

## **Response to reviewer #1**

We received another reviewer's comments on this manuscript, and extra modifications were made. Therefore, this version of response letter has some differences compared with the previously uploaded one, but your valuable suggestions were respected and the revisions were made according to your general and specific comments. Please take this version as the final confirmation. We are sorry for the inconvenience.

### **General comment:**

This paper compares the FY-4A/AGRI 0.65-um visible reflectance (O) with the model simulations generated from CMA-MESO forecasts using the RTTOV (B). The potential sources contributing to the differences between O and B, such as the unresolved aerosol processes, the ice scattering models, are analyzed.

The paper is relevant to the cloud remote sensing field, as the growing international fleet of next-generation geostationary imagers can be expected to aid in our understanding of the diurnal cycles of clouds and aerosols. Well understood and characterized the biases of their observations will therefore be well received by the community. However, the authors make what I think are several unsubstantiated assertions (see my detailed comments). I recommend major revisions before reconsidering for publication. My general and specific comments are below.

### **Our response:**

The authors thank the anonymous reviewer for the constructive comments. We made major revisions to the manuscript, including the evaluation of the forecasts of CMA-MESO, discussions on the spatial and seasonal variations of the O-B biases, corrections on some typo errors (e.g., abbreviations, inappropriate usage of "evaluation", etc.). As a result, the outline of the manuscript was changed compared with the initial version. Some of the revisions were made according to your valuable comments. Some were made according to

the comments of another reviewer. The point-by-point response to the reviewer's comments were provided below.

**General Comment 1:**

A comparison with the model simulations cannot be called an “evaluation”, especially when the model simulations are not as accurate as expected. Currently, the RTTOV forward-operator for clouds and/or within the visible and shortwave infrared spectral ranges is still questionable, and the forecasts from CMA-MESO also lack adequate evaluations.

**Our response:**

Thank you for pointing this out. It is true that the reflectance simulated from the forecasts of CMA-MESO model using RTTOV cannot represent the true reflectance due to the deficiencies of the CMA-MESO and RTTOV models. Therefore, the title of this manuscript was changed to “Exploring the characteristics of FY-4A/AGRI visible reflectance using the CMA-MESO forecasts and its implications to data assimilation”. (L1-2)

**General Comment 2:**

As (1), if the authors persist in characterizing the biases of AGRI reflectance observations by comparing with the model simulations, the performances of RTTOV forward-operator and the forecasts from CMA-MESO should be evaluated first.

**Our response:**

We agree with you that the evaluation of the CMA-MESO model and the RTTOV-DOM forward operator is necessary when addressing the O (FY-4A/AGRI visible reflectance) - B (model simulations generated from CMA-MESO forecasts using the RTTOV) differences. Accordingly, two major revisions were made.

(1) Evaluation of the forecasts of CMA-MESO (L195-227)

Firstly, the one-hour accumulated precipitation was compared with the observations provided by the multi-source observed precipitation products in Chinese mainland. Good agreement between the simulations and observations was revealed, except that the precipitation areas were overestimated by the CMA-MESO forecasts in Chinese mainland. In addition, the precipitation was overestimated by the CMA-MESO forecasts. The evaluation results suggest that despite general agreement between the observations and simulations were revealed, the CMA-MESO forecasts suffer from deficiencies especially over complex terrain areas.

Secondly, the Probability density Distribution Functions (PDFs) of one-month Brightness Temperature (BT) for the FY-4A/AGRI channel 13 (10.30  $\mu\text{m}$  – 11.30  $\mu\text{m}$ ) observations and simulations was analyzed. The results were shown in Fig. 4. The PDF was underestimated at the high-BT end. In contrast, it was overestimated at the low-BT end. Since channel 13 is an infrared window channel, BT in cloudy areas is directly related to cloud top height. Therefore, the PDF analysis implies that high-level clouds were underestimated by CMA-MESO whereas low-level clouds were overestimated. The evaluation suggested deficiencies of the CMA-MESO model in forecasting high-level clouds.

## (2) Discussions on the uncertainties of RTTOV (L66-80, L341-385)

We feel helpless to evaluate the performance of the RTTOV model. The largest challenge comes from a lack of accurate observation of reflectance corresponding to the real atmosphere state variables. Therefore, instead of evaluating the performance of RTTOV by comprehensive observations, the performance of RTTOV and the uncertainties of the forward operator were discussed in the revised manuscript.

RTTOV was widely used to generate synthetic visible images. The synthetic images were further compared with satellite observed visible images to better understand the observation errors and representativeness errors and to provide guidance for the improvements of NWP models and forward operators. To save computational cost, a method for fast satellite image synthesis (MFASIS) was developed based on a lookup table (LUT) computed with one-dimensional (1D) solver of RTTOV in rotated Cartesian coordinates to account for some three-dimensional (3D) radiative effects (Scheck et al., 2016; Scheck et al., 2018). Intercomparison of satellite visible reflectance and the equivalents derived from NWP models and MFASIS indicated generally good agreement, and the Bidirectional Reflectance Distribution Function (BRDF) of land surface derived from a monthly mean atlas generated reasonable results in cloud-free conditions (Lopez

and Matricardi, 2022). Data assimilation of satellite visible reflectance data based on the MFASIS suggested positive impacts in real-world cases (Scheck et al., 2020). Since March 2023, satellite visible reflectance data have been operationally assimilated in German Weather Service by using the MFASIS forward operator. Although most of the studies are based on the MFASIS solver, the error estimates derived for MFASIS present upper bounds for RTTOV-DOM since the latter is just an emulator for the latter used in this study. Therefore, it is expected that RTTOV could generate reliable visible images if the NWP models were well tuned and the configurations of RTTOV were optimized.

