Answer to Referee #2

We would like to thank Referee #2 for his/her positive and constructive comments and suggestions. We have studied comments carefully and made corrections, which we hope meet with approval. Comments and responses are listed as follows. In order to facilitate the reference to the questions and proposed changes, we use the following color coding:

Color coding:

Referee comment

Our answer

Proposed change in manuscript

Title: Without trying to be negative I would suggest leaving out the words "High-precision".

Our initial use of the term was based on the fact that UAV measurements can be made more closely to the funnel of ship, to obtain high-precision results. But on reflection, we think the precision will improve further, with the progress of technology and method. "High-precision" is not appropriate; we changed the title as:

"Monitoring of compliance with fuel sulfur content regulations through UAV measurements of ship emissions"

Abstract: As in the title I would suggest leaving out " high precision" in the last sentence. I would also mention the range of sulfur contents that were encountered in the study i.e. how many non-conformities were encountered. And I would like to mention more explicitly that the deviation of the estimated value for +FSC is less than 0.03% (m/m) at a level of 0.04 % to 0.24 % FSC. Note that in ECA areas, with a limit of 0.1%, an uncertainty of 0.03% is not very good. I would also suggest mentioning that in all cases the estimated FSC was always lower than the actual FSC derived from samples taken on board. This is an important aspect with a strong impact on the usefulness of the method in SECA areas with a 0.1% limit value.

"High-precision" has been leaving in whole manuscript. The range of sulfur contents is very important for this research, which should be mentioned in the abstract, result and conclusion. These parts have been rewritten. In addition, the discussion about underestimate of FSC has also add in the conclusion.

In abstract:

After more than 20 comparative experiments, the results show that, in general, the deviation of the estimated value for FSC is less than 0.03% (m/m) at an FSC level ranging from 0.035% (m/m) to 0.24% (m/m). Hence, UAV measurements can be used for monitoring of ECAs for compliance with FSC regulations.

In result:

As shown in Fig 6, the FSC in our experiments was mainly at a level of 0.035% (m/m) to 0.24% (m/m) (only one measurement of 0.37% (m/m), not enough for reference). The deviation of the estimated FSC value calculated using the proposed method was within 300 ppm (0.03% (m/m)), although there was some uncertainty.

In conclusion:

In general, the deviation of the estimated FSC value was within 0.03% (m/m) at an FSC level of 0.035% (m/m) to 0.24% (m/m). Because not all the sulfur in the fuel is emitted as SO_2, the estimated FSC is smaller than true value in many cases. Therefore, if the maritime department wants to take the estimated value as the basis for the preliminary judgment regarding whether the ship exceeds the emission standard, it needs to set an appropriate threshold and a confidence interval.

How many non-conformities were encountered.

In the result, we discuss that:

In addition, when the FSC of the target ship is low, for example, when the fuel used is light diesel fuel, the SO_2 observation values were mostly 0. When this happened, according to our experience, the FSC was generally lower than 200 ppm, and the ship was likely to meet the emission requirements.

In the conclusion, we discuss that:

1. In about 10% of the cases, the UAV did not measure the effective background value and peak value. This is mainly caused by the UAV missing the plume during its flight. Therefore, effective methods for finding and navigating to plumes using real-time sensor feeds need to be explored.
2. In about 10% of the cases, the absolute error was more than 0.03% (m/m), and even more than 0.05% (m/m) in rare cases. Unstable concentrations of SO_2 or CO_2 in the atmosphere just before the measurement may cause such errors. Furthermore, uncertainties, such as sensor uncertainty, exhaust uncertainty, measurement uncertainty, and calculation uncertainty, may hinder accurate measurement.

Paper: This could be a very useful paper with lots of detail. Especially the level of detail is useful since this is an area with a lot of development and sharing of these new results could very helpful to other scientists. I provide some comments that could help to make the paper a bit clearer in some areas. See my specific comments below.

Thank you for the comments, we are very encouraged.

Figure 1: I am not familiar with UAVs and in a first glance I thought the black

box mentioned in the text was the large flight case black box below the drone. Page 3 line 16. Not everybody may be familiar with the word "Pod".

Yes, this "Pod" was designed and customized by us. It's not a commercial product. At first glance it may indeed seem puzzling. We have explained it in more detail in the title of figure 1 and text.

Figure 1. Image of the modified UAV platform. The black box installed under the UAV is a pod which was designed and customized by us. It carries a gas pump (to collect the ship's exhaust), gas circuit, a filter (to remove water vapor), sensors for SO_2 and CO_2, a small motor (to provide energy for pumping), a camera, and communication modules.

In the experiment, we used the MATRICE 600 UAV (SZ DJI Technology Co., Ltd.), and modified it. We designed and customized a special pod, which was installed underneath the UAV, to carry sensors, communication circuit boards, gas circuit systems, and other modules, as shown in Fig.1.

