Reducing Southern Ocean Shortwave Radiation Errors in the ERA5 Reanalysis with Machine Learning and 25 Years of Surface Observations

MARC D. MALLET,[a] SIMON P. ALEXANDER,[b,a] ALAIN PROTAT,[c,a] AND SONYA L. FIDDES[a]

[a] *Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia*
[b] *Australian Antarctic Division, Kingston, Tasmania, Australia*
[c] *Bureau of Meteorology, Melbourne, Victoria, Australia*

(Manuscript received 2 June 2022, in final form 23 January 2023)

ABSTRACT: Earth system models struggle to simulate clouds and their radiative effects over the Southern Ocean, partly due to a lack of measurements and targeted cloud microphysics knowledge. We have evaluated biases of downwelling shortwave radiation in the ERA5 climate reanalysis using 25 years (1995–2019) of summertime surface measurements, collected on the Research and Supply Vessel (RSV) *Aurora Australis*, the Research Vessel (R/V) *Investigator*, and at Macquarie Island. During October–March daylight hours, the ERA5 simulation of SW_{down} exhibited large errors (mean bias = 54 W m^{-2}, mean absolute error = 82 W m^{-2}, root-mean-square error = 132 W m^{-2}, and R^2 = 0.71). To determine whether we could improve these statistics, we bypassed ERA5's radiative transfer model for SW_{down} with machine learning–based models using a number of ERA5's gridscale meteorological variables as predictors. These models were trained and tested with the surface measurements of SW_{down} using a 10-fold shuffle split. An extreme gradient boosting (XGBoost) and a random forest–based model setup had the best performance relative to ERA5, both with a near complete reduction of the mean bias error, a decrease in the mean absolute error and root-mean-square error by 25% ± 3%, and an increase in the R^2 value of 5% ± 1% over the 10 splits. Large improvements occurred at higher latitudes and cyclone cold sectors, where ERA5 performed most poorly. We further interpret our methods using Shapley additive explanations. Our results indicate that data-driven techniques could have an important role in simulating surface radiation fluxes and in improving reanalysis products.

SIGNIFICANCE STATEMENT: Simulating the amount of sunlight reaching Earth's surface is difficult because it relies on a good understanding of how much clouds absorb and scatter sunlight. Relative to summertime surface observations, the ERA5 reanalysis still overestimates the amount of sunlight entering the Southern Ocean. We taught some models how to predict the amount of sunlight entering the Southern Ocean using 25 years of surface observations and a small set of meteorological variables from ERA5. By bypassing the ERA5's internal simulation of the absorption and scattering of sunlight, we can drastically reduce biases in the predicted surface shortwave radiation. Large improvements in cold sectors of cyclones and closer to Antarctica were observed in regions where many numerical models struggle to simulate the amount of incoming sunlight correctly.

KEYWORDS: Southern Ocean; Shortwave radiation; Machine learning

1. Introduction

Numerical models struggle to simulate cloud properties and their effect on surface radiation over the Southern Ocean (Bodas-Salcedo et al. 2014; Schuddeboom and McDonald 2021; Protat et al. 2017). Deficiencies in simulated cloud cover and water content have resulted in large surface radiation biases in austral summer over this region. This has consequences for other components of the Earth system. For example, climate models that exhibit this surface radiation bias also have a positive bias in sea surface temperature (Hyder et al. 2018), and a poleward shift of the Southern Hemisphere jet (Ceppi et al. 2014). Importantly, a lack of knowledge about how clouds might respond to a changing climate stems from a poor understanding of clouds in extratropical regions such as the Southern Ocean (Zelinka et al. 2020; Tan et al. 2016). A key reason for the bias in simulated shortwave radiation over the Southern Ocean is a relative lack of long-term surface-based measurements and short-term intensive observational campaigns when compared with the Northern Hemisphere.

There has been a recent increase in efforts to collect more observations over the Southern Ocean (e.g., McFarquhar et al. 2021; Schmale et al. 2019; Kremser et al. 2021; Fossum et al. 2018). Observations are used to build and constrain parameterizations of cloud microphysical properties and to evaluate subsequent climate simulations. While issues in simulating Southern Ocean shortwave radiation in numerical weather prediction and climate models have been identified, this problem is also evident in reanalysis products (Trenberth and Fasullo 2010; Naud et al. 2014; Wang et al. 2020).

Reanalyses use resolved atmospheric variables and subgrid scale parameterizations to estimate gridscale cloud and radiation properties. The potential advantage of reanalyses is that

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/AIES-D-22-0044.s1.

Corresponding author: Marc D. Mallet, marc.mallet@utas.edu.au

DOI: 10.1175/AIES-D-22-0044.1 e220044

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

they assimilate a lot more data than weather forecast models, implying that cloud properties should potentially be better constrained. The accuracy of cloudy-sky radiation estimates strongly depends on how accurately the cloud properties are simulated. Unfortunately, even reanalysis products are prone to radiative biases in the Southern Ocean (Trenberth and Fasullo 2010), likely also due to the sparse data of cloud properties. Recently, shortwave radiation errors of up to 39 W m^{-2} in the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), were found using ship-based observations in the Southern Ocean during austral summer (Kuma et al. 2020). Similarly, daily errors of 60 W m^{-2} in the fifth major global reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ERA5) have also been reported in the Southern Ocean and associated with biases in cloud fraction and transmittance (Wang et al. 2020). Outputs from reanalyses are used to prescribe forcings (including downwelling shortwave radiation) for other Earth system models, such as ocean or ice models, that are not coupled to an atmospheric model. Therefore, while it is important to reduce biases in short- and long-term weather and climate models over the Southern Ocean, efforts to do the same for reanalysis products are also needed. A variety of approaches will be required in this endeavor.

Improving our understanding of the physical processes that underpin cloud–radiative interactions over the Southern Ocean remains a grand challenge (Paton-Walsh et al. 2022). The abundance of supercooled liquid water in the Southern Ocean is a particular challenge in numerical models (Vignon et al. 2021; Lenaerts et al. 2017). Furthermore, cold sectors of extratropical cyclones have also been identified as regions that exhibit the largest biases in simulated cloud and radiation properties, both in climate models and reanalyses (Bodas-Salcedo et al. 2014). The sources, transport and evolution of cloud condensation nuclei and ice nucleating particles, and the formation of ice crystals and precipitation over the Southern Ocean are processes that are still not fully understood (Twohy et al. 2021). Recent studies have also highlighted large differences in aerosol and cloud properties across the Antarctic Polar Front (Humphries et al. 2021; Mace et al. 2021a,b; McFarquhar et al. 2021; Krüger and Graßl 2011; Lang et al. 2018), likely due to complex spatial and seasonal variability in the sources of aerosol and interactions between aerosols, clouds, precipitation, and radiation. Until new measurements are collected and knowledge from these is integrated into improved cloud microphysics parameterizations in Earth system models, we should look to alternative approaches for modeling surface radiation. This study explores whether machine learning might be useful in this pursuit, using output from a reanalysis as a starting point.

There is a growing number of studies that are implementing machine learning methods in atmospheric and environmental sciences, including cloud microphysics parameterizations (Morrison et al. 2020). This trend can be attributed to the recent availability of very large datasets for training and advances in computational power and storage. This advancement has allowed us to collect, store and process extremely large amounts of data to advance scientific discovery in a range of fields including climate science (Karpatne et al. 2017). The development of open source tools for machine learning (Géron 2019; Lantz 2019; Boehmke and Greenwell 2019) has also expanded the capability of users outside of computer science to apply machine learning to a wide variety of applications, including those in environmental, atmospheric and climate related fields (Fleming et al. 2021). Machine learning algorithms based on decision trees and neural networks are popular across a variety of applications. Given enough data, these algorithms can learn complex, sometimes nonlinear, relationships between a set of inputs and a desired output. For example, gradient boosted regression, a decision-tree-based method, has been used to successfully model cloud fraction and cloud droplet radii over the southeast Atlantic (Fuchs et al. 2018). A number of studies have also used decision-tree algorithms to isolate meteorological influences on atmospheric composition (e.g., Grange et al. 2018; Grange and Carslaw 2019; Ryan et al. 2021; Mallet 2021). Machine learning techniques have also proven useful for predicting solar radiation for the purpose of solar energy production (Voyant et al. 2017).

The use of machine learning in weather and climate modeling is also showing promise. Krasnopolsky et al. (2013) used neural networks to learn a stochastic convection parameterization and successfully implemented this in the National Center for Atmospheric Research Community Atmosphere Model. Since then, machine learning has been used to parameterize moist convection (O'Gorman and Dwyer 2018) and emulate radiative transfer processes (Lagerquist et al. 2021; Beucler et al. 2021, and references within). Machine learning has also been recently used to develop data-driven weather and climate models, although this research is still in early stages (Dueben and Bauer 2018; Scher and Messori 2019; Weyn et al. 2019). To aid with this research, Rasp et al. (2020) developed a benchmark dataset to be able to consistently evaluate and compare data-driven methods with each other and with physically based models. Despite this, training data with high spatial and temporal coverage are crucial in these efforts, and so reanalyses are often used, which, as discussed previously, can themselves be biased in regions like the Southern Ocean.

While machine learning–based models can be used to make accurate or efficient predictions, it is also important that these models are trustworthy and interpretable (Miller 2019; Murdoch et al. 2019). If a model is interpretable, it should be clear for humans to understand how and why certain predictions are made. Explainable artificial intelligence is an emerging area of research, and a number of methods exist to interpret models built using machine learning (Molnar 2020). While permutation feature importance is a common tool for this purpose, this method can be misleading when multiple features are correlated with each other (Hooker and Mentch 2019). A good alternative is Shapley additive explanations (SHAP), a recently developed method that can quantify the individual feature contributions (Lundberg and Lee 2017). Aside from accounting for interactive effects between features, SHAP offers several other advantages: (i) SHAP values for each feature are calculated for every observation, which allows for a deeper investigation into understanding the drivers of good or bad predictions, or under any predefined condition. (ii) SHAP values for regression are expressed in the units of the predictor variable and are therefore easy to interpret and compare. (iii) SHAP values are additive, so if a number of predictor

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

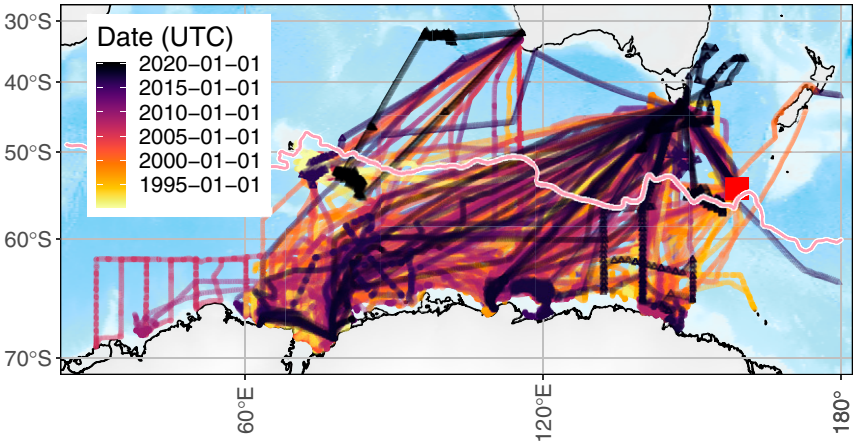

FIG. 1. (a) The coverage of the RSV *Aurora Australis* and R/V *Investigator* in the Southern Ocean between 1995 and 2020. Ship locations are displayed for each hour. Macquarie Island is shown with a red square. The climatological mean position of the Antarctic Polar Front (Orsi et al. 1995) is indicated with a pink line.

variables are strongly related to each other, the SHAP values for each of them can be added together. (iv) Exploring the relationships between each feature and its associated SHAP value can be used to understand the partial dependence between each feature and the predicted variable. Examples of SHAP used to better understand machine learning–based models in the context of atmospheric science include Stirnberg et al. (2021) and Guyot et al. (2022).

