Machine learning of cloud types shows higher climate sensitivity is associated with lower cloud biases

Peter Kuma[1], Frida A.-M. Bender[1], Alex Schuddeboom[2], Adrian J. McDonald[2], and Øyvind Seland[3]

[1]Department of Meteorology (MISU), Stockholm University, Stockholm, Sweden
[2]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, Aotearoa New Zealand
[3]Norwegian Meteorological Institute, Oslo, Norway

Correspondence: Peter Kuma (peter.kuma@misu.su.se)

Abstract. Uncertainty in cloud feedback in climate models is a major limitation in projections of future climate. Therefore, evaluation and improvement of cloud simulation is essential. We analyse cloud biases and cloud change with respect to global mean near-surface temperature (GMST) in climate models relative to satellite observations, and relate them to equilibrium climate sensitivity, transient climate response and cloud feedback. For this purpose, we develop a supervised deep convolutional artificial neural network for determination of cloud types from low-resolution (approx. $1^\circ \times 1^\circ$) daily mean top of atmosphere shortwave and longwave radiation fields, corresponding to the World Meteorological Organization (WMO) cloud genera recorded by human observers in the Global Telecommunication System. We train this network on a satellite top of atmosphere radiation observed by the Clouds and the Earth's Radiant Energy System (CERES), and apply it on the Climate Model Intercomparison Project phase 5 and 6 (CMIP5 and CMIP6) historical and abrupt-4xCO2 experiment model output and the ERA5 and the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) reanalyses. We compare these with satellite observations, link biases in cloud type occurrence derived from the neural network to change with respect to GMST to climate sensitivity, and compare our cloud types with an existing cloud regime classification based on the Moderate Resolution Imaging Spectroradiometer (MODIS) and International Satellite Cloud Climatology Project (ISCCP) satellite data. We show that there is a significant negative linear relationship between the root mean square error of cloud type occurrence derived from the neural network and model equilibrium climate sensitivity and transient climate response (Bayes factor 22 and 17, respectively). This indicates that models with a better representation of the cloud types globally have higher climate sensitivity. Factoring in results from other studies, there are two possible explanations: either high climate sensitivity models are plausible, contrary to combined assessments of climate sensitivity by previous review studies, or the accuracy of representation of present-day clouds in models is negatively correlated with the accuracy of representation of future projected clouds.

1 Introduction

Clouds are a major factor influencing the Earth's climate. They are highly spatially and temporally variable, with the top of atmosphere (TOA) radiation being particularly sensitive to cloud changes due to their high albedo and impact on longwave radiation. Of all climate feedbacks, cloud feedback is the most uncertain feedback in Earth system models (ESMs) (Zelinka

et al., 2020; Sherwood et al., 2020). Therefore, it is essential that climate models become more convergent on a correct representation of future clouds, but also on their representation of present-day clouds, which is a necessary (but not sufficient) condition for the fidelity of projected cloud change. The estimate of the 'likely' range (66%) of ECS has recently been refined in the 6th Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) to 2.5–4 K, from 1.5–4.5 K in AR5. Evidence for this estimate is only indirectly informed by the Climate Model Intercomparison Project phase 6 (CMIP6) models (Eyring et al., 2016, 2019), which have a multi-model mean of 3.7 K (Meehl et al., 2020). The combined assessment is based on paleoclimate and historical evidence, emergent constraints and process understanding. Notably, CMIP6 models predict a 16% higher multi-model mean than CMIP5 (3.2 K) (Meehl et al., 2020; Forster et al., 2020), and this fact has already been examined in a number of studies (Zelinka et al., 2020; Wyser et al., 2020; Schlund et al., 2020; Dong et al., 2020; Nijsse et al., 2020; Flynn and Mauritsen, 2020). It is also higher than the combined assessment central value of ECS of 3 K in AR6. The multi-model spread in CMIP6 is also larger than in CMIP5, with a standard deviation of 1.1 K in CMIP6 vs. 0.7 K in CMIP5. Modelling groups have prevailingly reported that the higher multi-model mean is due to changes in cloud representation in the recent generation of models (Meehl et al., 2020, Table 3), supported by the findings of Zelinka et al. (2020). Some authors have concluded that the high-ECS CMIP6 models are on average overestimating the ECS (Nijsse et al., 2020) and are not compatible with paleoclimatic records (Zhu et al., 2020, 2021). The high-ECS models are also not supported by the review studies Sherwood et al. (2020) and AR6. In contrast, Bjordal et al. (2020) argue that high ECS models might be plausible because of state-dependent cloud phase feedback in the Southern Ocean. Models which simulate too much ice in the Southern Ocean clouds, a common bias among CMIP models, are expected to have lower cloud feedbacks globally because of a spuriously enhanced negative feedback associated with cloud phase changes in that region. Recently, Volodin (2021) reported that changing cloud parametrisation in the Institute of Numerical Mathematics Coupled Model version 4-8 (INM-CM4-8) from Smagorinsky type to a prognostic type of Tiedtke (1993) resulted in more than doubling of ECS from 1.8 K to 3.8 K. This underscores the importance of cloud parametrisation in determining model climate sensitivity.

The effect of clouds on the climate comes primarily from cloud fraction and cloud optical depth, which are determined by factors and properties such as convection, mass flux, turbulence, atmospheric dynamics, cloud microphysics (cloud phase, cloud droplet and ice crystal size distribution, number concentration, ice crystal habit), vertical cloud overlap, cloud altitude, cloud cell structure, cloud lifetime and more. Simulation of clouds accurately is problematic not only because of the large number of properties, many of which are subgrid-scale in today's general circulation models, but also because compensating model biases may produce correct cloud radiative effect (CRE) while simulation of the individual properties is incorrect. This may be especially true for global radiation budget, as models processes are often tuned to achieve a desired radiation balance at TOA (Hourdin et al., 2017; Schmidt et al., 2017).

Cloud genera (WMO, 2021a) have been an established way of describing clouds for over a century. They broadly correspond or correlate with the individual cloud properties such as cloud altitude, optical depth, phase, overlap and cell structure. Therefore, they can be used as a metric for model evaluation which, unlike more synthetically-derived cloud classes, is easy to understand and has a very long observational record. So far, however, it has not been possible to identify cloud genera in low-resolution model output, because their identification depends on a high-resolution visual observation, generally from the

ground. Here, we show that it is possible to use a supervised deep convolutional artificial neural network (ANN) to identify cloud genera in low-resolution model output and satellite observations. Past classifications of cloud types or cloud regimes derived from satellite datasets have been based on cloud optical depth and cloud top pressure or height by simple partitioning (Rossow and Schiffer, 1991) or by statistical clustering algorithms (Jakob and Tselioudis, 2003; McDonald et al., 2016; Oreopoulos et al., 2016; Cho et al., 2021), and on active radar and lidar sensors (Cesana et al., 2019), which may only broadly correspond to human-observed cloud genera. More recently, deep ANNs have begun to be used (Zantedeschi et al., 2020).

We introduce a new method of quantifying cloud types in model and satellite data based on an ANN approach. Furthermore, we quantify their global distribution and change with respect to global mean near-surface temperature (GMST) in CMIP5 and CMIP6 models, the Clouds and the Earth's Radiant Energy System (CERES) satellite data and two reanalyses, European Centre for Medium-Range Weather Forecasts's (ECMWF) reanalysis ERA5 (Hersbach et al., 2020) and the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) (Gelaro et al., 2017). Convolutional artificial neural networks have been used before for cloud detection: Shi et al. (2017), Ye et al. (2017), Wohlfarth et al. (2018), Zhang et al. (2018), Liu and Li (2018) and Zantedeschi et al. (2020) used a convolutional ANN for identification of cloud genera in ground-based cloud images, and Drönner et al. (2018), Shendryk et al. (2019), Guo et al. (2020), Segal-Rozenhaimer et al. (2020) and Liu et al. (2021) developed a convolutional ANN for detecting cloudy pixels in high-resolution satellite imagery. While the determination of cloud types in model and satellite data and application of ANNs to identify cloud types are not new, in contrast to previous methods, we utilise cloud types with a direct correspondence to the established human-observed World Meteorological Organization (WMO) cloud genera. This dataset contains many decades of global cloud observations, recorded several times daily at a large number of stations. For this purpose, we develop an ANN which can be applied to input with low spatial and temporal resolution ($\approx 1^\circ$, daily mean). This is because most current climate models provide output with low resolution, which does not contain enough information to recognise individual clouds, but these can still be inferred statistically from large scale patterns. Likewise, the resolution of some satellite datasets such as CERES is on this spatial scale. Due to limitations of the method, the primary output of the ANN is global mean cloud type occurrence, but geographically-resolved cloud type occurrence with a resolution of approx. 4000×4000 km can be derived. We try to answer the question of whether cloud type biases and change with respect to GMST are related to cloud feedback, ECS and TCR in the CMIP models. Previously, Schuddeboom and McDonald (2021) found no link between ECS and shortwave CRE mean and compensating errors. The ANN and the associated code is made available under an open source license (Kuma et al., 2022).

2 Methods

2.1 Data

2.1.1 Satellite observations

We used satellite observations as a reference training dataset for the ANN. These were sourced from the Clouds and the Earth's Radiant Energy System (CERES) in years 2003–2020 (Wielicki et al., 1996; Doelling et al., 2013; Loeb et al., 2018).

In particular, we used the SYN1deg 1°-resolution geostationary-enhanced product Terra+Aqua Edition 4.1. Variables used in the training phase of ANN were the daily mean adjusted all-sky and clear sky shortwave and longwave fluxes at TOA, and shortwave (solar) insolation.

2.1.2 Climate models

CMIP5 and CMIP6 are the last two iterations of standardised global climate model experiments (Taylor et al., 2012; Eyring et al., 2016). We applied our ANN to the publicly-available model output of the *historical* and *abrupt-4xCO2* CMIP experiments in the daily mean products. An exception was the EC-Earth model, which did not provide the necessary variables in the historical experiment. For this model, we used data from the *hist-1950* experiment (model EC-Earth3P) of the High Resolution Model Intercomparison Project (HighResMIP) (Haarsma et al., 2016, 2020) as a substitute for historical. The model output resolution of EC-Earth3P is the same as EC-Earth. The CMIP model output used in our analyses were the variables *rsut* (TOA outgoing shortwave radiation), *rlut* (TOA outgoing longwave radiation), *rsutcs* (TOA outgoing clear-sky shortwave radiation), *rlutcs* (TOA outgoing clear-sky longwave radiation), *rsdt* (TOA incident shortwave radiation) and *tas* (GMST). In connection to CMIP models, we used estimates of the model ECS, TCR and cloud feedback from AR6, with missing values supplemented by Meehl et al. (2020), Zelinka et al. (2020), and ECS and TCR calculated with the ESMValTool version 2.4.0 (Righi et al., 2020). Here, we use a definition of cloud feedback adjusted for non-cloud influences as in Zelinka et al. (2020), Soden et al. (2008) and Shell et al. (2008). Table 1 lists CMIP5 and CMIP6 models used in our analysis, their ECS, TCR and cloud feedback. In total, we analysed 4 CMIP5 and 20 CMIP6 models, of which 18 had the necessary data in the historical experiment for comparison with CERES (years 2003–2014), and 22 had the necessary data in the abrupt-4xCO2 experiment. No selection was done on the models, i.e. all CMIP5 and CMIP6 models which provided the required fields were analysed here. The required NorESM2-LM model output was not available through the CMIP6 archives and was provided to us by the model developers. For some models, ECS, TCR or CLD were not available. For these models, the values were taken from the closest available model. The model developers of IPSL-CM6A-LR-INCA advised us that its TCR should be the same as IPSL-CM6A-LR (Olivier Boucher, personal communication).

2.1.3 Reanalyses

In addition to CMIP, we analysed the output of two reanalyses: ERA5 (Hersbach et al., 2020) and MERRA-2 (Gelaro et al., 2017). From MERRA-2, we used the *M2T1NXRAD* product: daily means of the variables *LWTUP* (upwelling longwave flux at TOA), *LWTUPCLR* (upwelling longwave flux at TOA assuming clear sky), *SWTDN* (TOA incoming shortwave flux), *SWTNT* (TOA net downward shortwave flux), *SWTNTCLR* (TOA net downward shortwave flux assuming clear sky). From ERA5, we used the *ERA5 hourly data on single levels from 1979 to present* product variables *tsr* (top net solar radiation), *tsrc* (top net solar radiation, clear sky), *ttr* (top net thermal radiation) and *ttrc* (top net thermal radiation, clear sky).

Table 1. Table of CMIP5, CMIP6 models and reanalyses used in our analysis and their CMIP phase, equilibrium climate sensitivity (ECS), transient climate response (TCR), cloud feedback (CLD), if they provided the necessary variables in the *historical* (hist.) (in the case of reanalyses *historical reanalysis*) and *abrupt-4xCO2* experiments ('●' – yes, '○' – no, '-' – not applicable), and model output resolution (Res.) as the number of longitude × latitude bins. Models are sorted by their ECS. ECS, TCR and CLD were sourced from AR6, Zelinka (2021) and Semmler et al. (2021).

