Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assessment of 21st-century planetary albedo changes in CMIP6 models

Aleksei Poliukhov

aeromsu@gmail.com

Lomonosov Moscow State University: Moskovskij gosudarstvennyj universitet imeni M V Lomonosova
https://orcid.org/0000-0001-6336-3376

Alexander Lutsenko

South-Russian State Polytechnic University

Research Article

Keywords: planetary albedo, CMIP6, CERES, model evaluation, model selection, Earth's energy budget

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10817668/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Assessment of 21st-century planetary albedo changes in CMIP6 models

Aleksei Poliukhov[1,2,*], Alexander Lutsenko[3]

[1] Lomonosov Moscow State University, Moscow, 119991, Russia

[2] Moscow Center for Fundamental and Applied Mathematics, Leninskie Gory 1, Moscow, 119991, Russia

[3] Platov South-Russian State Polytechnic University (NPI), Novocherkassk, 346400, Russia

* Corresponding author: Aleksei Poliukhov, aeromsu@gmail.com, Lomonosov Moscow State University, Moscow, 119991, Russia

Abstract

Planetary albedo is a key parameter of the planet's energy balance, and its sustained decline contributes appreciably to present-day warming. Using CERES EBAF Ed4.2.1 satellite data and the ERA5 reanalysis for March 2000 – December 2014, we evaluate the representation of planetary albedo by 51 CMIP6 models. According to CERES, global albedo declined by $(1.20 \pm 0.61)\times10^{-3}$ per decade, of which $(0.85 \pm 0.48)\times10^{-3}$ occurred under clear-sky conditions; the ERA5 trend is statistically indistinguishable from zero. The models systematically overestimate albedo (by 1.3% on average) and underestimate the observed darkening: over 2000–2014 the difference remains within internal variability, but over the extended 2000–2025 record it becomes statistically significant ($-(1.05 \pm 0.15)$ versus $-(2.46 \pm 0.33)\times10^{-3}$ per decade). An objective selection procedure based on median-normalized climatological metrics and model genealogy identifies a nine-model ensemble that reduces the mean-albedo bias to +0.1% while preserving the trend magnitude. In the selected ensemble's projections, albedo decline continues under all ScenarioMIP scenarios: by 2091–2100, global albedo falls 0.0046–0.0116 below its 2015–2024 level, equivalent to an additional ~1.6–3.9 W m^{-2} of absorbed solar radiation; the largest changes occur at polar latitudes, and only under SSP1-2.6 does the decline level off. Over 2015–2025 the scenarios remain statistically indistinguishable in albedo, while the observed decline lies at the lower edge of the ensemble spread: only two of the 36 scenario members darken at least as fast as observed. A faithful representation of cloud and cryospheric processes remains the primary prerequisite for reliable projections of Earth's radiation budget.

Keywords: planetary albedo; CMIP6; CERES; model evaluation; model selection; Earth's energy budget

1. Introduction

Planetary albedo is the fraction of incoming solar radiation reflected by the Earth back to space (on average about 29%); it determines the energy input to the climate system. Even a change in planetary albedo of a fraction of a percent can substantially affect the planet's energy balance (Stephens et al. 2015). The magnitude and variability of albedo are governed primarily by cloud cover, which accounts for the largest part of the reflection, and by the surface—above all snow and ice. At the same time, the surface contribution to planetary albedo is attenuated by the atmosphere by roughly a factor of three (Donohoe and Battisti 2011).

A quarter century of continuous CERES satellite measurements of the radiation budget (Wielicki et al. 1996; Loeb et al. 2018) has revealed a sustained "darkening" of Earth: reflected shortwave radiation is decreasing while absorbed radiation is increasing, and it is the shortwave component that drives the doubling of Earth's energy imbalance since the early 2000s (Stephens et al. 2022; Loeb et al. 2024; Mauritsen et al. 2025). The albedo decline is corroborated by independent earthshine observations (Goode et al. 2021). The record warmth of 2023 is largely associated with an unprecedentedly low planetary albedo (Goessling et al. 2025); proposed mechanisms include the contraction of storm-cloud zones (Tselioudis et al. 2025), declining aerosol emissions (Quaas et al. 2022; Hodnebrog et al. 2024), and cloud feedbacks. The causes of Earth's darkening, and whether climate models reproduce it, have become the subject of active debate (Schmidt et al. 2023; Hansen et al. 2025; Zelinka et al. 2026), making a systematic assessment of albedo in models particularly timely.

CMIP6 models reproduce the planetary albedo climatology markedly better than previous model generations (Jian et al. 2020; Wild 2020; Bock et al. 2020), yet the inter-model spread in cloud properties remains the dominant source of uncertainty in climate sensitivity (Zelinka et al. 2020), and the ability of models to reproduce observed trends in reflected radiation has been tested only on small model subsets and global means (Loeb et al. 2020a) or through the lens of hemispheric albedo symmetry (Datseris and Stevens 2021; Crueger et al. 2023). The recent review by Li et al. (2025) synthesized the observed characteristics and mechanisms of albedo change and presented ensemble-mean projections of reflected radiation. At the same time, existing assessments of CMIP6 albedo -including the climatology-based model ranking of Jian et al. (2020) -do not address the observed trends, the genealogical dependence among models, or ensemble selection for scenario projections; to our knowledge, an objective genealogy-aware model selection for planetary albedo and spatially resolved scenario projections of albedo are presented here for the first time. A separate methodological issue is on what grounds the "best" models should be identified: the practice of ensemble weighting and sub-selection requires reproducible criteria, allowance for model interdependence, and out-of-sample testing (Knutti et al. 2017; Abramowitz et al. 2019; Brunner et al. 2020), while the

presence of models with excessive climate sensitivity in CMIP6 makes uncritical use of the full multi-model ensemble (MME) risky (Hausfather et al. 2022). For the short satellite era, the role of internal variability is also fundamental: 15-yr trends in individual realizations of coupled models need not be in phase with observations (Deser et al. 2012; Marotzke and Forster 2015), so trend agreement is an unreliable criterion of model quality.

The aims of this study are to systematically evaluate the representation of planetary albedo by an ensemble of 51 CMIP6 models against CERES EBAF and ERA5 over 2000–2014; to construct an objective, reproducible model selection procedure based on climatological albedo characteristics that accounts for genealogical dependence among models; to decompose the observed and modeled trends into clear-sky and cloud contributions; to analyze albedo projections of the selected ensemble under four ScenarioMIP scenarios through 2100; and to perform an independent evaluation of the ensemble against 2015–2025 observations.

2. Data and methods

2.1. CERES EBAF

The evaluation is based on the CERES EBAF Ed4.2.1 product, "TOA and Surface" edition (Clouds and the Earth's Radiant Energy System, Energy Balanced and Filled; Loeb et al. 2018): monthly shortwave fluxes at the top of the atmosphere (TOA) on a 1°×1° grid, derived from measurements by the CERES scanning radiometers aboard the Terra, Aqua, and NOAA-20 satellites (Wielicki et al. 1996; Smith et al. 2011). The calibration uncertainty of the shortwave channel is ~1% (with stability better than 0.2% per decade); monthly regional fluxes carry additional uncertainty associated with changes in the accompanying imagers (Su et al. 2020). We use the variables toa_sw_all_mon (reflected shortwave flux under all-sky conditions), toa_sw_clr_t_mon (clear-sky flux computed for the total region -a definition consistent with the model diagnostic rsutcs; Loeb et al. 2020b), and solar_mon (incoming solar radiation) for March 2000 through December 2014; the alternative clear-sky estimate toa_sw_clr_c_mon (based on the cloud-free portions of each region) is used for a sensitivity check; for the scenario evaluation, the same product is used for January 2015 through December 2025.

2.2. ERA5

ERA5 is the global reanalysis produced by ECMWF (Hersbach et al. 2020). TOA radiative fluxes in ERA5 are not assimilated but are computed by the radiation scheme of the forecast model and therefore inherit its errors in cloud cover and aerosol climatology; reanalyses are known not to reproduce the CERES-observed trends in TOA fluxes and to diverge systematically from observations in the energy budget (Loeb et al. 2022; Wild and Bosilovich 2024). For this reason, ERA5 is used here not as a reference but as an

111 independent point of comparison. We use the monthly variables tisr (incoming solar
112 radiation at TOA) and tsr (net shortwave flux at TOA).

113 **2.3. CMIP6 and ScenarioMIP**

114 We use monthly rsut, rsdt, and rsutcs fluxes (Amon) from 51 CMIP6 models (Eyring et
115 al. 2016) in the historical experiment, with one realization (r1i1p1f1) per model (the full
116 list is given in Table S1). In the rXiYpZfW notation, the indices denote the initial
117 conditions (realization), initialization method, physics variant, and forcing variant;
118 r1i1p1f1 is the first, reference member of each model's ensemble. One realization per
119 model is standard practice for comparison with observations: the observations themselves
120 constitute a single realization of the climate system, and the "one model, one vote"
121 principle prevents models with large ensembles from dominating; averaging many
122 realizations would instead damp internal variability and make the trend comparison with
123 observations inconsistent. Projections are analyzed for four ScenarioMIP scenarios
124 (O'Neill et al. 2016) -SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 -which combine the
125 SSP socioeconomic pathways (Riahi et al. 2017) with target radiative forcing levels by
126 2100.

127 **2.4. Definition of planetary albedo and processing**

128 Planetary albedo is defined as the ratio of the reflected to the incoming shortwave
129 radiation flux at TOA. The global mean albedo is computed as the ratio of area-weighted
130 global-mean fluxes:

131 $$\mathrm{A} = \sum w\, SW^{\uparrow} \Big/ \sum wSW^{\downarrow} \qquad (1)$$

132 where $w = \cos\varphi$ are the area weights. This "energy-based" definition corresponds to the
133 fraction of energy reflected by the planet and differs from the area average of the field of
134 local albedos (the latter is systematically higher: 0.318 versus 0.292 for CERES). For
135 CERES, A = toa_sw_all_mon / solar_mon; for ERA5, A = (tisr − tsr)/tisr; for the models,
136 A = rsut/rsdt; clear-sky albedo is computed analogously from toa_sw_clr_t_mon and
137 rsutcs. In grid-cell calculations, polar-night cells ($SW^{\downarrow} < 10$ W m^{-2}) are masked. Model
138 and ERA5 fields are bilinearly interpolated onto the 1°×1° CERES grid. Seasons are
139 defined climatologically as DJF, MAM, JJA, and SON (for global means the calendar
140 seasons of the two hemispheres are opposite, so the international convention is adopted).

