Assimilation of Satellite Microwave Observations over the Rainbands of Tropical Cyclones

ISAAC MORADI,[a,b] K. FRANKLIN EVANS,[c] WILL MCCARTY,[b] MARANGELLY CORDERO-FUENTES,[b]
RONALD GELARO,[b] AND ROBERT A. BLACK[d]

[a] *Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland*
[b] *NASA Global Modelling and Assimilation Office, Greenbelt, Maryland*
[c] *Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, Colorado*
[d] *NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*

ABSTRACT: A novel Bayesian Monte Carlo integration (BMCI) technique was developed to retrieve geophysical variables from satellite microwave radiometer data in the presence of tropical cyclones. The BMCI technique includes three steps: generating a stochastic database, simulating satellite brightness temperatures using a radiative transfer model, and retrieving geophysical variables such as profiles of temperature, relative humidity, and cloud liquid and ice water content from real observations. The technique also provides uncertainty estimates for each retrieval and can output the error covariance matrix of selected parameters. The measurements from the Advanced Technology Microwave Sounder (ATMS) on board *Suomi National Polar-Orbiting Partnership* (*Suomi NPP*) and the Global Precipitation Measurement (GPM) Microwave Imager (GMI) were used as input. A new technique was developed to correct the ATMS and GMI observations for the beam-filling effect, which is due to small-scale variability of precipitation and clouds when compared with the instrument footprint and also the nonlinear relation between the brightness temperature and precipitation. In addition, the assimilation of the BMCI retrievals into the NASA GEOS model is discussed for Hurricane Maria. The results show that assimilating the BMCI retrievals can influence the dynamical features of the cyclone, including a stronger warm core, a symmetric eye, and vertically aligned wind columns. Two possible factors that may limit the impact of the BMCI retrievals include 1) the resolution of the model (about 25 km), which was too coarse to show the potential of the BMCI data in improving the representation of tropical storms in the model forecast, and 2) the data assimilation system not being able to consider vertically correlated observation errors.

KEYWORDS: Tropical cyclones; Radiative transfer; Microwave observations; Bayesian methods; Numerical weather prediction/forecasting; Data assimilation

1. Introduction

Despite the importance of clouds and their influence on atmospheric water and energy balance, numerical weather prediction (NWP) centers to a large extent exclude cloud information from the assimilation process and only assimilate clear-sky radiances (Janisková et al. 2012). In the clear-sky data assimilation systems, in order to ensure that only clear-sky radiances are assimilated, strict cloud detection thresholds are applied before the radiances are considered in the data assimilation (DA) solution. This process not only excludes a large portion of satellite radiances, but causes loss of information in the regions that are of high interest to meteorologists and are most challenging for weather forecasts (Errico et al. 2007; Haddad et al. 2015). Although, in recent years there has been great advances in the operational weather forecasting, the prediction of tropical cyclones, especially the intensity of tropical cyclones, remains challenging. According to Aksoy et al. (2013), in addition to the model deficiencies, another important factor that contributes to this challenge includes lack of observations in the peripheral environment (rainbands) of

tropical cyclones mainly because of the selective assimilation of existing observations. Satellite observations provide more than 90% of the input data for the initialization of NWP models, but more than 75% of satellite observations are discarded due to cloud contamination as well as land, snow, and ice emissivity issues (Bauer et al. 2010). As expected, excluding cloud contaminated observations causes a significant lack of satellite data in the rainbands of tropical cyclones. Measurements from infrared instruments are restricted in the presence of convective clouds and thus do not provide much information on the state of the atmosphere. However, microwave measurements are less sensitive to clouds and are capable of providing information even in the presence of deep convective clouds such as in the case of tropical cyclones.

Two techniques that can be used to assimilate satellite cloudy radiances into NWP models are (i) direct assimilation using a radiative transfer (RT) model, also known as a forward model, and (ii) first retrieving atmospheric state variables such as temperature and humidity from satellite radiances and then assimilating retrieved products.

a. Direct assimilation of all-sky radiances

The main advantage of direct assimilation is that observations from all satellite instruments can be assimilated using a single fast radiative transfer model without intermediate retrieval. Since the early twenty first century, direct assimilation

---

of clear-sky satellite radiances has become a routine practice at NWP centers owing to progresses in fast radiative transfer models. However, direct assimilation of cloudy radiances is limited by several factors, including inaccuracy in the radiative transfer scattering calculations, lack of a close first guess in cloudy conditions, nonlinearity in microphysical and radiative parameterizations, and displacement of clouds in the first guess provided by NWP models. In particular, assimilating observations from high frequency microwave channels operating at 89 GHz and above is more challenging than low frequency channels operating around 50 GHz and below, because the errors for radiative transfer scattering calculations are much larger for ice particle scattering that dominates the high frequency channels. However, previous studies show that assimilating such observations can lead to improvements in the dynamical state of the atmosphere (Geer et al. 2017).

Geer and Bauer (2010) developed a piecewise method for calculating the observation error in order to assimilate microwave all-sky observations from channels operating below 90 GHz into the ECMWF model. The model starts with a constant small error for the clear-sky radiances and a much higher constant error for the radiances affected by deep-clouds and rain, then interpolates between the two for other conditions. The upper limit of the errors reported in Table 1 of Geer and Bauer (2010) are relevant to tropical cyclone cases with deep convective clouds. If we ignore the channels with a very large observation error, which was used to avoid assimilating such channels, the error for the channels that have actually been assimilated is mostly above 20 K. The same observation model was used by Geer and Baordo (2014) and Zhu et al. (2016) for the assimilation of the Advanced Microwave Sounding Unit-A (AMSU-A) channels (channels 1–4, and 15), and Migliorini and Candy (2019) for the assimilation of AMSU-A channels 4 and 5. Zhu et al. (2016) used an observation error of 20, 18, 12, 3, and 15 K for channels 1, 2, 3, 4, and 15, respectively, when cloud liquid water path exceed a threshold (optically thick clouds). To reduce the negative impact from the assimilation of all-sky observations due to error in the input provided by the NWP model, error in RT calculations, and also interpolation error, Zhu et al. (2016) further inflated the observation error by the absolute of observation minus forecast (omf) values. Overall, all-sky assimilation increased the usage of AMSU-A radiances by 10%–12%. The results showed neutral to slightly positive impact on global forecast skills. Migliorini and Candy (2019) used a similar observation error model to assimilate AMSU-A Channels 4 and 5 into a global model. Migliorini and Candy (2019) reported deviation from Gaussianity in the omf distributions that can be corrected by eliminating cases with low or high liquid water path values. Migliorini and Candy (2019) reported generally some positive improvements and attributed the improvement to the exclusion of radiances in the presence of high liquid water path values (i.e., convective clouds).

In addition to inflating the observation error (Geer and Bauer 2010; Geer et al. 2014; Zhu et al. 2016), the variational bias correction schemes used at the NWP centers, which depend on the difference between calculated and observed values, exclude the observations when the forward model is not able to simulate them with enough accuracy (Baordo and Geer 2015; Bauer et al. 2010). This in some cases limits the assimilation of cloud affected radiances to nonprecipitating shallow clouds to avoid the adverse effect on the global forecast skills, which eliminates the majority of the observations over the rainbands of tropical cyclones (Zhu et al. 2016; Migliorini and Candy 2019).

As stated in Bauer et al. (2006) and Bauer et al. (2010), the limitations for the direct assimilation of all-sky radiances over the rainbands of hurricanes are more significant for high frequency microwave channels (e.g., channels operating at 89 GHz and above) than for lower frequency channels, therefore direct assimilation of observations over the rainbands of hurricanes using variational techniques is even much more challenging for the high frequency channels. For the current status of the assimilation of all-sky satellite observations at the NWP centers, the readers are referred to Geer et al. (2018). However, it should be noted that Geer et al. (2018) discusses assimilation of all-sky observations from a global perspective, while we discuss the specific case of assimilation of all-sky observations over the rainbands of tropical cyclones.

b. Assimilation of retrieved profiles

The second technique for assimilation of cloudy radiances is first retrieving state variables such as temperature and humidity, and then assimilating retrieved profiles into the NWP models. The main issue with this technique is that it requires two separate steps and the DA results depend on the accuracy of the retrieval technique as well as quantification of error covariance matrices for the retrievals. In the past, the DA communities have normally used variational techniques (known as 1D-Var) to retrieve geophysical variables from satellite observations. However, variational techniques depend on the minimization of differences between simulated and observed values, and thus suffer from the same issues mentioned for direct assimilation of cloudy radiances. Weng et al. (2007) used a 1D + 3D-Var system to assimilate the AMSU-A observations into the NOAA Global Data Assimilation System. They first retrieved temperature profiles from AMSU-A observations, then assimilated the retrieved profiles. Janisková et al. (2012) employed a 1D-Var method to retrieve atmospheric temperature and humidity profiles from *CloudSat* measurements then assimilated the retrievals into the ECMWF model. Another technique that has been used to retrieve geophysical variables from satellite observations is the Bayesian Monte Carlo integration (BMCI) technique. This technique relies on developing a comprehensive retrieval database; the method works satisfactory once such a database is generated (Kummerow et al. 2001, 2015; Rydberg et al. 2009; Evans et al. 2012, 2002). Some advantages of the BMCI approach over variational techniques are: (i) the retrievals only depend on the observations and prior assumptions, but not the NWP model fields, and thus displacement of clouds or biases in the NWP fields do not affect the results; (ii) the BMCI retrievals do not require Jacobians of the radiative transfer model or the adjoint of the NWP model, and therefore nonlinearity in the RT model does not affect the results; and (iii) the uncertainty (error) covariance matrix of retrieved profiles

of humidity and temperature can be estimated and used in the assimilation process as pseudo-observation error. The BMCI algorithm has successfully been used to retrieve cloud and humidity profiles from spaceborne and airborne microwave measurements (e.g., Evans et al. 2002; Rydberg et al. 2009; Evans et al. 2012). Kidd and Huffman (2011), Elsaesser and Kummerow (2015), Kummerow et al. (2015), and Duncan et al. (2018) used similar techniques to derive precipitation from passive microwave observations. The algorithm developed by Kummerow et al. (2015) is known as Goddard profiling algorithm (GPROF) and is used operationally at NASA to retrieve precipitation from microwave instruments in the GPM constellation.

We discuss the assimilation of all-weather observations of Advanced Technology Microwave Sounder (ATMS) and the GPM Microwave Imager (GMI) using the BMCI technique. We first retrieve atmospheric state and cloud variables from the ATMS/GMI observations and then assimilate the retrievals of temperature, humidity, and sea surface temperature (SST) along with the estimated uncertainty into NASA's Global Earth Observing System (GEOS) forecast system. Observations from GMI and ATMS microwave instruments are used to examine the technique. These instruments measure radiances at the frequencies in the range of 10–190 GHz. Previous studies have proven that even clear-sky radiances from these instruments can substantially improve the forecast for the track and intensity of hurricanes. For instance, Zou et al. (2013) demonstrated the advantage of the assimilation of ATMS clear-sky radiances into the Hurricane Weather Research and Forecasting (HWRF) for improving the track and intensity of four Atlantic hurricanes that made landfall in 2012. Zou et al. (2013) conducted two sets of assimilation experiments with and without ATMS data and indicated that, even in the case of Hurricane Sandy with an unusual track, the assimilation of ATMS observations systematically improved the forecast, especially in correcting an eastward bias in the forecast for the Hurricane Sandy track. Zou et al. (2013) reported that the forecast intensity error is reduced by assimilating ATMS clear-sky radiances. They also indicated that assimilation of ATMS clear-sky radiances reduces the error in both maximum wind speed and minimum sea level pressure forecasts. Caumont et al. (2010) used a similar technique to retrieve humidity profiles from radar reflectivities and then assimilated the retrieved profiles using a 3D-Var model. However, instead of building a comprehensive a priori database as we did in the proposed method, they employed the first guess provided by the NWP model to generate the retrieval database. This technique was later used by Wattrelot et al. (2014) for the assimilation of radar reflectivities, and Duruisseau et al. (2019) to assimilate observations from a passive microwave humidity sounder. The disadvantage of using a model-provided first guess to generate the retrieval database is that the model does not provide all of the inputs required for the RT calculations, and therefore it encounters the same issues mentioned before for the direct assimilation of radiances. In addition, Caumont et al. (2010) assumed spherical particles for most hydrometer types and also constant errors for the retrieved profiles during the assimilation process.