#	Model	Phase	ECS (K)	TCR (K)	CLD ($Wm^{-2}K^{-1}$)	hist.	abrupt-4xCO2	Res. (lon.×lat.)
1	INM-CM4-8	6	1.83	1.33	-0.09	●	●	180×120
2	INM-CM5-0	6	1.92	1.3	-0.06	●	●	180×120
3	NorESM2-LM	6	2.54	1.48	0.44	●	●	144×96
4	MRI-CGCM3	5	2.60	1.60	0.28	-	●	320×160
5	MPI-ESM-1-2-HAM	6	2.96	1.8	-0.16	●	●	192×96
6	MPI-ESM1-2-HR	6	2.98	1.66	0.27	●	●	384×192
7	MPI-ESM1-2-LR	6	3.00	1.84	0.18	●	●	192×96
8	MRI-ESM2-0	6	3.15	1.64	0.46	●	●	320×160
9	AWI-ESM-1-1-LR	6	3.29	2.11	0.29[a]	●	○	192×96
10	MPI-ESM-LR	5	3.63	2.00	0.44	-	●	192×96
11	IPSL-CM5A2-INCA	6	3.79	1.9	1.05	●	●	96×96
12	GFDL-CM4	6	3.89	2.10	0.64	○	●	144×90
13	IPSL-CM5A-MR	5	4.12	2.00	1.25	-	●	144×143
14	IPSL-CM5A-LR	5	4.13	2.00	1.18	-	●	96×96
18	IPSL-CM6A-LR-INCA	6	4.13	2.32[a]	0.43	●	●	144×143
15	CNRM-CM6-1-HR	6	4.28	2.48	0.59	●	●	720×360
16	EC-Earth3P	6	4.31[a]	2.62[a]	0.37[a]	●	○	512×256
17	IPSL-CM6A-LR	6	4.56	2.32	0.45	●	●	144×143
19	CNRM-ESM2-1	6	4.76	1.86	0.63	○	●	256×128
20	CNRM-CM6-1	6	4.83	2.14	0.61	●	●	256×128
21	UKESM1-0-LL	6	5.34	2.79	0.87	●	●	192×144
22	HadGEM3-GC31-MM	6	5.42	2.58	0.91	●	●	432×324
23	HadGEM3-GC31-LL	6	5.55	2.55	0.84	●	●	192×144
24	CanESM5	6	5.62	2.74	0.88	●	●	128×64
25	ERA5	-	-	-	-	●	-	1440×721
26	MERRA-5	-	-	-	-	●	-	576×361

[a]For some models, ECS, TCR or CLD were not available. For these models, the values were taken from the closest available model (CLD of AWI-ESM-1-1-LR as in AWI-CM-1-1-MR; ECS, TCR and CLD of EC-Earth3P as in EC-Earth3-Veg; and TCR of IPSL-CM6A-LR-INCA as in IPSL-CM6A-LR).

2.1.4 Station observations

In addition to satellite and model data, we used ground-based land and marine station data from the Historical Unidata Internet Data Distribution (IDD) Global Observational Data (Unidata, 2003). This dataset is a collection of the Global Telecommunication System (GTS) (WMO, 2021b) reports between 2003-05-19 and 2020-12-31 at standard synoptic times (00Z, 03Z, ..., 21Z), with only two time periods of missing data longer than three weeks in 2006 (whole year), December 2007 and 2008 (whole year). We used the cloud genus variables of the synoptic (SYNOP) and marine (BUOY) reports: low cloud (IDD variable 'cloudLow') based on the WMO *Code Table 0513*, middle cloud based on *Code Table 0515* (IDD variable 'cloudMiddle') and high cloud based on *Code Table 0509* (IDD variable 'cloudHigh') (WMO, 2011). Furthermore, we grouped the cloud genera into four categories to simplify our analysis:

1. *high*: cirrus, cirrostratus, cirrocumulus (C_H codes 1–9),
2. *middle*: altostratus, altocumulus (C_M codes 1–9),
3. *cumuliform*: cumulus, cumulonimbus (C_L codes 1–3, 8, 9),
4. *stratiform*: stratocumulus, stratus (C_L codes 4–7).

As an example of the geographical distribution of stations, Fig. 1a shows the location of SYNOP and BUOY station reports with cloud data available on 1 January 2010. Because the data come from operational weather stations, they are geographically biased to certain locations, especially land, extratropics and the Northern Hemisphere. Undersampled locations are ocean, the Southern Hemisphere and the poles. Cloud type information from stations in the USA is also not available in the WHO records. Because of partially missing data in 2003, 2006 and 2008, we excluded these years in the ANN training phase. Not all stations in the IDD database provide cloud type information, and such stations were excluded from our analysis. High and middle clouds can be obscured by underlying cloud layers. In such cases, the observation of high or middle clouds is recorded as missing in the IDD data and we exclude such stations from the calculation of statistics for the middle or high cloud types, respectively. This limitation of the dataset means it is less suitable for identifying middle and high clouds than low clouds in multi-layer cloud situations, and a similar, but reverse, limitation exists in spaceborne cloud observations (McErlich et al., 2021).

2.1.5 Historical global mean near-surface temperature

Historical GMST was sourced from the NASA Goddard Institute for Space Studies (GISS) Surface Temperature Analysis version 4 (GISTEMP v4) (Lenssen et al., 2019; GISTEMP Team, 2021). This dataset was used in combination with the CERES dataset to determine observed change of cloud type occurrence with respect to GMST.

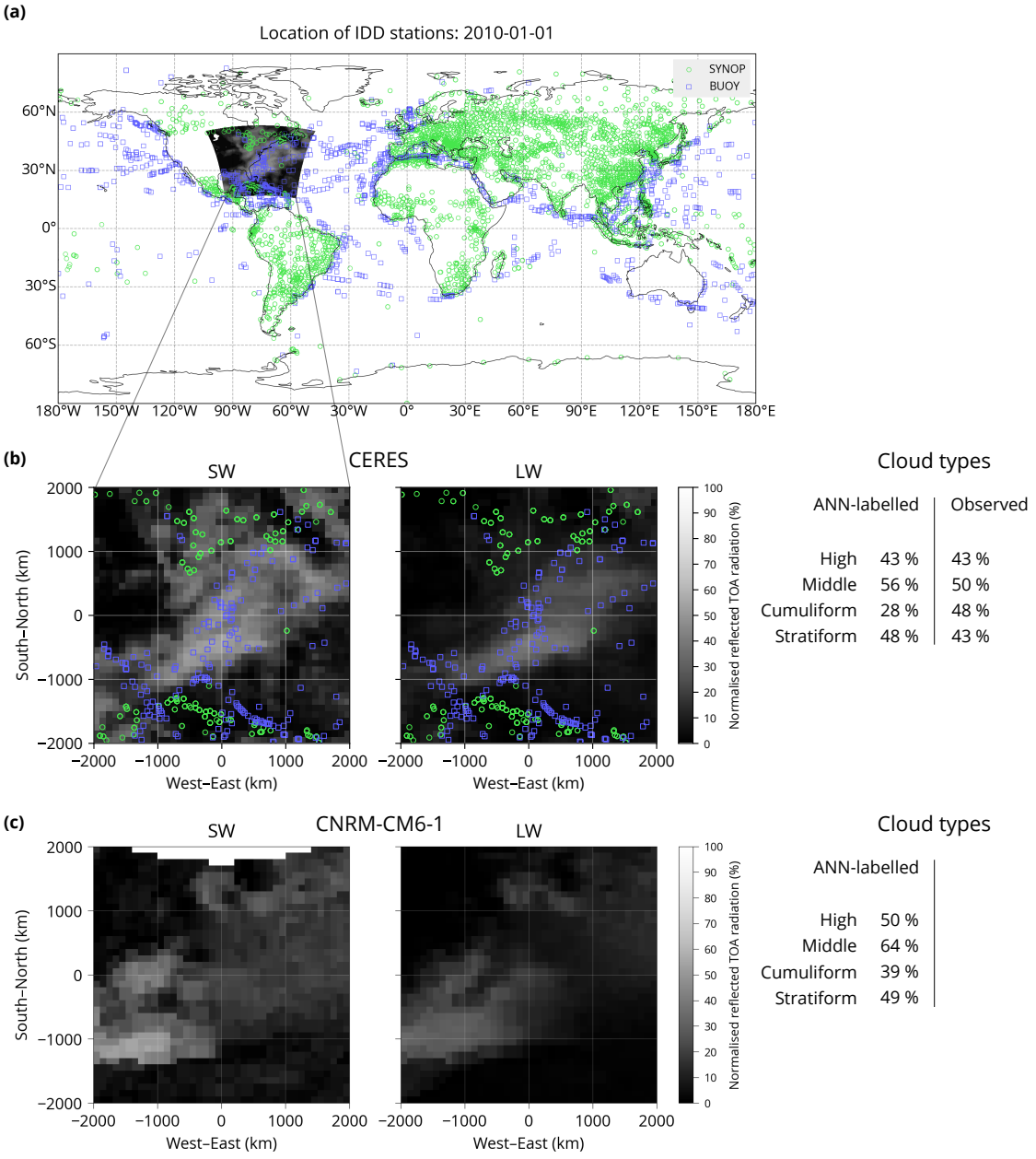

Figure 1. (a) A map showing the location of Internet Data Distribution (IDD) station reports containing cloud information on a single day. Shown is a single sample of size 4000×4000 km of the normalised top of atmosphere (TOA) shortwave spectrum. **(b)** The sample as in (a), but shown re-projected in a local azimuthal equidistant projection. Shown is the shortwave (SW) and longwave (LW) spectrum. **(c)** The same as (b), but for a single climate model (CNRM-CM6-1 in the historical CMIP6 experiment). The table on the right of (b) and (c) shows the ANN-labelled and observed cloud type occurrence. Note that all of the CMIP models are free-running, and therefore the cloud cover is not expected to be the same in (b) and (c). Part of the samples in the shortwave spectrum is missing due to incoming solar radiation below the threshold of 50 Wm^{-2}. Coastline data come from the public domain Global Self-consistent, Hierarchical, High-resolution Geography Database (Wessel and Smith, 1996, 2017).

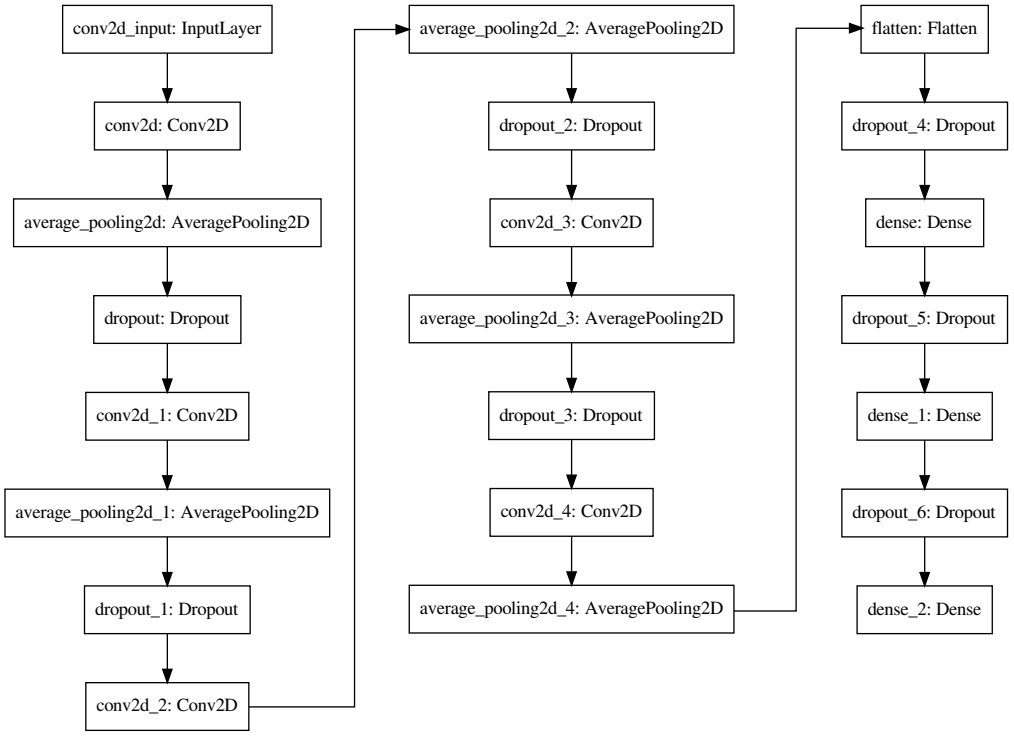

Figure 2. TensorFlow model graph.

2.2 Artificial neural network and data processing

TensorFlow is a machine learning framework for development of artificial neural networks (Abadi et al., 2016), supporting deep and convolutional neural networks. We used the Keras application programming interface (API) of TensorFlow (version 1.14), which provides a simple abstraction layer over TensorFlow, to define, train and apply an ANN on satellite and model data. We performed training of the ANN as follows, demonstrated schematically in Fig. 1. For each day, we generated 20 samples of 4000×4000 km composed of two channels (shortwave and longwave radiation) projected in a local azimuthal equidistant projection centred at globally stochastically random locations (Fig. 1b). The random locations on a sphere were generated using a multivariate normal distribution in three dimensions. More precisely, the channels were calculated from daily mean TOA all-sky shortwave and longwave radiation (*rsut* and *rlut*, respectively), clear sky shortwave and longwave radiation (*rsutcs* and *rlutcs*, respectively) as (1) shortwave CRE normalised to the incoming solar radiation, (2) longwave CRE normalised to clear sky outgoing longwave radiation:

$$\mathrm{CRE}_{\mathrm{SW,norm.}} = (\mathrm{rsut} - \mathrm{rsutcs})/\mathrm{rsdt}, \tag{1}$$

$$\mathrm{CRE}_{\mathrm{LW,norm.}} = (\mathrm{rlutcs} - \mathrm{rlut})/\mathrm{rlutcs}. \tag{2}$$

The normalisation was done so that the values are mostly in the [0, 1] interval, which is a more suitable input to the ANN than non-normalised values. In the shortwave spectrum, normalisation by incoming shorwave radiation was chosen so that the value represents the fraction of reflected incoming radiation due to clouds. In the longwave spectrum, such normalisation is not possible, and normalisation by outgoing clear sky longwave radiation was performed instead.