In this work, we focus on modeling the surface downwelling shortwave radiation flux (SW_{down}) in the Southern Ocean using machine learning techniques. To do this, we use gridscale meteorological variables provided by ERA5 as predictive features and 25 years of ship-based and ground-based downwelling shortwave radiation measurements as the ground truth. We interpret and scrutinize the relationships between our predictor variables and our predictions of SW_{down} using SHAP. This paper provides a method to bypass the issue of directly simulating cloud properties and implementing the radiative transfer modeling that ERA5 uses. Instead, we will show how the training of models using machine learning techniques with a set of imperfect gridscale meteorological predictor variables can improve predictions of SW_{down} when compared with measurements collected over the Southern Ocean between 1995 and 2019. We did not include predictors from external datasets (e.g., satellite data or other surface-based measurements). We chose to limit out predictive features to those available within ERA5 only. This limitation was imposed to demonstrate that a technique could be developed that would allow a trained model to be implemented in near-real time.

2. Methods

a. Surface measurements and reanalysis data

To evaluate the SW_{down} simulated by ERA5, we use measurements collected on Research and Supply Vessel (RSV) *Aurora Australis* between 1995 and 2020 (Reeve and Symons 1999), which made multiple transits of the Southern Ocean each austral summer, between Hobart (Australia) and the three Australian Antarctic stations; the Research Vessel (R/V) *Investigator*, which has undertaken a number of voyages in the Southern Ocean since 2015; and at Macquarie Island (54.5°S, 159°E) between 2016 and 2018 during an Atmospheric Radiation Measurement (ARM) deployment (McFarquhar et al. 2021). The uncertainty in the downwelling shortwave radiation measurements, collected with pyranometers, is on the order of ±4% (Hinkelman and Marchand 2020). The coverage of the ship tracks is displayed in Fig. 1. For the ship-based measurements of SW_{down}, the maximums of the port and starboard measurements were taken to mitigate the influence of shadows from ship structure (Protat et al. 2017). The data were then averaged for each hour so they could be used to coincide with ERA5 data. The number of observations collected each month from the RSV *Aurora Australis*, the R/V *Investigator*, and at Macquarie Island is shown in supplementary Fig. 1 in the online supplemental material.

Hourly ERA5 (Hersbach et al. 2020) data were retrieved from 1995 until 2019 across the domain covering all of the ship locations from 44° to 69.5°S and from 49.9° to 159.5°E. The single closest grid cell (0.25° by 0.25° resolution) to the mean hourly ship position were then extracted. A mixture of single-level and pressure-level fields was retrieved. The single-level variables included the surface downwelling shortwave radiation (SW_{down}), the clear-sky SW_{down} (sw_clear), the clear-sky $longwave_{down}$ (lw_clear), skin temperature (skt), 2 m temperature (t2m), sea surface temperature (sst), mean sea level pressure (mslp), boundary layer height (blh), total cloud cover (tcc), and low cloud cover (lcc). The pressure-level variables consisted of the temperature at 700, 850 and 1000 hPa (t700, t850, and t1000, respectively), the relative humidity at 950 and 850 hPa (r950 and r850, respectively), and the zonal and meridional wind components at 600 hPa (u600 and v600, respectively). The lower-tropospheric stability (lts),

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

estimated inversion strength (eis), and marine cold-air outbreak M parameter were also included and calculated (Naud et al. 2020) using a combination of these single-level and pressure-level variables. Many of these variables were selected because they had been used in a previous study using machine learning techniques to predict cloud properties over the southeast Atlantic (Fuchs et al. 2018), or because they were deemed important for distinguishing cloud and radiation properties in different seasons and latitudes of the Southern Ocean. A full list of these abbreviations and other acronyms used throughout this paper can be found in the appendix.

The measurements of SW_{down} were first used to evaluate the ERA5 simulation of SW_{down}. Following from this, all of the retrieved single-level, pressure-level and derived variables from ERA5 except the SW_{down} were used as predictor variables to tune, train and test models using various machine learning algorithms. The evaluation of the ERA5 SW_{down} against the measured SW_{down} was used as a benchmark against the performance of the models built using machine learning.

The most significant bias in ERA5's simulation of SW_{down} occurs in the late austral spring to early austral autumn, which also coincides with when most of the ship-based measurements were collected. We have therefore limited our study to try to accurately predict SW_{down} in the Southern Ocean using machine learning between October and March. Daylight hours only were used for training and testing by selecting periods when the clear-sky SW_{down} were nonzero, since this is a physical constraint on SW_{down}. This resulted in 45 262 hourly observations (83% from the RSV *Aurora Australis*, 12% from Macquarie Island, and 5% from the R/V *Investigator*) selected for model training, testing and evaluation.

b. *Predicting surface radiation with machine learning*

Three different algorithms were used to build models to predict SW_{down}. These included generalized linear regression modeling (GLM), random forest, and extreme gradient boosting (XGBoost). Although these algorithms are widely used in the application of machine learning, we will describe them briefly here. The random forest algorithm (Breiman 2001) fits many hundreds or thousands of decision trees to training data that can be used to make regression or classification predictions on new data. The decision trees are built independently from each other and the final decision for regression is the mean of each prediction from all of the decision trees. Diversity in decision trees is introduced by choosing different variables to split the tree further, as well as different thresholds for how the data is split down the tree. The XGBoost algorithm (Chen and Guestrin 2016) is similar, except the decision trees are grown sequentially, with each new decision tree attempting to predict the errors of the previous tree. Both the random forest and XGBoost algorithms are popular choices in regression and classification problems with many predictor variables due to their high performance and ability to easily handle nonlinearity between the predictor and predicted variables, as well as collinearity between predictor variables. Despite this, they can be time consuming to tune and train. Given this, generalized linear regression, based on ordinary linear regression, was also implemented in this study. Although GLM assumes a linear response between each predictor variable and the predicted variable, GLM is much faster to tune and train than decision-tree based algorithms.

c. *Model tuning, training, and evaluation strategy*

Tuning, training, and evaluating machine learning models in this study was done using the *tidymodels* collection of packages in *R* (Kuhn and Wickham 2020). Each algorithm has a small number of hyperparameters that can be tuned to optimize the performance of the trained models. For GLM, these hyperparameters are the amount of regularization applied as well as the degree of mixture between Lasso and Ridge regression (Tibshirani 1996). For random forest, the two tunable hyperparameters are the number of predictor variables available for splitting at each node of the decision tree, as well as the minimum number of samples required to split a node further. The number of decision trees can be considered a tunable parameter; however, once a sufficient number is selected, the performance of the trained model is generally stable. For XGBoost, the tunable hyperparameters include those from random forest, with several additional hyperparameters. These include the number of trees (too many can result in overfitting), the learning rate, which determines how influential each new sequentially built decision tree is, as well as the maximum depth of each tree. Further information about the hyperparameters associated with each algorithm can be found in a description of the R software package *parsnip* (Kuhn and Vaughan 2022).

To select the optimal hyperparameters and to fairly assess the performance of each model, we performed a 10-fold shuffle split of the data with a nested 10-fold cross validation for hyperparameter tuning via grid search (Kuhn and Silge 2022). Prior to each of the 10 shuffle splits, the data were grouped by month so that the smallest chunk of consecutive data in either set was one month. This step is important to avoid unwittingly overestimating the model performance by neglecting to account for temporal autocorrelation (the fact that each hourly observation is not independent of observations directly before or after each hour). Each of the 10 shuffle splits were produced by randomly selecting data from 75% of the full months spanning the 25 years of data. The 10-fold cross validation was performed on the training sets from each of the 10 shuffle splits for the purpose of hyperparameter tuning. In this step, data was separated into 10 splits grouped by month, and models were trained for hundreds of hyperparameter combinations using a grid search on 90% of the data and validated with the remaining 10% and then repeated for each of the folds. For each algorithm, the best combination of hyperparameters from the 10-fold cross-validation steps for each shuffle split was then selected based on the root-mean-square error (RMSE) performance metric. This final specification was then used to train models for each algorithm and for each of the 10 shuffle splits. The trained models were then fitted to the 25% of data in each of the 10 shuffle splits. The shuffle-split approach for the training and testing steps was chosen over k-fold cross validation because we were

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

bringing in SW_{down} datasets from different measuring platforms (i.e., Macquarie Island station, the R/V *Investigator*, and the RSV *Aurora Australis*) that had short periods when measurements were made near each other (e.g., near Hobart or when the RSV *Aurora Australis* was near Macquarie Island). This ensured that the data from these periods were not separated across the training and testing sets and any particular split.

The performance of each trained model on each of the test sets was then aggregated to extract the mean and standard deviation of the following metrics: mean bias error (MBE), mean absolute error (MAE), RMSE, and the coefficient of determination R^2. The same performance metrics were also calculated for ERA5's prediction of SW_{down}, which we use as a benchmark. All random selections in the training, tuning, validation and testing process were done with an assigned seed so that all data were split in the exact same way. This method ensures that comparing the aggregated performance metrics of the three machine learning methods as well as ERA5 is done in the fairest way, taking into account variability in performance due to differences in the selection of training and testing sets.

Last, we explored the performance of each model for different spatial regions. These spatial regions included data points north and south of the Antarctic Polar Front, as well as different positions relative to Southern Ocean extratropical cyclone centers. For the Antarctic Polar Front, the observations and model prediction data points were separated based on their position relative to the climatological mean position of the Antarctic Polar Front (Orsi et al. 1995). The mean latitudinal position of the Antarctic Polar Front does not vary substantially over time (Freeman et al. 2016), and therefore the mean climatological position was deemed appropriate. While the Antarctic Polar Front is defined as the maximal gradient in sea surface temperature latitudinally across the Southern Ocean, previous studies have shown that there are also distinct differences in cloud properties across this boundary (Mace et al. 2021a; McFarquhar et al. 2021). Cloud properties are also different in cold and warm sectors of cyclones, with cold sectors exhibiting most biases in associated cloud radiative properties in models, including reanalyses, over the Southern Ocean (Bodas-Salcedo et al. 2014). Here, we binned the observations and model predictions of SW_{down} according to their horizontal and vertical positions relative to cyclone centers for the period of 2016–18. The University of Melbourne cyclone tracking algorithm (Lim and Simmonds 2007) was used to identify cyclone centers using ERA5 data every 3 h following Lim and Simmonds (2007). These positions were interpolated to each hour and the latitudinal and longitudinal distances relative to the observations were then calculated. These distances were binned every 200 km within 1000 km of the cyclone centers in order to be able to calculate SW_{down} biases for each model in different sectors relative to cyclone centers.

d. *Interpreting and explaining predictions*

Interpretability is important for providing trust in the machine learning models as well as for gaining insight into how each predictor influences each individual prediction. The primary objective of this study is to predict the Southern Ocean SW_{down} using selected gridscale meteorological predictors from ERA5 (see section 2a). In this study, our focus is to demonstrate the predictive capability of machine learning models trained with surface observations and so we use SHAP (Lundberg and Lee 2017) in this context to ensure that our models are trustworthy and to explore the relationship between our predictors and SW_{down}. The SHAP analysis was done using the best-performing configuration of random forest hyperparameters on a model trained on the full dataset. SHAP was implemented with the *fastshap R* package (Greenwell 2020). This method approximates the contribution that each feature has on each prediction by taking only a subset of predictor permutations. We carried out 100 Monte Carlo repetitions so that the approximate SHAP values were representative. This SHAP analysis results in an estimate of how much each predictive feature changed each individual prediction. To gain a broad understanding of their impact, we aggregated these SHAP values across all predictions to provide the mean absolute impact of each predictive feature. Furthermore, we separated out predictions based on their position relative to the Antarctic Polar Front, to gain a better understanding of how the predictor variables impact changes at higher and lower latitudes. We then calculated the daily averages of each predictor variable and their corresponding SHAP values. This was done to investigate the relationship between the different values of each predictor and the influence that these have on the prediction of SW_{down}. Last, these SHAP values were also aggregated according to their position relative to cyclone centers.

3. Results and discussion

a. *The Southern Ocean shortwave radiation errors in ERA5*

The mean monthly bias in SW_{down} simulated by ERA5 in the Southern Ocean spanning all years that measurements were collected on the RSV *Aurora Australis*, R/V *Investigator*, and at Macquarie Island is shown in Fig. 2a. This demonstrates substantial errors in ERA5's simulation of SW_{down}, which is most prominent in the austral summer, with a mean monthly bias of up to 69 ± 43 W m^{-2} in December, and 73 ± 44 W m^{-2} considering daylight-only hours.

