

Response to reviewer comments on Enders et al., 2025: „Towards routine shipborne measurements of columnar CO₂, CH₄, CO, and NO₂: a case study for tracking regional-scale emission patterns“

Reply to reviewer 1:

The questions of the reviewer are shown in italics. The list of references can be found in the main manuscript.

I am grateful for the opportunity to review this manuscript, which contains thorough methodology and suggests a potential way to expand trace gas column measurements to the ocean. I believe that this study offers a novel contribution to this field with great promise for future developments that would be valuable for atmospheric remote sensing. The authors describe their efforts to test methods for shipborne observations with a modified EM27/SUN solar viewing spectrometer and coupled DOAS instrument. Both instruments share a uniquely designed solar tracker that has been described and tested in previous literature. They demonstrate the capability of these instruments to measure total column dry-air mole fractions of CO₂, CH₄ and CO (XCO₂, XCH₄, and XCO) with the EM27, as well as partial columns of NO₂ with the DOAS instrument, on board a ship that is in motion. Their methods for ensuring the performance of these instruments and the quality of the data collected are thorough and well established by previous work.

They demonstrate that their shipborne measurements may be useful for quantifying emissions from coastal urban areas through one instance of an observed enhancement. Although this may be sufficient as a proof of concept, a more robust analysis of multiple observed enhancements, perhaps requiring a longer campaign, is warranted to get a better idea of the efficacy. It is unclear how often the conditions would be right to observe enhancements that could be reasonably attributed to specific sources or what revisit times would be required. In the discussion here, the source attributions are fairly speculative and there may be other interpretations for the results.

While it is true that these types of shipborne measurements have the potential for use in satellite validation in coastal regions or even over the open ocean, this use of the data is not demonstrated in the paper. For the purposes of expanding these efforts for satellite validation over the open ocean, there are some other considerations for satellite validation that are not discussed, such as coincidence criteria. This should probably be noted as follow up for a future study, instead of claiming that it was proved in this manuscript.

Overall, I would suggest this manuscript for publication after some revisions and careful attention to the confidence with which some conclusions are presented.

We thank the reviewer for the appreciation of our work and the helpful comments.

In general, we would like to point out that the purpose of this publication is to discuss the instrumentation and the retrieval setup in detail and to showcase the potential of the setup for collecting local emission information through a case study. A major factor preventing a larger number of emission episodes is the fact that the ship mostly traveled during the night and was moored in the

harbor for loading operations during the day. Since we use the sun as a light source, we cannot measure at night. Thus, the current ship deployment is not optimal, and we are in the process of identifying better opportunities. Nonetheless, a more in-depth analysis of data from the whole campaign (including another deployment on the same ship in 2025) is currently ongoing. Such follow-up studies will refer to the present study for the technical details and put more focus on the use cases.

The suitability of the setup (without NO₂) for satellite validation has been shown in our previous studies (Klappenbach et al., 2015, Knapp et al., 2021), where we compared it to GOSAT and S5P/TROPOMI data. But, it is true that the present study does not add new material on satellite validation and thus, we removed the corresponding statements from the abstract and modified the “conclusion and outlook” section of the manuscript (see also reply to the question on P25, L472-473).

P 4-5, L90-91: I think there should be more explanation of how the beam is split between the two instruments.

We would like to thank the reviewer for this valuable suggestion. The following has been added to the manuscript, section 2.2: “This is possible since the beam from the solar tracker has a diameter approximately twice as large as the aperture of the FTS. The DOAS itself needs less than 0.15% of the total beam diameter. The aperture of both instruments is sufficiently far away from the fringes of the beam of the tracker mirror. The part of the solar beam needed for the DOAS is coupled into an optical fiber[...].”

P 8, L161: Could the vehicles being loaded and unloaded impact your observations when the ship is docked?

Loading and unloading of the vehicles may lead to vibrations. The whole setup is placed on damping mats to compensate for these vibrations, but also the much stronger vibrations caused by the engine of the ship. We have added the following clarification to line 84f.: “Mats have been placed underneath the instrument to damp it from vibrations caused by the engine of the ship.”

Loading and unloading of the ship may also impact the height at which the instrument is located through changes in the draft of the ship. Since the altitude of the instrument is, however, passed into the retrieval algorithm by using the GPS coordinates of each spectrum, changes in draft are considered in the retrieval.

