Climate models with moderate climate sensitivity best simulate the magnitude of Earth's energy imbalance

Kyriaki Bimpiri[1,2], Thomas Hocking[1,2], and Thorsten Mauritsen[1,2]

[1]Department of Meteorology, Stockholm University, Stockholm, Sweden
[2]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Correspondence: Kyriaki Bimpiri (kyriaki.bimpiri@misu.su.se)

Abstract. Recent studies have highlighted that state-of-the-art climate models are not able to simulate the large observed trend in Earth's energy imbalance. Here we evaluate climate models' ability to represent both the trend and the magnitude of the imbalance, while accounting for model energy leakage and remnant drift. As reference we use satellite observations and we find that every observed annual mean energy imbalance is within the range simulated by models, including the record year 2023, and when averaged over the 2001-2024 period, 15 out of 30 models simulate magnitudes of the imbalance that are statistically consistent with the observations. Models, however, generally underestimate the positive trend in the energy imbalance, albeit barely within the range of uncertainty. We suspected that a discontinuity in volcanic forcing between the historical and future scenario in 2014-2015 could have caused the underestimated trend, but only found evidence of such artifacts for a few models. Finally, we find a weak correlation between short-term decadal warming and energy imbalance, but a surprisingly close relationship between energy imbalance and equilibrium climate sensitivity. Based on observational constraints, the relationship suggests that models with moderate climate sensitivity are most realistic.

1 Introduction

Earth's energy imbalance, expressed as the difference between incoming solar and outgoing radiation at the top-of-atmosphere, is one of the most fundamental metrics of the climate system (Fourier, 1822; Arrhenius, 1896; Manabe and Strickler, 1964; Hansen et al., 2011; von Schuckmann et al., 2023). Today, the climate system is out of balance, mainly due to anthropogenic greenhouse gas emissions (Houghton et al., 2001), and a positive trend in the Earth's energy imbalance has been evident in satellite observations in recent decades (Loeb et al., 2024). The climate science community has indicated that the imbalance is rising faster than expected, and climate models are not able to reproduce this observed trend (Raghuraman et al., 2021; Hodnebrog et al., 2024; Olonscheck and Rugenstein, 2024; Mauritsen et al., 2025; Myhre et al., 2025). In this study we evaluate the performance of global climate models in simulating both the trend and the magnitude of the Earth's energy imbalance.

The value of observing the imbalance trend is indispensable for both science and climate policy. By definition the imbalance trend is the rate of change of accumulated energy in the climate system, so in a broad sense this trend determines the rate of climate change, with 89% of the excess heat stored in the ocean, 6% on land, 4% in the cryosphere while 1% ends up within the atmosphere (von Schuckmann et al., 2023). Specifically, the imbalance leads to rising temperatures, rising sea level, and
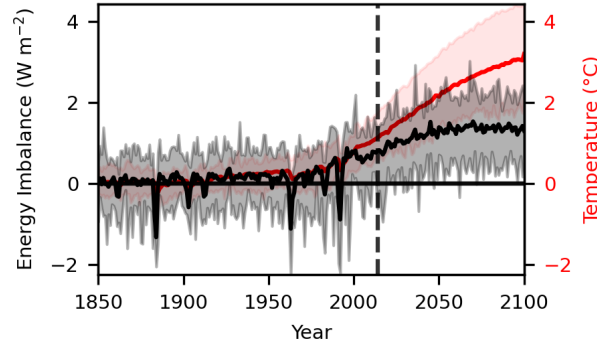
Figure 1. Energy imbalance (black line) and range of minimum and maximum imbalance in CMIP6 ensemble members (grey shaded). Temperature (red line) and ensemble range (red shaded area). The vertical dashed line corresponds to the transition year (2014) between the historical experiment and the SSP2-4.5 scenario.

more extreme weather. Climate models that run the Shared Socioeconomic Pathway SSP2-4.5 scenario reveal an increase in the imbalance, reaching a nearly constant level in the second half of the 21st century, while the temperature continues to rise through 2100 surpassing 3 °C (Fig. 1). Under more stringent mitigation scenarios that limit global warming to less than 2 °C, the energy imbalance is expected to peak already in the 2030s, several decades before the surface temperature stabilises (Mauritsen et al., 2025). Thus, Earth's energy imbalance can act as an indicator of future temperature change as a result of anthropogenic activity.

The Earth's energy imbalance (N) is influenced by a number of factors, including radiative forcing, feedbacks and internal variability. In a linearised framework the change in imbalance N can be expressed as:

$$N = F + \lambda \Delta T_s + \epsilon \tag{1}$$

in relation to the radiative forcing F, the feedback λ in response to a surface temperature change ΔT_s and the internal variability of the system ϵ. The three terms result from distinct processes:

- The dominant forcing originates from increased concentrations of greenhouse gases, primarily CO_2. Furthermore, anthropogenic aerosol emissions result in a negative forcing that translates to a cooling effect, due to an increase in the reflected shortwave radiation to space. There is considerable uncertainty in the strength and evolution of the aerosol cooling (Bellouin et al., 2020; Forster et al., 2021). However, there is evidence for a weakly decreasing aerosol cooling effect during approximately the past 20 years (Quaas et al., 2022; Forster et al., 2025), and this has contributed to the upward trend in simulated energy imbalance (Hodnebrog et al., 2024).

- The equilibrium climate sensitivity (ECS) of the system, which is the long-term temperature increase in response to a doubling of the atmospheric CO_2 concentration relative to pre-industrial values, can be defined in relation to the total

feedback parameter λ and the equivalent forcing $F_{2\times\mathrm{CO_2}}$:

$$\mathrm{ECS} = -\frac{F_{2\times\mathrm{CO_2}}}{\lambda} \tag{2}$$

A large negative λ is associated with a low climate sensitivity. As can be seen from Eq. 1, a large negative λ results in a strong dampening of the energy imbalance when temperatures rise, and Myhre et al. (2025) argue that low climate sensitivity models are not able to reproduce the trend in the shortwave component of the energy imbalance. Therefore, a high climate sensitivity is another candidate explanation for the rise in the energy imbalance.

– Finally, internal variability may be involved in causing the large trend in the Earth's energy imbalance, although it should be pointed out that internal variability alone cannot fully explain the observed trend without changes in external forcing (Raghuraman et al., 2021). Internal variability at the global scale can be caused both by atmospheric processes that lead to variations in the energy imbalance and by exchange of energy with the deep ocean, by as much as ± 0.2 $\mathrm{Wm^{-2}}$ on average over 15-year periods (Hedemann et al., 2017). Thus, internal variability can be an important factor when assessing the model-simulated magnitude and trend of the energy imbalance.

The central focus of this study is to evaluate models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) regarding the simulation of both the magnitude and the trend of the energy imbalance (Section 3). For this purpose, we subtract the pre-industrial control experiment imbalance to account for energy leakage and remnant model drifts (Section 2). We explore whether the implementation of volcanic aerosols in the future scenario could have played a role in the underestimated trend (Section 4). Finally, we investigate the relationship between the present-day imbalance and short-term to long-term global warming (Section 5).

2 Data and Methods

2.1 Observations

We use data from the Clouds and Earth's Radiant Energy System (CERES) mission as our observational reference, with complementary reconstructed data to extend our analysis period further back in the past. For the period 2001-2024, we use the CERES energy-balanced and filled (EBAF) product, version 4.2.1 (NASA/LARC/SD/ASDC, 2023; NASA, Langley Research Center, 2025). This data product combines radiation budget data from multiple satellites, which are known to have a bias of several $\mathrm{Wm^{-2}}$ in the global annual mean before adjustments (Fig. A1). These data are adjusted within their uncertainties to match the global annual mean imbalance for 2005-2015 as determined from measurements of ocean heat content, so that the final EBAF data do not exhibit this bias (Loeb et al., 2018).

For the period 1985-2000, we use data from the Diagnosing Earth's Energy Pathways in the Climate system (DEEP-C) project version 5.0, which is based on satellite observations from the Earth Radiation Budget Satellite (ERBS), CERES, atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ERA5) and model simulations from the
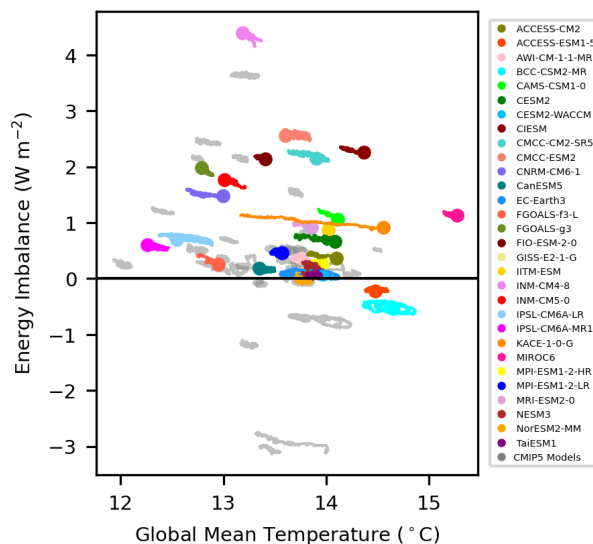
Figure 2. Energy imbalance versus near-surface air temperature for the piControl experiment. Lines correspond to the 50-year running mean of the full piControl simulation for CMIP6 models (colours) and CMIP5 models (grey). Circles correspond to the mean of the last 50 years.

