AUTHOR'S RESPONSES TO REFEREE #2:

We thank Referee #2 for the positive evaluation and for taking the time to read our paper and giving us valuable comments to improve the manuscript. Following the reviewer remarks, we addressed the comments and questions in detail below.

1)"The study evaluates the impacts of shipping emissions on the air quality in the region of the Iberian Peninsula and the Strait of Gibraltar, one of the busiest maritime routes in the world. This chemistry-transport modelling study makes use of shipping emissions generated by the STEAM 3 model that allocates ship activities via the Automatic Identification System operating on board the vessels. Among the valuable information presented in this manuscript are a comparison of ship emission intensities with those reported for ports in the Asian region and a calculation of the ship impact on exceedances of regulatory air quality limits. Unfortunately, it is not immediately apparent what the manuscript adds to already published chemistry transport modelling studies on the impact of ship emissions in Europe. Overall, the manuscript reads more like a good technical report than a research article, as the applied methods are not originally proposed and the uncertainties of model results are not comprehensively discussed and quantified."

<u>Answer:</u> Thank you for your comments. With this study, we try to give an overall view of the air quality impact of the shipping emissions over the Iberian Peninsula (specifically) using the STEAM, that is considered one of the most reliable methods to estimate shipping emissions (exhibits the highest spatially resolution in their emissions and a large number of secondary routes that do not appear in other inventories), and EMEP for the air quality modelling. Only another study estimated the concentrations for the Iberian Peninsula, but used different methodologies to do it (different inventory and CTM, referred in lines 69-78).

Furthermore, this is the first time that the modelling results of this new version of STEAM are discussed in detail for the Iberian Peninsula (this inventory was already discussed but from a global point of view, not specifically for the Iberian Peninsula) and the very first time that they were used to estimate the pollutant concentrations for the Iberian Peninsula.

Moreover, to support our results a model validation will be added. Model output $PM_{2.5}$, PM_{10} and NO_2 concentrations for the S-SCN scenario were compared with data from the monitoring stations of the EU Member States reported by the European Environmental Agency for 2015. Moreover, comparisons between the modelling reference results reported by EMEP for the year 2015 were also compared with the data from the monitoring stations. Annual mean concentrations observed in 139 stations for $PM_{2.5}$, 337 stations for PM_{10} and 446 stations for NO_2

were compared with the model results in time and space. Also, a comparison between the exceedances for the scenario including ship emissions and the levels from the monitoring stations was also made. We estimate the percentage of exceedances that our model found in relation to the exceedances detected with the concentrations of the monitoring stations. More information was added to the Uncertainties chapter to improve the discussion.

2) "The spatial resolution is the same as in prior studies that covered the whole of Europe. This is somewhat surprising, given that a prior study by Monteiro et al. (2018) in the same region used a finer resolution (3 km x 3km). The applied difference method for quantifying the shipping emission contribution is flawed since the effect of nonlinearities in ozone chemistry on the ship impact was not evaluated, despite the high photochemical activity in this region. For both daytime and night time, the instantaneous NOx lifetime in ship plumes is a strong function of the initial NOx concentration at ship stack, resulting in a very nonlinear loss rate for NOx in ship plumes (e.g. Song et al.,2003). Model procedures that shift ship plume levels by an order of magnitude, as can be expected for a 10-km wide grid cell, will quite likely overestimate NOx lifetime."

Answer: Thank you for your comments. As shipping is a major source of NOx to the troposphere, and especially because large amounts of these pollutants are often released from point sources into the relatively clean maritime atmosphere, since the version rv4.8 of the EMEP MSC-W a new pseudo-species "ShipNOx" has been introduced and NOx released by ships started to be treated differently and not like any other source of NOx. Like you said in your comment, in 3-D models NOx emissions are typically diluted into large grid volumes which can lead to large over-predictions in O3 production, and in the NOx lifetime. Tests with the EMEP model confirmed these issues and also that the early approach was not appropriate for the European area at least (considered ship-related NOx emissions like any other source of NOx). To prevent such effects, the model assigns 50% of shipping NO_x to the pseudo-species ShipNOx and the rest as NO and NO_2 as previously done. ShipNOx deposits as NO_2, but suffer simple atmospheric reactions:

ShipNOx + OH⇒HNO3 Reaction 1; ShipNOx⇒HNO3 Reaction 2

Reaction 1 proceeds with the same rate as the normal NO2+ OH reaction, thus proceeding faster in daylight and in areas with high-OH. Reaction 2 provides a minimum half-life of about 6 hours, accordingly to Vinken et al. (2011) results.

