

Geophysical Research Letters[®]



RESEARCH LETTER

10.1029/2022GL097716

Advanced Testing of Low, Medium, and High ECS CMIP6 GCM Simulations Versus ERA5-T2m

Nicola Scafetta¹ 

¹Department of Earth Sciences, Environment and Georesources, University of Naples Federico II, Complesso Universitario di Monte S. Angelo, Naples, Italy

Key Points:

- The Coupled Model Intercomparison Projects, global circulation models (GCMs) are very different from each other; their equilibrium climate sensitivity (ECS) large range needs to be greatly constrained for scientific and policy purposes
- We group the GCMs into three classes with low, medium, and high ECS; we test them against the ERA5-T2m records using spatial *t*-statistics
- We found that the high and medium-ECS GCMs run too hot; the low-ECS GCMs are not optimal yet, but also unalarming for the future

Correspondence to:

N. Scafetta,
nicola.scafetta@unina.it

Citation:

Scafetta, N. (2022). Advanced testing of low, medium, and high ECS CMIP6 GCM simulations versus ERA5-T2m. *Geophysical Research Letters*, 49, e2022GL097716. <https://doi.org/10.1029/2022GL097716>

Received 3 DEC 2021
Accepted 2 MAR 2022

Abstract The equilibrium climate sensitivity (ECS) of the CMIP6 global circulation models (GCMs) varies from 1.83°C to 5.67°C. Herein, 38 GCMs are grouped into three ECS classes (low, 1.80–3.00°C; medium, 3.01–4.50°C; high, 4.51–6.00°C) and compared against the ERA5-T2m records from 1980–1990 to 2011–2021. We found that all models with ECS > 3.0°C overestimate the observed global surface warming and that spatial *t*-statistics rejects the data-model agreement over 60% (using low-ECS GCMs) to 81% (using high-ECS GCMs) of the Earth's surface. Thus, the high and medium-ECS GCMs are unfit for prediction purposes. The low-ECS GCMs are not fully satisfactory yet, but they are also found unalarming because by 2050 they predict a moderate warming ($\Delta T_{preindustrial \rightarrow 2050} \lesssim 2^\circ\text{C}$).

Plain Language Summary The last-generation Coupled Model Intercomparison Projects (CMIP6) global circulation models (GCMs) are used by scientists and policymakers to interpret past and future climatic changes and to determine appropriate (adaptation or mitigation) policies to optimally address scenario-related climate-change hazards. However, these models are affected by large uncertainties. For example, their equilibrium climate sensitivity (ECS) varies from 1.83°C to 5.67°C, which makes their 21st-century predicted warming levels very uncertain. This issue is here addressed by testing the GCMs' global and local performance in predicting the 1980–2021 warming rates against the ERA5-T2m records and by grouping them into three equilibrium climate sensitivity (ECS) classes (low-ECS, 1.80–3.00°C; medium-ECS, 3.01–4.50°C; high-ECS, 4.51–6.00°C). We found that: (a) all models with ECS > 3.0°C overestimate the observed global surface warming; (b) Student *t*-tests show model failure over 60% (low-ECS) to 81% (high-ECS) of the Earth's surface. Thus, the high and medium-ECS GCMs do not appear to be consistent with the observations and should not be used for implementing policies based on their scenario forecasts. The low-ECS GCMs perform better, although not optimally; however, they are also found unalarming because for the next decades they predict moderate warming: $\Delta T_{preindustrial \rightarrow 2050} \lesssim 2^\circ\text{C}$.

1. Introduction

The World Climate Research Program (WCRP) Coupled Model Intercomparison Projects (CMIP) global climate models (GCMs) are used for interpreting and forecasting climate change (IPCC, 2007, 2013, 2021). The latest CMIP6 GCMs are still very different from each other since their equilibrium climate sensitivity (ECS)—the equilibrium warming induced by doubling the atmospheric CO₂ concentration from 280 to 560 ppm—ranges between 1.83°C and 5.67°C (IPCC, 2021). Consequently, they predict that the global surface temperature could warm between 1.0°C and 3.3°C above the pre-industrial period (1850–1900) even if the anthropocentric emissions stopped today (Huntingford et al., 2020).

The large ECS uncertainty is due to the poor physical understanding of various feedback mechanisms such as water vapor and cloudiness (IPCC, 2021; Knutti et al., 2017; Möller, 1963). Some studies suggested that high-ECS values are not supported by the observations (Jiménez-de-la Cuesta & Mauritsen, 2019; Nijse et al., 2020; Zelinka et al., 2020; Zhu et al., 2020); other studies supported only low ECS values, for example, 0.5–2.5°C (Bates, 2016; Christy & McNider, 2017; Lewis & Curry, 2018; Lindzen & Choi, 2011; Kluff et al., 2019; McKittrick & Christy, 2020; Monckton et al., 2015; Smirnov & Zhilyaev, 2021; Stefani, 2021; van Wijngaarden & Happer, 2020). Indeed, decadal and millennial climatic oscillations (Alley, 2004; Christiansen & Ljungqvist, 2012; Esper et al., 2012; Kutschera et al., 2017; Ljungqvist, 2010; Matskovsky & Helama, 2014; Moberg et al., 2005; Scafetta, 2014, 2020a) and additional solar/astronomical forcings are still debated (Connolly et al., 2021; Scafetta, 2012, 2013, 2021a; Scafetta et al., 2004, 2020; Scafetta & West, 2006; Svensmark

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

et al., 2017). Most GCMs overestimate the surface warming observed since 1980 (Scafetta, 2013, 2021a, 2021c; Tokarska et al., 2020; Wang et al., 2021) as well as that observed in the global (McKittrick & Christy, 2020) and tropical troposphere (Mitchell et al., 2020) and, in particular, at its top (200–300 hPa) where they predict an unobserved hot-spot (McKittrick & Christy, 2018). The above uncertainties prevent an accurate evaluation of the ECS.

Given the large ECS range and subsequent wide predictions of the GCMs, having information as to which models may be given more credibility would be of enormous value for policy. Furthermore, identifying the regions where the models most disagree from the data is important for improving them or correcting the data.

Scafetta (2021a) tested several CMIP6 GCMs against some temperature records and found that the data-model agreement improves for the models with lower ECS. Herein, we provide a complementary and more robust statistical approach by grouping the same models into three sub-ensembles according to their ECS: low-ECS, 1.8–3.0°C; medium-ECS, 3.01–4.50°C; and high-ECS, 4.51–6.0°C. We also adopt spatial Student *t*-statistics, which can optimally empathize regional dynamical divergences between observations and a specific ensemble of model predictions. Thus, the discrepancies being discussed cannot be interpreted simply as model noise, but should represent a significant model category failure covering the Earth's surface or specific regions over the historical period. Furthermore, we briefly discuss the implication of the findings regarding the reliability of the 21st-century predictions according to various emission scenarios.

2. Data and Method

The monthly reanalysis fields ERA5 Near-Surface Air Temperature (T2m) record (Huang et al., 2017) from 1980 to June 2021 and the surface air temperature (tas) records from 1850 to 2100 by 38 different CMIP6 GCMs (Table 1) were downloaded from the Koninklijk Nederlands Meteorologisch Instituut (KNMI) Climate Explorer (van Oldenborgh, 2020).

The ERA5-T2m record was preferred to other options (e.g., HadCRUT, GISTEM, and NOAA surface temperature records) because it covers the entire world surface and can be more properly used to test the simulations over the globe. The condition of spatial completeness is necessary because the global surface average temperature records will be also tested; this operation requires that both observed and modeled records are obtained by integrating over the same areas. The alternative temperature records poorly cover the polar, great forest, and desert regions for the lack of instrumental data (cf. Scafetta, 2021c). Their adoption would require specific model integrations over the same areas covered by the data but this exercise is left to a future work.