In order to exclude locations with low solar insolation, where $\mathrm{CRE_{SW,norm.}}$ might be ill-defined because of low values of the denominator, we excluded parts of $\mathrm{CRE_{SW,norm.}}$ where incoming solar radiation was lower than 50 $\mathrm{Wm^{-2}}$. A downside of this approach is that this may cause bias due to exclusion of wintertime polar regions. If a sample was missing any data points or if a training sample contained less than 50 stations, it was excluded from the analysis. The shortwave and longwave channels were the input to the ANN training phase. Furthermore, these samples were labelled with four numbers representing the fraction of ground station records positively reporting each of the four cloud types (cloud type occurrence) out of all stations capable of observing clouds at a given altitude (low, middle or high). The labels were the reference training expected output of the ANN in the training phase. We note that the cloud type occurrences do not have to sum to 100% due to the possibility of multiple cloud types observed at the same time at any given station. Stations which reported clear sky are also included in the calculation of the cloud type occurrence. We used a custom deep convolutional ANN defined in Keras composed of convolutional layers (Conv2D), pooling layers (AveragePooling2D), dropouts (Dropout), flattening (Flatten) and a dense sigmoid predictor (Dense) (Fig. 2 and Algorithm 1).

3 Results

3.1 Training and validation

We trained the ANN on CERES and IDD data in years 2004, 2005, 2007 and 2009–2020. The training was completed in 19 iterations, interrupted automatically once the validation loss function stopped improving for three iterations. Randomly selected samples amounting to 20% of the original dataset were reserved as a validation set, and the remaining 80% were the training set. We determined performance of the ANN by comparing it to an uninformative predictor, which for each sample predicted the fractions of cloud types equal to the all-sample average fraction of that cloud type. We calculated a total root mean square error (RMSE) as a square root of the average of four mean square errors of each of the four cloud types:

$$\text{total RMSE} = \sqrt{\frac{1}{4}\left(\sum_{i=1}^{4}\mathrm{MSE}_i\right)}. \tag{3}$$

Fig. 3a, b shows the result of the training phase. As can be seen in Fig 3a, the loss function reached about 9% after completing the training iterations. In comparison with this uninformative predictor, the total RMSE was reduced from about 17% to 9% (Fig. 3b). Relative to the uninformative predictor, the ANN was most successful in determining the cumuliform clouds (total RMSE reduced from 20% to 9%), and least successful in determining high clouds (total RMSE reduced from 18% to 11%). In summary, we conclude that the ANN is capable of explaining about 47% of the variance. While this number is relatively low, we

Algorithm 1 TensorFlow model definition of the artificial neural network. The function names and arguments are explained in the TensorFlow documentation (TensorFlow Developers, 2022). *nclasses* is the number of cloud types, equal to 4.

```
Conv2D(32, (3, 3), activation='relu', padding='same')
AveragePooling2D((2, 2))
Dropout(0.1)
Conv2D(32, (3, 3), activation='relu', padding='same')
AveragePooling2D((2, 2))
Dropout(0.1)
Conv2D(64, (3, 3), activation='relu', padding='same')
AveragePooling2D((2, 2))
Dropout(0.1)
Conv2D(64, (3, 3), activation='relu', padding='same')
AveragePooling2D((2, 2))
Dropout(0.1)
Conv2D(64, (3, 3), activation='relu', padding='same')
AveragePooling2D((2, 2))
Flatten()
Dropout(0.1)
Dense(64, activation='relu')
Dropout(0.1)
Dense(64, activation='relu')
Dropout(0.1)
Dense(nclasses, activation='sigmoid')
```

will show that this is enough to produce statistically significant results about present-day and future cloud type representation in climate models and its relation to climate sensitivity.

3.2 Geographical biases

We applied the ANN to CERES, climate models and reanalyses TOA radiation data. Next, we calculated geographical distributions of the cloud types predicted by the ANN. Fig. 4 and 5 show the geographical distribution of cloud types in CERES, CMIP models and reanalyses. Because this distribution is derived from the relatively large 4000×4000 km samples, this is also the approximate spatial resolution of the geographical distribution. Therefore, only large-scale features can be expected to be visible. Notably, a peak in high and middle cloud occurrence over the western equatorial Pacific and North Pacific (Fig. 4 a1, a2), high cumuliform cloud occurrence over the equator and low cumuliform cloud occurrence over the polar regions (Fig. 4 a3), and maxima of stratiform cloud occurrence over polar and extratropical regions and low stratiform cloud occurrence over equatorial and subtropical regions (Fig. 4 a4). Stratocumulus decks off western coasts of continents are only slightly visible

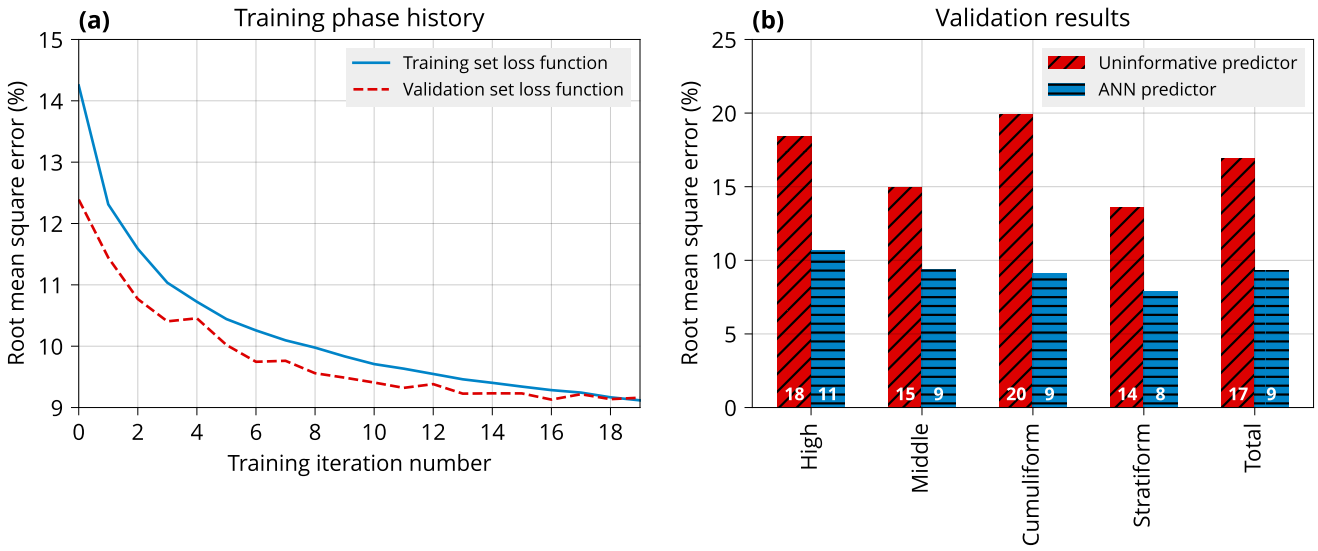

Figure 3. (a) Training phase loss function evolution for the training and validation set of samples. **(b)** Training results showing the performance of an uninformative predictor and the artificial neural network (ANN).

in the stratiform cloud type (Fig. 4 a4), likely due to low spatial resolution, or because this cloud type also includes cloud genera other than stratocumulus. A progression from large to low biases with increasing model ECS, with the exception of CanESM5 and INM-*, is quite notable (Fig. 4 d–k and Fig. 5 b–h). The model biases relative to CERES span a range of -60 to 60% regionally, with an RMSE up to 32% (NorESM2-LM, stratiform cloud type; Fig. 4 d4). The model with the lowest total RMSE was HadGEM3-GC31-MM (2.3%), and the model with the highest total RMSE was NorESM2-LM (23%). Models in the lower half of the ECS range show predominantly negative bias in the cumuliform type (Fig. 4 d3–j3) and positive bias in the stratiform type (Fig. 4 d4–j4). They also all have total RMSE above 4.2%. With the exception of CanESM5, models in the upper half of the ECS range show relatively small biases in individual cloud type RMSE below 9% (Fig. 5), and have total RMSE below 6%. Notable is also negative bias in the high and middle cloud type over western tropical Pacific, present in almost all models, co-located with a maximum in CERES (Fig. 4 a1).

Models which are closely related in their code (CNRM-*; ERA5 and EC-Earth3P; HadGEM3-* and UKESM1-0-LL; INM-*; IPSL-*; MPI-*) performed similarly in terms of the geographical distribution and scale of biases. This means that the ANN method is robust with respect to the model resolution, and also that the groups of related models represent clouds very similarly, presumably because this is to a large extent determined by cloud parametrisations in the atmospheric component of the model without much sensitivity to resolution. This gives us some confidence that this method is trustworthy, even though the ANN can explain only about 47% of the variance and resolution of the geographical distribution is limited by the very large samples of 4000×4000 km.

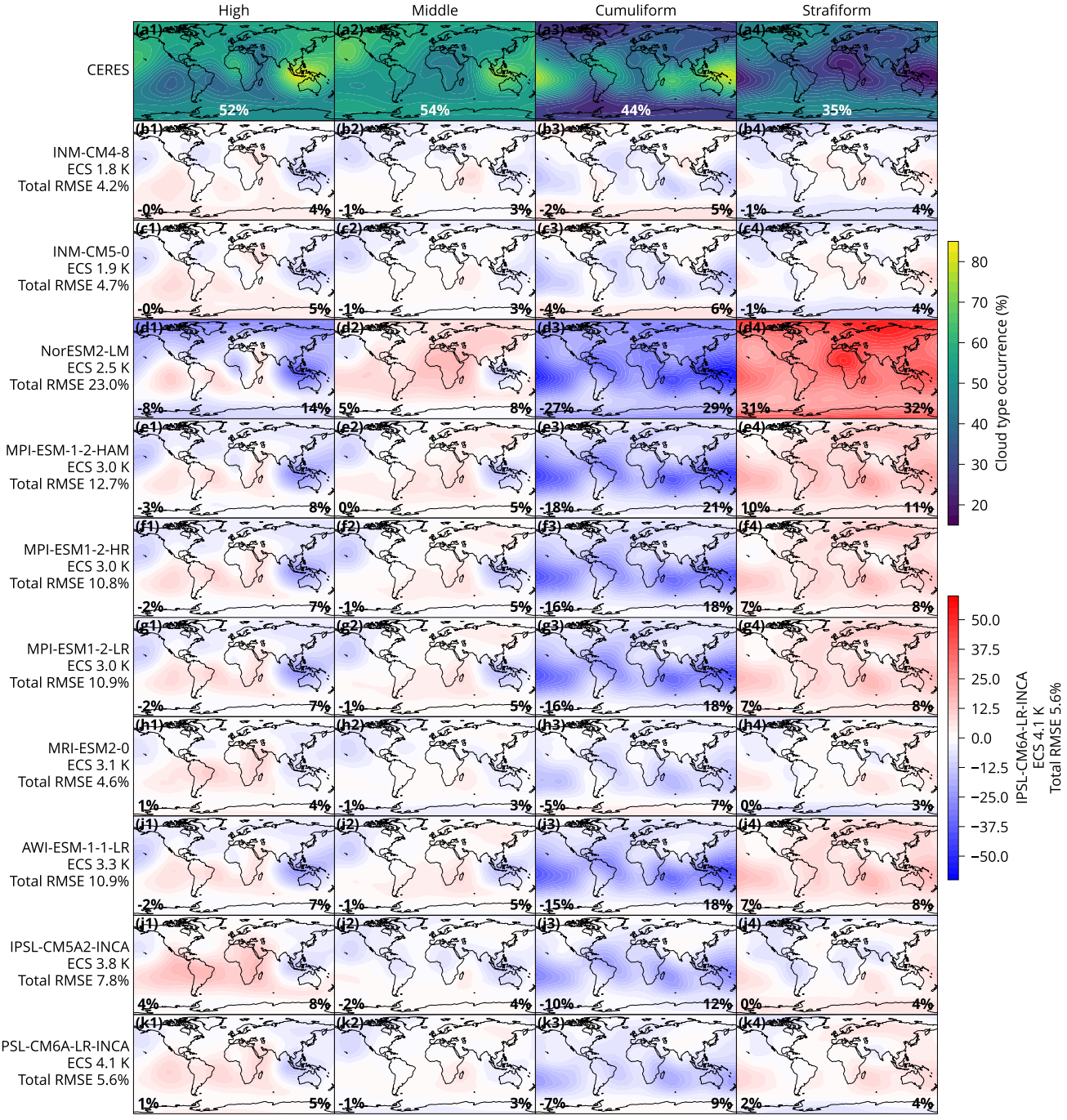

Figure 4. Geographical distribution of cloud type occurrence derived by applying the ANN on observed CERES satellite data in years 2003–2020 and on models output of the historical experiment of CMIP6 in years 2003–2014 relative to CERES. In the lower centre of the CERES plots is the geographical mean occurrence of the cloud type. On the lower left of the model plots is the mean error, and on the lower right is the geographical RMSE of the cloud type. Models are sorted by their ECS. Coastline data come from the public domain Global Self-consistent, Hierarchical, High-resolution Geography Database (Wessel and Smith, 1996, 2017).