Atmospheric Model Intercomparison Project in CMIP6 (Liu and Allan, 2022). We refer to the combined 1985-2024 dataset as

75 CERES EBAF extended.

As for temperature observations, the HadCRUT5 Analysis version 5.1.0.0 dataset from the Met Office Hadley Centre Climatic Research Unit, University of East Anglia is used (Morice et al., 2021).

2.2 Model experiments

We use output data from 239 ensemble members of 30 climate models participating in CMIP6 (Eyring et al., 2016). The models

80 are listed in Table A1. In our analysis we combine the historical experiment from 1850 to 2014 with the SSP2-4.5 scenario from 2014 to 2100 to produce time series of the annual global mean imbalance and near-surface air temperature from 1850 to 2100.

The top-of-atmosphere energy imbalance of a model that perfectly conserves energy is expected to be close to zero after the model has been run for thousands of years. However, most models do not perfectly conserve energy and some models are also

85 drifting in their piControl experiments (Mauritsen et al., 2012). Figure 2 shows the individual model energy imbalance values of the piControl experiment ranging from about -3 Wm^{-2} to more than 4 Wm^{-2}. A positive energy imbalance corresponds to energy leakage, and a negative energy imbalance is associated with an artificial input or source of energy. To account for this, we adjusted the imbalance time series to compensate for both artificial energy leakage and remnant drift by subtracting the global time-mean imbalance found in the piControl experiment for each model. It is noted that the drift in energy imbalance in

90 these 500 years or longer piControl experiments is relatively small, on the order of $0.1\,\mathrm{Wm^{-2}}$ in most models, such that the error caused by assuming a constant leakage is negligible.

2.3 The two-layer model

As outlined in the introduction, we expect a relationship between the simulated energy imbalance and climate sensitivity. To explore this relationship we will apply the widely used two-layer model with a pattern effect (Winton et al., 2010; Armour

95 et al., 2013; Geoffroy et al., 2013; Rohrschneider et al., 2019):

$$C\frac{dT}{dt} = F + \lambda T - \varepsilon\kappa(T - T_d)$$
$$C_d\frac{dT_d}{dt} = \kappa(T - T_d)$$
$$N = C\frac{dT}{dt} + C_d\frac{dT_d}{dt}$$

where N is the energy imbalance, F the effective radiative forcing, λ the feedback parameter, C and C_d are the heat capacities of the upper and deep layer, T and T_d the respective temperatures, ε the ocean heat uptake efficacy used to represent the forced pattern effect, and κ is the deep ocean heat uptake coefficient. The parameters $\varepsilon = 1.3$ and $\kappa = 0.7\,\mathrm{Wm^{-2}K^{-1}}$ are chosen such

100 that when ECS is set to 3.0 K the transient climate response (TCR) is approximately 1.8 K, which are the combined assessed best estimates of Forster et al. (2021). The model is driven by the historical and SSP2-4.5 best estimate effective radiative forcing time series (Smith et al., 2021).

2.4 Magnitude, trend and confidence intervals

We compute the imbalance magnitude and trend along with 5-95% confidence intervals for both quantities, for observations

105 and models. The magnitude of the imbalance is calculated as the time mean over the selected time period, while the trend in the imbalance is quantified as the slope of a linear regression of the annual global mean values. In the case of models with only one realisation, as well as for the single time series of observational reference data, these two quantities are straightforward to compute. In the case of models with multiple ensemble members, the magnitude and trend are first calculated separately for each member, and the mean of those values are used to represent the model. Uncertainties are presented as 5-95% confidence

110 intervals (CI), which are computed using a standard error σ:

$$CI = \pm 1.64 \times \sigma \tag{3}$$

The standard error is computed slightly differently for different models and for observations. For models with only a single realisation, the standard error in the imbalance magnitude is computed directly from the annual time series, after first removing the linear trend over the period. The standard error in the imbalance trend is given by the corresponding linear regression.

115 For models with multiple realisations, we use the mean of the standard errors calculated for the individual realisations. The standard error in the observation-based imbalance magnitude is computed in the same way as for the single-realisation models, and furthermore the standard error in the observed trend additionally includes the effect of an explicit observational uncertainty
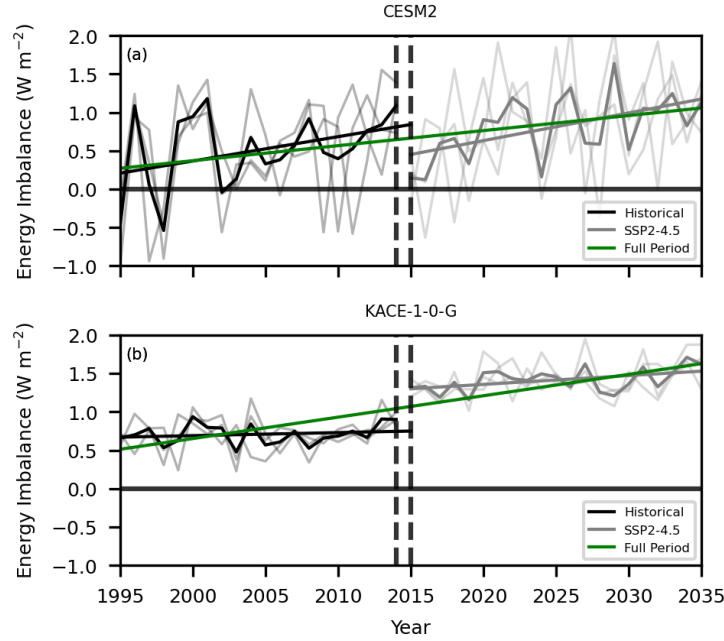
Figure 3. Energy imbalance in CMIP6 models CESM2 (a) and KACE-1-0-G (b) for 1995-2035. Time series of the historical experiment for 1995-2014 (black) and SSP2-4.5 future scenario for 2015-2035 (grey) along with a linear regression for each time period and the full time period (green). Year of transition between the two experiments in 2014-15 (black dashed vertical lines).

σ_{obs}. Specifically, we use $\sigma_{\mathrm{obs}} = 0.1 \ \mathrm{Wm}^{-2}$ and combine this with the standard error from a linear regression σ_{reg} as in Raghuraman et al. (2021) for the CERES EBAF product, but here we also include the same observational error for the CERES
120 EBAF extended dataset for simplicity:

$$\sigma = \sqrt{\sigma_{\mathrm{reg}}^2 + \sigma_{\mathrm{obs}}^2} \tag{4}$$

2.5 Difference in imbalance in the transition between experiments

To investigate how the overall calculation of the trend might be affected by the transition from the historical experiment to the SSP2-4.5 future scenario (Sect. 4), we quantify the difference in imbalance and clear-sky components of the net shortwave and
125 longwave radiation at the transition year 2015. We perform linear regressions for the periods immediately before (1995-2014) and after (2015-2035) the transition. Data from the historical experiment are used from 1995 to 2014, with an extrapolation for 2014-2015, while data from the SSP2-4.5 scenario are used from 2015 to 2035. To illustrate the method we show two examples in Fig. 3. For CESM2 there is a strong upward trend before and after the transition, but the negative shift leads to a weaker overall trend. For KACE-1-0-G there is a near-zero trend before and after the transition, but the upward shift, as we shall see,
130 leads to this model exhibiting the largest trend in the simulated energy imbalance.
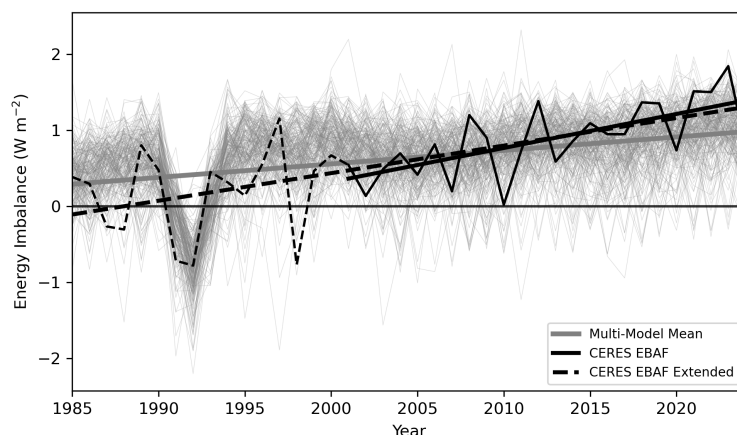
6

Figure 4. Energy imbalance simulated by CMIP6 models (grey lines) along with the multi-model mean regression line. The CERES EBAF values for the energy imbalance (black line) are also shown, in addition to the CERES EBAF Extended with DEEP-C data (dashed black line).

