https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023


Geoscientific Model Development Discussions

# Assimilation of the AMSU-A radiances using the CESM (v2.1.0) and the DART (v9.11.13)/RTTOV (v12.3)

Young-Chan Noh[1], Yonghan Choi[1], Hyo-Jong Song[2], Kevin Raeder[3], Joo-Hong Kim[1], and Youngchae Kwon[2]

[1] Korea Polar Research Institute, Incheon, 21990, South Korea
[2] Department of Environmental Engineering and Energy, Myongji University, Seoul, 17058, South Korea
[3] National Center for Atmospheric Research, CISL/DAReS, Boulder, CO, 80305, USA

*Correspondence to*: Yonghan Choi (yhdchoi@kopri.re.kr)

**Abstract.** To improve the initial condition ("analysis") for numerical weather prediction, we attempt to assimilate observations from the Advanced Microwave Sounding Unit-A (AMSU-A) on board the low-earth-orbiting satellites. The data assimilation system, used in this study, consists of the Data Assimilation Research Testbed (DART) and the Community Earth System Model as the global forecast model. Based on the ensemble Kalman filter scheme, DART supports the radiative transfer model that is used to simulate the satellite radiances from the model state. To make the AMSU-A data available to be assimilated in DART, preprocessing modules are developed, which consist of quality control and bias correction processes. In the quality control, three sub-processes are included: gross quality control, channel selection, and spatial thinning. The bias correction process is divided into scan-bias correction and air-mass-bias correction. As input data used in DART, the observation errors are also estimated for the AMSU-A channels. In the trial experiments, a positive analysis impact is obtained by assimilating the AMSU-A observations on top of the DART data assimilation system that already makes use of the conventional measurements. In particular, the analysis errors are significantly reduced in the whole troposphere and lower stratosphere over the Northern Hemisphere. Overall, this study demonstrates a positive impact on the analysis when the AMSU-A observations are assimilated in the DART assimilation system.

## 1 Introduction

Data assimilation is a numerical procedure for making the initial condition ("analysis") that is used as the starting point for a numerical weather prediction (NWP). In the data assimilation process, various observation data are combined with the short-term forecast ("background") derived from the NWP model, based on the error characteristics of the observations and model forecast (Kalnay, 2003). With the advances in the observation/computation technique and the improved data assimilation methodology, the quality of the initial condition significantly increases, which enhances the forecast skill. In particular, the initial condition has dramatically improved since the satellite observations started to be assimilated (Migliorini et al., 2008; Eyre et al., 2020; Eyre et al., 2022). It is because the satellites cover the regions where the conventional observations are sparse or absent. Among many types of satellite observations being assimilated, a significant forecast benefit mainly comes from the observations of the hyperspectral infrared and microwave sounders that provide unique information on the vertical structure of key atmospheric parameters (e.g., temperature and moisture) (Joo et al., 2013; Eresmaa et al., 2017; Menzel et al., 2018). For this reason, satellite observations are actively being assimilated into the data assimilation system in many operational NWP centers.

To advance the research related to data assimilation, a well-organized data assimilation system is essential, which consists of the forecast model, a data assimilation scheme, and flexible interfaces to use various types of observations. Operational NWP centers have well-constructed assimilation systems to use diverse types of available observations with up-to-date data assimilation schemes. However, researchers, who are not affiliated with the operational NWP centers, are restricted from accessing these data assimilation systems, because these operational NWP systems should be securely managed to

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



provide global weather forecasting to the forecasters and users on time. In addition, as most operational global NWP systems are installed in high-performing computation systems due to the huge computation resources required, it is practically impossible to handle the operational NWP system under the computation environment in which sufficient computation resources are not provided. Thus, a user-friendly global data assimilation system is needed for small numerical modeling communities to attempt challenging studies related to advancing the data assimilation quality.

The National Center for Atmospheric Research (NCAR) has developed an open-source data assimilation tool that is named the Data Assimilation Research Testbed (DART) for data assimilation research, development, and education (Anderson et al., 2009). DART has interfaces to diverse Earth system components (e.g., atmosphere, ocean, and cryosphere) developed by many modeling centers. For instance, the Community Atmospheric Model (CAM), the atmospheric component of the Community Earth System Model (CESM) developed by NCAR, can be used to provide the short-range forecast that is the background field in DART. DART is based on the ensemble data assimilation method instead of the variational method, which requires complicated software specific to a particular numerical prediction model (Anderson et al., 2009; Raeder et al., 2012). In addition, well-defined modules are included to make various types of observations available in the DART data assimilation process. Liu et al. (2012) investigated the impact of the Global Positioning System (GPS) Radio Occultation (RO) observations on the forecast of Hurricane Ernesto (2006) using the DART assimilation system. Coniglio et al. (2019) showed that additional forecast benefit is made by assimilating the measurements of ground-based wind profilers. In addition, a decade-long reanalysis was created with 80 ensemble members derived from DART, using ground-based data, satellite-based winds, GPS-RO observations, and temperature soundings retrieved from the Atmospheric Infrared Sounder (AIRS) observation (Raeder et al., 2021).

However, there are few studies of assimilating satellite-measured radiances in the DART data assimilation system, because the previous version of DART did not have the essential components, e.g., the radiative transfer model (RTM), needed to simulate the satellite radiances from the model state. Fortunately, in the recent version of DART (version 9.11.13), the RTM is included. The Radiative Transfer for TIROS Operational Vertical Sounder (RTTOV) version 12.3 is supported to map the model space into observation space in the data assimilation scheme (Saunders et al., 2018). In Zhou et al. (2022), the visible imagery of the Chinese geosynchronous-orbiting (GEO) satellite was assimilated in DART, but using the Observing System Simulation Experiment (OSSE) framework in which the visible imagery was simulated and then assimilated. Thus, it is of interest to assimilate the satellite-observed radiances using the DART data assimilation system, in order to know how the analysis derived from DART is affected by satellite observations.

Considering the fact that the analysis/forecast impact derived from the satellite radiances mainly comes from observations of hyperspectral infrared and microwave sounders (English et al., 2013; Joo et al., 2013; Kim and Kim, 2019), it is reasonable to assimilate the observations of both sounders first. Unfortunately, the use of hyperspectral infrared sounder observations was not supported in the recent version of DART. For this reason, we attempt to assimilate the radiances of the Advanced Microwave Sounding Unit-A (AMSU-A) temperature sounder within the DART data assimilation system coupled with the NCAR CESM. AMSU-A instruments are currently operating on board many low-earth-orbiting (LEO) satellite platforms, and thus a large amount of AMSU-A observation data is available for assimilation. In addition, as the microwave sounder observations are less sensitive to clouds than the infrared sounder observations, the data availability of AMSU-A is better than that of the infrared sounder. As the preprocessing modules (e.g., quality control, cloud detection, and spatial thinning) for AMSU-A observations are not provided in the DART package, they are developed in this study. In addition, the diagonal observation error covariance matrix is estimated using the method suggested by Desroziers et al. (2005), and the bias correction scheme is also developed based on the methods suggested by Harris and Kelly (2001). To assess the impact of assimilating AMSU-A observations on the analysis derived from DART, the assimilation experiments are conducted using the DART assimilation system coupled with the CESM as the forecast model system.

This paper is organized as follows. Section 2 provides the background information on the DART data assimilation

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



system and CESM. Observation data assimilated in DART are described in section 3. The developed preprocessing steps and the estimated observation errors are presented in sections 4 and 5, respectively. The setup of the assimilation experiments is explained in section 6. The results of the first-guess/analysis departure analysis and the analysis impact are explored in section 7, followed by a summary in section 8.

## 2 DART-CESM data assimilation system

### 2.1 Data Assimilation Research Testbed (DART)

DART is an open-source assimilation package that has been developed by NCAR since 2002 for data assimilation development, research, and education. DART can be coupled with full-complexity Earth system components due to the flexible interfaces provided. In addition, the DART package provides the modules to convert observation data from a variety of native formats, e.g., the Binary Universal Form for the Representation of meteorological data (BUFR) format and the Hierarchical Data Format (HDF), into the input format specified for the DART system (Anderson et al., 2009; Raeder et al., 2012). The recent version of DART (version 9.11.13) is capable of using the RTTOV, a fast RTM, for assimilating visible, infrared, and microwave satellite observations. Provided in RTTOV, many satellite instruments on board the GEO and LEO satellites are also supported in the DART assimilation package, but the hyperspectral infrared sounders, e.g., the Cross-track Infrared Sounder (CrIS) and the Infrared Atmospheric Sounding Interferometer (IASI), are excluded (Hoar et al., 2020). The main data assimilation technique provided by DART is the ensemble Kalman filter (EnKF) in which the forecast error covariance is estimated using short-range ensemble forecasts. The derived forecast error covariance is fully multivariate and depends on the synoptic situation.

