https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

Revisiting error models for the assimilation of all-sky infrared satellite radiances

Bingying Shi[1,2], Philipp Griewank[1], Florian Meier[3], Jinzhong Min[2], and Martin Weissmann[1]

[1]Institut für Meteorologie und Geophysik, Universität Wien, Vienna 1090, Austria
[2]School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3]GeoSphere Austria, Vienna 1190, Austria

Correspondence: Philipp Griewank (philipp.griewank@univie.ac.at)

Abstract.

The sensitivity of all-sky infrared radiances to both hydrometeor content and cloud height leads to a very non-Gaussian distribution of first-guess departures. This non-Gaussianity can be mitigated by the application of cloud-dependent error models that normalize departures by an estimate of the cloud-height effect via assigning increased errors in situations with high clouds that can lead to very large departures. In the current study, we systematically evaluate existing error models and propose a revised approach that leads to a better fit to a Gaussian distribution at no additional cost. Furthermore, the revised approach is physically better justified as the cloud effect is estimated by the maximum cloud effect of model and observations, which determines the largest possible departure. Preceding studies, in contrast, used the mean cloud effect of model and observations.

Our study is based on a one-month data set of infrared observations in two water vapor channels (6.2 and 7.3 μm) from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation satellite and corresponding simulations from the weather forecast model AROME (Application of Research to Operations at Mesoscale) over central Europe. The evaluation of different approaches revealed that near-Gaussian departures can be achieved with three different approaches for the estimation of the cloud effect: (1) deviation from a climatologically estimated value; (2) deviation from the clear-sky brightness temperature of a window channel; and (3) deviation from the clear-sky brightness temperature of the channel that is assimilated. For the lower-peaking channel (7.3 μm) with a larger cloud effect, best results were achieved with the first two options. For the 6.2 μm channel, the third option led to a slightly more Gaussian distribution. The third option, however, requires a quality control that eliminates about 10% of the observations.

1 Introduction

Satellite observations cover the whole globe and can provide crucial information on the atmospheric state in areas where conventional data are sparse. Nowadays, satellites have become the dominant source of observational information in global numerical weather prediction (NWP) models and their assimilation leads to significantly improved forecast skill (Eyre et al., 2020, 2022). Recent studies also showed significant improvements of assimilating satellite radiances in regional NWP models as they can provide crucial information on atmospheric clouds and convection (Zhu et al., 2023; Kugler and Weissmann, 2025b; Yang et al., 2025; Schomburg et al., 2026).

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

Traditionally, the assimilation of satellite observations has been restricted to clear-sky conditions due to, inter alia, systematic errors in the model representation of clouds (Geiss et al., 2021; Hu et al., 2023) and strong observation operator non-linearity (Kugler and Weissmann, 2025a). But as cloudy regions represent meteorologically particularly sensitive areas (McNally, 2002), substantial efforts have been made for developing all-sky assimilation techniques over the past decades. Owing to the relatively small degree of nonlinearity of microwave observations in cloudy conditions, early developments concentrated on all-sky microwave data assimilation (e.g., Geer et al., 2014; Yang et al., 2016; Shimizu et al., 2020; Candy and Migliorini, 2021). At some operational centers, such as the European Centre for Medium-range Weather Forecasts (Geer et al., 2017) and the National Centers for Environmental Prediction (Zhu et al., 2016), all-sky microwave radiances are nowadays assimilated operationally. More recent studies also documented benefits of all-sky infrared assimilation that involves larger non-linearity as clouds are mostly opaque in the infrared spectrum and the observed brightness temperature is strongly dependent on cloud height (Geer et al., 2019; Honda et al., 2018; Kugler et al., 2023; Schomburg et al., 2026).

Despite the success of operational all-sky microwave assimilation and demonstrated benefits of assimilating all-sky infrared radiances, the operational application of all-sky infrared assimilation is still quite limited (Li et al., 2022). One of the major challenges is the non-Gaussianity in the observation minus first-guess (FG departure) distribution (Gustafsson et al., 2018), implying that either observation and/or background errors are non-Gaussian (Desroziers et al., 2005; Pires et al., 2010). This violates a fundamental assumption used in most data assimilation schemes (Bocquet et al., 2010).

The non-Gaussian distribution of the departures is mainly caused by the sensitivity of brightness temperatures (BTs) to cloud height, in addition to their sensitivity to hydrometeor content and size distribution. This means that hydrometeor errors at high levels cause much larger departures than hydrometeor errors at lower levels. A commonly used approach to mitigate this non-Gaussianity, is to use a cloud-dependent observation error model that intends to normalize the departures by the estimated cloud-height effect via an adaptive assigned observation error. The underlying idea of this approach is that samples of BTs with a similar cloud effect have a Gaussian distribution of errors. Thus, after normalizing the FG departures using the cloud-dependent observation error, the distribution becomes nearly Gaussian (Hu et al., 2025). This method was first proposed for all-sky microwave radiances assimilation by Geer and Bauer (2011), who introduced a cloud effect predictor C based on the normalized 37 GHz polarization difference between the observation and its model equivalent. Subsequently, the observation error as a function of C was estimated from a climatology of first-guess departures. Following this study, Okamoto et al. (2014) (hereafter O_{14}), and Harnisch et al. (2016) (hereafter H_{16}) proposed similar observation error models for infrared radiances. Both approaches are based on the idea introduced by Geer and Bauer (2011), namely estimating the observation error as a function of the cloud effect averaged between observations and model equivalents. H_{16} introduced a climatologically determined limiting BT ($\mathrm{BT}_{\mathrm{lim}}$) and determined the cloud effect from deviations of observed and modeled BT from this $\mathrm{BT}_{\mathrm{lim}}$. In contrast, O_{14} determined the cloud effect from the deviations of observations and model equivalents from the model clear-sky BT that is obtained by using the observation operator without hydrometeors. Using the clear-sky BT has the advantage that it is aware of the current temperature and humidity profile in contrast to a climatologically determined $\mathrm{BT}_{\mathrm{lim}}$. The model state may, however, not be representative of the observed state. Consequently, model temperature and humidity errors can influence the estimate of the cloud effect. Both approaches can effectively mitigate non-Gaussianity as documented in the two original

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

studies and number of follow on studies that applied either the O_{14} method (e.g., Okamoto et al., 2021, 2024) or the H_{16} method (e.g., Xu et al., 2021; Lu and Wang, 2025) to infrared BTs.

As described by Geer and Bauer (2011), the purpose of determining C based on both the observations and model equivalents is to avoid a sampling bias that could occur when using only one data source. While the use of both data sources is obviously well-justified, we raise the question if averaging both is the best approach and proposed a revised approach that is based on the maximum cloud effect in both data sources, which is proportional to the largest possible departure. Additionally, some details of the two methods have not been investigated thoroughly so far. Firstly, although Okamoto et al. (2014) and Okamoto (2017) pointed out that necessary tuning and quality control (QC) are needed when using O_{14}, the motivation is still unclear, and some studies did not apply either of them (e.g., Hastuti and Min, 2023; Yang et al., 2025). Secondly, to our knowledge, there is no study directly comparing the different effects of the two approaches and the two flavors of O_{14} using the assimilated channel or a window channel for the estimate of the cloud effect. Lastly, the previous studies have not systematically quantified how close to Gaussianity the normalized FG departures are.

The main goals of this paper are:

1. To improve our understanding of the existing two cloud-dependent observation error models and their different flavors.
2. To propose a revised version of the two approaches.
3. To systematically evaluate all different approaches.

Our study is based on a 1-month dataset of infrared BT observations of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard Meteosat Second Generation (MSG). The limited-area numerical weather prediction (NWP) model AROME (Application of Research to Operations at MEsoscale) of GeoSphere Austria covering Central Europe is used to generate all-sky model equivalent BTs. To accomplish the goals, we revisit the O_{14} and H_{16} methods in full detail and specify necessary procedures for each method. Subsequently, we propose a revised version of the two methods that replaces the average C with the maximum C. We assess the effects of all methods on FG departures normalization by quantifying their distance and similarity to a Gaussian distribution. In addition, a bootstrap training and evaluation is conducted to provide a statistical assessment of their performance in terms of transforming the FG departure distribution to a near-Gaussian distribution.

Section 2 presents a description of the used observations, the AROME-Austria model and the simulated model equivalent BT. In Sect. 3, we assess the results of different existing cloud-dependent observation error models for all-sky infrared data. In Sect. 4, we propose an alternative way to estimate observation error by using the maximum of cloud effects from the model and the observations. Section 5 illustrates the statistical comparisons among different approaches. Lastly, we summarize the main findings in Sect. 6.

2 Data and methods

In this section, we describe the observation and model data as well as the two metrics used for assessing the Gaussianity of the departures.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

2.1 SEVIRI observations

Our study uses a one-month dataset of infrared BT observations in two water vapor channels and one window channel from the SEVIRI instrument onboard the geostationary MSG satellite in August 2023. The SEVIRI instrument on board the MSG satellite is a multi-spectral radiometer mainly utilized for the monitoring of atmospheric information every 15 minutes. It contains 12 channels covering wavelengths from 0.6 µm (visible) to 13.4 µm (infrared), with a nadir resolution of 3 km (Schmetz et al., 2002). In this paper, we use the observed BT (BT^{O}) in a domain over Central Europe (Figure 1) of two infrared channels at 6.2 µm (Ch6.2) and 7.3 µm (Ch7.3) that are sensitive to water vapor and only very weakly influenced by surface emission. Additionally, we used the window channel at 10.8 µm (Ch10.8) that is only weakly sensitive to water vapor as a reference for the calculation of the cloud effect in one method. The weighting functions of Ch6.2 and Ch7.3 in clear-sky conditions peak around 350 hPa and 500 hPa, respectively. For Ch10.8, it peaks near surface in clear-sky conditions. When clouds are present, the weighting functions shift upward and become truncated near the cloud top as clouds are mostly opaque at these wavelengths. The observation errors are computed for the two water vapor channels Ch6.2 and Ch7.3, while the window channel 10.8 is only used as a reference in the error calculation for the other channels.

2.2 Simulated AROME brightness temperature

For the calculation of model equivalents of the observations, we use simulations of the AROME-Austria forecast system with a 2.5 km grid spacing that was the operational modeling system of GeoSphere Austria in 2023,. AROME (Application of Research to Operations at MEsoscale (Seity et al., 2011)) is a spectral model with a non-hydrostatic dynamical core developed by Meteo-France and countries participating in the ACCORD consortium (A Consortium for COnvection-scale modelling Research and Development). The AROME-Austria domain shown in Figure 1 has 589 x 421 grid points and is still currently operational, but the GeoSphere plan to switch to a 1 km grid spacing some time in 2026. The first-guess (FG) fields have a 3-hr lead time and are used to compute the model equivalent BT for the MSG SEVIRI observations at Ch6.2, Ch7.3 and Ch10.8, with the radiative transfer model RTTOV (Radiative Transfer for TOVS) version 12.2 (Saunders et al., 2018) as the observation operator.