However, knowledge on the cloud optical properties is scarce, which may introduce some uncertainties to the simulated reflectance. For example, the pre-assumed cloud particle size distribution (PSD) inherent in the cloud schemes in RTTOV is inconsistent with that of NWP models, not to mention the representativeness of the pre-assumed PSD in real cases. To illustrate this problem, a sensitivity study was performed by RTTOV configured with two different ice schemes, i.e., the Baum and Baran schemes. Distinct differences were revealed for the simulated reflectance, which confirms the uncertainties in the cloud optical properties of the RTTOV model.

According to the evaluation of the forecasts of CMA-MESO and discussions on the uncertainties of RTTOV, B derived from the CMA-MESO+RTTOV simulations will inevitably suffer from errors. Therefore, we have to admit that the O-B method is a measure of last resort due to a lack of sufficient reference observations for comparing with the satellite observations. Whether the bias correction brings benefits to the numerical weather prediction should be tested by data assimilation in real-world weather systems and should be evaluated by comprehensive observations.

### **General Comment 3:**

The bias characteristics are not well analyzed. (1) How about the spatial distributions or seasonal variations of AGRI biases? (2) Do they have differences before and after the FY-4A satellite's U-turn at the vernal and autumnal equinoxes?

### **Our response:**

(1) Spatiotemporal variation of the O-B biases (L228-261)

To better characterizing the spatiotemporal variation of the O-B biases, extra simulations were performed for March, June, and December. We did not perform the simulations for the whole 2020 year mainly because the radiative transfer simulation is rather computationally expensive. On our linux cluster which is equipped with 2.20 GHz Xeon Silver 4214 CPU, it will take approximately 30 min ~ 1 hour (32 CPUs for parallel computation) for the RTTOV-DOM (V12.3) to generate a synthetic visible imagery which includes  $2501 \times 1671$  pixels (the CMA-MESO grids). We think the results for March, June, September, and December could reveal some seasonal variation characteristics of the O-B biases.

For the spatial distribution of O-B biases, systematic biases were revealed over the Southern foothills of the Himalayas, the Sichuan basin, and the Yunnan-Kweichow Plateau, both in September (Fig. 5, L255) and in other months (Fig. S2 in the supplementary material). On one hand, some areas of the Qinghai-Tibet Plateau were covered with snow. Reflectance simulated in these areas should be less accurate compared with other places since the BRDF atlas is questionable in snow-covered areas (Ji et al., 2022). On the other hand, the performance of the CMA-MESO model was reduced over complex terrain areas. The analysis of the spatial distribution of O-B biases suggested that the snow-covered areas and complex terrain areas should be excluded in the following analysis, mainly because that it is clear that one cannot get reasonable results in these areas.

Based on the four-month simulations over the CMA-MESO domain, the spatial (Fig. 5 for the September (L255) and Fig. S2 for March, June and December in the supplementary material) and temporal (Fig. 7 for the September (L331) and Fig. S3-Fig. S5 for March, June and December in the supplementary material) variation characteristics of the O-B biases were explored. The results indicate different spatiotemporal variations of the O-B biases for the four months, which is closely related to the spatiotemporal of the performance of the CMA-MEOS model, the variation of aerosol properties, etc.

(2) Characteristics during U–turn at the vernal and autumnal equinoxes (Fig. 7 in the revised manuscript (L331) and Fig. S3 in supplementary material)

In the Northern Hemisphere, the vernal and autumnal equinoxes falls about 20 March and September 22 or 23, respectively. During this period of time, the Sun crosses the celestial equator, leading to changes in the sun-satellite geometries. Checking through the time series of the O-B biases in March and September, we see no differences during these two days or around when compared with other dates. Therefore, a tentative conclusion could be drawn that the temporal variations of the O-B biases do not have differences before and after the FY-4A satellite's U-turn at the vernal and autumnal equinoxes.

Nevertheless, we found an interesting phenomenon for the temporal variation of the O-B biased in June. An abrupt change was revealed on June 21th (Fig. S4(b)). The abrupt change was caused by the annular solar eclipse on 06:00 UTC 21 June 2020, when the incoming solar radiance was sheltered by the moon over the west parts of the CMA-MESO domain. The annular solar eclipse caused an abrupt decrease of the photons received by the AGRI visible channel. As a result, the visible image was darkened. The darkened visible image was also revealed by the National Aeronautics and Space Administration (NASA) worldview project (<https://worldview.earthdata.nasa.gov/>). However, the annular solar eclipse was not considered when performing the radiative transfer simulations by RTTOV-DOM. Instead, the incoming solar irradiance was set to a constant, which caused an abrupt decrease of the O-B biased. (L324-330)

### **Specific Comment 1:**

Lines 16, 22, 33 and 72: The abbreviations (FY, TOVS, and so on) should be given full name when first appeared in the abstract and text.

