The Southern Ocean Radiative Bias, Cloud Compensating Errors and Equilibrium Climate Sensitivity in CMIP6 Models

A. J. Schuddeboom[1] and A. J. McDonald[1,2]

[1]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand
[2]Gateway Antarctica, University of Canterbury, Christchurch, New Zealand

Key Points:

- Mean and compensating shortwave cloud radiative effect errors in CMIP6 AMIP models are quantified
- A significant negative relationship between mean and compensating shortwave cloud radiative effect errors is found over the Southern Ocean
- Shortwave cloud radiative effect errors are shown to be unlikely to explain the high effective climate sensitivity in these CMIP6 models

Corresponding author: Alex Schuddeboom, `Alex.Schuddeboom@canterbury.ac.nz`

Abstract

Analysis of Coupled Model Intercomparison Project Phase 6 (CMIP6) models has focused on their higher equilibrium climate sensitivity (ECS) relative to CMIP5 simulations, with higher ECS values attributed to changing cloud feedbacks. We examine the simulation of the shortwave cloud radiative effect (SW CRE) in CMIP6 and explore a potential link between SW CRE errors and ECS. We derive mean and compensating errors in model data relative to satellite observations using a cloud clustering methodology. A statistically significant negative relationship between the mean and compensating errors in SW CRE over the Southern Ocean is identified. This relationship is observed elsewhere, but is only significant over the Southern Ocean. This implies model tuning efforts potentially hide biases in the representation of clouds in this region. High ECS models tend to have lower mean and compensating errors over the Southern Ocean, suggesting model biases are unlikely to explain the high ECS values.

Plain Language Summary

Climate models go through a continual process of evaluation and improvement. This process is based on using observational data and higher resolution models to detect errors within climate models and identifying the model elements that need to be improved. A major issue with the current generation of climate models is that the equilibrium climate sensitivity is higher than previous generations of models. This means that the current generation of models show a greater temperature change in response to increased carbon dioxide than previous generations. It is unclear if this high equilibrium climate sensitivity is physically sensible or whether hidden errors in models are the cause. The traditional evaluation of climate models focuses on the average error. However, this can obscure the true magnitude of errors since models often have compensating errors where errors of opposing signs cancel out when averaged. This paper uses a technique based on cloud typing that evaluates both mean and compensating errors in CMIP6 models. The relationship between these errors and the high equilibrium climate sensitivity in these models is investigated, with a specific focus on the Southern Ocean region. We find no relationship between these hidden errors and ECS.

1 Introduction

The refinement of climate models is a continual process driven by constant evaluation against observational measurements and more detailed models. With the recent release of Coupled Model Intercomparison Project Phase 6 (CMIP6) model data (Eyring et al., 2016), a new wave of model assessment is currently underway. A major focus in the research released so far has been the equilibrium climate sensitivity (ECS) of the CMIP6 models (Zelinka et al., 2020; Meehl et al., 2020). Simply put, the ECS describes the magnitude of temperature change associated with doubling the carbon dioxide and running a model to equilibrium (National Research Council, 1979; Gregory et al., 2004). Compared to the models in CMIP5, those in CMIP6 have higher average ECS values and many CMIP6 models have a greater ECS value than the largest values from CMIP5 (Meehl et al., 2020). Work by Zelinka et al. (2020) has attributed this change to variations in cloud feedbacks, with the largest changes related to the shortwave (SW) cloud feedback linked to decreasing extratropical low cloud coverage and albedo. It is important to note that this feedback has the largest impact over the Southern Ocean, a region that has long been associated with large model biases relative to observations (Wild et al., 1995; Bodas-Salcedo et al., 2012; Kay et al., 2012). Work detailed in Gettelman et al. (2019) has also identified that the atmosphere model in CESM2 has higher cloud feedbacks than in previous model versions and that changes to the stratiform cloud microphysics and ice nucleation processes are important, both of which have a large impact over the Southern Ocean.

While in general it appears that the models that increased their ECS values between CMIP5 and CMIP6 better match the observational data than their CMIP5 counterparts, it remains unclear if the high ECS models are more accurate than the low ECS models (Gettelman et al., 2019). One area of uncertainty is associated with the role played by compensating errors which can hide the true magnitude of model errors (Jakob, 2003; Hyder et al., 2018; Schuddeboom et al., 2019). Compensating errors potentially mean that an improvement to one aspect of a model can lead to poorer performance as the previous bias may have counteracted an oppositely signed error which is unaccounted for in model tuning. Additionally, Zelinka et al. (2020) suggests that the high ECS values in some CMIP6 models could be a result of compensating errors whereby improvements in the representation of the SW cloud feedbacks could mean that some unresolved model errors are no longer being cancelled out, leading to an artificially high ECS. It may also be possible that the improvements made over the Southern Ocean lead to larger errors in other regions.

To explore the cloud compensating errors in the CMIP6 models, the methodology developed in Schuddeboom et al. (2018) and Schuddeboom et al. (2019) is used in this study. Schuddeboom et al. (2018) used the self organizing map (SOM, Kohonen (1998, 2013)) clustering technique on cloud top pressure–cloud optical thickness (CTP-COT) histograms to generate a set of cloud regimes associated with different cloud types. This approach for identifying cloud regimes is well established (Jakob, 2003; Oreopoulos et al., 2014; Mason et al., 2015; McDonald et al., 2016). By examining errors in the representation of each cloud regime in terms of their frequency of occurrence and their radiative properties, we can derive both the mean and cumulative error in the shortwave cloud radiative effect (SW CRE). This then allows magnitude of compensating errors to be estimated (Schuddeboom et al., 2019). In Schuddeboom et al. (2019) this technique was used to compare the effect of applying different model parameterizations to the same model, here it is used to examine different CMIP6 model runs. Given the importance of the SW CRE errors over the Southern Ocean, this region was specifically analyzed in Schuddeboom et al. (2019) and will also be analyzed in this paper. Zelinka et al. (2012) used a radiative transfer model to derive the sensitivity of top of atmosphere fluxes to changes in each bin of ISCCP CTP-COT histograms, these sensitivities when multiplied by changes in cloud fraction forced by a doubling of carbon dioxide concentrations can be used to directly quantify cloud feedbacks. We therefore test whether potential issues in the representation of the current occurrence of different cloud regimes within CTP-COT histograms might impact ECS.

2 Data

2.1 Observational Data

The analysis in this paper relies on data generated using the CFMIP Observation Simulator Package (COSP, Bodas-Salcedo et al. (2011) and Swales et al. (2018)). COSP is a satellite simulator which allows models to simulate data that is directly comparable to observations. This is particularly useful for producing variables like CTP-COT histograms as simulation of the observation process can reproduce many observational biases. The clusters developed in Schuddeboom et al. (2018) were generated using CTP-

COT histograms from the Moderate Resolution Imaging Spectroradiometer (MODIS, Platnick et al. (2003, 2017)) dataset. However, the regular outputs from the CMIP6 model runs only include International Satellite Cloud Climatology Project (ISCCP, Rossow and Schiffer (1991, 1999)) CTP-COT histograms. As such, this paper uses cloud regimes derived from ISCCP observations developed in McDonald and Parsons (2018) and based on the earlier work of McDonald et al. (2016).

The first of the major differences between the ISCCP and MODIS clusters are that the two datasets have different resolutions which means that the generated clusters represent different spatial scales. The second difference is that the ISCCP regimes from McDonald and Parsons (2018) have a different number of members than used in Schuddeboom et al. (2018). We should also note that the ISCCP clusters also include a filter that removes all data over regions with an average ground elevation of 1 km above sea level and that McDonald and Parsons (2018) used data from both 2007 and 2008 rather than a single year. While these factors impact the interpretation of the individual clusters, the integrated nature of the error calculation process means that the methodology is unaffected. These differences do however mean that the ΔCRE and $|$CRE$|$ should not be directly compared to the values produced in Schuddeboom et al. (2019).

