Projections of Earth's Hottest Surface Temperatures in CMIP6

Anthony Wilson[1], **Nicholas Lutsko**[2], **Daniela Faggiani Dias**[3], **Arthur J. Miller**[2]

[1]Climate Sciences Program, Scripps Institution of Oceanography, UC San Diego, La Jolla, California, USA
[2]Scripps Institution of Oceanography, UC San Diego, La Jolla, California, USA
[3]Integration of Planetary and Biogeochemical Carbon Cycle Program and European Institute on Economics and the Environment, Euro-Mediterranean Center on Climate Change, Milan, Italy

Key Points

- CMIP6 captures the broad spatial pattern of $p99(T_{\mathrm{max}})$ vs Berkeley Earth, supporting projections of record-like heat.
- Tropical extremes amplify: T_{DM} and T_{DX} warm ∼30–40% faster than T_{TL} in SSP2-4.5 and abrupt 4×CO_2.
- Annual hottest-land "record-like" exceedances become common under SSP2-4.5, with odds increasing ∼35% per 1°C of tropical land warming.

Corresponding author: Anthony Wilson, `a4wilson@ucsd.edu`

Abstract

Earth's hottest near-surface air temperatures provide a benchmark for the upper tail of land heat extremes. Using CMIP6 simulations, we evaluate historical extremes against Berkeley Earth and project future changes under SSP2-4.5 and abrupt-4×CO_2. Historical fidelity is assessed from the spatial pattern of the 99th percentile of daily maximum temperature ($p99\,T_{\mathrm{max}}$) over land using centered RMSE, mean bias, and spatial R^2. We analyze three annual tropical land indices: tropical land-mean temperature (T_{TL}), daily-mean temperature on the hottest day each year (T_{DM}), and the maximum daily-maximum temperature (T_{DX}). Across models, T_{DM} and T_{DX} warm 30–40% faster than T_{TL} in transient trends and equilibrium responses, consistent with amplified upper-tail warming. Hotspot-frequency maps of the annual global land maximum show persistence in subtropical arid and semi-arid regions. We define a fixed-threshold record-like benchmark for hottest-land temperature and relate exceedance likelihood to tropical land warming; pooled fits show exceedance odds rise ∼35% per 1 °C.

Plain Language Summary

The hottest temperatures on Earth are changing as the planet warms. We use 21 CMIP6 models to study extreme heat over land, focusing on the hottest daytime temperature each year (T_{DX}). We analyze a mid-range future scenario (SSP2-4.5) and an experiment where carbon dioxide is quadrupled, sampling both transient and equilibrium responses. In both cases, hottest-day temperatures over tropical land rise faster than tropical-mean land temperature, showing that extremes intensify faster than the background climate. We also evaluate how well models reproduce the observed spatial pattern of past extreme heat using the Berkeley Earth dataset. Many models capture the broad pattern, but others have large warm or cold biases and place the hottest regions incorrectly. Tracking the single hottest land grid cell each year shows strong spatial persistence: record-setting heat remains concentrated in subtropical arid and semi-arid regions, especially the Middle East and adjacent South Asia. Finally, a fixed-threshold "record-like" benchmark indicates that exceeding the late-20th-century hottest-land temperature becomes more common as tropical land warms. These results show that mean warming alone does not determine where the most dangerous heat will occur or how often record-breaking extremes will happen there.

1 Introduction

Earth's hottest near-surface air temperatures provide a direct measure of the upper tail of the land-temperature distribution and a stringent test of how climate models represent land heat extremes. This study focuses on annual daytime maximum-temperature extremes and the geography of the hottest land grid cell each year. It is therefore distinct from heat-stress or heat-index studies, which additionally require humidity, nighttime temperature, exposure duration, and vulnerability information. A large body of literature investigates the controls on moist and dry heatwaves, projected changes in a warming climate, and the roles of meteorological variability and land-surface interactions (Seneviratne et al., 2021; Perkins-Kirkpatrick & Lewis, 2020; Raymond et al., 2020; Coffel et al., 2018; Fischer et al., 2007; Miralles et al., 2019).

One topic that has received less attention is the very warmest surface temperatures on Earth. While it is known that parts of the tropics are approaching thermal habitability limits (Sherwood & Huber, 2010; Raymond et al., 2020; Coffel et al., 2018), the upper bound of these temperatures has not been investigated in detail. Changes in maximum temperatures reveal how the tail of the global temperature distribution evolves and serve as a vivid indicator of a warming climate. Here, we use state-of-the-art climate models to examine how these warmest temperatures are projected to change over the 21st century.