### **Our response:**

The abbreviations were checked throughout the revised manuscript, and the full names were given the first time they appeared in the abstract and text. E.g., FY is the abbreviation of "Fengyun" (L17, L35), and RTTOV is the abbreviation of Radiative Transfer for the Television infrared observation satellite Operational Vertical Sounder, etc. (L67-68, etc.)

**Specific Comment 2:**

Line 85: Himawari-8 satellite should be introduced because not all readers know it is the first one of the Japanese next-generation geostationary satellite.

**Our response:**

Corrected. (L135-136)

**Specific Comment 3:**

Line 82: How about the spatial coverage of CMA-MESO, or the region of interest in this study?

**Our response:**

In the revised manuscript, the spatial coverage of CMA-MESO was shown by Fig. 1, which is also the region of interest in this study. (L114-116)

**Specific Comment 4:**

Lines 96 and 117? Here, the authors give two cloud mask definitions. Which one will be used for Tables 1 and 2?

**Our response:**

We are sorry for not making it clear here. In the original manuscript, two cloud masks are defined for the synthetic images (observations, O) and observed images (simulations, B) separately. For the observed images, cloud masks were directly derived from the cloud mask products. For the synthetic imageries, cloud masks were dragonized from the CWP with a threshold value of  $0.01 \text{ kg m}^{-2}$ . For spatiotemporally collocated observations and simulations, the O-B biases were calculated for the pixels which are designated to be cloudy and cloud-free for both O and B. The O-B biases for the cloudy and cloud-free pixels were further used to correct the systematic biases of the corresponding scenarios separately.

According to another reviewer's comment, different definitions of cloud mask for observations and simulations could introduce mismatch of cloudy or cloud-free scenarios in the observed and simulated visible images. Therefore, an equivalent criterion of cloud mask for observed and simulated images was introduced to the revised manuscript. In the

revised manuscript, cloud mask was determined by comparing the simulated and observed reflectance with the reflectance simulated by ignoring cloud impacts. (L278-303)

For the synthetic visible image, a pixel was designated to be cloudy if the simulated reflectance  $r_{sim}$  satisfies Equation (4). Otherwise, the pixel would be classified to be cloud-free.

$$r_{sim} > r_{sim,clear} \quad (4)$$

where  $r_{sim,clear}$  denotes the simulated reflectance when cloud contributions were ignored.

The aerosol contributions were neglected by the simulations since the CMA-MESO forecasts do not provide aerosol information explicitly, whereas the observed reflectance inevitably includes aerosol contributions. Considering the aerosol contributions to the reflectance, a pixel is assumed to be cloudy if the observed reflectance  $r_{obs}$  satisfies Equation (5),

$$r_{obs} > r_{sim,clear} + r_{aer}^{75} \quad (5)$$

where  $r_{aer}^{75}$  denotes the aerosol contributions to the reflectance of cloud-free pixels, which was set to the upper quartile of  $r_{obs,clear} - r_{sim,clear}$  for the preliminarily estimated cloud-free pixels.  $r_{obs,clear}$  denotes the observed reflectance for cloud-free pixels, which were preliminarily determined by the FY-4A CLM product. The second-step estimate of cloud-free pixels was determined Equation (6),

$$r_{obs} < r_{sim,clear} + r_{aer}^{25} \quad (6)$$

where  $r_{aer}^{25}$  denotes the aerosol contributions to the cloud-free reflectance. Similarly,  $r_{aer}^{25}$  was set to the lower quartile of  $r_{obs,clear} - r_{sim,clear}$  for the preliminarily estimated cloud-free pixels. The two-step estimate of cloud mask for observed images was performed to maintain equivalent criterion of the cloud mask for synthetic images. It is noted that the first-step estimate of cloud mask should have different representativeness compared with the cloud mask diagnosed from Equation (4). For example, the CLM cloud mask was generated by including extra infrared observations (Wang et al., 2019) that are much more sensitive to optically thin cloud, which may appear to be transparent in the visible band. Nevertheless, the quartile estimation should mitigate the impacts. On one hand, thin clouds which are transparent in the visible channel whereas are opaque in the infrared channels should contribute insignificant part to  $r_{obs}$ . On the other hand, the quartile estimation in Equations (4) and (5) discarded 25% samples in estimating the lower and upper quartiles of  $r_{obs,clear} - r_{sim,clear}$ .



**Specific Comment 5:**

Lines 201-203: I can't understand this sentence. Aren't the "microphysical properties therein" "cloud variables"?

**Our response:**

We were intended to say that "Compared with the infrared and microwave radiance observations, the visible reflectance is much more sensitive to cloud variables, regardless of the type of cloud hydrometeors or the vertical location of clouds. In contrast, the infrared radiance data are only sensitive to cloud-top properties due to strong absorption effects (Li et al., 2022)". This part was revised in the revised manuscript (L422-424)

**Specific Comment 6:**

Figure 6: Readers can hardly identify the differences between observed and model simulated reflectance. The authors are suggested using a different colormap or adding figures to show their differences.

