Estimating Nonraining Surface Parameters to Assist GPM Constellation Radiometer Precipitation Algorithms

F. JOSEPH TURK AND Z. S. HADDAD
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Y. YOU
Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

(Manuscript received 10 November 2015, in final form 8 April 2016)

ABSTRACT

The joint National Aeronautics and Space Administration (NASA) and Japanese Aerospace Exploration Agency (JAXA) Global Precipitation Measurement (GPM) is a constellation mission, centered upon observations from the core satellite dual-frequency precipitation radar (DPR) and its companion passive microwave (MW) GPM Microwave Imager (GMI). One of the key challenges for GPM is how to link the information from the single DPR across all passive MW sensors in the constellation, to produce a globally consistent precipitation product. Commonly, the associated surface emissivity and environmental conditions at the satellite observation time are interpolated from ancillary data, such as global forecast models and emissivity climatology, and are used for radiative transfer simulations and cataloging/indexing the brightness temperature (TB) observations and simulations within a common MW precipitation retrieval framework.

In this manuscript, the feasibility of an update to the surface emissivity state at or near the satellite observation time, regardless of surface type, is examined for purposes of assisting these algorithms with specification of the surface and environmental conditions. Since the constellation MW radiometers routinely observe many more nonprecipitating conditions than precipitating conditions, a principal component analysis is developed from the noncloud GMI–DPR observations as a means to characterize the emissivity state vector and to consistently track the surface and environmental conditions. The method is demonstrated and applied over known complex surface conditions to probabilistically separate cloud and cloud-free scenes. The ability of the method to globally identify “self-similar” surface locations from the TB observations without requiring any ancillary knowledge of geographical location or time is demonstrated.

1. Introduction

The joint National Aeronautics and Space Administration (NASA) and Japanese Aerospace Exploration Agency (JAXA) Global Precipitation Measurement (GPM) mission is a constellation mission, whereby the observations and precipitation profile estimates from the GPM core satellite dual-frequency (Ku/Ka band) precipitation radar (DPR) and the 13-channel (10–183 GHz) passive microwave (MW) GPM Microwave Imager (GMI) serve as a reference for the other constellation radiometer-only satellites (Hou et al. 2014). While each partner radiometer may have different channel sets, resolution, or orbit, the brightness temperatures (TB) have been intercalibrated within the context of the GPM mission (Biswas et al. 2013).

One of the key challenges for GPM is how to link the information from the single DPR across all sensors in the constellation, to produce physically consistent precipitation products from all of the constellation MW radiometers (Huffman et al. 2011). The DPR observations are available alongside the GMI TB observations within approximately the middle third of the GMI swath (the GMI, DPR Ku-band, and DPR Ka-band swaths are 885, 245, and 120 km, respectively). One proposed strategy involves utilizing the DPR precipitation profile retrievals (Munchak et al. 2016; Grecu et al. 2011) within this narrow swath, to forward simulate the TB observations using MW radiative transfer models and satellite
sensor simulators (Matsui et al. 2013). Ideally, these simulations should cover the full range of the variability in the underlying atmospheric and surface conditions, with sufficient skill to represent the multichannel TB that would have been measured by actual MW radiometer observations. Once created for each sensor in the GPM constellation, the observational “database” with its associated TB is available offline as the a priori knowledge for probabilistic, Bayesian-based precipitation retrievals, for subsequent precipitation retrievals separately from each MW radiometer (Ringerud et al. 2015; Kummerow et al. 2011). This strategy places a burden on the capability to model the expected range in atmospheric conditions and fraction of precipitating and nonprecipitating scenes with sufficient realism (Petty 2013; L’Ecuyer and Stephens 2002).

Since a MW radiometer alone contains a limited amount of information from the TB observations, a further challenge is how to devise a consistent inverse (i.e., retrieval) strategy that is adaptable to the underlying atmospheric and surface conditions present at (or as close as possible to) the observation time. The environmental conditions at the satellite observation time are typically interpolated from operational global numerical weather prediction (NWP) models, such as the Japanese Meteorological Agency (JMA) global spectral model (GSM), for operational processing, and from lengthy model reanalysis, such as ERA-Interim (Dee et al. 2011), for consistent reprocessing of both GPM and its predecessor Tropical Rainfall Measuring Mission (TRMM) precipitation products. For surface emissivity at typical MW channels between 10 and 90 GHz, the Tool to Estimate Land Surface Emissivities at Microwave frequencies (TELSEM) passive microwave-based surface classification (Prigent et al. 2006; Aires et al. 2011) provides a lookup table method to interpolate the emissivity mean and variance at a specified incidence angle and frequency, using a pre-calculated 0.25° gridded monthly mean emissivity climatology derived from Special Sensor Microwave Imager (SSM/I) observations. TELSEM has been successfully implemented into the current GPROF GPM radiometer algorithm (Kummerow et al. 2015) using a 15-class emissivity index to catalog the surface in the a priori database. This ancillary model and surface class information is attached to all a priori profiles, both precipitating and nonprecipitating. With these ancillary data for the environment and the surface, the current GPROF GPM Bayesian retrieval is constrained to weight a priori candidate profiles with the same classification index, and similar surface temperature ($T_{sfc}$) and total column water vapor (TWV) as the observation (Kummerow et al. 2015), although more recent studies have suggested the use of the 2-m air temperature ($T_{2m}$) (Sims and Liu 2015). Differences between forecast models can arise between the formulation (gridpoint spacing, or wave resolution in spectral models) and its temporal resolution. Furthermore, the surface properties that can be accurately represented by any model may be different than the stated model spatial and temporal resolution (Dirmeyer et al. 2016). To minimize representativeness errors, these factors need to be considered when interpolating models to locations of satellite observations, especially across finer-scale surface features such as water–land boundaries, and locations of fronts associated with moving weather systems.

