



Transitioning from CRD to CDRD in Bayesian retrieval of rainfall from satellite passive microwave measurements: Part 3 – Identification of optimal meteorological tags

E. A. Smith¹, H. W.-Y. Leung², J. B. Elsner³, A. V. Mehta⁴, G. J. Tripoli⁵, D. Casella⁶, S. Dietrich⁶, A. Mugnai⁶, G. Panegrossi⁶, and P. Sanò⁶

¹Center for Research on the Changing Earth System (CRCES) [HQ: Catonsville, MD 21228], Tallahassee, FL 32312, USA

²Dept. of Meteorology (MISU), Stockholm University, 106 91 Stockholm, Sweden

³Dept. of Geography, Florida State University, Tallahassee, FL 32306, USA

⁴NASA/Goddard Space Flight Center with Univ. of Maryland Baltimore County/Joint Center for Earth Systems Technology, Greenbelt, MD 20771, USA

⁵Dept. of Atmospheric and Oceanic Sciences (AOS), University of Wisconsin, Madison, WI 53706, USA

⁶Institute of Atmospheric Sciences and Climate (ISAC), Italian National Research Council (CNR), 00133 Rome, Italy

Correspondence to: E. A. Smith (easmith@crces.org)

Received: 31 October 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –
Revised: 11 February 2013 – Accepted: 18 February 2013 – Published: 16 May 2013

Abstract. In the first two parts of this study we have presented a performance analysis of our new Cloud Dynamics and Radiation Database (CDRD) satellite precipitation retrieval algorithm on various convective and stratiform rainfall case studies verified with precision radar ground truth data, and an exposition of the algorithm's detailed design in conjunction with a proof-of-concept analysis vis-à-vis its theoretical underpinnings. In this third part of the study, we present the underlying analysis used to identify what we refer to as the *optimal* meteorological and geophysical tags, which are the optimally effective atmospheric and geographic parameters that are used to refine the selection of candidate microphysical profiles used for the Bayesian retrieval. These tags enable extending beyond the conventional Cloud Radiation Database (CRD) algorithm by invoking meteorological-geophysical guidance, drawn from a simulated database, which affect and are in congruence with the observed precipitation states. This is guidance beyond the restrictive control provided by only simulated radiative transfer equation (RTE) model-derived database brightness temperature (TB) vector proximity information in seeking to relate physically consistent precipitation profile solutions to individual satellite-observed TB vectors. The first two parts of the study have rigorously demonstrated that the optimal tags effectively mitigate against solution ambiguity, where

use of only a CRD framework (TB guidance only) leads to pervasive non-uniqueness problems in finding rainfall solutions. Alternatively, a CDRD framework (TB + tag guidance) mitigates against non-uniqueness problems through improved constraints. It remains to show how these optimal tags are identified. By use of three statistical analysis procedures applied to a database from 120 North American atmospheric simulations of precipitating storms (independent of the 60 simulations for the European-Mediterranean basin region used in the Parts 1 and 2 studies), we examine 25 separate dynamical-thermodynamical-hydrological (DST) and geophysical parameters for their relationships to rainfall variables – specifically, surface rain rate and columnar liquid/ice/total water paths of precipitating hydrometeors. The analysis identifies seven optimal parameter tags which exceed all others in the strengths of their correlations to the precipitation variables but also have observational counterparts in the operational global forecast model outputs. The seven optimal tags are (1 and 2) vertical velocities at 700 and 500 hPa; (3) equivalent potential temperature at surface; (4) convective available potential energy; (5) moisture flux 50 hPa above surface; (6) freezing level height; and (7) terrain height, i.e., surface height.

1 Introduction

The first two parts of this series of investigations, i.e., the studies of Sanò et al. (2013) and Casella et al. (2013), have reported on the development of a new satellite precipitation retrieval algorithm which we refer to as the Cloud Dynamics and Radiation Database (CDRD) algorithm. The essence of the new algorithm is that it uses what we call optimal meteorological-geophysical parameter tags to assist in the process of guiding the algorithm in finding microphysical profile precipitation solutions that are congruent with the environments encompassing the satellite measurements used in the retrievals. The first of these studies conducted a performance evaluation and verification of the new algorithm involving a pair of case studies over the Lazio region of central Italy, for which precision ground-truth radar data were available from the Polar 55C Doppler C-band polarimetric radar facility located at the CNR/ISAC institute in Rome. The case studies included both weak and intense convective precipitation, as well as various degrees of intensity of stratiform precipitation. The performance of the algorithm justified the new algorithm design based on the very close agreement between the satellite radiometer and ground-truth radar retrievals. The second study proceeded to describe in detail the design of the algorithm, particularly the model underpinnings, and finished with a proof-of-concept analysis that confirmed the effectiveness of the algorithm in overcoming the prevailing ambiguity problem that plagued an earlier generation of algorithms which were formulated within a similar theoretical framework. In this third part of the study, we address the problem of how the optimal meteorological-geophysical parameter tags needed by the CDRD algorithm have been chosen.

The CDRD name we have chosen is not arbitrary. First of all it represents an extension of the name that has been used for the aforementioned earlier generation of algorithms called Cloud Radiation Database (CRD) algorithms (see Smith et al., 1994a; Mugnai et al., 2008) in which a simulation database produced by the combination of a cloud resolving model (CRM) and a radiative transfer equation (RTE) model is used for knowledge guidance to produce precipitation microphysics profile solutions from passive microwave (PMW) radiometer measurements obtained from space. The CRM is used to produce a large number (thousands to millions) of coincident meteorological and microphysical profile and scalar parameters, as well as key geophysical scalar parameters, situated within the environments of simulated precipitating storms. In sequence, the RTE model is used to calculate simulated brightness temperature (TB) vectors associated with the meteorological, microphysical-profile and geophysical conditions, that are said to represent what a satellite PMW radiometer would sense vis-à-vis top-of-atmosphere (TOA) upwelling TB vector quantities. Note the term TB vector denotes that a PMW TB measurement is multispectral (i.e., multi-channel) in nature, and thus must be treated as a

packet of values (i.e., a vector). Also, a microphysical profile is defined as a set of vertically distributed cloud and precipitation hydrometeor mixing densities consisting of multiple liquid and frozen hydrometeor categories (e.g., a 2-water/4-ice set might consist of cloud droplets, rain drops, pristine crystals, snow pellets/flakes, ice aggregates and graupel/hail particles).

The algorithm solution methodology is optional; relaxation approaches have been used (e.g., Smith et al., 1994b, c) as have Bayesian approaches (e.g., Kummerow et al., 1996; Pierdicca et al., 1996; Marzano et al., 1999; Mugnai et al., 2001) and others (e.g., Bauer et al., 2001). The key issue concerning CRD-type algorithms is that, in the process of obtaining a solution, an individual observed radiometer TB vector must be compared to the entire set of simulated TBs in the knowledge database in order to select candidate microphysical profile solutions (which may include all profiles in the database) based on the proximity of the *modeled* TB vector quantities to the corresponding *measured* TB vector quantity. The full set or subset of candidate profile solutions is then exported to the solver portion of the algorithm to determine a specific solution.

Although the CRD methodology has been partially successful, particularly in combination with Bayesian solver schemes (e.g., Evans et al., 1995), such algorithms are fraught with the problem of solution ambiguity (note the conventional CRD-type scheme being used for the operational processing of Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) measurements into Version 6 precipitation profile products is Bayesian; see Kummerow et al. (2001) for a description of the 2a12 TRMM TMI facility algorithm, also referred to as GPROF). The ambiguity stems from the fact that multiple solutions are possible because different vertical profile structures of microphysical hydrometeors can lead to exactly or nearly exactly the same TB vector. This represents the classical non-uniqueness problem inherent to multi-value mathematical functions, with the exception that within the framework of a CRD precipitation algorithm, the said function is the algorithm solver itself and the multi-value impediment stems from the many profiles situated in the CRD algorithm's database.

So the problem has existed for almost two decades of how to best overcome solution ambiguity within the framework of making precipitation retrievals with PMW radiometer measurements, i.e., the type of rain measuring instrument that has provided for many years, and continues to provide, global coverage. Of course since the launch of TRMM, which besides flying the TMI radiometer also flew the revolutionary 13.8 GHz incoherent scanning Precipitation Radar (PR), there have been a number of algorithms developed to combine TMI measurements with PR measurements in order to supplement the hydrometeor reflection information inherent to active microwave measurements with hydrometeor attenuation information intrinsic to passive microwave measurements. These began with the TRMM

facility combined PR-TMI algorithm 2b31 of Farrar (1997), Haddad et al. (1997) and Smith et al. (1997) and have continued to flourish (e.g., Marzano et al., 1999; Bauer et al., 2001; Grecu et al., 2004; Grecu and Olson, 2006 – see also a related study by Battaglia et al., 2003).

The TRMM radar has also inspired the development of PMW algorithms trained by PR profile measurements (e.g., Masunaga and Kummerow, 2005; Kummerow et al., 2006, 2011; Viltard et al., 2006; Munchak and Kummerow, 2011) (note the latter two papers, Kummerow et al. (2011) and Munchak and Kummerow (2011), constitute a description of the 2a12-v7 TMI (trained-by-PR) algorithm, which has replaced the Kummerow et al. (2001) 2a12-v6 TMI (CRD-type) algorithm, which had always suffered from database incompleteness and concomitant solution ambiguity problems). Of course, there is an underlying limitation with the training algorithms in that they only physically apply to the latitude belt over which the PR acquires observations (i.e., between 35° S to 35° N) governed by the TRMM observatory's orbit inclination. Moreover, given a high-frequency radar's fundamental inability to fully detect ice phase processes and the fact that starting at 35 GHz and upward, PMW observations are determined by constituents within the ice layers, training algorithms are missing some fundamental physics. This then begs the question of how to best improve on PMW-only algorithms, especially in the context of ambiguity, for applications outside the tropical and sub-tropical domain of TRMM. This will be the case until at least the ~2015 anticipated launch of the Global Precipitation Measuring (GPM) mission core satellite, which will fly a dual frequency 13.6/35.5 GHz incoherent radar package out to the 65° parallels – assuming all goes according to plan (see Smith et al., 2007 and <http://pmm.nasa.gov/>).

Thus, the purpose of the CDRD algorithm is to improve upon the conventional CRD-type algorithms, specifically to mitigate against ambiguity in the solutions. To do so we have extended the CRD methodology to include additional parameters that better isolate candidate profile solution subsets used by the CDRD algorithm's Bayesian solver to acquire unique solutions. As noted, the information used for the additional guidance is meteorological and geophysical in nature. The specific meteorological-geophysical parameters are drawn from those produced by the underlying CRM model, but with a restriction that they must have direct counterparts in the observational world such that the same procedure used on the TB vectors involving proximity testing can be employed with the meteorological-geophysical parameters when seeking to constrain the algorithm solution subsets. The meteorological-geophysical parameters we use and that have been described and analyzed in the Sanò et al. (2013) and Casella et al. (2013) Parts 1 and 2 studies are called optimal tags. This study focuses on the methodology we have used to acquire the optimal tags and on related analyses of the selected tags to ensure their effectiveness in establishing algorithm solution constraints.

The following divisions of the paper consist of a scientific background discussion (Sect. 2) that lays the scientific groundwork for the appearance of the optimal tags within the CDRD algorithm's framework, a methodology description (Sect. 3) that provides a thorough explanation of how the optimal tags are acquired, a scientific results presentation (Sect. 4) describing which tags reveal themselves as optimally effective insofar as their relationship to the rainfall quantities and the degree to which the relationships hold, and an account of the final conclusions (Sect. 5) providing an overview of the most important results of the analysis and their significance for this study and perhaps other independent studies which may find these results beneficial.

2 Scientific background

It has been clearly understood for over a half-century that the nature of precipitation reaching the surface is directly related to ambient meteorological, i.e., dynamical-thermodynamical-hydrological (DST) conditions – as well as various geophysical conditions. Processes and parameters generally accepted as correlating with precipitation intensity include the relative instability of the lower atmosphere, the amount of available moisture and degree of moisture convergence into the precipitation zone, the magnitudes of the mid- and upper-level vertical velocities for the case of convective precipitation, and the strength of upper level positive vorticity advection in creating upper level divergence that enables vertical storm development and outflow efficiency. Other factors such as the degree of orography and the convective inhibition have also been noted as strong correlators with precipitation intensity. In the context of the CDRD algorithm, as we have framed the problem, it is important to objectively and quantitatively determine, for a given satellite TB observation, which of the many dozens of possible DST and geophysical parameters (i.e., the possible optimal tags) are most appropriate to help isolate a constrained set of microphysical solution profiles from within an even larger set of profiles determined solely by TB vector proximity testing (generally a much larger set) – all such profiles residing in the CDRD's a priori database (currently consisting of approximately 2.5 million profiles). Determination of a manageable set of these tags gives rise to the optimal tags.

The emphasis here is finding a set of optimal tags that can serve to reduce the size of an initial set of solution profiles determined only by TB vector proximity testing, so as to end up with a final constrained set of profiles that are more congruent with the ambient meteorological-geophysical conditions associated with the TB observation. Such a result is then presumed to lead to the reduction or even elimination of ambiguity in the final solution as determined by the algorithm solver. This problem is not trivial. First of all, it is important that any optimal tag not be highly correlated with any of the TB components making up a TB vector, as this would render

that particular parameter no more valuable than the given TB component itself in acquiring the subsets of possible solution profiles. By the same token, partial correlations would be expected with at least some of the TB components since if there were none, that would suggest that the TB components themselves would be unrelated to precipitation intensity. That state of affairs would be in violation of the physics of the problem. Furthermore, it is important that any optimal tag not be highly correlated with any other optimal tag, as that would render the latter parameter tag ineffectual in effecting any further constraint. Again, partial correlations could be expected for the very reason that distinct meteorological processes and quantities in the atmosphere are simply not independent of one another.

The foundation that CRM used to produce the CDRD algorithm's a priori database contains within its physical and numerical formulations over 125 separate DST/geophysical parameters and rainfall variables. The CRM itself is the regional/mesoscale Nonhydrostatic Modeling System (NMS) of Tripoli (1992a) and Tripoli and Smith (2013a, b), run in CRM mode. Not all of the many NMS parameters have a direct relationship with rainfall. In fact, only about 20 % of the total number of parameters can be plausibly argued to exhibit a meaningful relationship. Using physically-based reasoning, we have identified 24 DST parameters from within the NMS equations that are likely to exhibit some degree of correspondence with rainfall. We also consider one geophysical parameter, that being the terrain elevation (i.e., surface height) in question. Table 1 provides a summary of the 25-parameter set with which the optimal tag analysis takes place, while Appendix A provides detailed definitions and explanations of the 25 parameters, including their underlying physical relationships with rainfall.

3 Methodology

The process of preparing the simulation database needed for the optimal tag analysis requires two modeling systems, a CRM model system and a RTE model system. These two models are used in combination to create a large simulation database involving thousands of members entailing meteorological-microphysical profile and scalar parameters along with a set of geophysical parameters, each member of which is associated with a TB vector calculated at preferred PMW frequencies and polarizations – specifically those corresponding to the PMW radiometer channels of interest. Note that the meteorological-microphysical information includes the main rainfall variables of interest to which the optimal tags are to be associated, specifically the surface rain rate (RR_{surf}) with an associated flag indicating whether the rainfall mass reaching the surface is liquid or frozen (referred to as the LF flag), and the columnar liquid/ice/total water paths (LWP/IWP/TWP) of precipitating hydrometeors. Once

the simulation database is prepared, three statistical analysis schemes are used in selection of the optimal tags.

