JGR Atmospheres

RESEARCH ARTICLE
10.1029/2021JD036061

An Objective Quality Control of Surface Contamination Observations for ABI Water Vapor Radiance Assimilation

Zhenglong Li[1], Zheng Ma[1,2,3], Pei Wang[1], Agnes H. N. Lim[1], Jun Li[4], James A. Jung[1], Timothy J. Schmit[5], and Hung-Lung Huang[1]

[1]Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, WI, USA, [2]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [3]University of Chinese Academy of Sciences, Beijing, China, [4]National Satellite Meteorological Center, China Meteorological Administration, Beijing, China, [5]Center for Satellite Applications and Research, NESDIS/NOAA, Madison, WI, USA

Key Points:

- Optimized quality control (QC) scheme minimizes the detrimental impact from surface contaminated infrared radiance assimilation
- A new machine learning-based QC scheme for surface contamination is developed for Advanced Baseline Imager (ABI) water vapor bands
- The new QC scheme improves the Hurricane Harvey (2017) forecast over existing QC schemes with added value from each ABI water vapor band

Correspondence to:
Z. Li and J. Li,
Zhenglong.Li@ssec.wisc.edu;
junli@cma.gov.cn

Citation:
Li, Z., Ma, Z., Wang, P., Lim, A. H. N., Li, J., Jung, J. A., et al. (2022). An objective quality control of surface contamination observations for ABI water vapor radiance assimilation. *Journal of Geophysical Research: Atmospheres*, *127*, e2021JD036061. https://doi.org/10.1029/2021JD036061

Received 19 OCT 2021
Accepted 5 JUL 2022

Abstract A quality control (QC) process which handles surface impacts is an important step toward successful assimilation of the Advanced Baseline Imager (ABI) water vapor (WV) band radiances. If the QC is too relaxed, many surface contaminated radiances get assimilated. If the QC is too stringent, useful radiances are rejected. Either way can result in reduced or even compromised observation impacts. A new machine learning-based QC scheme for the three ABI WV bands is developed and optimized to help understand the importance and effectiveness of the scheme. Unlike previous schemes which are dependent on the background, this scheme extracts and blends the surface information from 7 ABI bands (bands 8–10, 13–16) to determine if a WV radiance is affected by the surface. Simulation studies show that the new QC scheme is effective in retaining radiances that are either unaffected by the surface or have very small surface contamination. It is highly effective in rejecting radiances with large surface contamination. Numerical experiments from a single case study of Hurricane Harvey (2017) were carried out to optimize the QC and to understand the potential impacts on forecasts. The use of the new QC scheme shows that radiances from each WV band have substantially added value. Combining them has a positive impact on hurricane track forecasts compared with existing QC schemes. Hence, it is critical that an optimized QC scheme is used for infrared WV radiance assimilation. It provides a balance between positive impacts from useful radiances and negative impacts from surface contaminated radiances.

Plain Language Summary Surface affected (contaminated) infrared water vapor radiances are usually not assimilated into numerical weather prediction models, or are assimilated with reduced observation weights due to negative impacts on the analysis and forecast. The existing quality control schemes to find such radiances are not objective because they rely on background field. The quality control affects both the quantity (the number of observations) and the quality (the cleanness of the observations) of the radiances for assimilation. An objective quality control provides optimized balance between the quantity and quality to maximize the observation impacts. A machine learning-based quality control scheme is developed and used to understand the importance and effectiveness of optimized quality control. It is objective because it relies only on observations. Simulation studies show that stringent quality control improves the quality but reduces the quantity. Numerical experiments for Hurricane Harvey (2017) were carried out to find the optimized quality control. With the optimized quality control, radiances from each Advanced Baseline Imager water vapor band show added value on the forecast. Combining them shows improved forecast compared with existing schemes. It is therefore important to use optimized quality control for infrared water vapor radiance assimilation.

1. Introduction

Satellite radiance observations have been a dominant source of information used in Numerical Weather Prediction (NWP) models (Cardinali, 2009). The infrared (IR) water vapor (WV) bands from advanced imagers onboard the domestic and international constellation of geostationary satellites provide observations with near-global coverage of the tropics and mid-latitudes (Li et al., 2019). The radiances provide thermodynamic information with high temporal and spatial resolutions. Kazumori (2016) showed significantly improved moisture fields in analysis from the Advanced Himawari Imager (AHI) WV bands onboard Himawari-8 in the Japan Meteorological Agency (JMA) NWP system. Ma et al. (2017) reported neutral to marginal positive impacts on analyses and forecasts from AHI WV radiances assimilated in the National Centers for Environmental Prediction (NCEP) Global Data

Assimilation System Global Forecast System. Liu et al. (2019b) assimilated Advanced Baseline Imager (ABI) GOES-16 clear-sky WV radiances at NCEP and showed neutral to slightly positive impacts on forecast skills. Yin et al. (2021) emphasized that the thermodynamic information from the high temporal and spatial resolution from Geostationary Interferometric Infrared Sounder is critical for high-impact weather forecasts. Despite the use of these observations in NWP, WV radiance observations from geostationary satellites have smaller forecast error reductions compared to observations from polar-orbiting satellites (Cardinali, 2009). Further improvement in the usage of geostationary satellite WV radiances in NWP models is needed.

As increasing efforts are being made to improve the geostationary satellite WV radiances' impact in NWP models, it becomes apparent that the data quality control (QC) scheme plays a critical role in the success of the data assimilation and subsequent forecasts. For WV radiance assimilation, the QC scheme typically focuses on minimizing contamination from the surface and clouds. Contamination means that the radiance is affected by something other than the atmospheric temperature and moisture. For example, surface contamination means the radiance is affected by the surface related parameters such as surface skin temperature, emissivity and terrain height (Di et al., 2016). Radiances that are surface and/or cloud contaminated are either not assimilated or assimilated with lower weights. This study will focus on quality controlling observations that are contaminated by the surface. A relaxed QC can introduce many surface contaminated radiances into the assimilation system, whereas a stringent QC can prevent useful radiances from being assimilated. Neither is ideal as it may lead to reduced or even compromised observation impacts. The key is to find the optimized QC scheme that best balances the positive impact from useful radiances and the negative impact from contaminated radiances.

In the QC scheme used by the Gridpoint statistical interpolation (GSI), the surface skin temperature Jacobians used to determine if a WV radiance is affected by the surface (referred to as the GSI QC scheme) are calculated from the background fields. If the radiance is deemed to be surface contaminated, the observation error is inflated to account for surface emissivity and surface skin temperature uncertainty. This strategy works well for the ocean as the sea surface emissivity model is relatively accurate and the sea surface temperature is relatively stable. Over land, there are larger uncertainties as the land surface emissivity has much larger variability spatially, temporally, angularly, and spectrally (Li et al., 2010, 2011, and 2012). As a result, the ABI/GOES-16 WV radiances with terrain heights above 1 km are not assimilated operationally in the Finite-Volume Cubed-Sphere dynamical core Global Forecast System (FV3GFS). NCEP showed that the assimilation of ABI WV clear sky radiance in FV3GFS had neutral (500 hPa anomaly correlation) to slightly positive impacts (fit to selected observations) on the forecast (Liu et al., 2019a, 2019b).

Lee et al. (2019) demonstrated that the GSI QC scheme for ABI WV radiances is not adequate over land. The thresholds applied are not strict enough to identify surface contaminated observations, and a small percentage of surface contaminated observations still passed the GSI QC check. This may reduce or compromise the usefulness of the observations. The authors showed that using surface contaminated observations even with reduced weights is detrimental to the analysis. A better option is not to use such observations in the assimilation system. With that, the inverse problem is greatly simplified with only 2 types of unknowns (temperature and moisture) instead of 4 (temperature, moisture, surface skin temperature, and surface emissivity). An improved QC scheme that also uses the surface skin temperature Jacobians to examine the surface sensitivity, but with much stricter thresholds, was developed (referred hereto as the Lee QC scheme) and tested for hurricane forecasts. Despite being more stringent than the GSI QC scheme, the number of radiances assimilated is comparable. Furthermore, the Lee QC scheme showed improved hurricane track and intensity forecasts.

Both the GSI QC and the Lee QC schemes are not objective as they depend on the background fields, especially moisture. It is known that the NWP moisture field has larger uncertainty than the temperature field (Carminati et al., 2019). QC based on the background could be problematic in areas where there are large wet or dry biases. These regions are where data assimilation improvements are needed the most. Surface-sensitive WV radiances over regions where NWP moisture fields have large wet biases may be misclassified, resulting in surface contamination. Conversely, surface insensitive radiances over regions with large dry biases may be misclassified, resulting in them being rejected by the assimilation system. In addition, unwanted uncertainties may be introduced from the radiative transfer models to calculate the Jacobians in both the GSI and Lee QC schemes.

ABI is a 16-band radiometer onboard the new generation of GOES (GOES-R series) satellites, covering the visible (Bands 1–3), near IR (Bands 4–6) and IR (Bands 7–16) regions. The ABI imagery is widely used in

Table 1
The Center Wavelength, Weighting Function Peak Height, and the Nickname (For Typical Use) of the ABI 10 IR Bands

ABI band	Central wavelength length (μm)	Weighting function peak height		Nickname
		Temperature	Moisture	
7	3.9	Surface	853 hPa	Shortwave window
8	6.19	367 hPa	351 hpa	Upper-level moisture
9	6.95	451 hPa	399 hPa	Midlevel moisture
10	7.34	650 hPa	545 hPa	Lower-level moisture
11	8.5	Surface	815 hPa	Cloud top phase
12	9.61	Surface	866 hPa	Ozone
13	10.35	Surface	853 hPa	Clean longwave window
14	11.2	Surface	840 hPa	Longwave window
15	12.3	Surface	840 hPa	Dirty longwave window
16	13.3	Surface	815 hPa	CO_2 longwave

Note. The 1976 US standard atmospheric profiles are used to calculate the weighting functions.

monitoring weather, especially for the severe weather systems, such as hurricanes, local severe storms, natural hazards and etc. (Schmit et al., 2005, 2017). Table 1 shows the weighting function peak heights of the 10 IR bands. For 1976 US standard atmosphere, the three WV bands peak high above the surface. However, when the atmospheric moisture content is reduced, the WV band could be affected by the surface. Figure 1 shows temperature and moisture weighting functions of the three ABI WV bands for tropics, midlatitude winter, and subarctic winter. It is obvious that the peak of the weighting functions decreases with reduced moisture content. As a result,

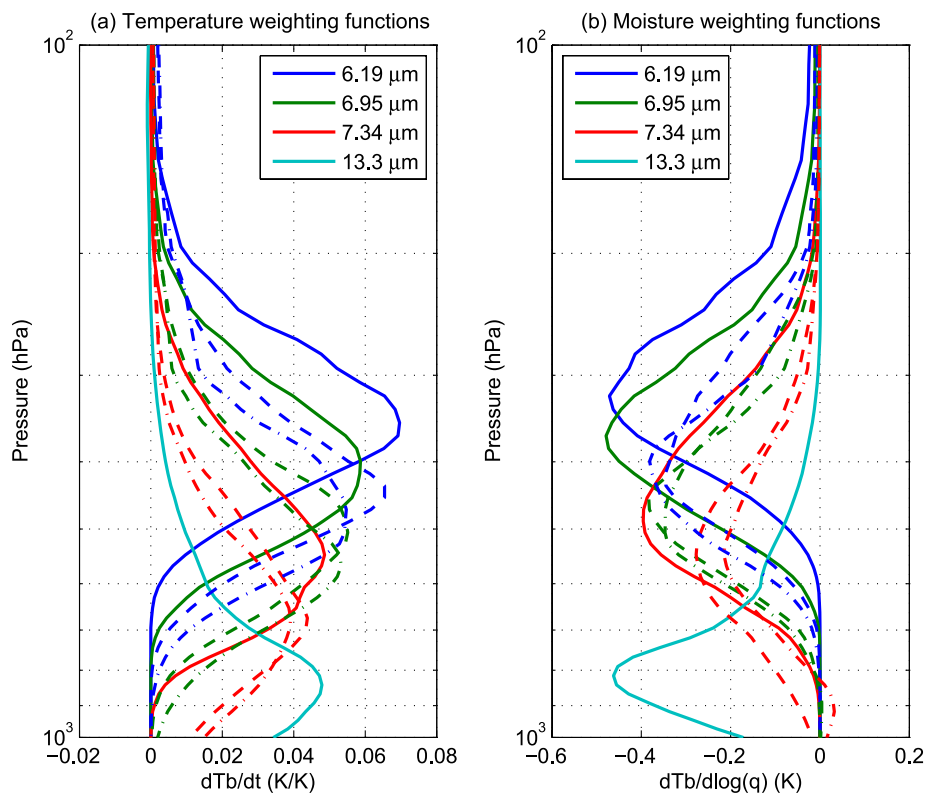

Figure 1. (a) The temperature and (b) water vapor weighting functions (Jacobian functions) of selected Advanced Baseline Imager (ABI) bands. The solid lines represent typical tropical profiles, dashed lines represent typical midlatitude winter profiles, and the dash-dotted lines represent typical subarctic winter profiles.

band 10 is affected by the surface for midlatitude winter and subarctic winter. Band 9 is also affected for subarctic winter. In the real world, atmospheric moisture content fluctuates with time. Even in tropics, there are regions with low moisture content. In addition, the information content of water vapor from water vapor absorption bands might be affected by terrain height (Di et al., 2016), and substantially affected by the contrast between surface air temperature and surface skin temperature (Li et al., 1994). Radiances, affected by the surface, should be identified and not assimilated. It is therefore necessary to develop an objective QC scheme to determine if an ABI WV radiance is affected by surface or not, based only on observations.

This study presents a machine learning-based QC scheme for ABI WV radiance assimilation that relies only on radiance observations. The new QC scheme determines if an ABI WV radiance is affected by the surface. Only radiances passing the QC will be made available for assimilation. Assimilation experiments are carried out to optimize the QC for each ABI WV band. The results will help better understand the importance and effectiveness of the optimized QC for improving hurricane forecasts. Specifically, three questions will be addressed through the impact studies: (a) how is the new QC optimized, (b) does each of the ABI three WV bands provide positive impacts, and (c) how does the new QC compare with existing schemes.

This paper is organized as follows. Section 2 introduces the machine learning technique with simulation studies. Section 3 presents the results from the application to real data. Section 4 shows an impact study of assimilating observations from the new QC for Hurricane Harvey (2017). A summary and discussion follow in Section 5.