The heat release from ship stack exhaust of large ships represents a buoyancy flux that may result in plume rise. Therefore, we can expect that a significant fraction of the shipping emissions are emitted at upper heights. The STEAM 3 model should be able to take into account plume rise of ship exhaust in generalized form. A

description of the treatment of the vertical distribution of shipping emissions and injection heights that are used for the corresponding vertical layers of the EMEP modelling system should be added to the method section. When shipping emissions have been fully transferred to the lowest vertical model layer, such a procedure has to be justified and the error due to this needs to be approximated.

Answer: Thank you for your comments. The estimated height distribution of emissions from STEAM are given in the figure below. According to these estimates, over 80% of emissions occur between 30-60 m height. The plume release height is a function of vessel type, size and plume rise. For the two former, vessel stack height is determined from photographs and existing data from IHS Markit. However, the airdraft (how high the vessel is from the water surface) or the keel-to-mast height are rarely available in commercially available databases. For this reason, we have used vessel scale drawings and photographs from Significant Ships publication serie, to link vessel types and sizes to stack height. There is a linear dependency between vessel stack height and vessel length, but these linear functions are vessel type specific. Regardless, STEAM does not consider plume rise, because exhaust temperature, exhaust velocity and funnel pipe diameter are not known. In principle, some typical values could be used, but currently STEAM only adds a constant value of 10m to stack height estimation to provide a primitive estimate of the plume rise.

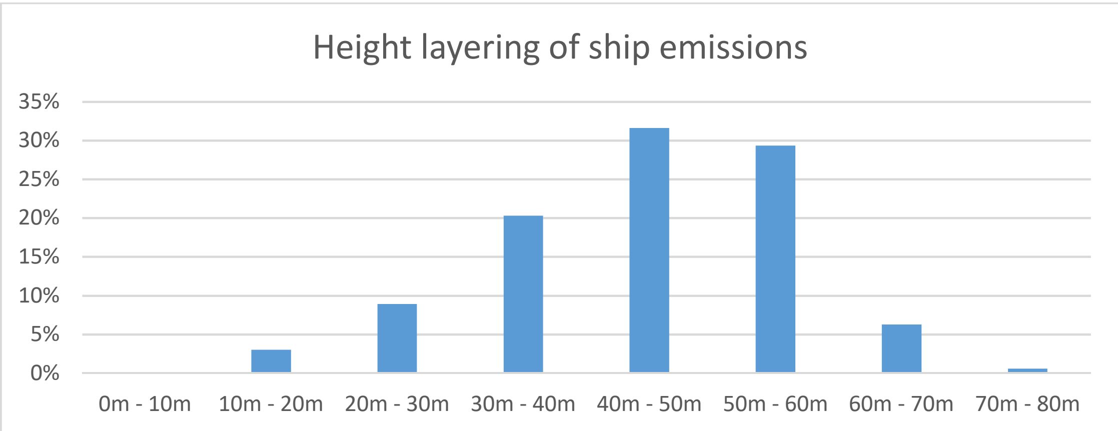

Figure 1 Estimated height profile for ship emissions around the Iberian Peninsula domain.

The detailed height profile for emissions obviously depends on the type of traffic operating in the area. Largest cruise vessels can have very high stacks, exceeding 70m height and plume rise may elevate the plumes even higher. However, STEAM does not currently consider meteorology during emission calculation. In our opinion, the plume rise issue in ship exhaust dispersion is most relevant for local scale air quality assessments, but less so for regional scale work. For this reason, the ship emissions from STEAM were allocated to the first model layer of the EMEP runs. Information related to this issue will be added to the Methods section as follows: "… having a thickness of 50 m. Assuming that the plume rise issue in ship exhaust dispersion is more relevant for local scale air quality assessments, and less for regional scale work. For this reason, the ship emissions from STEAM were allocated to the first model layer of the EMEP runs."

