Response to reviewer comments on Enders et al., 2025: „Towards routine shipborne measurements of columnar CO_2, CH_4, CO, and NO_2: a case study for tracking regional-scale emission patterns“

Questions of the reviewers are in italics. Full references can be found in the main manuscript.

Reviewer 1:

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 shipbourne 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 CO_2, CH_4 and CO (XCO_2, XCH_4, and XCO) with the EM27, as well as partial columns of NO_2 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 shipbourne 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 shipbourne 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 NO_2) 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 XCO_2 accounts for the known overestimation of the O_2-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 XCO_2, 1.003 for XCH_4, 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 O_2-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/CO_2) and urban gasoline or motorway gasoline emissions (with NO_x/CO_2 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/CO_2 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/CO_2 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/CO_2 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/CO_2 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/CO_2 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 $\Delta CO/\Delta NO_x$ 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/CO_2 ratio might point at mixing of the steel factory plume with air masses carrying low NO_x/CO_2 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/CO_2 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 CO_2, CH_4, 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 ΔCO_2, ΔCO, and ΔNO_2, 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 NO_2 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.”

Reviewer 2:

Enders et al. present an approach for distinguishing emission sources using the ratios of CO_2, CH_4 and CO dry-air mole fractions measured by a shipborne FTIR (EM27/Sun) instrument together with NO_x derived from NO_2 vertical column densities measured by a direct-sun DOAS instrument.

The instrumentation and methods are briefly explained in this manuscript and have been discussed in previous studies.

The general feasibility of the approach is demonstrated through a case study based on measurements conducted in Mikawa Bay, Japan, in 2023. The derived emissions from this case study are compared with the EDGAR and REAS inventory data sets and show good spatial agreement with the location of the emitters.

The paper is well written, the analyses are thorough, and the approach shows potential for validating satellite data sets in coastal regions.

We thank the reviewer for the appreciation of our work and the helpful comments. Please find our point-by-point reply below.

My two main concerns relate to the case study itself and its application to satellite validation:

1. The presented case study does not conveincingly demonstrate that the indentified emitters are responsible for the observed measurements. Although you mention that self-emissions are negligible, some of the recorded enhancements (particularly the strong signals after leaving Mikawa Bay) could be influenced by the vessel's own emissions. In addition, only wind direction is shown, while wind speed is missing. What was the vessel's cruising speed? Furthermore, why is only a single time series example presented, despite several months of available data?

To strengthen this analysis, I recommend including a map of wind trajectories for the selected case study, for instance using HYSPLIT or FLEXPART simulations.

The deployment on the Nichiyu Maru was a short-term opportunity within the SOOP program (section 3). The ship, however, is a suboptimal carrier for our purposes since it mostly travels during the night, while during the day, it is moored in the harbor for loading operations. This, together with other requirements such as good weather conditions, prevented us from collecting a more comprehensive dataset on emission episodes. Nonetheless, a more in-depth analysis of data from the whole campaign (including another deployment on the same ship in 2025) is currently ongoing. The goal of the present paper is to document the technical setup and the COCCON-compatible performance and to showcase the usefulness for emission studies. Follow-up studies will refer to the present study for the technical details and put more focus on the use cases.

Fig. A1 shows the wind field at a pressure level of 950 hPa for our case study at Nagoya Bay. The wind field (taken from ERA5) was rather constant in direction and speed throughout the region. In addition, the wind field was constant with altitude, as confirmed by analyzing the wind fields for the surface wind and the wind at 800 hPa. While we agree that the quantitative usage of a Lagrangian trajectory model (such as FLEXPART, STILT, HySPLIT) would be beneficial, we refrained from setting up such a system due to the technical complexity and the limited expected gain, as the wind fields were simple during the analyzed period, allowing for qualitative, straightforward interpretation

anyway. In particular, we expect that for quantitative interpretation, the resolution of publicly available wind fields (ERA5 ca. 0.25°) would not be sufficient for the local scales of our study, which would imply running a regional weather model first to generate the input for the trajectory model.

Regarding the self-emissions, we can point towards the ship's cruising speed, which was around 14 knots (approx. 7.2 m/s) after the ship left the port (see Fig. A2). Having had such a high cruising speed makes self-contamination virtually impossible if a wind with a constant speed of around 11 m/s blows from a perpendicular direction. This is the case during the time when the two main plumes were detected. We agree with the reviewer that self-contamination cannot be fully ruled out after the ship turned at the tip of Mikawa Bay and traveled in a direction parallel to the main wind direction. In that case, however, clear plume signatures should be visible, especially for NO_x. We find some larger variation in the NO_x data after 14:30 JST (see Fig. 8). Since, however, clouds started to become present (being the reason for the datapoints getting more scarce than before 14:30), they cannot clearly be interpreted as plume signatures, but might also be caused by variability due to scattering. In addition, for such a nearby source, a much stronger enhancement would be expected.

