Evaluating simulations of ship tracks in a high-resolution model

Anna Tippett, Paul R. Field, and Edward Gryspeerdt

Author Response

We thank both the reviewers for their comments, and for their time taken to consider this paper. We address the "Major comments" and "Minor comments" for each referee in turn. We provide both our response and details of the changes made in the text, with new text coloured in blue. Line numbers / figure numbers refer to the those in the **diff**.

Referee #1

Major Comments

	Comment	Response
1	The basic premise is that the only difference between the Eastern Pacific ship tracks (A-C) and the Central Pacific ship tracks (D & E) is the intensity of the precipitation, hence the difference in the longevity of the ship tracks is strictly a function of the precipitation. Both the satellite observations and simulations suggest that these two regions have vastly different types of MABL clouds with classic closed MCC in the Eastern Pacific and broken cumulus in the Central Pacific, potentially transitioning to trade cumulus or potentially open MCC. In terms of MABL clouds, you are comparing apples to oranges. There are differences in cloud fraction, MBL/cloud top height, entrainment from the free	This is a valuable comment and draws to light a potential oversimplification of the current manuscript. The referee is correct in that the two groupings of ship tracks (A-C and D&E) occur in different types of marine clouds, and simply attributing all differences in behaviour to differences in precipitation may be misleading. To account for this, we had modified our terminology when separating these two groupings accordingly. Ship tracks A-C are no longer labelled as "non-precipitating", but instead as ones occurring in the "Eastern Pacific" / "closed MCC". Similarly, ship tracks D&E are now referred to as those occurring in "Central Pacific"/ "broken cumulus", instead of "precipitating". In addition to this, discussion of the different processes occurring inside these different cloud types has been added to address potential mechanisms for the longevity of the ship track, other than the precipitation suppression. Despite there being other potential mechanisms for the different behaviours in these groupings of ship tracks that arise due to the different nature of the clouds they exist in, our results which demonstrate the significant precipitation suppression and large enhancements in LWP do suggest that precipitation is the likely source of discrepancies. However, we acknowledge that without more in-depth analysis of the processes in these ship tracks, such as entrainment at

troposphere, large scale subsidence, wind speed, estimated inversion strength, and sensible and latent heat fluxes off the ocean. Do we know if the Central Pacific MABL is decoupled or not? How do we know that the rate of collision & coalescence removing aerosols within CASIM isn't sensitive enough to turbulence in the MABL?

To simply attribute all the differences in the persistence of the ship track to precipitation alone, is overly simplistic.

While I find this a major concern, it can readily be addressed with a more comprehensive, robust discussion. Be upfront with the limitations of the analysis.

cloud top, or rate of collision and coalescence (which are not available in the current model output), we cannot make a definitive conclusion.

As such, our discussion has been updated to reflect these limitations in the analysis.

Changes made

- Fig 9.
    - Caption: "Percentage enhancements from the "unpolluted" references background state. (a)-(d) show observations of ship tracks ~~in non-precipitating (ships A, B and C)~~ in the Eastern Pacific region of the domain with closed MCC (ships A-C) and ~~precipitating~~ in the Central Pacific region with broken cumulus (ships D& E) conditions for Nd (left) and LWP (right)..."
    - Change column labels to "Broken cumulus" and "Closed MCC"
- Line 409:
    - Delete "In Fig. 9, we split our ship tracks into those that occur in different background conditions, as a means to investigate whether our model is able to capture varying responses, and to assess the impact on model performance. Since there is precipitation within the scene, we divide our tracks into precipitating / non-precipitating cases. We use direction of travel as a simple proxy for "initial condition'', since each grouping of ships is always within approximately 100km of each other, and therefore experiences largely similar meteorological conditions. Ships A-C follow routes from the coast of California, westwards into the cleaner ocean (see Fig. 4b). Ships D&E travel towards the coast and are more southerly than ships A-C. Ships A-C travel through primarily non-precipitating clouds, whereas ships D and E travel through drizzling clouds (at the surface; see Fig. S4). Dividing our enhancements into these two groupings, we see that the model is sensitive to the initial condition of the cloud before the ship passes through (Fig. 9)."
    - Replace with "In Fig. 9, we split our ship tracks into those that occur in different background conditions, as a means to investigate whether our model is able to capture varying responses, and to assess the impact on model performance. We use direction of travel as a simple proxy for "initial condition" of the cloud before the ship passes, since each grouping of ships is always within approximately 100 km of each other, and therefore experiences largely similar meteorological conditions. Ships A-C

follow routes from the coast of California, westwards through the Eastern Pacific (see Fig. 4b). Ships D and E travel towards the coast and are more southerly than ships A-C, through in the Central Pacific. The Eastern Pacific ships travel through classic closed mesoscale cellular convective clouds ("closed MCC"), where there are primarily non-precipitating clouds. In the Central Pacific, there are "broken cumulus" that potentially transition into trade cumulus or open MCC, and we see higher occurrence of drizzle at the surface from these clouds (see Fig. S1). Dividing our enhancements into these two different types of marine boundary layer clouds (Fig. 9), we see that the model is sensitive to the initial condition of the cloud before the ship passes through."

- Paragraph beginning Line 421: replace "non-precipitating" with "closed MCC"
- Paragraph beginning Line 434:
  - Delete: "Considering the ships that travel through precipitating clouds (Fig. 9 2nd and 4th column for Nd and LWP enhancements respectively), we see significant disagreement between the observations and the model. This suggests that the large disagreement in Fig. 8 is largely due to these ships in precipitating conditions. The Nd enhancement is too large, and shows no sign of diminishing even after 15 hours after the aerosol perturbation. This suggests that the model is not only overly sensitive to aerosol in precipitating conditions, but is not effective enough at removing aerosol from the cloud either. Additionally, the LWP response is a different sign between model and observations. This suggests that the model is far too keen to suppress precipitation, leading to large increases in LWP that are not observed in the satellite imagery."
  - Replace with: "Considering the ships that travel through broken cumulus clouds (Fig. 9 2nd and 4th column for Nd and LWP enhancements respectively), we see significant disagreement between the observations and the model. This suggests that the large disagreement in Fig. 8 is largely due to the ship tracks in these conditions where precipitation is more prevalent. The Nd enhancement is too large, and shows no sign of diminishing even after 15 hours after the aerosol perturbation. This suggests that the model is not only overly sensitive to aerosol in broken cumulus clouds, but is not effective enough at removing aerosol from the cloud either. Additionally, the LWP response is a different sign between model and observations. This suggests that the

model is far too keen to suppress precipitation in these conditions, leading to large increases in LWP that are not observed in the satellite imagery."

- Fig 10: caption replace "non-precipitating" and "precipitating" with "closed MCC" and "broken cumulus"
- Line 447: "Here, we see that for tracks D and E ~~(which occur in more precipitating environments~~) in the broken cumulus with larger precipitation rates, the precipitation is essentially completely shut off which is likely a far too strong response"
- Line 456: replace "precipitating" with "broken cumulus"
- Line 460:
  - Delete "In our simulations, we do see that there is some sensitivity to the initial conditions since we obtain different responses in precipitating / non-precipitating cases. We can conclude that the model is more suitable for simulations of high-concentration aerosol perturbations in non-precipitating conditions, at least for up to 5 hours. However in the case of precipitating clouds, current model parameterisation formulations are not suitable to accurately simulate these scenarios, and the lifetime of the response is far too long lived."
  - Replace with "In our simulations, we do see that there is some sensitivity to the initial conditions since we obtain different responses in different cloud structures (closed MCC vs. broken cumulus). We can conclude that the model is more suitable for simulations of high-concentration aerosol perturbations in closed MCC conditions where there is little precipitation, at least for up to 5 hours. However, in the case of broken cumulus clouds with higher occurrence of precipitation, current model parameterisation formulations are not suitable to accurately simulate these scenarios, and the lifetime of the response is far too long lived."
- Line 473: "...and find that in closed MCC (non-precipitating conditions) the..."
- Line 562: "Model-observation discrepancies become more pronounced in ~~precipitating environments~~ broken cumulus clouds, where there is greater precipitation."
- Addition of Section 4.1.1 to discuss the limitations of the "Definition of 'precipitating' and 'non-precipitating' clouds"

Comment	Response

2 "Observations of precipitation in the domain are confirmed with overpassed (sic) from the Cloud Profiling Radar (CPR) onboard CloudSat (Stephens et al., 2008)) during the simulation."