The satellite lower tropospheric temperature (TLT) measurements by Microwave Sounding Units (MSUs) (Mears & Wentz, 2016; Spencer et al., 2017) were not considered because their trends should be slightly scaled down to simulate the surface temperature (Christy et al., 2018; McKittrick & Christy, 2020; Mitchell et al., 2020). However, land-use and urban-changes impact the observed T2m records and, consequently, the actual global and local climatic warming trends may be significantly lower than those given by ERA5-T2m (cf.: Connolly et al., 2021; McKittrick & Tole, 2012; Scafetta, 2021b; Scafetta & Ouyang, 2019).

Simulations using historical forcings (1850–2014) and four shared socioeconomic pathway (SSP) scenarios (2015–2100) are available: SSP126 (low greenhouse gas emissions), SSP245 (intermediate emissions), SSP370 (high emissions), and SSP585 (very high emissions). The SSP curves are compatible with each other up to 2021. GCM tas levels from January 2011 to June 2021 were calculated by averaging the periods 2011–2020 and 2011–2021.

For each grid cell $j = 1, \dots, M$, and for an ensemble of models $i = 1, \dots, N$, we obtained the observed (ΔT_j^o) and modeled ($\Delta T_{i,j}^m$) temperature changes by comparing the 2011–2021 to the 1980–1990 means. Within each grid cell, the differences of the means between the models and observations are computed as:

$$\Delta T_j = \frac{1}{N} \sum_{i=1}^N \Delta T_{i,j}^m - \Delta T_j^o; \quad (1)$$

and tested using the spatial *t*-statistics

Table 1

(c1) The Adopted CMIP6 GCMs; (c2) Their ECS; (c3) SSP Simulation Analyzed in Section 3; (c3–c5) 2011–2021 Global Surface Warming (°C, 1980–1990 Anomalies); (c6–c8) 2040–2060 and (c9–11) 2080–2100 Global Surface Warming (°C, 1850–1900 Anomalies)

CMIP6 GCM	ECS (°C)	Used SSP	2011–2021 versus 1980–1990			2040–2060 versus 1850–1900			2080–2100 versus 1850–1900		
			SSP245	SSP370	SSP585	SSP245	SSP370	SSP585	SSP245	SSP370	SSP585
CIESM	5.67	585	0.74		0.71	2.72		3.07	3.66		6.28
CanESM5 p1	5.62	370	1.19	1.20	1.22	3.15	3.51	3.74	4.23	5.97	7.01
CanESM5 p2	5.62	370	1.22	1.22	1.22	3.17	3.56	3.79	4.27	6.04	7.07
CanESM5-CanOE p2	5.62	370	1.16	1.16	1.15	3.2	3.57	3.79	4.22	6.01	7.06
HadGEM3-GC31-LL f3	5.55	585	1.25		1.08	2.61		3.07	3.81		6.12
HadGEM3-GC31-MM-f3	5.42	585			0.86			3.01			5.99
UKESM1-0-LL f2	5.34	370	1.11	1.09	1.11	2.76	2.99	3.22	3.97	5.48	6.45
CESM2	5.16	370	0.75	0.76	0.80	2.2	2.28	2.77	3.22	4.17	5.52
CNRM-CM6-1 f2	4.83	370	0.65	0.66	0.68	2.12	2.24	2.54	3.28	4.41	5.65
CNRM-ESM2-1 f2	4.76	370	0.62	0.62	0.65	1.97	2.05	2.32	3.10	4.14	5.01
CESM2-WACCM	4.75	370	0.91	0.88	0.93	2.32	2.40	2.78	3.30	4.25	5.56
ACCESS-CM2	4.72	370	0.82	0.87	0.86	2.35	2.44	2.63	3.41	4.43	5.44
NESM3	4.72	585	0.95		1.02	2.13		2.75	2.88		4.98
IPSL-CM6A-LR	4.56	370	0.76	0.76	0.76	2.55	2.74	2.97	3.58	4.91	5.97
Average	5.17		0.93	0.92	0.93	2.56	2.78	3.03	3.61	4.98	6.01
Std. Dev.	0.42		0.23	0.23	0.20	0.42	0.59	0.47	0.46	0.81	0.71
KACE-1-0-G	4.48	370	0.95	0.93	0.96	3.01	3.13	3.34	3.74	4.85	5.67
EC-Earth3-Veg	4.31	370	0.86	0.84	0.87	2.44	2.52	2.80	3.44	4.56	5.42
EC-Earth3	4.3	370	0.76	0.89	0.88	2.34	2.53	2.73	3.39	4.50	5.44
CNRM-CM6-1-HR f2	4.28	370	0.71	0.71	0.73	2.62	2.68	2.96	3.82	4.71	5.76
GFDL-ESM4	3.9	370	0.73	0.67	0.66	1.65	1.75	1.90	2.28	3.24	3.70
ACCESS-ESM1-5	3.87	370	0.83	0.85	0.82	2.09	2.17	2.46	2.99	3.92	4.64
MCM-UA-1-0	3.65	370	0.83	0.79	0.90	2.08	2.20	2.63	2.96	3.89	4.68
CMCC-CM2-SR5	3.52	370	0.61	0.68	0.68	2.45	2.52	2.89	3.62	4.17	5.28
AWI-CM-1-1-MR	3.16	370	0.79	0.76	0.78	2.3	2.45	2.52	2.93	3.98	4.69
MRI-ESM2-0	3.15	370	0.71	0.63	0.80	2.04	2.13	2.41	2.70	3.53	4.31
BCC-CSM2-MR	3.04	370	0.64	0.65	0.66	1.87	2.14	2.32	2.45	3.64	3.95
Average	3.79		0.76	0.76	0.79	2.26	2.38	2.63	3.12	4.09	4.87
Std. Dev.	0.52		0.10	0.10	0.10	0.37	0.36	0.38	0.52	0.52	0.70
FGOALS-f3-L	3	370	0.68	0.70	0.69	2.24	2.44	2.58	2.83	3.89	4.62
MPI-ESM1-2-LR	3	370	0.57	0.57	0.55	1.82	2.00	2.10	2.40	3.38	3.98
MPI-ESM1-2-HR	2.98	370	0.57	0.59	0.57	1.8	2.00	2.07	2.44	3.38	3.92
FGOALS-g3	2.88	370	0.61	0.59	0.60	1.82	2.11	2.16	2.24	3.30	3.62
GISS-E2-1-G p1	2.72	370		0.70			2.12			3.39	
GISS-E2-1-G p3	2.72	370	0.58	0.40	0.45	1.97	2.10	2.33	2.58	3.46	4.10
MIROC-ES2L f2	2.68	370	0.59	0.56	0.56	1.75	1.86	2.15	2.42	3.20	3.97
MIROC6	2.61	370	0.48	0.47	0.51	1.57	1.70	1.93	2.16	2.96	3.75
NorESM2-LM	2.54	370	0.62	0.76	0.71	1.35	1.48	1.76	1.96	2.86	3.69
NorESM2-MM	2.5	370	0.77	0.67	0.71	1.56	1.61	1.86	2.12	3.00	3.77
CAMS-CSM1-0	2.29	370	0.42	0.38	0.41	1.43	1.59	1.64	1.98	2.69	3.11
INM-CM5-0	1.92	370	0.59	0.65	0.60	1.66	1.94	2.15	2.27	3.20	3.59

Table 1
Continued

CMIP6 GCM	ECS (°C)	Used SSP	2011–2021 versus 1980–1990			2040–2060 versus 1850–1900			2080–2100 versus 1850–1900		
			SSP245	SSP370	SSP585	SSP245	SSP370	SSP585	SSP245	SSP370	SSP585
INM-CM4-8	1.83	370	0.54	0.54	0.59	1.81	1.92	2.17	2.25	3.23	3.74
Average	2.59		0.58	0.59	0.58	1.73	1.91	2.07	2.31	3.23	3.82
Std. Dev.	0.38		0.09	0.12	0.09	0.24	0.26	0.25	0.25	0.30	0.36

$$t_j = \frac{|\Delta T_j|}{\sigma_j / \sqrt{N}}, \quad (2)$$

where σ_j is the standard deviation in the cell j among the N local values $\Delta T_{i,j}^m$. A data-model agreement is rejected (at the two-sided confidence level of 5%) when $t > t_0$, where t_0 (the 2.5% critical value) depends on the degree of freedom ($N - 1$). In our cases, t_0 is slightly larger than 2. The proposed statistics optimally emphasizes the regional differences between observations and a set of simulations.

