Interpreting mm-wave radiances over tropical convective clouds

Z.S. Haddad,[1,3] R.C. Sawaya,[2] S. Kacimi,[3] O.O. Sy,[1] F.J. Turk,[1] J. Steward[1,3]

Abstract. Few systematic attempts to interpret the measurements of mm-wave radiometers over clouds and precipitation have been made to date because the scattering signatures of hydrometeors at these frequencies are very difficult to model. The few algorithms that have been developed try to retrieve surface precipitation, to which the observations are partially correlated but not directly sensitive. In fact, over deep clouds, mm-wave radiometers are most sensitive to the scattering from solid hydrometeors within the upper levels of the cloud. In addition, mm-wave radiometers have a definite advantage over the lower-frequency window-channel radiometers in that they have finer resolution and can therefore explicitly resolve deep convection. The preliminary analyses (in particular of the NOAA MHS sounder brightness temperatures) summarized here indicate that the measurements are indeed very sensitive to the depth and intensity of convection. The challenge is to derive a robust approach to make quantitative estimates of the convection directly from the mm-wave observations, and conversely to derive a robust forward representation of the dependence of the radiances on the underlying moisture fields, to enable effective data assimilation. This is accomplished using a two-step semi-empirial approach: we first use the substantial amount of nearly-simultaneous coincident observations by mm-wave radiometers and orbiting atmospheric profiling radars in order to enforce unbiased consistency between the calculated brightness temperatures and the radar and radiometer observations; the second step is to explain the departure from the first-step mean empirical relations, in terms of the moisture variables, using cloud-resolving simulations with different microphysical schemes, including an original microphysical representation whose consistence with the remote sensing observations is superior to that of the other schemes. The results are a retrieval approach and a representation of the forward observation operator that are unbiased by construction and that quantify their uncertainties using the corresponding conditional variances.

1. Introduction

Very few algorithms to interpret the measurements of mm-wave radiometers over clouds and precipitation have been proposed to date because the scattering signatures of hydrometeors at these frequencies are very difficult to model. The few algorithms that have been developed, most notably D. Staelin's neural-network approach (*Staelin and Chen*, 2000), and a version of the Bayesian GPROF algorithm for the Global Precipitation Measurement (GPM) mission adapted to the GPM sounders (*Kidd et al.*, 2015), try to retrieve surface precipitation, to which the observations are partially correlated but not directly sensitive. Yet mm-wave radiometers *are* sensitive to the scattering from condensed water, especially from the solid hydrometeors towards the upper levels of the cloud. In addition, mm-wave radiometers have a definite advantage over the lower-frequency "window-channel" radiometers (e.g. TMI, AMSR), in that they have finer resolution and can therefore explicitly resolve deep convection. Preliminary analyses indicate that the measurements are indeed very sensitive to the depth and intensity of convection. Figures 1 and 2 illustrate this sensitivity to convection in the case of a Mesoscale Convective System that occurred in the early hours of 24 July 2014 straddling northern Burkina Faso and eastern Mali. The top panel of figure 1 shows the IR (12μ) temperatures measured by Meteosat at 02:00 UTC, featuring saturated cold temperatures at the top of the clouds, while the 31.4-GHz brightness temperatures measured by AMSU-A at 01:50 UTC (bottom panel) allow one to discern some horizontal variation in the amount of scattering within these clouds, though the resolution is too coarse to delineate horizontal structure. This is in contrast with figure 2, which shows the brightness temperatures measured by MHS (the version of AMSU-B that is flying on NOAA-19), clearly outlining individual convective cores within the IR-observed cold cloud tops.

Motivated by this case, the challenge is to derive a robust approach to make quantitative estimates of the convection within the MCS, for example the height and depth of the condensed water, directly from the mm-wave observations. Rather than base the estimation entirely on forward radiative transfer calculations applied to simulated convective systems, we seek as much as possible to exploit the substantial amount of nearly-simultaneous coincident observations by mm-wave radiometers and orbiting atmospheric profiling radars (most notably the Precipitation Radar on board the Tropical Rainfall Measuring Mission satellite – TRMM-PR, operated from November 1997 until July 2014) in order to enforce consistency between the microphysical assumptions that are necessary for the simulations (and for the radiative transfer) and the radar (and radiometer) observations. To do so, we use the nearly-simultaneous coincident observations that are currently the most abundant, namely intersections of MHS with TRMM-PR. Section 2 summarizes the empirical approach that we applied to the data from intersections of MHS and TRMM-PR, and section 3 uses the results of the empirical analysis to evaluate different microphysical assumptions.

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

[2]Department of Physics, UC Irvine, California, USA.

[3]Joint Institute For Regional Earth System Science and Engineering, UCLA, California, USA.

Copyright 2016 by the American Geophysical Union.
0148-0227/16/$9.00

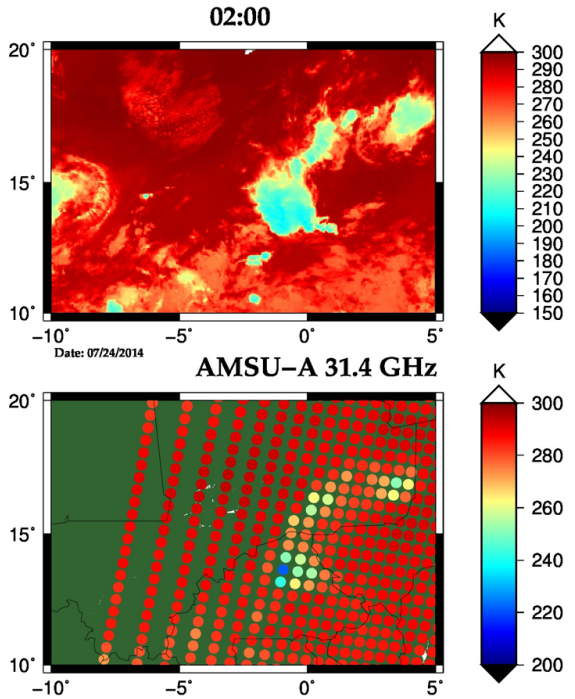

Figure 1. Meteosat IR (12μ) brightness temperatures over West Africa on 24 July 2014 at 02:00 UTC (top panel) with nearly simultaneous AMSU-A 31.4-GHz brightness temperatures measured from NOAA-19 at 01:50 UTC over the same domain.

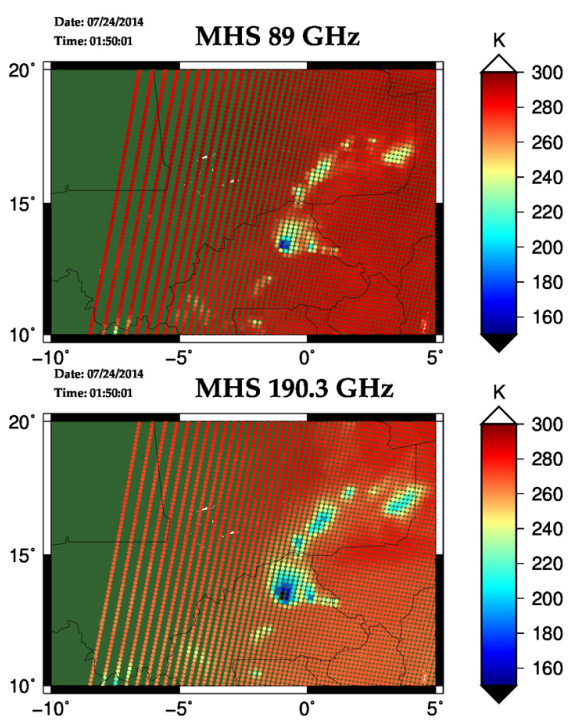

Figure 2. Brightness at 89 GHz (top panel) and 190.3 GHz (bottom panel) measured by the MHS on NOAA-19 over the same system as figure 1, observed on 24 July 2014 at 01:50 UTC.