The optical properties of aerosols and clouds (OPAC) cloud liquid water particle scheme (Hess et al., 1998) and the Baran 2014 cloud ice water particle scheme (Baran et al., 2014) are applied for the all-sky model equivalent BT ($\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}}$) computation. In the OPAC scheme, five water cloud types are required: stratus continental, stratus maritime, cumulus continental clean, cumulus maritime, and cumulus continental polluted. To classify these cloud types, we first distinguish continental and maritime conditions using the sea-land mask from the GLOBE dataset (Hastings and Dunbar, 1999), and then identify clouds with CAPE greater than 250 $\mathrm{J\,kg^{-1}}$ as cumulus. In our simulation, the cumulus continental polluted type is not considered. The cloud fraction, cloud liquid water, and cloud ice provided by the model are used as additional inputs for the $\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$ calculation.

The model data covers the month of August 2023. For the analysis, we use FG model fields valid at 06, 09, 12, 15, and 18 UTC, resulting in a total of 153 FG fields (two time points are missing due to data corruption). We initially use the first 16 days to train the observation error model and the remaining 15 days to verify the trained error model, but switch to a bootstrapping

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

approach in Sect. 5 where training and evaluation days are randomly allocated. Figure 1 illustrates the observed and simulated BT in the model domain for both water vapor channels at 12 UTC on 5 August 2023. To avoid the potential impact of model lateral boundaries, the subdomain shown with the green rectangle is selected for the analysis. Apart from missing convective clouds in the south-eastern part of the domain in the simulated BTs, the agreement between observations (Figs. 1b,d) and simulations (Figs. 1a,c) is overall reasonable. However, the BT^{O} is in general slightly lower than the $\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}}$ in cloudy regions for both channels, indicating that the simulated cloud tops are lower or optically thinner than their observed counterparts.

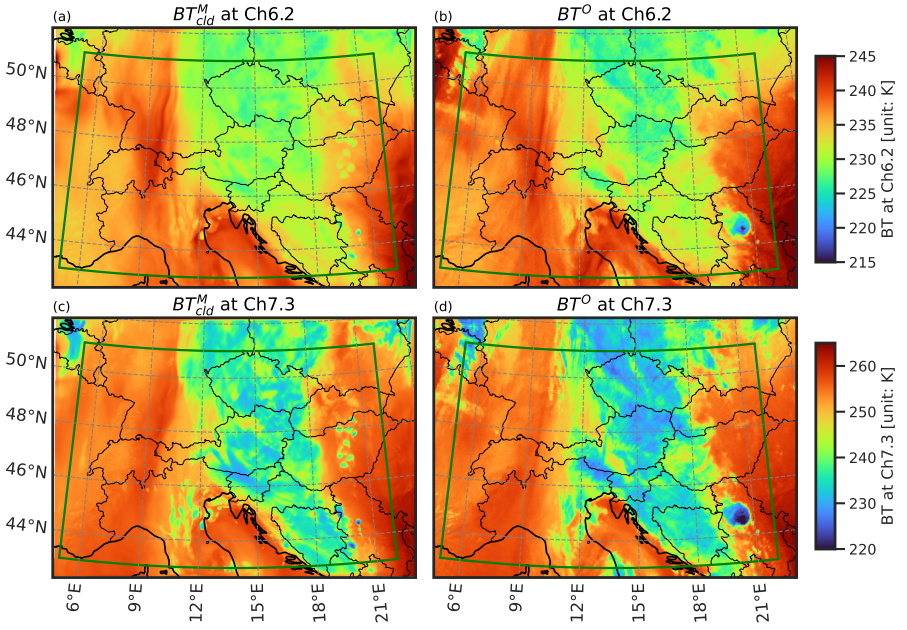

Figure 1. The distributions of (a)(c) $\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$ and (b)(d) BT^{O} at (a)(b) Ch6.2 and (c)(d) Ch7.3. The data is valid at 12 UTC 5th of August, 2023. The green rectangle is the domain used for further analysis.

Figure 2 shows a probability density distribution of the observations and model equivalents for both channels. In general, the simulations agree with the observations relatively well, but as seen in Fig. 1, the observations have more BTs lower than 230 K (Ch6.2) and 250 K (Ch7.3) than the simulations, indicating that the model underestimates the occurrence of high clouds. Additionally, there are more simulated than observed BTs around 230-235 (250-258) K at Ch6.2 (Ch7.3), which cloud be due to the model overestimating the occurrence of low/medium clouds, or it could be that with less high clouds occurring lower clouds have more influence on the simulated BTs. The mean biases at Ch6.2 and Ch7.3 are 0.426 K and -0.977 K, respectively.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

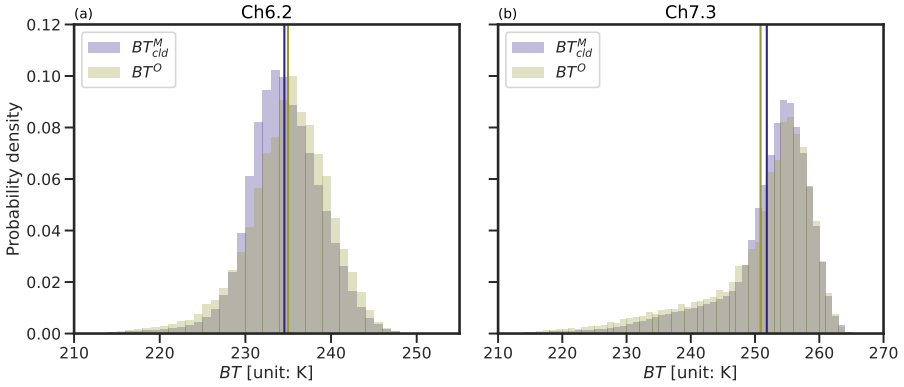

Figure 2. The probability density of $\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$ (deep indigo) and BT^{O} (olive gray) at (a) Ch6.2 and (b) Ch7.3. Samples are collected from the entire monitoring period. The mean values of the distribution are shown with vertical lines for $\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$ (deep indigo) and BT^{O} (olive gray), respectively.

2.3 Assessing Gaussianity

To objectively assess the Gaussianity of the FG departures normalized by the observation error, we use two metrics. The first one is the Kolmogorov–Smirnov statistic (KS_stat) (An, 1933; Smirnov, 1936), which is the maximum difference between the cumulative distribution function of the normalized FG departures ($\mathrm{F(x)}$) and a Gaussian distribution ($\mathrm{G}(x)$):

$$\mathrm{KS\_stat} = \sup_x |\mathrm{F}(x) - \mathrm{G}(x)| \tag{1}$$

where $\sup_x$ is the largest absolute difference between the $\mathrm{F}(x)$ and $\mathrm{G}(x)$ across all x values. The $\mathrm{KS\_stat}$ ranges from 0 to 1, with smaller values indicating that the normalized FG departures are close to a Gaussian distribution.

The second metric is the overlap rate between the probability density histograms of the normalized FG departures and a Gaussian distribution. In contrast to the $\mathrm{KS\_stat}$ which is defined by a single point of the distribution, the overlap rate quantifies the similarity between the target distribution and the standard Gaussian distribution over the whole distribution. Therefore, it's a good complementary index for $\mathrm{KS\_stat}$. The equation for the overlap rate is defined as:

$$\text{Overlap rate} = \frac{\sum_{i=1}^{N_{\mathrm{bin}}} \min(\mathrm{PD}_{\mathrm{gaussian}}(i), \mathrm{PD}_{\mathrm{normalized}}(i))}{\sum_{i=1}^{N_{\mathrm{bin}}} \max(\mathrm{PD}_{\mathrm{gaussian}}(i), \mathrm{PD}_{\mathrm{normalized}}(i))} \tag{2}$$

where $\mathrm{PD}_{\mathrm{normalized}}(i)$ and $\mathrm{PD}_{\mathrm{gaussian}}(i)$ are the probability density of normalized FG departures and of the standard Gaussian distribution at each bin, respectively. In contrast to $\mathrm{KS\_stat}$, the overlap rate is dependent on two settings, the number of bins (N_{bin}) and the range. Throughout this paper we compute the overlap rate for bins with a width of 0.1, and a range of -10 to 10. For our sample size and distributions, the overlap rate is not very sensitive to slight changes in bin width and range. Similar to $\mathrm{KS\_stat}$, the overlap rate also lies between 0 and 1, but has a reversed orientation with higher values for distributions closer to a Gaussian distribution.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

3 Cloud-dependent observation error models

In this section we introduce the error models of Harnisch et al. (2016) (H_{16}) and Okamoto et al. (2014) (O_{14}) and apply them to our monitoring period in August 2023. As Ch6.2 and Ch7.3 showed a similar behavior, we only illustrate the methods using Ch7.3 to keep the manuscript concise, but Sect. 5 includes a statistical comparison of all methods for both channels.

3.1 The H_{16} method

The first step of the H_{16} method is to derive the threshold BT (BT_{lim}) that will be used to determine the cloud effect predictor C. For deriving BT_{lim}, we calculate the mean difference between between BT^{M}_{cld} and the pseudo-cloud-free model equivalent BT (BT^{M}_{clr}) as a function of BT^{M}_{cld}. Subsequently, BT_{lim} is determined as the coldest BT^{M}_{cld} that has a mean difference between BT^{M}_{cld} and BT^{M}_{clr} of less than 0.1 K. Fig. 3a shows a 2D-histogram of BT^{M}_{cld} and the difference between BT^{M}_{cld} and BT^{M}_{clr}, the mean difference as a function of BT^{M}_{cld} (gray line) as well as BT_{lim} (yellow line). When $BT^{M}_{clr} \geq BT_{lim}$, differences larger than 2 K between BT^{M}_{clr} and BT^{M}_{clr} are very rare (note the logarithmic scaling in Fig. 3a).

$BT_{lim}$is also used to derive a cloud-free bias correction for the observations ($BC_{cloud-free}$). Specifically, we designate the FG departures with both $BT^{O} \geq BT_{lim} + 1K$ and $BT^{M}_{clr} \geq BT_{lim} + 1K$ as cloud-free FG departures. $BC_{cloud-free}$is then determined as the average over all FG departures for cloud-free observations. We add 1 K to $BT_{lim}$to ensure that the cloud effect is negligible. The $BC_{cloud-free}$for the training data at the two channels is 0.536 K for Ch6.2 and -0.040 K for Ch7.3, respectively. Hereafter, we remove $BC_{cloud-free}$from FG departures.