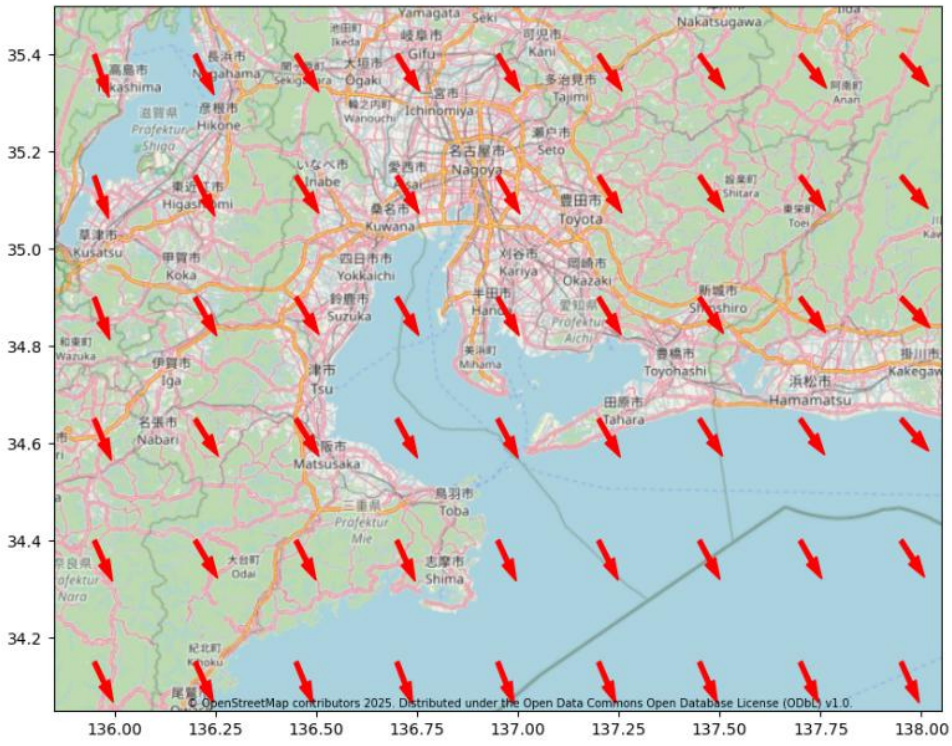


Figure A1: Wind vectors taken from the ERA5 reanalysis (Hersbach et al., 2020) for 14:00 JST on 13 November 2023. The length of the vectors is proportional to the wind speed. The wind field is for a pressure level of 950 hPa.

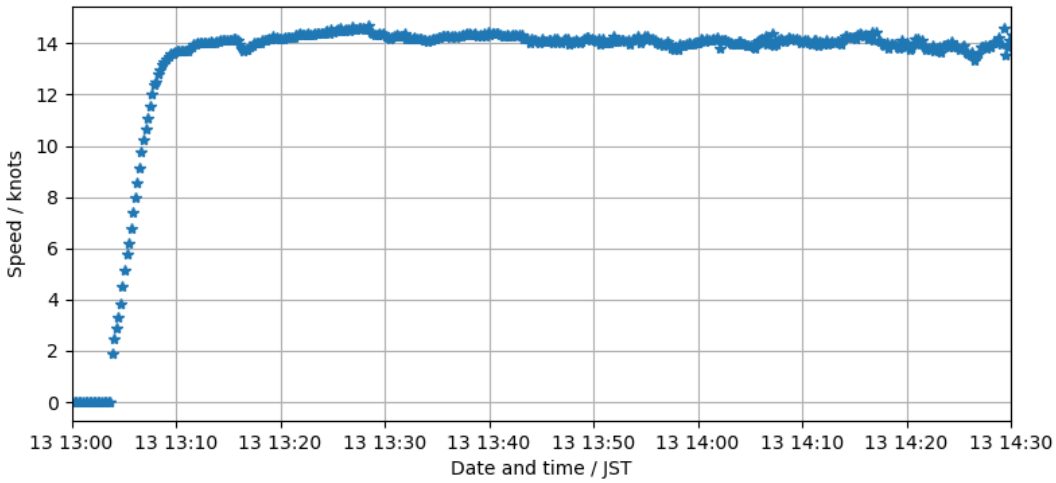


Figure A2: Speed of the vessel during the period depicted in the case study. Speed was measured with the GPS sensor onboard our instrument.

To make these points clearer, we have added the following to the manuscript:

Line 347ff.: “In addition, the chosen period is ideal for this task, since an almost constant, unidirectional wind field with a wind speed around 11 m/s (taken from the ERA5 reanalysis (Hersbach et al., 2020)) makes a simple interpretation of the dataset possible.”

Line 353ff.: “The ship Nichiyu Maru left Toyohashi port (Aichi prefecture) at 13:05 Japan Standard Time (JST), traveled with a speed of around 14 knots (7.2 m/s) westward along the coast of the Atsumi peninsula through Mikawa Bay, […].”

Line 357ff.: “Since the ship traveled in a direction perpendicular to the direction of the wind at a large speed, self-contamination by the plume of the Nichiyu Maru can be ruled out at least until 14:30 JST. Even afterward, no clear signatures of self-contamination can be found in the data (see Fig. 8).”

Line 464ff.: “The current study is limited by the approximate use of wind directions from the ERA-5 reanalysis (Hersbach et al., 2020), which is possible here because of the closely located sources and a very uniform wind field. These conditions are, of course, not warranted during all days of the deployment, and especially scenarios where the ship is further away from emission sources are of large scientific interest. To study such sources, the usage of wind trajectory models such as FLEXPART (Pisso et al., 2019) is needed. An optimized analysis framework including such trajectories is currently under development.”

2. While the main motivation for this work is to support the validation of satellite measurements, the manuscript does not explain how the proposed emitter estimation approach could be applied in that context. Please include a short discussion clarifying how this method could contribute to satellite data validation.

The authors would like to thank the reviewer for pointing out that this section was not clear. Following also suggestions from reviewer 1, we have restructured our “Conclusion and outlook” section. The main point is that the proposed emitter estimation approach in this paper is for studying

local emitters and not for satellite validation. Satellite validation is just another use case for the instrument, which was already demonstrated in our previous studies (Klappenbach et al., 2025; Knapp et al., 2021; Butz et al., 2022). The reason why satellite validation is mentioned in the outlook is that the new functions of the instrument (remote operation, added VIS-spectrometer) make it much easier to use the instrument for this use case, or make a validation of a broader selection of satellites possible. We hope that the modified version of the final section makes this clearer:

"[…] This shows that our shipborne measurements of CO_2, CH_4, 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 ΔCO_2, ΔCO, and ΔNO_2, 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, satellite validation over the open ocean, where currently almost no data exists (Müller et al., 2021), was already demonstrated for the ship-borne FTS, still requiring on-board personnel, in our previous publications (Klappenbach et al., 2015; Knapp et al., 2021). 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 NO_2 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."