2. Nearly-simultaneous coincidences of the MHS sounder with TRMM-PR

We compiled a database of coincident observations made by the TRMM-PR (single 13.8-GHz frequency) and the cross-track-scanning MHS radiometer with channels at 89 GHz, 157 GHz, 183.3±1 GHz, 183.3±3 GHz and 190.3 GHz. All cases where the two instruments observed within 3 minutes during 2009 and 2010 were kept. We obtained 1,027,325 sounder beams which the radar classified as "clear", and 79,090 "rainy" sounder beams. The corresponding linear discriminant was derived (to separate between clear and rainy sounder measurements):

$$-.021T_{89} - 0.003T_{157} + .0143T_{184} - .0503T_{186} + .0723T_{190} \quad < \text{(rain) or} > \text{(clear)} \; 3.2166 \qquad (1)$$

We tested this detector on the case of figure 2. The very imperfect result is shown in the left panel of figure 3. The middle panel shows the result after we subjectively reduced the threshold from 3.2166 down to 2.8309. In order not to have to adjust the threshold on a case by case basis, we re-applied a linear discriminant analysis after leaving out the first two channels, and augmenting the remaining three by appending the following three non-linear combinations of the brightness temperatures:

$$T_6 = k_6 \frac{T_{186} - T_{184}}{T_{186} + T_{184}} \qquad (2)$$

$$T_7 = k_7 \frac{T_{190} - T_{184}}{T_{190} + T_{184}} \qquad (3)$$

$$T_8 = k_8 \frac{T_{190} - T_{186}}{T_{190} + T_{186}} \qquad (4)$$

where k_6, k_7 and k_8 are scaling constants equal to the means of their respective denominators. The resulting discriminant is

$$- .1005T_{184} - .567T_{186} + .7153T_{190} + 17.15T_6 - 16.99T_7 + 15.96T_8 \; < \text{ or } > \; 12.158 \qquad (5)$$

The result of applying this detector to the case of figure 2 is shown in the right panel of figure 3. Evidenty, there is no need to adjust the threshold.

Continuing our approach to interpret the brightness temperatures over rain, in order not to have to worry about the background (specifically the surface emissivity and the surface temperature), we performed a principal component analysis on the 352,147 5-tuples of "clear" radiances in order to identify the direction(s) of the largest variability in clear skies, so that we can then work in the orthogonal complement in the rainy cases with measurements which do not vary much in clear air (so that all the variability in the measurements can reasonably be attributed to variations in the condensed water). This is very much in line with the scattering index concept first proposed by N. Grody (1991), which produced a combination of 89 GHz, 22 GHz and 19 GHz which acts similar to a *smallest-eigenvalue* principal component which varies the least in clear and, therefore, whose departures from the mean can be used to quantify the condensed water. In our case, the "clear" principal component analysis shows that the 4 combinations with lower clear-air eigenvalues are

$$\begin{aligned}
T'_2 &= -.086T_{89} + .357T_{157} + .542T_{184} + .538T_{186} + .531T_{190} \\
T'_3 &= -.305T_{89} + .75T_{157} - .535T_{184} - .174T_{186} + .169T_{190} \\
T'_4 &= .131T_{89} - .423T_{157} - .602T_{184} + .349T_{186} + .565T_{190} \\
T'_5 &= .025T_{89} - .132T_{157} + .236T_{184} - .756T_{186} + .608T_{190}
\end{aligned}$$

We then transform every "rainy" T_b 5-tuple into its corresponding quadruple T', and we perform a principal component analysis on the resulting T'. The result defines the

following four rainy principal components listing from the largest- to the smallest-eigenvalue eigenvector:

$$T_1'' = -.917T_2' - .393T_3' - .053T_4' - .04T_5' \quad (6)$$
$$T_2'' = .398T_2' - .91T_3' - .088T_4' - .069T_5' \quad (7)$$
$$T_3'' = -.012T_2' - .088T_3' + .983T_4' - .163T_5' \quad (8)$$
$$T_4'' = -.011T_2' - .095T_3' + .155T_4' + .983T_5' \quad (9)$$

We shall use T_1'' and T_2'' as the two main pieces of information in rain, which contain most of the information that the sounder can sense about the water in the column. The top left panel of figure 4 shows a scatterplot of these two variables, the values having been calculated on the rainy points in our database. Note that the two derived observations do appear to be uncorrelated, consistent with the fact that they were produced by a principal components analysis.

We can now look at the detailed radar information for the rainy cases in our database, and indeed calculate the conditional means of the radar-retrieved condensed water mass in the beam, conditioned on the pair (T_1'', T_2''). This would constitute a first retrieval algorithm, allowing one to look up the condensed water mass given a 5-tuple of actual sounder measurements. Because the condensed water mass varies over the vertical column, we first performed yet another principal components analysis, this time on the vectors of radar-retrieved condensed water mass (accepting that these radar retrievals do come with substantial uncertainty). The top right panel of figure 4 shows the coefficients of the first two condensed-water-mass principal components, and the bottom panel of figure 4 shows the conditional mean of the first condensed-water-mass principal components, given (T_1'', T_2''). The right panel of figure 4 shows the number of samples that fell in each bin (to indicate how many points were used for each conditional mean).

3. Forward simulations using WRF with 4 microphysical schemes

We can now test different microphysical assumptions to forward-calculate the expected brightness temperatures at the MHS frequencies and test a) if these calculated radiances do occupy the same space as the empirical ones do in the top left panel of figure 4, and b) if the conditional mean of condensed water mass, given the calculated radiances, is consistent with the empirical means as in the bottom left panel of figure 4.

We performed this evaluation on a single WRF simulation, of the tropical depression that gave rise to 2003's Hurricane Isabel. WRF was used in a 5-nested-grids configuration, with respective horizontal resolutions of 12km (for the outer grid), 4km, 1.333km (the 'd03' grid), 444m and 148m. The d03 grid roughly spanned a domain from longitudes 36°W to 26°W and from latitudes 6°N to 16°N. Initializing WRF with the GFS FNL analysis on 5 September 2003 at 12:00 UTC, WRF output was collected every 10 seconds from 06:00 to 06:05 on 6 September, then every 30 seconds from 06:05 to 07:00 UTC. The WRF WSM6 was used in the model simulations. Once the output files were obtained at the different times, the SOI forward radiative transfer model was applied to calculate MHS brightness temperatures, using four different microphysical assumptions. The first was the same WSM6 as was used in the model simulations, with hydrometeor scattering properties calculated using a T-matrix method. The second was the WRF Lin scheme, also with T-matrix scattering properties. The third was the WRF Thompson scheme, also with T-matrix scattering. The fourth was a set of microphysical assumptions chosen to mimic empirical correlations derived from in-situ observations and described in detail in Appendix A. The scattering properties for this fourth scheme were calculated using the discrete dipole approximation as described in the appendix. We will refer to this last scheme as the Oosy-ZSH scheme.

Figure 5 attempts to reproduce the scatter plot of the top left panel of figure 4, except that where figure 4 showed the joint behavior of the *observed* measurements during 2009 and 2010, this new figure shows the joint behavior of the *simulated* measurements for each of the 4 microphysical schemes for the single simulated storm. Note that not only do the synthetic observations in the three WRF schemes fail to occupy the same general region as the actual observations do, they also show a very marked linear correlation (where the real observations are manifestly uncorrelated). However the Oosy-ZSH scheme does essentially occupy the same domain as the observations. Figure 6 shows the corresponding retrieved condensed-water-mass principal component # 1. Again, the three WRF schemes do not compare very favorably with the empirical coincident observations (MHS measurements with TRMM-PR retrievals), though, again, the empirical Oosy-ZSH scheme produces conditional means that are remarkably similar to the observations.