Page 4 last sentence: electrochemistry method. Electrochemical method?

Electrochemical method, the term has been rewritten.

Page 5 line 12-13. These sentences are rather unclear. What is meant with 180 working hours apart? Each 180 working hours? It is not entirely clear what the actual accuracy is if it is 1% full scale.

This sentence has been rewritten. The accuracy is written as ±0.25 ppm for SO_2 and ±50 ppm for CO_2, respectively.

Sensor calibration is required when the equipment is used daily. The time interval for sensor calibration is three months or when the accumulated working time of the sensor exceeds 180 h. If either of these conditions is met, calibration will be carried out.

Page 7 line 16: correction should be corrected. Gradually establishing a quality management system.... Is rather vague what is meant. Please rephrase.

These sentences have been rewritten.

As for sensor uncertainty, the linear error is negligible and the nonlinearity of the two sensors should be no more than ±1%. It can be corrected through frequent calibrations with standard gases and gradually establishing a quality management system comprising sensor linearity, sensitivity, repeatability, hysteresis, resolution, stability, drift, and other attributes of the minimum requirements.

Page 7 line 22. Here 200 ppm is mentioned where in other places in the text 0.03 % (300 ppm) is mentioned. This should be explained or there should rather be only one number. Same place: the deviations mentioned in Balzani et al. (2014) were determined at FSC of 1%. It is not clear whether these deviations are still the same at 0.1% FSC. They could be lower at 0.1% FSC content. The authors should mention that or provide more information (which would be useful)

Yes, the measurement range of the FSC is very important information when discussing the measurement results. We supplement the information of measuring range when we discuss the relative precision. We have made the following description for the "200ppm".

Exhaust uncertainty arises because not all the sulfur in the fuel is emitted as SO_2. Preliminary studies showed that 1-19% of the sulfur in the fuel is emitted in other forms, possibly SO_3 or SO_4 (Schlager et al., 2006, Balzani Lööv et al., 2014). Hence, the assumption that all sulfur is emitted as SO_2 yields an underestimation of the true sulfur content in the fuel. Accordingly, this factor needs to be considered when setting the alarm threshold of the FSC. In our experiments, this uncertainty factor led to low FSC estimation results, and the deviation was generally not more than 200 ppm. This prediction is based on the fact that several measurements of some plumes were taken at particular times. Similar calculation results for FSC were obtained, but they were all less than the real value of 100–200 ppm. This is probably because not all the sulfur in the fuel is emitted as SO_2. This tendency of underestimation has also been found in previous studies (Johan, R et al. 2017).

[1] Balzani Lööv, J. M., Alfoldy, B., Gast, L. F. L., Hjorth, J., Lagler, F., Mellqvist, J., Beecken, J., Berg, N., Duyzer, J., Westrate, H., Swart, D. P. J., Berkhout, A. J. C., Jalkanen, J.-P., Prata, A. J., vander Hoff, G. R., and Borowiak, A.: Field test of available methods to measure remotely SOx and NOx emissions from ships, Atmos. Meas. Tech., 7, 2597–2613, doi:10.5194/amt-7-2597-2014, 2014.
[2] Johan, R., Conde, V., Beecken, Jörg and Ekholm, J.: Certification of an aircraft

and airborne surveillance of fuel sulfur content in ships at the SECA border, CompMon (https://compmon.eu/), 2017.
[3] Schlager, H., Baumann, R., Lichtenstern, M., Petzold, A., Arnold, F., Speidel, M., Gurk, C., and Fischer, H.: Aircraft-based Trace Gas Measurements in a Primary European Ship Corridor, proceedings TAC-Conference, 83–88, 2006.

Page 7 last paragraph. To me it is not clear how errors in determination of the peak height is propagated in the total error and it is not clear how this is done. The error of 300 ppm is (it seems) related to the comparison with the on-board samples. And not from error propagation analysis as far as I can tell. It would be nice to show the error propagation numbers as well and see how well these two approaches match. In general, I think that the uncertainty discussion could be more quantitative.

The results of the FSC are derived from the calculation of four data. Therefore, errors or incorrect selection of these four values can affect the results of the FSC. Therefore, the law of error propagation can explain the uncertainty. I have supplemented the error propagation formula of the FSC formula to illustrate this problem. Currently, the data we can obtain are FSC estimates (derived from four measurements) and FSC true values (derived from chemical validation of the fuel). Currently, only multiple measurements of the same plume or multiple peaks using the same measurement can be used to analyze its uncertainty.