### 2.2 Community Earth System Model (CESM)

CESM version 2 (CESM v2.1.0) is used as the model component of the ensemble data assimilation system. CESM2 is the latest generation of coupled climate/earth modeling system developed by NCAR, consisting of the atmosphere, land surface, ocean, sea-ice, land-ice, river, and wave models. These component models can be coupled to exchange states and fluxes (Hurrell et al., 2013; Kay at al., 2015). In this study, atmosphere and land component models are actively coupled, but the ocean component (sea surface temperature) and sea ice coverage are specified by data read from files. As the atmosphere model of CESM2, CAM version 6 (CAM6) is an atmospheric general circulation model (AGCM) with the Finite Volume (FV) dynamical core (Danabasoglu et al., 2020). CAM6 provides the short-term forecast (6-h forecast) of the atmospheric state, which is used as the background state in the DART assimilation scheme. The land model is the Community Land Model version 5 (CLM5). The atmospheric variables are directly updated by the information derived from the observations ingested in the DART assimilation process, while the land state is affected interactively by the updated atmosphere state because the two component models are coupled. The two active models (CAM6 and CLM5) are run with a nominal 1° (1.25° in longitude and 0.95° in latitude) horizontal resolution. CAM6 has 32 vertical levels from the surface level to the top at 3.6 hPa (about 40 km).

## 3 Observations

### 3.1 NCEP PrepBUFR data

The baseline observation data are obtained from the National Centers for Environmental Prediction (NCEP) Automated Data Processing (ADP) global upper air and surface weather observations that are available from the NCAR Research Data Archive (NCAR RDA) (https://rda.ucar.edu/). These data are produced in the PrepBUFR format for assimilation

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023

Geoscientific Model Development Discussions
Open Access

in the diverse NCEP NWP systems, and mainly consist of ground-based observations and satellite-based wind retrievals. The ground-based observations include land and marine surface reports, aircraft reports, radiosonde, and pilot balloon (pibal) measurements, which are transmitted via the Global Telecommunications System (GTS) coordinated by the World Meteorological Organization (WMO). The satellite-based retrievals are provided from the National Environmental Satellite Data and Information Service (NESDIS). They include oceanic wind derived from the Special Sensor Microwave Imager (SSMI) and upper wind from the LEO and GEO satellites. As the NCEP ADP dataset is provided in the BUFR format, it must be converted to the data format available in the DART assimilation system, using the modules provided in the DART data assimilation package.

### 3.2 AMSU-A data

AMSU-A is the microwave temperature sounder that is currently on board diverse sun-synchronous satellite platforms e.g., MetOp satellites (MetOp-A, -B, and -C), three satellites of the National Oceanic and Atmospheric Administration (NOAA), and the National Aeronautics and Space Administration (NASA) research satellite *Aqua*. These three LEO satellite constellations provide near-global coverage, even in data assimilation that has a sub-daily assimilation window; NOAA satellites circle in an early-morning orbit (around 0600 local time), MetOp satellites have a mid-morning orbit (around 0900 local time), and Aqua has an afternoon orbit (around 1300 local time). As a cross-track scanning sounder, the AMSU-A instrument has a total of 15 channels that consist of 12 channels (AMSU-A channels 3–14) over the 50–58 GHz oxygen ($O_2$) absorption band and three window channels (AMSU-A channels 1, 2, and 15) at 23.8, 31.4, and 89 GHz. The instrument measures 30 pixels in each swath with a spatial footprint size of 48 km in nadir. The channels over the $O_2$ absorption band mainly provide information about the vertical structure of tropospheric and stratospheric temperature (Mo, 1999; Goldberg et al., 2001). In this study, observations of AMSU-A instruments on board four LEO satellites (i.e., NOAA-19, Aqua, MetOp-A, and MetOp-B) are assimilated within the DART data assimilation system.

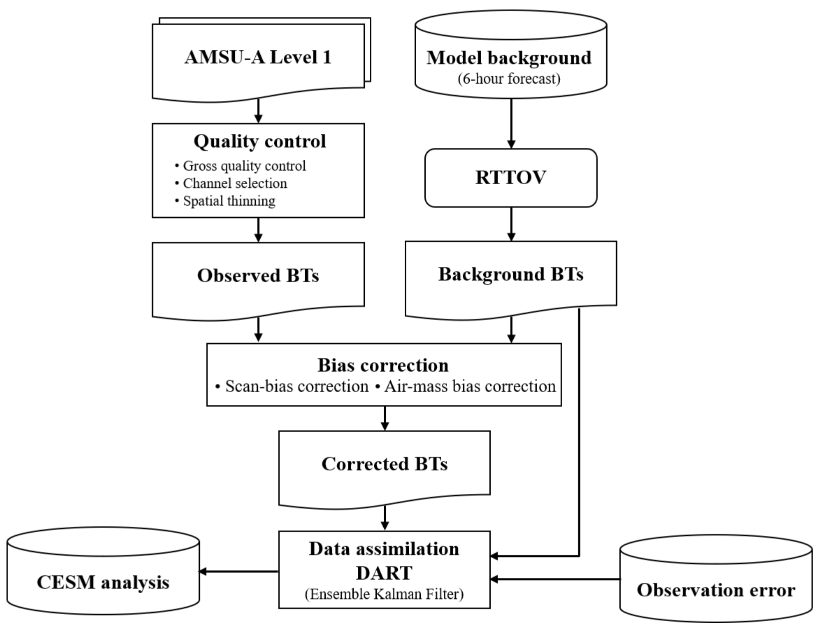


**Figure 1.** Flowchart of the preprocessing system for AMSU-A brightness temperatures (BTs).

## 4 Preprocessing AMSU-A observations

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023


Geoscientific Model Development
Discussions

Prior to assimilating the AMSU-A observations into DART, the AMSU-A observations should be passed through a preprocessing stage. Figure 1 shows the flowchart of the preprocessing stage for the AMSU-A observations as well as the DART assimilation step. In the preprocessing, two main steps are included: quality control and bias correction. Quality control consists of three sub-processes: gross quality control, channel selection, and spatial thinning. If the difference between the observed AMSU-A brightness temperature and the forward-modeled brightness temperature derived from the model background (6-h forecast) is larger than three times the square root of the sum of the observation error variance and the prior background error variance, the AMSU-A observation is not assimilated (called gross quality control). More detailed information on the other two sub-processes (i.e., channel selection and spatial thinning) of the quality control and the bias correction process is described in sections 4.1, 4.2, and 4.3, respectively.

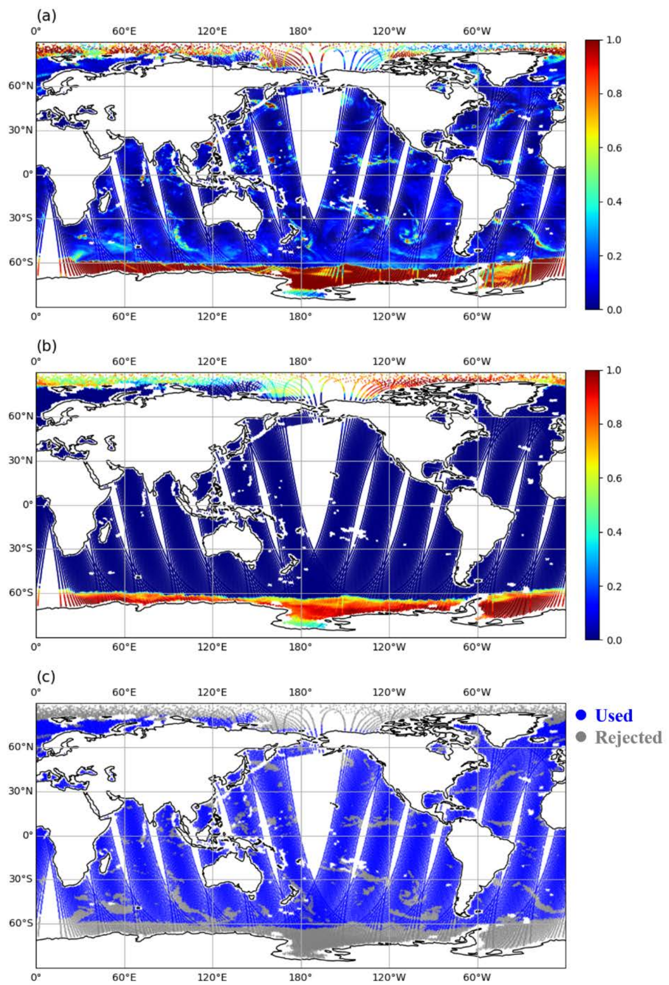


**Figure 2.** Spatial distribution of (a) cloud liquid water (CLW, mm), (b) sea-ice index (SII) retrieved from AMSU-A observations on board NOAA-19, and (c) quality flag of AMSU-A channel 5 (53.6 GHz) from NOAA-19 on 12 August 2014.

### 4.1 Channel selection

As each AMSU-A channel has distinct spectral characteristics, it is necessary to carefully choose the channels to be assimilated in the DART data assimilation system. First, the three AMSU-A channels at 23.8, 31.4, and 89 GHz (i.e., channels 1, 2, and 15), distributed over the window region of the microwave spectrum, are not assimilated. These three window channels

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



are mostly affected by the emitted radiances from the surface under clear-sky conditions, so there is almost no information about the atmosphere. However, AMSU-A channels 1 (23.8 GHz) and 2 (31.4 GHz) are highly sensitive to clouds, so they are used for the quality control in which clouds are detected. In addition, even though the AMSU-A channels 3 (50.3 GHz) and 4 (52.8 GHz) are located over the $O_2$ absorption band used for the temperature sounding, they have a strong sensitivity to the surface, so they are not used in DART. Considering that the upper parts of the weighting function of AMSU-A channels 12 (57.29±0.322±0.022 GHz), 13 (57.29±0.322±0.010 GHz), and 14 (57.29±0.3222±0.0045 GHz) are above the top of the atmosphere (i.e., 3.6 hPa) in the CAM6, these three channels are also removed to prevent vertical interpolation errors that may occur in the forward modeling using the RTM. This leaves channels 5-11 (53.596±0.115, 54.4, 54.94, 55.5, 57.29, 57.29±0.217, and 57.29±0.322±0.048 GHz) as the ones which may be assimilated.

