Review of "Detection of large-scale cloud microphysical changes and evidence for decreasing cloud brightness within a major shipping corridor after implementation of the International Maritime Organization 2020 fuel sulfur regulations" by Diamond

This study looks at how cloud microphysical properties (effective radius and brightness) have responded to the 2020 IMO regulation of sulfur in shipping fuels. They examine changes in the southeast Atlantic shipping corridor using statistical methods to construct counterfactual fields of cloud properties without shipping aerosols. They show that only 3 years of data are necessary to already see robust changes in effective radius and cloud albedo due to the sulfur regulation, despite large interannual variability.

Overall, this is a very interesting study. The paper is well written and clear. In general, I felt that the discussion of methodology and results could be expanded a bit to make the paper more readable, especially for those unfamiliar with the previous Diamond et al. (2020) work. I have the following comments, which I think would improve this manuscript.

Major comments:

- It would be nice to include more information about the methods in the main text rather than relegating it to the appendix. I found it necessary to read the methodology first in many cases to understand the figures and key points of the manuscript. In particular, the sections on "Universal kriging" and "Statistical significance testing" would be most beneficial to include before the presentation of results.
- Fig 4:
    - A more substantive question that arose from this figure: Why are data only reported from non-overlapping 3-year time windows? Could, for example, the analysis be done and this figure be made showing a 3-year running-mean over the time period? How would that change the calculation of IMO effect via the persistence method?
    - Can error bars be added to this figure to show an estimate of the confidence that the NoShip effective radii are in fact larger.
    - A technical point: The bracket on 0.3 um should extend the full height of the dotted line.

Minor comments:

- I recommend making a shorter, more direct title. Maybe just remove the phrase "and evidence for decreasing cloud brightness within a major shipping corridor"
- L9: add "*may* come with an undesired" because this is the question the paper sets out to prove or disprove
- L33-34: Delete "Challenges in" & change "pollution are" to "pollution *is*"
- L46: change to "Yuan et al. (2022) found *smaller* Nd *increases*" to parallel the phrasing of "greater re decreases"
- L67: Add a sentence here explaining the choice of season, or why SON features the strongest shipping signal. I assume it is because Sc are most prominent during this season, so there is more baseline cloud which has the potential to be brightened, but this would be helpful to make explicit.

- Fig 1/2: More descriptive labels on the figures would be helpful. E.g. instead of just labeling the years, add “Pre-regulation climatology (2002-2019)”, “3 years pre-regulation (2017-2019)”, “3 years post-regulation (2019-2022)”. And instead of using the ambiguous names “Ship” and “NoShip” these columns could be labeled as “Measurements” and “Inferred Counterfactual”
- Fig 3: This figure is great, and very rich. It warrants more than 1 paragraph of discussion in the text. In particular, it would be nice to include some more detail on how pfield is calculated and then interpretation of what the pfield values mean. Is it significant that the 2020-2022 years are the only 3-year mean that has a change in re with pfield > 0.0001?
- L120: It would be interesting to put this section on compliance into more context in the geophysical literature. This is not the first time that geophysical data have been useful in assessing compliance with policy regulations (e.g. remote sensing monitoring of CFCs and methane leakage from oil and gas). How does your work fit into that bigger picture?
- L164: Does the difference between 2 W/m2 and 0.5 W/m2 in the seasonal vs annual mean give an estimate of how the cloud susceptibility to aerosols varies seasonally? This could be an interesting idea to pursue quantitatively in the context of MCB.
- L167: Can you put this estimate of 0.4 W/m2 into more context? First, what is the baseline value (the total IRF_ACI) this shipping term is modifying? Second, how does this compare to the IRF_ARI from shipping?
- Fig 5: 1) Put “Twomey effect (W m-2)” as the x-label rather than in the subplot titles. 2) Either define the mathematical expressions in the legend in the caption, or (even better) change the legend labels to something more interpretable, 3) consider using more B/W-friendly colors for this plot.
- L192: Please elaborate on the assumptions made for these data products. What bias does assuming the constant cloud and meteorological properties over the diurnal cycle introduce?
- L194: Assuming a constant clear-sky albedo of 0.1 seems like it would ignore the presence of aerosols (dust or smoke). Is this a problem for this region? How much does this bias your results?
- L209: It could be helpful to include a map of the EDGAR SO2 overlaid with AIS ship tracks to illustrate the discussion of the section on “Shipping corridor identification”
- L219: Please add some references on the kriging algorithm, for its development, and also a bit of discussion for how this algorithm is used by others in the literature. Just from reading this section is sounds as if Diamond et al. (2020) was the first/only study to use this method, but of course, this is a fairly common geostatistical technique.
- L226: What is the physical reasoning for including lat^2, lon^2, and lat*lon as predictors in the multiple linear regression?
- L232: This one sentence is the only mention of Figures S5-S8. These figures are not explained anywhere in the supplement and difficult to interpret alone. If they are not going to be discussed in the paper, I would suggest removing them entirely because they do not aid understanding without much more explanation about what an empirical variogram is, how it is computed, and what they show us about the kriging method.

This paper was reviewed by Clare E. Singer, including discussion with Emily K. de Jong.