Calculation uncertainty lies in selecting the background and peak values of SO_2 and CO_2. According to the law of error propagation (widely used in surveying, mapping, and statistics), the relationship between the deviation in the measurement values and that in the FSC can be obtained. The FSC calculation results are functions of independent observations $SO_{2,peak}$, $SO_{2,bkg}$, $CO_{2,peak}$, and $CO_{2,bkg}$ as in formula 1. The relationship between the observation error ($\Delta SO_{2,peak}$, $\Delta SO_{2,bkg}$, $\Delta CO_{2,peak}$, and $\Delta CO_{2,bkg}$) and function error (ΔFSC) can be approximated using the full differential of the function as follows:

$$\Delta FSC = \frac{\partial f}{\partial SO_{2,peak}} \Delta SO_{2,peak} + \frac{\partial f}{\partial SO_{2,bkg}} \Delta SO_{2,bkg} + \frac{\partial f}{\partial CO_{2,peak}} \Delta CO_{2,peak} + \frac{\partial f}{\partial CO_{2,bkg}} \Delta CO_{2,bkg} \quad (2)$$

In our study, this deviation was generally in the order of hundreds of ppm, as explained in section 4.

Page 9 line 16: "this makes the FSC value relatively larger than that of CO_2". It is not clear what is meant here.

These sentences have been rewritten.

$$FSC[\%] = \frac{S[kg]}{fuel[kg]} = \frac{SO_2[ppm]\cdot A(S)}{CO_2[ppm]\cdot A(C)} \cdot 87[\%] = 0.232 \frac{\int (SO_{2,peak} - SO_{2,bkg}) dt[ppb]}{\int (CO_{2,peak} - CO_{2,bkg}) dt[ppm]} [\%] \quad (1)$$

As in Eq. (1), a higher SO_2 peak leads to a higher FSC estimate, while a higher CO_2 peak leads to a lower FSC estimate. As discussed in section 3.3, not all the sulfur in the fuel is emitted as SO_2, which will result in a lower estimate value. This selection allows the estimate to be relatively close to the true value.

Page 9 line 6: were synchronized is rather vague. Please explain Page 9 In general, the data treatment is unclear to me. Why are peak values taken to compare SO_2 and CO_2? Or is it the surface area? The S-content may be derived from any set of concentrations. Taking the pea area ins just of way of averaging. It seems to me now that the peak position and its height is depending on the performances of the sensors (especially response time) and the accidental position in the plume. This could lead to uncertainties especially if the peak height only is used. This should be explained better. Especially the "approach" could be elaborated more. Sometimes I am in doubt whether peak means the highest point in the concentration or the peak area.

We made the following explanation in the manuscript:

The response time of both sensors is less than 1s. Even if the sampling rates of the two sensors are set to be consistent, the two sensors cannot be completely synchronized. This makes it difficult to calculate the ratio of SO_2 and CO_2. Our approach is that the sensor sends the average measurement value of the last 10 s to the receiver at an interval of 10 s. Therefore, the interval of integration in formula (1) is 10 s. We determined that taking the mean of measurements directly or at shorter intervals leads to too many narrow peaks in one measurement process. This makes it difficult to select the peak value, and the calculation results are unstable. At the same time, the interval should not be set too long, which will make the crest very inconspicuous or too flat. Therefore, we selected 10 s as the empirical parameter value after several experiments.

Also, in the description of result:
After the measurement of plume 5, the communication module was fault when we wanted to adjust sampling rate. We consequently replaced the communication protocol "HTTP protocol" with the "TCP/IP protocol". The main changes involved adjusting the data sampling rate from 10 to 2 s to make it easier to find the peak value (the sensor sends the average measurement value of the last 10 s to the receiver at an interval of 2 s), and the sensors were consequently recalibrated by standard mixture gas.

Page 9 in general: what exactly is "selected". This should be made clear. Now it seems a bit arbitrary. Of course, full range values are not used. But what are dramatic changes? Would be useful to explain.

These sentences have been rewritten:

The global maximum values are selected as peak values for calculating the FSC.

The peak values resulting from dramatic changes (for instance, the change in CO_2 exceeded 500 ppm, or SO_2 changes by more than 500 ppb) in continuous observations are ruled out. This may be because of exhaust uncertainty.

Page 9 line 21: 300 ppm at what level??

These sentences have been rewritten:

However, the final deviation generally does not exceed 0.03% (m/m) at an FSC level of 0.04% (m/m) to 0.24% (m/m).

Page 10: Figure 5. Sometimes background values of SO_2 are 400 ppb? That is very high. Why not subtract the background? Also in plume 6 the background seems to fluctuate very much. makes interpretation of peaks uncertain. please discuss.