3. Results

This section analyzes 38 CMIP6 simulations referring to the SSP scenario listed in Table 1-c3, one for each model.

Figure 1 shows the 38 synthetic global surface temperature records (1980–1990 anomalies, red curves) against ERA5-T2m (blue). Three panels depict separately the records from 13 low-ECS, 11 medium-ECS, and 14 high-ECS GCMs. The right panels compare the 2011–2021 mean temperatures of the models (red dots) against the data (blue dots) and show that only the low-ECS GCMs are consistent with the observed warming, while both the medium and high-ECS GCMs definitely exceed it.

Figure 2 shows the world distributions of the 2011–2021 warming (1980–1990 anomalies) of the GCM ensemble mean (top panel) and the ERA5-T2m record (bottom panel), and the relative latitudinal temperature profiles (left panels) for the land (brown), ocean (blue) and land + ocean (black) regions. The simulation predicts a worldwide diffused warming. The high latitudes (60–90°N) and the continents generally warmed more than other regions because of albedo changes related to sea-ice melting and the lower heat capacity of land versus ocean. ERA5-T2m too shows a diffused warming, however, contrary to the simulation, it also presents vast cooling regions (blue) around Antarctica, over the tropical Pacific and North Atlantic oceans, and over some land regions between the USA and Canada and in North-West Australia.

Figure 3 shows the temperature maps produced by the high, medium and low-ECS GCMs, the latitudinal profiles of the single GCMs (blue curves), and their distribution (black, mean curve; blue, $\pm\sigma$ range) versus the ERA5-T2m latitudinal temperature profile (red). The figure also shows the correspondent means of the t -test values (Equation 2) calculated from both the latitudinal profiles using the mean and a latitudinal-cosine weighted mean, and the correspondent threshold value t_0 (black). A statistical model-data agreement is not found (at the confidence level of 5%) although the low-ECS GCMs get the closest to the data ($t = 2.54 > t_0 = 2.18$).

Figure 4 (bottom) shows the spatial t -value distribution. The statistics rejects the simulations over 81% of the Earth's surface. Statistical compatibility occurs in the cyan areas ($t < t_0$); yellow-red areas indicate incompatibility ($t > t_0$). Several interesting patterns are observed. For example, a better agreement appears in the Northern Hemisphere and on the continents, although this might be coincidental because land regions are likely affected by non-climatic warming biases (cf.: Connolly et al., 2021; McKittrick & Tole, 2012; Scafetta, 2021b; Scafetta & Ouyang, 2019). The spatial t -statistics also emphasizes a large divergence over the oceanic main current gyres, in particular in the inter-tropical Pacific and Atlantic, and in the Antarctic Circumpolar region. The result complements Scafetta (2021c) where single model simulations and less robust ensemble statistics were adopted.

Figure 5 (left panels) shows the differences between the high, medium and low-ECS GCMs and ERA5-T2m, and their relative latitudinal temperature profiles. The high-ECS GCMs overestimate the warming over most of the

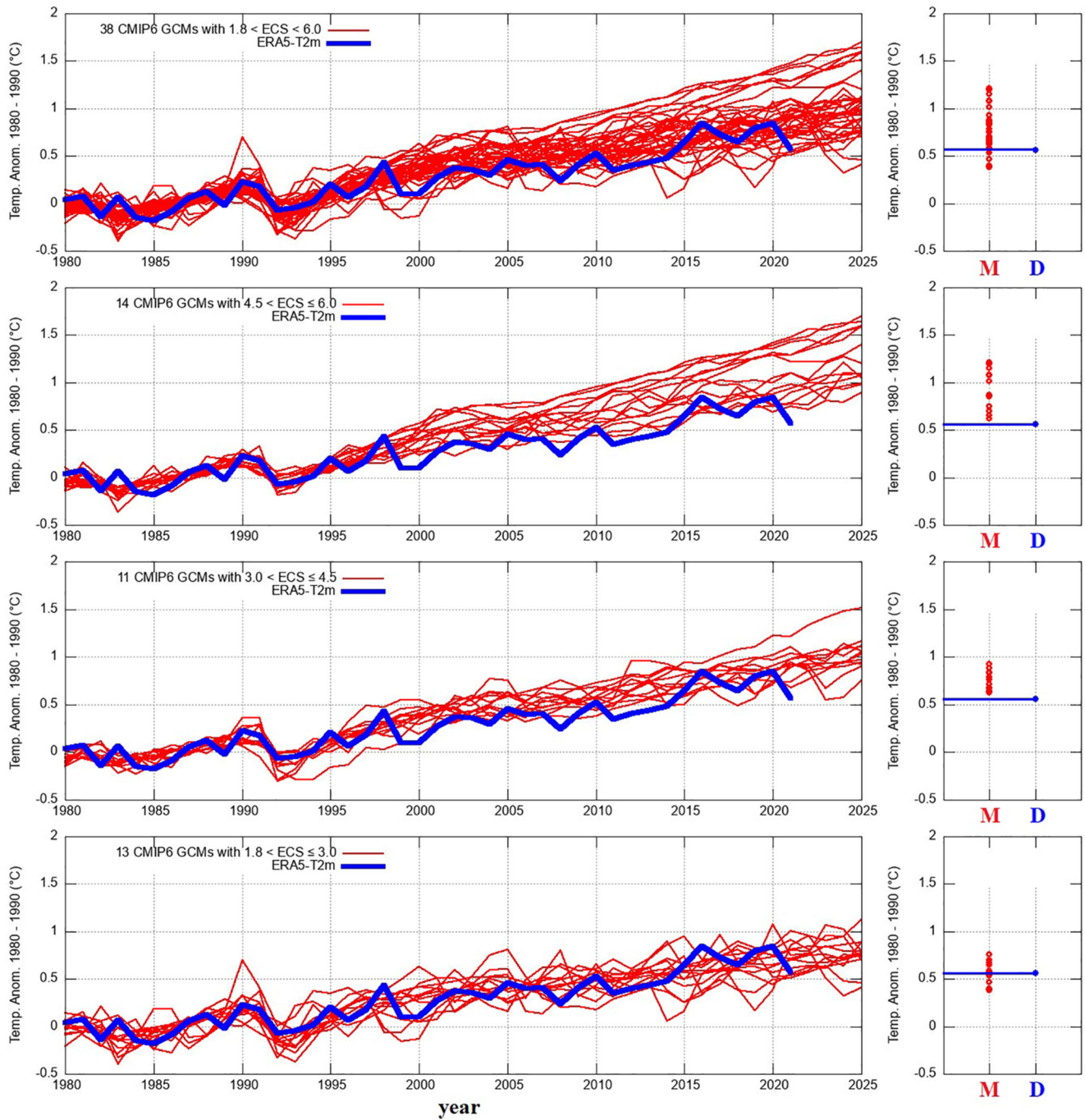


Figure 1. Left: Global surface temperature simulations (red) against ERA5-T2m (blue). Right: modeled (M, red) and observed (D, blue, $\Delta T \approx 0.56^\circ\text{C}$) mean global surface warming 2011–2021 (1980–1990 anomalies).

globe (cf. Figure 1). The situation slightly improves with the medium-ECS GCMs. Yet, it remains unsatisfactory also using the low-ECS GCMs because, for example, the ocean warming generally appears overestimated while the land one underestimated. Figure 5 (right panels) shows the spatial t -statistics between the high, medium and low-ECS GCMs versus ERA5-T2m with their relative latitudinal temperature profiles. The t -test rejects the agreement over 60% (using the low-ECS GCMs) to 81% (using the high-ECS GCMs) of the Earth's surface.

