Supplement of Atmos. Chem. Phys., 24, 8295–8316, 2024
https://doi.org/10.5194/acp-24-8295-2024-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

A systematic evaluation of high-cloud controlling factors

Sarah Wilson Kemsley et al.

Correspondence to: Sarah Wilson Kemsley (s.wilson-kemsley@uea.ac.uk)

S1. CMIP models

Eighteen CMIP5/6 models are available having run the ISCCP simulator. We use the first available, historical (AMIP) realisations for CCFs and cloud radiative anomalies for the years 1981 – 2000 from the following models:

- **CMIP5**: CanESM2, CNRM-EM2-1*, HadGEM2-ES*, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC5, MIROC-ESM, MPI-ESM-LR, MRI-CGCM3,
- **CMIP6**: CanESM5, CNRM-CM6-1*, GFDL-CM4, HadGEM3-GC31-LL, IPSL-CM6A-LR*, MIROC6, MIROC-ES2L, MRI-ESM2-0, UKESM1-0-LL.

*Models where daily air temperature and/or humidity at standard pressure levels is not available, therefore no CAPE or CIN data has been calculated.

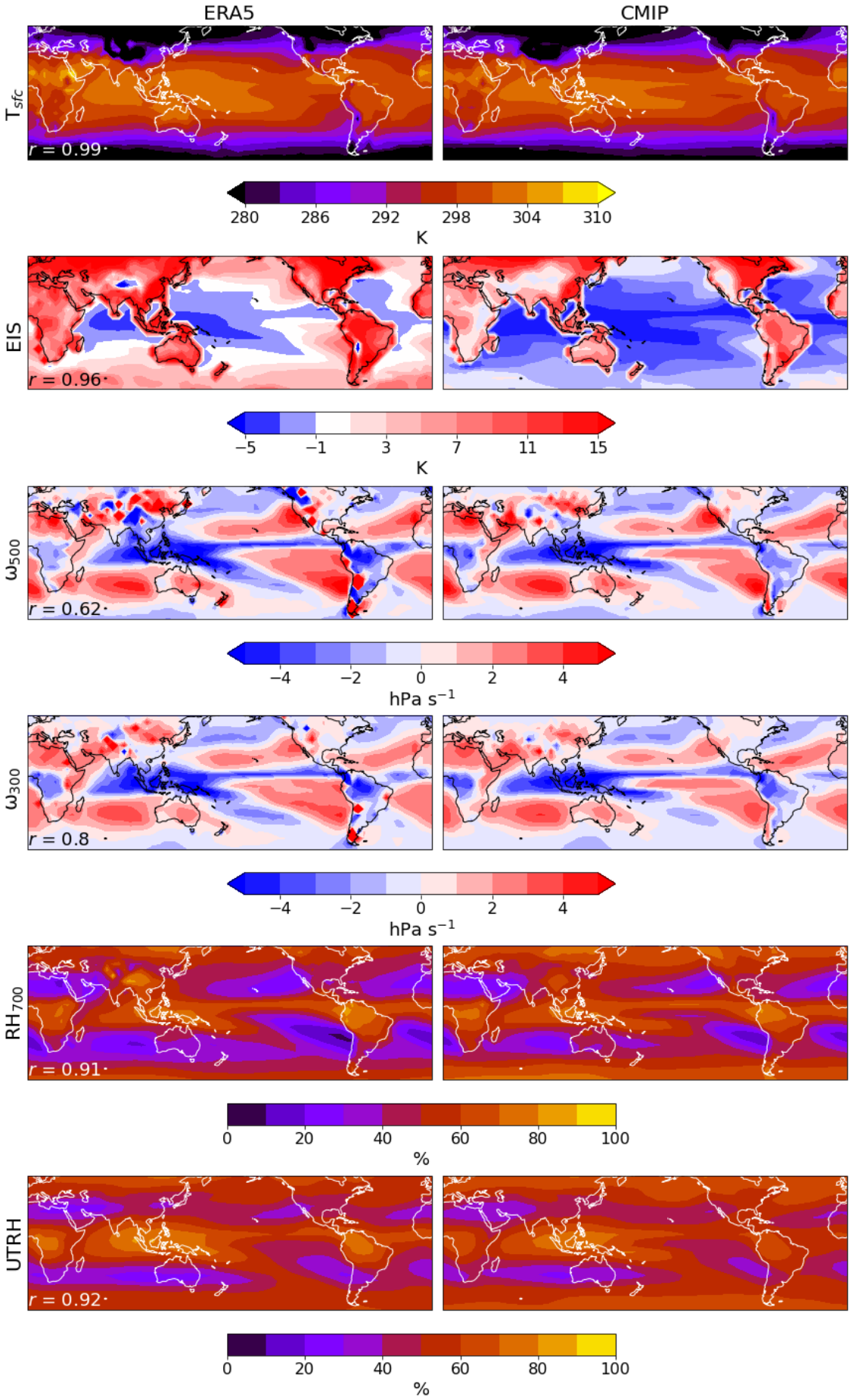
S2. Standard deviation weighting

We note that not all regions are equally cloudy, which leads to differently high levels of variance in R_{LW} and, thus, signals for the regression to learn from. As a result, performance metrics tend to be lower in subsidence regions where there are few high clouds. For the global performance averages, we have addressed this issue by weighting grid-cells based upon their climatological mean R_{LW} in addition to the cosine of their latitude, to avoid penalising average metrics from low scores in regions with relatively little signal.

S3. CCF importance at different spatial scales

Here we expand on our analysis of complex interplay between spatial domain, model dimensions, and predictive skill. Though the most important predictor, UTRH's explained variance fraction (EVF) robustly decreases as a function of domain size (shown in Fig. S6c-f). Inspection of composite spatial sensitivities reveals that the magnitudes of the regression coefficients are largest close to the target grid cells (typically within the local optimal 7x3 domain), with noisy patterns further afield (shown in Fig. S6a-b). We find this is also true for ω_{300} and (though to a lesser extent) RH_{700}. However, for a few of the predictors (ΔU_{300}, T_{sfc} and S_{UT}), adding non-local data adds valuable information, underlining the advantages of including non-local predictors in general, also showing increasing EVF with domain size. Note that there are several mechanisms that may be associated with non-local sensitivities, including remote SST pattern effects for deep convection (Fueglistaler, 2019), the transferral of cloud from one grid-cell to another, or upstream/downstream advection of the meteorological drivers.

We thus propose that a trade-off in predictive skill exists; while increasing domain size may add more relevant information for "non-local" predictors (such as T_{sfc}), superfluous information may be added for more "local" predictors (such as UTRH and ω_{300}) which average out at the globally-aggregated scales, all while increasing the number of model dimensions and thus supressing local predictive skill. The visible patterns in the regression coefficients thus underpin the empirical result that too distant information does not provide additional predictive skill, at least to the degree that it would outweigh the corresponding increase in dimensionality of the regression problem. Though this may suggest that different predictors have different optimal domain sizes, we use equal domain sizes to ensure coefficients are weighted equally (owing to collinearity of the predictors).