Thus, we do not expect an impact of the unloading and loading operations. But, as noted in the manuscript, we are cautious about using measurements while docked since there are many sources in the immediate vicinity, and the ship plume itself might contaminate the measurements.

P 11, L217: Table 6 is referenced before Tables 3, 4, and 5. This can make the paper a little confusing to navigate.

We have removed the reference to Table 6 in line 217, since the sentence already refers to Section 4.3, and there is no need to additionally reference the Table here.

P11, L222-224: If the X_{gas} values are calculated from the VCD, as stated in Eq. 2, why would you need to have an additional correction applied to X_{gas} ?

This has pure technical reasons. Inside the post-processing code, X_{gas} is calculated prior to performing the bias correction. This is done to allow for visual inspection of the uncorrected data. The bias is therefore not corrected twice, but rather separately for VCDs and mixing ratios.

P11, L232-235: It would be helpful to state the typical time period for the 100-spectra co-addition.

Since the exposure time changes according to illumination conditions, there is no general typical time period, but rather a range. We added the typical time range for 100 spectra to lines 232ff.:
“Afterwards, typically 100 spectra (corresponding to between 30 and 90 seconds of measurements depending on the exposure time) are co-added.”

P12, L242-243: Would the prevalence of high-altitude clouds in this area impact the usefulness of similar campaigns in the future, since this would likely limit the amount of time you could actively observe? These clouds may also affect the accuracy and precision of the EM27/SUN measurements.

High-altitude clouds are indeed a limiting factor for our setup. As outlined in Section 4.1, brightness fluctuations – a typical signature of high-altitude clouds – are already filtered out in the EM27/SUN dataset during the pre-processing. The filter thresholds were set as in previous publications (Klappenbach et al., 2015; Knapp et al., 2021), where it was demonstrated that these thresholds were reasonable and led to a good data quality. We therefore do not expect issues caused by high-altitude clouds in our final EM27/SUN dataset.

For the DOAS, large saturation fluctuations are filtered out during the coadding process. Since this filter is, however, rather sensitive to large brightness fluctuations caused by tracking errors than smaller fluctuations caused by clouds, the DOAS is potentially more sensitive to these variations. Therefore, Ring effect and CLD need to be taken into consideration, as mentioned in lines 241-243. Measurements severely impacted by high-altitude clouds are mostly filtered out for the DOAS, too, when the combined EM27+DOAS dataset is generated, since only measurements not being flagged in the pre-processing of the EM27/SUN are included in the final dataset.

Overall, rainy conditions and low, thick clouds were a much more important reason for measurement days being flagged than high-altitude clouds. The number of days being flagged was, however, in general not larger than during other campaigns with similar instruments being conducted by our group in other regions of the world.

P14, L275: What time intervals are used for averaging?

We added in line 275: “This averaging results in one datapoint every 2.5 to 3 minutes.”

Section 4.3: It seems like somewhere in this section you should report on the variance in the co-added spectra from the EM27/SUN while the ship is in motion. All of your performance metrics are based on observations taken while the instrument is stationary, but the objective of this paper is to show that the spectrometer can perform on a moving ship.

The variance in the co-added spectra while the ship is in motion was already discussed in our previous publications (Knapp et al., 2021; Butz et al., 2022). It was demonstrated there that the EM27/SUN, together with our custom-built tracker assembly, delivers precise and repeatable measurements. In the present study, we aim to evaluate the performance metrics that are typically assessed for qualifying the dataset for the COCCON standards. Thus, we put emphasis on the stability

of the spectral response function (Fig. 6) and the calibration factors with respect to a COCCON reference instrument. The latter can only be deployed on land, and the spectral response function can only be measured in the laboratory (under the required quality-controlled conditions). Thus, we cannot evaluate these metrics while moving. But we do evaluate these metrics for several episodes (before/after ship deployments) and show that they are stable over time.

We, however, added an additional reference to our previous publications in section 4.3 to make clear that tests on the measurement (and especially tracking) precision while the ship was in motion were already described there: “For additional performance tests, including an evaluation of the precision of the tracking system while the ship was in motion, the reader is referred to Knapp et al. (2021) and Butz et al. (2022).”

P16, L328: How can you be sure that the larger ratios are caused by differences in the retrieval algorithms? Can you elaborate on which specific differences are most likely contributing?

