We would like to thank Michael Diamond and the two Anonymous Referees for their insightful reviews. Their corrections and suggestions improved the clarity of the manuscript significantly. Following are our point-by-point responses with the Referee comments in blue font color and italicized.

Reply to Anonymous Referee #3

The manuscript documents a SEVIRI based analysis of the potential effects of ship emissions on low-cloud properties. The uniqueness of this analysis is that the authors make use of geostationary retrievals. However, it is hard to discern what is new, especially from the abstract, as the findings are comparable to those described in Diamond et al. (2020). While some differences between Benas et al and other studies are attributed to the use of Aqua MODIS and Terra MODIS and plausible calibration issues (which is far from being certain), the main findings of Benas et al. are a corroboration of Diamond et al and other studies. The novel result of this study is the analysis of the diurnal cycle, but the ship track signal is weak, except for cloud fraction. In addition, the trend analysis for the entire SEVIRI record is not conclusive, and the following sentence in the abstract is not supported by the evidence "The long-term analysis reveals a weakening of the shipping corridor effect on Nd and re presumably following the International Maritime Organization's 2020 stricter regulations on sulfur emissions…" I appreciate the effort but, again, the authors need to more clearly state what is new of their study.

Reply: We have edited part of the abstract and main text, to highlight novel aspects of this study. We agree that a large part of our findings corroborate Diamond et al and findings from other studies, but we also claim that the novel aspects in our study are not limited to the diurnal cycle analysis; the full-year seasonal analysis and the time series analysis, both based on monthly resolved data, are also novel in their temporal coverage and resolution. Regarding the diurnal cycle of the corridor effect: we agree that the magnitude of the diurnal cycle is not large for N_d, r_e and W, while ambiguities appear for cloud fraction. This is discussed in the text, and shown in supplementary Fig. S4 (S5 in the revised version). However, we would argue that finding a weak diurnal variation is also a result worth showing and discussing.

Regarding the long-term analysis, Figs. 9b and 9d clearly show a previously unseen weakening of the corridor effect on N_d and r_e after 2020, so we don't see why the relevant sentence in the abstract is not supported. In fact, we attempted to carefully phrase this sentence in order not to conclusively attribute this change to the IMO regulations.

The novel aspects of this study can be summarized as follows:

- use of cloud property retrievals from geostationary observations for the first time in a study of ship emission effects on clouds (and the ensuing diurnal cycle analysis)
- full seasonal cycle analysis
- high temporal resolution in the time series analysis.

As mentioned above, we try to emphasize more clearly these new aspects in various parts of the revised manuscript. Please note, however, that we also consider corroboration of previous findings, based on a different sensor and methodology, as a valuable outcome of our analysis.

Other comments

Difference between SEVIRI and MODIS: While I agree with the authors about issues with some channels for Terra MODIS, I am not aware of any quality degradation of the retrievals for Terra

MODIS relative to Aqua (line 270-271 should be rephrased because the results in Benas et al. do not indicate significant uncertainties between the 2 MODIS instruments). Moreover, sensor issues can also be invoked for SEVIRI, as more than 1 SEVIRI instrument provided the data for this study. While I would agree with the authors that SEVIRI is a stable sensor, I am more interested in learning about pixel resolution differences between MODIS and SEVIRI and how they would affect the findings. More specifically, what would be the impact of the 3-km (nadir) pixel resolution of SEVIRI versus MODIS (1 km). This could be the single most important difference between this study and Yuan et al./Diamond et al. A coarser resolution would certainly impact the cloud mask identification, and cloud retrievals (I would expect larger droplet effective radii and smaller optical depth as the pixel resolution is degraded). Is the 3-km pixel resolution sufficient for detecting ship tracks? Possibly, the pixel degradation with viewing zenith angle is minimized for the study region (which is good); however, there is no quantitative description of a solar zenith angle threshold.

Reply: We agree with the Referee that the statement in lines 270-272 is not a result of our analysis; it was meant to highlight the non-conclusive results of the MODIS-based studies mentioned in lines 266 and 270. However, other factors may also play a role in the reported differences between these studies, so the statement was removed, and the sentence was rephrased accordingly.

A coarser resolution will indeed lead to decreased τ and increased r_e values; this is the result of an increased probability of pixels containing a mix of cloudy and clear sky, which will lead to lower reflectance values in both the visible and SWIR channels. This can lead to biases with corresponding MODIS results, due to the higher spatial resolution of the latter. It is not expected, however, to affect the corridor effect significantly: the effect is estimated based on retrievals from the corridor and surrounding areas, which are similarly affected (biased) in SEVIRI compared to MODIS.