We focus on the assimilation of retrieved profiles over the rainbands of Hurricane Maria. Since separating rainbands from the relatively calm regions surrounded by the rainbands is not possible, we use all of the observations over both the heavy precipitating and relatively calm regions in our assimilation process.

2. Satellite microwave instruments

The BMCI technique is applied to level-1b microwave radiances from GPM/GMI and *Suomi National Polar-Orbiting Partnership* (*Suomi NPP*) ATMS. ATMS has 22 channels operating roughly between the frequency range of 23–190 GHz (Kim et al. 2014) and GMI has 13 channels operating roughly between 10 and 190 GHz (Draper et al. 2015). The ATMS beamwidth is 5.2° for the first two channels, 2.2° for Channels 3–16, and 1.1° for Channels 17–22. The beamwidth for GMI changes from 1.75° for the first two channels to 0.4° for Channels 12 and 13. The ATMS and GMI orbits have different altitudes and inclination so that for GMI each degree of beamwidth corresponds to about a 10-km footprint size, but for ATMS each degree is roughly equal to a 15-km footprint size.

Overall, observations from the 183 GHz absorption feature are sensitive to tropospheric humidity, ice clouds and frozen precipitation and observations from 150 GHz are sensitive to scattering from precipitation-sized ice particles. As the frequency decreases, below 90 GHz, the measurements become less sensitive to ice clouds and are mostly affected by hail and raindrops.

ATMS observations from the temperature sounding channels are used to better constrain the retrievals of atmospheric temperature. The weighting functions for the ATMS channels operating in the frequency range of 50–60 GHz mostly reside above precipitating clouds (due to oxygen absorption), therefore they are not important for rain absorption or ice scattering. In addition, only very large ice particles found in deep convection have significant brightness temperature depressions around 30 GHz, though ATMS channel 3 operating at 50.3 GHz has a significant contribution from raindrops and hail. The GMI channels on the other hand are generally designed to have maximum sensitivity to precipitation and clouds. A drawback of the GMI channel selection is that there are no independent temperature sounding channels, but some of the channels indirectly provide information on the atmospheric temperature. For more information about the scattering properties in microwave frequencies, the reader is referred to Kulie et al. (2010) as well as Ekelund and Eriksson (2019).

ATMS/NPP generally provides two overpasses each day for a hurricane, one during the ascending and one during the descending phase of the orbit. However, since GPM inclination is 65°, depending on the location of the storm, we may have more than two overpasses per day. All GMI channels were used in the BMCI retrievals; however, Channels 13, 14, and 15 of ATMS were not used because they have essentially no tropospheric contribution to their weighting functions. We limited the study to observations over ocean only, due to difficulties in calculating emissivity over land at microwave frequencies. In addition, calculating surface emissivity over frozen waters is

not trivial, which introduces errors in near-surface retrievals such as sea surface temperature, and therefore we limited the study to the latitude band between 45°S and 45°N. This should not introduce any problem for the current study as the tropical cyclones normally form and dissipate within this zone.

3. The 1D Bayesian retrieval method

a. Overview of method

The 1D Bayesian retrieval algorithm developed in Evans et al. (2012) was extensively modified for use with lower microwave frequency sensing of tropical cyclones. The two main parts of this algorithm are 1) the construction of a high-dimensional prior probability density function (pdf) of atmospheric properties, and 2) the hybrid BMCI and optimization retrieval method.

The prior pdf is derived from satellite radar reflectivity profiles combined with cloud and precipitation statistics from in situ microphysical probes. Temperature and relative humidity profiles associated with each *CloudSat* radar profile are obtained from reanalysis datasets, though the humidity is modified in the presence of hydrometeors. Relative humidity adjustments in the presence of ice particles is described in Evans et al. (2012); in the presence of liquid cloud, the relative humidity varies between 1.0 and 1.002 linearly as a function of cloud LWC until LWC reaches $2\,\mathrm{g\,m^{-2}}$, thus preserving correlation with cloud LWC. Radar reflectivity, including attenuation corrections, is combined with microphysical statistics to stochastically generate one or more profiles of hydrometeor properties for each radar profile.

The prior pdf is in the form of cumulative distribution functions (CDFs) of each variable (e.g., profiles of temperature, relative humidity, and microphysical properties) and empirical orthogonal functions (EOFs) made from rank correlations between the variables. Using a function for the prior pdf instead of the original radar profiles allows many more profiles (e.g., 10^6) in the Monte Carlo integration and is the only option for using an optimization-based retrieval. The atmospheric profile is represented by temperature and humidity at specified levels and several microphysical variables for three hydrometeor types in appropriate layers. The atmospheric profile and surface properties are denoted by $x_i^{(k)}$ for variable i and radar derived profile k. Each variable x_i is sorted independently over the profiles to make the CDFs $D_i(x_i)$, which range from 0 to 1. A small amount of the information in the joint distribution among all the x_i is preserved in a kind of rank correlation matrix between the variables. The ranks or probabilities representing the variables are converted to standard Gaussian deviates using

$$\xi_i = \Phi^{-1}[D_i(x_i)], \tag{1}$$

where Φ is the cumulative distribution function of the standard normal distribution. The covariance matrix from which the EOFs are derived is calculated from these Gaussian deviates. Since the Gaussian deviates ξ_i have zero mean and unit variance, the covariance matrix is also the correlation matrix:

$$\mathbf{C}_{ij} = \frac{1}{N_{\mathrm{prof}}} \sum_{k=1}^{N_{\mathrm{prof}}} \xi_i^{(k)} \xi_j^{(k)}. \tag{2}$$

The eigenvectors of the correlation matrix $\mathbf{C}_{ij}$ are the EOFs, and the EOF amplitudes are the square root of the eigenvalues.

Variables are generated stochastically from the CDF–EOF prior pdf by 1) creating a vector of independent standard Gaussian random deviates ($\boldsymbol{\xi}$), 2) linearly transforming this vector to correlated Gaussian variables $\boldsymbol{\zeta}$ using the EOF amplitudes and eigenvectors, and 3) transforming the correlated Gaussian variables using the CDFs to the geophysical variables. For example, the temperature and relative humidity at each level is obtained from

$$x_i = D_i^{-1}[\Phi(\zeta_i)], \tag{3}$$

where D_i^{-1} is the inverse of the CDF for temperature or relative humidity for a particular level.

The microphysical variables that specify the gamma size distribution for each hydrometeor type are the mass content, mean mass-weighted equivalent sphere diameter D_{me}, and the mass weighted dispersion. The cloud droplets and raindrops are assumed to be spherical. There are four types of ice hydrometeors (Evans et al. 2012): hexagonal plates and small aggregates of plates, aggregates of spheres representing frozen droplets and graupel, aggregates of 2D dendrites representing snowflakes, and large solid spheres representing hail. Nonspherical particle scattering is calculated using the discrete dipole approximation. One-dimensional radiative transfer is calculated for radiometers using "SHDOMPPDA" (Evans 2007) with four discrete ordinates. Molecular absorption over the channel spectral responses is calculated using "MonoRTM" (Clough et al. 2005).

The Bayes's theorem conditional pdf of the radiometer brightness temperatures given an atmospheric profile $\mathbf{x}$ is assumed to be an uncorrelated Gaussian distribution, with a specified width σ_j for each channel, around the observation vector simulated with a radiative transfer model:

$$p_{\mathrm{cond}} \propto \exp\left(-\frac{1}{2}\chi^2\right) \chi^2 = \sum_{j=1}^{N_{\mathrm{chan}}} \frac{[O_j - H_j(\mathbf{x})]^2}{\sigma_j^2}, \tag{4}$$

where O_j is the observation for channel j and $H_j(\mathbf{x})$ is the radiative transfer observational operation. The highly efficient MCI method distributes the atmospheric cases according to the prior pdf, precomputes the radiative transfer, and stores the simulated observations in a retrieval database. Since the atmospheric cases are distributed according to the prior pdf, the Bayesian posterior pdf is proportional to the conditional pdf p_{cond}. The MCI retrievals are the mean of the desired variables weighted by the Bayesian posterior pdf, which is simply a weighted sum over the retrieval database cases:

$$\hat{x} = \frac{\sum_i \mathbf{x}_i \exp\left(-\frac{1}{2}\chi_i^2\right)}{\sum_i \exp\left(-\frac{1}{2}\chi_i^2\right)}. \tag{5}$$

Uncertainties for each retrieved variable are similarly calculated using the weighted standard deviation.

When an MCI retrieval does not have enough database cases that match the observation vector (to within a specified χ^2), then an optimization is performed to maximize the posterior pdf. This usually happens when the observation is from deep convection. The optimization minimizes a least squares cost function using gradient information, which is much slower than MCI. The highly non-Gaussian prior pdf is included in the χ^2 calculation using a control variable transformation, $\mathbf{x} = G(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is the control vector and $G(\boldsymbol{\xi})$ is the function represented by the algorithm described in the paragraph surrounding Eq. (3). This allows the retrieval to be performed with the efficient Levenberg–Marquardt least squares minimization method for the function $H_j[G(\boldsymbol{\xi})]$.

b. The BMCI modifications

The major changes in the 1D retrieval algorithm from Evans et al. (2012) are listed below.

- The melting/melted particle hydrometeor component has been replaced with spherical raindrops, so the three components are ice particles, raindrops, and liquid cloud droplets.
- The FASTEM microwave ocean surface emissivity model (Liu et al. 2011), with its adjoint (gradient) was added. Skin temperature and surface wind speed were added to the control vector, and prior information for these variables was obtained from the *CloudSat* ECMWF-AUX(2) data files.
- The temperature profile is now a retrieved variable.
- In addition to the 1-sigma uncertainty for each variable, parts of the full error covariance matrix may be retrieved.
- ERA-Interim profiles of stratospheric temperature and water vapor matched to *CloudSat* times and locations are included to complement the *CloudSat* ECMWF-AUX profiles (which only reach 24 km) in the prior pdf.
- The control vector now includes a hydrometeor masking variable for ice, rain, and liquid cloud. This allows for clear layers, exactly matching the hydrometeor occurrence probabilities for each layer, and substantially improves the integrated mass content (e.g., IWP) representation of the CDF-EOF prior pdf. The masking variable M_k for a particular hydrometeor type and layer is defined by the differentiable indicator function:

$$M_k = \frac{1}{1 + \exp[-10\,000(\zeta_k - t_k)]}, \qquad (6)$$

where ζ_k is the correlated Gaussian element for the masking variable and t_k is the threshold that results in the correct probability of occurrence for the hydrometeor type and layer. The masking variable multiplies the corresponding hydrometeor water content variable:

$$x_i = M_k D_i^{-1}[\Psi_i(\zeta_i)] \quad \text{for water content} \quad \text{and} \qquad (7)$$

$$x_i = D_i^{-1}[\Psi_i(\zeta_i)] \quad \text{for } D_{\text{me}} \quad \text{and dispersion}, \qquad (8)$$

where D_i^{-1} is the inverse of the CDF for the hydrometeor property and layer, $\Psi_i(\zeta_i)$ is a function that translates the Gaussian distributed ζ_i to a probability (0 to 1) in a manner that corrects the biased distribution [if $\Phi(\zeta_i)$ is used] caused by the correlation between the masking (ζ_k) and hydrometeor

property (ζ_i) control vector elements. Special procedures are also used to generate the correlation matrix ($\mathbf{C}_{ij}$) elements pertaining to hydrometeor mask and property variables due to 1) not including input profiles in the sum when a layer does not have a particular type of hydrometeor and 2) the hydrometeor mask variables having binary (0 or 1) values that do not map onto a Gaussian distribution. Last, a hydrometeor properties and mask correlation tuning procedure is available (using rejection of certain input hydrometeor profiles) to improve the agreement between the radar derived and CDF-EOF generated water path distributions.