The RemoTeC algorithm does not perform any scaling, while PROFFAST scales all results by a common calibration factor with respect to TCCON, which itself is scaled to WMO standards. As stated in the paper, PROFFAST includes a common scaling factor for ground-based FTS with respect to the standards of the World Meteorological Organization (WMO). This WMO-scaling factor of, e.g., 0.989 for XCO₂ accounts for the known overestimation of the O₂-column by ground-based FTS (Wunch et al., 2010). If we multiply our results in Table 6 by these general scaling factors (reported in Table 5 of Wunch et al. (2010)), which are not included in RemoTeC, the discrepancy to the COCCON instrument decreases to 1.016 for XCO₂, 1.003 for XCH₄, and 1.073 for XCO.

In addition, offsets caused by the usage of different spectroscopic parameters are well-known from the literature (e.g., Malina et al. (2022), Sha et al. (2020)). We used the HITRAN2016 linelist in our RemoTeC retrieval, while the PROFFAST retrieval uses the standard TCCON linelist. Deviations on the same order of magnitude as our deviations are expected, as is, e.g., shown in Malina et al. (2022).

The following has been modified in the paper to make this more clear: “Most importantly, PROFFAST corrects for a spectroscopic error in the O₂-column, which accounts for a bias of about 2% (Wunch et al., 2010), and already has an overall scaling factor with respect to the standards of the World Meteorological Organization (WMO) included, while RemoTeC leaves these corrections to the post-processing. In addition, our RemoTeC retrieval uses a different spectroscopic database (see sect. 4.1) than PROFFAST, which leads to well-known offsets (Malina et al., 2022; Sha et al., 2020).”

Figure 9: The peak enhancement ratio in panel e appears to occur between the two enhancement events. This would cast some doubt on the validity of this ratio. If this point were discounted the overall range of the CO/NO_x ratios would be closer to 25-40, and this might change your assessment of the likely emission sources. In addition, the ratios after the enhancement are not very different from those during the event.

We agree that attributing the datapoint at 13:48 JST in panel e as part of episode 1 is debatable. However, even if the cut is made at 13:45 JST for episode 1, an overall range of 25-40 would not change the interpretation. As explained in the main text, the plume is attributed to a mixing of the steel factory plume with traffic contributions and contributions from gas-fired power plants. Since no large motorways or cement factories are upwind of our measurement, these two source types can be ruled out. In addition, urban traffic with gasoline cars only leads to a CO/NO_x ratio of 20 (Fontaras et

al., 2014). Therefore, even if only the range of 25-40 is considered, the interpretation of the measurement does not change. (Please also note the response to your comment on the text corresponding to the figure further below.)

P21, L404-406: The uncertainty in the retrievals should also be considered here.

The propagated retrieval uncertainty is shown in panels d-f of Figure 9 as error bars. As can be seen, the retrieval uncertainty is much lower than the variability of the enhancement ratios caused by atmospheric effects. We therefore neglect it in the discussion. We added to the caption of Figs. 8 and 9: "Error bars in all panels show the respective fit uncertainty from the retrieval and reflect precision only."

Table 7: I suggest ordering the columns of the ratios to match the order of plots in Figure 9.

The authors would like to thank the reviewer for this very helpful suggestion.

P24, L419: Your conclusion that this is an enhancement from the steel factory is based on the designation of the high CO grid cell being labeled as steel manufacturing in the inventories and wind direction; however, your enhancement ratios are very different from those reported by the inventories for this category. You also pointed out that you confirmed with the inventory that they had misplaced the steel factory. All of this disagreement suggests that there may be another explanation for your results. From the ratios you cite in Table 7, the sources that match the best are cement factories (with CO/CO₂) and urban gasoline or motorway gasoline emissions (with NO_x/CO₂ and CO/NO_x). You do mention later that there is likely a contribution from transportation emissions, but you do not discuss a possible contribution from cement manufacturing. While you do not show any cement industry emissions in Figure 7, have you investigated whether there is any cement manufacturing not reported by the Climate TRACE coalition or the inventories. As this is an industrial area, I would be surprised if there were no cement factories in the region.