In addition to the CTP-COT histograms from ISCCP, shortwave radiative fluxes from the Clouds and the Earth's Radiant Energy System (CERES, Wielicki et al. (1996)) dataset are used. In particular, this analysis uses top of atmosphere radiative fluxes from the synoptic 1° (SYN1deg) Edition 4.1 CERES dataset (Minnis et al., 2020). As in Schuddeboom et al. (2018) the CERES all sky and clear sky SW radiative fluxes are used to calculate the SW CRE. When used to analyze the cloud clusters, the CERES data is interpolated from a 1 degree by 1 degree equal angle grid to a 2.5 degree by 2.5 degree equal angle grid to match the resolution of the ISCCP data.

2.2 Model Data

There are a wide range of different model experiments included within the CMIP6 framework. In this research only model runs that correspond to the Atmospheric Model Intercomparison Project (AMIP, Gates et al. (1999) and Eyring et al. (2016)) are utilized. In the AMIP simulations sea surface temperature, sea ice and CO_2 concentrations are all prescribed. This specific experiment was chosen as it covers the same historical period as the observations, it is also forced in a way which makes the models more comparable to observations and includes model runs from several different models which simulated all of the variables required for the analysis. The variables used in this study include the COSP generated daily ISCCP COT-CTP histograms (*clisccp*) and the clear sky and all sky SW radiative fluxes (*rsut* and *rsutcs*). While the historical experiment runs also met this criteria, fewer model runs simulated all of these variable and by focusing specifically on one experiment we can ensure that the models are compared in a consistent manner. To match the observational data, the cluster based analysis of the model data covers a period between 2007 and 2008 and removes all points over topography greater than 1 km above sea level.

Once these requirements have been applied there are eight models available through the Earth System Grid Federation (ESGF) online system which have daily output for the set of variables required. The eight suitable models are CESM2 (Danabasoglu, 2019; Gettelman et al., 2019), CNRM-CM6-1 (Voldoire, 2018), CNRM-ESM2-1 (Seferian, 2018), GFDL-CM4 (Guo et al., 2018), HadGEM3-GC31-LL (Ridley et al., 2019), IPSL-CM6A-LR (Boucher et al., 2018), MRI-ESM2-0 (Yukimoto et al., 2019) and UKESM1-0-LL (Tang et al., 2019). There are two models which produce the *clisccp* variable but are not included in this analysis, BCC-CSM2-MR and GISS-E2-1-G, this is associated with an output error and a lack of other data, respectively. Both the HadGEM3 (5 runs) and IPSL (2 runs) models include multiple runs, for these models the clustering is applied to all runs and the ensemble mean displayed unless otherwise indicated. In later analysis, we use the effective climate sensitivity values for each of these models taken from Meehl et al. (2020) to represent model ECS. While there are differences between effective climate sensitivity and equilibrium climate sensitivity, these are generally small. These values are calculated from coupled atmosphere–ocean model simulations which is different then our analysis based on AMIP simulations. A lack of data means that this inconsistency can not be avoided without significant extra work. However, given that the errors we consider are associated with cloud feedbacks and these have been identified in previous work as critical to changes in ECS, we believe that the analysis of atmosphere-only models in AMIP is still valid. Brief testing (not included in this study) of a small subset of historical runs shows that they generate similar results to their equivalent AMIP runs, differing by less than one third of a Wm^{-2} in mean error and less than one Wm^{-2} in compensating error.

3 Methodology

The cluster based approach used to analyze the model runs in this paper was developed in Schuddeboom et al. (2019) and was heavily influenced by the work in Williams and Webb (2009), Williams and Bodas-Salcedo (2017) and Hyder et al. (2018). This approach is based on using the rate at which a given cluster occurs, known as the relative frequency of occurrence (RFO), and the cluster average SW CRE to calculate the difference in SW CRE between models and observations. Two values are calculated, the ΔCRE which is the mean difference between the model and observations and the |CRE| which is the difference if the errors associated with each cluster are summed rather than allowed to cancel. Therefore, |CRE| allows us to calculate the magnitude of the cloud cluster compensating errors. While there will be compensating errors not captured by this approach, by using the cloud clusters we are at least able to estimate the magnitude of compensating errors. ΔCRE and |CRE| can be calculated with equations 1 and 2 where C represents the average CRE of cluster N and R represents the RFO of cluster N.

$$\Delta\text{CRE} = |\sum_N C_N^{\text{Model}} R_N^{\text{Model}} - C_N^{\text{CERES}} R_N^{\text{ISCCP}}| \tag{1}$$

$$|\text{CRE}| = \sum_N |C_N^{\text{Model}} R_N^{\text{Model}} - C_N^{\text{CERES}} R_N^{\text{ISCCP}}| \tag{2}$$

4 Results and Discussion

Before examining the models using the ΔCRE and |CRE|, the zonal mean SW CRE is studied. Figure 1 (a) shows the zonal mean SW CRE for the CERES data and for each model and also includes the anomaly for each of these models from the CERES observations in figure 1 (b). The SW CRE zonal means are generally consistent across the different models with a clear peak over the Southern Ocean. There are however relatively large differences between the models over the tropics. These results can be compared to figure 9.5d from Flato et al. (2013), which shows a corresponding plot for the CMIP5 participants. In general, the shape of the distributions in CMIP5 and CMIP6 are similar, although the CMIP6 models have a smaller range of SW CRE over the Southern Ocean and also match better with the CERES mean SW CRE.

The differences between the models are further explored in figure 1 (b) which shows the zonal mean anomalies from the CERES measurements. The zonal average anoma-