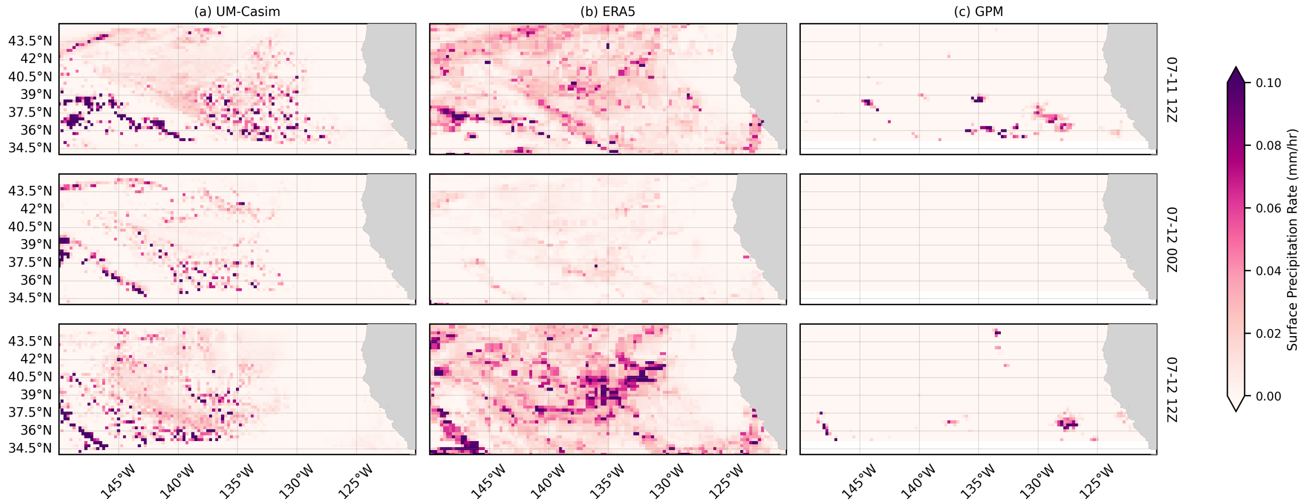
Are you really trying to say that since CloudSat (which CloudSat product, the 2C-column-precip or the 2C-rain-profile?) recorded precipitation somewhere along it's overpass, that the simulated precipitation is reasonable?

In all sincerity, I was inclined to recommend rejecting the manuscript on that single sentence alone.

This is a major weakness in research, the simulated precipitation has not been evaluated in any way. Given all the challenges we have in estimating precipitation in shallow convection over the remote ocean and the significant differences commonly found between various precipitation products, both from reanalyses and satellite-based

Thank you to the referee for this comment. To address this comment (and some comments from the second referee), we have added a section to our results (new Section 3.2) which contains an evaluation of the simulated precipitation (and aerosol).

New Fig. 5:

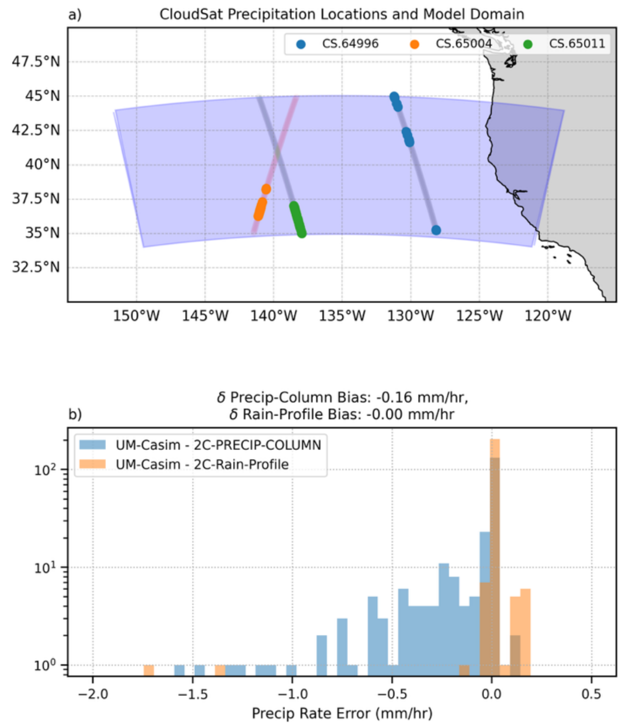


We evaluate our domain-wide surface precipitation in the simulation (UM-CASIM) against ERA5 and GPM-IMERG at 12 hourly intervals in Fig. 5 The UM-CASIM simulation is regridded to the same resolution as ERA5 for ease of comparison. ERA5 roughly agrees with the magnitude and location of precipitation in the simulated domain, whereas GPM is unable to capture the very low rain rates (<0.1mm/hr). This is to be expected for such light drizzle and GPM (Skofronick-Jackson et al., 2017).

products, it needs to be shown that simulated precipitation has some measure of skill. The simulated precipitation, after all, underpins the analysis.

I strongly recommend that a section on the evaluation of the simulated precipitation be added to this manuscript. Ideally, the evaluation would be made against both ERA5 and GPM-IMERG for the full domain, commenting on the level of skill for both the classic stratocumulus over the Eastern Pacific and the broken cumulus of the Central Pacific. In addition, an evaluation should be made along the CloudSat overpass, against both the 2C-RP and 2C-CP products.

New Fig. 6 (caption: “Evaluation of surface against CloudSat CPR products. a) Location of CloudSat overpasses during the 48 hour simulation, with bold points showing locations where surface precipitation from the 2C-Precip-Column product are greater than zero. Associated times of overpass are roughly as follows: CS.64996 at 2200Z 11th July, CS.65004 at 1200Z on 12th July, and CS.65011 at 2300Z on 12th July”)



As suggested by the referee, we also evaluate our simulated precipitation against both the 2C-Precip-Column and 2C-Rain-Profile CloudSat products. In Fig. 6a we show the CloudSat overpasses of our domain, within our simulation runtime. There are only 3 overpasses that intersect our domain within 48 hours, and do not have significant coverage of the area in the Central Pacific (where the simulated precipitation is occurring). Therefore, any rigorous evaluation of the domain-wide precipitation against CloudSat products is not possible, however we show in Fig. 6b the bias of UM-CASIM against these available intersections. Our model seems to underestimate precipitation when compared to 2C-Precip-Column and is roughly in agreement with 2C-Rain-Profile.

As the referee mentions, measuring light precipitation over the remote ocean is very difficult, with different products providing significantly different values. This evaluation demonstrates that there is some uncertainty in the simulated precipitation within the domain, and further discussion has been added to acknowledge this source of uncertainty. We have also added further references to CASIM papers which contain some evaluation of the performance of CASIM in simulating precipitation more generally (Field et al., 2023; Bush et al., 2025).

Additionally, for the increased LWP effect to occur, precipitation only needs to be suppressed at cloud base. We cannot easily obtain observations of precipitation at cloud base, making this effect very difficult to evaluate against observations. Further discussion of this point is added to the manuscript.

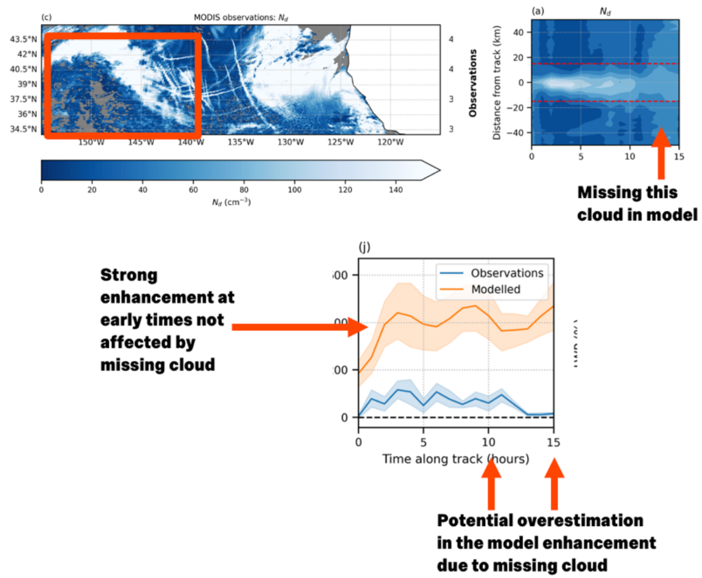
References:
Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini, R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., and Wilheit, T.: The Global Precipitation Measurement (GPM) Mission for Science and Society, Bulletin of the American Meteorological Society, 98, 281679–1695, https://doi.org/10.1175/BAMS-D-1500306.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2017.

Changes made

- New Section 3.2 – Evaluation of the control,
  - Evaluation of precipitation against GPM, ERA5 and CloudSat.
  - Performance of CASIM discussed more generally
  - Section 2.4 moved into this section.
  - New Figures 5 and 6.

- Line 581: “As discussed in Section 3.2, there is some uncertainty in the realism of the simulated precipitation in the domain, since it is difficult to obtain conclusive precipitation measurements across different satellite and reanalysis products. In order for the enhancement of LWP in a ship track to occur, precipitation must only be suppressed at cloud base, and therefore it is also possible that no impact in surface precipitation could occur. This highlights the difficulty in evaluating not only precipitation in models, but specifically the precipitation suppression effect.”

	Comment	**Response**
3	There appears to be a difference between the simulations and the satellite imagery in the cloud structure/morphology in the Central Pacific. This difference is evident in Figure 5. Do ship tracks D & E pass through this cloud field? If so, please comment on how this may affect the analysis.	Thank you for this valuable comment. There is a difference in cloud structure between our simulations and MODIS observations, with a large area of high droplet number concentration missing from our model in the Central Pacific. This is a result of our constant initialised aerosol fields which we use to simplify our analysis; therefore, we are missing a source of background aerosol here. Ship tracks D&E do pass through this cloud structure, but it is only the 10-20 hours along the ship track that will be impacted by this. As a result, we only consider the first 15 hours of time evolution along the track. Clarification of this point is added to the manuscript.



