a \left(\vec{y} - \vec{T}b_m \right)$, 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 $\left\{ \left(\vec{T}b_i, \vec{r}_i \right) \right\}_{i=1}^k$ of each query vector \vec{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 $\left\{ \vec{T}b_i \right\}_{i=1}^{k_1}$ and the corresponding ancillary information in the database. Then, a binary decision is made to label the vector \vec{y} as a precipitating observation, when the number of precipitating neighbors n_p is greater than $p_1 \times k_1$. For precipitating \vec{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 $\left\{ \left(\vec{T}b_j, \vec{r}_j \right) \right\}_{j=1}^{n_p}$ 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 \vec{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 $\left\{ \vec{T}b_i \right\}_{i=1}^{k_2}$. If those conditions are not satisfied, the algorithm continues similarly to find if the phase of \vec{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 curves (ROC, Fig. 5), which characterize the tradeoff 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 $\frac{a}{a+c}$ and $\frac{b}{b+d}$, 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 k -nearest neighbor algorithm (KNN) 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 in Figs. 7,8 and 9. We also compare the algorithm outputs with the precipitation phase products of the Multi-Radar/Multi-Sensor System (MRMS) on a seasonal basis (Figs. 10 and 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, high-resolution ground-based radars (Figs. 12 and 13), and simulated by the Weather Research and Forecasting (WRF) model (Fig. 14) during the Olympic Mountain Experiment (OLYMPEX, Houze et al. 2017).

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

407 a probability continuum of phase transition from the liquid (0) to solid (1), at the grid resolution
408 0.1-degree. These results are mapped where the precipitation is detected only by the DPR for
409 two seasons. The seasons are defined as summer (June-October 2015 and May 2016) and winter
410 (November 2015 to April 2016) of the Northern Hemisphere.

411 Overall, since the phase of the passive product is dictated by the reanalysis data, the results
412 mostly follow the climatology patterns of near-surface wet-bulb temperature and are smoother than
413 those of the active product (Fig. 6 a-d). The smoothness of the GPROF retrievals could also be
414 due to its ability in retrieving the light precipitation with intensities below the minimum detectable
415 rate by the DPR ($< 0.2 \text{ mm h}^{-1}$), as the GPROF also uses precipitation data from MRMS ground-
416 based radar in its a priori database to increase the retrieval sensitivity to snowfall. Comparison of
417 the official passive and active products remains outside the scope of this research; however, there
418 seem to be notable differences in the spatial patterns of precipitation phases in these products.
419 The difference in the source of ancillary data could be a reason for the observed discrepancies,
420 which largely exist over mountainous terrains such as the Andes, Tibetan highlands, Rockies,
421 Scandinavian mountains, Alps, and Zagros Mountains (Fig. 6 e, f) — where precipitation is mostly
422 triggered by topographic features.

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

Fig. 6 i,j shows the results of the KNN algorithm in summer and winter and compare them with the REF map (Fig. 6 k,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 (Fig. 6 k,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. 7 b). 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 Squared Error (RMSE), and the Kullback-Leibler divergence (KL) in Fig. 1. The KL divergence $KL(P \parallel Q) = \sum_{i=1}^n \frac{P(i)}{Q(i)}$ is a non-symmetric and non-negative 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 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 to 0.91 in winter and summer, indicating that the algorithm is not excessively sensitive to the sea-

sonal 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 $HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}$ (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 and Figs 7 and 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. 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 score are observed over the snow cover, tropical forests, and under-sampled deserts such as Sahara, respectively (Fig. 7c).

477 The conditional probability of hit and false alarm are calculated for liquid, mixed, and solid
 478 phases (Fig. 8), with respect to the REF product. For separating the errors of the precipitation
 479 and phase detection, the probabilities are obtained assuming that the precipitation is correctly
 480 detected by the KNN algorithm. Similar to the precipitation detection, the algorithm displays
 481 improved phase detection capabilities over snow-covered surfaces (Fig. 8). The probability of hit
 482 for the liquid, mixed, and the solid phase is mostly greater than 0.85 and reaches 0.95 over the
 483 high-altitudes of North America. However, we observe a relatively lower detection rate of around
 484 0.74 for liquid precipitation over the tropical and subtropical regions such as the rainforest of
 485 Amazonian and Central Africa. The results show that the low probability of detection for the liquid
 486 phase is mostly because the algorithm detects some false mixed phase precipitation (Fig. 8 d).
 487 We speculate that this error could be partly attributed to the reduced skill of the algorithm over
 488 vegetated surfaces. The reduced detection skill of the algorithm could also be partly due to warm
 489 rain occurrences over the heterogeneous land surface of tropical and subtropical regions where
 490 cloud ice scattering is not significant.

491 To understand the reasons for improved retrievals over snow-covered surfaces, the averaged
 492 values of the probability of hit and false alarm are stratified based on precipitation occurrence (\mathcal{D}_1)
 493 at liquid (\mathcal{D}_2), mixed (\mathcal{D}_3) and the solid (\mathcal{D}_4) phase over different land surface types $\{\mathcal{L}\}_{s=1}^3$,
 494 where $s = 1$ to 3 denotes the ground, wet, and dry snow cover (Table 2). The probability of
 495 precipitation detection increases by almost 11% from the ground to the dry snow cover, and 3%
 496 from wet to dry snow. An increase of 8 to 11% is also observed in the probability of hit in
 497 detection of solid and liquid phase over dry snow, where the largest detection rate of 94% is
 498 obtained for the snowfall. The results show that the probability of false alarm also increases in
 499 detection of precipitation occurrence over snow cover, whereas it decreases when it comes to
 500 the detection of its phase. Because, once precipitation is detected, due to significant differences

501 between the signatures of rain and snowfall, the probability of false alarm is markedly reduced.
502 Table 2 quantifies the dependency of the probability of hit and false alarm on the annual percentage
503 of the dry snow cover. For precipitation detection, the probability of hit increases by about 10%
504 as the annual percentage of dry snow increases from zero to more than 70%, while the probability
505 of false alarm increases between 2–4%. As is evident, for precipitation phase detection, both
506 probabilities show improvements of around 4%.

507 *b. Comparison with the ground-based radar*

508 1) COMPARISON WITH MULTI-RADAR/MULTI-SENSOR SYSTEM

509 To further evaluate the performance of the KNN algorithm, we compare its outputs against a
510 precipitation product derived from the Multi-Radar/Multi-Sensor System (MRMS) (Zhang et al.
511 2011, 2016). MRMS mosaics three-dimensional volume scan observations from 146 S-band dual-
512 polarization Doppler Weather Surveillance Radar-1988 (WSR-88D) and 31 C-band single polar-
513 ization Canadian radars. The product optimally integrates the radar observations with simula-
514 tions of atmospheric models as well as hourly gauge data to produce seamless precipitation rate
515 and phase estimates over the CONUS, at spatial resolution of 1 km at every 2 min. The MRMS
516 products are further quality-controlled and gauge-adjusted at fine scale following the procedure
517 described in Kirstetter et al. (2012) to derive a consistent and high quality surface precipitation.

518 To determine the precipitation phase, MRMS uses thresholds on the wet and dry bulb temper-
519 atures. Specifically, the precipitation is labeled as snowfall when the radar reflectivity exceeds
520 5 dBZ, the surface temperature is below 2 °C, and the surface wet bulb temperature is below 0 °C
521 (Zhang et al. 2016). Thus the MRMS rain-snow delineation is subject to similar uncertainties as
522 in the passive GPM data (Chen et al. 2016). However, the uncertainties in detecting precipitation
523 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).

Fig. 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 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 ($< 40^\circ$ 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.

Fig. 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 the 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

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

555 It is surprising that in orbits #10412 and #12149 (Fig. 12), the DPR reports the phase as liquid
556 where the GPROF classifies the phase largely as solid since the discrepancy is often in the other
557 direction. Based on the atmospheric temperature profile derived from environmental ancillary data
558 (2A-DPRENV) used in the active retrieval algorithm, we conclude that there is a temperature
559 inversion when the storm is happening (see Fig. 13). In this case, liquid precipitation can refreeze
560 near the surface and may not be captured by the DPR.

561 2) COMPARISON WITH THE WRF SIMULATIONS DURING THE OLYMPEX

562 The MRMS data lacks coverage over mountainous regions, thus we need a venue with rich
563 ground-based observations for further evaluation of the presented approach. There is a wealth
564 of orographic precipitation data during the GPM Olympic Mountains Experiment (OLYMPEX,
565 Houze et al. 2017) from November 1 to December 23, 2015. The Olympic Mountains are lo-
566 cated in the northwestern corner of the Washington State, United States (Fig. 14) with a dominant
567 orographic precipitation regime. This regime is a result of the abrupt uplift of moisture-laden
568 southwest airflow coming from the mid-latitude baroclinic storm systems. A few high-elevation
569 snow and precipitation gauges were used during the OLYMPEX field campaign. However, the
570 coarse temporal resolution of the DPR (i.e., 117 partial overpasses), relative to the 56-day du-

571 ration of OLYMPEx hamper their use for our purpose. Therefore, we choose the outputs of a
572 high-resolution (1.33 km) hourly WRF simulation by the Northwest Modeling Consortium over
573 the Olympic Mountains (Mass et al. 2003).

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

583 Fig. 14 illustrates the precipitation phase for the DPR, GPROF, KNN, and the WRF for three
584 GPM orbits (#9722, #09773, #10019). We observe that at high-elevation regions, the KNN de-
585 tects mixed phase over the areas that exhibit phase discrepancies between the GPROF and DPR.
586 We see that these KNN results are in a good agreement with the WRF simulations. However, it is
587 important to note that the precipitation phase partitioning in the WRF outputs is based on cloud mi-
588 crophysical parameters in the atmospheric boundary layer, and thus its mixed-phase precipitation
589 is physically different than the defined mixed-phase category in KNN retrievals.