- Prior pdf statistics for warm cloud droplets and raindrops in tropical cyclones were made by analyzing in situ cloud and rain microphysical data from 10 flights provided by the NOAA Hurricane Research Division (HRD). Liquid water content (LWC) and D_{me} for cloud droplets were derived from the Cloud Droplet Probe (CDP) size distributions (averaged to 1700-m distance), and LWC, D_{me}, and dispersion for raindrops were derived from Precipitation Imaging Probe (PIP) and Cloud Imaging Probe (CIP) size distributions (all three probes were made by Droplet Measurement Technologies).

- The original *CloudSat* reflectivity profile-based prior pdf generation was modified to also use GPM Dual-frequency Precipitation Radar (DPR) reflectivity profiles. When the simulated hydrometeor 94-GHz two-way attenuation for a *CloudSat* profile reaches 5 dB or the layer is below the freezing level (and the simulated Ku reflectivity is above a threshold), a matching DPR profile is found from a dataset of 233 882 profiles. The DPR profile most closely matching the simulated Ku reflectivity and transition layer temperature (from collocated ERA-Interim reanalysis) is selected. At freezing temperatures, the original ice hydrometeor generation procedure is used with *Cloudsat* reflectivity above the transition level and with the DPR Ku reflectivity below the transition level. At warmer temperatures, the DPR operationally retrieved raindrop LWC and D_{me} are used.

- When operationally retrieved DPR rain profiles are used, warm liquid cloud profiles consistent with the DPR rain properties are generated stochastically from statistical lookup tables. These lookup tables, which include the probability of liquid cloud occurrence and covarying statistics for LWC and D_{me} as a function of rain LWC, are created from the HRD microphysical samples with a CDF-EOF procedure.

- When *CloudSat* radar reflectivity is not too attenuated, then it is used to stochastically generate liquid cloud and rain (actually drizzle) properties in warm layers from statistical lookup tables. The three related lookup tables (and corresponding probability of occurrence) are for cloud only, rain only, and both cloud and rain, and are a function of 94-GHz unattenuated *CloudSat* reflectivity. The liquid cloud and/or rain properties are generated with a CDF-EOF procedure from the HRD microphysical samples

- Supercooled liquid cloud parameters in the presence of ice particles are generated with a new procedure using a file of in situ derived microphysical samples from the Tropical Composition, Cloud and Climate Coupling experiment (TC4) (Toon et al. 2010), rather than the previous method of using a
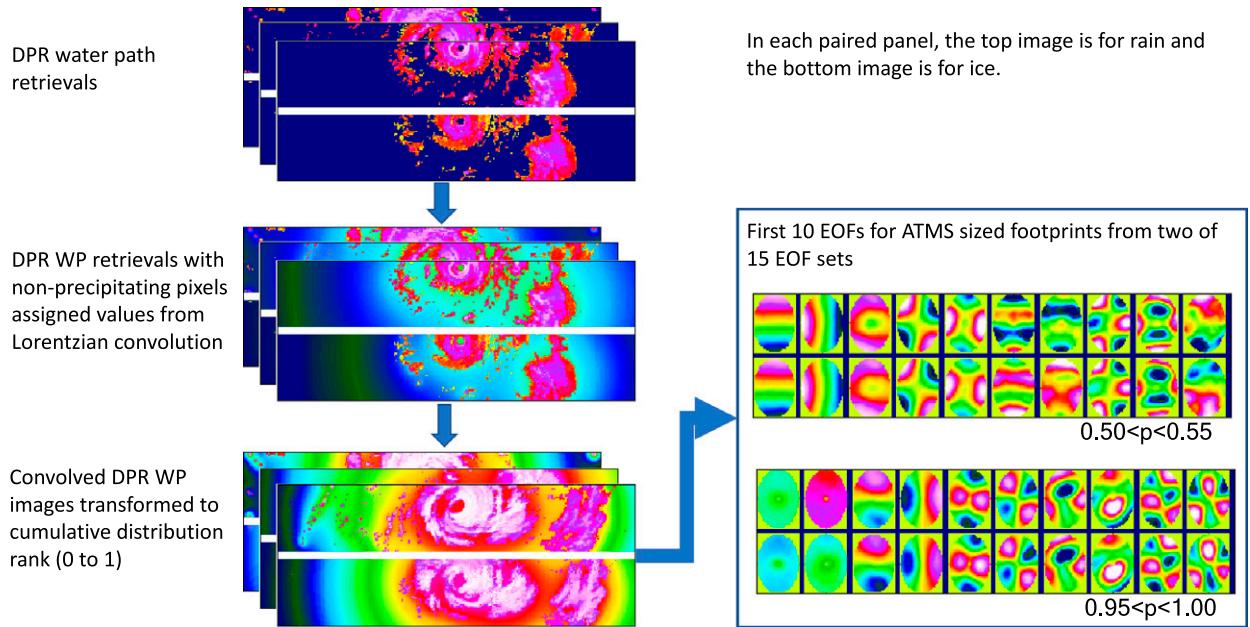
FIG. 1. A flowchart of the procedure used to generate spatial statistics of rain and ice water path (in the form of EOFs used for the beam-filling correction) from DPR retrievals for 218 hurricane overpasses. Each EOF set is for a particular interval (e.g., $0.95 < p \leq 1.00$) in the central pixel rainwater path cumulative distribution.

poorly fit Gaussian distribution for T, $\ln(\text{IWC})$, $\ln(\text{LWC})$, and $\ln(D_{\text{me,liq}})$.

- The stochastic generation of thin ice cloud layers detected by *CALIPSO* lidar but not by *CloudSat* radar was removed because they are completely unobservable by microwave radiometer channels below 300 GHz.
- The specific radar profiles used here are from 114 *CloudSat* tropical cyclone granules in the Atlantic and eastern Pacific Ocean with overpasses within 50 km of the cyclone centers and 218 2 A-DPR files subsetted for hurricanes in the Atlantic and eastern Pacific.

c. Beam-filling correction

The beam-filling problem in microwave remote sensing of precipitation (e.g., Kummerow 1998) is caused by the small-scale variability of precipitation compared to microwave radiometer footprint sizes, combined with the nonlinear relationship between precipitation and microwave brightness temperatures. This results in biased estimates of precipitation when assuming uniformly filled footprints. It is especially important for sounding instruments, such as ATMS with nadir footprint sizes ranging from 16 to 75 km.

A beam-filling bias correction procedure was developed to statistically account for the radiative transfer effects of realistic rain and ice horizontal variability over large passive microwave footprints. The source of precipitation horizontal variability statistics is DPR retrievals of liquid (rain) and frozen (ice) precipitation water path images (49 pixel wide swaths at 5.0 km resolution) for 218 hurricane overpasses. There are two parts to the beam-filling correction procedure: 1) encoding of the DPR water path spatial variability statistics in the form of

EOFs over a maximally sized footprint, and 2) using the water path EOFs and the 1D profile prior pdf (described above) to generate stochastic 3D fields of hydrometeor properties in each footprint. Independent column radiative transfer with beam weighting over the 3D fields in a footprint is done to calculate brightness temperatures with hydrometeor variability. Uniform footprint brightness temperatures are calculated from the atmospheric columns averaged over a suitably defined target footprint. The statistics of the difference between the variable and uniform footprint brightness temperatures are used to remove the beam-filling bias and increase the assumed uncertainties in the observed brightness temperatures.

The encoding of the DPR water path spatial variability statistics is done only once to prepare a footprint water path EOF file for input to the retrieval program. This procedure is outlined as a graphical flowchart in Fig. 1. Since the DPR water path values will be sorted and converted to CDF ranks (or probabilities between 0 and 1), the water path values for non-precipitating pixels are scaled based on the distance from the precipitating pixels, rather than all being set to zero. Specifically, the rain and ice water path images for nonprecipitating clouds are convolved with a Lorentzian function as follows:

$$\hat{W}_{\text{rain/ice}}(x_i, y_j) = 0.0002 \frac{\sum_{k,l} [R_c^2 + r^2]^{-1} W_{\text{rain/ice}}(x_k, y_l)}{\sum_{k,l} [R_c^2 + r^2]^{-1}} \quad \text{for}$$

$$W(x_i, y_j) < 1 \,\text{g m}^{-2} \quad \text{and} \tag{9}$$

$$\hat{W}_{\text{rain/ice}}(x_i, y_j) = W_{\text{rain/ice}}(x_i, y_j) \quad \text{for} \quad W(x_i, y_j) \geq 1 \,\text{g m}^{-2},$$
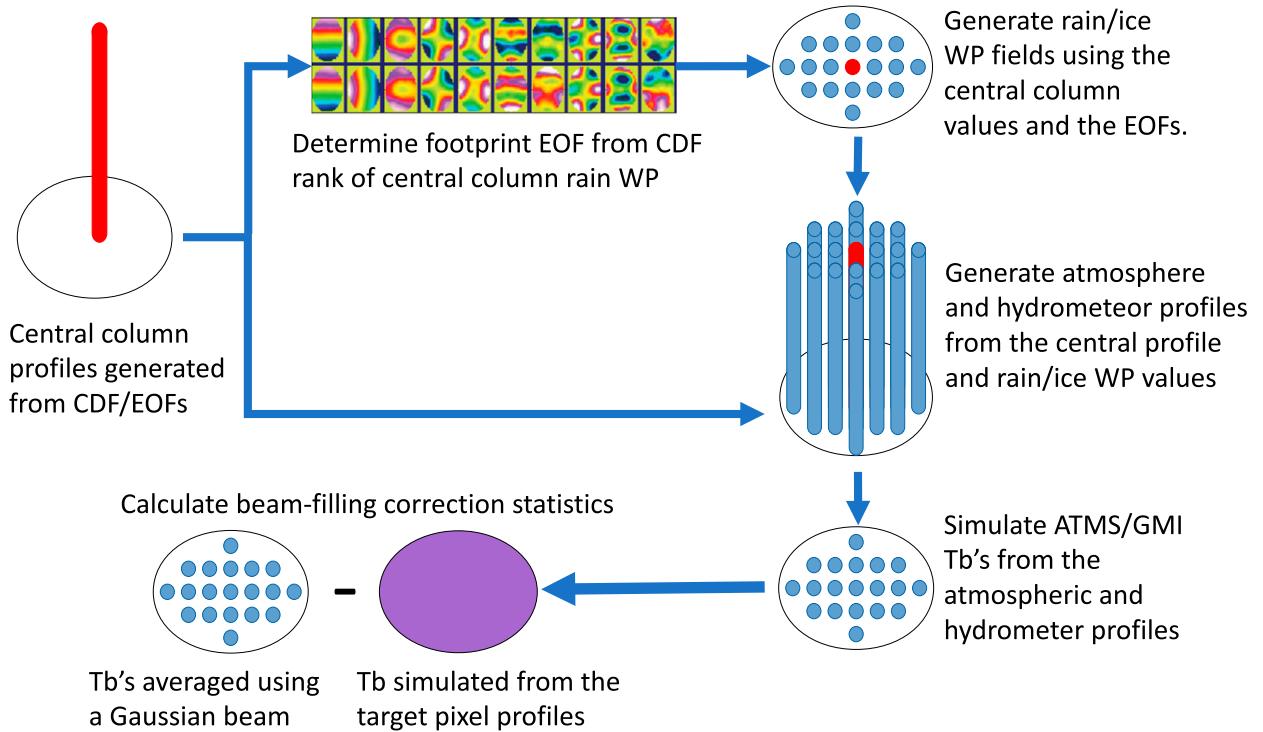
$$\tag{10}$$