L4: Please change to: "direct-sun DOAS"

Changed.

L12: I recommend replacing "concentration" with "volume mixing ratio" throughout the manuscript, as only VMR (and VCD) values are presented in the figures.

We thank the reviewer for pointing out the ambiguity in the use of the term "concentration." Since "concentration" was used in several sentences of the manuscript to refer to both VCD and VMR, we decided to replace it with the general term "trace gas abundance" or "abundance" to avoid any ambiguity. The following changes have been made:

Line 12: "concentration" -> "trace gas abundance"

Line 19: "concentrations" -> "abundances"

Line 20: "concentration" not replaced, since it is used in terms of "concentration gradients" needed for top-down estimates.

Line 21: "concentration measurements" -> "column measurements"

Line 26: "concentrations" -> "abundances"

Line 43ff: Sentence removed.

Line 45: “concentration” -> “vertical column densities (VCD)”

Line 56: “concentration” -> “trace gas abundance”

Line 255: “Concentrations” was not exchanged, since it is the correct wording here for this general discussion of Langley’s method.

Line 365: “background concentrations” -> “background column”

Line 389: “concentrations” -> “abundances”

Line 407: “concentration enhancements" -> “column enhancements"

L14: You mention emission monitoring but do not provide emission estimated for the case study. It should be possible to calculate emissions using the VCD data together with wind/vessel speed vectors.

The authors thank the reviewer for this suggestion. However, we would like to point out that this is not as easy as it seems at first glance. The authors think that the reviewer here points at the mass balance method, which was, e.g., also used with a similar setup by our group in Luther et al. (2019). This method is, however, only applicable for isolated sources. The latter is not true for the two plumes in the case study, which we have identified as a superposition of two main and several minor sources. It is therefore not possible to give a simple emission estimate within the scope of this study. Modelling of several plumes with an inversion framework (e.g., based on Lagrangian footprint calculations) would be needed for that, but we deem it outside the scope of the present study.

We have nevertheless tried the mass balance approach (as outlined in Luther et al. (2019)) and calculated emissions of around 6 Mt/yr of CO_2, 6.7 kt/yr of NO_2, and 370 kt/yr of CO for the whole double plume structure. If the emissions of all cells in the Nagoya harbor area are summed up, EDGAR v6.1 gives us emissions of approximately 3.5 Mt/yr of CO_2, 4.8 kt/yr of NO_2, and 200 kt/yr of CO. While it is a good indicator for the overall performance of our instrument that the emissions are in the same order of magnitude, a direct comparison is not possible, since it is not clear which grid cells of EDGAR contributed to the actual plume and which sources are only included in the background which was subtracted. In addition, as discussed above, these calculations do not help us with respect to source attributions, which is the main goal of the case study. We therefore decided not to include them in the paper.

L90: Which part of the light beam is used for the DOAS measurements? Please elaborate on this point.

The following has been added to the manuscript, also following suggestions by reviewer 1: “The DOAS instrument uses part of the light beam from the solar tracker, which is not needed for the FTS. 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[…].”

Section 2.2: Since stable temperature is crucial for DOAS performance, what was the average detector temperature variation during the case study? How might any temperature variations have influenced your results?

During the case study, the detector temperature of the DOAS varied by less than 0.0015 K (see Fig. A3). The temperature inside the DOAS box (i.e., between the first and the second cooling stage) varied by less than 0.2 K (see Fig. A3). Similar variations were found during most other days. All measurements where temperature variations exceeded 0.5 K were manually flagged and excluded from further analysis. Due to the low detector temperature variations, we do not expect any influence on our results.

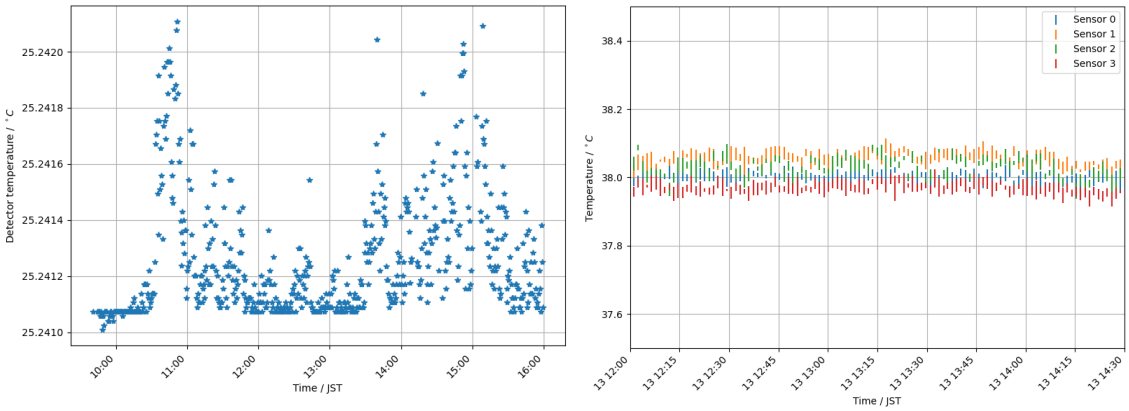


Figure A3: Temperature of the DOAS detector (left) and temperature of the DOAS box measured with four different sensors (right) on November 13 during the period of the case study.