To explore the latitudinal variability of these errors further, the mean bias in the ERA5 SW_{down} was calculated in 2° latitude bands between 44° and 70°S (Fig. 2b) during late austral spring to early austral autumn (October–March). There is a distinct increase in ERA5 SW_{down} bias south of 56°S. To the north of this boundary, the mean daylight ERA5 SW_{down} bias is 29 ± 103 W m^{-2}, while south of 56°S, it is 58 ± 115 W m^{-2}. As discussed in the introduction, addressing this radiation bias is an important and active area of research, which involves integrating new detailed observations of aerosol and cloud properties into improved model parameterizations of cloud droplet and ice formation and precipitation (McFarquhar et al. 2021). The machine learning approach used in this study takes a different approach and only requires observations of SW_{down}.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

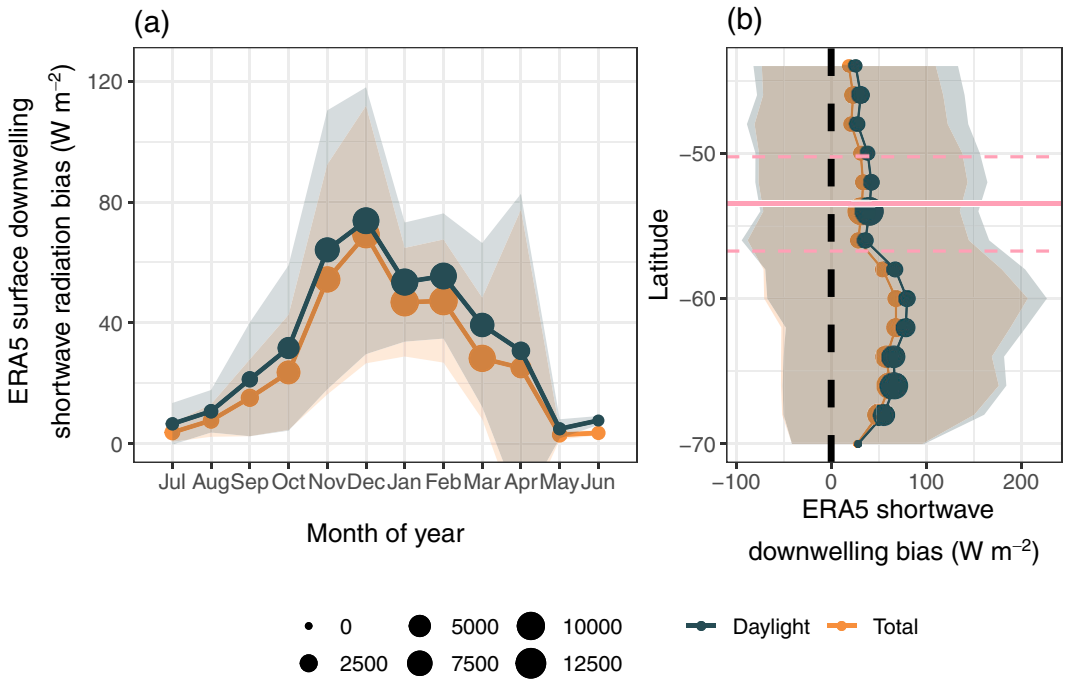

FIG. 2. (a) The MAE in shortwave radiation from ERA5 using the surface observations averaged for each month, and (b) the mean bias in shortwave radiation from ERA5 at different latitudes during October–March. The size of the filled circles is proportional to the number of hours of observations, and the shading represents the standard deviation across months in (a) and 2° latitude bins in (b). Orange colors represent data from all hours throughout the day and night, and dark colors represent daylight hours only, defined as hours when the clear-sky SW_{down} from ERA5 is nonzero. The horizontal solid pink line represents the mean latitude of the Antarctic Polar Front, and the dashed pink lines indicate the standard deviation of the latitudinal position of the Antarctic Polar Front across the longitudes included in this study.

In the following section, we will investigate whether machine learning–based predictions can outperform ERA5.

b. *Model performance*

The models built with both the random forest and XGBoost algorithms show large improvements relative to ERA5's simulation of SW_{down} (see Fig. 3). Consistent with the previous discussion and Fig. 2, the mean bias error in ERA5 across all daylight hours for October–March months for the 10 data splits is 54 ± 4 W m^{-2}. North of the Antarctic Polar Front, ERA5's MBE is 34 ± 4 W m^{-2}, while south of the Antarctic Polar Front it is 64 ± 4 W m^{-2}. This positive mean bias error is not present within the predictions from the generalized linear, random forest, and XGBoost based models. The MBE for the three machine learning–based models is within ±8 of 0 W m^{-2}. This demonstrates that the machine learning–based models do not exhibit the persistent positive bias present within ERA5. However, because positive and negatives biases can cancel out in the MBE, it is necessary for robust evaluation to also consider other performance metrics such as RMSE, MAE and the coefficient of determination, R^2.

Across the 10 splits, the ERA5 SW_{down} MAE and RMSE are 81 ± 3 W m^{-2} and 130 ± 8 W m^{-2}, respectively, with an R^2 of 0.72 ± 0.03. The random forest–based models' performance of SW_{down} MAE and RMSE are 61 ± 3 W m^{-2} and 97 ± 6 W m^{-2}. Relative to ERA5, this is a 25% ± 3% reduction in both the MAE and RMSE metrics. The random forest–based models also have a higher R^2 value of 0.75, a 5% ± 1% increase relative to ERA5.

The XGBoost-based models have a very similar performance to the random forest–based models and outperform ERA5 over all three metrics, also with a 25% ± 2% reduction in MAE, 25% ± 2% reduction in RMSE and a 5% ± 1% increase in R^2. The similarity in performance is not surprising considering both XGBoost and random forest algorithms rely on decision trees. Because XGBoost is a more complicated model (i.e., has a greater number of tunable hyperparameters), we might expect a superior performance if more observations and/or predictors were included. However, in this study, it is apparent that the predictions from the random forest–based models and XGBoost-based models converged, with both being able to capture the underlying relationships between the selected predictors and the surface downwelling shortwave radiation.

The generalized linear models result in more of a mixed performance, with reductions in the MAE and RMSE of 2% ± 3% and 14% ± 4%, but a decrease in the R^2 of 6% ± 2% relative to ERA5. This performance relative to the tree-based methods can likely be attributed to the inability of linear models to capture nonlinear relationships between the gridscale meteorological predictors and surface downwelling SW_{down}.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

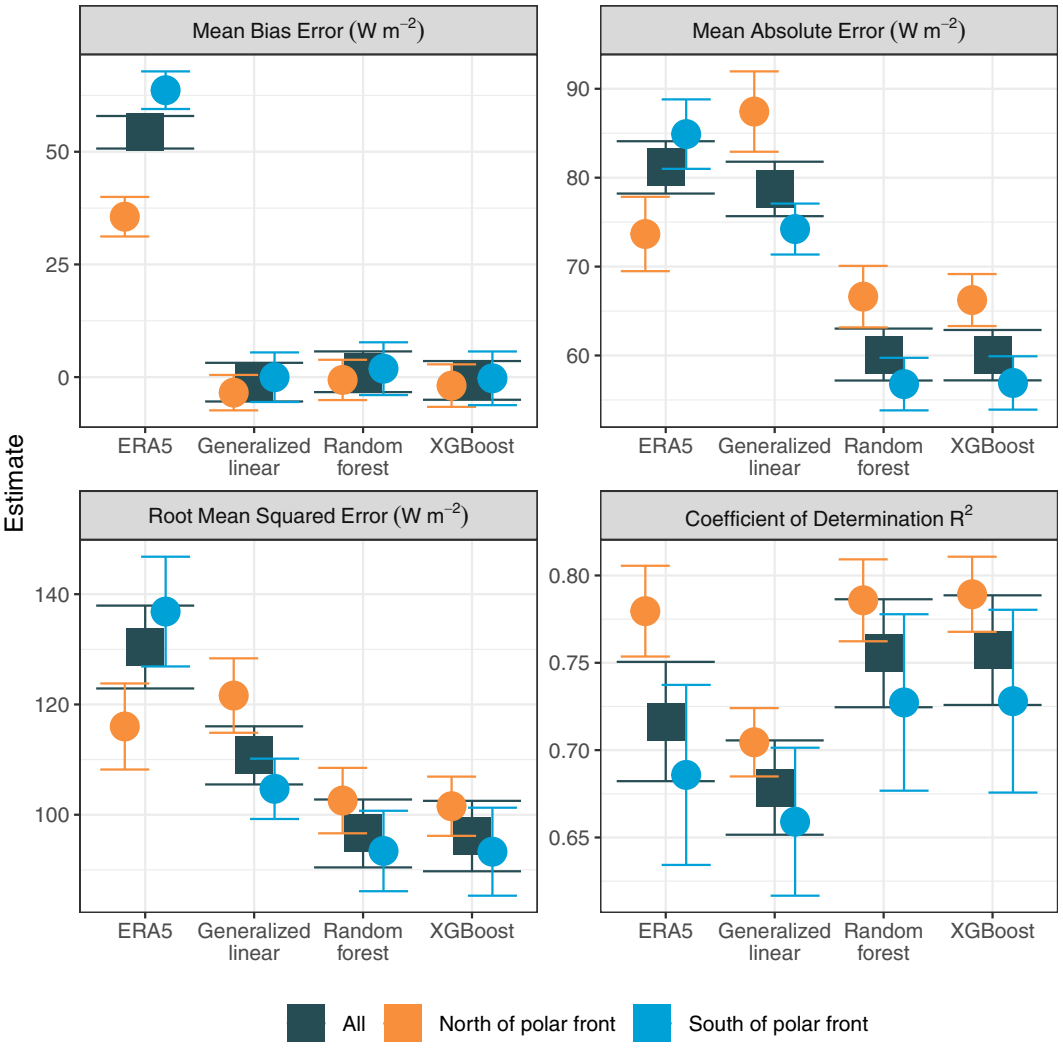

FIG. 3. The MBE, MAE, RMSE, and R^2 in SW_{down} for each model in comparison with surface measurements for October–March and daylight hours. The colored symbols represent the mean performance of each model north and south of the climatological mean position of the Antarctic Polar Front. The black symbols and error bars indicate the mean and standard deviation of each metric across the 10 outer folds.

Comparing the performance of these models for predictions both north and south of the climatological mean position of the Antarctic Polar Front is revealing. While the position of the Antarctic Polar Front varies with longitude, the mean and median latitude of the climatological Antarctic Polar Front is near 54°S. As shown in both Figs. 2 and 3, it is latitudes south of this that show the largest biases in the ERA5 SW_{down}, likely due to deficiencies in simulated cloud properties unique to high latitudes of the Southern Ocean (McFarquhar et al. 2021). South of the Antarctic Polar Front is also where the random forest and XGBoost-based models show the largest improvements in performance. South of the Antarctic Polar Front, the random forest–based models results in a 33% ± 2% reduction in MAE, a 31% ± 2% reduction in RMSE, and a 6% ± 1% increase in R^2 relative to ERA5. Although the random forest–based models also perform better than ERA5 north of the Antarctic Polar Front, these improvements are less pronounced, with a 9% ± 3% reduction in MAE, an 11% ± 2% reduction in RMSE, and a 1% ± 1% increase in R^2. The changes in performance with respect to ERA5 for the XGBoost-based models across the Antarctic Polar Front are nearly the same as the random forest–based models.

Splitting the performance of the models across the Antarctic Polar Front highlights further limitations with the generalized linear models. While south of the Antarctic Polar Front, the generalized linear models outperform ERA5, with an 11% ± 2% reduction in MAE and a 22% ± 3% reduction in RMSE, they do show a 5% ± 2% decrease in R^2. North of the Antarctic Polar Front, the performance of the generalized linear models suffers, with a 19% ± 5% increase in MAE, a 5% ± 5% increase in RMSE, and a 10% ± 2% decrease in R^2.