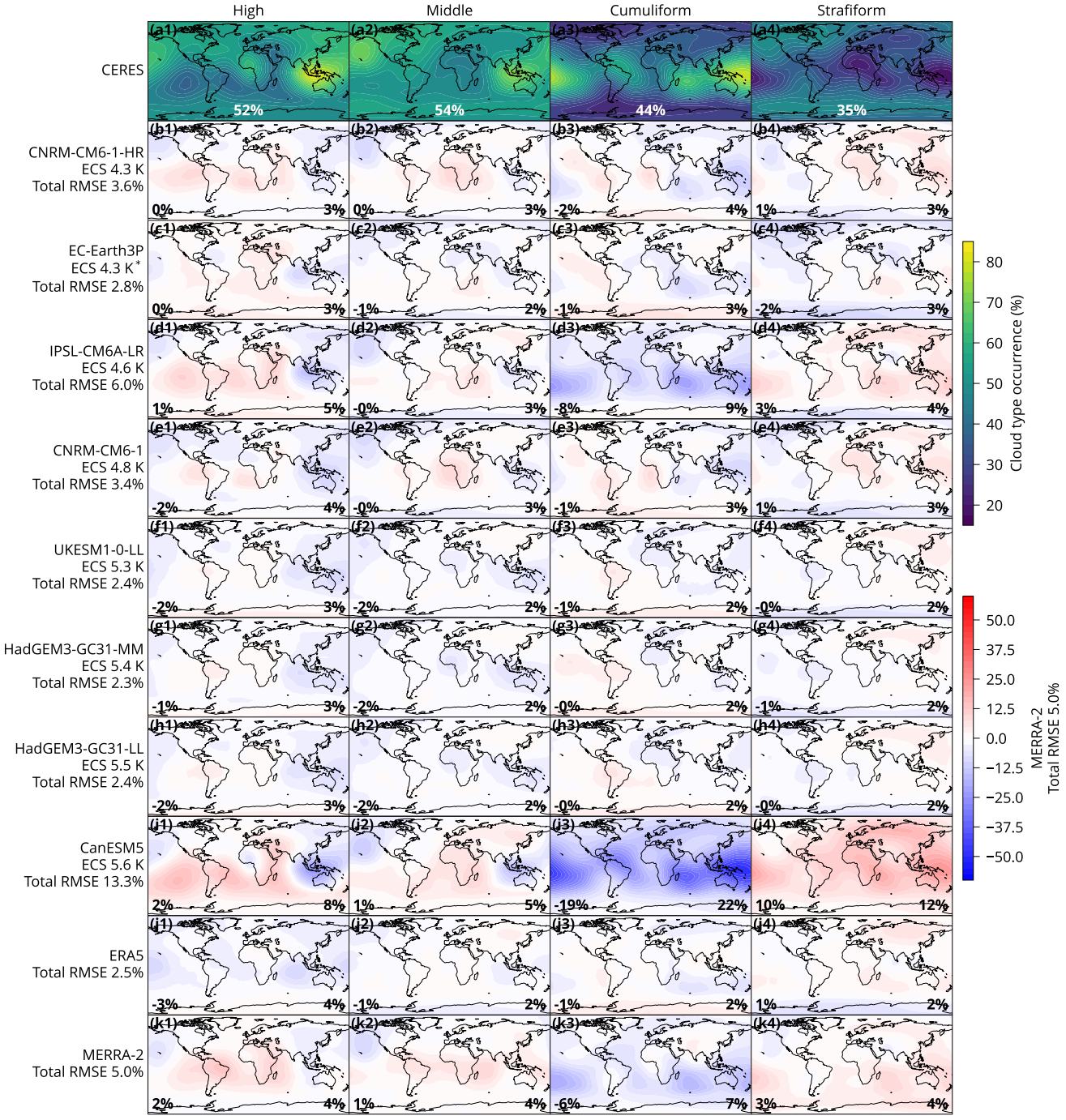

Figure 5. Fig. 4 continued. *For some models, ECS was not available and was taken from the closest available model (see Table 1).

3.3 Optical properties and vertical distribution

From the ANN-labelled samples we calculated joint histograms of cloud optical depth and cloud top pressure (Fig. 6). This type of histograms relates to previous work on cloud classification Rossow and Schiffer (1999), Hahn et al. (2001), Oreopoulos et al. (2016), Schuddeboom et al. (2018) and others. The diagrams in Fig. 6 show cloud type occurrence binned by cloud optical depth and cloud top pressure for the four types as a difference from the mean of the four types. We show anomalies from the mean because the values are relatively close to the mean, which is due to the relatively large spatial averaging of the method (all four types are are well represented in any 4000×4000 km sample). The high cloud type difference from the mean is characterised, as expected, by a maximum occurrence at low pressure (200–400 hPa) (Fig. 6b, f). Interestingly, it is also associated with a local maximum in mid-level high optical depth clouds (Fig. 6b), which can potentially be caused by their co-incidence with cumulonimbus. The middle cloud type difference from the mean is characterised by high optical depth clouds between 300 and 1000 hPa, and is associated with a relatively small local maximum of high optically thin clouds (Fig. 6c). The cumuliform and stratiform cloud types have much stronger deviations from the mean than the high and middle clouds. The cumuliform cloud type difference from the mean has a maximum in optically thin clouds between 400 and 1000 hPa and another maximum in mid-to-high optical depth clouds at pressure below 400 hPa. This may be explained by a frequent co-occurrence of cumuliform clouds with high clouds, such as cumulonimbus and associated anvil clouds. The cumuliform cloud type is identified with the maximum occurrence in the western equatorial Pacific region (Fig. 4 a3), and this regions also has a maximum of occurrence of the high cloud type (Fig. 4 a1). The stratiform cloud histogram is also close to an inverted version of the cumuliform histogram, characterised by a maximum in mid-to-high optical depth clouds between 300 and 1000 hPa.

3.4 Comparison with MODIS and ISCCP cloud clusters

To understand how the cloud types determined by the ANN relate to more traditionally generated cloud clusters, we compare them to clusters generated using cloud top pressure–cloud optical depth joint histograms from the MODIS Schuddeboom et al. (2018) and the ISCCP McDonald and Parsons (2018) datasets. Specifically, the ANN-generated values for each of the four types are calculated given that a MODIS/ISCCP cluster also occurs. The ANN geographical distribution is generated on a $5^\circ \times 5^\circ$ global grid, while the equivalent ISCCP data is on a $2.5^\circ \times 2.5^\circ$ and MODIS on a $1^\circ \times 1^\circ$ grid. To account for this difference in spatial resolution all of the MODIS/ISCCP grid cells that fall within a corresponding ANN geographical distribution grid cell are considered as having the same occurrence values. This will overestimate the similarity between the clusters, as the small cloud structures that can be identified in the higher resolution dataset will be merged.

The 12 MODIS clusters identified in Schuddeboom et al. (2018) and the 15 ISCCP clusters derived in McDonald and Parsons (2018) are displayed as an ordered grid in Fig. 7. One interesting feature of the self-organising map (SOM) algorithm used to derive these clusters is that the most distinct clusters occur in opposite corners of those grids due to the ordering. For example, for the ISCCP grid the top row relates to clouds with low cloud top pressures and the lower row to high cloud top pressures. In general, the variability between the different MODIS/ISCCP clusters occurrence within an ANN cloud type is small. For

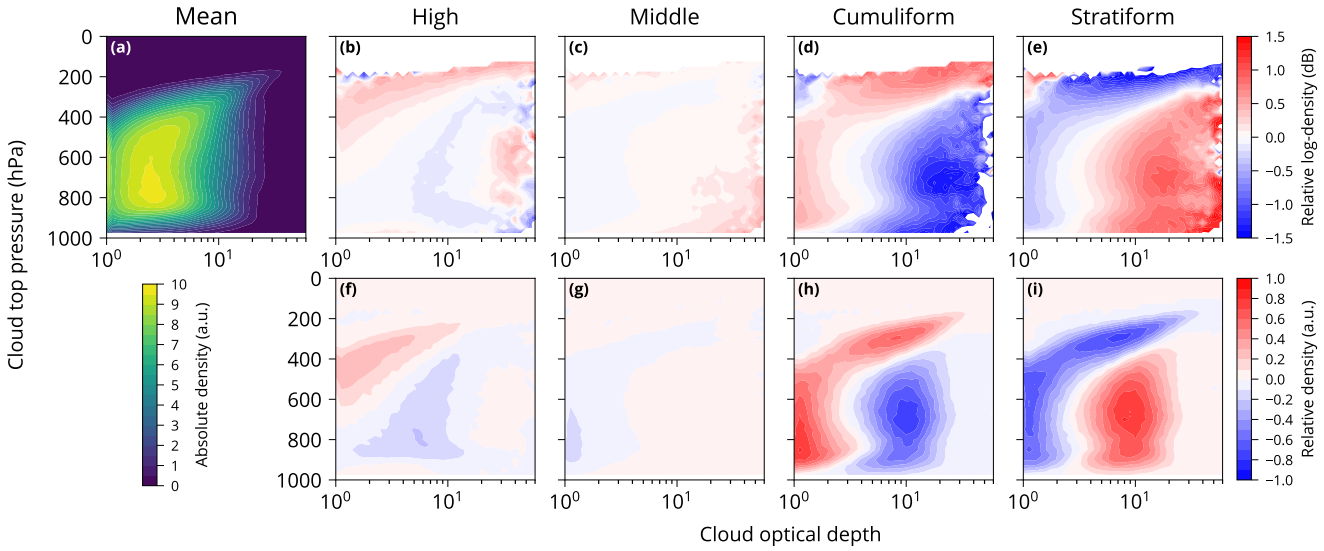

Figure 6. Histogram of cloud optical depth and cloud top pressure of the cloud types derived from CERES observations (2003–2020). **(a)** Mean of the four cloud types. **(b–e)** Histograms relative to the mean on a logarithmic scale (decibels). **(f–i)** The same as (b–e), but on a regular (non-logarithmic) scale.

example the high cloud type shows values ranging from 0.52 to 0.58 in the ISCCP clusters and values between 0.53 and 0.59 for the MODIS clusters. This means that all of the cloud clusters in the MODIS and ISCCP classifications are associated with largely the same amount of high cloud with only minor variations between clusters evident. Similar scale differences are seen for the middle and stratiform cloud type, however the cumuliform cloud type shows much stronger separation in both the MODIS and the ISCCP clusters.

To better understand these biases, the distributions of each of the ANN cloud types is shown in Fig. 8. The comparison with ISCCP and MODIS was done on the level of a daily geographical distribution derived from ANN samples (Fig. 8i–l). The original training samples (Fig. 8a–d) and ANN-labelled samples (Fig. 8e–h) show very similar distributions, while the geographical samples (Fig. 8i–l) have a narrower distribution as a result of the geographical averaging. Comparing these results to the cluster specific results shown in Fig. 7, the range of high cloud occurrence values over the clusters only covers a relatively small portion of the distribution of occurrence values. The same can also be said for the middle and stratiform types, but not the cumuliform type. As such Fig. 8i–l shows that the different cloud scenarios within the high, middle and stratiform types are not well separated in the ISCCP and MODIS clusters. However, the cumuliform cloud type does appear to be well captured by the ISCCP and MODIS clusters. These results may be the result of the aforementioned differences in spatial resolution between the ANN output and the ISCCP and MODIS data.

Looking at the individual clusters, they show the relationships that are expected. First a broad examination at these figures shows clear transitions between the different corners of the SOM. As mentioned earlier these corners should represent the

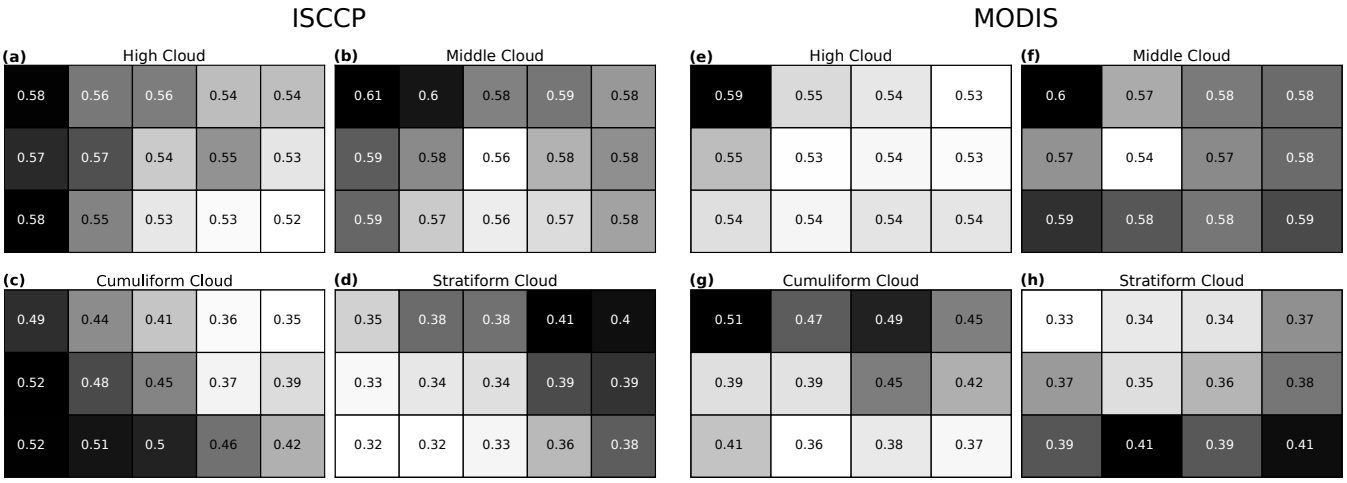

Figure 7. The mean occurrence value for each ANN defined cloud type given that a corresponding **(a–d)** ISCCP and **(e–h)** MODIS cluster also occurs. These clusters are defined in McDonald and Parsons (2018) using cloud top pressure–cloud optical depth joint histograms. For ISCCP, these clusters are generated from a self-organising map (SOM) that consists of 15 clusters gridded on a 5 by 3 grid. For MODIS, these clusters are generated from a SOM that consists of 12 clusters gridded on a 4 by 3 grid. The values shown are matched to the grid that was used when the clusters were defined with cluster 1 in the top left cell and then counting up as going left and then down the grid. Note the colour-scale is reset for each grid and used to distinguish between the different clusters.

most distant clusters, so these transitions are a good sign of cluster separation. Looking at specific clusters, the ISCCP clusters with the largest amount of high cloud are ISCCP clusters 1, 6 and 11 and MODIS cluster 1. An examination of these clusters shows that these all correspond to clusters noted for the presence of high-level clouds. The same relationship is apparent in the clusters with low amounts of high-level clouds in ISCCP, although the clusters are too similar in MODIS for these clouds to be sufficiently distinguished. Similar but weaker results are also true for the middle cloud type. The cumuliform type does not have an obvious physical analogue in the clusters however, basing the analysis on the histograms shown in Fig. 6 shows strong agreement with the established clusters. Finally, the results for the stratiform cloud type show very good agreement with the interpretation of the MODIS clusters but weaker relationships for the ISCCP clusters. Overall there is good agreement between the physical interpretation of the ANN cloud types and the ISCCP and MODIS clusters despite the separation within each ANN cloud type being small.