2. The Machine Learning-Based QC

Observed radiances from IR WV bands have three components: the atmospheric upwelling component $R^{\uparrow}$, the surface emission component R_s, and the atmospheric downwelling component reflected by the surface $R^{\downarrow\uparrow}$, as shown in Equation 1

$$R_o = R^{\uparrow} + R_s + R^{\downarrow\uparrow} \tag{1}$$

All three components are non-negative. The surface emission and the surface-reflected downwelling components are attenuated by the atmosphere before reaching the satellite. The surface contamination component R_{sc} is the sum of the surface emission and surface-reflected downwelling components and can be calculated using Equation 2

$$R_{sc} = R_s + R^{\downarrow\uparrow} = R_o - R^{\uparrow} \tag{2}$$

If the surface emission component does not reach the satellite, neither does the surface-reflected downwelling component. In this case, there is no surface contamination, and the observed radiance is the same as the upwelling component. Otherwise, the observed radiance is surface contaminated and is larger than the upwelling component. So mathematically, surface contamination happens when both the surface emission and the surface-reflected downwelling components are greater than 0.

2.1. The Multi-Layer Perceptron Network-Based QC Scheme

There are two ways to study the surface contamination problem using machine learning: regression or classification. For regression, the basic assumption is that the atmospheric upwelling component can be predicted from observed radiances of the 7 ABI IR bands, which are bands 8–16 but excluding band 11 (8.4 μm, due to complication caused by surface emissivity) and band 12 (9.6 μm, due to sensitivity to ozone). The ABI IR radiance observations (Schmit et al., 2005, 2017) contain important thermodynamic information about the atmosphere and surface. Some bands are more sensitive to the atmosphere, while others are more sensitive to the surface. There are signal correlations between different spectral bands, especially in the radiance domain. These signal correlations make it possible to predict the upwelling component of each ABI WV band from the IR radiance observations. Note that the error correlations between different spectral ABI bands are small and neglectable because each band is independent. Equation 2 can be rewritten as

$$R_{sc} = R_o - R_p^{\uparrow} \tag{3}$$

where $R_p^{\uparrow}$ is the predicted atmospheric upwelling radiance. Surface contamination is present if the predicted upwelling radiance is smaller than the ABI WV radiance observation, which then should not be used for assimilation.

A simulation study was carried out to demonstrate the above concept. Clear sky ABI WV radiances were simulated using the Community Radiative Transfer Model V2.2.3 with Optical Depth in Pressure Space coefficients (Chen et al., 2012) with the SeeBor database Version 5.0 (Seemann et al., 2003, 2008). The database contains 15,704 global clear sky atmospheric profiles of temperature, moisture, and ozone, as well as surface temperature and emissivities. Two types of radiances are simulated: the synthetic ABI radiance observations from the seven IR bands and the upwelling radiance components of the 3 WV bands (6.2, 6.9, and 7.3 μm). The simulation was carried out at 11 different local zenith angles (LZAs): 0°, 24.62°, 33.56°, 39.72°, 44.42°, 48.19°, 51.32°, 53.97°, 56.25°, 58.24°, and 60.00°. These 11 LZAs ensure that all angles have similar weights in the secant space between 0° and 60° (Li et al., 2020). LZAs larger than 60.00° are not included because those observations will not be assimilated in GSI (Liu et al., 2019a, 2019b). Random Gaussian noise based on the sensor's pre-launch noise-equivalent delta radiance (NEdR) requirements is added. The pre-launch NEdR is used instead of the on-orbit internal calibration target NEdR to allow for additional uncertainties in the radiative transfer modeling. In the end, a simulation data set with 172,744 samples is prepared. Each sample has synthetic ABI radiance observations from the seven IR bands and the atmospheric upwelling radiances of the 3 WV bands. These radiances will be used to develop the machine learning-based QC schemes.

The multi-layer perceptron (MLP) network is adopted to develop the regression-like prediction relationship. The MLP QC scheme is based on the comparison of the observed radiance with the predicted upwelling component. MLP is a supervised machine learning algorithm consisting of one input layer, one output layer, and one or more hidden layers, with nodes arranged on different layers receiving inputs from the preceding layer and passing outputs only to the next layer (Hinton, 1990; Murtagh, 1991). Through nonlinear activation functions and back-propagation, MLP can learn a nonlinear function approximator given a set of features (inputs) and target values (outputs) (Baum, 1988; Blackwell & Chen, 2009; Glorot & Bengio, 2010). In this study, a five-layer MLP was developed with seven nodes in the input layer. These represent the features of radiances from the seven ABI bands (8–10 and 13–16). Three nodes in the output layer represent the relevant atmospheric upwelling radiances from bands 8–10 to be predicted from the regression. The neuron number of the three hidden layers are 400, 200, and 100, and the activation function and loss function are Rectified Linear Unit (RELU, Nair & Hinton, 2010) and mean squared error (MSE), respectively. These configurations were determined after comparative experiments using two and four hidden layers, more and fewer hidden nodes, and with activation functions of Sigmoid and Logistic. The comparisons take into account both the accuracy and computational efficiency. Adam, the stochastic gradient-based optimizer (Kingma & Ba, 2017), was used in the framework to optimize the weights on different nodes during the training process. Also, an L2 regularization term was added to the weights to limit the chance of overfitting by penalizing weights with large magnitudes (Girosi et al., 1995; Krogh & Hertz, 1992). The final loss function that is optimized in the training process can be described by Equation 4 as

$$L = L_0 + \frac{\lambda}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} W_{i,j}^2 \tag{4}$$

where L_0 stands for the original MSE loss function calculated using the predicted upwelling radiances and their corresponding truth values from simulations. The second term represents the regularization term with m and n representing the number of layers and the dimension of weights on each layer respectively. $W_{i,j}$ stands for the weight on the jth neuron of the ith layer. λ is the penalty term deciding how much to penalize the weights and is adopted as 0.01 in this study.

The simulation data set was first separated randomly into a training set consisting of 80% of all samples, and an independent validation set consisting of the remaining 20%. The samples in the training set were shuffled and scaled before being split again into a 90% training subset that was fed into the MLP model, and a 10% testing subset used for monitoring the model performance during the training process. The early stopping method was applied during the training to avoid overfitting by automatically terminating the training when the model's performance on the testing subset did not show an improvement over 200 epochs.

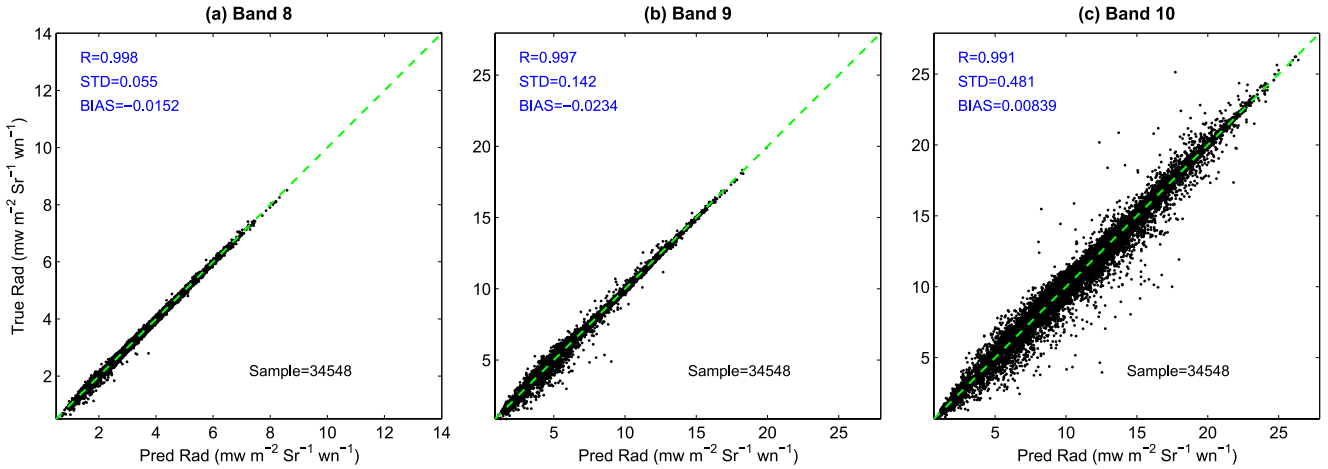

Figure 2. The validation of the predicted atmospheric upwelling radiances (mw · m^{-2} · Sr^{-1} · wn^{-1}) for Advanced Baseline Imager (ABI) (a) band 8, (b) band 9, and (c) band 10 using the multi-layer perceptron network. The SeeBor database is used to simulate synthetic ABI observed radiances as input and the upwelling radiances of the 3 Water Vapor (WV) bands as output. 80% of the simulated data is randomly selected for training while the remaining 20% is for validation.

After training, the MLP model was validated using the independent validation set. Figure 2 shows that ABI WV upwelling radiances can be predicted from the 7-band synthetic ABI IR radiance observations. For all three bands, the prediction shows close-to-zero biases. Band 8 is best predicted with a standard deviation (STD) of the differences of 0.055 mw · m^{-2} · Sr^{-1} · wn^{-1}, on par with the NEdR of 0.056 mw · m^{-2} · Sr^{-1} · wn^{-1}. Bands 9 and 10 are more difficult to predict. The STD of band 9 is 0.142 mw · m^{-2} · Sr^{-1} · wn^{-1}, which is substantially larger than the NEdR of 0.082 mw · m^{-2} · Sr^{-1} · wn^{-1}. And it is 0.481 mw · m^{-2} · Sr^{-1} · wn^{-1} for band 10, significantly larger than its NEdR of 0.095 mw · m^{-2} · Sr^{-1} · wn^{-1}. The main reason for the decreased prediction accuracy for bands 9 and 10 is the increased chance of surface contamination, especially for band 10 whose weighting function peaks closest to the surface among the three WV bands.

To help understand the importance of the window bands on the prediction, the permutation importance of each of the four window bands on the MLP regression model is calculated. This is done by calculating the difference in overall MSE with random permutation applied to the specific feature of the validation data set and repeating for 20 times. Mean and standard deviation (STD) scores are calculated based on the MSE changes over the 20 times. Shown by Table 2 are the rankings of permutation importance of the four window bands. Results show that Band 15 has the highest importance among the four, while Band 13 shows the lowest importance on the MLP. This is consistent with the fact that Band 15 has the most sensitivity to water vapor among the four, while Band 13, in contrast, has the least water vapor absorption and is regarded as the 'clean window band' (see Table 1).

The validation can be further separated into two groups: surface contaminated and surface uncontaminated. The prediction is highly accurate for the surface uncontaminated group with small STDs of 0.051, 0.074, and 0.096 mw · m^{-2} · Sr^{-1} · wn^{-1} for the three WV bands, indicating that the MLP can accurately predict the upwelling component for surface uncontaminated radiances. The small STDs suggest that the MLP QC will be able to identify a majority of the surface uncontaminated ABI WV radiances, even though the technique is trained on radiances both contaminated and uncontaminated. Based on these results, the MLP QC scheme is developed; if the surface contamination is smaller than the NEdR, the ABI WV radiance passes the MLP QC test.

Table 2
Permutation Importance Rankings of Four Window Bands From ABI (Unit: [mw · m^{-2} · Sr^{-1} · wn^{-1}]^2)

Ranking	ABI band no.	Mean score	Score STD
1	15	8.881	0.078
2	14	5.513	0.046
3	16	4.125	0.036
4	13	2.618	0.019

The prediction is less accurate for the surface contaminated group, with STDs of 0.069, 0.184, 0.494 mw · m^{-2} · Sr^{-1} · wn^{-1}. Using the same MLP QC, a substantial percentage of the surface contaminated samples pass the test. The majority of these samples have surface contributions smaller than the observation noise. These WV radiances, although surface contaminated, can be assimilated as surface uncontaminated. However, some of the samples which passed the MLP QC have a surface contribution larger than the observation

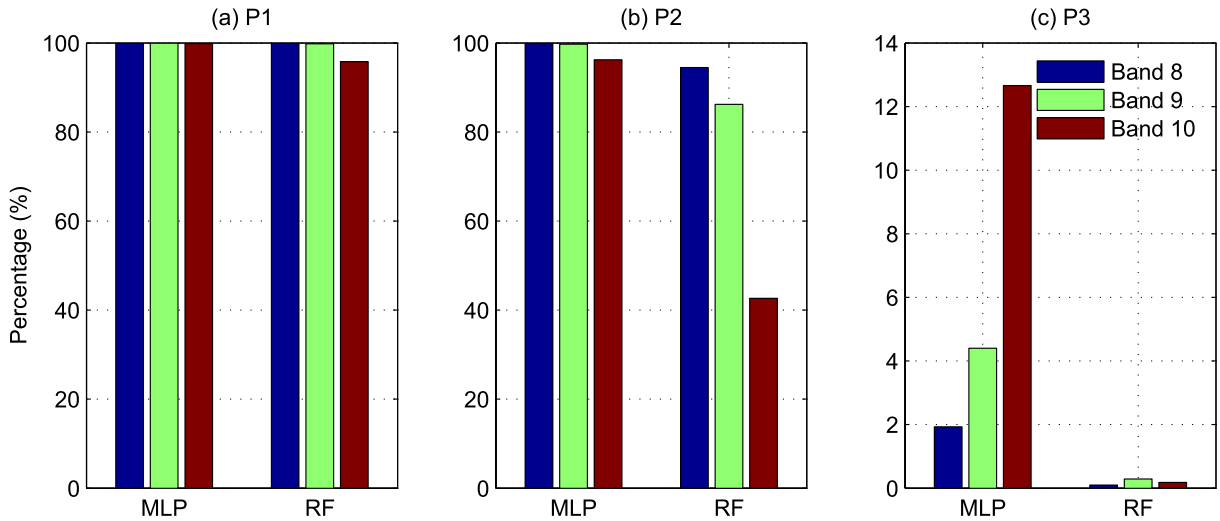

Figure 3. The performance of multi-layer perceptron (MLP) network and random forest (RF) quality control (QC) tests. P1 shows the percentage of the surface uncontaminated samples passing QC, P2 shows the percentage of the surface contaminated samples with contribution less than noise passing the QC, and P3 shows the percentage of residual surface contamination, or samples with surface contribution larger than noise, within all those passing the QC. The results are based on analyzing the validation results with the validation data set.

noise. For these radiances, assimilating them may potentially be detrimental. Therefore, additional QC on top of the MLP QC is beneficial to help further minimize surface contamination.