3.1 Description of Nonhydrostatic Modeling System (NMS)

The foundation that CRM used for this investigation is the Nonhydrostatic Modeling System (NMS), originally developed by Tripoli (1992a) with more recent major improvements concerning the model's dynamical conservation properties and its unique variable step topography (VST) surface coordinate system described by Tripoli and Smith (2013a, b). The NMS is a 3-dimensional, nonhydrostatic, nested, scalable regional-mesoscale prognostic model. It is able to simulate atmospheric phenomena across all relevant scales from the microscale, involving such phenomena as turbulence or smoke plumes, up through mesoscale addressing phenomena such as water spouts, severe storms and tornadoes, out to the regional/synoptic scales where weather systems such as tropical and extra-tropical cyclones, frontal systems and massive high pressure events can be addressed. This model is chosen because of its ability to achieve accuracy in simulating scale-interaction processes through imposition of conservation on mass, energy, momentum, vorticity and enstrophy throughout model integration. The underlying model framework uses quasi-compressible closure formulated on an Arakawa "C" grid, cast on multiple-nest rotated spherical grids using multiple two-way nesting. The model employs non-Boussinesq dynamics, two-way grid nesting exchanges, and a unique terrain-following VST vertical coordinate system at its lower boundary. The two-way interactive nesting scheme allows increased resolution in focused areas. VST coordinates are able to capture the dynamical consequences of either steep inclinations or subtly varying terrain features without sacrificing accuracy for any type of terrain-induced slope flows at any scale as shown in Tripoli and Smith (2013a, b). A variable ice/liquid water potential temperature is used as the predictive thermodynamic variable in the model (Tripoli and Cotton, 1981). The advantage in using this quantity is its conserved properties for all phase changes. In so doing, potential temperature, water vapor and cloud water are all treated as diagnostic variables

Physical turbulence associated with diagnosed down gradient sub-grid scale motion is represented by either level 1 or level 2 closure based on the schemes of Redelsperger and Sommeria (1982) or Tripoli (1992b), respectively. Fluxes of heat, moisture and momentum resulting from exchanges with the surface enter the simulation domain as vertical surface boundary fluxes due to physical turbulence. Surface fluxes are determined from optional 1-dimensional surface layer parameterizations of varying complexity (e.g., Louis, 1979; Businger, 1982; Smith et al., 1993). The NMS has been designed to function with any plug-compatible shortwave and longwave radiation parameterizations (e.g., the shortwave models of Ackerman and Stephens, 1987; Liou et al., 1988;

Table 1. Meteorological and geophysical parameters considered for selection of optimal tags.

#	Parameter Name	Symbolic Name	Nominal Units
Meteorological Parameters (24):			
Incompatible/Non-observable – both GFS & ECMWF Models (N^2 , Ri , Fr , H_{PBL} , LH, LCL)			
Incompatible/Non-observable – only ECMWF model (CIN) SPACE			
1	Brunt–Väisälä Frequency Squared	N^2	s^{-2}
2	Richardson Number	Ri	unitless
3	Froude Number	Fr	unitless
4	Planetary Boundary Layer Height	H_{PBL}	km
5	Convective Available Potential Energy	CAPE	$J\ kg^{-1}$
6	Convective Inhibition	CIN	$J\ kg^{-1}$
7	Surface Equivalent Potential Temperature	θe_{surf}	K
8	Surface Skin Temperature	T_{skin}	K
9	Lifted Index	LI	K
10	Lapse Rate from 500 to 850 hPa	$\Gamma_{500-850}$	$^{\circ}C\ km^{-1}$
11	Latent Heating Rate of Column	LH	$^{\circ}C\ day^{-1}$
12	Lifting Condensation Level	LCL	km
13	Freezing Level Height	H_{FL}	km
14	Sensible Heat Flux from Surface	Φh_{surf}	$W\ m^{-2}$
15	Moisture Flux 50 hPa above Surface	$\Phi q_{\Delta 50}$	$g\ m^{-2}\ s^{-1}$
16	Positive Vorticity Advection at 500 hPa	ξ_{500}	s^{-2}
17	Positive Vorticity Advection at 700 hPa	ξ_{700}	s^{-2}
18	Vertical Velocity at 700 hPa	ω_{700}	$hPa\ s^{-1}$
19	Vertical Velocity at 500 hPa	ω_{500}	$hPa\ s^{-1}$
20	Divergence at Surface	DI_{surf}	s^{-1}
21	Divergence at 700 hPa	DIV_{700}	s^{-1}
22	Thickness from 500 to 1000 hPa	$\Delta Z_{500-1000}$	m
23	Thickness from 700 to 1000 hPa	$\Delta Z_{700-1000}$	m
24	Vertical Wind Shear in Lower Troposphere	$\Delta_z V_{LT}$	$m\ s^{-1}\ km^{-1}$
Geophysical Parameter (1)			
25	Surface Height	H_{surf}	km

Morcrette, 1991; and the longwave models of Morcrette, 1991; Schwarzkopf and Fels, 1991; Chou and Suarez, 1994) and cumulus parameterizations (e.g., the models of Kuo, 1974; Betts, 1986; Emanuel, 1991; Kain and Fritsch, 1993).

The microphysics parameterization is a bulk scheme developed progressively, beginning with the work of Cotton et al. (1982) and Flatau et al. (1989) and more recently improved by Panegrossi (2004) and Tripoli (2005). This scheme handles initiations, growth processes and inter-hydrometeor mass exchanges of six individual hydrometeors. Each NMS grid volume can contain any combination of hydrometeors, with mass exchanges between hydrometeors treated as a localized set of processes, formulated into a 6-dimensional upper, off-diagonal matrix of physics interactions. The investigation of Panegrossi (2004) was important in that by use of data assimilation procedures using aircraft-based in situ microphysics and PMW TB measurements, it was possible to obtain more realistic exchange coefficients for mass build-up of snow and graupel particles to prevent an original problem with the NMS bulk microphysical parameterization related to

its tendency to overly-complicate and over-produce graupel mass. The six individual hydrometeors considered in the 2-water/4-ice bulk microphysics scheme are as follows (noting their associated mixing densities are given in the parenthetical expressions): (1) cloud droplets (q_c), (2) rain drops (q_r), (3) pristine crystals (q_p), (4) snow (q_s) (representing snow flakes, rimed crystals and snow pellets), (5) ice aggregates (q_a) and (6) high density graupel/hail particles (q_g). Of these six, all but cloud droplets and pristine crystals precipitate.

The size distributions, density factors and habit issues of the hydrometeors assumed in the NMS simulations, characteristics that have a large impact on the RMS calculations, require close attention. Cloud drops are assumed monodisperse except with respect to the formulations for auto-conversion and ice splintering, in which they are cast in the form of a modified Gamma distribution; see Tripoli and Cotton (1981). Their mixing ratio is diagnosed, while their concentration is specified a priori since cloud water nucleation is not explicitly considered. The typical characteristic diameter (D_C) of cloud droplets is 0.02 mm with the

Table 2. Values of α and β parameters for different values of ice crystal mass used in density formulation given in Eq. (2) for $K = 1$ g.

Crystal Mass m [g]	α	β [cm]
$1.7 \text{ e}^{-7} \leq m \leq 1.0 \text{ e}^{-5}$	0.5	19.2
$m < 1.7 \text{ e}^{-7}$	0.5	51.5
$m > 1.0 \text{ e}^{-5}$	0.419	8.89

density of pure water (ρ_w), i.e., 1 g cm^{-3} . The term D_C is given by the first moment of the size distribution and can be considered a weighted-mean parameter; see Panegrossi et al. (1998, 2004).

Pristine crystals represent newly nucleated cloud ice and are also considered mono-disperse. Both their concentration and mixing ratio are predicted, therefore their mass and size actually change at each grid point, with a typical D_C of ~ 0.24 mm. The density (ρ_p) is derived according to Flatau et al. (1989), starting with a mass–diameter (m – D) relationship:

$$D = \beta \left(\frac{m}{K} \right)^\alpha \quad (1)$$

where α is a non-dimensional exponential factor, β is a size scale factor (in cm) and K is a mass normalization factor (in g). For an equivalent volume sphere, the density becomes

$$\rho_p = \frac{6}{\pi} K \beta^\alpha D^{1/\alpha-3} \quad (2)$$

The three parameters, α , β and K , depend on the crystal mass in such a manner that as the mass (or size) increases, the density decreases. Table 2 gives values of α and β for different regimes of crystal mass for $K = 1$ g.

For the other four hydrometeor categories, size distributions are described with an exponential function:

$$n(D) = A \exp(-B D), \quad (3)$$

where either a constant intercept (A) or a constant slope (B) is assumed for each hydrometeor. In this study, constant intercepts of 0.08 cm^{-4} , 0.014 cm^{-4} and 0.071 cm^{-4} have been assumed for rain drops, snow and graupel particles, respectively, while a constant slope of 3 cm^{-1} (i.e., the inverse of the associated characteristic diameter) is used for aggregates.

As with cloud droplets, rain drops have the density of pure water. Snow represents soft, low-density ice forming when pristine crystals or aggregates become heavily rimed, with their density (ρ_s) formulated according to Macklin (1962):

$$\rho_s = a \left(-\frac{\hat{r} \hat{U}_{\text{imp}}}{T_s} \right)^b \quad (4)$$

in which T_s is the surface temperature of the ice substrates (in $^\circ\text{C}$), \hat{r} is a weighted averaged radius (in μm) and \hat{U}_{imp}

is the weighted average impact velocity of cloud droplets and rain drops (in m s^{-1}). The values for the a and b coefficients are 0.23 g cm^{-3} and 0.44 , respectively, as reported by Prodi et al. (1991). The \hat{r} is calculated by averaging the radii of the cloud droplet and rain drop diameters, weighted by their respective mixing ratios. The \hat{U}_{imp} is calculated by averaging the impact velocity between rain drops and snow (i.e., the difference between the terminal velocities of rain drops and snow) with the impact velocity between cloud droplets and snow (i.e., approximately the terminal velocity of snow alone), again weighted by the mixing ratios of cloud droplets and rain drops. The resultant snow density typically covers a range of values from 0.05 – 0.9 g cm^{-3} . An aggregate, formed by either the collisions of two pristine crystals, two existing aggregates or a pristine crystal and an existing aggregate, has a density (ρ_a) given by Eqs. (2) and (3) for the case of a large crystal mass, as expressed in Table 2 (i.e., $\alpha = 0.419$ and $\beta = 8.89$ cm). The resultant expression is $\rho_a(D_C) = 0.015/[D_C^{0.6}] \text{ g cm}^{-3}$; see Panegrossi et al. (1998). The D_C is given by the characteristic diameter of aggregates (i.e., the inverse of the associated constant slope). Graupel particles are considered hard, high-density ice forming with a fixed density (ρ_g) of 0.9 g cm^{-3} – close to that of pure ice (i.e., 0.91 g cm^{-3}).

In this study's simulation framework, three two-way nested grids are configured, within which the two innermost nests are run at CRM resolution, while the outer nest uses the cloud parameterization scheme of Emanuel (1991) in order to generate convection and stratiform cloudiness. The vertical grid extends to 17 km divided into 36 levels with variable, height-dependent grid spacing. The horizontal grid configuration is comprised of (1) an outer domain of 4500×4500 km at 50-km resolution, (2) a first interior domain of 900×900 km at 10-km resolution, and (3) a second interior and innermost domain of 500×500 km at 2-km resolution. The horizontal and vertical mesh dimensions plus horizontal resolutions and domain sizes for these three grids are summarized in Table 3. In general, initial data for the outer grid can be interpolated from another global model such as the NOAA National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS) model or the European Center for Medium Range Forecasts (ECMWF) model, or from, e.g., imposed horizontally homogenous or inhomogeneous balanced states. For this study, we have used GFS initial data fields to stipulate initial boundary conditions. Simulation cases are selected to ensure thorough sampling over an extensive manifold of multi-channel TBs and across a wide range of meteorological and microphysical conditions containing precipitation.

3.2 Description of RTE Model System (RMS)

An accurate multiple scattering RTE model is needed to transform the CRM-generated meteorological-microphysical information into upwelling passive microwave TB vectors

Table 3. Horizontal and vertical mesh dimensions, horizontal resolutions and domain sizes of nested grid configuration used for NMS simulations.

Grid	Horizontal Mesh	Vertical Mesh	Horizontal Resolution (km)	Horizontal Domain Size (km)
Outer	90 × 90	36	50	4500 × 4500
Middle	90 × 90	36	10	900 × 900
Innermost	250 × 250	36	2	500 × 500

for inclusion in the simulation database. Also, in addition to the RTE model, a single scatter model and a surface emissivity module are needed to complete what we call the CDRD's RTE Model System (RMS). It is the RMS which enables the link between simulated and observed TBs.

The 3-dimensional adjusted plane parallel RTE model developed by Roberti et al. (1994) is used for the multiple scattering calculations. For its relevant gaseous absorption calculations, the RTE model uses the millimeter wave propagation model of Liebe (1985, 1987, 1989), which is designed to accurately calculate O₂ and H₂O absorption coefficients for microwave frequencies up to 1000 GHz, these being the two principal active gases in the cm–mm radiation spectrum. The calculations of absorption, scattering, extinction and phase function properties of any cloud-precipitation medium are arrived at through a variety of applications of a prudently modified Mie scattering model. It is noted that the underlying success of the CDRD's RMS in attaining accurate simulations of any arbitrary microphysical cloud-precipitation situation is attributable to using much more realistic renditions of the optical properties of the multiple types of hydrometeors inherent to the actual precipitation process than have been used in past studies. This is particularly so for the frozen hydrometeors, and the adjustment of classical Mie theory to account for, in a relatively unadorned fashion, the general properties of non-sphericity associated with specific ice habits. The calibration-level accuracy of the CDRD RMS is described and explained in the Part 2 study of Casella et al. (2013).

Plane parallel cloud structures are generated from the cloud model paths but in which the RTE calculations are taken with respect to a 53° slant path typical (\pm a few degrees) of a conical-scanning PMW radiometer. Notably, the RMS is designed for flexibility in selecting, for any simulated PMW radiometer, the channel frequencies and polarizations, the channel spectral widths and spectral response functions, the channel instantaneous-field-of-view (IFOV) elliptical dimensions, the channel noise properties, the radiometer viewing angle and the radiometer antenna(e) response pattern(s) needed for accurate simulations of the upwelling TBs (note, radiometers necessarily carry multiple antennas if they measure over an extended cm–mm spectrum). First, monochromatic upwelling radiances are calculated for each radiometer channel at the same resolution of the NMS inner grid (i.e.,

2 km). Once a set of RMS calculations is complete, instrument transfer functions are used to calculate the final TB vector components associated with the preferred channels. This is accomplished by first integrating the upwelling monochromatic radiances over the channel spectral widths considering each channel's spectral response function. It is then necessary to integrate the channel upwelling radiances over each channel's IFOV, considering all grid elements of the CRM that are included in an IFOV, and taking into account the radiometer antenna pattern and ambient radiometric noise.