In Fig. 8a, this cloud field outside of the ship track can be seen between 10-15 hours, whereas it is not visible in the simulation (Fig. 8e). As a result, it is possible that the enhancement from the background/control state in the simulation is overestimated for the simulated tracks D&E over 10 hours (Fig. 9j). However, this is clearly not the source for the overestimation in Nd enhancement, since this large Nd enhancement is seen before 10 hours along the track. Acknowledgement and discussion of this potential source of uncertainty is added to the manuscript.

Changes made

- Line 375: "Due to the missing high Nd region in the Central Pacific part of the domain, ship tracks are only considered up to 15 hours along their length, since times longer than this are within this missing cloud and therefore comparison between model and observations are not suitable."
- Line 442: "There is some uncertainty that may be introduced due to the missing cloud structure in the Central Pacific region in our simulation, causing our background state to look cleaner than it should be. This missing cloud structure, however, will only influence the ship track over 10 hours along the track, and therefore is not the source of the large overestimation in Nd enhancement seen before 10 hours along the track in Fig. 9j."

Minor Comments

	Comment	Response

1,2	Line 32: "large-scale" is ambiguous to me in this sentence.
	
	Line 32: The whole sentence confuses me. Why are you isolating large-scale turbulence here? Are you saying the large-scale turbulence within coarse resolution CGMs are used in activation of CCN in boundary layer clouds?
	
	Line 34: "certainty"? We will never have certainty. But we can have some measure of confidence.
		Thank you for these comments.
		
		We agree that these sentences are confusing and disjointed, and the isolation of "large-scale turbulence" is unnecessary for the point that we are attempting to make. In the revised manuscript we restructure the paragraph as detailed below for better conciseness and clarity.
		
		Changes made
		
		• Line 33:
		    ○ Deleted: While large-scale turbulence is resolved in these models, many of the same parameterisations are used in GCMs (e.g., for activation or precipitation processes). Therefore, it is critical that we have certainty in our regional representations of MCB.
		    ○ Replaced by: "This is a particularly challenging task for climate models. While recent models are able to adequately simulate cloud patterns of cloudiness (Tselioudis et al., 2021), their simulation of aerosol-cloud interactions is more varied (Gryspeerdt et al., 2020). However, these evaluations are typically done on large-scale temporal or spatial averages - assessments of MCB efficacy require accurate simulations of the model response to individual clouds. This is not well evaluated by current techniques. In addition, as many of the same parameterisations for cloud microphysics are used across both coarse and finer resolution models (e.g. activation and precipitation processes), greater confidence in our regional representation of MCB should inform our global modelling capability, avoiding the equifinality problem in larger-scale studies (Mülmenstädt and Feingold, 2018)."

	Comment	Response
4	Line 47: This sentence is out of place to me. The changes to the LWP happen after the	To improve the flow of this paragraph this sentence is removed, since the following sentences describe the same thing. **Changes made**

suppression of precipitation (Albrecht effect), which is in the next sentence.

- Line 53: After the initial droplet number (Nd) perturbation, clouds respond to aerosol perturbations over longer timescales through effects known as "cloud adjustments". ~~Changes in liquid water path (LWP) - the amount of liquid in a cloud column - are particularly important because they influence not only cloud radiative properties but also precipitation processes~~. A precipitation suppression effect (Albrecht, 1989; Rosenfeld, 2000), where smaller cloud droplets take longer to form precipitation, increasing liquid water path (LWP), tends to result in additional cooling. Alternatively, smaller droplets can also enhance the mixing of dry air above cloud-top into the cloud (a process known as entrainment) and reduce LWP, resulting in a warming effect (Ackerman et al., 2004; Bretherton et al.,2007).

	Comment	**Response**
5	Line 72: "short timescales" is ambiguous	The ambiguity of "short" and "longer" timescales in this sentence is addressed by adding approximate timescale values, as well as citations for said values.

Changes made

- Line 78: "This helps us evaluate model process representation, such as the activation of cloud droplets (which occurs on the order of seconds to minutes; Arabas and Shima, 2016), or the autoconversion of cloud droplets into rain droplets (occurring on longer timescales, on the order of minutes to hours; Stephens and Haynes, 2007)."

	Comment	**Response**
6	Line 80: The simulated ship locations must be prescribed whether from an actual ship or not.	This is a useful comment, and a slight oversight in our description. Simulated ship locations are always prescribed in the model, however in this study we simulate the exact ship trajectories that are associated with each observed ship tracks. In previous work, tracks have been simulated on days that contained ship tracks, but not using the same ship trajectories as the actual ships. As seen in this work, the location and cloud structures that the ships travel through have a significant influence on the track evolution. To improve the clarity and accuracy of this point, we make the following changes:

Changes made

- Line 93: deleted "…prescribed and…"
  - "…yet the simulated ship locations are not from the actual ships that caused the observed ship tracks, which limits the ability to assess the accuracy of MCB simulations."

	Comment	Response
7	Line 115: It would be nice to add text stating that the domain covers the Eastern and Central Pacific, off the coast of California.	Thank you for this suggestion. We implement it as described below: **Changes made** • Line 135: sentence added "Our domain covers the Eastern and Central Pacific, off the coast of California (Fig. 1.)"

	Comment	Response
8	Line 120: Do you need four references here?	Thank you for this suggestion, we agree that four references is somewhat overkill, therefore we keep only two references to double-moment cloud microphysics schemes **Changes made** • Line 143: "…Double-moment microphysics schemes prognose both the mass and number concentration of each hydrometeor species (e.g. Ferrier, 1994; Seifert and Beheng, 2006)…"

	Comment	Response
9	Line 175: Is 'constrain' the right word here? You aren't assimilating any of the satellite observations, are you?	This is a useful comment, and the referee is correct that we are not "constraining" our model. We improve our terminology, as detailed below, to correctly describe how we are evaluating our model (not constraining it). **Changes made** • Line 225: "To constrain our model simulations" replaced by "To evaluate our model simulations"

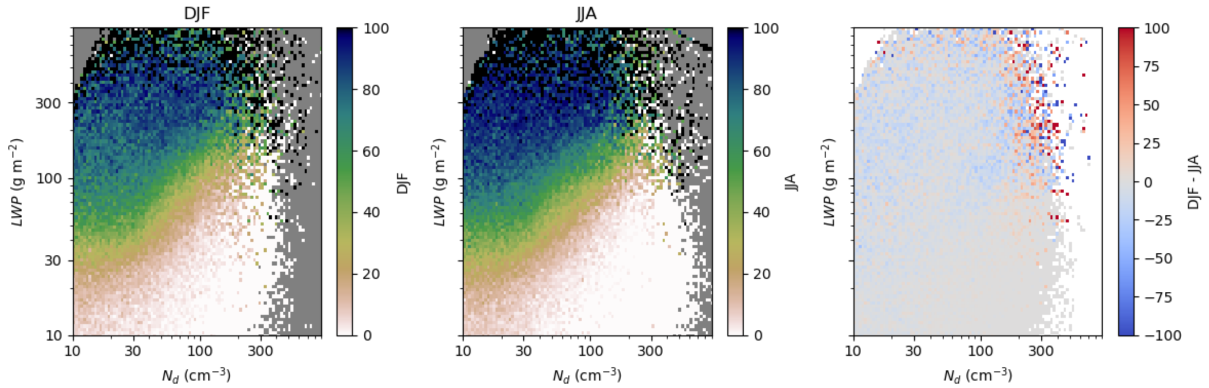
	Comment	Response
10	Line 182: Which CloudSat product?	We modify this paragraph as a result of our addition of the new "Evaluation of the control simulation" section to include our newly included data sources, and associated references. **Changes made** • Line 232: ○ Delete: "Observations of precipitation in the domain are confirmed with overpassed from the Cloud Profiling Radar (CPR) onboard CloudSat (Stephens et al.,2008) ○ Add: "Simulated precipitation in the domain is evaluated against ERA5 surface precipitation (Hersbach et al., 2020), as well as Global Precipitation

