AUTHOR'S RESPONSES TO REFEREE #3:

We thank Referee #3 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:

Major comment

My main criticism of the manuscript is that the results on additional exceedances of air quality standards due to ship emissions (which are potentially of interest to policymakers) depend strongly on the quality of the modelled fields. Only if they give a good representation of the actual air quality and exceedances, then the difference between the results with and without ship emissions can be trusted. I therefore believe that the authors need to include a comparison of the modelled concentrations and exceedances for the scenario including ship emissions to those measured by in-situ air quality net-works to demonstrate that they are close enough to reality to make interpretation of delta exceedances worthwhile.

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 the reference results reported by EMEP. Similar quality indicators

were obtained for the comparison of 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 make our results acceptable."

Table 1. Model quality indicators for the present study estimations and 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

Moreover, a comparison between the exceedances for the modelled scenario including ship emissions and those calculated with the data from the monitoring stations was also made. We were able to compare only the exceedances for $PM_{2.5}$, PM_{10} and NO_2 since we didn't have daily SO_2 concentrations data from the monitoring stations. Information about the comparison of the exceedances found with the model and with the data from the stations will be added in the Methods section as follows: "…and PM_{10} (20 μg m^{-3} for annual mean) (European Comission, 2018; WHO, 2018). To support the results of the present study, $PM_{2.5}$, PM_{10} and NO_2 exceedances found for the S-SCN scenario were compared with those calculated with data from the monitoring stations of the EU Member States (EEA, 2020). For $PM_{2.5}$, the exceedances to the WHO guideline found with the modelled data represented more than 60% of the exceedances calculated with the data from the stations. Regarding PM_{10}, a small agreement was found, with only 11% of the exceedances found for the modelled data. However, for NO_2 all the exceedances were estimated with the modelled data. According to these results, the model seems to predict with good reliability the exceedances of $PM_{2.5}$ and NO_2. For PM_{10} the results need to be used with caution."

Minor comments

While the manuscript is overall well written, it would benefit from proof reading by a native speaker.

Answer: Suggestion attended. The manuscript will be review by a native speaker.

page 4, line 120: Are Sahara dust emissions and NOx from lightning really taken from the NCAR fire inventory?

Answer: No. Only the forest fire emissions were taken from the NCAR fire inventory. To improve the comprehension of the sentence we decided to delete the part of "from the Fire INventory from NCAR version 1.5" and keep the reference of the inventory and add the Simpson et al. (2012) reference where more information about these emissions can be found.

page 7, line 204 and figures: I think it is stated nowhere that when you talk about concentrations, that always means at the surface (I assume)

Answer: Thank you for your comment. The concentrations are surface concentrations. Modifications will be introduced in the line referred and the figure legend.

page 8, line 251 and following: I'm a bit confused by this discussion of the origins of the seasonality. It sounds as if it is not really clear what the origin is, but don't you have all the information on the magnitude of emissions from STEAM so that you can give clear answers on what drives the seasonality?

Answer: Thank you for your comment. This analysis was related to the concentrations. Although we had all the information on the magnitude of emissions from STEAM, there are other factors that can influence the concentrations over the seasons. To understand if statistically significant differences in concentrations between the various seasons exist, the non-parametric test Kruskal-Wallis test was used to compare multiple samples (the four seasons) and the non-parametric Wilcoxon signed-rank test was used to compare related samples (two by two). Moreover, as our previous analyses claim it was during spring and summer that were registered the highest emission amounts. Information about the statistical analyses above referred will be added in the Methods section as follows: "The non-parametric test Kruskal-Wallis for multiple samples (the four seasons) and the non-parametric Wilcoxon signed-rank test for two by two samples analyses performed at the 95% confidence interval level were used to detect statistically significant variations for all pollutants in the seasonal concentration data." Moreover, we will change the Results section as follows: "Regarding the seasonal concentration data, statistically significant variations were found for all pollutants across all seasons (p-values < 0.05). In fact, according to the model results, the higher contributions of shipping emissions to the concentrations levels were registered during spring and summer periods (warm season). This pattern seems to be related to the increase in ship traffic during summer due to better meteorological conditions that allow better navigation conditions, which increases the traffic and subsequently the emissions and atmospheric pollution. Moreover, during summer months,

the number of passenger ships tends to increase (due to recreational travel), especially in the Mediterranean Sea, which led to an increase of shipping emissions and their contributions to the pollutant's concentration levels."

page 10, line 299: The discussion on uncertainties and limitations is very general indeed and mainly lists the obvious. I think that the comparison to real data will make this section also more relevant.

Answer: Thank you for your comment. Comparison with real data will be added according to a previous comment.

Figure 1: I'm not sure that it makes really sense to show all these figures here – they all look the same with the colour scale chosen and I do not see what I can learn from 8 figures which I cannot already see in the first.

Answer: Thank you for your comment. In a previous version of the manuscript, we had the same scale for all pollutants which allowed to see the differences in the quantities of each pollutant. We will change it back to that version.