General considerations

It looks like some major effort has been done to make this article acceptable for publication, but still some major issues remain. I did not have the impression that all changes were included in the last version (version 4), although they have been mentioned to be solved in the author responses. I would request a thorough revision before providing the manuscript again to the reviewers.

Analysis of the responses of the author

Comment 1 from reviewer 1: Line 22 : Typo in first emission factor is still not corrected

Response to the answers from reviewer 3

Comment 1: The authors do not mention some highly relevant projects, studies and operations that have been executed, or are ongoing in Europe whether or not with UAV systems on the subject of airborne and remote ship emission monitoring. Although the study has some interesting and innovative aspects, the use of UAV systems for emission monitoring is not new and should not be resented as such.

Answer 1: The authors considered the Reviewer comment, yet the emphasis was intended to be on the fact that EFPN of ships has never been evaluated with UAVs. The updated manuscript has been modified in multiple lines to clarify this.

Response1: The manuscript is still not clear about the focus on EFPN in several places still "ship emissions" in the broader sense are mentioned without emphasising on EFPN (e.g. in the abstract line 23 and in the conclusion line 336)

Comment 2: Line 23: The authors indicate that emissions were assessed during real world conditions. This is not assessed as such as all measurements were performed from and for one ship. Besides the measured RV is a relatively small vessel (94m) while average merchant vessels are in the order of 200-400m. The RV was also running on ultra low sulphur marine diesel fuel while in reality only a fraction of the international merchant vessels use this fuel type. Different factors may influence the successful assessment of ship emissions among others are: ship-type, ship-age, ship-size, ship-shape, shipactivity, fuel-type, funnel height, funnel shape, wind conditions, inversion layers, etc For a realistic assessment during real world conditions these factors should have been elaborated. Furthermore for this study the flight path was based on the ship position, in real life ship position is not known in detail, AIS only provides basic navigation info e.g. there is no information on the location and shape of the funnel on the ship. The limited autonomy, range and payload of the UAV make this UAV not suitable for realistic operational measurements at sea during real world conditions, the study can therefore hardly be used as a proof of concept. For actual (cost-) effective operations offshore, much more robust fixed- or rotary-wing UAV systems should be used, these systems have other specifications (speed, manoeuvrability etc.) than the one used in this study.

Answer 2: The phrase "real world conditions" is intended to indicate that rather than in a lab or simulated conditions, the UAV was launched on a ship performing operations at sea and measured the exhaust plume. The focus of this paper is on a proof on concept of the methodology. It is not a proof of concept for widespread deployment of this methodology in the field for regulatory or commercial use. That is far beyond the scope of this manuscript. The authors disagree that ship type, class, fuel type, and other differing factors would prevent this methodology from being used. Provided there is an exhaust plume which can be intercepted by the UAV, this methodology can be used to assess emission factors of PNC. The wording of the paper has been changed in multiple places to highlight this is a proof on concept.

Response 2: An explanations on "the real world conditions" and "proof of concept" should be included in the text, as the text is not clear on this phrasing. The reviewer agrees that as a proof of concept this method is promising, but stating that this method "provides a reliable inexpensive and accessible way to asses and potentially regulate ship emissions" as mentioned in line 23 is a premature conclusion (see Comment 3)

Comment 3: Line 24: The authors indicate that for the first time ship emissions can be assessed and regulated on a reliable and inexpensive way. This is incorrect, as emissions from ships are already assessed and regulated from both airborne, land based and shipborne sensors in Belgium, The Netherlands, Denmark, Germany and Finland since 2015 at a large scale and on a reliable and cost efficient manner. The use of the UAV's is not necessarily more cost-effective, especially if operated from a ship, and often more time-consuming with less operational output capacity per flight hour. Clearly more information is required to establish cost-effectiveness (platform cost, number of ship measurements per hour, personnel involved, robustness of platform in offshore conditions, ...) . Furthermore the use of UAV's for emission monitoring operations is not new, in 2016 EMSA ordered a feasibility study, granted to CLS, concerning the use of RPAS for emission monitoring (STEAM project), in addition the Danish company EXPLICIT performed some successful emission measurements with small drones. The only aspect which might be innovative in this study is the measurement of PM emissions from ships using drones, but as this is not yet regulated by international law, this has (currently) only academic use.