As this study aims to assimilate the AMSU-A observations under the clear-sky condition, the cloud-affected channels are filtered out in the quality control step. In other words, the tropospheric channels (channels 5–7) whose peak of the weighting function is below 200 hPa are rejected if the AMSU-A pixel is determined to be a cloud-affected pixel. To determine this, we calculate the cloud liquid water (CLW) derived from observations of AMSU-A channels 1 and 2 over the ocean, using the retrieval methodology suggested by Grody et al. (2001). The CLW is defined as follows:

$$\mathrm{CLW} = \cos\theta\,[D_0 + 0.754\ln(285.0 - BT_{23}) - 2.265\ln(285.0 - BT_{31})] \quad (1)$$

$$D_0 = 8.240 - (2.622 - 1.846\cos\theta)\cos\theta \quad (2)$$

where $\theta$ is the satellite viewing zenith angle. $BT_{23}$ and $BT_{31}$ are the brightness temperature of AMSU-A channels 1 and 2, respectively. If the retrieved CLW is larger than 0.2 mm, the AMSU-A pixel is judged to be cloud-contaminated, then the three tropospheric channels (channels 5–7) are rejected.

In this study, seven candidate AMSU-A channels (i.e., channels 5–11) are assimilated differently, depending on the surface type. Channels 5, 6, and 7 are the main tropospheric channels. Their weighting functions peak below 200 hPa, but also have a bit of sensitivity to the surface because of the broad vertical shape of the weighting functions. Thus, the quality of the analysis can be degraded by assimilating the three tropospheric channels over the land and sea-ice types whose surface information (e.g., surface temperature and surface spectral emissivity) is uncertain. For this reason, AMSU-A channels 5–7 are not assimilated over the land and sea ice. To identify sea-ice area, the sea-ice index (SII) is retrieved from observations of AMSU-A channels 1 and 3 over the high latitude region (poleward of 50 degrees), using the retrieval algorithm suggested by Grody et al. (1999). The SII is derived as follows:

$$\mathrm{SII} = 2.85 + 0.20\,BT_{23} - 0.028\,BT_{50} \quad (3)$$

where $BT_{50}$ is the brightness temperature of AMSU-A channel 3. Three tropospheric channels are turned off if the SII is larger than 0.1 in the latitudes beyond 50 degrees. However, as the surface information over the ocean is relatively reliable, seven candidate AMSU-A channels are assimilated under the clear-sky condition. The AMSU-A channel list for DART is summarized in Table 1. As an example, Figures 2a and b present the spatial distribution of the CLW and the SII retrieved from AMSU-A on board NOAA-19 on 12 August 2014. It is found that many regions over the ocean are covered by cloud-related systems (CLW > 0.2 mm) and also sea-ice (SII > 0.1) exists near the north and south pole regions. Observations of AMSU-A channel 5 over the cloud region and sea-ice areas are rejected (Fig. 2c). The channel selection process is also applied to the other two AMSU-A channels (channels 6 and 7) which are likely affected by clouds and sea ice. In the pre-trial runs, it was found that the analysis quality is degraded if the AMSU-A observations are assimilated over Antarctica during the Southern Hemisphere

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



winter season. It seems to be due to the complex topography of the Antarctic continent, extreme cold weather conditions, and large errors in the numerical model. Thus, AMSU-A observations are not used over the high latitude region (> 60°S) during the Southern Hemisphere winter season, in order to prevent the degradation of the analysis quality.

**Table 1.** AMSU-A channel list for the DART data assimilation.

<table>
<tr><th>Satellite platform</th><th>Type</th><th>CH5</th><th>CH6</th><th>CH7</th><th>CH8</th><th>CH9</th><th>CH10</th><th>CH11</th></tr>
<tr><td rowspan="3">Aqua</td><td>Land/Sea-ice</td><td rowspan="3">N/A*</td><td>X</td><td rowspan="3">N/A</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Ocean</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Cloud</td><td>X</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td rowspan="3">NOAA-19</td><td>Land/Sea-ice</td><td>X</td><td>X</td><td>X</td><td rowspan="3">N/A</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Ocean</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Cloud</td><td>X</td><td>X</td><td>X</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td rowspan="3">MetOp-A</td><td>Land/Sea-ice</td><td>X</td><td>X</td><td rowspan="3">N/A</td><td rowspan="3">N/A</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Ocean</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Cloud</td><td>X</td><td>X</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td rowspan="3">MetOp-B</td><td>Land/Sea-ice</td><td>X</td><td>X</td><td>X</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Ocean</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
<tr><td>Cloud</td><td>X</td><td>X</td><td>X</td><td>O</td><td>O</td><td>O</td><td>O</td></tr>
</table>

*N/A; not available due to the malfunction in August and September 2014. O; assimilated. X; excluded.

### 4.2 Spatial thinning

In addition to the inter-channel error correlation (refer to section 5), spatial error correlation between the observations at a close distance also exists due to different representativeness of the observed radiances and the model state, and the uncertain quality control process such as cloud detection (Ochotta et al., 2005; Bormann and Bauer, 2010). Thus, the analysis is likely to be sub-optimized if highly dense observations are assimilated without considering the spatial error correlations. A common treatment to counteract the spatial error correlation is spatial thinning which is widely used in data assimilation systems operated by the NWP centers. In this study, the AMSU-A observations are spatially thinned at an interval of about 290 km that was empirically estimated with multiple pre-trial runs.

### 4.3 Bias correction

The biases mainly come from systematic errors: instrument calibration errors, inaccuracies of the RTM, and uncertain preprocessing (e.g. cloud detection errors). The biases tend to change with time (diurnal or seasonal), the scan position of the instrument, and air mass. While random errors are considered by defining the observation errors used in the assimilation process, the biases should be removed before assimilating the satellite observations. In general, the biases are estimated using the time averaged departures between the observed radiances and the simulated radiances from the spatiotemporally collocated model field (background), because of the absence of reference data suitable to compare the satellite observations. To estimate the systematic biases coming from diverse error sources, in this study, two bias correction processes are performed separately: scan-bias correction and air-mass-bias correction, using the statistical bias correction methods suggested by Harris and Kelly (2002).

As a cross-track microwave sounder, the AMSU-A scans 30 field of views (FOVs) per scan line, which are distributed symmetrically about the nadir. The scan angles of 30 FOVs range between ±48.33°. Thus, the observed radiance varies depending on the scan angle even though the observation point is the same. The variation of AMSU-A radiance is due to the

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



change in the optical path length between the earth and the satellite instrument, called the limb effect. The variation of radiance along with the scan angle can be simulated in the RTTOV embedded in DART. However, the mean first-guess departures between the AMSU-A observed radiances and forward-modeled radiances still increase with an increasing scan angle on the center of two near-nadir FOVs (15 and 16) (Fig. 3a), meaning that the residual scan-angle-dependent biases exist for each AMSU-A channel. Thus, the scan-bias correction is required to correct the residual scan bias for each AMSU-A channel. In this study, the scan-bias correction is performed using the pre-computed residual scan bias for each AMSU-A channel. There are two steps to estimate the residual scan bias for AMSU-A channels assimilated. First, the mean bias of the departure between the AMSU-A observed radiances and forward-modeled radiances for each FOV is made with the data assimilation results derived from the pre-trial run. Second, the averaged residual scan bias is obtained by removing the mean bias of two near-nadir FOVs (15 and 16) from the bias of the departure for each FOV (1–30). In addition, as shown in Fig. 3b, it is also found that the residual scan biases have different patterns depending on the latitude band for AMSU-A channel 6 (not shown for other channels), suggesting that the use of globally averaged scan bias is likely to deteriorate the quality of AMSU-A data assimilation. Thus, the residual scan ($b^{scan}$) bias for each AMSU-A channel is subdivided into 15 latitude bands as follows:

$$b_i^{\mathrm{scan}}(\theta,\phi) = [y - H(x_{\mathrm{b}})]_i(\theta,\phi) - [y - H(x_{\mathrm{b}})]_i(\theta = 0,\phi) \tag{4}$$

where the subscript $i$ denotes the AMSU-A channel number ($i$=1, 2, . . . , 15), $\theta$ is the satellite scan angle, $\phi$ is the latitude band at an interval of 10 degrees, $y$ is the AMSU-A radiance, $x_b$ is the background model state, and H is the observation operator. Prior to the air-mass-bias correction, the observed brightness temperatures of each AMSU-A channel are corrected using the estimated scan bias coefficients.