	Comment	Response
		Measurement (GPM-IMERG; Huffman et al., 2015) and overpasses from the Cloud Profiling Radar (CPR) onboard CloudSat (both the 2C-Precip-Column and 2C-Rain-Profile products; Stephens et al., 2008; Haynes et al., 2009; L'Ecuyer and Stephens, 2002)."
	Comment	**Response**
11	Line 183: I skimmed through Kato et al. (2010) but could find no mention of a precipitation product being produced, nor LWP or Nd. Is this reference correct? If so please add the detail necessary to from the referenced material to the products used in this paper.	The CCCM product collocates CERES, CloudSat, CALIPSO and MODIS observations. The probability of precipitation is found using the CloudSat CPR 2C-Precip-Column precip flag, whereas the Nd and LWP retrievals from MODIS are used, Kato et al., 2010 discusses the Cloud Top and Base merging strategy used in the CCCM combined product, whereas Kato et al., 2011 describes in more detail the collocation with MODIS (where LWP and Nd are obtained from). The correct citation is added to the revised manuscript.
		Changes made
		• Line 236: "...We investigate the probability of precipitation across Nd- LWP space using observations from the CCCM (CERES–CloudSat–CALIPSO–MODIS) combined product (Kato et al., 2010, 2011). We use the CloudSat CPR precipitation flag (from 2C-Precip-Column; Haynes et al., 2009) to define a probability of precipitation as counts of liquid precipitation or drizzle divided by all counts, per Nd-LWP bin (from MODIS Aqua).
	Comment	**Response**
12	Line 190: Ship plumes are advected in wind fields, not positions.	Yes, in reality the ship plumes are advected in the wind fields, however for the purpose of our ship track prediction methodology we use the ship locations to predict where the clouds above the ship (at the time of aerosol injection) are advected over time. In this sense, we do advect the ship positions to predict the ship track positions. We appreciate the slight nuance in this point, therefore we make the following modifications to address this with greater clarity.
		Changes made
		• Line 244:

	Comment	Response
		○ Delete "Ship positions (from AIS data) are advected in wind fields over time to obtain a ship track which consists of not only location, but the associated time since that location experienced the aerosol perturbation." ○ Replace with "Using ship positions (from AIS data) we infer the locations of clouds which experienced an aerosol perturbation (neglecting the time taken for aerosol to reach cloud base). These positions are advected in wind fields over time to predict ship tracks which consist of not only location, but the associated time since that location experienced the aerosol perturbation."
	Comment	**Response**
13	Line 254: Should you also define eL here?	We apologise for the oversight – eL has not been defined anywhere in the manuscript. Gryspeerdt et al., 2019b only fits a curve to the enhancement in Nd, not LWP therefore we do not define it here as this could be misleading. Instead, we add a sentence to define both eN and eL in Section 2.6 ("Quantifying enhancements").
		Changes made
		• Line 270: add "…percentage difference to obtain enhancements in Nd and LWP (ϵN and ϵL)."
	Comment	**Response**
14	Figure 1: This may work better if you present the control (left column) and the difference (right column.	Thank you for this comment. We have modified Fig. 2 (which used to be Fig. 1) to now show the control (left) and difference from the control for the ships simulation (right).
		Changes made
		• Fig 4 new right hand column and caption.
	Comment	**Response**
15	Figure 9b: Is this for the full four-year time period, as stated in line 186? If so, why is it meaningful to compare a four-year average, that contains winter and summer, against a simulation of a single day	Figure 11b (previously 9b) shows the composite of four years for CloudSat-MODIS collocated data, which is evidently a much longer time frame than our 48-hour simulation run time. The expectation is that the distribution of probability of precipitation within this space is a representation of the microphysical processes and shouldn't differ significantly seasonally or day-to-day. Assuming all processes have been observed, the probability of precipitation as a function of Nd and LWP should be the same in winter as for summer.

during July? Please address this issue.

Splitting up all DJF and JJA days across the four-year period, season has no significant impact on the distribution of the probability of precipitation within this space:



There is a small decrease in the probability of precipitation, perhaps, in DJF compared to JJA (possibly due to reduced sea surface temperatures), however the representation of PoP in the space is largely the same – the binary nature of the PoP in our model run is unlikely to be due to the timeframe of the observations we compare against. This figure is added to the supplementary and discussion added to the text.

Additionally, the use of four years of data is necessary to obtain significant counts for each bin. CloudSat has a small ground footprint of 1.5km, meaning that an overpass through our domain will contain fewer than 1000 pixels. Conversely, in our model domain we have 1750x750 pixels of 1.5km resolution, corresponding to over a million pixels per time step.

The distinct binary representation that we obtain in the model is not likely due to the different time periods covered, as even seasonally we do not see major differences in this space (and we can't compare day-to-day due to lack of coverage). We address these points in Section 3.3.3.

Changes made

- Line 450: "We demonstrate this in Fig. 11, where the probability of precipitation (PoP) is plotted in Nd- LWP space for (a) our 48 hour model simulation and (b) four years of collocated CloudSat-MODIS observations (following a similar methodology to Gryspeerdt et al., 2022). The distribution of PoP within this Nd-LWP space is a representation of

	Comment	Response
		microphysical processes, and shouldn't differ significantly seasonally or day-to-day, therefore a comparison between a two-day simulation and four years of observational data should represent the same distribution (see Fig. S2)."
	Comment	**Response**
16	Final comment: This research focusses the impact of aerosol loading (the ship track) on the clouds and precipitation. It doesn't explore the processes where clouds and precipitation impact the aerosol loading, i.e., collision-coalescence and wet-deposition. It may strengthen your findings to consider this in the discussion.	Thank you for this comment. In our coupled UKCA-CASIM configuration, aerosol can be removed (irreversibly) from the cloud via precipitation or wet-deposition – autoconversion and accretion rates are summed and fed back into the aerosol scheme to determine the removal or aerosol by rain (wet-scavenging). Evaporation of cloud droplets does not return aerosol into the environment. This may explain why, once precipitation is slightly suppressed, there is no removal of aerosol from the cloud and therefore the ship track lifetime is too long. We add to our discussion to address the cloud/precipitation impacts on aerosol loading. **Changes made** • Within New Section "4.1 Limitations", we discuss in subsection "4.1.2 Aerosol Scheme" the limitations introduced from our aerosol scheme, and the impact on our findings. ○ Line 525: "Additionally, our coupling between CASIM and GLOMAP-mode aerosols (as described in Gordon et al., 2020) means that the removal of aerosol from the cloud is only possible through wet-scavenging. Autoconversion and accretion rates determine the rate of removal of aerosol by precipitation, however cloud microphysical processing of the aerosol is not implemented, meaning that there is no reduction of Na by collision-coalescence, nor an increase in Na through droplet evaporation. This severely limits the mechanisms by which aerosol can be removed from the cloud and may explain our long-lived Nd enhancements within our ship tracks. Once precipitation is even slightly suppressed, aerosol removal from the cloud is also significantly reduced, and therefore ship track lifetimes are too long. In Fig. 11 we see that very small increases in Nd are able to cause an almost complete suppression of precipitation, which in turn will feed back on the lifetime of the Nd response, since aerosol will continue to exist in the cloud. This indicates that cloud processing is important for accurately simulating ship tracks, as to avoid this precipitation suppression feedback."

Referee #2

Major Comments

	Comment	Response
1	Overall, I found the manuscript to be quite disjointed. For example, Section 2.3 is a Methods section, however it contains results, before going back into Methods in Section 2.4. Then in the Results section, Equation 2 is outlining a method. A lot of Section 4.1 could also be moved to Methods.	We thank the reviewer for this comment, and we appreciate how the previous structure of the manuscript was not the most readable. We have undertaken extensive restructuring of the revised manuscript as a result, in the hopes that the story is now easier to digest. The changes we have made are listed below. **Changes made** • Section 2.3 has been moved to new Results Section 3.2 ("Evaluation of the control") • Much of Section 4.1 has been removed from the Discussion, and a modified version has been absorbed into Section 2.6 ("Quantifying Enhancements"). • The remaining discussion from old Section 4.1 has now been absorbed into new Section 4.1.3 ("Definition of the "control" region" within Section 4.1 "Limitations"). • Equation 2 is removed from the Results section, and a new methods Section 2.7 "Expected enhancements" has been added for this method.

	Comment	Response
2	Title and use of 'high-resolution': The phrase 'high-resolution' might mean different things to different readers. Perhaps consider using a more specific phrase, like 'convection-permitting' or 'convection-resolving' through the manuscript.	Thank you for this comment. We appreciate that the phrase "high-resolution" has subjective meaning to different parts of the atmospheric physics community, and as such have made clarifications within the manuscript's introduction and discussion as listed below. To rename the entire paper as "Evaluating simulations of ship tracks in a convection-resolving model" may draw too much focus onto convection as a theme, which is not discussed in this paper. For the avoidance of doubt, we change the title of the paper to "Evaluating simulations of ship tracks in a km-scale model" as we believe this is the most accurate description. Furter modifications are made to the text to refer to the model as 'km-scale' also. **Changes made**

	Comment	Response
		• Title: "Evaluating simulations of ship tracks in a km-scale model" • Line 97: "We utilise a double-moment cloud microphysics scheme - Cloud AeroSol Interacting Microphysics (CASIM; Field et al., 2023) in the Met Office Unified Model (UM) in regional configuration (allowing for convection to be resolved), and compare directly to satellite observations." • Line 467: "In this study, we produce convection permitting simulations of ship tracks using a regional model…" • Line 643: "and aerosol forcing estimates from ~~high resolution~~ km-scale models"

	Comment	Response
3	The manuscript's language and sentence structure could be tightened up and polished. Doing so would make the manuscript considerably easier to read. I have tried to highlight various places in Other comments	We thank the reviewer for taking the time to highlight specific places within the manuscript where greater care is needed with language and sentence structure. Efforts have been made throughout the manuscript to improve on the readability of the text, and specific points from the reviewer are addressed in the "Minor comments" section below.
		Changes made
		Improvements have been made throughout, please see "Minor comments" section for exact line numbers and changes, or refer to the diff PDF.