ERA5
CMIP
T_{sfc}
r = 0.99
280
286
292
298
304
310
K
EIS
r = 0.96
−5
−1
3
7
11
15
K
ω_{500}
r = 0.62
−4
−2
0
2
4
hPa s^{-1}
ω_{300}
r = 0.8
−4
−2
0
2
4
hPa s^{-1}
RH_{700}
r = 0.91
0
20
40
60
80
100
%
UTRH
r = 0.92
0
20
40
60
80
100
%

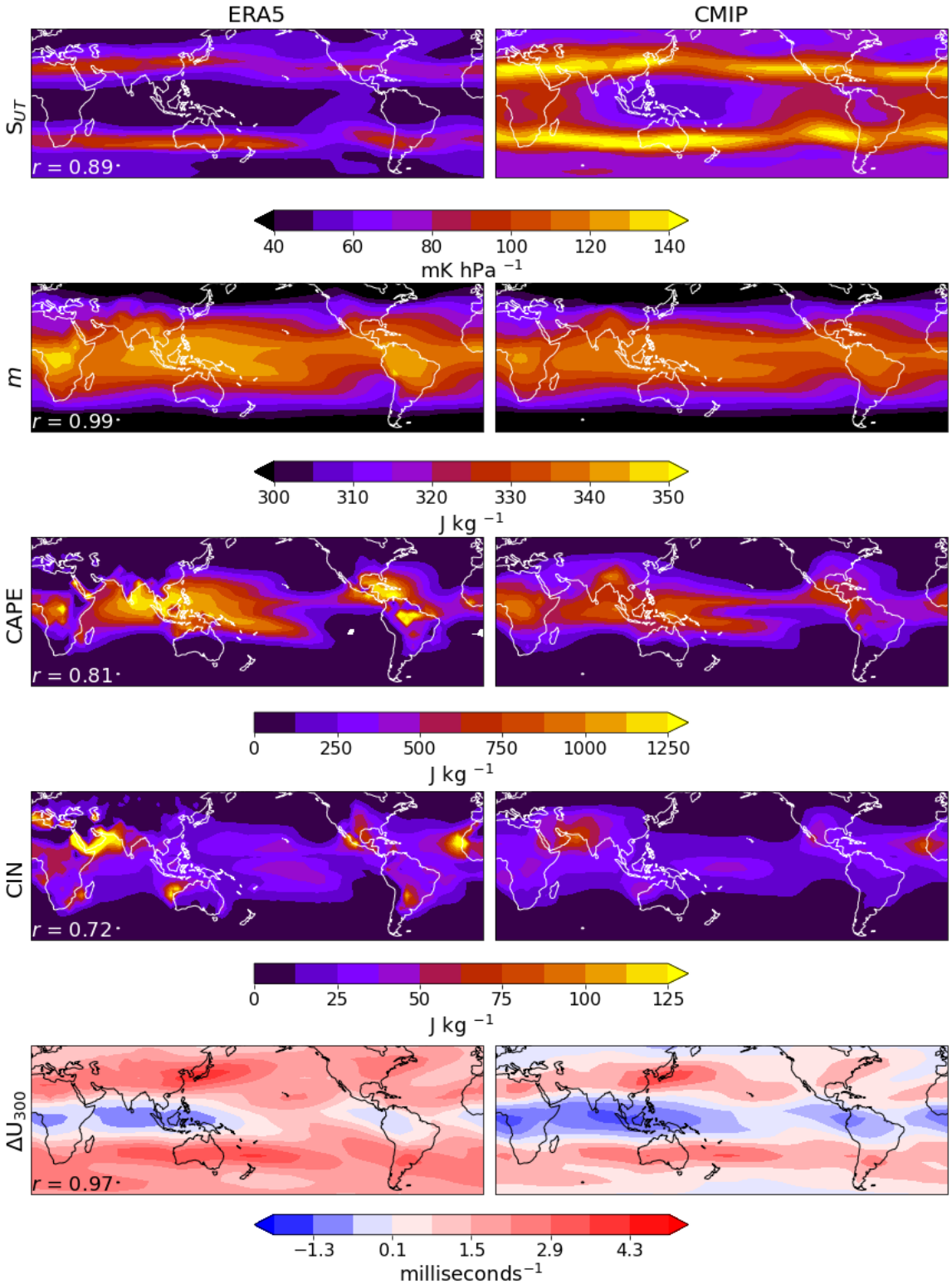


Figure S1. ERA5 (left column) and multi-model mean CMIP (right column) monthly climatological mean for the CN21 and candidte CCFs (from top to bottom, surface temperature, estimated inversion strength, vertical velocity at 500hPa and 300hPa, relative humidity at 700hPa and in the upper troposphere, static stability in the upper troposphere, boundary layer moist static energy, convective available potential energy, convective inhibition, and easterly wind shear. Note different scales and units on the colourbars for each variable. The Pearson r correlation coefficient for the spatial distribution of ERA5 and multi-model CMIP mean values is shown in the bottom left of each ERA5 figure.