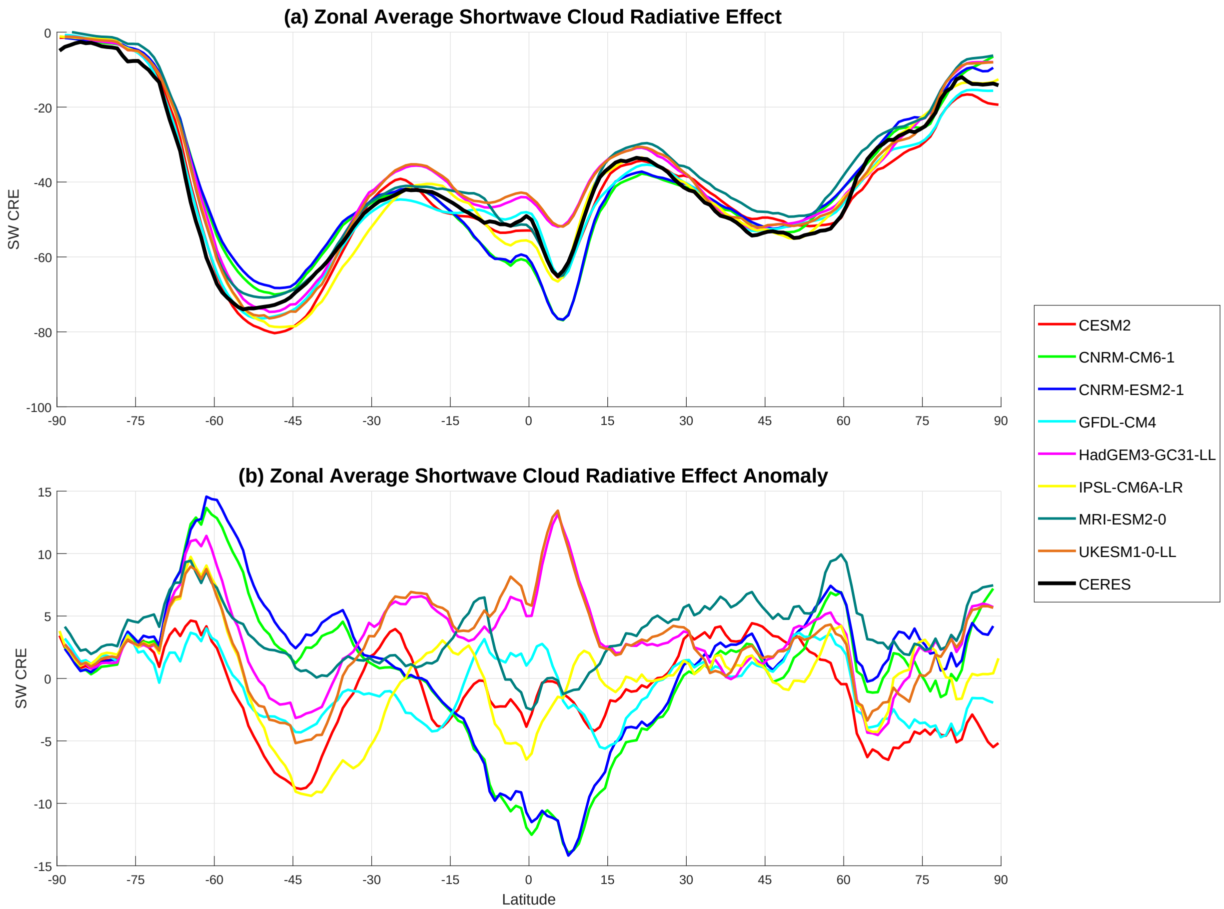


Figure 1. Zonal mean SW CRE for the CERES observations and each of the included CMIP6 model runs (a) and SW CRE anomalies between each of the models and CERES observations (b). The observational data used to generate the anomalies is from the CERES dataset and the sign convention used for the anomalies is model minus observations.

lies show a wide spread of values, ranging from +15 Wm^{-2} to -15 Wm^{-2}. At latitudes between 30^oN and 60^oN the models appear relatively consistent with small biases, with a wider range of biases north of 60^oN. Between 30^oS and 30^oN there are a wide range of behaviours. The GFDL, IPSL, MRI and CESM models all show relatively small biases, while the CNRM models show large negative biases and HADGEM3 and UKESM models show large positive biases. Finally, clear differences between the models are observed south of 30^oS, specifically between 45^oS and 70^oS. Over this region the models display the same shape with positive anomalies between 50^oS and 70^oS with a peak around 60^oS and then steadily decreasing anomaly values from 60^oS to 45^oS in every case. While the form of the curve is similar in each case the magnitude of the bias differs significantly between the models. This difference can reach up to 15 Wm^{-2} meaning that some models have a mostly positive bias over this region while others are mostly negative. The nature of this bias is similar to a result identified in Schuddeboom et al. (2019) where the Southern Ocean featured two different regional biases, although the result in this paper is much weaker than shown in the previous work.

Next the ΔCRE and |CRE| are calculated with equations 1 and 2 and plotted in figure 2. Figure 2 (a) shows results averaged globally for the cluster analysis, while figure 2 (b) only examines the Southern Ocean region. In figure 2 the Southern Ocean is defined as the region between $40°S$ and $70°S$ to ensure it includes both of the large spikes visible in figure 1 (b). For the models with multiple runs, each model member is shown along with the ensemble mean. Examination of these runs highlights that they show little difference from the ensemble mean. Additionally, to determine if these results are representative, the same analysis was completed using data from 2009 and 2010 (see figure S1 in supplementary information) and differences are very minor. Also included in both sub-figures is the Pearson correlation coefficient, however this should be interpreted carefully as the values are not statistically significant.

The global values in figure 2 (a) show little coherent structure between the different models. In the mean error there appear to be two groups of models, MRI-ESM, HadGEM and UKESM which show higher mean errors of around 5 Wm^{-2} and all of the other models which show mean errors around 1 Wm^{-2}. The compensating errors show less structure with values over a large range. Notably, the IPSL model shows compensating errors that are around 10 Wm^{-2} larger than the next largest model. Overall, it seems that CESM2, CNRM-ESM2 and CNRM-CM6 are the standout models when both metrics are considered with MRI-ESM2 also having notably small compensating errors. This figure also demonstrates the importance of considering the compensating errors as both IPSL and GFDL appear to be amongst the most accurate models in the mean error, but are the least accurate in the compensating error. In all models the compensating errors are significantly larger than the mean errors suggesting significant work is still required to improve the representation of clouds in climate models. For example, consider the best performing model in |CRE|, CESM2. CESM2 has a mean error of around 1 Wm^{-2} and a total compensating error of around 34 Wm^{-2}. This suggests that CESM2 actually has compensating errors of 16.5 and 17.5 Wm^{-2} that are cancelling out to give the mean error of 1 Wm^{-2}.

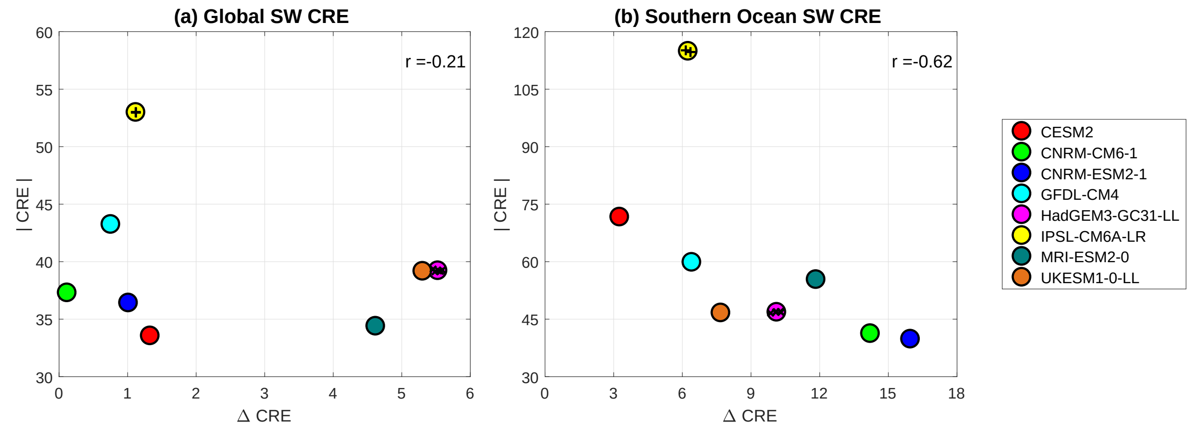


Figure 2. The ΔCRE and $|$CRE$|$ for each of the models. Subplot (a) shows the global average values while subplot (b) covers the entire Southern Ocean region, defined here as between $40^{\circ}S$ and $70^{\circ}S$. For models with multiple runs, IPSL and HadGEM3, the individual runs are indicated with a + symbol for IPSL and $\times$ symbol for HadGEM3. To aide interpretation of this figure recall that ΔCRE can be considered the mean SW CRE error and $|$CRE$|$ the magnitude of the SW CRE compensating errors. This means that an ideal model would minimize both ΔCRE and $|$CRE$|$ and appear near the origin. The variables on each of the axis are described by equations 1 and 2. The Pearson correlation coefficients are included in the top right of each subplot with any results that are statistically significant in bold.

The Southern Ocean values shown in figure 2 (b) display a completely different structure compared to the global values in figure 2 (a). Over the Southern Ocean, the models show a clear structure in which an increase in the mean error corresponds to a reduction in the compensating errors. There is one major outlier, the IPSL model, which as in the global data has significantly larger compensating errors than other models. Looking at the other models, they appear to fall on a line between CESM2 and CNRM-ESM2. CESM2 has the lowest mean error of around 3 Wm^{-2}, but the largest (excluding IPSL) $|$CRE$|$ of around 70 Wm^{-2}. While CNRM-ESM2 has the largest mean error of 16 Wm^{-2} and the smallest $|$CRE$|$ of 40 Wm^{-2}. This suggests compensating errors of 33.5 and 36.5 Wm^{-2} for CESM2 and 12 and 28 Wm^{-2} for CNRM-ESM2. Special mention should also be made of the UKESM and HadGEM models as they have relatively small ΔCRE and $|$CRE$|$. The relationship between the mean and compensating biases identified in figure 2 (b) appears to be related to ordering of the models over the Southern Ocean in figure 1 (b). This suggests a relationship between the geographic nature of the SW CRE biases and the magnitude of compensating errors. We also note that the errors derived over the Southern Ocean are larger than those observed globally which is associated with the fact that the occurrence of specific cloud clusters is more flawed when averaged over the Southern Ocean than when averaged over the globe.