With its ±65° orbit latitude range, the GPM scenes contain many mixed conditions—inland water fraction, vegetation and forest structure, snow and ice edges that can change over the period of a few days, and overall wider seasonal extremes in air temperature, water, and land surface temperature and associated water vapor conditions—that contribute to a physically complex surface emissivity background state. In recent years, a number of modeling and observationally oriented surface emissivity studies have been carried out (Tian et al. 2015) to estimate or model the land surface emissivity structure (Ringerud et al. 2015), or its impact upon the precipitation retrieval (You et al. 2015; Carr et al. 2015). Inputs from land surface models have been used to calibrate microwave emissivity models, such as the Community Radiative Transfer Model (CRTM), over limited areas (Harrison et al. 2016). Over well-drained soil types, transient soil moisture is introduced upon the onset of intermittent rain events (Brocca et al. 2014; Turk et al. 2015). In general, the instantaneous radiometer fields of view contain unknown fractions of subpixel variability whose radiometric signatures are currently too complex to physically model for everywhere they are used within the GPM latitude range.

An earlier study (Elsaesser and Kummerow 2008) demonstrated that the surface and atmospheric properties controlling the multichannel microwave surface emissivity state over ocean (wind speed, surface temperature, and water vapor column) could model the TB with sufficient accuracy for MW radiometer retrieval applications. In this manuscript, the feasibility of a more timely update to the surface emissivity state at or near the satellite observation time is examined, regardless of surface type, for the intent of guiding or assisting the current specification of the surface and environmental conditions used for Bayesian-based radiometer precipitation retrieval algorithms. Since the constellation MW radiometers routinely observe many more nonprecipitating than precipitating conditions, the information content within the TB observations themselves is studied as a means to characterize the
emissivity state and track the associated environmental conditions. In this study, we focus on conically scanning MW radiometers with low-frequency (89 GHz and below) capabilities and near-constant Earth incidence angle. Over one full year of combined GMI–DPR data, together with the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Reichle et al. 2011) model reanalysis, was used to examine and devise relationships between the TB under ‘no cloud’ conditions (defined later), and the associated surface emissivity, surface temperature, and column water vapor. To illustrate intermodel differences, comparisons between MERRA and ERA-Interim were made. Examples are shown from time series of GPM overpasses over specific target locations across the continental United States. The ability of the method to self-identify global ‘similar’ surface emissivity and environmental conditions from the TB observations is shown for selected complex land surface conditions.

2. Methodology and extensions to GPM

The framework used is an extension of a recently published method by the authors (Turk et al. 2014a), which was based upon TRMM TMI TB and the matched PR rainfall estimates, to which the reader is referred for a detailed explanation. Figure 1 shows a flowchart of the method, and the extensions to GPM (discussed in the next paragraph) are indicated in red text.

Previous studies have demonstrated the complex relationships between physical surface parameters, such as soil type, vegetation, surface roughness, etc., and the microwave surface emissivity state (Ringerud et al. 2014; Turk et al. 2014b), and furthermore that the surface emissivity at typical radiometer bands below 100 GHz are highly correlated (Aires et al. 2011). This suggests that a relatively small number of controlling factors may be sufficient to describe the surface emissivity state. The method here does not explicitly consider these physical properties; rather, it is based upon two hypotheses. The first (top of Fig. 1) is that the physical conditions can be carried via relationships between the principal components (PC) [often denoted by empirical orthogonal function (EOF)] of the surface emissivity state vector in terms of linear and nonlinear TB combinations. If the emissivity principal components can be well represented this way, then the emissivity state vector can be well reconstructed from the TB
observations. However, this relationship would be strictly valid only when the scene is cloud free, and separating nonraining (or more generally, no-cloud conditions) from TB observations is a long-standing difficult challenge (You et al. 2015; Munchak and Skofronick-Jackson 2013; Ferraro et al. 1998; Grody 1991). Therefore, the second hypothesis is that the emissivity state vector PC elements will diverge or “separate” from their clear-scene values when applied to cloud-affected observations and thereby serve to discriminate the scene as “cloud affected” in a probabilistic fashion; that is, the threshold between clear and cloud-affected TB scenes is not a yes/no answer. Rather, it is given as a probability, which can be set to a desired minimum no-cloud detection rate. Similar probabilistic reasoning was applied by Laviola and Levizzani (2011) for delineation of raining areas in 183-GHz-based precipitation retrievals.