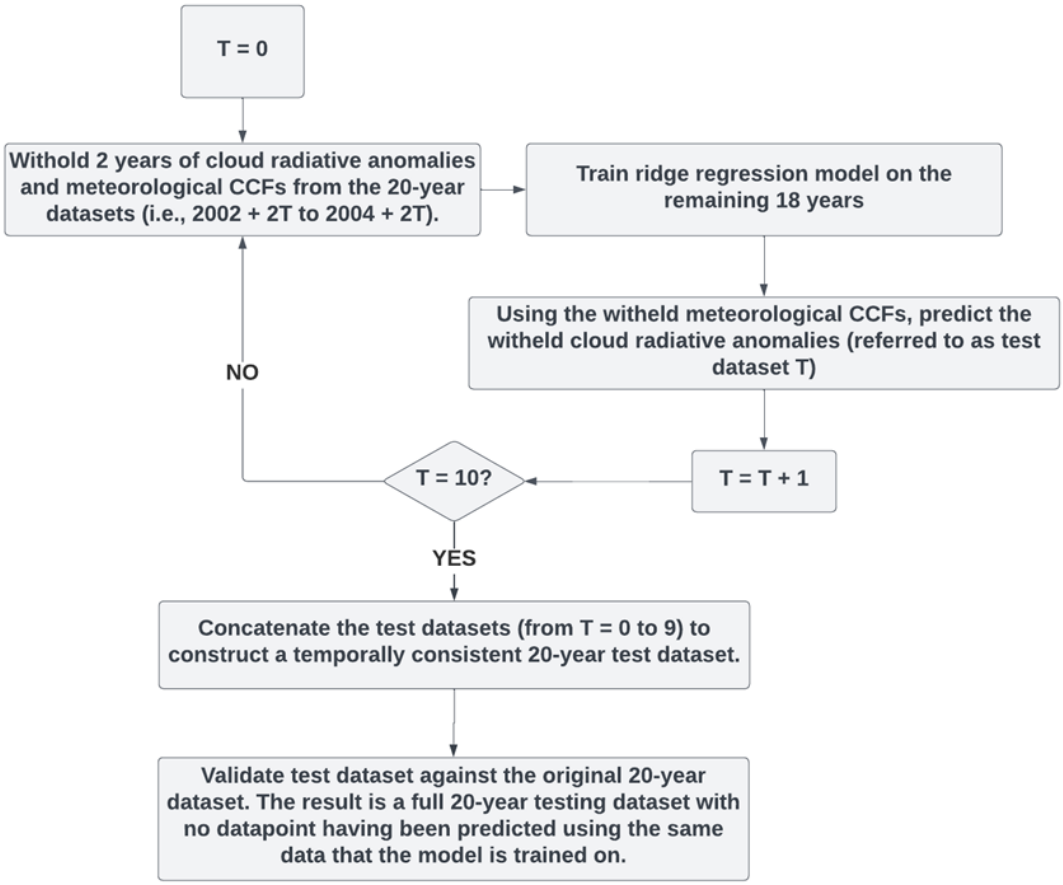


Figure S2. Schematic showing the process used to train and test the cloud controlling factor configurations. This method results in a continuous 20-year validation dataset, with no data having been predicted using the corresponding months in the training data.

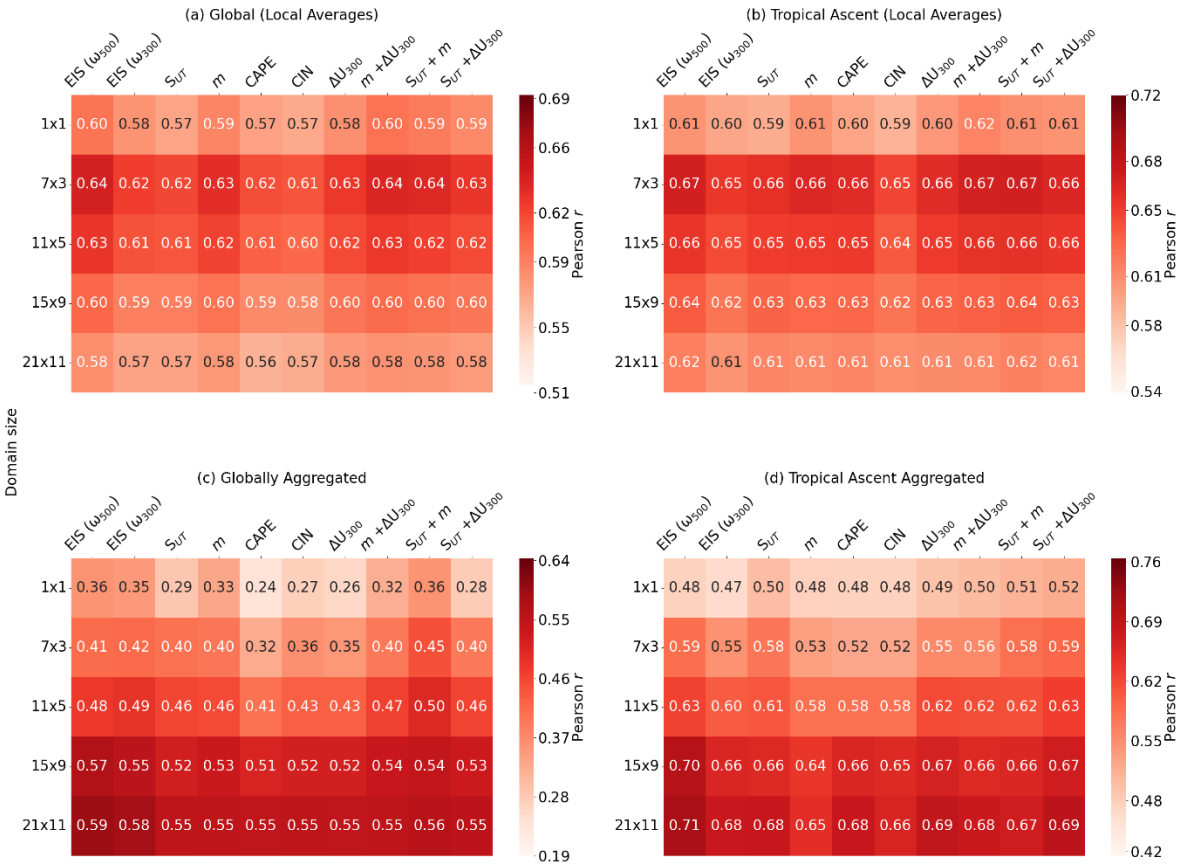


Figure S3. Matrices showing Pearson *r* for predictions made for the observed R_{NET} time series at each domain size using different "CCF configurations". A "CCF configuration" refers to the selection of cloud controlling factors used to predict R_{NET}. Each configuration uses T_{sfc}, RH_{700}, UTRH and ω_{300} (with the exception of the first column, where ω_{500} is used instead) and a candidate CCF(s) (e.g., S_{UT}), which is used to label each column. Predictions are made locally, with the Pearson r averaged (a) globally and (b) in tropical ascent regions defined as grid-cells with observed climatological EIS < 1 K, $\omega_{500} < 0$ hPa s^{-1}. Metrics are weighted by the cosine of latitude and monthly standard deviation of R_{LW} of each grid-cell. Pearson *r* is also shown for aggregated predictions, (c) globally and (d) in the tropical ascent regions, and compared to similarly aggregated observations. Note different scales for each colorbar.