The vertical profiles of hydrometeor-specific liquid/ice water contents (LWC/IWC) mixing densities referred to as ($q_c, q_r, q_p, q_s, q_a, q_g$) for cloud droplets, rain drops, pristine crystals, snow pellets/flakes, ice aggregates and graupel/hail particles, respectively, together with vertical temperature-moisture profiles [$T(z)$ and $q(z)$], surface height (H_{surf}) and surface skin temperature (T_{skin}), are required parameters for the RMS RTE model calculations. Furthermore, for any given RTE model calculation, additional inputs are needed for the surface emissivity model concerning the emissive-reflective properties of the given type of surface (TY_{surf}) under consideration, properties actually dependent on the surface's biogeophysical features. For water surfaces the relevant radiometric properties are a function of sea/fresh water surface temperature ($SST = T_{\text{skin}}$), salinity (S) for sea water, and surface roughness height (z_R) for either fresh or sea water – with z_R controlled by near-surface wind speed (V_{surf}); for unfrozen land surfaces the notional radiometric properties are largely controlled by near-surface soil moisture content (S_{MC}), near-surface soil quartz content (S_{QC}), canopy areal index (C_{AI}) and canopy water content (C_{qc}); and finally, for frozen surfaces the radiometric properties are generally dependent on the age of the snow or ice (Δ_{SI}), its granularity condition (G) and its melt-state (M_S). The surface emissivity module of the RMS, described in the next sub-section, intrinsically includes the features embodied by the $S, z_R, S_{\text{MC}}, S_{\text{QC}}, C_{\text{AI}}, C_{\text{qc}}, \Delta_{\text{SI}}, G$ and M_S parameters.

Thus, the RMS RTE model requires the LWC/IWC profiles of the six hydrometeors, profiles of $T(z)$ and $q(z)$, and scalar quantities $H_{\text{surf}}, TY_{\text{surf}}, T_{\text{skin}}$ and V_{surf} . The left-land two columns of Table 4 summarize this collection of NMS-generated scalar and vector (profile) parameters needed by the RMS, along with the foremost NMS-generated scalar rainfall variables. The factors in this table represent the minimal meteorological-microphysical-geophysical information packet for a given database profile member needed by a conventional CRD algorithm. It is the expansion of this packet with additional meteorological and geophysical information (either scalar or vector parameter form) that enables extending a algorithm database into a fully defined CDRD database equipped with optimal tags.

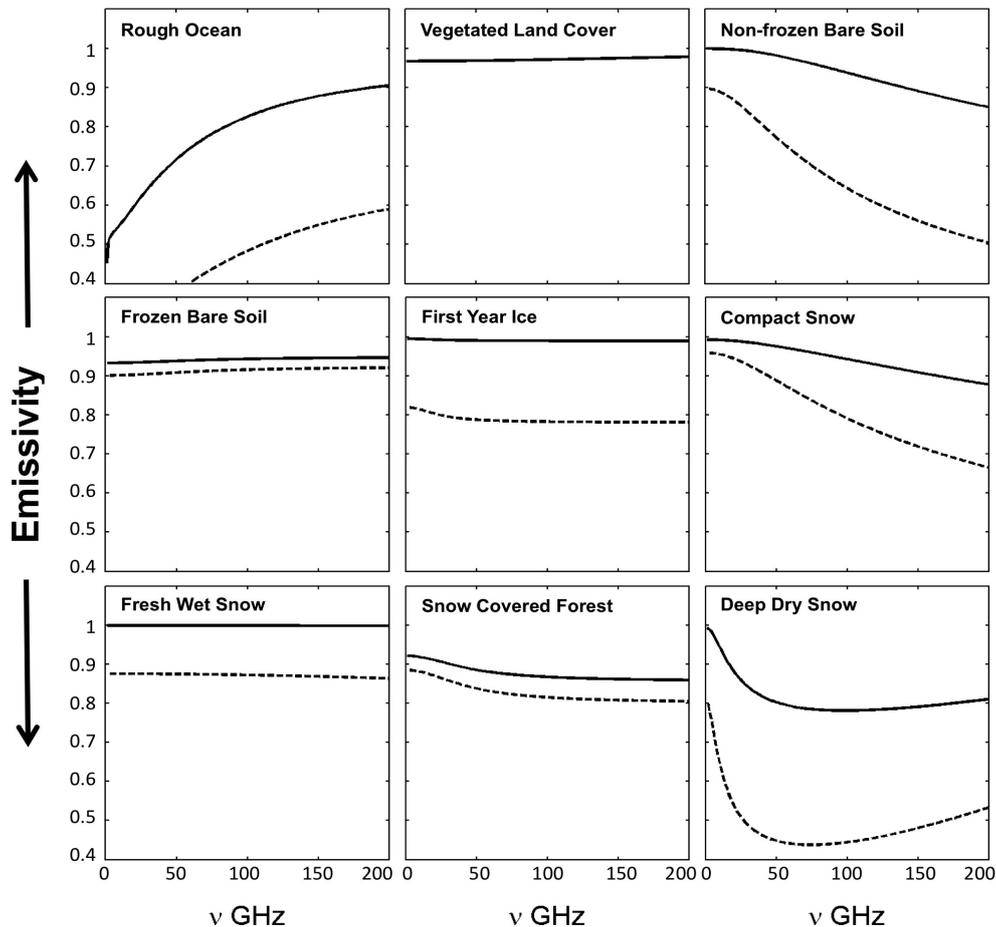


Fig. 1. Emissivity properties for nine surfaces, specifically rough ocean, vegetated land cover, non-frozen bare soil, frozen bare soil, first year ice and four types of snow cover (compact snow, fresh wet snow, snow-covered forest, deep dry snow) – as function of frequency from 0 to 200 GHz. Solid (dashed) lines show vertically (horizontally) polarized emissivities, respectively, with viewing angle fixed at 53-deg. Rough ocean properties are with respect to sea surface temperature of 283 K, surface salinity of 35 ppt and above-surface wind velocity of 2 m s^{-1} (thus roughening surface). Vegetated land cover is assumed to completely depolarize scattered and emitted radiation.

3.3 Description of Surface Emissivity Module (SEM)

To accommodate the various surface backgrounds being used for the CDRD retrieval algorithm, a consistent and quantitative means to acquire characteristic surface emissivities (reflectances) for variable satellite view angles and for both horizontal and vertical polarizations is essential. Thus, we have developed a 9-member surface emissivity module (SEM). For a rough ocean (i.e., an ocean surface undergoing above-surface winds), the SEM employs the ocean emissivity model of English and Hewison (1998); see also Schluesel and Luthardt (1991) and Hewison and English (2000). This scheme calculates accurate estimates of open sea emissivity between 10 and 200 GHz for observation angles up to 60 degrees and winds between 0 and 20 m s^{-1} . For non-frozen land emissivities, we have adopted two different surface emissivity models from Hewison (2001), specifically models for “other forestry” and “bare soil”, which we refer to

as “vegetated land cover” and “non-frozen bare soil”, respectively. For frozen surfaces we have adopted six surface emissivity models from Hewison and English (1999) consisting of “frozen bare soil”, “snow-covered forest”, “first year ice”, “compact snow”, “fresh wet snow”, and “deep dry snow” – noting we have imposed various minor name changes from the originals for the frozen surface cases. We also note that the latter four frozen surfaces may be applied to either ocean or land areas.

Figure 1 shows results from the nine surface emissivity component models adopted for the SEM as a function of frequency (from 0 to 200 GHz) and for H and V polarizations. For the purpose of this diagram, the satellite view angle is held constant at 53° . A number of remarks are pertinent. In the case of rough ocean, the calculations are taken with respect to a sea surface temperature of 283 K, a surface salinity of 35 ppt and an above-surface wind velocity of 2 m s^{-1} (thus producing the roughened surface). Note that HV polarization

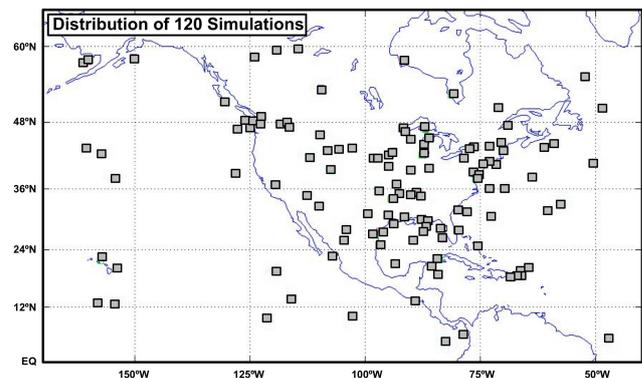
Table 4. NMS-generated microphysical, meteorological and geophysical scalar/vector parameters used for RMS calculations (left-hand two columns) and NMS-generated scalar rainfall variables (right-hand column).

NMS-generated hydrometeor-specific profile parameters used for RMS calculations	Other NMS-generated profile & scalar parameters used for RMS calculations	NMS-generated rainfall-specific scalar variables
Cloud Droplet LWC q_c Profile (g cm^{-3})	Atmospheric Temperature Profile [$T(z)$ in K]	Liquid/Frozen Flag for Surface Rain [LF]
Rain Drop LWC q_r Profile (g cm^{-3})	Atmospheric Water Vapor Mixing Ratio Profile [$q(z)$ in g kg^{-1}]	Surface Rain Rate [RR_{surf} in mm h^{-1}]
Pristine Crystal IWC q_p Profile (g cm^{-3})	Surface Height [H_{surf} in km]	Liquid Water Path [LWP in kg m^{-2}]
Snow Pellet IWC q_s Profile (g cm^{-3})	Type of Surface [TY_{surf}]	Ice Water Path [IWP in kg m^{-2}]
Ice Aggregate IWC q_a Profile (g cm^{-3})	Surface Skin Temperature [T_{skin} in K]	Total Water Path [TWP in kg m^{-2}]
Graupel Particle IWC q_g Profile (g cm^{-3})	Near-surface Wind Speed [V_{surf} in m s^{-1}]	

differences are large but fairly constant regardless of any variation of emissivity itself with respect to frequency. In the case of vegetated land cover, it is evident that the model predicts emissivity close to the value of 1.0 with no variation in regard to polarization state, nearly constant with frequency, and larger than any ocean emissivity across the analyzed frequency range. Similar to rough ocean, non-frozen bare soil and first year ice also exhibit large HV polarization differences whereas frozen bare soil exhibits very small and nearly constant differences (it is noted that we use non-frozen bare soil to represent desert surfaces). For snow cover, the H and V emissivities vary significantly depending on snow conditions (because of different grain sizes and whether melted water is present) both in average values and HV polarization state differences, with deep dry snow exhibiting the greatest differences. It is important to recognize that in using the RMS, the SEM is applied according to the corresponding surface type defined by the NMS in which any snow conditions that are assigned are actually inferred from the NMS's soil temperatures and geographical locations.

3.4 Generation of simulation database

For a one year period from November 2007 to October 2008, 120 individual NMS simulations using the nested grid scheme described in Table 3 are made over the North American region for pre-selected precipitating storm events. GFS optimal analysis data are used to define initial conditions for the NMS and to describe the outer boundary conditions for the outer grid, updated every 6 h of simulation time throughout the individual simulation runs. Simulation runs are integrated from 18 to 36 h, prefaced by a 12-h spin-up pe-

**Fig. 2.** Locations of 120 simulations generated over North American region for creation of simulation database; 68 over water and 52 over land.

riod, depending on the development and dissipation periods of the particular weather systems. The 12-h spin up time is required to allow local forcing to develop and stabilize gradually. The 120 simulation are uniformly distributed over time such that 10 simulations are produced for each month. The spatial distribution of simulation domains is shown in Fig. 2. Within this distribution, 68 simulations take place over water while 52 take place over land.

After the spin-up period and during the simulations, meteorological-microphysical profile and scalar parameters, geophysical parameters and TB vectors over all grid points in the inner grid nest are saved hourly in the database whenever any single point in the inner domain contains a liquid or frozen form RR_{surf} of at least 0.01 mm h^{-1} . By saving data hourly in simulation time means that a great variety

of meteorological-microphysical-geophysical-radiation realizations describing precipitation systems at different development, maturity and dissipation stages become included in the database. Profile data are saved at all 36 vertical levels for each grid point along with all essential scalar parameters. There are a total of 128 meteorological DST profile parameters besides various other meteorological scalars, including the four scalar rainfall variables and liquid/frozen LF flag. There are also five scalar geophysical parameters that can be used for constraint tags and/or algorithm logic; these are described in Sanò et al. (2013).

Finally, the microwave frequencies of 10.65, 19.35, 22.23, 23.8, 36.5, 85.5, 89.0 and 150 GHz form the principal elements of the TB vectors, noting all frequencies are calculated at horizontal (H-pol) and vertical (V-pol) polarizations with the exception of 22.23 and 23.8 GHz, which are just calculated at V-pol. These frequencies-polarizations are selected because they correspond to a number of radiometer channels now in use on current satellites. Upon completion, some 2.14 and 1.84 million database members are obtained over water and land, respectively.

As noted at the end of Sect. 2, the 24 DST parameters and one surface height geophysical parameter are identified in Table 1 and described in detail in Appendix A. These parameters have been chosen based on their potential to provide diagnostic constraint information in helping to differentiate atmospheric states that can initiate and support various types of precipitation events and thus serve as optimal tags. These parameters provide information concerning (a) the stability of the atmosphere, (b) the amount of mesoscale forcing (such as by local surface convergence, mid-level vertical velocity and wind shear), (c) the amount of large scale dynamical forcing (by mid- to upper-level divergence governed by potential vorticity advection), (d) the degree to which low-level moisture is available to support convection, (e) the vertical levels of the PBL, convection onset and freezing, (f) the surface fluxes of heat and moisture and the release of latent heat, (g) the surface pressure tendencies as defined by low and deeper level pressure-layer thicknesses, and (h) the influence of topography in promoting orographic lift as measured by the elevation of the terrain. Note that all the parameters in Table 1 are defined in the simulation database at 50-km outer grid resolution – so as to make them congruent with counterpart observed parameters enabled by the optimally assimilated gridded meteorological datasets used for initialization of the GFS or ECMWF operational forecast models. It is also noted that Table 1 identifies seven parameters that cannot, in fact, be used in conjunction with operational CDRD algorithms because they do not have compatible or actual counterparts, in relationship to the NMS, insofar as observed meteorological parameters enabled by the operational models. These are (1) Brunt–Väisälä Frequency Squared, (2) Richardson Number, (3) Froude Number, (3) Planetary Boundary Layer Height, (5) Latent Heating Rate of Column, (6) Lifting Condensation Level – no coun-

terparts from either GFS or ECMWF models, and (7) Convective Inhibition – no counterpart from just ECMWF model.

Over the last 30 yr, with higher operational global forecast model resolutions now prevalent and with gradually improving physical parameterizations and continuous data assimilation techniques, the initial condition accuracies of operational prediction models and subsequent quality of predictions out from 6 to 48 h have significantly improved. However, the operational models cannot resolve small-scale processes due to turbulence, convection and other cloud processes such as microphysical mass and latent heat exchanges and the concomitant radiative flux variations. Although at times they can produce reasonably accurate renditions of the flux exchanges of enthalpy, heat, moisture and momentum, they must do so by relying on physical parameterization schemes that are more practical than realistic. That is why turning to a CRM to simulate cloud processes down to the horizontal scales that are actually at work in generating clouds and precipitation is so important when trying to interpret satellite PMW information that arrives at relatively high resolutions, depending on the actual PMW frequency.