After determining BT_{lim} and $BC_{cloud-free}$, the cloud effect predictor C of H_{16} ($C^{mean}_{BT_{lim}}$) for each channel is calculated with the following equation:

$$C^{mean}_{BT_{lim}} = \frac{\max(BT_{lim} - BT^{M}_{cld}, 0) + \max(BT_{lim} - (BT^{O} - BC_{cloud-free}), 0)}{2} \tag{3}$$

The two terms, $\max(BT_{lim} - BT^{M}_{cld}, 0)$ and $\max(BT_{lim} - (BT^{O} - BC_{cloud-free}), 0)$, represent the cloud effect in model and observation space, respectively. The observation error for a chosen channel is described as

$$\sigma^{O} = \begin{cases} \sigma^{O}_{min}, & C \leq C_{min} \\ f(C), & C_{min} < C < C_{max} \\ \sigma^{O}_{max}, & C \geq C_{max} \end{cases} \tag{4}$$

where σ^{O}_{min} is the minimum observation error assigned in case of a small cloud effect lower than C_{min}, $f(C)$ is an increasing function of C, and σ^{O}_{max} is the assigned observation error for regions with a strong cloud effect. To obtain σ^{O}_{min}, C_{max}, σ^{O}_{max}, and the function $f(C)$, the FG departures of the training data set are first sorted in ascending order of cloud effect C. Subsequently, the sorted data is divided into bins of equal sample size and the standard deviation (Std) of the FG departures is computed for each bin. We chose to bin the data into bins of equal sample size rather than equal C width, which was used by Okamoto et al. (2014) and Harnisch et al. (2016), as there are far fewer samples with a high than with low C (Fig. 3a).

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

In the H_{16} approach, $C^{\mathrm{mean}}_{\mathrm{BT_{lim}}} = 0$ indicates no cloud effect for the given pixel. Therefore, the first bin includes all values with $C^{\mathrm{mean}}_{\mathrm{BT_{lim}}} = 0$, and $\sigma^{\mathrm{O}}_{\mathrm{min}}$ is the Std of the FG departures in that bin. The remaining data is then divided into 50 bins, each containing the same number of samples. The Std in each bin from Ch7.3 is shown by yellow dots in Fig. 3b, where the x-axis position is the average value of C in each bin. The results demonstrate that the Std increases monotonically before reaching its peaks around $C = 16.5$ K and decreases afterwards. As stated by Harnisch et al. (2016), the decrease in Std for the largest C follows from the definition of C. A very large C can only be achieved when C in both model and observation space is very large, which limits the potential difference of the two. Following Harnisch et al. (2016) and preceding studies, we assign the value of the peak of the function ($\sigma^{\mathrm{O}}_{\mathrm{max}}$) to all bins with a higher C than at the peak. This was motivated by higher errors of the observation operator in cloudy conditions. For $\mathrm{f}(C)$ we interpolate between the values of the individual bins. The resulting observation error curve is illustrated in Fig. 3b by the solid yellow line. In this context, it should be noted that the approach ignores the contribution of model error to the departure, which leads to a slightly inflated observation error estimate. Sensitivity tests of Harnisch et al. (2016) using the ensemble spread as proxy for model error, however, showed that ignoring model error only has a minor effect on the estimate.

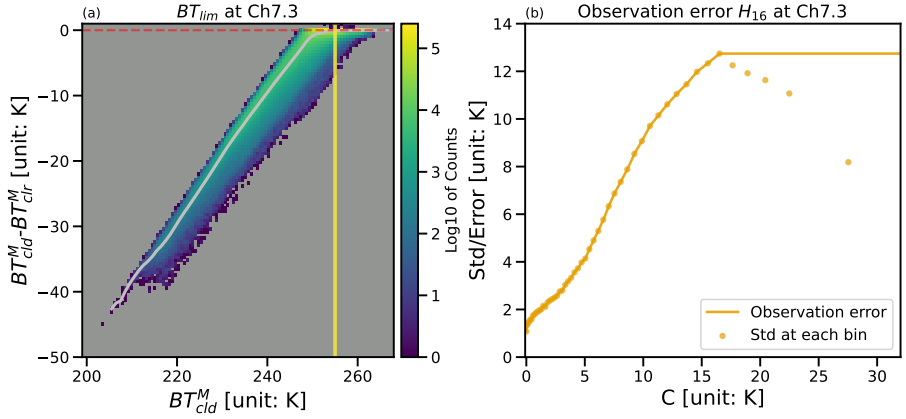

Figure 3. (a) 2D-histogram of $\mathrm{BT^{M}_{cld}}$ minus $\mathrm{BT^{M}_{clr}}$ vs $\mathrm{BT^{M}_{cld}}$ and mean difference as a function of $\mathrm{BT^{M}_{cld}}$ (gray line). The yellow line represents $\mathrm{BT_{lim}}$, and the red dashed line is $\mathrm{BT^{M}_{clr}} - \mathrm{BT^{M}_{cld}} = 0$. (b) Std of the FG departure (dots) and the modified observation error (line) as a function of the cloud effect predictor C for Ch7.3.

Figure 4 shows the effect of normalizing the FG departures with the H_{16} observation error estimates on the Gaussianity of the departures compared to using a constant error. We find that H_{16} successfully reduces the pronounced peak and long tails that occur when a constant error is used, but the distribution still has more values around zero and slightly higher tails than the Gaussian distribution. Our two error metrics quantify the improvement, with the overlap rate increasing from 42 to 85 %, and $\mathrm{KS_{stat}}$decreasing from 0.22 to 0.04.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

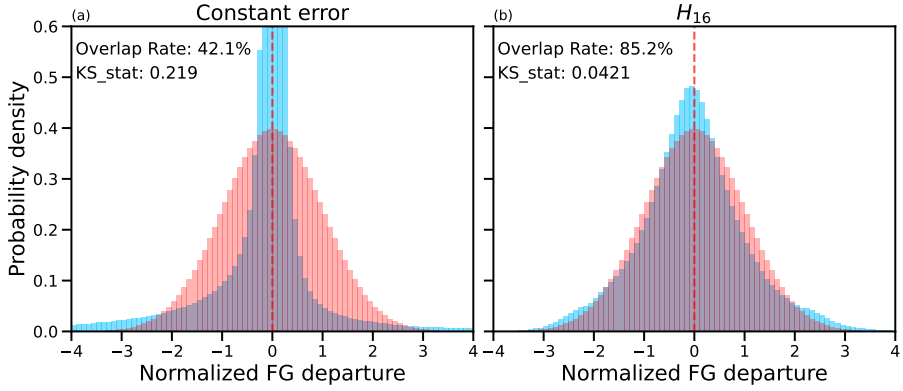

Figure 4. Probability density of a standard Gaussian distribution (red) and the FG departures of Ch7.3 normalized by the observation error (light blue) using (a) a constant value and (b) the H_{16} method. Samples are collected over the whole verification period of August 2023. The overlap rate and KS_{stat} are measures of Gaussianity as described in Subsect. 2.3.

3.2 The O_{14} method

While the O_{14} method has many similarities to H_{16}, it has two crucial differences. The first one is that instead of a climatological BT_{lim} value, the individual model clear-sky brightness temperature (BT^{M}_{clr}) of the assimilated channel is used as a reference to compute the cloud effect. This has the advantage that a situation-dependent reference value is used that accounts for the actual temperature and humidity profile. However, BT^{M}_{clr} is only a near-perfect reference for the cloud effect in model space, while the observed profile may differ due to model error. The estimate of the cloud effect in observation space may therefore suffer from model error. The second difference is that absolute differences are used instead of setting negative differences to zero. The use of absolute differences results in:

$$C^{\mathrm{mean}}_{\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}} = \frac{|\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}} - \mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}| + |\mathrm{BT}^{\mathrm{O}} - \mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}|}{2} \tag{5}$$

The two terms in the numerator, $|\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}} - \mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}|$ and $|\mathrm{BT}^{\mathrm{O}} - \mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}|$, are C in model and observation space, respectively. In contrast to $C^{\mathrm{mean}}_{\mathrm{BT}_{\mathrm{lim}}}$ which is zero whenever $\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}}$ and BT^{O} are both higher than $\mathrm{BT}_{\mathrm{lim}}$, $C^{\mathrm{mean}}_{\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}}$ can only be zero when $\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}}$ is exactly equal to BT^{O}.

With $C^{\mathrm{mean}}_{\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}}$, the observation error can also be estimated using the same way as described for H_{16} in Sect. 3.1. As the model-derived $\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$ does not once have the exact same value as the observed BT^{O} in our dataset, there are no samples with $C^{\mathrm{mean}}_{\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}}$ equal zero. Thus, the full training data set is divided into 50 bins of $C^{\mathrm{mean}}_{\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}}$ with equal sample size. C_{min} and $\sigma^{\mathrm{O}}_{\mathrm{min}}$ are then determined by the Std of the FG departures in the first bin. The way to derive C_{max}, $\mathrm{f}(C)$ and $\sigma^{\mathrm{O}}_{\mathrm{max}}$ is the same as that of H_{16}.

The resulting Std and observation error are illustrated in Fig. 5a and histograms of normalized departures in Fig. 5b. Figure 5b implies that the O_{14} method performs poorly without additional procedures that will be discussed in the subsequent section. It produces a bimodal shape with two peaks located around ± 1. As Okamoto et al. (2014) noted, their approach doesn't work well for cloud free samples (see Fig. 10a of Okamoto et al. (2014)). The underlying reason why the O_{14} approach fails when

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

no clouds are present is that by computing the observed cloud effect against the clear-sky model, the observed cloud effect is inflated by the absolute model and observation errors. Looking only at clear-sky cases ($\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}}$=$\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$), if the true BT were BT^{T} and the model error ϵ^{M} and the observation error ϵ^{O}, the cloud effect becomes:

$$2C^{\mathrm{mean}}_{\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}} = |\mathrm{BT}^{\mathrm{T}} + \epsilon^{\mathrm{M}} - \mathrm{BT}^{\mathrm{T}} - \epsilon^{\mathrm{M}}| + |\mathrm{BT}^{\mathrm{T}} + \epsilon^{\mathrm{O}} - \mathrm{BT}^{\mathrm{T}} - \epsilon^{\mathrm{M}}| = 0 + |\epsilon^{\mathrm{O}} - \epsilon^{\mathrm{M}}| \tag{6}$$

If many cloud-free samples are binned together by cloud effect, the FG departure distribution of that bin will be a two-point distribution, with an Std of two times the cloud effect if the FG departures are unbiased. As a result, when using the Std_j as the observation error σ^{O}_j at $\mathrm{bin_j}$, the normalized FG departure in clear-sky conditions is:

$$\text{Normalized FG departures} = \frac{\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}} - \mathrm{BT}^{\mathrm{O}}}{\sigma^{\mathrm{O}}_j} = \frac{2C_{\mathrm{j}}}{\pm 2C_{\mathrm{j}}} = \pm 1 \tag{7}$$

As shown in Fig. 5a, the Std of the FG departures (orange line) almost overlaps with the black dotted line showing twice the cloud effect ($2C$) for small values of C below 2 K, showing that almost no clouds are present in these bins. These bins, which contain a majority of all samples, causes the two peaks of the normalized FG departures around ± 1, resulting in the bimodal distribution shown in Fig. 5b. Note that this bimodal distribution will be much weaker when the cloud effect binning is wider.