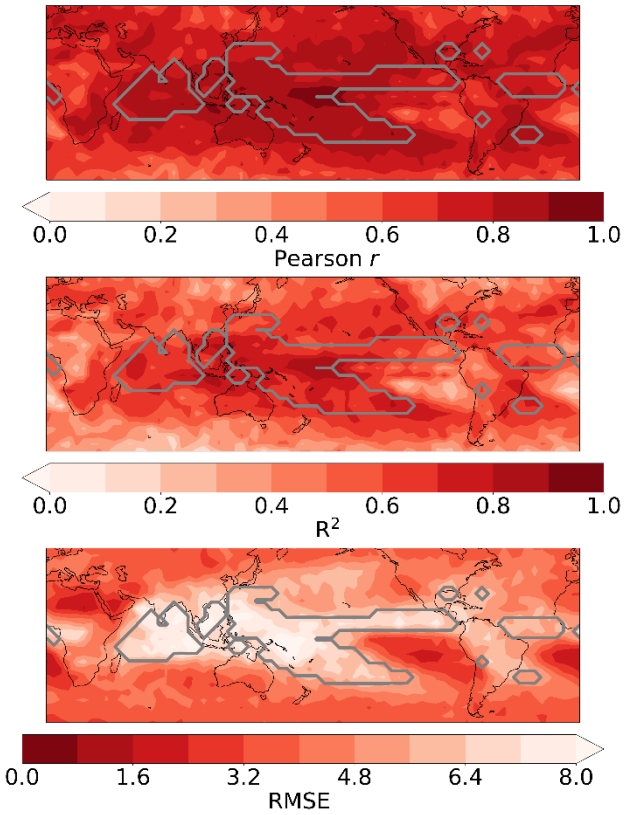


Figure S4. Multi-model median skill metrics (from top to bottom: Pearson r, R^2 score, and RMSE) for predictions of R_{LW} time series using the configuration $S_{UT} + \Delta U_{300}$ (also with T_{sfc}, RH_{700}, UTRH and ω_{300}), with CCFs within a spatial domain of 21x11. Grey contours show the tropical ascent regions. This figure is analogous to Figure 3 in the main text, where skill metrics for the observations are shown instead. Predictive skill is strongest in the tropics where high clouds are ubiquitous and there is a large standard deviation in monthly R_{LW}, resulting in a strong signal for the regression model to learn from. Lower scores are present in the subsidence regions and the Southern Ocean. In addition to weaker R_{LW} signals, poorer performance in the Southern Ocean may be attributable to a reduced quality of reanalysis data, arising from fewer observations available for assimilation (Mallet et al., 2023). Performance metrics are high in the tropics, leaving less room for improvement. Regardless, we find most of our configurations increase (decrease) local Pearson r (RMSE; see Fig. S5) compared to EIS (ω_{300}). Larger improvements are present in the extratropics, where ΔU_{300} leads to improvements.

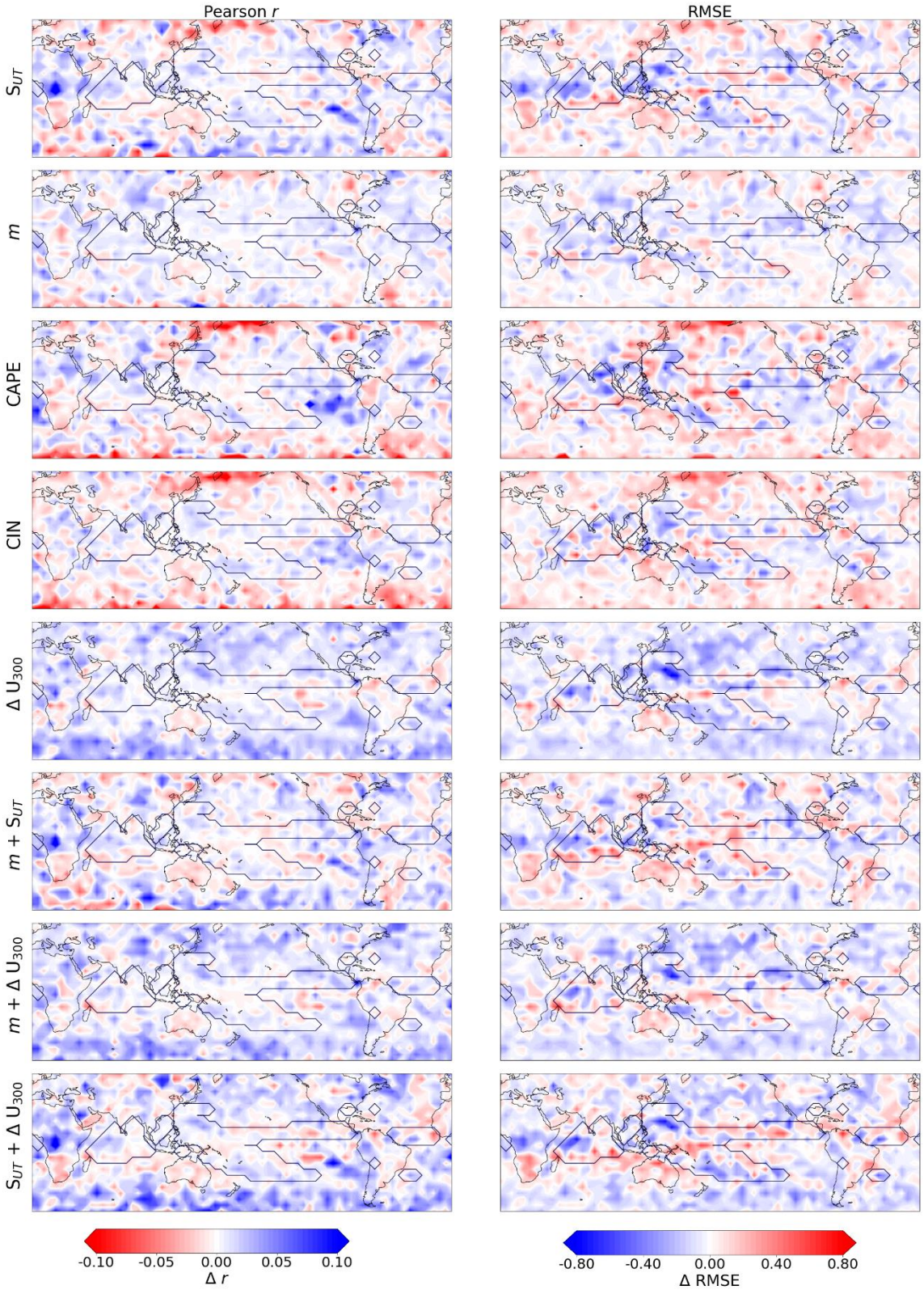


Figure S5. The % difference in local Pearson r (left column) and RMSE (right column) scores for predictions made of observed R_{LW} compared to configuration *EIS* (also with $T_{sfc,}$ RH_{700}, UTRH and ω_{300}), using a 21x11 domain for each configuration. A "CCF configuration" refers to the selection of cloud controlling factors used to predict R_{LW}. Each configuration uses T_{sfc}, RH_{700}, UTRH and ω_{300} and a candidate CCF(s) (e.g., S_{UT}), which is used to label each row. Red indicates regions where the configuration labelled on the LHS preforms better than *EIS* (ω_{300}) (and visa versa for blue). Note the reversed colorbars for each column. This is to indicate where performance is improved, corresponding to an increase in Pearson r or decrease in RMSE.

