General.

We would like to thank the anonymous Referee #2 for providing very valuable comments to improve and clarify our manuscript. Many of the questions regarding the uncertainty calculation are related. Therefore, our responses to various questions contain cross-references. Please find our responses to the specific comments and questions below. Our response is written in bold. The revised parts of the manuscript are highlighted in bold red.

Comments of Referee #2 and our responses to them

General comments

The authors develop approaches for generating XCH4 time series over ocean combining ship and aircraft measurements with model data. The observation-based XH4 data are compared with independent TCCON measurements and finally used for evaluating GOSAT measurements. The paper is well written and within scope of AMT.

I have some minor comments related to the uncertainty calculation that should be addressed in a revised manuscript:

Specific comments

Comment 1

L82: What would be the required accuracy for a dataset assess the accuracy of trends and variations in XCH4 satellite observations over oceans? Moreover, what is the accuracy that is achieved with the dataset presented in this study?

Response

In order that our dataset is useful for accessing the accuracy of trends and variations in satellite data, the uncertainty of our reference dataset should be lower than that of the satellites.

For GOSAT, launched in 2009, the target for CH_4 was a relative accuracy of 2% for 3-month averaged data within a 1,000^2 km^2 grid. The target was achieved in 2010. This accuracy is suitable for research on global phenomena and for getting a better understanding of carbon cycles (Nakajima et al., 2010).

The mission targets for TROPOMI for the total column of CH_4 are a systematic error (bias) of less than 1.5% and 1% precision (ESA, 2017). Because the accuracy is determined by both, the bias and precision, it would be in the range of 1.8% using Gaussian Error propagation. This corresponds to concentrations of around 30 ppb.

Higher accuracy is required for the estimation of regional sources and sinks, for example for political decision making related to global warming countermeasures. Thresholds are given for land observations, and they are much higher with a precision of < 34 ppb for a single observation and < 11 ppb for monthly averaged data within 1000^2 km^2 grid. The systematic error after bias correction should be < 10 ppb (Buchwitz et., al, 2020).

In this context, GOSAT-2 was launched in 2018 with the aim for improved concentration precision of 5 ppb for monthly averaged CH_4 data at 500^2 km^2 grid over land and 2000^2 km^2 grid over the ocean (Nakajima et al., 2017).

Given the above mission targets of GOSAT and TROPOMI, our dataset with a conservative estimated uncertainty of 16 ppb fulfils the requirement.

Besides the uncertainty, the long-term availability of a reference dataset is important in areas where no other long-term datasets are available. Therefore, in regions like the open ocean, a reference dataset with even a high uncertainty is useful to fill in gaps where other highly precise reference data, such as from measurement campaigns like HIPPO flights or TCCON stations, are not available. Even though the reference has a relative high uncertainty, spatial pattern and large-scale long-term trends can be evaluated. Our dataset is not suitable for detecting small scale variations like those from point sources and sinks.

We clarified the requirement as follows:

Lines 83–87: *We propose a new approach to assess the accuracy of satellite derived XCH_4 trends and variations over open ocean regions by combining commercial ship and various aircraft observations with the help of atmospheric chemistry models.* ***We are targeting an accuracy better than that required for the GOSAT and TROPOMI mission of <35 ppb (<2%) (ESA, 2017; Nakajima et al., 2010). Our*** *approach was successfully applied to the evaluation of satellite XCO_2 previously (Müller et al., 2021).*

Lines 621–623: ***ESA, European Space Agency: Sentinel-5 Precursor Calibration and Validation Plan for the Operational Phase, Issue 1, Revision 1, 26 pp., https://sentinel.esa.int/documents/247904/2474724/Sentinel-5P-Calibration-and-Validation-Plan.pdf, 2017, accessed on 28 November 2023.***

Lines 690–692: ***Nakajima, M., Kuze, A., Kawakami, S., Shiomi, K., and Suto, H.: Monitoring of the greenhouse gases from space by GOSAT, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 38, 94–99, 2010.***

Comment 2

Figure 1 could already be mentioned in the beginning of Section 2.