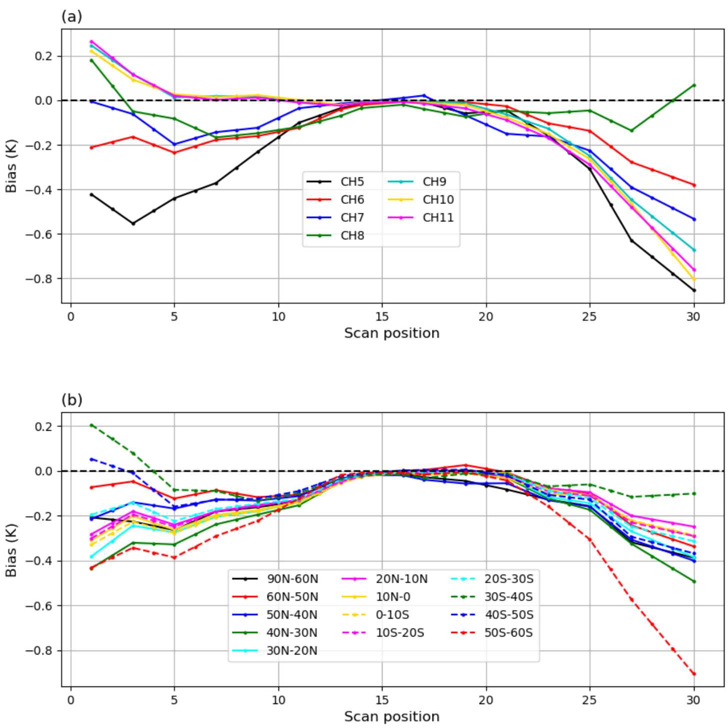


**Figure 3.** (a) Globally averaged, residual scan bias of AMSU-A channels 5–11 and (b) the regionally averaged, residual scan bias depending on 14 latitude bands for AMSU-A channel 6 on board MetOp-B during the period from 11 August to 25 August 2014.

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



The air-mass bias ($b^{airmass}$) is predicted using the multivariate regression method. The biases are mainly due to uncertainties in the RTM, which tend to vary with the air mass and surface characteristics. The predictors, used in the regression method, come from the model variables (i.e., 1000–300 hPa thickness, 200–50 hPa thickness, and surface temperature) that include information on air mass and surface characteristics. The predictors regress to the first-guess departure between the satellite radiances and forward-modeled radiances as follows:

$$b_i^{airmass} = \beta_{i,0} + \sum_{j=1}^{N} \beta_{i,j} p_{i,j} \tag{5}$$

where $\beta_{i,0}$ indicates the constant component of bias $b_i$, and $\beta_{i,j}$ are the bias correction coefficients of the predictor $p_{i,j}$. The subscripts $i$ and $j$ denote the AMSU-A channel number and the predictor number (i.e., $j$ = 1,2…N), respectively.

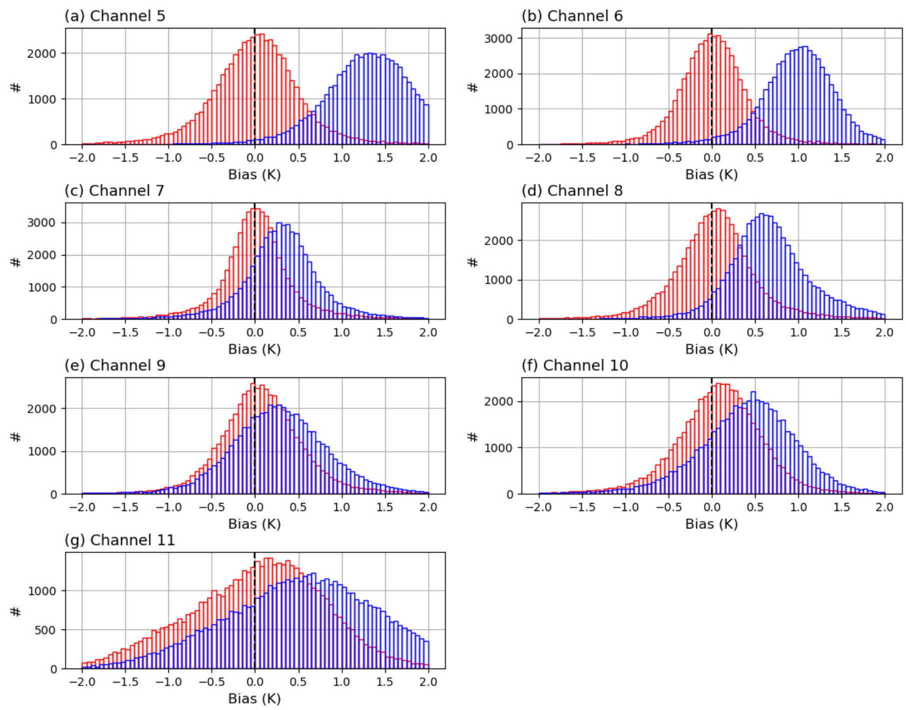


**Figure 4.** Histogram of the first-guess departures between the observations of the MetOp-B AMSU-A channels 5–11 and the corresponding model background (6-h forecast). Colors indicate the results before the bias correction (blue) and after the bias correction (red), respectively.

For the tropospheric AMSU-A channels (channels 5–7), the air mass bias is estimated with two model variables (i.e., 1000–300 hPa thickness and surface temperature), because the peak of the channel weighting function is positioned below the 200 hPa pressure level, and these channels have a bit of sensitivity to the surface. However, 200–50 hPa thickness is only employed for other upper-tropospheric and stratospheric AMSU-A channels (channels 8–11) whose peak of the weighting function is above 200 hPa. As the biases fluctuate with time, it is reasonable to update the regression coefficients and an intercept point periodically, rather than using the climatological-based coefficients that are estimated using the long-term model outputs. In this study, at each data assimilation cycle, the regression coefficients and an intercept point for each AMSU-A channel are computed using DART outputs for the last four cycles and then used to predict the air-mass biases. As shown in Fig. 4, the histograms of the first-guess departures of the MetOp-B channels 5–11 show a positive bias and a Gaussian distribution if the AMSU-A observations are not bias-corrected. In particular, channels 5 and 6 have a large positive bias of

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023

Geoscientific Model Development Discussions

1.0–1.5K. However, the positive biases are almost removed through the bias correction process, meaning that the bias correction scheme works well.

## 5 AMSU-A observation errors

As well as the model background error, the observation errors play an important role in determining the weight of the observations in the data assimilation system. Thus, it is an important step to define the observation errors so that the observations are suitably blended with the model background, which is a 6-hour forecast derived from the CAM6, in order to provide the optimal initial condition to the numerical model. In this study, a diagonal observation error covariance matrix is used for the AMSU-A channels, meaning that the inter-channel error correlation is not considered. In fact, the use of the diagonal observation error covariance matrix may be problematic because the inter-channel error correlation definitely exists for the infrared and microwave sounders (Bormann and Bauer, 2010; Stewart et al., 2014; Weston et al., 2014; Campbell et al., 2017). Unfortunately, the recent version of DART (version 9.11.13) does not support the use of a full observation error covariance matrix in which the diagonal and off-diagonal components are fully defined. For this reason, the diagonal observation errors are empirically inflated to counteract the effect of error correlation between different AMSU-A channels. In other words, the inflated diagonal observation errors take account of the inter-channel error correlation as well.

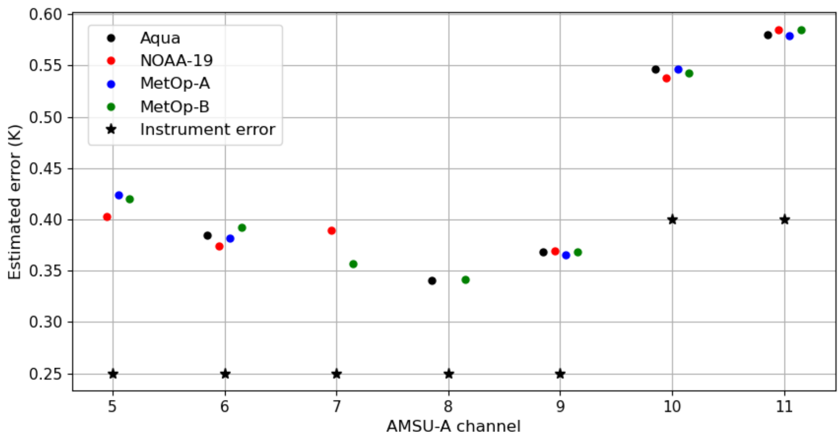


**Figure 5.** Estimated observation errors (K) for AMSU-A channels on board Aqua, NOAA-19, MetOp-A, and MetOp-B satellite platforms. Black asterisks indicate the instrument noise errors for AMSU-A channels.

To estimate the diagonal components (called variances) of the observation error covariance matrix (R) for AMSU-A channels, we use a diagnostic procedure suggested by Desroziers et al. (2005) in which the error variances are made with two departures, i.e., the background innovation (O-B) between the observation (y) and the model background ($x_b$) and the analysis innovation (O-A) between the observation and the model analysis ($x_a$), using the expression in Eq. (6).

$$R = E[\{y - H(x_\mathrm{b})\}\{y - H(x_\mathrm{a})\}^\mathrm{T}] \tag{6}$$

where $E$ is the statistical expectation operator and the superscript "T" indicates the matrix transpose. To compute the observation error variances of AMSU-A channels on board four satellite platforms (i.e., Aqua, NOAA-19, MetOp-A, and MetOp-B), the background and analysis innovations were derived from the pre-trial run. In the pre-trial run, the instrument noise errors were initially used as the observation errors within DART. Then, the observation error variances were estimated

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023

Geoscientific Model Development Discussions

using the Eq. (6).