FIG. 2. A flowchart of the procedure used to generate stochastic 3D fields of atmospheric profiles in radiometer footprints, and the use of these footprint fields to calculate beam-filling corrections to the brightness temperatures.

where $W_{\mathrm{rain/ice}}$ are the original water path values in images indexed by i, j or k, l, $r^2 = (x_k - x_i)^2 + (y_l - y_i)^2$, and $R_c = 8\,\mathrm{km}$ is the half-width at half-maximum convolving radius. The convolved water path values $\hat{W}_{\mathrm{rain/ice}}$ are sorted over all the DPR overpass images to obtain the CDF function $D_{\mathrm{rainWP/iceWP}}(\hat{W})$ and the corresponding rank values (p_{ij} from 0 to 1) for each pixel. These CDF rank values p_{ij} are transformed to Gaussian distributed values with $\Xi_{ij} = \Phi^{-1}(p_{ij})$ [as above, $\Phi(\Xi)$ is the cumulative distribution function of the standard normal distribution].

The spatial variability EOFs are designed to generate a 2D field of water path in the rest of a footprint, given the central pixel rain/ice water path Gaussian rank values $\Xi_{kl,\mathrm{cent}}$. That is, the EOFs are calculated for the differences in Gaussian rank values from the central pixel values:

$$\boldsymbol{\delta\Xi}_{i-k,j-l} = \Xi_{ij} - \Xi_{kl,\mathrm{cent}} \qquad (11)$$

The Gaussian rank differences ($\boldsymbol{\delta\Xi}$) obviously depend on the values of the central pixels (Ξ_{cent}). For example, the rainwater path field surrounding a very high central column rainwater path will tend to trend downward with distance from the footprint center, though at a suitably large distance it will depend very little on the central value. Therefore, multiple EOF sets are made that depend on the CDF rank of the central column (p_{cent}). Here 15 EOF sets are defined with 5 p_{cent} intervals from 0 to 0.5 and 10 p_{cent} intervals from 0.5 to 1.

By definition, EOFs are the eigenvectors of a covariance matrix calculated from a sequence of vectors. Here the vectors

are the $\boldsymbol{\delta\Xi}_{i-k,j-l}$ for the pixels inside a specified ellipse centered on the central pixel, but not including the central pixel itself, for the rain and ice images combined. The mean vector and covariance matrix for each EOF set are accumulated over all the feasible central pixels for the specified size ellipse and over all the DPR overpasses. The EOFs in each set are the eigenvectors of the covariance matrix, but the mean $\boldsymbol{\delta\Xi}$ vector and the eigenvalues are also needed to generate stochastic rain and ice water path fields. For ATMS with a maximum nadir footprint size of 75 km (full-width at half-maximum) and allowing zenith viewing angles up to about 45°, the EOF footprint size is 280 km by 200 km. The elliptical footprints are 57 by 41 pixels with 1757 columns and an EOF vector length of 3512. For 99% of the variance, 2017–2432 EOFs are required (depending on the set), resulting in a very large file size for all 15 EOF sets. Undoubtedly, more efficient statistics could be created by assuming certain symmetries.

The footprint water path EOFs are used in the retrieval program to generate stochastic 3D atmosphere/hydrometeor fields inside footprints for the radiative transfer calculation (see Fig. 2 for a flowchart overview of the procedure). As described previously, the water path EOFs are in terms of the CDF rank differences from the central column. The rain and precipitating ice water path CDFs used with the EOFs are calculated from 10 000 randomly generated profiles from the 1D CDF-EOF prior, and the water path CDFs derived from the DPR retrievals are not used. Precipitating ice is defined here as the ice water path from 5 to 12 km. The first step in the footprint column generation procedure is to use the 1D

CDF-EOF profile generation algorithm (see section 3a) for the central column.

The second step is to stochastically generate the rain and ice water paths for the rest of the footprint columns using the water path EOFs. The rainwater path is calculated for the central column, converted to a CDF rank ($p_{\text{cent,rain}}$), and then used to interpolate between two EOF sets. A Gaussian random deviate vector is weighted by the EOF amplitudes, multiplied by the EOF eigenvector matrix, and added to the mean EOF vector to generate a stochastic vector of Gaussian rank differences, $\boldsymbol{\delta\Xi}_{\text{rainWP/iceWP}}(\Delta i, \Delta j)$. These Gaussian rank differences are added to the central column values and transformed with the water path CDF functions to make the 2D field of rain and ice water path:

$$W_{\text{rain/ice}}(\Delta i, \Delta j) = D_{\text{rainWP/iceWP}}^{-1}\{\Phi[\Xi_{\text{cent,rainWP/iceWP}} + \boldsymbol{\delta\Xi}_{\text{rainWP/iceWP}}(\Delta i, \Delta j)]\}. \quad (12)$$

The third step is to generate stochastic profiles for the rest of the footprint columns from the central profile 1D control vector ($\boldsymbol{\xi}_{\text{cent}}$) and the rain and ice water path Gaussian rank differences of the columns [$\boldsymbol{\delta\Xi}_{\text{rainWP/iceWP}}(\Delta i, \Delta j)$]. First, the rain or ice hydrometeor mask profile control elements for a column [$\xi_{\text{rainMask/iceMask}}(z, \Delta i, j)$] are calculated by perturbing the central column mask control elements [$\xi_{\text{cent,rainMask/iceMask}}(z)$] by an offset ($g_{\text{rain/ice}}$). If the desired water path is zero, then the mask offset value is set to assure that no layers have hydrometeors. Otherwise, the offset $g_{\text{rain/ice}}$ is the difference between the mean hydrometeor mask control elements [$\overline{g}(W)$] for the desired water path (obtained from the 10 000 cases in the CDF calculation) and the vertical mean of the central column control elements:

$$g_{\text{rain/ice}} = \overline{g}_{\text{rain/ice}}[W_{\text{rain/ice}}(\Delta i, \Delta j)] - \text{Mean}[\xi_{\text{rainWC/iceWC}}(z)] \quad \text{and} \quad (13)$$

$$\xi_{\text{rainMask/iceMask}}(z, \Delta i, \Delta j) = \xi_{\text{cent,rainMask/iceMask}}(z) + g_{\text{rain/ice}}. \quad (14)$$

Then the control vector elements for the rain and ice water content in a column [$\xi_{\text{rainWP/iceWP}}(z, \Delta i, \Delta j)$] are obtained by optimizing a factor $f_{\text{rain/ice}}$ that multiplies the 2D field water path Gaussian rank differences [$\boldsymbol{\delta\Xi}_{\text{rainWP/iceWP}}(\Delta i, \Delta j)$] and is added to the central column LWC/IWC control elements [$\xi_{\text{cent,rainWC/iceWC}}(z)$]:

$$\xi_{\text{rainWC/iceWC}}(z, \Delta i, \Delta j) = \xi_{\text{cent,rainMask/iceMask}}(z) + f_{\text{rain/ice}}\boldsymbol{\delta\Xi}_{\text{rainWP/iceWP}}(\Delta i, \Delta j). \quad (15)$$

These two single variable ($f_{\text{rain/ice}}$) optimizations assures that the column rain and precipitating ice water paths exactly match the desired values from the 2D stochastic footprint water paths. Finally, most of the other control variable elements (for temperature, water vapor, and hydrometeor properties and masks in each layer) in the column are adjusted from the central column values using the variable's correlation (ρ in the Gaussianized rank space) with the rain or ice water content of that layer and the Gaussian rank differences for rain or ice water path:

$$\xi_{\text{par}}(z, \Delta i, \Delta j) = \xi_{\text{par,cent}}(z) + \rho_{\text{par,rainWC}}(z)f_{\text{rain}}\boldsymbol{\delta\Xi}_{\text{rainWP}}(\Delta i, \Delta j), \quad (16)$$

where par stands for temperature, relative humidity, rain D_{me}, liquid cloud mask, and cloud liquid water content. The rainwater path Gaussian differences ($\boldsymbol{\delta\Xi}_{\text{rainWP}}$) are used for the layers where rain is possible. The surface temperature and wind speed similarly use the correlation with rain LWC in the lowest layer. The ice water path Gaussian differences ($\boldsymbol{\delta\Xi}_{\text{rainWP}}$) are used for the layers where ice is possible or using the highest "precipitating ice" layer:

$$\xi_{\text{par}}(z, \Delta i, \Delta j) = \xi_{\text{par,cent}}(z) + \rho_{\text{par,iceWC}}(z)f_{\text{ice}}\boldsymbol{\delta\Xi}_{\text{iceWP}}(\Delta i, \Delta j), \quad (17)$$

where par stands for temperature, relative humidity, ice D_{me}, cloud liquid mask and water content, and ice mask and ice water content for layers above 12 km. The control vector elements for the ice and rain size distribution dispersion and the cloud D_{me}, which have small effects on microwave radiation, use the central column values.

Figure 3 illustrates hydrometeor and water vapor water paths for 10 stochastically generated footprints. These footprint images show the detailed horizontal structure of the hydrometeors and the correlations between the different hydrometeor types.

Having generated the atmospheric and hydrometeor profiles for all the columns in a footprint, the radiative transfer can proceed. The retrieval quantities (e.g., profiles of water vapor and ice water content) are obtained from the 3D field of footprint columns using the concept of a target beam. Typically this target beam will match the footprint (or beam) of some of the instrument channels at some intermediate scan angle. In this work the target beam size is 36 km for ATMS and 12 km for GMI.

Two sets of brightness temperatures (for each instrument channel and viewing angle) are calculated and output to the retrieval database: 1) the footprint averaged brightness temperature calculated with 1D radiative transfer on each column in the footprint (the independent column approximation), and 2) a single 1D radiative transfer calculation for the average profile over the target beam. The brightness temperatures for each column are weighted according to an elliptical Gaussian beam pattern appropriate for each channel and viewing angle. Only those columns with weights above some value (e.g., 0.06) enter into the radiative transfer calculation for a particular channel.