Without examining other regions it is hard to contextualize the results over the Southern Ocean. As such, figure 3 shows the ΔCRE and |CRE| over six different regions of the globe. An alternative representation which shows the mean and compensating errors as a function of latitude is also included in the supplementary information as figure S2. The region that covers the majority of the Southern Ocean, the midlatitude Southern Hemisphere, shows a strongly negative statistically significant relationship between ΔCRE and |CRE|. The correlation coefficients also suggest negative relationships in the Northern Hemisphere regions. The two bands that cover the Southern Ocean also show considerably larger mean and compensating errors than the other geographic regions examined. Overall, these results show that the negative relationship identified in figure 2 (b) is strongest over the Southern Ocean and is only statistically significant over that region.

We now examine whether the mean and compensating cloud errors, which quantify the misrepresentation in the CTP-COT histograms in the models relative to satellite observations, have any relationship to ECS (figure 4). Figure 4 displays the relationship between mean and compensating errors with ECS for the globe and the Southern Ocean. This enables an exploration of whether compensating errors over the Southern Ocean are a driving factor in the higher low-level cloud feedbacks which drive high ECS values (Zelinka et al., 2020). As with figure 2, this figure is recreated using data from 2009 and 2010 and is included in the supplementary information as figure S3. Once again the differences between these figures are minimal. To more quantitatively evaluate the trends presented figures 2 and 4 alternate versions of these figures with a line of best fit determined by linear regression are included in the supplementary information as figure S4 and S5.

The global mean errors shown in figure 4 (a) suggest only a weak relationship between global mean SW CRE error and ECS. As described earlier, MRI-ESM, HadGEM and UKESM show much larger global mean errors than the other models however the MRI-ESM is the smallest ECS model used in this paper and the HadGEM and UKESM

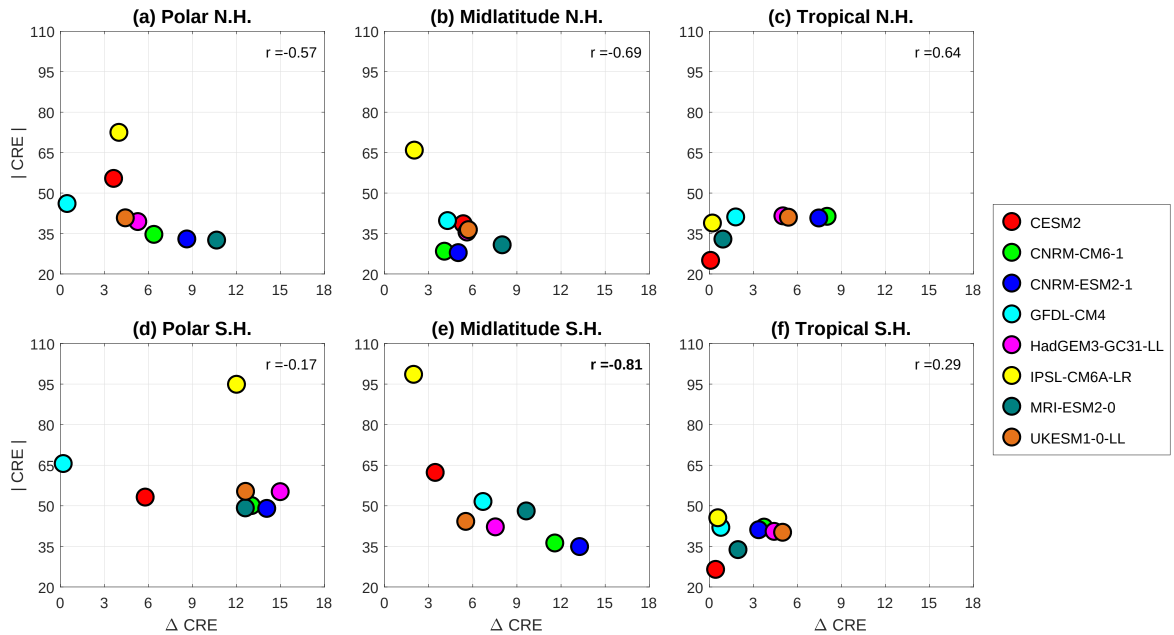


Figure 3. The ΔCRE and $|$CRE$|$ for each of the models. The subplots show regions defined as following: (a) $90^\circ N$ to $60^\circ N$, (b) $60^\circ N$ to $30^\circ N$, (c) $30^\circ S$ to $0^\circ N$, (d) $90^\circ S$ to $60^\circ S$, (e) $60^\circ S$ to $30^\circ S$ and (f) $30^\circ S$ to $0^\circ S$. The variables on each of the axis are described by equations 1 and 2. The Pearson correlation coefficients are included in the top right of each subplot with any results that are statistically significant in bold.

are the two largest. The global compensating errors in figure 4 (b) show more structure, but again do not show a clear relationship between compensating errors and ECS and the correlation coefficient suggests no relationship. At first glance it appears that the highest ECS models have the lowest compensating errors, but this is mostly due to the poor performance of the IPSL and GFDL models. Ultimately there might be more coherent patterns for both these figures, but the limited number of models makes it unclear and further work should add more models into this analysis as the data becomes available.

The Southern Ocean errors in figure 4 (c) and (d) display some level of coherence between the different models. The mean errors in figure 4 (c) display a complicated structure that can be be interpreted in two different ways. While the general tendency is that lower mean error models have higher ECS, this interpretation is complicated by the average performance of HadGEM and UKESM. An alternative interpretation is to group the models into three groups. First, CESM by itself, then GFDL, IPSL, HadGEM and