As the surface-sensitive channels and upper-stratospheric channels are not assimilated in this study (see section 4.1), Figure 5 shows the observation errors of seven AMSU-A channels (channels 5–11) as well as the instrument noise errors employed in the pre-trial run. As some channels (i.e., channels 5 and 7 for Aqua, channel 8 for NOAA-19, and channels 7 and 8 for MetOp-A) malfunctioned during the trial period (11 August – 30 September 2014), the errors for these channels were not needed or estimated. The estimated errors are larger than the instrument noise errors because various error sources (e.g., the radiative transfer modeling errors, representative errors, and systematic errors) are considered as well as the instrument noise errors. The estimated errors for the tropospheric and upper-tropospheric channels (channels 5–9) are smaller than the errors for the stratospheric channels (channels 10–11). This error pattern is also presented for the instrument noise errors. As aforementioned, the estimated observation errors were inflated by a factor of two that was empirically estimated by the multiple pre-trial runs, in order to counteract the inter-channel error correlation. Then, the inflated observation errors, two times the estimated observation errors, were employed for the trial experiments aiming at assessing the analysis impact of assimilating the AMSU-A observations.

## 6 Trial experiment design

To diagnose the analysis impact of assimilating the AMSU-A observations into the DART global data assimilation system, two assimilation experiments were conducted: (a) a control run (CNTL) where the conventional observations (i.e., ground-based observations and satellite-derived winds) were assimilated, and (b) "AMSU-A run", where the AMSU-A observations from four LEO satellite platforms (i.e., Aqua, NOAA-19, MetOp-A, and MetOp-B) were assimilated as well as the conventional data that were assimilated in the CTRL run. For the AMSU-A run, the developed preprocessing steps (e.g., channel selection, thinning, and bias correction) were applied to the AMSU-A observed radiances and then the pre-computed AMSU-A observation errors were employed in the DART data assimilation process.

For two trial runs, available observation data were assimilated within a 6-h assimilation window from -3 to +3 h centered at the nominal analysis time (0000, 0600, 1200, and 1800 UTC). All trial runs were carried out four times a day for the trial period from 0000UTC 11 August to 1800UTC 30 September 2014. The CAM6 forecast model was run with a nominal 1° horizontal resolution (1.25° in longitude and 0.95° in latitude) and 32 vertical levels. The initial ensembles that are available at the NCAR RDA (https://rda.ucar.edu/ datasets/ds345.0/) were obtained from the DART reanalysis. To adjust the effect of initial ensembles, a two-week spin-up period (0000UTC 11 August to 1800UTC 24 August 2014) was included in the trial period. In this study, the ensemble adjustment Kalman filter (EAKF) is applied, which is a variation of the EnKF (Anderson, 2001). Twenty ensemble members were integrated to compute the flow-dependent background error covariance and the correlation between the DART state variables and observations.

All EnKF-based assimilation techniques have the sampling error that is induced by the limited size of the ensemble. In particular, the sampling error is likely to be large when the absolute value of correlation between the DART state variables and the observations is small. To remove the spurious correlation induced by limited ensemble size in DART, the correlation is multiplied by a localization factor that decreases from 1 to 0 with the physical distance between the model state variables and the observations. In DART, the localization half-width can be user-defined, which is half of the distance to where the localization factor is zero. In this study, the horizontal/vertical localization half-width of 0.075 radians was employed to prevent the use of erroneous correlation. However, as the model top height is quite lower than the Earth's horizontal scale, the localization half-width in the vertical is normalized by the user-defined scale height, which is equivalent to one radian. In DART, the difference in scale height between the model top (360 Pa) and the standard surface pressure (101325 Pa) is 5.73. In this study, the normalization scale height of 1.5, a default value in DART, was used, which is assumed to be equal to one radian. Thus, the localization half-width of 0.075 radians is converted into the scale height of 0.11, meaning that the localization

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



cutoff can be an ellipsoid that is flat horizontally.

In addition to the reduction of localization half-width (compared to the default value of 0.15), the sampling error correction algorithm was applied, which uses pre-defined information about the horizontal distribution of the correlation between the model state variables and the observations as a function of ensemble size. Detailed information on the sampling error correction algorithm is described in Anderson (2012).

In addition, the EnKF technique has a risk of underestimation of the ensemble spread, meaning that the ensemble estimates are too confident. If the ensemble spread becomes too small, the observation data are ignored in the data assimilation process, resulting in an ensemble collapse (Anderson et al., 2009; Gharamti et al., 2019). To mitigate the underestimation issue of the ensemble spread, the uncertainty in the ensemble estimate is inflated by linearly moving each ensemble member away from the ensemble mean. It means that the standard deviation of the ensemble spread increases by applying the inflation value in a way that the ensemble mean is unchanged. In DART, the ensemble spread varies spatiotemporally, as a function of the evolving observation network and the chosen inflation algorithm. These experiments use a spatiotemporally varying inflation algorithm. More detailed information on the inflation algorithm adopted in DART is presented in Gharamti et al. (2019).

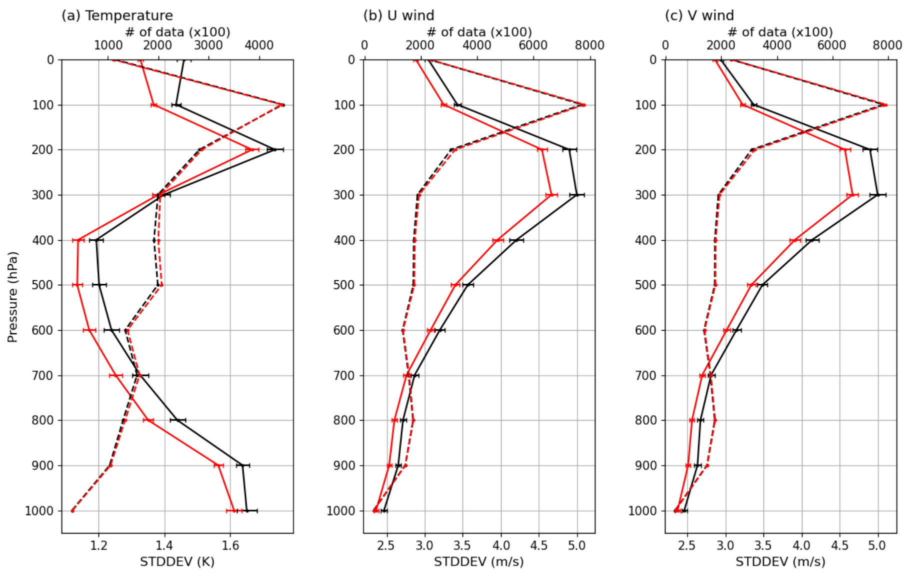


**Figure 6.** The standard deviation (STDDEV) of the first-guess departures for the radiosonde (a) temperature, (b) zonal wind, and (c) meridional wind for the control (CNTL run; black line) and experiment (AMSU-A run; red line) runs. Solid and dashed lines indicate the STDDEV and the number (top axis) of radiosonde measurements assimilated, respectively. Horizontal bars indicate 99% confidence intervals.

## 7 Results

### 7.1 Assessment of first-guess departure and analysis departure

As the same conventional radiosonde measurements were assimilated in the two trial runs (i.e., CNTL and AMSU-A), the first-guess departure statistics between the radiosonde measurements and the spatiotemporally-collocated background states (6-h forecast) can be used to assess the impact of the AMSU-A observations to the short-range forecast. Figure 6 shows the vertical structure of the standard deviation (STDDEV) of the first-guess departure from the radiosonde temperature, zonal wind, and meridional wind as well as the number of the radiosonde measurements used.

For the temperature, the first-guess departure errors are significantly reduced below 300 hPa for the AMSU-A runs as compared with the errors for the CNTL run (Fig. 6a). Because the AMSU-A channels provide vertical information about

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023




the air temperature, the temperature error reduction is the direct impact derived by assimilating the AMSU-A observations in the AMSU-A run. In addition to the radiosonde temperature, the first-guess departure errors decrease for the two wind components (i.e., zonal and meridional winds) (Figs. 6b and c). In particular, the STDDEVs of the two winds at the 300 hPa level are reduced by up to about 4.7m/s in the AMSU-A run, compared to the error of about 5.1m/s for the CNTL run. As the model background error covariance includes the multivariate correlation between different model parameters (e.g., temperature and winds), a change in one model parameter can change another model parameter in the assimilation process. In addition, model parameters are linked in the governing equations and the physical parameterizations, which are embedded in the CAM6. That is, the change in one parameter results in the adjustment of another parameter in the model time integration. Thus, the error reduction of the wind components is the indirect impact of the improved temperature field by assimilating the AMSU-A observations.

In addition to the first-guess departure analysis of radiosonde, the assimilation impact of the AMSU-A observations can be diagnosed by comparing the first-guess departures of the AMSU-A with the analysis departures between the AMSU-A observations and the model analysis state. In general, if the observations are successfully assimilated, the STDDEV of the analysis departure is smaller than that of the first-guess departure, because the background fields are improved by assimilating the observations. As shown in Fig. 7, the STDDEVs of the analysis departure are significantly smaller than that of the first-guess departure for AMSU-A assimilated channels (channels 5–11) regardless of the satellite platforms, meaning that the AMSU-A observations have a positive analysis impact. In particular, the gap between the STDDEVs of two departures is large for the stratospheric AMSU-A channels (channels 9–11).