2.6 Climate sensitivity

Climate sensitivity values are retrieved from (Myhre et al., 2025), derived from abrupt $4xCO_2$ simulations over 150 years. Following Gregory et al. (2004), they calculate the effective climate sensitivity through regression of the energy imbalance and surface temperature in CMIP6 models. Because Myhre et al. (2025) do not include the CMCC-CM2-SR5 and CMCC-ESM2 models, we also exclude these models in Section 5.

3 Evaluation of CMIP6 modelled energy imbalance magnitude and trend

The evolution of the energy imbalance simulated by CMIP6 models was introduced earlier for the complete time series (Fig. 1), and in the following paragraphs we focus on the time period in which satellite observations are available.

First, when inspecting the full time series we find that model-simulated values of the energy imbalance in the beginning of the observational time period are higher than those observed, whereas the opposite is the case in the end of this period (Fig. 4). As a result the trend of the observed imbalance is steeper than the multi-model mean. Variability of the annual mean imbalance is evident within the different ensemble members, with some of the members simulating values ranging from about -2.0 Wm^{-2}, mainly during and right after the 1991 Pinatubo eruption, up to more than +2.0 Wm^{-2} in some other years. All the observed annual means, including the record peak value of 1.8 Wm^{-2} in 2023, are within the range simulated by the CMIP6 ensemble members. The fact that the model range captures the peak observations is related to internal variability. For comparison Myhre et al. (2025) found that the 2023 CERES EBAF value is higher than that of any of the CMIP6 models, but their study uses a single realisation for each model.
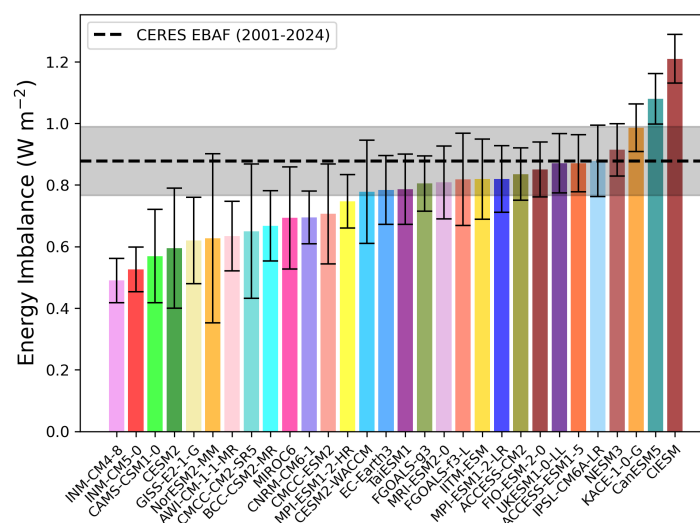
Figure 5. Energy imbalance magnitude of CMIP6 models (coloured bars) and observations (dashed black line) for the 2001-2024 CERES EBAF time period. The error bars for the models and the grey shaded area for the observational data correspond to the 5-95% confidence interval.
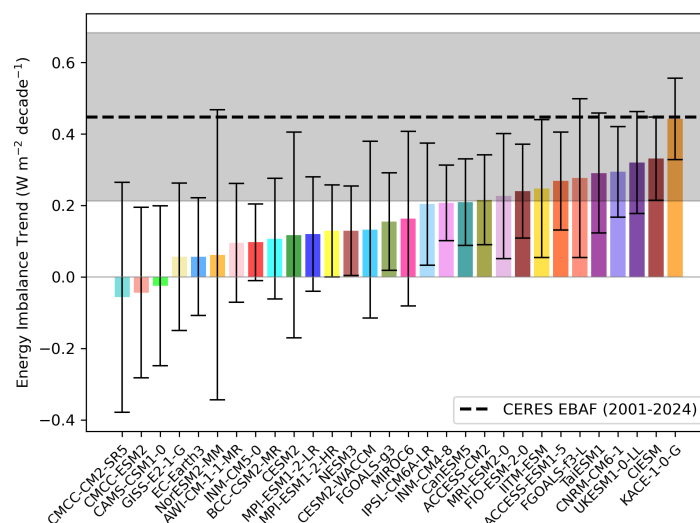

Figure 6. Energy imbalance trend of CMIP6 models (coloured bars) and observations (dashed black line) for the 2001-2024 CERES EBAF time period. The error bars for the models and the grey shaded area for the observational data correspond to the 5-95% confidence interval.

Second, we evaluate the temporal mean magnitude by comparing CMIP6 simulated values with the CERES EBAF data (Fig. 5). Our results indicate a good agreement between models and observations, and specifically 15 out of the 30 models are in agreement with the observational data within the 5-95% confidence interval. Two models simulate values above, while 13 models are below the observational confidence interval. An alternative is to use the model estimated confidence intervals, which vary substantially between models due to different levels of internal variability. Nevertheless, in this case a slightly different subset of 15 models have confidence intervals that overlap with the CERES EBAF mean. It is worth noting that even though many models are consistent with the observed magnitude of the energy imbalance, Fig. 4 shows that this is partly the result of larger simulated values at the beginning of the period, combined with smaller than observed model simulated values towards the end of the time period.

Third, we turn to the CMIP6 simulated trend of the energy imbalance. In line with earlier studies (Raghuraman et al., 2021; Hodnebrog et al., 2024; Myhre et al., 2025), we find that the majority of the models fail to reproduce the large observed trend in the energy imbalance (Fig. 6). The individual models exhibit a wide range in their calculations of the trend: Of the 30 CMIP6 models considered in this study, 11 are consistent with the observational confidence interval, while 3 models even simulate negative trends in their energy imbalance. If we instead use the individual model confidence intervals only 6 of them show an overlap with the CERES EBAF mean trend. This difference in the number of models that are consistent is partly due to varying internal variability in the models and the inclusion of measurement uncertainty in the observational confidence interval (Eq. 4). Regardless of how it is viewed, however, it seems clear that the majority of models are not simulating trends that are consistent with the CERES EBAF observations.

To investigate these trend results further, we evaluate CMIP6 model performance regarding the trend of the energy imbalance compared with all available observational data using two separate approaches. For both methods, 1985 is chosen as the baseline year, and the trend of the energy imbalance is computed over increasingly longer intervals, for consecutive end years of the time series from 1995 to 2024. Thus, the shortest trend is calculated over 11 years and the longest over 40 years.

In the first approach, we evaluate model performance in a binary way of categorization, using an agreement/disagreement method in which each individual model is considered consistent with observations when the model ensemble mean calculated trend (Sect. 2.4) falls within the 5-95% confidence interval of the observations. Initially, as illustrated in Fig. 7, agreement is found for most of the CMIP6 models, with a few exceptions in the beginning of the time period. This is only natural as 10 percent of the models should be outside the 5-95 percentiles of the observations. However, a nearly continuous disagreement is apparent in several models starting in 2008, and for trends calculated over the entire 1985-2024 period, 21 of the 30 models are inconsistent with the observations.