	Comment	Response
4	The model description needs to be extended to give the reader basic and more accurate information about the model configuration. Specifically: 1. The CASIM description (currently two sentences) doesn't detail what cloud microphysics processes are represented. 2. The nomenclature used for UKCA and GLOMAP is incorrect. UKCA is the	Thank you to the reviewer for bringing this issue to light. We agree that more specific information on the CASIM and model set up is necessary, and we have expanded Section 2.2 as a result. This modified section now contains more details on the cloud microphysics represented within CASIM. As for the description of UKCA and GLOMAP, we apologise for any inaccuracies in the nomenclature. This paragraph in Section 2.2 has been modified with the correct (expanded) description of the coupling between CASIM and GLOMAP-mode. In terms of non-ship aerosol emissions, our simulations contain only the aerosol sources from GLOMAP-mode. We clarify this in the updated text.
		Changes made

framework used in the UM for handling atmospheric composition, whether aerosol or gas-phase chemistry. GLOMAP, specifically GLOMAP-mode, is the microphysical aerosol scheme available in the UM. 3. There is no information about what non-ship aerosol emissions are included	• Line 153: add "Each cloud species' particle size distribution (PSD) is described using a generalised gamma function with constant shape parameters. The autoconversion of cloud droplets into rain droplets and accretion of cloud water by rain are parametrised in CASIM following Khairoutdinov and Kogan (2000) .CASIM also contains other warm cloud processes, such as aggregation (rain collecting rain), and we refer the reader to Field et al. (2023) for details on these parametrised processes" • Line 158: ○ Delete: "Our configuration of CASIM allows Nd to be derived from aerosol (rather than prescribed), and we couple it to the United Kingdom Chemistry and Aerosols (UKCA) aerosol microphysics scheme. UKCA is a full prognostic two moment aerosol microphysics scheme, based on GLOMAP (Mann et al., 2010), and provides both aerosol mass and number information for the activation of cloud droplets." ○ Replace with: "We couple CASIM to the double moment modal Global Model of Aerosol Processes aerosol microphysics scheme (GLOMAP-mode; Mann et al., 2010), within United Kingdom Chemistry and Aerosols (UKCA) sub-model. Coupling to GLOMAP-mode in CASIM allows for aerosol mass and number concentration to influence the activation of cloud droplets. Activation of aerosol into cloud droplets is parametrised by Abdul-Razzak and Ghan (ARG; 2000)." • Line 163: ○ "To simplify our simulations, we only consider aerosol ~~only~~ in the soluble accumulation mode and initialise our simulation with a constant sulphate aerosol number concentration of 200 cm−3, characteristic mass 3×10−18 kg, and aerosol diameter of ≈ 0.1μm, which is typical of sulphate aerosols (Noone et al., 2000). ~~The aerosol activation scheme used is that of Abdul-Razzak and Ghan (ARG; 2000)~~. This assumption that we only have aerosol in the soluble accumulation mode enables us to consider differences between control and ships runs as solely a function of increasing aerosol concentration, and the effects of aerosol chemistry are neglected. Differences in hydroscopicities in other aerosol-modes are assumed to be a less important factor in droplet activation than the mass/diameter of the mode."

		• Line 195: add "Non-ship aerosol sources are described in Mann et al. (2010), with emissions from marine phytoplankton, SO2 from volcanoes, fires, and industrial sources. These sources can exist across different aerosol modes (not just the accumulation mode)."
	Comment	**Response**
5, 7, 8	***We combine Referee #2 points 5,7,8 here since they are addressed in one response*** **5)** The simplified GLOMAP-mode aerosol assumption (accumulation mode only) is a non-standard configuration. Has it been described and tested elsewhere? While it may be computationally cheaper, I have significant concerns about whether it is actually scientifically valid. The simulated aerosol in this simplified configuration is not representative of real world aerosol, which calls into question the simulated aerosol-cloud interactions. For example: 1. Other components of marine aerosol are completely neglected, e.g. sea salt, which has a different hygroscopicity to sulfate, which will affect cloud droplet formation	Thank you for these valuable comments. As the referee remarks, the simplified aerosol configuration of this study, whereby we assume all aerosol is in the accumulation mode, is non-standard and not an entirely realistic representation of aerosol within this domain. For the purposes of this study, a simplified aerosol configuration is necessary not only for computational cost, but also for the ability to isolate the causal impact of inputting our ship aerosol. We are largely concerned with the difference between the control and ship simulations, where the difference will be the very high concentration of input ship aerosol. It is likely that the background aerosol field will have some impact due to the different hydroscopicities of other modes, yet we only aim to explore the impact of the ship aerosol. We assume that the ship aerosol concentrations will be sufficiently high such that the background composition is relatively irrelevant. It is clear that this aerosol configuration is not a perfect representation of reality, as seen in Fig. 7 where we can see that our background CDNC field is missing two large cloud formations. This is likely due to our initialisation of the aerosol field at a constant value of 200cm-3, and highlights a more significant uncertainty introduced from our simplified aerosol configuration. Ideally, a more complex aerosol configurations could be implemented, where multiple components with different hydroscopicities are considered. However, this is deferred to future work. Whilst these components will have different hydroscopicities to sulphate, we believe that the size of the mode remains the most important for droplet formation (Dusek et al., 2006).

2. The other aerosol size modes can also contribute to cloud droplet formation, and possibly also impact the initiation of precipitation. **7)** Has any evaluation of the revised aerosol configuration been undertaken? In terms of aerosol number concentrations, or impacts on the cloud fields of modifying the aerosol? **8)** With the above three points in mind, there is no discussion on what the consequences of the revised aerosol configuration might have on the results, say in comparison to the standard full aerosol configuration.	This study aims to demonstrate that even in a simplified aerosol configuration, we can learn about model representations of aerosol perturbations, and isolate areas where improvements are needed. There are multiple sources of uncertainty that we believe are greater than the assumption of solely accumulation mode aerosol, for example: the uncertainty from ARG activation scheme's non-monotonic behaviour, the selection of parameters in the cloud scheme, or the lack of in-cloud aerosol processing in CASIM. In light of this, we add a new section in our Discussion (Section 4.1.2) to discuss the limitations and uncertainty introduced with this simplified aerosol configuration, and how this might impact our results.
	Changes made
	• New Section 4.1.2 "Aerosol scheme" which contains a paragraph describes the limitations of our analysis with respect to this simplified aerosol configuration and discusses the consequences with respect to our results: • "In this work, we only consider aerosols in the accumulation mode, and neglect other aerosol size modes such as sea salt and organic matter. These other modes have different hydroscopicities, which can affect cloud droplet formation and the initiation of precipitation. For example, sea salt can modulate the activation of sulphate nuclei into cloud droplets (Fossum et al., 2020) therefore inclusion of even small concentrations of sea salt aerosol may reduce Nd concentrations. This effect occurs because of the different hydroscopicity of sea salt, but also its larger size, making it an effective CCN. The neglecting of other aerosol modes with different hydroscopicities should not significantly impact cloud droplet number concentrations or initiation of precipitation, however the neglecting of aerosols with different sizes may impact our analysis, since the size of the aerosol mode remains the more important factor for cloud nucleating ability of aerosol particles (Dusek et al., 2006). We assume that when considering the differences between our *ships* and *control* simulations this is not a large impact, since the aerosol perturbation is much greater than the background, however this is an uncertainty that is introduced by this simplified aerosol configuration."

	Comment	Response
6	It is stated that aerosol number concentrations are initialised to 200 cm-3. Does the model reinitialise these concentrations every time it cycles? What happens at the boundaries? Are the simulated aerosol number concentrations allowed to increase or deplete in response to aerosol microphysical processes? Or are they tightly constrained?	The aerosol number concentrations are initialised at the start of the simulation at 200 cm-3, and then allowed to evolve freely after that with the ship emissions, and other model aerosol sources (e.g. industrial sources, or organic marine sources). At the boundaries, the global driving model (initialised from UM analysis) determines the aerosol boundary conditions. The simulated aerosol number concentrations can change in response to aerosol microphysical processes that are in GLOMAP-mode (such as coagulation), however since our coupling between CASIM and GLOMAP-mode is one directional, cloud microphysics does not alter aerosol mass/size/composition except through removal via wet-scavenging (Gordon et al., 2020). We expand our description of the aerosol scheme in Section 2.2
		Changes made
		Line 176: "To simplify our simulations, we only consider aerosol in the soluble accumulation mode and initialise our simulation with a constant sulphate aerosol number concentration of 200 cm−3, characteristic mass 3×10−18 kg, and aerosol diameter of ≈0.1μm, which is typical of sulphate aerosols (Noone et al., 2000). Aerosol is allowed to evolve freely after the model simulation begins, with the boundaries determined by the global driving model. As described in Gordon et al., 2020, aerosol concentrations can vary as determined by the microphysical processes in GLOMAP-mode, however coupling to CASIM does not have the capability to track aerosol tracers within hydrometeors and therefore cloud microphysics does not alter aerosol mass/size/composition except through removal via wet-scavenging."