141 **2.5. Statistical analysis**

142 Trends in the monthly global series are estimated by ordinary least squares after removal
143 of the mean seasonal cycle; significance and 95% confidence intervals are computed
144 accounting for autocorrelation of the residuals through the effective sample size

145 $$n_{eff} = n(1 - r_1)/(1 + r_1) \qquad (2)$$

where r_1 is the lag-1 autocorrelation (equivalent to a t-test with a reduced sample; Santer et al. 2000). Because monthly anomalies can carry memory beyond one lag (ENSO), the confidence intervals are additionally cross-checked with a moving-block bootstrap (24-month blocks, 2000 replicates): the resulting intervals are 20–35% wider than the AR(1) estimates (±0.73 versus ±0.61 for the CERES trend over 2000–2014; ±0.45 versus ±0.33 over 2000–2025), yet all conclusions about trend significance hold. The robustness of the trends is verified with the nonparametric Theil–Sen estimator and the Mann–Kendall test applied to the annual series. Grid-cell trends are computed from annual means; significance on the maps is displayed with false discovery rate (FDR) control (α_FDR = 0.10) following Wilks (2016) -without this correction, stippling systematically overstates the area of significant change. Spatial similarity of fields is characterized by area-weighted centered correlations and normalized standard deviations (Taylor diagrams; Taylor 2001); correlations are reported with their sign.

2.6. Model selection procedure

Models for the scenario analysis were selected in four objective steps based entirely on the climatological characteristics of 2000–2014; trends were deliberately excluded from the selection criteria, because the 15-yr trend of a single realization of a coupled model is largely governed by the phase of internal variability, and its agreement with observations does not characterize the model physics (Deser et al. 2012; Marotzke and Forster 2015; Abramowitz et al. 2019).

Eligibility: a complete historical record for March 2000 through December 2014 and availability of rsut/rsdt for all four ScenarioMIP scenarios. Notably, this step excludes the two models with the lowest values of the D_i score introduced below: GFDL-CM4 (D_i = 0.82; only SSP2-4.5 and SSP5-8.5 are available on ESGF) and SAM0-UNICON (0.85; no ScenarioMIP simulations); the GFDL family is represented in SEL9 by GFDL-ESM4, which shares the AM4 atmosphere (D_i = 0.86).

Performance: for each model, three area-weighted errors relative to CERES were computed -the RMSE of the annual-mean albedo field (E_1), the RMSE of the climatological seasonal cycle (12 monthly anomaly fields relative to the annual mean, E_2), and the RMSE of the field of interannual standard deviation of annual means (E_3). Each error was normalized by the ensemble median, and the aggregate score D_i is the mean of the three normalized errors (median normalization follows the performance metrics of Gleckler et al. (2008); the aggregate score is analogous to the weighting schemes of Knutti et al. (2017)). Models with $D_i < 1$, i.e., better than the median model, were retained (22 models without step (i), 15 with it; six of these are subsequently displaced by a better-performing member of their family).

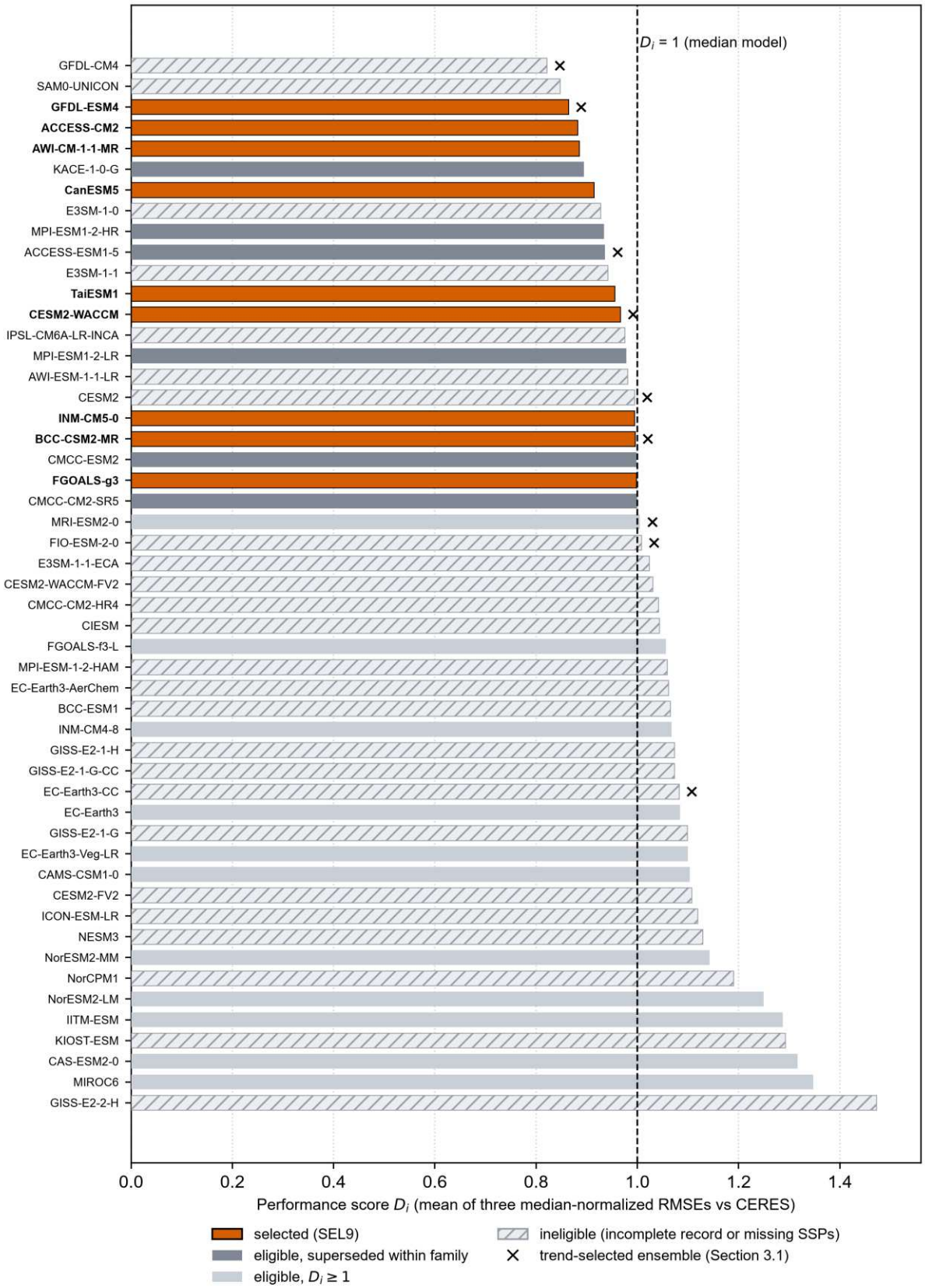
Independence: the models were grouped into model families sharing a common atmospheric component (CAM5, CAM6, ECHAM6, MetUM, IFS/EC-Earth, GISS,

GFDL, INM, and others; cf. Sanderson et al. 2015; Brunner et al. 2020), and one representative with the best D_i was retained from each family. The matrix of inter-model distances (the same metric, model against model) independently confirms this grouping: all nearest pairs in the ensemble are family relatives (Table S2).

Sensitivity: both the threshold and the metric weights were tested (Table S3). Tightening the threshold to $D_i < 0.9$ leaves only three models -too few for projections, while their trend (-0.41×10^{-3} per decade) and bias (-0.25%) remain consistent with the qualitative conclusions; relaxing it to $D_i < 1.1$ adds three models (MRI-ESM2-0, FGOALS-f3-L, EC-Earth3) with a virtually unchanged trend (-0.65×10^{-3} per decade) and bias ($+0.01\%$). Replacing the equal weights of the three metrics with pairwise combinations (E_1+E_2, E_1+E_3, E_2+E_3) yields ensembles of 7–9 models with trends from -0.40 to -0.96×10^{-3} per decade and biases from -0.60 to $+0.11\%$, and the E_1+E_3 variant reproduces the SEL9 composition exactly. We note that E_3 is the noisiest of the three metrics (the sampling error of the interannual standard deviation over 15 years is large), yet excluding it (the E_1+E_2 variant) yields an ensemble in which six of the seven members belong to SEL9. All variants preserve the sign of the trend, its underestimation relative to CERES, and the severalfold reduction of the bias relative to the MME; the median used for normalizing the errors is in all cases taken over the full set of 51 models.

The resulting selected nine-model ensemble (hereafter SEL9) comprises GFDL-ESM4 ($D_i = 0.86$), ACCESS-CM2 (0.88), AWI-CM-1-1-MR (0.89), CanESM5 (0.91), TaiESM1 (0.96), CESM2-WACCM (0.97), INM-CM5-0 (0.99), BCC-CSM2-MR (0.996), and FGOALS-g3 (0.998). The full ranking of all 51 models is given in Table S1.

The ranking of all 51 models by D_i, together with eligibility and family membership, is shown in Fig. 1: the nine best eligible representatives of their families constitute SEL9; crosses mark the alternative trend-selected ensemble (the nine models with the highest trend-field correlations; Section 3.1).


D_i = 1 (median model)
GFDL-CM4
SAM0-UNICON
GFDL-ESM4
ACCESS-CM2
AWI-CM-1-1-MR
KACE-1-0-G
CanESM5
E3SM-1-0
MPI-ESM1-2-HR
ACCESS-ESM1-5
E3SM-1-1
TaiESM1
CESM2-WACCM
IPSL-CM6A-LR-INCA
MPI-ESM1-2-LR
AWI-ESM-1-1-LR
CESM2
INM-CM5-0
BCC-CSM2-MR
CMCC-ESM2
FGOALS-g3
CMCC-CM2-SR5
MRI-ESM2-0
FIO-ESM-2-0
E3SM-1-1-ECA
CESM2-WACCM-FV2
CMCC-CM2-HR4
CIESM
FGOALS-f3-L
MPI-ESM-1-2-HAM
EC-Earth3-AerChem
BCC-ESM1
INM-CM4-8
GISS-E2-1-H
GISS-E2-1-G-CC
EC-Earth3-CC
EC-Earth3
GISS-E2-1-G
EC-Earth3-Veg-LR
CAMS-CSM1-0
CESM2-FV2
ICON-ESM-LR
NESM3
NorESM2-MM
NorCPM1
NorESM2-LM
IITM-ESM
KIOST-ESM
CAS-ESM2-0
MIROC6
GISS-E2-2-H
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Performance score D_i (mean of three median-normalized RMSEs vs CERES)
selected (SEL9)
eligible, superseded within family
eligible, $D_i \geq 1$
ineligible (incomplete record or missing SSPs)
trend-selected ensemble (Section 3.1)

210

Fig. 1. Ranking of the 51 CMIP6 models by the performance score D_i (mean of three median-normalized area-weighted errors relative to CERES). Orange bars denote the selected SEL9 ensemble; gray bars denote eligible models outperformed by a member of their own family; hatched bars denote ineligible models (incomplete historical record or missing r1i1p1f1 output for at least one of the four scenarios); crosses mark the alternative trend-selected ensemble (Section 3.1).