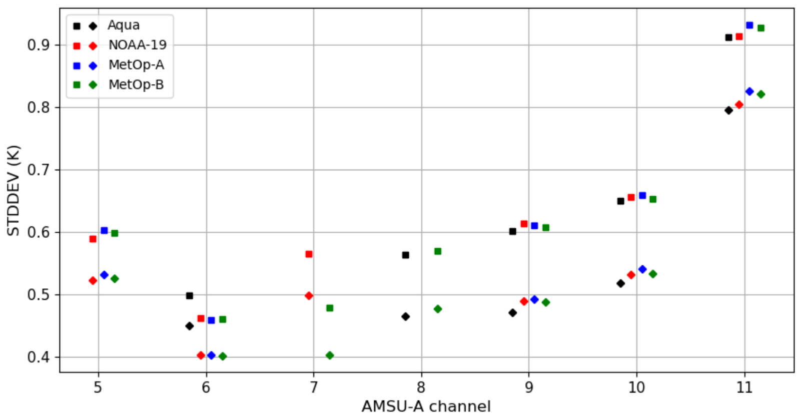


**Figure 7.** The standard deviations (STDDEVs) of the first-guess departure (square) and analysis departure (diamond) for AMSU-A channels on board Aqua, NOAA-19, MetOp-A, and MetOp-B satellites.

### 7.2 Analysis impact of AMSU-A observations

To assess the impact of the AMSU-A observations on the analysis derived from the DART data assimilation system, the analysis errors are computed between the DART analysis and the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis version 5 (ERA5) as the reference data. As the ERA5 is made through the assimilation of all available observation data in the ECMWF data assimilation system and provides consistent maps without spatial gaps, the ERA5 is employed to assess the model-derived output. For four primary atmospheric parameters (i.e., 500 hPa geopotential height, temperature, zonal wind, and meridional wind), the analysis errors are computed. In particular, the skill score of 500 hPa geopotential height is widely used as one of the key indicators to assess the overall performance of the model-derived output, because large-scale atmospheric motion in the middle troposphere (500 hPa) is closely linked with lower-level atmospheric motion.

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023




Figure 8 describes the mean bias and STDDEV of 500 hPa geopotential height for the CNTL and AMSU-A run, depending on the latitudinal regions. Detailed error values are described in Table 2. For two trial runs, overall negative mean bias occurs, reaching up to about -18m. However, the bias difference varies depending on the latitudinal regions. Over the Northern Hemisphere (30°N–90°N), the AMSU-A run has a larger negative bias than the bias for the CNTL run. However, over the tropics (30°S–30°N) and Southern Hemisphere (30°S–90°S), the CNTL run has a larger negative bias than the bias for the AMSU-A run. Thus, similar global mean bias (about -18m) for two trial runs is caused by the offsetting between regionally different bias patterns.

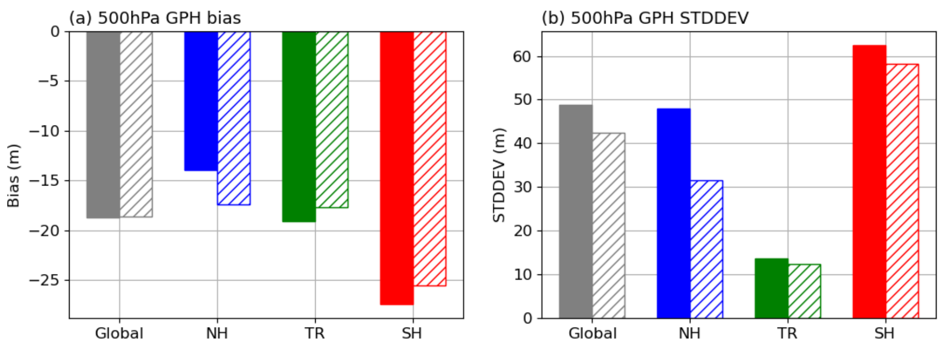


**Figure 8.** (a) Mean bias and (b) standard deviation (STDDEV) of 500 hPa geopotential height over the global (grey), Northern Hemisphere (NH; blue), tropics (TR; green), and Southern Hemisphere (SH; red). Filled and hatched bars indicate the results for the control (CNTL) and experiment (AMSU-A) run, respectively.

Considering that the geopotential height is a primary function of the average air temperature between the surface and the pressure level, we assumed that the model temperature has a cold bias at least below the 500 hPa pressure level. As expected, it is found that a negative bias is presented in the temperature field for both two trial runs (not shown). In addition, as shown in Fig. 9, the first-guess departure of the radiosonde temperature for the two trial runs has large positive values, implying that a cold bias exists in the model temperature fields (6-h forecast). In Raeder et al. (2021), it was noted that the CAM6/DART-derived reanalysis has a cold bias in the troposphere. However, it is still unclear the reason why the CAM6-based temperature fields have a cold bias. The bias issue in CAM6 will be an interesting study in future work.

**Table 2.** Error statistics of 500 hPa geopotential height (m) for the control (CNTL run) and experiment (AMSU-A run) run. Better values are bolded.

| **Trial name** | **Bias** | | | | **STDDEV** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Global** | **NH** | **TR** | **SH** | **Global** | **NH** | **TR** | **SH** |
| CNTL | -18.70 | **-13.90** | -19.05 | -27.45 | 48.82 | 48.02 | 13.55 | 62.54 |
| AMSU-A | **-18.59** | -17.39 | **-17.73** | **-25.51** | **42.42** | **31.55** | **12.41** | **58.29** |

Even though the AMSU-A observations, including the temperature information, are additionally assimilated in the AMSU-A run, the AMSU-A run has a negative temperature bias that occurs in the CNTL run. It is related to the bias correction applied to the AMSU-A observations in DART. As mentioned in section 4.3, the AMSU-A radiances are corrected by eliminating the biases based on the departure between the observed radiances and the forward-simulated radiances from the model background field. In addition, in this study, the bias correction coefficients were even updated at each cycle, using the DART outputs from the last four cycles. Thus, the information on the model bias is included in the biases derived from the

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023


Geoscientific Model Development
Discussions

correction scheme, which gradually fits the observations to the model background over the sequent assimilation cycles. As a result, the model bias still exists in the AMSU-A run as well as the CNTL run.

However, the global-mean STDDEV of 500 hPa geopotential height for the AMSU-A run is reduced to about 42 m as compared with the STDDEV (about 49 m) for the CNTL run, meaning that the 500 hPa geopotential height predictions are improved by assimilating the AMSU-A observations (Table 2). In particular, the error is significantly reduced over the Northern Hemisphere. As shown in Figs. 10a and b, a positive impact mainly occurs in the high-latitude region (> 60°N). In contrast, over the tropics and Southern Hemisphere, the error reduction is relatively smaller than over the Northern Hemisphere. In the tropics, the analysis error (about 14 m) is quite small for the CNTL run, as compared with the large errors of about 48 m and 63 m in the Northern Hemisphere and Southern Hemisphere, respectively. The small STDDEV over the tropics in the CNTL run (shown in Fig. 10a) suggests that the assimilation of the conventional data has brought the model ensembles into an agreement with the AMSU-A observations, so less improvement is there compared to the extratropics.

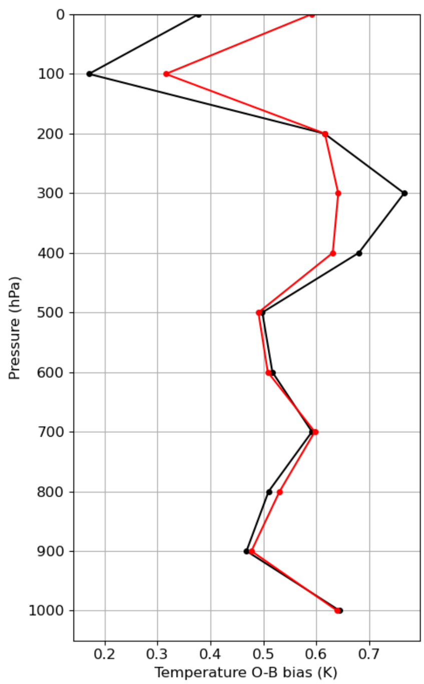


**Figure 9.** Mean bias of the first-guess departure for the radiosonde temperature measurements for the control (CNTL run; black line) and experiment (AMSU-A run; red line) runs.

It is noted that the AMSU-A assimilation impact is neutral in the high-latitude region (> 60°S) over the Southern Hemisphere. In contrast, in the high-latitude region (> 60°N) over the Northern Hemisphere, the assimilation impact is significant. It is because the AMSU-A observations were not assimilated in the high latitude region (> 60°S) over the Southern Hemisphere during the Southern Hemisphere winter season when the trial runs were conducted (mentioned in section 4.1), resulting in the neutral analysis impact. Thus, if the high-latitude regions (i.e., 60°S-90°S and 60°N-90°N) are extracted in the error computation over both hemispheres, the assimilation impact is comparable (not shown). It is still a challenging issue to assimilate the satellite radiances over the Antarctic continent, because of the complex topography, extreme weather condition, and large errors in the numerical model. In particular, as the conventional observations are quite sparse in the high latitude region, the model errors are relatively larger than the other latitudinal regions (i.e., the tropics and mid-latitude region, shown

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023

Geoscientific Model Development
Discussions
Open Access

in Fig. 10a). In addition, the trial period (11 August – 30 September 2014) is the Southern Hemisphere winter season when the Antarctic continent was under extremely cold weather conditions. In fact, in the pre-trial run, we found that the analysis field was degraded near the Antarctic continent by assimilating the AMSU-A observations. Thus, to prevent the analysis degradation, the AMSU-A observations were rejected over the high latitude region (> 60°S) in the Southern Hemisphere. The assimilation of the AMSU-A observation in the Antarctic region will be handled in future work.