**Our response:**

Thanks for pointing this out. Since O-B is positively biased for the selected cases, reflectance of the bias-corrected visible image should be reduced by a factor of  $\gamma$ , which denotes the bias correction coefficient (Equation (4), L435). The bias-corrected visible image remains general characteristics of the original manuscript. As a result, the contour maps of the original image and bias-corrected image would be rather similar, except that the bias-corrected image was slightly darker than the original one. Therefore, it is difficult to differentiate the original and bias-corrected visible images in contour maps, and the contour maps in the original manuscript were deleted in the revised manuscript.

Instead of presenting the results by contour maps, the O-B biases with and without bias correction were summarized in Table 1 (based on deterministic forecasts) and Table 2 (based on ensemble forecasts). The results should be shown in a clearer way in the revised manuscript. (L454-458)

## Response to reviewer #2

### General Comments:

In this manuscript, differences between observed visible satellite images from the AGRI instrument onboard FY-4A and synthetic images computed with the RTTOV forward operator package from deterministic forecasts of the CMA-MESO model are discussed and a bias correction method is proposed. The authors address a relevant question. Fast forward operators have become available for the up to now underused visible satellite channels in the last years, several services have started monitoring experiments for these channels and at the German weather service visible reflectances are assimilated operationally since March 2023. So there is a clear interest in exploiting the cloud and aerosol information contained in visible channels for data assimilation and understanding systematic errors is an important first step.

While the topic is clearly relevant, the methods employed by the authors are in my opinion not really sufficient to get information on the origin of reflectance errors. The latter could be caused by the instrument (e.g. calibration problems), the model (e.g. deficiencies in the representation of clouds) or the forward operator (e.g. albedo errors or 3D effects). In particular, I think the authors skipped one first, important step: Comparing reflectances histograms for O and B. In contrast to O-B statistics, histograms are not sensitive to the location of the clouds (to correct that is the job of data assimilation) and can provide a lot of information (see e.g. Geiss et al 2021). I would also suggest to distinguish not only between cloudy and clear pixels, but also between regions for which we would expect different error characteristics. Given the CMA-MESO domain contains the Himalaya and oceans, it could make sense to distinguish between land, sea (lower albedo error) and extreme orography (potentially larger model and albedo errors). And finally, it would be important to exclude cases from the statistics, for which it is clear that one cannot get reasonable results. This would e.g. be pixel for which the BRDF atlas contains contribution from snow and thus no reliable albedo information is available.

The bias correction proposed in the last part of the manuscript is in principle interesting. However, the results are not discussed in a sufficiently clear way and I do not understand the motivation for the "ensemble" version.

### **Our response:**

Thank you for the instructive comments. According to your comments, major revisions were made to the manuscript, and the revisions include the following three main aspects.

(1) We added the analyses of the Probability distribution Density Functions (PDFs) for the reflectance and brightness temperature in the revised manuscript. (L200-216, L234-261)

The brightness temperature was analyzed for the FY-4A/AGRI channel 13 (10.30  $\mu\text{m}$  – 11.30  $\mu\text{m}$ ), which is an infrared window channel. Therefore, the brightness temperature could reflect cloud top height in cloudy regions. For the BT images generated by the CMA-MESO+RTTOV-DOM simulations, the PDF was underestimated at the high-BT end and overestimated at the low-BT end, implying that high clouds were underestimated or low clouds were overestimated. For the visible images generated by the CMA-MESO+RTTOV-DOM simulations, the low-reflectance end of the PDF was overestimated, which should be related to the underestimated cloud cover and neglected aerosol contributions. In the medium reflectance range, the PDF was underestimated by the synthetic visible images, which should be related to the deficiencies of the NWP models and the forward operators (potential biases in the cloud optical properties).

(2) In the revised manuscript, the O-B biases were explored for different scenarios by distinguishing not only between cloudy and cloud-free pixels, but also between sea and land surfaces. In addition, three sets of threshold tests were applied to exclude cases where reasonable results cannot be expected. (L262-334)

The first threshold test is terrain height test. Due to the deficiencies of the CMA-MESO model over the complex terrain areas, especially over the Qinghai-Tibet Plateau, a threshold test of terrain height ( $> 4.0$  km) was applied to exclude these areas.

The second threshold test is the land surface albedo test. Over land, the snow-covered areas were screened out by applying a threshold test of the surface BRDF (0.2/3.14). Here 0.2 is the surface albedo of typical snow-covered surfaces, and 0.2/3.14 denotes the BRDF for a Lambertian radiator

The third threshold test is the cloud-free reflectance test over sea areas. The highly reflective areas over ocean surface (reflectance  $> 0.1$ ) were excluded to reduce the sun-glint impacts.

(3) For the data assimilation of FY-4A visible reflectance data in real-world cases, it is necessary to correct the systematic biases in the observations in order to comply with the unbiased Gaussian PDF of the observation errors. By applying the first-order bias correction with a deterministic forecast, the bias correction was not that effective since the Gaussianness of the PDF was not well respected. In fact, some irregular structures were revealed in the PDF. We think an important reason is that the synthetic images were accompanied with some random errors in the deterministic forecasts from CMA-MESO. An ensemble forecast should mitigate some of the random errors in the synthetic images. After bias correction based on an ensemble forecast, the Gaussianness of the PDF for O-B differences were better respected, and the biases were reduced more evidently compared with the bias-corrected results based on a deterministic forecast. This is the reason why an ensemble forecast was introduced in comparison with the results with a deterministic forecast. (L417-470)

**Specific Comment:**

- l. 24: How do you mean "negatively biased"? In Fig. 2 and Tab. 1 O-B is in general positive.