The brightness temperature difference between the uniform target beam $T_{b,\text{uniform}}$ and the variable footprint average $T_{b,\text{variable}}$, denoted by ΔT_{bf}, is a measure of the beam-filling error of assuming uniformity for a particular stochastic footprint field. The standard deviation and mean of ΔT_{bf} over all cases in the retrieval database for each channel and viewing angle are used for the correction. The beam-filling corrections are applied differently for the two types of retrievals (MCI and optimization). In both cases the Bayesian σ for each channel is calculated from a quadrature average of the input uncertainties and the beam-filling σ. Monte Carlo integration retrievals use
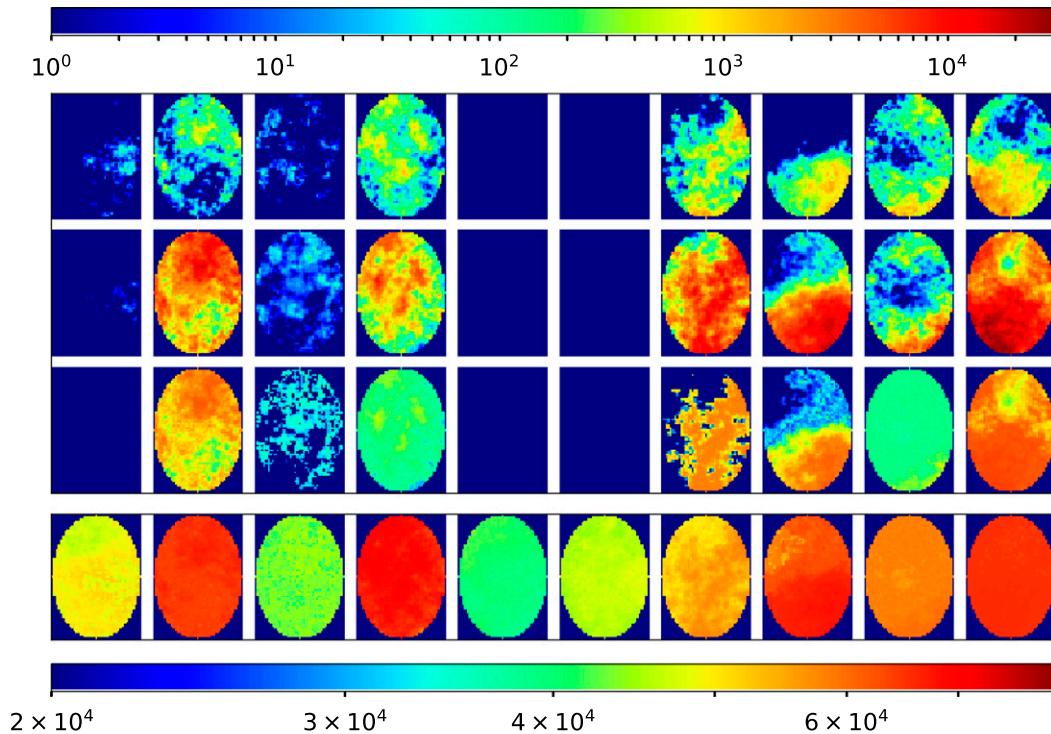
FIG. 3. Water path $(g\,m^{-2})$ images for 10 stochastically generated footprints: (top) rain, (top middle) ice, (bottom middle) cloud liquid, and (bottom) water vapor. The bottom color bar is for water vapor images, and the top color bar is for the rest of the fields.

the footprint averaged brightness temperatures ($T_{b,\text{variable}}$) in the retrieval database, which include the subfootprint variability, and therefore the bias correction does not need to be applied separately. For optimization retrievals, the beam-filling bias is subtracted from the input observed brightness temperatures. The optimization retrievals use the 1D profile generation procedure, because the control vector does not include elements that drive the stochastic footprint variability (since subfootprint variability cannot be retrieved).

Figure 4 shows beam-filling brightness temperature difference between the uniform target beam brightness temperatures and the variable footprint average brightness temperatures (ΔT_{bf}) statistics as a function of either rain or ice water path for selected GMI channels and ATMS channels at about 32° viewing zenith angle. In general, the beam-filling errors are larger for the lower frequency channels with larger beamwidths (see section 2) and for channels with larger radiative effect from rain or ice water path. The beam-filling effect is generally positive for rain-sensitive channels and negative for channels sensitive to ice particles. The standard deviations and the means (biases) for the 5000 case blocks can be quite large, often more than 10 K for the lowest frequency channels. The beam-filling errors are smaller when the target beam size is closer in size to the channel beamwidth. Larger biases are also associated with larger standard deviations, which can be explained by the size of the beamwidth. The larger the footprint is, the larger is the possibility for mixtures of intense precipitation and clear sky within the field of view. ATMS

temperature sounding channels (channels 4–15) are only slightly sensitive to precipitation, and therefore are not shown in Fig. 4.

4. Data assimilation experiments

The second part of the work was to investigate how the temperature and humidity profiles retrieved from the BMCI method can be assimilated into the GEOS forecast system. Results are shown for several data assimilation experiments focused on Hurricane Maria (Table 1), a category-5 storm that affected the Caribbean Sea and southeastern United States during September and October 2017. For all of the experiments, the GEOS model was run at c360 resolution, corresponding to a horizontal spacing of approximately 25 km at the equator, with 72 vertical levels from the surface to 0.01 hPa (Putman and Lin 2007). The data assimilation was performed at a horizontal resolution of 50 km using a hybrid four-dimensional ensemble-variational (4D-EnVar) algorithm with a 6-h update cycle and 32 ensemble members (Todling and El Akkraoui 2018). The analysis increment was applied as a correction to the background state using an incremental analysis update procedure.

The control experiment (FP-Ctrl) used the same observing system configuration as the GEOS real-time production system at the time of Hurricane Maria, which included clear-sky satellite microwave and infrared radiances from multiple platforms (including ATMS and GMI), conventional observations from radiosondes and other sources, Advanced Scatterometer
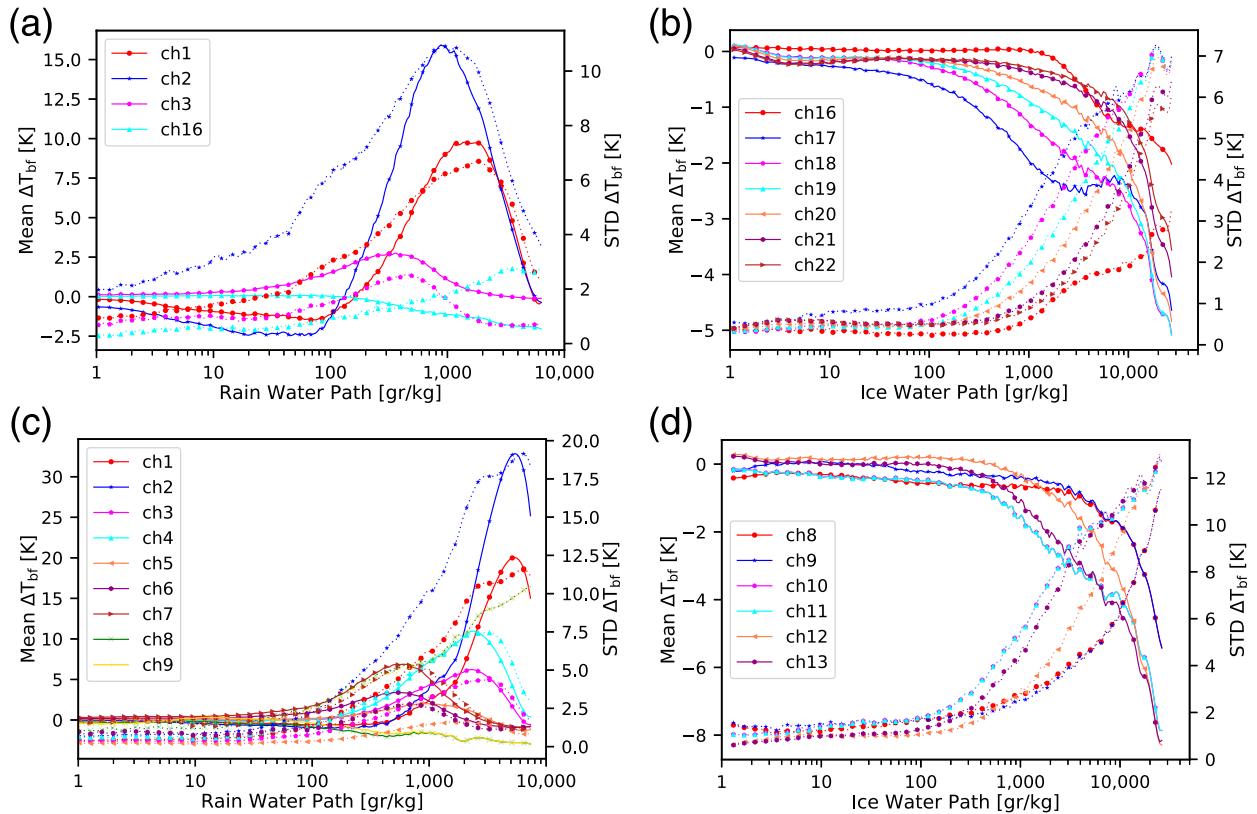
FIG. 4. The beam-filling impact on (a),(b) ATMS and (c),(d) GMI for (left) lower-frequency channels sensitive to liquid and (right) higher-frequency channels sensitive to ice particles. The solid lines are for bias, and the dashed lines for standard deviation.

(ASCAT) surface winds, and Global Positioning System radio occultations (GPS-RO).

In addition to the control experiment, three experiments were conducted assimilating BMCI retrievals of temperature T, water vapor q, and SST. The first of these (FP+Both) assimilated BMCI retrievals from both ATMS and GMI in addition to the observations assimilated in FP-Ctrl. In this experiment no horizontal thinning was performed for the BMCI retrievals, but the water vapor profiles were thinned in the vertical at a resolution of 50 hPa to diminish the effects of correlated observation error (discussed later in this section). The second experiment (FP+Both-Thin) also assimilated BMCI retrievals from ATMS and GMI, but differed in the way the retrieved observations were thinned. In this case, the retrievals for all variables were thinned horizontally at a resolution of 50 km, while the retrieved profiles of temperature

and water vapor were thinned in the vertical direction at 50 and 100 hPa, respectively. The third experiment with BMCI retrievals (FP+ATMS-Thin) is similar to the second experiment, but the GMI retrievals were excluded from the assimilation process. A final experiment was conducted assimilating only conventional data (Conv-Only) to evaluate the impact of satellite observations in general on the forecasts of Maria.

a. Observation error

One of the advantages of the BMCI technique is that an estimate of the observation error can be obtained during retrieval process (Evans et al. 2012). These estimates were used during the assimilation process to determine the observation error. Figure 5 shows an example of the retrieved observation error correlation matrix for clear-sky and precipitating cases. As shown in Fig. 5, the retrieved atmospheric temperature

TABLE 1. List of data assimilation experiments along with the observations assimilated into each experiment as well as information about horizontal and vertical thinning.