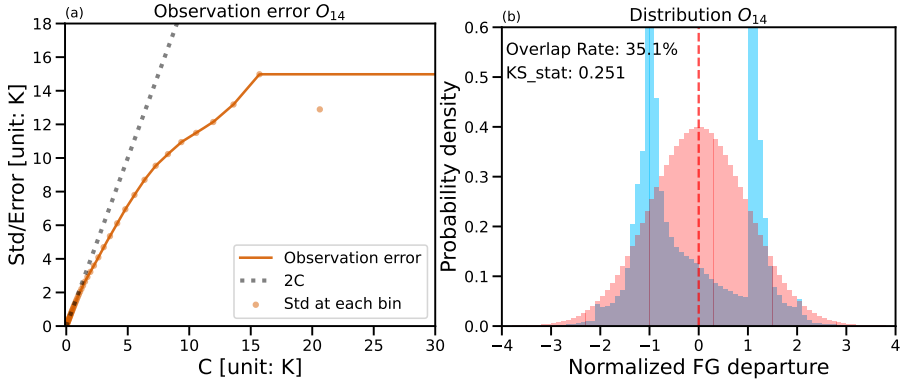

Figure 5. (a) The Std at each bin (orange dots) and the modified observation error (orange solid line) using O_{14} for Ch7.3 from the training period. The black dotted line indicates two times the cloud effect ($2C$). (b) Probability density of a standard Gaussian distribution (red) and the normalized FG departures (light blue) using the O_{14} method for Ch7.3 in the verification period.

3.2.1 The O_{14} method with additional quality control and C_{min}

To solve the bimodality issue, Okamoto et al. (2014) suggested to use another σ^{O} for pixels with small C. For this modification, observation errors are tuned by resetting the values of C_{min} and $\sigma^{\mathrm{O}}_{\mathrm{min}}$. Specifically, we redefine the first bin as $[0, C_{\mathrm{min}}]$ and use the Std of the samples in this bin as the $\sigma^{\mathrm{O}}_{\mathrm{min}}$. The remaining data is then divided into 50 equally sized bins and the observation error model is derived using the same steps described before. Here, $C_{\mathrm{min}} = 2$ is set for further analysis.

The resulting observation error curve is shown in Fig. 6a. With this tuning , the observation error avoids being close to $2C$ for $C \leq C_{\mathrm{min}}$. In general, data with tuned observation error accounts for over 60% of the whole samples at Ch7.3. After tuning,

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

the bimodal shape in Fig. 5b disappears, and the distribution of the normalized FG departures becomes more Gaussian (Fig. 6b). However, there are still two small "ears" around ± 1.

In order to remove the two "ears", a so-called quality control (QC) procedure proposed by Okamoto (2017) is applied. Data is rejected with $|\mathrm{FG\ departure}| > 1.8C$ (the black dotted lines in Fig. 6c) when $C > C_{\min}$ (the white dotted line in Fig. 6c). As Fig. 6d shows, the two "ears" are removed successfully by the QC. Although the peak of the distribution becomes too large after QC, both the overlap rate and the KS_stat are improved compared to the distribution in Fig. 6b. However, with this QC, around 12% data is removed and the removed data is not necessarily of low observational quality. We can understand which data is removed by looking at when the absolute FG departure is twice the cloud effect. This can occur when no clouds are present in the model ($\mathrm{BT}^{\mathrm{M}}_{\mathrm{cld}}$=$\mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$), resulting in an FG departure exactly twice the cloud effect, or when $\mathrm{BT}^{\mathrm{O}} \approx \mathrm{BT}^{\mathrm{M}}_{\mathrm{clr}}$, resulting in an FG departure approximately minus twice the cloud effect. Accordingly, the samples removed by the QC include all situations when only one of the observations or model is cloud free. Arguably, these samples are the most valuable to assimilate, helping to remove clouds that should not be there or to create clouds when they are missing.

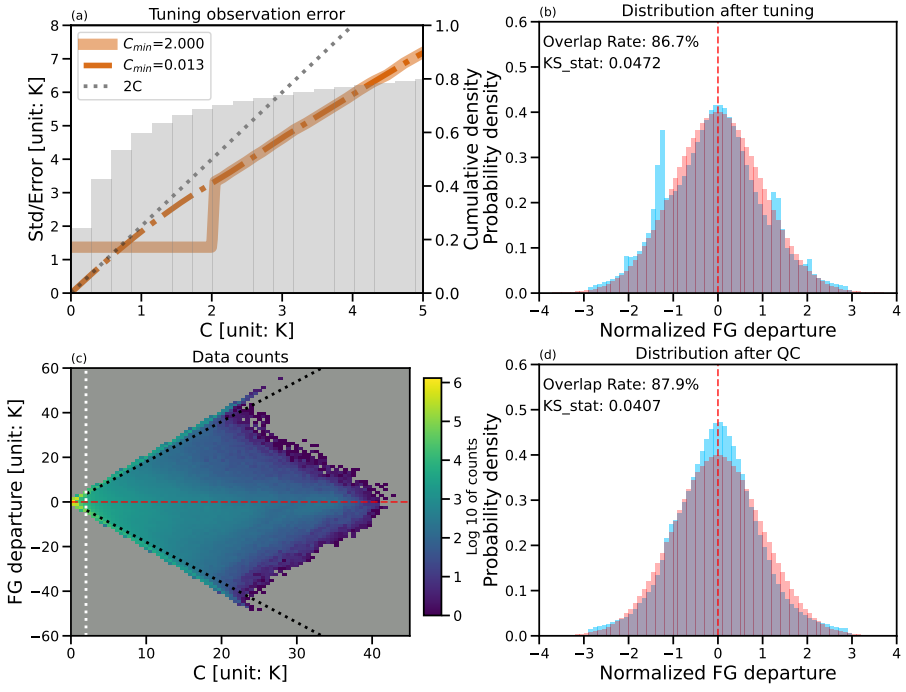

Figure 6. (a) The assigned observation error as a function of C before (orange dashed dotted line) and after tuning (orange solid line) of the O_{14} method, with cumulative density histogram overlaid in gray. (c) Density scatterplot of C vs FG departure. The black dotted lines represents $|\mathrm{FG\ departure}| = 1.8C$ and the white dotted line $C = 2$. (b)(d) Probability density of a standard Gaussian distribution (red) and the normalized FG departures (light blue) (b) after tuning the observation error and (d) after additional QC. All samples are collected for Ch7.3 in the verification period .

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

To evaluate the impact of the choice of $C_{\min}$, Fig. 7 demonstrates the KS_stat, the overlap rate and the data utilization rate as functions of $C_{\min}$. Overall, the results are similar for both channels. Regarding the impact on the distribution (Figs. 7a-b), both metrics, the KS_stat and the overlap rate, show similar results. Both indicating that the distribution becomes closer to Gaussianity as $C_{\min}$ increases, reaching the optimal distribution at approximately $C_{\min} = 1.5$ for Ch6.2 and $C_{\min} = 1.1$ for Ch7.3. Beyond these points, the KS_stat (overlap rate) gradually increases (decreases) with $C_{\min}$. For data utilization, the rates for both channels increase rapidly for $C_{\min} < 1$. For higher values of $C_{\min}$, the rate of increase become more moderate.

The relationships of the distribution, data utilization, and $C_{\min}$ indicate that, even after observation error tuning and QC, it's advisable to set $C_{\min}$ bigger than 1 K. Additionally, there is a trade-off between the Gaussianity of the distributions and data utilization rate.

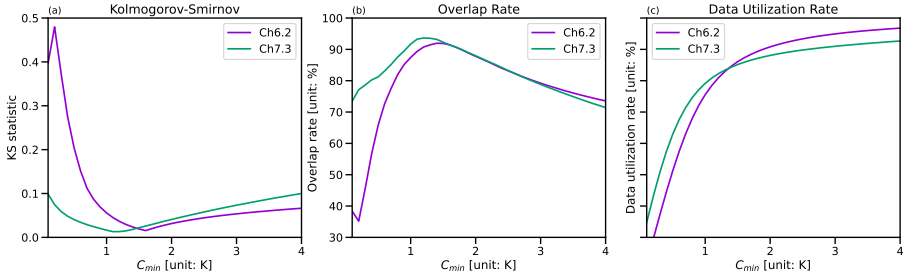

Figure 7. The (a) KS_stat ,(b) overlap rate and (c) data utilization rate as a function of $C_{\min}$. The violet and teal green lines show the results from Ch6.2 and Ch7.3 during verification period, respectively. (b) and (c) share the same y-axis range.

3.2.2 The O_{14} method with use of a window channel

Some studies applied the O_{14} approach with a cloud effect estimated using a window channel instead of the assimilated channel (Geer et al., 2019; Feng and Pu, 2022). Following this, we also test the impact of applying the O_{14} approach using the window channel Ch10.8. For simplicity, this approach will be named O_{14}^{win} in the following. The cloud effect of the assimilated water vapor channels and the window channel are named $C_{\text{wv}}^{\text{mean}}$ and $C_{\text{win}}^{\text{mean}}$, respectively. In this case, all training samples are divided into 50 bins with equal sample size without any extra pre-processing.