590 We calculate and compare the average phase outputs of the DPR, GPROF, KNN, and WRF data
591 for all 117 coincident DPR overpasses. We found that compared with the average phase probability
592 of WRF, the KNN precipitation phase is positively biased by about 28% (i.e., KNN captures more
593 solid phase than WRF, Fig. 14). However, this bias is about 31% at elevations above 800 meters,
594 while reduced to about 24% for lower elevations. Additionally, the results show that over areas

with elevations higher than 800 meters, the KNN phase bias is significantly smaller compared to both DPR (positive bias $\sim 48\%$) and GPROF (negative bias $\sim 56\%$). At elevations below 800 meters, the KNN is less biased than the positively biased DPR ($\sim 41\%$); however, about 9% more biased than GPROF with a negative bias $\sim 19\%$. 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 to 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 $\sim 94\%$. 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

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

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

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

Acknowledgments. The authors acknowledge the support from the National Aeronautics and Space Administration (NASA) through a Precipitation Measurement Mission award (NNX16AO56G) and a New Investigator Program award (80NSSC18K0742). Zeinab Takbiri acknowledges the support provided by the Minnesota's Discovery, Research, and Innovation Economy (MnDRIVE 2017) fellowship. Pierre-Emmanuel Kirstetter acknowledges support from the NASA Precipitation Science Program (NNX16AE39G) and from the GPM mission Ground Validation Program (NNX16AL23G). The contributions from F. Joseph Turk were performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. The GPM data (version 4) is provided courtesy of the NASA Precipitation Processing System at the Goddard Space Flight Center (<https://pmm.nasa.gov/data-access/>). The MERRA-2 and MODIS data are from the Goddard Earth Sciences and Information Service Center (<https://disc.sci.gsfc.nasa.gov/mdisc/>) and the Land Processes Distributed Active Archive Center by the USGS (https://lpdaac.usgs.gov/data_access/data_pool). The authors would also like to thank Ryan Currier from the University of Washington for providing the data from the Weather Research and Forecasting Model over the Olympic Mountains.

References

- Baggi, S., and J. Schweizer, 2009: Characteristics of wet-snow avalanche activity: 20 years of observations from a high alpine valley (dischma, switzerland). *Natural Hazards*, **50** (1), 97–108, doi: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 Science*, **38** (4), –, doi:10.1029/2002RS002626.
- Chen, F. W., and D. H. Staelin, 2003: Airs/amsu/hsb precipitation estimates. *IEEE Transactions on Geoscience and Remote Sensing*, **41** (2), 410–417, doi: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. *Bulletin of the American Meteorological Society*, **97** (2), 187–193, doi: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. *Journal of Hydrometeorology*, **18** (10), 2681–2703, doi: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. *Weather and Forecasting*, **5** (4), 576–585, doi:10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.
- Ebtehaj, A. M., R. L. Bras, and E. Foufoula-Georgiou, 2015: Shrunk locally linear embedding for passive microwave retrieval of precipitation. *IEEE Transactions on Geoscience and Remote Sensing*, **53** (7), 3720–3736, doi:10.1109/TGRS.2014.2382436.

667 Ebtehaj, A. M., R. L. Bras, and E. Foufoula-Georgiou, 2016: Evaluation of sharp passive rainfall
668 retrievals over snow-covered land surfaces and coastal zones. *Journal of Hydrometeorology*,
669 **17 (4)**, 1013–1029, doi:10.1175/JHM-D-15-0164.1.

670 Ebtehaj, A. M., and C. D. Kummerow, 2017: Microwave retrievals of terrestrial precipitation
671 over snow-covered surfaces: A lesson from the GPM satellite. *Geophysical Research Letters*,
672 **44 (12)**, 6154–6162, doi:10.1002/2017GL073451.

673 Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and
674 Applications, Version 2 (MERRA-2). *Journal of Climate*, **30 (14)**, 5419–5454, doi:10.1175/
675 JCLI-D-16-0758.1.

676 Gong, J., and D. L. Wu, 2017: Microphysical properties of frozen particles inferred from Global
677 Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements. *At-*
678 *mospheric Chemistry and Physics*, **17 (4)**, 2741–2757.

679 Grody, N. C., 1991: Classification of snow cover and precipitation using the special sensor mi-
680 crowave imager. *Journal of Geophysical Research*, **96 (D4)**, 7423, doi:10.1029/91JD00045.

681 Grody, N. C., 2008: Relationship between snow parameters and microwave satellite measure-
682 ments: Theory compared with Advanced Microwave Sounding Unit observations from 23
683 to 150 GHz. *Journal of Geophysical Research: Atmospheres*, **113 (D22)**, –, doi:10.1029/
684 2007JD009685, d22108.

685 Hall, D. K., G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo, and K. J. Bayr, 2002:
686 MODIS snow-cover products. *Remote Sensing of Environment*, **83 (1-2)**, 181–194, doi:10.1016/
687 S0034-4257(02)00095-0.

688 Hallikainen, M. T., F. T. Ulaby, and M. Abdelrazik, 1986: Dielectric properties of snow in the 3
689 to 37 GHz range. *IEEE Transactions on Antennas and Propagation*, **AP-34** (11), 1329–1340,
690 doi:10.1109/TAP.1986.1143757.

691 Hallikainen, M. T., F. T. Ulaby, and T. E. Van Deventer, 1987: Extinction Behavior of Dry Snow in
692 the 18- to 90-GHz Range. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-25** (6),
693 737–745, doi:10.1109/TGRS.1987.289743.

694 Hanley, J. A., and B. J. McNeil, 1982: The meaning and use of the area under a receiver operating
695 characteristic (roc) curve. *Radiology*, **143** (1), 29–36, doi:10.1148/radiology.143.1.7063747.

696 Harpold, A. A., M. L. Kaplan, P. Z. Klos, T. Link, J. P. McNamara, S. Rajagopal, R. Schumer, and
697 C. M. Steele, 2017: Rain or snow: hydrologic processes, observations, prediction, and research
698 needs. *Hydrology and Earth System Sciences*, **21** (1), 1, doi:doi:10.5194/hess-21-1-2017.

699 Hou, A. Y., and Coauthors, 2014: The global precipitation measurement mission. *Bulletin of the*
700 *American Meteorological Society*, **95** (5), 701–722, doi:10.1175/BAMS-D-13-00164.1.

701 Houze, R. A. J., and Coauthors, 2017: The olympic mountains experiment (olympex). *Bulletin of*
702 *the American Meteorological Society*, **98** (10), 2167–2188, doi:10.1175/BAMS-D-16-0182.1.

703 Iguchi, T., S. Seto, R. Meneghini, N. Yoshida, J. Awaka, and T. Kubota, 2010: GPM/DPR
704 Level-2 algorithm theoretical basis document authors. Technical report, NASA Goddard Space
705 Flight Center, Greenbelt, MD, USA. [Available online at [https://pmm.nasa.gov/sites/default/](https://pmm.nasa.gov/sites/default/files/document_files/ATBD_GPM_DPR_n3_dec15.pdf)
706 [files/document_files/ATBD_GPM_DPR_n3_dec15.pdf](https://pmm.nasa.gov/sites/default/files/document_files/ATBD_GPM_DPR_n3_dec15.pdf).].

707 Iguchi, T., and Coauthors, 2012: An overview of the precipitation retrieval algorithm for the dual-
708 frequency precipitation radar (dpr) on the global precipitation measurement (gpm) mission's

core satellite. *International Society for Optics and Photonics*, **8528 (9)**, 85 281C, doi:10.1117/12.977352.

Karl, T. R., P. Y. Groisman, R. W. Knight, and R. R. H. Jr., 1993: Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations. *Journal of Climate*, **6 (7)**, 1327–1344, doi:10.1175/1520-0442(1993)006<1327:RVOSCA>2.0.CO;2.

Kidd, C., and G. Huffman, 2011: Global precipitation measurement. *Meteorological Applications*, **18 (3)**, 334–353, doi:10.1002/met.284.

Kienzle, S. W., 2008: A new temperature based method to separate rain and snow. *Hydrological Processes*, **22 (26)**, 5067–5085, doi: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. *Journal of Geophysical Research*, **113 (D9)**, D09 201, doi:10.1029/2007JD008589.

Kirstetter, P. E., 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, D. Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, G. Huffman, W. Kustas, P. V. Oevelen, T. Painter, J. Parajka, M. Rodell, and R. C. Eds., John Wiley & Sons, Inc, 61–79, doi:10.1002/9781118872086.ch4, URL <http://dx.doi.org/10.1002/9781118872086.ch4>.

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.

Journal of Hydrometeorology, **13** (4), 1285–1300, doi:10.1175/JHM-D-11-0139.1.

Kongoli, C., H. Meng, J. Dong, and R. Ferraro, 2015: A snowfall detection algorithm over land utilizing high-frequency passive microwave measurements—Application to ATMS. *Journal of Geophysical Research: Atmospheres*, **120** (5), 1918–1932.

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). *Geophysical research letters*, **30** (14).

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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **7** (9), 3931–3944, doi: 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. *Journal of the Atmospheric Sciences*, **67** (11), 3471–3487, doi:10.1175/2010JAS3520.1.

Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The tropical rainfall measuring mission(TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15** (3), 809–817, doi:10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2.

Kummerow, C. D., 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. *Journal of Atmospheric and Oceanic Technology*, **32** (12), 2265–2280, doi: 10.1175/JTECH-D-15-0039.1.

- 752 Levizzani, V., S. Laviola, and E. Cattani, 2011: Detection and measurement of snowfall from
753 space. *Remote Sensing*, **3** (1), 145–166, doi:10.3390/rs3010145.
- 754 Li, J.-L. F., and Coauthors, 2008: Comparisons of satellites liquid water estimates to ecmwf and
755 gmao analyses, 20th century ipcc ar4 climate simulations, and gcm simulations. *Geophysical*
756 *Research Letters*, **35** (19), doi:10.1029/2008GL035427.
- 757 Liu, G., 1998: A fast and accurate model for microwave radiance calculations. *J. Meteorol. Soc.*
758 *Japan*, **76** (2), 335–343, URL <http://cat.inist.fr/?aModele=afficheN&cpsidt=2379327>.
- 759 Liu, G., 2008: A Database of microwave single-scattering properties for nonspherical ice par-
760 ticles. *Bulletin of the American Meteorological Society*, **89** (10), 1563–1570, doi:10.1175/
761 2008BAMS2486.1.
- 762 Liu, G., and E.-K. Seo, 2013: Detecting snowfall over land by satellite high-frequency microwave
763 observations: The lack of scattering signature and a statistical approach. *Journal of geophysical*
764 *research: atmospheres*, **118** (3), 1376–1387.
- 765 Lundquist, J. D., P. J. Neiman, B. Martner, A. B. White, D. J. Gottas, and F. M. Ralph, 2008: Rain
766 versus snow in the Sierra Nevada, California: Comparing Doppler profiling radar and surface
767 observations of melting level. *Journal of Hydrometeorology*, **9** (2), 194–211.
- 768 Mankin, J. S., D. Viviroli, D. Singh, A. Y. Hoekstra, and N. S. Diffenbaugh, 2015: The potential
769 for snow to supply human water demand in the present and future. *Environmental Research*
770 *Letters*, **10** (11), 114 016, doi:10.1088/1748-9326/10/11/114016.
- 771 Mass, C. F., and Coauthors, 2003: Regional environmental prediction over the pacific north-
772 west. *Bulletin of the American Meteorological Society*, **84** (10), 1353–1366, doi:10.1175/
773 BAMS-84-10-1353.