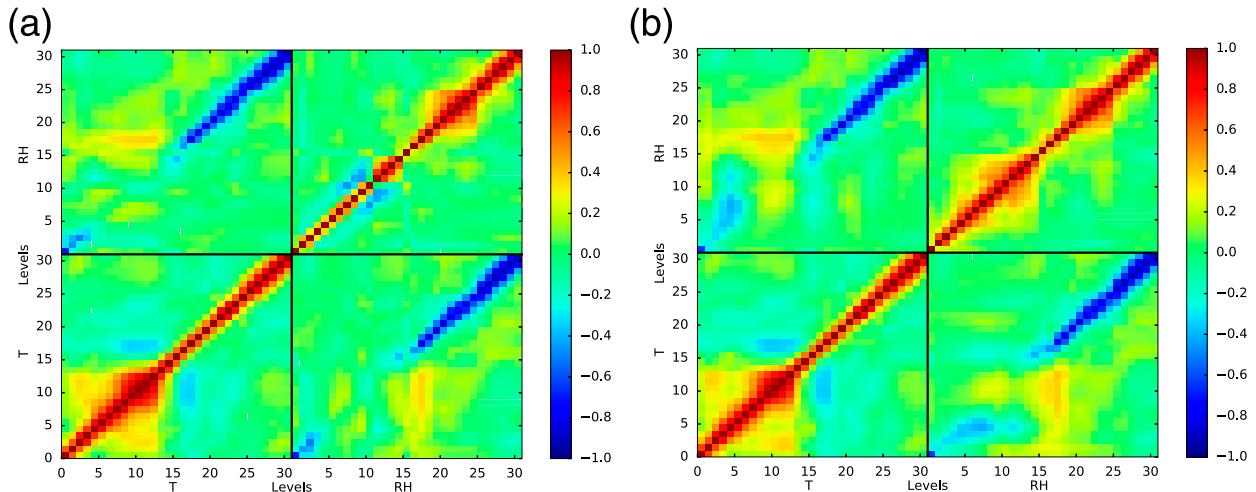| Expt	Obs	Horizontal thinning	Vertical thinning
FP-Ctrl	Standard FP setup	140-km satellite obs	
FP+Both	FP plus ATMS and GMI retrievals		50 hPa for q
FP+Both-Thin	FP plus ATMS and GMI retrievals	50 km for t, q, and SST	100 hPa for q and 50 hPa for t
FP+ATMS-Thin	FP plus ATMS retrievals	50 km for t, q, and SST	100 hPa for q and 50 hPa for t
ConvOnly	Conventional only		

FIG. 5. Averaged retrieved error correlation matrix for temperature and humidity profiles for (a) 942 clear-sky profiles and (b) 383 profiles with $3000 < \text{IWP} < 10\,000 \, \text{g m}^2$. The levels represent altitude in kilometers.

profiles are vertically correlated over a longer distance in the middle and upper troposphere than in the stratosphere. For relative humidity errors in clear profiles there is negative correlation for levels a few km apart in the troposphere (except around the tropopause), but for precipitating profiles the positive correlation extends to large vertical distances. Consequently, in cloudy conditions, the errors for relative humidity profiles become highly vertically correlated in both lower and upper troposphere. We assimilated the BMCI retrievals similar to the way the GSI assimilates radiosonde profiles. GSI currently considers radiosonde observations from different levels as single independent observations, therefore only the diagonal elements of the observation error covariance matrix are used. This means that vertically correlated observation errors cannot be taken into account when assimilating these profiles into GSI. One way to avoid the negative impact from the vertically correlated observation error was to thin the profiles at different levels. The DA system only allows uniform thinning for the entire atmosphere, otherwise the thinning can be performed differently for lower and upper troposphere.

Figure 6 shows the SST omf values for retrievals assimilated over the rainband of Hurricane Maria and as shown the differences are mostly between -2 and -3 K. Although, it is not possible to validate the BMCI retrievals given the lack of any reliable datasets over the rainband of tropical cyclones, it is expected that the first guess for SST is reasonably accurate given that a large amount of SST information and satellite observations sensitive to the surface temperature are assimilated into GEOS. However, because of large errors in calculating sea surface emissivity over cold waters, we discarded all the ATMS and GMI observations beyond 45°N before any processing.

The difference between the observations and first guess (known as innovation) for the air temperature is shown in Fig. 7 for the experiments where the BMCI retrievals were assimilated. The differences range mostly between -1 and 1 K. The innovations are generally positive in lower level, especially

over the rainbands. In mid and upper troposphere, the innovations are only slightly positive over the rainbands, but negative elsewhere. The experiments with thinned observations show similar pattern for omf values in mid and lower troposphere, but in upper troposphere the innovations turn to be mostly positive. Note that the innovations in Fig. 7 are only from the ATMS observations, because there was no GMI observations for the storm during that particular cycle (at 1800 UTC 24 September 2017).

b. Impact on tropical cyclone forecast

The analyzed minimum sea level pressure (SLP) for Hurricane Maria for the period from 17 September 2017 to 2 October 2017 is shown in Fig. 8. The analyzed minimum SLP was based on the minimum pressure in a window expanding 2° in east–west and north–south directions from the storm location provided in the Tropical Cyclone Vitals (TCVitals) (Trahan and Sparling 2012). The storm locations in TCVitals and analysis may not necessary match, but the displacement of the cyclones in the analysis is expected to be small, less than 100 km as later shown in this section. Therefore, a window expanding 2° in each direction should be sufficient to determine the cyclone's center in the analysis.

Figure 8 shows the minimum SLP for different experiments versus the minimum SLP provided in the TCVitals. According to the TCVitals, the storm initially experienced a rapid intensification such that, just 3 days after the storm was developed, its center SLP reached 900 hPa on 20 September. The storm intensity then rapidly decreased so that a day later the storm minimum SLP was recorded to be 960 hPa. The cyclone slightly intensified thereafter and reached around 940 hPa on 25 September and then slowly de-intensified until dissipated in early October. There is a very small difference between different DA experiments in terms of the minimum SLP. In the DA experiments, the storm slowly intensified and the minimum SLP only reached 960 hPa on 21 September. The model missed the initial rapid intensification that happened before
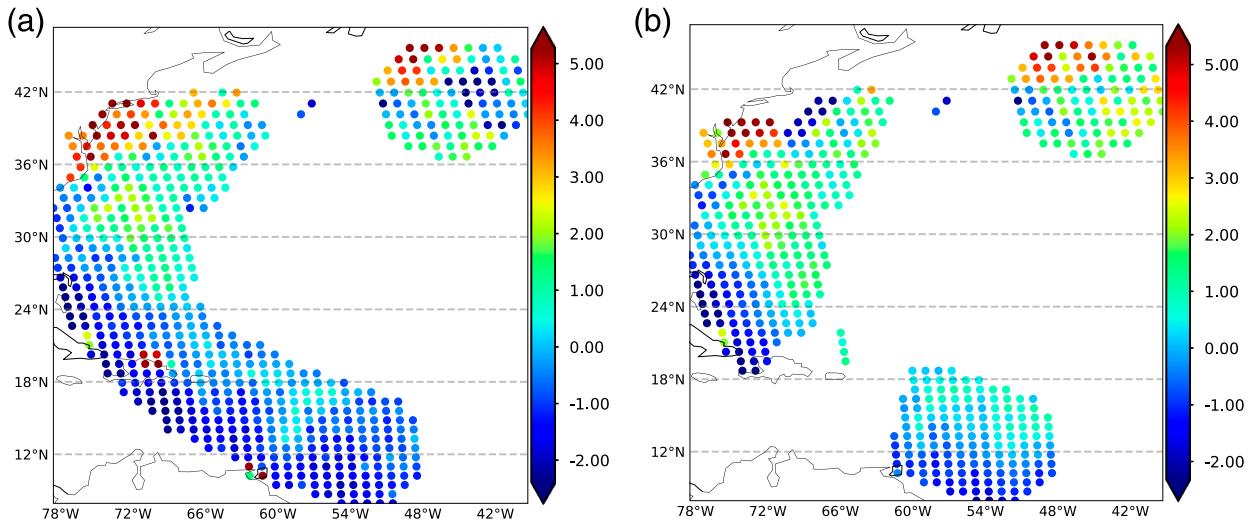
FIG. 6. Difference between BMCI retrieved sea surface temperature and first guess provided by GEOS over the rainbands of Hurricane Maria: (a) FP+Both and (b) FP+ATMS-Thin.

20 September. The storm in the DA experiments then slightly weakened with a minimum SLP reaching almost 980 hPa on 22 September then (after a slight intensification) the storm intensity slightly decreased until it dissipated in the DA experiments in early October. Overall the difference between the minimum SLP in the control experiments and other experiments is less than 10 hPa. The difference between ConvOnly in which only conventional data are assimilated and FP-Ctrl
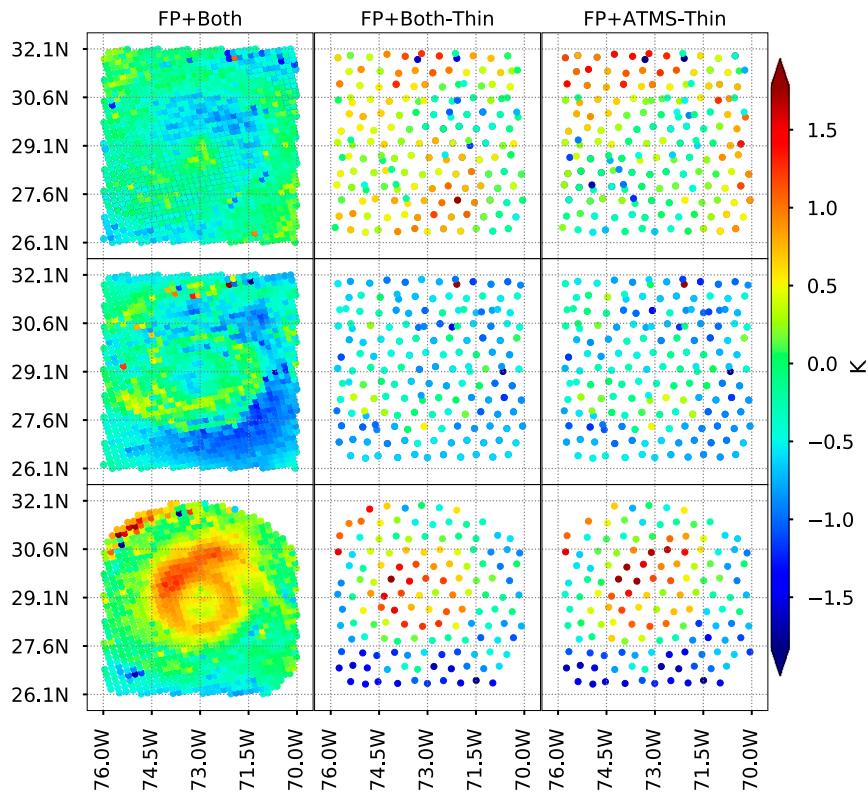
FIG. 7. Observation minus first guess for the BMCI temperature retrievals in different layers of the atmosphere: (top) 100–70, (middle) 500–400, and (bottom) 1000–925 hPa at 1800 UTC 24 Sep 2017.
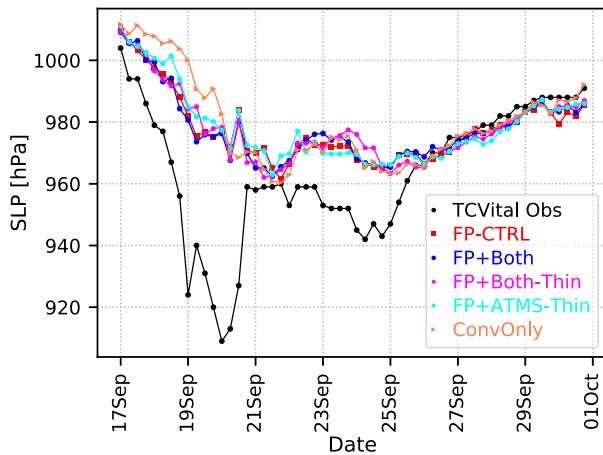
FIG. 8. Minimum sea level pressure in the GEOS analysis for Hurricane Maria.

in which satellite data are also assimilated is not as large as was expected. The main reason probably is that the initiation of storm is affected by mesoscale atmospheric features that are well captured when conventional data are assimilated. There are of course some periods where assimilating satellite observations relatively lowers the minimum SLP. Generally, running the experiments at higher resolutions is expected to produce a stronger storm (lower the center SLP), which may also increase the impact of the BMCI retrievals on the forecasts for the storm intensity or track.