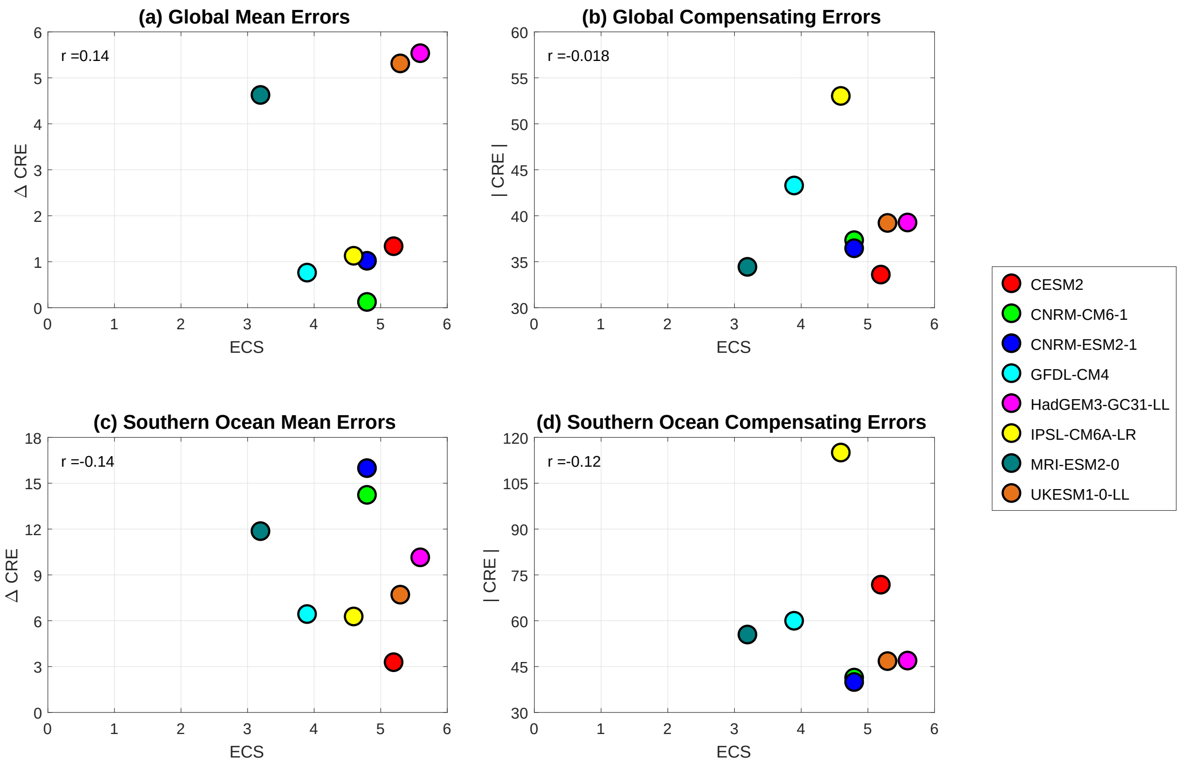


Figure 4. The ECS and the mean and compensating SW CRE errors of the included CMIP6 models. This includes both global values and values over the Southern Ocean, defined here as between $40^\circ S$ and $70^\circ S$. Subplot (a) shows the global ΔCRE, subplot (b) shows global |CRE|, subplot (c) shows the Southern Ocean ΔCRE and subplot (d) shows Southern Ocean |CRE|. ΔCRE and |CRE| are described in equations 1 and 2 and the ECS values are taken from Meehl et al. (2020). The Pearson correlation coefficients are included in the top left of each subplot with any results that are statistically significant in bold.

UKESM and finally, MRI-ESM and the two CNRM models. With this interpretation a greater mean error would correspond to a larger ECS. The line of best fit in supplementary figure S5 and the correlation coefficient suggest the negative relationship is more likely, but it is not statistically significant. Figure 4 (d) shows the clearest negative relationship with the highest ECS models showing the smallest compensating errors, although the relationship is weak. These results suggest a further examination of both the IPSL and CESM2 models as they have a clear position as outliers.

5 Conclusion

This paper focuses on eight models from CMIP6 which have released AMIP runs with the appropriate data for the SW CRE to be evaluated with the techniques introduced in Schuddeboom et al. (2019). This methodology examines cloud cluster in terms of their frequency of occurrence and their radiative properties to derive both the mean and cumulative error in the shortwave cloud radiative effect (SW CRE). Initially the zonal distribution of SW CRE was compared across all the models. This identified that there was generally good agreement between the models and CERES observations, although there were two regions, the Tropics and the Southern Ocean, where clear model disagreement with the observations occurred. The issues over the Tropics are not explored further in the current work, but is an area worthy of further investigation.

Next the mean and compensating errors in SW CRE were analyzed for each of the eight models using ΔCRE and $|$CRE$|$. Both globally and over the Southern Ocean the compensating errors are shown to be large relative to the mean errors. While there was no clear relationship between mean and compensating errors in the global results, there is a strong negative relationship over the Southern Ocean. Models with small mean errors over the Southern Ocean consistently have large compensating errors and vice versa. This suggests that the models with the high mean errors may better simulate the Southern Ocean region but errors that are being cancelled out in other models increase the mean value. This behaviour possibly stems from a geographical bias identified in the zonal mean SW CRE.

In addition to the Southern Ocean, other geographical regions were examined. While several relationships are suggested between mean and compensating errors, the strongest and only statistically significant relationship was observed over the Southern Ocean. The mean and compensating errors were also directly evaluated against the ECS taken from Meehl et al. (2020). Globally, there was no clear relationship between model ECS and its mean or compensating errors. Over the Southern Ocean there appears to be a weak relationship that suggested that a model with a strong ECS was more likely to have lower mean and compensating errors, though we emphasise that this relationship is not statistically significant. The exact strength of these relationships are uncertain due to the limited number of available models.

The main result of this study is that the mean and compensating errors related to clouds in these models display a negative relationship over polar and midlatitude regions, most strongly observed over the Southern Ocean. This suggests that recent model improvements targeted at removing shortwave radiative biases over the Southern Ocean relative to satellite observations have only partially corrected the issues and that other errors still exist when the distribution of cloud types and their radiative properties are considered. Unfortunately, the small number of models with suitable outputs is a major limitation on this study. We hope that as a wider array of models release appropriate data it will become possible to reach more definitive conclusions about the relationship between SW CRE errors over the Southern Ocean and ECS. In addition to adding extra models, the authors also plan on using this approach to evaluate the impact of changes made to models between CMIP5 and CMIP6. This could potentially lead to the identification of the parts of the model that need further improvements.

Acknowledgments

The authors would like to acknowledge the teams behind the ISCCP and CERES datasets. The CERES data were obtained from https://ceres.larc.nasa.gov/data/, and the ISCCP data were obtained from https://www.ncei.noaa.gov/data/international-satellite-cloud-climate-project-isccp-h-series-data/access/isccp-basic/hgg/. We would also like to acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. Finally, we would like to thank the Deep South National Science Challenge for providing funding for this research.

References

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., … John, V. O. (2011, 8). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/2011BAMS2856.1` doi: 10.1175/2011BAMS2856.1

Bodas-Salcedo, A., Williams, K. D., Field, P. R., & Lock, A. P. (2012). The surface downwelling solar radiation surplus over the southern ocean in the met office model: The role of midlatitude cyclone clouds. *Journal of Climate*, *25*(21), 7467–7486. doi: 10.1175/JCLI-D-11-00702.1

Boucher, O., Denvil, S., Caubel, A., & Foujols, M. A. (2018). *IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.5113` doi: 10.22033/ESGF/CMIP6.5113

Danabasoglu, G. (2019). *NCAR CESM2 model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.7522` doi: 10.22033/ESGF/CMIP6.7522

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of Climate Models. In Intergovernmental Panel on Climate Change (Ed.), *Climate change 2013 - the physical science basis* (Vol. 9781107057, pp. 741–866). Cambridge: Cambridge University Press. Retrieved from `https://www.cambridge.org/core/product/identifier/CBO9781107415324A028/type/book_part` doi: 10.1017/CBO9781107415324.020

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., ... Williams, D. N. (1999, 1). An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, *80*(1), 29–55. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/1520-0477(1999)080%3C0029:AOOTRO%3E2.0.CO;2http://journals.ametsoc.org/doi/abs/10.1175/1520-0477%281999%29080%3C0029%3AAOOTRO%3E2.0.CO%3B2` doi: 10.1175/1520-0477(1999)080⟨0029:AOOTRO⟩2.0.CO;2

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., ... Mills, M. J. (2019, 7). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, *46*(14), 8329–8337. Retrieved from `https://onlinelibrary.wiley`

.com/doi/abs/10.1029/2019GL083978 doi: 10.1029/2019GL083978

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., ... Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3), 2–5. doi: 10.1029/2003GL018747

Guo, H., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., ... Zhang, R. (2018). *NOAA-GFDL GFDL-CM4 model output amip.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.8494 doi: 10.22033/ESGF/CMIP6.8494

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., ... Belcher, S. E. (2018, 12). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, *9*(1), 3625. Retrieved from http://dx.doi.org/10.1038/s41467-018-05634-2http://www.nature.com/articles/s41467-018-05634-2 doi: 10.1038/s41467-018-05634-2

Jakob, C. (2003, 10). An Improved Strategy for the Evaluation of Cloud Parameterizations in GCMS. *Bulletin of the American Meteorological Society*, *84*(10), 1387–1402. Retrieved from http://journals.ametsoc.org/doi/10.1175/BAMS-84-10-1387 doi: 10.1175/BAMS-84-10-1387