While it is informative to evaluate ERA5 and the machine learning–based models north and south of Antarctic Polar Front, it is also useful to investigate their performance in different positions relative to cyclone centers. The mean SW_{down}

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

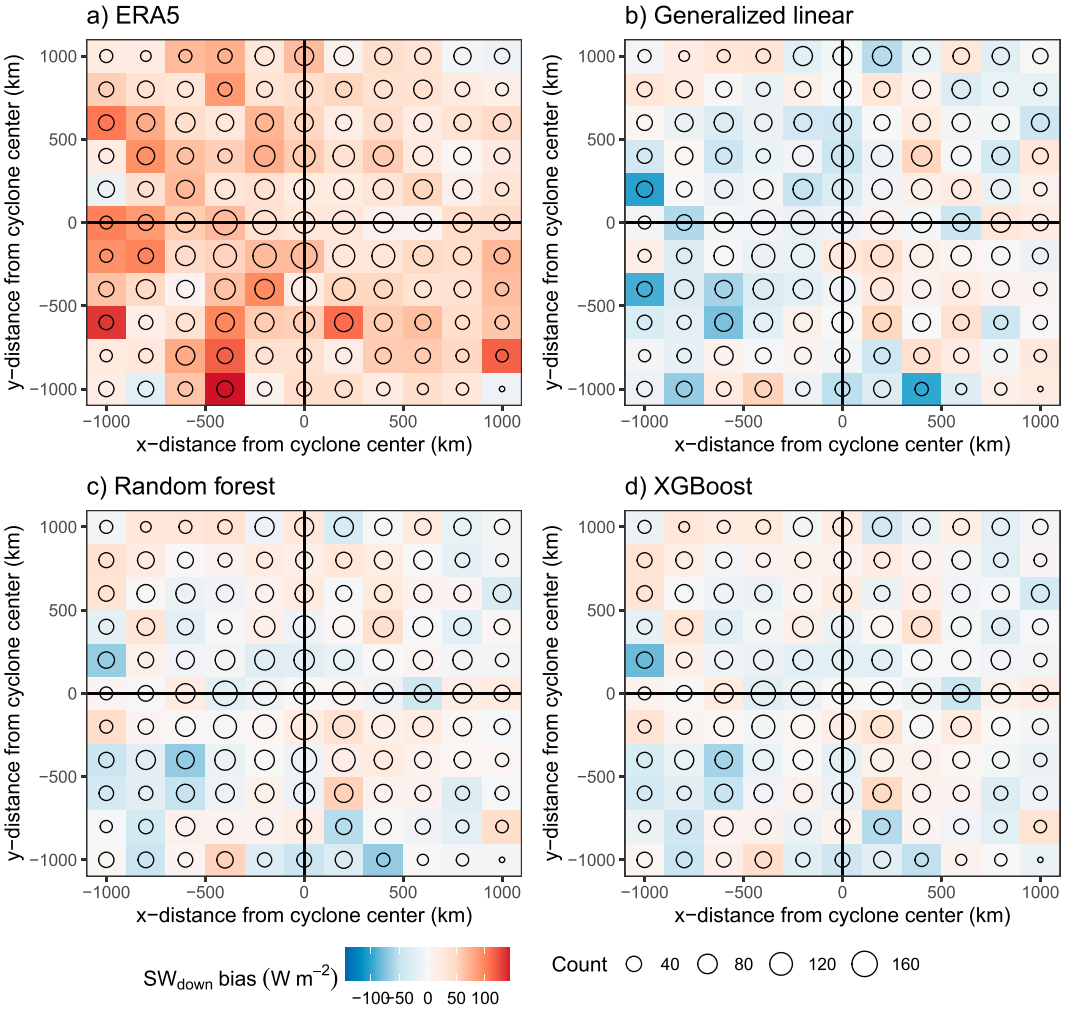

FIG. 4. The mean bias in SW_{down} for (a) ERA5, (b) GLM, (c) random forest, and (d) XGBoost for different positions relative to cyclone centers between 2016 and 2018. The size of the circles corresponds to the number of hourly observations in each position. In the Southern Ocean, the northeast quadrant typically represents warm moist air, whereas the other three quadrants represent cooler, drier air.

bias for each model in different positions within 1000 km of cyclone centers between 2016 and 2018 is shown in Fig. 4. We can see that ERA5 exhibits positive biases in all four quadrants of the cyclone composite. The largest biases occurring in the cold sectors, with average biases in the northwest quadrant of 44 ± 116 W m^{-2}, 59 ± 131 W m^{-2} in the southwest quadrant, and 48 ± 112 W m^{-2} in the southeast quadrant. ERA5 exhibits a smaller but still positive bias of 29 ± 92 W m^{-2} in the northeast quadrant. In contrast, both the random forest and XGBoost based models do not show any consistent pattern across the quadrants, with averages biases less than 9 W m^{-2}. While the persistent positive bias has been eliminated in the generalized linear model predictions, the southwest and northwest quadrants show negative biases of −17 ± 123 W m^{-2} and −22 ± 122 W m^{-2}, respectively. Previous studies have shown that large shortwave errors in climate models in the cyclone cold sectors over the Southern Ocean correspond to periods of high occurrences of supercooled liquid clouds and underestimated cloud liquid water paths (Bodas-Salcedo et al. 2016), although the underlying reasons for these deficiencies can differ across climate models, depending on how they represent clouds. Some models that exhibit positive shortwave biases in cyclone cold sectors partially compensate for this with negative shortwave biases in the warm sector (Bodas-Salcedo et al. 2014). The random forest and XGBoost-based models in this study did not exhibit any compensating biases in the cold and warm cyclone sectors, giving us further confidence in their improved performance relative to ERA5.

To further explore the performance of the three machine learning algorithms against the benchmark ERA5, it is important to explore the relationship between the predictions and their associated residuals to identify any systematic errors. These relationships are shown in Fig. 5. A good performing model would contain residuals centered on zero with a narrow spread across the range of predicted values. As discussed in section 3a and highlighted in Figs. 2–4, ERA5 overpredicts SW_{down}, particularly at high latitudes, cyclone cold sectors,

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

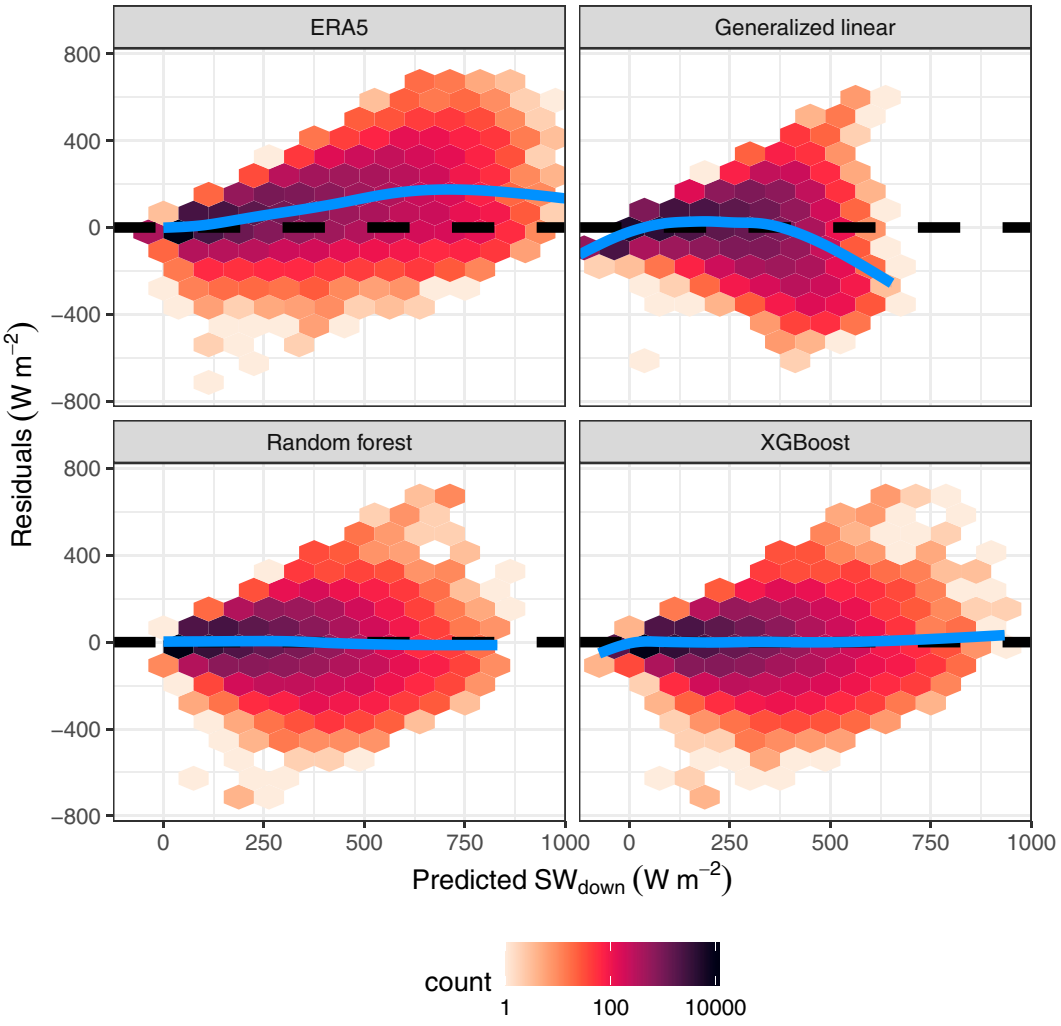

FIG. 5. Heat maps of the prediction residuals of SW_{down} from ERA5, GLM, random forest, and XGBoost for October–March and daylight hours, as a function of the predicted SW_{down}. Blue lines represent the smoothed fit for each model type.

and in austral summer. Figure 5 breaks this down further, showing the tendency for ERA5 to produce positive residual values, shifting to even more positive residuals at higher values of predicted SW_{down}. The residuals for the generalized linear model leave much to be desired. In contrast to the other models, the generalized linear model does not make many predictions of SW_{down} larger than 500 W m^{-2}. The residuals at smaller values of predicted SW_{down} also exhibit a larger spread than the other models and are shifted toward negative values. The distribution of residuals for the random forest and XGBoost models again show better performance than ERA5 and the generalized linear model. Their residuals are centered around zero across the entire range of predicted values. Combined with the overall improvement in the MBE, MAE, RMSE, and R^2 metrics discussed previously and summarized in Fig. 3, these prediction residuals give us further confidence in the reliability of the random forest and XGBoost methods used in this study.

For this study, the most successful model is based on the random forest algorithm, with very similar performance from XGBoost. It is also important to consider the computational cost of tuning and training these models. Due to the higher number of hyperparameters to tune with XGBoost, it generally takes at least as long as random forest to tune (on the order of tens of hours in this study) and train (on the order of minutes in this study). In contrast, the generalized linear models were nearly 100 times faster to tune and train than the random forest–based models in this study. Because the generalized linear models did not offer a significant improvement in performance relative to ERA5, however, the trade-off in computation time for a significantly improved performance from random forest was justified for this dataset. Once these models are trained, the computational time to make predictions is trivial. The time spent training them would therefore be nondetrimental for using them in an operational capacity. The machine learning–based models would only need to be retrained when new versions of the underlying forecast model become available, or if new observational data were to be incorporated for training.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