**Our response:**

We made a typo error in the original manuscript. This mistake was corrected in the revised manuscript. (L25-26)

**Specific Comment:**

- l. 27: The standard deviations are actually not reduced significantly and that is also not what you would expect from a bias correction

**Our response:**

It is true that the effects of the bias correction method on the standard deviations were trivial. As a result, discussions on the standard deviations were deleted in the revised manuscript.

**Specific Comment:**

- Introduction: There are two papers comparing synthetic and observed visible reflectances that should be cited here, Geiss et al. (2021) and Lopez & Matricardi (2022).

**Our response:**

Thank you for recommending the two papers which are closely related to the topic in our manuscript. The two papers were cited in the introduction in other parts of the manuscript.

(L72-76)

**Specific Comment:**

- l. 57: "The key assumption [...] is that the model equivalents do not generate systematic biases" Well, this is just not a reasonable assumption! Systematic errors due to model and forward operator cannot be assumed to be smaller than instrument/calibration errors. What you can assume is that the different error contributions have different spatial and temporal characteristics (e.g. the jump in Fig. 2 is very likely related to the calibration changes), which can be used to identify likely error sources.

**Our response:**

We agree with you on this comment. Currently, it is difficult to illuminate the representativeness errors which are related to the NWP models and forward operators. One example of the representativeness error in B was caused by the annular solar eclipse on 21 June 2020 (Fig. S4(b)). The forward operator neglected the abrupt decrease of incoming solar irradiance, leading to underestimated reflectance in the synthetic image. Therefore, the sentence was replaced by "One assumption of the inter-comparison method is that the spatiotemporal characteristics of different error contributions differ so that the O-B analysis can be used to identify different error sources" in the revised manuscript. (L81-82)

**Specific Comment:**

- l. 61: "NWP model errors could be alleviated [] by temporally averaging several instants over a long period of time". Averaging over a long period can help to detect systematic errors, but it cannot "alleviate" them. No DA system works with time-averaged states...

**Our response:**

Thanks for pointing this out. This sentence was deleted in the revised manuscript. Averaging several instants over a long period of time is helpful to detect some systematic errors of the NWP models. Therefore, the spatial distribution of one-month mean precipitation and O-

B biases of reflectance was analyzed. The analysis revealed systematic errors of the CMA-MESO in overestimating the precipitation areas (Fig. 3, L209). In addition, analysis of the one-month mean O-B differences revealed systematic biases of CMA-MESO+RTTOV-DOM simulations over complex terrain areas (Fig. 5, L255).

**Specific Comment:**

- 2.1: The parameterization for unresolved sub-grid clouds was found to be very important in Geiss et al. (2021). How is CMA-MESO handling this and are the subgrid contributions included in the input data for RTTOV?

**Our response:**

Previous studies suggested that the parameterization for unresolved sub-grid clouds was critical to the simulated reflectance (Scheck et al., 2018; Geiss et al., 2021). In this study, the sub-grid clouds were approximated by the meso-SAS shallow-convective cumulus parameterization. The tendency equations of the grid-box mean moist static energy, water vapour mixing ratio, and vertical velocity were related to the transfer equations of related variables at sub-grid scale. The mixing ratio of cloud hydrometeors at sub-grid scale was generated by convective condensation with interactions to grid-scale processes considered. The spatial coverage of the sub-grid clouds within a grid box was depicted by cloud cover, which was diagnosed from the grid-scale humidity following Xu and Randall (1996). The cloud cover derived from the CMA-MESO forecast was included in the RTTOV input to account for the sub-grid contributions and the radiative transfer was solved by using the maximum random overlap method. (L117-125)

**Specific Comment:**

- 1. 95/ Fig. 5: For the Baum ice clouds you need also the effective ice particle radius -- which parameterization did you use? Do you have an idea why the effect visible in Fig. 5 is so much stronger than what Geiss et al. found for modified ice optical properties?

**Our response:**

The liquid and ice cloud optical properties in RTTOV were parameterized by the “Deff” scheme (Mayer and Kylling, 2005) and the Baran et al. (2014) scheme, respectively. The

effective radius of liquid water clouds ( $Re_{liq}$ ) was calculated following Thompson et al. (2004) and Yao et al. (2018). The effective radius of ice clouds ( $Re_{ice}$ ) was not calculated explicitly since the ice scheme developed by Baran et al. (2014) does not have a dependence on  $Re_{ice}$ . (L155-159)