As shown in Figure 7, Climate TRACE lists a cement factory in Inabe. This factory is, however, not upwind of the measured plume. In addition, its emissions are much smaller than those of, e.g., the steel factory. No other large cement factories are listed in the region. While we can certainly not fully rule out that Climate TRACE (or also EDGAR) misses out on a small cement plant in the Nagoya harbor area, a plant large enough to cause such a significant plume would hardly be missed by two inventories, especially in a country like Japan with transparent emission reporting. In addition, we have checked the Nagoya harbor area for sources linked to cement industry using Google Maps. Two small complexes in the vicinity of the steel factory are listed as "cement service station". Since we do not recognize the characteristic chimneys of shaft kilns on the satellite images of Google Maps, we concluded that these facilities are only for storage or secondary processing of cement. Because of this, we can rule out the contribution of emissions from the cement industry with a very high degree of certainty.

We have added the following in lines 420ff.: "Other source types with large CO emissions (see Table 7), such as shaft kilns from the cement industry or motorways, can be ruled out based on Climate TRACE coalition (2022)."

P24, L420-421: The CO/CO₂ emission ratios reported by Schneising et al. are an order of magnitude larger than what you observed. How does this support your argument?

As can be seen in Fig. 9, most of our observed CO/CO₂ enhancement ratios are above 1×10^{-2} , which means that the ratios observed by Schneising et al. (2024) are not an order of magnitude larger, as stated by the reviewer, but only a factor of 2-3. We added this factor to the manuscript in line 420 to make it clearer how different our ratios are from the ones in the Schneising et al. (2024) publication. We also softened the statement in line 419, stating that the steel factory is an “important contributor” instead of the “most likely source”. We clarified further that we expect mixing with “other sources with lower CO content” (line 422), which leads to a measured enhancement ratio lower than the one from the literature. However, we would like to point out that such a high CO/CO₂ enhancement ratio as in our observation can only be reached if the steel factory contributes to the plume, since “other source types with large CO emissions [...] can be ruled out” (added to lines 420f.). Nevertheless, we removed phrases overstating the attribution to the steel factory, as shown in the revised paragraph shown further below (answer to P24, L430-433).

P24, L423-424: The Van der Maas ratios for CO/NO_x are also quite a bit larger than what you observed and this is especially true if you exclude the dubious peak in CO/NO_x ratios that occurs at the end of the first enhancement event.

We added the actual factor between the literature and our observations, stating that the observed ratios are “a factor of 2 to 4 smaller” (line 423) than the ratios from van der Maas (2019). As already mentioned above, we also added that we expect mixing with “other sources with lower CO content” (line 422), leading to an enhancement ratio deviating from the literature. In general, the paragraph about the steel factory source attribution was revised, as outlined in the answer to P24, L430-433.

P24, L428-429: While I think all the information provided is interesting and should be considered, it seems like you are overstating the certainty that one should have about the steel factory source attribution.

This is answered in combination with the comment on P24, L430-433, below.

P24, L430-433: Glad to see this finally being discussed.

We have revised the paragraph about the steel factory attribution in order to have the discussion in lines 420-429 better aligned with lines 430ff. to make sure that we clearly point out the steel factory as certainly one, but not the only contributor to plume episode A. The revised paragraph is as follows:

“As pointed out previously, the large CO content of the plume in episode A suggests a steel factory as an important contributor. Other source types with large CO emissions (see Table 7), such as shaft kilns from the cement industry or motorways, can be ruled out based on Climate TRACE coalition (2022). According to a new study based on satellite data (Schneising et al., 2024), blast furnace steel production can lead to CO/CO₂ emission ratios of 3.24×10^{-2} on average, being larger than our measured enhancements by a factor of 2 to 3. While our measured ratios are lower than typical blast furnace CO/CO₂ ratios, the presence of the Nippon Steel Nagoya factory, operating two large blast furnaces and two basic-oxygen furnaces (Nippon Steel Corporation, 2024), indicates that steel production is a plausible, important contributor. Also, the large $\Delta\text{CO}/\Delta\text{NO}_x$ enhancement ratios are, by a factor of 2 to 4, smaller than ratios expected for steel factories (Van der Maas, 2019) (see also Table 7). For steel factories, Van der Maas (2019) found a CO/NO_x emission ratio between 60 and 122 based on data from the TROPOMI satellite for five blast furnaces. They point out that these findings