A good QC scheme should (a) retain as many of the surface uncontaminated samples (perfect data) as possible, (b) retain as many of the surface contaminated samples with a surface contribution less than the observation noise (good data) as possible, and (c) reject as many surface contaminated samples that have a surface contribution larger than the noise (bad data) as possible. Suppose the numbers of samples for the three portions before the QC test are A, B, and C. Accordingly the numbers of samples passing the QC are a, b, and c. Three parameters are used to characterize the effectiveness of a QC scheme.

$$P1 = \frac{a}{A} \times 100\% \tag{5}$$

$$P2 = \frac{b}{B} \times 100\% \tag{6}$$

$$P3 = \frac{c}{a+b} \times 100\% \tag{7}$$

where $P1$ is the perfect data retention rate or the percentage of perfect data passing the QC, $P2$ is the good data retention rate or the percentage of good data passing the QC, and $P3$ is the residual surface contamination rate or the percentage of bad data passing the QC. Note that $P3$ is calculated from the ratio concerning all samples passing the QC. For both $P1$ and $P2$, a value closer to 100% means a higher retention rate and indicates more skill in the QC scheme. In contrast, for $P3$, a value closer to 0 means a lower residual surface contamination rate and indicates more skill in the QC scheme. Figure 3 shows the MLP QC retains more than 99% of the perfect data, and more than 96% of the good data. However, there is a significant residual surface contamination rate within those samples passing the MLP QC, or 1.9%, 4.4%, and 12.7% for the three ABI WV bands respectively. Lower bands have more chances of surface contamination than upper bands. These surface contaminated WV radiances should be identified and not be assimilated.

2.2. The Random Forest-Based QC

The random forest (RF) (Breiman, 2001) is adopted to develop a classification-like prediction relationship. The same 7-band synthetic IR radiance observations are used as the features. Unlike the MLP QC, where the prediction is compared with the WV radiance to determine if it is surface contaminated, the RF QC directly predicts if

a WV radiance is surface contaminated. Before training, the threshold for classification is set to 0.001 mw · m^{-2} · Sr^{-1} · wn^{-1} between the observed radiance of a band and the upwelling component from the simulation. When the difference exceeds the threshold, it is considered surface contaminated (marked as 0); otherwise it would be considered uncontaminated (marked as 1).

The RF has been used for many classification problems (Belgiu & Drăguţ, 2016; Liu et al., 2019; Min et al., 2019; Pal, 2005). In this study, the classification model is developed as a chain classifier (Read et al., 2011). It is comprised of three RF-based single-output classifiers predicting the three output indicators of the three ABI WV bands. The order that the outputs are predicted is band 10, band 9 and band 8. During the prediction, every single classifier in the model gives predictions based on all the input predictors of the model and the predictions of models that are earlier in the chain. This methodology is adopted on the basis that the order of the peaks of the weighting functions for the three WV bands is Band 10, Band 9, and Band 8, from low to high in vertical levels (Li et al., 2020). If the band 10 radiance of a pixel is not surface contaminated (i.e., predicted as 1), radiances from bands 9 and 8 for the same pixel should not be contaminated, thus the prediction should also be 1. Similarly, if the band 8 radiance of a pixel is surface contaminated, so are the radiances from bands 9 and 10 of the same pixel. Compared to the traditional multi-target classifiers, the chain model used in this study ensures the internal physical connection among outputs while keeping the non-linearity and efficiency of RF training. The final configurations of RF (n_estimators = 400, max_depth = 20, and max_features = 4) were determined by a series of exhaustive searches using the GridSearchCV tool provided by Scikit-learn (Pedregosa et al., 2011) with a 5-fold cross-validation (Kohavi, 1995) applied to the training data set just like the MLP regression model.

Validation of the RF model was performed on the same validation data set as MLP. Figure 3 shows the RF QC results. Similar to the MLP QC, the RF QC can retain a majority (more than 95%) of the perfect data. The RF QC is less effective in retaining the good data. Band 8 and band 9 have a retention rate of 95% and 86% respectively. For band 10, only 43% is retained. However, the RF QC is much more effective in eliminating residual surface contamination. Band eight only has a residual surface contamination rate of 0.10%. This means that only 1 out of 1000 band 8 WV radiances which passed the RF QC has surface contamination larger than the noise. Bands 9 and 10 have slightly larger rates with 0.29% and 0.18% respectively.

2.3. The Dual MLP and RF QC Scheme

Results from using the MLP QC scheme indicate that the MLP QC is effective in retaining a majority of the perfect and good data but is less effective in eliminating bad data. The RF QC is effective in retaining the perfect data and eliminating bad data, but less effective in retaining the good data. Combining both schemes, the dual MLP and RF (DMR) QC scheme, may further improve the QC and eliminate bad data. An ABI WV radiance will be flagged as passing the QC and suitable for assimilation if both the MLP and RF QC tests are passed. Figure 4 shows that the DMR QC retains a majority of radiances passing the RF QC. The perfect data has retention rates of more than 99% for bands 8 and 9 and more than 95% for band 10. The good data has retention rates of more than 94% for band 8, more than 86% for band 9, and more than 42% for band 10. These retention rates are similar to or slightly smaller than the RF QC. However, the residual surface contamination rates are substantially reduced from the RF QC, especially for band 8 by 20% and band 9 by 14%. These results indicate that the DMR QC is useful in reducing the residual surface contaminations, at the expense of losing good data.

Using smaller thresholds for MLP QC can further reduce the residual surface contamination rates. For example, when the thresholds are halved, the residual surface contamination rates are reduced to 0.07%, 0.22%, and 0.17% for bands 8, 9, and 10 respectively (Figure 4). Meanwhile, the data yields are reduced as well. An average loss of about 8% in useable data is seen. When threshold values are reduced to 1/8 of the original ones, the average residual surface contamination rates are significantly reduced by about 51% for the three WV bands, and the average data yields are significantly reduced by about 66%. The reduced residual surface contamination rates increase the cleanness and thus the confidence in the quality-controlled data, which favors the assimilation study. The data yield reduction on the other hand does not. When processing real data, different thresholds are used to classify radiances that passed the QC. ABI WV radiances passing stricter thresholds are assigned with a higher quality category (Figure 4). For example, the quality flag is set as 1 if original thresholds are used, 2 for half, 3 for a quarter, and 4 for 1/8. This provides flexibility in tightening or relaxing the QC in the assimilation study to find the optimized QC category for each band.

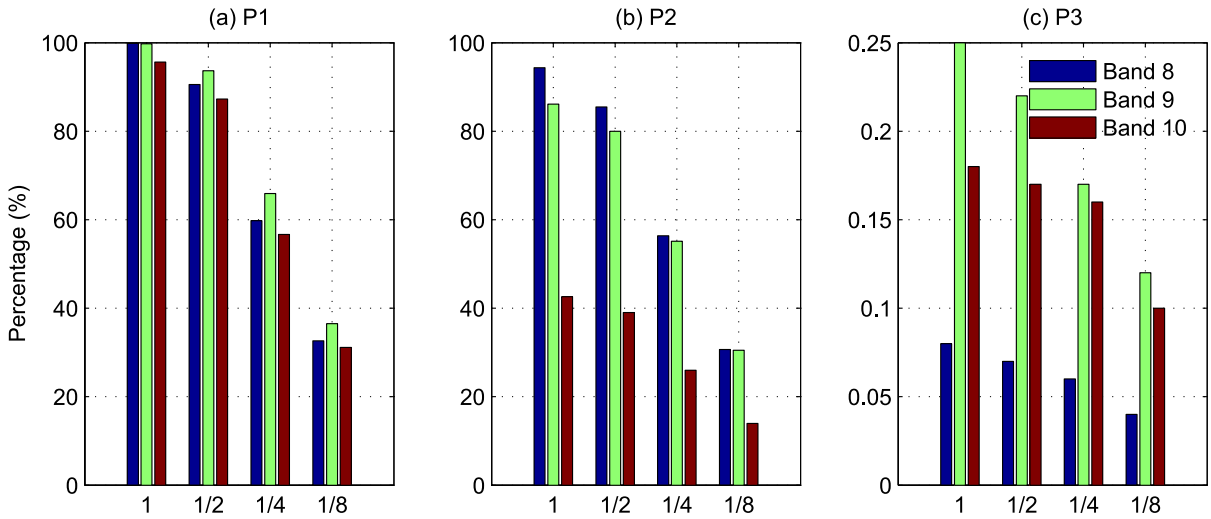

Figure 4. Similar to Figure 3 but for the dual multi-layer perceptron (MLP) and RF (DMR) QC test. The number on the *x*-axis indicates the thresholds used for the MLP QC test. One means the original threshold values are used. 1/2 means half of the threshold values are used, etc.

Figure 4 confirms that the residual surface contamination rate is small. The distribution of the surface contamination for all radiances passing the QC is examined. For those radiances to be assimilated, it is better to have minimum surface contamination. Figure 5 shows that the vast majority of the samples passing the QC have very little surface contamination. Band 8 has more than 96% of the samples with surface contributions less than 10% of the noise, band 9 has more than 95%, and band 10 has more than 98%. These results indicate that the DMR QC is highly effective in rejecting radiances with large surface contamination. ABI WV radiances passing the QC, even with surface contamination, are less likely to have surface contamination larger than the observation noise.

3. Application to Real Data

The DMR QC is applied to the case of Hurricane Harvey (2017) from 0000 UTC of 17 August to 2100 UTC of 26 August 2017. The full disk data are processed every 3 hr to generate the QC flags. Since the real ABI observations are more complicated than the simulated data, tuning of the DMR QC scheme is needed. Analyses were

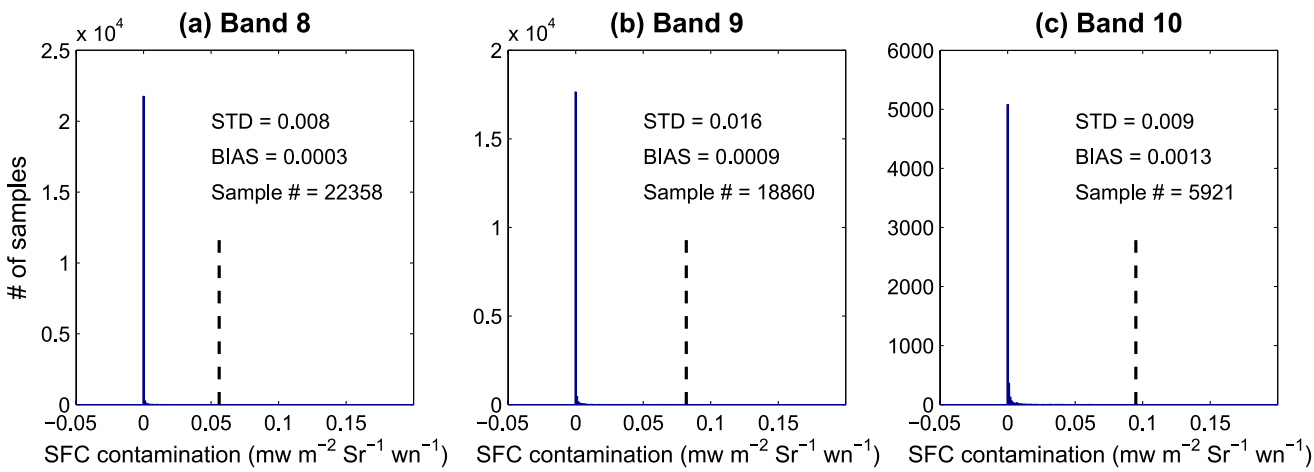

Figure 5. Histograms of the surface contamination, or the surface emission plus the surface-reflected atmospheric downwelling components, of the Advanced Baseline Imager (ABI) Water Vapor (WV) radiances that pass the DMR QC for (a) band 8, (b) band 9, and (c) band 10. The histograms appear to be a solid vertical line at 0 because most samples have no or little surface contamination. The vertical dashed lines show the observation noise of each band. The same validation database is used as in Figures 2–4.

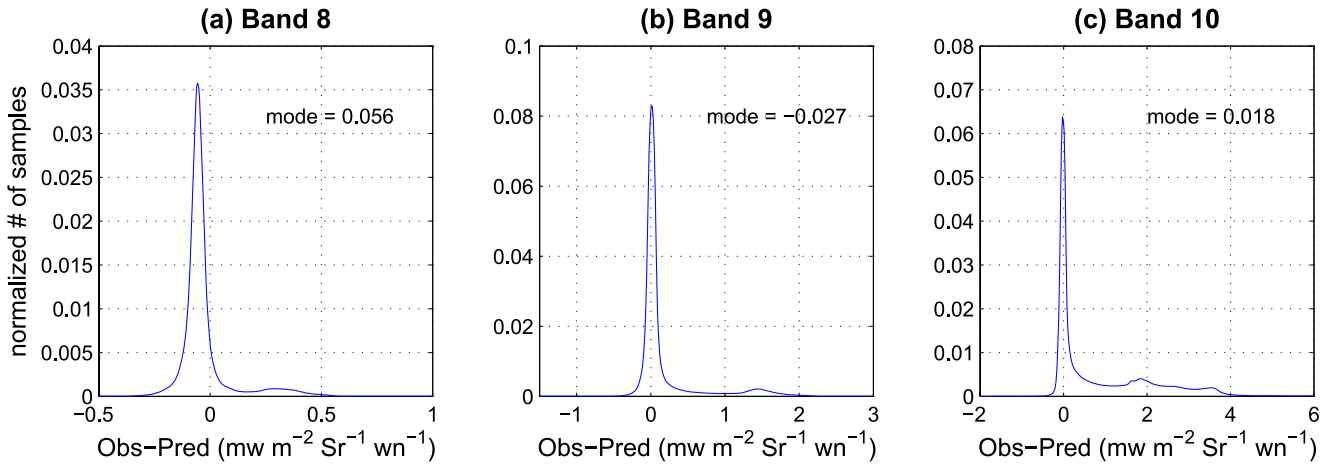

Figure 6. Normalized histograms of observed radiances minus predicted upwelling radiances for (a) band 8, (b) band 9, and (c) band 10 for ABI full disk observations of Hurricane Harvey from 00 UTC of 17 August to 21 UTC of 26 August 2017. Note that the modes, or the peak values, are not zero, which indicates biases might exist in either the observed or the predicted radiances or both. The mode must be removed before applying the MLP QC test. Note that the x and y ranges vary for each band.

also carried out to understand how the QC is affected by the atmospheric moisture content, and how the data yield is affected by different QC categories.