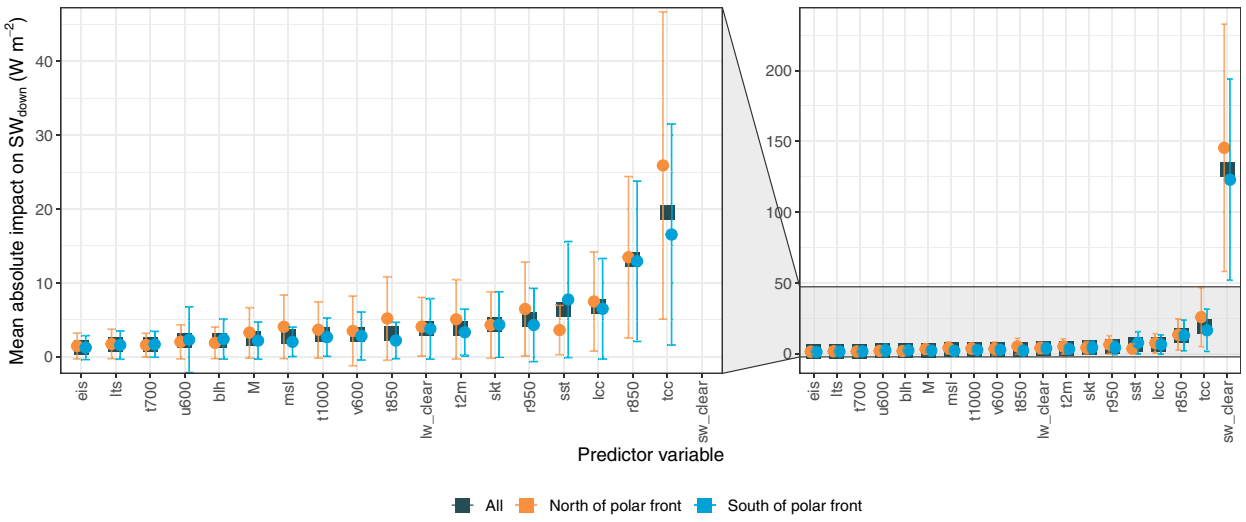

FIG. 6. The mean absolute SHAP value for each predictor variable calculated over the entire dataset as well as for grid points north and south of the Antarctic Polar Front. The detailed names for each variable are described in section 2a.

c. *Predictor contributions*

While the previous section discusses the performance of our machine learning–based models, we now turn our attention to interpreting our predictions using SHAP. The mean SHAP values (i.e., the mean impact that each feature has on the predicted SW_{down}) for each feature over all observations for the best-performing random forest model fitted over the entire dataset is shown in Fig. 6. As expected, information about the clear-sky SW_{down} variable has the largest impact on the SW_{down} prediction, with a mean impact of 130 W m^{-2} (with a standard deviation of the clear-sky SW_{down} SHAP values of ±151 W m^{-2}). This corresponds to an average of 56% ± 59% of the total contribution to the predictions (see supplementary Fig. 2 in the online supplemental material). This clear-sky SW_{down} feature contains combined information about the time of day, latitude, and season. The next most impactful feature was the total cloud cover, changing the model prediction by an average 20 ± 14 W m^{-2} (or 9% ± 7% of the relative change). While the relative humidity at 850 hPa, sea surface temperature, low cloud cover and skin temperature also had marginal impacts on the order of 7–13 W m^{-2}, many of the other features only changed the model prediction by less than 5 W m^{-2}.

The SHAP values calculated for observations either north or south of the Antarctic Polar Front are similar, but with some notable differences. North of the Antarctic Polar Front, the SHAP value of the clear-sky SW_{down} was higher (145 ± 141 W m^{-2}), which can be expected due to the higher solar zenith angles at lower latitudes. The SHAP values for total cloud cover were lower and for sea surface temperature were higher south of the Antarctic Polar Front (16 ± 12 W m^{-2} and 8 ± 5 W m^{-2}, respectively). A sharp gradient in sea surface temperature exists over the Antarctic Polar Front, and these differences in the SHAP values indicate that the random forest algorithm was using this information to constrain the SW_{down} predictions. To explore this further we can investigate local, rather than global, SHAP values for each predictor.

The SHAP analysis also allows to investigate the impact that each predictor has on every individual prediction. The so-called SHAP dependence plots are shown in Fig. 7 (daily averages are shown for clarity), with data north and south of the climatological mean position of the Antarctic Polar Front colored yellow and blue, respectively. It is useful to recall that the SHAP values tell us how much knowledge of each predictive feature changes our prediction away from the "base" (i.e., average) prediction. These SHAP dependence plots are consistent with physically based assumptions about the nature of the relationship between each predictor and the surface SW_{down}. As expected, higher values of clear-sky SW_{down} drive the predictions of SW_{down} to higher values (Fig. 7a). Conversely, higher values of total cloud cover drive the prediction of SW_{down} to lower values, dropping sharply as the total cloud cover approaches 1 (Fig. 7b). Similar relationships can be observed for relative humidity at 850 hPa (Figs. 7c,f) and low cloud cover (Fig. 7d). Interestingly, while sea surface temperature values spanning from 275 to 295 K have very little influence on the SW_{down} prediction, cool sea surface temperatures (<275 K) contribute to higher predictions of SW_{down}. These cool sea surface temperatures occur south of the Antarctic Polar Front and contribute toward the good performance of the random forest model at these higher latitudes. Similarly, high values of sea ice concentration were associated with higher predictions of SW_{down} (Fig. 7m), suggesting a reduced surface shortwave cloud radiative effect (CRE) over sea ice.

Other predictors that can be linked to cloud cover over the Southern Ocean, such as the marine cold-air outbreak *M* parameter (Fig. 7o) and the lower-tropospheric stability (Fig. 7s) (Naud et al. 2020), influence the prediction of SW_{down} in a physically realistic way. Higher values of lower-tropospheric stability are generally associated with more cloud cover and hence lower surface SW_{down}, while the opposite is true for the

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

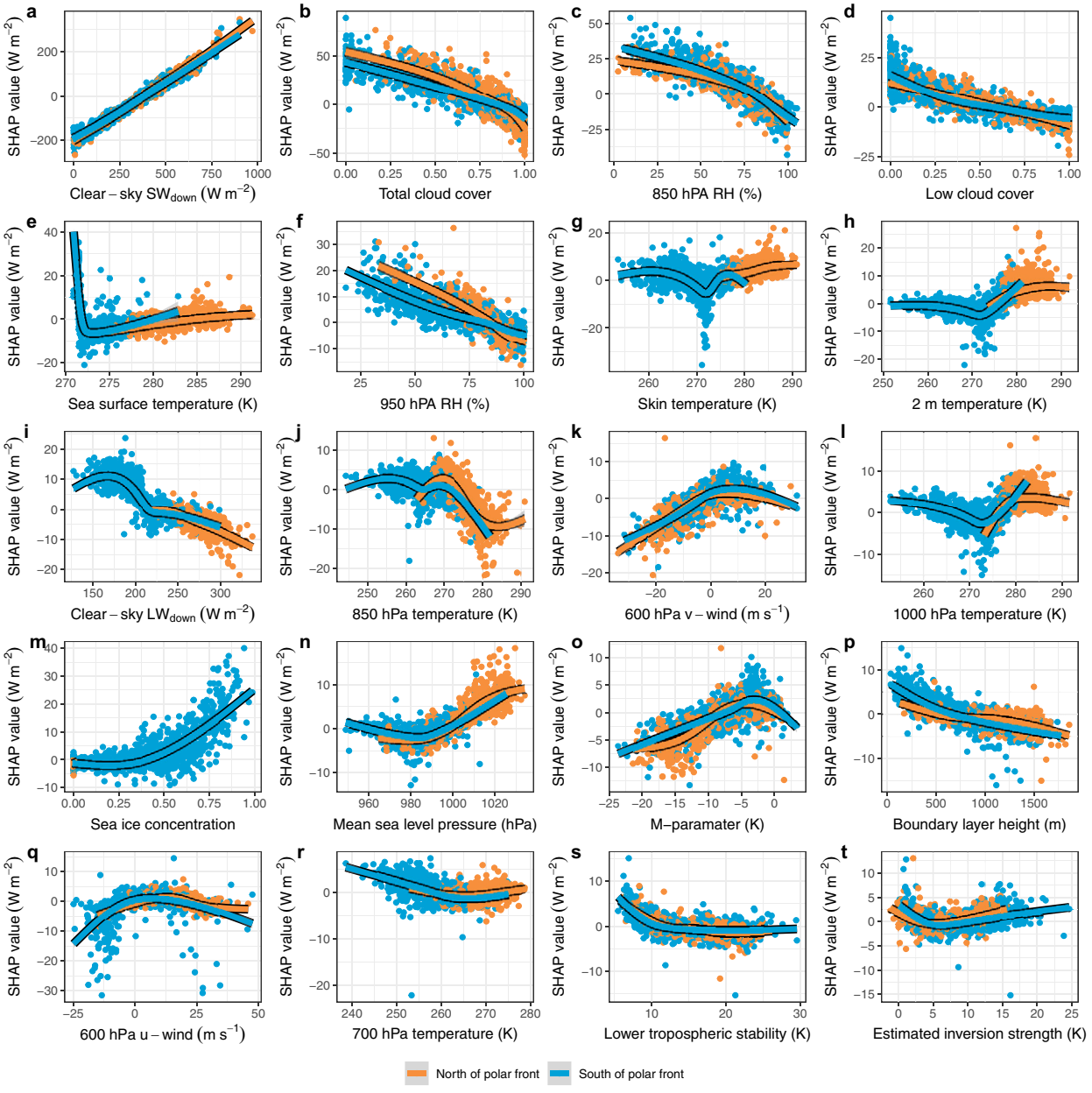

FIG. 7. The SHAP dependence plots for each predictor. Each SHAP value represents the mean impact that each predictor had on daily averaged predictions of SW_{down}. Orange and blue points represent predictions made north and south of the climatological mean position of the Antarctic Polar Front, respectively. The plots are arranged from the most influential predictor in the top left to the least influential predictor in the bottom right.

M parameter. The estimated inversion strength (Fig. 7t) was the least influential predictor. While it is also related to cloud cover (Wood and Bretherton 2006; Naud et al. 2020), it is apparent in this study that any extra information knowledge of the estimated inversion strength provides is negligible, and likely already accounted for by other predictors (e.g., total and low cloud cover, *M* parameter, lower-tropospheric stability, as well as the features used in its calculation). Overall, only a small number of predictor variables are responsible for most of the impact on predicting SW_{down} in this study. Physically, there are complex explanations on how radiation and cloud property vary across the Antarctic Polar Front, including the presence of supercooled liquid water (McFarquhar et al. 2021; Mace et al. 2021a). Despite that complexity, it is evident that differences in key predictor variables used in this study, such as the clear-sky SW_{down}, total cloud cover, and sea surface temperature, contain the most information about how SW_{down} varies across the Antarctic Polar Front.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