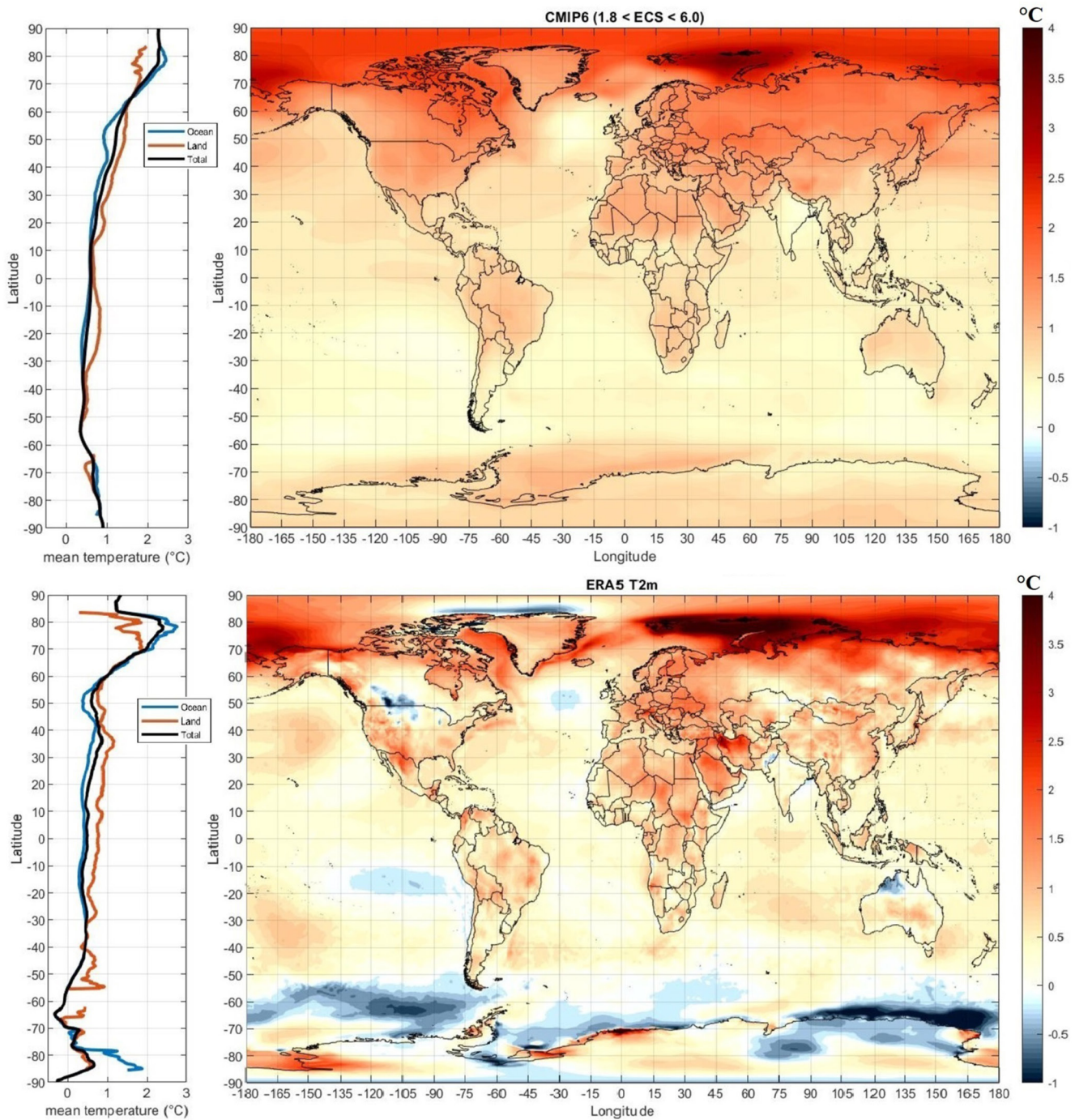


Figure 2. Warming patterns from 1980 to 1990 to 2011–2021 predicted by the global circulation models as ensemble average record (top) and ERA5-T2m (bottom). Left-sides: latitudinal temperature profiles for land, ocean and land + ocean regions.

4. Discussion and Implication for Policy

It has previously been shown that the model trends of the last decades are too steep both at the surface and in the troposphere (e.g.: McKittrick & Christy, 2018, 2020; Mitchell et al., 2020; Scafetta, 2013, 2021a, 2021c; Tokarska et al., 2020; Wang et al., 2021). This analysis focuses on the 1980–2021 period at the surface level and shows that, while no model group succeeds in reproducing the observed surface warming patterns, the high ECS models systematically do worse. Figure 1 compares the surface temperature changes for stratified ECS values and shows