In the simulation study, the histogram of the surface contamination defined by Equation 3 peaks at zero, as shown in Figure 5, for accurately predicted surface uncontaminated samples. Due to the uncertainties in radiance observations and the radiative transfer model simulation, the histogram does not necessarily peak at zero for real data. Figure 6 shows that all three ABI WV bands have one mode significantly higher than other modes. The peak values of the modes are near zero for all three ABI WV bands. It is occurred when the predicted upwelling radiances are close to the observed, which is determined as surface uncontaminated. However, none of them are zero. This is likely due to biases in the observed or the predicted radiances, or both. Biases in the predicted upwelling radiances may come from the radiative transfer model simulation used in the training or the observed radiances that were used as inputs. Since the mode values for the three ABI WV bands are comparable to the observation noise, a bias correction is needed to remove the peak values before applying the MLP QC. The bias correction step is dynamic as the peak values could be different at different times. For example, for each time step, the histogram of the surface contamination or observation minus prediction is examined, and the highest mode is determined as the radiance bias. The bias is different for different time step due to scene dependency. The radiance bias is corrected simply by removing the bias from the observation minus prediction. The MLP QC is applied to bias-corrected data. The second tuning is based on the fact that if band 8 is surface contaminated, bands 9 and 10 must be contaminated as well. Similarly, if band 9 is contaminated, band 10 must be contaminated. This has been accounted for in the RF QC, but not the MLP QC, which will help reduce the residual surface contamination in bands 9 and 10. In addition, all radiances with local zenith angles larger than 60° are ignored since they are not assimilated (Liu et al., 2019a, 2019b).

The simulation study shows that stricter thresholds used in the MLP QC scheme lead to a smaller residual surface contamination rate, at the expense of the overall yield. In the training, random noise is added based on ABI pre-launch specifications to account for the uncertainties in the forward model simulation and the observed ABI radiances. It is not clear how large such uncertainties should be for real data. Providing QC flags based on different thresholds makes it possible to find the optimized QC category for each WV band. Figure 7 shows the quality flag imagery in clear skies for the three ABI WV bands, along with level-2 total precipitable water imagery (Jin et al., 2008). Band 8, being the strongest absorption band of the three, is the least contaminated by the surface. Only radiances with very high terrain and/or an extremely dry atmosphere in the central Andes' mountain region are masked as surface contaminated. Other regions that have high terrain, such as the Rocky Mountains are not identified as surface contaminated because they are likely not as dry. These radiances may be suitable for assimilation. Overall, more than 97% of the clear sky radiances for band 8 pass the DMR QC test for this time step (Figure 8a). As the threshold becomes stricter, fewer data pass the QC test. When the strictest threshold is applied, only 51% of the data is retained. Band 9 sees a similar pattern but is more sensitive to the threshold used for the

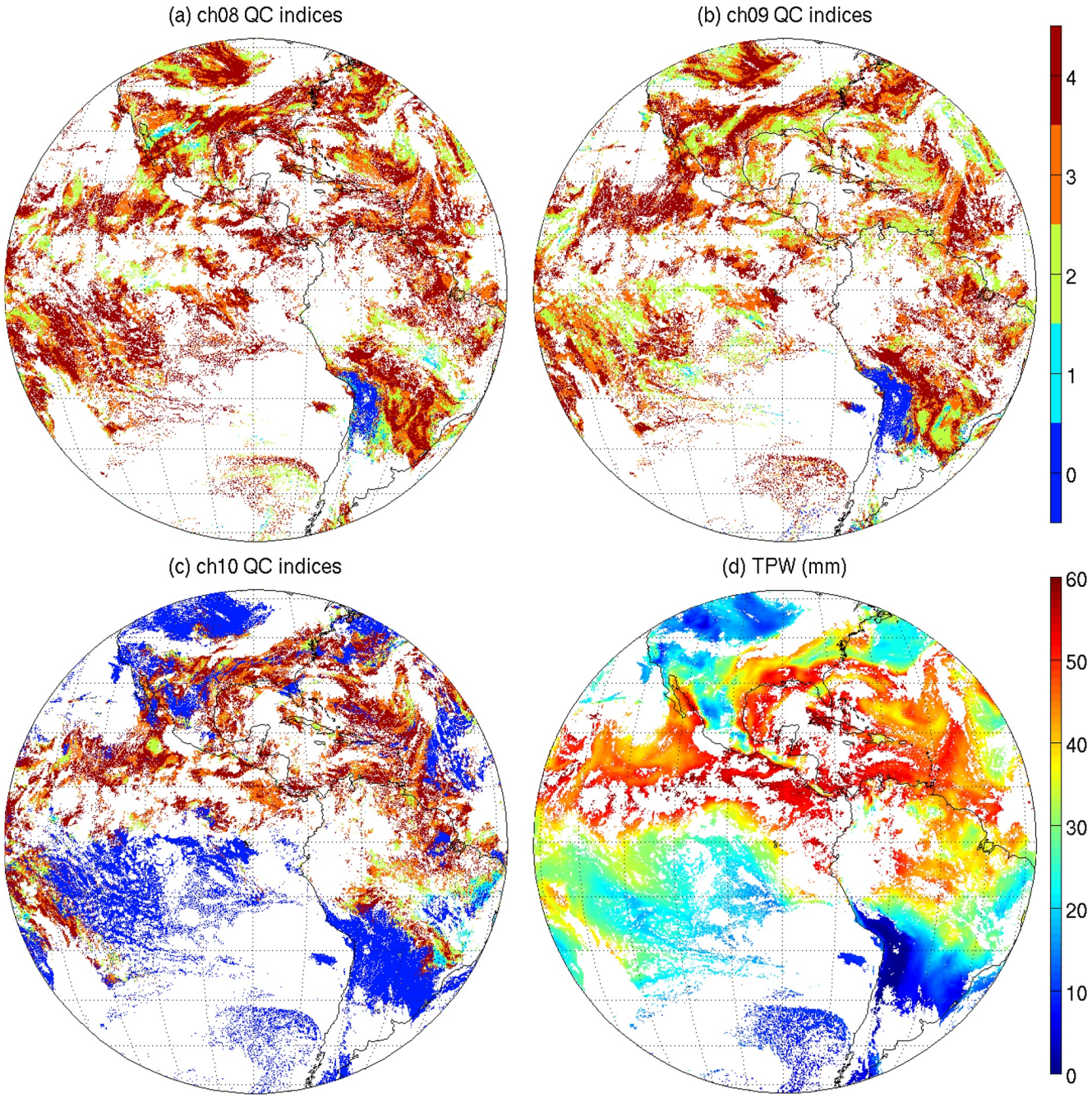

Figure 7. The quality control (QC) category imagery for (a) band 8, (b) band 9, and (c) band 10 of Advanced Baseline Imager (ABI)/GOES-16 at 1500 UTC on 22 August 2017, as well as (d) the level-2 total precipitable water imagery. 0 means surface contamination, 1 to 4 means no surface contamination with 1 using the least strict thresholds, and 4 using the strictest thresholds. Only the clear sky based on the ABI cloud mask is shown. The three QC category imagery use the same color bar on the top. Note that the spatial coverage of total precipitable water is slightly different from the QC imagery because the former is obtained with a spatial resolution of 10 km or a 5 × 5 box, and the retrievals are done when there are five or more clear pixels within the box.

MLP QC test. The yield is reduced from more than 96% to less than 37% with the tighter threshold values. Band 10, being the weakest absorption band, has the smallest yield. With the most relaxed threshold, only 56% of the radiances pass the QC test. The strictest threshold reduces the small yield by half, or 28% are retained. Figure 8b shows the mean yield of all time steps between 17 and 26 August 2017. The yield percentages are smaller than the case demonstration from 1500 UTC of 22 August 2017. The overall patterns are the same; band 8 is less prone to surface contamination than band 9, and band 9 is less so than band 10; and stricter thresholds result in

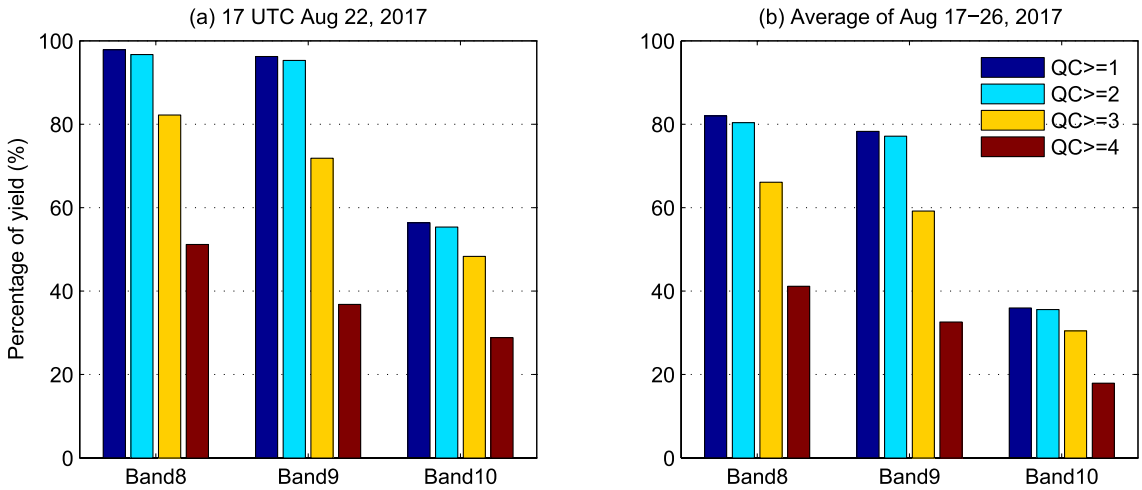

Figure 8. The data yields in percentage (%) of different DMR QC indices for the three Advanced Baseline Imager (ABI) water vapor bands for (a) the case demonstration at 1500 UTC on 22 August 2017, and (b) the average yields for the period from 17–26 August 2017 are also shown.

reduced data yields. Even with the strictest thresholds, bands 8, 9, and 10 still retain more than 41%, 32%, and 17% of the observations for assimilation, respectively. These reduced yields should have smaller residual surface contamination rates, as demonstrated by the simulation studies in Section 2.3.

Figure 7 also shows that the moisture content is the most important factor affecting surface contamination. Both bands 8 and 9 show strong surface contamination over the central Andes mountains, where the high terrain results in extremely dry air. And band 9 shows slightly more surface contamination because of weaker absorption. There are other areas that have a dry atmosphere, such as the southern Pacific off the coast of Chile, northern Argentina and southern Brazil, the high latitudes in the northwestern part of the United States, and southwestern part of Canada. However, these areas have relatively large local zenith angles. For the same amount of moisture, there is more absorption than with smaller local zenith angles because of the longer slant path. Since these areas are not as dry as the central Andes mountains, they are masked as uncontaminated by the surface.

Among the three absorption bands, band 10 is most affected by surface contamination. The regions, with surface contamination, appear to be in good agreement with the atmospheric moisture content. In August, the Southern Hemisphere experiences winter and has less atmospheric moisture than the Northern Hemisphere in summer. As a result, the moisture deficient areas in the Southern Hemisphere including much of the Southern Pacific Ocean and the southern part of South America are masked as surface contaminated. Note that surface contamination can happen in the tropics over the ocean. In the Northern Hemisphere, there are more spatial variations in the atmospheric moisture.

There are three regions with drier air: the high latitudes (greater than 40°N) in the northwestern United States and southwestern Canada, the Rocky Mountains, and off the coast of the northeastern United States. The majority of the high latitudes are masked as surface contaminated because this region is drier than the other two. Some of the southern boundaries of the high latitude region, with slightly more moisture, are masked as surface uncontaminated. Some areas in the Rocky Mountains and the northern Atlantic Ocean are masked as surface contaminated. Figure 7 show that these areas either have a dry atmosphere or a drier atmosphere than its surroundings. It is worth noting that the vertical distribution of moisture also plays an important role in the absorption of radiation. Due to the pressure broadening (Petty, 2006), the same amount of moisture in the lower atmosphere leads to more absorption than in the mid or upper atmosphere. While the lower atmosphere usually has the largest contribution to the total precipitable water, a different vertical distribution of the moisture may lead to different degrees of absorption. Therefore, it is possible that areas with slightly more moisture may appear to be surface uncontaminated while areas with slightly less moisture may appear to be surface contaminated. However, compared with its surroundings, the areas classified as surface contaminated appear to have less moisture.

As pointed out in previous sections, all radiances passing the DMR QC test are given a category value from 1 to 4, with 1 being the least stringent and 4 being the most strict. A higher QC category means higher confidence in

the quality-controlled data, however, the yield is smaller. A lower QC category may increase the yield but may contain more residual surface contamination, which may reduce or compromise the data impact. It is not clear which QC category best serves the radiance assimilation. Numerical experiments will be carried out to explicitly determine the optimized QC category, which will provide the optimized balance to maximize the positive impact from surface unaffected radiances and minimize the negative impact from surface contaminations. For this study, the objective is to maximize the overall impact of ABI WV radiance observations in assimilation rather than the number of surface unaffected radiances for assimilation.

4. Impact Study

As mentioned earlier, there is no direct way to validate the DMR QC when it is applied to real data since there is no truth value (e.g., which observation is surface contaminated) for comparison. The impact of the DMR QC can be evaluated indirectly through NWP impact studies. The radiances passing the DMR QC test will be assimilated in comparison with the GSI and Lee QC schemes. A successful QC is expected to have a positive impact on weather forecasts. This study mainly focuses on developing the QC methodology using the machine learning techniques and using it to better understand the importance and effectiveness of optimized QC. One single case demonstration with Hurricane Harvey (2017) in a regional model with the three dimensional variational (3DVar) assimilation approach is presented. In the following sections, three questions will be addressed: (a) how the optimized QC categories are determined, (b) whether each of the ABI three water vapor bands provides positive impacts, and c) how the DMR QC compares with existing schemes. Note that radiances from the CO_2 band (13.3 μm) is not assimilated in this study because it is also affected by the surface as shown in Figure 1.