L146: If ancillary in situ data are available, please specify what kind of data these are and explain why they were not included in the manuscript.

The in situ data collected onboard the Nichiyu Maru within the framework of SOOP consists of meteorological data and trace gas data. The meteorological data were partly used within this manuscript, since one of the in situ pressure and temperature sensors of our instrument failed during the deployment. Therefore, the “in situ ground-based pressure and temperature measurements” mentioned in lines 190f. used for the EM27/SUN retrieval are taken from the SOOP dataset. In principle, the in situ sensors onboard the Nichiyu Maru also collect wind data, which is a capability our instrument does not have. Unfortunately, due to some issues with sensor location, the in situ sensors did not correctly compensate for the motion of the ship when measuring the wind direction. We therefore refrained from using the in situ wind data and used the ERA5 dataset instead.

The sensors onboard the Nichiyu Maru also measure in situ trace gas abundances of, e.g., CO_2, CH_4, CO, and NO_x using standard gas analyzers. Since the in situ instruments have, in general, a different sampling footprint than the remote sensing spectrometers, a simple comparison is, however, not possible within the scope of this publication. A second manuscript focusing on a comparison between the in situ and remote sensing data, including a discussion of caveats and reasons for differences between the instruments, is currently under preparation.

We have added the words “meteorological and chemical in situ data” in line 146 of the manuscript to make clear what type of data is collected.

L153: It seems unlikely that no ship plumes were detected during several months of measurements. The wind and vessel velocity vectors in the first figure even suggest possible self-emissions. Please elaborate on how you ensured that signals from the vessel’s own emissions were excluded.

The Nichiyu Maru usually travels at a high speed compared to other ships of around 14 knots (approx. 7.2 m/s). Unless a strong wind comes from the opposite direction of the ship's travel direction, self-contamination is not possible since the chimneys are around 100 m behind our instrument. In order to also rule out self-contamination during unfavorable wind conditions, all recorded days were manually checked for suspicious plume signatures, such as plumes with large amplitude, but small width, or repeating patterns of such plumes. No such suspicious signatures were found while the ship was moving. However, as also mentioned in the manuscript, suspicious signatures were found while the ship was docked in harbors. An example is shown in Fig. A4.

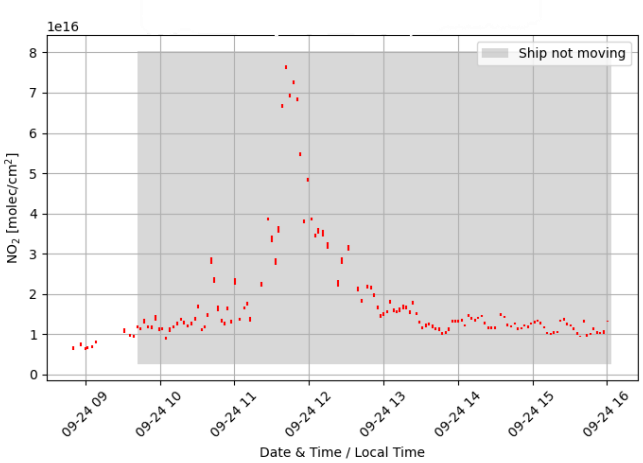


Fig. A4: NO_2 VCD recorded on 24 September 2023, where the ship was in the harbor most of the day. Since the wind direction was changing during this day, several plumes swept through the field of view of the spectrometer. These plumes can either be interpreted as self-contamination from the ship or as the signature of nearby sources.

It should be noted that contamination by other ships, e.g., travelling in the opposite direction, cannot be ruled out. However, such scenarios would lead to plume signatures with very small widths. Such signatures were found in the data by manual analysis and excluded from further interpretation.

L162: How many days of daytime measurements are available, and how many of these contained useful data?

In total, 112 days of daytime measurements are available. On 57 days, the weather conditions were favorable enough to have useful data produced at least during part of the day.

Following also suggestions from reviewer 1, we added after line 163: “In total, daytime measurements are available for 112 days of the deployment. On 57 of these days, the weather conditions were favorable enough to have useful data produced at least during part of the day (see Sect. 4 for information on quality filters included in the retrieval).”

Table 2: Which windows were used? Why do you show all windows here?

The RemoTeC algorithm uses all the retrieval windows listed in Table 2 simultaneously. If multiple windows contain a certain molecule as a target absorber, its abundance is fitted simultaneously across all the windows. For example, CO_2 is retrieved from two windows simultaneously. Splitting the

spectrum into several windows is useful to exclude uninformative spectral regions in between and to avoid overly complex parameterizations of the background radiance polynomials.

L204: What percentage of your measurements was filtered based on the pressure comparison? What is the maximum deviation between surface and measured pressure values?

Since the percentage depends heavily on the weather conditions, we did not calculate a number for the whole campaign. On the day of our case study (a day with good weather conditions), for example, 1.6% of our measurements got flagged by the pressure filter, and 2.7% got flagged by the DC filter. The mean deviation between the surface and spectroscopic pressure is 29.3 mbar on the day of the case study. The maximum deviation is 34.1 mbar, leading to a mean offset-subtracted maximum deviation of 4.8 mbar.

L218: Is this correction also applied to the NO_2 VCDs? Typically, such effects are accounted for using air mass factors derived from radiative transfer models. Have you considered using model-based AMFs, and how would you expect the results to differ from those obtained with the geometric AMF, including the applied correction? The geometric AMF does not account for aerosols, which were likely present. Please discuss how aerosol effects may influence your results and the implications for the AMF approach.

The airmass-dependent bias in EM27/SUN measurements is not the same as the bias that is corrected for using model-based AMFs in a DOAS retrieval. The airmass-dependent bias is most likely related to instrumental artifacts, from which all direct-sun viewing FTS suffer (see Wunch et al., 2011).