a. Adaptations to GPM

Fifteen months (from March 2014 through May 2015) of GMI level 1B and DPR level 2A (2A.GPM.DPR) data were merged, over all surfaces. GMI scan conically at a rate of 32 revolutions per minute, collecting data on 140° of each scan. During each scan, all 13 channels are collected for 221 locations per scan, but the swath of the four S2 channels [166 horizontal/vertical (H/V), 183 ± 3, 183 ± 8 GHz] is slightly narrower than—and not beam centered with—the nine S1 channels (10.7H/V, 18.7H/V, 23.8V, 36.5H/V, and 89H/V GHz). A nearest-neighbor search was used to locate the closest S2 location within 6 km of each S1 location. The uncorrected reflectivity profiles for the three DPR radar beams—the normal scan (NS) Ku band, the matched swath (MS) Ka band, and the high-sensitivity (HS) Ka band—were interrogated from the 3 × 3 (9 DPR beams) surrounding each GMI pixel, illustrated on the left side of Fig. 2. The NS and MS profiles each contain 176 range gates (often referred to as bins) numbered from top to bottom, spaced 250 m along the beam; the HS profile has 88 range bins spaced 500 m apart. To avoid classifying surface clutter as cloud, only radar bins less than the value of binClutterFreeBottom are examined [provided in the NASA Precipitation Processing System (PPS) 2A.GPM.DPR radar product for each of the three radars]. This approach is a trade-off and may miss some shallow low clouds near the surface. The most recent analysis has shown the NS minimum detectable reflectivity to be near 13–14 dBZ and the MS near 17–18 dBZ (Hamada and Takayabu 2016). In principle, the HS should be 6 dB more sensitive than the MS (6 dB arises from having twice the pulse length and half of the noise), but actual data reveal this value to be near 3–4 dB (R. Meneghini 2016, personal communication). Within this 3 × 3 set of beams, the number of bins within intervals of 15–20 dB, 20–25 dB, and greater than 25 dBZ were counted. The surface temperature, 2-m air temperature, and the 42-level temperature and specific humidity profile from MERRA were interpolated to the location and time of the observation. The observations where all radar bins from all three radars were < 15 dB were set aside as the no-cloud database. For these scenes, the surface emissivity at the nine S1 (10–89 GHz) channels was determined from an emission-only radiative transfer calculation (Turk et al. 2014a). The choice of MERRA as the ancillary model data was not intentional; the ancillary data could just as well have been taken from a different model. In this method, this model becomes the reference by which the TB observations are trained. Therefore, for consistency it is desirable to use the same model for the forward TB simulations and for the precipitation retrieval procedure. One feature of MERRA is that its surface temperature datasets are output on a frequent hourly update cycle, which mitigates to some extent errors introduced due to temporal interpolation of coarser (3 or 6 hourly) time updates (Aires et al. 2004), and during solar times of rapidly changing near-surface skin temperature (Norouzi et al. 2012).
The adaptations to GPM in Fig. 1 are 1) the associated MERRA surface temperature $(T_{sfc})$ and TWV are added to the nine (10–89 GHz) emissivity for an augmented (length 11) emissivity state vector. The emissivity state vector $\mathbf{e}$ can be broken down into its 11 PCs (denoted by $\mathbf{u}$) via a transformation expressed by an orthogonal matrix $\mathbf{E}$, whose columns are the eigenvectors of the emissivity covariance matrix. 2) All nonlinear TB co-terms and cross-terms were used in the estimation of each state vector PC element; that is, Eq. (1) in Turk et al. (2014a) for the first PC element was replaced with

$$u_i = a_0 + \sum_{i=1}^{N} b_i T_{h,i} + \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} T_{h,i} T_{h,j} + \sum_{i=1}^{3} d_i PD$$

$$N = 9 \text{ or } 11 \text{ (GMI), } 7 \text{ (SSMIS)} \text{ (1)}$$

(similar for the other 10 PC elements), where three PD terms refer to the polarization difference $PD = (T_{h,B} - T_{h,V})(T_{h,B} + T_{h,H})$ at 10, 18, and 36 GHz. The term $N = 9$ refers to the nine S1 channels, and $N = 11$ when the 166H/V-GHz channels are added.

The efficacy of this methodology will also be examined with one other MW sensor in the GPM constellation, the Special Sensor Microwave Imager/Sounder (SSMIS), which has been operating since 2003 on (currently) four Defense Meteorological Satellite Program (DMSP) satellites ($F16, 17, 18,$ and $19$). This sensor does not have 10-GHz capabilities, and only seven channels (19–85 GHz) and two polarization differences (19 and 37 GHz) will be used in Eq. (1). Since these satellites do not have DPR-like precipitation radar capabilities, all 15-min orbit coincidences between the GPM and $F17$ satellites were processed, and for each SSMIS 19-GHz pixel location, the nearby $3 \times 3$ DPR radar profiles were identified and processed using the same procedures in Figs. 1 and 2. Obviously this will not provide the same data density as GMI–DPR, but over 1700 orbit intersections were identified up to the beginning of May 2015, and owing to GPM’s asynchronous orbit provided approximately $10^7$ SSMIS–DPR points spread across the globe.

No information from the DPR rain retrieval was used, since the intention is to use the native (not corrected for attenuation) radar observables (total number of bins exceeding progressively increasing reflectivity thresholds) as a proxy for discerning gradually “nonclear” column conditions within the radiometer view, rather than separating by a retrieved rain rate at the surface. Any clouds whose intrinsic reflectivity falls below the sensitivity of the DPR will be undetected; therefore “clear” scenes are defined relative to the low-end sensitivity of the DPR radar beams. Some nonprecipitating cloud water will likely be present in many scenes and whose vertical profile is unknown. On the other hand, this methodology wraps the effect of undetected cloud water into the estimation of the emissivity state. This may have an unintended consequence of mitigating uncertainties in the corrections for the attenuation due to the vertical location and amount of nonprecipitating cloud water, which is typically parameterized from cloud-resolving model simulations (Iguchi et al. 2009), and the competition to the net upwelling TB (Bennartz and Bauer 2003; Muller et al. 1994).