3. Results and discussion

3.1. Global mean albedo and its trend, 2000–2014

According to CERES, the global mean albedo over March 2000 – December 2014 was 0.292, with a robust decline of $(1.20 \pm 0.61)\times10^{-3}$ per decade ($p < 0.001$ accounting for lag-1 autocorrelation; Theil–Sen estimate -1.0×10^{-3} per decade, $p_MK < 0.01$). Importantly, a significant decline is also found in the clear-sky albedo $-(0.85 \pm 0.48)\times10^{-3}$ per decade (the clr_c estimate, based only on cloud-free areas, gives a nearly identical value: (0.90 ± 0.49)): a substantial part of the observed planetary darkening over this period is associated not with a reorganization of cloud cover but with changes in the cloud-free atmosphere and at the surface -the declining aerosol load (Quaas et al. 2022; Hodnebrog et al. 2024) and the retreating cryosphere (Flanner et al. 2011; Pistone et al. 2014). The cloud contribution to the trend (all-sky minus clear-sky) amounts to -0.35×10^{-3} per decade.

The observed trend should also be weighed against the instrumental uncertainty budget. The stated calibration stability of the CERES shortwave channel (better than 0.2% per decade; Loeb et al. 2018) is equivalent to $\sim0.6\times10^{-3}$ albedo per decade -half of the 2000–2014 trend -so the 15-yr trend taken alone requires cautious interpretation. However, the sign and magnitude of the observed darkening are corroborated independently: by the growth of Earth's energy imbalance inferred from ocean heat content (Loeb et al. 2022, 2024), by earthshine observations (Goode et al. 2021), and by the consistent decline of the aerosol burden (Quaas et al. 2022). Over the extended 2000–2025 record, the trend (-2.46×10^{-3} per decade) exceeds the possible drift fourfold: even subtracting the full drift changes neither the sign nor the significance of the conclusion.

ERA5 underestimates the mean albedo by 1.6% relative to CERES, and its trend, $(+0.42 \pm 0.66)\times10^{-3}$ per decade, is statistically indistinguishable from zero. The opposite signs of the ERA5 and CERES trends are consistent with the documented inability of reanalyses to reproduce the observed trends in the radiation budget, owing to inhomogeneities in the assimilated observations and the model origin of the TOA fluxes (Loeb et al. 2022; Wild and Bosilovich 2024). At the same time, in the spatial structure of the trends, ERA5 is substantially closer to CERES than any individual model (see Section 3.2): assimilation of the actual synoptic evolution constrains the geography of the changes, but not their global balance.

The full CMIP6 ensemble systematically overestimates the albedo: the MME mean bias is +1.3% (individual models range from −3.2 to +10.5%; Table S1), extending earlier assessments of the CMIP6 radiation budget (Wild 2020; Jian et al. 2020). The ensemble-mean trend, $-(0.72 \pm 0.22)\times10^{-3}$ per decade, has the correct sign; the central estimate amounts to ~60% of the observed value, although over a 15-yr window the difference is formally indistinguishable from internal variability (the CERES trend corresponds to the 35th percentile of the distribution of individual-model trends (Fig. 2), and the MME estimate lies within the 95% CI of the observations); the selected nine-model ensemble (SEL9) yields essentially the same magnitude ($-(0.68 \pm 0.31)\times10^{-3}$). The underestimation becomes statistically significant over the extended 2000–2025 record (Section 3.3); it is consistent with reported model underestimates of the trends in Earth's energy imbalance (Schmidt et al. 2023; Mauritsen et al. 2025) and holds for any ensemble composition, pointing to a common, rather than model-specific, deficiency in process representation. Over the longer 1940–2014 record (Fig. 3), the modelled global albedo varies within a narrow range (MME annual means 0.294–0.304); the 15-yr running trends of the MME reach $+2.9\times10^{-3}$ per decade for windows centred on the late 1950s, the period of rapid growth of anthropogenic aerosol, and -4.3×10^{-3} per decade for windows centred on the late 1990s, the recovery from the Pinatubo spike, so the modelled 2000–2014 decline (-0.7×10^{-3} per decade) is moderate against this background.

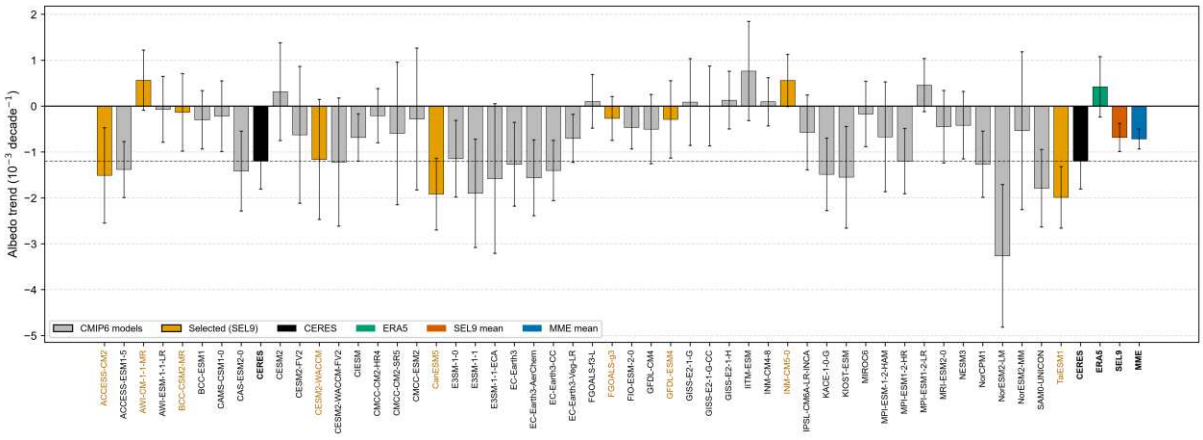


Fig. 2. Global albedo trends for 2000–2014 (10^{-3} per decade) for the 51 CMIP6 models, SEL9, MME, CERES, and ERA5; error bars denote 95% confidence intervals accounting for lag-1 autocorrelation. CERES is shown twice for reference: in the common row and in the right-hand group.

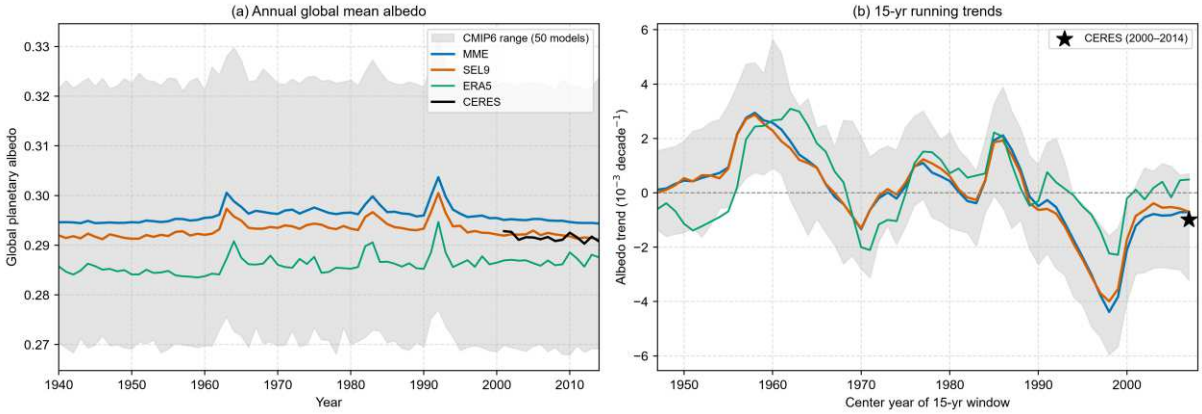


Fig. 3. (a) Annual global mean albedo for 1940–2014: the 50 CMIP6 models with complete 1940–2014 records (range; GISS-E2-2-H excluded), MME, SEL9, ERA5 (from 1940), and CERES (from 2000); (b) 15-yr running trends of the same series. Albedo spikes are associated with the El Chichón (1982) and Pinatubo (1991) eruptions (Robock 2000; Minnis et al. 1993); the post-war rise and subsequent decline of the trend reflect the evolution of anthropogenic aerosol (global dimming and brightening; Wild 2009).

The main effect of the model selection appears in the mean state: the SEL9 bias is only +0.11%, versus +1.31% for the MME -an order of magnitude smaller (Table 1). This improvement is partly built into the selection by construction: the E_1 metric directly penalizes the mean-field error relative to CERES over the same period, so the estimate is in-sample. An independent check is provided by the period outside the training window: over 2015–2025 (historical+SSP2-4.5), the SEL9 bias relative to CERES is +0.9%, and the increase from +0.1% almost entirely reflects the accumulated underestimation of the observed trend (Section 3.3) rather than a degradation of the ensemble climatology. Seasonal SEL9 biases do not exceed 1.4% in magnitude (DJF −0.61%, MAM +0.25%, JJA +1.40%, SON −0.59%); ERA5 biases are negative in all seasons (−0.8…−2.6%). Importantly, selection does not improve the global trend: SEL9 reproduces ~57% and the MME ~60% of the observed magnitude. Moreover, if models are deliberately selected by trend agreement -taking the nine models with the highest trend-field correlations with CERES -the mean-trend correlation of such an ensemble reaches 0.37, versus 0.26 for SEL9, with an almost identical mean-state bias (+0.12% versus +0.11%). This advantage is illusory, however: agreement of 15-yr trends of individual realizations with observations reflects a chance coincidence of internal-variability phases rather than model quality (Deser et al. 2012; Marotzke and Forster 2015); in addition, the trend-selected ensemble is genealogically degenerate: it contains two pairs of direct relatives -CESM2 and CESM2-WACCM, GFDL-CM4 and GFDL-ESM4 -as well as FIO-ESM-2-0 from the CAM5 lineage, and four of its models lack the full set of scenarios and are unusable for projections (Fig. 1, crosses). This result provides a quantitative argument

against selecting models by short-term trends, complementing the methodological cautions of Abramowitz et al. (2019).

Table 1. Biases of the global mean albedo relative to CERES (2000–2014), % of the CERES value: SEL9, MME, and ERA5 by season and for the annual mean.

Dataset	DJF	MAM	JJA	SON	Annual
SEL9	-0.0018 (-0.61%)	+0.0007 (+0.25%)	+0.0040 (+1.40%)	-0.0017 (-0.59%)	+0.0003 (+0.11%)
MME	+0.0003 (+0.08%)	+0.0058 (+2.02%)	+0.0078 (+2.75%)	+0.0012 (+0.41%)	+0.0038 (+1.31%)
ERA5	-0.0077 (-2.57%)	-0.0046 (-1.58%)	-0.0022 (-0.77%)	-0.0046 (-1.59%)	-0.0048 (-1.63%)

3.2. Spatial structure of albedo and its trends, 2000–2014

The spatial structure of the mean albedo (Fig. 4, left column) is reproduced by both ERA5 and the ensemble with correlations above 0.98 (Fig. 5a; Table 2); this level of agreement, however, largely reflects the latitudinal insolation gradient and the pole-to-equator contrast and is therefore a weak discriminator of model quality ($r \geq 0.87$ for the annual fields of all 51 models, Fig. 5a, and $r \geq 0.96$ for the nine selected models, Table 2). The trend fields are more informative (Fig. 4, right column; Fig. 5b): CERES exhibits a significant (FDR-controlled) albedo decrease over regions of sea-ice loss in the Arctic and Antarctic, together with cloud-driven changes of both signs over the oceans; ERA5 reproduces this geography with a correlation of 0.67, whereas SEL9 reaches only 0.26, and the ensemble-mean trend field is damped in amplitude by a factor of two (normalized standard deviation 0.50) because of the averaging of the unsynchronized interannual variability of nine realizations.

