Supplement of Geosci. Model Dev., 17, 2641–2662, 2024
https://doi.org/10.5194/gmd-17-2641-2024-supplement

Geoscientific Model Development
Open Access

Supplement of

A machine learning approach for evaluating Southern Ocean cloud radiative biases in a global atmosphere model

Sonya L. Fiddes et al.

Correspondence to: Sonya L. Fiddes (sonya.fiddes@utas.edu.au)

1 Multiple Linear Regression

The MLR was able to predict between 42-43% of the variance (when tested on different summers, in the same way as described for the XGBoost training and testing data sets), with a predicted mean of $12.4\,\mathrm{W\,m^{-2}}$, median of $13.0\,\mathrm{W\,m^{-2}}$ and a standard deviation of $29.2\,\mathrm{W\,m^{-2}}$, compared to mean of $12.4\,\mathrm{W\,m^{-2}}$, median of $11.7\,\mathrm{W\,m^{-2}}$ and standard deviation of $44.4\,\mathrm{W\,m^{-2}}$ for the true values. Figure S1 shows the true spatial SWCRE_{TOA} bias (a), the MLR predicted bias (b), residual (c) and a histogram of the residual against the prediction (d). In this final subplot, a more symmetrical concentration of residuals (y-axis), centred around zero, and a narrow range of predictions is an indicator of a well performing model (x-axis). Here we can see that the values are skewed towards more negative negative values for the residual, while the prediction is more evenly distributed. This indicates that the model is tending to under predict the SWCRE_{TOA} bias more frequently than it over predicts it, regardless of what the prediction value is.

While LWP has the largest linear relationship with the SWCRE (as shown in Appendix Figure A3), in a normalised multiple linear regression, the liquid water cloud optical depth (TauL) has the largest co-efficient, then LWP, IWP, CFI, CFL, CTP and TauI. This result is somewhat different to what the SHAP features indicate (as discussed in Section 4.1). Unlike with the SHAP analysis, we cannot then look at how the individual contributions of each predictor contribute to the final result of individual or averaged predictions, limiting our ability to analyse this further without repeating this analysis across individual components. This is another reason why we think that the XGBoost/SHAP method presented is superior, given it can be modelled using the entire dataset, and then analysed for its individual components.

These results suggests that linear assumptions are less suitable for a problem such as this, and that weak-moderate multi-collinearity must be taken into account. Furthermore, beyond evaluation of the correlation values, mean statistics, and co-efficients, the MLR cannot then be further split into differing spatial, temporal (or both - e.g. cloud types) to further understand its predictions. Performing individual MLRs on each cloud regime (or even individual lat/lon cells even) can over come this to a degree, but this results in the compartmentalisation of a system, where we are looking for a method that can consider the system and its complexities as a whole.

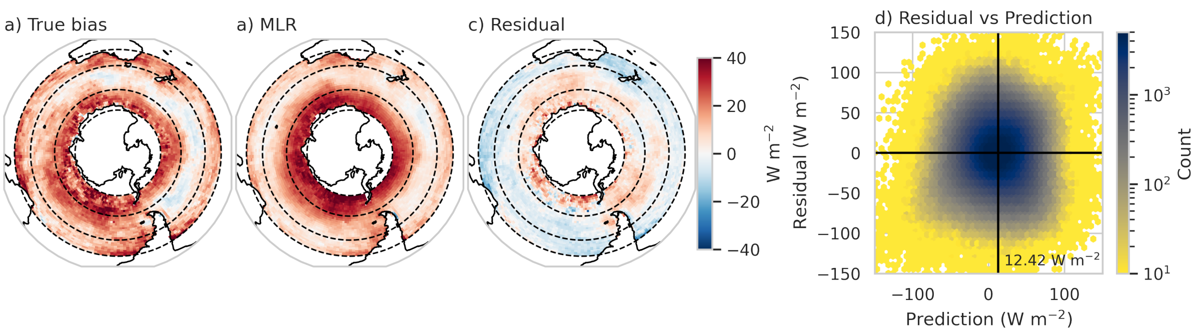

Figure S1. The true (a) and multi-linear regression predicted (b) DJF SWCRE$_{TOA}$ bias (ACCESS-AM2 minus CERES-Syn1D) averaged over time; (c) shows the residual difference between the predicted and true biases. The dashed lines represent the three regions of interest, mid-latitudes (30-43°S), sub-polar (43-58°S) and the polar (58-69°S) regions. In (d) a histogram of the residual against the prediction, the black lines are represent 0 W m^{-2} for the residual and the mean prediction of 12.42 W m^{-2}. All units are in W m^{-2}.