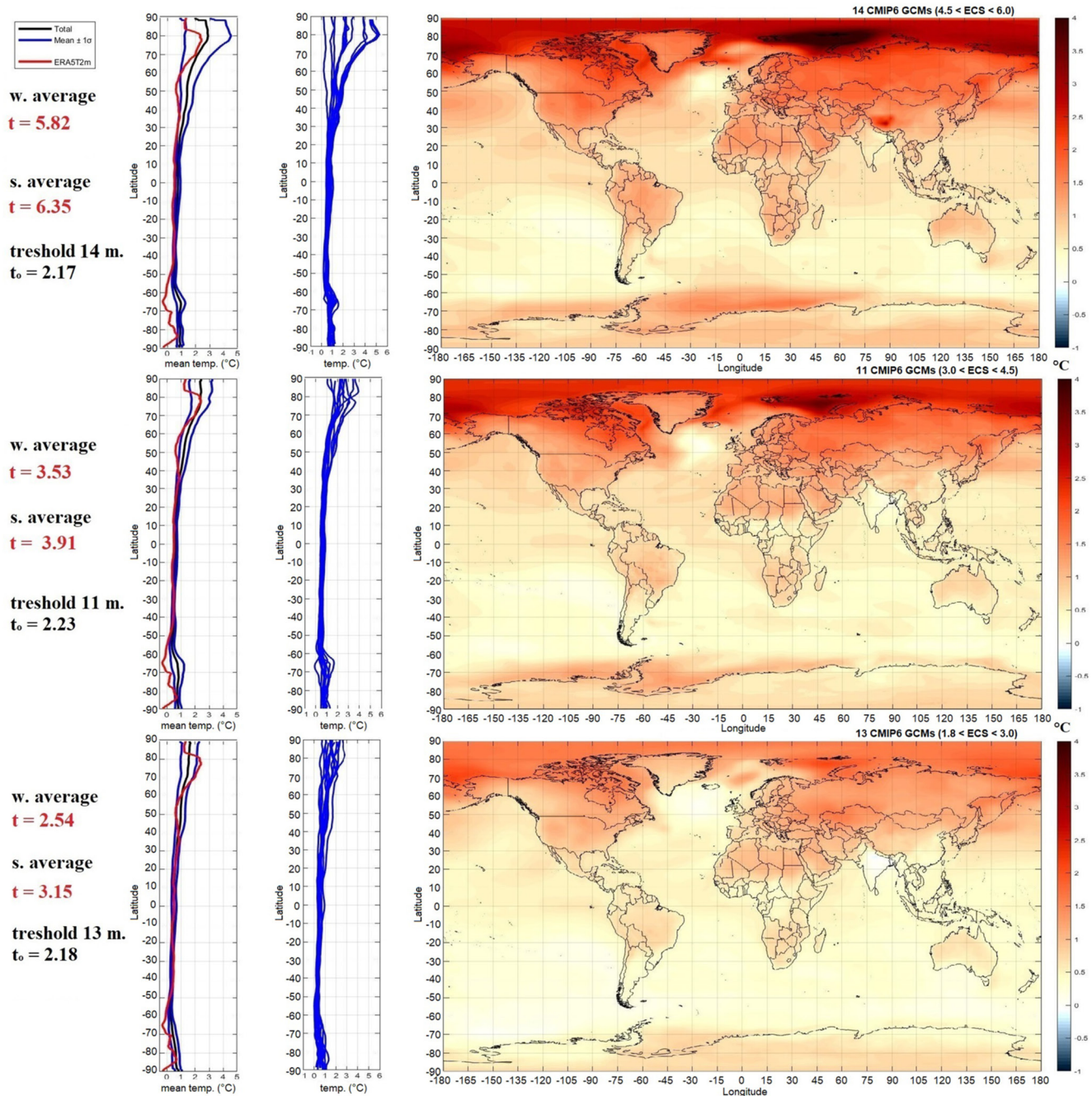


Figure 3. Spatial warming patterns from 1980 to 1990 to 2011–2021 produced by high, medium and low-equilibrium climate sensitivity (ECS) global circulation models. Left-panels: latitudinal temperature profiles for each model (blue curves); their statistical distribution against the ERA5-T2m (red) profiles; the correspondent simple and weighted mean Student's t values and threshold limit t_0 (Equation 2).

that medium and high ECS GCMs exceed the observed warming. Figures 2–5 suggest where, on the globe, the problems may be greatest. However, the model-data agreement between CMIP6 GCMs and ERA5-T2m is in general very poor both locally and globally, suggesting that the global trend alignment with the low-ECS GCMs (Figure 1) may be coincidental because it is not supported by the regional analysis (Figure 5).

Accurately reproducing regional temperature differences over the past 40+ years is beyond the capability of climate model simulations, and even fails for major ocean basins and continents. The result suggests the existence of major issues with all models and/or with the ERA5-T2m record, which may also be affected by uncorrected

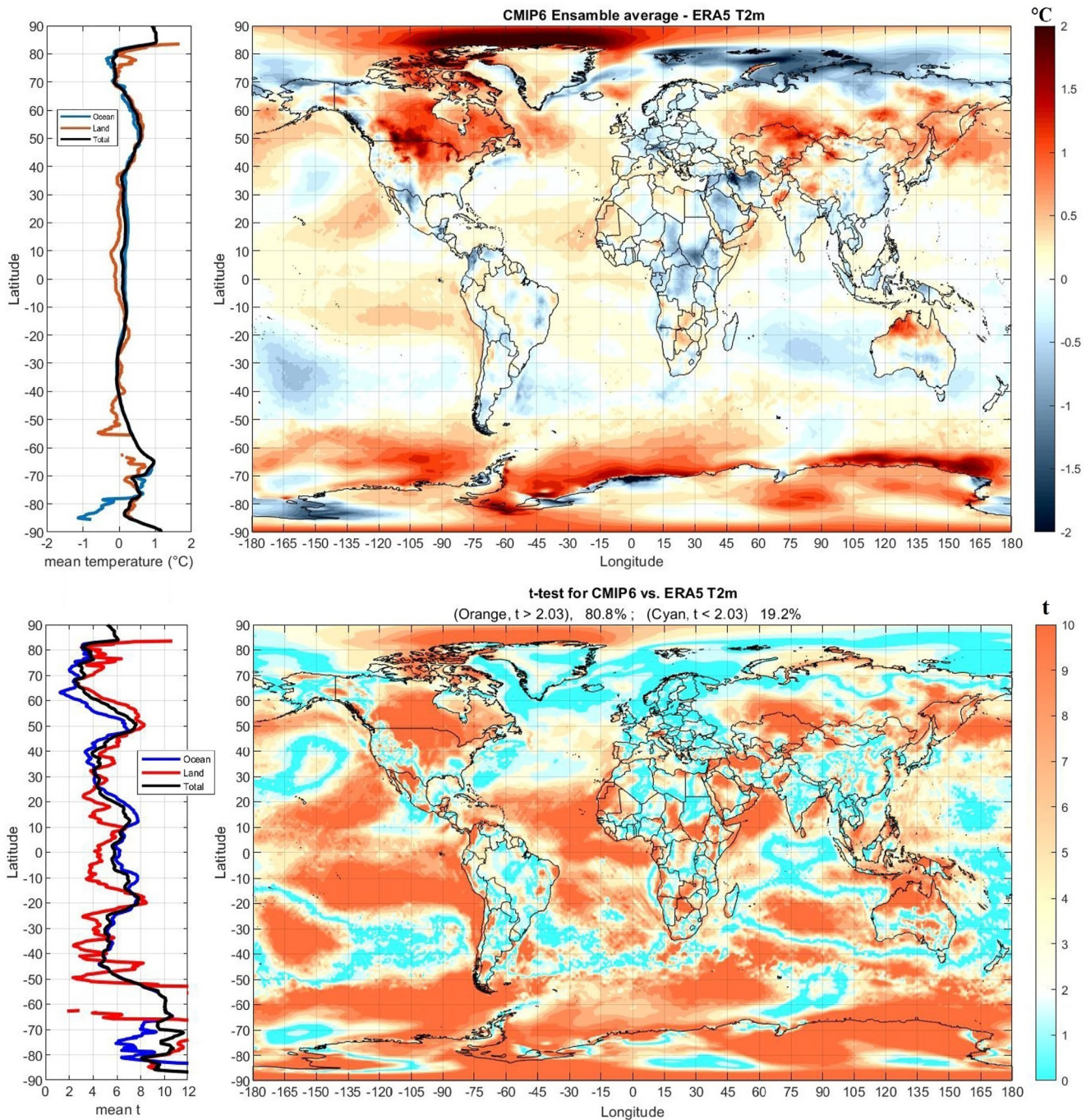


Figure 4. (Top) Differences between the global circulation models ensemble average tsa and ERA5-T2m (Equation 1). (Bottom) Spatial t-statistics between the 38 models and ERA5-T2m (Equation 2). Model-data compatibility (at the confidence level of 5%) occurs in the cyan color areas ($t < t_0 = 2.03$) while yellow-red areas indicate statistical incompatibility ($t > t_0 = 2.03$). Left-panels: relative latitudinal profiles.

local non-climatic biases (e.g., over land: Connolly et al., 2021; McKittrick & Tole, 2012; Scafetta, 2021b; Scafetta & Ouyang, 2019)

The evidence here presented indicates that the CMIP6 GCMs do not reproduce well both the global and regional (i.e., country-size) responses to enhanced greenhouse gases over the past 40+ years, which calls into question model-based attribution of climate responses to anthropogenic forcing.