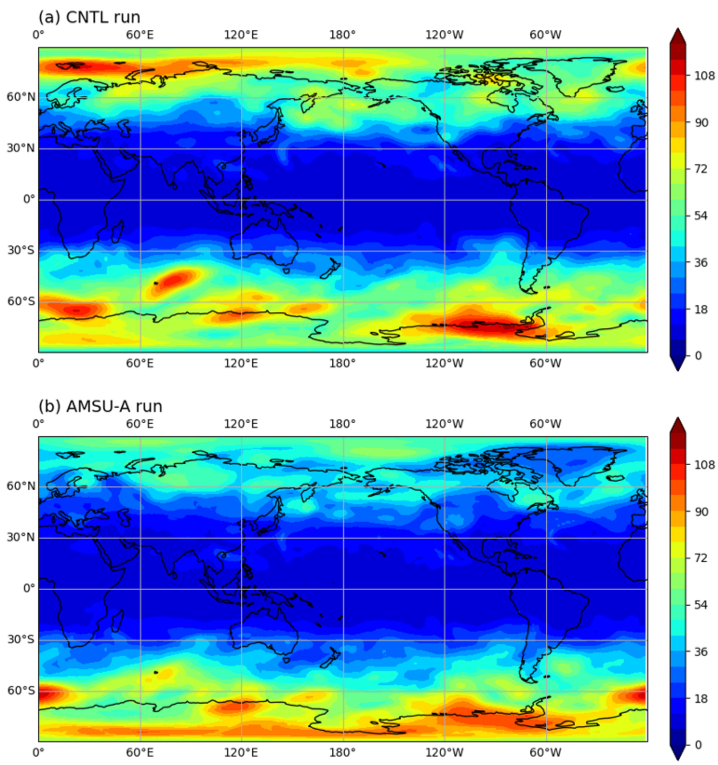


**Figure 10.** Spatial distribution of the standard deviation (STDDEV) of the 500 hPa geopotential height for the (a) control run (CNTL) and (b) experiment (AMSU-A) runs.

Figure 11 shows the normalized difference of STDDEV of temperature, zonal wind, and meridional wind between the AMSU-A run and CNTL run, depending on the latitudinal regions (i.e., global, Northern/Southern Hemispheres, and tropics). The STDDEV difference is normalized by the STDDEV for the CNTL run. A negative value means that assimilating the AMSU-A observations provide analysis benefit. In contrast, a positive value indicates that the analysis error increases for the AMSU-A run compared with the error for the CNTL run, implying a negative analysis impact of the AMSU-A observations.

For the temperature, the global-mean analysis errors are significantly reduced in the whole troposphere and lower stratosphere for the AMSU-A run, as compared with the CNTL run. Large error reduction occurs in the lower stratosphere (-28% and -21% in 100 hPa and 200 hPa, respectively), which is consistent with the large gap between the STDDEVs of the first-guess departure and the analysis departure for the stratospheric AMSU-A channels (channels 9–11) whose peak of the weighting function is above 200 hPa (shown in Fig. 7). Similar to the results of the 500 hPa geopotential height, a strong error reduction mainly occurs in the Northern Hemisphere where the error reduces up to about 28% in the 500 hPa pressure level (Fig. 11a). The error decrease trends are consistent with the trends of the first-guess departure errors of the radiosonde temperature measurements in which a significant error decrease occurs in the 500 hPa layer (Fig. 6a). However, in the lower

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023



stratosphere (100 hPa pressure level), the analysis error decreases up to about 45% in the Southern Hemisphere.

For two wind components (i.e., zonal and meridional winds), similar to the results of the temperature, the global-mean analysis errors for the AMSU-A run overall decrease in the whole troposphere and lower stratosphere (Figs. 11b and c). It is noted that the magnitude of the error decrease tends to increase with height, reaching about -13% in the 100 hPa for the zonal and meridional wind. Moreover, most analysis impact is made in the Northern Hemisphere, except in the 100 hPa where the maximum error decrease occurs in the Southern Hemisphere. However, over the Southern Hemisphere, the analysis errors for the AMSU-A runs are larger than the errors for the CNTL run in the middle and lower troposphere. For the spatial pattern of the STDDEV of two wind components (not shown), it is found that the error increment mainly occurs in the high latitude region (> 60°S) where the AMSU-A data were not assimilated for the AMSU-A run. Considering that the temperature fields above the latitude of 60°S were only updated by the AMSU-A assimilation, the analysis degradation is possibly due to the discontinuity of the latitudinal temperature gradient near the latitude of 60°S.

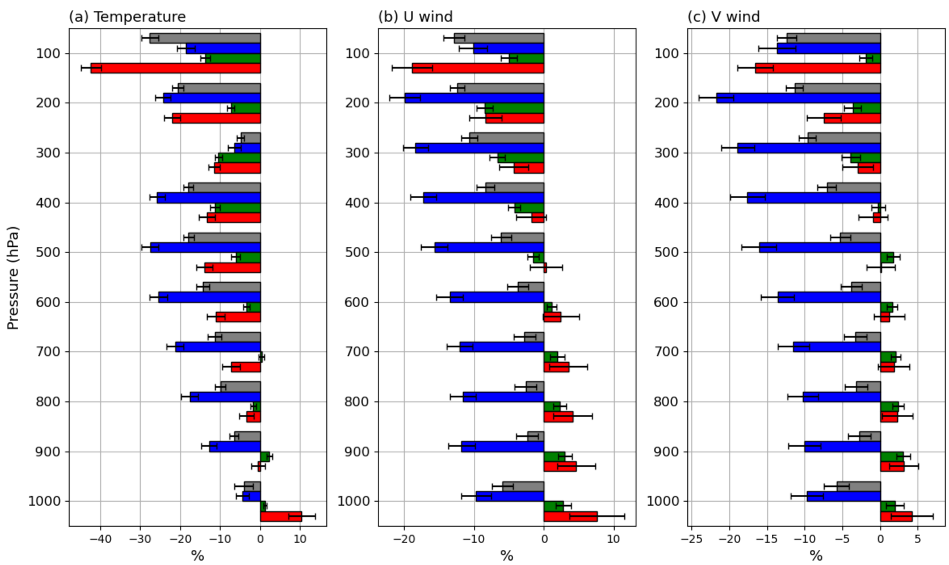


**Figure 11.** Normalized difference of the standard deviation (STDDEV) of (a) temperature, (b) zonal wind, and (c) meridional wind between the experiment (AMSU-A) run and the control (CNTL) run. Colors indicate the latitude regions (global; grey, Northern Hemisphere; blue, tropics; green, and Southern Hemisphere; red). Horizontal bars indicate 99% confidence intervals.

## 8 Summary

In this study, we attempted to assimilate the AMSU-A observations using the global data assimilation system consisting of DART and CESM. To make the AMSU-A data available to be assimilated, preprocessing steps were developed, which include quality control (i.e., gross quality control, channel selection and spatial thinning) and bias correction (i.e., scan-bias correction and air-mass-bias correction). In addition, the observation error covariance matrix was estimated, but only its diagonal components were employed in DART because the inter-channel error correlation is not considered in the current version of DART. To counteract the inter-channel error correlation, the diagonal components were inflated.

To assess the impact of the AMSU-A observations on the DART-derived analysis, trial experiments were conducted from 11 August to 30 September 2014. The derived analysis fields were verified using the ERA5 as the reference. For the primary atmospheric parameters (i.e., 500 hPa geopotential height, temperature, zonal wind, and meridional wind), an additional analysis benefit is provided by assimilating the AMSU-A observations on top of the DART data assimilation system which already makes use of the conventional ground-based observations. In particular, a large analysis impact is shown in the

Northern Hemisphere where the analysis errors of the temperature and two wind components are significantly reduced in the whole troposphere. However, in the tropics, the analysis impact is relatively small due to the small model errors. Compared with the Northern Hemisphere, the number of assimilated AMSU-A data is small over the Southern Hemisphere, because the AMSU-A data are not assimilated in the harsh condition of high latitude regions (> 60°S) during the Southern Hemisphere winter season, resulting in a relatively small analysis impact over the Southern Hemisphere.

**Code and data availability.** DART version 9.11.13 is available at https://github.com/NCAR/DART. CESM version 2.1.0 is released at https://github.com/ESCOMP/CESM/tree/release-cesm2.1.0. Atmospheric initial conditions and the baseline observations at the BUFR format are obtained from the NCAR RDA (https://rda.ucar.edu). AMSU-A observations from Aqua satellite are downloaded via the NASA data center (https://www.earthdata.nasa.gov). AMSU-A observations from NOAA-19, MetOp-A, and MetOp-B satellites are offered from the EUMETSAT data store (https://www.eumetsat.int/eumetsat-data-store). The ECMWF ERA5 is available at the Climate Data Store (https://cds.climate.copernicus.eu). As well as the software codes, the model outputs are available at https://doi.org/10.5281/zenodo.7714755.