In order to plot O_{14} against O_{14}^{win} despite them using different cloud effect predictors, we bin the observation errors of O_{14}^{win} by the cloud effect of O_{14}. The resulting observation errors for O_{14}^{win} is shown with pink dots in Fig. 8a, and the observation error curve calculated with O_{14} is shown by the orange curve for comparison. To take a closer look at the difference between the two methods when the value of $C_{\text{wv}}^{\text{mean}}$ is small, the range of $C_{\text{wv}}^{\text{mean}} \in [0, 5]$ in Fig. 8a is zoomed in Fig. 8b. The two plots indicate that the observation errors are similar for $C_{\text{wv}}^{\text{mean}}$ values approximately within the range of 1 to 9. Larger differences occur for a larger $C_{\text{wv}}^{\text{mean}}$, but those values only account for about 10% of the total sample. For $C_{\text{wv}}^{\text{mean}} < 1$, the O_{14}^{win} approach avoids assigning an observation error close to $2C_{\text{wv}}^{\text{mean}}$ in contrast to O_{14}. As a result, the O_{14}^{win} approach also resolves the bimodality issue and leads to a fairly Gaussian distribution (Fig. 8c) without removing data by an additional QC.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

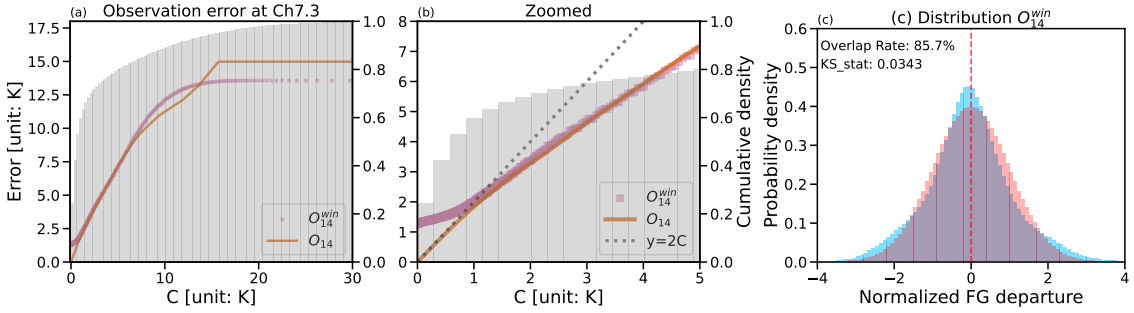

Figure 8. (a) The assigned observation errors a function of $C_{\mathrm{wv}}^{\mathrm{mean}}$, overlaid with cumulative density histogram in gray. The orange line represents assigned observation errors calculated by using O_{14}, and the pink square dots are the mean observation errors calculated by using $\mathrm{O}_{14}^{\mathrm{win}}$. (b) The same as (a), but only showing the enlarged region of $0 \leq C \leq 5$. The black dotted line illustrates $y = 2C_{\mathrm{wv}}^{\mathrm{mean}}$. (c) Probability density of a standard Gaussian distribution (red) and the FG departures of Ch7.3 normalized by the observation errors computed by using $\mathrm{O}_{14}^{\mathrm{win}}$ (light blue). All results are derived from the verification period.

4 Revised error models

4.1 Replacing mean by maximum when computing CA

While the approaches of H_{16}, O_{14} with additional QC, and $\mathrm{O}_{14}^{\mathrm{win}}$ can all lead to fairly Gaussian distributions of the normalized departures and while these methods have been widely applied, the use of a mean C does not seem to be fully justified. Using a mean C, the same cloud effect is estimated for (1) a high cloud in one data source and clear sky in the other and (2) mid level clouds in both data sources. The first case, however, would lead to a very large effect of hydrometeor errors on the FG departure, while the latter would lead to a comparably small one. To address this issue, we propose a revised approach that is based on the maximum cloud effect in both data sources, which is proportional to the largest possible departure for the same model error in hydrometeors. Thereby, the method should better mitigate the nonlinear cloud height effect and make the normalized departures more linearly dependent on errors in cloud hydrometeors.

As the revised approach would also require an adaptation of the error tuning and QC for the O_{14} approach, we will only evaluate it for the H_{16} and $\mathrm{O}_{14}^{\mathrm{win}}$ approaches. Nevertheless, we believe that it could also be applied in the O_{14} approach with some minor adaptations.

The modified C for the H_{16} method ($C_{\mathrm{BT_{lim}}}^{\mathrm{max}}$) and the $\mathrm{O}_{14}^{\mathrm{win}}$ method ($C_{\mathrm{win}}^{\mathrm{max}}$) at each channel are defined as:

$$C_{\mathrm{BT_{lim}}}^{\mathrm{max}} = \max(\max(\mathrm{BT_{lim}} - \mathrm{BT_{cld}^{M}}, 0), \max(\mathrm{BT_{lim}} - (\mathrm{BT^{O}} - \mathrm{BC_{cloud-free}}), 0)) \quad (8)$$

$$C_{\mathrm{win}}^{\mathrm{max}} = \max(|\mathrm{BT_{cld}^{M}} - \mathrm{BT_{clr}^{M}}|, |\mathrm{BT^{O}} - \mathrm{BT_{clr}^{M}}|) \quad (9)$$

$C_{\mathrm{BT_{lim}}}^{\mathrm{max}}$ is calculated using the corresponding data from the assimilated channel while $C_{\mathrm{win}}^{\mathrm{max}}$ is computed using the data from the Ch10.8 window channel. The revised version of H_{16} will be named $\mathrm{S}_{26}^{\mathrm{H}}$ and the revised version of $\mathrm{O}_{14}^{\mathrm{win}}$ will be $\mathrm{S}_{26}^{\mathrm{O}}$.

Figures 9a-b show the observation error curves using the mean and the maximum for computing C. The maximum approach results in a monotonic increase of the error with the cloud effect, while errors using the mean approach decrease after a certain

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

value of C. This decrease of errors can easily be explained by the fact that both model and observation must be cloudy for a very large C in the mean approach, which limits the potential FG departure. In practice, however, the decrease was usually ignored and errors were kept at the largest value for a large C in the mean approach (Harnisch et al., 2016; Schomburg et al., 2026) due to large expected operator and representation errors in case of clouds. The maximum approach avoids this ad-hoc adjustment given the monotonic increase of the error with C which seems to be physically better justified given larger errors in cloudy conditions and the intention to mitigate the nonlinear cloud height effect.

Most importantly, the normalized FG departures of $\mathrm{S}_{26}^{\mathrm{H}}$ and $\mathrm{S}_{26}^{\mathrm{O}}$ (Figs. 9c-d) also show a more Gaussian distribution than applying the ones of H_{16} and $\mathrm{O}_{14}^{\mathrm{win}}$ (Fig. 4b, Fig. 8c) with overlap rates increasing from 85 % (86%) to 93 % (90 %) for the H_{16} and the $\mathrm{O}_{14}^{\mathrm{win}}$ approaches, respectively.

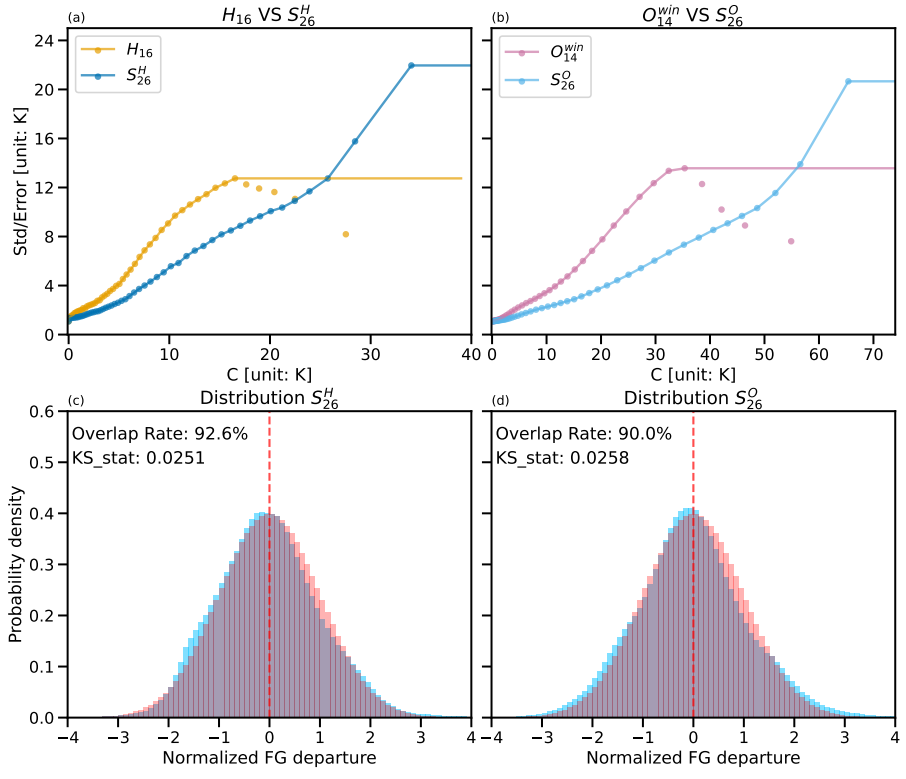

Figure 9. (a)(b) The Std at each bin (dots) and the adjusted observation error (line) for Ch7.3 from the training period using (a) H_{16} (yellow) and $\mathrm{S}_{26}^{\mathrm{H}}$ (blue), and (b) $\mathrm{O}_{14}^{\mathrm{win}}$ (purple) and $\mathrm{S}_{26}^{\mathrm{O}}$(light blue). (c)(d) Probability density of a standard Gaussian distribution (red) and the normalized FG departures (light blue) using observation errors calculated with (c) $\mathrm{S}_{26}^{\mathrm{H}}$ and (d) $\mathrm{S}_{26}^{\mathrm{O}}$ for the verification period.

To understand why the maximum approach results in a more Gaussian normalized FG departure distribution, let us first look at where the normalized FG departures of the $\mathrm{O}_{14}^{win}$and H_{16}approaches depart from Gaussianity (Fig. 4b & 8c). They both have too many small departures in the range of -0.5 to 0.5, too many large departures beyond ± 2, and too few in between. As can be seen in Fig. 10, the maximum approach results in lower errors when FG departures are small (along the diagonal) and larger

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

ones further away from the diagonal. For example, using the mean approach, if the observed $\mathrm{BT^O}$ was 8 K below $\mathrm{BT_{lim}}$ and the simulated $\mathrm{BT^M_{cld}}$ was 10 K lower, the mean cloud effect would be 9 K and the assigned error would be ≈ 8 K (Fig. 9a), resulting in a normalized departure of ≈ 0.25. Using the max approach the cloud effect for the same values would be 10 K, and the assigned error would be ≈ 5 K, resulting in a normalized FG departure of ≈ 0.4. This explains why the using the max helps reduce the number of small normalized FG departures. For large FG departures the opposite occurs. If $\mathrm{BT^O}$ was equal to $\mathrm{BT_{lim}}$ and $\mathrm{BT^M_{cld}}$ was 18 K lower, $C^{\mathrm{mean}}_{\mathrm{BT_{lim}}}$ would again be 9 K, with the same assigned error of ≈ 8 K, resulting in a normalized FG departure of ≈ 2.25. In contrast, $C^{\mathrm{max}}_{\mathrm{BT_{lim}}}$ would be 18, the assigned error ≈ 10 K, and the normalized departure ≈ 1.8.