	Comment	Response
9	With the limits of ARG acknowledged, was any consideration given to implementing the Nenes et al., 2014 activation scheme?	This is an important consideration; we thank the reviewer for this comment. In Fig. 3a, the activated droplet number concentration as a function of aerosol number concentration is plotted for both the ARG and Nenes schemes. In this simple adiabatic parcel model, it is found that the Nenes et al., 2014 scheme would not suffer from the same non-monotonic behaviour as ARG. However, the resultant cloud droplet number concentrations from the Nenes scheme would be significantly overestimated, with Nd of roughly 800 cm-3 predicted, whereas the expected concentrations were of the order of 200-300 cm-3. Therefore, the decision was made to modify the concentrations input into the

	Comment	Response
		model in the ARG scheme, rather than implementing the Nenes scheme. It is worth noting that there is a good agreement between both schemes for aerosol number concentrations less the 1000 cm-3.
		Changes made
		• Line 312: "NS, whilst not containing this non-monotonic behaviour, would contain a large over estimation of cloud droplets at the realistic ship production number and therefore would also be unsuitable with these."

	Comment	Response
10	The results in Section 3.2.2 (Timescales of response) are important, but don't seem to be discussed further. Why is the response different between model and observations?	Thank you to the reviewer for this comment. The results in Section 3.2.2 (now Section 3.3.2) reveal that the timescale of the model response to aerosol in ship tracks is substantially different from that seen in observations. We identify the reason for this difference being due to the representation of the precipitation and therefore is discussed in the following section. Our "Sensitivity to initial condition" Section (3.3.3) describes how tracks that occur in the Eastern Pacific non-precipitating clouds match more closely to observations. In the revised manuscript we extend Section 3.3.2 to make it clear that this is discussed in the following section.
		Changes made
		• Line 399: add "Both the lifetime of the Nd and LWP responses are found to be too long-lived, which suggests that aerosol is not being efficiently removed from the clouds (Wood et al., 2012). This prompts further investigation into which processes are acting unrealistically, and in Section 3.3.3 we explore the realism of the simulated ship tracks in two different cloud types: closed MCC and broken cumulus." • Further discussion in Section 4.1.2 (around line 530) to address potential reasons for the long-lived Nd response.

Minor Comments

	Comment	**Response**

1	Abstract, line 3-4: ‘A good natural analogue for investigating MCB is analysis of ship tracks’. Ship tracks are not natural, and describing their suitability as ‘good’ is over-stated. Perhaps ‘possible’?	We note that whilst the community sometimes describes ship tracks as “opportunistic experiments”, “natural experiments” remains the term used in other fields. The term does not require only nature to be involved (e.g. the London Cholera Outbreak is a natural experiment which is not solely natural).
		
		We modify our terminology in the abstract to use both “natural” and “opportunistic” as to appeal to all audiences.
		Changes made
		• Line 3-4:
		    ○ Delete “A good natural analogy for investigating MCB is…”
		    ○ Replace with “A “natural", or “opportunistic", experiment for investigating MCB is analysis…”

	Comment	**Response**
2	Abstract, line 13: ‘unrealistically easy suppression’. By ‘easy’, perhaps ‘rapid’ or ‘early’ is meant?	In this work, the precipitation of suppression is found to be unrealistic in terms of existence and strength, rather than in timescale (although this would likely be unrealistic too). Therefore, we don’t think “rapid” or “early” wouldn’t be the suitable terminology, however, we agree that “easy” is also not entirely suitable here. We decide to describe the precipitation suppression as “unrealistically strong” to convey the ease at which precipitation is suppressed when aerosol is added into the clouds. **Changes made** • Line 13: “unrealistically easy” changed to “unrealistically strong”.
	Comment	**Response**
3	Line 37: Delete ‘and’ after the second comma?	Thank you for noticing this error. **Changes made** Line 44: Deleted “and”.
	Comment	**Response**
4		We apologise for this mistake in the prose.

	Line 44: ‘…and can be useful they can be…’?	**Changes made**
		Line 51: Delete extra words “and can be useful ~~they can be~~ in assessing”.
	Comment	**Response**
5	Line 60: ‘…an cloud…’?	Apologies for this typo.
		Changes made
		Line 67: replace “an” with “a”.
	Comment	**Response**
6	Line 67: ‘best estimate’, perhaps better referred to as a ‘proxy’?	Thank you for this suggestion, we have made the following modification to the text.
		Changes made
		Line 74: replace “best estimate” with “proxy”.
	Comment	**Response**
7	Line 68: ‘This makes them’, please be more specific about what is being referred to.	Thank you for this comment, we clarify the following point as detailed below:
		Changes made
		Line 75: delete “This makes them…” and replace with “Ship tracks are therefore…”
	Comment	**Response**
8	Line 73: ‘logistical’? Perhaps meant ‘logical’?	“Logistical” was in reference to questions surrounding MCB experiment design, therefore we have made this more clear in the text.
		Changes made
		Line 82: “answer logistical questions relating to MCB” replaced by. “answer questions relating to MCB experimental design”
	Comment	**Response**
9	Line 80: ‘… yet the simulated ship locations are prescribed and not from the actual ships that caused the observed ship tracks’. Can you expand on what is meant here? I don’t follow, surely in all simulations ship locations are prescribed? The model is not	This comment has been addressed in Referee #1 comment #6.
		Changes made
		Please refer to the changes in Referee #1 comment #6.

	explicitly simulating the movement of the ship.	

	Comment	Response
10	Line 84: ‘infer’, mean ‘simulate’?	Thank you for drawing attention to this sentence, as it is not quite clear. The following modification has been made **Changes made** Line 98: Replace “We infer changes in cloud properties and the processes involved from changes in Nd and LWP, and the time evolution of these responses” with “We simulate changes in cloud properties (Nd and LWP), investigating the processes involved as well as the time evolution of the response.”

	Comment	Response
11	Section 2.1: the title ‘Evaluating ship tracks’ isn’t very descriptive. I suggest something like ’Criteria for evaluating ship tricks’. Additionally, some commentary on how the criteria are applied, i.e. what is considered ‘correct’?	Thank you to the reviewer for this comment. We agree that the previous title of this subsection was not very descriptive and now have modified it as suggested. Additionally, we add to this section to discuss how the criteria are applied. **Changes made** • Section title renamed to “Criteria for evaluating ship tracks” • Line 129: add “We apply the above criteria by considering temporal evolutions (treating the length of a ship track as a time since aerosol injection axis) of inside track Nd and LWP, and the associated enhancements from the background (the calculation of which is explained in Section 2.6). We discuss the “correctness" with respect to each criterion in terms of the similarities and differences between the observed responses compared to the simulated responses.”

	Comment	Response
12	Line 115: in describing the domain, it would be helpful to state where the domain is	This is a useful point, and has also been addressed in Referee #1 comment #7. Since we have moved old Fig. 1 to new Section 3.1 (and the figure is now Fig. 2), the old Fig. 2 is now Fig. 1. Therefore, we describe the domain in Section 2.2, and refer to Fig. 1 to illustrate it.

	alongside the co-ordinates, e.g. eastern North Pacific Ocean. A map would also be helpful for orientation; Figure 2 would be a good candidate if it was provided earlier on in the manuscript.	**Changes made** Line 135: Domain described: “Our domain covers the Eastern and Central Pacific, off the coast of California (Fig. 1)” New Fig. 1 shows the domain shaded in blue.

	Comment	**Response**
13	Line 139: what year is simulated?	We have added the year to these dates (now in Section 3.1). **Changes made** Line 332: “from 0000Z 11th July 2018 to 0000Z 13th July 2018”

	Comment	**Response**
14, 15, 16	Line 148: could you refer to Fig. 1c here to show the cellular convection? Line 149: reference to ‘a drizzling scene’ corresponds to Fig. 1g? Line 149: using the word ‘scene’ to describe a geographic domain is confusing. That terminology is useful when considering satellite images, but not appropriate for modelling.	These are useful comments to help improve the clarity of this sentence. The following changes have been made: **Changes made** Line 341: • Delete: The coupled UM-CASIM configuration is able to reproduce open/closed cellular convection in this marine stratocumulus, with a drizzling scene (surface rain rates on average of roughly 0.5 mm/hr). • Add: “The coupled UM-CASIM configuration is able to reproduce open/closed cellular convection in this marine stratocumulus (Fig. 4c in the southwest of the domain), with the presence of drizzle (surface rain rates on average of roughly 0.5 mm/hr; see Fig. 4g)”

	Comment	**Response**
17	Line 153: Reference to Fig. 5 suggests figures need to be re-ordered	Due to the reorganisation of figures and sections, all figure numbers have now been corrected such that they are referenced in order. **Changes made** • N/A