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., ... Ackerman, T. P. (2012). Exposing global cloud biases in the Community Atmosphere Model (CAM) using satellite observations and their corresponding instrument simulators. *Journal of Climate*, *25*(15), 5190–5207. doi: 10.1175/JCLI-D-11-00469.1

Kohonen, T. (1998, 11). The self-organizing map. *Neurocomputing*, *21*(1-3), 1–6. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0925231298000307 doi: 10.1016/S0925-2312(98)00030-7

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65. Retrieved from http://dx.doi.org/10.1016/j.neunet.2012.09.018 doi: 10.1016/j.neunet.2012.09.018

Mason, S., Fletcher, J. K., Haynes, J. M., Franklin, C., Protat, A., & Jakob, C. (2015). A hybrid cloud regime methodology used to evaluate Southern Ocean cloud and shortwave radiation errors in ACCESS. *Journal of Climate*, *28*(15), 6001–6018. doi: 10.1175/JCLI-D-14-00846.1

McDonald, A. J., Cassano, J. J., Jolly, B., Parsons, S., & Schuddeboom, A. (2016, 11). An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states. *Journal of Geophysical Research: Atmospheres*, *121*(21), 009–13. Retrieved from `http://doi.wiley.com/10.1002/2016JD025199` doi: 10.1002/2016JD025199

McDonald, A. J., & Parsons, S. (2018, 7). A Comparison of Cloud Classification Methodologies: Differences between Cloud and Dynamical Regimes. *Journal of Geophysical Research: Atmospheres*, *123*(19), 173–11. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018JD028595?af=Rhttp://doi.wiley.com/10.1029/2018JD028595` doi: 10.1029/2018JD028595

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., ... Schlund, M. (2020, 6). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, *6*(26), eaba1981. Retrieved from `https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aba1981` doi: 10.1126/sciadv.aba1981

Minnis, P., Sun-Mack, S., Chen, Y., Chang, F.-L., Yost, C. R., Smith, W. L., ... Xie, Y. (2020). CERES MODIS Cloud Product Retrievals for Edition 4–Part I: Algorithm Changes. *IEEE Transactions on Geoscience and Remote Sensing*, 1–37. doi: 10.1109/tgrs.2020.3008866

National Research Council. (1979). *Carbon Dioxide and Climate*. Washington, D.C.: National Academies Press. Retrieved from `https://www.nap.edu/catalog/12181` doi: 10.17226/12181

Oreopoulos, L., Cho, N., Lee, D., Kato, S., & Huffman, G. J. (2014, 7). An examination of the nature of global MODIS cloud regimes. *Journal of Geophysical Research: Atmospheres*, *119*(13), 8362–8383. Retrieved from `http://doi.wiley.com/10.1002/2013JD021409` doi: 10.1002/2013JD021409

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C., & Frey, R. A. (2003). The MODIS cloud products: Algorithms and examples from terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2 PART 1), 459–472. doi: 10.1109/TGRS.2002.808301

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., ... Riedi, J. (2017). The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples from Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525. doi: 10.1109/TGRS.2016.2610522

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019). *MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.5853` doi: 10.22033/ESGF/CMIP6.5853

Rossow, W. B., & Schiffer, R. A. (1991, 1). ISCCP Cloud Data Products. *Bulletin of the American Meteorological Society*, *72*(1), 2–20. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/1520-0477%281991%29072%3C0002%3AICDP%3E2.0.CO%3B2` doi: 10.1175/1520-0477(1991)072⟨0002:ICDP⟩2.0.CO;2

Rossow, W. B., & Schiffer, R. A. (1999, 11). Advances in Understanding Clouds from ISCCP. *Bulletin of the American Meteorological Society*, *80*(11), 2261–2287. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/1520-0477%281999%29080%3C2261%3AAIUCFI%3E2.0.CO%3B2` doi: 10.1175/1520-0477(1999)080⟨2261:AIUCFI⟩2.0.CO;2

Schuddeboom, A., McDonald, A. J., Morgenstern, O., Harvey, M., & Parsons, S. (2018, 4). Regional Regime-Based Evaluation of Present-Day General Circulation Model Cloud Simulations Using Self-Organizing Maps. *Journal of Geophysical Research: Atmospheres*. Retrieved from `https://doi.org/10.1002/2017JD028196http://doi.wiley.com/10.1002/2017JD028196` doi: 10.1002/2017JD028196

Schuddeboom, A., Varma, V., McDonald, A. J., Morgenstern, O., Harvey, M., Parsons, S., ... Furtado, K. (2019, 3). ClusterBased Evaluation of Model Compensating Errors: A Case Study of Cloud Radiative Effect in the Southern Ocean. *Geophysical Research Letters*, *46*(6), 3446–3453. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL081686` doi: 10.1029/2018GL081686

Seferian, R. (2018). *CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from `https://`

doi.org/10.22033/ESGF/CMIP6.3924 doi: 10.22033/ESGF/CMIP6.3924

Swales, D. J., Pincus, R., & Bodas-Salcedo, A. (2018, 1). The Cloud Feedback Model Intercomparison Project Observational Simulator Package: Version 2. *Geoscientific Model Development*, *11*(1), 77–81. doi: 10.5194/gmd-11-77-2018

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., ... Jones, C. (2019). *MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.5857 doi: 10.22033/ESGF/CMIP6.5857

Voldoire, A. (2018). *CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.3922 doi: 10.22033/ESGF/CMIP6.3922

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bulletin of the American Meteorological Society*, *77*(5), 853–868. doi: 10.1175/1520-0477(1996)077⟨0853:CATERE⟩2.0.CO;2

Wild, M., Ohmura, A., Gilgen, H., & Roeckner, E. (1995, 5). Validation of General Circulation Model Radiative Fluxes Using Surface Observations. *Journal of Climate*, *8*(5), 1309–1324. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1520-0442(1995)008<1309:VOGCMR>2.0.CO;2 doi: 10.1175/1520-0442(1995)008⟨1309:VOGCMR⟩2.0.CO;2

Williams, K. D., & Bodas-Salcedo, A. (2017). A multi-diagnostic approach to cloud evaluation. *Geoscientific Model Development*, *10*(7), 2547–2566. doi: 10.5194/gmd-10-2547-2017

Williams, K. D., & Webb, M. J. (2009). A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics*, *33*(1), 141–157. doi: 10.1007/s00382-008-0443-1

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., ... Adachi, Y. (2019). *MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP amip.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.6758 doi: 10.22033/ESGF/CMIP6.6758

Zelinka, M. D., Klein, S. A., & Hartmann, D. L. (2012, 6). Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part I: Cloud

Radiative Kernels. *Journal of Climate*, *25*(11), 3715–3735. Retrieved from http://journals.ametsoc.org/doi/10.1175/JCLI-D-11-00248.1 doi: 10.1175/JCLI-D-11-00248.1

Zelinka, M. D., Myers, T. A., McCoy, D. T., PoChedley, S., Caldwell, P. M., Ceppi, P., ... Taylor, K. E. (2020, 1). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, *47*(1). Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085782 doi: 10.1029/2019GL085782

Figure 1.