4. semi-empirical (forward) observation operator for MHS

The analysis in the previous two sections concerned the retrieval of vertical information about the condensed water mass, from mm-wave brightness temperatures. One can just as easily address the converse problem, namely the derivation of the forward operator producing the mean brightness temperatures that correspond to a given vertical column of condensed water and water vapor. In that converse problem, it is natural to want to use the sample conditional means, derived from the coincident TRMM+MHS data, to characterize the mean behavior and thus remove the bias that typically plagues forward operator representations for a scattering atmosphere. Adopting this approach, if we successfully represent the mean forward representation directly using the empirical database and thus eliminate any bias, we can then use simulated data to represent the variability about the empirical mean or, rather, the dependence of this variability on the atmospheric variables, including the specific humidity inside the cloud. Indeed, in a converse effort to the analysis in section 2, we first use the TRMM+MHS data to obtain the sample conditional mean functions e_1 and e_2 given by

$$e_1(CWM) = \mathsf{E}\{T_1'' \,|\, PC_1(CWM), PC_2(CWM)\} \quad (10)$$
$$e_2(CWM) = \mathsf{E}\{T_2'' \,|\, PC_1(CWM), PC_2(CWM)\} \quad (11)$$

(so that (10) corresponds to the converse of the function illustrated in figure 6). Then we turn to the simulated data, and fit the difference $T_1'' - e_1(CWM)$ to the first vertical principal components of condensed water mass and water vapor (and similarly for the difference $T_2'' - e_2(CWM)$. The left panel of figure 7 shows the definition of the first two principal components of the log of the condensed water mass, while the right panel of the figure shows the corresponding coefficients for the relative humidity in the simulated clouds. And figures 8 and 9 illustrate the sensitivity of the differences $T_1'' - e_1(CWM)$ (in blue) and $T_2'' - e_2(CWM)$ (in red) separately to the water vapor (figure 8, with the left panel showing the sensitivity of the first-principal-component departure from the mean to the first vertical principal component of the relative humidity, and the right panel showing the sensitivity departure of the second vertical principal component to the relative humidity), and to the condensed

water (figure 9). While the scatter diagrams are not perfect curves, they are not isotropic either, especially when the variable is the first vertical principal component of the relative humidity. In fact, one can now apply the correlation-optimization approach demonstrated in *Haddad et al.* 2015, to obtain a very efficient representation of (T_1'', T_2'') in terms of e_1, e_2, $PC_1(\log(RH))$, and $PC_2(\log(RH))$. As figure 8 confirms, the dependance on the relative humidity is negligible for low values and saturates at the high end, so the actual function has to be the (nonlinear) conditional mean $\mathcal{M}_i$ of T_i'' (for i =1 or 2), to capture the non-linearity. The final result will have the form

$$T_i'' \;=\; e_i(CWM) \;+\; \mathcal{M}_i(RH_1', RH_2'), i = 1, 2 \qquad (12)$$

where "RH_1'" and "RH_2'" denote the first two principal components of $\log(RH)$. The significance of this analysis is not so much the specific form of the forward representation, as it is the fact that the departures from the empirical conditional means of the derived observations do correlate consistently with the underlying water vapor variable. This correlation implies that the mm-wave radiances over the cloud do indeed depend on the specific humidity as well as the condensed water. The resulting expression (12) should then be effective in variational assimilation to analyze the water vapor directly, as well as the condensed water.

Conversely, the retrieval approach does not need to be limited to the estimation of the first vertical principal components of the condensed water mass. Indeed, one can use the coincident database to estimate any of the radar-derived geophysical quantities. Figure 10 shows the estimates of the conditional mean of the top of the condensation that exceeds a nominal detection threshold, as was done for the case of AF 447 (Haddad and Park, 2010), in this case the highest altitude in the column where the water content exceeds approximately 0.05 g/m^3. The corresponding r.m.s. uncertainty (given by the conditional standard deviation) depends on the estimated height, with values mostly around 900m (right panel of figure 10). The significance of these estimates in this case stems from the fact that Air Algérie flight AH 5017 passed over the convective core of this system between 01:29:14Z (when it was near 1.311W 13.27N, altitude 22854 ft) and 01:35:24Z (when it was near 1.495W 14N, altitude 29114 ft), before crashing very shortly thereafter at 01:47Z near 1.079W 15.134N.

5. Conclusions and upcoming work

The empirical TRMM-PR + NOAA19-MHS coincidence dataset that we have used is an excellent testbed for our approach to relate the measured mm-wave brightness temperatures to the underlying vertical profile of condensed water and water vapor inside the cloud. Not only can it be used to derive a robust detector (5) for clouds using mm-wave brightness temperatures, it can be used to derive an empirical expression for the forward observation operator expressing the mean value of the brightness temperatures given the underlying condensed water (10). Moreover, our microphysical scheme (described in the appendix) can then be used to relate the departure of the brightness temperatures from their mean, on one hand, to the underlying water vapor in the cloud, on the other hand (12). The resulting observation operator then constitutes a semi-empirical forward operator that has the dual merits of being unbiased by design and of effectively expressing the dependence of the observations on one of the most important prognostic variables, the specific humidity, as well as the condensed water.

There are two main opportunities for improvement in this approach. First, the mismatch between the sensitivity of the radiometer and that of the radar makes this particular dataset sub-optimal. Indeed, the mm-wave sounders are more sensitive to the solid condensation above the melting layer, while the radars aboard TRMM (and GPM-core) do not have enough sensitivity to detect the lighter concentrations of solid hydrometeors aloft (witness the quickly decreasing magnitudes above 6 km of the coefficients of the first two principal components of the condensed water mass, in figure 7, compared to the magnitudes between 1 and 5.75 km AMSL). To remedy this problem, we are now building a more complete database of CloudSat reflectivities coincident with either the TRMM-PR or the GPM-DPR radar measurements.

In addition, there is a more subtle sensitivity issue that requires special attention for observations over the ocean. The left panel of figure 11 shows the joint behavior of the 89-GHz brightness temperature and the 157-GHz temperature over land, confirming that the scattering in rain produces colder values at both frequencies, consistently. The right panel of figure 11 shows the cases over the ocean. Note the substantial number of cases toward the center left, where the 89-GHz measurement is cold even though the 157-GHz measurement is warm. Closer examination of the radar estimates in these rainy cases having unexpectedly mismatched brightness temperatures (figure 12) reveals that these anomalous observations are most likely over shallow clouds (the green values, for the solid condensation that is mostly above the freezing level, are below the nominal detection threshold of the radar, and are therefore quite likely to be noise). The boundary-layer condensation is sufficient to cool the 89 GHz upwelling, but its effects at 157 GHz are likely obscured by water vapor. These cases should be easy to detect using CloudSat.

In conclusion, the approach above capitalizes on the coincident radar+sounder dataset to a) build a Bayesian retrieval algorithm that estimates the first vertical principal component of the condensed water mass and the height of the convection (as defined by the radar detection threshold), and b) formulate a robust forward observation operator that takes a two-step approach to represent the dependence of the sounder brightness temperatures on the underlying condensed water and water vapor (the first step relying exclusively on the coincident data to capture the mean relation, and the second step relying on cloud-resolving simulations to complete the empirical relation with the dependence on the first vertical principal component of the relative humidity). This 2-step approach solves the bias problem between model simulations and observations (which has so far plagued the approaches that rely on model simulations alone), and relies on the model to represent the sensitivity to one of the main prognostic variables, the specific humidity.

Acknowledgments. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The work was partially supported by the National Oceanic and Atmospheric Administration's Hurricane Forecasting Improvement Project.

References

Gravner, J., and D. Griffeath, 2009: Modeling snow-crystal growth: A three-dimensional mesoscopic approach. *Phys. Rev. E* **79**, DOI: 10.1103/PhysRevE.79.011601

Grody, N.C., 1991: Classification of snow cover and precipitation using the Special Sensor Microwave Imager. *J. Geophys. Res. D* **96**, 7423-7435

Haddad, Z.S., J.P. Meagher, S.L. Durden, E.A. Smith, and E. Im, 2007: Drop Size Ambiguities in the Retrieval of Precipitation Profiles from Dual-Frequency Radar Measurements. *J. Atmos. Sci.* **63**, 204-217.

Haddad, Z.S., and K.W. Park, 2010: Vertical profiling of tropical precipitation using passive microwave observations, and its implications regarding the crash of Air France 447. *J. Geophys. Res. A* **115**, D12129, doi:10.1029/2009JD013380.