	Comment	Response
18	Figure 1: the 'ships' column would be better presented as a difference plot for contrast to the 'control' column. For example, differences in (g) and (h) are very hard to see.	Thank you for this suggestion (also suggested by Referee #1 in comment #14), we have modified the old Fig. 1 (now Fig. 2) to show the difference in the second column. **Changes made** • New Fig. 4 right hand column
	Comment	**Response**
19	Figure 1: at what level are Na and Nd shown? Surface?	Na and Nd are shown as averages across the first 15 model levels (below ~1km). There is some element of choice in how the number concentrations are calculated such that they are most comparable to the concentrations retrieved from satellites; therefore, this point is made clearer in the text and this uncertainty is acknowledged. **Changes made** • Fig. 4 caption added: "…Nd and Na are mean values across model level numbers below 1km."
	Comment	**Response**
20	Figure 1 caption: do the dotted lines in (b) show the ship location histories rather than just their locations. If so, please state this in the caption.	The dotted lines in Fig. 4b show the historical ship trajectories (i.e. the simulated ship routes in the previous 44 hours of the simulation). We update our caption to state this. **Changes made** • Fig. 4 caption replace "Panel (b) contains the locations of the ships simulated during this study, up until the time step of this figure" with "Panel (b) contains the historical simulated ship routes for this study, up until the time step of this figure."
	Comment	**Response**
21, 22	Line 156: is there any significance attached to them being container ships? Is this an indirect proxy for size? Line 163: is there any logic for emissions being added at 10 m?	We simulate the aerosol perturbations from container ships, as these produce the most visible ship tracks in our observations of this case study (due to their high emissions and large size). We also specify that they are the same type of ship such that the use of the same ship emissions when simulating them is valid. Other than that, there is no significance to the ships being container ships, and we make the changes listed below to improve the clarity of this topic.

	Comment	Response
		Container ships exhaust stack heights are much taller than 10m (roughly 50m), so 10m is not the height at which aerosol should necessarily be added. We left this as it is because the vertical transport of aerosol to cloud base within our simulations happens almost instantaneously, and we do not see deposition at the bottom model level, therefore the simulations are relatively insensitive to the input height.
		Changes made
		• Line 203: add "Container ships are selected because of their large size and high emissions, meaning that the resulting ship tracks are the most distinct in the observations. • Line 211: add "Aerosol is almost immediately transported up to cloud base, and there is little deposition of aerosol at the bottom model level, therefore the 10m injection height (whilst an underestimation) is approximately correct."
	Comment	**Response**
23, 24	Line 169: are the authors aware of the difference between condensation nuclei and cloud condensation nuclei? It's not clear why a CN range is quoted, without further explanation of why 5x10^14 was then chosen. Line 170: 'issues with activation schemes', is there a reference to support this?	This point relates to our selection of input aerosol concentrations. We quote that typical ship production numbers are in the range 10^{16}-10^{18}, yet than state that we use $5x10^{14}$ in this work, without explaining clearly why this is the case. We select $5x10^{14}$ for this study due to the activation parameterisation issues that occur with larger aerosol concentrations than this. However, in an attempt to not discuss results in a methods section the reasoning for this choice is left unexplained. Therefore, we rephrase this section in terms of investigations into two aerosol concentrations ($5x10^{14}$ and $1x10^{16}$). Section 3.1 therefore describes why we use $5x10^{14}$, and why values in the range 10^{16}-10^{18} cannot be simulated with this activation parameterisation in its current form.
		Changes made
		• Line 214: "Realistic emissions of condensation nuclei (CN) from ships fall in the range 10^16 to 10^18 kg/s (Taylor and Ackerman, 1999; Hobbs et al., 2000; Berner et al., 2015). ~~In this work we use 5 × 10^14 due to issues with activation schemes at very high aerosol concentrations.~~ Connolly et al., 2014 find that the ARG activation parameterisation (Abdul-Razzak and Ghan, 2000) exhibits a sharp "drop off" in activated fraction at increasing aerosol concentrations (albeit for NaCl), therefore we investigate the use of a

		high ship aerosol case (ship production number 10^16) and a lower ship aerosol case (ship production number 5 x 10^14) in this study (Section 3.1)."

	Comment	Response
25, 26	Line 172: 'so that all aerosols within the simulation are consistent', can this statement be clarified? Does consistent mean homogeneous? In space and/or time? Line 174: I don't understand how this approach helps to 'isolate causal aerosol effects'	We apologise for the lack of clarity in this paragraph, and we address these two comments together. When we state "consistent" we mean that all aerosols are in the same mode throughout the domain (our assumption of accumulation-mode only aerosol) – both our background and ship aerosols consist only of accumulation mode aerosol. This helps to isolate causal aerosol effects, as we can attribute changes between *control* and *ships* runs as solely due to the increased concentration of aerosol at the ship location. With different aerosol species / modes present, it would not be as simple to isolate the ship impacts on the cloud processes (e.g. due to interactions between different aerosol species). We acknowledge that this is a major simplification in terms of the realism of the aerosol scheme, and therefore this is addressed in the manuscript.
		Changes made
		• Line 219: We add the aerosol into the accumulation soluble mode with a mass 3 × 10−18 kg, the same as background aerosol, ~~so that all aerosols within the simulation are consistent~~ meaning that both our background and ship aerosols are only in the accumulation mode with the same mass. This is done to simplify simulations and isolate causal aerosol effects in our analysis, since differences between the *ships* and *control* run can be solely attributed to the increased aerosol concentration at the ship locations."

	Comment	Response
27, 35, 40	• Line 226: reference to Fig. S3 should be S2? • Line 305: Fig. S4 should be S3? • Line 355: Fig. S3 should be S2?	Thank you for pointing this out, supplementary figure labels have been rectified in the final version, since some plots have been moved to the main text.
		Changes made
		• N/A

	Comment	Response
28	Line 228: I don't follow the logic that 'relatively clean' equates to low background variability, can you expand?	Thank you for drawing to attention this point. The referee is correct in that a relatively clean background does not equate to low background variability. The point that was attempting to be made was that, compared to the high concentration of aerosol in a ship track, any background variability in a clean scene will not have as much of an influence as background variability in a polluted scene, however we agree that the logic of this sentence is unclear. We rephrase this point now in terms of the background variability, however, to make this point clearer.
		Changes made
		• Line 285: "~~This is due to the model background being relatively clean, and therefore is a good representation of the control cloud since there is not much background variability.~~ Our model background has little background variability, and any variability is much smaller than the large perturbation inside the ship track, therefore both methods provide similar results. This would not necessarily be the case in a model simulation with greater background variability."
	Comment	**Response**
29	Line 234: 'cm-3' is this a flux? i.e. per unit area per time?	Our apologies, describing this as a perturbation is incorrect. These are just the aerosol number concentrations inside the ship track. The sentence is modified as below:
		Changes made
		Line 300: "we obtain an aerosol perturbation of roughly 20,000 cm-3" changed to "we obtain aerosol number concentrations within the ship track of roughly 20,000 cm-3"
	Comment	**Response**
30	Line 240: 'implementing simple' should be 'implementing a simple'?	Thank you for identifying this typo, we have fixed this now.
		Changes made
		Line 307: "implementing a simple"
	Comment	**Response**
31	Line 247: 'has be documented' should be 'has been documented'?	This is fixed in the updated manuscript.
		Changes made
		Line 314: "has be documented" replaced with "has been documented"

	Comment	Response
32	Figure 4 caption: please state what the axis labels are in unabbreviated form	Thank you for this comment, we have updated the caption accordingly. **Changes made** Fig. 3 caption "(a) Number of aerosol particles (Na) activated to cloud droplets (Nd) in different activation parameterisations calculated in a simple adiabatic cloud parcel model. Confidence intervals given by the range of temperature, pressure and updrafts simulated in our case study. (b) The expected enhancement in cloud droplets (ϵN) as a function of aerosol mass emission rate (Me), based on equation (1) from Gryspeerdt et al. (2019b). Vertical lines are the emissions from different ship production numbers. Horizontal lines are the enhancements produced from our simulations for each individual ship. Confidence interval is given by the range of background droplet number concentrations in our simulation."
	Comment	**Response**
33	Figure 6: I think it would also be valuable to see the simulated enhancements calculated using the lateral offset method	We remove the supplementary section which discusses the two different methodologies, and the enhancements from the lateral offset method to Fig. 8 panels j and l as an orange dashed line. There are various modifications needed in the manuscript as a result of this **Changes made** • Fig. 8: orange dashed line added for panels j and l • Fig. 8 caption: "(a)-(d) Observed Nd and LWP across all observations of our 5 ship tracks, as well as the percentage enhancement of the "inside track" region (defined by <15 km from the centre). Percentage enhancement is calculated from the mean of the region between 30-45 km away from the centre of the track. (e)-(h) The same as above, but for modelled ship tracks. The percentage enhancement is instead calculated as the percentage change from the control model run (solid orange line). Percentage enhancement using the lateral offset method (as in the observations) is shown in orange dashed line..." • Line 281: "Fig S2" replaced with "Fig 8j,l"
	Comment	**Response**
34	Line 299: are conditions matched between precipitating and non-precipitating conditions in the model? Or selected based on	This comment echoes that of major comments #1 and #2 from Referee #1. For more details of the response please refer to these comments (page 2-3 of this document). In summary, we do select our "precipitating" and "non-precipitating" conditions based on model precipitation, and we determine this precipitation to be somewhat accurate. However,

	observed conditions? If the latter, are the simulated conditions representative of the observed conditions?	inability of GPM to capture light drizzle, and scarce coverage from CloudSat overpasses means that there remains some uncertainty, which we acknowledge in our discussion of the results.
		Changes made

		See above comments from Referee #1 (page 2-3 of this response)
		
		In summary:
		• New Section 3.2 – Evaluation of the control,
		    ○ Evaluation of precipitation against GPM, ERA5 and CloudSat.
		    ○ Performance of CASIM discussed more generally
		    ○ Old section 2.4 moved into this section.
		    ○ New Figures 5 and 6.
		