b. Discrimination of the surface emissivity state vector

The top row of Fig. 3 shows 2D histograms between the actual 36H-GHz emissivity, the associated MERRA $T_{sfc}$ and TWV from all observations in the cloud-free database, and their corresponding values emissivity state vector estimates, everywhere without any regard to surface, using the nine GMI S1 channels (10–89 GHz) in the regression equations for each PC [Eq. (1)]. The low and high maxima in the 36H emissivity correspond to the overwater and overland cases, respectively, with coastal and other mixed pixels in between. The best agreement in surface temperature is above 273 K and degrades for increasingly cold overland conditions. For TWV, the agreement is very good for values below 20 mm, after which the points split into two branches, mainly the overwater scenes close to the diagonal axis and a “knee” for the mainly warm land surfaces, which are biased low as TWV increases. In an attempt to improve the TWV performance, Eq. (1) was extended by adding in the 166H/V channels (i.e., $N = 11$). The results are shown in the middle row of Fig. 3, where the overall agreement at the high TWV values is slightly improved, and the corresponding change to the emissivity (not all shown) was negligible. The bottom row of Fig. 3 shows the histograms for these same parameters when the emissivity state vector PCs are estimated from the seven-channel SSMIS TB observations. Despite the reduced data density from the GPM–$F17$ coincidences, the methodology appears to reproduce the 36H emissivity and $T_{sfc}$ with slightly increased error, but it fares considerably worse for TWV (the knee for the overland surfaces is much more evident, suggesting the importance of 10-GHz capability for estimating land surface parameters).

As shown in the middle of Fig. 1, the discrimination is based upon intentionally applying the PC estimation coefficients (which were derived purely from the no-cloud database) to the cloud-affected scenes. As the overall level of cloud contamination increases, the PCs (and the reconstructed emissivity state vector) should therefore take on unrealistic values not represented in
the clear-scene database, and any resultant displacement in the PC structure may be exploited to design a discrimination between the no-cloud and cloud-affected scenes (Turk et al. 2014a). The top row of Fig. 4 shows how the histograms of the fourth, sixth, and seventh PCs are distributed when Eq. (1) is applied to both databases, over all surfaces (of the 11 total PCs, these best separated the two populations, the others had little separation and are not shown). The histograms of the clear-scene data are shown in black, and the histograms of the cloudy-scene data are shown in a series of colors, where each line color indicates a gradually increasing threshold on the total number ($N$) of NS-Ku-band reflectivity bins exceeding 20 dB, as illustrated on the right side of Fig. 2. The red line is for a threshold of $N > 0$, the loosest threshold in the sense that this cloud-affected population contains all of the heavy cloud/precipitation scenes and many of the cloudy/weak or nonprecipitating scenes. Since only a single reflectivity bin exceeding 20 dB is needed, there is practically no separation from the clear scenes. The orange line represents $N > 20$ bins exceeding 20 dB, which improves the separation. The blue line is a threshold of $N > 100$ bins, the tightest threshold where cloudy/weak scenes are removed, leaving only the heavier cloud/precipitating scenes. Since this $N > 100$ threshold represents substantial clouds within the radiometer view, there is still some PC overlap but also significant separation outside of the
clear-scene ranges. Using these PCs (4, 6, and 7) in a linear discriminant to optimize the separation of the clear and cloud-affected scenes [using the same procedure described in Turk et al. (2014a)], the bottom row of Fig. 4 shows the resultant relative operating curves (ROC) for each of these same thresholds, for (left to right) the 9-channel (10–89 GHz) case, the 11-channel (10–166 GHz) case, and the 7-channel (19–85 GHz) SSMIS case. An ROC curve expresses the hit rate [or probability of detection (POD)] as a function of the false alarm rate, for different values of a threshold or “discrimination” term (Wilks 2006). Since the separation of no-cloud and cloud-affected scenes is a fuzzy separation, the optimal value for the discriminant lies near the knee of each of these curves, where the POD (no cloud) is sufficiently high at the expense of a reasonably low corresponding false alarm rate. In the Turk et. al. (2014) formulation, the linear discriminant is expressed as a linear combination of the observed TB; so, once an operating curve is chosen, the corresponding POD (no cloud) is found by calculating the discriminant from the observed TB and locating where the value lies along the ROC curve.

Note that the discrimination performance gets slightly worse for the $N = 11$ case, even though the $N = 11$ case was previously shown to improve the estimation of TWV (Fig. 3). The reason for this is not clear, but it may be because for most extreme cloudy/rainy atmospheric conditions, the surface is opaque to the 166-GHz channels and thereby these channels are essentially adding noise to the estimation of the surface emissivity principal components. Since the nine (10–89 GHz) GMI channels are available across the full swath of GMI and also for AMSR-2, for the remainder of this manuscript the discrimination of no-cloud conditions is set whenever the POD (no cloud) > 0.9, using the ROC curve.
based on a 20-dB threshold value of $N > 50$ bins and the $N = 9$ formulation.