Haddad, Z.S., J. Steward, H.-C. Tseng, T. Vukicevic, S.-H. Chen, and S. Hristova-Veleva, 2015: A data assimilation technique to account for the nonlinear dependence of scattering on microwave observations of precipitation. *J. Geophys. Res. A* **120**, 5548-5563, doi:10.1002/2015JD023107

Heidinger, A.K., C. O'Dell, R. Bennartz, and T. Greenwald, 2006: The Successive-Order-of-Interaction radiative transfer model. Part I: model development. *J. Appl. Meteor. Clim.* **45**, 1388-1402.

Heymsfield, A.J., C. Schmitt, A. Bansemer, and C.H. Twohy, 2010: Improved representation of ice particle masses based on observations in natural clouds. J. Atmos. Sci. **67**, 3303-3318.

Hong, S.Y., and J.-O. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Kor. Met. Soc.* **42**, 129-151.

Kidd, C., T. Matsui, J. Chern, K. Mohr, C. Kummerow, and D. Rundel, 2015: Global precipitation estimates from cross-track passive microwave observations using a physically-based retrieval scheme. *J. Hydromet.*, DOI: 10.1175/JHM-D-15-0051.1

Kuo, K.-S., W. S. Olson, B. T. Johnson,M. Grecu, L. Tian, T. L. Clune,B. H. van Aartsen, A. J. Heymsfield, L. Liao and R. Meneghini, 2016: The Microwave Radiative Properties of Falling Snow Derived from Nonspherical Ice Particle Models. Part I: An Extensive Database of Simulated Pristine Crystals and Aggregate Particles, and Their Scattering Properties. *J. Appl. Meteorol. Clim.* **55**, 691-708, doi:10.1175/JAMC-D-15-0130.1

Lin, Y.L., R.D. Farley, and H. D. Orville, 1983: Bulk parametrization of the snow field in a cloud model. *J. Clim. Appl. Meteor.* **22**, 1065-1092.

Liu, G., 2008: A database of microwave single-scattering properties for nonspherical ice particles. *Bull. Amer. Meteor. Soc.* **89**, 1563-1570.

Petty, G.W., and W. Huang, 2010: Microwave backscatter and extinction by soft ice spheres and complex snow aggregates. *J. Amtos. Sci.* **67**, 769-787.

Schmitt, C.G., and A.J. Heymsfield, 2010: The dimensional characteristics of ice crystal aggregates from fractal geometry. *J. Amtos. Sci.* **67**, 1605-1616.

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

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.* **136**, 5095-5115.

Turk, F.J., K.-W. Park, Z.S. Haddad, P. Rodriguez and D. Hudak, 2011: Constraining CloudSat-based snowfall profiles using AMSU/MHS and C-band ground radar. *J. Geophys. Res. A* **116**, DOI: 10.1029/2011JD016126

Corresponding author: Z. S. Haddad, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109-8099, USA. (zsh@jpl.nasa.gov)

Appendix: creating and parametrizing distributions of solid hydrometeors

Much progress has been made recently on modeling realistically shaped solid hydrometeors, including pristine ice crystals, (*Gravner and Griffeath*, 2009; *Liu*, 2008), and aggregates and graupel (*Petty and Huang*, 2010; *Kuo et al.*, 2013). We start with the modeled hydrometeors created by Kuo. Using Gravner and Griffeath's Snowfake, Kuo constructed about 10,000 realistic solid hydrometeors, including both pristine and aggregate types, with maximum diameter ranging from about 100 microns to about 15 mm. Then, using the discrete-dipole-approximation code DDSCAT, Kuo calculated the signatures of each individual hydrometeor (extinction efficiency, absorption efficiency, scattering efficiency, scattering asymmetry parameter, and radar backscattering cross-section) at the different relevant frequencies, to build a database consisting of one entry for each synthesized hydrometeor, specifying the hydrometeor's defining physical characteristics and the associated radiometric signatures. The problem then is to assemble these hydrometeors into realistic "candidate" distributions.

This appendix describes an effective, efficient and elegant approach that we took to create, from Kuo's database, physically realistic hydrometeor distributions, as one would encounter in a volume element within a modeled cloud or within the beam of a radiometer (or radar). The requirements are that the distribution be efficiently parametrized (to allow one to estimate the parameters from remote-sensing observations), and that the parameters reflect the correlations that have been empirically determined to exist from in-situ observations.

Our approach is inspired by the solution of the far simpler case of spherical hydrometeors: in that case, the single parameter that differentiates the hydrometeors is their diameter, and empirical distributions obtained by fitting analytic expressions to a population of diameters sampled over a reasonably short interval have led modelers to adopt different flavors of the Γ distribution as the probability law governing the diameter – with different parameter-estimation methods used by different groups to quantify the parameters of the Γ distribution that best fits a given sample of drops. In order for this approach to be applicable to the more complex case of distributions of solid hydrometeors, one needs to address the fact that the latter are not spherical, and indeed come in different habits whose characterization requires more than a single "diameter" variable. Indeed, examining the relation between the maximum dimension D_M of individual observed hydrometeors and each hydrometeor's apparent mass m, A. Heymsfield's extensive analyses and literature reviews (*Heymsfield et al.*, 2010; *Schmitt and Heymsfield*, 2010) point to different mean power-law relations that seem to be verified between D_M and m and that do appear to characterize the particle habit – the power being the fractal dimension of the hydrometeors in question.

For retrievals, we need to represent not just the mean relation, but rather the spread about this mean too, and this spread would then correspond to the range of hydrometeor habits. Therefore, analytically, we assume that

$$m = aD_M^b \tag{A.1}$$

where b is the fractal dimension characterizing the type of solid hydrometeor (*Schmitt and Heymsfield*, 2010), and where the ranges of values that a and b can take are

$$0.004 < a < 0.009 \quad \text{and} \quad 1.8 < b < 2.4. \tag{A.2}$$

Rather than force possibly unrealistic constraints a priori, the **first step** of our Hydrometeor Size Distribution (HSD) database construction starts with the database of individual synthetic hydrometeors (pristine and aggregates), loops over combinations of values of (a, b) as above, and for each empirical power-law we pick out all the individual hydrometeors that obey that fractal power law. Thus the first step avoids any a priori assumptions about the correlation between the coefficient a and the fractal dimension b (realistic constraints will be applied in the final step 3 of the construction). To obtain distributions that are representative of the mean relations in Heymsfield's analyses, we use
a = either .003, .004, .005, .006, .007, .008, .009, or .01
b = either 1.8, 2, 2.2, or 2.4
independently, knowing that not all of the 8×4=32 (a, b) pairs will survive the consistency test in step 3. Still, at the end of step 1, we have classified each hydrometeor as having one of the 32 habits parametrized by A.1+A.2.

The **second step** consists in associating, with each collection of hydrometeors, a count for each hydrometeor representing the relative frequency of that hydrometeor versus the others in the collection. This is akin to the Γ distribution of spherical raindrops, and this step is indeed implemented by applying Γ distributions

$$N(D) = N_0 D^{\mu} e^{-\Lambda D} \tag{A.3}$$

(with different shape parameters μ and Λ) to each of the 32 collections obtained in step 1. By the end of this second step, each one of the (a, b) pairs from step 1 will have produced $M \times L$ candidate distributions, M and L being the number of values of μ and Λ that we use – the third parameter, the overall count multiplier N_0, only comes into play in the next (and final) step.