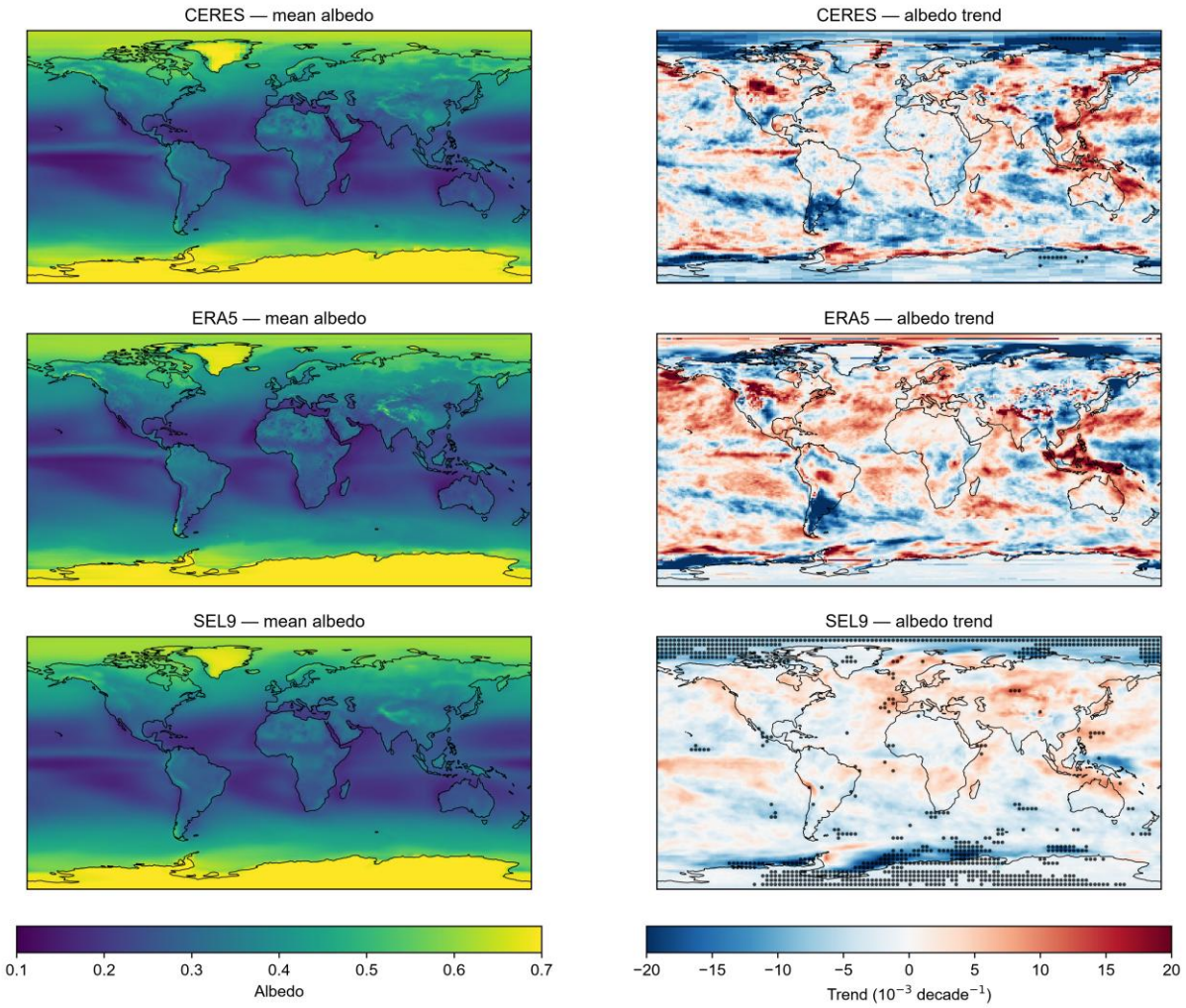
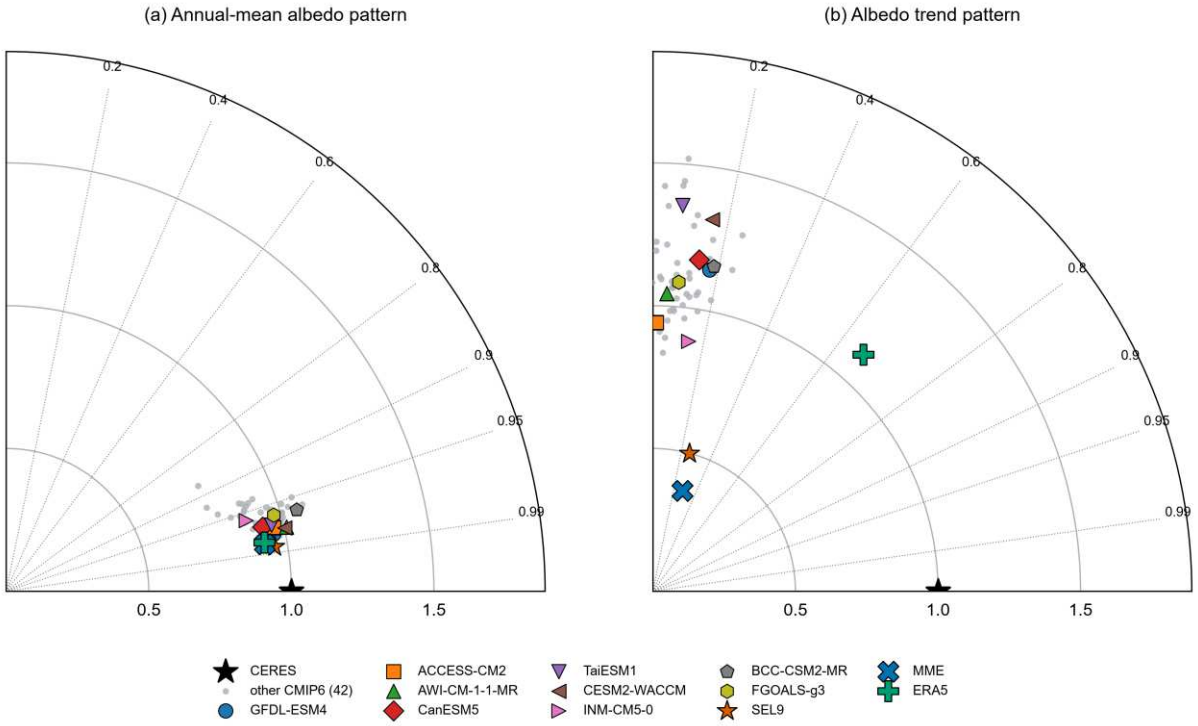


Fig. 4. Mean albedo (left) and albedo trends from annual means (right), 2000–2014: CERES, ERA5, SEL9. Stippling marks grid cells with significant trends (FDR control, α_FDR = 0.10; Wilks 2016); the test is applied to each of the three trend panels, for SEL9 to the trend of the ensemble mean. The test is passed by 0.4% of the area for CERES (sea-ice margins), by 4.8% for SEL9 and by no grid cell for ERA5: at the grid-cell scale, 15-year trends are almost everywhere indistinguishable from interannual variability, and the larger significant area for SEL9 reflects the suppression of internal variability by averaging nine realizations rather than a stronger signal.



328

329 *Fig. 5. Taylor diagrams: (a) annual-mean albedo fields, (b) trend fields; the nine SEL9*
330 *models (colored symbols), the remaining 42 CMIP6 models (gray dots), SEL9, MME, and*
331 *ERA5 relative to CERES.*

332 *Table 2. Area-weighted spatial correlations (with sign) of the annual and seasonal fields*
333 *of mean albedo and its trends with CERES: the nine SEL9 models, the SEL9 and MME*
334 *means, and ERA5. High mean-field correlations are expected -they are set by the pole-to-*
335 *equator contrast (annual fields yield $r \geq 0.96$, seasonal fields $r \geq 0.92$); it is the trend*
336 *fields that discriminate among models.*

Model	**Annual**	**DJF**	**MAM**	**JJA**	**SON**
GFDL-ESM4	0.978 / 0.174	0.950 / 0.180	0.962 / 0.000	0.958 / 0.113	0.973 / 0.063
ACCESS-CM2	0.973 / 0.013	0.948 / 0.087	0.960 / -0.032	0.950 / 0.033	0.967 / 0.016
AWI-CM-1-1-MR	0.975 / 0.046	0.956 / 0.053	0.964 / 0.030	0.948 / -0.064	0.972 / 0.065
CanESM5	0.970 / 0.138	0.949 / -0.004	0.964 / 0.141	0.937 / 0.093	0.962 / 0.108
TaiESM1	0.970 / 0.077	0.951 / -0.033	0.956 / 0.092	0.951 / -0.044	0.960 / 0.138
CESM2-WACCM	0.975 / 0.158	0.960 / 0.027	0.967 / -0.051	0.945 / 0.092	0.968 / 0.114
INM-CM5-0	0.959 / 0.142	0.943 / 0.071	0.950 / 0.038	0.919 / -0.020	0.949 / 0.027
BCC-CSM2-MR	0.963 / 0.185	0.946 / 0.051	0.947 / 0.015	0.927 / 0.160	0.955 / 0.108
FGOALS-g3	0.962 / 0.083	0.949 / 0.013	0.951 / 0.006	0.917 / 0.036	0.948 / -0.001
ERA5	0.983 / 0.665	0.963 / 0.686	0.980 / 0.716	0.972 / 0.698	0.983 / 0.708
SEL9	0.986 / 0.257	0.977 / 0.116	0.979 / 0.086	0.969 / 0.132	0.982 / 0.197
MME	0.984 / 0.281	0.975 / 0.132	0.976 / 0.048	0.968 / 0.114	0.980 / 0.357

Difference maps of the mean fields (Fig. 6) show that both ERA5 and SEL9 overestimate the albedo over Greenland, the Antarctic interior, and the Himalayas (over the shelf seas and West Antarctica the difference changes sign) and underestimate it over regions of marine stratocumulus; the most extensive ensemble error is a band of underestimation over the Southern Ocean (50–70°S), the well-known cloud-related shortwave bias of climate models (Hyder et al. 2018). These are known systematic errors in the representation of the cryosphere and the boundary layer (Stephens et al. 2012; Jian et al. 2020). The SEL9 inter-model spread (Fig. 7) is largest in the subpolar regions and over the Tibetan Plateau -the regions of snow and ice albedo -consistent with the snow–albedo feedback spread persisting since CMIP3 (Qu and Hall 2014; Thackeray et al. 2018) and indicating exactly where climatology-based selection does not remove structural uncertainty.