"The significance of the modelled ship contribution was not validated with measurements. Although the Norwegian Meteorological Institute regularly validates the air quality predictions with the EMEP MSC-W model for Europe, it is not sufficient to simply refer to this. The manuscript should include a validation of the modelled concentrations in the subdomain region with monitoring data from stations in Portugal, Spain and France for 2015 (EMEP network, EEA AirBase, EBAS database). The comparison should include model data from both runs with and without shipping emissions."

Answer: Thank you for your comments. To support our results, model output $PM_{2.5}$, PM_{10} and NO_2 concentrations for the S-SCN scenario were compared with data from the monitoring stations of the EU Member States reported by the European Environmental Agency for 2015. Moreover, comparisons between the modelling reference results reported by EMEP for the year 2015 were also compared with the data from the monitoring stations. Annual mean concentrations observed in 139 stations for $PM_{2.5}$, 337 stations for PM_{10} and 446 stations for NO_2 were compared with the model results in time and space. Information about model validation will be added in the Methods section as follows: "…and recent studies that used the model to assess the effects of shipping emissions (Jonson et al., 2015, 2017; Turner et al., 2017). To support the results of the present study, model output $PM_{2.5}$, PM_{10} and NO_2 concentrations for the S-SCN scenario were compared with data from the monitoring stations of the EU Member States reported by the European Environmental Agency for 2015 (EEA, 2020). Moreover, comparisons between the modelling reference results reported by EMEP for the year 2015 (Norwegian Meteorological Institute, 2019) were also compared with the data from the monitoring stations. Annual mean concentrations observed in 139 stations for $PM_{2.5}$, 337 stations for PM_{10} and 446 stations for NO_2 were compared with the model results in time and space. Table 1 summarizes the model quality indicators (Pearson correlation coefficient (Pearson's r), Mean Bias Error (MBE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)), for the present study estimations and for the reference results reported by EMEP. Similar quality indicators were obtained for the comparison the results of the present study and the reference results of EMEP, which indicates that the model simulations were well executed. Although the correlations obtained were moderate positive correlations (Pearson's r > 0.5) for all pollutants, the errors obtained were smaller than those reported in the literature (Monteiro et al., 2018), which makes our results acceptable."

Table 1. Model quality indicators for the present study estimations and for the reference results reported by EMEP.

Indicators	This study			EMEP reference		
	$PM_{2.5}$	PM_{10}	NO_2	$PM_{2.5}$	PM_{10}	NO_2
Pearson's r	0.57	0.55	0.70	0.64	0.55	0.67
MBE [a]	1.32	19.51	5.78	0.34	18.70	5.19
MAE [b]	2.86	19.55	8.70	2.81	18.74	9.18
RMSE [c]	3.62	20.83	11.24	3.59	20.11	11.90

[a] Mean Bias Error; [b] Mean Absolute Error; [c] Root Mean Square Error

Specific Comments:

1.) P.1 lines 24-26: Many studies can be found about the impacts of shipping emissions on air quality and health. It would be a good place here to discuss deviations and contradictions in the literature concerning the relevance of shipping for health impacts, and specifically the roles of primary versus secondary particulate matter.

Answer: Suggestion attended. We will change the following sentences: "Marine traffic has been identified as a relevant source of pollutants especially nitrogen oxides (NO_x), sulphur oxides (SO_x) and particulate matter (PM), which may lead to known negative effects on air quality and health, being its contribution to human health degradation still not well documented (Brandt et al., 2013; Corbett et al., 2007; Nunes et al., 2017b; Sofiev et al., 2018). Studies have been reporting that shipping contributions to ambient PM in port areas are mainly secondary particles (around 60 to 70% of PM_{10} and $PM_{2.5}$ mass concentrations). Despite this, studies have been suggesting that could be more advantageous to reduce shipping-related primary particle emissions than precursors of secondary particles (NO_x and SO_x), which are the target of current international regulations (Viana et al., 2014)."