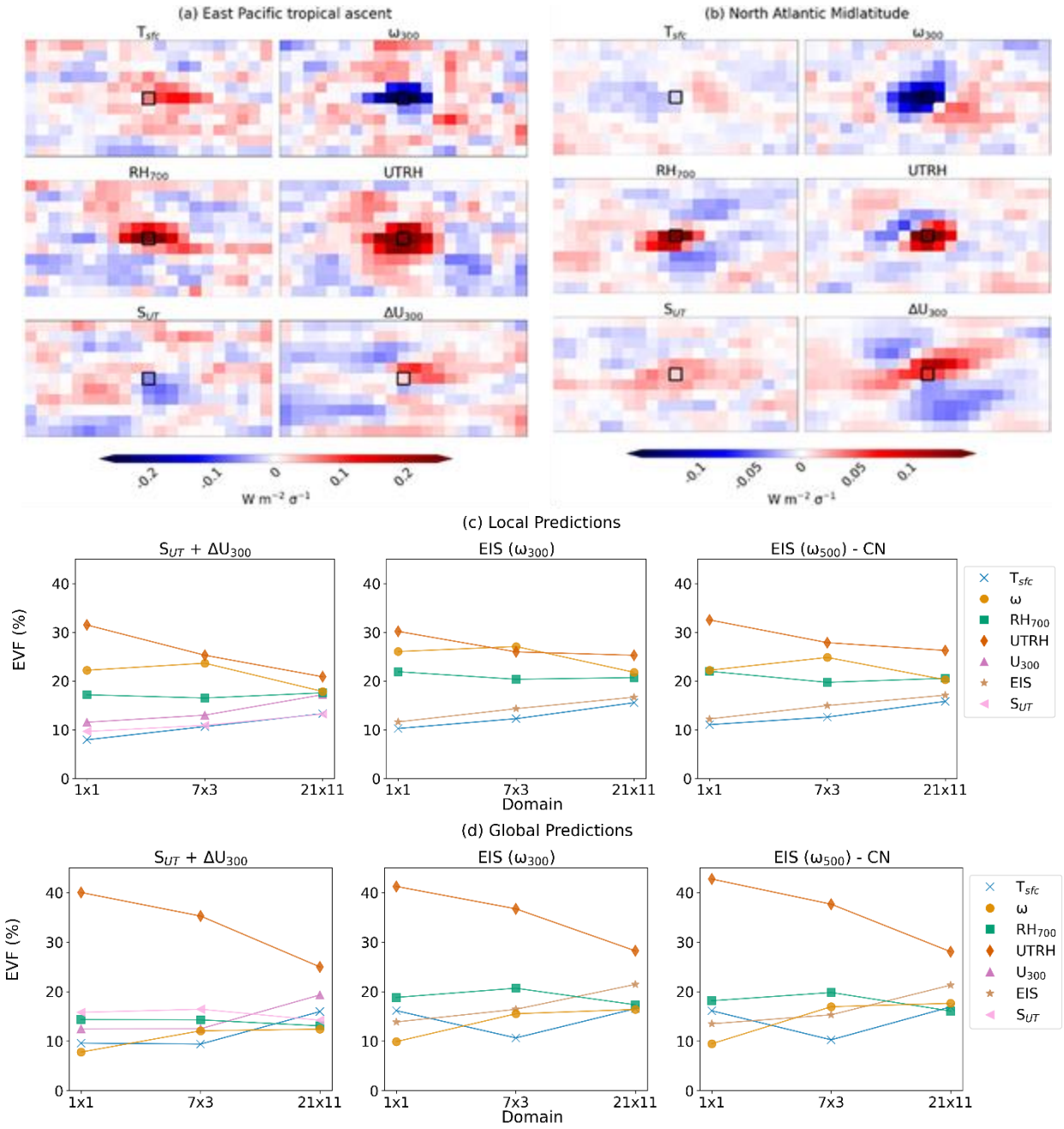


Figure S6. Composite spatial sensitivities using the 21x11 domain and configuration $S_{UT} + \Delta U_{300}$ (with additional core CCFs T_{sfc}, RH_{700}, UTRH, and ω_{300}) in (a) tropical ascent grid-cells (defined by climatological mean EIS < 1 K, and ω_{500} < 0 hPa s^{-1}) in the East Pacific (130°W to 80°W) and (b) North Atlantic (60°W to 10°E, latitudes north of 30°N) midlatitude clouds (climatological mean EIS > 1 K, and ω_{500} < 1.5x10^{-4} hPa s^{-1}). Panel (c) shows the global mean EVF as a function of cloud controlling factor and domain size for local predictions. Note that the global mean EVF has only been weighted based on latitude, and not as a function of R_{LW} standard deviation. Panel (d) shows the EVF for globally-aggregated predictions.