3. Time series at different surfaces

To demonstrate, the technique described above was applied to GPM overpasses that covered specific locations, for one full year beginning in May 2014. All examples use the identical figure layout for ease of comparison. The first example is an overwater case where the technique is expected to perform the best, followed by lightly vegetated land, open/frozen lake, forest snow cover, and seasonal wetland.

a. Gulf of Mexico

The top panel of Fig. 5 depicts the time series of the GMI 10H-, 89H-, and 166H-GHz TB observations (black, green, and red points, respectively) whenever the closest GMI pixel was located within 6-km distance of a point 50 km offshore of Pensacola, Florida (29.9°N 87.2°W), which is over water but within range of the coastal NEXRAD. The blue impulses represent the 1-h rainfall accumulations (mm) from the NEXRAD Multi-Radar Multi-Source (MRMS) QPE radar product (Zhang et al. 2011). There are several overpasses in the summer months during strong convective rain conditions when the 166H-GHz TB < 89H-GHz TB, and in the winter months the 89H-GHz TB exhibits a wider range of variability due to frequent southerly surges of dry, cold air. Figure 5b plots the POD(no cloud) based on the discriminant for each observation, where values above and below a value of 0.9 are shown in black and red, respectively (red indicating reduced confidence that the scene is cloud free, and therefore the estimate of the emissivity state vector is to be disregarded or treated with caution). These points appear to be consistent with the presence of rain at the overpass time, but it is difficult to discern whether the overpass occurred in between rain events, or whether shallow rain was present that was missed by the MRMS processing.

To illustrate the differences between two models, Fig. 5c shows the estimated $T_{sfc}$ (red points) plotted on top of the corresponding values from the MERRA (black, every hour) and ERA-Interim (blue, every 6 h) model reanalysis. In this case the two models agreed nearly perfectly, and the estimated $T_{sfc}$ followed the overall trend. The red points with open circles correspond to low confidence [red points in Fig. 5b, where POD(no clouds) < 0.9], which are generally the outlier values. Figures 5d and 5e show the difference between $T_{sfc}$ and the 2-m air temperature, and the TWV for these same two models, which shows the periods of cold, dry air surges several times during November 2015–March 2015, when the 2-m air temperature was 15° or more degrees colder than the surface. From Fig. 5e, the TWV agree very closely between the two models, and the estimated values (red points) closely track the models, across all seasons. Figure 5f shows the estimated 10H-GHz (black points) and 89H-GHz (red points) emissivity, where the open circles indicate reduced confidence in the no-cloud conditions highlighted in Fig. 5b.

These results are neither surprising nor unique; simple statistical techniques based on SSMI channel combinations have been long demonstrated for overocean water vapor (Alishouse et al. 1990), and the associated radiative transfer models are well developed (Meissner and Wentz 2012). Rather, the emissivity state vector estimation will be useful if it can be demonstrated to equally apply to more complex surfaces, without the need for land/sea masks or geographical considerations.

b. West Texas

Figure 6 shows the identical plot format as Fig. 5, except it represents GMI overpasses and model data just west of Lubbock, Texas (33.4°N 102°W). The early summer rains during May 2014 are evident (Fig. 6a), including cases where the 10H-GHz channel TB is radiometrically colder than the 89H-GHz channel, likely owing to reduced emissivity and increased soil moisture in the upper 2 cm, which is the range of in situ–measured skin depths (Owe and Van de Griend 1998). Overpasses where POD(no cloud) < 0.9 appear to be coincident with rain at the overpass time. A close-up of the May 2014 period (not shown) shows that the majority of these cases occurred when overpass time happened during rain events. In general, the estimated $T_{sfc}$ and TWV track the seasonal model trends, and the diurnal range in $T_{sfc}$ often exceeds 25 K when the daily ascending/descending GPM local observing times (which repeat about every 60 days at this latitude) occur near the early morning and late afternoon. Over this type of surface and during dry periods, the surface emissivity at 10H and 89H GHz is fairly steady near 0.9, but a rapid decrease in the 10H-GHz emissivity is detected following individual rain events in May and early June, followed by a gradual increase during the subsequent dry-down period. The estimated $T_{sfc}$ also tracks well the diurnal $T_{sfc}$ trends in both models, but for TWV there are overpasses where the estimated TWV is biased either high or low by about 25% relative to these two models. There are also periods when MERRA and ERA-Interim TWV differ by about 15%–20%, and the ERA-Interim $T_{sfc}$–$T_{2m}$ difference is often 5–10 K larger than MERRA during the warmest periods of each day. Therefore, the variability in the retrieved $T_{sfc}$ is not that much larger than the intermodel variability. This ability to capture the rain-induced
FIG. 5. One year (May 2014–May 2015) time series of GPM overpasses over a region south of Pensacola (29.9°N, 87.2°W). (a) GMI 10-GHz TB (black), 89H (green), and 166H (red). The blue impulses are the 1-hourly accumulations (mm) from the NEXRAD MRMS (negative values indicate periods of missing NEXRAD data); (b),(f) As in (a). (b) POD of no-cloud conditions, where black and red colors indicate a value above and below a value of 0.9, respectively. (c) Surface temperature $T_{sfc}$ from the emissivity state vector estimation (red points), where the values associated with POD(no cloud) < 0.9 are shown as open circles. The lines indicate the values interpolated from MERRA (black) and ERA-Interim (green). (d) Difference between $T_{sfc}$ and the 2-m air temperature (K) from the MERRA (black) and ERA-Interim (green) models. (e) TWV (mm) from the emissivity state vector estimation (red points), where the values associated with POD(no cloud) < 0.9 are shown as open circles. The lines indicate the TWV values interpolated from MERRA (black) and ERA-Interim (green). (f) 10H- (black) and 89H-GHz (red) emissivity from the emissivity state vector estimation, where the values associated with POD(no-cloud) < 0.9 are shown as open circles.
FIG. 6. As in Fig. 5, but for a region west of Lubbock (33.4°N, 102.0°W).
Fig. 7. As in Fig. 5, but for a region in Lake Superior (48.0°N, 87.3°W).
dynamic emissivity change in this type of low-vegetation, well-drained soil, using the same PC and discriminant coefficients as the overwater case above, is an encouraging result for the overall robustness of the emissivity state vector estimation over both water and land surfaces.