774 Matsuo, T., Y. Sasyo, and Y. Sato, 1981: elationship between Types of Precipitation on the Ground
775 and Surface Meteorological Elements. *Journal of the Meteorological Society of Japan*, **59** (4),
776 462–476.

777 Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier, 2005: Declining Mountain Snow-
778 pack in Western North America. *Bulletin of the American Meteorological Society*, **86** (1), 39–49,
779 doi:10.1175/BAMS-86-1-39.

780 Motoyama, H., 1990: Simulation of seasonal snowcover based on air temperature and precipita-
781 tion. *Journal of Applied Meteorology*, **29** (11), 1104–1110.

782 Noh, Y.-J. J., G. Liu, A. S. Jones, T. H. V. Haar, and T. H. Vonder Haar, 2009: Toward snowfall
783 retrieval over land by combining satellite and in situ measurements. *Journal of Geophysical*
784 *Research Atmospheres*, **114** (24), 1–15, doi:10.1029/2009JD012307.

785 Panegrossi, G., J. F. Rysman, D. Casella, A. C. Marra, P. Sanò, and M. S. Kulie, 2017: CloudSat-
786 based assessment of GPM Microwave Imager snowfall observation capabilities. *Remote Sens-*
787 *ing*, **9** (12), 1263.

788 Pederson, G. T., and Coauthors, 2011: The unusual nature of recent snowpack declines
789 in the north american cordillera. *Science*, **333** (6040), 332–335, doi:10.1126/science.
790 1201570, URL <http://science.sciencemag.org/content/333/6040/332>, <http://science.sciencemag.org/content/333/6040/332.full.pdf>.

792 Petty, G. W., W. Huang, G. W. Petty, and W. Huang, 2010: Microwave backscatter and extinction
793 by soft ice spheres and complex snow aggregates. *Journal of the Atmospheric Sciences*, **67** (3),
794 769–787, doi:10.1175/2009JAS3146.1.

795 Prigent, C., W. B. Rossow, and E. Matthews, 1997: Microwave land surface emissivities estimated
796 from SSM/I observations. *Journal of Geophysical Research*, **102 (D18)**, 21 867–21 890,, doi:
797 10.1002/1099-0690(200204)2002:7<1143::AID-EJOC1143>3.0.CO;2-G.

798 Radić, V., A. Bliss, A. C. Beedlow, R. Hock, E. Miles, and J. G. Cogley, 2014: Regional and global
799 projections of twenty-first century glacier mass changes in response to climate scenarios from
800 global climate models. *Climate Dynamics*, **42 (1-2)**, 37–58, doi:10.1007/s00382-013-1719-7.

801 Shi, J. J., and Coauthors, 2010: Wrf simulations of the 20–22 january 2007 snow events over east-
802 ern canada: Comparison with in situ and satellite observations. *Journal of Applied Meteorology*
803 *and Climatology*, **49 (11)**, 2246–2266, doi:10.1175/2010JAMC2282.1.

804 Sims, E. M., and G. Liu, 2015: A Parameterization of the Probability of Snow–Rain Transition.
805 *Journal of Hydrometeorology*, **16 (4)**, 1466–1477, doi:10.1175/JHM-D-14-0211.1.

806 Skofronick-Jackson, G., and B. T. Johnson, 2011: Surface and atmospheric contributions to pas-
807 sive microwave brightness temperatures for falling snow events. *Journal of Geophysical Re-*
808 *search*, **116 (D2)**, D02 213, doi:10.1029/2010JD014438.

809 Skofronick-Jackson, G., M.-J. Kim, J. Weinman, and D.-E. Chang, 2004: A physical model to
810 determine snowfall over land by microwave radiometry. *IEEE Transactions on Geoscience and*
811 *Remote Sensing*, **42 (5)**, 1047–1058, doi:10.1109/TGRS.2004.825585.

812 Skofronick-Jackson, G., and Coauthors, 2017: The global precipitation measurement (GPM) mis-
813 sion for science and Society. *Bulletin of the American Meteorological Society*, **98 (8)**, 1679–
814 1695, doi:10.1175/BAMS-D-15-00306.1.

815 Skofronick-Jackson, G. M., B. T. Johnson, and S. J. Munchak, 2013: Detection thresholds of
 816 falling snow from satellite-borne active and passive sensors. *IEEE Transactions on Geoscience
 817 and Remote Sensing*, **51** (7), 4177–4189, doi:10.1109/TGRS.2012.2227763.

818 Staelin, D. H., and F. W. Chen, 2000: Precipitation observations near 54 and 183 ghz using the
 819 noaa-15 satellite. *IEEE Transactions on Geoscience and Remote Sensing*, **38** (5), 2322–2332,
 820 doi:10.1109/36.868889.

821 Stiles, W. H., and F. T. Ulaby, 1980: The active and passive microwave response to snow
 822 parameters: 1. Wetness. *Journal of Geophysical Research*, **85** (C2), 1037, doi:10.1029/
 823 JC085iC02p01037.

824 Takbiri, Z., A. M. Ebtehaj, and E. Foufoula-Georgiou, 2017: A multi-sensor data-driven method-
 825 ology for all-sky passive microwave inundation retrieval. *Hydrology and Earth System Sciences*,
 826 **21** (6), 2685–2700, doi:10.5194/hess-21-2685-2017.

827 Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based
 828 precipitation measurements. *Geophysical Research Letters*, **37** (24), n/a–n/a, doi:10.1029/
 829 2010GL046008.

830 Ulaby, F. T., and W. H. Stiles, 1980: The active and passive microwave response to
 831 snow parameters: 2. Water equivalent of dry snow. *Journal of Geophysical Research:
 832 Oceans*, **85** (C2), 1045–1049, doi:10.1029/JC085iC02p01045, URL [http://dx.doi.org/10.1029/
 833 JC085iC02p01045](http://dx.doi.org/10.1029/JC085iC02p01045).

834 Wan, Z., 2014: New refinements and validation of the collection-6 MODIS land-surface temper-
 835 ature/emissivity product. *Remote Sensing of Environment*, **140**, 36–45, doi:10.1016/j.rse.2013.
 836 08.027.

- 837 Wang, Y., G. Liu, E. K. Seo, and Y. Fu, 2013: Liquid water in snowing clouds: Implications for
838 satellite remote sensing of snowfall. *Atmospheric Research*, **131** (April), 60–72, doi:10.1016/j.
839 atmosres.2012.06.008.
- 840 Weng, F., B. Yan, and N. C. Grody, 2001: A microwave land emissivity model. *J. Geophys. Res.*
841 *Atmos.*, **106** (D17), 20 115–20 123, doi:10.1029/2001JD900019, URL [http://doi.wiley.com/10.](http://doi.wiley.com/10.1029/2001JD900019)
842 1029/2001JD900019.
- 843 Ye, H., J. Cohen, M. Rawlins, H. Ye, J. Cohen, and M. Rawlins, 2013: Discrimination of Solid
844 from Liquid Precipitation over Northern Eurasia Using Surface Atmospheric Conditions. *Jour-*
845 *nal of Hydrometeorology*, **14** (4), 1345–1355, doi:10.1175/JHM-D-12-0164.1.
- 846 You, Y., N.-Y. Wang, and R. Ferraro, 2015: A prototype precipitation retrieval algorithm over land
847 using passive microwave observations stratified by surface condition and precipitation vertical
848 structure. *Journal of Geophysical Research: Atmospheres*, **120** (11), 5295–5315, doi:10.1002/
849 2014JD022534, URL <http://dx.doi.org/10.1002/2014JD022534>, 2014JD022534.
- 850 You, Y., N.-Y. Wang, R. Ferraro, and P. Meyers, 2016: A prototype precipitation retrieval al-
851 gorithm over land for atms. *Journal of Hydrometeorology*, **17** (5), 1601–1621, doi:10.1175/
852 JHM-D-15-0163.1.
- 853 You, Y., N.-Y. Wang, R. Ferraro, and S. Rudlosky, 2017: Quantifying the snowfall detection
854 performance of the GPM Microwave Imager channels over land. *Journal of Hydrometeorology*,
855 **18** (3), 729–751, doi:10.1175/JHM-D-16-0190.1.
- 856 Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: de-
857 scription, results, and future Plans. *Bulletin of the American Meteorological Society*, **92** (10),
858 1321–1338, doi:10.1175/2011BAMS-D-11-00047.1.

859 Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation es-
860 timation: initial operating capabilities. *Bulletin of the American Meteorological Society*, **97** (4),
861 621–638, doi:10.1175/BAMS-D-14-00174.1.

LIST OF TABLES

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 Squared Difference (RMSD), Spearman's correlation (ρ), and the Kullback-Leibler divergence (KL).	41
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, $s = 1$ to 3 represents the ground, wet, and dry snow-covered surfaces, $i = 1$ denotes the detection of precipitation occurrence, and $i = 2$ to 4 represents 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.	42
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 Root Mean Squared Difference (RMSD), Spearman's correlation (ρ), and the Kullback-Leibler divergence (KL).	43

877 TABLE 1. Quality metrics obtained by comparing the annual probability of phase transition between the KNN
878 results and the reference product (REF). Shown statistics are the normalized Root Mean Squared Difference
879 (RMSD), Spearman's correlation (ρ), and the Kullback-Leibler divergence (KL).

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

880 TABLE 2. The annual probability of hit and false alarm for the KNN results over different land surface types
881 $\{\mathcal{L}\}_{s=1}^3$ and detection classes $\{\mathcal{D}\}_{i=1}^4$. Here, $s = 1$ to 3 represents the ground, wet, and dry snow-covered
882 surfaces, $i = 1$ denotes the detection of precipitation occurrence, and $i = 2$ to 4 represents the detection of liquid,
883 mixed, and solid phase, respectively. The results over the dry snow cover (\mathcal{L}_3) are further stratified based on the
884 annual percentage of the snow.

