Major Points

Hyperparameters

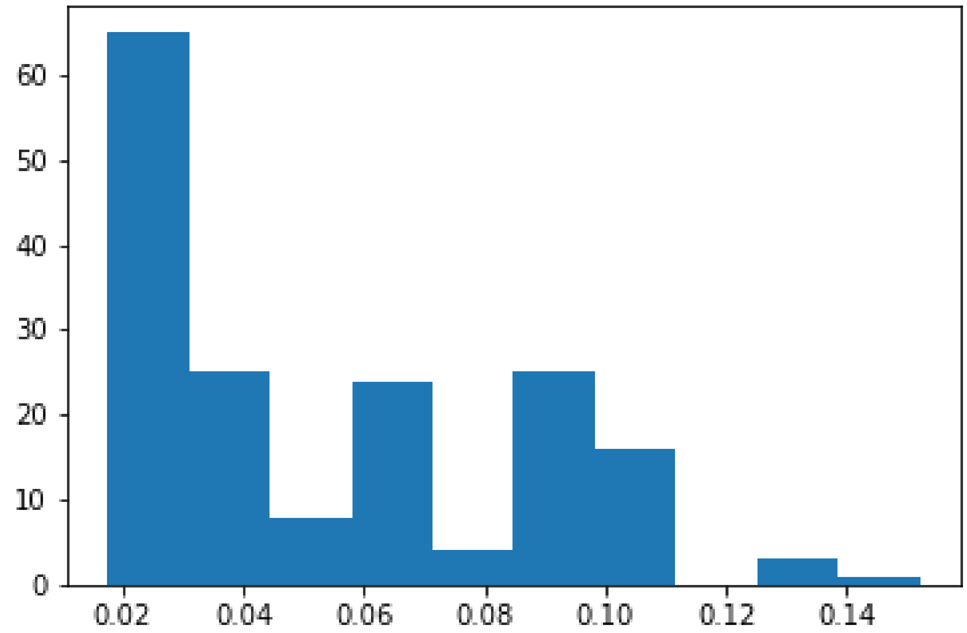
Based on comments from this review and other reviewers, we have added details on the setup of each of the models including whether the model used subsets to help with overfitting.

“All (RF, SGB, XG, and MVLR) models are only trained on 80% of the data (Train data), leaving out 20% of the clean scenes as a testing sample (test data). The RF hyperparameters include having bootstrapped training, meaning the RF is trained on a random subset of observations over each iteration, the number of iterations was limited to 125, and the depth limited to 30. The SGB hyperparameters are a max depth of 30, a learning rate of 0.1, a minimum number of samples per leaf as 3, subsampling at 0.8, and the number of iterations as 200. The XG hyperparameters are max depth of 30, subsampling of 0.8, ‘GBTree’ as the booster type, a histogram tree method, a learning rate of 0.8, and the max number of iterations to train as 300. These hyperparameters were chosen by using a grid search over each model using as subset of the total, global data. Subsampling over each iteration, as the XG and SGB allow, reduces the chances of overfitting as this forces the trees to generalize; the RF uses random sampling over each iteration for similar purposes. The models are cross validated by re-training on a different subset of 80% of clean observations over ten iterations as a cross validation step to reduce possible sampling bias.”

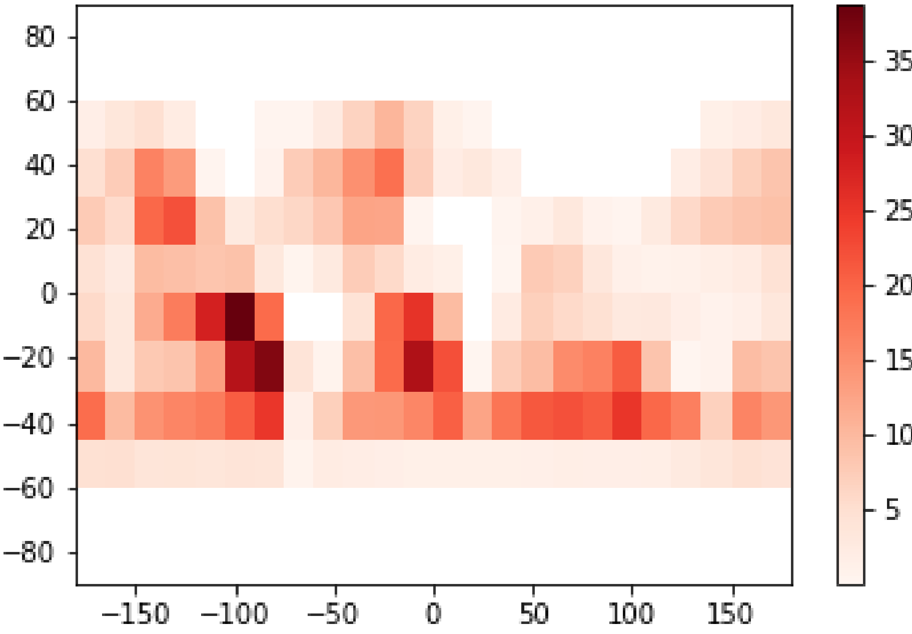
If the models were overfit and simply memorized the training data as you suggest, this should have greatly decreased their performance during cross validation. Further, the models do not experience a decrease in their mean squared error or explained variance scores in the test set relative to the training set until the test set is >40% of observations. The other ML based studies predicting cloud fraction cited (Fuchs et al. and Chen et al.) use vastly different cloud fraction estimates. Our observations, as explained in the data section of the methods, come from an active satellite radar, which is much better at separating low, warm clouds from other cloud layers and only reports its data in along-satellite track segments of ~1 km x 1 km. MODIS observations are from a passive sensor, which changes the nature of the observations, their accuracy in defining warm clouds vs. other cloud types, and the accuracy of defining cloud vs. aerosol scenes. It is likely these differences lead to the difference in skills. In Cumulo: A Dataset for Learning Cloud Classes, we achieved 89% explained variance scores using only a single year of CloudSat/CALIPSO data when predicting a much more convolved, intricate problem. Higher resolution observations lead to a more precise, accurate estimation of the environmental drivers of cloudiness.

