Response to Reviewer #2

We thank the reviewer for his/her comments, which have helped us to improve the manuscript. Our responses to the specific queries follow below. Our responses are given in blue text.

Although the presented data is very important and valuable interpretation needs to be more focused. In the present form, the data has been used just to compare among the ship vessels, but causes of variability is not very clear. The differences in EFs between different categories are small compared to variability in a given type of ship/engine exhaust.

We are unable to establish precisely the reason for the variability. However, we note that it is expected that engines can exhibit different emission factors depending on their particular operating conditions, age and level of maintenance. We do not have information on the particular operating conditions of each vessel, and believe that doing this is beyond the scope of the manuscript (not to mention, near impossible given that most ships would be unlikely to provide us with this information). Our intention is to simply capture the actual variability in emission factors for vessels operating in the real world, to provide a complement and contrast with more targeted studies that investigate emissions from either single engines or single ships.

The paper does not highlight: how such small changes in EFs with respect to type of ship/engine can be useful to develop inventory or for air quality study?

It is not our intention to actually develop an emission inventory with this manuscript. However, the raw emission factors provided for each vessel (in the supplemental material) can be used by those interested in the technical details of emission inventory development.

Sulfur content has been the issue of the discussion in the beginning but later its role in EFs (or variability) has been sidelined.

There is an entire section dedicated to understanding the role of fuel quality (i.e. sulfur content) on BC emissions, as assessed through comparison with other measurements from both in situ real-world operation and targeted single vessel or engine studies. (See section "Literature comparison and influence of fuel quality on EFBC.") So, it is unclear to us in what way this is "sidelined."

Several BC instruments have been inter- compared, which is good, but on what way it is required or how it is used to estimate final BC emission factor is not very clear.

It is not clear to us what is meant by "on what way it is required." It is "required" in the sense that it is important to understand how well different technologies can measure the same property under different conditions (here BC in ship plumes).

We have clarified that all averages have been calculated excluding the SP2 measurements, since it was established that this technique can exhibit a negative bias when measuring BC in fresh ship plumes.

Several key figures are presented as suppl. which can very much be part of the draft as there are only 5 Figures. Overall, Figure representation also needs some improvements. A discussion on general importance of EF in context of regional air quality and climate change is also necessary. Several of recently published works related to BC inter-comparison and microphysical properties of BC over California deserve citation.

We have moved the two figures showing the cruise ship tracks and locations of ship plume intercepts to the main text (new Figure 1). Regarding the comment that "Figure representation also needs some improvements," without further guidance as to how to improve the figures we have retained them as originally presented.

We have added a reference to the use of EFs in general. Specifically, "Measurements of emission factors for different sources, such as ships, are required for development of emission inventories, such as the US Environmental Protection Agencies NONROAD model (U.S. EPA, 2005)."

We have added a citation and explicit comparison with Sahu et al. (2012) in Fig. 3, who reported rBC size distributions in fossil fuel plumes over CA in June 2008.

In the phrase ": : :which measure the mass concentration of refractory BC directly", I do not think that all these instruments measure the mass concentration DIRECTLY.

This is a question of semantics and what one thinks of as a "direct" measurement. Both the SP2 and SP-AMS are calibrated to particular BC standards based on the mass of the individual BC particles. We believe this constitutes a direct measurement, and thus retain this terminology.

Moreover, importance of data measured using four independent instruments is not clear. I mean, why the data from one good instrument (for BC) is not sufficient? At least for such studies.

We believe that one should take advantage of multiple measurement methodologies, if they are available. Especially if a particular technology being used has not be specifically applied to a particular setting previously (such as for ship plume sampling).

The interception of vessels is random in that case general comparison between CalNex and TexAQS is not worthy, some specific aspects can be highlighted.

It is unclear what "specific aspects" the reviewer is referring to. It is difficult to respond without having further details. However, to the broader point, yes, the sampling is random. This is why we have performed a rigorous statistical analysis of the observations from both campaigns. This allows us to make a general comparison between results from these two separate campaigns.

Page 24677 Line 10-20: In this region (California coast), little more details of for in-use ships (fractions of fuel consumed on annual basis, etc.) using HFO, MGO and MDO will provide wider acceptance of such case studies.

Current regulations make it such that HFO is not used in the California coastal waters. We do not have statistics on the particular breakdown between MGO and MDO available. We have modified the appropriate sentence in the introduction to read "Similarly, in 2009 California began regulating the sulphur content in fuels used by ships travelling within 24 nautical miles of the California coast, lowering FS to ≤ 1.5% (MGO) or ≤ 0.5% (MDO) in 2009, to ≤ 1% (MGO) as of August 2012 and to ≤ 0.1% for both MGO and MDO by January 2014, with HFO usage not allowed (CARB, 2011a)."

Page 24679 Line 10-20: However, without a reference species, how the measurement of SO2 alone can confirm that the type of vessel uses LSFs or HSFs? This information is also important considering the distance of plume source from Atlantis varied hence the travel time and dilution of plumes.

The reference species is CO2. We have used the method of Williams et al. (2009) to determine the fuel sulfur content. We have modified the sentence as follows: "…confirmed through in situ measurement of SO2 relative to CO2 within plumes)."

I do not find anything new in this section than already reported in other studies. And the contents do not deserve a separate (subsection). However, key information can be merged to subsection "2.3.2".

Although this does not cover new ground, because we are using multiple measurement methodologies that detect different "types" of black carbon, we believe this to be useful for the reader who is not as familiar with the different terminology.

There can be a realistic situation, when the distances between intercepted vessels are not very far. In this case, a measured plume may have got mixed, then how the plume is identified and separated? Is the wind direction/speed has been used to filter the data? How about the plume-to-plume variation in ABC and ACO2, do these vary significantly?

As was stated, "Plumes were identified based on the relative ship positions and winds, and by a noticeable increase in [CO2], babs or [BC] above background." We have updated this to state "Plumes were identified based on the relative ship positions and wind direction and speed, and by a noticeable increase in [CO2], babs or [BC] above background." Also, yes, the absolute values do vary significantly between plumes. We have added a sentence "The absolute values of [CO2], babs and [BC] vary from plume to plume."

In the context of inter comparison between different BC instruments I recommend to cite following works.

We now cite the Kondo et al. (2011) paper and added the sentence "Some previous studies have investigated the comparability of the different measurement methods (Kondo et al., 2011;Cross et al., 2010), however such investigations were not performed for BC particles in "fresh" ship plumes."

In Figure 2, Size distributions separately for the categories of slow speed diesel (SSD), medium speed diesel (MSD) and high speed diesel (HSD) will be interesting and informative than just average.

We now include complementary figures for the engine type categories.

A brief comparison with respect to the fresh vehicular exhaust (size distribution) in California will be important. In this context I would suggest to compare the size distributions of BC presented in following study for fossil fuel category over California. It suggested to cite following paper which also presents BC size distribution and emission ratio of BC for two different types of plumes over California.

We now compare with the rBC size distributions observed in “fossil fuel” plumes from Sahu et al. (2012), explicitly showing the comparison in Fig. 3.