By the same token, when seeking to constrain the CDRD algorithm's selection of potential Bayesian solution profiles using observed meteorological-geophysical guidance parameters, we remain cognizant of the fact that the notional observed meteorological-geophysical parameters that we apply as optimal tags are not yet available at cloud resolving scales – currently only at synoptic scales (i.e., order 50 km). For example, the current NOAA/NCEP Hybrid EnKF Global Ensemble Forecast System (GEMS) and the current ECMWF Cycle 38r1 Ensemble Prediction System (EPS) operational model resolutions are 35 km and 30 km, respectively (note GEMS is the 22-member ensemble version of the GFS forecast model suite while EPS is the 51-member ensemble version of the ECMWF forecast model suite). We are almost certainly on the order of a decade away from the time when cloud resolving global forecast models become a reality and the type of algorithm we are now exploiting will be able to conduct its observational data gathering tasks at high resolution scales, i.e., presumably at or below 2-km resolution.

3.5 Description of statistical analyses used to identify optimal tags

In determining the optimal tags, three statistical analysis procedures are used: (i) the 1st is a parameter-by-parameter linear correlation analysis with respect to three principal rainfall variables – RR_{surf} , LWP, and IWP – applied on a seasonal basis; (ii) the 2nd is a 2-dimensional cross-tabulated histogram analysis used to confirm how well-behaved the strongly correlating parameters are insofar as their statistical relationships to the rainfall variables; and (iii) the 3rd is a multi-linear regression analysis used to determine the effectiveness of the optimal tags, in combination with TBs, in

regressively determining estimates of the rainfall variables. This 3rd analysis helps justify the claims in the Part 1 and 2 papers of Sanò et al. (2013) and Casella et al. (2013) that the optimal tags can be used effectively for improving CDRD rainfall retrievals.

4 Results of analysis

The simulation database is first divided into over-water and over-land partitions. For the purpose of the 1st and 2nd statistical procedures, the parameter-by-parameter linear correlation analysis and the 2-dimensional cross-tabulated histogram analysis, both partitions are analyzed. However, for the 3rd procedure, the multi-linear regression analysis, only the over-water partition is analyzed. The latter restraint is essential in preventing significant errors appearing in the regression results due to the fact that detection of rainfall for over-land backgrounds using PMW TB measurements is far more uncertain than with respect to over-water backgrounds. This is because TB-delineated precipitation in its degree of contrast to over-land surfaces is much less detectable than in conjunction with over-water surfaces. Therefore, in seeking to determine rainfall using *statistically* formulated multi-linear regression relationships involving use of TB predictor variables for the over-land case, it is difficult to obtain unambiguous conclusions concerning the effectiveness of the optimal tags within the regressions. Note the emphasis here is on *statistical* regression relationships; these types of relationships are actually not used in the physically and Bayesian formulated CDRD algorithm – they are used here simply for demonstrating the potential of optimal tags as Bayesian solution constraints.

Note that the linear correlation analysis is performed on a seasonal basis to prevent blurring of the relative relationships. This is important since the optimal tags should be effective for all seasons, not just particular seasons. For the over-water (over-land) partitions of the database, the percentages of database members for the winter, spring, summer and autumn seasons are 39.1 % (23.8 %), 29.4 % (26.4 %), 14.0 % (31.1 %) and 17.5 % (18.7 %), respectively. Since H_{surf} is zero for the over-water partition, it is not included in the associated linear correlation analysis. However, for the over-land partition, H_{surf} is used and emerges as one of the seven optimal tags. The process of identifying the six meteorological optimal tags is as follows.

4.1 Parameter-by-parameter linear correlation analysis

The 1st statistical analysis procedure uses parameter-by-parameter linear correlation analysis to identify which of the 25 target parameters correlate best with the RR_{surf} , LWP, and IWP rainfall variables. This procedure is first applied to the four over-water seasonal groups of 24 meteorological parameters, which have been summarized in Table 1; note that H_{surf}

plays no role with the sea level database members. Within each group, the individual parameters are assigned totalized correlation scores based on summing the absolute values of their respective linear correlation coefficients in conjunction with the three rainfall variables, then ordered according to the magnitudes of their total scores from highest to lowest. Next, parameters are discarded for further consideration if they do not indicate at least one absolute value of correlation coefficient exceeding 0.25, in conjunction with the three rainfall variables and the winter, spring and autumn seasons, with the threshold lowered to 0.15 for the summer season when all correlation coefficients systematically decrease. Furthermore, for a parameter to remain in play, it must pass the threshold test for all four seasons – otherwise it is discarded. Finally, the parameters passing these two tests are grouped into two sets, one for which counterpart observational parameters are available and a second for which there are no counterpart parameters. This process is then repeated for the over-land partition which now includes surface height as a possible parameter.

For each of the over-water and over-land partitions, six meteorological parameters survive the two correlation tests for which observational counterparts are available. These are as follows: (1 and 2) vertical velocities at 700 and 500 hPa (ω_{700} , ω_{500}), (3) equivalent potential temperature at surface (θe_{surf}), (4) convective available potential energy (CAPE), (5) moisture flux 50 hPa above surface ($\Phi q_{\Delta 50}$), and (6) freezing level height (H_{FL}). Table 5 provides an example of the over-water correlation results for the winter season. It is interesting why both ω_{700} and ω_{500} score well in the correlation tests. The Table 5 results show that ω_{700} is slightly more strongly correlated than ω_{500} in conjunction with RR_{surf} and LWP, whereas ω_{500} is much more strongly correlated with IWP. The reason for this reversal is related to scale. At 700 hPa, vertical circulations are closely associated with finer scale eddies and the roots of convection (see Appendix A), that is, they are tied in with cloud bases and condensation processes. Alternatively, at 500 hPa, which is typically above the freezing level, vertical motions are more wave like (e.g., vertical Rossby wave propagation) and tend to describe the vertical ascents needed to support ice formation. Thus, for the CDRD algorithm, the ω_{500} parameter can be discarded if only RR_{surf} in either liquid or frozen form is the sole interest, but retained if LWP/IWP are also of interest.

For the over-land partition, similar results are obtained but with the additional result that H_{surf} also passes its linear correlation tests. This is because elevated terrain is so effective in mechanically stimulating vertical motions and possibly convection and precipitation. Thus, the linear correlation procedures applied to the two partitions identify six meteorological parameters and one geophysical parameter (i.e., ω_{700} , ω_{500} , θe_{surf} , CAPE, $\Phi q_{\Delta 50}$, H_{FL} and H_{surf}) as possible optimal tags, contingent on their performance in conjunction with the 2-dimensional cross-tabulated histogram analysis and the multi-linear regression analysis.

Table 5. For cases of winter season and over-water, summary of parameter-by-parameter linear correlation results for 6 optimal tags.

#	Parameter	RR _{surf}	LWP	IWP
	Optimal			
1	ω_{700}	−0.40	−0.42	−0.40
2	ω_{500}	−0.39	−0.41	−0.67
3	$\theta_{e_{surf}}$	0.25	0.31	0.40
4	CAPE	0.21	0.25	0.30
5	$\Phi_{q\Delta 50}$	0.33	0.37	0.18
6	H_{FL}	0.22	0.28	0.36

Of further interest are three parameters which score well in the linear correlation tests but for which observable counterparts from the operational global forecast model initial data analyses are simply not available. These are the Brunt–Väisälä Frequency Squared (N^2), the Richardson Number (Ri) and the Planetary Boundary Layer Height (H_{PBL}). Each of these parameters is strongly correlated with the three rainfall variables. These three parameters remain as tantalizingly useful information for future applications, although as shown in the next sub-section, it is possible that Ri might not serve as an effective constraint. This is because of its complex statistical relationship with TWP, denoting that strong correlations by themselves may hide statistical features in particular parameters that belie their usability as optimal tags.

4.2 Two-dimensional cross-tabulated histogram analysis

The 2nd statistical analysis procedure is used to confirm, by qualitative examination of 2-dimensional histograms of relationships between the contingent optimal tag parameters and the rainfall variables, whether the individual tags are physically acceptable for use as CDRD algorithm constraints. This check is necessary in determining whether strong correlative behaviors as determined in the first test are not simply pathological in nature, but instead indicative of tag-rainfall relationships that contain relatively continuous and uniform value-to-value associations. To accomplish this, we cross-tabulate the contingent optimal tag quantities with rainfall variable quantities using normalized frequency counts within small, discrete 2-dimensional bin intervals to produce histogram array diagrams, in which color is used to illustrate normalized bin frequencies. Figure 3a presents nine of these histograms representing selections from all four seasons of the year for the over-water database partition, each panel illustrating relationships between selections from all six contingent meteorological tags (abscissa values) and from all three rainfall variables (ordinate values). The ω_{700} , H_{FL} and CAPE tags are repeated for two seasons each. Both abscissas and ordinates are plotted in terms of \log_{10} scales. Histogram arrays and presented in color (refer to rhs color bars), for which summed histogram counts are transformed into either

un-scaled normalized frequencies (i.e., indicated above color bars by NF) or percent-scaled normalized frequencies (i.e., indicated by NF%).

The various diagrams shown in Fig. 3a are typical of the types of relationships found between the tag parameters and the rainfall variables for the different seasons, in which generally the clearest relationships are found with respect to LWP, followed by RR_{surf} and TWP, and finally by IWP. Certain of the tags' clearest relationships are in conjunction with IWP, dominated by the case of the ω_{500} tag (see panel 3 in Fig. 3a), whereas a few others for specific seasons tend to demonstrate their clearest relationships with respect to TWP, e.g., as with CAPE for autumn (see panel 9 in Fig. 3a). It is noted that some of the relationships are relatively linear and sloped in nature, some more nonlinear, some tending to lie parallel along either the abscissa or the ordinate for part of or all of their distribution domain and all of them show some degree of scatter along whatever underlying functional form is evident. It is important to recognize that no specific aspect of these various tag-rainfall relationship features in any way inhibits their effectiveness in the CDRD algorithm methodology, as described by Sanò et al. (2013) and Casella et al. (2013). Nor does it matter in the context of the algorithm that the statistical relationships vary between different rainfall variables. This is because the types of statistical relationships presented in Fig. 3a are not used in the algorithm, and thus cannot govern the algorithm's constraint methodology.

In the constraint methodology, simulated optimal tag values are actually used in Euclidian norm measures with respect to observed optimal tag values. Their purpose is to help identify and eliminate microphysical-meteorological profile members from Bayesian profile solution subsets that have been initially selected on the basis of TB information only but do not pass muster insofar as their likelihood of being congruent with the meteorological observations. The Euclidian norm measures are actually cast within a probability framework, and therefore for a number of important reasons, can entertain any or all of the characteristic relationships seen in Fig. 3a as well as more compacted versions of these relationships, as we have found in some of the histogram patterns (diagrams not shown). What the tag-rainfall relationships cannot do is be highly discontinuous or highly complex – such as the complex behavior illustrated in Fig. 3b in conjunction with the Richardson Number (discussed below). In other words, the constraint methodology of the CDRD algorithm is impervious to compactness, scatter and/or zero or infinite slope behaviors in the tag-rainfall relationships, because the methodology treats tag values more in the form of probabilistic discriminant measures than in the form of mathematical functions. Accordingly, we find that the six contingent optimal tags, for both database partitions, satisfy the 2nd of the statistical analysis procedures.

An examination of Fig. 3b showing the tag-rainfall relationships for the three non-observable parameters N^2 , H_{PBL} and Ri (again for the over-water partition), indicates that the

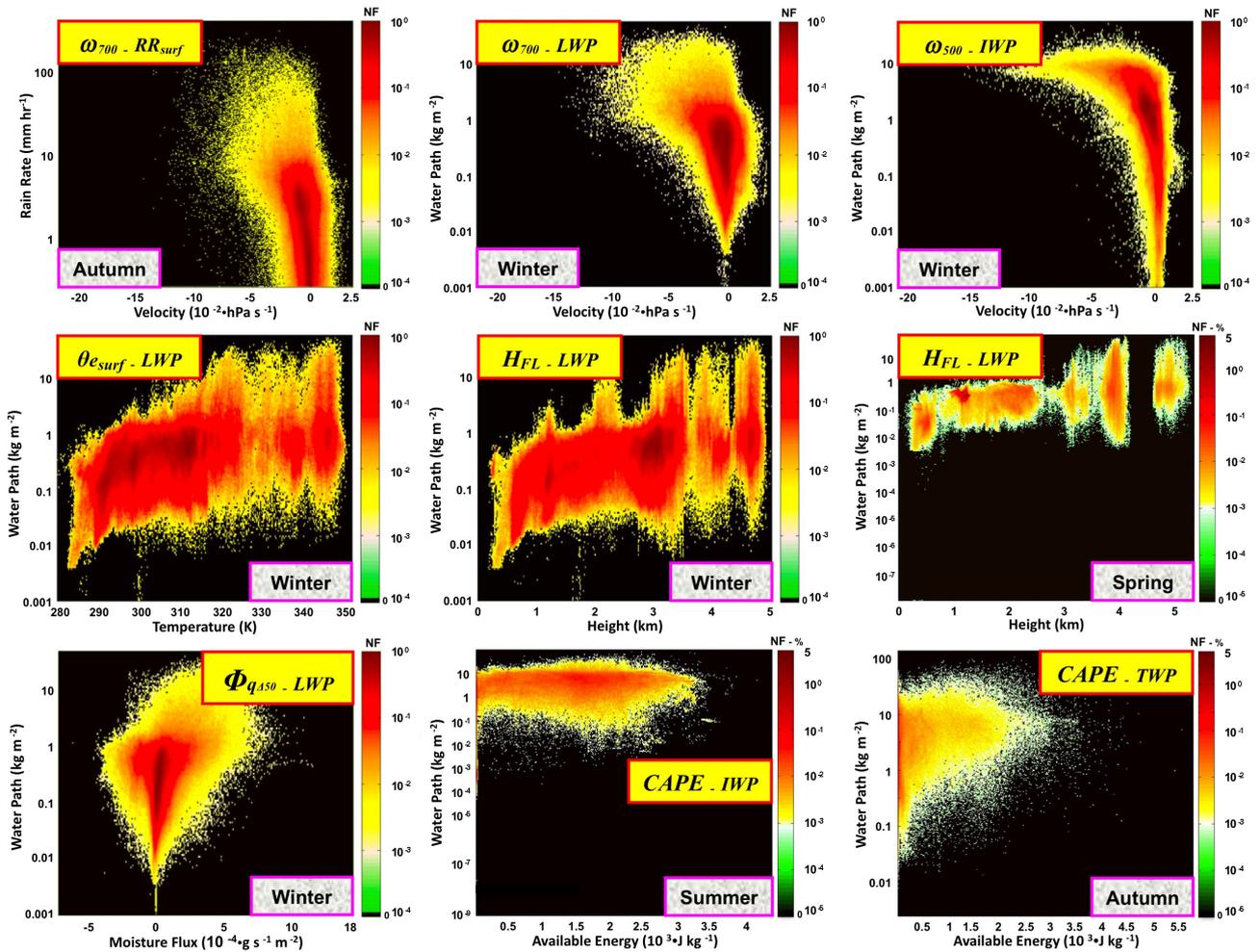


Fig. 3a. Nine 2-dimensional histograms (over-water partition) of selections from all four seasons, illustrating relationships between selections from all six contingent optimal meteorological tags (abscissas) and from all three rainfall variables (ordinates), noting ω_{700} , H_{FL} and CAPE parameters are repeated for two seasons each. Both abscissas and ordinates are given on \log_{10} scales and either unscaled normalized frequencies (NF) or percent-scaled normalized frequencies (NF%) present histogram frequencies over color ranges described by individual color bars.

first two of these parameters might serve effectively as optimal tags, but that the third, i.e., Ri , likely would not. It is evident that in panel 3 of Fig. 3b that there are underlying strands of association loci that describe how Ri is functionally related to TWP, all of which emanate out from the abscissa value of 0.25, i.e., the value of the critical Richardson Number (Ri^C). There are various possible reasons for this, but are speculative. The relevance of the diagram is that the Ri parameter, were it available as an observable, would likely not be a wise choice for a CDRD algorithm constraint tag because its behavior is too complex to serve effectively in a discriminant-based framework.