In the second approach, we use the entire set of individual model ensemble members and evaluate the observational data relative to the full spread of the models (Fig. 8). We find that the energy imbalance trend derived from the observational data lies mostly within the range of the CMIP6 ensemble members. However, from 2014 onwards the observed calculated trend is found to be at the upper bound of the model ensemble range, on the verge of exceeding the maximum values of the CMIP6 ensembles. Consequently, both of these methods indicate that many models fail to reproduce the observed trend after about 2010-2015.
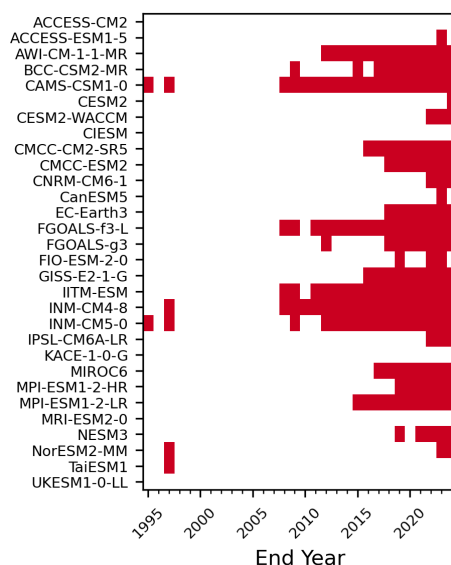
Figure 7. Agreement between CMIP6 models and CERES EBAF Extended observational data. Agreement (white shaded area), when the ensemble mean value of the energy imbalance trend lies within the 5-95% confidence interval of the observations. By contrast, disagreement (red shaded), when the model ensemble mean of the energy imbalance trend falls outside the 5-95% confidence interval.
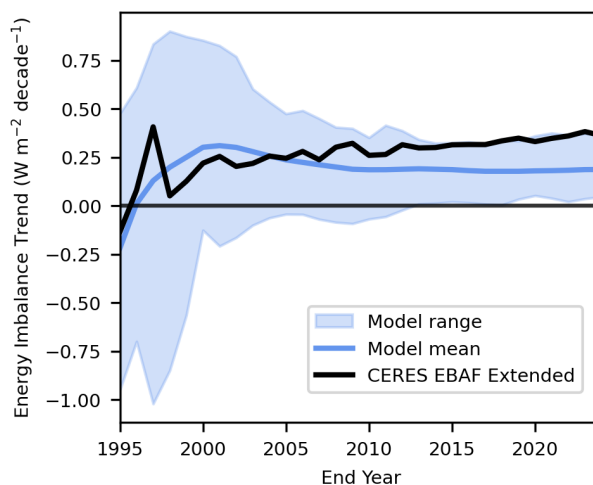

Figure 8. Energy imbalance trend per decade for starting year in 1985 and for consecutive end years from 1995 to 2024. The blue shaded area corresponds to the minimum and maximum of the ensemble members of CMIP6, the blue line is the ensemble mean, and the black line refers to the CERES EBAF extended observational dataset.

All in all, we can say that 1) on an individual annual mean basis, none of the observations are outside the range simulated by CMIP6 models, 2) the magnitude of the observed imbalance is also in line with CMIP6 models, but 3) most models systematically underestimate the observed trend. In the next section we will investigate whether the underestimated trend in recent years can be caused by implementation issues when going from the historical to future scenario experiments.

4 A potential source of the CMIP6 underestimated trend

Here, we investigate one possible cause, related to volcanic aerosol forcing in the future scenario, that could be responsible for the underestimation of the CMIP6 simulated energy imbalance trend in recent decades. Our idea is that the introduction of a constant background volcanic aerosol in future scenarios from 2015 onward, following the CMIP6 experimental protocol as described by O'Neill et al. (2016), would manifest as a negative shift or discontinuity in the imbalance between 2014 and 2015. Because the average historical volcanic aerosol loading is larger than the actual loading during 2015-2024, implementing this would have a negative impact on the model-simulated energy imbalance trend, and to a lesser extent also result in a lower magnitude of the energy imbalance. A discrepancy attributable to the volcanic aerosol forcing in the future scenarios should be most evident in the clear-sky component of the shortwave radiation.

We find, however, little evidence that models have generally implemented volcanic aerosols in the scenario according to protocol (Figs. 9 and A2). Instead, the majority of models show relatively small discontinuities, and most models actually exhibit positive shifts, which is inconsistent with the idea that the underestimated trend in energy imbalance is caused by too much volcanic aerosols in the future scenario.

Some models do show a negative discontinuity that appears consistent with protocol (Fig. A2). To corroborate this further, we inspect the clear-sky net shortwave and longwave fluxes (Figs. A3 and A4), here defined as positive downwards to be consistent with their impact on the energy imbalance. We do find negative shifts also in the clear-sky net shortwave radiation in CESM2, CESM2-WACCM, CMCC-CM2-SR5, CMCC-ESM2, GISS-E2-1-G, suggesting that these models followed the experimental protocol. Consequently, these models could have shown a larger positive trend in their energy imbalance if they had used more realistic volcanic aerosols after 2015. For example, CESM2 would have had a trend of about $0.34 \ \mathrm{Wm}^{-2}$ per decade over the period 1995 to 2035, twice as much as the model's average trend, which is more in line with the observed trends.

For the remaining models, there is either no discontinuity, or in some cases even positive shifts in both the energy imbalance and the clear-sky net shortwave radiation. The most pronounced cases of the latter is the IITM-ESM and KACE-1-0-G models. This could suggest that these models do not apply volcanic aerosols in the future scenario. For IITM-ESM this leads to an overall positive trend, despite exhibiting negative trend in the historical experiment. And for KACE-1-0-G, the discontinuity is likely the reason it appears to be very close to the observed trend (Fig. 6), despite having only weak trends before and after the transition (Figs. 3b and 9v).

Overall, our analysis suggests that the general CMIP6 model underestimation of the observed energy imbalance trend cannot be attributed to the implementation of the volcanic aerosol in the future scenario, since only a handful of models exhibit a
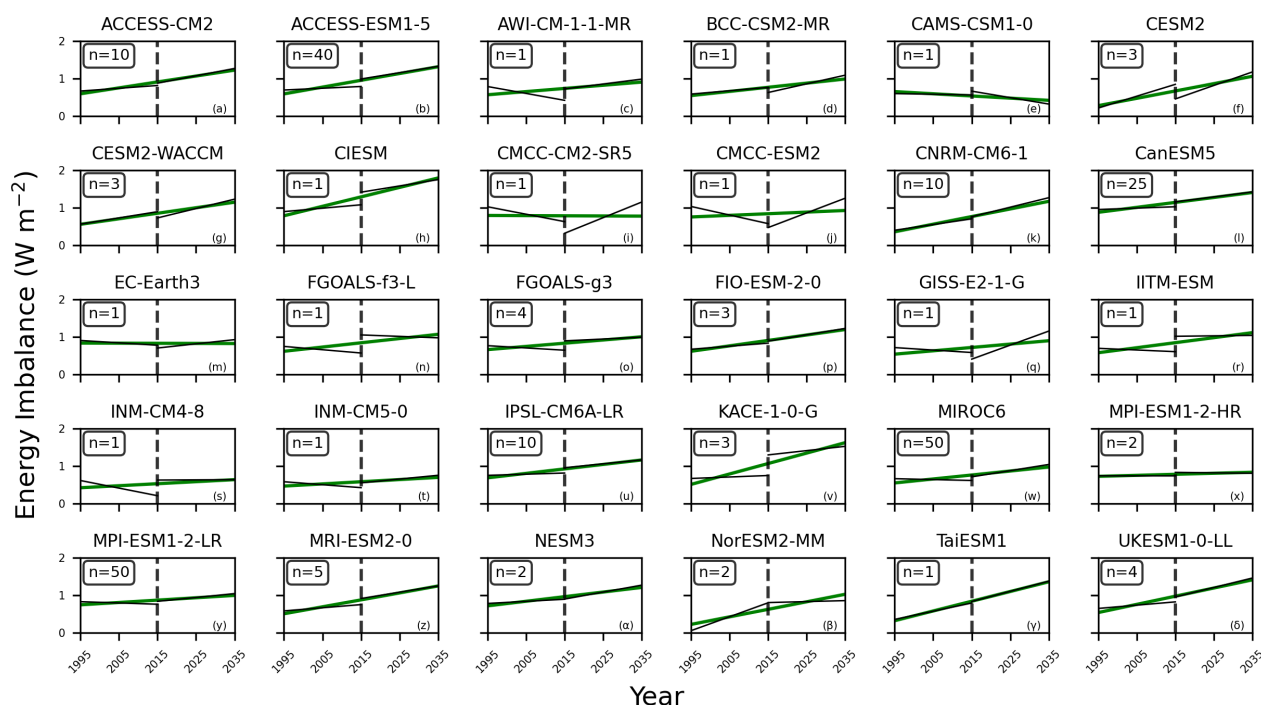
Figure 9. Linear fits (black) to the energy imbalance on 20-year time intervals before and after the year of transition (2015) from the historical experiment to the SSP2-4.5 scenario for the CMIP6 models. Green lines show linear fits on the combined 1995-2035 period.

negative shift in the clear-sky net shortwave flux around 2014-2015. On the contrary, some models exhibit a positive shift, suggesting that they do not apply volcanic aerosol in the scenario, making the simulated trend in the energy imbalance larger than what it should have been.