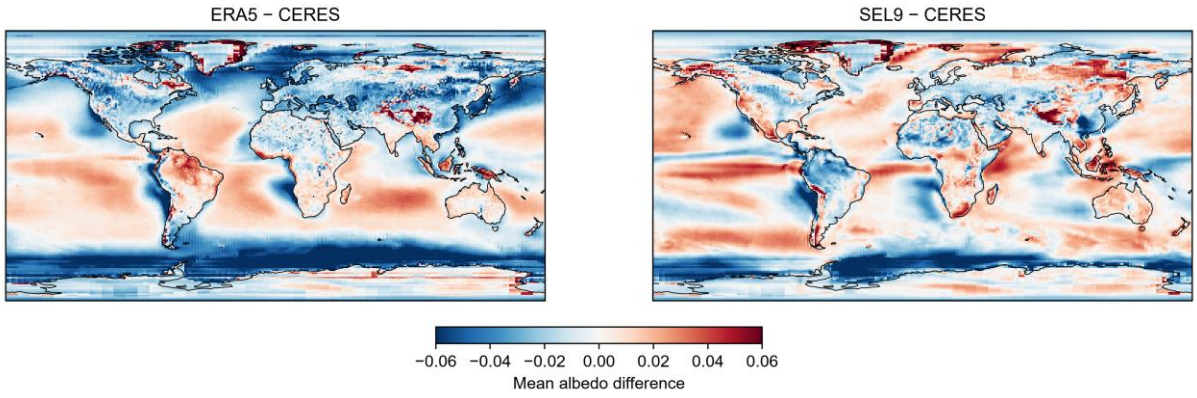


Fig. 6. Differences in mean albedo, ERA5 − CERES and SEL9 − CERES 2000–2014.

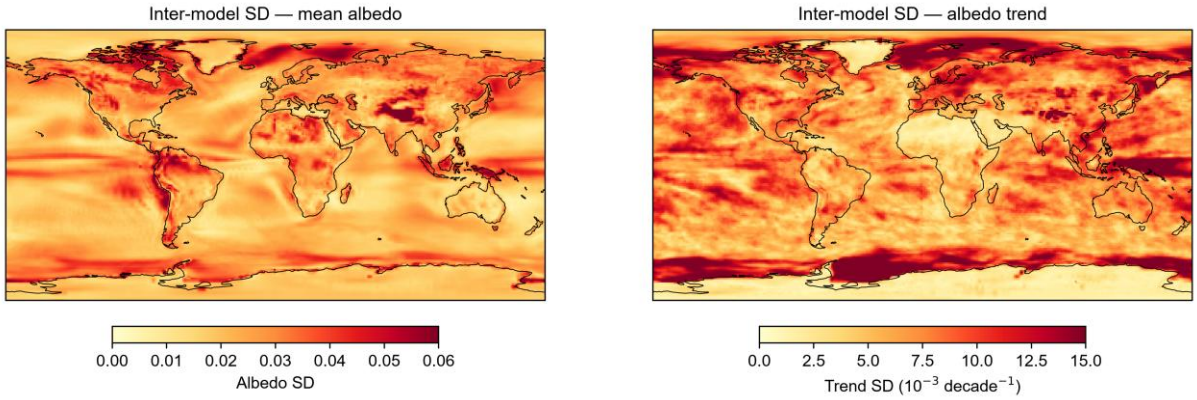


Fig. 7. Inter-model standard deviation of SEL9: mean albedo and trends, 2000–2014.

Decomposing the trends into clear-sky and cloud contributions (Fig. 8) shows that a substantial part of the observed 2000–2014 darkening occurs under clear-sky conditions. In CERES, the clear-sky trend ($(0.85 \pm 0.48)\times10^{-3}$ per decade) is concentrated at polar latitudes: over the Arctic (70–90°N) it reaches -19×10^{-3} per decade and exceeds the all-

sky trend (-12×10^{-3}) in magnitude -cloud cover partly masks the surface signal of retreating ice and snow. Over East Asia, clear-sky albedo was still increasing during this period ($+2.1\times10^{-3}$ per decade): the regional aerosol burden peaked only in the early 2010s, and its reduction largely postdates the period considered here.

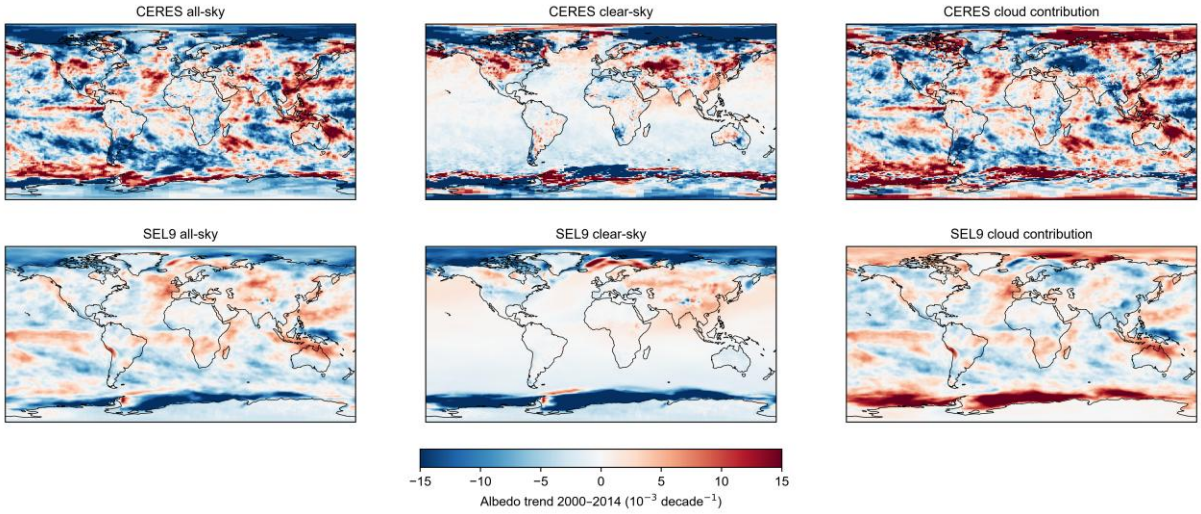


Fig. 8. Albedo trends for 2000–2014: all-sky, clear-sky, and the cloud contribution (all-sky minus clear-sky) from CERES (top) and the SEL9 ensemble (bottom).

The SEL9 ensemble reproduces the cloud contribution to the trend closely (-0.27×10^{-3} versus -0.35×10^{-3} per decade in CERES) but captures less than half of the clear-sky trend (-0.41 ± 0.58 versus -0.85 ± 0.48; the ensemble-mean estimate is insignificant, and in three of the nine models the clear-sky trend is even positive). The models' underestimation of the observed darkening is thus concentrated in the clear sky -in cryospheric and aerosol mechanisms -whereas the cloud-driven part of the change is reproduced on average; this echoes the finding that models systematically underestimate the observed trend in Earth's energy imbalance (Mauritsen et al. 2025). These conclusions do not depend on the choice of the CERES clear-sky estimate: the global clear-sky trends from clr_t (total-region, model-consistent) and clr_c (cloud-free areas) differ by less than 0.06×10^{-3} per decade -far less than their confidence intervals.

3.3. Early out-of-sample evaluation, 2015–2025

The 2015–2025 period, which was not used in the model selection, serves as an independent evaluation of the ensemble. The extension of the SEL9 series with the SSP2-4.5 experiment (the scenarios differ negligibly over 2015–2025) is shown in Fig. 9. Over 2000–2025, CERES albedo declined by $(2.46 \pm 0.33)\times10^{-3}$ per decade -twice as fast as over 2000–2014 -with the absolute minimum in 2023 (Goessling et al. 2025); the ensemble yields $(1.05 \pm 0.15)\times10^{-3}$, i.e., only ~43% of the observed trend is reproduced over the extended record, and after 2015 the observations shift toward the lower edge of

the ensemble spread. Over the full 26-yr record the darkening is also spatially significant (Fig. 9b–d): the FDR test is passed by 15% of the area for CERES (94% of it with negative trends, mostly over the mid-latitude oceans of both hemispheres and the polar sea-ice zones; 75% of the significant area lies poleward of 30°) and by 14% for ERA5 (74% negative), whereas over the 15-yr window of Fig. 4 almost no grid cell passes the same test. For the SEL9 ensemble mean the significant area is larger (29%) despite trend amplitudes about half as large, because averaging nine realizations suppresses internal variability; the significance of an ensemble-mean trend therefore characterizes the consistency of the forced response across models rather than its magnitude and is not directly comparable to that of a single observational record.

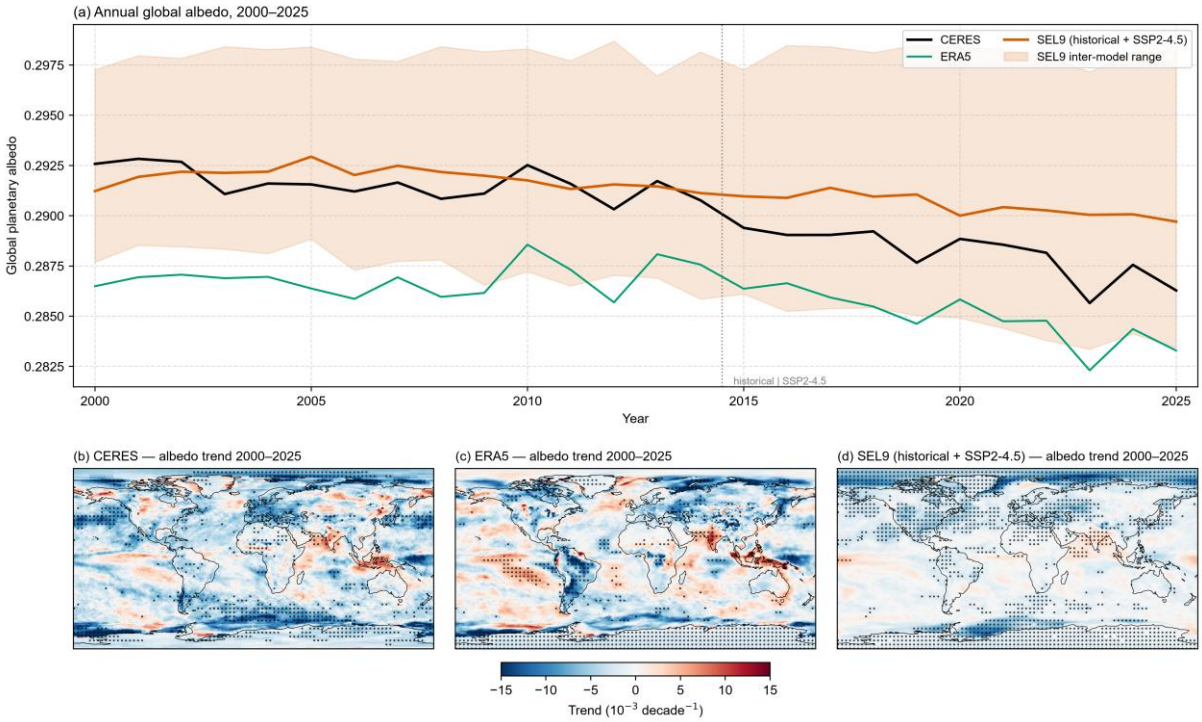


Fig. 9. (a) Annual global albedo for 2000–2025: CERES, ERA5, and SEL9 (historical through 2014, SSP2-4.5 thereafter; shading – inter-model range). (b–d) Albedo trends from annual means over 2000–2025 for CERES, ERA5 and the SEL9 ensemble mean; stippling marks grid cells with significant trends (FDR control, α_FDR = 0.10; Wilks 2016), covering 15% of the area for CERES (94% of it with negative trends), 14% for ERA5 and 29% for the SEL9 ensemble mean, where averaging nine realizations suppresses internal variability.