Probability of hit								
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
\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								
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
\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

885 TABLE 3. Quality metrics obtained by comparing the annual probability of phase transition between the
886 KNN retrievals and MRMS observations. Shown statistics are the normalized Root Mean Squared Difference
887 (RMSD), Spearman's correlation (ρ), and the Kullback-Leibler divergence (KL).

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

LIST OF FIGURES

Fig. 1.	Variation of mean brightness temperatures (June 2014 to May 2016) in response to changes in precipitation intensity for different land-atmosphere classes including the liquid (a-i) and solid (j-r) precipitation phase over the ground (no snow), 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.	46
Fig. 2.	Average variations of the cloud liquid water path (LWP) (a,/d); the cloud ice water path (IWP) (b,/e); the water vapor path (WVP) (c,/f); the skin temperature (T_s) (g,/i); and latitudes (h,j) 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 (Wan 2014, MOD11A1) during June 2014 until May 2016. The intensity bins are the same as Fig. 1.	47
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 liquid (a-c) and solid phase (d-f) 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.	48
Fig. 4.	Algorithm flowchart of the proposed weighted k -nearest neighbor (KNN) algorithm for detection of precipitation occurrence and phase.	49
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 no-snow covered surfaces (top row: a–c) and snow-covered surfaces (bottom row: d–f). The optimal values of k and the detection probabilities p are given in each subplot.	50
Fig. 6.	Seasonal probability of the precipitation phase change. The average phase of the DPR (a, b), GPROF (c, d), REF (merged) (g, h) and the KNN algorithm (i, j) as well as the differences between the DPR and GPROF (e, f) and the REF and KNN products (k, l). The differences are shown where both products detect the precipitation occurrence.	51
Fig. 7.	The mean annual map of the probability of hit (a), probability of false alarm (b), and Heidke score (c) obtained by comparing the pixel-level results of the KNN algorithm with the REF product for the detection of precipitation occurrence. The map of snow-cover fraction (d) is also obtained from the MODIS data (MOD10A1 Hall et al. 2002) coincident with GPM inner-swath overpasses from June 2015 to May 2016.	52
Fig. 8.	The mean annual map of the probability of hit and false alarm by the KNN algorithm for the detection of the liquid phase (a, b), mixed phase (c, d) and solid phase (e, f). The results are obtained for all GPM inner-swath overpasses from June 2015 to May 2016.	53
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.	54

929	Fig. 10.	Mean seasonal maps of the probability of precipitation phase change from liquid ($p = 0$) to	
930		solid ($p = 1$) for KNN in winter (a) and summer (b), and for the MRMS in winter (c) and	
931		summer (d), from June 2015 to May 2016	55
932	Fig. 11.	The zonal mean of the probability of precipitation phase change from liquid ($p = 0$) to solid	
933		($p = 1$) by the KNN and MRMS products in (a) winter (November-April) and (b) summer	
934		(May-October) — from June 2015 to May 2016.	56
935	Fig. 12.	Orbital-level precipitation phase detection from the KNN, DPR, GPROF, and MRMS for	
936		a few GPM overpasses including #10412 on 2015/12/28 (top row), #10796 on 2016/01/22	
937		(second row), #12149 on 2016/04/18 (third row), and #12155 on 2016/04/18 (bottom row). . . .	57
938	Fig. 13.	Inversion of the air temperature at orbit #10412 on 2015/12/28 (top row) and orbit #12149	
939		on 2016/04/18 (second row). The data (2A-DPRENV) are presented at four ranges from 0	
940		(surface) to 1.5 km.	58
941	Fig. 14.	The digital elevation map (DEM) of the Olympic Mountains (top panel), and the precipi-	
942		tation phase by the DPR, GPROF, WRF, and KNN for orbit numbers #9722 (2015/11/14,	
943		second row), #9773 (2015/11/17, third row), and #10019 (2015/12/03, fourth row). The	
944		bottom panel is the average probability of phase for 117 GPM inner-swath overpasses from	
945		November 1 to December 23, 2015. The last column shows the 2-meter air temperature	
946		from the WRF simulations.	59

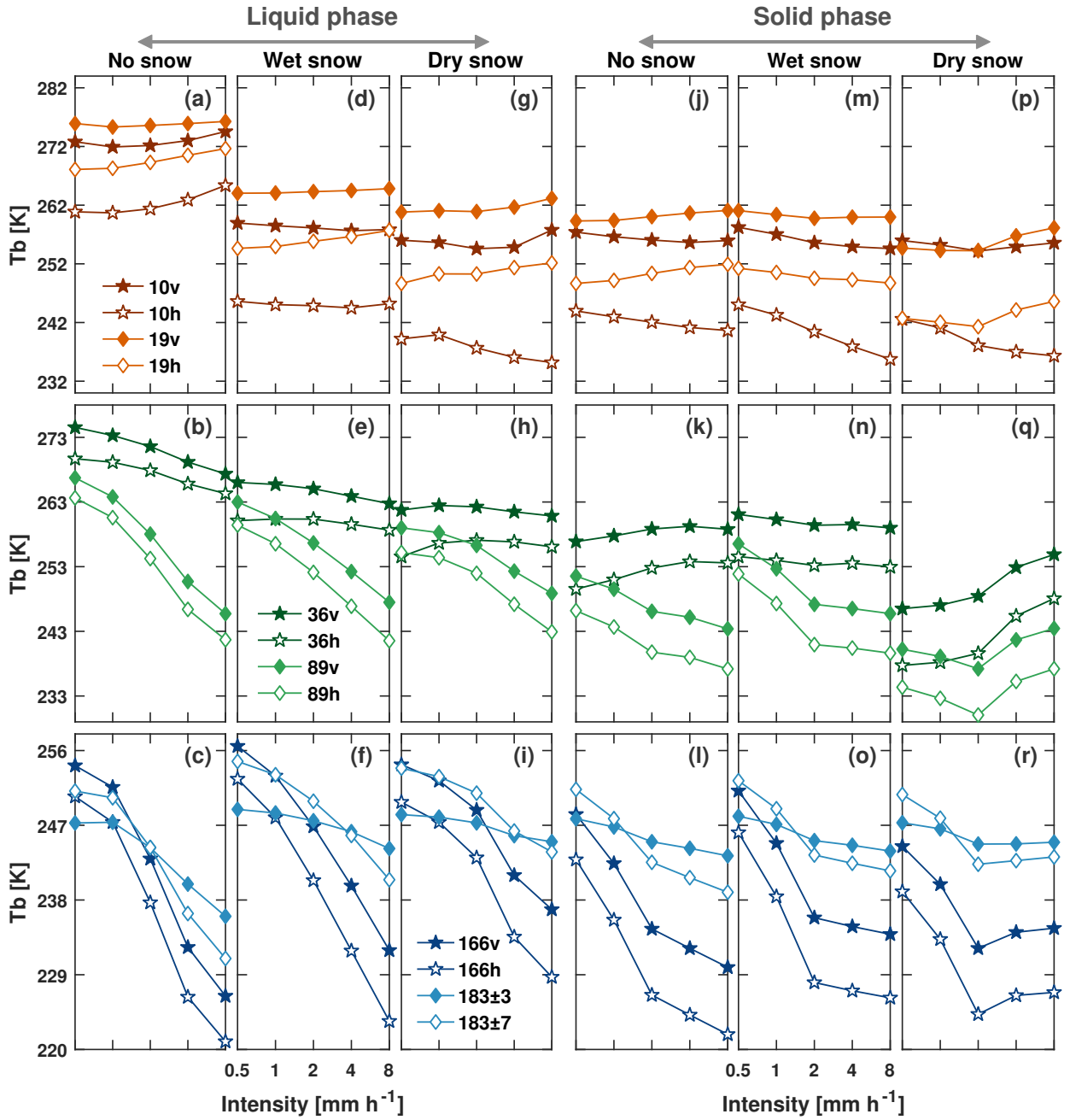
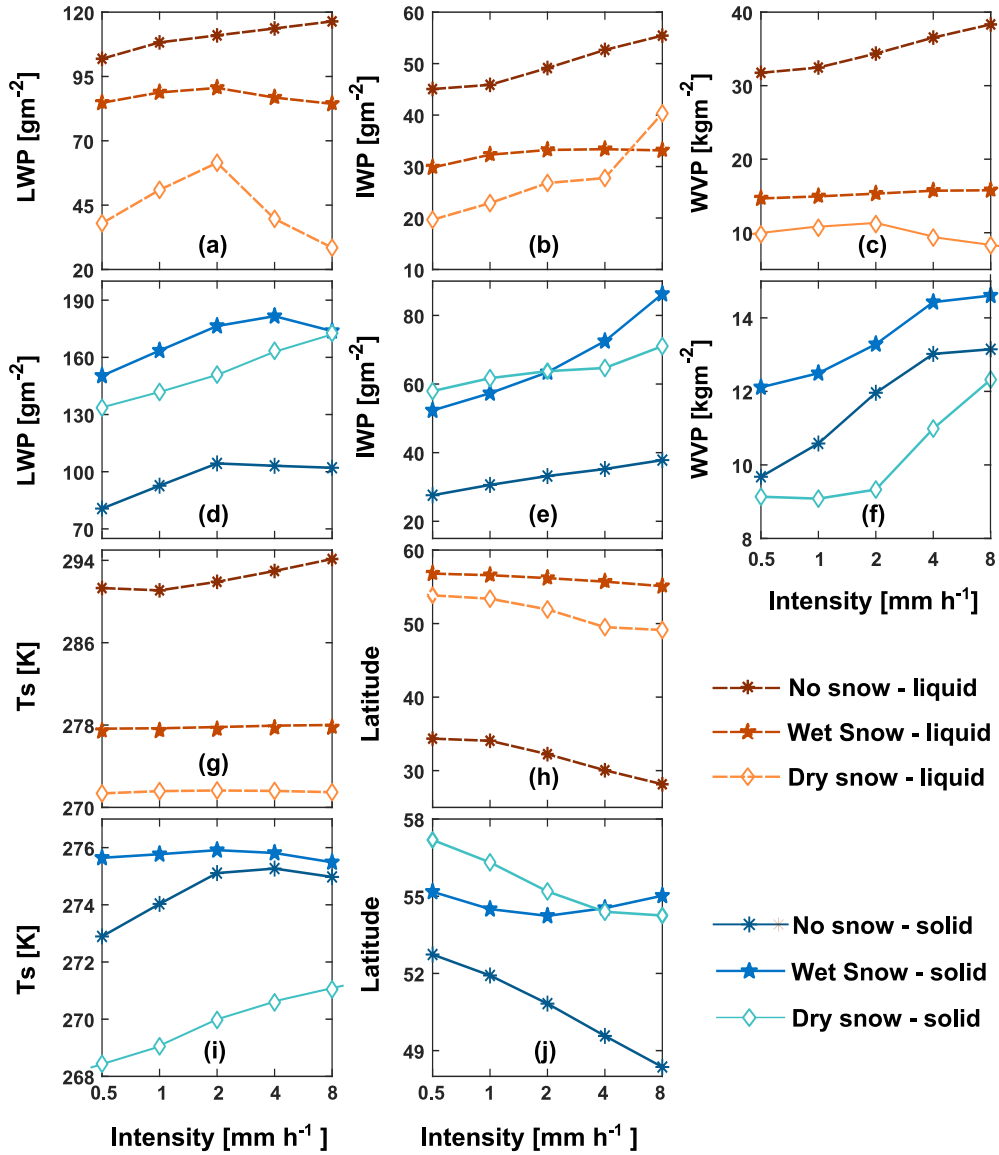


FIG. 1. Variation of mean brightness temperatures (June 2014 to May 2016) in response to changes in precipitation intensity for different land-atmosphere classes including the liquid (a-i) and solid (j-r) precipitation phase over the ground (no snow), wet and dry snow cover. The bins are five logarithmically (base 2) spaced intervals with the width of 1 mm h^{-1} in the log-space.