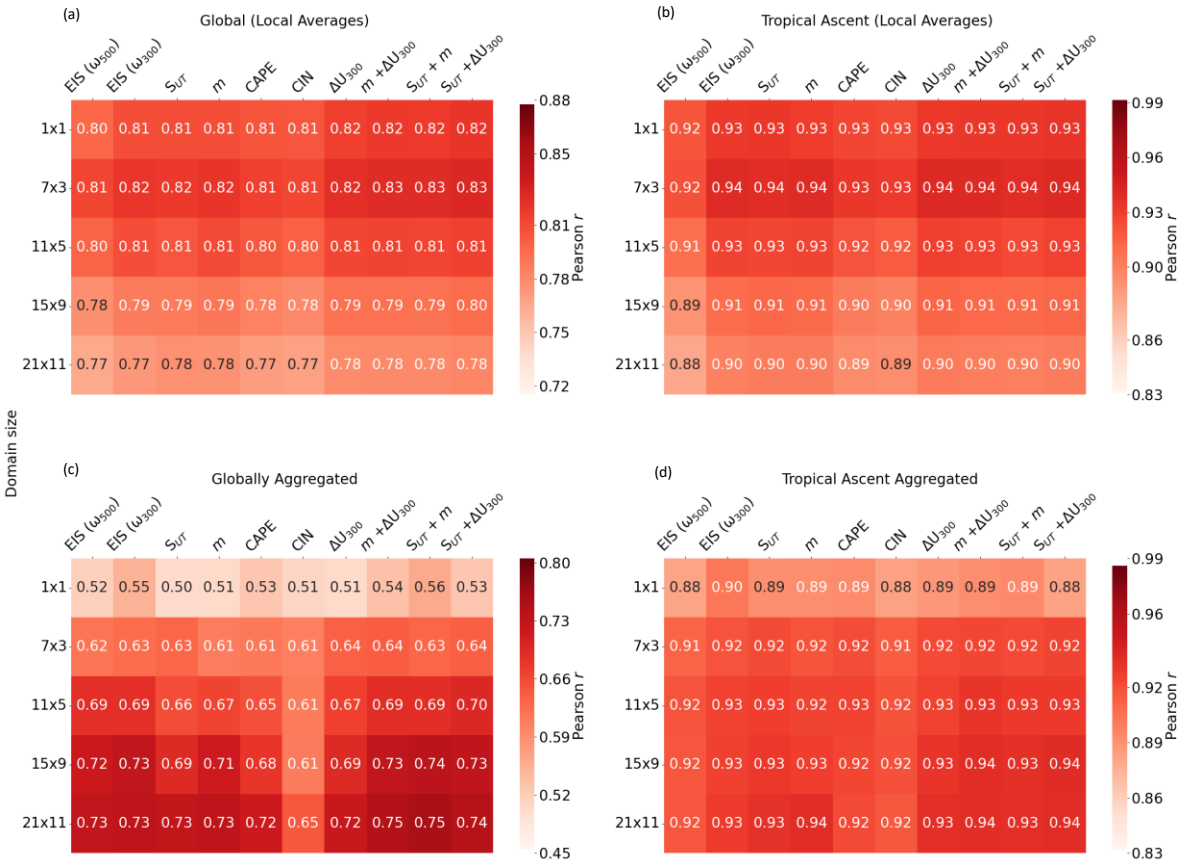


Figure S7. Matrices showing multi-model median Pearson *r* for predictions made for CMIP R_{LW} time series at each domain size using different "CCF configurations". A "CCF configuration" refers to the selection of cloud controlling factors used to predict R_{NET}. Each configuration uses T_{sfc}, RH_{700}, UTRH and ω_{300} (with the exception of the first column, where ω_{500} is used instead) and a candidate CCF(s) (e.g., S_{UT}), which is used to label each column. Predictions are made locally, with the Pearson r averaged (a) globally and (b) in tropical ascent regions defined as grid-cells with observed climatological EIS < 1 K, $\omega_{500} < 0$ hPa s^{-1}. Metrics are weighted by the cosine of latitude and monthly standard deviation of R_{LW} of each grid-cell. Pearson *r* is also shown for aggregated predictions, (c) globally and (d) in the tropical ascent regions, and compared to similarly aggregated observations. Note different scales for each colorbar.

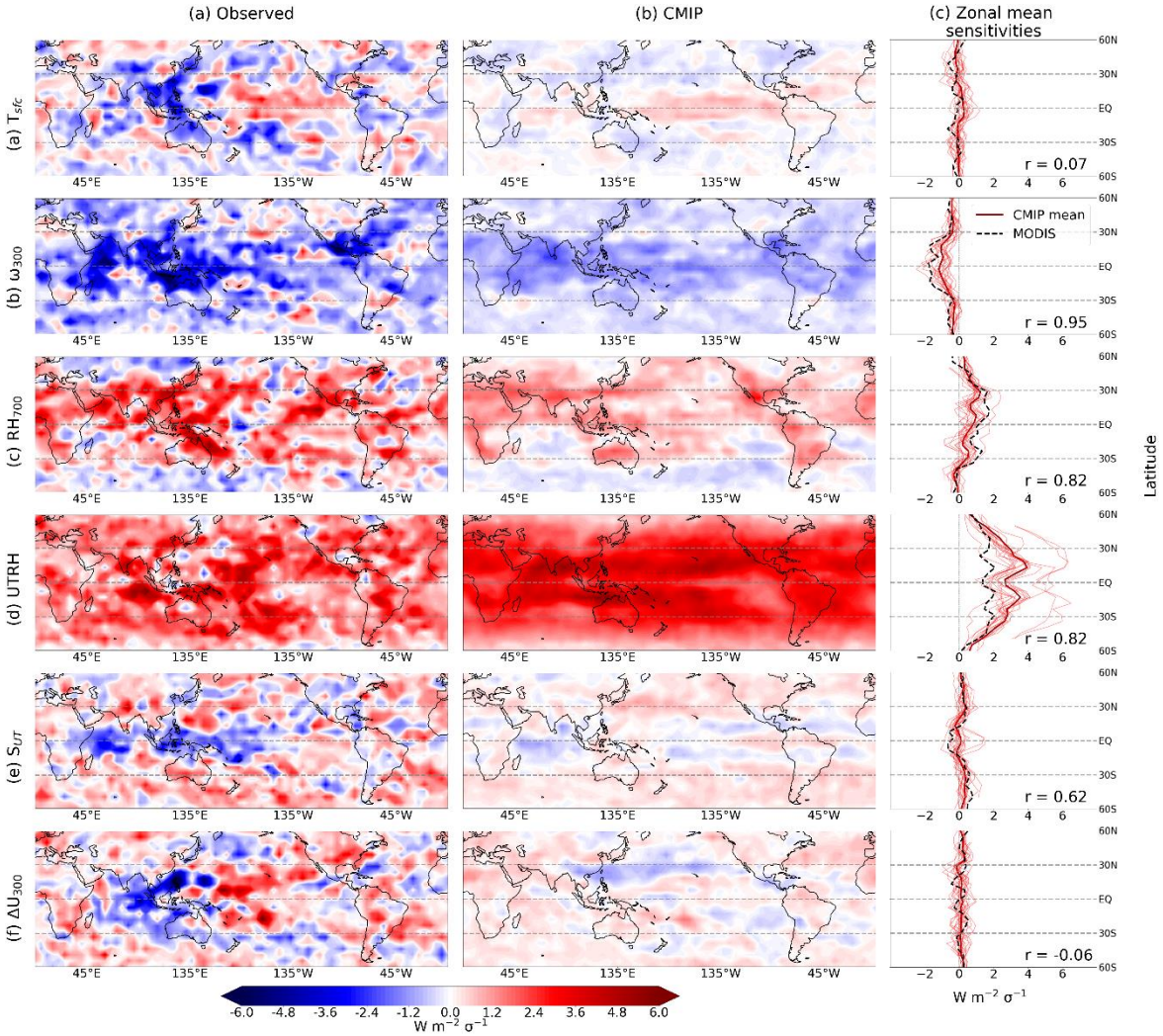


Figure S8. Sensitivities ($\sum \boldsymbol{\theta_i}$) to the cloud controlling factors in configuration $\boldsymbol{S_{UT} + \Delta U_{300}}$(also with T_{sfc}, RH_{700}, UTRH and ω_{300}) for longwave radiative anomalies caused by changes in cloud fraction, derived using a 21x11 domain and defined for a one-standard deviation anomaly in each CCF. To produce the maps, we sum all elements of the sensitivity vectors at each point r. Sensitivities are shown for the observations (first two columns) and the multi-model mean (last two panels). The same colorbar has been used as Figure S9 to show the relative strengths of the sensitivities.