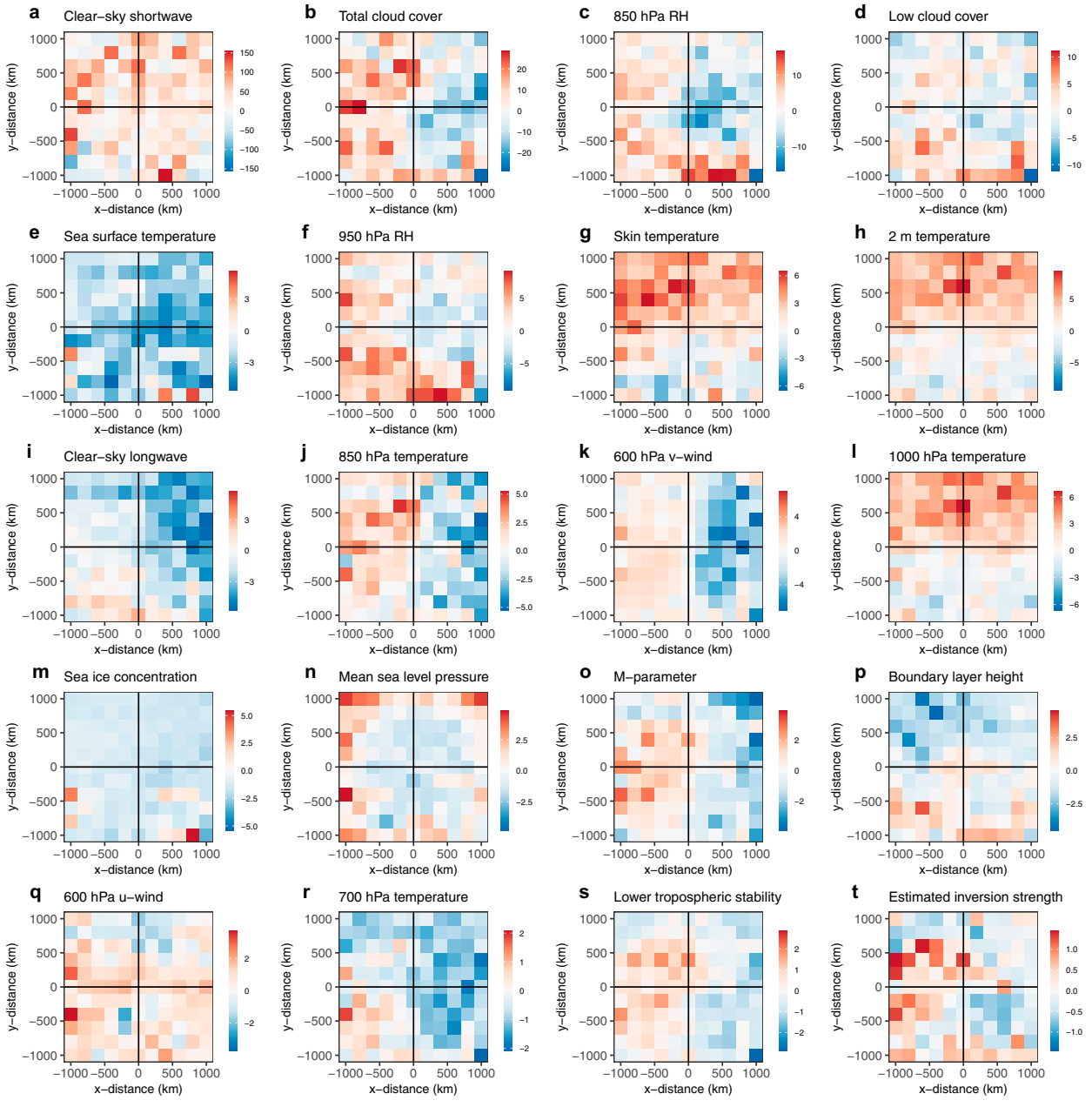

FIG. 8. Cyclone composites of the mean predictor contributions (i.e., SHAP values) for each ERA5 predictor between 2016 and 2018. All units are in watts per meter squared. The number of data points for each square is the same as those shown in Fig. 4. The plots are arranged from the most influential predictor in the top left to the least influential predictor in the bottom right.

We emphasize that the global and local SHAP values for each predictor are an indication of how much each feature impacts the *prediction* of the underlying model, given the coalition of predictors. How meaningful these SHAP values are depends on the performance of the underlying model. Excluding (or including new) predictors will also change the SHAP values of other predictors. The degree to which these SHAP values can change will depend on the correlation and relationship between predictors. The relationships between each predictor and its corresponding SHAP values should therefore not be confused for the true relationship between each predictor and SW_{down}. However, investigating the relationships between the predictor values and their corresponding SHAP values and ensuring that these relationships are qualitatively consistent with an expected physical relationship is still useful. To furthermore demonstrate this, we have aggregated the mean SHAP values for each predictor variable according to their relative distance to cyclone centers. These are shown in Fig. 8, and for additional context the cyclone composites of the predictor variable values themselves are shown in supplementary Fig. 3 in the online supplemental material.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

As highlighted in Fig. 4, ERA5 displays positive SW_{down} biases in all four quadrants around cyclone centers, with the largest biases occurring in the northwest, southwest, and southeast, which loosely correspond to cyclone cold sectors, consistent with other Earth system models over the Southern Ocean (Bodas-Salcedo et al. 2014). It is in these three quadrants that we see the largest improvements from our random forest–based model. Exploring the SHAP values for each predictor variable in these quadrants provides additional trust in our predictions. For example, we can see that variables that are most strongly linked with cloud properties (i.e., total cloud cover, low cloud cover, relative humidity at both 850 and 950 hPa) show consistent patterns in their SHAP contributions, with more positive values in the cyclone cold (dry) sectors. This highlights that drier air with lower values of cloud cover in this cold dry sector drives the prediction of SW_{down} toward higher values and is consistent with the SHAP dependence plots shown in Fig. 7. Similarly, temperature-related predictor variables at lower pressures (i.e., higher altitudes) show a similar pattern. Other predictor variables related to temperature near the surface (temperature at 2 m, temperature at 1000 hPa, and skin temperature), all show expected patterns with the northern (equatorward) quadrants displaying more positive SHAP values. Other predictors, such as the meridional wind speed at 600 hPa, sea ice concentration, and mean sea level pressure, all display patterns that are consistent with our expectations. For example, sea ice concentrations above zero only occur at high latitudes near Antarctica and are associated with more positive SHAP values (i.e., higher values of predicted SW_{down}), and this is reflected in the southern quadrants. As shown in Fig. 7, low values of mean sea level pressure are associated with lower SHAP values, and this is evident at distances closest to the cyclone centers. Our most influential predictor, the clear-sky SW_{down} did not show any obvious spatial pattern across the cold and warm sectors of cyclones, although the two northern quadrants do generally display higher SHAP values, which is expected at lower latitudes. Overall, the SHAP analysis, visualized by the importance and dependence plots and the cyclone composites shown here demonstrate that our best machine learning model (based on random forest) is trustworthy.

d. *Demonstrated application*

The improved prediction of SW_{down} relative to ERA5 justifies the potential application of machine learning–based methods. The robust performance of the best-performing RF model makes it a good candidate to predict SW_{down} across the Southern Ocean. To demonstrate this, we applied our best-performing random forest model (1000 trees, 8 variables randomly sampled as candidates for each node split, and a minimum of 10 observations required to split a node further) over a large domain and calculated the effective shortwave cloud radiative effect (SW_{CRE}). The model was trained on all of the available daylight data between October and March. ERA5 features were taken during all October–March months every 6 h from 0300 UTC each day (to reduce computation time) from 44° to 69.43°S and from 49.86° to 159.46°E between 2016 and 2018. The trained model was then used to predict shortwave downwelling radiation over this domain and period. Predictions during nondaylight hours when the ERA5 clear-sky SW_{down} were zero were then physically constrained to zero. The net shortwave cloud radiative effect was then calculated for both ERA5 and every random forest–based prediction. This was done by calculating the difference between the shortwave downwelling radiation and the clear-sky shortwave downwelling radiation and multiplying this by $(1 - a)$, where a is the estimated albedo. For open ocean, an albedo value of 0.055 was used (Fairall et al. 2008). Albedo values for nonzero sea ice concentrations were scaled linearly from 0.055 up to 0.8 for 100% sea ice concentration (Brandt et al. 2005). The simulated clear-sky shortwave downwelling radiation from ERA5 was taken for this calculation for both the ERA5 and machine learning–based predictions.

Figure 9 shows the calculated mean net shortwave cloud radiative effect over the East Antarctic sector of the Southern Ocean for austral summer months from 2016 to 2018 inclusive for ERA5 (Fig. 9a) and the best-performing RF model (Fig. 9b). Figure 9c shows the difference between ERA5 and the RF model. The calculated SW_{CRE} from ERA5 and the RF model show stark differences south of the Antarctic Polar Front. On average, the random forest–based model net SW_{CRE} was 56 ± 15 W m^{-2} less than ERA5 south of the Antarctic Polar Front, while north of the Antarctic Polar Front, the random forest–based model was 30 ± 8 W m^{-2} lower than ERA5. The low values of SW_{CRE} near the Antarctic coast are a result of the presence of sea ice, which has a high albedo, therefore reducing the net SW_{CRE}. The SHAP analysis shown in Fig. 6 indicates that these differences in the RF model are mostly due to differences in the clear-sky SW_{down} (more impactful north of the Antarctic Polar Front), total cloud cover (more impactful north of the Antarctic Polar Front) and sea surface temperature (more impactful south of the Antarctic Polar Front). From the evaluation of ERA5's SW_{down} using surface observations, we know this high latitude region exhibits the largest biases (Fig. 2b).

Resolving these biases in ERA5 and other Earth system models and reanalyses will require a deeper understanding and implementation of cloud droplet and ice crystal formation across different latitudes of the Southern Ocean where cloud condensation nuclei (Humphries et al. 2021) and ice nucleating particle (McCluskey et al. 2018) sources and processes likely differ. Variability in sea surface temperatures in the Southern Ocean have also been associated with variability in cloud microphysical and precipitation properties (Huang et al. 2016), and causal links between these dynamics needs to be further explored. Until this is achieved by integrating new observations with model development, it is necessary to search for alternative approaches. While significant efforts should still be made to improve our physical understanding of cloud radiative effects over the Southern Ocean and the implementation of this into climate models, it will be important to recognize, and integrate, machine learning methods in this pursuit. Our study demonstrates that a machine learning–based approach that uses potentially biased features from ERA5, but trained with observations of SW_{down} could provide an additional tool to address the Southern Ocean surface shortwave radiation problem, particularly in reanalyses products.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

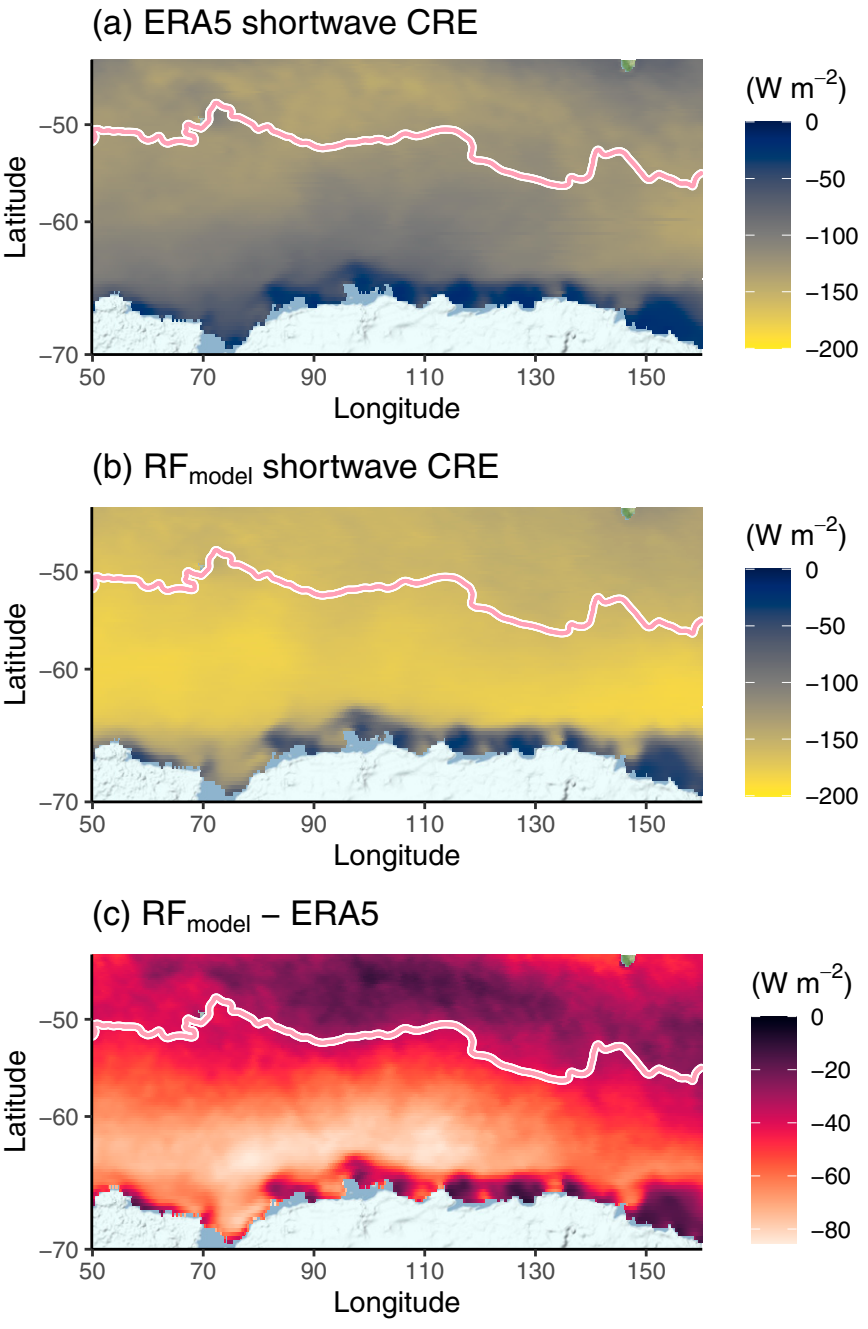

FIG. 9. (a) The mean net shortwave CRE calculated from ERA5, (b) the mean shortwave CRE calculated from the best-performing random forest model, and (c) the difference between (a) and (b). The pink line indicates the mean position of the Antarctic Polar Front. Low values near the coast are a result of high albedo over sea ice.