951 FIG. 2. Average variations of the cloud liquid water path (LWP) (a,/,d); the cloud ice water path (IWP) (b,/,e);
 952 the water vapor path (WVP) (c,/,f); the skin temperature (T_s) (g,/,i); and latitudes (h,j) against the precipitation
 953 intensity. The ice and liquid water paths are extracted from the MERRA-2 data (M2I1NXASM, Gelaro et al.
 954 2017) and the surface temperature data are from MODIS (Wan 2014, MOD11A1) during June 2014 until May
 955 2016. The intensity bins are the same as Fig. 1.

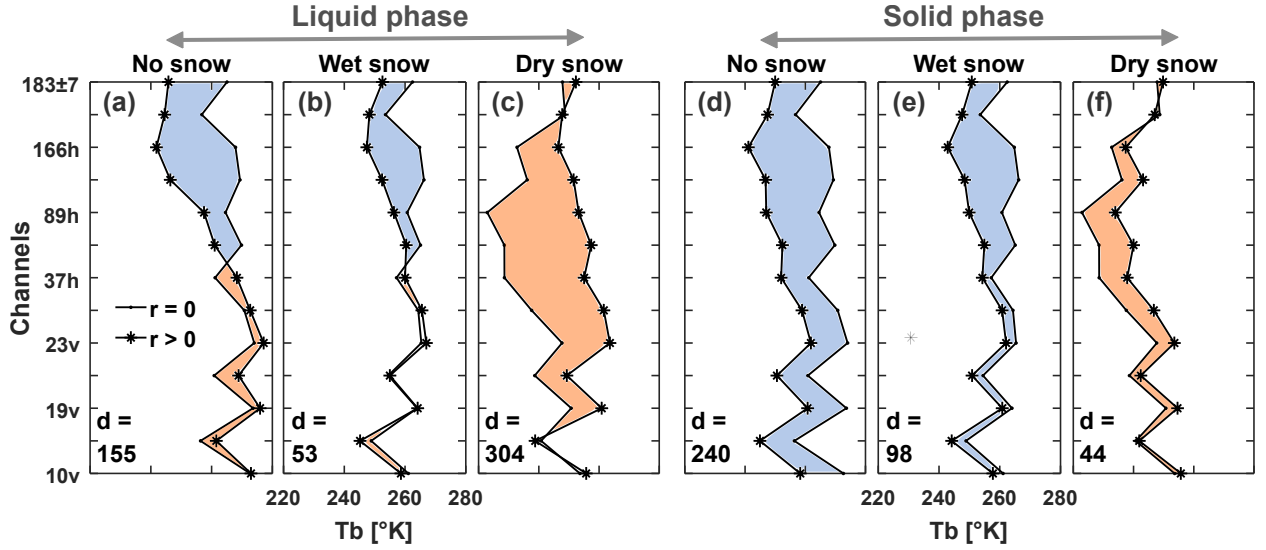


FIG. 3. Average distance between vectors of mean brightness temperatures in the database from June 2014 to May 2016 for a clear-sky ($r = 0$) and a near-surface precipitating atmosphere ($r > 0$) with liquid (a-c) and solid phase (d-f) overland precipitation with no snow cover, wet snow cover, and dry 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.

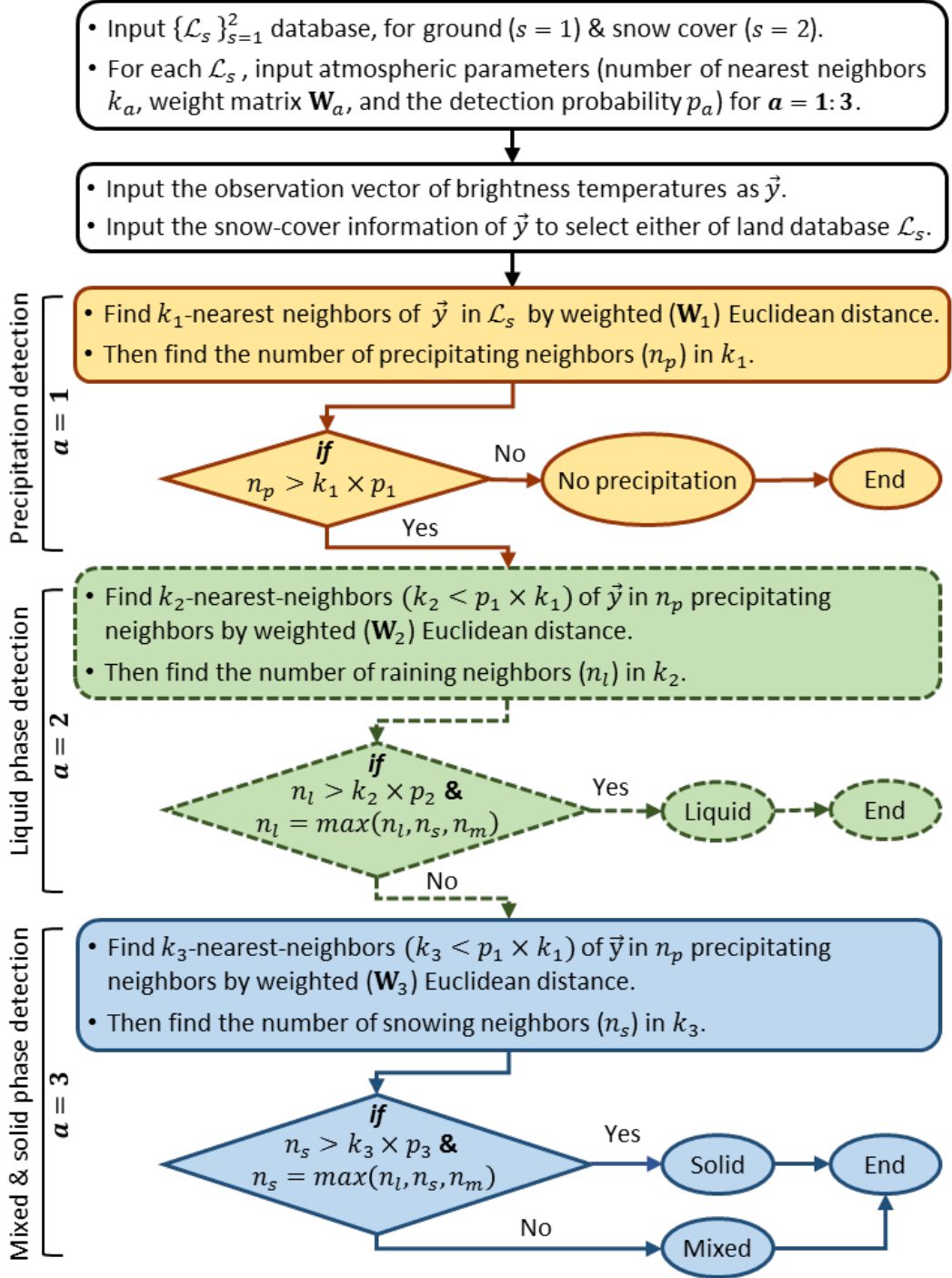


FIG. 4. Algorithm flowchart of the proposed weighted k -nearest neighbor (KNN) algorithm for detection of precipitation occurrence and phase.

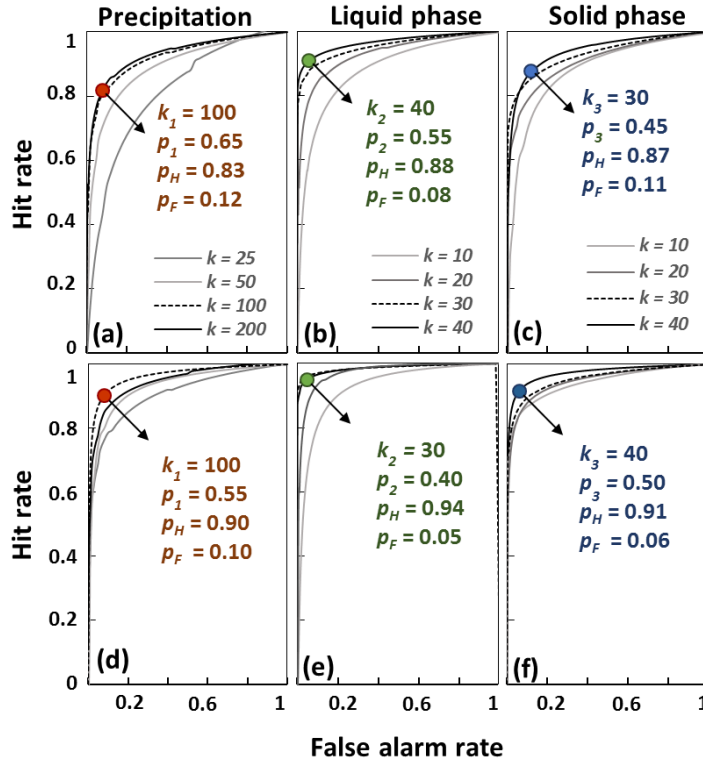


FIG. 5. Trade-off curves between the probability of hit (p_H) and false alarm (p_F) calculated with different numbers of k -nearest neighbors for detection of the precipitation occurrence and phase over no-snow covered surfaces (top row: a–c) and snow-covered surfaces (bottom row: d–f). The optimal values of k and the detection probabilities p are given in each subplot.

