

Atmos. Chem. Phys. Discuss., referee comment RC1
<https://doi.org/10.5194/acp-2022-688-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2022-688

Anonymous Referee #1

Referee comment on "Possible evidence of increased global cloudiness due to aerosol-cloud interactions" by Alyson Rose Douglas and Tristan L'Ecuyer, Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2022-688-RC1>, 2022

Review of "Possible evidence of increased global cloudiness due to aerosol-cloud interactions" Douglas & L'Ecuyer (2022).

General Comments:

The study used three machine learning models and multivariate linear regression to study the impact of aerosols on cloud fraction. The authors used the synergy of reanalysis and satellite observations to create a dataset combining cloud and environmental information for comparing actual (polluted) cloud fraction to pre-industrial (clean) cloud fraction. The results show that the different machine learning models agree that cloud fraction is generally higher for present day compared to pre-industrial (except for one region). The different models agree well with each other. The authors also estimate the change in shortwave forcing at the top of the atmosphere and find a general cooling. The paper addresses relevant scientific questions within the scope of ACP. The paper is generally well constructed except for some minor editing problems (equation 2 missing, title of section 2 misplaced...). Nevertheless, I have some major concerns about the method and the conclusion on the radiative forcing. Machine learning models are great tools to understand the link between the parameters but the study does not really explore that. I think a deeper analysis on the impact of the different parameters are needed. Therefore I am not convinced that the method is better than looking at clean and polluted cloud fraction constraining for meteorological parameters. Therefore I recommend major revisions before publication. I will detail all the different points below.

Major comments:

- The authors considered so-called pre-industrial scenes but in my understanding they considered pristine present day scenes. I do not understand the label "pre-industrial" since meteorological parameters are the present day conditions and pre-industrial seems to rely on low AI only. I think it would be better to refer to the regimes as clean. Also I am not sure how it would be different to just do a direct comparison of cloud fraction between clean and polluted cases constraining for the 4 parameters (stability, humidity, vertical motion and sea surface temperature). I am wondering if the author did look at that, what would be the conclusions, and how would it differ from their analysis?
- The authors trained the models with 4 parameters: boundary layer stability, relative humidity, vertical motion, and sea surface temperature. Why did they use these 4 parameters? Are they relevant for all the different regions? I think that there are too few parameters to determine the cloud fraction.
- The authors considered that aerosols next to the clouds are similar to within the clouds. I am usually very skeptical about this method. If there is a cloud where it is observed, it might be for a reason (maybe aerosols if this is what we are interested in). They interpolated from surrounding clear sky (taking care of 3D effect or hygroscopic growth). Do they consider all clear sky pixels around the clouds to do the interpolation, or the closest clear-sky pixels to the considered cloudy pixels? If the clouds are horizontally extended, is there a maximum distance between the cloudy pixel and the clear-sky pixels?
- The authors performed the training of the models for different regions. Some regions are constantly under the influence of anthropogenic aerosols or maybe not after specific events (e.g., precipitation) but is it safe to consider this as pre-industrial cases? Also, the regions are never properly defined and I do not know what they refer to.
- All the considered results are shown averaging over time and space, are the models still good with a single prediction? There is no measure to determine if the results are statistically robust or not even on the figures. Are all pixels show a statistically significant change in CF?
- The authors state that they assumed a constant cloud albedo and therefore cannot account for the Twomey effect. Therefore I am wondering why estimating the radiative forcing at all? I advise to omit this part of the study or to clearly highlight everywhere that this result should be handled very carefully. The abstract for example is very misleading to me. In general I would limit the study to cloud fraction.
- I do not understand how the method is validated, nor the statistical models. If the dataset is bad and the method is bad then the cross validation among the models can still be good. I am not saying that the method and the dataset are bad but how the validation is presented does not validate the results but rather the performance of the models. Therefore I am not sure about the validation of the results from the authors and it would need a careful discussion.
- In the figures, I do not understand why the resolution of 12km and 96 km have not been plotted instead of the 30 degree x 30 degree boxes. Also the changes in cloud fraction are shown but I am wondering if the changes are statistically significant for every box (see point 5).
- I think the ML techniques are useful for a deeper analysis as presented in Figure 10 to understand the correlation between the different parameters. Otherwise I feel that this paper is presenting a promising method (then ACP might not be the best journal) with interesting results on CF but I really would like to see an explanation on the reasons for the CF changes, if one parameter is more important than others, if it is the same for all regions...
- I am not sure the study mentions exactly what a region is, is it a 30x30 degrees region ?
- Machine learning models require some parameters as input, for example the depth of the tree, boosting iterations, learning rate for random forest regressor. The current study does not mention the parameters used and how they are decided. Some details

are required.

Minor comments:

- There are several typo that need to be proofread, I am not a native speaker but some sounds strange, I am just highlighting few here:
 - line 1 "uncertainty"-> "uncertainties" or "the uncertainty",
 - line 15 "Aerosol" should not be plural?, if aerosol is singular, then I think it should be "nucleus" on line 16
 - line 25 "aerosol" should be plural?
 - line 54 remove parenthesis for the publications
 - line 106 "in" is missing I think "as important as in regression based..."
 - line 122 "effect" -> "affect"
 - line 145: "below the a" -> "below a"
 - line 254 "we chose to use stick" -> "We chose to stick"
- The authors referred to the IPCC AR5 (Boucher et al., 2013, line 11), a newer version would be better.
- line 15: "Aerosol enters a cloud" I do not know if the phrasing is correct. I suggest "interacts".
- l.75, there is a "2 Methods" that should not be here
- Data section: The authors omitted to include the references for the products they used
- I am wondering how the case of multilayer cloud scenes are treated. Are they removed from the analysis?
- line 132: The authors considered cloudy pixels as $CF > 0$, I am wondering if they checked their method for $CF > 0.9$ for example.
- line 156: The authors referred to $CF_{\{warm\}}$ but it is the first time it appears, is it the same as CF from Equation 1? If so, the annotation should be consistent.
- Equation 2 is missing.
- line 169: "we believe the magnitude of changes in these regions is much smaller than those found in warm topped, liquid phase clouds we evaluate within". I am not sure how it is justified. I would like some more discussion about that.
- l.190 I do not understand how taking 80% of the data avoids overfitting. I do not know if I do not understand how the sentence is constructed or the method.
- The results presented from 12 and 96 kms are very similar and I am wondering if it is necessary to put two different figures. Stating that the results are similar might be enough in my opinion.
- The Figures are never presented but introduced in parenthesis.
- The part "Validation of ML Model Results" arrives at the end of the paper. Should it not be more fitted at the beginning of the result section? Or it should be stated that a discussion on the validation is later when presenting the method.
- Figure 6: on Figure 6a for example, on the y axis, the number of regions is 160 for XG for 0%CF but on the zoomed subplot it is $\sim 40\%CF$. Why is it different?
- From lines 301 to 310: I do not understand the paragraph. Can the authors rephrase it?
- The result section contains a lot of discussion, I suggest to merge the result and discussion sections for a "Results & discussions" section
- Figure 11: I am not sure to understand what are the x-axes for these two figures