SPRINTARS to define pre-industrial

A previous version of this same study set the pre-industrial AI to 0.08 as the limit. The reviewers on that version of this paper did not agree with setting a single AI as the pre-industrial threshold. With that feedback in mind, in this version of the paper we implemented a moving threshold based on the regional pre-industrial aerosol index from the SPRINTARS model. This is the same pre-industrial AI used in previous studies to define the forcing from ACI (Douglas & L'Ecuyer 2019, 2020) and others looking direct effects (Matus et al. 2019). A more map of the AI used can be found in Douglas & L'Ecuyer 2019. This threshold is much lower than the previous limit of 0.08 and led to a slight increase in the change in cloud fraction estimates.



The regions with a higher limit are those off of the coasts of desert regions, such as near saharan



Africa and western United States.
The number of observations coincides with where low clouds lie. This is also what leads to the greater weighting of the southern hemisphere as all forcings are weighted by occurrence.

Change in CF

The change in cloud fraction is found for all scenes, however since the predictors are trained on the pristine scenes, those predictions would therefore “cancel” each other out as the predicted pristine and actual observed pristine would be the same. The change in cloud fraction is then not overly weighted towards the polluted observations.

MVLR

We did implement a simple (two layer ANN) and complicated (4 layer, dense) neural networks as comparisons when beginning this work. The neural networks performed worse than the decision trees. The drawbacks of adding layers and steps which decreases our interpretability of the neural network were outweighed by their decrease in accuracy compared to the more simple decision tree architectures chosen. Neural networks excel at images, segmentation, and classification. Decision tree models are still the state-of-the-art standard however for tabular data. It is possible that for some of the passive satellites such as MODIS or GOES, which create data “images” rather than segments like CloudSat, a neural network would perform better given the type of data.

Minor Points

L8: Units of Wm-2 added.

L28 - 31: We have added: “And though cloud fraction is a physical state that can be directly observed by our naked eye, our sensors aboard satellites must still assume thresholds on values like optical depth to determine if a scene is truly”cloudy” (Mieslinger et al. 2022).”

L157: Fixed.

L171 - 174: Reworded to: “Likely the pristine scene conditions are minimally, if at all, affected by aerosol direct effects, therefore the only uncertainty or bias direct effects may pose are in extremely polluted scenes which experience a decrease in cloud fraction due to aerosol absorption and atmospheric heating.”

L176 - 178: Fixed the strange wording: “A random forest regressor (RF), stochastic gradient boosted regressor (SGB), and an extreme gradient boosted regressor (XG) are compared against results from a multivariate linear regression.”

L229-231: With feedback from this review and others, we have reworded our explanation on why we only predict for clouds which have formed. In essence, we treated aerosol effects on cloudiness as two distinct problems: how aerosol may alter the amount of cloud (which we can quantify) and how aerosol may alter *when* cloud forms (which is harder to prove). Further, altering when cloud forms can be influenced by large scale feedbacks. This now reads: “We quantify only the changes for scenes which were already cloudy, as we believe delineating how cloud fraction may have increased is different in nature than how cloud occurrence may have changed due to anthropogenic aerosols.”

L231-232: We have added a reference to a similar study using the same data from CloudSat which found a reduced cloud albedo in the south Atlantic. A recent study by Zhang and Feingold 2022 (*in discussion*, https://acp.copernicus.org/articles/23/1073/2023/) has also found distinct decreases in albedo and extent in the south Atlantic. Our past work (Douglas & L’Ecuyer, 2020), using a different framework to separate the effects of aerosol on cloud extent vs. brightness, also found a decrease in extent in the south Atlantic.

L319: Simplified to: “The MVLR, similar to historical methods of estimating the sensitivity of a cloud property to a CCN proxy, displays variability in the sign and magnitude of the change in cloud fraction compared to the ML models.”

Southern Hemisphere: We have added more discussion on the magnitude including how our weighting leads to these results in the discussion.

Specific points:

“environment” We have clarified: “In our context, we term the environment to mean the local meteorology of the cloud without considering aerosol as an environmental feature.”

“to constraint” fixed to “to constrain.”

2 Methods 3 Data fixed

“split into by 20%” removed by

“until our we” removed our

L321: fixed to read stability

References

Douglas, A., & L’Ecuyer, T. (2019). Quantifying variations in shortwave aerosol–cloud–radiation interactions using local meteorology and cloud state constraints. Atmospheric Chemistry and Physics, 19(9), 6251-6268.

Douglas, Alyson, and Tristan L’Ecuyer. “Quantifying cloud adjustments and the radiative forcing due to aerosol–cloud interactions in satellite observations of warm marine clouds.” Atmospheric Chemistry and Physics 20.10 (2020): 6225-6241.

Matus, Alexander V., Tristan S. L’Ecuyer, and David S. Henderson. “New estimates of aerosol direct radiative effects and forcing from A-train satellite observations.” Geophysical Research Letters 46.14 (2019): 8338-8346.