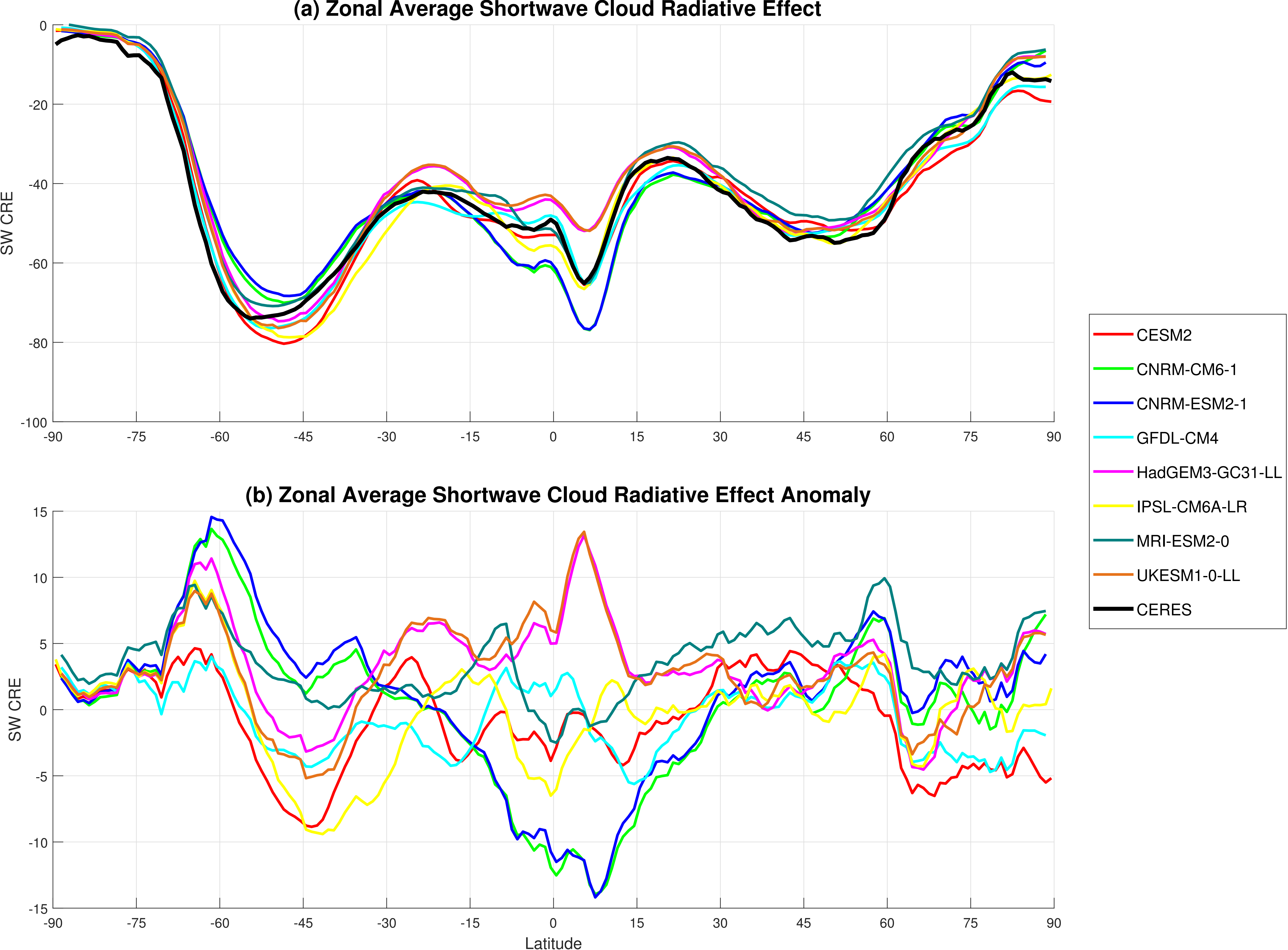
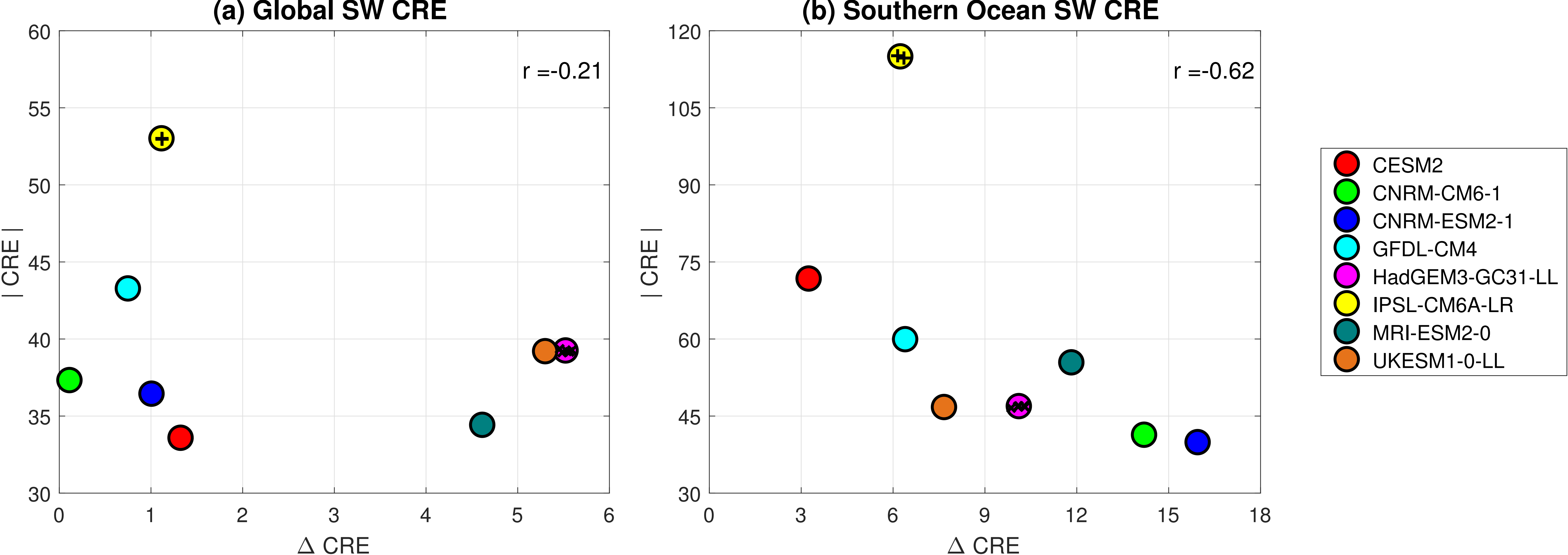

(a) Zonal Average Shortwave Cloud Radiative Effect
SW CRE
(b) Zonal Average Shortwave Cloud Radiative Effect Anomaly
SW CRE
Latitude
CESM2
CNRM-CM6-1
CNRM-ESM2-1
GFDL-CM4
HadGEM3-GC31-LL
IPSL-CM6A-LR
MRI-ESM2-0
UKESM1-0-LL
CERES

Figure 2.



(a) Global SW CRE
r =-0.21
| CRE |
Δ CRE
(b) Southern Ocean SW CRE
r =-0.62
| CRE |
Δ CRE
CESM2
CNRM-CM6-1
CNRM-ESM2-1
GFDL-CM4
HadGEM3-GC31-LL
IPSL-CM6A-LR
MRI-ESM2-0
UKESM1-0-LL

Figure 3.