Figure 9 shows the mean track and intensity error for all 5-day forecasts started at 0000 UTC for the period of 17–30 September 2017. Since the location of storm in forecasts can be significantly different from the real storm location, the storm minimum SLP was determined as minimum SLP in a window expanding 5° in each direction from the storm location provided in TCVitals. Unlike intensity, assimilating satellite observations remarkably improves the storm track in the forecast, so that the storm track error is reduced more than 50 km in the first day of the forecast when satellite data are

assimilated (FP-Ctrl vs ConvOnly). The positive impact of satellite observations increases as the forecast proceeds such that track error is reduced more than 100 km on average in the day-5 forecast. The change in track error when the BMCI retrievals are assimilated is very small, less than 25 km, which is comparable to the model grid size. For instance, FP+Both shows increases in the track error (less than 25 km) during the first 3 days of the forecasts, then after 3 days, it starts showing some small improvements compared with the control experiment. The experiment with thinned observations (FP+Both-Thin and FP+ATMS-Thin) perform better than the experiment FP+Both in the forecasts for up to 3 days. This is probably due to the fact that horizontal and especially more vigorous vertical thinning reduced correlated errors, which shows that better tuning of the errors may increase the impact of the BMCI retrievals on the track and intensity of tropical cyclones. Figure 9 also shows the mean intensity error for all forecasts between 17 and 30 September 2017. Again, assimilating satellite observations (FP-Ctrl) only slightly reduces the intensity error in the forecast, about 5–10 hPa. In terms of intensity, the experiments only slightly differ from the control experiment in the first 3 days (72 h) of the forecasts. After 3 days, a large variance is observed among the forecasts, and the FP-Ctrl shows a smaller SLP bias when compared with the TCVitals, nevertheless the difference between the experiments is less than 10 hPa. The model coarse resolution (almost 25 km) may play a role in only seeing some small improvements in intensity forecast when BMCI retrievals are assimilated.

Other important feature of tropical cyclones are the high magnitude wind speed and the warm core anomaly near the center of cyclone that occurs because of the release of latent heat from the water vapor lifted by the convection. The warm core can be detected using the difference between observed temperature and a reference temperature representing the environmental temperature in the absence of the cyclone (Haurwitz 1935; Durden 2013). We averaged the data over a 10° band on each side of cyclone to calculate the reference environmental temperature. The zonal wind cross sections along with the corresponding cross sections of temperature anomaly
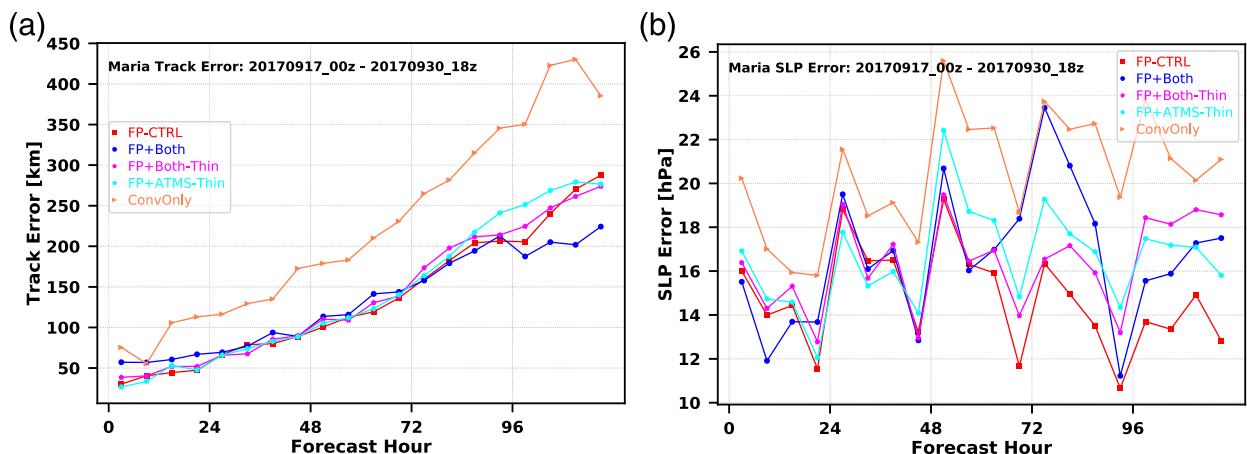

FIG. 9. Mean (a) track and (b) intensity error for Hurricane Maria for all 5-day forecasts started at 0000 UTC between 17 and 30 Sep 2017.
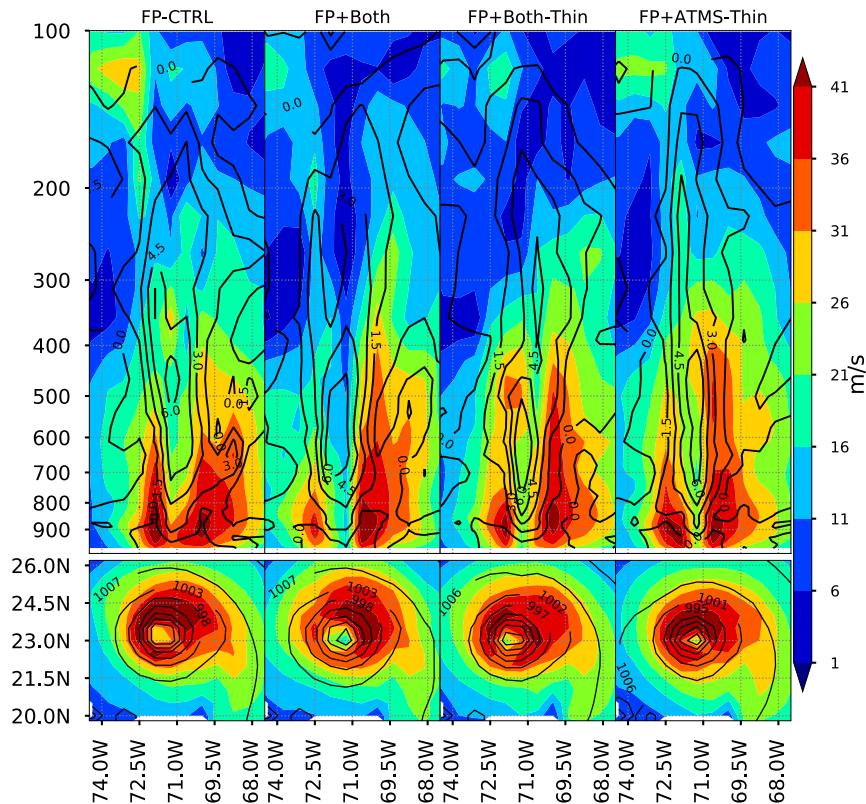
FIG. 10. Vertical cross section of (top) wind magnitude (m s^{-1}; shaded) and temperature anomaly (K) as well as (bottom) 850-hPa wind speed (color shaded) and sea surface pressure (hPa; contours). These are for the 2100 UTC 22 Sep 2017 cycle.

and horizontal wind magnitude at 850 hPa for Hurricane Maria are shown in Fig. 10. Comparing the experiments where the BMCI retrievals are assimilated with the control experiment shows that the storm is intensified in terms of both wind speed and temperature anomaly when the BMCI retrievals are assimilated.

In Fig. 10, the control experiment shows a shallow warm core, while all other experiments show a much deeper warm core. Overall, a warmer anomaly is seen in the FP+ATMS-Thin with anomalies of 4.5 K reaching upper levels. In order for model to generate a stronger storm, the warm core anomaly has to reach the higher levels (200–300 hPa) and the experiment FP+ATMS-Thin performs better than other experiments from that perspective. Although, the eye in FP+Both is "clear," however, the eye is too wide to represent a category-5 hurricane, while in the FP+ATMS-Thin a smaller and a more symmetric eye is seen (see lower panels in Fig. 10). The narrow eye is also consistent with the storm seen in the visible images from geostationary satellites for Hurricane Maria. In a strong tropical cyclone, winds have to be symmetric and the wind columns need to be vertically aligned, which is the case for FP+ATMS-Thin. The west side of the storm in FP+Both and FP+Both-Thin is weaker than in FP+ATMS-Thin, not representing a well formed category-5 hurricane. The improvement in the forecast skills for track and intensity

depends on how model can handle the improvement in the initial conditions. Both experiments with horizontal thinning show some improvement compared with the experiment when no thinning was done. This can be due to reducing either vertically or horizontally correlated error for temperature and water vapor profiles.

5. Summary and conclusions

Satellite observations from microwave and infrared instrument are the largest source of observations for the NWP data assimilation systems. Infrared observations are very sensitive to clouds so that clouds are largely opaque in the infrared spectrum in the presence of convection (e.g., in the case of hurricanes' rainbands), but microwave observations tend to provide useful information on the state of the atmosphere even in the rainbands of tropical cyclones. Although direct assimilation of satellite observations in clear-sky conditions is now routinely performed at the NWP centers owing to advances in developing fast radiative transfer models, the direct assimilation of all-sky microwave observations is limited to mostly shallow clouds. The main limiting factors for direct assimilation of cloud contaminated observations include (i) NWP models tend not to provide a close first guess for the radiative transfer calculations, (ii) oversimplification of radiative

transfer models such as assuming spherical particles for the clouds, and (iii) lack of inputs such as particles size and shape and their distribution required by radiative transfer models to perform accurate scattering calculations. Previous efforts to retrieve atmospheric state variables from satellite observations, and then assimilate the retrieved profiles have also focused on using optimal estimation techniques, which encounter the same issue as variational data assimilation techniques.

We have introduced a novel Bayesian Monte Carlo integration technique that is capable of retrieving geophysical variables such as temperature, water vapor, cloud liquid and ice water content as well as sea surface skin temperature (SST) and wind speed from passive microwave observations even in the presence of deep-convective clouds. The main advantages of the BMCI technique over optimal estimation techniques is that it does not depend on the first guess provided as background information and also no minimization between simulated and real observations is performed during the retrieval process. The BMCI technique can be summarized in three steps: (i) generating a stochastic database using information derived from ERA-Interim reanalysis, several spaceborne active radar measurements, and cloud probe information from several campaigns, (ii) simulating satellite observations for selected instruments and including them in a retrieval database, and (iii) retrieving atmospheric and surface information such as profiles of temperature, relative humidity and cloud liquid and ice water content as well as sea surface temperature using the BMCI technique and real observations. The main limitations of the BMCI when compared with direct assimilation of all-sky radiances include (i) the need to perform a separate step to retrieve the products and then assimilate them into the system and (ii) the fact that the retrieval database generation must be changed if the method is to be used in other regions of the world.

The method was originally developed by Evans et al. (2012) to retrieve relative humidity and ice particle parameters from an airborne submillimeter radiometer. This work extends the method to also retrieve temperature profiles, rain parameters, as well as SST and wind speed in tropical cyclones environments from spaceborne passive microwave observations. In addition to including new geophysical variables to retrieved parameters, other major improvements over the method developed by Evans et al. (2012) include (i) adding a sea surface emissivity model (FASTEM-6) and its adjoint to be able to retrieve surface parameters, (ii) using in situ cloud and rain microphysical data from different hurricane field campaigns to generate the prior pdf statistics for warm cloud droplets and raindrops in tropical cyclones, (iii) modifying the prior PDF generation to use GPM DPR reflectivity profiles when *CloudSat* reflectivity is too attenuated, and (iv) implementing a beam-filling bias correction technique to account for the horizontal variability of hydrometers within the large field-of-views of microwave radiometers.