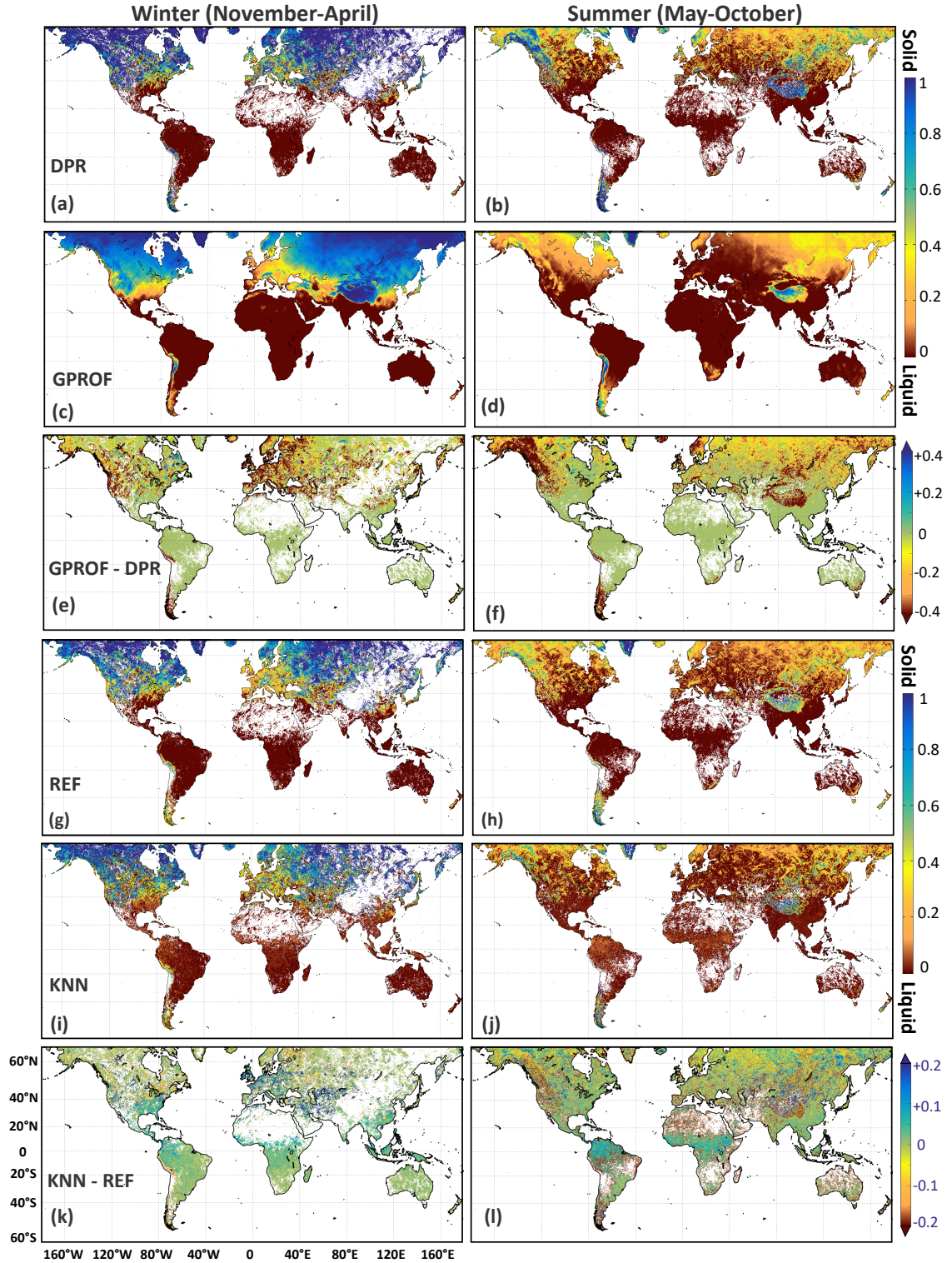


FIG. 6. Seasonal probability of the precipitation phase change. The average phase of the DPR (a, b), GPROF (c, d), REF (merged) (g, h) and the KNN algorithm (i, j) as well as the differences between the DPR and GPROF (e, f) and the REF and KNN products (k, l). The differences are shown where both products detect the precipitation occurrence.

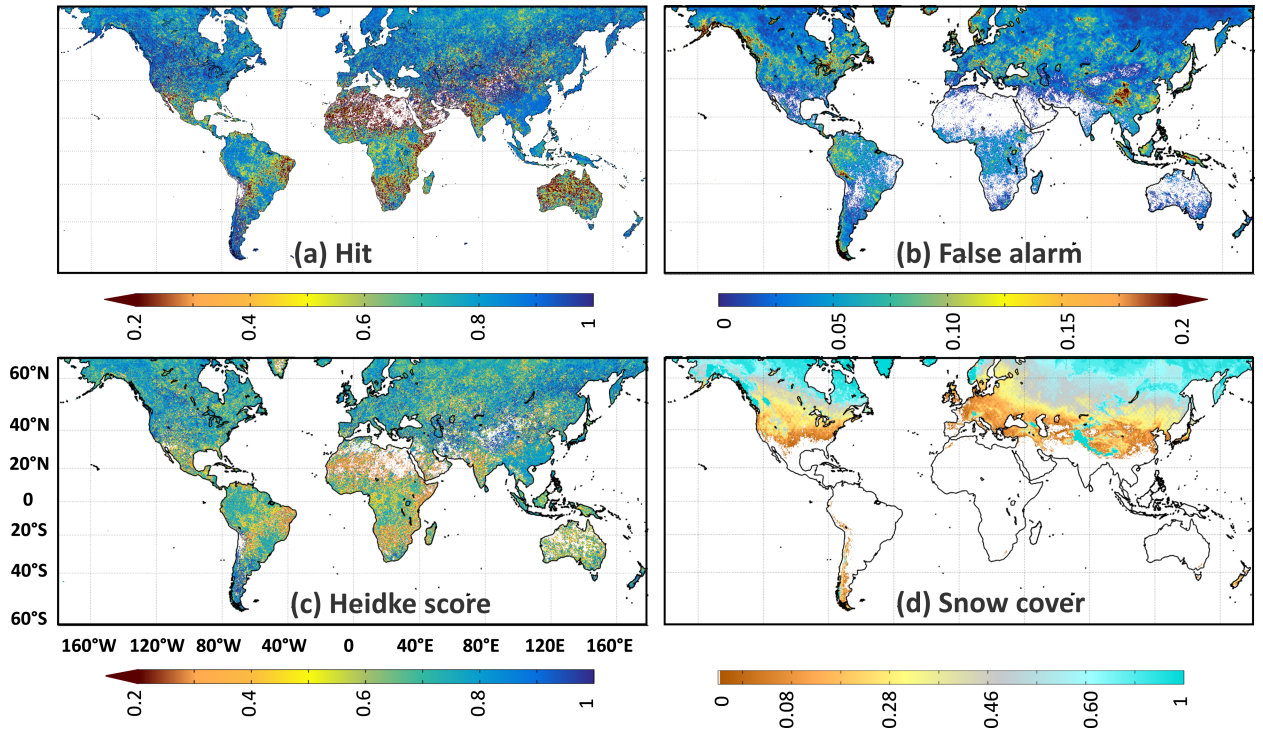


FIG. 7. The mean annual map of the probability of hit (a), probability of false alarm (b), and Heidke score (c) obtained by comparing the pixel-level results of the KNN algorithm with the REF product for the detection of precipitation occurrence. The map of snow-cover fraction (d) is also obtained from the MODIS data (MOD10A1 Hall et al. 2002) coincident with GPM inner-swath overpasses from June 2015 to May 2016.