3.5 Cloud type climatology and its change with global mean near-surface temperature

We analysed the global cloud type occurrence across the CMIP historical experiment models and the reanalyses, and change with respect to GMST in the abrupt-4xCO2 experiment. The abrupt-4xCO2 experiment was chosen because (1) it is commonly used for the determination of ECS and cloud feedback, (2) provides a strong forcing by greenhouse gases and therefore a strong signal in cloud change due to increasing GMST, (3) a large number of models provide the necessary data in this experiment in the CMIP5 and CMIP6 archives. As shown in Fig. 9a comparing model global cloud type occurrence relative to CERES,

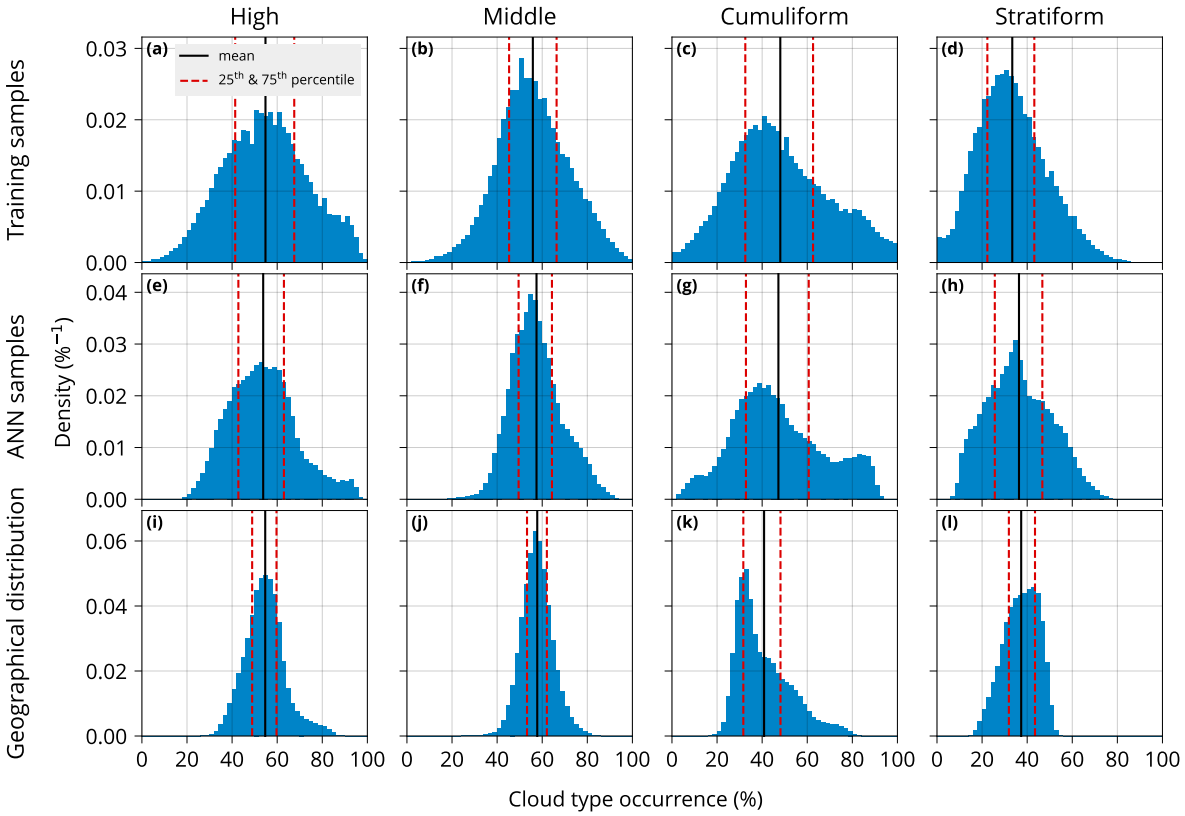

Figure 8. Histograms of cloud type occurrence calculated from **(a–d)** the training samples as determined directly from the Global Telecommunication System (GTS) stations (years 2004, 2005, 2007, 2009–2020), **(e–h)** artificial neural network (ANN)-labelled CERES samples (years 2003–2020), and **(i–l)** a daily mean geographical distribution on a $5^\circ \times 5^\circ$ longitude–latitude grid derived from ANN-labelled CERES samples (year 2007).

the models exhibit a broad range of biases. A progression from large biases to low biases with increasing model ECS is quite notable, with the exception of INM-* and CanESM5. In particular, many models underestimate the cumuliform cloud type (NorESM2-LM, CanESM5, MPI-ESM*, AWI-ESM-1-1-LR, IPSL-* and MRI-ESM2-0) and the same models usually overestimate the stratiform cloud type, possibly as a compensating bias between the two cloud types. To a smaller extent, some models underestimate the high cloud type (NorESM2-LM and MPI-ESM*). Models with small biases in this comparison were predominantly models with ECS above 4.3 K such as CNRM-CM6-*, EC-Earth3P, UKESM1-0-LL and HadGEM3-*, with biases below about 2%. The reanalyses (ERA5 and MERRA-2) have some of the best agreement with CERES compared to other models. This may be partly explained by the fact that their representation of the atmosphere and ocean closely follows observed real state. At same time, EC-Earth3P, a model related in its code base to ERA5, has a similar magnitude of biases as ERA5.

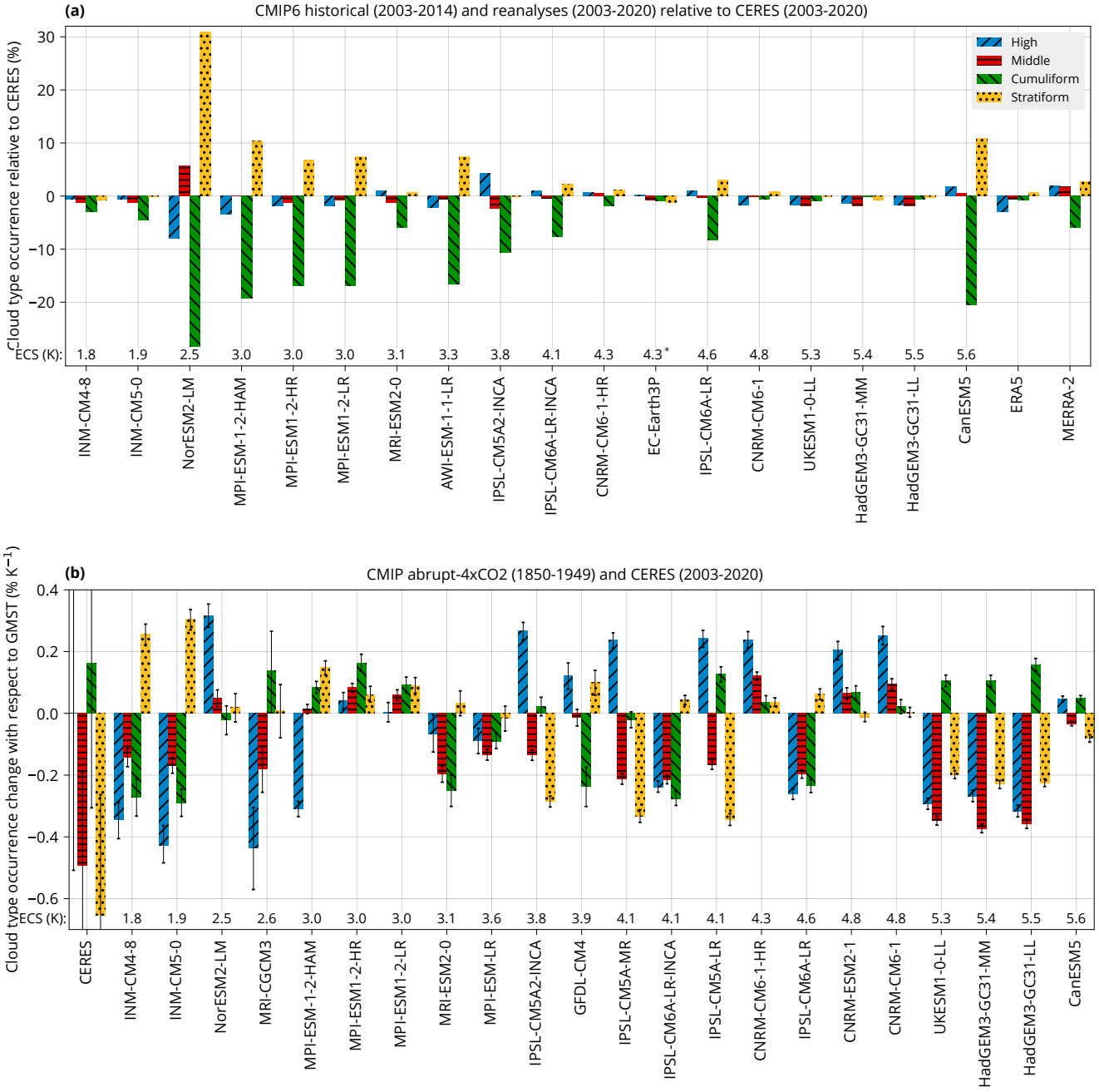

Figure 9. **(a)** Global average of cloud type occurrence in CMIP6 models and reanalyses relative to CERES in the historical experiment (historical reanalysis in the case of reanalyses). **(b)** Global average of cloud type occurrence change with respect to global mean near-surface air temperature (GMST) in CERES+GISS, CMIP5 and CMIP6, calculated by linear regression. Confidence intervals in (b) represent 68% range. Models are sorted by their equilibrium climate sensitivity (ECS). *For some models, ECS was not available and was taken from the closest available model (see Table 1).

We analysed cloud type occurrence change with respect to GMST, defined as the slope of a linear regression of cloud type occurrence as a function of GMST in units of $\% \, \mathrm{K}^{-1}$, shown in Fig. 9b. It was calculated over the time period 2003–2020 for CERES and 1850–1949 for CMIP abrupt-4xCO2. This comparison lacks a reliable observational reference because the CERES record is too short to accurately determine the slope of regression. The abrupt-4xCO2 experiment is also not directly comparable to reality due to the different CO_2 and aerosol forcing. We can, however, note that the models show a broad range of values. Some models display a similar combination of values, such as the HadGEM models and UKESM1-0-LL having decreasing high, middle and stratiform and increasing cumuliform cloud type with increasing GMST. The CNRM models and NorESM2-LM have increasing high, middle and cumuliform clouds. The IPSL-CM5A* models have increasing high clouds and decreasing middle and stratiform clouds. CERES has a negative slope in the middle and stratiform cloud type (68% confidence interval) and is inconclusive in the other cloud types.

3.6 Climate sensitivity

We analysed how the cloud type occurrence change with respect to GMST relate to climate sensitivity. ERA5 and MERRA-2 are excluded from the analysis in this section because climate sensitivity and feedbacks are not estimated for reanalyses. Fig. 10a shows a linear regression of ECS as a function of model's cloud type occurrence change with respect to GMST. Statistically significant is only the regression for the stratiform cloud type (Bayes factor 9), meaning that models which predict increasing stratiform clouds have lower ECS. This is in line with the known negative radiative forcing of low clouds, as demonstrated by global mean cloud radiative kernels (Zelinka et al., 2012).