A beam-filling correction technique was developed to statistically account for the radiative transfer effects due to horizontal variability of realistic rain and ice over large passive microwave footprints. First, the 1D stochastic atmosphere

profile generation algorithm is used to generate the central column of the footprint from the CDF-EOF prior pdf. The central column ice and water path and spatial statistics of DPR retrieved rain and ice water path are then used to generate stochastic rain and ice water path for the rest of the columns (each 5 km wide) inside the footprint. All other variables defining the vertical structure of each column (temperature, water vapor, and rain, ice, and liquid cloud parameters) are generated using the 1D generation algorithm combined with the stochastic rain and ice water paths of the column. The radiative transfer calculations are performed individually for each column, and then the column brightness temperatures are averaged by weighting by a Gaussian footprint pattern. The statistics of the differences between the variable and uniform footprint brightness temperatures are used to account for the beam-filling bias and to increase the assumed uncertainties in the observed brightness temperatures.

We have discussed the development of the BMCI retrieval package and the assimilation of retrieved profiles into a state-of-the-art NWP model. We explain how the errors associated with the retrievals can be used to improve the assimilation of such profiles into NASA's GEOS forecast system. The main purpose of this paper was to discuss the improved BMCI retrieval method and demonstrate its utility in a modern data assimilation system and in the context of the full global observing system. Analyzing the uncertainty values for the temperature and humidity profiles revealed a significant difference between the error correlations for the RH profiles for clear and cloudy conditions. So that, thinning the water vapor profiles in the height dimension based on the error correlation matrix significantly improved the impact of assimilating water vapor profiles into GEOS. We performed several experiments with and without the BMCI retrievals including thinning the retrieved profiles at different vertical and horizontal resolutions. Observations from ATMS and GMI were used to evaluate the impact of the retrieved profiles on the horizontal and vertical structure as well as the track and central pressure of Hurricane Maria in GEOS analyses and forecasts. The experiments with the BMCI profiles assimilated generally showed a much stronger storm in terms of both wind speed and storm warm core. However, the impact on the track and central pressure was minimal. Possible factors that may limit the impact of the assimilation of the BMCI retrievals include, the resolution of the model (about 25 km), which was too coarse to show the potential of the BMCI retrievals in improving the representation of tropical cyclones in the model forecast and also the data assimilation system not being able to handle the correlated observation errors. We are currently conducting data assimilation experiments for several other hurricanes to evaluate the impact of assimilating such observations for different tropical cyclones.

anonymous reviewers for their comments, which helped to substantially improve the original paper.

REFERENCES

Aksoy, A., S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 vortex-scale analyses. *Mon. Wea. Rev.*, **141**, 1842–1865, https://doi.org/10.1175/MWR-D-12-00194.1.

Baordo, F., and A. Geer, 2015: All-sky assimilation of SSMIS humidity sounding channels over land within the ECMWF system. EUMETSAT/ECMWF Fellowship Programme Research Rep. 38, 28 pp.

Bauer, P., P. Lopez, A. Benedetti, D. Salmond, and E. Moreau, 2006: Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. *Quart. J. Roy. Meteor. Soc.*, **132**, 2277–2306, https://doi.org/10.1256/qj.05.189.

——, A. J. Geer, P. Lopez, and D. Salmond, 2010: Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Quart. J. Roy. Meteor. Soc.*, **136**, 1868–1885, https://doi.org/10.1002/qj.659.

Caumont, O., V. Ducrocq, E. Wattrelot, G. Jaubert, and S. Pradier-Vabre, 2010: 1D+3DVar assimilation of radar reflectivity data: A proof of concept. *Tellus*, **62A**, 173–187, https://doi.org/10.1111/j.1600-0870.2009.00430.x.

Clough, S., M. Shephard, E. Mlawer, J. Delamere, M. Iacono, K. Cady-Pereira, S. Boukabara, and P. Brown, 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer*, **91**, 233–244, https://doi.org/10.1016/j.jqsrt.2004.05.058.

Draper, D. W., D. A. Newell, F. J. Wentz, S. Krimchansky, and G. M. Skofronick-Jackson, 2015: The Global Precipitation Measurement (GPM) Microwave Imager (GMI): Instrument overview and early on-orbit performance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **8**, 3452–3462, https://doi.org/10.1109/JSTARS.2015.2403303.

Duncan, D. I., C. D. Kummerow, B. Dolan, and V. Petković, 2018: Towards variational retrieval of warm rain from passive microwave observations. *Atmos. Meas. Tech.*, **11**, 4389–4411, https://doi.org/10.5194/amt-11-4389-2018.

Durden, S. L., 2013: Observed tropical cyclone eye thermal anomaly profiles extending above 300 hPa. *Mon. Wea. Rev.*, **141**, 4256–4268, https://doi.org/10.1175/MWR-D-13-00021.1.

Duruisseau, F., P. Chambon, E. Wattrelot, M. Barreyat, and J.-F. Mahfouf, 2019: Assimilating cloudy and rainy microwave observations from SAPHIR on board Megha Tropiques within the ARPEGE global model. *Quart. J. Roy. Meteor. Soc.*, **145**, 620–641, https://doi.org/10.1002/qj.3456.

Ekelund, R., and P. Eriksson, 2019: Impact of ice aggregate parameters on microwave and sub-millimetre scattering properties. *J. Quant. Spectrosc. Radiat. Transfer*, **224**, 233–246, https://doi.org/10.1016/j.jqsrt.2018.11.013.

Elsaesser, G. S., and C. D. Kummerow, 2015: The sensitivity of rainfall estimation to error assumptions in a bayesian passive microwave retrieval algorithm. *J. Appl. Meteor. Climatol.*, **54**, 408–422, https://doi.org/10.1175/JAMC-D-14-0105.1.

Errico, R. M., P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.*, **64**, 3785–3798, https://doi.org/10.1175/2006JAS2044.1.

Evans, K. F., 2007: SHDOMPPDA: A radiative transfer model for cloudy sky data assimilation. *J. Atmos. Sci.*, **64**, 3854–3864, https://doi.org/10.1175/2006JAS2047.1.

——, S. J. Walter, A. J. Heymsfield, and G. M. McFarquhar, 2002: Submillimeter-wave cloud ice radiometer: Simulations of retrieval algorithm performance. *J. Geophys. Res.*, **107**, 4028, https://doi.org/10.1029/2001JD000709.

——, and Coauthors, 2012: Ice hydrometeor profile retrieval algorithm for high-frequency microwave radiometers: Application to the CoSSIR instrument during TC4. *Atmos. Meas. Tech.*, **5**, 2277–2306, https://doi.org/10.5194/amt-5-2277-2012.

Geer, A. J., and P. Bauer, 2010: Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation. ECMWF Tech. Memo. 620, 41 pp.

——, and F. Baordo, 2014: Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmos. Meas. Tech.*, **7**, 1839–1860, https://doi.org/10.5194/amt-7-1839-2014.

——, ——, N. Bormann, and S. J. English, 2014: All-sky assimilation of microwave humidity sounders. ECMWF Tech. Memo. 741, 57 pp.

——, and Coauthors, 2017: The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quart. J. Roy. Meteor. Soc.*, **143**, 3189–3206, https://doi.org/10.1002/qj.3172.

——, and Coauthors, 2018: All-sky satellite data assimilation at operational weather forecasting centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1191–1217, https://doi.org/10.1002/qj.3202.

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

Haurwitz, B., 1935: The height of tropical cyclones and of the "eye" of the storm. *Mon. Wea. Rev.*, **63**, 45–49, https://doi.org/10.1175/1520-0493(1935)63<45:THOTCA>2.0.CO;2.

Janisková, M., P. Lopez, and P. Bauer, 2012: Experimental 1D + 4D-Var assimilation of *CloudSat* observations. *Quart. J. Roy. Meteor. Soc.*, **138**, 1196–1220, https://doi.org/10.1002/qj.988.

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

Kim, E., C.-H. J. Lyu, K. Anderson, R. Vincent Leslie, and W. J. Blackwell, 2014: S-NPP ATMS instrument prelaunch and on-orbit performance evaluation. *J. Geophys. Res. Atmos.*, **119**, 5653–5670, https://doi.org/10.1002/2013JD020483.

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

Kummerow, C., 1998: Beamfilling errors in passive microwave rainfall retrievals. *J. Appl. Meteor.*, **37**, 356–370, https://doi.org/10.1175/1520-0450(1998)037<0356:BEIPMR>2.0.CO;2.

——, and Coauthors, 2001: The evolution of the Goddard Profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801–1820, https://doi.org/10.1175/1520-0450(2001)040<1801:TEOTGP>2.0.CO;2.

——, D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015: The evolution of the

Goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280, https://doi.org/10.1175/JTECH-D-15-0039.1.

Liu, Q., F. Weng, and S. J. English, 2011: An improved fast microwave water emissivity model. *IEEE Trans. Geosci. Remote Sens.*, **49**, 1238–1250, https://doi.org/10.1109/TGRS.2010.2064779.

Migliorini, S., and B. Candy, 2019: All-sky satellite data assimilation of microwave temperature sounding channels at the Met Office. *Quart. J. Roy. Meteor. Soc.*, **145**, 867–883, https://doi.org/10.1002/qj.3470.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, https://doi.org/10.1016/j.jcp.2007.07.022.

Rydberg, B., P. Eriksson, S. A. Buehler, and D. P. Murtagh, 2009: Non-Gaussian Bayesian retrieval of tropical upper tropospheric cloud ice and water vapour from Odin-SMR measurements. *Atmos. Meas. Tech.*, **2**, 621–637, https://doi.org/10.5194/amt-2-621-2009.

Todling, R., and A. El Akkraoui, 2018: The GMAO hybrid ensemble-variational atmospheric data assimilation system: Version 2.0. NASA Goddard Space Flight Center,Tech. Rep. NASA/TM-2018-104606/VOL50, GSFC-E-DAA-TN54363, 184 pp., https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20180002172.pdf.

Toon, O. B., and Coauthors, 2010: Planning, implementation, and first results of the tropical composition, cloud and climate coupling experiment (TC4). *J. Geophys. Res.*, **115**, D00J04, https://doi.org/10.1029/2009JD013073.

Trahan, S., and L. Sparling, 2012: An analysis of NCEP tropical cyclone vitals and potential effects on forecasting models. *Wea. Forecasting*, **27**, 744–756, https://doi.org/10.1175/WAF-D-11-00063.1.

Wattrelot, E., O. Caumont, and J.-F. Mahfouf, 2014: Operational implementation of the 1D+3D-Var assimilation method of radar reflectivity data in the AROME model. *Mon. Wea. Rev.*, **142**, 1852–1873, https://doi.org/10.1175/MWR-D-13-00230.1.

Weng, F., T. Zhu, and B. Yan, 2007: Satellite data assimilation in numerical weather prediction models. Part II: Uses of rain-affected radiances from microwave observations for hurricane vortex analysis. *J. Atmos. Sci.*, **64**, 3910–3925, https://doi.org/10.1175/2006JAS2051.1.

Zhu, Y., and Coauthors, 2016: All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Mon. Wea. Rev.*, **144**, 4709–4735, https://doi.org/10.1175/MWR-D-15-0445.1.

Zou, X., F. Weng, B. Zhang, L. Lin, Z. Qin, and V. Tallapragada, 2013: Impacts of assimilation of ATMS data in HWRF on track and intensity forecasts of 2012 four landfall hurricanes. *J. Geophys. Res. Atmos.*, **118**, 11 558–11 576, https://doi.org/10.1002/2013JD020405.