4.1. Experiment Configuration

The GSI system supported by the Development Testbed Center version 3.7 is used as the data assimilation system. The 3DVar data assimilation method is used to evaluate the assimilation of the ABI WV bands. In addition, we are using the NCEP North American Mesoscale (NAM) Forecast System background error covariance matrix (B matrix). The NAM B matrix is not tuned for the domain used in this study. So it is not necessarily optimal for hurricane forecasting and may degrade the analysis and forecast results. However, all experiments use the same configurations (including the NAM B matrix), settings, and observations except the ABI WV radiances from different QC schemes. The differences between different experiments should reflect the differences in those QC schemes. The impact of the sub-optimal B matrix should be small when comparing different numerical experiments. For the satellite radiance bias correction, we use the enhanced bias correction method (Zhu et al., 2014). The thinning box for the satellite radiances is 60 km. The Advanced Research Weather Research and Forecasting Model (WRF-ARW) version 3.9.1 is used as the regional NWP model. The horizontal resolution is 12 km with 400 × 400 grid points. There are 52 vertical layers from surface to 10 hPa. For the model simulation, the NCEP Final operational global analysis data (0.25° × 0.25°) provides the initial and boundary data. The WSM6 is chosen as the microphysics scheme (Hong & Lim, 2016), the Rapid Radiative Transfer Model for Global Climate Models as the longwave and shortwave radiation scheme (Iacono et al., 2008), Kain-Fritsch as the cumulus scheme (Kain, 2004), and the Yonsei University as the boundary scheme (Hong et al., 2006).

In Figure 7, only clear sky results are shown to avoid the complication caused by cloud contamination. However, in the assimilation experiments, no cloud mask was applied. The assimilation experiments will rely on the GSI's cloud detection to screen out possible cloud contamination. This allows more WV radiances to be assimilated because not all of them in the cloud region are affected by clouds, and also to be consistent with the GSI and Lee QC schemes. The assimilation of ABI radiances follows the work by Lee et al. (2019).

Hurricane Harvey (2017) was one of the strongest hurricanes that led to heavy precipitation for the coastal regions of Texas. Harvey was significantly affected by the land surface, especially after landfall, where the forecast became more difficult. Three experiments are carried out to understand the impact of the DMR QC scheme. Each of the three experiments assimilates a baseline of observations plus the ABI/GOES-16 WV radiances. Observations included in the baseline are the Global Telecommunication System conventional data and commonly assimilated satellite radiances, such as the Advanced Microwave Sounding Unit-A from NOAA-15, NOAA-18, NOAA-19, Metop-a/b, the Advanced Technology Microwave Sounder from Suomi National Polar-orbiting Partnership, the Infrared Atmospheric Sounding Interferometer from Metop-a/b, and the Cross-track Infrared

Table 3
The Number of Radiance Observations Assimilated by Different QC Categories for Each ABI WV Band at 1200 UTC on 23 August 2017

	QC = 2	QC = 3	QC = 4	GSI QC scheme	Lee QC scheme
Band 8	3,443	**3,053**	1,817	3,265	3,289
Band 9	3,409	**2,451**	1,073	3,128	3,149
Band 10	2,343	2,150	**1,396**	3,092	2,606

Note. Bold numbers show the optimized QC category for each band. The numbers from the GSI and Lee QC schemes are also shown.

Sounder from Suomi National Polar-orbiting Partnership. In Experiment 1 (EXP1), the ABI WV radiances are assimilated using the GSI QC scheme. Experiments 2 (EXP2) and 3 (EXP3) assimilate ABI WV radiances using the Lee and the DMR QC schemes to remove surface contamination, respectively. For each of the 3 experiments, the cycling assimilation runs start at 1200 UTC on 23 August 2017, and end at 1200 UTC on 24 August 2017. Each run assimilates the ABI WV radiance data, which is followed by a 72-hr forecast. The 6-hourly forecast output is used as the background for the next assimilation run. In total there are 5 groups of cycling runs. Mean error and root mean square error (RMSE) statistics will be used for the impact study, focusing on the track and intensity forecasts.

4.2. Optimized QC Categories

Figure 8 shows that different QC categories have different data yields and likely different residual surface contamination rates. The optimized QC is determined explicitly through impact studies using EXP3 with different configurations of QC categories. Starting with band 8, each QC category from 2 to 4 was tested. Category 1 is not tested because there are too many observations available for assimilation and the coverage reduction is small from 1 to 2 (Figure 7). Three experiments were carried out for each of the QC categories (2–4) for band 8. Based on the comparison of the 72-hr mean error and RMSE of the hurricane track error, minimum sea level pressure (SLP), and maximum wind speed (SPD), QC category 3 was found to provide the best forecast. This process was repeated for bands 9 and 10, each time also assimilating the band where an optimized QC category was found. The optimized QC category for band 9 was also 3. Band 10, the lowest peaking band of the three, is highly prone to surface contamination. The strictest QC category four is found to provide the best forecast impact.

Table 3 shows the number of radiances assimilated by different QC categories for each band at 12 UTC on 23 August 2017, as well as those from the GSI and Lee QC schemes for the whole experiment domain. Note that band 9 is tested with band 8 assimilated (QC category 3), and band 10 is tested with bands 8 and 9 assimilated (QC category 3 for both). There are 3,053 band 8 radiances assimilated in the domain. It is substantially less than the 3,443 from QC category 2, and significantly larger than 1,817 from QC category 4. QC category 3 likely has the best balance between surface unaffected radiances and residual surface contamination, maximizing the impact from the former while minimizing the impact from the latter. Band 9, peaking lower than band 8, using QC category 3, has substantially less optimized radiances (2,451) assimilated than band 8. And band 10, the lowest peaking band has only 1,396 optimized radiances assimilated using QC category 4.

Compared with the GSI QC scheme (similar to the Lee QC scheme), the DMR QC scheme has 6% fewer radiances assimilated for band 8, 22% fewer for band 9%, and 55% fewer radiances for band 10. This indicates that there are significant differences between the DMR QC scheme and the GSI and Lee QC schemes. It is not clear if this is because the DMR QC scheme results in a smaller residual surface contamination rate than the GSI and Lee QC schemes. However, the DMR QC scheme is solely based on observations and does not prescribe the surface impact based on surface type or elevation. Impact studies with optimized QCs will be carried out to show that each one of the three WV bands has additional added value in the hurricane track and intensity forecast and that the DMR QC scheme with all three ABI WV bands can improve the hurricane forecast over both the GSI and Lee QC schemes. Also note that there are no significant differences between GSI and Lee QC schemes for bands 8 and 9. But for band 10, Lee QC scheme uses 16% fewer radiances than GSI QC.

One of the benefits of using a machine learning method to detect surface contamination is its independence from the model background. As long as there is enough moisture in the atmosphere, making the surface opaque to the sensor, the radiance can be assimilated. There is no need to classify land and ocean. And there is no need to consider the surface elevation. Figure 9 shows the radiances from ABI band 10 with optimized QC category four of the DMR QC scheme that are assimilated cover both ocean and land, including many of the southeastern states. However, the DMR QC scheme assimilates much fewer radiances than the GSI and Lee QC schemes, rejecting many radiances over both ocean and land. Band 8, the highest-peaking ABI WV band, has more WV radiances assimilated over high terrain in Mexico. As a comparison, NCEP rejected all ABI WV radiances with terrain height above 1 km over land in FV3GFS.

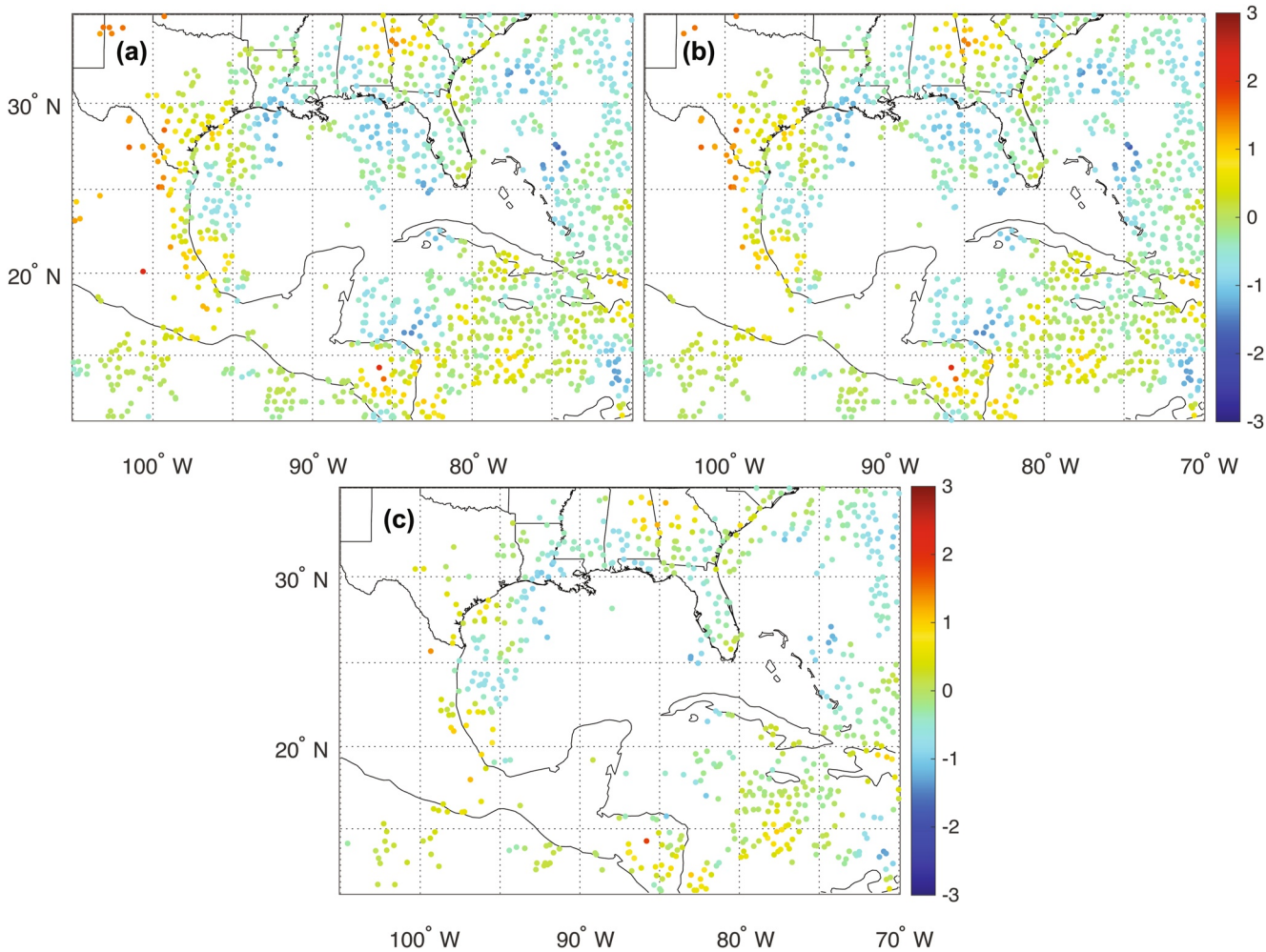

Figure 9. The Advanced Baseline Imager (ABI)/GOES-16 band 10 analysis minus background (AmB) (K) from assimilated with (a) Gridpoint statistical interpolation (GSI) QC scheme, (b) Lee QC scheme, and (c) the optimized QC category 4 of the DMR QC scheme at 12 UTC, on 23 August 2017. (a and b) represent the calculated brightness temperatures based on the analysis and background temperature and moisture fields.

Figure 10 shows that QC category 3 has substantial improvements for the 72-hr track forecast compared to QC categories 2 or 4 for Band 9. The mean error and RMSE are the smallest. QC category 4 shows slight improvements on the track forecast for RMSE when compared with QC category 2, but not as much as QC category 3. The improvements become larger as the forecast hour increases. The largest improvement is seen after 72 hr of forecast; the RMSE of the track forecast error is improved by around 35 km from QC category 2 to 3, and less than 16 km for all forecast hours from QC category 2 to 4. These results indicate that QC of the surface contamination is critical to the successful assimilation of ABI WV radiances for Hurricane Harvey. Table 3 shows that QC category 2 has many more radiances assimilated, meaning better data coverage, thus more potential positive impacts. The simulation study shows radiances passing QC category 2 are more prone to residual surface contamination than category 3. Negative impacts from a larger residual surface contamination rate may outweigh the positive impacts from the uncontaminated data. QC category 4 has much fewer radiances assimilated, meaning reduced data coverage, thus reduced potential positive impacts. The potential negative impacts from the surface contaminated radiances are likely reduced as well because of the reduced residual surface contamination rate. The overall impacts from QC category 2 to 4 are slightly positive. QC category 3, on the other hand, provides the optimized balance to maximize the positive impacts from surface uncontaminated radiances and minimize the negative impacts from the radiances with residual surface contamination. Note that the impact on hurricane intensity is either slightly positive or neutral. A possible reason is that the assimilated ABI radiances are clear, which mostly come from the environmental region and contain little information about the cloudy regions over