In our manuscript, the reference run was configured with the Baran scheme, while the experiment run was configured with the Baum scheme which includes a general habit mixture (GHM) of ice crystals. In Geiss et al. (2021), the reference run was configured with the GHM model developed by Baum et al. (2014), whereas the experiment run was configured with the solid column scheme based on ice optical properties of Yang et al. (2005). The sensitivity study in our manuscript indicated that the impacts were especially apparent for optically thin clouds (reflectance < 0.2) (Fig. 8(b)) and extended to optically thick clouds. However, Geiss et al. (2021) suggested that changing the ice scheme from GHM to the solid-column scheme only affected the high-reflectance end of the PDF. We did not conduct an inter-comparison study of ice cloud schemes between the solid columns and GHM. But Baum et al. (2014) compared the ice cloud optical thickness retrieved based on the GHM and solid columns and indicated good consistency between two ice models due to their similar asymmetry parameters. However, there should exhibit distinct differences of the optical properties between the Baran scheme and Baum scheme. For example, the Baum scheme was developed based on nine basic ice habits whereas the Baran scheme involves only six ice habits. In addition, the PDFs and the mixing ratio of each habit are different between the two ice schemes. Therefore, the distinct differences between the Baran and Baum schemes should be the main cause to the larger differences than Geiss et al. (2021) between the reference run and experiment run. (L361-380)

### **Specific Comment:**

-1. 98/117: How well does "cloudy" in the CLM product correspond to  $CWP > 0.01\text{kg/m}^2$ ? This probably rather unclear (no pun intended). The CLM algorithm may even use infrared channels and could therefore "see" clouds that are actually transparent in the visible range (I have not checked that). A more reliable way to define cloudy pixels would be to use reflectance > threshold value (e.g.  $r > 0.2$  as in Geiss et al. 2021) or even better reflectance > clear sky

reflectance + threshold value. Exactly the same criterion could be applied to the observed and the synthetic reflectances.

**Our response:**

Thank you for the constructive suggestion to build an equivalent criterion of cloud mask for the simulations and observations. In the revised manuscript, cloud mask was determined by comparing the simulated and observed reflectance with the reflectance simulated by ignoring cloud impacts. (L278-303)

For the synthetic visible image, a pixel was designated to be cloudy if the simulated reflectance  $r_{sim}$  satisfies Equation (4). Otherwise, the pixel would be classified to be cloud-free.

$$r_{sim} > r_{sim,clear} \quad (4)$$

where  $r_{sim,clear}$  denotes the simulated reflectance when cloud contributions were ignored.

The aerosol contributions were neglected by the simulations since the CMA-MESO forecasts do not provide aerosol information explicitly, whereas the observed reflectance inevitably includes aerosol contributions. Considering the aerosol contributions to the reflectance, a pixel is assumed to be cloudy if the observed reflectance  $r_{obs}$  satisfies Equation (5),

$$r_{obs} > r_{sim,clear} + r_{aer}^{75} \quad (5)$$

where  $r_{aer}^{75}$  denotes the aerosol contributions to the reflectance of cloud-free pixels, which was set to the upper quartile of  $r_{obs,clear} - r_{sim,clear}$  for the preliminarily estimated cloud-free pixels.  $r_{obs,clear}$  denotes the observed reflectance for cloud-free pixels, which were preliminarily determined by the FY-4A CLM product. The second-step estimate of cloud-free pixels was determined Equation (6),

$$r_{obs} < r_{sim,clear} + r_{aer}^{25} \quad (6)$$

where  $r_{aer}^{25}$  denotes the aerosol contributions to the cloud-free reflectance. Similarly,  $r_{aer}^{25}$  was set to the lower quartile of  $r_{obs,clear} - r_{sim,clear}$  for the preliminarily estimated cloud-free pixels. The two-step estimate of cloud mask for observed images was performed to maintain equivalent criterion of the cloud mask for synthetic images. It is noted that the first-step estimate of cloud mask should have different representativeness compared with the cloud mask diagnosed from Equation (4). For example, the CLM cloud mask was generated by including



extra infrared observations (Wang et al., 2019) that are much more sensitive to optically thin cloud, which may appear to be transparent in the visible band. Nevertheless, the quartile estimation should mitigate the impacts. On one hand, thin clouds which are transparent in the visible channel whereas are opaque in the infrared channels should contribute insignificant part to  $r_{obs}$ . On the other hand, the quartile estimation in Equations (4) and (5) discarded 25% samples in estimating the lower and upper quartiles of  $r_{obs,clear} - r_{sim,clear}$ .

### **Specific Comment:**

- End of 2.1: Maybe here would be a good place to add some more information on RTTOV. In contrast to Anonymous Reviewer #1 I do not think RTTOV in the visible range "is still questionable". No forward operator is perfect, but there are several studies (Geiss et al 2021, Lopez & Matricardi 2022) that demonstrate the capabilities and discuss limitations of RTTOV for visible satellite channels. Moreover, RTTOV is used for the operational assimilation of the visible Meteosat SEVIRI channel at DWD, which is a clear indication that RTTOV produces reasonable results. While most of them are based on the MFASIS solver (Scheck et al. 2016, 2022), the latter is just an emulator for the DOM solver used in this study, so error estimates derived for MFASIS present upper bounds for DOM errors. (I guess MFASIS coefficients are not yet available for FY-4A).