Our DOAS spectrometer collects direct sunlight with a narrow field-of-view (1.15°). Direct sunlight is very bright, and single-scattered light that enters the telescope has almost the same lightpath as direct sunlight. Thus, it requires light that has at least been scattered twice to induce significant lightpath changes. Given that all our data have reasonably large sun elevations (SZA<70°) and that the sun tracking requires clear-sky conditions anyway, we consider the contamination by second-order and higher-order scattering negligible.

In general, scattering by aerosols leads to a higher DC-variability in the EM27/SUN data. Since we perform a filtering for this as detailed in Sect. 4.1, it is not expected that our filtered EM27/SUN dataset still contains spectra with artifacts caused by scattering at aerosols. A similar filter is included in the processing of the DOAS data, which filters for large detector saturation variations (i.e., brightness fluctuations), as detailed in Sect. 4.2. We, therefore, also do not expect that spectra perturbed by aerosol scattering influence our final, filtered dataset.

Table 3: Why did you include IO in your fit? Did you detect any tropospheric IO above the instrument's detection limit?

Figure A5 shows the retrieved IO dSCDs for the day of the case study. It can be seen that IO dSCDs around the detection limit have been found. In order to confirm or exclude IO detection, we would need to set up a retrieval dedicated to IO (not NO_2), e.g., for the optimal spectral range. Running a retrieval specifically targeting IO is out of scope for this publication. We have run sensitivity studies with various DOAS configurations, including with and without IO as an interfering absorber. These sensitivity studies show that retrieved NO_2 dSCDs are robust against changes in the DOAS configuration, as can be seen in Fig. A6.

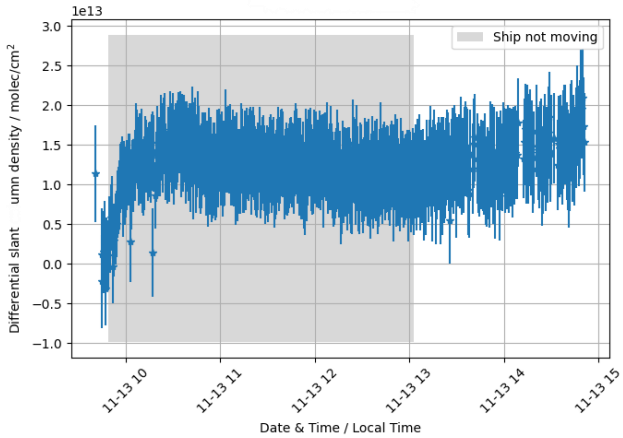
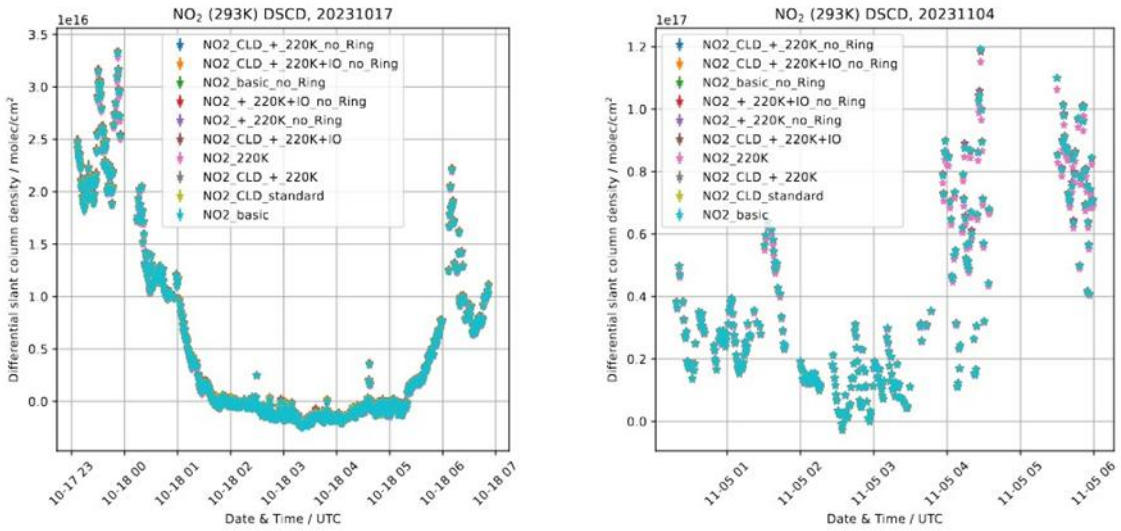


Figure A5: dSCD for IO measured on 13 November 2023. The error bars correspond to the fit error.



(a) October 17, 2023. (b) November 4, 2023.

Figure A6: Final NO_2 dSCDs retrieved with all retrieval scenarios included in the sensitivity study. While the left panel shows results for a day with good weather conditions, the right panel shows results for a day with large brightness fluctuations caused by high-altitude clouds. Absorber cross-sections were used as defined in Table 4. Within the case studies, the additional absorbers “Ring”, “CLD”, “IO”, and “NO_2 at 220K”, which are used in the retrieval scenario outlined in Table 4, were partly switched off as detailed in Table A1. Within the main manuscript, the scenario “NO2_CLD_+_220K+IO” is used.

Name	NO_2 at 220K	Ring effect	CLD	IO
NO2_CLD_standard	No	Yes	Yes	No
NO2_basic_no_Ring	No	No	No	No
NO2_CLD_+_220K+IO	Yes	Yes	Yes	Yes
NO2_CLD_+_220K_no_Ring	Yes	No	Yes	No
NO2_220K	Yes	Yes	No	No
NO2_basic	No	Yes	No	No
NO2_CLD_+_220K+IO_no_Ring	Yes	No	Yes	Yes
NO2_CLD_+_220K	Yes	Yes	Yes	No
NO2_+_220K_no_Ring	Yes	No	No	No
NO2_+_220K+IO_no_Ring	Yes	No	No	Yes

Table A1: Definition of the retrieval scenarios included in the sensitivity study shown in Fig. A6. The scenario used in the main publication is "NO2_CLD_+_220K+IO".