The effect of the SEVIRI coarser resolution (compared to MODIS) on the results cannot be completely disentangled from other differences, based on the current setups of this study and the one by Diamond et al. (2020). The definition of the shipping corridor and the methodology to estimate its effect on clouds are probably the two major differences of these other differences. While we consider the comparisons with MODIS a valuable part of our analysis, and our study was largely inspired by the MODIS-based Diamond et al. (2020) study, a full analysis of the mentioned differences would be beyond scope. However, in Sect. 3.1 we have added a relevant discussion on possible differences due to the different spatial resolution.

Additionally, we attempt here a comparison with part of the results presented in Table 1 of Diamond et al. (2020), by adjusting our temporal coverage and months examined to match those of Diamond et al. (2020) as closely as possible (2004-2015 and September-October-November, respectively). As mentioned above, we only consider the SEVIRI time slots closest to the Terra and Aqua overpasses. The mean corridor value of τ appears consistently lower in CLAAS-3 compared to MODIS/Terra and Aqua values, as would be expected due to a coarser resolution. This is not the case with r_e, where results are similar for Terra and lower in CLAAS-3 compared to Aqua. Differences increase in the cases of W and especially N_d, for which additional assumptions/filters are used in the MODIS retrievals compared to CLAAS-3 (Bennartz and Rausch, 2017). Overall, and considering the factors that can lead to deviations, the results of the two studies, including corridor effect values, appear consistent. Note that corridor effect values of τ, although reported here, should probably not be considered meaningful, as explained in our response to the next comment.

		MODIS/Terra (D20)	SEVIRI CLAAS-3 at Terra overpass	MODIS/Aqua (D20)	SEVIRI CLAAS-3 at Aqua overpass
r_e [μm]	Mean Ship value	10.83	10.98	11.41	10.26
	Absolute Ship – NoShip difference	-0.28	-0.31	-0.29	-0.27

	Relative Ship - NoShip difference (%)	-2.61	-2.86	-2.52	-2.59
τ [-]	Mean Ship value	11.07	9.93	8.73	7.13
	Absolute Ship - NoShip difference	0.24	0.23	0.05	0.09
	Relative Ship – NoShip difference (%)	2.13	2.36	0.58	1.23
W [g m^{-2}]	Mean Ship value	85.23	77.19	66.08	51.70
	Absolute Ship – NoShip difference	-0.49	-0.56	-1.32	-0.80
	Relative Ship - NoShip difference (%)	-0.57	-0.73	-2.00	-1.55
N_d [cm^{-3}]	Mean Ship value	-	125.24	93.25	131.78
	Absolute Ship - NoShip difference	-	9.08	4.87	8.40
	Relative Ship - NoShip difference (%)	-	7.25	5.22	6.38

The question on the sufficiency of the SEVIRI 3 km resolution for detecting (individual) ship tracks is an interesting one, which we plan to investigate in the future. The answer, however, would not affect the results of the present study. Compared to individual ship tracks, the shipping corridor studied here is characterized by a continuous provision of emitted aerosols by many ships crossing the region, while their movement in both directions and spread in space result in a wider affected area. These conditions lead to effects on clouds that are discernible from SEVIRI.

As the Referee mentions, this region is favorable in terms of viewing conditions, with low viewing zenith angles and minimum expected pixel degradation. Regarding the solar zenith angles, the CLAAS-3 algorithm retrieves τ and r_e for values of $\vartheta_0 < 75°$, which is used as threshold to define day light conditions. This is included in Sect. 2.1 of the revised manuscript.

Statistical analysis: Considering the tiny changes in microphysical properties, a robust statistical method for testing the hypothesis is essential and should be highlighted throughout the article. I am also somewhat concerned about the methodology to construct Fig 3c-f. Particularly for cloud fraction (f_c, Fig 3f), I don't understand why the pattern is undulating. This makes me speculate that the methodology is not ideal for identifying the effect of ship tracks because cloud spatial variability is likely dominating the signal. The same comment applies to optical depth (tau) in the supplement; the tau pattern does not make sense. If tau were unperturbed by ship emission, deltatau as a function of the corridor distance should be a flat curve, right?

Reply: Throughout the manuscript we report propagated uncertainties along with all variables. These uncertainties allow evaluating whether differences between variables – inside/outside the corridor or before/after 2020 – are significant.