Response

We added references to Figure 1 in section 2 as follows:

Lines 93–95: *As part of Japan's Comprehensive Observation Network for Trace gases by Airliner, CONTRAIL, air samples of* CH_4 *are collected by the Automatic air Sampling Equipment (ASE) and Manual air Sampling Equipment (MSE) about twice a month between Japan, Hawaii, and Australia since 2005.* ***The sampling locations of the CONTRAIL data are shown in Fig. 1.***

Lines 122–124: *In this study, we used* CH_4 *observations by the cargo ship Trans Future 5 (TF5, Toyofuji Shipping Co., Ltd.), which sails between Japan, Australia, and New Zealand* ***(Fig. 1)****.*

Comment 3

Section 3.3.1: The calculation of the tropospheric uncertainty of XCH4 is difficult to judge mainly because no profiles are shown in the manuscript. Please add a figure comparing the constructed CH4 profiles from measurements and the MIROC4-ACTM model with the HIPPO profiles.

Response

Thank you for the comment.

Below, we show the comparison between MIROC4-ACTM and HIPPO 4 profiles under a) in Fig. 2 and Table 1.

The comparison between MIROC4-ACTM and obs. CH_4 **profiles is shown under b) in Fig. 3 using some example profiles, and Table 2.**

However, we cannot show the direct comparison between HIPPO 4 profiles and the constructed obs. CH_4 **profiles. You can find our explanation under the response to Comment 4.**

For the comparison, we selected 8 HIPPO profiles, which are within 2000 km distance of the centre location of the bounding box g2. See Fig. 1 below. We like to clarify that in the original manuscript mistakenly only 6 profiles were selected. This is corrected in the revised manuscript, and changes resulting from this are listed under our response to Comment 6.

More details about the reason for the selection of these 8 profiles are found under our response to Comment 5.

a) MIROC4-ACTM versus HIPPO 4

Figure 1 below shows the location of the profiles within the 2000 km buffer, and Fig. 2 the HIPPO 4 profiles in comparison with those for the MIROC4-ACTM on July 3 and 6, 2011, respectively. We added Fig. 2 to Appendix A

of the manuscript as Fig. A1. The average difference ± standard deviation of the differences and root-mean-square error (RMSE) between the profiles were 6 ± 5 ppb (RMSE = 8), 6 ± 10 ppb (RMSE = 12), 6 ± 12 ppb (RMSE = 13) for the altitude ranges 0-1500 m, 1500-6000 m, and 6000-11000m, respectively (Table 1).

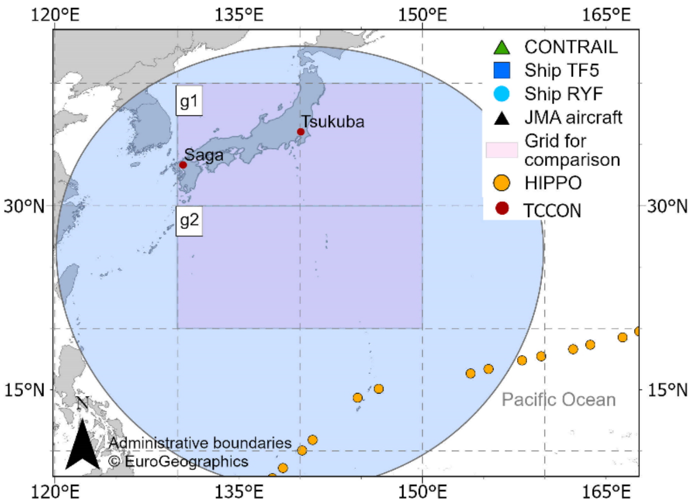

Figure 1. Location of selected HIPPO 4 profiles on July 3 and 6, 2011 within 2000 km distance of the centre location of bounding box g2.

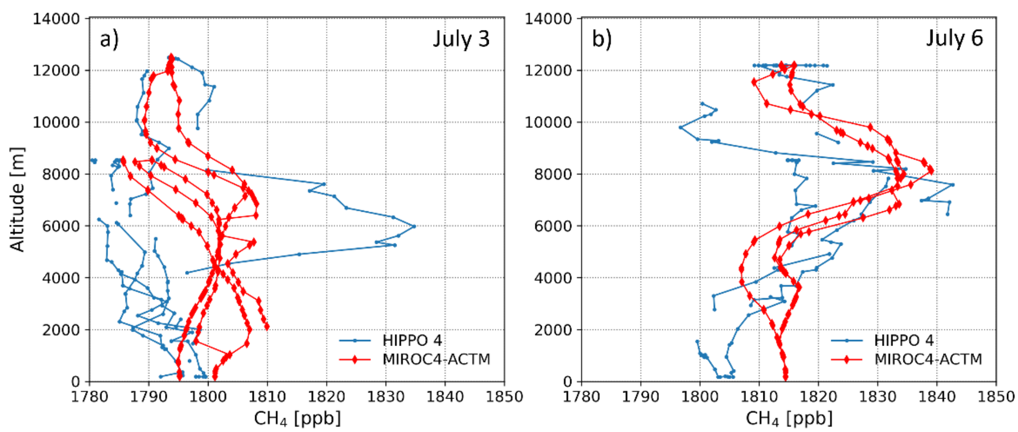

Figure 2. Comparison between HIPPO 4 (blue) and MIROC4-ACTM profiles (red) on July 3 (a) and 6 (b), 2011.