5 The relation between the imbalance and short-term and long-term warming

220 The energy imbalance is directly related to the accumulation of energy in the Earth system, so it is a natural constraint on the current rate of warming. But can it be used to constrain warming in the coming decade, or the long-term global warming? Myhre et al. (2025) found an emergent constraint relationship between the trend in absorbed shortwave radiation and end-of-century warming. Here we take a different approach and instead use the magnitude of the imbalance, based on the physical justification that the total imbalance is the actual energy accumulation rate.

225 We first focus on short-term decadal warming. Since we have 24 years of CERES EBAF data, we divide the time series in two, using the energy imbalance in the first half as input, and global warming between the first and the second period as the quantity to be predicted (Fig. 10). This approach allows us to verify the resulting emergent constraint against observations, something that is rarely possible in climate research. The fitted regression line to all model runs (black) shows a steep and
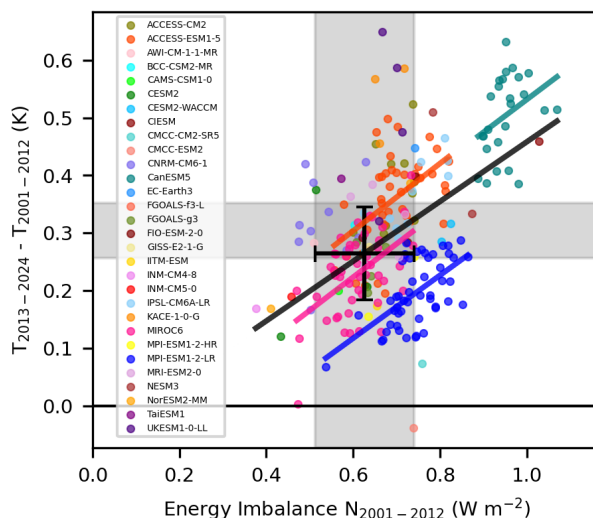
Figure 10. Scatter plot of mean energy imbalance 2001-2012 and change in near surface air temperature from 2001-2012 to 2013-2024 as simulated by CMIP6 models. Linear regressions are shown for individual models (coloured lines) as well as for the multi-model ensemble (black line). The shaded regions show the observed temperature change and energy imbalance. The error bar corresponds to the predicted temperature change as per the emergent constraint between temperature change and energy imbalance.

physically reasonable relationship with an intercept close to zero, such that no imbalance leads to a prediction of no warming. The line also intersects the actual observed imbalance and global warming, and therefore correctly predicts the observed decadal warming. However, although the multi-model relation successfully predicts the observed global warming, individual model ensembles are far off, with one ensemble member showing global cooling and another showing more than twice as fast warming as observed, both cases with an imbalance within the observed confidence interval.

To investigate this problem further, we take advantage of the four largest single-model ensembles available in CMIP6: ACCESS-ESM1-5, CanESM5, MIROC6 and MPI-ESM1-2-LR. For each of these models we fit a separate regression line in colours (Fig. 10). The slopes are surprisingly close to that of the entire CMIP6 ensemble, and in three of the four cases the regression line intersects the observed imbalance and warming. This includes CanESM5, a high-sensitivity model, even though it has all its ensemble members simulating both more imbalance and warming than observed. Thus, although an outlier with respect to the observations, the model provides useful information for constraining decadal warming. The relationship for MPI-ESM1-2-LR is shifted below the observations, although some of its individual ensemble members matched observations. Figure 9y shows a small upward shift and change in trend for this model, although this does not explain why that model under-predicted the decadal warming.

We interpret these results to mean that the skill in predicting decadal global warming arises from internal variability, given that the same slope is found in all single-model ensembles. The idea is that ensemble members that are colder than expected in the first period will show a larger imbalance as the negative feedback term ($\lambda \Delta T_s$) in Eq. 1 is small, and these realisations
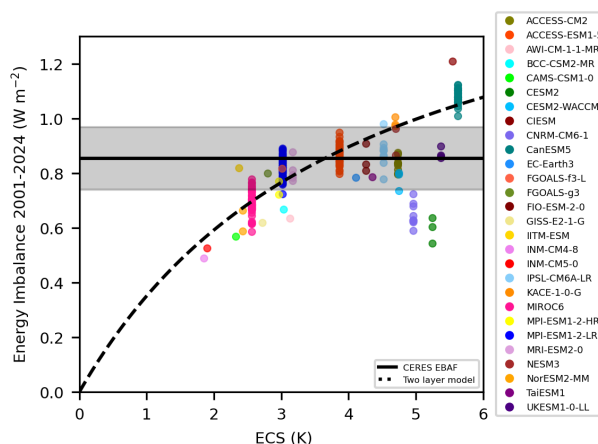
Figure 11. Scatter plot of mean energy imbalance and ECS over the CERES EBAF time period for CMIP6 models (colours) and a two-layer model (black dashed line). CERES EBAF mean energy imbalance (black solid line) along with the 5-95% confidence interval (grey shaded).

therefore warm relatively more in the second period. Nevertheless, some models show offsets that are too large to be purely associated with internal variability. As such the analysis shows that there is great potential to predict decadal-scale global warming, but such efforts are hampered by systematic model issues.

Moving focus to long-term warming, we choose to focus on the relationship between the current energy imbalance and equi-
250 librium climate sensitivity (ECS= $-F_{2x}/\lambda$). Here the rationale is that if all models applied the same forcing, then models with a large ECS will have a small negative $\lambda\Delta T$ and therefore a larger mean imbalance during the ongoing transient warming. The expected relationship is shown by varying only ECS in the two-layer model in Fig. 11. To obtain the maximum signal-to-noise ratio we use the entire CERES EBAF period. We see that indeed the vast majority of CMIP6 models closely follow the expected behaviour with low ECS models showing a low imbalance, and high ECS models simulating more energy accumulation.
255 The observed imbalance favours models with moderate ECS, and shows that some models with very low or very high ECS are less consistent with the observations, lending further support for recent assessments (Sherwood et al., 2020; Forster et al., 2021). The notable exceptions are the CNRM-CM6-1 and the CESM2 models, and to a lesser extent the CESM2-WACCM model. Inspecting Fig. A2, it is clear that both of the CESM2 model versions show a negative shift in the transition between the historical experiment and the future scenario, and if that had not been the case, both models would have been in line with the
260 energy imbalance simulated by the two-layer model. For the CNRM-CM6-1 ensemble we have not found a viable explanation.

In summary, we find that CMIP6 models in general show only a weak relationship between present-day imbalance and decadal warming, whereas model internal variability shows some skill in predicting short-term warming. A surprisingly close relationship between present-day energy imbalance and long-term warming as represented by ECS suggests that models with moderate climate sensitivity are more realistic.

14

6 Conclusions

In this study we have investigated the ability of climate models to simulate the magnitude and trend of the energy imbalance at the top-of-atmosphere, while accounting for model energy leakage and remnant drift. We find that climate models exhibit a robust agreement with the observed magnitude of the energy imbalance (5-95% confidence), and that the observed individual annual mean imbalance is within the range simulated by models over the complete time series, including the extreme year 2023.

Regarding the trend, and in line with several recent studies (Raghuraman et al., 2021; Hodnebrog et al., 2024; Myhre et al., 2025), we find that most models underestimate the trend of the energy imbalance in relation to observations. Extending the analysis using the DEEP-C data we also show that this underestimation becomes more pronounced in the most recent decade. That said, despite the general underestimation, a subset of the CMIP6 models do simulate trends that are consistent with observations: either the simulated trends with internal variability are consistent with the observed trend, or the simulated trends are within the observational uncertainty.

A particular concern we had was that the introduction of the historical mean volcanic aerosol forcing in the future scenario, as recommended by the CMIP6 experiment protocol, could have caused a low bias in the simulated energy imbalance trend since no major volcanoes erupted after 2014. Although we find evidence that a few models could have done this in a negative discontinuity of the clear-sky shortwave radiation between 2014-15, this does not appear to be a general issue, and some models even show the opposite behaviour. Efforts to simulate the top-of-atmosphere energy imbalance on a more continuous basis could help studies like this in the future (Schmidt et al., 2023).

Finally, we investigated whether the energy imbalance can be used to constrain short-term and long-term global warming. For short-term decadal warming, models show a weak relationship with present imbalance, but internal variability from single model ensembles does show some potential skill in predicting short-term warming. We do, however, find a surprisingly close relationship between present-day imbalance and equilibrium climate sensitivity, which, when combined with the observed CERES EBAF imbalance suggests that models with moderate climate sensitivity, and hence moderate long-term warming (Grose et al., 2018; Huusko et al., 2021), are more realistic.