4.3 Multi-linear regression analysis

The 3rd statistical analysis procedure is used to confirm whether the contingent optimal tags actually exert influence

in improving explained variances when an

$$RV_n = \text{MLR}^{\text{nom}}(\text{TB}_i) \quad (5)$$

expression, is extended to an

$$RV_n = \text{MLR}^{\text{ext}}(\text{TB}_i + \text{OT}_j) \quad (6)$$

expression, where RV_n is one of three rainfall variables (i.e., RR_{surf} , LWP or IWP), MLR^{nom} and MLR^{ext} are nominal (nom) and extended (ext) multi-linear regression functional relationships for which $i = 1$, NF is a sequence of TB parameters in conjunction with a set of frequency-polarization states and $j = 1$ NT is an optimal tag (OT) sequence for partial or complete inclusion in stepwise fashion into the MLR^{ext} . The stepwise regression process is supervised by a Bayesian information criteria (BIC) scheme (see Schwarz,

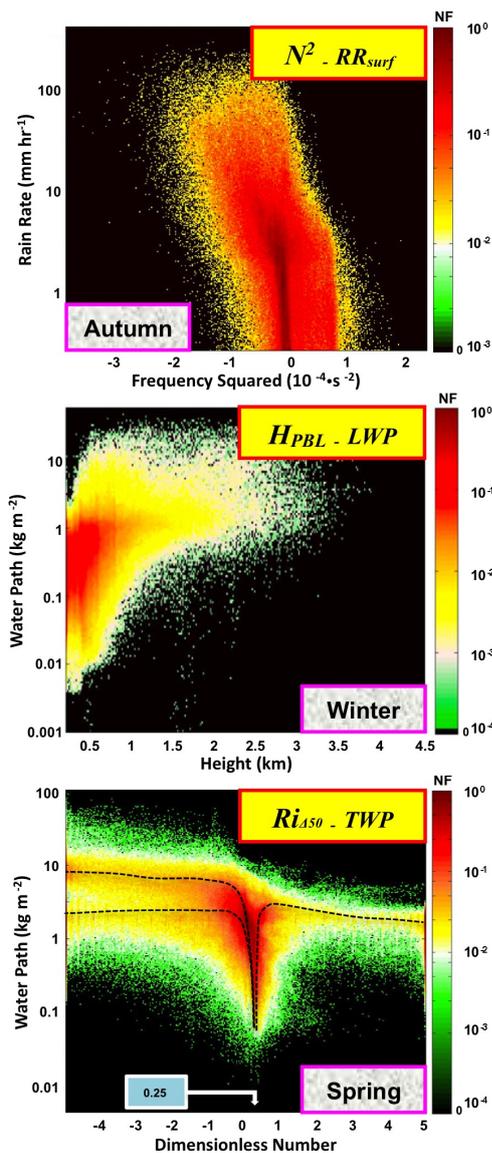


Fig. 3b. Similar to Fig. 3a except for three selections of N^2 , H_{PBL} and Ri , noting that N^2 and Ri calculations take place 50 hPa above surface.

1978), which objectively, but not necessarily sequentially, inserts OT_j parameters into the MLR^{ext} according to their ability to boost the explained variance at each new step. As parameters from the OT_j set are included in the order selected by the BIC scheme (denoted by index k), the percentage explained variances ($\% \varepsilon_k^2$) associated with the individual OT_k s of the associated sequence of MLR_k^{ext} relationships are tallied and analyzed for the effectiveness of the OT_k s in improving upon the MLR^{nom} nominal percentage explained variance ($\% \varepsilon_0^2$). If the $\% \varepsilon_k^2$ sequence indicates steady and significant improvement relative to the ε_0^2 value for a given RV_n , then the associated OT_k set is confirmed as effective for CDRD algorithm applications for that particular RV_n .

In the design of the multi-linear regression process, the nominal MLR^{nom} regression relationships are established in *optimized linearized form* by \log_{10} transformations of the response RV_n (thus adjusting for their positive skew toward small water paths or light rain rates) and transformation of the TB parameters to normalized polarization difference (NPD) predictor parameters (a unitless quantity). Following Petty (1994), the form of an NPD is taken as follows:

$$NPD \equiv \frac{T_V - T_H}{T_{V,O} - T_{H,O}} \quad (7)$$

where T_V and T_H are the observed vertical (V-pol) and horizontal (H-pol) linearly polarized TBs at a given frequency, and $T_{V,O}$ and $T_{H,O}$ are the clear sky (cloud-free) V-pol/H-pol TBs for the same scene (most effectively determined climatologically from an independent TB dataset).

When viewed at an oblique angle over water as measured by conical scanning PMW radiometers, the observed difference between T_V and T_H is dominated by the polarized emissivity of the water surface such that $NPD = 1$ represents a completely cloud-free situation, regardless of a precipitation radiometer's microwave frequency. On the other hand, for a completely opaque situation due to cloud and precipitation hydrometeors (both liquid and frozen phases), NPD will become nearly 0 for the lower radiometer frequencies ($f \leq 60$ GHz) over which hydrometeor scattering is negligible (but not zero) and does not induce meaningful polarization, while for the higher frequencies ($f > 60$ GHz), NPD will become small (order 0.01–0.1) and be entirely dependent on the hydrometeor scattering-induced polarization signal controlled by water phases, hydrometeor size distributions, ice densities and hydrometeor shapes/habits. By normalizing the observed polarization difference with the cloud-free polarization difference, the TBs' sensitivity to cloud water in a column is isolated from its sensitivity to the ocean surface emissivity such that an NPD is nearly directly proportional to column transmittance. An NPD parameter thus exhibits a monotonic relationship to increasing optical depth due to the increase of hydrometeor densities, and for most of the microwave frequency range applicable to precipitation, retrieval ($f \leq 100$ GHz) is only weakly sensitive to the scattering effects of the cloud within both liquid and ice layers. These advantages make NPDs better predictor parameters than the underlying TBs since the former correlate better with the \log_{10} transformations of the rainfall variables. Note that NPDs can only be calculated for polarized channels; it is important to note that the NPD idea is really a simplification of the idea behind the polarization corrected temperature (PCT) of Spencer et al. (1989), a parameter that was developed for simplifying the delineation of precipitation using H-pol/V-pol differentiated conical-scan radiometer channels, with the added advantage that a PCT responds nearly linearly to rain rate for either low or high frequencies.

The nominal MLR^{nom} models can be tested for their linearity, based on two assumptions: (a) the associated residuals are independent and (b) the residuals are normally distributed with zero mean and constant variance. The $\% \varepsilon_0^2$ is the best indicator to identify and assess the goodness-of-fit of a MLR^{nom} model. For the over-water database partition, based on $\% \varepsilon_0^2$ calculations with the mix of possible NPDs calculated for the database, NPDs at 10.65, 19.35 and 36.5 GHz frequencies are found to be the most effective predictor parameters for \log_{10} transformations of the response RV_n . A quantile-quantile (Q-Q) plot (also called a normal probability plot) can be used to graph the quantiles for a distribution of residuals taken for a given MLR^{nom} model against equivalent quantiles from a theoretical normal distribution (note, quantiles are points taken at regular intervals from a cumulative distribution function (CDF) of a random variable). This is effective in checking how well the probability distribution of the model residuals agrees to that for a perfect normal distribution.

As an example, the left-hand panel of Fig. 4 shows that the distribution of individual NPD residuals of the over-water MLR^{nom} [$\log_{10}(\text{LWP}) \approx \text{NPD}(10.65, 19.35, 36.5)$] model exhibits nearly constant variance over the range of fit of $\log_{10}(\text{LWP})$. Then it is shown with a Q-Q plot in the right-hand panel of Fig. 4 that over the ± 3.5 quantile intervals of the normal distribution (nearly all of the data) that the quantiles from the MLR^{nom} model residual distribution are in relatively close agreement with equivalent quantiles from a perfect normal distribution. The overall behavior of the Q-Q plot corroborates the intrinsic degree of linearity in the MLR^{nom} model (proximity of the Q-Q plot to the $x = y$ reference line), with associated near-agreement in scale, skewness and extensions of the distribution tails. In fact, we find this to be a far better degree of linearity than with respect to the associated non-transformed MLR^{nom} [$\log_{10}(\text{LWP}) \approx \text{TB}(10.65 \text{ H/V}, 19.35 \text{ H/V}, 36.5 \text{ H/V})$] model. Similar results are found for the over-water RR_{surf} and IWP variables and for the full RV_n set of the over-land partition.

As noted, the BIC supervisory scheme attempts to determine a regression model that best explains the predictor and response data (goodness-of-fit), noting it attempts to do so with a minimum combination of variables (i.e., a model with the optimal combination of OT_k predictor parameters that results in maximum response variable precision). Generally, in selecting predictor parameters for regression models through the technique of maximum likelihood estimation, an essential technique for the BIC scheme to operate, overfitting may result (maximum likelihood refers to the *probabilities* of the observed results being as large as possible). This is because it is possible to increase the likelihood of the estimates by simply adding more and more parameters. The advantage of the BIC scheme is that it not only awards the goodness-of-fit, but also penalizes overfitting gauged by the BIC log likelihood

expression:

$$\text{BIC}_{\text{LL}} = -2 \cdot [n + n \log(2\pi) + n \log(\text{RSS}/n)] + (p + 1) \cdot \log(n), \quad (8)$$

where n is the number of observations, p is the number of parameters used in the model, RSS is the residual sum of squares and RSS/n is the maximum likelihood estimate. Note that the second term on the rhs of Eq. (8) is the penalty term. The preferred model is the one with the smallest BIC_{LL} value. After a training dataset is extracted from the partition in question (25 % of the samples are randomly selected for this purpose with the remainder used for the independent test dataset), the BIC scheme is applied using the open source statistical package R (<http://www.r-project.org/>). The scheme invokes a forward selection procedure that starts with the MLR^{nom} regression equation as the base model, then extends to the MLR^{ext} model by adding the contingent optimal tag parameters one at a time (producing an OT_k sequence), until no further parameter addition significantly improves the BIC_{LL} fit, with the relative significance evaluations of the OT_k sequence assessed by the values of the associated $\% \varepsilon_k^2$ sequence.

After MLR^{ext} regression coefficients are evaluated for the RV_n of the over-water and over-land database partitions using training datasets, the results are prepared for a final experiment. The hypothesis of this experiment is that at least a few optimal tags contained within the MLR^{ext} models will improve upon (increase) the $\% \varepsilon_0^2$ s significantly when the regression model coefficients are used in conjunction with test datasets (mutually exclusive from the training datasets). This hypothesis must be confirmed to demonstrate that the set of contingent optimal tags can be credibly proposed for use in the CDRD algorithm as effective constraint parameters. In fact, the results from the experiment confirm the hypothesis, actually somewhat better than expected. Three to five tags are consistently added to the MLR^{ext} models before overfitting begins, with the summations of the sets of $\% \varepsilon_k^2$ s around 25 % over and above the associated $\% \varepsilon_0^2$ s, regardless of the rainfall parameter. For example, the $\% \varepsilon_k^2$ sequence associated with RR_{surf} for the winter over-water experiment member is 16.54 %, 5.25 %, 3.61 %, 1.35 % and 0.40 %. The overall results of the experiment are impressive considering only one adjustment has been made for non-linearities in the original relationships between individual optimal tags and the rainfall variables represented on \log_{10} scales, as exemplified by the patterns seen in Fig. 3a (the exception is that ω_{700} can be introduced as is or as the square of its value).

5 Conclusions

In developing an extensive dual-model generated CDRD database, underpinned by the Tripoli Nonhydrostatic Modeling System and implemented according to explanations given in Sects. 3.1 and 3.4 of this paper, and the RTE Model System

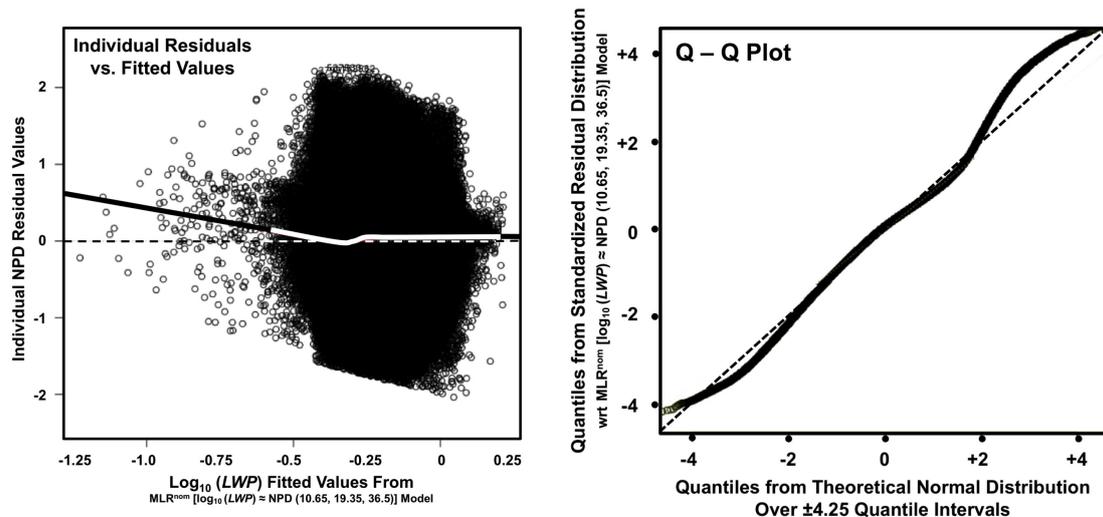


Fig. 4. Left-hand panel shows scatter of individual NPD residuals with respect to range of \log_{10} (LWP) from over-water MLR^{nom} [\log_{10} (LWP) \approx NPD (10.65, 19.35, 36.5)] model. Zero residual axis is seen as thin black-white dashed line while best fit to residual distribution is seen as thick black-white solid line. Right-hand panel shows associated Q-Q plot over ± 4.25 quantile interval of quantiles from same MLR^{nom} model residual distribution to equivalent quantiles from perfect normal distribution. Reference $x = y$ line is seen as dashed bisector.

(RMS) developed in the Part 2 study and implemented according to descriptions given in Sects. 3.2, 3.3 and 3.4 also of this paper, we have identified seven optimal meteorological-geophysical tags for purposes of algorithm constraint. A variety of these tags have already been used in the Part 1 and Part 2 studies of this multi-part investigation. Three statistical analysis procedures have been used to identify and confirm the reliability of the seven parameters for use in the CDRD algorithms. The 1st procedure is a parameter-by-parameter linear correlation analysis conducted seasonally on 25 candidate meteorological-geophysical parameters and the first three of the four principal associated rainfall variables, RR_{surf} , LWP, IWP and TWP. The 2nd procedure is a 2-dimensional cross-tabulated histogram analysis used to confirm that the distribution properties of the strongest correlating parameters in conjunction with their statistical relationships to the rainfall variables are well-behaved. The 3rd procedure is a multi-linear regression analysis used to determine the effectiveness of the optimal tags, in combination with TBs, in regressively determining estimates of the rainfall variables, and by so doing, confirming their potential effectiveness as physical constraints in CDRD type rainfall retrieval algorithms.