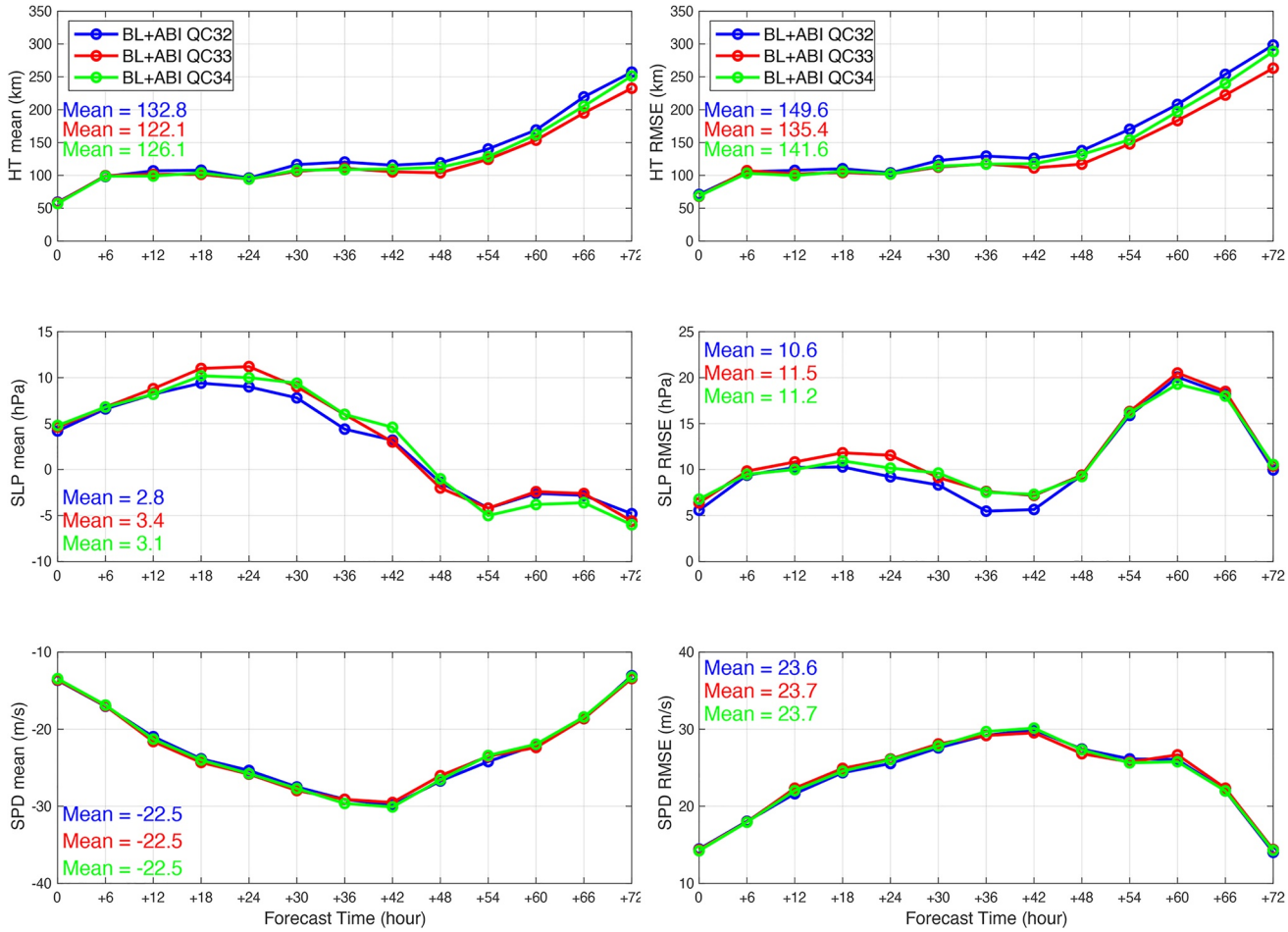

Figure 10. The 72-hr forecast impacts on (left) the mean error and (right) root mean square error (RMSE) of (top) hurricane track (HT), (middle) minimum seal level pressure (SLP), and (bottom) maximum wind speed (SPD). Different colors represent different experiments assimilating the baseline measurements plus Advanced Baseline Imager (ABI) water vapor radiances with different combinations of the QC categories for ABI bands 8 and 9, that is, ABI QC32 means DMR QC categories of 3 and 2 are used for bands 8 and 9 respectively. Note that QC33 has the smallest mean and RMSE track errors and QC category 3 is determined as optimized for band 9 accordingly.

the hurricane domain. The information from the environmental region can change the temperature or moisture fields in the large scale to improve the hurricane track forecast. The impacts on the intensity forecast are usually not obvious and neutral. Improving intensity forecast through satellite radiance assimilation needs further investigation. Selecting the optimized QC category for band 9 depends more on the track forecast than the intensity.

4.3. Added Value From Individual Bands

The three ABI WV bands are highly correlated because of the significant overlap in their Jacobians. However, in clear skies they peak at different heights. Each of them should be able to provide the thermodynamic information for different layers of the atmosphere, which may lead to improvements in hurricane track and intensity forecasts. Figure 11 shows that the mean error and RMSE of track forecasts are improved when additional bands are added. When band 9 with QC category three is added on top of band 8 (also with QC category 3), the RMSE of the track forecast error is reduced. The improvements are greater as the forecast hour increases. The largest improvement, or reduction of RMSE, at 72-hr forecast, is around 32 km, or 11%. Over the course of the 72 hr of the forecast, the average improvement in RMSE is around 10 km, or 7%. Similarly, band 10 with QC category 4 also helps improve the track forecast with both bands 8 and 9 assimilated. The largest RMSE reduction at the 72-hr forecast is 43 km, or 17% when band 10 is added, and the average improvement in RMSE over the course of 72 hr is 12 km or 9%. These results suggest that the assimilation of WV radiances from each of the three WV bands with

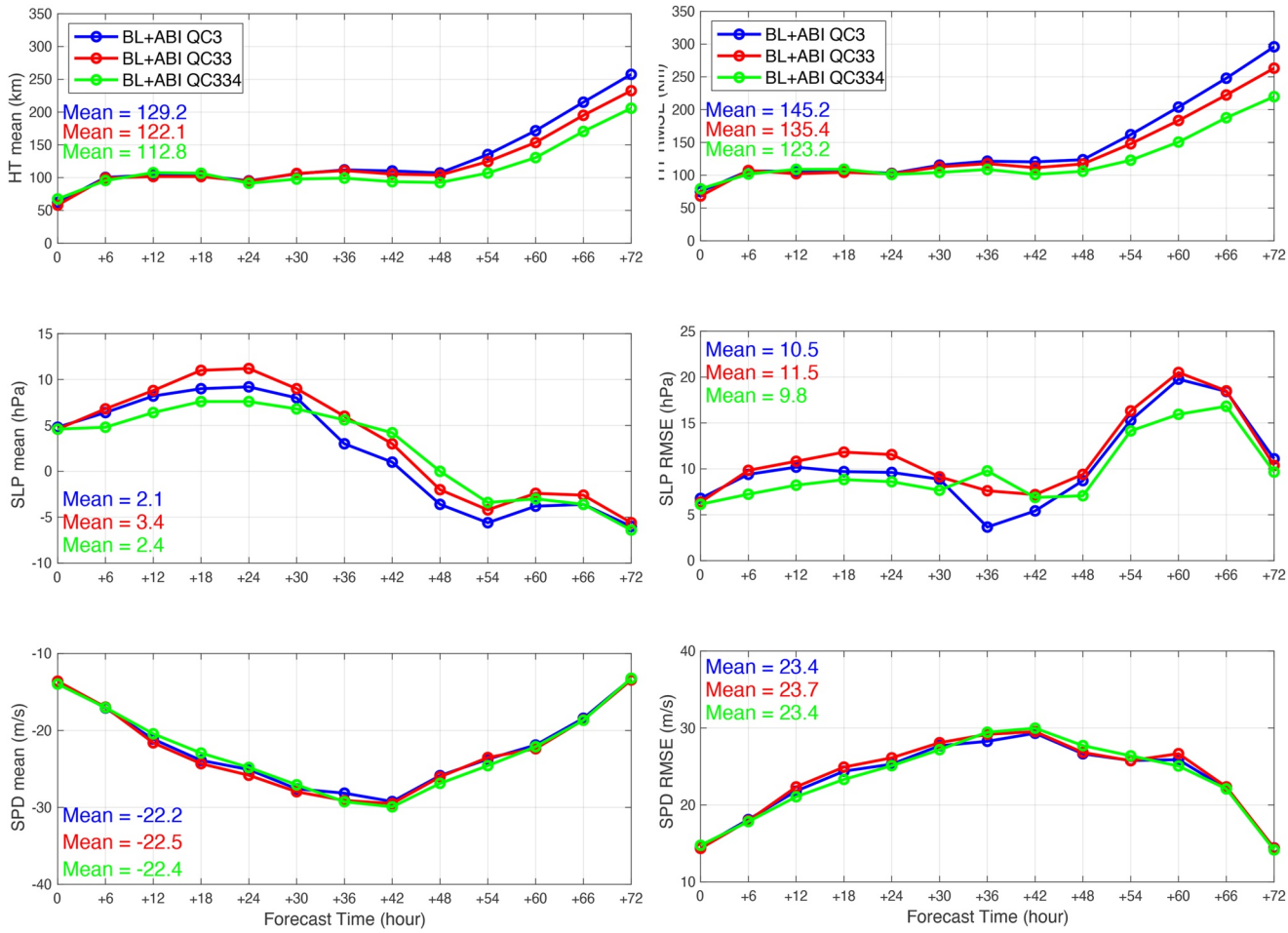

Figure 11. Similar to Figure 10 but for three different experiments assimilating the baseline measurements plus Advanced Baseline Imager (ABI) water vapor radiances from (blue) band 8, (red) bands 8 and 9, and (green) bands 8, 9, and 10. The quality control (QC) categories for bands 8, 9, and 10 are 3, 3, and 4, respectively. Note how bands 9 and 10 each provide additional added value on the Hurricane Harvey track forecast.

the DMR QC scheme for surface contamination provides independent added value to the Hurricane Harvey track forecast, and the combination of the three bands has the largest impact. Again, the impact on hurricane intensity is either slightly positive or neutral.

4.4. Comparison With GSI and Lee QC Schemes

Once the optimized QC category for each band is determined (3, 3, and 4 for bands 8, 9, and 10, respectively), EXP3 was carried out and the forecasts were compared with those from EXP1 and EXP2, again focusing on track and intensity forecasts. Figure 12 shows the Lee QC scheme improves the track forecast when compared with the GSI QC scheme. The improvement is consistent for almost all forecast hours. The 72-hr forecast has the largest improvement with RMSE reduced by around 36 km, or 11%, with the average improvement over the 72 hr of the forecast around 15 km, or 10%. The mean errors are also reduced. There are some impacts on the intensity, but the overall impact is either slightly positive or neutral.

The DMR QC scheme, which relies solely on observations for QC, can show larger improvements in the hurricane forecasts compared with the Lee QC scheme. The track forecast is significantly improved compared with the GSI QC scheme. The largest reduction in RMSE, again at the 72-hr forecast, is 108 km, or 33%. This is around 72 km (or 25%) better than the Lee QC scheme. The impact is neutral for the first 24 hr of the forecast. Beyond that, both the mean error and RMSE are consistently smaller than the Lee QC scheme. Over the course of 72 forecast hours, the average improvement in RMSE over the GSI QC scheme is 33 km or 21%. That is 18 km (or

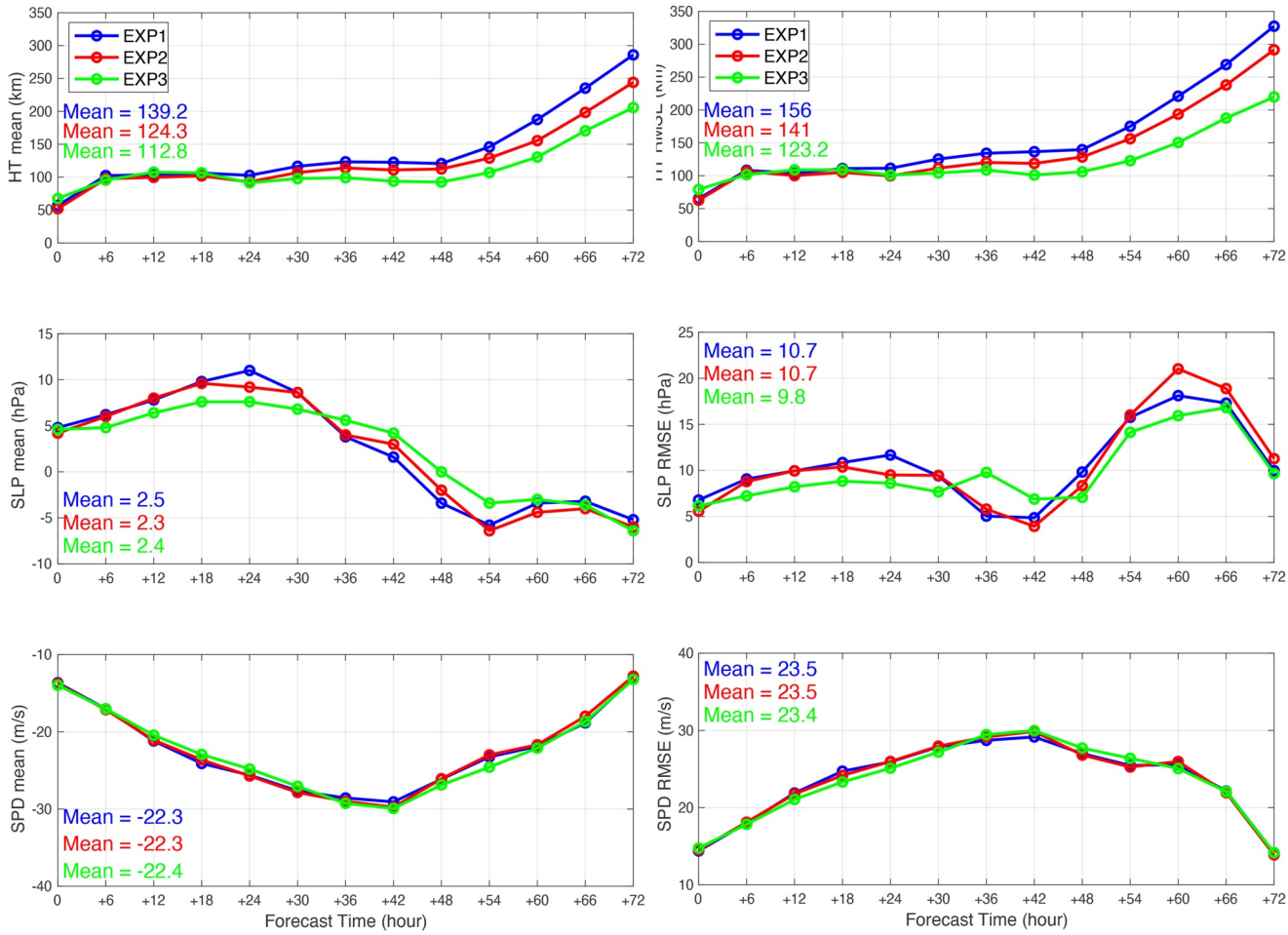

Figure 12. Similar to Figures 10 and 11 but for three different experiments assimilating the baseline measurements plus Advanced Baseline Imager (ABI) water vapor radiances from (blue) EXP1, (red) EXP2, and (green) EXP3. For EXP3, the QC categories for bands 8, 9, and 10 are 3, 3, and 4, respectively. Note that the DMR QC in EXP3 significantly improves the track forecast.

13%) better than the Lee QC scheme. The DMR QC scheme has some positive impacts on the minimum sea level pressure. For most forecast hours, the mean error and RMSE from the DMR QC scheme are the smallest. Over the 72 hr, the average RMSE is 0.9 hPa (or 8.4%) smaller than the GSI QC scheme, and 0.9 hPa (or 8.6%) smaller than the Lee QC scheme. The impact on the maximum wind speed is slightly positive, but not significant.