Indeed, **step 3** consists in sifting through all the candidate distributions created in step 2, and retaining only those distributions that meet the observed relations between the first three sample moments, namely the water content q $(=\int m(D)N(D)dD)$, the mass-weighted mean melted-equivalent-sphere diameter D_m $(=\int D\, m(D)N(D)dD/\int m(D)N(D)dD)$, and its mass-weighted standard deviation σ_m, more specifically the relations between these three as derived from in-situ data. Indeed, *Turk et al.* (2011) derived empirical D_m–q and D_m–σ_m relations of the form

$$D_m = \alpha q^{0.17} \quad \text{and} \quad \sigma_m = \beta D_m^{1.3} \tag{A.4}$$

where the coefficients α and β turn out to lie in the ranges

$$0.2 < \alpha < 2, \quad \text{and} \quad 0.15 < \beta < 0.6,$$

respectively. Pending analysis of data from different regimes, step 3 therefore consists in calculating, for every candidate distribution produced by step 2, its first three moments (q, D_m, σ_m), and removing those candidate HSDs whose ratios $D_m/q^{0.17}$ and $\sigma_m/D_m^{1.3}$ are not within the observed ranges. This effectively removes distributions whose (a, b) pair was not realistic to begin with, or whose Γ-distribution parameters are not consistent with the bulk correlations derived from our observations.

The main advantage of this approach is that it leverages what reliable in-situ observations we do have (of the correlation between the first three moments of the observed hydrometeor distributions, which forms the basis for step 3 in the approach) to produce a realistic effective parametrization of the HSDs that can be used, in the form of a database, for the retrievals.

Figure 13 confirms the potential of this approach and illustrates its issues. It shows the maximum D_M of the members of our input single-hydrometeor database, plotted against their mass (or, actually, the melted-equivalent sphere's diameter). The curves show a sampling of the (a, b) power-laws proposed above, the pink curves corresponding to a=0.003 (with b=1.8, 2.0, 2.2 and 2.4 from left to right), the green to a=0.005, the red to a=0.007, and the single blue curve to a=0.009 with b=2.4. Note that the simulated hydrometeors are generally consistent with the proposed relation, though they do appear to be biased towards the denser side, implying that we need to include aggregates that are synthesized with larger center-to-center aggregation distance. Conversely, the simulation-derived relations do appear generally consistent with the simulated 3D hydrometeors, though the smaller fractal coefficients may well be the result of a too-generous 2D-to-3D fractal-dimension scaling factor ("S" in *Schmitt and Heymsfield*, 2010).

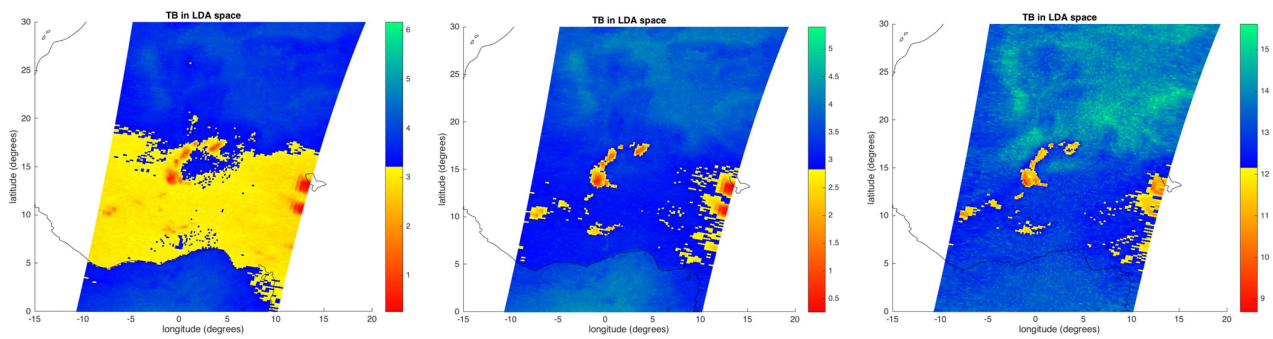

Figure 3. Left panel: Detector defined by equation 1 applied to the MHS measurements in the case of figure 2; center panel: same detector defined by equation 1 but applied with a reduced threshold of 2.8309 (instead of 3.2166); right panel: Detector defined by equation 5.

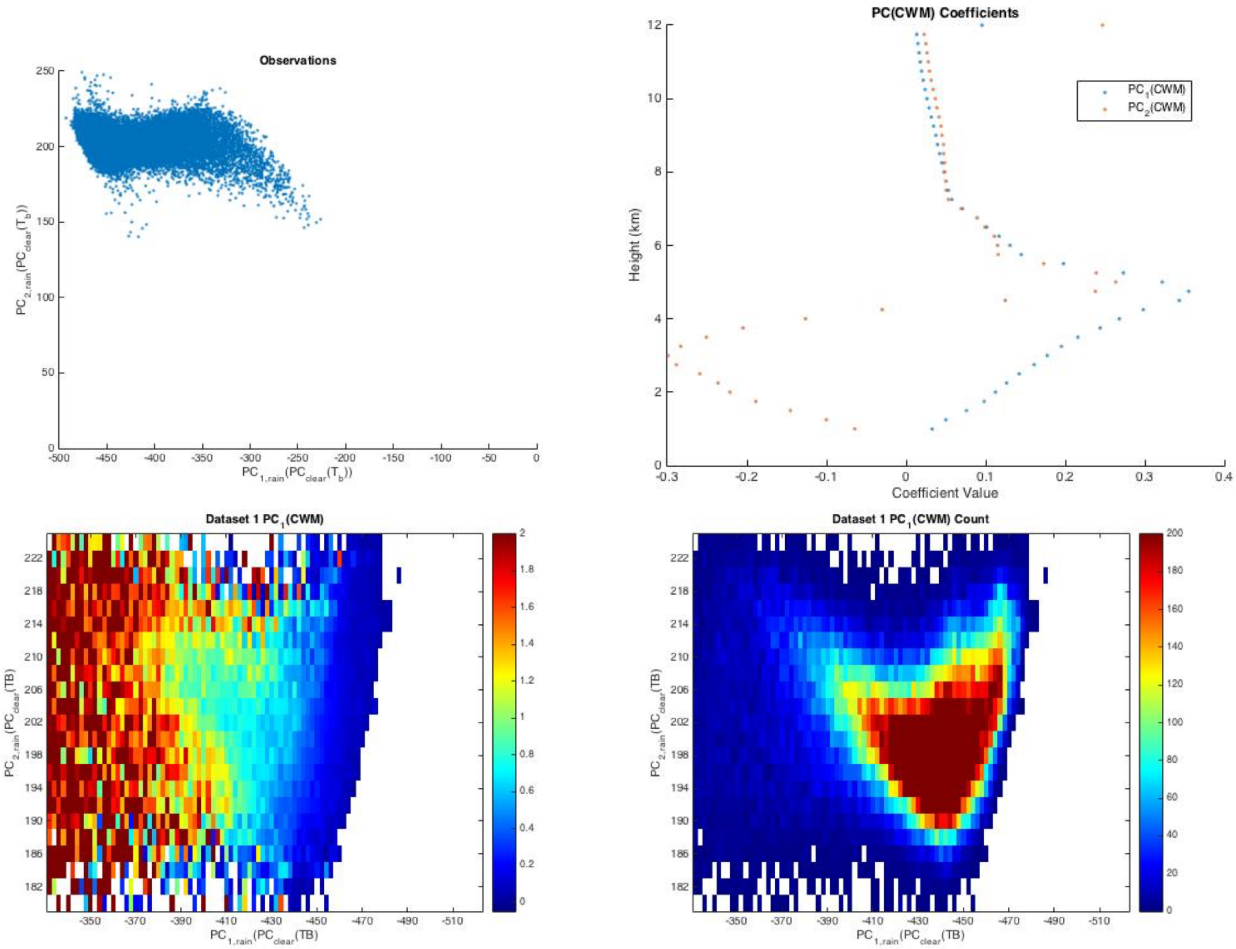

Figure 4. Top left: Scatter plot of the pair of derived sounder observation made up of the two rainy principal components T_1'' and T_2'', defined by equations (6) and (7), applied to our database (of coincident MHS and TRMM-PR observations). **Top right:** Coefficients of the first two vertical principal components of the condensed water mass (the latter having been retrieved from the TRMM-PR observed profile of radar reflectivity factors, by the standard TRMM PR+TMI retrieval algorithm 2B31). **Bottom left:** Zoom in on the area of the joint observations in the top left panel, showing the conditional mean of the first vertical principal component of the condensed water mass, conditioned on the pair (T_1'', T_2'') of sounder brightness temperature principal components, for our empirical database. **Bottom right:** Counts for the binned data in the left panels.