We have made the following discussion in the manuscript:

The background value of CO_2 in plumes 1-4 exceeded 300 ppm, but the global background CO_2 was approximately 400 ppm. Meanwhile, the background value of SO_2 exceeded 400 ppb at some time. This was due to sensor calibration, which did not affect the final result. This kind of situation did not happen again after we recalibrated the sensors by standard mixture gas. In some cases, background values seem to fluctuate very much. This is mainly because the UAV took off from the dock, where multiple ships were berthed and wind speeds were high. Therefore, we used the flight procedure given in section 3.1 to minimize this impact.

Page 11 table. Why is not a graph provided? Such as true value (x-axis) against estimated value (y-axis). Then also a correlation coefficient could be calculated. Also a good measure of quality. In general: The results section could improve in clarity if some structure was used: data treatment; FSC observed etc. For example, the issues with sampling rate etc. (page 1 top) are perhaps important but mixed here with the results. To increase clarity this could be treated separately Conclusions High precision is not reasonable to state in view of the rather large underestimations.

This suggestion is very helpful; we added the result graph of all the data.

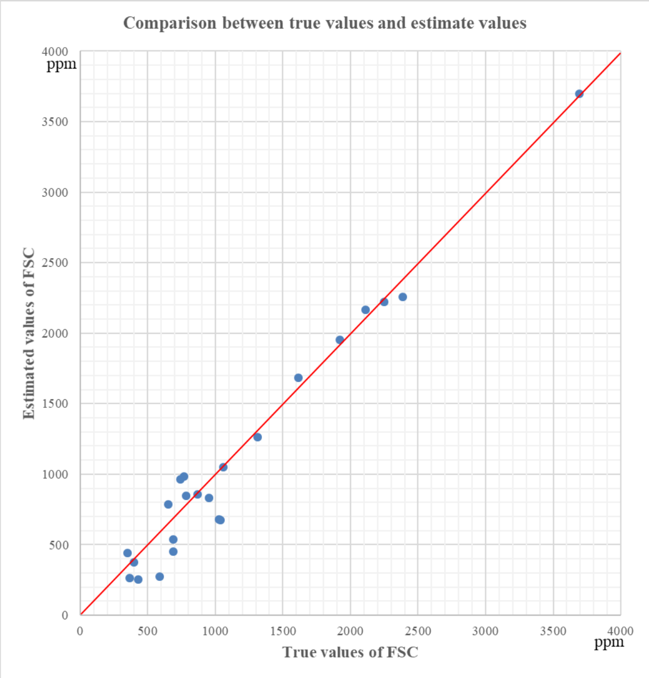

Figure 6. Comparison between the true values of FSC (x-axis) against the estimated values of FSC (y-axis) of 23 times measurement.

As shown in Fig 6, the FSC in our experiments was mainly at a level of 0.04% (m/m) to 0.24% (m/m) (only one measurement of 0.37% (m/m), not enough for reference). The deviation of the estimated FSC value calculated using the proposed method was within 0.03% (m/m), although there was some uncertainty. Considering the uncertainties listed in section 3.3, the proposed method provides accurate results. Overall, the estimated FSC is smaller than the true value in many cases. This is because 1–19% of the sulfur in the fuel is emitted in other forms, possibly SO_3 or SO_4.

In general: The results section could improve in clarity if some structure was used: data treatment; FSC observed etc. For example, the issues with sampling rate etc. (page 1 top) are perhaps important but mixed here with the results. To increase clarity this could be treated separately

The structure has been adjusted. 4.Results: 4.1 Data treatment; 4.2 FSC estimation.

Page 12: in Conclusions something might be said on the effect of SO_3 and SO_4
Specific comments:

As mentioned above, we explained it in the results. At the same time, we make the following explanation in the conclusion.

The estimated results were compared with the FSC values determined at certified laboratories. In general, the deviation of the estimated FSC value was within 0.03%

(m/m) at an FSC level of 0.035% (m/m) to 0.24% (m/m). Because not all the sulfur in the fuel is emitted as SO_2, the estimated FSC is smaller than true value in many cases. Therefore, if the maritime department wants to take the estimated value as the basis for the preliminary judgment regarding whether the ship exceeds the emission standard, it needs to set an appropriate threshold and a confidence interval.

I am not a native speaker, but the English seems fine with me in general. Some specific text could be altered: - on ships the "chimney" is often called the "funnel" - "ship" is normally "vessel". - Culled is not a word that is often used Page 8 line 3: English: none of the monitored ships were fitted with exhaust cleaning equipment

The overall language of the manuscript has been enhanced; thus, any language and grammar mistakes have been corrected to the greatest extent possible.

Some words were changed as follows:

"Chimney" or "funnel": funnel.

"ship" or "vessel": In the relevant literatures, "ship" seems to be used more frequently.

"Culled": Replace with "be ruled out"

"none of the monitored ships were fitted with exhaust cleaning equipment": It has been changed.

In the end, we thank the Referee #2 for his/her positive and constructive comments.