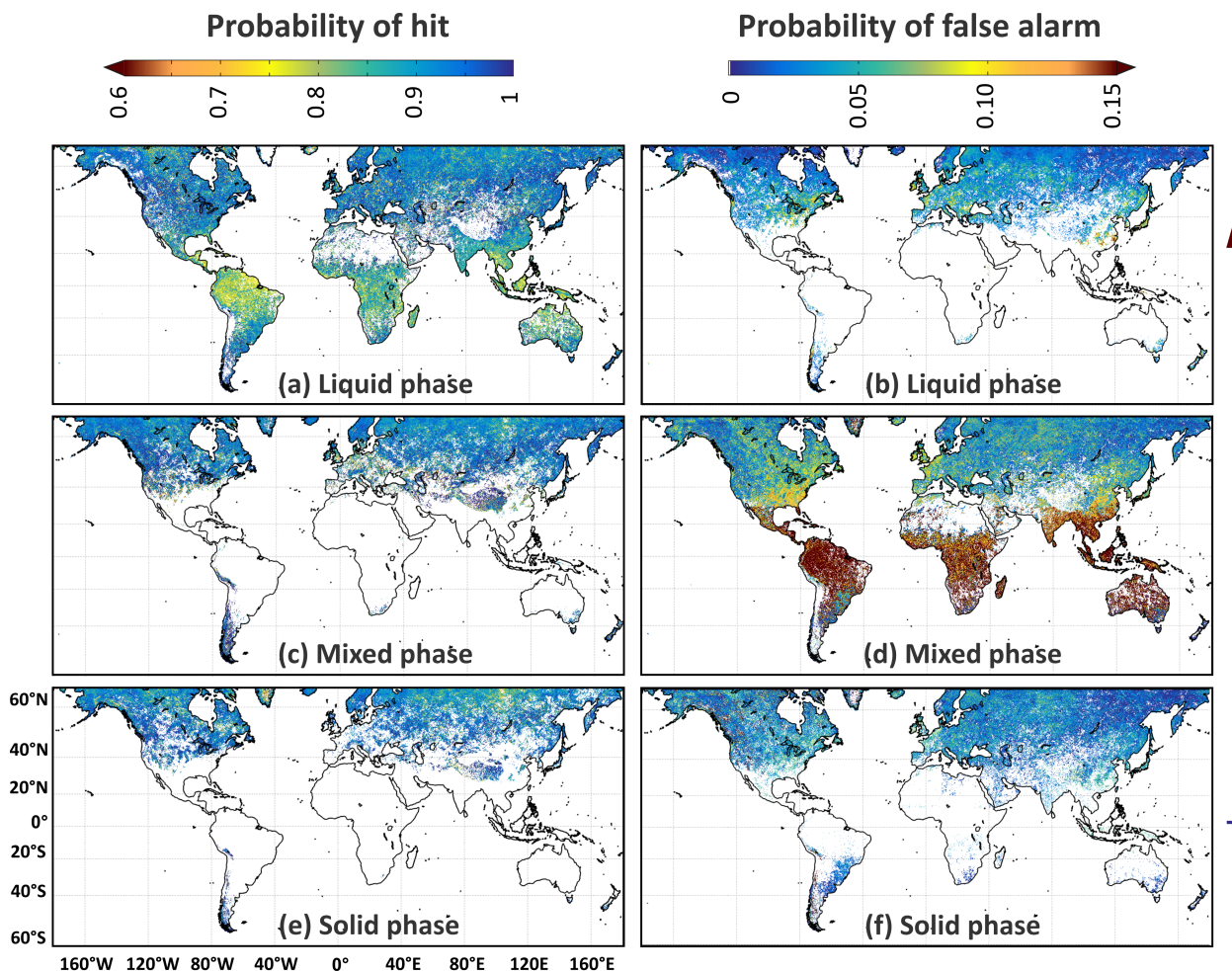


FIG. 8. The mean annual map of the probability of hit and false alarm by the KNN algorithm for the detection of the liquid phase (a, b), mixed phase (c, d) and solid phase (e, f). The results are obtained for all GPM inner-swath overpasses from June 2015 to May 2016.

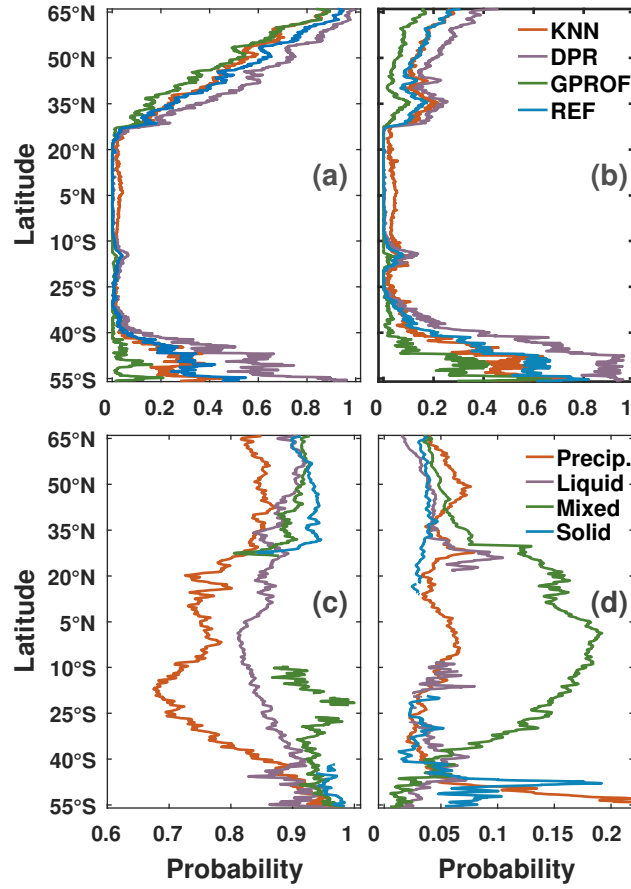


FIG. 9. Zonal mean values of the probability of precipitation phase change from liquid ($p = 0$) to solid ($p = 1$) by the KNN, DPR, GPROF, and REF products in (a) winter (November-April) and (b) summer (May-October). Zonal mean values of (c) probability of hit and (d) false alarm for the detection of the precipitation occurrence and its phase change by comparing the KNN results with the REF product.

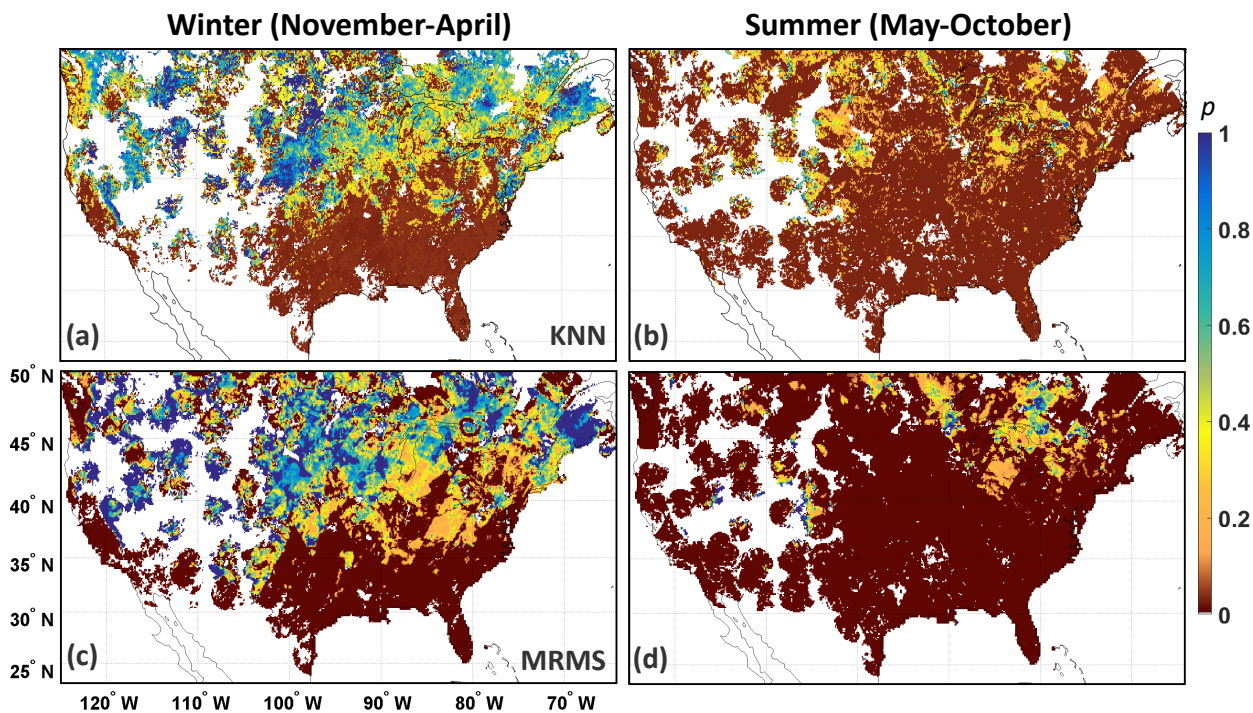


FIG. 10. Mean seasonal maps of the probability of precipitation phase change from liquid ($p = 0$) to solid ($p = 1$) for KNN in winter (a) and summer (b), and for the MRMS in winter (c) and summer (d), from June 2015 to May 2016

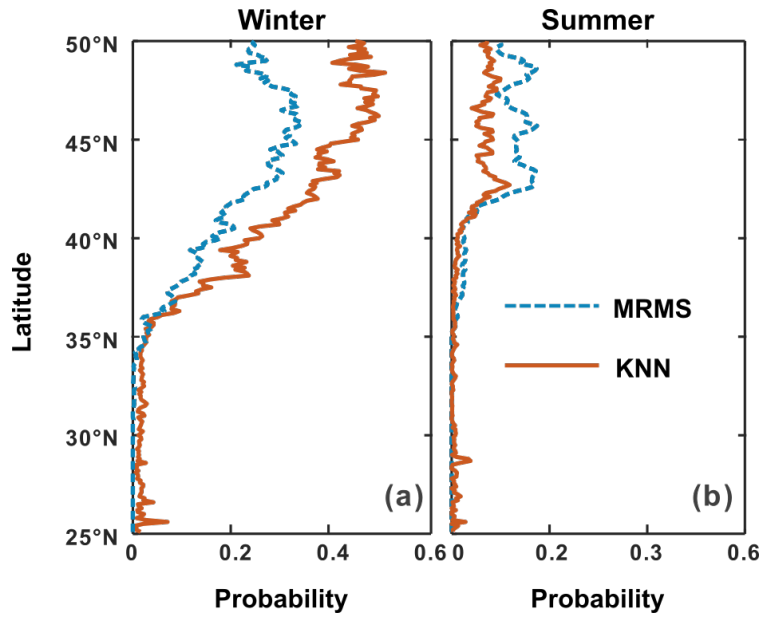


FIG. 11. The zonal mean of the probability of precipitation phase change from liquid ($p = 0$) to solid ($p = 1$) by the KNN and MRMS products in (a) winter (November-April) and (b) summer (May-October) — from June 2015 to May 2016.