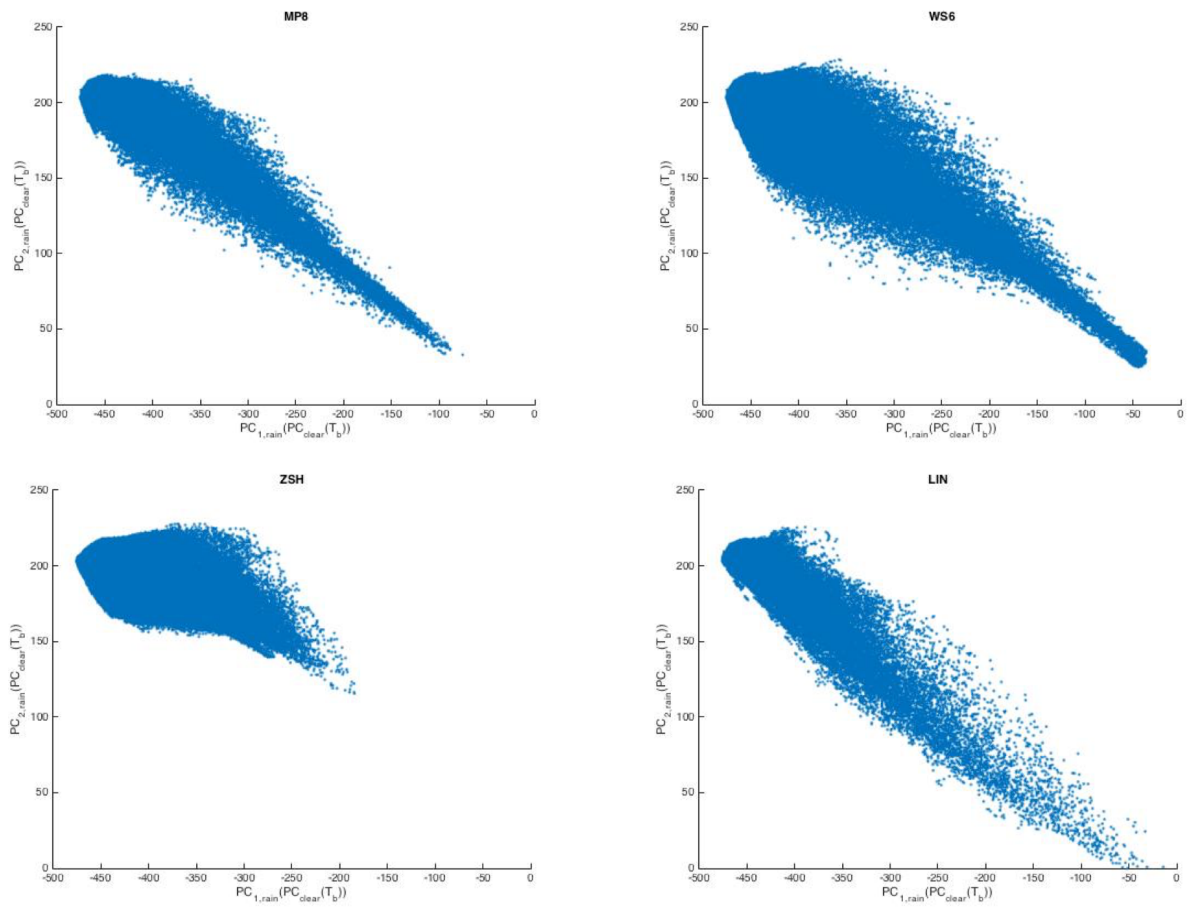

Figure 5. Scatter plot as in figure 4, but for the WRF simulation with brightness temperatures calculated using the Thompson scheme (top left), the WSM6 scheme (top right), the Oosy-ZSH scheme (bottom left), and the Lin scheme (bottom right). This figure is to be compared to the corresponding figure for the actual observations, the top left panel of figure 4.

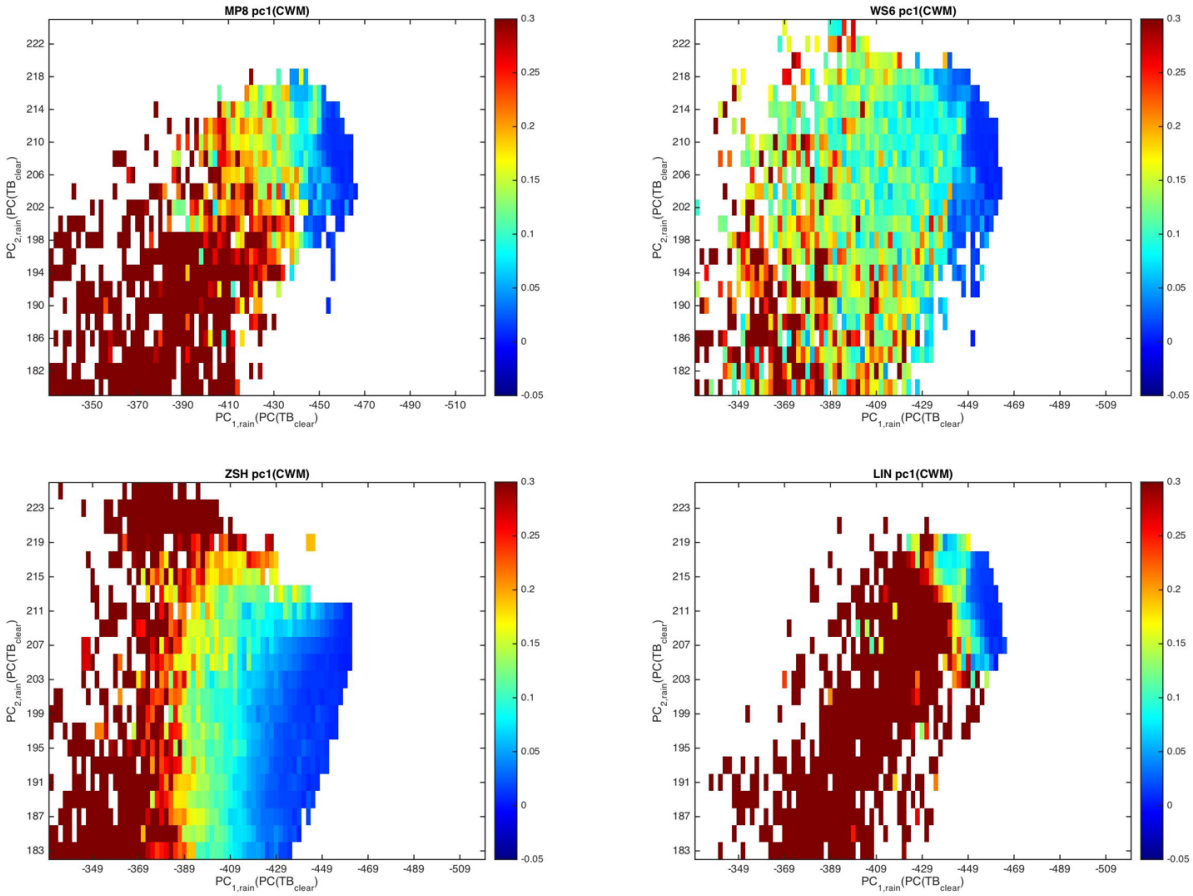

Figure 6. Zoom in on the area of the joint synthetic observations of figure 5, showing the conditional mean of the condensed water mass first-principal-component (defined by the coefficients shown in the top right panel of figure 4), but for the simulated data with the Thompson scheme (top right), the WSM6 scheme (top right), the Oosy-ZSH scheme (bottom left), and the Lin scheme (bottom right). This figure is to be compared to the corresponding figure for the actual observations, the bottom left panel of figure 4.