Table 1. Average difference between MIROC4-ACTM (ACTM) and HIPPO 4 data (mean difference ± standard deviation of differences) and root-mean-square error (RMSE) at different altitude ranges.

Altitude [m]	ACTM – HIPPO 4 [ppb]	RMSE [ppb]
0-1500	6 ± 5	8
1500-6000	6 ± 10	12
6000-11000	6 ± 12	13

b) MIROC4-ACTM versus obs. CH4 profiles

Figure 3 shows the MIROC4-ACTM (black) and the constructed obs. CH_4 profiles. For illustration, only some examples are shown. Table 2 lists the mean difference ± standard deviation of differences and the RMSE for each altitude range.

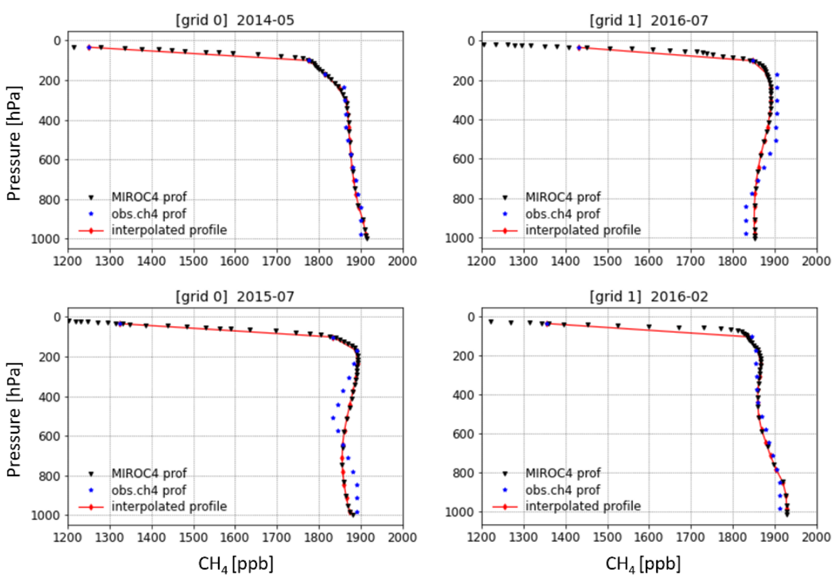

Figure 3. Comparison MIROC4-ACTM (MIROC4 prof) with obs. CH_4 profiles. The profiles are examples. The red curve is the MIROC4-ACTM profile interpolated on the pressure grid of the obs. CH_4 profile.

Table 2. Average difference between MIROC4-ACTM (ACTM) and obs. CH_4 profile data (mean difference ± standard deviation of differences) and root-mean-square error (RMSE) at different altitude ranges

Altitude [m]	ACTM – obs. CH_4	RMSE [ppb]
0-1500	4 ± 18	18
1500-6000	−3 ± 17	17
6000-11000	5 ± 16	18

Comment 4

It would also be interesting to see how well your approaches can reconstruct a HIPPO profile when taking the three measurements (2 aircraft + 1 ship) from the HIPPO profile.

Response

We agree with the referee #2 that this comparison would be very interesting and important to evaluate our approach. However, there are several reasons why we cannot provide a reasonable comparison.

The reasons are as follows.:

- **The HIPPO profiles are obtained only on 2 days of July 2011 (July 3 and 6) at specific locations.**
- **In contrast, our approach is based on monthly averaged data within a 10° latitude by 20° longitude grid. If we increase the sampling frequency to, for example, ± 2 days within 1 degree of the HIPPO profiles or higher, we won't have enough in situ data to apply our approach.**
- **Furthermore, our current data processing was for the years 2014–2017 for the latitude range north of the HIPPO flights. One reason for selecting that location was that we have additional JMA data for ship and aircraft. These data are missing at the location of the HIPPO flights, which makes the number of in situ data even less.**
- **If we apply our approach of monthly averages using the 10° latitude by 20° longitude grid for July 2011, we will not be able to reproduce the strong variation of a specific HIPPO profile of a single flight or day.**

However, if in future more in situ data are available, and new profile flights are performed, we agree that this comparison is very important!