2.) P.4 lines 111-114: Which boundary conditions of the chemical concentrations were used for the subdomain runs?

Answer: Information about boundary conditions used in the chemical transport model for the subdomain runs will be added as: "…400 km from the Iberian Peninsula coast. The initial and the lateral boundary conditions for most of the chemical compounds were defined by functions defining concentrations in terms of latitude and time, based on measurements and/or model calculations, providing robustness which chemical transport model results sometimes lack. More information about the EMEP/MSC-W configuration for initial and boundary concentrations used in this study can be found in Simpson et al. (2012)."

3.) P. 5 line 157 to P. 6 line 173: Suggest to transfer the information of annual average emission intensities (per pollutant and per port/sea area) into a table to facilitate the comparison with shipping activity in the Asian region.

Answer: Suggestion attended. We will change the following sentences: “The annual average intensities of ash, CO, CO_2, EC, NO_x, OC, sulphate and SO_x emissions were 9.0E-04 tonnes/yr/km^2, 1.38E-02 tonnes/yr/km^2, 8.47 tonnes/yr/km^2, 1.27E-01 tonnes/yr/km^2, 1.97E-01 tonnes/yr/km^2, 3.16E-03 tonnes/yr/km^2, 8.04E-03 tonnes/yr/km^2 and 1.01E-01 tonnes/yr/km^2, respectively. The annual average and highest intensities for NO_x and SO_x reported for the Asian Region are present in Table 3 (Chen et al., 2016a, 2017; Fan et al., 2016). In general, the average intensities that were reported for Asia were considerably higher than those found in this study. It was possible to identify in the present study two main hubs given the high emissions intensity: Valencia Port and the Strait of Gibraltar. At Valencia Port, ash, CO, EC and OC had the highest values, respectively, 1.46E-01 tonnes/yr/km^2, 1.85 tonnes/yr/km^2, 1.99E-01 tonnes/yr/km^2 and 5.09E-01 tonnes/yr/km^2. At the Strait of Gibraltar, CO_2, NO_x, sulphate and SO_x had the highest values, respectively, 1330 tonnes/yr/km^2, 24 tonnes/yr/km^2, 1.03 tonnes/yr/km^2 and 11.6 tonnes/yr/km^2. In accordance to what was referred above, in the Asian Region maxima intensities were also higher than those here estimated (Chen et al., 2016b; Fan et al., 2016; Ng et al., 2013).“ We will also add Table 3.

Table 3. Annual average and highest intensities of NO_x and SO_x (in tonnes/yr/km^2) reported from researches in Asian Region.

Study	Port/sea area	NO_x		SO_x	
		Annual average	Highest value	Annual average	Highest value
Chen et al. (2016a)	Tianjin Port	5.06	1.51E+03	7.14	1.79E+03
Chen et al. (2017)	Qingdao Port	1.83	-	1.42	-
Fan et al. (2016)	East China Sea	1.0	1.0E+04	1.90	1.30E+03
Ng et al. (2013)	Hong Kong	-	1.1E+02		2.0E+02

4.) Impact on Air Quality: Suggest to divide section 3.2 in topical subsections; for example “Annual average concentrations” (P. 7 lines 202 - 218), “Comparison with previous studies in the region” (P. 7 line 219 to P. 8 line 250), “Seasonal variation”

(P.8 lines 251 - 260), “Possible health impacts” (P. 8 line 261 to P.9 line 297). Some passages could be shortened.

Answer: Suggestion attended. We will subdivide the section 3.2.

5.) P. 9 lines 276-287: Suggest to illustrate the contribution of shipping emissions to the exceedances of limit values in form of a bar diagram, i.e. showing the increment of number of exceedances (NO2, PM2.5, PM10) and number of days of exceedances(SO2) due to ship traffic for the major ports of the Iberian Peninsula.

Answer: Thank you for your comment. We believe that with a bar diagram the spatial distribution (illustrated with Figure A2) will be lost.