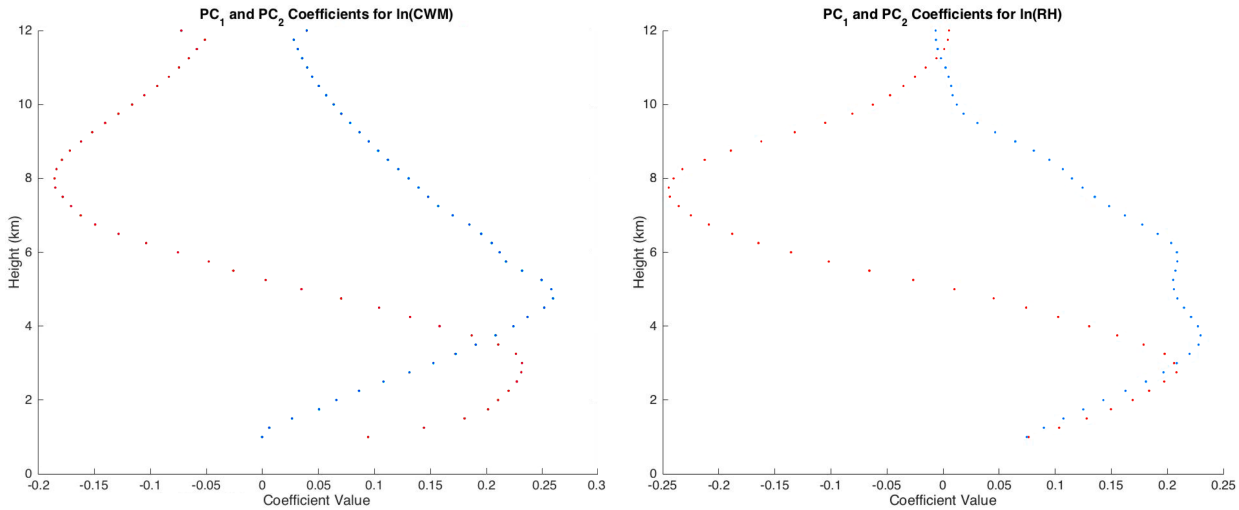

Figure 7. Definition of the first two vertical principal components of the condensed water mass profiles obtained from the coincident TRMM+MHS data (left), and of the first two vertical principal components of the relative humidity profiles obtained from the Isabel simulation (right). The coefficients of the first principal components are shown in blue, those of the second are shown in red.

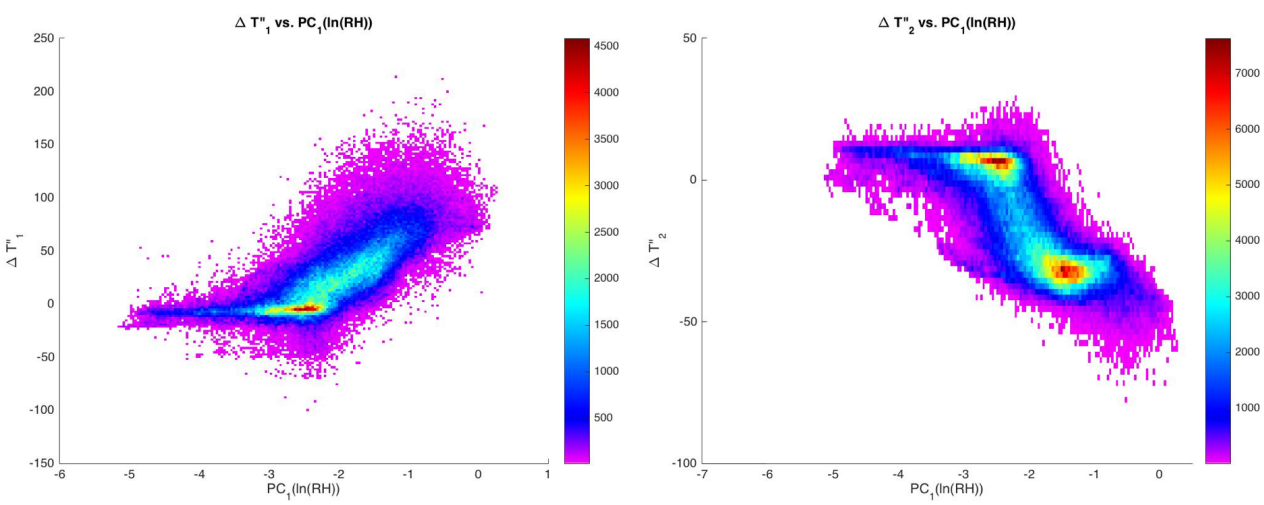

Figure 8. Scatter diagram illustrating the relation between $T_1'' - \text{mean}[T_1''|PC_1(CWM), PC_2(CWM)]$ and $PC_1(log(RH))$ (in blue), and between $T_2'' - \text{mean}[T_2''|PC_1(CWM), PC_2(CWM)]$ and $PC_1(log(RH))$ (in red), as derived from the Isabel simulation. The "conditional mean" functions are the sample conditional means derived from the coincident TRMM+MHS data.

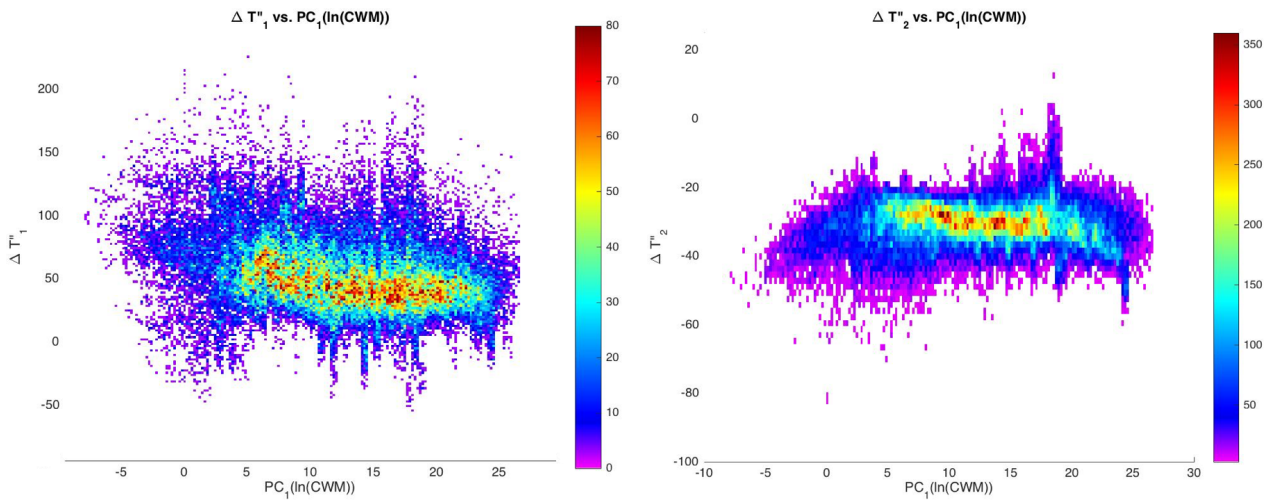

Figure 9. Scatter diagram illustrating the relation between $T_1'' - \text{mean}[T_1''|PC_1(CWM), PC_2(CWM)]$ and $PC_1(log(CWM))$ (in blue), and between $T_2'' - \text{mean}[T_2''|PC_1(CWM), PC_2(CWM)]$ and $PC_1(log(CWM))$ (in red), as derived from the Isabel simulation. The "conditional mean" function are the sample conditional means derived from the coincident TRMM+MHS data.