L246: Did you test alternative noon spectra, and if so, how did this affect your reference SCD?

The reference SCD for NO_2 is determined according to the Langley method described in section 4.2 and illustrated in Fig. 5. We chose our reference noon spectrum such that the reference SCD is small, i.e., such that the dataset (e.g., used in Fig. 5) has few negative NO_2 dSCDs. For determining the reference SCD, however, the fit uses all the 20% lowest dSCDs (orange in Fig. 5). Choosing a different reference spectrum would simply shift the zero crossing of the Langley fit and thus, yield the same total SCDs.

L314: Please provide an explanation for the higher mean error (ME).

Between the SRF measurement in February and May 2024, the instrument was moved with a crane twice to be set up and taken down from the ship. During one of these moves, the crane heavily bumped the instrument against parts of the ship. We assume that this led to a change in detector alignment, which is the reason for the change in ME. It should be noted that only the SRF from channel 1 was used in the final retrieval setup, since it is considered more reliable in general.

We have added the following sentence to the manuscript: "It is assumed that this change is caused by an incident where the crane bumped the spectrometer against parts of the ship when it was lifted to the upper deck in February 2024, leading to a slight misalignment of the second detector channel. Since we consider the SRF measurement based on channel 1 more reliable, we have used an SRF from that channel for the final retrieval described in Sect. 4.1."

L330: It is unclear to me why an additional scaling is necessary given that a correction was already applied earlier. Could you please include an example (perhaps in the appendix) illustrating how your corrections and scaling steps affect the final results?

It is not clear to the authors what the reviewer refers to when asking about additional scaling. As outlined in lines 216ff., all measurements are referenced to TCCON (and by that to WMO standards) by a single scaling factor. This procedure is standard, and, e.g., is also applied to all data of the COCCON network. In line 216, it is referenced to Sect. 4.3 regarding a detailed description of how these scaling factors are calculated for our instrument. The procedure is then discussed in lines 321ff. in detail. Lines 330ff. only state that the factors, whose calculation is outlined here, are used for the scaling as described in Sect. 4.1.

Figure 7: Please include wind speed information. If the wind speed remained relatively constant, state this in the figure caption. If not, update the figure to show the wind speed variations. Additionally, include the vessel's velocity in the caption and discuss the potential contamination from the vessel's own emissions. Could you estimate a characteristic gas signature of the vessel's plume?

Please refer to our reply to general comment 1.

L364: Were any CH_4 enhancements observed in the measurements? Please also explain the decreasing trend shown in Figure 8.

On some days of the deployment, enhancements of up to 20 ppb were observed. The case day presented in the manuscript is therefore not ideal to discuss methane enhancements. Because of that and because methane is not needed for the proposed analysis approach, we have omitted a detailed discussion in the publication.

It can also only be speculated about the cause of the downward trend of around 6 ppb shown in Fig. 8. There are several LNG terminals in the harbor of Nagoya and, in addition, Nagoya has a large city gas network (https://www.tohogas.co.jp/lang/en/corporate/document/annual/pdf/2023toho_tougou_en_print.pdf, page 71). Urban plumes caused by fugitive emissions from natural gas infrastructure are known from other studies (e.g., Jones et al., 2021). Such a scenario would also fit the observation that the methane column decreases while the ship moves away from Nagoya. In order to confirm this interpretation, measurements at a larger distance from Nagoya would, however, be needed to check whether the trend continues.

Figure 10: Adjust the color scale so that weaker emitters are visible. It would also be helpful to include a comparable figure illustrating CO_2 emissions.

The authors would like to thank the reviewer for this suggestion. We have changed the color scale and also set all grid cells with CO emissions below 1000 t/yr to grey in order to make weaker emitters more visible. The updated figure is shown below:

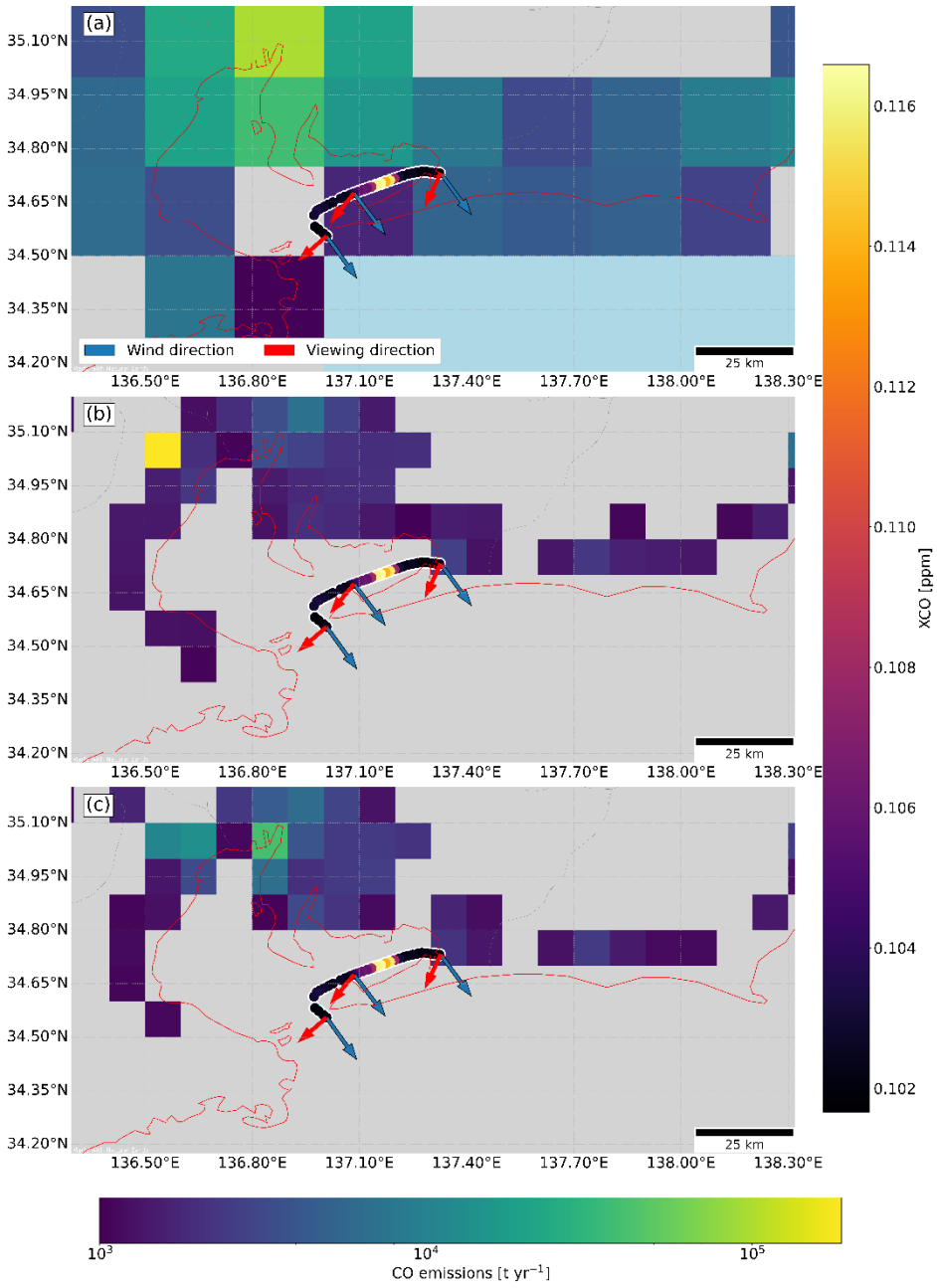


Figure A7: Updated version of Fig. 10 from the manuscript.

The following has been added to the caption: “Cells with emissions below 1000 t yr^{-1} according to the respective inventory are shown in grey.”

Figure A8 shows the same plot for CO_2 as suggested by the reviewer. However, as stated in the manuscript, only one source with strong CO but not that strong of CO_2 emissions has been misplaced. Therefore, there are no major differences between the two versions of EDGAR visible. Since all main sources in the Nagoya area are point sources, which are already visualized in Fig. 7, we do not see any added value in adding Fig. A8 to the manuscript.