Earth's warmest surface temperatures are typically found in subtropical deserts, where daily-mean temperatures can reach up to ∼44°C and daily-maximum temperatures the mid-50s°C ((Mildrexler et al., 2011; Stachelski, 2013), Figure 1). Surface, or skin, temperatures can be even warmer, with satellites measuring values as high as 80.8°C (Zhao et al., 2021). Subtropical deserts have all the ingredients needed to generate extreme temperatures: high levels of insolation, extreme surface dryness and the relatively low surface emissivity of sand. Low moisture levels are crucial for generating high temperatures, as latent heat fluxes can efficiently cool the surface (Seneviratne et al., 2010), while sand's relatively weak emissivity (∼0.85-0.9, compared to >0.9 for most other surface types, e.g., (Ogawa & Schmugge, 2004; Jin & Liang, 2006; Huang et al., 2016; Chen et al., 2019)) requires warmer temperatures to achieve the same emission as other surfaces.

Measuring these extreme temperatures is difficult because they often occur in remote deserts (e.g., the Sahara, Gobi, Sonoran, and Lut), where station coverage is sparse. Satellites provide continuous monitoring of skin temperature but not near-surface air temperature; the 80.8°C estimate is therefore based on a relatively short satellite record (July 2002–December 2019; (Zhao et al., 2021)). Historical air-temperature records include disputed measurements of 56.7°C (Death Valley, 1913) and 57.8°C ('Aziziya, 1922) (Spencer et al., 2026; El Fadli et al., 2013), while more recent values near 54°C have been reported in Kuwait and Pakistan (Merlone et al., 2019; World Meteorological Organization, 2019).

Studying extreme temperatures also poses methodological challenges because they are, by definition, rare events, making it difficult to obtain robust statistics for trend detection and model evaluation. Moreover, focusing on these extremes adds an additional dimension of complexity: the locations where maximum temperatures occur may themselves shift, requiring models to correctly capture spatial changes as well as the physical drivers of temperature extremes.

To address these issues, we first evaluate the ability of state-of-the-art climate models to simulate the observed spatial pattern of historical extreme land heat. We then examine how strongly the hottest land temperatures amplify relative to tropical-mean land warming under both transient (SSP2-4.5) and equilibrium (abrupt-4×CO_2) forcing, and where the world's hottest land-temperature hotspots occur. Finally, we quantify changes in exceedance frequency using a fixed-threshold "record-like" metric based on the annual hottest-land temperature, and estimate how exceedance likelihood scales with tropical land warming in pooled fits. Together, these analyses clarify where record-setting heat is most likely to recur and how rapidly record-like extremes become more common as the climate warms.

2 Methods

2.1 Datasets and preprocessing

We analyze daily near-surface air temperature extremes in observations and CMIP6 models. Observed daily maximum near-surface air temperature (`tasmax`) is taken from Berkeley Earth (Rohde & Hausfather, 2020). Unless otherwise noted, historical evaluation uses 1980–2014 to match the satellite-era reference period and to maximize overlap with the CMIP6 `historical` experiment. As an observational-product sensitivity check, we also compare the Berkeley Earth historical $p99(T_{\max})$ field with an ERA5-derived $p99(T_{\max})$ field over 1980–2014 (Fig. S7); the results are reported in Section 3.1.

Model simulations are drawn from CMIP6 (Eyring et al., 2016). We define a core ensemble of $N_{\text{core}} = 21$ models (Table 1; Table S1), and this same set is used for the historical skill metrics (Table 1) and for the primary multi-model analyses in Fig. 1. Model variants are listed in Table S1; where multiple variants were available, we used a single member per model to avoid weighting any modeling center more than once. For projections, we analyze SSP2-4.5 simulations over 2015–2100. Future projections are not bias-corrected; historical evaluation is used to contextualize model fidelity, but projected changes are interpreted as CMIP6 ensemble behavior rather than as a bias-corrected impacts product.