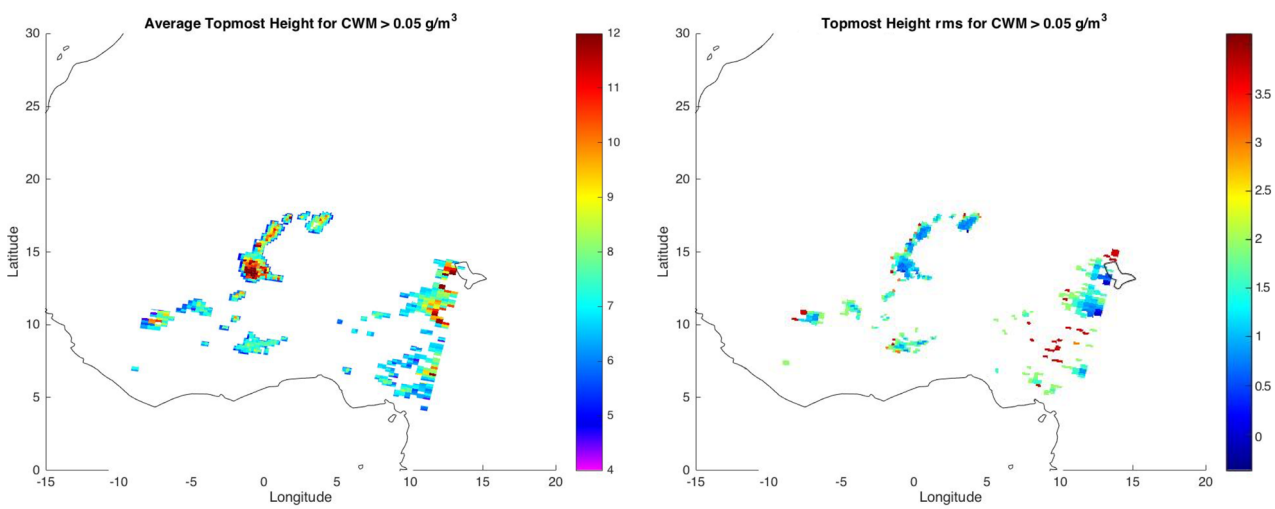

Figure 10. Mean heights (left) and height uncertainty (right) of the condensed water mass exceeding 0.05 g/m^3 for the case of figures 1 and 2.

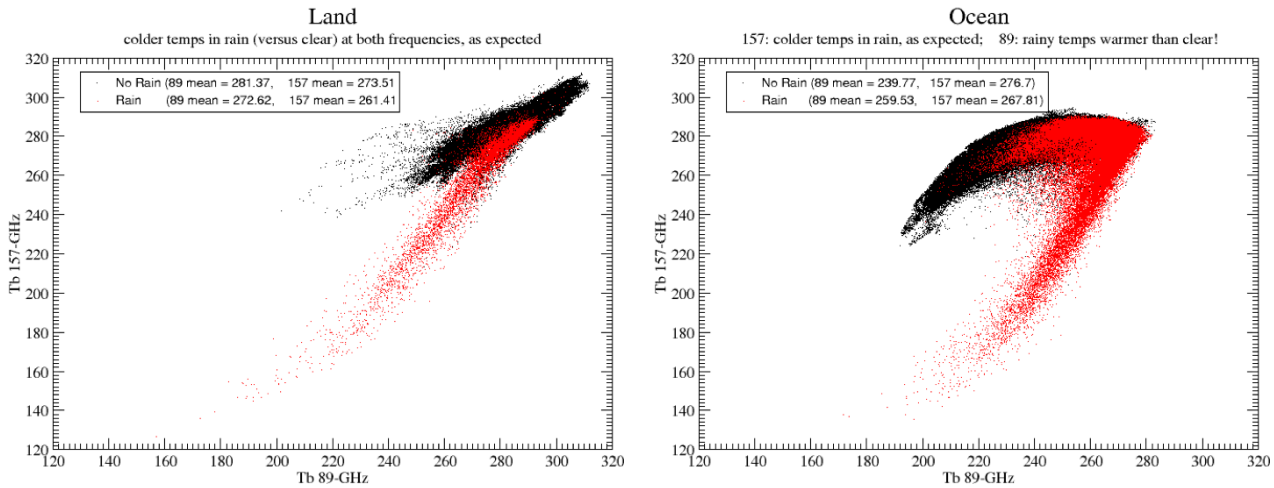

Figure 11. Scatter plot of MHS 157-GHz and 89-GHz brightness temperatures over land (left panel) and over ocean (right panel). Black points indicate measurements radar-classified as "clear", red points indicate "rain").

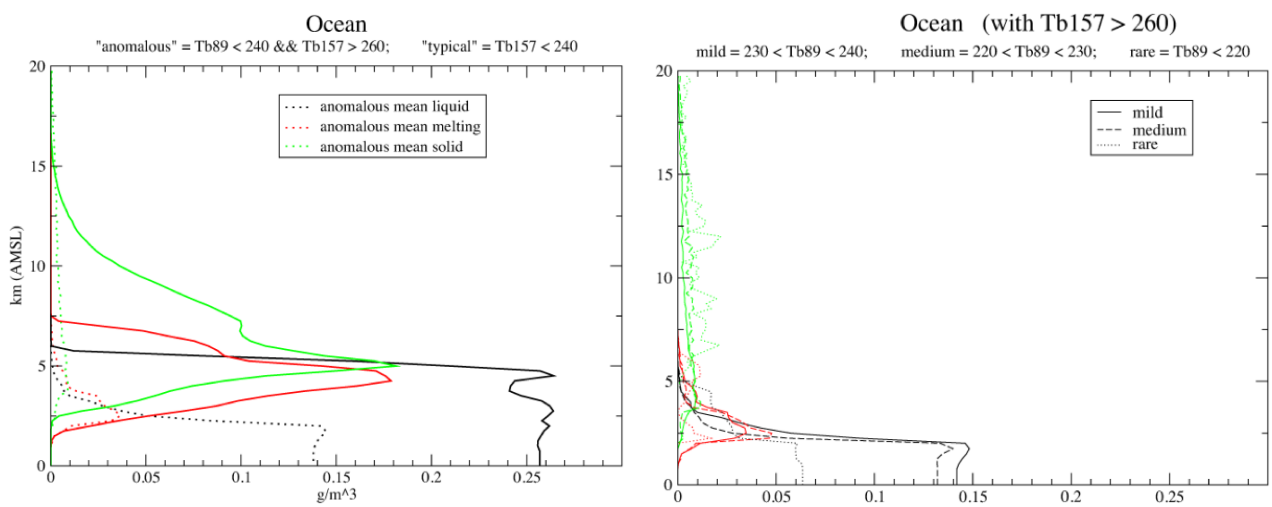

Figure 12. Left panel: Mean anomalous rainy profiles (dotted lines) along with the means over all the rainy profiles. **Right panel:** Means of three anomalous ocean+rainy classes (with cold 89-GHz obs but warm 157-GHz obs).

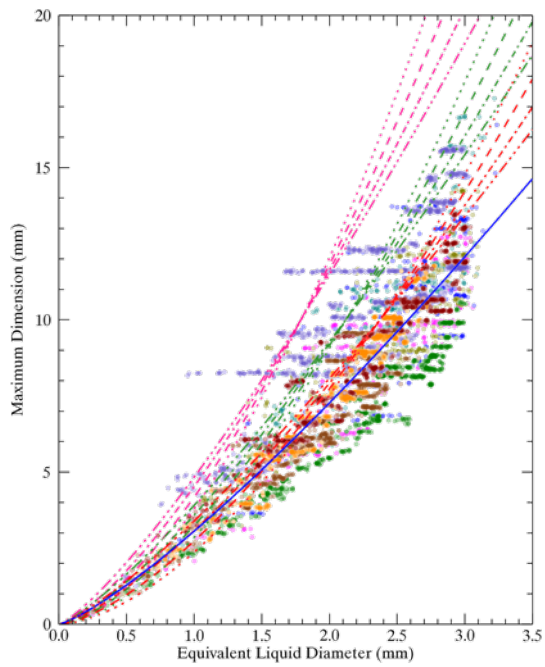

Figure 13. Illustration of the m-D_M relations (the curves) superposed on a scatter diagram of the Kuo database's single-crystal hydrometeors' (m,D_M) pairs, with D_M replaced here by the melted diameter