**Author contributions.** YN and YC conceptualized the research idea. YN and YC developed the methods with assistance from HS and YK. YN led the writing of the manuscript with support from YC, HS, and KR. YC, HS, KR, and JK involved in writing the final version of the manuscript, whereas YK provided feedback on it.

**Competing interests.** The authors declare no conflicts of interest.

**Acknowledgements.** This project is sponsored by Korea Polar Research Institute (KOPRI) grant funded by the Ministry of Oceans and Fisheries (KOPRI PE23010). NCAR is supported by the U.S. National Science Foundation (NSF). Any opinions expressed here are not necessarily those of NCAR or the NSF. Hyo-Jong Song and Youngchae Kwon are supported by Korea Environment Industry & Technology Institute (KEITI) through “Climate Change R&D Project for New Climate Regime”, funded by Korea Ministry of Environment (MOE) (2022003560006).

Geoscientific Model Development
Discussions
Open Access
EGU

**References**

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Mon. Weather Rev., 129, 2884-2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2, 2001.

Anderson, J. L.: Localization and sampling error correction in ensemble Kalman filter data assimilation, Mon. Weather Rev., 140, 2359-2371, doi:10.1175/MWR-D-11-00013.1, 2012.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A.: The data assimilation research testbed: A community facility, B. Am. Meteorol. Soc., 90, 1283-1296, doi:10.1175/2009BAMS2618.1, 2009.

Bormann, N., and Bauer, P.: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data, Q. J. Roy. Meteor. Soc., 136, 1036-1050, doi:10.1002/qj.616, 2010.

Campbell, W. F., Satterfield, E. A., Ruston, B., and Baker, N. L.: Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system, Mon. Weather Rev., 145, 1019-1032, doi:10.1175/MWR-D-16-0240.1, 2017.

Coniglio, M. C., Romine, G. S., Turner, D. D., and Torn, R. D.: Impacts of targeted AERI and Doppler lidar wind retrievals on short-term forecasts of the initiation and early evolution of thunderstorms, Mon. Weather Rev., 147, 1149-1170., doi: 10.1175/MWR-D-18-0351.1, 2019.

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al.: The community earth system model version 2 (CESM2), J. Adv. Model. Earth Sy., 12, 1-35, doi:10.1029/2019MS001916, 2020.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. Roy. Meteor. Soc., 131, 3385-3396, doi:10.1256/qj.05.108, 2005.

English, S., McNally, T., Bormann, N., Salonen, K., Matricardi, M., Horanyi, A., Rennie, M., Janisková, M., Michele, S. D., Geer, A., Tomaso E. D., Cardinali, C., Rosnay, P., Sabater, J. M., Bonavita, M., Albergel, C., Engelen, R., Thépaut, J.: Impact of Satellite Data, ECMWF Technical Memorandum, 711, ECMWF Reading, UK, 2013.

Eresmaa, R., Letertre-Danczak, J., Lupu, C., Bormann, N., and McNally, A. P.: The assimilation of Cross-track Infrared Sounder radiances at ECMWF, Q. J. Roy. Meteor. Soc., 143, 3177-3188, doi:10.1002/qj.3171, 2017.

Eyre, J. R., Bell, W., Cotton, J., English, S. J., Forsythe, M., Healy, S. B., and Pavelin, E. G.: Assimilation of satellite data in numerical weather prediction. Part II: Recent years, Q. J. Roy. Meteor. Soc., 148, 521-556, doi:10.1002/qj.4228, 2022.

Eyre, J. R., English, S. J., and Forsythe, M.: Assimilation of satellite data in numerical weather prediction. Part I: The early years, Q. J. Roy. Meteor. Soc., 146, 49-68, doi:10.1002/qj.3654, 2020.

El Gharamti, M., Raeder, K., Anderson, J., and Wang, X.: Comparing adaptive prior and posterior inflation for ensemble filters using an atmospheric general circulation model, Mon. Weather Rev., 147, 2535-2553, doi:10.1175/MWR-D-18-0389.1, 2019.

Goldberg, M. D., Crosby, D. S., and Zhou, L.: The limb adjustment of AMSU-A observations: Methodology and validation, J. Appl. Meteorol. Clim., 40, 70-83, doi:10.1175/1520-0450(2001)040<0070:TLAOAA>2.0.CO;2, 2001.

Grody, N., Weng, F., and Ferraro, R.: Application of AMSU for obtaining water vapor, cloud liquid water, precipitation, snow cover, and sea ice concentration, 10th International TOVS Study Conference, International TOVS Working Group (ITWG), 1999.

Grody, N., Zhao, J., Ferraro, R., Weng, F., and Boers, R.: Determination of precipitable water and cloud liquid water over oceans from the NOAA 15 advanced microwave sounding unit, J. Geophys. Res., 106, 2943-2953, doi: 10.1029/2000JD900616, 2001.

Harris, B. A. and Kelly, G.: A satellite radiance-bias correction scheme for data assimilation, Q. J. Roy. Meteor. Soc., 127, 1453-1468, doi:10.1002/qj.49712757418, 2001.

Hoar, T. J., Raeder, K., Anderson, J. L., Steward, J., El Gharamti, M., Johnson, B. K., Romine, G., Ha, S., and Mizzi, A. P.:

https://doi.org/10.5194/gmd-2023-60
Preprint. Discussion started: 26 April 2023

Geoscientific Model Development Discussions
Open Access

DART: Empowering Geoscience with Improved Ensemble Data Assimilation, 2020 AGU Fall Meeting, American Geophysical Union, 2020.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P.J., et al.: The community earth system model: a framework for collaborative research, B. Am. Meteorol. Soc., 94, 1339-1360, doi:10.1175/BAMS-D-12-00121.1, 2013.

Joo, S., Eyre, J., and Marriott, R.: The impact of MetOp and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method, Mon. Weather Rev., 141, 3331-3342, doi:10.1175/MWR-D-12-00232.1, 2013.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, 2003.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, B. Am. Meteorol. Soc., 96, 1333-1349, doi:10.1175/BAMS-D-13-00255.1, 2015.

Kim, S. M. and Kim, H. M.: Forecast sensitivity observation impact in the 4DVAR and hybrid-4DVAR data assimilation systems, J. Atmos. Ocean. Tech., 36, 1563-1575, doi:10.1175/JTECH-D-18-0240.1, 2019.

Liu, H., Anderson, J., and Kuo, Y. H.: Improved analyses and forecasts of Hurricane Ernesto's genesis using radio occultation data in an ensemble filter assimilation system, Mon. Weather Rev., 140, 151-166, doi:10.1175/MWR-D-11-00024.1, 2012.

Menzel, W. P., Schmit, T. J., Zhang, P., and Li, J.: Satellite-based atmospheric infrared sounder development and applications, B. Am. Meteorol. Soc., 99, 583-603, doi:10.1175/BAMS-D-16-0293.1, 2018.

Migliorini, S., Piccolo, C., and Rodgers, C. D.: Use of the information content in satellite measurements for an efficient interface to data assimilation, Mon. Weather Rev., 136, 2633-2650, doi:10.1175/2007MWR2236.1, 2008.

Mo, T.: AMSU-A antenna pattern corrections, IEEE Trans. Geosci. Remote Sens., 37, 103-112, doi: 10.1109/36.739131,1999.

Ochotta, T., Gebhardt, C., Saupe, D., and Wergen, W.: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods, Q. J. Roy. Meteor. Soc., 131, 3427-3437, doi:10.1256/qj.05.94, 2005.

Raeder, K., Anderson, J. L., Collins, N., Hoar, T. J., Kay, J. E., Lauritzen, P. H., and Pincus, R.: DART/CAM: An ensemble data assimilation system for CESM atmospheric models, J. Clim., 25, 6304-6317, doi:10.1175/JCLI-D-11-00395.1, 2012.

Raeder, K., Hoar, T. J., El Gharamti, M., Johnson, B. K., Collins, N., Anderson, J. L., Steward, J., and Coady, M.: A new CAM6+ DART reanalysis with surface forcing from CAM6 to other CESM models, Sci. Rep., 11, 1-24, doi: 10.1038/s41598-021-92927-0, 2021.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Gear, A., Bormann, N., and Lupu, C.: An update on the RTTOV fast radiative transfer model (currently at version 12), Geosci. Model Dev., 11, 2717-2737, doi:10.5194/gmd-11-2717-2018, 2018.

Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R., and Cameron, J.: Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system, Q. J. Roy. Meteor. Soc., 140, 1236-1244, doi:10.1002/qj.2211, 2014.

Weston, P. P., Bell, W., and Eyre, J. R.: Accounting for correlated error in the assimilation of high-resolution sounder data, Q. J. Roy. Meteor. Soc., 140, 2420-2429, dpi:10.1002/qj.2306, 2014.

Zhou, Y., Liu, Y., Huo, Z., and Li, Y.: A preliminary evaluation of FY-4A visible radiance data assimilation by the WRF (ARW v4. 1.1)/DART (Manhattan release v9. 8.0)-RTTOV (v12. 3) system for a tropical storm case, Geosci. Model Dev., 15, 7397-7420, doi: 10.5194/gmd-15-7397-2022, 2022.