Comparison of the scenarios with the 2015–2025 observations (Table 3) confirms that the first decade of the projections cannot be used to "identify the pathway": the mean-field correlations are identical across scenarios (0.98), the trend-field correlations are small (0.03–0.15), and the inter-model ranges of the global trends overlap widely -from [−5.2;

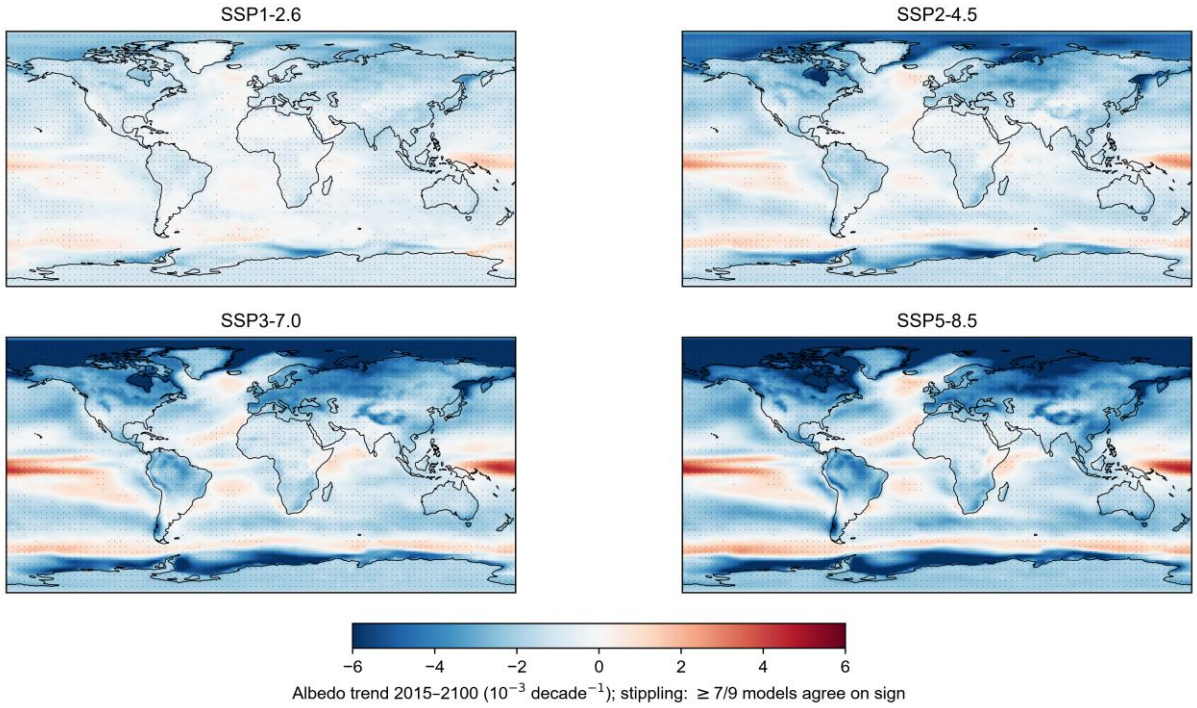
−0.4]×10⁻³ per decade under SSP1-2.6 to [−2.4; +0.8] under SSP3-7.0 (Table 3); the ensemble means already differ, but the pathway cannot be identified from a single realization of the observations. The observed decline ((2.97 ± 1.32)×10⁻³ per decade) is faster than all scenario means (−0.52 to −1.65×10⁻³), and only two of the 36 scenario members (both SSP1-2.6) darken at least as fast as observed; notably, the strongest near-term decline is produced by SSP1-2.6 -the scenario with the fastest reduction of aerosol emissions -while the reduction in shipping sulfur emissions since 2020, absent from all scenarios, further amplified the observed darkening (Yuan et al. 2024). The ERA5 trend field, by contrast, correlates with CERES more strongly than the models do (0.80), and its global trend over 2015–2025 (−(3.30 ± 1.17)×10⁻³ per decade) is statistically indistinguishable from the observed one (Table 3) -as expected for a system that assimilates the observed surface and atmospheric state. This does not overturn the conclusion of Section 3.1, however: over 2000–2014 the ERA5 trend had the opposite sign, so the agreement of a reanalysis with observed TOA-flux trends is not guaranteed and must be verified for each specific period.

Table 3. Comparison of the scenarios with CERES over 2015–2025: correlations of mean and trend fields, global trends of the ensemble means with 95% confidence intervals, and inter-model ranges across the nine members.

Dataset	r (mean field)	r (trend field)	Trend 2015–2025 (10^{-3} dec^{-1})	95% CI (±)	Member range (10^{-3} dec^{-1})
CERES	1.0	1.0	-2.97	1.32	—
ERA5	0.977	0.802	-3.3	1.17	—
SSP1-2.6	0.984	0.034	-1.65	0.38	[−5.24; −0.35]
SSP2-4.5	0.985	0.087	-1.41	0.49	[−2.87; −0.09]
SSP3-7.0	0.984	0.145	-0.52	0.47	[−2.38; +0.78]
SSP5-8.5	0.984	0.136	-1.38	0.48	[−2.78; +1.01]

3.4. Projected albedo changes, 2015–2100

Maps of SEL9 albedo trends for 2015–2100 (Fig. 10) show a monotonic amplification of the changes from SSP1-2.6 to SSP5-8.5 with a common spatial pattern. The strongest decline occurs at polar latitudes of both hemispheres -the zone of retreating sea ice and snow cover, where at least seven of the nine models agree on the sign of the trend; a robust decline also covers much of the mid-latitude oceans. All scenarios exhibit a band of increasing albedo over the equatorial Pacific, consistent with the expected cloud response to the El Niño-like pattern of surface ocean warming (for the role of the warming pattern in radiation budget changes, see Loeb et al. 2020a); under strong forcing, areas of weak increase appear in the subtropics. The ensemble shows no persistent albedo increase over South Asia: trends there are weak and lack sign agreement, which is expected given the non-monotonic aerosol emission trajectories of the scenarios (Lund et al. 2019) -continued emission growth is plausible only under SSP3-7.0.



437

438 *Fig. 10. SEL9 albedo trends for 2015–2100 under the four ScenarioMIP scenarios;*
439 *stippling marks areas where at least 7 of the 9 models agree on the sign of the trend; the*
440 *criterion reflects sign agreement rather than amplitude and, under SSP1-2.6, also covers*
441 *areas of weak trends.*

442 Ensemble-mean global trends over 2015–2100 are −0.58 [−1.47 to −0.08], −0.88, −1.18,
443 and −1.51 [−3.19 to +0.11] $\times 10^{-3}$ per decade for SSP1-2.6, SSP2-4.5, SSP3-7.0, and
444 SSP5-8.5, respectively (inter-model ranges in brackets). By 2091–2100, global albedo
445 falls below its 2015–2024 level by 0.0046 (SSP1-2.6) to 0.0116 (SSP5-8.5), i.e., by 1.6–
446 4.0% in relative terms; with a mean incoming solar flux of ~340 W m^{-2}, this is equivalent
447 to an additional ~1.6–3.9 W m^{-2} of absorbed solar radiation.

448 The 15-yr running trends (Fig. 11) show that the inter-model ranges of the scenarios
449 overlap for essentially the entire century -only the ensemble means are clearly
450 distinguishable: the decadally smoothed means of SSP1-2.6 and SSP2-4.5 separate
451 robustly after ~2055, SSP2-4.5 and SSP3-7.0 after ~2060, and only the difference
452 between SSP3-7.0 and SSP5-8.5 keeps its sign from the beginning of the century. Only
453 under SSP1-2.6 does the decline level off: after ~2065 the ensemble 15-yr trends oscillate
454 around zero (mostly below 0.2×10^{-3} per decade in magnitude, with episodic
455 strengthening to -0.6×10^{-3} in the late 2070s); under SSP2-4.5 the trend weakens to
456 roughly -0.7×10^{-3} by the end of the century, under SSP3-7.0 it persists near -1.2×10^{-3},
457 and under SSP5-8.5 it steepens to -2.1×10^{-3} per decade.

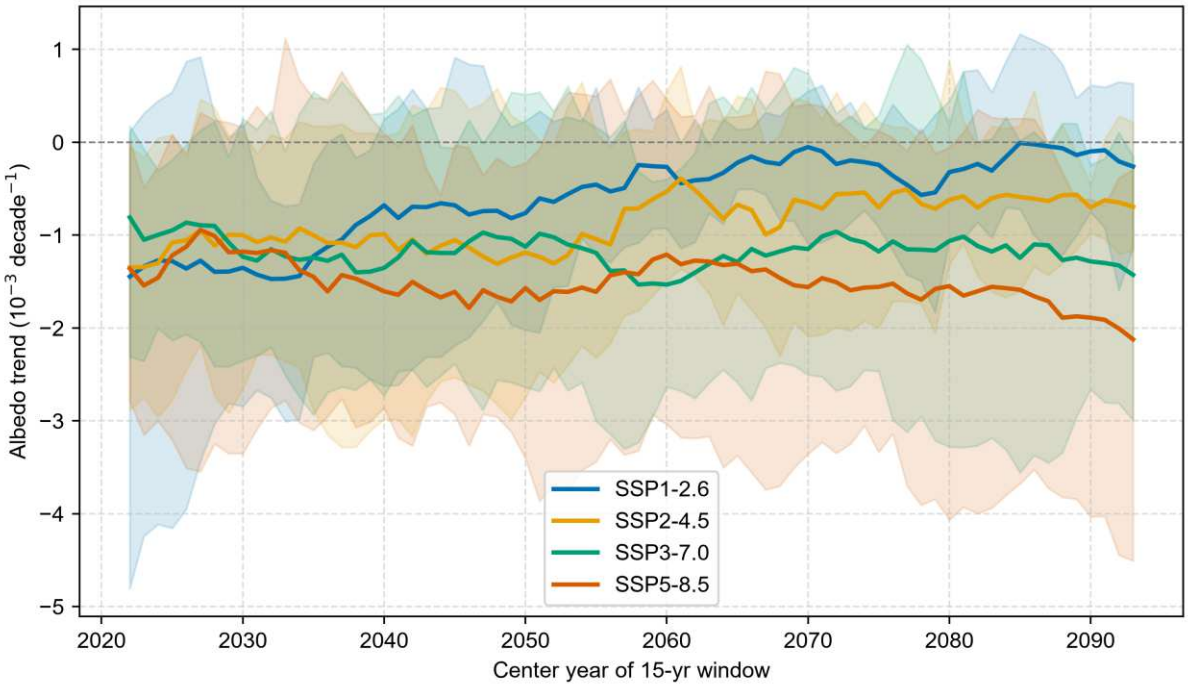


Fig. 11. 15-yr running trends of SEL9 global albedo through 2100 by scenario (lines - ensemble means, shading -inter-model range).