Additionally, since maximum approach gives larger observation errors to pixels strongly affected by clouds, it assigns observations with large FG departures lower weights (Fig. 9b) compared to the mean approach (Fig. 9a). By doing so, it reduces potential negative effects of outliers. Meanwhile, these observations are easier to pass the background QC when getting larger errors. Thus, it also helps to retain potential valuable observations as for example the use of a Huber norm for QC (Tavolato and Isaksen, 2015).

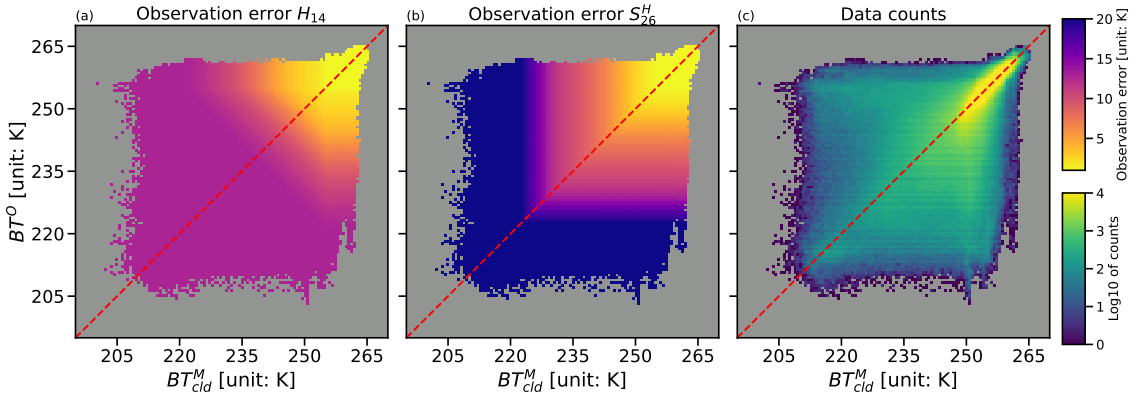

Figure 10. The (a)(b) 2D histograms of the observation error as a function of $\mathrm{BT^O}$ and $\mathrm{BT^M_{clr}}$ when computed by using (a) $\mathrm{H_{16}}$ and (b) $\mathrm{S^H_{26}}$. (c) 2D histogram of the number of observations as a function of $\mathrm{BT^O}$ and $\mathrm{BT^M_{clr}}$. Results are for Ch7.3 during the verification period. The red dashed lines represents $\mathrm{BT^O}$ = $\mathrm{BT^M_{clr}}$.

4.2 The possibility of abandoning $\mathrm{BT_{lim}}$ in the formulation of $\mathrm{H_{16}}$ and $\mathrm{S^H_{26}}$

In this section, we discuss the effect of replacing the empirically determined parameter $\mathrm{BT_{lim}}$ in the formulation of $\mathrm{H_{16}}$ and $\mathrm{S^H_{26}}$ by a constant value, which would simplify the required training. Figure 11 shows the KS_stat and overlap rate as a function of the chosen $\mathrm{BT_{lim}}$ for the $\mathrm{H_{16}}$ and the $\mathrm{S^H_{26}}$ approach. To ensure consistency, we still use the initial $\mathrm{BT_{lim}}$ value to compute the cloud-free bias correction, but this could be done differently in an operational system. When a small $\mathrm{BT_{lim}}$ is chosen (left side of each subplot), C equals zero for all pixels. From this follows that the Std of the whole sample is used as observation error for all observations. When $\mathrm{BT_{lim}}$ is large (right side of each subplot), $\max(\mathrm{BT_{lim}} - \mathrm{BT^M_{cld}}, 0)$ is always equal to $\mathrm{BT_{lim}} - \mathrm{BT^M_{cld}}$ and $\max(\mathrm{BT_{lim}} - (\mathrm{BT^O} - \mathrm{BC_{cloud-free}}), 0)$ is always equal to $\mathrm{BT_{lim}} - (\mathrm{BT^O} - \mathrm{BC_{cloud-free}})$, which means that we bin the data solely by a function of $\mathrm{BT^O}$ and $\mathrm{BT^M_{cld}}$. In this case, $\mathrm{BT_{lim}}$ does not have any impact on the observation error estimate and

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

could therefore be replaced by a constant parameter instead of an empirically determined parameter that may have a seasonal dependence.

Overall, the KS_stat and the overlap rate show consistent results. For H_{16}, we get the best results when $\mathrm{BT}_{\mathrm{lim}} = 229\ (253)$ for Ch6.2 (Ch7.3), revealing that an empirically determined $\mathrm{BT}_{\mathrm{lim}}$ is beneficial for this approach. Interestingly, best results were achieved for a $\mathrm{BT}_{\mathrm{lim}}$ that is slightly colder than in the default setup. This means that H_{16} could be slightly improved by adding an offset on $\mathrm{BT}_{\mathrm{lim}}$ or modifying it's derivation, but we did not investigate this further as the $\mathrm{S}_{26}^{\mathrm{H}}$ approach leads to even better results.

The $\mathrm{S}_{26}^{\mathrm{H}}$ approach is not sensitive on the choice of $\mathrm{BT}_{\mathrm{lim}}$ as long as the chosen $\mathrm{BT}_{\mathrm{lim}}$ is warmer than the empirically determined $\mathrm{BT}_{\mathrm{lim}}$. This means that the empirically determined $\mathrm{BT}_{\mathrm{lim}}$ could be replaced by a constant parameter or could even be completely removed in the formulation without any effect on the KS_stat or the overlap rate. Nevertheless, we decided to keep $\mathrm{BT}_{\mathrm{lim}}$ in the formulation of the $\mathrm{S}_{26}^{\mathrm{H}}$ approach for this study to have a more direct comparison to the other approaches from preceding studies.

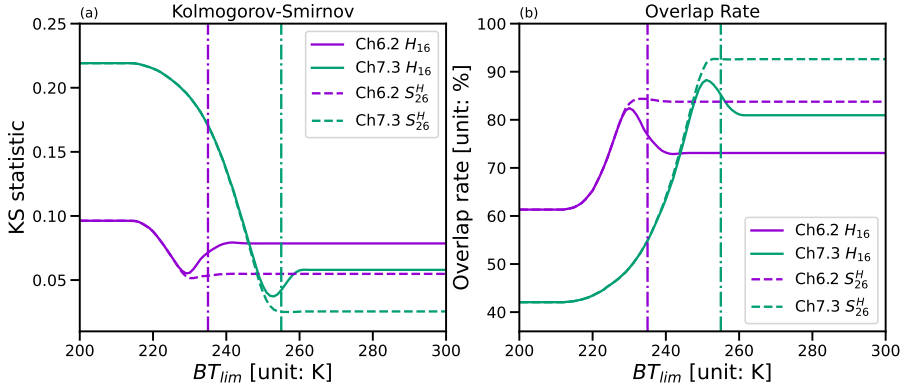

Figure 11. (a) KS_stat and (b) overlap rate as a function of $\mathrm{BT}_{\mathrm{lim}}$. The violet and green lines show the results for Ch6.2 and Ch7.3 during the verification period, respectively. The solid and dashed lines are the results of H_{16} and $\mathrm{S}_{26}^{\mathrm{H}}$. The dash–dot lines represent the default $\mathrm{BT}_{\mathrm{lim}}$ values that are used in all other parts of our study.

5 Statistical comparison of all approaches

Figure 12 presents all normalized FG departure distributions for Ch7.3 together to facilitate their direct comparison. For the O_{14} approach, we set $C_{\mathrm{min}} = 2$ and use additional QC here and thereafter. Compared to using a constant error (Fig. 12a), all methods obtain a clearly more Gaussian distribution, which means that all of them can act as a viable solution for all-sky IR data assimilation. Among the approaches, the revised approaches using the maximum for calculating C (Figs. 12e-f) lead to the best results. The revised approaches lead to a better match with the Gaussian distribution both around the peak and the tails compared to the approaches using the mean (Figs. 12b-c). The better match is also revealed by a higher overlap rate and lower KS_stat. The two methods using the mean for C and no QC (H_{16} and $\mathrm{O}_{14}^{\mathrm{win}}$) show similar distributions, but $\mathrm{O}_{14}^{\mathrm{win}}$ has a slightly lower KS_stat and higher overlap rate than H_{14}, mainly owing to a better distribution around the peak. Among the

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

methods using the mean for C, the O_{14} approach with $C_{\min} = 2$ and additional QC leads to the best overlap rate, but $\mathrm{KS\_stat}$ lies in between the other two approaches mainly due to too many values around the peak of the distribution. Additionally, the QC removes a significant fraction of the data that may contain valuable information. Specifically, in Fig. 12d the QC removes normalized FG departures with an absolute value greater than 4.