While the cloud type change with respect to GMST is too uncertain in the observational reference (CERES) to be be useful for quantifying the accuracy of models in the representation of this value, and also the abrupt-4xCO2 experiment assessed here is not directly comparable to reality, we can link present-day cloud biases to climate sensitivity. In Fig. 10b, c, d we show that the total RMSE of cloud type occurrence is significantly linearly related to the model climate sensitivity and transient climate response, respectively, but not to cloud feedback. Assuming Cauchy-distributed error, the Bayes factor of a linear regression model (M_1) relative to a model with no slope (M_0), is 22 and 17 for ECS and TCR, respectively (see Appendix 1 for the definition of M_1 and M_0). Models with the lowest total RMSE (UKESM1-0-LL and the HadGEM3 models with total RMSE of about 2.4%) have the highest equilibrium climate sensitivity (up to 5.55 K), transient climate response (up to 2.79 K) and cloud feedback (up to 0.91 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$). Models with the high end of the total RMSE, such as NorESM2-LM, AWI-ESM-1-1-LR and MPI-ESM* have some of the lowest ECS (down to 2.54 K), TCR (down to 1.48 K) and cloud feedback (down to -0.16 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$), but CanESM5 is a clear outlier with relatively high sensitivity, and NorESM2-LM has a mid-range cloud feedback at 0.44 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$. Other models show a strong linear correlation between the sensitivity indicators (ECS and TCR) and the total RMSE. However, there are some models which do not conform to this relationship: (1) CanESM5 has a high total RMSE (13.3%) and high sensitivity (ECS 5.62 K, TCR 2.74 K and cloud feedback 0.88 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$), (2) NorESM2-LM has a high total RMSE (23%) and low-to-mid sensitivity (ECS 2.54 K, TCR 1.48 and cloud feedback 0.44 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$), and (3) the INM models have very low sensitivity (ECS 1.83–1.92 K and TCR 1.3–1.33 K) and low total RMSE (4.2–4.7%).

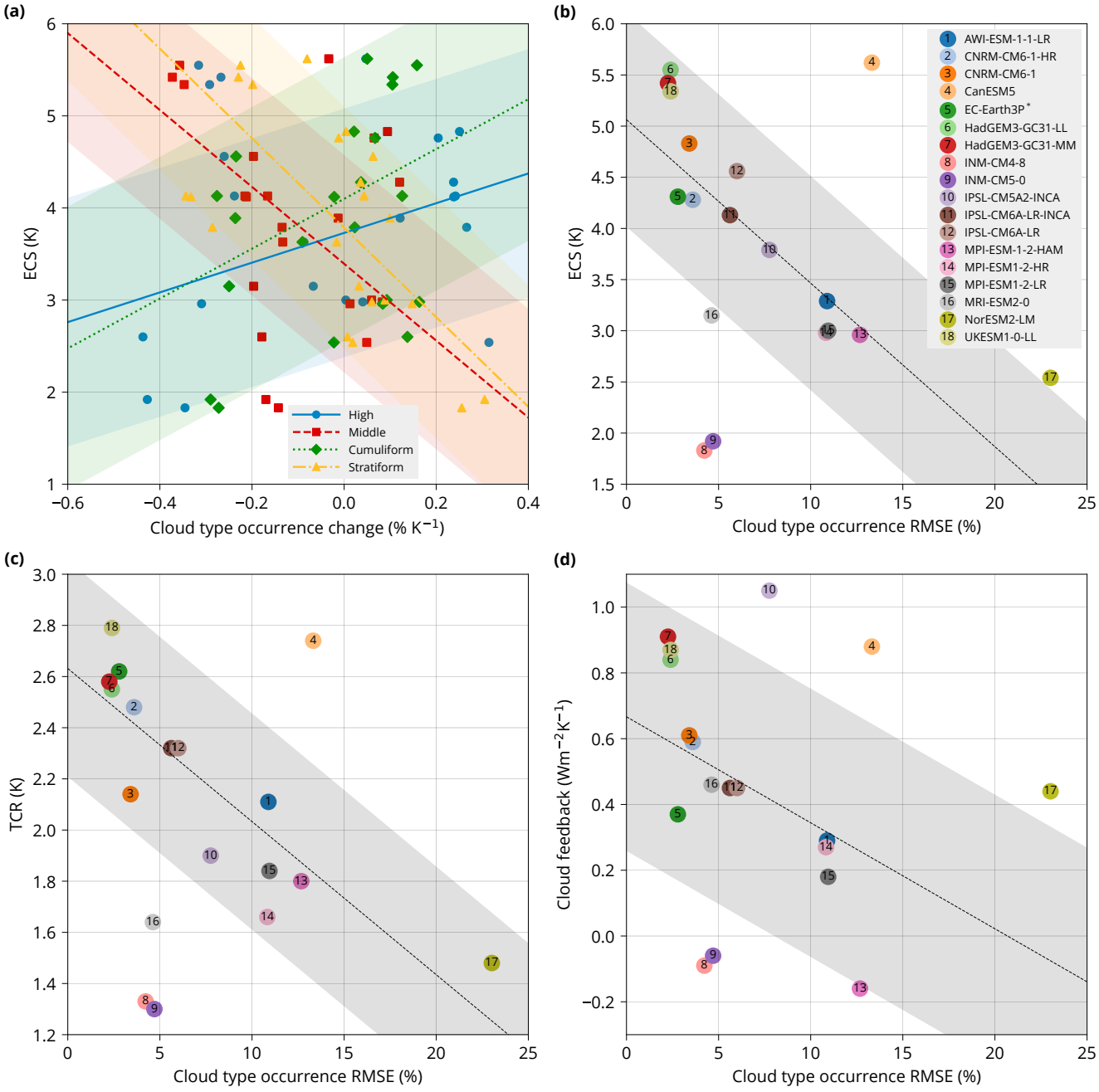

Figure 10. **(a)** Dependence of model equilibrium climate sensitivity (ECS) on the cloud type occurrence change with respect to GMST. **(b)** Dependence of equilibrium climate sensitivity (ECS), **(c)** transient climate response (TCR) and **(d)** climate feedback of CMIP6 models on the model total cloud type root mean square error (RMSE) relative to CERES, calculated from the geographical distribution (as in Fig. 4 and 5). Confidence bands represent 68% range. Linear regression is calculated using Bayesian simulation assuming Cauchy error distribution (Appendix 1). *For some models, ECS was not available and was taken from the closest available model (see Table 1).

4 Discussion

An important finding of our analysis is that cloud type biases in models show that more sensitive models are more consistent with observations in this metric. Zelinka et al. (2022) have also recently found that the mean-state radiatively-relevant cloud properties in CMIP5 and CMIP6 models are strongly and significantly correlated with total cloud feedback, and in particular that better simulating present-day cloud properties is associated with larger cloud feedbacks, and which is in contrast with the the expert judgement of Sherwood et al. (2020). They call this an 'open question for future research'. The most natural assumption is that the models with better representation of present-day clouds, with cloud being the main determinant of model ECS (Wang et al., 2021, Fig. 1a), are also better in their ECS and TCR. The reasons for this assumption are twofold. Firstly, if these models represent clouds well across present-day spatially and temporally variable conditions, it can be expected that even with a shift of global conditions to higher GMST, the models will still be able to simulate clouds well in conditions which have an equivalent in the present-day climate. Secondly, an accurate representation of the starting point of change of cloud properties is important. If, for example, future cloud types should shift from stratiform to cumuliform in a certain region due to increase in atmospheric temperature, the model needs to represent the initial state well, such as stratiform clouds in the present day, to be able to project decrease in this cloud type. Likewise, if extratropical clouds should shift from ice phase to liquid phase, the ice phase cloud fraction needs to be well represented in the initial state for the model to be able to project a decrease in ice content (Bjordal et al., 2020). However, as also discussed by Zelinka et al. (2022), it is possible that models with better representation of present-day clouds are worse in their projection of future cloud changes. This explanation would be more consistent with the findings of other studies pointing to lower ECS, such as Sherwood et al. (2020), based largely on observational evidence, and notably the combined assessment of AR6, which estimates ECS of 2.5–4 K (likely range). In summary, either (1) models with better present-day representation of clouds (in the limited context of the analysis presented here and in Zelinka et al. (2022)) are also better in their representation of ECS and TCR, and therefore high climate sensitivity above the assessed range of Sherwood et al. (2020) and AR6 is plausible, or, (2) paradoxically, models with better present-day representation of clouds are poorer in their representation of ECS and TCR, despite a significant linear relationship between these two properties (Fig. 10b, c). This is an important dilemma worth further investigation. We note that by Ockham's razor (Jaynes, 2003, Chapter 20), (1) is a more likely explanation.

During the 20–21th century historical record, the effect of rising greenhouse gas concentrations is obscured by the rapid rise of aerosol concentration since the 1950s (Dittus et al., 2020; Tokarska et al., 2020; Smith and Forster, 2021), masking the effect of greenhouse gases by a direct radiative effect and a highly uncertain indirect radiative effect through changing cloud formation, microphysics and lifetime. Paleoclimatic constraints are limited by the accuracy of paleo records, and the sometimes different state of the climate. Mann (2021) cautions that ECS involves feedbacks which are not the same for warm and cold climate. ECS derived from past cold climate states is not the same as ECS of the present-day and future warm climate, which is thought to be higher than the cold climate ECS (Caballero and Huber, 2013; Shaffer et al., 2016; Schneider et al., 2019) due to non-linear feedback related to ice, permafrost and clouds. Mann (2021) argues that as a result, the upper-end constraint on ECS of Sherwood et al. (2020) by paleoclimatic evidence is unreliable, stating that 'Neither the cooling during

the largest volcanic eruptions of the Common Era nor the cooling during the Last Glacial Maximum can provide any constraint on feedback processes that are specific to hothouse climates. Even the most sophisticated statistical analysis cannot account for physical responses that lie outside the range of the data analyzed.'

Recent understanding of climate sensitivity is represented by diverging results relative to the CMIP6 multi-model mean. de-la Cuesta and Mauritsen (2019) and Tokarska et al. (2020) estimate low ECS based on the historical record, Renoult et al. (2020) estimate low ECS based on paleoclimatic evidence from the last glacial maximum and mid-Pliocene warm period, Bjordal et al. (2020), on the other hand, show that high ECS is possible due to transition to higher Southern Ocean cloud phase change feedback with warming climate. One way in which high sensitivity models may be better in their representation of cloud types, while still being biased too high in their climate sensitivity is due to the 'too few, too bright' problem (Kuma et al., 2020; Nam et al., 2012; Klein et al., 2013; Wall et al., 2017). If relative to CMIP5, CMIP6 models improved their cloud type representation, while remaining biased too high in cloud brightness, a decreasing middle and stratiform clouds with increasing GMST (as shown in Fig. 9b for UKESM1-0-LL and the HadGEM models, for example) would cause a too strong positive feedback, relative to a situation when the cloud brightness is consistent with reality. Such positive feedback could then be the cause of an unrealistically high climate sensitivity. Preliminary results presented in Schuddeboom and McDonald (2021) also suggest this possibility as changes to stratocumulus cloud representation in CMIP6 that have lead to higher occurrence rates, particularly over the Southern Ocean. However, changes to the brightness of stratocumulus clouds was not directly investigated preventing a definitive conclusion.

4.1 Limitations

The artificial neural network method introduced in this study comes with a number of limitations, which may be improved in future versions of the ANN. Firstly, the ANN can only explain about 47% of the variance. While this is far from ideal, we think this is still enough to detect significant biases in cloud type representation and its change with respect to GMST in CMIP models. Secondly, because we use a simpler method of labelling whole samples as opposed to pixel-wise labelling, and we have to use relatively large samples (4000×4000 km) for accurate detection, the spatial accuracy of the result is on about the same magnitude of scale. Thus, the resulting distribution is not well-resolved geographically, and unable to resolve more detailed regions such as the marine stratocumulus decks on the western sides of continents. The CERES dataset is too short (2003–2020) to reliably detect change with respect to GMST, other than being able to determine that the cumuliform and middle cloud type is likely (68% confidence interval) decreasing. This means that an observational reference for change with respect to GMST is lacking. While the NOAA and ESA satellite series provide much longer time series, datasets derived from these satellites, the Climate Change Initiative Cloud project (Cloud_cci) (Stengel et al., 2020), the Pathfinder Atmospheres Extended (PATMOS-x) (Foster and Heidinger, 2013) and the 'CM SAF Cloud, Albedo And Surface Radiation dataset from AVHRR data' (CLARA-A2) (Karlsson et al., 2017), appear to be unreliable for determining change with respect to GMST, most likely due to changing orbit and instrument sensors. We suggest that future research on this topic could focus on improving the predictive capability of the ANN. Some of the aforementioned limitations could be addressed by improving the ANN by, for example, increasing the number of layers, or by training two ANNs separately for the land and marine areas. These have a markedly different cloud

coverage and density of ground-based observations. Therefore, the training process is biased in favour of land stations. Pixel-wise classification, whereby instead of labelling whole samples, each pixel is assigned cloud type fractions, could improve the geographical resolution of cloud type occurrence distribution. Potentially, other satellite data products from ISCCP, Multi-angle Imaging SpectroRadiometer (MISR), MODIS, CloudSat and the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) could be utilised through the use of the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP).

5 Conclusions

We developed a deep convolutional artificial neural network for the purpose of determining cloud types in observed and simulated TOA shortwave and longwave radiation imagery, trained on global historical records of human observations of WMO cloud genera. This ANN could explain about 47% of the variance relative to an uninformative predictor. When compared to the results produced by past clustering approaches based on MODIS and ISCCP, the ANN shows good agreement on the physical properties of the clouds. However, the output from the ANN does show smaller separation in the other sets clusters than expected. This result is in conceptual agreement with the result from the uninformative predictor as the ANN captures much of the cluster variability but with clear limitations.

We then applied the ANN to satellite observations from CERES, climate model and reanalysis output to derive geographical distribution, global mean of cloud type occurrence and its change with respect to GMST. The models displayed a variety of biases relative to satellite observations, most notably a negative bias in cumuliform clouds and a positive bias in stratiform clouds, with models such as NorESM2-LM, CanESM5, MPI-ESM* and AWI-ESM-1-1-LR having the largest biases, and HadGEM3-*, UKESM1-0-LL, EC-Earth3P and ERA5 having the lowest biases. Models related in their code base often showed the same pattern and magnitude of biases, highlighting the utility of this analysis. The set of all models had highly variable cloud type occurrence change with respect to GMST.