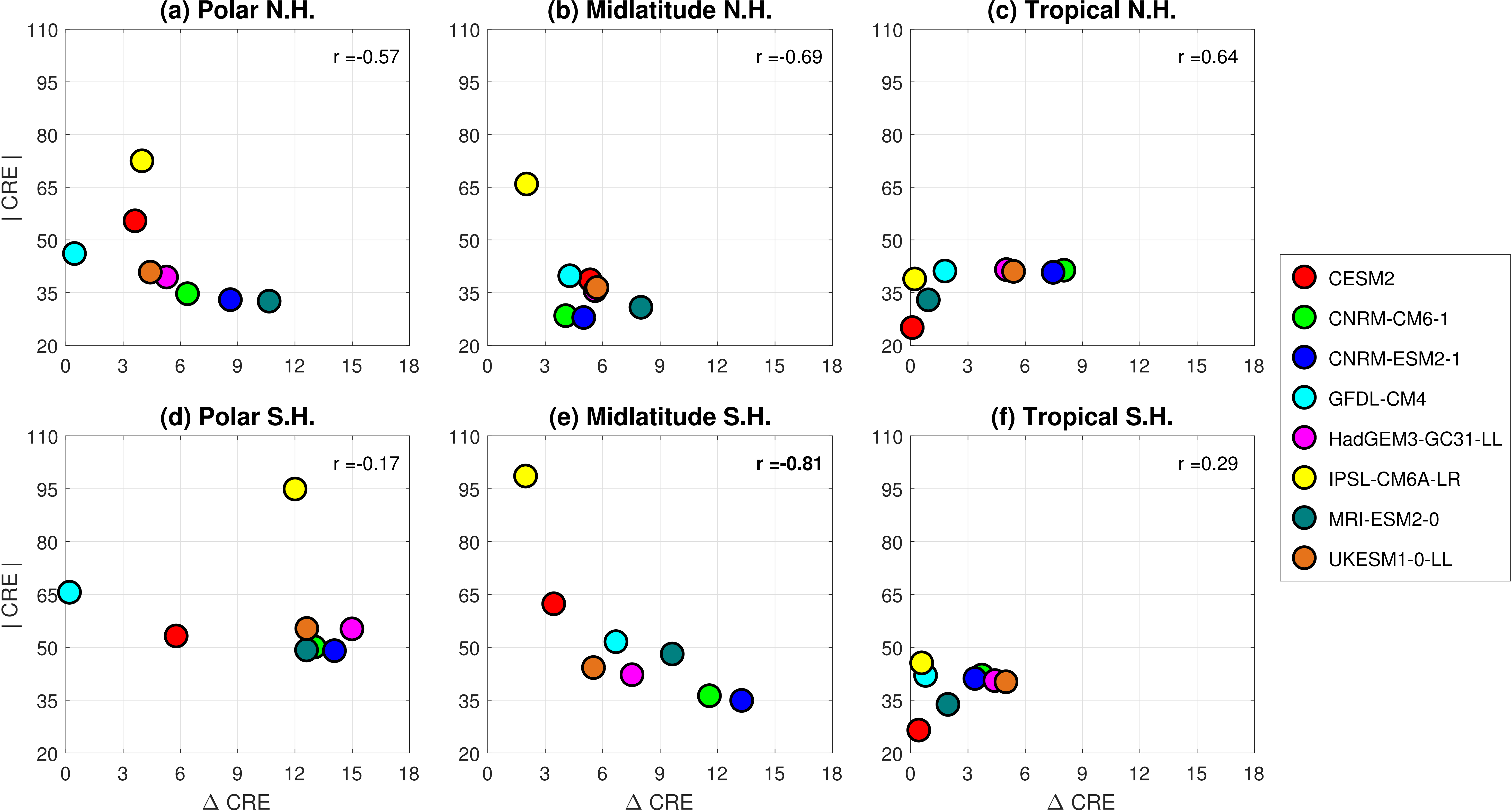
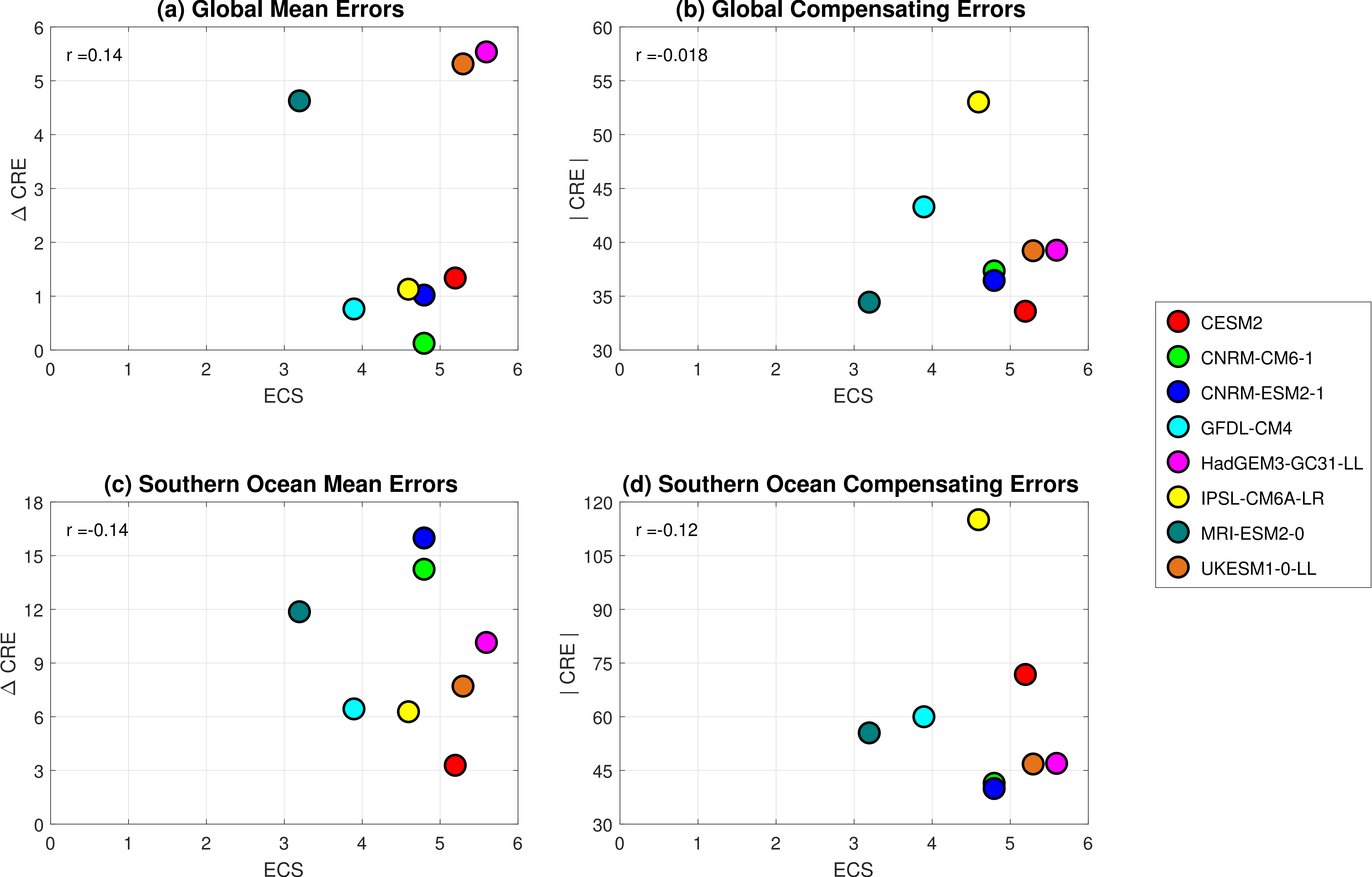

(a) Polar N.H.
r =-0.57
(b) Midlatitude N.H.
r =-0.69
(c) Tropical N.H.
r =0.64
(d) Polar S.H.
r =-0.17
(e) Midlatitude S.H.
r =-0.81
(f) Tropical S.H.
r =0.29
| CRE |
Δ CRE
CESM2
CNRM-CM6-1
CNRM-ESM2-1
GFDL-CM4
HadGEM3-GC31-LL
IPSL-CM6A-LR
MRI-ESM2-0
UKESM1-0-LL

Figure 4.


(a) Global Mean Errors
r =0.14
Δ CRE
ECS
(b) Global Compensating Errors
r =-0.018
| CRE |
ECS
(c) Southern Ocean Mean Errors
r =-0.14
Δ CRE
ECS
(d) Southern Ocean Compensating Errors
r =-0.12
| CRE |
ECS
CESM2
CNRM-CM6-1
CNRM-ESM2-1
GFDL-CM4
HadGEM3-GC31-LL
IPSL-CM6A-LR
MRI-ESM2-0
UKESM1-0-LL