The parameter-by-parameter linear correlation analysis has identified six meteorological parameters and one geophysical parameter that would best exert constraint leverage on potential microphysical profile Bayesian solution subsets for a CDRD algorithm. Note that each of these optimal tags has or supports calculations of counterpart observational tags within the optimally assimilated initial meteorological datasets used for forecasting by the GFS and ECMWF op-

erational models. The seven optimal tags are (1 and 2) vertical velocities at 700 and 500 hPa (ω_{700} , ω_{500}), (3) equivalent potential temperature at surface ($\theta_{e_{\text{surf}}}$), (4) convective available potential energy (CAPE), (5) moisture flux 50 hPa above surface ($\Phi_{q_{\Delta 50}}$), (6) freezing level height (H_{FL}) and (7) surface height (H_{surf}). Note that ω_{700} and ω_{500} should be considered dynamical tags, $\theta_{e_{\text{surf}}}$ and CAPE thermodynamical tags, $\Phi_{q_{\Delta 50}}$ and H_{FL} hydrological tags, and H_{surf} a geophysical tag.

It is further shown that three additional non-observable parameters might be effective as optimal tags, if global operational forecast model initial data analyses could provide or could support the calculations of such parameters. As noted in Sect. 4, these three parameters consist of the Brunt-Väisälä Frequency Squared (N^2), the Richardson Number (Ri) and the Planetary Boundary Layer Height (H_{PBL}). All three of these quantities exhibit strong correlative relationships with the rainfall variables, with the caveat that the distribution of the Ri realizations with respect to TWP entails complexities that suggest it may be difficult to use as a constraint tag.

The 2-dimensional histogram analysis confirms well-behaved uniform and continuous relationships between the optimal tags and the rainfall variables. Some of the relationships are relatively linear and sloped, some more non-linear, some tend to lie parallel along either the abscissa or the ordinate for part of or all of their distribution domain and all of them indicate scatter over their functional structure. We stress that none of the underlying relationships inhibit their effectiveness as constraint parameters with regards to the CDRD algorithm methodology.

The multi-linear regression analysis corroborates the claims made in the Parts 1 and 2 papers of Sanò et al. (2013) and Casella et al. (2013) that the optimal tags applied as Bayesian solution constraints in a CDRD framework serve to improve satellite PMW estimates of rainfall. This is shown by a detailed study of the effectiveness of the optimal tags in increasing percentage explained variances (consistently around 25 %) within multi-linear regression models that relate from three to five optimal tag parameters, in combination with normalized polarization difference (NPD) parameters, to logarithmic transformations of the rainfall variable quantities. These results are impressive because they attempt to account only weakly for standing non-linearities in the tag-rainfall relationships, while using as base models well-linearized regressions between just the NPD parameters and the \log_{10} transformations of the rainfall variable. It is emphasized that the multi-linear regression results do not provide a quantitative measure of the expected performance of the optimal tags as applied in the CDRD algorithm, because there they are used in an entirely different discriminant-type fashion within a probabilistic framework. These results simply buttress our argument why algorithm performance would improve.

Finally, it is stressed that in any implementation of a CDRD-type algorithm, the retrieval outcomes strongly depend on the quality of the observational tags and the physical compatibilities between the simulation tags and their observational counterparts. This point is important because the global models are hydrostatic in nature and the NMS model is nonhydrostatic in nature. Thus, there always exists the possibility of counterpart parameter incompatibility when experimenting with different sets of optimal tags. This, of course, does not diminish any such parameter as a future possible optimal tag at such time that nonhydrostatic models at CRM resolutions emerge – models that would provide meteorological observables on a global basis through assimilation procedures – although, as we have noted, such a state of affairs may be a decade or more away.

Appendix A

25 Selected meteorological and geophysical parameters

In the following, the formulations of the 25 NMS variables selected for testing as possible correlator parameters to the three rainfall variables (RR_{surf} , LWP, IWP) are described in the manner they are calculated in the NMS code. Note that the first four of these parameters (N^2 , Ri , Fr , H_{PBL}) are calculated differently than what would be found in meteorological textbooks, but, in fact, are simply more advanced formulations perfectly suited to high resolution models.

A1 Brunt–Väisälä frequency squared (N^2)

Brunt–Väisälä Frequency Squared (N^2) is a measure of buoyant stability, given in units of s^{-2} . Whereas N^2 can be calculated at any level of the atmosphere, for the purpose of this study it is calculated at 50 hPa (~ 500 m) above ground level (AGL). Its formulation requires different equations for three cases: (Case 1a) a dry parcel, (Case 1b) a saturated parcel with respect to liquid and/or ice water, and (Case 1c) an unsaturated parcel for which liquid and/or frozen precipitation is present in the parcel.

Case 1a. N^2 for a dry parcel:

$$N^2 = \left(\frac{g}{\theta_0} \right) \frac{\partial \theta_v}{\partial z} \quad (\text{A1})$$

For this case, θ_0 is the dry, hydrostatic reference value of potential temperature (θ) given by $\theta = (c_p/\pi) T$, where c_p is specific heat capacity of dry air at constant pressure, g is acceleration due to gravity, and $\theta_v = \theta (1 + 0.61 q_v)$ is the virtual potential temperature of an air parcel in which $\partial \theta_v/\partial z$ is its vertical derivative and q_v is the specific humidity of water vapor.

Case 1b. N^2 for a saturated parcel:

$$N^2 = \left(\frac{g}{\theta_0} \right) \frac{\partial \theta_{\text{ev}}}{\partial z} \quad (\text{A2})$$

For this case, θ_v is replaced by the virtual equivalent potential temperature (θ_{ev}) given by $\theta_{\text{ev}} = \theta_v [1 + (l_p^w q_s^w)/(c_p T)]$, where l_p^w is a weighted latent heat of phase change given by $l_p^w = \alpha \cdot l_{\text{vc}} + (1 - \alpha) \cdot l_{\text{sd}}$ in which $\alpha = \max[0, \min(1, (T - 253.16)/20)]$ is the weighting factor and l_p^w is defined using the latent heat of vaporization-condensation (l_{vc}) for $T \geq 273.16$ and the latent heat of sublimation-deposition (l_{sd}) for $T \leq 253.16$ as limiting values, with the weighted value linearly changing between the two limits with respect to temperature change over the intervening 20 °C interval 273.16 K to 253.16 K, and where q_s^w is a weighted saturation specific humidity using the same weighting scheme $q_s^w = \alpha \cdot q_{\text{vs}} + (1 - \alpha) \cdot q_{\text{is}}$, defined using the saturation specific humidities with respect to liquid and ice surfaces (q_{vs} , q_{is}), respectively.

Case 1c. N^2 for an unsaturated parcel:

$$N^2 = \left(\frac{g}{\theta_0} \right) \frac{\partial \theta_v}{\partial z} - P_1 \frac{\partial q'}{\partial z} \left(\frac{l_p^w g}{c_p T} \right) \quad (\text{A3})$$

For this case, evaporation cooling and melting are taken into effect, where P_1 is a precipitation evaporation/melting factor pertaining to the size of the precipitating hydrometeors (set to 0.5 because the larger hydrometeors fail to maintain saturation) and $\partial q'/\partial z$ is replaced by a delta expression in which $\Delta q'$ approximates the total evaporating and/or sublimating

condensate:

$$\Delta q' \cong P_1 \min \left[\frac{gq_v}{RT} \Delta z \frac{\frac{\varepsilon}{c_p} \frac{I_p^w}{T} - 1}{\frac{\varepsilon}{c_p R} \frac{I_p^w}{T^2} q_v + 1} + q_{vs} - q_v, q_{cl} + q_{ci} \right] \quad (\text{A4})$$

where R is the ideal gas constant ($8.314 \text{ J K}^{-1} \text{ mol}^{-1}$), $\varepsilon = 0.622$ (ratio of molecular weight of water to that of air) and q_{cl} , q_{ci} are the specific humidities of cloud liquid condensate and cloud ice condensate, respectively.

Typically at 50 hPa AGL, a parcel is dry or unsaturated, although if the latter it may contain liquid or frozen precipitation. Note that N^2 is positive for a statically stable environment. Thus, the greater a positive value of N^2 , the more stratified the atmosphere becomes and the greater the atmospheric stability. For a positive N^2 , its square root N is defined as the Brunt–Väisälä frequency (after David Brunt and Vilho Väisälä), i.e., the frequency at which the parcel oscillates when vertically displaced in the statically stable environment. On the other hand, N^2 is negative for a statically unstable environment, which means the rhs of the relevant N^2 expression (i.e., either Eqs. A1, A2 or A3) is negative (e.g., for Case 1a, $\partial\theta_v/\partial z$ is negative). In this instance, N must be considered complex, but in reality is not defined because there is no oscillation frequency – instead, simply run away acceleration, convection and overturning. Thus, as N^2 becomes more negative, the greater the probability there is for precipitation. It is important to note that N^2 is useful in determining gravity wave behavior, such as how likely gravity waves will grow or dampen. Significantly, gravity waves can alter the vertical moisture gradient and trigger convective instability. Note also that latent heating can possibly change the sign of $\partial\theta_v/\partial z$, $\partial\theta_{ev}/\partial z$ or even $\partial q'/\partial z$, and thus the sign of N^2 .

A2 Richardson number (Ri)

Richardson number (Ri) is the ratio of the buoyant production of turbulence to the shear production of turbulence, and is thus a dimensionless quantity. Whereas Ri can be calculated at any level of the atmosphere, for the purpose of this study it is calculated at 50 hPa ($\sim 500 \text{ m}$) a.g.l. Its formulation requires the calculation of N^2 and is meaningful for either signed quantity.

$$Ri = \frac{N^2}{D_{ij}^2} \quad (\text{A5})$$

where D_{ij} is the 9-component deformation tensor, expressed by

$$D_{ij} = \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \quad (\text{A6})$$

where $i = 1, 3$ and $j = 1, 3$ represent indices for the zonal, meridional and vertical velocities (u , v , w) expressed by u_i

or u_j , and the associated coordinate axes (x , y , z) expressed by x_i or x_j . Ri is used to evaluate the dynamic stability of a specified region of the atmosphere. The critical value of Ri (Ri^C) is found at 0.25, the value below which flow is unstable and turbulent, signifying that wind shear is strong enough to overpower static stability. Thus, values of Ri below Ri^C can be indicative of the presence of convection and precipitation.

A3 Froude number (Fr)

Froude number (Fr) is the ratio of the inertial resistance to lifting a parcel flowing at horizontal velocity (V) at a specified height above the surface (h) to the resistance against lifting due to static stability in the PBL, and is thus a dimensionless quantity. Whereas Fr can be calculated at any level of the atmosphere, for the purpose of this study it is calculated at 50 hPa ($\sim 500 \text{ m}$) a.g.l. Its formulation requires the calculation of N and is only meaningful for static stability.

$$Fr = V/Nh \quad (\text{A7})$$

where h marks the top of the statically stable PBL and N is the average Brunt–Väisälä frequency over depth h (since Fr only applies to the statically stable PBL, N^2 is guaranteed positive and thus N is real). Values of $Fr < 1$ are called subcritical (flow velocity $<$ wave velocity), while values of $Fr > 1$ are called supercritical (flow velocity $>$ wave velocity). Flows that are forced to go around an obstacle have smaller values of Fr (i.e., subcritical), while flows that tend to rise over the top of an obstacle have larger values of Fr (i.e., supercritical). In actual atmospheric applications, generally small values of Fr would be associated with resistance to orographic or frontal lifting and thus indicative of conditions unfavorable for convection and precipitation, while large values of Fr would be associated with preference for orographic or frontal lifting and thus indicative of conditions favorable for forced convection and precipitation.

A4 Planetary boundary layer height (H_{PBL})

Planetary boundary layer height (H_{PBL}) is defined as the top of the planetary boundary layer (PBL), and is generally considered the lowest level of the free atmosphere not in frictional contact with the surface. The H_{PBL} generally varies during the day and can be found as the level over which the PBL can be mixed until its bulk Ri becomes the Ri^C , noting the bulk Ri incorporates a value of N^2 defined between the surface and the mixing level; see Troen and Mahrt (1986) for a description of this methodology for defining H_{PBL} . This means that for a very deep statically unstable atmosphere (such as over an intensely heated desert during daytime), the associated H_{PBL} itself can become very deep (theoretically up to troposphere depth). After sunrise, assuming negligible horizontal advection and undisturbed conditions, turbulent eddies and rising thermals in the mixed layer deepen the PBL depth by the process of entrainment from above. After

sunset, turbulence dissipates with the mixed layer, transforming to a residual layer. Afterwards, during nighttime, the surface cools by outgoing longwave radiation, promoting the formation of a stable PBL. For moist boundary layers, under greater PBL depths, moisture from the surface is mixed to higher levels, increasing the likelihood of condensation and precipitation.

A5 Convective available potential energy (CAPE)

Convective available potential energy (CAPE) is the potential energy a parcel of air near the surface would have, in excess of its environment, if lifted along a specified vertical path, given in units of J kg^{-1} . Its formulation requires an integral along the path for which the parcel is buoyant.

$$\text{CAPE} = \int_{z_f}^{z_e} g \left(\frac{T_{v\text{par}} - T_{v\text{env}}}{T_{v\text{env}}} \right) dz \quad (\text{A8})$$

where z_f and z_e are the heights of the level of free convection (LFC) and the equilibrium level (EL) (also referred to as the level of neutral buoyancy), respectively, $T_{v\text{par}}$ is the virtual temperature (T_v) of the specified parcel and $T_{v\text{env}}$ is the T_v of the environment. It is the positive buoyancy of the air parcel as an indicator of atmospheric instability, which makes it valuable in predicting severe weather. In general, CAPE should be considered as a form of fluid instability found in thermally stratified atmospheres, in which a colder fluid overlies a warmer one. When an air parcel is statically unstable, it is displaced upwards by its buoyancy and then accelerated by the pressure differential between the displaced air and the ambient air at the higher altitude to which it is displaced. CAPE is also realized under the condition that when a moist parcel is forced to rise to its LFC, excess heating will be released through condensation and/or freezing and thus will additionally warm the parcel to temperatures in excess of its environment. This can eventually lead to thunderstorms and precipitation. Notably, CAPE can be created by various processes, particularly those which can cause cooling above and moistening and warming below. Even if the air is cooler on the surface, there may still be warmer air at mid-levels, which would rise to upper levels. However, if there is insufficient water vapor, there can be no significant condensation, meaning no clouds or precipitation.

A6 Convective inhibition (CIN)

Convective inhibition (CIN) is the amount of positive energy needed to lift an air parcel vertically from the surface to its LFC, given in units of J kg^{-1} . Its formulation resembles that of CAPE.

$$\text{CIN} = - \int_{z_s}^{z_f} g \left(\frac{T_{v\text{par}} - T_{v\text{env}}}{T_{v\text{env}}} \right) dz \quad (\text{A9})$$

where z_s and z_f are the heights of the surface and the LFC, respectively, and the T_v parameters are equivalent to those used for the definition of CAPE. The negative sign makes the definition a positive quantity, because the integral term, by itself, represents negative energy. Typically, for conditions in which CIN becomes a relevant parameter, the larger its value, the stronger the capping inversion, which temporarily suppresses the development of convection and thunderstorms. The capping inversion is an important element of severe weather because it is the layer which separates warm, moist air below from cool, dry air above. Thus, for a strong capping inversion (i.e., large CIN), surface convective instability will continue to build up under ongoing heating and moistening of the near surface air, until convective elements break through the cap later in the day, enabling the development of severe weather and precipitation. On the other hand, if CIN becomes too large, the inversion is too strong and prevents convection altogether. Conversely, if CIN is too small, it may indicate that there is insufficient CAPE for convection to develop, with or without a capping inversion. Thus, there tends to be an intermediate large value of CIN (i.e., a *Goldilocks* value), that may depend on other parameters such as wind shear, for which severe weather and precipitation are enabled.