c. Lake Superior

Figure 7 shows the identical plot format as Fig. 5, except it represents GMI overpasses and model data in the middle of Lake Superior (48.0°N, 87.3°W), which is a large enough water body to avoid coastal land contamination in the largest (10 GHz) field of view, and a location that freezes during winter on average about 80%. The open water season is noted from the cold 10H-GHz TB, except for a 3-month period between February and April when the lake experienced freezing conditions. By coincidence, 2014 and 2015 were anomalous years when the lake was nearly entirely frozen by late February. The typically annual precipitation in this location is about 750 mm with persistent drizzle conditions. While such generally persistent light rain conditions are evident from the MRMS values in Fig. 7b (note the change in scale to 0–5 mm h⁻¹), this region is about 150 km offshore of the nearest NEXRAD (KMQT, Marquette, Michigan) and shallow rain or snow events may be misrepresented. Nonetheless, during the open water period, the lower confidence [POD(no cloud) < 0.9] is well represented with periods of solid precipitation, although MRMS likely underreports the freezing precipitation in winter months. Beginning with the onset of freezing, the confidence drops to a sustained value near 0.6. The emissivity of the surface quickly rises and falls over the course of just a few days, with the onset and end of the freeze period. Although many of these values are flagged as low confidence of no-cloud conditions, the values do appear reasonable for ice and snow cover (Willmes et al. 2014), but more extensive tests over these surface conditions are needed.

The two models agree very well in TWV, but they show marked differences in Tₛfc (MERRA is larger than ERA-Interim in summer months, and the opposite in winter months), as well as the Tₛfc–T₂₉m difference. This difference has been suggested as one means to separate the precipitation phase (Sims and Liu 2015), whether it is positive or negative gives different interpretations. While this could be due to different model physics, another possibility is the model resolution referred to in section 1. At the horizontal scale of these models, which have different grid postings (0.75° for ERA-Interim, and 0.5° × 0.66° for MERRA), one model could have a very different land/water (or land/ice) fraction than the other. This illustrates the limitations of interpolating ancillary model data near surface edges, especially water/land, in identifying the surface environmental conditions. Use of a consistent model throughout simulations and retrievals would alleviate some of the difference.

d. Northern Minnesota forest

Figure 8 shows the identical plot format as Fig. 5, except it represents GMI overpasses and model data in a northern Minnesota fairly dense coniferous forest area (47.7°N, 92.3°W). This surface is characterized by similar and fairly stable TB between 10 and 89 GHz throughout the year, punctuated by occasional summertime convective rain events and winter snow cover. This decreases the 89-GHz TB owing to scattering processes within the snow grain media, which are modulated by the dry, cold environmental conditions (Markus et al. 2006). For these conditions, the no-cloud discrimination setting of 0.9 at the NKu > 50 bins level is too high to capture the many no-cloud overpasses. As in section 3c, this shows the challenge in discriminating cloud/no-cloud scenes from this type of forest/snow conditions (Shahroudi and Rossow 2014; Cordisco et al 2006). Since the discrimination is formulated probabilistic, it will operate at a reduced confidence. One could optionally operate at the tighter NKu > 100 bins threshold (Fig. 4), which would identify the heavier cloud conditions but also assign a larger number of cloud-affected scenes as cloud free (increased false detection). Even so, the estimated emissivity state vector values for Tₛfc and TWV track the model values very well, even into the winter season with Tₛfc near 240 K, and the decrease in the 89H emissivity values appear reasonable for snow cover. As above, more examples are needed before quantitative conclusions can be made. In this case the two models agree extremely well in Tₛfc and TWV, but the MERRA Tₛfc–T₂₉m difference is occasionally 10 K smaller than ERA-Interim in winter months.

Other forest regions in the western and eastern United States were studied and are not shown. More open forest areas exhibited increased variability in emissivity from overpass to overpass, depending upon the canopy density and fraction of tree cover within the radiometer fields of view. For example, the dense deciduous forest found in the Appalachian Mountains in western North Carolina exhibited a steady emissivity near 0.95 regardless of frequency or rainfall.

e. Meghna seasonal wetland

Figure 9 shows the identical plot format as Fig. 5, except it represents GMI overpasses and model data within a seasonal wetland in the “haor” area of the Meghna River (24.5°N, 91.0°E). During monsoon onset, it becomes a large sediment-filled water body with a rich stock of fish, almost like a freshwater lake. As the
FIG. 8. As in Fig. 5, but for a region in northern Minnesota (47.7°N, 92.3°W).
FIG. 9. As in Fig. 5, but for a region in the Ganges–Brahmaputra–Meghna delta (24.5°N, 91.0°E). No MRMS data are available in this area.
monsoon recedes, it reverts back to its dry state. The seasonal river flows are apparent in the trend of the 10H-GHz TB, as the location gradually transitions from fairly dry land in May to surface water by July and slowly back to land after November. The use of passive MW observations for estimating the extent of inundated land has a long heritage (Salvia et al. 2011). Since this area is mainly sediment-filled freshwater (and not saline ocean water as in Fig. 5, which had 10H TB near 90 K for similar T_{sfc} values), there are correspondingly different emissivity values. During the summer months from June to October, the monsoon-related convective rain events are apparent from the radiometrically cold 89H and 166H TB values, and the discrimination (Fig. 9b) identifies these events, and also likely nonconvective rain events, although there are no ground radar data available to verify. In this example, the estimated emissivity state vector tracked the model data better during the October–May months, capturing a brief period of warm weather in early January. In the peak summer months (June–September), the estimated values were biased $\approx$5 K high in T_{sfc} and $\approx$10 mm low in TWV, relative to either model. Even for the events where the POD (no cloud) fell to 0.8, the 10H and 89H emissivity tracked the changing land conditions. Over the course of May alone, the 10H emissivity dropped from 0.9 to 0.6, and in June from 0.6 to 0.3. This is nowhere as dynamic as the rapid rain-induced emissivity changes shown in Fig. 6, but it illustrates that monthly emissivity climatology may underrepresent the surface change for many coastal, wetland, or seasonal water/land mixed pixels regions.