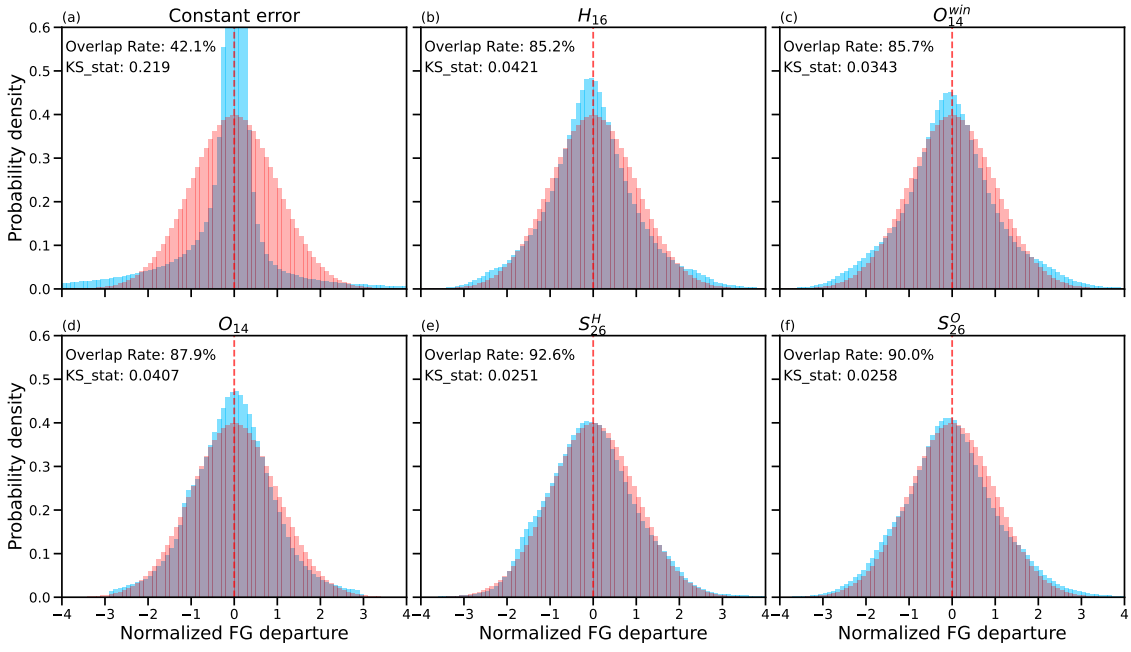

Figure 12. Probability density of a standard Gaussian distribution (red) and the normalized FG departures (light blue) using (a) a constant value as the observation error, (b) H_{16} method, (c) $\mathrm{O}_{14}^{\mathrm{win}}$ method, (d) O_{14} method, (e) $\mathrm{S}_{26}^{\mathrm{H}}$ and (f) $\mathrm{S}_{26}^{\mathrm{O}}$. Results are for Ch7.3 during verification period.

All the analysis presented so far is based on the results from a single training-verification run, which may not be sufficiently representative. Thus, we conducted multiple training-verification runs to achieve a more robust evaluation of the different approaches. For this purpose, we conducted 1000 training-verification runs that randomly select 16 days in August as the training period and use the remaining 15 days for verification.

Table 1 illustrates the average $\mathrm{KS\_stat}$, overlap rate and data utilization rate of this bootstrapping with 1000 runs. For both channels, the maximum versions (lines 2 and 4) performed better than the mean versions (lines 1 and 3) of the respective methods in all 1000 bootstrapped runs, which is consistent with the conclusions from Fig. 12. Overall, the mean versions without QC, $\mathrm{O}_{14}^{\mathrm{win}}$ and H_{16}, perform fairly similarly with no clear winner for both channels and metrics. O_{14} including QC (line 5) performs best in terms of the overlap rate for Ch6.2 and is among the best three in terms of $\mathrm{KS\_stat}$ for Ch6.2. For Ch7.3, however, results are only in the mid-range and it should be noted that about 11 % of the data is lost due to the QC.

Apart from the five approaches discussed above, we also conducted a corresponding bootstrap experiments, $\mathrm{S}_{26}^{\mathrm{H}}\_\mathrm{noBTlim}$ (lines 6), that use a constant $\mathrm{BT}_{\mathrm{lim}}$ value of 350 K. This constant $\mathrm{BT}_{\mathrm{lim}}$ is large enough that the results would be the same

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

if BT_{lim} were completely removed from the formulation of the approach. S_{26}^{H}_noBTlim shows even slightly better statistical results than S_{26}^{H}. Thus, removing BT_{lim} seems advisable as it also simplifies the training procedure. S_{26}^{H}_noBTlim even shows the best results for KS_stat and the best overlap rate for Ch7.3. Only in terms of the overlap rate for Ch6.2, O_{14} shows slightly better results, but this comes at the expense of losing 8 % of the observations due to QC.

Table 1. Mean KS_stat, overlap rate and data utilization rate for the verification period after bootstrapping the training/verification 1000 times with a look-up-table observation error model.

Exp names	KS_stat	Overlap rate (%)	Data utilization rate (%)	KS_stat	Overlap rate (%)	Data utilization rate (%)
	Ch6.2			Ch7.3		
H_{16}	0.071	77.4	100.0	0.089	81.3	100.0
S_{26}^{H}	0.058	82.1	100.0	0.076	85.0	100.0
O_{14}^{win}	0.074	77.7	100.0	0.090	80.6	100.0
S_{26}^{O}	0.060	82.4	100.0	0.078	84.0	100.0
O_{14}	0.060	85.2	92.1	0.085	81.4	88.6
S_{26}^{H}_noBTlim	0.056	82.6	100.0	0.075	85.0	100.0

6 Discussion and summary

In this study, we contrast existing cloud-dependent error models for the assimilation of all-sky infrared satellite radiances, perform a thorough testing of the methods and propose a conceptually revised approach. The testing is based on a one-month data set of SEVIRI observations at two water vapor sensitive infrared channels (Ch6.2 and Ch7.3) in a Central European domain and corresponding model equivalent brightness temperatures calculated from the output of the regional NWP model AROME-Austria. As evaluation metrics for Gaussianity, we used the Kolmogorov–Smirnov statistic and the overlap rate with a Gaussian distribution. Furthermore, we conduct 1000 bootstrapping experiments for training and verification.

The main findings of our study can be summarized as follows:

1. Overall, all observation error models can successfully transform the very non-Gaussian distribution of FG-departures of all-sky IR radiances into a near-Gaussian distribution of normalized departures.

2. For the approach of Okamoto et al. (2014), however, successful transformation requires either an additional quality control that eliminates about 10 % of the data or the use of a different channel (e.g. a window channel in our case) for the estimation of the cloud effect.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

3. Despite conceptual differences in the calculation of the cloud effect, the approaches proposed by Harnisch et al. (2016) and Okamoto et al. (2014) perform fairly equally in terms of Gaussianity. For the lower-peaking channel (7.3 μm), which has a larger cloud effect than the other higher peaking channel (6.2 μm), best results were achieved with the approach of Harnisch et al. (2016) or the window-channel version of Okamoto et al. (2014). For the 6.2 μm channel, the own-channel version of Okamoto et al. (2014) led to a slightly more Gaussian distribution, but at the expense of removing observations with a QC procedure.

4. The approach of Harnisch et al. (2016) relies on an empirically determined limiting brightness temperature, which likely requires local and seasonal training. Such geographic and seasonal effects will be investigated thoroughly in a subsequent study. The approach of Okamoto et al. (2014) has the advantage that it takes a situation-dependent reference for calculating the cloud effect, the model clear sky equivalent brightness temperature. This reference, however, is only a perfect reference for the model state and not necessarily for the observed state. The estimated cloud effect therefore incorporates clear-sky model error which can lead to artifacts in the distribution (e.g. a bimodal distribution) if not implemented carefully.

5. We propose a revised formulation for both the method of Harnisch et al. (2016) and the window-channel version method of Okamoto et al. (2014). In principle, the revised approach should also be applicable to the own-channel version of Okamoto et al. (2014) with additional QC, but that would require some additional testing.

6. We are convinced that the revised approach, which defines the cloud effect as the maximum of the model and observation cloud effects, is physically better justified, as the resulting cloud effect is approximately proportional to the maximum possible FG departure. Thereby, the revised approach should better mitigate the cloud height effect that leads to the very non-Gaussian distribution of FG departures.

7. In line with a better physical justification, we demonstrate that the revised approaches consistently lead to more Gaussian distributions than their original counterparts at no additional cost.

8. For the revised version of Harnisch et al. (2016), the climatologically determined parameter limiting brightness temperature ($\mathrm{BT_{lim}}$) can actually be replaced by a constant, or even completely removed from the formulation without any deterioration of the results. In our tests, results even improved slightly without $\mathrm{BT_{lim}}$. It should be noted, however, that the removal only simplifies the formulation and the training procedure without changing the potential seasonal and geographic dependence of the training of the error model.

9. In the original version of Harnisch et al. (2016), the parameter $\mathrm{BT_{lim}}$ should be kept in the formulation, but there is indication that adding an offset in it's derivation or slightly revising its derivation would be beneficial.

In regards to the computational cost of the methods proposed, Harnisch et al. (2016) is the cheapest as it does not require computing any clear-sky values during run time, while using the window channel for the Okamoto et al. (2014) approach is the most expensive as it requires calculating both the all-sky and clear-sky window channel. If the various channels are

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

calculated for other purposes anyway, all the methods discussed should not differ noticeably. All our experiments stored the error model as a 50 value look up table. The look up table can be replace through an analytical function, but we recommend that a second-order polynomial be used for the max cloud effect approaches as a linear function offers a poor approximation.

Our results are based on a limited data set of a regional domain in one month in summer while some error properties may vary with location and season (Geer et al., 2019). These effects will be evaluated thoroughly in a follow study that includes data of various season and various climate zones, but preliminary results indicate that the qualitative results hold true elsewhere.

Code and data availability. The simulated BT and departures are available via (https://zenodo.org/records/18980206), and the full AROME fields used to simulate the BTs from are avialable at (https://zenodo.org/records/18955720). The code used to analyse and plot the results are currently freely available at: (https://gitlab.phaidra.org/dataassimilation/RTTOV_AROME/-/tree/sby_v2/ir_plot/Revisiting_OE_for_IR?ref_type=heads). Once the review process is finalized the final scripts will be added to the zenodo archive containing the BT. The SEVIRI observations are freely available via the EUMETSAT's data store (https://user.eumetsat.int/data-access/data-store) The RTTOV code is maintained, archived, and distributed by NWP-SAF (https://nwp-saf.eumetsat.int/site/software/rttov/), v12.2 was used in this study.

Author contributions. BS: Software, Formal analysis, Visualization, Investigation, Writing (original draft preparation); PG: Methodology, Visualization, Writing (review and editing), Validation: FM: Data curation, Resources; JM: Funding acquisition, Supervision; MW: Conceptualization, Methodology, Writing (review and editing), Funding acquisition, Supervision

Competing interests. There are no competing interests.

Acknowledgements. We would like to thank Sandy Chkeir for providing the scripts for BT simulation and Leonhard Scheck for helping to apply RTTOV to AROME data. The open Access funding provided by University of Vienna is acknowledged for covering the article processing charges, the China Scholarship Council for providing financial resources for the visit of BS and the FFG for supporting the research through the ASAP project CloudyRadiances (FFG-Nr.: 892666). Additionally, we'd like to thank Chun Yang from NUIST as well Adhithiyan Neduncheran and Christoph Wittmann from the GeoSphere for their advice.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

References

An, K.: Sulla determinazione empirica di una legge didistribuzione, Giorn Dell'inst Ital Degli Att, 4, 89–91, 1933.

Baran, A. J., Hill, P., Furtado, K., Field, P., and Manners, J.: A coupled cloud physics–radiation parameterization of the bulk optical properties of cirrus and its impact on the Met Office Unified Model Global Atmosphere 5.0 configuration, Journal of Climate, 27, 7725–7752, https://doi.org/10.1175/JCLI-D-13-00700.1, 2014.