A7 Surface equivalent potential temperature ($\theta_{e\text{surf}}$)

Surface equivalent potential temperature ($\theta_{e\text{surf}}$) is the θ of an air parcel raised from the surface to a level that all latent heat content has been released and q_v removed, then returned adiabatically to the surface reference level. Therefore, it is a thermodynamic measure of the combined θ and q_v of a parcel. $\theta_{e\text{surf}}$ is conserved as a parcel rises and can be compared to the values above to determine the heights to which a parcel is capable of rising during moist convection. Large values of $\theta_{e\text{surf}}$ are helpful in distinguishing warm-moist tropical air mass sources from cool-dry temperate sources. Increases in surface temperature (T_{surf}) and surface dew point temperature ($T_{d\text{surf}}$) will result in larger values of $\theta_{e\text{surf}}$, and thus will increase the potential height to which an air parcel can rise. Also, $\theta_{e\text{surf}}$ is strongly influenced by surface pressure and elevation, as it depicts the effect of elevation or lower surface pressure on increasing the potential for air to rise to greater atmospheric heights via convective processes. Thus, larger values of $\theta_{e\text{surf}}$ raise the probabilities for clouds and precipitation.

A8 Surface skin temperature (T_{skin})

Surface skin temperature (T_{skin}) is defined as the T at the top of the Earth's surface (whether it be land or sea), assuming radiative equilibrium of an idealized thin-membrane molecular boundary. Notably, T_{skin} responds differently for different types of surfaces. For example, it can be used to distinguish vegetated from soil surfaces, salinated sea water from fresh

sea water surfaces, or even tropical latitude from temperate latitude surfaces (under certain assumptions). Moreover, it conditionally can be used to determine the degree of surface heating and thus be an indicator of convection and possibly precipitation.

A9 Lifted index (LI)

Lifted index (LI) is the difference between the T of a parcel that has been adiabatically lifted to the 500 hPa level from the surface and the environmental T at 500 hPa. It measures the stability of the troposphere with respect to convection that has originated from the surface. If LI is negative, it indicates instability and the likelihood for convection. Thus it can be used as a severe weather parameter and a potential indicator of precipitation.

A10 Lapse rate from 500 to 850 hPa ($\Gamma_{500-850}$)

Lapse rate from 500 to 850 hPa ($\Gamma_{500-850}$) is the atmospheric temperature gradient with height at mid levels. Since $\Gamma_{500-850}$ can be altered by horizontally differential temperature advection and vertically differential diabatic heating, any change in this parameter has a direct effect on CAPE and CIN, and thus can be used to assess atmospheric stability. When $\Gamma_{500-850}$ is less than $\sim 6^\circ\text{C km}^{-1}$, conditions are generally stable with little to no chance for the formation of clouds and precipitation. Alternatively, when $\Gamma_{500-850}$ approaches $\sim 10^\circ\text{C km}^{-1}$, conditions are typically absolutely unstable, indicative of the formation of convection and precipitation. Between these two lapse rate values, conditions are considered to be conditionally unstable, and thus possibly indicative of cloud and precipitation formation.

A11 Latent heating of column (LH)

Latent heating of column (LH) is the heat released or taken up by the phase change(s) of water. This is an important parameter in determining the growth and development of both synoptic and mesoscale circulations. The actual mechanisms producing LH release or uptake and thus the magnitude and sign of LH, are the explicit counterpoised microphysical processes of evaporation–condensation ($e-c$), melting–freezing ($m-f$) and sublimation–deposition ($s-d$). Accordingly, any net production of ($e + m + s$) would lead to LH uptake (diabatic cooling) with a reduction in the possibility of precipitation, while any net production of ($c + f + d$) would lead to LH release (diabatic heating) with an increase in the possibility of precipitation.

A12 Lifting condensation level (LCL)

Lifting condensation level (LCL) is the height at which an air parcel reaches its saturation level by lifting. It is used to estimate cloud base heights in which smaller LCLs would generally indicate larger probabilities for clouds and precipitation.

A13 Freezing level height (H_{FL})

Freezing level height (H_{FL}) is the height at which the value of T along its vertical axis reaches a value of 0°C . In the presence of condensates, ice formation will generally occur above the H_{FL} . It can be used to estimate the amount of ice and water in a column. If the H_{FL} is low, the cloud column generally contains more ice and less water; alternatively, if H_{FL} is high, the reverse occurs. In this sense, H_{FL} represents an estimator of surface precipitation. Since H_{FL} depends on the T profile of the atmosphere, it changes due to temperature advection, convection and evaporation–cooling from precipitation itself. It is often used to differentiate tropical from higher latitude environments because the H_{FL} generally decreases as latitude extends poleward.

A14 Sensible heat flux from surface (Φh_{surf})

Sensible heat flux from surface (Φh_{surf}) is defined as the vertical transport of heat by turbulent eddies with respect to the surface, using the sign convention that conducting heat away from the surface into the atmosphere is positive. Thus, larger values of Φh_{surf} can be indicative of greater probabilities for convective destabilization, storms, clouds and precipitation.

A15 Moisture flux 50 hPa AGL ($\Phi q_{\Delta 50}$)

Moisture flux 50 hPa AGL ($\Phi q_{\Delta 50}$) is defined as the vertical transport of water vapor at a level of 50 hPa above the surface, using the sign convention that upward directed flux is positive. $\Phi q_{\Delta 50}$ actually combines vertical motion with the vertical flux of moisture content near the surface into a single parameter. It effectively integrates the effect of moisture convergence over the friction layer, which can be the source of moisture lifted to form clouds when lifting is forced from below such as with Ekman pumping or orographic forcing. Thus, larger values of $\Phi q_{\Delta 50}$ are indicative of greater probabilities for clouds and precipitation.

A16 Positive vorticity advection at 500 hPa (ξ_{500})

Positive vorticity advection at 500 hPa (ξ_{500}) is defined as $\xi = -V \nabla (\zeta + f)$ at the 500 hPa level, where $\zeta + f$ is the absolute vorticity, given by the sum of relative and planetary vorticity, respectively [$\zeta = \partial v / \partial x - \partial u / \partial y$, i.e., curl of V ; $f = 2\Omega \sin \varphi$], given in units of s^{-2} , and noting that Ω is the rotation rate of the Earth ($7.2921 \times 10^{-5} \text{ rad s}^{-1}$) and φ is the specified Earth latitude. For quasi-geostrophic (QG) motions in middle latitudes (requiring geostrophic and hydrostatic balance), positive vorticity advection (PVA) is produced when parcels of air move from higher to lower values of vorticity. For such QG flow with a unimodal vertical profile of PVA, the difference between PVA at an upper level and at the surface signifies the forcing of vertical motion to maintain geostrophic balance. Because the PVA at the surface tends to be much smaller than that at upper levels,

the surface value is typically neglected such that a value of PVA at upper levels, by itself, represents the integrated effect of QG lifting from below. It is a convention to assess PVA at 500 hPa because this is approximately the level of non-divergence for deep tropospheric divergence. QG theory also shows that PVA-induced rising motion is usually enhanced if there is coincident warm air advection at lower levels. Therefore, elevated values of ξ_{500} in middle latitudes are indicative of inertially forced vertical motion in the lower troposphere and possibly the development of clouds and precipitation.

A17 Positive vorticity advection at 700 hPa (ξ_{700})

Positive vorticity advection at 700 hPa (ξ_{700}) is equivalent to the definition for ξ_{500} , except that it is calculated in the middle of the lower troposphere at 700 hPa. For shallow weather systems or for the lower portions of deep weather systems, ξ_{700} assesses the potential of PVA to induce lifting motions in the lower troposphere. This gives a more focused measure of QG lifting in the lowest portions of the atmosphere where moisture is most abundant. Therefore, in comparison to large values of ξ_{500} , large values of ξ_{700} indicate lifting focused in the moist layer and thus may better isolate the potential for vertical motion that would lead to clouds and precipitation.

A18 Vertical velocity at 700 hPa (ω_{700})

Vertical velocity at 700 hPa (ω_{700}) is a measure of the vertical air motion in p-coordinates in the middle of the lower half of the troposphere. Because pressure decreases with height, negative values mean rising motion while positive values mean descending motion. Notably, 700 hPa is at a level where precipitation production is typically at its maximum, and thus large negative ω_{700} values can represent the amount of precipitation being actively formed. Generally, such lifting is maximized in deep convective systems, frontal regions and deep orographic clouds. Unlike PVA, ω_{700} is a direct measure of vertical motion rather than a QG assessment of the potential for vertical motion. Hence, it represents not only QG forcing, but unbalanced motion (in NMS, this includes resolved convective motions and gravity wave motions). Therefore, large negative ω_{700} values are indicative of greater probabilities for clouds and precipitation.

A19 Vertical velocity at 500 hPa (ω_{500})

Vertical velocity at 500 hPa (ω_{500}) is a measure of vertical air motion in p-coordinates in the middle of the troposphere. Although the 500 hPa level is above the height where precipitation is typically produced, upward motion at this level signifies very deep tropospheric overturning, which in turn supports deep convection. Therefore, large negative values of ω_{500} are indicative of greater probabilities for clouds and precipitation.

A20 Divergence at surface (DIV_{surf})

Divergence at surface (DIV_{surf}) is negative for air mass convergence at the surface and thus, for this situation, represents a mechanical lifting mechanism rooted at the surface for creating rising air. Unlike lower tropospheric divergence above the surface, surface divergence is strongly affected by local boundaries such as density currents left by convective systems, low level deformations produced by cold fronts, sea breeze fronts, or thermal contrasts created by, e.g., the separation of river and field areas, desert and grassland landscapes, or snow-cover and snow-free regions. This type of convergence can be important in triggering conditionally unstable air to rise to its LFC, thus producing convection and possibly precipitation. On the other hand, it is less important for long-term lifting of stable air to form stratiform clouds, except possibly over sloped terrain. In the case of lifting stable air over sloped terrain, divergence should be diagnosed at deeper upper levels, for example, at 700 mb. This is because deep convergence is more important for lifting a deep layer, rather than for triggering shallow vertical motion.

A21 Divergence at 700 hPa (DIV_{700})

Divergence at 700 hPa (DIV_{700}) is a measure of how much mass is being deposited in the lower troposphere and thus indicates the integrated effect of downward (divergent) or upward (convergent) motion above that level. The effect at 700 hPa is most pronounced in the vicinity of fronts, where isentropic frontal lifting is manifested as the 700 hPa convergence of flow. In addition, along major mountain barriers such as, e.g., the Rockies, Andes, Alps or Himalayas, 700 mb convergence (divergence) results from mid-level flow normal to the mountain ridge, resulting in upslope (downslope) flow on the windward side. Thus, negative DIV_{700} (convergence) is indicative of frontal or orographic lifting and the concomitant formation of clouds and possibly precipitation, while positive DIV_{700} is indicative of frontal or orographic subsidence and the concomitant suppression of any cloudiness.

A22 Thickness from 500–1000 hPa ($\Delta Z_{500-1000}$)

Thickness from 500–1000 hPa ($\Delta Z_{500-1000}$) is the separation length between the 500 and 1000 hPa pressure levels. It is proportional to the mean T of this layer, and typically matches the 700 hPa thermal pattern. It can be used to signify a cold versus warm lower troposphere. It can also signify the possibility of precipitation reaching the surface as frozen snow versus conditions of melting and thus precipitation reaching the surface as liquid rain. Whereas it is difficult to relate $\Delta Z_{500-1000}$ to precipitation because such a relationship is dependent on the geography and ambient climatology of the region; in general, the greater its value, the greater the possibility for snow.

A23 Thickness from 700–1000 hPa ($\Delta P_{700-1000}$)

Thickness from 700–1000 hPa ($\Delta Z_{700-1000}$) is the separation length between the 700 and 1000 hPa pressure levels. As with $\Delta Z_{500-1000}$, it is proportional to the mean T of this layer and can be used to identify cold or warm air masses centered in the low troposphere. In addition, as with $\Delta Z_{500-1000}$, it is difficult to relate $\Delta Z_{700-1000}$ to precipitation because such a relationship is dependent on the geography and ambient climatology of the region; in general, the greater its value, the greater the possibility for snow.

A24 Vertical wind shear in lower troposphere ($\Delta_z V_{LT}$)

Vertical wind shear in lower troposphere ($\Delta_z V_{LT}$), is defined as the vertical velocity gradient in the lower troposphere between the surface and the 6 km level. It is a measure of how vertically-layered dynamic shear stresses can affect the organization, type, longevity and severity of storms. At mid-latitudes, in considering thunderstorms, moderate levels of $\Delta_z V_{LT}$ can help tilt the vertical storm structures such that precipitation is able fall away from the updraft regions, thus preventing updraft suppression. Moreover, $\Delta_z V_{LT}$ can produce rotating convective storms such as supercells, which are long-lived, severe and containing intense precipitation. However, if $\Delta_z V_{LT}$ becomes too large, it can destroy weak updrafts and prevent storm development. By the same token, at lower latitudes in tropical cyclone environments, any meaningful $\Delta_z V_{LT}$ is generally considered a menace to the efficient organization of the cyclone outflow layers, serving to tear apart disturbances before they can reach or maintain cyclone strength. Therefore, regardless of the fact that $\Delta_z V_{LT}$ is related to how precipitating storms develop, it is difficult to set meaningful thresholds for this parameter insofar as defining practical guides for the probability of precipitation.

A25 Surface height (H_{surf})

Terrain elevation, referred to as surface height (H_{surf}), is an effective parameter for defining the potential for mechanical lifting of horizontal airflow, which if sufficiently strong, can produce condensation, cloud formation and possibly precipitation. It also defines the potential for producing enhanced equivalent potential temperature at the surface ($\theta_{e,surf}$) (resulting from lowered surface pressure due to the terrain-induced lifting), a parameter also noted for its relationship to precipitation.

Acknowledgements. Smith and Tripoli, Mehta and Ms. Leung have been supported through NASA's TRMM-GPM science program. Elsner has been supported by Florida State University. Casella, Dietrich, Mugnai, Panegrossi and Sanò have been supported by a group of six European research programmes: (1) EUMETSAT through the project "Satellite Application Facility on Support to Operational Hydrology and Water Management" (H-SAF); (2) by the European Commission Sixth Framework Programme through

the project "Observation, Analysis and Modeling of Lightning Activity in Thunderstorms, for use in Short Term Forecasting of Flash Floods" (FLASH); (3) by the European Community Initiative INTERREG III-B ARCHIMED through the project "Weather Risk Reduction in the Central and Eastern Mediterranean" (RISKMED); (4) by the Italian Civil Protection Department (DPC) through the Operative Agreement DPC-ISAC; (5) by the Italian FISR-MIUR Programme "Sustainable Development and Climate Changes" through the project "Aerosol Effects on Clouds and Climate" (AEROCLOUDS); and (6) by the Italian Space Agency (ASI) through the pilot project "Civil Protection from Floods: The Nowcasting".