Stepping back, based on the results presented here we cannot clearly distinguish whether the generally underestimated trend is due to incorrect model forcing, feedbacks or an expression of internal variability. Nevertheless, the fact that models do simulate the magnitude of the imbalance, first above and then below the observations, favours internal variability or incorrectly applied forcing, over an underestimated climate sensitivity. To be more concrete, the 1998-2012 period is known as the global warming hiatus period during which the global mean temperature rose slower than expected (Modak and Mauritsen, 2021), which was dominated by La Niña conditions in the Pacific, causing more low-level cloudiness with a dampening effect on the Earth's energy imbalance by reflecting sunlight (Zhou et al., 2016). Subsequently, the Pacific has returned to more normal conditions with a mix of both El Niños and La Niñas, which could have contributed to the upward trend in absorbed shortwave radiation. Certainly, future studies should be able to bring more clarity on the actual cause of the trend with more data in the coming years.

Author contributions. KB carried out most of the study, including the data analysis. All authors contributed to the study and the writing.

Competing interests. The authors declare no competing interests.

References

Armour, K. C., Bitz, C. M., and Roe, G. H.: Time-Varying Climate Sensitivity from Regional Feedbacks, Journal of Climate, 26, 4518 –
4534, https://doi.org/10.1175/JCLI-D-12-00544.1, 2013.

Arrhenius, S.: XXXI. On the influence of carbonic acid in the air upon the temperature of the ground, The London, Edinburgh, and Dublin
Philosophical Magazine and Journal of Science, 41, 237–276, https://doi.org/10.1080/14786449608620846, 1896.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-
L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T.,
McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourde-
val, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding Global Aerosol Radiative Forcing of Climate Change, Reviews
of Geophysics, 58, e2019RG000 660, https://doi.org/https://doi.org/10.1029/2019RG000660, e2019RG000660 10.1029/2019RG000660,
2020.

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.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D., Mauritsen, T., Palmer, M., Watanabe, M., Wild,
M., and Zhang, H.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity, pp. 943–1054, Cambridge University Press,
Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.009, 2021.

Forster, P. M., Smith, C., Walsh, T., Lamb, W. F., Lamboll, R., Cassou, C., Hauser, M., Hausfather, Z., Lee, J.-Y., Palmer, M. D., von
Schuckmann, K., Slangen, A. B. A., Szopa, S., Trewin, B., Yun, J., Gillett, N. P., Jenkins, S., Matthews, H. D., Raghavan, K., Ribes, A.,
Rogelj, J., Rosen, D., Zhang, X., Allen, M., Aleluia Reis, L., Andrew, R. M., Betts, R. A., Borger, A., Broersma, J. A., Burgess, S. N.,
Cheng, L., Friedlingstein, P., Domingues, C. M., Gambarini, M., Gasser, T., Gütschow, J., Ishii, M., Kadow, C., Kennedy, J., Killick, R. E.,
Krummel, P. B., Liné, A., Monselesan, D. P., Morice, C., Mühle, J., Naik, V., Peters, G. P., Pirani, A., Pongratz, J., Minx, J. C., Rigby,
M., Rohde, R., Savita, A., Seneviratne, S. I., Thorne, P., Wells, C., Western, L. M., van der Werf, G. R., Wijffels, S. E., Masson-Delmotte,
V., and Zhai, P.: Indicators of Global Climate Change 2024: annual update of key indicators of the state of the climate system and human
influence, Earth System Science Data, 17, 2641–2680, https://doi.org/10.5194/essd-17-2641-2025, 2025.

Fourier, J.: Théorie analytique de la chaleur, Didot, 1822.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., and Tytéca, S.: Transient Climate Response in a Two-Layer Energy-
Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs, Journal of
Climate, 26, 1859 – 1876, https://doi.org/10.1175/JCLI-D-12-00196.1, 2013.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and
Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31,
https://doi.org/https://doi.org/10.1029/2003GL018747, 2004.

Grose, M. R., Gregory, J., Colman, R., and Andrews, T.: What Climate Sensitivity Index Is Most Useful for Projections?, Geophysical
Research Letters, 45, 1559–1566, https://doi.org/10.1002/2017GL075742, 2018.

Hansen, J., Sato, M., Kharecha, P., and von Schuckmann, K.: Earth's energy imbalance and implications, Atmospheric Chemistry and
Physics, 11, 13 421–13 449, https://doi.org/10.5194/acp-11-13421-2011, 2011.

Hedemann, C., Mauritsen, T., Jungclaus, J., and Marotzke, J.: The subtle origins of surface-warming hiatuses, Nature Climate Change, 7, 336–339, https://doi.org/10.1038/nclimate3274, 2017.

Hodnebrog, O., Myhre, G., Jouan, C., Andrews, T., Forster, P., Jia, H., Loeb, N., Olivié, D., Paynter, D., Quaas, J., Raghuraman, S. P., and Schulz, M.: Recent reductions in aerosol emissions have increased Earth's energy imbalance, Communications Earth Environment, 5, https://doi.org/10.1038/s43247-024-01324-8, 2024.

Houghton, J., Ding, Y., Griggs, D., Noguer, M., van der Linden, P., Dai, X., Maskell, M., and Johnson, C.: Climate Change 2001: The Scientific Basis, vol. 881., p. 881, 2001.

Huusko, L. L., Bender, F. A.-M., Ekman, A. M. L., and Storelvmo, T.: Climate sensitivity indices and their relation with projected temperature change in CMIP6 models, Environmental Research Letters, 16, 064 095, https://doi.org/10.1088/1748-9326/ac0748, 2021.

Liu, C. and Allan, R.: Reconstructions of the radiation fluxes at the top of atmosphere and net surface energy flux: DEEP-C Version 5.0, https://doi.org/10.17864/1947.000347, dataset, 2022.

Loeb, N., Ham, S.-H., Allan, R., Thorsen, T., Meyssignac, B., Kato, S., Johnson, G., and Lyman, J.: Observational Assessment of Changes in Earth's Energy Imbalance Since 2000, Surveys in Geophysics, 45, 1757–1783, https://doi.org/10.1007/s10712-024-09838-8, 2024.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product, Journal of Climate, 31, 895 – 918, https://doi.org/10.1175/JCLI-D-17-0208.1, 2018.

Manabe, S. and Strickler, R. F.: Thermal Equilibrium of the Atmosphere with a Convective Adjustment, Journal of Atmospheric Sciences, 21, 361 – 385, https://doi.org/10.1175/1520-0469(1964)021<0361:TEOTAW>2.0.CO;2, 1964.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, Journal of Advances in Modeling Earth Systems, 4, https://doi.org/https://doi.org/10.1029/2012MS000154, 2012.

Mauritsen, T., Tsushima, Y., Meyssignac, B., Loeb, N. G., Hakuba, M., Pilewskie, P., Cole, J., Suzuki, K., Ackerman, T. P., Allan, R. P., Andrews, T., Bender, F. A.-M., Bloch-Johnson, J., Bodas-Salcedo, A., Brookshaw, A., Ceppi, P., Clerbaux, N., Dessler, A. E., Donohoe, A., Dufresne, J.-L., Eyring, V., Findell, K. L., Gettelman, A., Gristey, J. J., Hawkins, E., Heimbach, P., Hewitt, H. T., Jeevanjee, N., Jones, C., Kang, S. M., Kato, S., Kay, J. E., Klein, S. A., Knutti, R., Kramer, R., Lee, J.-Y., McCoy, D. T., Medeiros, B., Megner, L., Modak, A., Ogura, T., Palmer, M. D., Paynter, D., Quaas, J., Ramanathan, V., Ringer, M., von Schuckmann, K., Sherwood, S., Stevens, B., Tan, I., Tselioudis, G., Sutton, R., Voigt, A., Watanabe, M., Webb, M. J., Wild, M., and Zelinka, M. D.: Earth's Energy Imbalance More Than Doubled in Recent Decades, AGU Advances, 6, e2024AV001 636, https://doi.org/https://doi.org/10.1029/2024AV001636, e2024AV001636 2024AV001636, 2025.

Modak, A. and Mauritsen, T.: The 2000–2012 Global Warming Hiatus More Likely With a Low Climate Sensitivity, Geophysical Research Letters, 48, e2020GL091 779, https://doi.org/https://doi.org/10.1029/2020GL091779, 2021.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/https://doi.org/10.1029/2019JD032361, e2019JD032361 2019JD032361, 2021.

Myhre, G., Øivind Hodnebrog, Loeb, N., and Forster, P. M.: Observed trend in Earth energy imbalance may provide a constraint for low climate sensitivity models, Science, 388, 1210–1213, https://doi.org/10.1126/science.adt0647, 2025.

NASA, Langley Research Center: $CERES_E BAF_E d4.2 and Ed4.2.1 Data Quality Summary Version 5$, , 2025.