Answer 3: The novel aspect of this paper is the measurement of PM emission factors using a relatively inexpensive UAV. This is primarily is for academic purposes. However, PM has been identified as critical to both health and climate and thus developing the basis for tools which may have potential regulatory of PM emissions is important.

Response 3: The abstract in version 4 of the manuscript was not changed on this matter, the reviewer request rephrasing the abstract with emphasis on the novel aspects of PM monitoring using UAV.

-UAV's have and are being used for ship emission measurements, so no claim can be made that this is done for the first time.

-Accessibility can hardly be claimed as all operations were performed from an RV to measure the RV itself (it would have been possible if test were conducted from shore to measure ships inbound of ports)

-Reliability has to be assessed on a much wider sample size, this was not assed, neither mentioned in this proof of concept, so no conclusion on reliability can be drawn on reliability.

-No cost efficiency analysis has been made to use this proof of concept for ship emission monitoring (others than the ship it is operating from), therefore it is premature to call this method inexpensive

-The author mentions "potentially regulate", as no regulation is currently in place for PM from ships, it is premature to mention this

Comment 4: Line 64: The authors make the assumption that manned aircraft are not feasible for airborne measurements of ship emissions, although the EU funded CompMon project clearly showed the feasibility of manned aircraft for operational emission regulatory airborne surveillance (e.g. operations in Belgium with >2500 monitored ships in 3 years and operations in Denmark with >1000 monitored ships in 2 years).

Answer 4: The UAV-based methodology detailed in this manuscript offers an operational setup with orders of magnitude less upfront and operational costs than manned aircraft. The project listed is of a far larger scale and budget than typical research projects.

Response 4: Line 69 still mentions the limited feasibility of manned aircraft for ship emission monitoring. Note that in Belgium more than 1000 ships per year are monitored for FSC (via measurement of SO_2 and CO_2). So claiming that manned aircraft are not feasible is not correct. Deploying drones from vessels to inspect the same amount of ships in the same area (in shipping lanes at ca. 60 km from shore) would have been possible, but at significant higher cost. The Belgian aircraft is able to provide a measurement for less than 200€ per inspection, navigating a ship to deploy a drone will cost a multitude of this amount. Drones have a commentary value in for instance the monitoring of inland waters, ports, or when shipping lanes cross near shore, this nuance was not made, simply stating that aircraft are expensive compared to drones is not correct as this requires a description of the operational framework and the application. Furthermore, large scale drones with sufficient radius require similar budgets than manned aircraft and face similar limitations in matter of flying restrictions.

The authors mention risks as well as a reason why manned aircraft have limited feasibility, like for all flying operations risks are involved and have to be assessed in a risk analysis, therefore the risks have to be considered indeed, but on the other hand do not necessarily limit the feasibility.

Comment 5: Line 142: Sensitivity range for CO_2 is 50ppm, this is important as this is same order of magnitude as the delta CO_2 for measurements at 100m, this aspect should be discussed further in the article in an overall assessment of the margin of error, which is currently missing.

Answer 5:The updated manuscript addresses instrumentation sensitivities and error margins. In particular this comment has been discussed in lines 313-317.