4. Conclusions and outlook

a. Summary of findings

We built models with machine learning algorithms to predict surface downwelling shortwave radiation, SW_{down}, in summer over the Southern Ocean using a set of gridscale meteorological predictor variables from ERA5. These models were trained with surface-based observations of SW_{down} spanning 25 years between the late austral spring to early austral autumn months from October to March using a 10-fold shuffle split for training, tuning, and testing. We were able to predict SW_{down} with an average MAE and RMSE that is 25% lower than ERA5's own simulation of SW_{down}, in comparison with observations. This improvement was achieved with a model

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

built using the random forest algorithm. The XGBoost-based model had a similar performance to the RF model, offering a mean 25% reduction in RMSE and MAE relative to ERA5. Both the random forest and XGBoost-based models were also able to explain up to 5% more of the variance in SW_{down} than ERA5. Generalized linear regression models offered slight improvements in the errors relative to ERA5 with an average reduction in the MAE and RMSE of 2% and 14%, respectively, however, were not able to explain as much of the variance as ERA5 or the RF and XGBoost based models. Furthermore, while ERA5 contains a positive mean bias error of 54 W m^{-2}, this has been nearly eliminated for the three machine learning based models, with the MBE for the test sets from the 10 data splits being within ± 8 W m^{-2} of 0 W m^{-2}.

The shuffle-split and nested-cross-validation process that we used was important in yielding the most unbiased assessment of the performance of each model and to account for the variability in performance across different splits in the training and testing sets. Overestimating performance without due diligence in the training and testing stages of a machine learning workflow could have severe consequences if machine learning–based models are applied in practice. The success of our machine learning–based models in this study hinged on their performance relative to ERA5, therefore these steps were crucial.

The machine learning based models offered the largest improvements south of the Antarctic Polar Front and in the cold sectors of cyclones, where ERA5 contained the largest biases in SW_{down}. This is likely due to inadequate parameterizations of cloud formation and precipitation properties leading to biases in cloud radiative properties. Nonetheless, this demonstrates that machine learning algorithms can be trained on potentially biased predictors using real observations to make more accurate predictions.

Ultimately the power of machine learning comes from having sufficient data to train and test models. The success in more accurately simulating SW_{down} in this study can be attributed to the long-term measurements collected on the RSV *Aurora Australis*, R/V *Investigator*, and at Macquarie Island. Utilizing or establishing other existing long-term datasets of cloud properties in conjunction with techniques from the rapidly evolving field of machine learning could significantly improve our ability to understand and simulate these complex atmospheric processes and systems in regions like the Southern Ocean. Using machine learning tools to improve reanalysis products is a useful starting point in this pursuit.

b. *Recommendations for future work*

We have demonstrated that machine learning techniques can be a useful tool for improving the simulation of SW_{down} over the Southern Ocean in a reanalysis. The final trained model using the random forest algorithm could be implemented over the Southern Ocean to offer a prediction of SW_{down} as a new product. Although ERA5 variables were used as the predictors in the trained model, these could easily be replaced with meteorological data from other models (e.g., other reanalyses, numerical weather prediction models, or general circulation models). Furthermore, there is significant scope to expand the method outlined in this paper to include other regions and seasons. By improving the simulation of SW_{down} using mostly thermodynamic properties from ERA5, the trained model has implicitly learned the shortwave cloud radiative effect (as demonstrated in Fig. 9). It is therefore also likely that this method could also be extended to predict the longwave cloud radiative effect and cloud properties such as cloud optical depth, liquid water path, and/or cloud-base height. An alternative approach to using machine learning to predict these quantities themselves could be instead to predict biases in these quantities in reanalysis, numerical weather prediction models or general circulation models. In regions and periods when there are sufficient measurements to use as training data, this could be a useful tool in correcting, understanding, and identifying model biases. This will be explored in future work using a range of reanalysis models, as well as numerical weather prediction and general circulation versions of the Australian Community Climate and Earth system simulator.

In this work we used generalized linear models, random forest and XGBoost due to their relative ease of implementation, previously demonstrated success and ability to explore the relationships between the predictor and predicted variables. This method could be extended on larger temporal or spatial scales and applied to produce a new product. There are several possibilities that could potentially further improve performance, although this would come at the cost of interpretability. In this study we only used a small number of selected predictor variables that were likely to influence the shortwave downwelling radiation. It would be possible to use all of the thermodynamic and cloud-related variables across all pressure levels as predictors, with the possibility of using dimensionality reduction (e.g., principal component analysis). Furthermore, neural network techniques, while more difficult to interpret, have been demonstrated to achieve high predictive capabilities and potentially remove the need to do to feature selection, as was done in this study.

While this study used surface-based observations of SW_{down}, future work could use satellite-based products as the "truth." While some of these satellite-based products also exhibit small biases themselves over the Southern Ocean (Hinkelman and Marchand 2020), the advantage would be a significant expansion of the spatial coverage and sample size. The primary motivation for our study was to demonstrate the utility and performance of a data-driven approach to parameterizing surface downwelling shortwave radiation in the ERA5 reanalysis. To this end, we limited our set of predictor variables to those already included as output from ERA5 even though some of these are prone to biases themselves (e.g., cloud cover, which is not assimilated from observations). The method we have used here could be extended to include satellite observations of a variety of cloud properties, rather than their simulated counterpart. Furthermore, in addition to using machine learning to predict surface shortwave radiation, as done here, it could be used to predict these cloud properties as well, using a set of thermodynamic predictor variables. This would open the door to using a machine learning approach to explore how cloud properties might be perturbed with changes in thermodynamics that are to be expected with climate change.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

Our study is limited to demonstrating that a more accurate prediction of surface shortwave downwelling radiation is possible in a reanalysis using machine learning. Applying this technique to produce a data-driven parameterization would be valuable in applications that rely on accurate reanalysis products to drive other Earth system models that do not simulate the atmosphere (e.g., an ocean or biogeochemical model). Furthermore, because we have robustly evaluated the performance of our machine learning–based models and show large improvements relative to ERA5, predicted surface shortwave radiation from a data-driven approach as shown here could be a valuable product in evaluating other Earth system models (e.g., historic runs of climate models) or satellite-based products that measure top-of-atmosphere radiative fluxes. Finally, a data-driven parameterization as shown in this study could be implemented in a weather forecast model to refresh the radiation efficiently and regularly. These possibilities will be explored in future work.

Acknowledgments. This project received grant funding from the Australian government as part of the Antarctic Science Collaboration Initiative program. The Australian Antarctic Program Partnership is led by the University of Tasmania and includes the Australian Antarctic Division, CSIRO Oceans and Atmosphere, Geoscience Australia, the Bureau of Meteorology, the Tasmanian state government, and Australia's Integrated Marine Observing System. The positions of Southern Ocean cyclone centers were provided by Yi Huang. Technical support and logistical support for the Macquarie Island field experiment were provided by the AAD through Australian Antarctic Science project 4292, and we thank Terry Egan, Nick Cartwright, and Ken Barrett for their assistance. Macquarie Island shortwave radiation data were obtained from the Atmospheric Radiation Measurement (ARM) program sponsored by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, and Climate and Environmental Sciences Division.

Data availability statement. ERA5 data are publicly available from the Copernicus Climate Change Service (C3S; https://cds.climate.copernicus.eu/cdsapp#!/home/), *Aurora Australis* underway data are publicly available from the Australian Antarctic Data Centre (AADC; https://data.aad.gov.au), R/V *Investigator* underway data are publicly available from the CSIRO Marlin portal (https://marlin.csiro.au/), and measurements from Macquarie Island are available from ARM (https://adc.arm.gov/).

APPENDIX

Acronym List

ARM	Atmospheric Radiation Measurement
blh	Boundary layer height
CRE	Cloud radiative effect
eis	Estimated inversion strength
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	Fifth major global reanalysis produced by ECMWF
GLM	Generalized linear model
lcc	Low cloud cover
lts	Lower-tropospheric stability
lw_clear	Longwave clear sky
M	Marine cold-air outbreak parameter
MAE	Mean absolute error
MBE	Mean bias error
MERRA-2	Modern-Era Retrospective Analysis for Research and Applications, version 2
mslp	Mean sea level pressure
r850	Relative humidity at 850 hPa
r950	Relative humidity at 950 hPa
RMSE	Root-mean-square error
RSV	Research and Supply Vessel
R/V	Research Vessel
SHAP	Shapley additive explanations
siconc	Sea ice concentration
skt	Skin temperature
sst	Sea surface temperature
sw_clear	Shortwave clear sky
SW_{CRE}	Surface shortwave cloud radiative effect
SW_{down}	Surface shortwave downwelling radiation
t1000	Temperature at 1000 hPa
t2m	Temperature at 2 m
t700	Temperature at 700 hPa
t850	Temperature at 850 hPa
tcc	Total cloud cover
u600	Zonal wind speed at 600 hPa
v600	Meridional wind speed at 600 hPa

REFERENCES

Beucler, T., I. Ebert-Uphoff, S. Rasp, M. Pritchard, and P. Gentine, 2021: Machine learning for clouds and climate (invited chapter for the AGU Geophysical Monograph Series "Clouds and Climate"). ESS Open Archive, 10506925.1, https://doi.org/10.1002/essoar.10506925.1.

Bodas-Salcedo, A., and Coauthors, 2014: Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models. *J. Climate*, **27**, 41–56, https://doi.org/10.1175/JCLI-D-13-00169.1.

——, T. Andrews, A. V. Karmalkar, and M. A. Ringer, 2016: Cloud liquid water path and radiative feedbacks over the Southern Ocean. *Geophys. Res. Lett.*, **43**, 10 938–10 946, https://doi.org/10.1002/2016GL070770.

Boehmke, B., and B. M. Greenwell, 2019: *Hands-on Machine Learning with R*. CRC Press, 488 pp.

Brandt, R. E., S. G. Warren, A. P. Worby, and T. C. Grenfell, 2005: Surface albedo of the Antarctic sea ice zone. *J. Climate*, **18**, 3606–3622, https://doi.org/10.1175/JCLI3489.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Ceppi, P., M. D. Zelinka, and D. L. Hartmann, 2014: The response of the Southern Hemispheric eddy-driven jet to future changes in shortwave radiation in CMIP5. *Geophys. Res. Lett.*, **41**, 3244–3250, https://doi.org/10.1002/2014GL060043.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, ACM, 785–794, https://doi.org/10.1145/2939672.2939785.

Dueben, P. D., and P. Bauer, 2018: Challenges and design choices for global weather and climate models based on machine learning. *Geosci. Model Dev.*, **11**, 3999–4009, https://doi.org/10.5194/gmd-11-3999-2018.

Fairall, C. W., T. Uttal, D. Hazen, J. Hare, M. F. Cronin, N. Bond, and D. E. Veron, 2008: Observations of cloud, radiation, and surface forcing in the equatorial eastern Pacific. *J. Climate*, **21**, 655–673, https://doi.org/10.1175/2007JCLI1757.1.

Fleming, S. W., J. R. Watson, A. Ellenson, A. J. Cannon, and V. C. Vesselinov, 2021: Machine learning in Earth and environmental science requires education and research policy reforms. *Nat. Geosci.*, **14**, 878–880, https://doi.org/10.1038/s41561-021-00865-3.

Fossum, K. N., and Coauthors, 2018: Summertime primary and secondary contributions to Southern Ocean cloud condensation nuclei. *Sci. Rep.*, **8**, 13844, https://doi.org/10.1038/s41598-018-32047-4.

Freeman, N. M., N. S. Lovenduski, and P. R. Gent, 2016: Temporal variability in the Antarctic Polar Front (2002–2014). *J. Geophys. Res. Oceans*, **121**, 7263–7276, https://doi.org/10.1002/2016JC012145.

Fuchs, J., J. Cermak, and H. Andersen, 2018: Building a cloud in the southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning. *Atmos. Chem. Phys.*, **18**, 16 537–16 552, https://doi.org/10.5194/acp-18-16537-2018.

Géron, A., 2019: *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 718 pp.