NASA/LARC/SD/ASDC: CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data in netCDF
385   Edition 4.2, https://doi.org/10.5067/TERRA-AQUA-NOAA20/CERES/EBAF_L3B004.2, 2023.

Olonscheck, D. and Rugenstein, M.: Coupled Climate Models Systematically Underestimate Radiation Response to Sur-
face Warming, Geophysical Research Letters, 51, e2023GL106 909, https://doi.org/https://doi.org/10.1029/2023GL106909,
e2023GL106909 2023GL106909, 2024.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F.,
390   Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP)
for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Quaas, J., Jia, H., Smith, C., Albright, A. L., Aas, W., Bellouin, N., Boucher, O., Doutriaux-Boucher, M., Forster, P. M.,
Grosvenor, D., Jenkins, S., Klimont, Z., Loeb, N. G., Ma, X., Naik, V., Paulot, F., Stier, P., Wild, M., Myhre, G., and Schulz,
M.: Robust evidence for reversal of the trend in aerosol effective climate forcing, Atmospheric Chemistry and Physics, 22,
395   12 221–12 239, https://doi.org/10.5194/acp-22-12221-2022, 2022.

Raghuraman, S. P., Paynter, D., and Ramaswamy, V.: Anthropogenic forcing and response yield observed positive trend in
Earth's energy imbalance, Nature Communications, 12, 4577, https://doi.org/10.1038/s41467-021-24544-4, 2021.

Rohrschneider, T., Stevens, B., and Mauritsen, T.: On simple representations of the climate response to external radiative
forcing, Climate Dynamics, 53, 3131–3145, https://doi.org/10.1007/s00382-019-04686-4, 2019.

400   Schmidt, G. A., Andrews, T., Bauer, S. E., Durack, P. J., Loeb, N. G., Ramaswamy, V., Arnold, N. P., Bosilovich, M. G., Cole,
J., Horowitz, L. W., Johnson, G. C., Lyman, J. M., Medeiros, B., Michibata, T., Olonscheck, D., Paynter, D., Raghuraman, S. P.,
Schulz, M., Takasuka, D., Tallapragada, V., Taylor, P. C., and Ziehn, T.: CERESMIP: a climate modeling protocol to investigate
recent trends in the Earth's Energy Imbalance, Frontiers in Climate, Volume 5 - 2023, https://doi.org/10.3389/fclim.2023.1202161,
2023.

405   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, 2020.

410   Smith, C., Hall, B., Dentener, F., Ahn, J., Collins, W., Jones, C., Meinshausen, M., Dlugokencky, E., Keeling, R., Krummel, P.,
Mühle, J., Nicholls, Z., and Simpson, I.: IPCC Working Group 1 (WG1) Sixth Assessment Report (AR6) Annex III Extended
Data (v1.0) [Data set], https://doi.org/10.5281/zenodo.5705391, 2021.

von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast, G., Adusumilli, S., Straneo, F., Ablain, M., Allan,
R. P., Barker, P. M., Beltrami, H., Blazquez, A., Boyer, T., Cheng, L., Church, J., Desbruyeres, D., Dolman, H., Domingues,
415   C. M., García-García, A., Giglio, D., Gilson, J. E., Gorfer, M., Haimberger, L., Hakuba, M. Z., Hendricks, S., Hosoda, S.,
Johnson, G. C., Killick, R., King, B., Kolodziejczyk, N., Korosov, A., Krinner, G., Kuusela, M., Landerer, F. W., Langer,

M., Lavergne, T., Lawrence, I., Li, Y., Lyman, J., Marti, F., Marzeion, B., Mayer, M., MacDougall, A. H., McDougall, T., Monselesan, D. P., Nitzbon, J., Otosaka, I., Peng, J., Purkey, S., Roemmich, D., Sato, K., Sato, K., Savita, A., Schweiger, A., Shepherd, A., Seneviratne, S. I., Simons, L., Slater, D. A., Slater, T., Steiner, A. K., Suga, T., Szekely, T., Thiery, W., Timmermans, M.-L., Vanderkelen, I., Wjiffels, S. E., Wu, T., and Zemp, M.: Heat stored in the Earth system 1960–2020: where does the energy go?, Earth System Science Data, 15, 1675–1709, https://doi.org/10.5194/essd-15-1675-2023, 2023.

Winton, M., Takahashi, K., and Held, I. M.: Importance of Ocean Heat Uptake Efficacy to Transient Climate Change, Journal of Climate, 23, 2333 – 2344, https://doi.org/10.1175/2009JCLI3139.1, 2010.

Zhou, C., Zelinka, M., and Klein, S.: Impact of decadal cloud variations on the Earth's energy budget, Nature Geoscience, 9, https://doi.org/10.1038/ngeo2828, 2016.

Model	Ensemble members
ACCESS-CM2	10
ACCESS-ESM1-5	40
AWI-CM-1-1-MR	1
BCC-CSM2-MR	1
CAMS-CSM1-0	1
CanESM5	25
CESM2	3
CESM2-WACCM	3
CIESM	1
CMCC-ESM2	1
CMCC-CM2-SR5	1
CNRM-CM6-1	10
EC-Earth3	1
FGOALS-f3-L	1
FIO-ESM-2-0	4
GISS-E2-1-G	1
INM-CM4-8	1
INM-CM5-0	1
IPSL-CM6A-LR	10
IITM-ESM	1
KACE-1-0-G	3
MIROC6	50
MPI-ESM1-2-HR	2
MPI-ESM1-2-LR	50
MRI-ESM2-0	5
NESM3	2
NorESM2-MM	2
TaiESM1	1
UKESM1-0-LL	4