4. Identification of self-similar surface conditions

The previous section demonstrated the ability to track the joint surface emissivity and environmental conditions over a range of expected surface conditions using the estimated emissivity state principal component structure, with varying levels of skill depending upon the scene type and acceptable level of the false alarm ratio. Since the emissivity state vector is estimated from the observed no-cloud TB via a few simple matrix transformations, this suggests that the PC structure may be useful to assist a priori database indexing and selection criteria discussed in section 1. For example, in the current GPM radiometer algorithm, given a TB observation and surface class index, it is not clear how far to “extend” the a priori search in “nearby” T_{sfc} and TWV space, or to extend the search to other surface classes if needed, to find a suitable number of candidate solutions (Kummerow et al. 2015).

To demonstrate how the formulation described here could be used, a single GMI observation was extracted from Fig. 5, taken at 0215 UTC 17 June 2014. The associated T_{sfc} and TWV interpolated from MERRA were 301.2 K and 36.9 mm, respectively. The clear-scan database was scanned for entries with a self-similar principal component structure by computing a simple inverse standard deviation–weighted distance metric $d$ for each entry,

$$d = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{(u_{i}^{obs} - u_{i}^{sim})/\sigma_{i}^{sim}}{2} \right]^{1/2} \right\},$$

where $u_{i}^{obs}$ and $u_{i}^{sim}$ are the $i$th principal components of the observation and the database entries, respectively; $\sigma_{i}^{sim}$ is the $i$th associated standard deviation of the database; and $N$ is the number of leading principal components to compare ($N = 3$ accounts for $\approx$99% of the emissivity variability) and then sort the values from smallest to largest and retain the values where $d < 0.25$. The geographical locations of the resultant nearby entries are plotted in the top panel of Fig. 10, where the 3-month interval of the points is indicated by color code (blue = DJF, green = MAM, red = JJA, gold = SON, where the capital letters indicate the first letter of each month). Most notably, the candidate entries appear at self-similar locations on a map, located over similarly warm ocean temperature locations, regardless of hemisphere. In the three bottom panels of Fig. 10, the 2D histograms of TWV versus T_{sfc}, $e_{37H}$ versus $e_{10H}$, and $e_{89H}$ versus $e_{37H}$ are shown, left to right. These panels show that the majority of the selected candidate profiles cover a wider range in TWV (25–40 mm) than in T_{sfc} (295–302 K), with corresponding 37H emissivity between 0.25 and 0.35.

Figure 11 demonstrates another example, using a GMI observation extracted from Fig. 8, taken at 1425 UTC 3 March 2015. The associated T_{sfc} and TWV interpolated from MERRA were 260.9 K and 2.5 mm, respectively, representing below-freezing surface conditions. Using the same database search as above, the nearby entries in PC space cluster themselves in winter Northern Hemisphere months (blue), where snow cover is common, including mountain ranges; a small number of locations in the Andes in the Southern Hemisphere winter months (red); and the Himalayas, where snow is present year-round. In this case, the associated range of variability is larger in T_{sfc} (240–275) than in TWV (less than 10 mm). Figure 12 shows the locations of self-similar emissivity vector conditions, using an observation from Fig. 6 on 21 August 2014, where the associated T_{sfc} and TWV are 295.9 K and 29.8 mm, respectively. The entries cluster themselves in similar soil conditions in the western United States and Mexico, much of the Indian Peninsula, and known dry regions in Brazil and Africa. The associated range in T_{sfc} (295–302 K) is smaller than the range observed during the snow-covered month of March.
lies between 285 and 305 K and a wide extent in TWV is clustered between 15 and 50 mm.

It is emphasized that throughout this entire manuscript, the only instances where the latitude, longitude, or observation time was consulted was to plot points on a map. The information that can be extracted from the no-cloud TB observations may provide meaningful information to guide the direction and extent of the a priori database search. Furthermore, by performing the search in PC space, each independent PC in Eq. (2) is weighted by the associated standard deviation, rather than by largely ad hoc channel weights that are used when searching in TB space.

5. Application to individual radiometer overpasses

Figures 13a and 13b show the observed 10H- and 89H-GHz TB from the GPM overpass over the central United States, centered on Oklahoma and Texas at 0252 UTC 11 May 2015. Several regions of reduced TB are noted over regions affected by the persistent rains that fell in this area throughout May–June. The associated emissivity for these channels estimated from the state vector method is shown in Figs. 13c and 13d and is blanked out (white color) in the areas where the discriminant analysis indicated POD(no cloud) < 0.9. Using the emissivity values in clear-scene TB simulations
(section 2a), the simulated 10H and 89H TB are shown in Figs. 13e and 13f, respectively. Simulations are not performed in the regions where the POD(no cloud) < 0.9 (the TB simulations would not make sense without including the scattering properties of the hydrometeor profiles). Figures 13c and 13d show the same at 89 GHz, which shows good correspondence with the discrimination at this level and the regions of radiometrically cold 89-GHz TB.