		• Line 581: "As discussed in Section 3.2, there is some uncertainty in the realism of the simulated precipitation in the domain, since it is difficult to obtain conclusive precipitation measurements across different satellite and reanalysis products. In order for the enhancement of LWP in a ship track to occur, precipitation must only be suppressed at cloud base, and therefore it is also possible that no impact in surface precipitation could occur. This highlights the difficulty in evaluating not only precipitation in models, but specifically the precipitation suppression effect."

	Comment	**Response**
36	Line 309: 'On short timescales', meaning 'Shortly after the ship passes' or similar?	Thank you for this suggestion, we make the modification as detailed below:
		Changes made
		Line 422: "On short timescales" replaced with "Shortly after the ship passes"
	Comment	**Response**

37	Figure 8 caption: Second sentence should begin with something like 'The figure shows…'?	Thank you for this suggestion, we have implemented it in the Fig. 10 caption. **Changes made** Fig. 10 caption "The figure shows clear precipitation suppression…"
	Comment	**Response**
38	Line 314: 'therefore there must be some process within the model…'. This is a bit unfulfilling. Can you review the output diagnostics? Or offer some supporting commentary linked to the processes included in the model?	This is a valuable comment and requires further discussion within the manuscript. Unfortunately, additional output diagnostics are not available for the simulation of this study, therefore we can offer only a commentary on the processes present and potential reasons for the long-lived Nd perturbation in these clouds, and any further analysis would only be possible in future work. We add the following discussion below: **Changes made** • Line 427: "We would expect the response from weaker aerosol perturbations to reduce over time, and therefore there must be some process within the model allowing for this increased Nd perturbation lifetime. A potential source of this long-lived Nd enhancement could be inefficient removal of cloud droplets by entrainment into the surrounding cloud. In these clouds, we do see larger simulated LWP enhancements from 10-15 hours after the ship passes, therefore weak entrainment rates could correspond to the large Nd enhancements at these times also. Additionally, even in these "non-precipitating" conditions, if cloud droplets cannot become rain droplets via accretion or autoconversion, or if larger droplets cannot sediment out, then the removal of aerosol will be underestimated and reinforce the Nd perturbation. A thorough analysis of the output process rates is needed to fully identify the source of this long-lived Nd perturbation, and beyond the scope of this work."
	Comment	**Response**
39	Line 341: I don't think it is accurate to say that a parameterisation has been tuned.	This is a fair point, since we do not actually tune the parameterisation at all, and only change the input aerosol concentrations. We rephrase the sentence as follows: **Changes made** Line 471: "With some tuning of parameterisations…" replaced by "Adjusting ship emissions to 5 x10^14 particles per second as to avoid unphysical behaviour in the activation parameterisation…"
	Comment	**Response**

	Comment	Response
41	Line 367: ‘simple background’ meaning homogeneous background?	Thank you for this suggestion, this vocabulary is more suitable. **Changes made** Line 548: replace “simple” with “homogeneous”
	Comment	**Response**
42	Line 371: ‘We find that the commonly used activation…’. Wasn’t this already known? Perhaps the study has confirmed that finding on Abdul-Razzak & Ghan.	We rephrase this sentence to discuss how our findings demonstrate the implications of these previously found results when attempting to simulate ship tracks. **Changes made** • Line 553: “We find that” replaced by “We confirm that” • Line 554: add “We demonstrate the potential result of the non-monotonicity of the ARG activation when simulating ship tracks - ship tracks appear split (Fig. 2)”
	Comment	**Response**
43	Line 375: ‘at long timescales’. I would argue that the ‘workaround’ undermines the physical realism at all timescales.	This is a fair point, and therefore we acknowledge this in the text. **Changes made** Line 558: “it undermines the physical realism of the simulation at long timescales” replaced by “it undermines the physical realism of the simulation’s temporal evolution”
	Comment	**Response**
44	Line 377: ‘which governs the conversion of cloud droplets to rain…’. Is the KK00 param the approach used in CASIM? A more thorough model introduction would make this clear.	Thank you for this comment, we agree that a more in-depth model description is needed. This is addressed in major comment #4. **Changes made** Please refer to the changes made in major comment #4.
	Comment	**Response**
45	Line 383: ‘an increases of Nd’, should be ‘an increase’?	Thank you for identifying this typo. **Changes made** Line 570: “an increases” replaced by “a small increase”

	Comment	Response
46	Line 396: 'where high-resolution global cloud resolving models are being employed'. Global climate models are not yet approaching resolutions that can be considered high nor cloud resolving?	Thank you to the reviewer for this comment, we have rectified it as follows to be more in line with the current state of the field.
		Changes made
		Line 587: "This has important implications with respect to ~~climate modelling, where work towards high-resolution global cloud resolving models are being employed~~ modelling across many scales, not just regional simulations. The development of global cloud resolving models would also benefit from improved cloud microphysics parameterisations (Satoh et al.,2019)."

	Comment	Response
47	Line 399: Do you have any insights or suggestions as to how autoconversion schemes could be improved?	This is an interesting point for discussion. In this study we find that the rate of production of rain droplets is essentially zero in these increased aerosol concentrations, and therefore our suggestion would be that the autoconversion rate is not allowed to drop to zero as quickly. As seen in Fig. 11a, there is a very distinct transition between precipitating and non-precipitating clouds, and this line is heavily influenced by the autoconversion parameterisation. For a given LWP, small increases in Nd are extremely efficient at suppressing precipitation, therefore a autoconversion parameterisation which enables a slower transition between precip/non-precip would be beneficial.
		Changes made
		Line 593: "Improved autoconversion schemes that remain valid at high Nd (such as a scheme where autoconversion rate does not tend to zero as quickly with increases in Nd for a given LWP, allowing for a more diffuse PoP within Nd-LWP space)…"

	Comment	Response
48	Line 404: 'In order to evaluate the model representation of these processes…'. I would suggest that the model representation hasn't been evaluated, rather the simulated responses to a	Thank you for this suggestion, we implement it as detailed below.
		Changes made
		Line 601: replace "In order to evaluate the model representation of these processes" with "In order to evaluate the simulated response to an aerosol perturbation…"

	Comment	Response
	perturbation have been evaluated.	
	Comment	**Response**
49	Line 410: ‘very clean background conditions’. I would make the argument that the region in question cannot be considered very clean background. As the study shows, there is considerable ship traffic in the area, and with aerosol lifetimes in the range of days to weeks, there is a strong possibility that the background conditions are anthropogenically influenced.	Thank you for this suggestion, we agree that it is unlikely this region is “very clean” given the high frequency of ship traffic. In our model simulation, which does not contain this ship traffic, the domain is very clean, however the same cannot be said for the observations. We update the text accordingly. **Changes made** Line 608: “The ship tracks occur within a marine stratocumulus deck under moderately clean background conditions, including pockets of drizzling cloud”
	Comment	**Response**
50	Paragraph beginning line 412: Can you suggest alternatives, or suggest how ARG can be revised / improved?	This is a useful comment, and we add further discussion to this paragraph as a result. Only the scheme from Nenes et al., 2003 has been evaluated in this work, and it is also found to be unsuitable at our aerosol concentrations. Another possible scheme is that of Shipway and Abel, 2010, yet we do not evaluate it in this work since it is not implemented in *pyrcel*. We suggest that a more comprehensive evaluation of activation parameterisations must be taken place to determine whether modifications to preexisting schemes are possible, or development of new schemes are necessary. This is beyond the scope of this work but would interesting and important for those interested in modelling MCB impacts. **Changes made** Line 613: “An alternative activation parameterisation, from Nenes and Seinfeld (2003, NS), is found to overestimate expected Nd at high aerosol concentrations when evaluated in a simple adiabatic parcel model. Further work is necessary to fully evaluate other activation parameterisations (such as that of Shipway and Abel, 2010) at these high concentrations and determine whether modifications can be made to existing schemes, or if novel ones must be developed.”

		Line 618: remove "However,"

	Comment	Response
51	Line 438: 'This has particular relevance'... It is not clear what 'this' refers to, especially as it used at the start of a paragraph.	Thank you for pointing this out, we modify the start of this paragraph to be clearer about what we are referring to.
		Changes made
		Line 642: replace "This..." with "Any model overestimation of the aerosol effect..."

	Comment	Response
52	Acknowledgements: are incomplete	Thank you for identifying this, this has now been modified.
		Changes made
		Acknowledgements: removed "~~Paul Field acknowledges~~..."