Grange, S. K., and D. C. Carslaw, 2019: Using meteorological normalisation to detect interventions in air quality time series. *Sci. Total Environ.*, **653**, 578–588, https://doi.org/10.1016/j.scitotenv.2018.10.344.

——, ——, A. C. Lewis, E. Boleti, and C. Hueglin, 2018: Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. *Atmos. Chem. Phys.*, **18**, 6223–6239, https://doi.org/10.5194/acp-18-6223-2018.

Greenwell, B., 2020: fastshap: Fast approximate Shapley values. R Project, https://CRAN.R-project.org/package=fastshap.

Guyot, A., A. Protat, S. P. Alexander, A. R. Klekociuk, P. Kuma, and A. McDonald, 2022: Detection of supercooled liquid water clouds with ceilometers: Development and evaluation of deterministic and data-driven retrievals. *Atmos. Meas. Tech.*, **15**, 3663–3681, https://doi.org/10.5194/amt-15-3663-2022.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hinkelman, L. M., and R. Marchand, 2020: Evaluation of CERES and CloudSat surface radiative fluxes over Macquarie Island, the Southern Ocean. *Earth Space Sci.*, **7**, e2020EA001224, https://doi.org/10.1029/2020EA001224.

Hooker, G., and L. Mentch, 2019: Please stop permuting features: An explanation and alternatives. arXiv, 1905.03151v2, https://arxiv.org/abs/1905.03151.

Huang, Y., S. T. Siems, M. J. Manton, D. Rosenfeld, R. Marchand, G. M. McFarquhar, and A. Protat, 2016: What is the role of sea surface temperature in modulating cloud and precipitation properties over the Southern Ocean? *J. Climate*, **29**, 7453–7476, https://doi.org/10.1175/JCLI-D-15-0768.1.

Humphries, R. S., and Coauthors, 2021: Southern Ocean latitudinal gradients of cloud condensation nuclei. *Atmos. Chem. Phys.*, **21**, 12 757–12 782, https://doi.org/10.5194/acp-21-12757-2021.

Hyder, P., and Coauthors, 2018: Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nat. Commun.*, **9**, 3625, https://doi.org/10.1038/s41467-018-05634-2.

Karpatne, A., and Coauthors, 2017: Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.*, **29**, 2318–2331, https://doi.org10.1109/TKDE.2017.2720168.

Krasnopolsky, V. M., M. S. Fox-Rabinovitz, and A. A. Belochitski, 2013: Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model. *Adv. Artif. Neural Syst.*, **2013**, 485913, https://doi.org/10.1155/2013/485913.

Kremser, S., and Coauthors, 2021: Southern Ocean cloud and aerosol data: A compilation of measurements from the 2018 Southern Ocean Ross Sea marine ecosystems and environment voyage. *Earth Syst. Sci. Data*, **13**, 3115–3153, https://doi.org/10.5194/essd-13-3115-2021.

Krüger, O., and H. Graßl, 2011: Southern Ocean phytoplankton increases cloud albedo and reduces precipitation. *Geophys. Res. Lett.*, **38**, L08809, https://doi.org/10.1029/2011GL047116.

Kuhn, M., and H. Wickham, 2020: Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles. Tidymodels, https://www.tidymodels.org.

——, and J. Silge, 2022: *Tidy Modeling with R*. O'Reilly Media, 381 pp.

——, and D. Vaughan, 2022: Parsnip: A common API to modeling and analysis functions. R Project Doc., 80 pp., https://cran.r-project.org/web/packages/parsnip/parsnip.pdf.

Kuma, P., and Coauthors, 2020: Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations. *Atmos. Chem. Phys.*, **20**, 6607–6630, https://doi.org/10.5194/acp-20-6607-2020.

Lagerquist, R., D. Turner, I. Ebert-Uphoff, J. Stewart, and V. Hagerty, 2021: Using deep learning to emulate and accelerate a radiative-transfer model. *J. Atmos. Oceanic Technol.*, **38**, 1673–1696, https://doi.org/10.1175/JTECH-D-21-0007.1.

Lang, F., Y. Huang, S. T. Siems, and M. J. Manton, 2018: Characteristics of the marine atmospheric boundary layer over the Southern Ocean in response to the synoptic forcing. *J. Geophys. Res. Atmos.*, **123**, 7799–7820, https://doi.org/10.1029/2018JD028700.

Lantz, B., 2019: *Machine Learning with R: Expert Techniques for Predictive Modeling*. Packt Publishing, 437 pp.

Lenaerts, J. T. M., K. Van Tricht, S. Lhermitte, and T. S. L'Ecuyer, 2017: Polar clouds and radiation in satellite observations, reanalyses, and climate models. *Geophys. Res. Lett.*, **44**, 3355–3364, https://doi.org/10.1002/2016GL072242.

Lim, E.-P., and I. Simmonds, 2007: Southern Hemisphere winter extratropical cyclone characteristics and vertical organization observed with the ERA-40 data in 1979–2001. *J. Climate*, **20**, 2675–2690, https://doi.org/10.1175/JCLI4135.1.

Lundberg, S., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. arXiv, 1705.07874v2, https://doi.org/10.48550/arXiv.1705.07874.

Mace, G. G., A. Protat, and S. Benson, 2021a: Mixed-phase clouds over the Southern Ocean as observed from satellite and surface based lidar and radar. *J. Geophys. Res. Atmos.*, **126**, e2021JD034569, https://doi.org/10.1029/2021JD034569.

——, and Coauthors, 2021b: Southern Ocean cloud properties derived from CAPRICORN and MARCUS data. *J.*

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC

Geophys. Res. Atmos., **126**, e2020JD033368, https://doi.org/10.1029/2020JD033368.

Mallet, M. D., 2021: Meteorological normalisation of PM_{10} using machine learning reveals distinct increases of nearby source emissions in the Australian mining town of Moranbah. *Atmos. Pollut. Res.*, **12**, 23–35, https://doi.org/10.1016/j.apr.2020.08.001.

McCluskey, C. S., and Coauthors, 2018: Observations of ice nucleating particles over Southern Ocean waters. *Geophys. Res. Lett.*, **45**, 11 989–11 997, https://doi.org/10.1029/2018GL079981.

McFarquhar, G. M., and Coauthors, 2021: Observations of clouds, aerosols, precipitation, and surface radiation over the Southern Ocean: An overview of CAPRICORN, MARCUS, MICRE, and SOCRATES. *Bull. Amer. Meteor. Soc.*, **102**, E894–E928, https://doi.org/10.1175/BAMS-D-20-0132.1.

Miller, T., 2019: Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, **267**, 1–38, https://doi.org/10.1016/j.artint.2018.07.007.

Molnar, C., 2020: *Interpretable Machine Learning*. Lulu, 328 pp.

Morrison, H., and Coauthors, 2020: Confronting the challenge of modeling cloud and precipitation microphysics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001689, https://doi.org/10.1029/2019MS001689.

Murdoch, W. J., C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, 2019: Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA*, **116**, 22 071–22 080, https://doi.org/10.1073/pnas.1900654116.

Naud, C. M., J. F. Booth, and A. D. Del Genio, 2014: Evaluation of ERA-Interim and MERRA cloudiness in the Southern Ocean. *J. Climate*, **27**, 2109–2124, https://doi.org/10.1175/JCLI-D-13-00432.1.

——, ——, K. Lamer, R. Marchand, A. Protat, and G. M. McFarquhar, 2020: On the relationship between the marine cold air outbreak *M* parameter and low-level cloud heights in the midlatitudes. *J. Geophys. Res. Atmos.*, **125**, e2020JD032465, https://doi.org/10.1029/2020JD032465.

O'Gorman, P. A., and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.*, **10**, 2548–2563, https://doi.org/10.1029/2018MS001351.

Orsi, A. H., T. Whitworth III, and W. D. Nowlin Jr., 1995: On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Res. I*, **42**, 641–673, https://doi.org/10.1016/0967-0637(95)00021-W.

Paton-Walsh, C., and Coauthors, 2022: Key challenges for tropospheric chemistry in the Southern Hemisphere. *Elementa*, **10**, 00050, https://doi.org/10.1525/elementa.2021.00050.

Protat, A., E. Schulz, L. Rikus, Z. Sun, Y. Xiao, and M. Keywood, 2017: Shipborne observations of the radiative effect of Southern Ocean clouds. *J. Geophys. Res. Atmos.*, **122**, 318–328, https://doi.org/10.1002/2016JD026061.

Rasp, S., P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, 2020: WeatherBench: A benchmark data set for data-driven weather forecasting. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002203, https://doi.org/10.1029/2020MS002203.

Reeve, J., and L. Symons, 1999: Underway voyage data collected from Australian Antarctic Division chartered ships, version 1. Australian Antarctic Data Centre, accessed 21 January 2021, https://data.aad.gov.au/metadata/records/underwayship%20data.

Ryan, R. G., J. D. Silver, and R. Schofield, 2021: Air quality and health impact of 2019–20 Black Summer megafires and COVID-19 lockdown in Melbourne and Sydney, Australia. *Environ. Pollut.*, **274**, 116498, https://doi.org/10.1016/j.envpol.2021.116498.

Scher, S., and G. Messori, 2019: Weather and climate forecasting with neural networks: Using general circulation models (GCMs) with different complexity as a study ground. *Geosci. Model Dev.*, **12**, 2797–2809, https://doi.org/10.5194/gmd-12-2797-2019.

Schmale, J., and Coauthors, 2019: Overview of the Antarctic Circumnavigation Expedition: Study of Preindustrial-like Aerosols and their Climate Effects (ACE-SPACE). *Bull. Amer. Meteor. Soc.*, **100**, 2260–2283, https://doi.org/10.1175/BAMS-D-18-0187.1.

Schuddeboom, A. J., and A. J. McDonald, 2021: The Southern Ocean radiative bias, cloud compensating errors, and equilibrium climate sensitivity in CMIP6 models. *J. Geophys. Res. Atmos.*, **126**, e2021JD035310, https://doi.org/10.1029/2021JD035310.

Stirnberg, R., and Coauthors, 2021: Meteorology-driven variability of air pollution (PM_1) revealed with explainable machine learning. *Atmos. Chem. Phys.*, **21**, 3919–3948, https://doi.org/10.5194/acp-21-3919-2021.

Tan, I., T. Storelvmo, and M. D. Zelinka, 2016: Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, **352**, 224–227, https://doi.org/10.1126/science.aad5300.

Tibshirani, R., 1996: Regression shrinkage and selection via the Lasso. *J. Roy. Stat. Soc.*, **58B**, 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Trenberth, K. E., and J. T. Fasullo, 2010: Simulation of present-day and twenty-first-century energy budgets of the Southern Oceans. *J. Climate*, **23**, 440–454, https://doi.org/10.1175/2009JCLI3152.1.

Twohy, C. H., and Coauthors, 2021: Cloud-nucleating particles over the Southern Ocean in a changing climate. *Earth's Future*, **9**, e2020EF001673, https://doi.org/10.1029/2020EF001673.

Vignon, É., and Coauthors, 2021: Challenging and improving the simulation of mid-level mixed-phase clouds over the high-latitude Southern Ocean. *J. Geophys. Res. Atmos.*, **126**, e2020JD033490, https://doi.org/10.1029/2020JD033490.

Voyant, C., G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, 2017: Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, **105**, 569–582, https://doi.org/10.1016/j.renene.2016.12.095.

Wang, H., A. R. Klekociuk, W. J. R. French, S. P. Alexander, and T. A. Warner, 2020: Measurements of cloud radiative effect across the Southern Ocean (43°S–79°S, 63°E–158°W). *Atmosphere*, **11**, 949, https://doi.org/10.3390/atmos11090949.

Weyn, J. A., D. R. Durran, and R. Caruana, 2019: Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J. Adv. Model. Earth Syst.*, **11**, 2680–2693, https://doi.org/10.1029/2019MS001705.

Wood, R., and C. S. Bretherton, 2006: On the relationship between stratiform low cloud cover and lower-tropospheric stability. *J. Climate*, **19**, 6425–6432, https://doi.org/10.1175/JCLI3988.1.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

Brought to you by BATTELLE PACIFIC NW LAB | Unauthenticated | Downloaded 05/15/23 12:39 PM UTC