We do not claim that the methodology we use for identifying the effect of the shipping corridor on cloud properties is ideal. It assumes that the corridor manifests as a deviation from an otherwise smooth background, and this is not always the case. This is stated in Sect. 2.3 of the revised manuscript. When this assumption is not valid, results are not meaningful. This is clearly true for τ, but also for parts of the seasonal and diurnal analyses. For this reason, we always examine individual profiles of the estimated corridor effects (shown in Figs. S4, S5 and S6 in the revised supplement),

when reporting average values. Cloud spatial variability may indeed dominate the corridor signal in these cases. In most cases, however, this approach provides meaningful results. For example, the corridor signal is prominent in N_d, r_e and W. Based on another review comment, we estimate the no-ship scenario uncertainties by repeating the calculations while varying the distance of the assumed unaffected ranges from the corridor center. This revised approach shows that our method is robust for calculating the corridor effect on these three variables, while it highlights its limitations in the case of f_c (see e.g. Figs. S2 and S6 in the revised supplement). As mentioned above, in the case of τ the underlying smoothness assumption does not hold, so the results are not meaningful. $\Delta\tau$ is not shown, so we are not sure what the Reviewer refers to. Based on our methodology, if τ were unperturbed, and its across-corridor distribution were (perfectly) smooth, $\Delta\tau$ would be zero.

Last paragraph of page 7 should be discussed in section 2.3

Reply: Section 2.3 is meant to discuss the method for estimating the shipping corridor effect, not the results. Thus, our references to Fig. 2 in this section were confusing. In the revised manuscript we have removed all references to results from Sect. 2.3.

Equation 1: I am not sure why the combination of 2 dissimilar variables (standard deviation and retrieval uncertainty) can be used to produce the uncertainty of averaged data. At least, it does not seem to be mathematically correct.

Reply: This equation is a short version of Eq. (5) in Stengel et al. (2017). It considers both the retrieval uncertainty, which is an output of the CPP algorithm based on input data uncertainties, and natural variability. According to Eq. (5) in Stengel et al. (2017):

$$\sigma^2_{\langle x\rangle} = \frac{1}{N}\sigma^2_{natural} + c\langle\sigma_i\rangle^2 + (1-c)\frac{1}{N}\langle\sigma_i^2\rangle$$

where $\sigma_{natural}$ is the natural variability, $\langle\sigma_i\rangle = \frac{1}{N}\sum_{i=1}^{N}(\sigma_i)$ and $\langle\sigma_i^2\rangle = \frac{1}{N}\sum_{i=1}^{N}(\sigma_i^2)$. Given that $\sigma^2_{natural} = \sigma^2_{std} - (1-c)\langle\sigma_i^2\rangle$ (Eq. (4) in Stengel et al. (2017)), we get:

$$\sigma^2_{\langle x\rangle} = \frac{1}{N}\sigma^2_{std} - \frac{1}{N}(1-c)\langle\sigma_i^2\rangle + c\langle\sigma_i\rangle^2 + (1-c)\frac{1}{N}\langle\sigma_i^2\rangle = \frac{1}{N}\sigma^2_{std} + c\langle\sigma_i\rangle^2$$

Line 171 You mean "determine" instead of "simulate"

Reply: Yes, replaced.

Page 9 "This result suggests that two opposite tendencies, namely an increase in τ due to the Twomey effect and a decrease due to the decreasing W cancel each other out." This is an overinterpretation of the satellite data because from a remote sensing perspective droplet effective radius is the key variable, not LWP (LWP is indirectly estimated from effective radius and optical depth).

Reply: Indeed, W is estimated indirectly from r_e and τ. We don't see, however, why this statement cannot be possible, due to this retrieval feature. Our purpose was to present it as a possibility, not as an outcome of our analysis. We clarify this point in the revised manuscript.

Line 238: "it is difficult to draw any conclusion on the shipping corridor effect on fc, day." Is there a sentence missing here?

Reply: No, "$f_{c,\ day}$" denotes the daytime cloud fraction. The confusion was likely caused by an unfortunate line break between 'c' and 'day'.

Introduction: Lines 22-43. Future readers already know this. For brevity, the authors should

Reply: This sentence is incomplete. Does the Referee mean to say that this part of the introduction is redundant? We agree that this is common knowledge in our field, but we also think it is good to start such a study with a broader overview, which serves the purpose of placing the specific questions addressed here in a wider context.