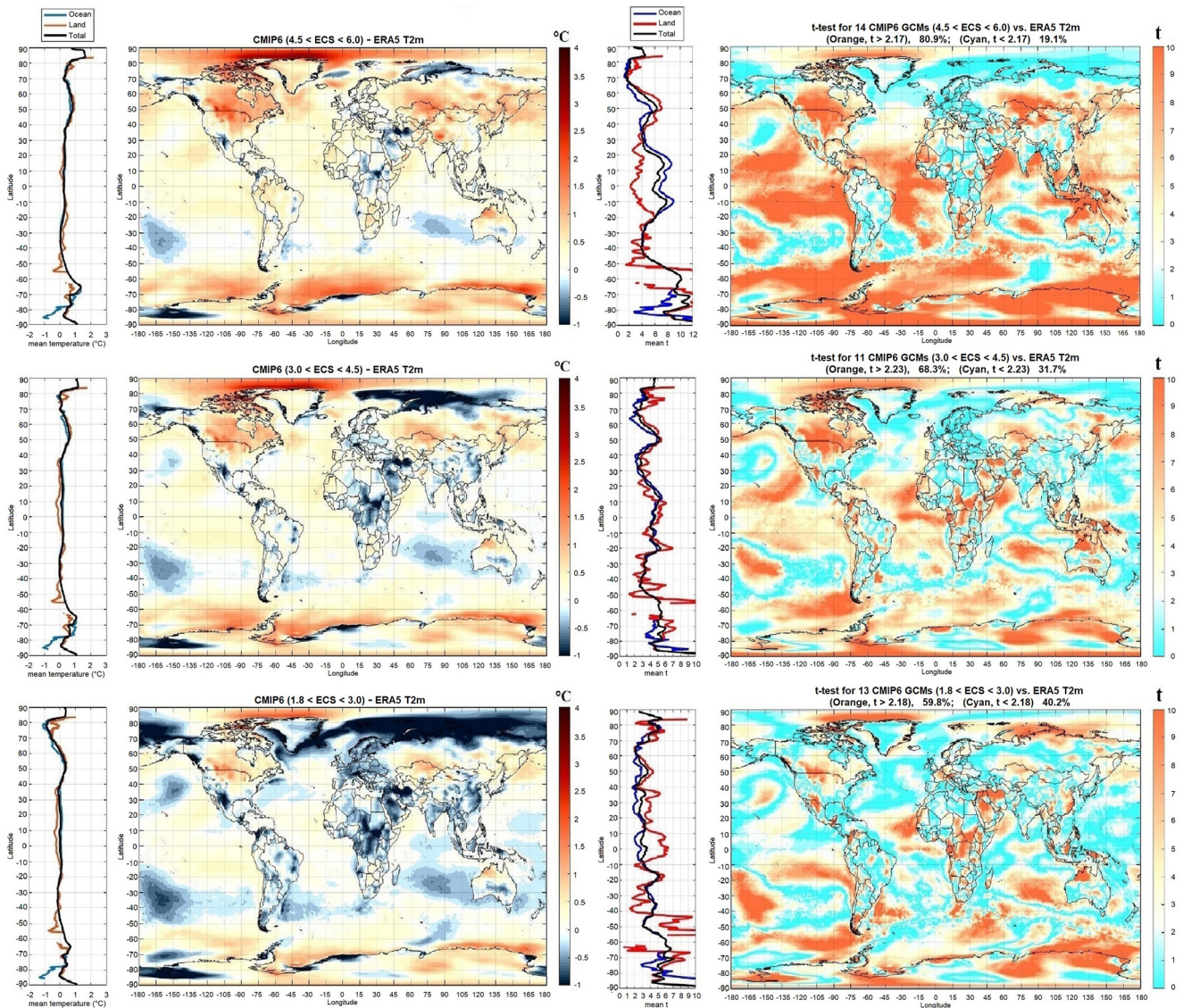


Figure 5. (Left) Differences between the high, medium and low-equilibrium climate sensitivity (ECS) global circulation models (GCM) ensemble average tsa and ERA5-T2m (Equation 1). (Right) Spatial t-statistics between high, medium and low-ECS GCMs and ERA5-T2m (Equation 2). Left-panels: relative latitudinal profiles. Model-data compatibility (at the confidence level of 5%) occurs in the cyan color areas ($t < t_0$) while yellow-red areas indicate statistical incompatibility ($t > t_0$).

The fact that the low-ECS models perform relatively better while representing post-1980 surface temperature responses has important social and political implications. Policy plans are based on projections like those in Figure 6, which compares all available CMIP6 GCM simulations for each available SSP relative to the pre-industrial era (1850–1900) up to 2100. The curves are differently colored according to the GCM ECS value: low-ECS (blue), medium-ECS (green) and high-ECS (red). The right panels show the correspondent warming levels in 2040–2060 and 2080–2100. The GCM predicted warming increases with ECS.

The prospect of experiencing a high level of warming associated with high ECS models has led to costly international efforts to reduce net greenhouse gas emissions to zero on a rapid timetable. However, the low-ECS GCMs predict average warming by 2040–2060 close to or lower than 2°C even for the SSP585 scenario (Table 1), which is considered highly unlikely (IPCC, 2021). The fact that the high and medium-ECS GCMs do not appear to be consistent with the observations over the past 40+ years imply that their projections should not be used as a basis for policy. The low-ECS GCMs are the closest to the data but they are unalarming because they predict

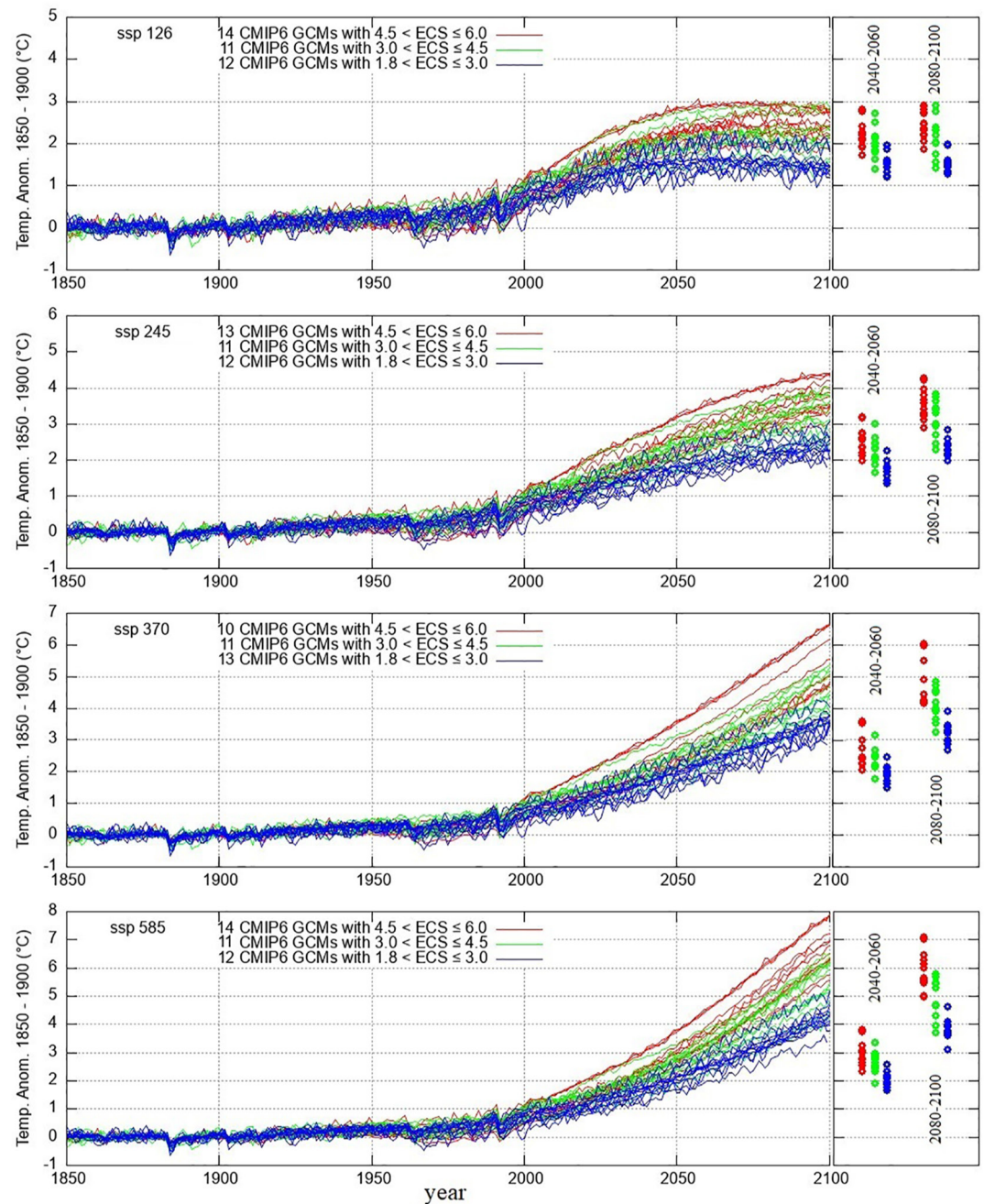


Figure 6. All available Coupled Model Intercomparison Projects global circulation models (GCM) simulations for different SSP: low-equilibrium climate sensitivity (ECS) (blue), medium-ECS (green) and high-ECS (red) GCMs. Right-panels: model mean warming levels in 2040–2060 and 2080–2100. See Table 1.

moderate warming ($\Delta T_{\text{preindustrial} \rightarrow 2050} \lesssim 2^\circ\text{C}$). Thus, inexpensive adaptation policies should be preferred because they should be sufficient to address most hazards related to future climate changes.