6.) Uncertainties and limitations: The uncertainties of the emission factors of pollutants from different ship types could easily dominate the uncertainty of the evaluated contribution from shipping. With the STEAM 3 model at hand, it should be possible to estimate the overall uncertainty in the modelled concentrations due to uncertain emission factors. To arrive at a more reliable margin of the contribution of shipping emissions in this region, my request to the authors is that they perform shipping emission calculations with the respective lower and upper bound of the emission factors of NOx, SOx and primary particulates, then repeating the runs with EMEP MSC-W using the lower and higher emission dataset.

Answer: Thank you for your comments. In STEAM, there are several sources of uncertainty which can have an impact on the accuracy of the results. These could be classified in three categories:

a) Gaps in input data (incomplete AIS coverage, missing IHS Markit data)
b) Power prediction (weather contributions, Hollenbach resistance inaccuracy, fouling, squat, sea currents, aux engine power profiles, engine load estimation, power transmission, propeller properties)
c) Emission factors (specific fuel oil consumption, fuel type, fuel sulphur content allocation, engine generation)

Each of these three categories can have multiple contributions, indicated by various error sources in parenthesis. STEAM has mechanisms to mitigate most of the uncertainties listed above and some are features, like weather, are currently developed, but will be reported separately at a later stage. Uncertainties concerning emission factors may be larger for products of incomplete combustion, like CO, NMVOC, OC and EC, than for CO_2, or NO_x, because these are strongly related to engine load, engine generation and service history. The emission factors may also depend on the fuel type assignment and fuel sulphur content, which are estimated based on engine properties and maximum sulphur content allowed in each region at the time period of the study. However, the emission factors for incomplete combustion products may be affected by engine service history and thus are notoriously difficult to estimate. We are not currently aware of any study which would provide uncertainty evaluations for all emission sectors and emission factors included in regional air quality modelling and it seems curious to us to demand one for shipping, only. To conduct such a study would require many computer simulations and significant additional effort. Even if these tests would be limited to uncertainty evaluations of three air pollutants modelled by STEAM, it would still require low- and high-bound runs with STEAM and consecutive analysis with the EMEP model. Even then, the uncertainty evaluation would not be just about emission factors, because primary particulates would also require adjusting the assumptions concerning how the fuel sulphur content was assigned and what are its consequences on PM components. This will multiply the work required by at least a factor of six, which is not currently possible due to limitations in available research funding.

Technical Corrections:

P. 1 lines 17-18: "ktonnes y-1" is not a SI unit.

Answer: Given that the values are high, it is usual to present the results in above referred units. We maintained the units to be more coherent with the literature.

P. 1 line 27: on a global scale?

Answer: Suggestion attended. Yes, it is on a global scale. This information will be added.

P. 1 lines29-30: reference(s) for this statement missing.

Answer: Suggestion attended. The reference will be added.

P. 3 lines 66-67: suggest to reference the study of Ramacher et al. (2019) on local scale for Baltic Sea ports.

Answer: Suggestion attended. The reference will be added.

P. 5 line 150: "ash" – what is this chemically? Please define.

Answer: Suggestion attended. "Ash" will be defined.

P. 8 line 232: please replace "lower increases contributions" by "lower positive contributions".

Answer: Suggestion attended.

Conclusions: the word "verify" is used several times in the conclusions section (P. 10,line 321; P. 10, line 324; P. 11, line 330). Verification implies the comparison of model results to the true values, which are not known. Please change wording.

Answer: Suggestion attended. The word “verify” will be changed by “observe” and “detect”.

P. 11 line 340: what about the code availability of STEAM 3? Please include a statement here.

Answer: Suggestion attended. The following statement will be added:” STEAM model is intellectual property of the Finnish Meteorological Institute and is not publicly available”.

P. 15 lines 469-471: the citation of Marelle et al. is incomplete.

Answer: Suggestion attended. The reference will be changed.

Table 1 and Table 2: “tonne y-1” is not SI unit.

Answer: Given that the values are high, it is usual to present the results in above referred units. We maintained the units to be more coherent with the literature.

Figure 1 and Figure 2: please use SI units in labels, axis annotations and captions.

Answer: Given that the values are high, it is usual to present the results in above referred units. We maintained the units to be more coherent with the literature.

Figure 4f: what is the cause for high O3 values along the North African coast overwater?

Answer: Thank you for your comment. There are no O3 sinks in this region.