### **Our response:**

Thank you for this suggestion. We added some comments on this topic in the introduction. (L64-80)

“The inter-comparison method was also applied to satellite visible channels (Geiss et al., 2021; Lopez and Matricardi, 2022; Lopez et al., 2022) to better understand the observation errors and representativeness errors and to provide guidance for the improvements of NWP models and forward operators. Most of the studies performed the radiative transfer simulations based on a software package termed the Radiative Transfer for the Television infrared observation satellite Operational Vertical Sounder (TOVS) (RTTOV) (Saunders et al., 2018). To save computational cost, a method for fast satellite image synthesis (MFASIS) was developed based on a lookup table (LUT) computed with one-dimensional (1D) solver of

RTTOV in rotated Cartesian coordinates to account for some three-dimensional (3D) radiative effects (Scheck et al., 2016; Scheck et al., 2018). To better simulate the tangent linear and adjoint models, a neural network-based forward operator was also developed based on RTTOV simulations (Scheck et al., 2021). Intercomparison of satellite visible reflectance and the equivalents derived from NWP models and MFASIS indicated generally good agreement, and the Bidirectional Reflectance Distribution Function (BRDF) of land surface derived from a monthly mean atlas generated reasonable results in cloud-free conditions (Lopez and Matricardi, 2022). However, neglecting aerosol contributions in the radiative transfer simulations would lead to systematic biases both in cloudy and cloud-free conditions (Geiss et al., 2021). Data assimilation of satellite visible reflectance data based on the MFASIS suggested positive impacts in real-world cases (Scheck et al., 2020). Since March 2023, satellite visible reflectance data have been operationally assimilated in German Weather Service by using the MFASIS forward operator. Existing studies imply the promising expectation that RTTOV could generate reliable visible images if the NWP models were well tuned and the model configurations were optimized.”

**Specific Comment:**

- 1. 109: It would be good to provide the wavelength

**Our response:**

Done. (L93)

**Specific Comment:**

- 1. 116: If you produce output with CMA-MESO every 15 minutes and FY-4A/AGRI starts scanning from the north of the disk to the south at the same times (I am not sure about the scanning strategy), the time difference is probably smaller than 7.5 minutes. Please check...

**Our response:**

The full-disk scanning cycle of AGRI is 15 minutes and the scanning usually starts at 00:00 UTC. In addition, the CMA-MESO forecasts were produced at hourly intervals (e.g., 04:00, 05:00, 06:00, ...). Therefore, the maximum allowable time differences between the FY-

4A observations and CMA-MESO forecasts are within 15 minutes to ensure the temporal match.  
(L166-170)

**Specific Comment:**

- Fig 2: Is all = cloudy + cloud-free? Then I do not understand how the bias for all can be larger than for cloudy. Or is the third category "uncertain" missing in the plot? Then it would be good to add it. Moreover, the number of pixels in cloudy/cloud-free/uncertain as a function of time would be interesting. Maybe this is helpful for the interpretation: If the jump on 9th of September is related to a changed calibration factor (the constant used to convert the number of detected photons to a radiance) then the change in reflectance bias in cloudy/cloud-free should be proportional to the mean reflectance in cloudy/cloud-free. The fact that the cloud-free bias increases indicates that either the calibration has in fact become worse on September 9th or that before a calibration-related bias compensated a clear-sky bias (e.g. due to an albedo bias).

**Our response:**

(1) In the original manuscript, all = cloudy + cloud-free + uncertain. In Fig. 7 of the revised manuscript, all = cloudy + cloud-free to avoid misleading. Fig. 7 includes three subpanels. The first panel showed the biases for all pixels (cloudy + cloud-free for both land and sea surfaces) and the corresponding number of pixels. The second panel showed the biases and the number of pixels for land surface, where results for cloudy and cloud-free pixels were presented separately. The third panel showed the biases and the number of pixels for sea surface, where results for cloudy and cloud-free pixels were presented separately. (L331-334)

(2) The abrupt change of the bias from September 8th to 9th was caused by the measurement calibration processes, which were confirmed by two facts. First, the O-B biases were positively related to the observed reflectance which is proportional to the calibration coefficient. Therefore, an abrupt of the domain-averaged observed reflectance was also revealed. Second, the calibration correction coefficients of FY-4A/AGRI channel 2 were updated by the National Satellite Meteorological Center (NSMC) of CMA at 02:00 UTC on 9 September 2020 (<http://www.nsmc.org.cn/nsmc/cn/news/103609.html>) (remember that both the observations and simulations were deployed at 06:00 UTC). (L316-321)

(3) Cloud-free biases were reduced after the calibration correction coefficients were updated (Fig. 7(b)), which confirms the effectiveness of the calibration processes. (L321-322)

**Specific Comment:**

- 1. 145: If I understand this correctly, your "ensemble" is just seven deterministic model states for different lead times. This is not what anybody in data assimilation would call an ensemble and you are definitely not evaluating an "ensemble forecast" (l. 146). This is really misleading and I also do not see why this "ensemble" should be interesting. If you compute the average of synthetic images for different lead times you will of course get a blurred version of the 3h lead-time image and increase the cloud cover. But no operational data assimilation system I know uses time-averaged states. Is the point here that in a real ensemble DA system you would use the "real ensemble" for the bias correction but as you have none available here you are using the set of deterministic model states, because it has similar properties (the clouds are not exactly at the same locations in all of the members)?