Comment 5

L249: Can you explain why you only used only six profiles for assessing the MIROC4-ACTM simulations, while 20 profiles seem to be available in the study region?

Response

First, we want to clarify that 8 HIPPO profiles should have been included and not 6, which was a mistake in our calculation previously. Second, we noticed that the root-mean-square error (RMSE), which is as a measure of the differences between the model and the in situ observations, is a better and straight forward measure to access the uncertainty instead of using the average ± standard deviation of the differences. Therefore, the uncertainty numbers changed in the revised manuscript. The changes made are listed under the response to Comment 6.

Our study region is influenced by the continental emission outflow. These conditions are expected to be more challenging to be represented by the model correctly. To assess the uncertainty of the MIROC4-ACTM for conditions similar to our study region, we selected HIPPO flights within 2000 km of the centre location of the grid box g2. We chose the 2000 km threshold as a balance between closeness of the profile flights and number (Fig. 1).

Comment 6

L255: Finally, it is unclear how the uncertainty in the profiles translate to XCH4 uncertainties. Do you do an error propagation or sensitivity study as for the tropopause uncertainty?

Response

Using the 8 selected HIPPO 4 profiles as a reference, we can only access the tropospheric uncertainty of our constructed profiles indirectly. We used Gaussian Error Propagation as described as follows:

1)

First, we assess the uncertainty of the MIROC4-ACTM profiles by calculating the difference between MIROC4-ACTM and HIPPO 4 to derive the RMSE as measure for the uncertainty of the model. Here we call it "ACTM_unc".

2)

Second, we estimate the uncertainty of our obs. CH_4 profile in 2 steps:

a) We calculate the difference between obs. CH_4 profile and MIROC4-ACTM profile and obtain the RMSE as part of the total tropospheric uncertainty of the obs. CH_4 profile.

b) The total uncertainty consists of the partial uncertainty a) + that of the ACTM model (ACTM_unc) from step 1. It can be calculated using Gaussian Error Propagation. The results are shown in Table 3.

Table 3. Root-mean-square error of the difference between MIROC4-ACTM (ACTM) and HIPPO 4, and MIROC4-ACTM and obs. CH_4 profile data at different altitude ranges. Last column shows the total uncertainty after Gaussian Error propagation. Uncertainties applied to approach 3 are shown in bold.

Altitude [m]	**ACTM – HIPPO 4**	**ACTM – obs. CH_4**	**Total uncertainty**
0-1500	8	18	**20**
1500-6000	12	17	**21**
6000-11000	**13**	18	22

For approach 3, MIROC4-ACTM data are used at the altitude range 6000-11000 m. Therefore, no error propagation was applied and only the ACTM_unc was used for that altitude range (RMSE = 13). Uncertainties of approach 3 are shown bold in Table 3.

We revised Table 1 in the manuscript as follows:

Table 1: Uncertainty assessment of the obs. CH_4 profiles at the troposphere. Top rows: average concentration range of CH_4 within each HIPPO 4 profile (mean variability ± standard deviation). Bottom rows: **Root-mean-square error (RMSE) of the difference between MIROC4-ACTM (ACTM) and HIPPO 4, and MIROC4-ACTM and obs. CH_4 profile data at different altitude ranges. The last column shows the total uncertainty after Gaussian Error propagation. Uncertainties applied to approach 3 are shown in bold.**

HIPPO 4 profile range [m]	Variation within profiles [ppb]
~300–~13000	24 ± 17

Altitude [m]	ACTM – HIPPO 4 [ppb]	ACTM – obs. CH_4 [ppb]	Total uncertainty [ppb]
0–1500	8	18	20
1500–6000	12	17	**21**
6000–11000	**13**	18	22

A detailed description of our uncertainty estimation in the troposphere is added as follows:

Lines 260–274: *Second, we assessed* ***the uncertainty of the constructed CH_4 profiles in 3 steps with the help of the MIROC4-ACTM. In the first step, we investigate*** *how good the MIROC4-ACTM reproduces the variation of HIPPO profiles* ***for similar conditions to our study region, which is influenced by the continental emission outflow (Appendix A, Fig. A1). Therefore, we selected 8 profiles within 2000 km of the center location of g2*** *(**Fig. 1**). We choose the MIROC4-ACTM to be consistent with our previous study (Müller et al., 2021). We distinguished the altitude range 0–1500 m, corresponding to the boundary layer, 1500–6000 m, corresponding to the middle troposphere between the extrapolated ship and JMA aircraft data, and 6000–11000 m, corresponding to the upper troposphere between the JMA and CONTRAIL aircraft data.* ***As model uncertainty, we obtain t****he* ***root-mean-square error (RMSE) of the*** *difference between the MIROC4-ACTM and the HIPPO profiles* ***with 8*** *ppb,* ***12*** *ppb, and* ***13*** *ppb for the altitude ranges 0–1500 m, 1500–6000 m, and 6000–11000 m, respectively (**Table 1**).* ***In the second step, we compare the MIROC4-ACTM with our obs. CH_4 profiles and obtain the RMSE (Table 1, ACTM – obs. CH_4). Because the model itself has an uncertainty as obtained in step 1, the tropospheric uncertainty of the constructed profile of each altitude range is 20 ppb, 21 ppb, and 22 ppb using Gaussian Error propagation (Table 1, Total uncertainty). As a result****, we added* ***21*** *ppb uncertainty between the extrapolated ship and JMA data* ***in approach 2 and 3****, and* ***22*** *ppb* ***and 13 ppb*** *between the JMA data and up to the TROPPB in approach 2 and 3****, respectively****.*

The updated uncertainty values are added as follows:

Line 23: *Uncertainties were 22 ppb for approach 1,* ***20 ppb*** *for approach 2****,*** *and* ***16 ppb for approach*** *3.*

Lines 328–329: *The uncertainty range of the simple obs. XCH_4 (22 ppb) is by* ***2 and 6*** *ppb larger than those of the obs. XCH_4* ***(20 ppb)*** *and blended obs. XCH_4 (1**6** ppb)**, respectively** (**section 3.3**).*

Lines 365–367: *Given the lower maximal possible averaged difference between TCCON and approach 2 and 3 compared to approach 1,* ***and given the lowest uncertainty range of approach 3,*** *the latter approach* ***is*** *preferable for future applications.*

Lines 430–431: *The retrievals mostly lie in the uncertainty range (1**6** ppb) of the blended obs. XCH_4.*

Lines 454–455: *Uncertainties of the calculated XCH_4 were reduced by* ***2*** *ppb* ***and 6 ppb*** *from 22 ppb (approach 1) to* ***20*** *ppb for approach 2 and* ***16 ppb for approach 3.***

The Figure numbers of the Appendix changed as follows:

Lines 180–182: *A comparison with the RemoTeC v2.4.0 full-physics retrieval operated at Heidelberg University is shown in* ***Appendix A (Fig. A4)****.*

Lines 296–297: *GOSAT NIES CH_4 observations have a higher sensitivity in the stratospheric column as compared to CO_2 (averaging kernel >0.8 in the stratosphere,* ***Appendix A, Fig. A2****).*

Lines 308–309: *CAMS was positively biased by 138 ± 9 ppb, and 165 ± 15 ppb at 30–40° N and 20–30° N, respectively (**Table 2, Appendix A, Fig. A3 (a), (b)**).*

Lines 311–313: *The highest average difference occurred in June (30–40° N: 37 ± 6 ppb, 20–30° N: 44 ± 3 ppb), the lowest in October (4 ± 0.6 ppb) at 30–40° N, and January (5 ± 5 ppb) and February (3 ± 13 ppb) at 20–30° N (**Appendix A, Fig. A3 (c), (d)**).*

Furthermore, Table S5 and S6 of the supplement are updated with the new uncertainties of the obs. XCH_4 of 20 ppb, and the blended obs. XCH_4 of 16 ppb.

Comment 7

L337: If you have three simulations and two agree with each other, it is not valid to conclude that the agreeing models are correct. The conclusion in this paragraph need therefore to be argued more carefully using previous results from literature (as done in Section 3.3.3) or conducting additional analyses (e.g., comparison with independent measurements).