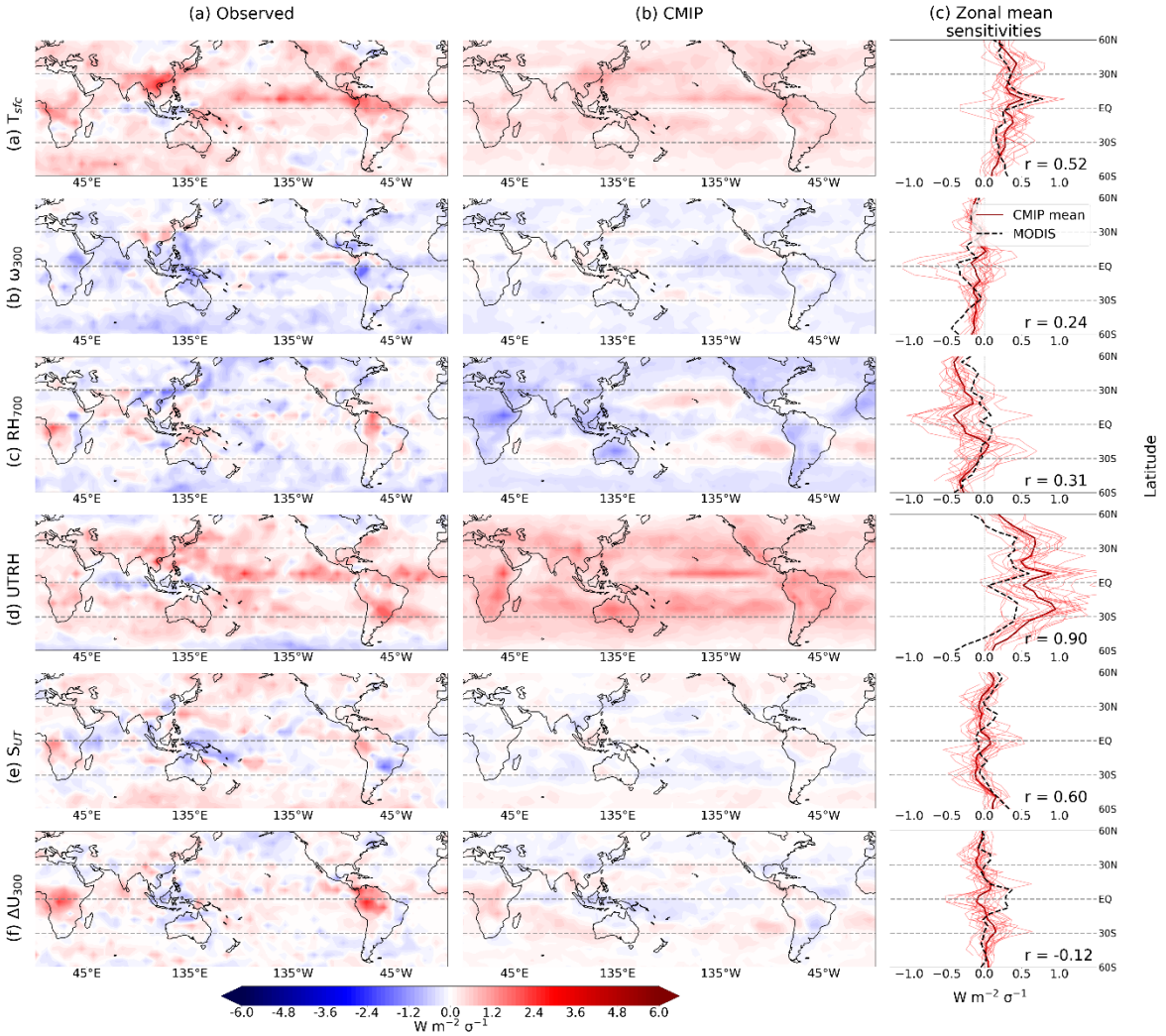


Figure S9. Sensitivities ($\sum \theta_i$) to the cloud controlling factors in configuration $S_{UT} + \Delta U_{300}$(also with T_{sfc}, RH_{700}, UTRH and ω_{300}) for longwave radiative anomalies caused by changes in cloud top pressure, derived using a 21x11 domain and defined for a one-standard deviation anomaly in each CCF. To produce the maps, we sum all elements of the sensitivity vectors at each point r. Sensitivities are shown for the observations (first two columns) and the multi-model mean (last two panels). The same colorbar has been used as Figure S8 to show the relative strengths of the sensitivities.

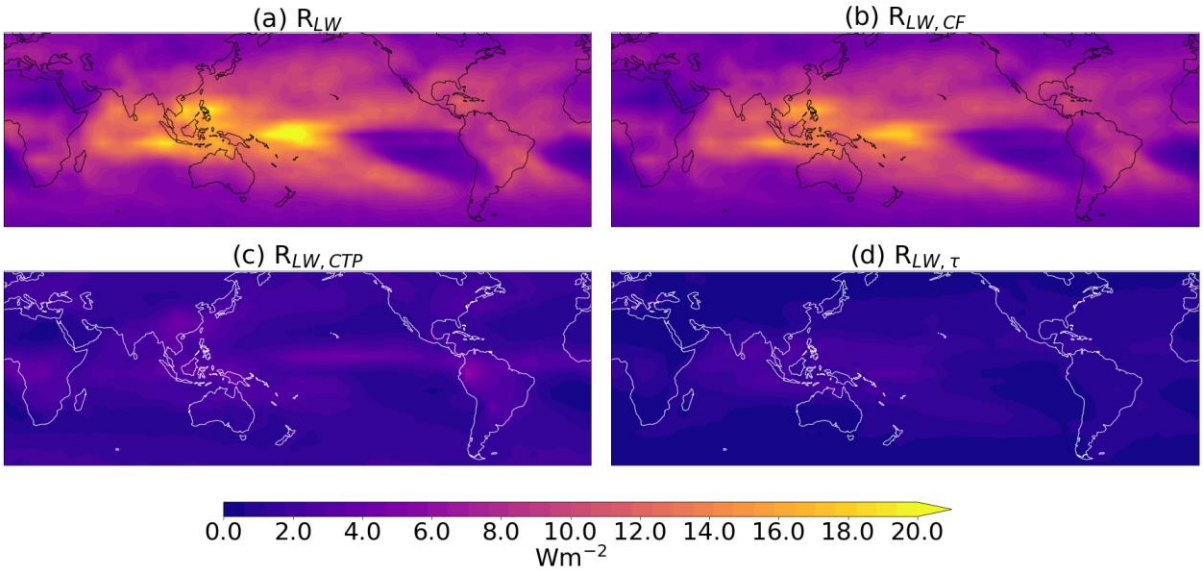


Figure S10. Observed monthly climatologies for (a) longwave high-cloud radiative anomalies and the decompositions into radiative anomalies induced by changes in (b) cloud fraction (c) cloud top pressure and (d) optical depth.