Response: The discussion in the text from line 324 to 328 now sufficiently describes this issue

Comment 6: Line146: Significantly more detailed information should be provided on the calibration method (references samples, calibration-factors, offset, : : :). It is also not clear if a calibration was performed before (and after) every measuring day, this should have been done to ensure the validity of the data. Line 147 (Figure S1): More information is required for the comparison of the CPC with the DISC, it is not clear what kind of air samples were used for the comparison, it looks like this is just done based on continuous ambient air measurements on board of the RV, for a proper validation a comparison should be made with real emissions. A comparison of the IAQ with the PICARO is completely missing here. If only a comparison (validation) is possible in a lab, this comparison should at least be done during similar conditions as during the field measurement (exposure time, concentration, temperature,), this is clearly not the case as the particle concentrations is very low in this comparison. It looks like the intercept of the linear regression is not put at zero, why is this, was a zero calibration performed? Especially for CO2 it is important to perform the calibration in the same range as the measurement range as the IR absorption is nonlinear, no comments were made on this aspect in the article. Furthermore it should be noted that a linear regression is not an ideal method to compare 2 sensors, the Bland Allman method is more appropriate (Statistical Methods for Assessing Agreement Between Two

Methods of Clinical Measurement," by JM Bland and DG Altman, The Lancet, February 8, 1986, 307-310).

Answer 6: Methodology has been expanded upon significantly in the updated manuscript and a C02 picaro comparison is provided in supplementary material.

Response6: Figure S2 does not use a linear regression like for the comparison between the PM sensors. An explanation is missing why a different comparison method was used.

Line 151 in the CO2 section refers to Figure S1 instead of figure S2.

Figure S2 shows the extensive variation and measurement difference between the Picaro (393ppm) and the IAQ-Calc (486.5ppm), this issue should have been elaborated in the text (see Comment 6). Please explain why the authors did not choose to conduct a calibration with a set of reference span gasses (including zero gas) as this is the most common way to calibration gas analysing sensors?

The sensor comparison in Figure S1 uses a linear regression (that intercepts at 0) therefore a calibration can be done of the DISC data using the regression coefficient as the main calibration factor, for the IAQ only an offset is used (93) and no calibration factor, please explain why this different approach was used.

Comment 7: Line 158: Flight speed is here expressed as 1.5m/s, it is not clear if this is the airspeed or ground speed. If this is the airspeed, the actual ground speed will depend on the wind conditions, therefore the flight speed through the plume is dependent on the wind conditions too. During the first day, the wind was cross on the ship heading. The plume would be expected at 180° if transect were flown with alternating heading 250° and 70° (perpendicular to the ship heading), the transect with heading 250° would have been flown with a significant different ground speed (ca. 6.5 m/s instead of 1.5 m/s), no mention is made of this in the article.

Answer 7:Flight speed listed is the airspeed. Whilst the wind conditions will influence the ground speed, the only influence on the measurements will be a variation in the amount of data points captured inside the plume during transect. The discussion of the amount of in plume data points in a transect and its importance is in the updated manuscript in lines 277-281.

Response7: Wind conditions will have an impact in the usability of this method as they impact the number of data points but also the concentrations in the plume (dilution) and therefore depending on the wind condition the measurement could fall outside sensor sensitivity. Furthermore the transect and flight path have to be adjusted depending on the wind conditions, eg. in case of cross wind the measurements behind the ship would not make any sense.

Comment 8: Line 208: I would suggest adding an indication of the resulting plume location and flight pattern on the graphs. These graphs would also visualise the different airspeed between the transects (see comment line 158).

Answer 8:The emphasis in the graphs is on the clear detection of the plume by each instrument. The authors do not believe that plume locations would not provide any further information and would overcomplicate the graphs.

Response 8: Plotting the estimated location of the smoke plume could visualise the issue that was explained in response 7.

Comment 9: Line220: Only 9 times the plume was sampled, very few statistical conclusions can be made based on this small sample size, especially the linear regression on line 277 is questionable.

Answer 9:The methodology has been updated in the updated manuscript.

Response 9: The reviewer argues that the limited number of successful measurements was not sufficiently considered in the manuscript (see comment 3).

Comment 10: Line 229: The distance (25m) is missing in this sentence.

Answer 10:This has been clarified in the updated manuscript.