To help understand how the assimilation of ABI radiances with the DMR QC scheme improves the track forecast, the analysis minus background (AmB) of band 10 at 1200 UTC on 23 August 2017 in Figure 9 is further analyzed. A and B represent the calculated brightness temperatures based on the analysis and background temperature and moisture fields. This analysis time is chosen because it is the first cycle and all three experiments can be fairly compared with the same background. All three experiments show similar AmB pattern. A is colder than B over much of the ocean for all three experiments, especially around Hurricane Harvey (2017) over the Gulf of Mexico. Further away from the hurricane, A is mostly warmer than B. However, there are visible differences in the three experiments. EXP2 assimilates slightly less band 10 brightness temperatures than EXP1, and EXP3 assimilates less than EXP2. Figure 9 shows the reduction from EXP1 to EXP2 is not much, and happens mostly over land, such as northwestern Texas and central Mexico. From EXP1 to EXP3, the reduction is much more significant. Over land, most of the observations leading to positive AmB are rejected by EXP3. Over ocean, much of the observations leading to large negative AmB are rejected.

These differences in AmB from different experiments lead to differences in the analysis fields of temperature and moisture. For example, because EXP2 rejects some observations that would lead to increased A over northern

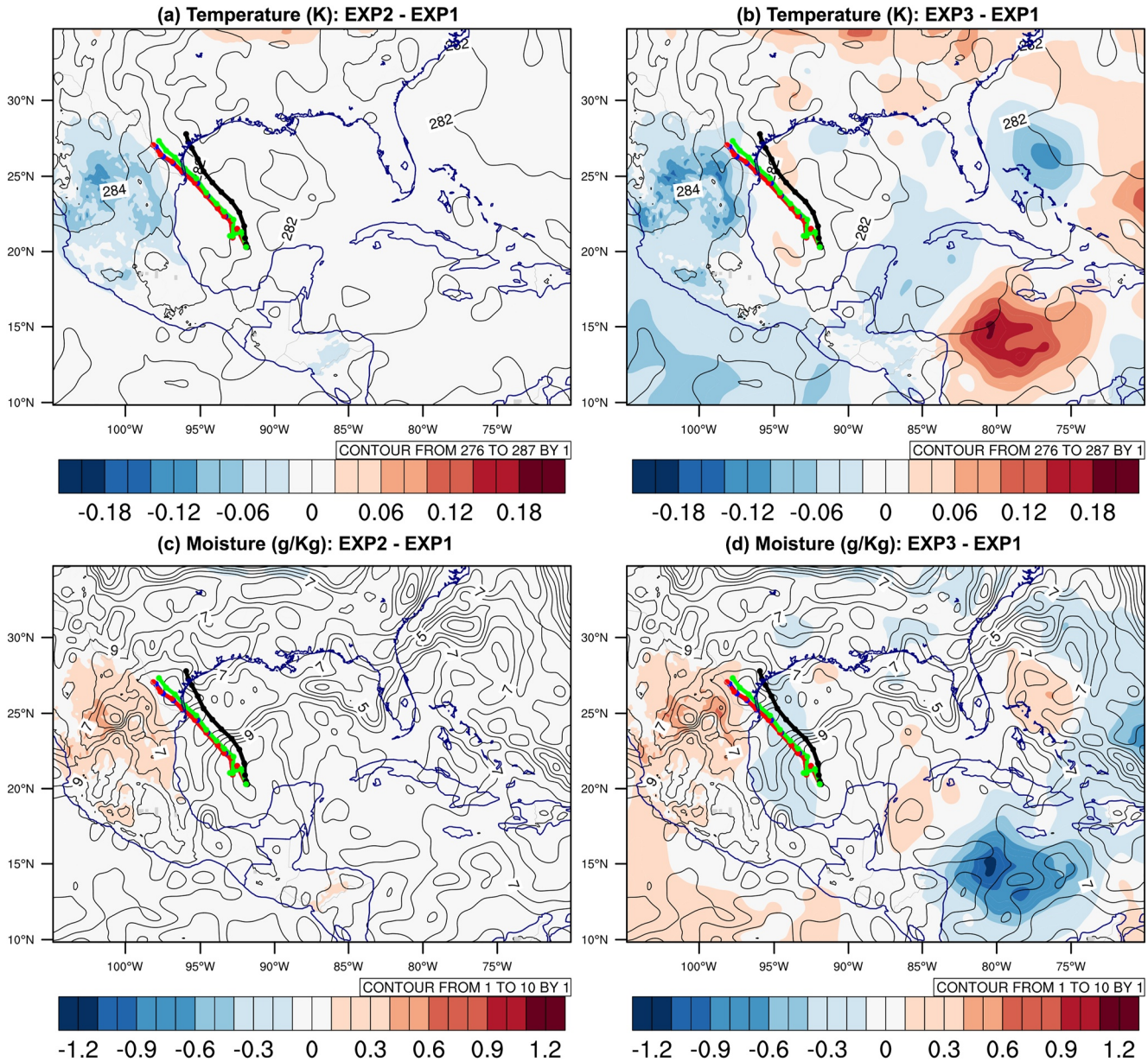

Figure 13. The temperature (a and b), unit: K) and moisture (c and d, unit: g/kg) fields of EXP2 (a and c) and EXP3 (b and d) in black contour. Overlaid is the difference (shaded) compared with EXP1, as well as the best track (black dotted line) and the 72-hr forecasts track (EXP1 in blue, EXP2 in red, EXP3 in green) of Hurricane Harvey (2017). Analysis time is 1200 UTC on 23 August 2017.

Mexico (Figures 9a and 9b), the analysis temperature at 700 hPa is increased from EXP1 to EXP2 (Figure 13a) over the region. The moisture analysis, on the other hand, is decreased from EXP1 to EXP2 (Figure 13c). This opposite effect is due to the temperature Jacobian being positive and the moisture Jacobian being negative for WV bands. Similarly, in Figures 9a and 9c, there are more observations rejected over northern Mexico and southern Texas from EXP1 to EXP3. As a result, the temperature analysis decrease and moisture increase from EXP1 to EXP3 are larger than from EXP1 to EXP2 (Figure 13b). Over the Caribbean, many of the observations that would lead to decreased A are not assimilated in EXP3. As a result, the analysis temperature at 700 hPa is increased from EXP1 to EXP3 (Figure 13b), and the analysis moisture is reduced over the region (Figure 13d).

The temperature and moisture analysis field differences in different experiments lead to differences in geopotential height (GPH) field. Figure 14a shows that the 700 hPa GPH is reduced over northern Mexico from EXP1 to EXP2. The GPH reduction is even larger from EXP1 to EXP3 (Figure 14b). And the region affected is also larger,

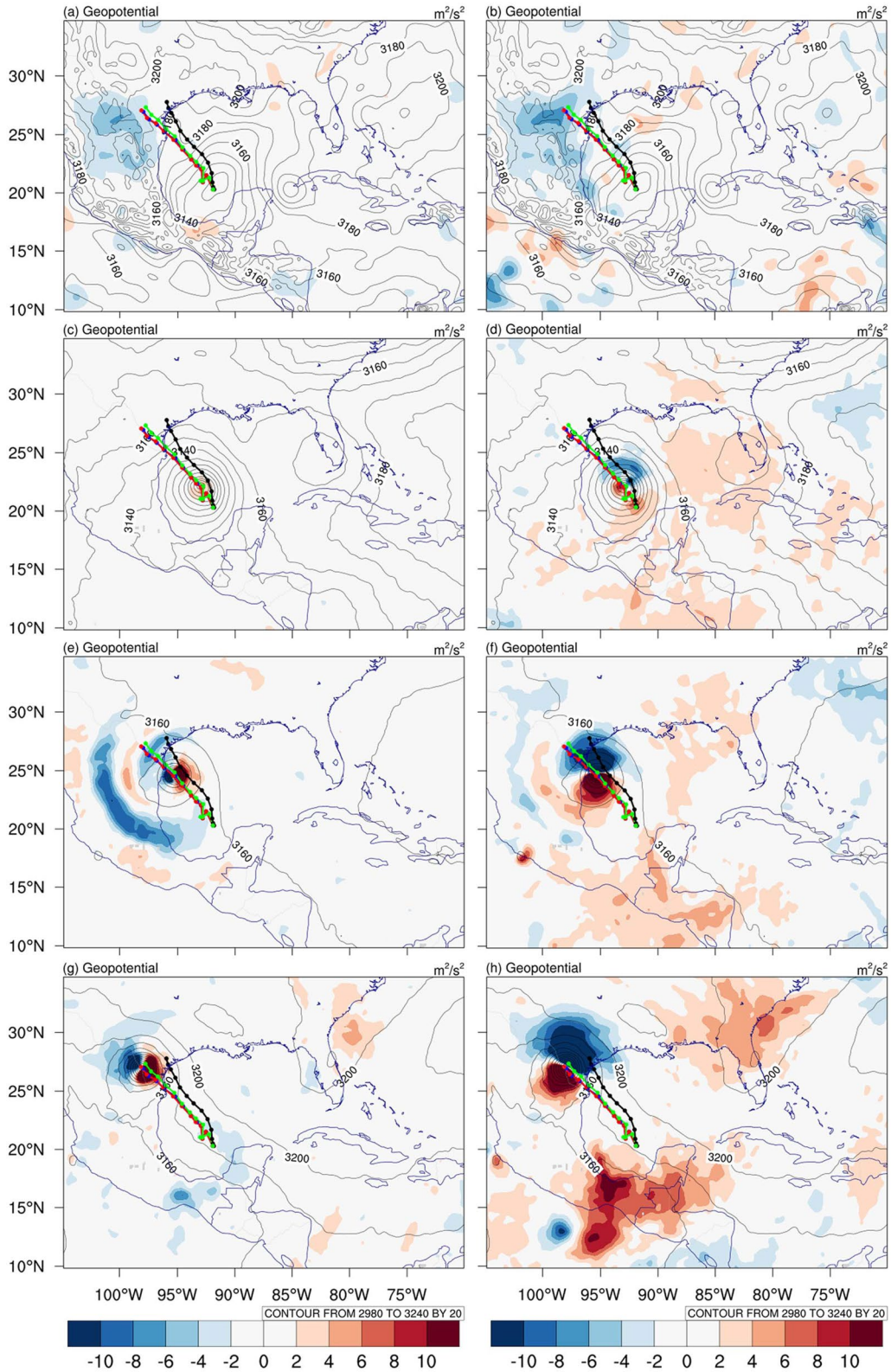

Figure 14. The geopotential height of EXP2 (left: a, c, e, and g) and EXP3 (right: b, d, f, and h) in black contour (m^2/s^2), overlaid with its difference (shaded) compared with EXP1, as well as the best track (black dotted line) and the 72-hr forecasts track (EXP1 in blue, EXP2 in red, EXP3 in green) of Hurricane Harvey (2017). The 4 rows from the top to the bottom represent different times: analysis time at 1200 UTC on 23 August 23 (a and b), 24-hr forecast at 12 UTC on 24 August (c and d), 48-hr forecast at 12 UTC on 25 August (e and f), and 72-hr forecast at 12 UTC on 26 August (g and h).

covering both northern Mexico and southern Texas. Over the Caribbean in the southeast corner of the domain, the GPH is increased from EXP1 to EXP3 (Figure 14b). The location of GPH differences in Figures 14a and 14b does not appear to match perfectly with the temperature and moisture field changes in Figure 13. This is because GPH has a positive relationship with both temperature and moisture. The opposite effect of assimilating WV band brightness temperatures on the temperature and moisture fields reduces their impacts on GPH. However, the two regions with large GPH change (such as northern Mexico and the Caribbean), are where large temperature and moisture changes are. The GPH change at this analysis time is directly from the assimilated data set. Therefore, the GPH changes in Figures 14a and 14b reflect the impact of the assimilated ABI WV radiances due to different QC schemes at the analysis time.

For the 24-hr forecast, GPH differences between EXP2 and EXP1 are weak and mostly near the hurricane area, while those between EXP3 and EXP1 are more substantial. There exist gradients from south to north in the environment and the hurricane region from EXP3 to EXP1. No obvious gradients are seen from EXP2 to EXP1. In the 48-hr forecasts, EXP2–EXP1 shows overall gradients from east to west whereas EXP3–EXP1 gradients are stronger. EXP3–EXP1 gradients maintain the same direction from south to north and are over a large domain, covering both the hurricane and the environment. In the 72-hr forecasts, both EXP2–EXP1 and EXP3–EXP1 gradients grow stronger. EXP3–EXP1 gradients are still stronger than EXP2–EXP1 and with a larger domain, again covering the hurricane and the environment. EXP2–EXP1 gradients show a direction from southeast to northwest, while EXP3–EXP1 gradients maintain the same direction from south to north, although slightly toward east.

The differences in the domain and the strength of GPH gradients have a direct impact on the track forecast. For example, EXP2–EXP1 gradients are weaker, so the EXP2 track is close to EXP1. On the other hand, EXP3–EXP1 gradients are stronger and maintain the direction from south to north, so the EXP3 track is pushed northwardly and is closer to the best track. 500 hPa has similar results where GPH gradients push the hurricane closer to the best track.

5. Summary

Infrared water vapor radiances sensitive to surface contamination are usually not assimilated, or assimilated with reduced weight into NWP models due to their reduced or even compromised impacts on the analysis and forecast. The quality control (QC) procedure used to handle these observations is critically important for the successful assimilation of infrared water vapor radiances. Existing schemes are not objective because of they rely on the background. They can be problematic in areas where there are large background errors, such as the moisture field. A new QC scheme for surface contamination in ABI WV radiances based on machine learning is developed. The scheme is built upon two machine learning techniques, the MLP network and the random forest (RF), named the dual MLP and RF QC, or DMR QC scheme. Unlike the existing QC schemes, the DMR QC scheme is an objective method because it is solely based on ABI radiance observations. The DMR QC scheme is used to better understand the importance and effectiveness of optimized QC, which provides optimized balance between maximizing the positive impact of surface unaffected water vapor radiances and minimizing the negative impact of surface contaminated water vapor radiances.

The MLP QC treats the problem as a regression problem. Radiances from 7 ABI infrared bands (bands 8–16, excluding bands 11 and 12) are used as input. The atmospheric upwelling radiances of the 3 ABI water vapor bands are predicted with MLP. If the prediction is close enough to the observed water vapor radiance, then the observation is deemed to be free from surface influence and is classified as surface uncontaminated. The simulation study shows that the MLP QC can retain more than 99% of surface uncontaminated radiances and more than 96% of radiances with surface contamination smaller than the observation noise. However, significant residual surface contamination (larger than observation noise) is found in those radiances that passed the MLP QC. The three ABI water vapor bands are found to have residual surface contamination rates of 1.9%, 4.4%, and 12.7% with the MLP QC.