Edited by: G. Boni

Reviewed by: two anonymous referees

References

- Ackerman, S. A. and Stephens, G. L.: The absorption of solar radiation by cloud droplets: An application of anomalous diffraction theory, *J. Atmos. Sci.*, 44, 1574–1588, 1987.
- Battaglia, A., Kummerow, C., Shin, D.-B., and Williams, C.: Constraining microwave brightness temperatures by radar brightband observations, *J. Atmos. Ocean. Technol.*, 20, 856–871, 2003.
- Bauer, P., Amayenc, P., Kummerow, C. D., and Smith, E. A.: Over-ocean rainfall retrieval from multisensor data of the Tropical Rainfall Measuring Mission. Part II: Algorithm implementation, *J. Atmos. Ocean. Technol.*, 18, 1838–1855, 2001.
- Betts, A. K.: A new convective adjustment scheme. Part I: Observational and theoretical basis, *Quart. J. Roy. Meteor. Soc.*, 112, 677–692, 1986.
- Businger, J. A.: The fluxes of specific enthalpy, sensible heat and latent-heat near the earth's surface, *J. Atmos. Sci.*, 39, 1889–1892, 1982.
- Casella, D., Panegrossi, G., Sanò, P., Mugnai, A., Smith, E. A., Tripoli, G. J., Dietrich, S., Formenton, M., Di Paola, F., Leung, H. W.-Y., and Mehta, A. V.: Transitioning from CRD to CDRD in Bayesian retrieval of rainfall from satellite passive microwave measurements: Part 2. Overcoming database profile selection ambiguity by consideration of meteorological control on microphysics, *IEEE T. Geosci. Remote Sens.*, in press, 2013.
- Chou, M.-D. and Suarez, M. J.: An efficient thermal infrared radiation parameterization for use in general circulation models, NASA Tech. Memo. 104606, NASA/Goddard Space Flight Center, Greenbelt, MD, 85 pp., 1994.
- Cotton, W. R., Stephens, M. A., Neuhorn, T., and Tripoli, G. J.: The Colorado State University three-dimensional cloud/mesoscale model. Part II: An ice phase parameterization, *J. Res. Atmos.*, 16, 295–320, 1982.
- Emanuel, K. A.: A scheme for representing cumulus convection in large-scale models, *J. Atmos. Sci.*, 48, 2313–2329, 1991.
- English, S. J. and Hewison, T. J.: A fast generic millimetre-wave emissivity model, in: *Microwave Remote Sensing of the Atmosphere and Environment*, edited by: Hayasaka, T., Wu, D. L., Jin, Y. Q., and Jiang, J. S., 3503, Bellingham, SPIE International Society of Optical Engineering, 288–300, 1998.
- Evans, K. F., Turk, J., Wong, T., and Stephens, G. L.: A Bayesian approach to microwave precipitation profile retrieval, *J. Appl. Meteor.*, 34, 260–279, 1995.

- Farrar, M. R.: Combined Radar-Radiometer Rainfall Retrieval for TRMM Using Structure Function-based Optimization, Ph.D. Dissertation, (Smith, E. A., Major Prof.), Dept. of Meteorology, The Florida State University, Tallahassee, FL, 185 pp., 1997.
- Flatau, P., Tripoli, G. J., Berlinder, J., and Cotton, W. R.: The CSU-RAMS Cloud Microphysics Module: General Theory and Code Documentation, Atmospheric Sciences Report 451, Colorado State University, Dept. of Atmospheric Sciences, Fort Collins, CO 80523, 88 pp., 1989.
- Greco, M., Olson, W. S., and Anagnostou, E. N.: Retrieval of precipitation profiles from multiresolution, multifrequency active and passive microwave observations, *J. Appl. Meteor.*, 43, 562–575, 2004.
- Greco, M. and Olson, W. S.: Bayesian estimation of precipitation from satellite passive microwave observations using combined radar–radiometer retrievals, *J. Appl. Meteor. Climatol.*, 45, 416–433, 2006.
- Haddad, Z. S., Smith, E. A., Kummerow, C. D., Iguchi, T., Farrar, M. R., Durden, S. L., Alves, M., and Olson, W. S.: The TRMM “Day-1” radar/radiometer combined rain-profiling algorithm, *J. Meteor. Soc. Japan*, 75, 799–809, 1997.
- Hewison, T. J. and English, S. J.: Airborne retrievals of snow and ice surface emissivity at millimeter wavelengths, *IEEE Trans. Geosci. Remote Sens.*, 37, 1871–1879, 1999.
- Hewison, T. J. and English, S. J.: Fast models for land surface emissivity. Radiative Transfer Models for Microwave Radiometry, Directorate-General for Research, European Commission, Brussels, Belgium, 117–127, 2000.
- Hewison, T. J.: Airborne measurements of forest and agricultural land surface emissivity at millimeter wavelengths, *IEEE Trans. Geosci. Remote Sens.*, 39, 393–400, 2001.
- Kain, J. S. and Fritsch, J. M.: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. The Representation of Cumulus Convection in Numerical Models, *Am. Meteorol. Soc. Meteor. Monogr.*, 46, 165–170, 1993.
- Kummerow, C., Olson, W. S. and Giglio, L.: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors, *IEEE Trans. Geosci. Remote Sens.*, 34, 1213–1232, 1996.
- Kummerow, C., Hong, Y., Olson, W. S., Yang, S., Adler, R. F., McCollum, J., Ferraro, R., Petty, G., Shin, D.-B., and Wilheit, T. T.: The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors, *J. Appl. Meteor.*, 40, 1801–1820, 2001.
- Kummerow, C., Berg, W., Thomas-Stahle, J., and Masunaga, H.: Quantifying global uncertainties in a simple microwave rainfall algorithm, *J. Atmos. Ocean. Technol.*, 23, 23–37, 2006.
- Kummerow, C. D., Ringerud, S., Crook, J., Randel, D., and Berg, W.: An observationally generated a priori database for microwave rainfall retrievals, *J. Atmos. Ocean. Technol.*, 28, 113–130, 2011.
- Kuo, H. L.: Further studies of the parameterization of the influence of cumulus convection on large-scale flow, *J. Atmos. Sci.*, 31, 1232–1240, 1974.
- Liebe, H. J.: An updated model for millimeter wave propagation in moist air, *Radio Sci.*, 20, 1069–1089, 1985.
- Liebe, H. J.: A contribution to modeling atmospheric mm-wave properties, *Frequenz.*, 41, 31–36, 1987.
- Liebe, H. J.: MPM: An atmospheric mm-wave propagation model, *Int. J. Infrared Milli.*, 10, 631–650, 1989.
- Liou, K.-N., Fu, Q., and Ackerman, T. P.: A simple formulation of the delta-four-stream approximation for radiative transfer parameterizations, *J. Atmos. Sci.*, 45, 1940–1948, 1988.
- Louis, J. F.: A parametric model of vertical eddy fluxes in the atmosphere, *Bound. Layer Meteor.*, 17, 187–202, 1979.
- Macklin, W. C.: The density and structure of ice formed by accretion, *Quart. J. Roy. Meteor. Soc.*, 88, 30–50, 1962.
- Marzano, F. S., Mugnai, A., Panegrossi, G., Perdicca, N., Smith, E. A., and Turk, J.: Bayesian estimation of precipitating cloud parameters from combined measurements of space-borne microwave radiometer and radar, *IEEE Trans. Geosci. Remote Sens.*, 37, 593–613, 1999.
- Masunaga, H. and Kummerow, C. D.: Combined radar and radiometer analysis of precipitation profiles for a parametric retrieval algorithm, *J. Atmos. Ocean. Technol.*, 22, 909–929, 2005.
- Morcrette, J.-J.: Radiation and cloud radiative properties in the ECMWF operational weather forecast model, *J. Geophys. Res.*, 96D, 9121–9132, 1991.
- Mugnai, A., Di Michele, S., Marzano, F. S., and Tassa, A.: Cloud-model based Bayesian techniques for precipitation profile retrieval from TRMM microwave sensors, *Proc. ECMWF/EuroTRMM Workshop on Assimilation of Clouds and Precipitation*, ECMWF, Reading, UK, 323–345, 2001.
- Mugnai, A., Smith, E. A., Tripoli, G. J., Dietrich, S., Kotroni, V., Lagouvardos, K., and Medaglia, C. M.: Explaining discrepancies in passive microwave cloud-radiation databases in microphysical context from two different cloud-resolving models, *Meteor. Atmos. Phys.*, 101, 127–145, doi:10.1007/s00703-007-0281-4, 2008.
- Munchak, S. J. and Kummerow, C. D.: A modular optimal estimation method for combined radar-radiometer precipitation profiling, *J. Appl. Meteor. Climatol.*, 50, 433–448, 2011.
- Panegrossi, G., Dietrich, S., Marzano, F. S., Mugnai, A., Smith, E. A., Xiang, X. W., Tripoli, G. J., Wang, P. K., and Baptista, J.: Use of cloud model microphysics for passive microwave-based precipitation retrieval: Significance of consistency between model and measurement manifolds, *J. Atmos. Sci.*, 55, 1644–1673, 1998.
- Panegrossi, G.: Validation of Microphysics Parameterization in Cloud Resolving Models using Passive Microwave Measurements, Ph.D. Dissertation, Univ. of Wisconsin, Madison, WI, 186 pp., 2004.
- Petty, G. W.: Physical retrievals of over-ocean rain rate from multi-channel microwave imagers. Part I: Theoretical characteristics of normalized polarization and scattering indices, *Meteor. Atmos. Phys.*, 54, 79–99, 1994.
- Pierdicca, N., Marzano, F. S., d’Auria, G., Basili, P., Ciotti, P., and Mugnai, A.: Precipitation retrieval from spaceborne microwave radiometers based on maximum a posteriori probability estimation, *IEEE Trans. Geosci. Remote Sens.*, 34, 831–846, 1996.
- Prodi, F., Levi, L., Nasello, O. B., and Lubart, L.: Morphology and density of ice accreted on cylindrical collectors at low values of impaction parameter. II: Rotating deposits, *Quart. J. Roy. Meteor. Soc.*, 117, 783–801, 1991.
- Redelsperger, J. L. and Sommeria, G.: Methode de representation de la turbulence associe aux precipitations dans un modele tridimensionnel de convection nuageuses, *Bound.-Layer Meteor.*,

- 24, 231–252, 1982.
- Roberti, L., Haferman, J., and Kummerow, C.: Microwave radiative transfer through horizontally inhomogeneous precipitating clouds, *J. Geophys. Res.*, 99, 707–716, 1994.
- Sanò, P., Casella, D., Mugnai, A., Schiavon, G., Smith, E. A., and Tripoli, G. J.: Transitioning from CRD to CDRD in Bayesian retrieval of rainfall from satellite passive microwave measurements: Part 1. Algorithm description and testing, *IEEE Trans. Geosci. Remote Sens.*, 1–25, doi:10.1109/TGRS.2012.2227332, 2013.
- Schluessel, P. and Luthardt, H.: Surface wind speeds over the North-Sea from Special Sensor Microwave Imager observations, *J. Geophys. Res.-Ocean.*, 96, 4845–4853, 1991.
- Schwarz, G. E.: Estimating the dimension of a model, *Ann. Stat.*, 6, 461–464, 1978.
- Schwarzkopf, M. D. and Fels, S. B.: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes, *J. Geophys. Res.*, 96, 9075–9096, 1991.
- Smith, E. A., Crosson, W. L., Cooper, H. J., and Heng-Yi, W.: Estimation of surface heat and moisture fluxes over a prairie grassland. Part 3: Design of a hybrid physical/remote sensing biosphere model, *J. Geophys. Res.*, 98, 4951–4978, 1993.
- Smith, E. A., Kummerow, C., and Mugnai, A.: The emergence of inversion-type precipitation profile algorithms for estimation of precipitation from satellite microwave measurements, *Remote Sens. Rev.*, 11, 211–242, 1994a.
- Smith, E. A., Mugnai, A., and Tripoli, G.: Theoretical foundations and verification of a multispectral, inversion-type microwave precipitation profile retrieval algorithm, in: *Passive Microwave Remote Sensing of Land-Atmosphere Interactions (599-621)*, VSP Press, Utrecht – The Netherlands, 685 pp., 1994b.
- Smith, E. A., Xiang, X., Mugnai, A., and Tripoli, G. J.: Design of an inversion-based precipitation profile retrieval algorithm using an explicit cloud model for initial guess microphysics, *Meteorol. Atmos. Phys.*, 54, 53–78, 1994c.
- Smith, E. A., Turk, J., Farrar, M., Mugnai, A., and Xiang, X.: Estimating 13.8 GHz path integrated attenuation from 10.7 GHz brightness temperatures for TRMM combined PR-TMI precipitation algorithm, *J. Appl. Meteor.*, 36, 365–388, 1997.
- Smith, E. A., Asrar, G., Furuhashi, Y., Ginati, A., Mugnai, A., Nakamura, K., Adler, R. F., Chou, M.-D., Desbois, M., Durning, J. F., Einaudi, F., Entin, J. K., Ferraro, R. R., Guzzi, R., Houser, P. R., Hwang, P. H., Iguchi, T., Joe, P., Kakar, R., Kaye, J. A., Kojima, M., Kummerow, C. D., Kuo, K.-S., Lettenmaier, D. P., Levizzani, V., Lu, N., Mehta, A. V., Morales, C., Morel, P., Nakazawa, T., Neeck, S. P., Okamoto, K., Oki, R., Raju, G., Shepherd, J. M., Simpson, J., Sohn, B.-J., Stocker, E. F., Tao, W.-K., Testud, J., Tripoli, G. J., Wood, E. F., Yang, S., and Zhang, W.: International Global Precipitation Measurement (GPM) Program and Mission: An overview, *Measuring Precipitation from Space: EU-RAINSAT and the Future*, edited by: Levizzani, V., Bauer, P., and Turk, F. J., *Advances in Global Change Research (Vol. 28)*, Springer, Dordrecht, NETHERLANDS, 611–653, 2007.
- Spencer, R. W., Goodman, H. M., and Hood, R. E.: Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal, *J. Atmos. Ocean. Technol.*, 6, 254–273, 1989.
- Tripoli, G. J. and Cotton, W. R.: The use of ice-liquid water potential temperature as a thermodynamic variable in deep atmospheric models, *Mon. Weather Rev.*, 109, 1094–1102, 1981.
- Tripoli, G. J.: A nonhydrostatic model designed to simulate scale interaction, *Mon. Weather Rev.*, 120, 1342–1359, 1992a.
- Tripoli, G. J.: An explicit three-dimensional nonhydrostatic numerical simulation of a tropical cyclone, *Meteorol. Atmos. Phys.*, 49, 229–254, 1992b.
- Tripoli, G. J.: Numerical study of the 10 January 1998 lake-effect bands observed during Lake-ICE, *J. Atmos. Sci.*, 62, 3232–3249, 2005.
- Tripoli, G. J. and Smith, E. A.: Introducing variable-step topography (VST) coordinates within dynamically constrained Nonhydrostatic Modeling System (NMS), Part 1: VST formulation within NMS host model framework, *Dyn. Atmos. Oceans*, submitted, 2013a.
- Tripoli, G. J. and Smith, E. A.: Introducing variable-step topography (VST) coordinates within dynamically constrained Nonhydrostatic Modeling System (NMS). Part 2: VST performance on orthodox obstacle flows, *Dyn. Atmos. Oceans*, submitted, 2013b.
- Troen, I. and Mahrt, L.: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation, *Bound.-Layer Meteorol.*, 37, 129–148, 1986.
- Viltard, N., Burlaud, C., and Kummerow, C. D.: Rain retrieval from TMI brightness temperature measurements using a TRMM PR-based database, *J. Appl. Meteor. Climatol.*, 45, 455–466, 2006.