Bocquet, M., Pires, C. A., and Wu, L.: Beyond Gaussian statistical modeling in geophysical data assimilation, Monthly Weather Review, 138, 2997–3023, https://doi.org/10.1175/2010MWR3164.1, 2010.

Candy, B. and Migliorini, S.: The assimilation of microwave humidity sounder observations in all-sky conditions, Quarterly Journal of the Royal Meteorological Society, 147, 3049–3066, https://doi.org/10.1002/qj.4115, 2021.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, 2005.

Eyre, J., Bell, W., Cotton, J., English, S., Forsythe, M., Healy, S., and Pavelin, E.: Assimilation of satellite data in numerical weather prediction. Part II: Recent years, Quarterly Journal of the Royal Meteorological Society, 148, 521–556, https://doi.org/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, Quarterly Journal of the Royal Meteorological Society, 146, 49–68, https://doi.org/10.1002/qj.3654, 2020.

Feng, C. and Pu, Z.: A bias correction scheme with the symmetric cloud proxy variable and its influence on assimilating all-sky GOES-16 brightness temperatures, Monthly Weather Review, 150, 3305–3323, https://doi.org/10.1175/MWR-D-21-0333.1, 2022.

Geer, A., Baordo, F., Bormann, N., Chambon, P., English, S., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K., and Lupu, C.: The growing impact of satellite observations sensitive to humidity, cloud and precipitation, Quarterly Journal of the Royal Meteorological Society, 143, 3189–3206, https://doi.org/10.1002/qj.3172, 2017.

Geer, A. J. and Bauer, P.: Observation errors in all-sky data assimilation, Quarterly Journal of the Royal Meteorological Society, 137, 2024–2037, https://doi.org/10.1002/qj.830, 2011.

Geer, A. J., Fabrizio, B., Bormann, N., and English, S.: All-sky assimilation of microwave humidity sounders, vol. 140, European Centre for Medium-Range Weather Forecasts Reading, UK, https://doi.org/10.21957/obosmx154, 2014.

Geer, A. J., Migliorini, S., and Matricardi, M.: All-sky assimilation of infrared radiances sensitive to mid-and upper-tropospheric moisture and cloud, Atmospheric Measurement Techniques, 12, 4903–4929, https://doi.org/10.5194/amt-12-4903-2019, 2019.

Geiss, S., Scheck, L., de Lozar, A., and Weissmann, M.: Understanding the model representation of clouds based on visible and infrared satellite observations, Atmospheric Chemistry and Physics, 21, 12 273–12 290, https://doi.org/10.5194/acp-21-12273-2021, 2021.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., et al.: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres, Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256, https://doi.org/10.1002/qj.3179, 2018.

Harnisch, F., Weissmann, M., and Periáñez, Á.: Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system, Quarterly Journal of the Royal Meteorological Society, 142, 1797–1808, https://doi.org/10.1002/qj.2776, 2016.

Hastings, D. A. and Dunbar, P. K.: Global land one-kilometer base elevation (GLOBE), 1999.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

Hastuti, M. I. and Min, K.-H.: Impact of assimilating GK-2A all-sky radiance with a new observation error for summer precipitation forecasting, Remote Sensing, 15, 3113, https://doi.org/10.3390/rs15123113, 2023.

Hess, M., Koepke, P., and Schult, I.: Optical properties of aerosols and clouds: The software package OPAC, Bulletin of the American meteorological society, 79, 831–844, https://doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2, 1998.

Honda, T., Miyoshi, T., Lien, G.-Y., Nishizawa, S., Yoshida, R., Adachi, S. A., Terasaki, K., Okamoto, K., Tomita, H., and Bessho, K.: Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015), Monthly Weather Review, 146, 213–229, https://doi.org/10.1175/MWR-D-16-0357.1, 2018.

Hu, C.-C., Geer, A., and van Leeuwen, P. J.: Introducing non-Gaussian observation errors into incremental variational data assimilation methods, Quarterly Journal of the Royal Meteorological Society, p. e5050, https://doi.org/10.1002/qj.5050, 2025.

Hu, G., Dance, S. L., Bannister, R. N., Chipilski, H. G., Guillet, O., Macpherson, B., Weissmann, M., and Yussouf, N.: Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 November 2021, Atmospheric Science Letters, 24, e1130, https://doi.org/10.1002/asl.1130, 2023.

Kugler, L. and Weissmann, M.: Effects of Observation-Operator Nonlinearity on the Assimilation of Visible and Infrared Radiances in Ensemble Data Assimilation, Quarterly Journal of the Royal Meteorological Society, 151, e4970, https://doi.org/10.1002/qj.4970, 2025a.

Kugler, L. and Weissmann, M.: The synergy of assimilating visible and infrared radiances and radar observations, Quarterly Journal of the Royal Meteorological Society, 151, e4938, https://doi.org/10.1002/qj.4938, 2025b.

Kugler, L., Anderson, J. L., and Weissmann, M.: Potential impact of all-sky assimilation of visible and infrared satellite observations compared with radar reflectivity for convective-scale numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 149, 3623–3644, https://doi.org/10.1002/qj.4577, 2023.

Li, J., Geer, A. J., Okamoto, K., Otkin, J. A., Liu, Z., Han, W., and Wang, P.: Satellite all-sky infrared radiance assimilation: Recent progress and future perspectives, Advances in Atmospheric Sciences, 39, 9–21, https://doi.org/10.1007/s00376-021-1088-9, 2022.

Lu, X. and Wang, X.: Assimilating GOES-16 ABI all-sky brightness temperature into the HAFS dual-resolution self-consistent EnVar DA system: Methods for observation error estimation and impact on Hurricane Laura (2020), Earth and Space Science, 12, e2024EA004 058, https://doi.org/10.1029/2024EA004058, 2025.

McNally, A.: A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 128, 2551–2556, https://doi.org/10.1256/qj.01.206, 2002.

Okamoto, K.: Evaluation of IR radiance simulation for all-sky assimilation of Himawari-8/AHI in a mesoscale NWP system, Quarterly Journal of the Royal Meteorological Society, 143, 1517–1527, https://doi.org/10.1002/qj.3022, 2017.

Okamoto, K., McNally, A., and Bell, W.: Progress towards the assimilation of all-sky infrared radiances: An evaluation of cloud effects, Quarterly Journal of the Royal Meteorological Society, 140, 1603–1614, https://doi.org/10.1002/qj.2242, 2014.

Okamoto, K., Hayashi, M., Hashino, T., Nakagawa, M., and Okuyama, A.: Examination of all-sky infrared radiance simulation of Himawari-8 for global data assimilation and model verification, Quarterly Journal of the Royal Meteorological Society, 147, 3611–3627, https://doi.org/10.1002/qj.4144, 2021.

Okamoto, K., Ishibashi, T., Okabe, I., and Shimizu, H.: Extension of all-sky radiance assimilation to hyperspectral infrared sounders, Quarterly Journal of the Royal Meteorological Society, 150, 5472–5497, https://doi.org/10.1002/qj.4883, 2024.

Pires, C. A., Talagrand, O., and Bocquet, M.: Diagnosis and impacts of non-Gaussianity of innovations in data assimilation, Physica D: Nonlinear Phenomena, 239, 1701–1717, https://doi.org/10.1016/j.physd.2010.05.006, 2010.

https://doi.org/10.5194/egusphere-2026-1463
Preprint. Discussion started: 14 April 2026

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., et al.: An update on the RTTOV fast radiative transfer model (currently at version 12), Geoscientific Model Development, 11, 2717–2737, https://doi.org/10.5194/gmd-11-2717-2018, 2018.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: An introduction to Meteosat second generation (MSG), Bulletin of the American Meteorological Society, 83, 977–992, https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, 2002.

Schomburg, A., Schraff, C., Köpken-Watts, C., Stephan, K., Schühl, L., and Faulwetter, R.: Operational all-sky assimilation of geostationary infrared water-vapour channels in the regional ICON-D2 model with an ensemble Kalman filter, Quarterly Journal of the Royal Meteorological Society, p. e70113, https://doi.org/10.1002/qj.70113, 2026.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective-scale operational model, Monthly Weather Review, 139, 976–991, https://doi.org/10.1175/2010MWR3425.1, 2011.

Shimizu, H., Kazumori, M., and Kadowaki, T.: Implementation of all-sky microwave radiance assimilation in JMA's global NWP system, Research activities in Earth system modeling. Working Group on Numerical Experimentation, pp. 1–20, 2020.

Smirnov, N.: Sui la distribution de w2 (Criterium de MRv Mises), Comptes Rendus (Paris), 202, 449–452, 1936.

Tavolato, C. and Isaksen, L.: On the use of a Huber norm for observation quality control in the ECMWF 4D-Var, Quarterly Journal of the Royal Meteorological Society, 141, 1514–1527, https://doi.org/10.1002/qj.2440, 2015.

Xu, D., Liu, Z., Fan, S., Chen, M., and Shen, F.: Assimilating all-sky infrared radiances from Himawari-8 using the 3DVar method for the prediction of a severe storm over North China, Advances in Atmospheric Sciences, 38, 661–676, https://doi.org/10.1007/s00376-020-0219-z, 2021.

Yang, C., Liu, Z., Bresch, J., Rizvi, S. R., Huang, X.-Y., and Min, J.: AMSR2 all-sky radiance assimilation and its impact on the analysis and forecast of Hurricane Sandy with a limited-area data assimilation system, Tellus A: Dynamic Meteorology and Oceanography, 68, 30 917, https://doi.org/10.3402/tellusa.v68.30917, 2016.

Yang, C., Zhong, T., and Min, J.: A preliminary study of FY-4B AGRI all-sky assimilation by WRFDA for tropical cyclones, Atmospheric Research, p. 108377, https://doi.org/10.1016/j.atmosres.2025.108377, 2025.

Zhu, L., Xue, M., Kong, R., and Min, J.: Direct assimilation of all-sky GOES-R ABI radiances in GSI EnKF for the analysis and forecasting of a mesoscale convective system, Monthly Weather Review, 151, 737–760, https://doi.org/10.1175/MWR-D-21-0293.1, 2023.

Zhu, Y., Liu, E., Mahajan, R., Thomas, C., Groff, D., Van Delst, P., Collard, A., Kleist, D., Treadon, R., and Derber, J. C.: All-sky microwave radiance assimilation in NCEP's GSI analysis system, Monthly Weather Review, 144, 4709–4735, https://doi.org/10.1175/MWR-D-15-0445.1, 2016.