The RF QC treats the problem as a classification problem. Using the same input—that is radiances from the 7 ABI infrared bands—the RF QC directly determines if a water vapor band radiance is surface contaminated. The simulation studies show that the RF QC can retain more than 95% of surface uncontaminated water vapor radiances for all three ABI water vapor bands. It is less effective in catching water vapor radiances with surface

contamination less than the observation noise (94% for band 8, 86% for band 9%, and 42% for band 10). However, the RF QC has significantly smaller residual surface contamination rates than the MLP QC. The rate for band 8 is only 0.10%, while the rates for bands 9 and 10 increase slightly to 0.29% and 0.18%, respectively.

Combining both MLP and RF QC schemes, the DMR QC scheme shows improved performance of the QC. The simulation studies show that the DMR QC is still capable of retaining a high percentage of surface uncontaminated water vapor radiances and more importantly, reducing the residual surface contamination rates. The residual surface contamination rate can be further reduced by using stricter thresholds for the MLP QC. When the thresholds are reduced to 1/8 of their original values, the residual surface contamination rate is reduced by about 50% for the three ABI water vapor bands, but at the expense of data yields (an average of 66% loss).

The DMR QC is applied to a case study of Hurricane Harvey (2017) to study the scheme's potential impact on assimilation. ABI level-2 total precipitable water products show that surface contamination is most sensitive to atmospheric moisture. Where there is little moisture in the atmosphere, the water vapor band radiances may be sensitive to the surface, such as over high terrain and at high latitudes. Using stricter thresholds again shows reduced data yields. ABI has a fine spatial resolution of 2 km and thus provides significantly more data than most data assimilations can handle. Stricter thresholds may be beneficial as they reduce the residual surface contamination rates, even though the data yield is reduced. Results show that the optimized QC categories to use are 3, 3, and 4 for bands 8, 9, and 10, respectively. This combination provides the largest impact on Hurricane Harvey track forecasts. It was also found that sequentially adding the ABI water vapor bands provided added value. Comparing with the existing GSI and Lee QC schemes showed that the DMR QC scheme is beneficial for track forecasts. When using the DMR QC scheme, the RMSE of the 72-hr track forecast error is reduced by 108 km, or 33% compared with the GSI QC scheme, and by 72 km (or 25%) compared with the Lee QC scheme. The average RMSE reduction over the course of a 72-hr forecast is 33 km, or 21% compared with the GSI QC scheme, and 18 km (or 13%) compared with the Lee QC scheme. The DMR QC scheme also has some slightly positive or neutral impacts on the minimum sea level pressure and the maximum wind speed forecasts. The number of radiances assimilated for the DMR QC scheme is found substantially smaller for all three bands than the GSI and Lee QC schemes. The improved hurricane forecast by the DMR QC is likely a result of the reduced residual surface contamination rate.

These results confirmed that surface contamination QC for ABI water vapor radiances in data assimilation is critically important. It is generally true that assimilating more radiances is beneficial, but the quality of the observations is important. Assimilating water vapor radiances with an increased possibility of surface contamination may reduce or even compromise the impact from these observations. A balance to maximize the positive impact of surface unaffected radiances and minimize the negative impact of surface contaminated radiances is necessary. While results are in favor of the new QC scheme, the impact studies presented are not statistically significant because they are based on one single case demonstration. More extensive assimilation experiments conducted using global NWP models such as the Finite-Volume Cubed-Sphere Global Forecast System will provide more statistically significant results.

The DMR QC scheme could be affected by the inconsistencies between the simulation studies and the application to real data. For example, the radiative transfer modeling uncertainties are unknown in the simulation, the true observation noises are unknown although the on-orbit estimates are a good representative, and the observation radiance biases are unknown. Three steps were carried out in this study to mitigate or minimize the potential impacts from these inconsistencies. First, the pre-launch NEdR is used instead of the on-orbit in the simulation. The larger NEdR partially accounts for the uncertainties in the radiative transfer calculation and the observation. Second, different thresholds are used to generate different QC categories. The numerical experiments are used to find the optimized QC category for each band. And third, the radiance biases between the observations and the radiative transfer simulations are removed before applying the DMR QC scheme.

The DMR QC scheme can be applied to the water vapor radiances of other infrared sensors, such as current polar-orbiting hyperspectral sounders, like the Atmospheric Infra-Red Sounder, the Infrared Atmospheric Sounding Interferometer, and the Cross-track Infrared Sounder, or, in theory, planned operational hyperspectral geostationary sounders. Hyperspectral sounders have many more spectral channels and much improved vertical resolution than the ABI. The correlations between those channels are stronger than ABI's 7 infrared bands. Such correlations are beneficial for both the MLP and RF QC schemes. It is possible that the DMR QC may work

better with hyperspectral infrared sensors. In addition, the DMR QC scheme can be applied to CO_2 channels on hyperspectral infrared sensors.

Data Availability Statement

The ABI data are available from National Oceanic and Atmospheric Administration (NOAA) Comprehensive Large Array-data Stewardship System (NCEI, 2017). The NCEP FNL data are available from NCAR/UCAR Research Data Archive (NCEP GDAS/FNL, 2015). The Global Telecommunication System conventional data and commonly assimilated satellite radiances are available from NCEP ADP Global Upper Air and Surface Weather Observations (2008) and NCEP GDAS Satellite Data 2004-continuing (2009). The best track of Hurricane Harvey (2017) is based on National Hurricane Center (NHC) data archive (NHC, 2017).

Acknowledgments
This work is partly supported by NASA/NOAA ROSES (NA20NES4320003) and SSEC2022 internal funding. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

References

Baum, E. B. (1988). On the capabilities of multilayer perceptrons. *Journal of Complexity*, *4*(3), 193–215. https://doi.org/10.1016/0885-064x(88)90020-9

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Blackwell, W. J., & Chen, F. W. (2009). *Neural networks in atmospheric remote sensing*. Artech House.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Cardinali, C. (2009). Monitoring the observation impact on the short-range forecast. *Quarterly Journal of the Royal Meteorological Society*, *135*(638), 239–250. https://doi.org/10.1002/qj.366

Carminati, F., Migliorini, S., Ingleby, B., Bell, W., Lawrence, H., Newman, S., et al. (2019). Using reference radiosondes to characterise NWP model uncertainty for improved satellite calibration and validation. *Atmospheric Measurement Techniques*, *12*(1), 83–106. https://doi.org/10.5194/amt-12-83-2019

Chen, Y., Han, Y., & Weng, F. (2012). Comparison of two transmittance algorithms in the community radiative transfer model: Application to AVHRR. *Journal of Geophysical Research*, *117*(D6), D06206. https://doi.org/10.1029/2011JD016656

Di, D., Ai, Y., Li, J., Shi, W., & Lu, N. (2016). Geostationary satellite-based 6.7 μm band best water vapor information layer analysis over the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *121*(9), 4600–4613. https://doi.org/10.1002/2016jd024867

Girosi, F., Jones, M., & Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, *7*(2), 219–269. https://doi.org/10.1162/neco.1995.7.2.219

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256). JMLR Workshop and Conference Proceedings.

GOES-R Calibration Working Group and GOES-R Series Program. (2017). *NOAA GOES-R Series Advanced Baseline Imager (ABI) Level 1b Radiances*. NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5BV7DSR

Hinton, G. E. (1990). Connectionist learning procedures. In *Machine learning* (pp. 555–610). Elsevier.

Hong, S.-Y., & Lim, J.-O. J. (2016). The WRF single-moment 6-Class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, *42*(2), 129–151.

Hong, S.-Y., Noah, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13), D13103–D13108. https://doi.org/10.1029/2008JD0099944

Jin, X., Li, J., Schmit, T. J., Li, J., Goldberg, M. D., & Gurka, J. J. (2008). Retrieving clear-sky atmospheric parameters from SEVIRI and ABI infrared radiances. *Journal of Geophysical Research*, *113*(D15), D15310. https://doi.org/10.1029/2008JD010040

Kain, J. S. (2004). The Kain-Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, *43*(1), 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:tkcpau>2.0.co;2

Kazumori, M. (2016). Assimilation of Himawari-8 clear-sky radiance data in JMA's NWP systems. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, *46*. 01.15-01.16.

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. ArXiv:1412.6980 [Cs]. Retrieved from http://arxiv.org/abs/1412.6980

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, *14*, 1137–1145.

Krogh, A., & Hertz, J. A. (1992). A simple weight decay can improve generalization. In *Advances in neural information processing systems* (pp. 950–957).

Lee, J.-R., Li, J., Li, Z., Wang, P., & Li, J. (2019). ABI water vapor radiance assimilation in a regional NWP model by accounting for the surface impact. *Earth and Space Science*, *6*(9), 1652–1666. https://doi.org/10.1029/2019EA000711

Li, J., Fengxian, Z., & Qingcun, Z. (1994). Simultaneous non-linear retrieval of atmospheric temperature and absorbing constituent profiles from satellite infrared sounder radiances. *Advances in Atmospheric Sciences*, *11*(2), 128–138. https://doi.org/10.1007/bf02666541

Li, J., Li, Z., Jin, X., Schmit, T. J., Zhou, L., & Goldberg, M. (2011). Land surface emissivity from high temporal resolution geostationary infrared imager radiances: Methodology and simulation studies. *Journal of Geophysical Research*, *116*(D1), D01304. https://doi.org/10.1029/2010JD014637

Li, J., Li, Z., & Schmit, T. J. (2020). ABI legacy atmospheric profiles and derived products from the GOES-R series. In *The GOES-R series* (pp. 63–77). Elsevier.

Li, Z., Li, J., Gunshor, M., Moeller, S.-C., Schmit, T. J., Yu, F., & McCarty, W. (2019). Homogenized water vapor absorption band radiances from international geostationary satellites. *Geophysical Research Letters*, *46*(17–18), 10599–10608. https://doi.org/10.1029/2019GL083639

Li, Z., Li, J., Jin, T. J. S., Borbas, E., & Goldberg, M. (2010). An objective methodology for infrared land surface emissivity evaluation. *Journal of Geophysical Research*, *115*(D22), D22308. https://doi.org/10.1029/2010JD014249

Li, Z., Li, J., Li, Y., Zhang, Y., Schmit, T. J., Zhou, L., et al. (2012). Determining diurnal variations of land surface emissivity from geostationary satellites. *Journal of Geophysical Research*, *117*(D23), D23302. https://doi.org/10.1029/2012JD018279

Li, Z., Menzel, W. P., Jung, J., Lim, A., Li, J., Matricardi, M., et al. (2020). Improving the understanding of CrIS full spectral resolution nonlocal thermodynamic equilibrium radiances using spectral correlation. *Journal of Geophysical Research: Atmospheres*, *125*(16), e2020JD032710. https://doi.org/10.1029/2020JD032710

Liu, H., Collard, A., Derber, J. C., & Jung, J. A. (2019a). Clear-sky radiance (CSR) assimilation from geostationary infrared imagers at NCEP, international TOVS study Conference XXII, Saint-Sauveur, Québec, Canada.

Liu, H., Collard, A., Derber, J. C., & Jung, J. A. (2019b). Evaluation of GOES-16 clear-sky radiance (CSR) data and preliminary assimilation results at NCEP, 2019 Joint Satellite Conference.

Liu, Z., Min, M., Li, J., Sun, F., Di, D., Ai, Y., et al. (2019). Local severe storm tracking and warning in pre-convection stage from the new generation geostationary weather satellite measurements. *Remote Sensing*, *11*(4), 383. https://doi.org/10.3390/rs11040383

Ma, Z., Maddy, E. S., Zhang, B., Zhu, T., & Boukabara, S. A. (2017). Impact assessment of Himawari-8 AHI data assimilation in NCEP GDAS/GFS with GSI. *Journal of Atmospheric and Oceanic Technology*, *34*(4), 797–815. https://doi.org/10.1175/jtech-d-16-0136.1

Min, M., Bai, C., Guo, J., Sun, F., Liu, C., Wang, F., et al. (2019). Estimating Summertime precipitation from Himawari-8 and global forecast system based on machine learning. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(5), 2557–2570. https://doi.org/10.1109/TGRS.2018.2874950

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, *2*(5–6), 183–197. https://doi.org/10.1016/0925-2312(91)90023-5

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning* (pp. 807–814). ICML-10.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. (2008). updated daily. NCEP ADP Global Upper Air and Surface Weather Observations (PREPBUFR format). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/Z83F-N512

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. (2009). updated daily. NCEP GDAS Satellite Data 2004-continuing. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/DWYZ-Q852

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. (2015). updated daily. NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D65Q4T4Z

National Hurricane Center data archive. (2017). Retrieved from https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Petty, G. W. (2006). *A first course in atmospheric radiation* (pp. 126–127). Sundog.

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. *Machine Learning*, *85*(3), 333–359. https://doi.org/10.1007/s10994-011-5256-5

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebair, W. J. (2017). A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society*, *98*(4), 681–698. https://doi.org/10.1175/BAMS-D-15-00230.1

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., & Bachmeier, A. S. (2005). Introducing the next-generation advanced baseline imager on GOES-R. *Bulletin of the American Meteorological Society*, *86*(8), 1079–1096. https://doi.org/10.1175/bams-86-8-1079

Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., & Huang, H.-L. (2008). Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multi-spectral satellite radiance measurements. *Journal of Applied Meteorology and Climatology*, *47*(1), 108–123. https://doi.org/10.1175/2007JAMC1590.1

Seemann, S. W., Li, J., Menzel, W. P., & Gumley, L. E. (2003). Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances. *Journal of Applied Meteorology and Climatology*, *42*(8), 1072–1091. https://doi.org/10.1117/12.466686

Yin, R., Han, W., Gao, Z., & Li, J. (2021). Impact of high temporal resolution FY-4A Geostationary Interferometric Infrared Sounder (GIIRS) radiance measurements on Typhoon forecasts: Maria (2018) case with GRAPES global 4D-Var assimilation system. *Geophysical Research Letters*, *48*(15), e2021GL093672. https://doi.org/10.1029/2021GL093672

Zhu, Y., Derber, J., Collard, A., Dee, D., Treadon, R., Gayno, G., & Jung, J. A. (2014). Enhanced radiance bias correction in the National centers for environmental Prediction's Gridpoint statistical interpolation data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1479–1492. https://doi.org/10.1002/qj.2233