By linking the cloud type change with respect to GMST to climate sensitivity, we showed that in line with the negative radiative effect of low clouds, models with an increasing stratiform cloud type with GMST tended to have lower ECS. We investigated the link between present-day cloud biases and equilibrium climate sensitivity, transient climate response and cloud feedback. We found that model cloud bias is significantly correlated with its ECS and TCR but not cloud feedback, manifested by a strong linear relationship between the total RMSE of cloud type occurrence of a model and its ECS and TCR (Bayes factor of 22 and 17, respectively). Models with the lowest total RMSE were the most climate sensitive models in terms of ECS and TCR, and also models with high total RMSE were the least climate sensitive.

We therefore conclude that model cloud biases, investigated through cloud type occurrence corresponding to the WMO cloud genera, are predictive of climate model sensitivity. While the cloud type representation is not the only determinant of climate sensitivity, it appears to suggest that cloud type biases in CMIP models cause underestimation of climate sensitivity, and high sensitivity models are more correct in their representation of present-day clouds types. We caution, however, from concluding that high sensitivity models are more correct in general, because other cloud properties such as cloud phase, droplet

and ice crystal concentration, lifetime and height also determine the cloud radiative effect. These properties are included in our analysis implicitly through their collective effect on cloud genera, but they are not analysed explicitly. Also the link between the quality of present-day and future cloud representation is not explored in our analysis, but is an important open question (Zelinka et al., 2022).

Limitations of our study are the predictive strength of the ANN (explaining only about 47% of the variance), relatively low spatial resolution on the order of 4000 km, and the lack of a good observational reference for cloud type occurrence change with respect to GMST.

We suggest that our results about high sensitivity models being more correct in their cloud type representation should be considered in the context of other factors influencing the sensitivity in a multiple-factor analysis (Bretherton and Caldwell, 2020; Sherwood et al., 2020), especially when constraining equilibrium climate sensitivity or transient climate response using emergent constraints. An important dilemma requiring further investigation is that either models with better cloud representation are also more accurate in their ECS, and therefore high ECS is plausible and in contradiction with Sherwood et al. (2020) and AR6, or models with better cloud representation are less accurate in their representation of ECS, despite the significant linear relationship between cloud representation quality and climate sensitivity identified here (Fig. 10b, c) and in Zelinka et al. (2022).

Appendix A

The linear regression model M_1 representing the alternative hypothesis and the null hypothesis model M_0 are defined as:

$$\mathrm{M}_1: \quad y = \alpha x + \beta + \epsilon, \tag{A1}$$

$$\mathrm{M}_0: \quad y = \beta + \epsilon, \tag{A2}$$

$$\alpha = \tan(\varphi), \tag{A3}$$

$$\varphi \sim \text{Uniform}(-\pi/2, \pi/2), \tag{A4}$$

$$\beta \sim \text{Uniform}(-100, 100), \tag{A5}$$

$$\epsilon \sim \text{Cauchy}(0, \gamma), \tag{A6}$$

where x is a vector of the independent variables, y is a vector of the dependent variables, α and β are the slope and intercept, respectively, ϵ is a Cauchy-distributed random error, γ is the scale parameter of the Cauchy distribution, and φ is the angle of the slope. φ and β come from a continuous uniform prior distribution. The statistical distribution of the free parameters φ, β and the Bayes factor ($P(\mathrm{M}_1|x,y)/P(\mathrm{M}_0|x,y)$) were determined using the Metropolis algorithm (Metropolis et al., 1953), and simulated with the Python library PyMC3 version 3.11.2 (Salvatier et al., 2016). The prior probability of M_0 and M_1 was assumed to be equal: $P(\mathrm{M}_0) = P(\mathrm{M}_1) = 0.5$. Before running the simulation, the variables x and y were normalised by their mean and standard deviation.

Code and data availability. The datasets used in our analysis are publicly available: CERES (https://ceres.larc.nasa.gov/data/), GISTEMPv4 (https://data.giss.nasa.gov/gistemp/, CMIP5 (https://esgf-node.llnl.gov/search/cmip5/), CMIP6 (https://esgf-node.llnl.gov/projects/cmip6/), MERRA-2 (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/), ERA (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5), and IDD (https://rda.ucar.edu/datasets/ds336.0/). The code used in our analysis is open source and available on GitHub (https://github.com/peterkuma/ml-clouds-2021) and Zenodo (Kuma et al., 2022).

Author contributions. PK participated on conceptualisation and methodology development, developed the artificial neural network model, performed the data analysis and wrote the manuscript; FB participated on conceptualisation and methodology development, review and editing of the manuscript, funding acquisition and project administration. AS and AM performed the comparison with MODIS and ISCCP cloud regimes and reviewed the manuscript. ØS participated on NorESM2-LM data preparation and acquisition, consultation of the analysis and reviewed the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was conducted as part of the FORCeS project: 'Constrained aerosol forcing for improved climate projections' (https://forces-project.eu), funded by the European Union's Horizon 2020 research and innovation programme under grant agreement number 821205. We also acknowledge funding from the Swedish e-Science Research Centre (SeRC). AS and AM acknowledge funding from New Zealand's Deep South National Science Challenge 'Cloud and Aerosol Measurements for Improved Climate Projections'. We thank Hossein Azizpour for his consultation and advice on the artificial neural network development. We acknowledge the CERES dataset provided by the NASA Langley Research Center, the IDD dataset provided by Unidata and the University Corporation for Atmospheric Research through the Research Data Archive at the National Center for Atmospheric Research, the GISTEMP dataset provided by the NASA Goddard Institute for Space Studies, the ERA5 dataset provided by the ECMWF through the Copernicus Climate Change Service, and the MERRA-2 dataset provided by the Global Modeling and Assimilation Office, NASA Goddard Space Flight Center Greenbelt. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP5 and CMIP6. We thank the climate modelling 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 CMIP5, CMIP6 and ESGF. We acknowledge open source software used in our analysis: TensorFlow (Abadi et al., 2016), Python, NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Matplotlib, cartopy (Met Office, 2010), PyMC3 (Salvatier et al., 2016), parallel (Tange et al., 2011), Pandas (The pandas development team, 2020), pyproj, Cython (Behnel et al., 2011), aria2 and Devuan GNU+Linux.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: A System for Large-Scale Machine Learning, in: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, OSDI'16, pp. 265–283, USENIX Association, USA, 2016.

Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D. S., and Smith, K.: Cython: The Best of Both Worlds, Computing in Science Engineering, 13, 31–39, https://doi.org/10.1109/MCSE.2010.118, 2011.

Bjordal, J., Storelvmo, T., Alterskjær, K., and Carlsen, T.: Equilibrium climate sensitivity above 5 °C plausible due to state-dependent cloud feedback, Nature Geoscience, 13, 718–721, https://doi.org/10.1038/s41561-020-00649-1, 2020.

Bretherton, C. S. and Caldwell, P. M.: Combining Emergent Constraints for Climate Sensitivity, Journal of Climate, 33, 7413–7430, https://doi.org/10.1175/JCLI-D-19-0911.1, 2020.

Caballero, R. and Huber, M.: State-dependent climate sensitivity in past warm climates and its implications for future climate projections, Proceedings of the National Academy of Sciences, 110, 14 162–14 167, https://doi.org/10.1073/pnas.1303365110, 2013.

Cesana, G., Del Genio, A. D., and Chepfer, H.: The Cumulus And Stratocumulus CloudSat-CALIPSO Dataset (CASCCAD), Earth System Science Data, 11, 1745–1764, https://doi.org/10.5194/essd-11-1745-2019, 2019.

Cho, N., Tan, J., and Oreopoulos, L.: Classifying Planetary Cloudiness with an Updated Set of MODIS Cloud Regimes, Journal of Applied Meteorology and Climatology, 60, 981–997, https://doi.org/10.1175/JAMC-D-20-0247.1, 2021.

de-la Cuesta, D. J. and Mauritsen, T.: Emergent constraints on Earth's transient and equilibrium response to doubled CO2 from post-1970s global warming, Nature Geoscience, 12, 902–905, https://doi.org/10.1038/s41561-019-0463-y, 2019.

Dittus, A. J., Hawkins, E., Wilcox, L. J., Sutton, R. T., Smith, C. J., Andrews, M. B., and Forster, P. M.: Sensitivity of Historical Climate Simulations to Uncertain Aerosol Forcing, Geophysical Research Letters, 47, e2019GL085 806, https://doi.org/https://doi.org/10.1029/2019GL085806, e2019GL085806 10.1029/2019GL085806, 2020.

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., Wielicki, B. A., Young, D. F., and Sun, M.: Geostationary Enhanced Temporal Interpolation for CERES Flux Products, Journal of Atmospheric and Oceanic Technology, 30, 1072–1090, https://doi.org/10.1175/JTECH-D-12-00136.1, 2013.

Dong, Y., Armour, K. C., Zelinka, M. D., Proistosescu, C., Battisti, D. S., Zhou, C., and Andrews, T.: Intermodel Spread in the Pattern Effect and Its Contribution to Climate Sensitivity in CMIP5 and CMIP6 Models, Journal of Climate, 33, 7755–7775, https://doi.org/10.1175/JCLI-D-19-1011.1, 2020.

Drönner, J., Korfhage, N., Egli, S., Mühling, M., Thies, B., Bendix, J., Freisleben, B., and Seeger, B.: Fast Cloud Segmentation Using Convolutional Neural Networks, Remote Sensing, 10, https://doi.org/10.3390/rs10111782, 2018.

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

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and

Williamson, M. S.: Taking climate model evaluation to the next level, Nature Climate Change, 9, 102–110, https://doi.org/10.1038/s41558-018-0355-y, 2019.

Flynn, C. M. and Mauritsen, T.: On the climate sensitivity and historical warming evolution in recent coupled model ensembles, Atmospheric Chemistry and Physics, 20, 7829–7842, https://doi.org/10.5194/acp-20-7829-2020, 2020.

Forster, P. M., Maycock, A. C., McKenna, C. M., and Smith, C. J.: Latest climate models confirm need for urgent mitigation, Nature Climate Change, 10, 7–10, https://doi.org/10.1038/s41558-019-0660-0, 2020.

Foster, M. J. and Heidinger, A.: PATMOS-x: Results from a Diurnally Corrected 30-yr Satellite Cloud Climatology, Journal of Climate, 26, 414–425, https://doi.org/10.1175/JCLI-D-11-00666.1, 2013.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

GISTEMP Team: GISS Surface Temperature Analysis (GISTEMP), version 4, https://data.giss.nasa.gov/gistemp/, last access: 7 December 2021, 2021.

Guo, Y., Cao, X., Liu, B., and Gao, M.: Cloud Detection for Satellite Imagery Using Attention-Based U-Net Convolutional Neural Network, Symmetry, 12, https://doi.org/10.3390/sym12061056, 2020.

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Corti, S., Davini, P., Exarchou, E., Fabiano, F., Fladrich, U., Fuentes Franco, R., García-Serrano, J., von Hardenberg, J., Koenigk, T., Levine, X., Meccia, V. L., van Noije, T., van den Oord, G., Palmeiro, F. M., Rodrigo, M., Ruprich-Robert, Y., Le Sager, P., Tourigny, E., Wang, S., van Weele, M., and Wyser, K.: HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – description, model computational performance and basic validation, Geoscientific Model Development, 13, 3507–3527, https://doi.org/10.5194/gmd-13-3507-2020, 2020.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hahn, C. J., Rossow, W. B., and Warren, S. G.: ISCCP Cloud Properties Associated with Standard Cloud Types Identified in Individual Surface Observations, Journal of Climate, 14, 11–28, https://doi.org/10.1175/1520-0442(2001)014<0011:ICPAWS>2.0.CO;2, 2001.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning, Bulletin of the American Meteorological Society, 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1, 2017.

Jakob, C. and Tselioudis, G.: Objective identification of cloud regimes in the Tropical Western Pacific, Geophysical Research Letters, 30, https://doi.org/https://doi.org/10.1029/2003GL018367, 2003.

Jaynes, E. T.: Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, UK, https://doi.org/10.1017/CBO9780511790423, 2003.

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R.: CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmospheric Chemistry and Physics, 17, 5809–5828, https://doi.org/10.5194/acp-17-5809-2017, 2017.

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., and Gleckler, P. J.: Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator, Journal of Geophysical Research: Atmospheres, 118, 1329–1342, https://doi.org/10.1002/jgrd.50141, 2013.

Kuma, P., McDonald, A. J., Morgenstern, O., Alexander, S. P., Cassano, J. J., Garrett, S., Halla, J., Hartery, S., Harvey, M. J., Parsons, S., Plank, G., Varma, V., and Williams, J.: Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations, Atmospheric Chemistry and Physics, 20, 6607–6630, https://doi.org/10.5194/acp-20-6607-2020, 2020.

Kuma, P., Bender, F. A.-M., Schuddeboom, A., McDonald, A. J., and Seland, Ø.: Code accompanying the manuscript "Machine learning of cloud types shows higher climate sensitivity is associated with lower cloud biases", https://doi.org/10.5281/zenodo.6164982, 2022.

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP uncertainty model, J. Geophys. Res. Atmos., 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Liu, C., Yang, S., Di, D., Yang, Y., Zhou, C., Hu, X., and Sohn, B.-J.: A Machine Learning-based Cloud Detection Algorithm for the Himawari-8 Spectral Image, Advances in Atmospheric Sciences, https://doi.org/10.1007/s00376-021-0366-x, 2021.

Liu, S. and Li, M.: Deep multimodal fusion for ground-based cloud classification in weather station networks, EURASIP Journal on Wireless Communications and Networking, 2018, https://doi.org/10.1186/s13638-018-1062-0, 2018.