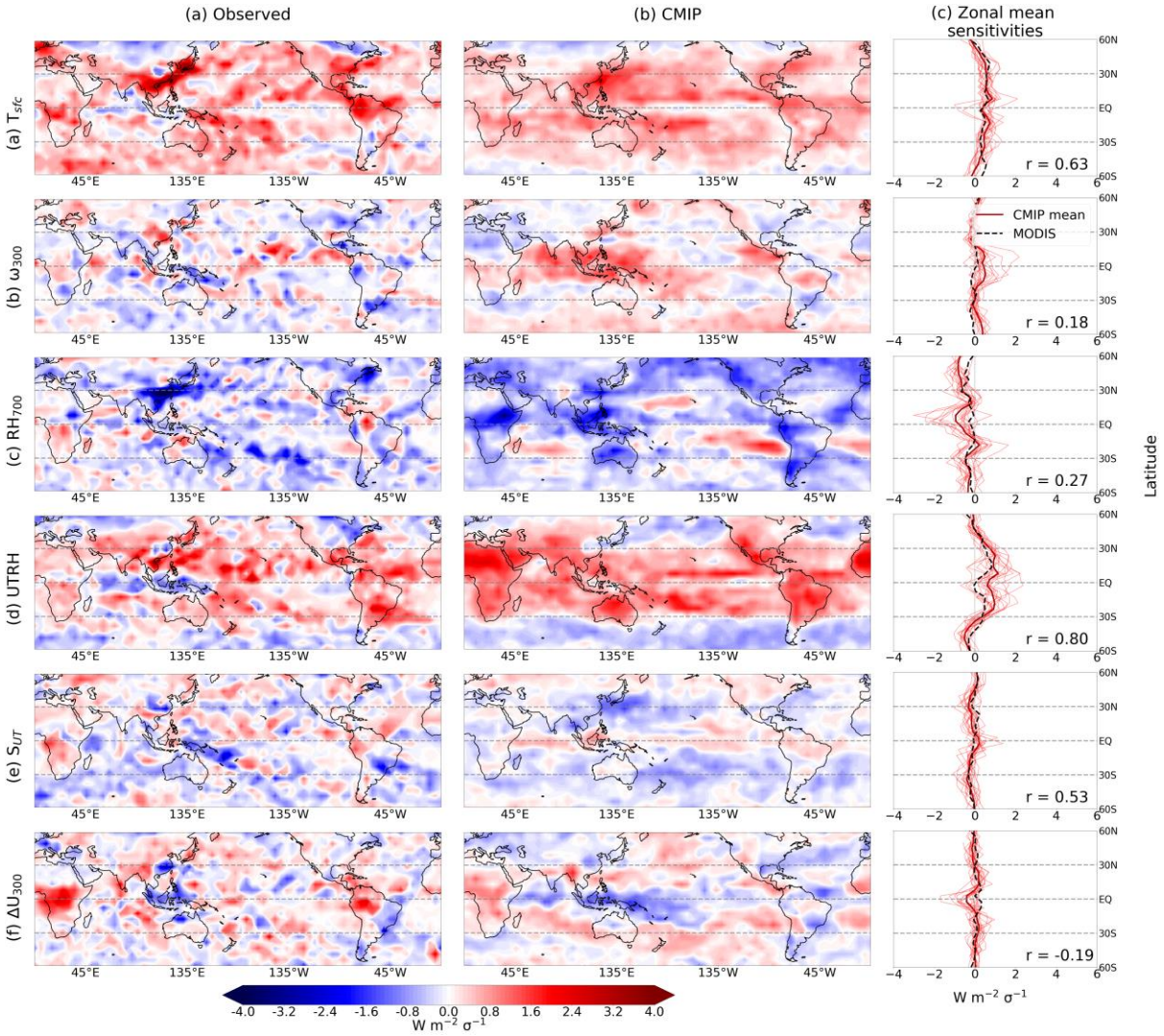


Figure S11. Sensitivities ($\sum \boldsymbol{\theta_i}$) to the cloud controlling factors in configuration $\boldsymbol{S_{UT} + \Delta U_{300}}$(also with T_{sfc}, RH_{700}, UTRH and ω_{300}) for net radiative anomalies caused by changes in cloud fraction, derived using a 21x11 domain and defined for a one-standard deviation anomaly in each CCF. To produce the maps, we sum all elements of the sensitivity vectors at each point r. Sensitivities are shown for the observations (first two columns) and the multi-model mean (last two panels).

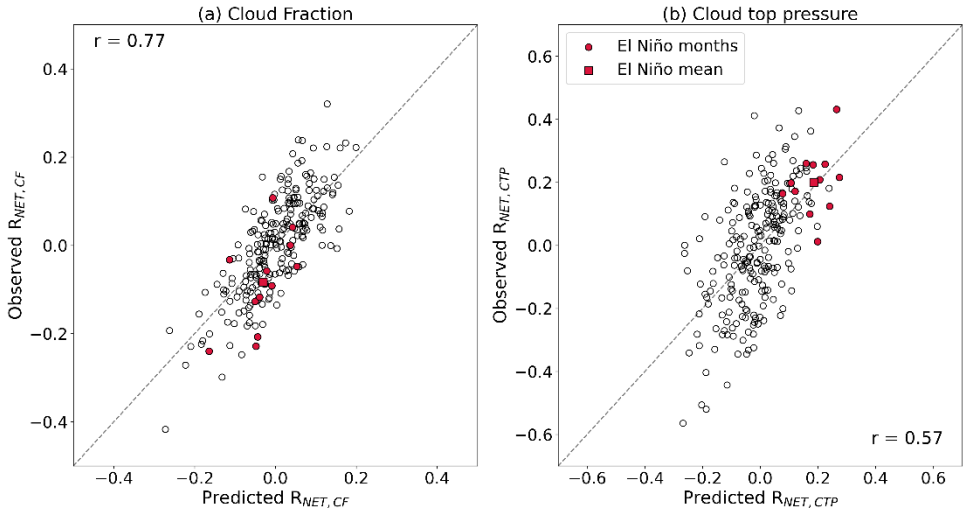


Figure S12. Scatter plot showing the correlation between observed and predicted monthly globally-aggregated (a) $R_{NET,CF}$ and (b) $R_{NE,CTP}$ time series using configuration $S_{UT} + \Delta U_{300}$ (in addition to T_{sfc}, RH_{700}, UTRH, and ω_{300}) and a 21x11 domain. El Niño months are shown using coloured circles, with the annual mean shown using a coloured square. Solid lines show y = x, and the dashed lines show the line-of-best fit through the points.

References

Fueglistaler, S.: Observational Evidence for Two Modes of Coupling Between Sea Surface Temperatures, Tropospheric Temperature Profile, and Shortwave Cloud Radiative Effect in the Tropics, Geophysical Research Letters, 46, 9890–9898, https://doi.org/10.1029/2019GL083990, 2019.

Mallet, M. D., Alexander, S. P., Protat, A., and Fiddes, S. L.: Reducing Southern Ocean Shortwave Radiation Errors in the ERA5 Reanalysis with Machine Learning and 25 Years of Surface Observations, Artificial Intelligence for the Earth Systems, 2, https://doi.org/10.1175/AIES-D-22-0044.1, 2023.