**Our response:**

It is noted that the ensemble forecast here could not represent a real ensemble in any operational ensemble DA systems. On one hand, the number of ensemble members is too small to fully represent the uncertainties of atmosphere states. On the other hand, a more commonly used way to generate an ensemble forecast is to add perturbations to the ICs and LBCs or to combine several forecasts with different combination of microphysical schemes (Li et al., 2015). The simplified ensemble forecast in this study was used mainly because none of a well-tuned ensemble forecast is currently available for the selected area. Nevertheless, synthetic visible images derived from the ensemble forecast should be accompanied with increased cloud cover since clouds are not exactly overlapped for different ensemble members. As a result, the number of matched pixels which are cloudy both for the observations and simulations would be increased, which benefited the bias correction in cloudy regions (see Section 5 for more details). In a real ensemble DA system, a real ensemble would be adopted for the bias correction. (L139-148)

**Specific Comment:**

- 1. 147 / 211: Using the overbar for both ensemble mean and spatial average is confusing.

**Our response:**

In the revised manuscript, the overbar specifically denotes the domain average. Synthetic image generated by an ensemble forecast was explicitly noted in the text. (L431, L473-484)

**Specific Comment:**

- Table 1 (deterministic forecast): I am missing a better discussion of the results. The cloudy biases are reduced by 40-60%, but in all cases the bias correction actually \*increases\* the bias for the clear pixels. For SEP 15 & 17 the bias correction changes the sign of the bias for the clear pixels to negative. As the bias is positive for the cloudy pixel, this leads to compensating biases for the "all" category. Why is the bias correction ineffective for the clear pixels?

**Our response:**

The bias correction method used the systematic biases derived from the cloud-free (or cloudy) pixels for both O and B to estimate error characteristics of cloud-free (or cloudy) pixels only for O. Apparently, the cloud-free (or cloudy) pixels both for O and B are only a subset of those only for O. Therefore, the performance of the bias correction is determined by the representative of the subset of cloud-free (or cloudy) pixels to the corresponding cloud-free (or cloudy) pixels only in the observed images. (L437-442)

The bias correction was tested by two selected cases on September 15th and 17th, 2020. The case on September 1 was deleted due to the errors in observations before September 8<sup>th</sup>. For the ensemble forecast, the synthetic image was generated by averaging the seven visible images simulated from seven ensemble members. Cloud mask was determined by Equation (4) except that  $r_{sim}$  and  $r_{sim,clear}$  denotes the reflectance of the ensemble mean synthetic image. For the bias correction based on deterministic forecasts, O-B biases were reduced in most cases, but increased biased were also revealed on September 17<sup>th</sup> for cloudy regions over sea (Table 1). In contrast, the bias reduction was especially effective when B was derived from ensemble forecasts (Table 2). Since the synthetic image for an ensemble forecast would increase cloud

cover compared with a deterministic forecast, the number of the matched cloudy pixels was increased for an ensemble forecast. As a results,  $\gamma$  derived from ensemble forecasts should represent cloudy bias characteristics better than a deterministic forecast and vice versa, which explains why the biases were increased in some cases based on deterministic forecasts. In cloud-free regions, the original O-B biases were trivial, and the bias correction in cloud-free regions reduced the O-B biased to almost zero. (L443-458)

**Specific Comment:**

- l. 253: What does "cloudy" mean for the ensemble? That in the ensemble mean state  $CWP > 0.01 \text{ kg/m}^2$ ? So potentially clear member pixels contribute to  $\bar{B}_{\text{cld}}$ ?

**Our response:**

For the ensemble forecast, the synthetic image was generated by averaging the seven visible images simulated from seven ensemble members. Cloud mask was determined by Equation (4) (L284) except that  $r_{\text{sim}}$  and  $r_{\text{sim,clear}}$  denotes the simulated reflectance from the ensemble mean synthetic image. (L444-446)

**Specific Comment:**

- l. 304ff: There are two YZs -- maybe use the full names.

**Our response:**

The first YZ is Yongbo Zhou, and the second YZ is Yuefei Zeng. We used the full names in the author contribution. (L536-541)

**References:**

- Geiss et al. (2021): "Understanding the model representation of clouds based on visible and infrared satellite observations", ACP, Volume 21, issue 16, <https://doi.org/10.5194/acp-21-12273-2021>

- Lopez & Matricardi (2022): "Validation of IFS+RTTOV/MFASIS0.64- $\mu\text{m}$  reflectances against observations from GOES-16, GOES-17, MSG-4 and Himawari-8", ECMWF Technical

memorandum, DOI 10.21957/14u0f56lm, <https://www.ecmwf.int/en/elibrary/81322-validation-ifsrttovmfasis064-mm-reflectances-against-observations-goes-16-goes-17>

- Scheck et al. (2016): "A fast radiative transfer method for the simulation of visible satellite imagery", *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 175, 54-67, <https://doi.org/10.1016/j.jqsrt.2016.02.008>.

- Scheck, L., Weissmann, M., and Bernhard, M. (2018): "Efficient Methods to Account for Cloud-Top Inclination and Cloud Overlap in Synthetic Visible Satellite Images", *J. Atmos. Ocean. Tech.*, 35, 665-685, doi: 10.1175/JTECH-D-17-0057.1

- Scheck (2020): "A neural network based forward operator for visible satellite images and its adjoint", *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 274, November 2021, 107841, <https://doi.org/10.1016/j.jqsrt.2021.107841>

**Our response:**

All the papers you recommended were cited in the revised manuscript.