5. Conclusions

The large ECS range (between 1.8°C and 5.7°C) of the CMIP6 GCMs indicates that these models are intrinsically very different from each other. To constrain it, we grouped the models into three different classes (low-ECS, 1.80–3.00°C; medium-ECS, 3.01–4.50°C; and high-ECS, 4.51–6.00°C) and tested their predicted temperature

changes from the 1980–1990 to 2011–2021 periods against ERA5-T2m. The proposed methodology compares single GCM runs and specific ensemble averages versus the observations using both the temperature changes (Equation 1) and spatial t-statistics (Equation 2) maps. This technique optimally highlights both the global and local discrepancies between the observations and GCM ensemble predictions. Figures 4 and 5 show applications of the proposed methodology and provide much information regarding the performance of the CMIP6 GCMs in simulating the data.

High and medium-ECS GCMs overestimate the observed warming. The low-ECS GCMs appear to agree with the observations better on average (Figure 1), but they still perform poorly when synoptic temperature patterns are analyzed (Figure 5). In general, spatial t-statistics demonstrated that over more than 60% of the world surface the CMIP6 GCM predictions are incompatible with the temperature records (at the confidence level of 5%). Various Northern-Southern hemispheric and land-ocean asymmetries, and important dynamical patterns—such as those related to the main oceanic currents of the Pacific, the Atlantic and around the Antarctic Circumpolar region—are observed. The results suggest poor modeling of heat transfer, ocean and atmospheric circulation, and Arctic and Antarctic sea-ice processes. Furthermore, over land, ERA5-T2m is likely affected by numerous non-climatic biases that could have stressed local warming trends (McKittrick & Tole, 2012; Scafetta, 2021b; Scafetta & Ouyang, 2019).

The result has important implications for policy as well because the medium and high-ECS GCMs are not sufficiently reliable at the largest scale (global average) and all GCMs fail to provide confident forecasts of regional responses to enhanced greenhouse gases so that also the impacts of policy options on the regional climate are highly uncertain. The models that best match the post-1980 observations imply that by 2050 the global surface warming should remain moderate ($\Delta T \lesssim 2^\circ\text{C}$) compared to preindustrial temperatures even under an extremely high emissions growth scenario with no mitigation effort.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Data from KNMI Climate Explorer: (a) ERA5-T2m, http://climexp.knmi.nl/selectfield_rea.cgi?id=someone@somewhere; (b) CMIP6 GCMs, http://climexp.knmi.nl/selectfield_cmip6.cgi?id=someone@somewhere.

References

- Alley, R. B. (2004). GISP2 ice core temperature and accumulation data. *IGBP PAGES/World data center for paleoclimatology data contribution series #2004-013*. NOAA/NGDC Paleoclimatology Program
- Bates, J. R. (2016). Estimating climate sensitivity using two-zone energy balance models. *Earth and Space Science*, 3, 207–225. <https://doi.org/10.1002/2015EA000154>
- Christiansen, B., & Ljungqvist, F. C. (2012). The extra-tropical Northern hemisphere temperature in the last two millennia: Reconstructions of low-frequency variability. *Climate of the Past*, 8, 765–786. <https://doi.org/10.5194/cp-8-765-2012>
- Christy, J. R., & McNider, R. T. (2017). Satellite bulk tropospheric temperatures as a metric for climate sensitivity. *Asia-Pacific Journal of Atmospheric Sciences*, 53(4), 511–518. <https://doi.org/10.1007/s13143-017-0070-z>
- Christy, J. R., Spencer, R. W., Braswell, W. D., & Junod, R. (2018). Examination of space-based bulk atmospheric temperatures for climate research. *International Journal of Remote Sensing*, 39, 3580–3607. <https://doi.org/10.1080/01431161.2018.1444293>
- Connolly, R., Soon, W., Connolly, M., Baliunas, S., Berglund, J., Butler, C. J., et al. (2021). How much has the Sun influenced Northern hemisphere temperature trends? An ongoing debate. *Research in Astronomy and Astrophysics*, 21, 131. <https://doi.org/10.1088/1674-4527/21/6/131>
- Esper, J., Frank, D. C., Timonen, M., Zorita, E., Wilson, R. J. S., Luterbacher, J., et al. (2012). Orbital forcing of tree-ring data. *Nature Climate Change*, 12, 862–866. <https://doi.org/10.1038/nclimate1589>
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed sea surface temperature version 5 (ERSSTv5), upgrades, validations, and intercomparisons. *Journal of Climate*, 30, 8179–8205. <https://doi.org/10.1175/JCLI-D-16-0836.1>
- Huntingford, C., Williamson, M. S., & Nijse, F. J. M. M. (2020). CMIP6 climate models imply high committed warming. *Climatic Change*, 162, 1515–1520. <https://doi.org/10.1007/s10584-020-02849-5>
- Intergovernmental Panel on Climate Change (IPCC). (2007). *Working group I contribution to the Third Assessment Report (AR3), climate change 2007: The physical science basis*. Retrieved from <https://www.ipcc.ch/assessment-report/ar3/>
- Intergovernmental Panel on Climate Change (IPCC). (2013). *Working group I contribution to the Fifth Assessment Report (AR5), climate change 2013: The physical science basis*. Retrieved from <https://www.ipcc.ch/assessment-report/ar5/>
- Intergovernmental Panel on Climate Change (IPCC). (2021). *Working group I contribution to the Sixth Assessment Report (AR6), climate change 2021: The physical science basis*. Retrieved from <https://www.ipcc.ch/assessment-report/ar6/>