In the bottom row, Figs. 13i and 13j depict \( T_{sfc} \) and TWV interpolated from MERRA to this observation time. Figures 13k and 13l are the associated depictions from the emissivity state vector, except in areas where POD(no cloud) < 0.9. The estimated patterns are noisy, and the higher values of TWV are underestimated relative to the model (consistent with the TWV variability shown in Fig. 3), but the overall environmental patterns are captured.

As an example of current research, examination of the day-to-day emissivity trends in Figs. 5–9 suggests that a realistic estimate of the current emissivity under the “cloudy” conditions might often be obtained by looking backward in time for the most recent TB observations that “passed” the no-cloud discrimination test. Using a computationally fast daily lookup table approach to search GMI observations backward in time for any global location, Figs. 13g and 13h display the 10H and 89H surface emissivity, respectively, where the missing areas are filled in with the first no-cloud observation encountered (the look-back is currently set to 5 days maximum). While the resultant patterns appear realistic, there is no way to further evaluate this approach.
without introducing the hydrometeor profiles in the TB simulations, which is beyond the scope of this study. In this case, the backward lookup did not attempt to separate the previous observations by ascending and descending nodes, which view the earth under very different $T_{sfc}$ and TWV conditions. This method would fail to capture very fast changes to surface emissivity from rain events that may have occurred between overpasses, but nonetheless it may be a good initial estimate from which to vary using the emissivity range patterns demonstrated in Figs. 10b, 11b, and 12b.

6. Conclusions

In this manuscript, the feasibility of a more timely update to the surface emissivity state at or near the satellite observation time was examined, for the purpose of assisting the specification of the surface and environmental conditions used for MW radiometer precipitation retrieval algorithms. The method is designed around estimation of the principal components of the MW surface emissivity state vector [defined by the emissivity at the nine channels (10–89 GHz), the surface temperature, and the total column water vapor] from the TB observations over any surface, trained on global combined GMI–DPR observations and NASA MERRA to specify the associated surface and environmental conditions. The model choice is important since it is the basis to which the method is trained, and it attempts to replicate from the “no cloud” TB observations. The number of clouds detected within the profile of the three DPR radar beams (NS-Ku, MS-Ka, and HS-Ka bands) was used to
probabilistically separate cloud and no-cloud conditions. Therefore, no-cloud scenes were defined relative to the low-end sensitivity of the DPR radar beams, wrapping the effect of any undetected cloud water directly into the estimation of the emissivity state. While the quantitative effect of this is currently unknown, it may have the unintended benefit of mitigating uncertainties in how nonprecipitating cloud water is apportioned in TB simulations. Comparisons between MERRA and ERA-Interim were illustrated out to explain the differences between the estimated state vectors, and to emphasize the need for a consistent model in both the simulations and during the inversion, especially over land/water/snow edges, such as inland lakes and seasonal wetlands.

The technique demonstrated its ability to identify global “self-similar” surface emissivity and environmental conditions by searching for nearby observations in principal component space. The observations obtained this way provide a measure of the natural intervariability among the emissivity, surface temperature, and total water vapor conditions. Knowledge of this joint variability may be useful to assist existing retrieval methods in directing the extent of their a priori database search for candidates with similar joint surface emissivity, surface temperature, and column water vapor conditions. Since the emissivity state vector estimated this way is strictly valid for no-cloud scenes, when a cloud-affected scene is encountered, an efficient reverse-time

Fig. 13. (a),(b) Observed 10H- and 89H-GHz TB from the GPM overpass over the central United States at 0252 UTC 11 May 2015. (c),(d) Estimated 10H and 89H emissivity, where the unfilled white regions within the swath indicate locations where POD (no cloud) < 0.9. (e),(f) Clear-scene (no hydrometeors) 10H and 89H TB simulations using the emissivity values shown in (c),(d). No simulations are carried out for regions where POD (no cloud) < 0.9. (g),(h) As in (c),(d), except the unfilled locations within the swath are replaced with the emissivity calculated from the closest previous-time GMI overpass where POD (no cloud) > 0.9. (i) The term $T_{sfc}$ and (j) TWV interpolated from MERRA to the satellite overpass time. (k) The term $T_{sfc}$ and (l) TWV estimated from the emissivity state vector, where the unfilled white regions within the swath indicate locations where POD (no cloud) < 0.9.
search was studied to locate the most recent no-cloud TB observation at the desired no-cloud discrimination threshold level. While this would be self-consistent and timelier than falling back to climatology, it would miss emissivity change owing to precipitation in between overpasses. However, more quantitative radiative simulations, including the proper specification of the hydrometeor profiles, would need to be carried out before any conclusions can be made.

**Acknowledgments.** The authors acknowledge support through the NASA Precipitation Measurement Missions (PMM) science team. GPM data were obtained via the NASA Precipitation Processing System (PPS), and MERRA data from the Modeling and Assimilation Data and Information Services Center (MDISC). The authors would like to acknowledge their colleagues from the PMM Land Surface Working Group, and Wes Berg, Mark Kulie, Pierre Kirstetter, Robert Meneghini, and Faisal Hussain for their constructive comments. The work of FJT and Z.S.H. was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

**REFERENCES**