Loeb, N., Su, W., Doelling, D., Wong, T., Minnis, P., Thomas, S., and Miller, W.: 5.03 - Earth's Top-of-Atmosphere Radiation Budget, in: Comprehensive Remote Sensing, edited by Liang, S., pp. 67–84, Elsevier, Oxford, https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.10367-7, 2018.

Mann, M. E.: Beyond the hockey stick: Climate lessons from the Common Era, Proceedings of the National Academy of Sciences, 118, https://doi.org/10.1073/pnas.2112797118, 2021.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., eds.: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom, in press, 2021.

McDonald, A. J. and Parsons, S.: A Comparison of Cloud Classification Methodologies: Differences Between Cloud and Dynamical Regimes, Journal of Geophysical Research: Atmospheres, 123, 11,173–11,193, https://doi.org/https://doi.org/10.1029/2018JD028595, 2018.

McDonald, A. J., Cassano, J. J., Jolly, B., Parsons, S., and Schuddeboom, A.: An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states, Journal of Geophysical Research: Atmospheres, 121, 13 009–13 030, https://doi.org/https://doi.org/10.1002/2016JD025199, 2016.

McErlich, C., McDonald, A., Schuddeboom, A., and Silber, I.: Comparing Satellite- and Ground-Based Observations of Cloud Occurrence Over High Southern Latitudes, Journal of Geophysical Research: Atmospheres, 126, e2020JD033 607, https://doi.org/https://doi.org/10.1029/2020JD033607, e2020JD033607 2020JD033607, 2021.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M.: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Science Advances, 6, eaba1981, https://doi.org/10.1126/sciadv.aba1981, 2020.

Met Office: Cartopy: a cartographic python library with a Matplotlib interface, Exeter, Devon, https://scitools.org.uk/cartopy, 2010.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, The Journal of Chemical Physics, 21, 1087–1092, https://doi.org/10.1063/1.1699114, 1953.

Nam, C., Bony, S., Dufresne, J.-L., and Chepfer, H.: The 'too few, too bright'tropical low-cloud problem in CMIP5 models, Geophysical Research Letters, 39, https://doi.org/10.1029/2012gl053421, 2012.

Nijsse, F. J. M. M., Cox, P. M., and Williamson, M. S.: Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models, Earth System Dynamics, 11, 737–750, https://doi.org/10.5194/esd-11-737-2020, 2020.

Oreopoulos, L., Cho, N., Lee, D., and Kato, S.: Radiative effects of global MODIS cloud regimes, Journal of Geophysical Research: Atmospheres, 121, 2299–2317, https://doi.org/https://doi.org/10.1002/2015JD024502, 2016.

Renoult, M., Annan, J. D., Hargreaves, J. C., Sagoo, N., Flynn, C., Kapsch, M.-L., Li, Q., Lohmann, G., Mikolajewicz, U., Ohgaito, R., Shi, X., Zhang, Q., and Mauritsen, T.: A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP, Climate of the Past, 16, 1715–1735, https://doi.org/10.5194/cp-16-1715-2020, 2020.

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview, Geoscientific Model Development, 13, 1179–1199, https://doi.org/10.5194/gmd-13-1179-2020, 2020.

Rossow, W. B. and Schiffer, R. A.: ISCCP Cloud Data Products, Bulletin of the American Meteorological Society, 72, 2–20, https://doi.org/10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2, 1991.

Rossow, W. B. and Schiffer, R. A.: Advances in Understanding Clouds from ISCCP, Bulletin of the American Meteorological Society, 80, 2261–2288, https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2, 1999.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C.: Probabilistic programming in Python using PyMC3, PeerJ Computer Science, 2, e55, https://doi.org/10.7717/peerj-cs.55, 2016.

Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., and Eyring, V.: Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6?, Earth System Dynamics, 11, 1233–1258, https://doi.org/10.5194/esd-11-1233-2020, 2020.

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., Molod, A., Neale, R. B., and Saha, S.: Practice and philosophy of climate model tuning across six US modeling centers, Geoscientific Model Development, 10, 3207–3223, https://doi.org/10.5194/gmd-10-3207-2017, 2017.

Schneider, T., Kaul, C. M., and Pressel, K. G.: Possible climate transitions from breakup of stratocumulus decks under greenhouse warming, Nature Geoscience, 12, 163–167, https://doi.org/10.1038/s41561-019-0310-1, 2019.

Schuddeboom, A., McDonald, A. J., Morgenstern, O., Harvey, M., and Parsons, S.: Regional Regime-Based Evaluation of Present-Day General Circulation Model Cloud Simulations Using Self-Organizing Maps, Journal of Geophysical Research: Atmospheres, 123, 4259–4272, https://doi.org/https://doi.org/10.1002/2017JD028196, 2018.

Schuddeboom, A. J. and McDonald, A. J.: The Southern Ocean Radiative Bias, Cloud Compensating Errors, and Equilibrium Climate Sensitivity in CMIP6 Models, Journal of Geophysical Research: Atmospheres, 126, 1–16, https://doi.org/10.1029/2021JD035310, 2021.

Segal-Rozenhaimer, M., Li, A., Das, K., and Chirayath, V.: Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN), Remote Sensing of Environment, 237, 111 446, https://doi.org/https://doi.org/10.1016/j.rse.2019.111446, 2020.

Semmler, T., Jungclaus, J., Danek, C., Goessling, H. F., Koldunov, N. V., Rackow, T., and Sidorenko, D.: Ocean Model Formulation Influences Transient Climate Response, Journal of Geophysical Research: Oceans, 126, e2021JC017 633, https://doi.org/https://doi.org/10.1029/2021JC017633, e2021JC017633 2021JC017633, 2021.

Shaffer, G., Huber, M., Rondanelli, R., and Pepke Pedersen, J. O.: Deep time evidence for climate sensitivity increase with warming, Geophysical Research Letters, 43, 6538–6545, https://doi.org/https://doi.org/10.1002/2016GL069243, 2016.

Shell, K. M., Kiehl, J. T., and Shields, C. A.: Using the Radiative Kernel Technique to Calculate Climate Feedbacks in NCAR's Community Atmospheric Model, Journal of Climate, 21, 2269–2282, https://doi.org/10.1175/2007JCLI2044.1, 2008.

Shendryk, Y., Rist, Y., Ticehurst, C., and Thorburn, P.: Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 157, 124–136, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2019.08.018, 2019.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., von der Heydt, A. S., Knutti, R., Mauritsen, T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D.: An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence, Reviews of Geophysics, 58, e2019RG000 678, https://doi.org/https://doi.org/10.1029/2019RG000678, e2019RG000678 2019RG000678, 2020.

Shi, C., Wang, C., Wang, Y., and Xiao, B.: Deep Convolutional Activations-Based Features for Ground-Based Cloud Classification, IEEE Geoscience and Remote Sensing Letters, 14, 816–820, https://doi.org/10.1109/lgrs.2017.2681658, 2017.

Smith, C. J. and Forster, P. M.: Suppressed Late-20th Century Warming in CMIP6 Models Explained by Forcing and Feedbacks, Geophysical Research Letters, 48, e2021GL094 948, https://doi.org/https://doi.org/10.1029/2021GL094948, e2021GL094948 2021GL094948, 2021.

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., and Shields, C. A.: Quantifying Climate Feedbacks Using Radiative Kernels, Journal of Climate, 21, 3504–3520, https://doi.org/10.1175/2007JCLI2110.1, 2008.

Stengel, M., Stapelberg, S., Sus, O., Finkensieper, S., Würzler, B., Philipp, D., Hollmann, R., Poulsen, C., Christensen, M., and McGarragh, G.: Cloud_cci Advanced Very High Resolution Radiometer post meridiem (AVHRR-PM) dataset version 3: 35-year climatology of global cloud and radiation properties, Earth System Science Data, 12, 41–60, https://doi.org/10.5194/essd-12-41-2020, 2020.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., eds.: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Tange, O. et al.: Gnu parallel-the command-line power tool, The USENIX Magazine, 36, 42–47, 2011.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

TensorFlow Developers: TensorFlow, https://www.tensorflow.org, last access: 18 February 2022, 2022.

The pandas development team: pandas-dev/pandas: Pandas, https://doi.org/10.5281/zenodo.3509134, 2020.

Tiedtke, M.: Representation of Clouds in Large-Scale Models, Monthly Weather Review, 121, 3040 – 3061, https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2, 1993.

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, Science Advances, 6, eaaz9549, https://doi.org/10.1126/sciadv.aaz9549, 2020.

Unidata, U. C. f. A. R.: Historical Unidata Internet Data Distribution (IDD) Global Observational Data, https://doi.org/10.5065/9235-WJ24, 2003.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and and, Y. V.-B.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Volodin, E.: The Mechanisms of Cloudiness Evolution Responsible for Equilibrium Climate Sensitivity in Climate Model INM-CM4-8, Geophysical Research Letters, 48, e2021GL096 204, https://doi.org/https://doi.org/10.1029/2021GL096204, e2021GL096204 2021GL096204, 2021.

Wall, C. J., Hartmann, D. L., and Ma, P.-L.: Instantaneous linkages between clouds and large-scale meteorology over the Southern Ocean in observations and a climate model, Journal of Climate, 30, 9455–9474, https://doi.org/10.1175/JCLI-D-17-0156.1, 2017.

Wang, C., Soden, B. J., Yang, W., and Vecchi, G. A.: Compensation Between Cloud Feedback and Aerosol-Cloud Interaction in CMIP6 Models, Geophysical Research Letters, 48, e2020GL091 024, https://doi.org/https://doi.org/10.1029/2020GL091024, e2020GL091024 2020GL091024, 2021.

Wessel, P. and Smith, W. H. F.: A global, self-consistent, hierarchical, high-resolution shoreline database, Journal of Geophysical Research: Solid Earth, 101, 8741–8743, https://doi.org/https://doi.org/10.1029/96JB00104, 1996.

Wessel, P. and Smith, W. H. F.: Global Self-consistent, Hierarchical, High-resolution Geography Database Version 2.3.7, https://www.soest.hawaii.edu/pwessel/gshhg/, last access: 14 February 2022, 2017.

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

WMO: Manual on Codes - International Codes, Volume I.1, Annex II to the WMO Technical Regulations: part A - Alphanumeric Codes, World Meteorological Organization (WMO), 2019 edition edn., 2011.

WMO: International Cloud Atlas: Manual on the Observation of Clouds and Other Meteors (WMO-No. 407), https://cloudatlas.wmo.int, 2021a.

WMO: Global Observing System, https://public.wmo.int/en/programmes/global-observing-system, 2021b.

Wohlfarth, K., Schröer, C., Klaß, M., Hakenes, S., Venhaus, M., Kauffmann, S., Wilhelm, T., and Wohler, C.: Dense Cloud Classification on Multispectral Satellite Imagery, in: 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), pp. 1–6, https://doi.org/10.1109/PRRS.2018.8486379, 2018.

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., and Döscher, R.: On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6, Geoscientific Model Development, 13, 3465–3474, https://doi.org/10.5194/gmd-13-3465-2020, 2020.

Ye, L., Cao, Z., and Xiao, Y.: DeepCloud: Ground-Based Cloud Image Categorization Using Deep Convolutional Features, IEEE Transactions on Geoscience and Remote Sensing, 55, 5729–5740, https://doi.org/10.1109/TGRS.2017.2712809, 2017.

Zantedeschi, V., Falasca, F., Douglas, A., Strange, R., Kusner, M. J., and Watson-Parris, D.: Cumulo: A Dataset for Learning Cloud Classes, 2020.

Zelinka, M. D.: Tables of ECS, Effective Radiative Forcing, and Radiative Feedbacks, https://github.com/mzelinka/cmip56_forcing_feedback_ecs, last access: 26 January 2022, 2021.

Zelinka, M. D., Klein, S. A., and Hartmann, D. L.: Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part I: Cloud Radiative Kernels, Journal of Climate, 25, 3715–3735, https://doi.org/10.1175/JCLI-D-11-00248.1, 2012.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of Higher Climate Sensitivity in CMIP6 Models, Geophysical Research Letters, 47, e2019GL085 782, https://doi.org/https://doi.org/10.1029/2019GL085782, e2019GL085782 10.1029/2019GL085782, 2020.

Zelinka, M. D., Klein, S. A., Qin, Y., and Myers, T. A.: Evaluating Climate Models' Cloud Feedbacks Against Expert Judgment, Journal of Geophysical Research: Atmospheres, 127, e2021JD035 198, https://doi.org/https://doi.org/10.1029/2021JD035198, e2021JD035198 2021JD035198, 2022.

Zhang, J., Liu, P., Zhang, F., and Song, Q.: CloudNet: Ground-Based Cloud Classification With Deep Convolutional Neural Network, Geophysical Research Letters, 45, 8665–8672, https://doi.org/https://doi.org/10.1029/2018GL077787, 2018.

Zhu, J., Poulsen, C. J., and Otto-Bliesner, B. L.: High climate sensitivity in CMIP6 model not supported by paleoclimate, Nature Climate Change, 10, 378–379, https://doi.org/10.1038/s41558-020-0764-6, 2020.

Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Poulsen, C. J., Tierney, J. E., Lofverstrom, M., and DiNezio, P.: Assessment of Equilibrium Climate Sensitivity of the Community Earth System Model Version 2 Through Simulation of the Last Glacial Maximum, Geophysical Research Letters, 48, e2020GL091 220, https://doi.org/https://doi.org/10.1029/2020GL091220, e2020GL091220 2020GL091220, 2021.