3.5. Limitations

Using a single realization per model does not allow the forced response to be separated from internal variability at the level of individual models; this is partly mitigated by the fact that the selection criteria are deliberately restricted to climatological characteristics, with trends used for evaluation only. Genealogical thinning reduces, but does not eliminate, ensemble interdependence (Sanderson et al. 2015). The evaluation is confined to the period of stable CERES operation since March 2000; the extension of the comparison beyond 2014 relies on composite historical+SSP2-4.5 series and inherits the differences in CMIP6 forcings after 2014. Finally, the ensemble includes models with overestimated equilibrium sensitivity (CanESM5, TaiESM1; Sherwood et al. 2020; Hausfather et al. 2022), which should be borne in mind when interpreting the absolute magnitudes of future changes.

Several methodological caveats should also be stated. First, the calibration stability of CERES (~0.2% per decade, i.e., ~0.6×10^{-3} albedo per decade) is comparable to half of the observed 2000–2014 trend, and only the extension of the record to 2025 makes the conclusion that the models underestimate the darkening robust to a possible drift (Section 3.1). Second, the threshold $D_i < 1$, the equal weights of the three metrics, and the median normalization are a priori choices; sensitivity to the threshold and the weights has been tested (Section 2.6, Table S3), but a different combination of choices could yield a somewhat different ensemble composition with the same qualitative conclusions. Third,

the reduction of the mean-state bias is an in-sample result by construction; its out-of-sample check over 2015–2025 gives +0.9% (Section 3.1). Finally, the clear-sky decomposition relies on the total-region CERES estimate (clr_t), consistent with the model diagnostic rsutcs; the sensitivity to the alternative clr_c estimate is negligible (Section 3.2).

Sensitivity to the comparison domain was tested by excluding latitudes poleward of 60°: the observed trend changes by less than 10%, the trend-field correlations by no more than 0.02, and the fraction of the trend reproduced by the ensemble even decreases (~44% versus ~57% under the same definition of the mean), i.e., the models' underestimation of the darkening is not an artifact of CERES polar data. The composition of the selected ensemble partly changes (five of the nine members), as expected: the criteria are optimized for the global (planetary) albedo, and for regional applications the selection should be repeated for the target domain. Furthermore, an ESGF catalog check confirms that requiring all four scenarios specifically under r1i1p1f1 does not affect the ensemble composition: only CESM2 has complete scenario sets under another member of the same ensemble (r4i1p1f1 and others), and even if admitted, it would still be superseded as the CAM6 family representative by CESM2-WACCM with its better D_i; GISS-E2-1-G and GISS-E2-1-H have complete sets only under different physics or forcing configurations (p3, f2), i.e., effectively different model versions, and neither passes the $D_i < 1$ threshold in any case; no other ineligible model has a complete scenario set under any variant (Table S1).

4. Conclusions

According to CERES EBAF Ed4.2.1, the global planetary albedo declined significantly over 2000–2014, at $(1.20 \pm 0.61) \times 10^{-3}$ per decade, with about 70% of this magnitude occurring under clear-sky conditions; the ERA5 trend is statistically indistinguishable from zero, reflecting the model origin of its TOA fluxes.

CMIP6 models reproduce the correct sign of the trend but underestimate its magnitude: 57–60% of that observed over 2000–2014 (within internal variability) and less than half over 2000–2025 -regardless of ensemble composition; the full multi-model ensemble (MME) systematically overestimates the mean albedo by +1.3%.

The proposed objective selection procedure (data completeness → performance on three climatological metrics → independence across model families → sensitivity analysis) identifies the selected nine-model ensemble (SEL9), which reduces the mean-albedo bias to +0.1% with the other characteristics remaining comparable; the apparent advantage of ensembles selected for agreement of 15-yr trends with observations is not reproduced under independent climatology-based selection and reflects internal variability.

Global albedo continues to decline in all ScenarioMIP scenarios -on ensemble average from -0.58×10^{-3} per decade under SSP1-2.6 to -1.51×10^{-3} under SSP5-8.5 over 2015–2100, with the largest changes at polar latitudes (by 2091–2100 equivalent to an additional ~1.6–3.9 W m^{-2} of absorbed solar radiation); only under SSP1-2.6 does the decline level off by the end of the century.

Over 2015–2025 the scenarios are statistically indistinguishable in albedo, so this period is suitable for independent evaluation of the ensemble but not for identifying the pathway; the observed decline ($(2.97 \pm 1.32)\times10^{-3}$ per decade) lies at the lower edge of the ensemble spread (only two of the 36 scenario members darken at least as fast as observed), and over 2000–2025 the ensemble reproduces less than half of the observed trend, which is significant at the grid-cell level over 15% of the globe.

Statements and Declarations

Funding: This work was supported by the Ministry of Education and Science of the Russian Federation as part of the program of the Moscow Center for Fundamental and Applied Mathematics under the Agreement No. 075-15-2025-345.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions: Aleksei Poliukhov: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Supervision, Funding acquisition. Alexander Lutsenko: Conceptualization, Investigation, Software, Data curation, Visualization, Writing – original draft.

Data availability: CERES EBAF Ed4.2.1 (TOA and Surface) [dataset]: NASA LaRC ASDC (NASA/LARC/SD/ASDC 2023), https://doi.org/10.5067/TERRA-AQUA-NOAA20/CERES/EBAF_L3B004.2.1 (accessed August 2026). ERA5 monthly averaged data on single levels [dataset]: Copernicus Climate Data Store (Hersbach et al. 2023), https://doi.org/10.24381/cds.f17050d7 (accessed August 2026). CMIP6 model output [dataset]: Earth System Grid Federation (ESGF) / CEDA Archive; the list of models and variant labels is given in Table S1 (accessed August 2026).

References

Abramowitz G, Herger N, Gutmann E, Hammerling D, Knutti R, Leduc M, Lorenz R, Pincus R, Schmidt GA (2019) ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. Earth Syst Dyn 10:91–105. https://doi.org/10.5194/esd-10-91-2019

Bock L, Lauer A, Schlund M, Barreiro M, Bellouin N, Jones C, Meehl GA, Predoi V, Roberts MJ, Eyring V (2020) Quantifying progress across different CMIP phases with

the ESMValTool. J Geophys Res Atmos 125:e2019JD032321. https://doi.org/10.1029/2019JD032321

Brunner L, Pendergrass AG, Lehner F, Merrifield AL, Lorenz R, Knutti R (2020) Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth Syst Dyn 11:995–1012. https://doi.org/10.5194/esd-11-995-2020

Crueger T, Schmidt H, Stevens B (2023) Hemispheric albedo asymmetries across three phases of CMIP. J Clim 36:5267–5280. https://doi.org/10.1175/JCLI-D-22-0923.1

Datseris G, Stevens B (2021) Earth's albedo and its symmetry. AGU Adv 2:e2021AV000440. https://doi.org/10.1029/2021AV000440

Deser C, Phillips A, Bourdette V, Teng H (2012) Uncertainty in climate change projections: the role of internal variability. Clim Dyn 38:527–546. https://doi.org/10.1007/s00382-010-0977-x

Donohoe A, Battisti DS (2011) Atmospheric and surface contributions to planetary albedo. J Clim 24:4402–4418. https://doi.org/10.1175/2011JCLI3946.1

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9:1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Flanner MG, Shell KM, Barlage M, Perovich DK, Tschudi MA (2011) Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. Nat Geosci 4:151–155. https://doi.org/10.1038/ngeo1062

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res Atmos 113:D06104. https://doi.org/10.1029/2007JD008972

Goessling HF, Rackow T, Jung T (2025) Recent global temperature surge intensified by record-low planetary albedo. Science 387:68–73. https://doi.org/10.1126/science.adq7280

Goode PR, Pallé E, Shoumko A, Shoumko S, Montañés-Rodríguez P, Koonin SE (2021) Earth's albedo 1998-2017 as measured from earthshine. Geophys Res Lett 48:e2021GL094888. https://doi.org/10.1029/2021GL094888

Hansen JE, Kharecha P, Sato M, Tselioudis G, Kelly J, Bauer SE, Ruedy R, Jeong E, Jin Q, Rignot E, Velicogna I, Schoeberl MR, von Schuckmann K, Amponsem J, Cao J, Keskinen A, Li J, Pokela A (2025) Global warming has accelerated: Are the United Nations and the public well-informed? Environ Sci Policy Sustain Dev 67:6–44. https://doi.org/10.1080/00139157.2025.2434494

Hausfather Z, Marvel K, Schmidt GA, Nielsen-Gammon JW, Zelinka M (2022) Climate simulations: recognize the "hot model" problem. Nature 605:26–29. https://doi.org/10.1038/d41586-022-01192-2

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 RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J-N (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146:1999–2049. https://doi.org/10.1002/qj.3803

Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut J-N (2023) ERA5 monthly averaged data on single levels from 1940 to present [dataset]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.f17050d7

Hodnebrog, Ø., Myhre G, Jouan C, Andrews T, Forster PM, Jia H, Loeb NG, Olivié DJL, Paynter D, Quaas J, Raghuraman SP, Schulz M (2024) Recent reductions in aerosol emissions have increased Earth's energy imbalance. Commun Earth Environ 5:166. https://doi.org/10.1038/s43247-024-01324-8

Hyder P, Edwards JM, Allan RP, Hewitt HT, Bracegirdle TJ, Gregory JM, Wood RA, Meijers AJS, Mulcahy J, Field P, Furtado K, Bodas-Salcedo A, Williams KD, Copsey D, Josey SA, Liu C, Roberts CD, Sanchez C, Ridley J, Thorpe L, Hardiman SC, Mayer M, Berry DI, Belcher SE (2018) Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. Nat Commun 9:3625. https://doi.org/10.1038/s41467-018-05634-2

Jian B, Li J, Zhao Y, He Y, Wang J, Huang J (2020) Evaluation of the CMIP6 planetary albedo climatology using satellite observations. Clim Dyn 54:5145–5161. https://doi.org/10.1007/s00382-020-05277-4

Knutti R, Sedláček J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017) A climate model projection weighting scheme accounting for performance and interdependence. Geophys Res Lett 44:1909–1918. https://doi.org/10.1002/2016GL072012

Li R, Jian B, Li J, Li J, Cao Z, Wang Y, Wang Y, Huang J (2025) Planetary albedo and reflected shortwave flux: Basic characteristics, mechanisms of change and future projections. Earth-Sci Rev 271:105274. https://doi.org/10.1016/j.earscirev.2025.105274

Loeb NG, Doelling DR, Wang H, Su W, Nguyen C, Corbett JG, Liang L, Mitrescu C, Rose FG, Kato S (2018) Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 data product. J Clim 31:895–918. https://doi.org/10.1175/JCLI-D-17-0208.1