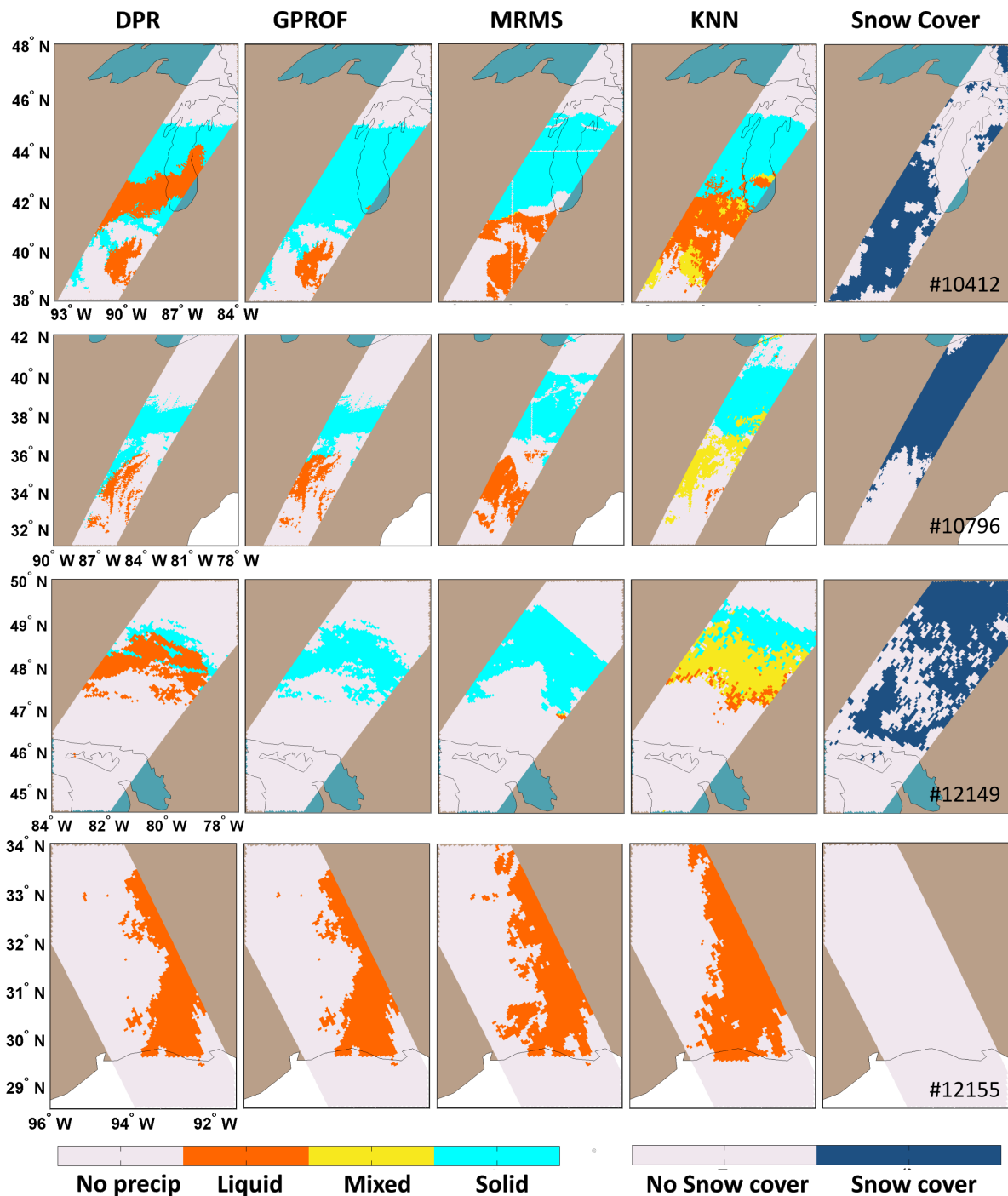


FIG. 12. Orbital-level precipitation phase detection from the KNN, DPR, GPROF, and MRMS for a few GPM overpasses including #10412 on 2015/12/28 (top row), #10796 on 2016/01/22 (second row), #12149 on 2016/04/18 (third row), and #12155 on 2016/04/18 (bottom row).

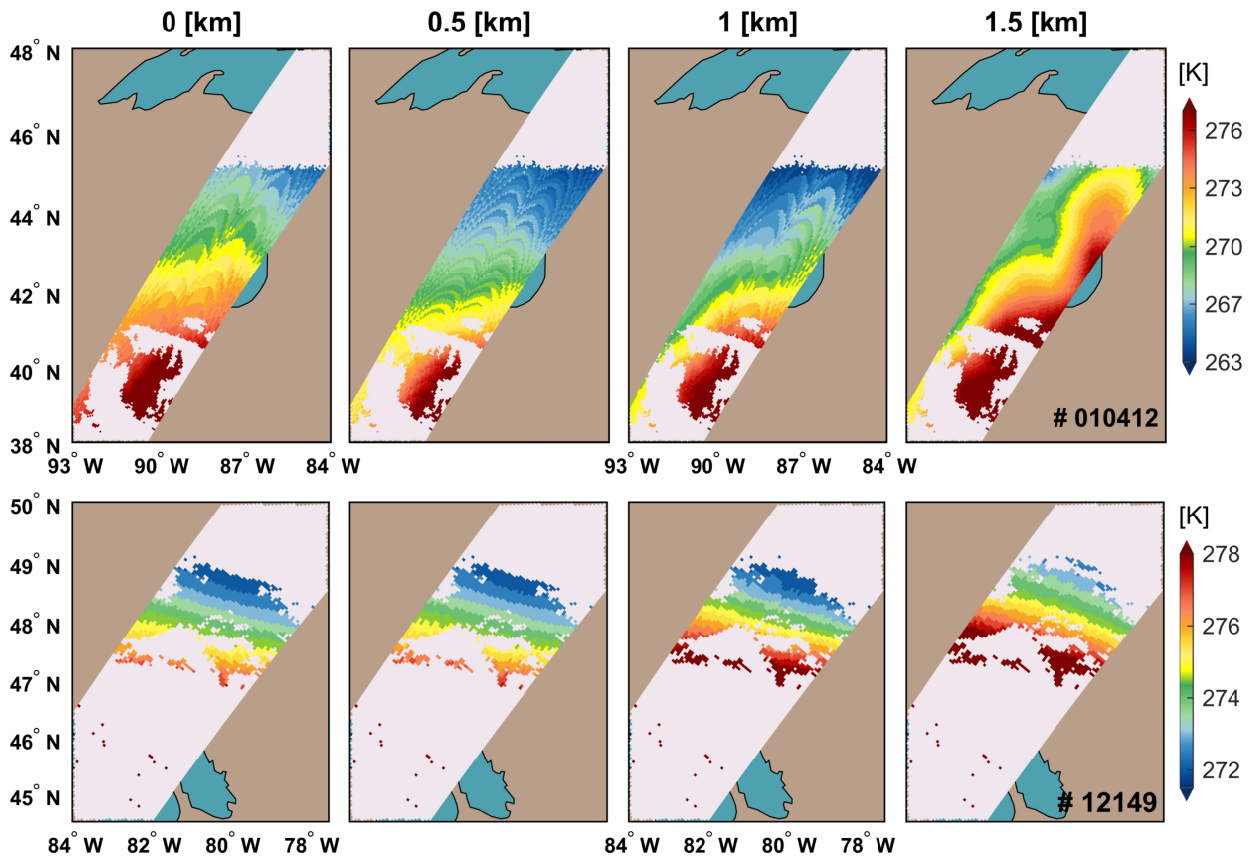


FIG. 13. Inversion of the air temperature at orbit #10412 on 2015/12/28 (top row) and orbit #12149 on 2016/04/18 (second row). The data (2A-DPRENV) are presented at four ranges from 0 (surface) to 1.5 km.

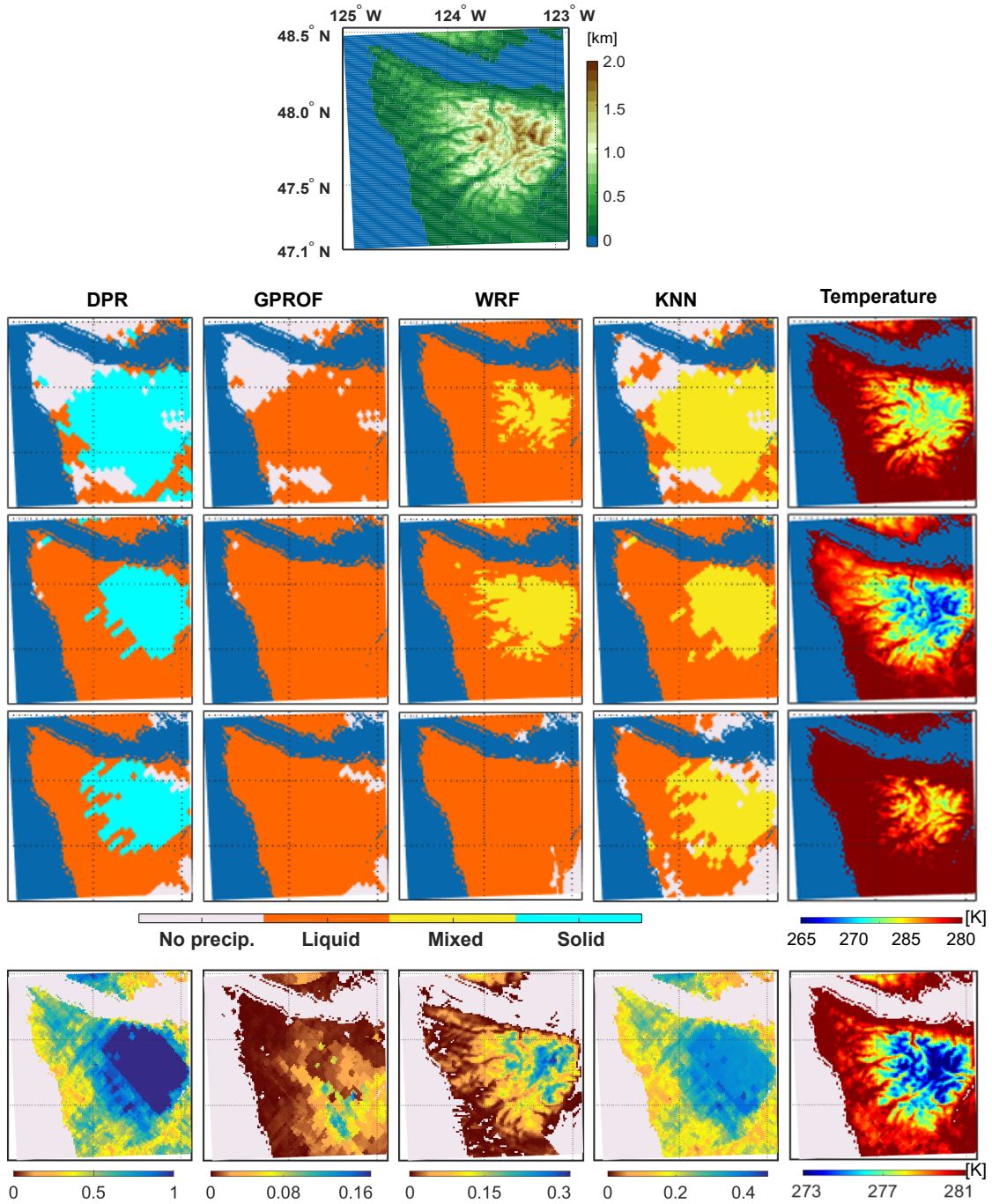


FIG. 14. The digital elevation map (DEM) of the Olympic Mountains (top panel), and the precipitation phase by the DPR, GPROF, WRF, and KNN for orbit numbers #9722 (2015/11/14, second row), #9773 (2015/11/17, third row), and #10019 (2015/12/03, fourth row). The bottom panel is the average probability of phase for 117 GPM inner-swath overpasses from November 1 to December 23, 2015. The last column shows the 2-meter air temperature from the WRF simulations.