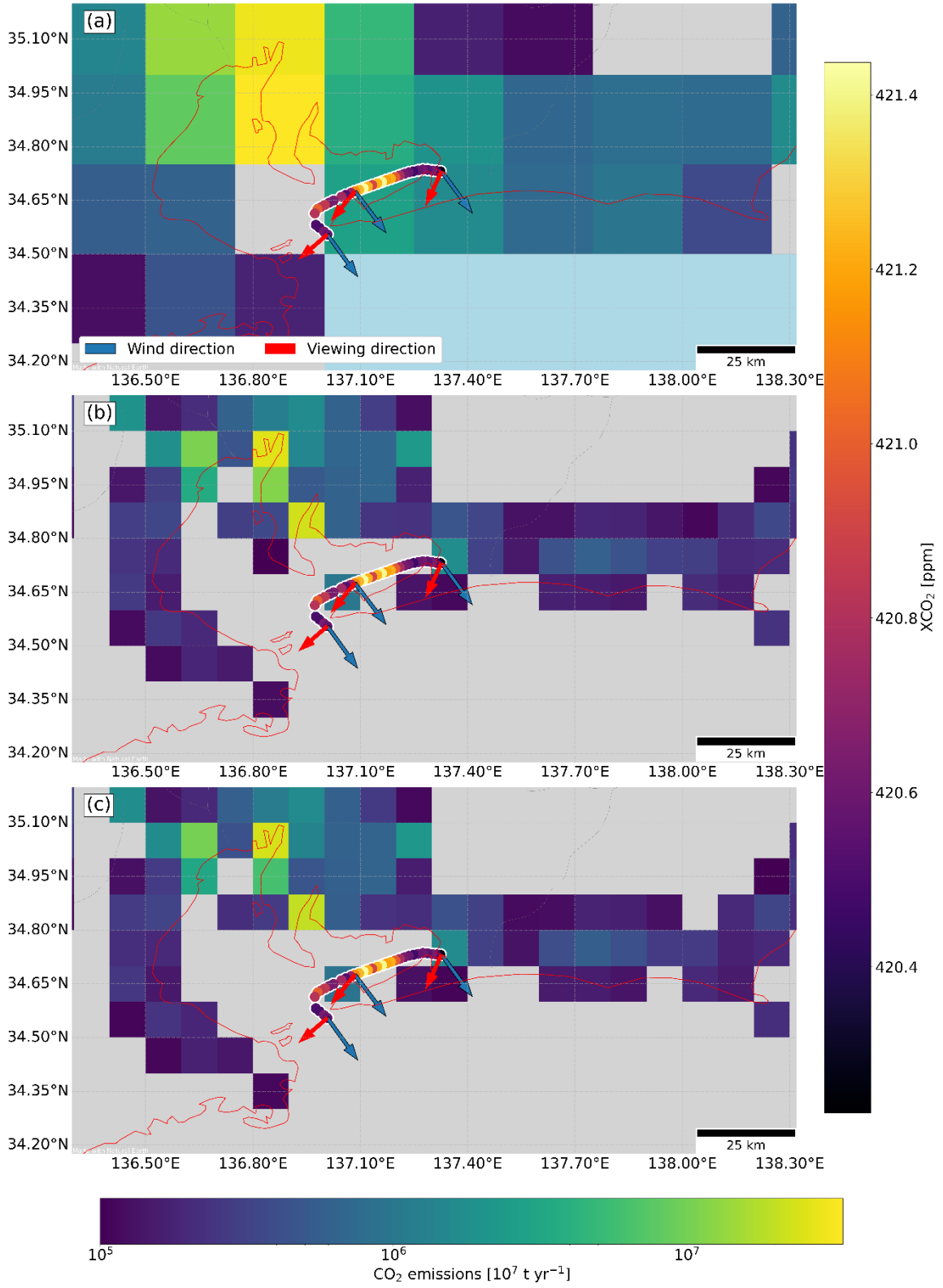


Figure A8: Comparison of measured XCO_2 enhancements to CO_2 emission patterns from different inventories: (a) REAS inventory for 2015 (Kurokawa et al., 2020); (b) EDGAR v6.1 for 2018 (Crippa et al., 2018); (c) EDGAR v8.1 for 2022 (Crippa et al., 2023). Cells with emissions below 10^5 t yr^{-1} according to the respective inventory are shown in grey. Wind directions (blue arrows), interpolated to the location and time of the measurement, are taken from ERA-5 (Hersbach et al., 2020). Viewing direction is shown by the red arrows. Note that the REAS inventory (a) only includes land grid cells.

L396: From Table 7 it seems that you refer to oil rather than lignite because the NO_x/CO_2 ratio is larger for oil.

We thank the reviewer for this comment. We have revised the wording to avoid implying a direct causal relationship between fuel composition and the observed NO_x/CO_2 ratios. The key point here is that lignite contains much more oxygen and much less carbon than crude oil. The excess oxygen (above the optimum ratio) decreases the combustion efficiency and, by that, reduces the combustion temperature. Because of the temperature-dependence of the Zeldovich reaction, this in the end reduces the NO_x production. This leads to a lower NO_x/CO_2-ratio for lignite. It should be noted that all these mechanisms are non-linear.

In order to avoid this ambiguity, we have rephrased the sentence in the following way: "Concerning fuel type, for example, lignite coal contains, on average, much less carbon, much more oxygen, and about the same amount of nitrogen as crude oil (Klemm and Hoppe, 1980), leading to a higher NO_x/CO_2-ratio for oil-fired compared to coal-fired power plants (Guevara et al., 2024) through non-linear reaction mechanisms." Note that we also changed the wording from "more nitrogen" to "about the same amount of nitrogen", since the difference in nitrogen content is almost negligible compared to the difference in oxygen and carbon content.

L414: meaured --> measured

Changed.

L463: Please add wind trajectories to the discussion.

Please refer to our reply to "general comment 1", where changes to the manuscript are also listed.

Erratum:

We were notified by an anonymous colleague via email about a mistake in Fig. 7 and Fig. 10, which we would like to correct in the final version of the manuscript. Unfortunately, there was an offset of around 30° in the plotted viewing directions due to a bug in the plotting code that went undetected. The corrected version of Fig. 10 is shown as Fig. A7. The corrected version of Fig. 7 is shown below as Fig. A8. Please note that this shift in viewing direction does not change any of the conclusions drawn in the manuscript.

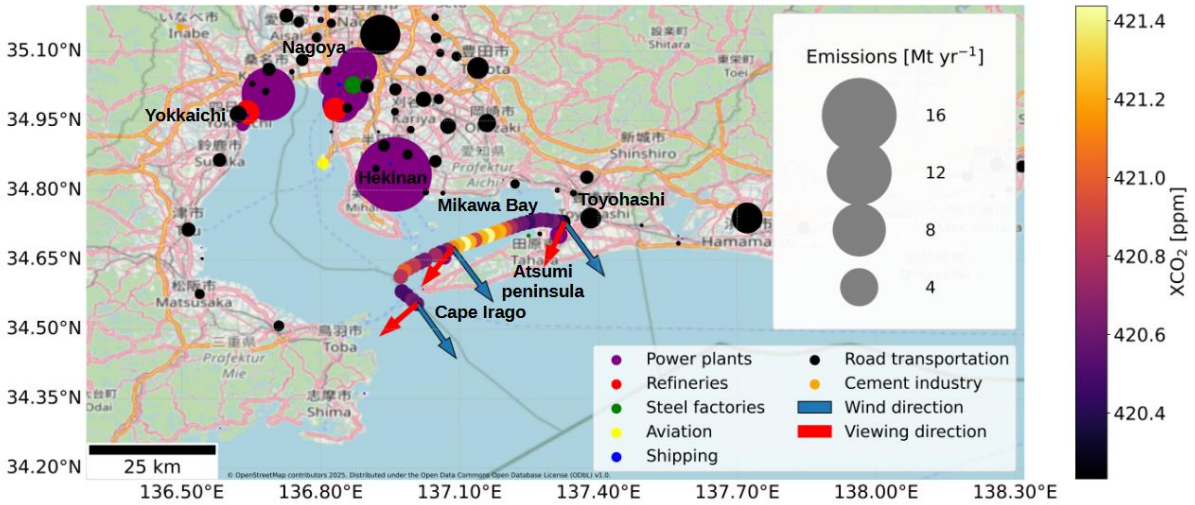


Figure A8: Corrected version of Fig. 7.