disagree by up to one order of magnitude with the EDGAR v6.1 inventory (Crippa et al., 2018; Janssens-Maenhout et al., 2019), which reports emission ratios between 2.8 and 16.8 for the same blast furnaces. This discrepancy is partly attributed to EDGAR relying on end-of-pipe measurements at the chimney and not including fugitive emissions within steel factories. Based on the emission ratios found by Van der Maas (2019) and Schneising et al. (2024), a blast furnace as the major source, with mixing from other sources with lower CO content, would fit our measured enhancement ratio quite well. The relatively low measured NO_x/CO₂ ratio might point at mixing of the steel factory plume with air masses carrying low NO_x/CO₂ ratios, such as caused by traffic (2.5×10^{-4} (Fontaras et al., 2014), assuming EURO 5 engines, urban driving conditions, and a vehicle fleet dominated by gasoline engines, which is typical for Japan (IEA-AMF, 2023)) or gas-fired power plants. Indeed, the city center of Nagoya, with a high traffic density, and several gas-fired power plants are located in the vicinity of the steel factory (Climate TRACE coalition, 2022). The latter are equipped with low-NO_x burners or denitration equipment (Jera Corporation, 2025), which reduces the NO_x/CO₂ ratio of a gas-fired power plant listed in Table 7 by up to 90% (Crippa et al., 2018). Although the EDGAR inventory at the grid cell level cannot fully explain the enhancement signature of episode A, a fitting source composition dominated by steel factory emissions and contributions from gas-fired power plants with low-NO_x burners and traffic can be found based on the measured enhancement ratios.”

P25, L461: Should this say “bottom-up” instead of “top-down”?

The authors would like to thank the reviewer for spotting this error.

P25, L468: I would suggest replacing “without permanent human attendance” with “with only part-time remote oversight” or something similar. It was my understanding that the instrument operations were supervised remotely. Did you encounter any problems that required intervention throughout the course of the campaign or did the instruments run automatically for the whole time without incident?

The instrument operated fully automatically and, in principle, would not have needed remote oversight. (Only the on-site maintenance every six weeks was needed to change the hard drives, and the exposure time settings needed to be changed twice per year manually to change between winter and summer.) The remote access was just installed to allow remote interventions in case of problems. There were indeed a few incidents that required human intervention, especially during the first month of the deployment. After that, the instrument operated mostly stably for the remainder of the deployment. On average, one incident requiring human intervention (usually a reboot) occurred per month. At the beginning, we logged in usually once per day to check whether everything was working correctly. We subsequently decreased the login frequency until the end of the deployment. It should be pointed out that these logins were just sanity checks to make sure we are not missing out on an incident and were not needed for the operation of the instrument. Since the instrument would also have fully operated without these remote checks during most of the deployment, we would like to keep the wording “without permanent human attendance”.

P25, L472-473: I don’t believe you proved that your setup is suitable for satellite validation because you did not actually compare to any satellite measurements.

Satellite validation was not part of this study. To make this clearer, we have revised our “conclusion and outlook” section as outlined in the answer to the question on P26, L481, below.

P26, L481: You did not operate your instrument over the open ocean.

We indeed did not perform satellite validation or measurements over the open ocean in the current study. We have therefore restructured our “Conclusion and outlook” section, combining the suggestions of reviewers 1 and 2. Through this restructuring, we distinguish more between what was actually done in this study and what is only part of the outlook. Please find the modifications below:

“[...] This shows that our shipborne measurements of CO₂, CH₄, and CO are compatible with the standards of the COCCON and, in consequence, they can be adjusted to the TCCON scale through a multiplicative adjustment. Thus, our setup is suitable for several potential use cases.

Here, we have demonstrated, in a case study, the detection of plume enhancements of ΔCO₂, ΔCO, and ΔNO₂, carrying the outflow from the heavily populated Nagoya region. [...] Building on the demonstrated potential of the instrument, long-term shipborne installations will enable systematic emission monitoring along coastal hotspots.

A second use case is satellite validation over open oceans, where almost no data currently exists (Müller et al., 2021). This use case has already been demonstrated for a shipborne FTS in our previous publications (Klappenbach et al., 2015; Knapp et al., 2021), where the system still required on-board personnel. Having now an upgraded instrument that can operate fully remotely for several weeks and that is compatible with COCCON standards, as demonstrated within this study, makes future routine deployments for satellite validation over the open ocean possible. In addition, the added capability to simultaneously measure NO₂ makes our instrument an ideal platform for the validation of the newest generation of satellites measuring this air pollutant along with the greenhouse gases, if an appropriate route for the ship is chosen.”