Response 10: This is sufficiently clarified by the authors

Comment 11: Line 232: It is mentioned that the CO_2 is up to 100 ppm higher in the plume, this is not clear on the graph (only 50-75 ppm), this will be the part for integration to amount to the delta CO_2. Furthermore it should be noted that the peaks for CO_2 at a distance of 100 m is of the same order of magnitude of the sensor accuracy.

Answer 11:The C02 is up to 144ppm counts above background inside the plume in graph 4(a). The graph has been replotted with background removed in the updated manuscript to clarify this. The short 100m transect data has also been discussed in more detail.

Response 11: This is sufficiently clarified by the authors

Comment 12: Line 262: Another flight transect could have been used where the UAV would be flown at the same speed and heading as the RV and hovered in the plume, this would require a transmission of measurement info to the control station to adjust flight altitude and pattern to successfully find the plume and measure the plume for longer periods.

Answer 12: The focus of this project was the measurement of EFPN through transects of the ship plume. Due to time constraints alternative methodologies could not be investigated, though this suggestion is one of the recommendations for further research listed in the manuscript.

Response 12: This is sufficiently clarified by the authors

Comment 13: Line280: Instead of a comparison between calculated emission factors and the emission factors from previous studies a comparison with the emission factors calculated based on a plume measurement with the other equipment on board of the RV (e.g. Picaro) would have made more sense.

Answer 13: There was no possibility of accessing the plume with the larger instrumentation such as the picaro or CPC. This is one of the primary advantages of UAV-based platforms. A future validation study would look into this. This is a recommendation in the updated manuscript.

Response 13: This is sufficiently clarified by the authors

Comment 14: Line 312: Generalization and misconception that the use of UAV systems would consist of a reduced cost. It is definitely not presented in this article that UAV systems could provide a real cost effective alternative to other surveillance methods as no cost benefit comparison was made between different surveillance methods (both fixed stations and airborne sensors; operational output capacity; personnel and supporting platform etc.) and all missions were carried out from a vessel, which has a higher operational cost per hour as an aircraft and a lower speed and therefore a much lower cost efficiency. Note that a higher cost efficiency could maybe be acquired with this setup where these operations would be combined with other task carried out by patrol vessels, pilot ships or research vessels assuming that these vessels would operate within 2 km of shipping lanes. This was not mentioned in the article.

Answer 14: The authors have addressed this concern in Answer 4, the setup and operational costs of this UAV system are orders of magnitude less than manned aircraft. The focus of this manuscript was on the development of the methodology. Whilst some suggestions for future applications are made, it is premature and beyond the scope of this paper to recommend wide-scale deployments of UAVs and cost benefit comparisons with other methodologies.

Response 14: The reviewer does not want to argue that this cost efficiency comparison should have been made, but claiming that UAV's are less expensive and more effective than other methods also requires a description off the operational framework. The reviewer agrees that for the concept of this research a UAV was most likely less expensive than an aircraft as the RV was monitored during another campaign. Concerning the proximity to the plume, note that the Danish company Explicit is conducting helicopter measurements and is also involved in the development of sensors for drones, in their operational procedures drones fly further away from the ships than the helicopter. The text from line 335 to 339 should be changed as this is based on unfounded assumptions that do not reflect the reality of airborne measurements.

Comment 15: Line 326: SO_2 is completely missing here, SO_2 is the only emission regulation which is effectively monitored using airborne platforms at this moment and should therefore at least be included in the discussion.

Answer 15: The focus of this study was on PN emissions. SO_2 would be an interesting alternate application. To the authors knowledge the main challenge for such a system would be that fast and accurate SO_2 meters are significantly above the payload of any lightweight UAV, include fixed wings.

Response 15: The sensor-system from Explicit is able to measure SO_2, CO_2 and NO. This sensor system is in the order of magnitude of a few kg and is deployed on drones and helicopters (slow time response requires monitoring in smoke plumes for up to 30 sec). The sensor used by the Belgian coastguard measures SO_2 and CO_2, weights 40 kg and has conducted more than 3000 measurements, this sensor uses a Thermo 43I TLE that has been specially modified for faster time response (1-2 sec)