Table A1. CMIP6 models and number of ensemble members used. Following the official CMIP6 data record we include realizations (r=1 to 50) according to availability, initialisation (i=1), physics (p=1) and forcing (f=1), while for UKESM1-0-LL and CNRM-CM6-1 we used f=2.
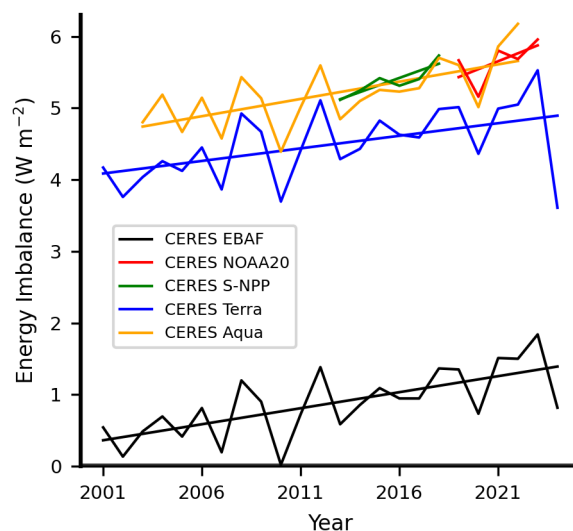
Figure A1. Annual time series of Earth's energy imbalance from the CERES EBAF and several SSF (single-satellite footprint) data products.
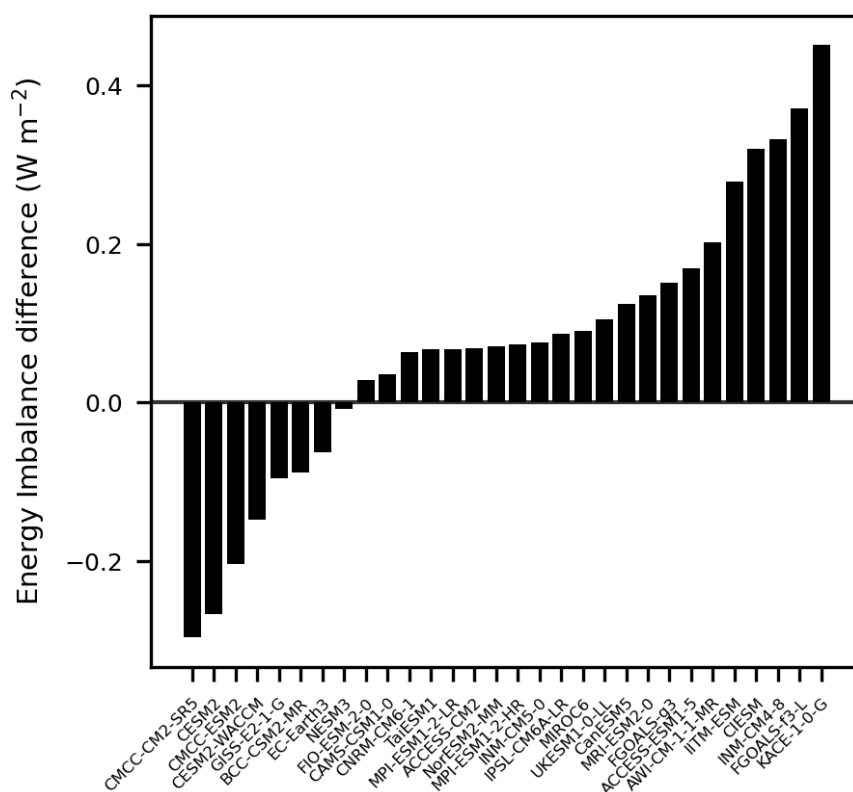
Figure A2. Bar plot of the energy imbalance difference between historical experiment in 2014 and SSP2-4.5 future scenario in 2015 for CMIP6 models.
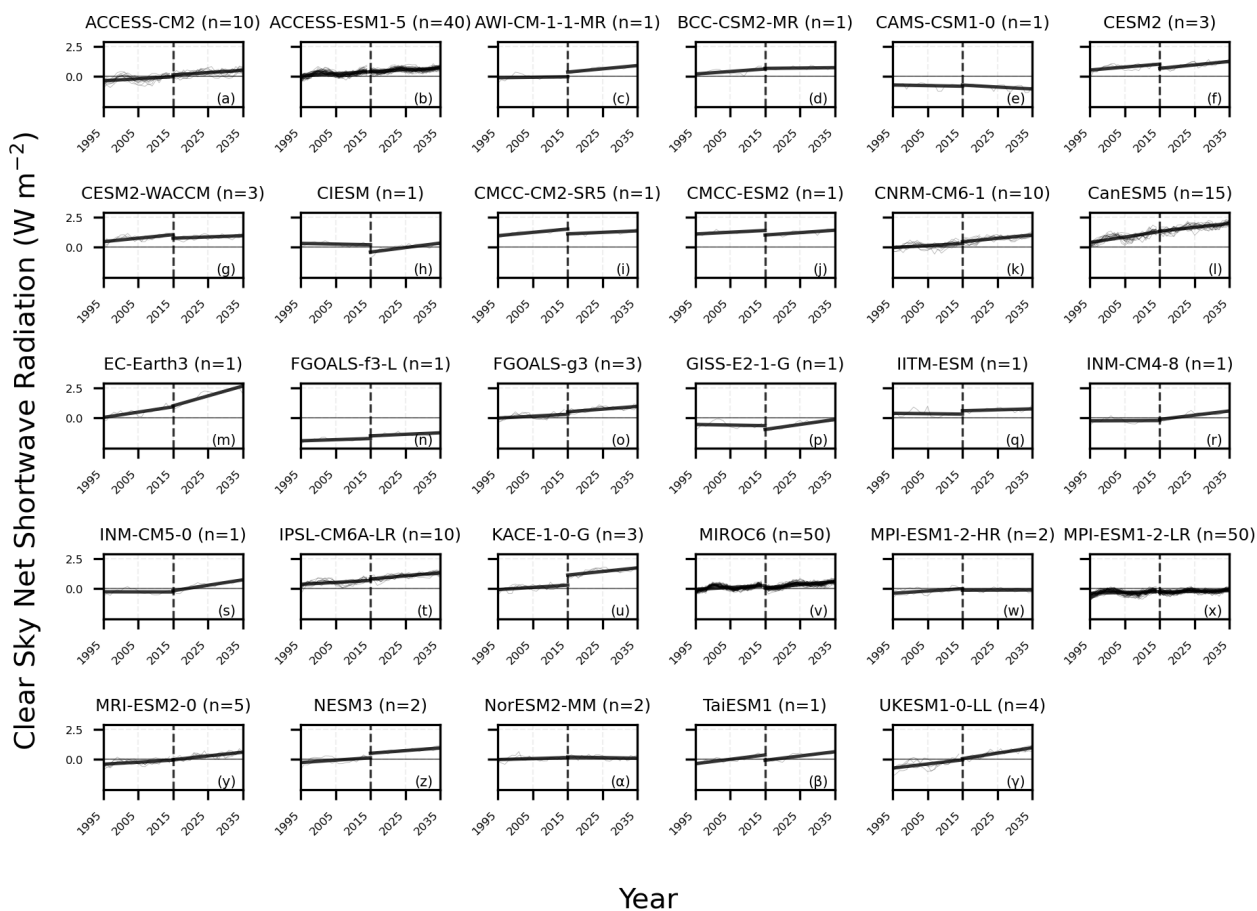
Figure A3. Clear-sky net shortwave radiation time series (grey) and linear fit (black) for historical experiment 1995-2014 and SSP2-4.5 scenario 2015-2035.
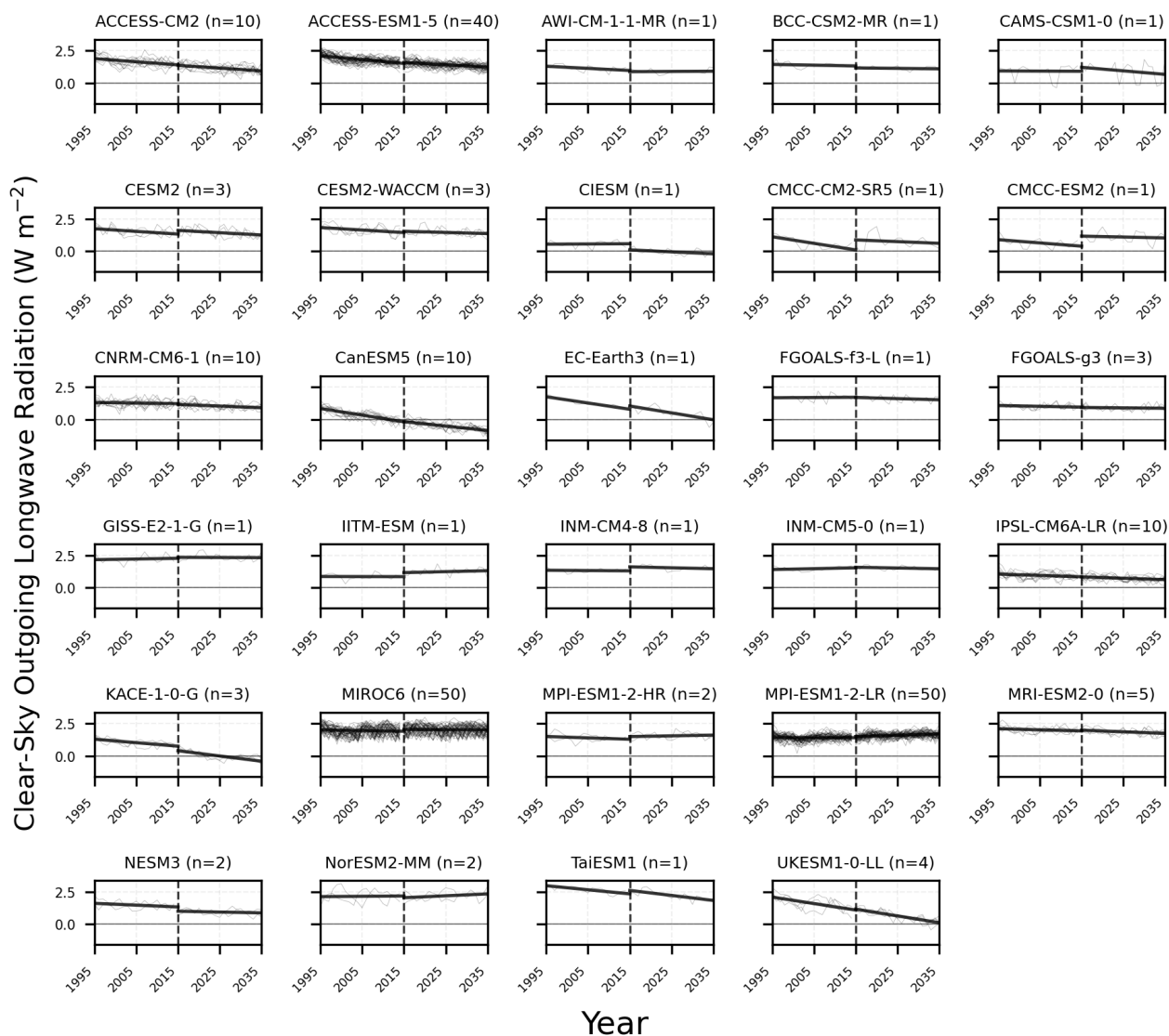
Figure A4. Clear-sky outgoing longwave radiation time series (grey) and linear fit (black) for historical experiment 1995-2014 and SSP2-4.5 scenario 2015-2035.