Response

We agree with the referee that our argumentation needs to be more careful.
We added the clarification as follows:

Lines 370–380: ***Figure 5*** *shows the comparison of the blended obs. XCH_4 (approach 3) using the MIROC4-ACTM, CAMS, and CAMSinv for the stratospheric column (**section 3.3.3**), denoted as $ACTM_{XCH4}$, $CAMS_{XCH4}$, and $CAMSinv_{XCH4}$. Using $ACTM_{XCH4}$ as reference, $CAMS_{XCH4}$ is highly biased at both latitude ranges by 12 ± 5 ppb (0.6 ± 0.2%) in total. In contrast, $CAMSinv_{XCH4}$ shows a small negative total bias of −5 ± 3 ppb (−0.3 ± 0.2%).* ***CAMS has a known large positive stratospheric CH_4 bias (Agustí-Panareda et al., 2023). MIROC4-ACTM and $CAMS_{inv}$ account for stratospheric CH_4 loss and the modelled stratosphere is comparable as discussed in section 3.3.3.*** *The similarity of* ***the $ACTM_{XCH4}$ and $CAMS_{invXCH4}$*** *and their differences to $CAMS_{XCH4}$* ***indicate*** *the strong impact of the stratospheric part on the derived XCH_4 and highlights the importance to make an appropriate model choice.* ***Considering the large uncertainty of CAMS and the fact that the other two products are better optimized for modelling CH_4 in the stratosphere****, we suggest using either the MIROC4-ACTM or CAMSinv to model the stratospheric column.*

Furthermore, we made the following correction in section 3.3.3. CAMSinv is "inversion-optimized for greenhouse fluxes and concentrations" but doesn't use a better atmospheric transport model than CAMS. We revised the sentence as follows:

Lines 315–317: *Compared to CAMS, both the MIROC4-ACTM and CAMSinv account for chemical losses in the stratosphere****. Additionally, MIROC4-ACTM uses an optimized atmospheric transport model (Patra et al., 2018)***.

Lines 468–471: *MIROC4-ACTM and CAMSinv consider chemical losses in the stratosphere,* ***where MIROC4-ACTM additionally uses an optimized atmospheric transport model. We conclude that for accurately deriving XCH_4, a well modelled stratosphere is necessary that includes CH_4 sinks. Therefore,*** *either CAMSinv or MIROC4-ACTM is suitable for our approach of which CAMSinv is publicly available.*

Reference

Agustí-Panareda, A., Barré, J., Massart, S., Inness, A., Aben, I., Ades, M., Baier, B. C., Balsamo, G., Borsdorff, T., Bousserez, N., Boussetta, S., Buchwitz, M., Cantarello, L., Crevoisier, C., Engelen, R., Eskes, H., Flemming, J., Garrigues, S., Hasekamp, O., Huijnen, V., Jones, L., Kipling, Z., Langerock, B., McNorton, J., Meilhac, N., Noël, S., Parrington, M., Peuch, V.-H., Ramonet, M., Razinger, M., Reuter, M., Ribas, R., Suttie, M., Sweeney, C., Tarniewicz, J., and Wu, L.: Technical note: The CAMS greenhouse gas reanalysis from 2003 to 2020, Atmos Chem Phys, 23, 3829–3859, https://doi.org/10.5194/acp-23-3829-2023, 2023.

Buchwitz, M., Chevallier, F., and Marshall, J.: User Requirements Document (URD) - Greenhouse Gases (GHG), Simulation, 42, 2020.

Müller, A., Tanimoto, H., Sugita, T., Machida, T., Nakaoka, S., Patra, P. K., Laughner, J., and Crisp, D.: New approach to evaluate satellite-derived XCO2 over oceans by integrating ship and aircraft observations, Atmos. Chem. Phys., 21, 8255–8271, https://doi.org/10.5194/acp-21-8255-2021, 2021.

Nakajima, M., Kuze, A., Kawakami, S., Shiomi, K., and Suto, H.: Monitoring of the greenhouse gases from space by GOSAT, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 38, 94–99, 2010.

Nakajima, M., Suto, H., Yotsumoto, K., Shiomi, K., and Hirabayashi, T.: Fourier transform spectrometer on GOSAT and GOSAT-2, in: International Conference on Space Optics — ICSO 2014, 2, https://doi.org/10.1117/12.2304062, 2017.

Patra, P. K., Takigawa, M., Watanabe, S., Chandra, N., Ishijima, K., and Yamashita, Y.: Improved Chemical Tracer Simulation by MIROC4.0-based Atmospheric Chemistry-Transport Model (MIROC4-ACTM), SOLA, 14, 91–96, https://doi.org/10.2151/sola.2018-016, 2018.

ESA, European Space Agency: Sentinel-5 Precursor Calibration and Validation Plan for the Operational Phase, Issue 1, Revision 1, 26 pp., https://sentinel.esa.int/documents/247904/2474724/Sentinel-5P-Calibration-and-Validation-Plan.pdf, 2017, accessed on 28 November 2023.