Loeb NG, Wang H, Allan RP, Andrews T, Armour K, Cole JNS, Dufresne J-L, Forster P, Gettelman A, Guo H, Mauritsen T, Ming Y, Paynter D, Proistosescu C, Stuecker MF, Willén U, Wyser K (2020a) New generation of climate models track recent unprecedented changes in Earth's radiation budget observed by CERES. Geophys Res Lett 47:e2019GL086705. https://doi.org/10.1029/2019GL086705

Loeb NG, Wang H, Rose FG, Kato S, Smith WL Sr, Sun-Mack S (2020b) Toward a consistent definition between satellite and model clear-sky radiative fluxes. J Clim 33:61–75. https://doi.org/10.1175/JCLI-D-19-0381.1

Loeb NG, Mayer M, Kato S, Fasullo JT, Zuo H, Senan R, Lyman JM, Johnson GC, Balmaseda M (2022) Evaluating twenty-year trends in Earth's energy flows from observations and reanalyses. J Geophys Res Atmos 127:e2022JD036686. https://doi.org/10.1029/2022JD036686

Loeb NG, Ham S-H, Allan RP, Thorsen TJ, Meyssignac B, Kato S, Johnson GC, Lyman JM (2024) Observational assessment of changes in Earth's energy imbalance since 2000. Surv Geophys 45:1757–1783. https://doi.org/10.1007/s10712-024-09838-8

Lund MT, Myhre G, Samset BH (2019) Anthropogenic aerosol forcing under the Shared Socioeconomic Pathways. Atmos Chem Phys 19:13827–13839. https://doi.org/10.5194/acp-19-13827-2019

Marotzke J, Forster PM (2015) Forcing, feedback and internal variability in global temperature trends. Nature 517:565–570. https://doi.org/10.1038/nature14117

Mauritsen T, Tsushima Y, Meyssignac B, Loeb NG, Hakuba MZ, Pilewskie P, Cole J, Suzuki K, Ackerman TP, Allan RP, Andrews T, Bender FA-M, Bloch-Johnson J, Bodas-Salcedo A, Brookshaw A, Ceppi P, Clerbaux N, Dessler AE, Donohoe A, Dufresne J-L, Eyring V, Findell KL, Gettelman A, Gristey JJ, Hawkins E, Heimbach P, Hewitt HT, Jeevanjee N, Jones C, Kang SM, Kato S, Kay JE, Klein SA, Knutti R, Kramer R, Lee J-Y, McCoy DT, Medeiros B, Megner L, Modak A, Ogura T, Palmer MD, 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 MJ, Wild M, Zelinka MD (2025) Earth's energy imbalance more than doubled in recent decades. AGU Adv 6:e2024AV001636. https://doi.org/10.1029/2024AV001636

Minnis P, Harrison EF, Stowe LL, Gibson GG, Denn FM, Doelling DR, Smith WL Jr (1993) Radiative climate forcing by the Mount Pinatubo eruption. Science 259:1411–1415. https://doi.org/10.1126/science.259.5100.1411

O'Neill BC, Tebaldi C, van Vuuren DP, Eyring V, Friedlingstein P, Hurtt G, Knutti R, Kriegler E, Lamarque J-F, Lowe J, Meehl GA, Moss R, Riahi K, Sanderson BM (2016) The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geosci Model Dev 9:3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

NASA/LARC/SD/ASDC (2023) CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly Means Data in netCDF Edition 4.2.1 [dataset]. NASA Langley Atmospheric Science Data Center DAAC. https://doi.org/10.5067/TERRA-AQUA-NOAA20/CERES/EBAF_L3B004.2.1

Pistone K, Eisenman I, Ramanathan V (2014) Observational determination of albedo decrease caused by vanishing Arctic sea ice. Proc Natl Acad Sci USA 111:3322–3326. https://doi.org/10.1073/pnas.1318201111

Qu X, Hall A (2014) On the persistent spread in snow-albedo feedback. Clim Dyn 42:69–81. https://doi.org/10.1007/s00382-013-1774-0

Quaas J, Jia H, Smith C, Albright AL, Aas W, Bellouin N, Boucher O, Doutriaux-Boucher M, Forster PM, Grosvenor D, Jenkins S, Klimont Z, Loeb NG, Ma X, Naik V, Paulot F, Stier P, Wild M, Myhre G, Schulz M (2022) Robust evidence for reversal of the trend in aerosol effective climate forcing. Atmos Chem Phys 22:12221–12239. https://doi.org/10.5194/acp-22-12221-2022

Riahi K, van Vuuren DP, Kriegler E, Edmonds J, O'Neill BC, Fujimori S, Bauer N, Calvin K, Dellink R, Fricko O, Lutz W, Popp A, Crespo Cuaresma J, KC S, Leimbach M, Jiang L, Kram T, Rao S, Emmerling J, Ebi K, Hasegawa T, Havlik P, Humpenöder F, Da Silva LA, Smith S, Stehfest E, Bosetti V, Eom J, Gernaat D, Masui T, Rogelj J, Strefler J, Drouet L, Krey V, Luderer G, Harmsen M, Takahashi K, Baumstark L, Doelman JC, Kainuma M, Klimont Z, Marangoni G, Lotze-Campen H, Obersteiner M, Tabeau A, Tavoni M (2017) The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Glob Environ Change 42:153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Robock A (2000) Volcanic eruptions and climate. Rev Geophys 38:191–219. https://doi.org/10.1029/1998RG000054

Sanderson BM, Knutti R, Caldwell P (2015) A representative democracy to reduce interdependency in a multimodel ensemble. J Clim 28:5171–5194. https://doi.org/10.1175/JCLI-D-14-00362.1

691 Santer BD, Wigley TML, Boyle JS, Gaffen DJ, Hnilo JJ, Nychka D, Parker DE, Taylor
692 KE (2000) Statistical significance of trends and trend differences in layer-average
693 atmospheric temperature time series. J Geophys Res 105:7337–7356.
694 https://doi.org/10.1029/1999JD901105

695 Schmidt GA, Andrews T, Bauer SE, Durack PJ, Loeb NG, Ramaswamy V, Arnold NP,
696 Bosilovich MG, Cole J, Horowitz LW, Johnson GC, Lyman JM, Medeiros B, Michibata
697 T, Olonscheck D, Paynter D, Raghuraman SP, Schulz M, Takasuka D, Tallapragada V,
698 Taylor PC, Ziehn T (2023) CERESMIP: a climate modeling protocol to investigate recent
699 trends in the Earth's Energy Imbalance. Front Clim 5:1202161.
700 https://doi.org/10.3389/fclim.2023.1202161

701 Sherwood SC, Webb MJ, Annan JD, Armour KC, Forster PM, Hargreaves JC, Hegerl G,
702 Klein SA, Marvel KD, Rohling EJ, Watanabe M, Andrews T, Braconnot P, Bretherton
703 CS, Foster GL, Hausfather Z, von der Heydt AS, Knutti R, Mauritsen T, Norris JR,
704 Proistosescu C, Rugenstein M, Schmidt GA, Tokarska KB, Zelinka MD (2020) An
705 assessment of Earth's climate sensitivity using multiple lines of evidence. Rev Geophys
706 58:e2019RG000678. https://doi.org/10.1029/2019RG000678

707 Smith GL, Priestley KJ, Loeb NG, Wielicki BA, Charlock TP, Minnis P, Doelling DR,
708 Rutan DA (2011) Clouds and Earth Radiant Energy System (CERES), a review: Past,
709 present and future. Adv Space Res 48:254–263. https://doi.org/10.1016/j.asr.2011.03.009

710 Stephens GL, Li J, Wild M, Clayson CA, Loeb N, Kato S, L'Ecuyer T, Stackhouse PW
711 Jr, Lebsock M, Andrews T (2012) An update on Earth's energy balance in light of the
712 latest global observations. Nat Geosci 5:691–696. https://doi.org/10.1038/ngeo1580

713 Stephens GL, O'Brien D, Webster PJ, Pilewski P, Kato S, Li J-l (2015) The albedo of
714 Earth. Rev Geophys 53:141–163. https://doi.org/10.1002/2014RG000449

715 Stephens GL, Hakuba MZ, Kato S, Gettelman A, Dufresne J-L, Andrews T, Cole JNS,
716 Willén U, Mauritsen T (2022) The changing nature of Earth's reflected sunlight. Proc R
717 Soc A 478:20220053. https://doi.org/10.1098/rspa.2022.0053

718 Su W, Liang L, Wang H, Eitzen ZA (2020) Uncertainties in CERES top-of-atmosphere
719 fluxes caused by changes in accompanying imager. Remote Sens 12:2040.
720 https://doi.org/10.3390/rs12122040

721 Taylor KE (2001) Summarizing multiple aspects of model performance in a single
722 diagram. J Geophys Res Atmos 106:7183–7192. https://doi.org/10.1029/2000JD900719

723 Thackeray CW, Qu X, Hall A (2018) Why do models produce spread in snow albedo
724 feedback? Geophys Res Lett 45:6223–6231. https://doi.org/10.1029/2018GL078493

Tselioudis G, Rémillard J, Jakob C, Rossow WB (2025) Contraction of the world's storm-cloud zones the primary contributor to the 21st century increase in the Earth's sunlight absorption. Geophys Res Lett 52:e2025GL114882. https://doi.org/10.1029/2025GL114882

Wielicki BA, Barkstrom BR, Harrison EF, Lee RB III, Smith GL, Cooper JE (1996) Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System experiment. Bull Am Meteorol Soc 77:853–868. https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2

Wild M (2009) Global dimming and brightening: A review. J Geophys Res Atmos 114:D00D16. https://doi.org/10.1029/2008JD011470

Wild M (2020) The global energy balance as represented in CMIP6 climate models. Clim Dyn 55:553–577. https://doi.org/10.1007/s00382-020-05282-7

Wild M, Bosilovich MG (2024) The global energy balance as represented in atmospheric reanalyses. Surv Geophys 45:1799–1825. https://doi.org/10.1007/s10712-024-09861-9

Wilks DS (2016) "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. Bull Am Meteorol Soc 97:2263–2273. https://doi.org/10.1175/BAMS-D-15-00267.1

Yuan T, Song H, Oreopoulos L, Wood R, Bian H, Breen K, Chin M, Yu H, Barahona D, Meyer K, Platnick S (2024) Abrupt reduction in shipping emission as an inadvertent geoengineering termination shock produces substantial radiative warming. Commun Earth Environ 5:281. https://doi.org/10.1038/s43247-024-01442-3

Zelinka MD, Myers TA, McCoy DT, Po-Chedley S, Caldwell PM, Ceppi P, Klein SA, Taylor KE (2020) Causes of higher climate sensitivity in CMIP6 models. Geophys Res Lett 47:e2019GL085782. https://doi.org/10.1029/2019GL085782

Zelinka MD, Myers TA, Qin Y, Chao L-W, Klein SA, Po-Chedley S, Ma P-L, Wall CJ, Ceppi P, Gettelman A (2026) Recent cloud trends and extremes reaffirm established bounds on cloud feedback and aerosol-cloud interactions. Commun Earth Environ 7:615.https://doi.org/10.1038/s43247-026-03461-8

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.docx