Acknowledgments

Open Access Funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

- Jiménez-de-la Cuesta, D., & Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO₂ from post-1970s global warming. *Nature Geoscience*, *12*, 902–905. <https://doi.org/10.1038/s41561-019-0463-y>
- Klufft, L., Dacie, S., Schmidt, H., & Stevens, B. (2019). Re-examining the first climate models: Climate sensitivity of a modern radiative-convective equilibrium model. *Journal of Climate*, *32*, 8111–8125. <https://doi.org/10.1175/JCLI-D-18-0774.1>
- Knutti, R., Rugenstein, M. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, *10*, 727–736. <https://doi.org/10.1038/ngeo3017>
- Kutschera, W., Patzelt, G., Steier, P., & Wild, E. M. (2017). The Tyrolean iceman and his glacial environment during the holocene. *Radiocarbon*, *59*, 395–405. <https://doi.org/10.1017/RDC.2016.70>
- Lewis, N., & Curry, J. (2018). The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *Journal of Climate*, *31*, 6051–6071. <https://doi.org/10.1175/JCLI-D-17-0667.1>
- Lindzen, R. S., & Choi, Y.-S. (2011). On the observational determination of climate sensitivity and its implications. *Asia-Pacific Journal of Atmospheric Sciences*, *47*, 377–390. <https://doi.org/10.1007/s13143-011-0023-x>
- Ljungqvist, F. C. (2010). A new reconstruction of temperature variability in the extratropical northern hemisphere during the last two millennia. *Geografiska Annaler: Series A, Physical Geography*, *92*, 339–351. <https://doi.org/10.1111/j.1468-0459.2010.00399.x>
- Matskovsky, V. V., & Helama, S. (2014). Testing long-term summer temperature reconstruction based on maximum density chronologies obtained by reanalysis of tree-ring data sets from northernmost Sweden and Finland. *Climate of the Past*, *10*, 1473–1487. <https://doi.org/10.5194/cp-10-1473-2014>
- McKittrick, R., & Christy, J. (2018). A test of the tropical 200 to 300 hPa warming rate in climate models. *Earth and Space Science*, *5*, 529–536. <https://doi.org/10.1029/2018ea000401>
- McKittrick, R., & Christy, J. (2020). Pervasive warming bias in CMIP6 tropospheric layers. *Earth and Space Science*, *7*, e2020EA001281. <https://doi.org/10.1029/2020ea001281>
- McKittrick, R. R., & Tole, L. (2012). Evaluating explanatory models of the spatial pattern of surface climate trends using model selection and Bayesian averaging methods. *Climate Dynamics*, *39*, 2867–2882. <https://doi.org/10.1007/s00382-012-1418-9>
- Mears, C. A., & Wentz, F. J. (2016). Sensitivity of satellite-derived tropospheric temperature trends to the diurnal cycle adjustment. *Journal of Climate*, *29*, 3629–3646. <https://doi.org/10.1175/jcli-d-15-0744.1>
- Mitchell, D. M., Lo, Y. T. E., Seviour, W. J. M., Haimberger, L., & Polvani, L. M. (2020). The vertical profile of recent tropical temperature trends: Persistent model biases in the context of internal variability. *Environmental Research Letters*, *15*, 1040b4. <https://doi.org/10.1088/1748-9326/ab9af7>
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., & Karlén, W. (2005). Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, *433*, 613–617. <https://doi.org/10.1038/nature03265>
- Möller, F. (1963). On the influence of changes in the CO₂ concentration in air on the radiation balance of the Earth's surface and on the climate. *Journal of Geophysical Research*, *68*, 3877–3886. <https://doi.org/10.1029/jz068i013p03877>
- Monckton, C., Soon, W.-H., Legates, D. R., & Briggs, W. M. (2015). Why models run hot: Results from an irreducibly simple model. *Science Bulletin*, *60*, 122–135. <https://doi.org/10.1007/s11434-014-0699-2>
- Nijse, F. J. M. M., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, *11*, 737–750. <https://doi.org/10.5194/esd-11-737-2020>
- Scafetta, N. (2012). Testing an astronomically based decadal-scale empirical harmonic climate model versus the IPCC (2007) general circulation climate models. *Journal of Atmospheric and Solar-Terrestrial Physics*, *80*, 124–137. <https://doi.org/10.1016/j.jastp.2011.12.005>
- Scafetta, N. (2013). Discussion on climate oscillations: CMIP5 general circulation models versus a semiempirical harmonic model based on astronomical cycles. *Earth-Science Reviews*, *126*, 321–357. <https://doi.org/10.1016/j.earscirev.2013.08.008>
- Scafetta, N. (2014). Discussion on the spectral coherence between planetary, solar and climate oscillations: A reply to some critiques. *Astrophysics and Space Science*, *354*, 275–299. <https://doi.org/10.1007/s10509-014-2111-8>
- Scafetta, N. (2020). Solar oscillations and the orbital invariant inequalities of the solar system. *Solar Physics*, *295*(2), 33. <https://doi.org/10.1007/s11207-020-01599-y>
- Scafetta, N. (2021a). Detection of non-climatic biases in land surface temperature records by comparing climatic data and their model simulations. *Climate Dynamics*, *56*(9–10), 2959–2982. <https://doi.org/10.1007/s00382-021-05626-x>
- Scafetta, N. (2021b). Reconstruction of the interannual to millennial scale patterns of the global surface temperature. *Atmosphere*, *12*, 147. <https://doi.org/10.3390/atmos12020147>
- Scafetta, N. (2021c). Testing the CMIP6 GCM simulations versus surface temperature records from 1980–1990 to 2011–2021: High ECS is not supported. *Climate*, *9*(11), 161. <https://doi.org/10.3390/cli9110161>
- Scafetta, N., Grigolini, P., Imholt, T., Roberts, J. A., & West, B. J. (2004). Solar turbulence in Earth's global and regional temperature anomalies. *Physical Review E*, *69*, 026303. <https://doi.org/10.1103/physreve.69.026303>
- Scafetta, N., Milani, F., & Bianchini, A. (2020). A 60-year cycle in the meteorite fall frequency suggests a possible interplanetary dust forcing of the Earth's climate driven by planetary oscillations. *Geophysical Research Letters*, *47*(18), e2020GL089954. <https://doi.org/10.1029/2020gl089954>
- Scafetta, N., & Ouyang, S. (2019). Detection of UHI bias in China climate network using Tmin and Tmax surface temperature divergence. *Global and Planetary Change*, *181*, 102989. <https://doi.org/10.1016/j.gloplacha.2019.102989>
- Scafetta, N., & West, B. J. (2006). Phenomenological solar contribution to the 1900–2000 global surface warming. *Geophysical Research Letters*, *33*, L05708. <https://doi.org/10.1029/2005gl025539>
- Smirnov, B. M., & Zhilyaev, D. A. (2021). Greenhouse effect in the standard atmosphere. *Foundations*, *1*, 184–199. <https://doi.org/10.3390/foundations1020014>
- Spencer, R. W., Christy, J. R., & Braswell, W. D. (2017). UAH Version 6 global satellite temperature products: Methodology and results. *Asia-Pacific Journal of Atmospheric Sciences*, *53*, 121–130. <https://doi.org/10.1007/s13143-017-0010-y>
- Stefani, F. (2021). Solar and anthropogenic influences on climate: Regression analysis and tentative predictions. *Climate*, *9*, 163. <https://doi.org/10.3390/cli9110163>
- Svensmark, H., Enghoff, M. B., Shaviv, N. J., & Svensmark, J. (2017). Increased ionization supports growth of aerosols into cloud condensation nuclei. *Nature Communications*, *8*, 2199. <https://doi.org/10.1038/s41467-017-02082-2>
- Tokarska, K. B., Hegerl, G. C., Schurer, A. P., Forster, P. M., & Marvel, K. (2020). Observational constraints on the effective climate sensitivity from the historical period. *Environmental Research Letters*, *15*, 034043. <https://doi.org/10.1088/1748-9326/ab738f>
- van Oldenborgh, G. J. (2020). *The KNMI climate explorer*. © 2020 KNMI (results, code)/World Meteorological organization (WMO) (design). Retrieved from <https://climexp.knmi.nl/start.cgi>

- van Wijngaarden, W. A., & Happer, W. (2020). *Dependence of Earth's thermal radiation on five most abundant greenhouse gases* (pp. 1–38). Cornell University. Retrieved from <https://arxiv.org/abs/2006.03098>
- Wang, C., Soden, B. J., Yang, W., & Vecchi, G. A. (2021). Compensation between cloud feedback and aerosol-cloud interaction in CMIP6 models. *Geophysical Research Letters*, 48, e2020GL091024. <https://doi.org/10.1029/2020gl091024>
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, 47, e2019GL085782. <https://doi.org/10.1029/2019gl085782>
- Zhu, J., Poulsen, C. J., & Otto-Bliesner, B. L. (2020). High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nature Climate Change*, 10, 378–379. <https://doi.org/10.1038/s41558-020-0764-6>