Several analyses require different underlying daily variables and/or complete time coverage, so the effective sample size varies by diagnostic. In Fig. 2 we analyze three indices derived from daily temperature: the tropical land-mean temperature (T_{TL}), the daily-mean temperature on the hottest day (T_{DM}), and the annual maximum daily maximum temperature (T_{DX}). Because availability of the processed SSP2-4.5 time series differs by index, the projection sample sizes are $N_{TL} = 24$, $N_{DM} = 21$, and $N_{DX} = 17$ (Table S1), and we report the corresponding N in the figure legend. For equilibrium warming (Fig. 2c), we further restrict to models with both abrupt-4×CO_2 and the corresponding piControl simulations available for the index; the final sample size is reported in the panel.

In Fig. 3, the record-exceedance analysis requires both daily `tasmax` and daily `tas` over `historical` and SSP2-4.5 in order to define the fixed historical record and relate exceedances to a warming predictor. This yields $N = 17$ models (Table S1); any additional model-specific exclusions (e.g., insufficient years in the regression fit window) are noted alongside the analysis output and do not alter the data-availability subset.

All temperatures are converted to degrees Celsius (°C; Kelvin minus 273.15). Unless otherwise stated, analyses are restricted to land grid boxes within 45°S–45°N. Land points are identified using a land mask on each model grid. A land grid box is included when its land-area fraction exceeds 50%, using the model-specific land-fraction field where available. When comparing model fields with observations, fields are put on a common latitude–longitude grid using bilinear interpolation prior to computing spatial metrics; the interpolation direction is specified in each figure caption or table note.

We use the following temperature indices throughout:

1. T_{TL}: annual-mean near-surface temperature averaged over tropical land (45°S–45°N). We denote the tropical-land spatial mean with an overline, $\overline{T}_{TL}$, and define tropical land warming relative to a baseline period as

$$\Delta\overline{T}_{TL} = \overline{T}_{TL} - \overline{T}_{TL}^{\mathrm{ref}},$$

   where $\overline{T}_{TL}^{\mathrm{ref}}$ is the mean over 2005–2014 unless stated otherwise.
2. T_{DM}: annual maximum of daily-mean near-surface air temperature (computed from daily `tas`).
3. T_{DX}: annual maximum of daily-maximum near-surface air temperature (computed from daily `tasmax`).
4. T_{X5D}: annual maximum 5-day mean daily maximum temperature, used as a compact sensitivity test for multi-day heat exposure (Fig. S8).
5. T^*: annual hottest-land temperature, defined as the annual maximum of the daily maximum over all land grid boxes within 45°S–45°N. In practice, we compute the spatial maximum of daily `tasmax` over land and then take the annual maximum of that daily series.

2.2 Historical model evaluation

We evaluate historical model fidelity using the climatological 99th percentile of daily maximum temperature, denoted $T_{DX,99}$, computed from daily `tasmax` over 1980–2014. For each model and for Berkeley Earth, $T_{DX,99}$ is computed at each land grid box and then compared on the common Berkeley Earth grid.

We summarize spatial agreement using three complementary pattern metrics computed over all land grid boxes with latitude-area weights. Let M_i and O_i denote the model and observed $T_{DX,99}$ at grid box i, and let $w_i = \cos(\phi_i)$ be the latitude weight at latitude ϕ_i, normalized by the sum of weights over valid land grid boxes. Weighted spatial means are denoted by angle brackets, e.g., $\langle M \rangle_w = \sum_i w_i M_i / \sum_i w_i$, and primes denote anomalies relative to the weighted spatial mean:

- **Area-weighted spatial bias**:

$$\mathrm{Bias}_w = \frac{\sum_i w_i (M_i - O_i)}{\sum_i w_i}.$$

- **Area-weighted centered RMSE (cRMSE)**:

$$\mathrm{cRMSE}_w = \sqrt{\frac{\sum_i w_i \left[(M_i - O_i) - \mathrm{Bias}_w\right]^2}{\sum_i w_i}}.$$

- **Area-weighted spatial pattern** R^2 (squared weighted spatial correlation):

$$R_w^2 = \frac{\mathrm{Cov}_w(M', O')^2}{\mathrm{Var}_w(M')\ \mathrm{Var}_w(O')},$$

  where covariance and variance are computed using the same latitude weights.

Together, these metrics separate (i) systematic warm/cool offsets (Bias), (ii) mismatch in the amplitude of the spatial pattern (cRMSE), and (iii) agreement in pattern structure (R^2). Model skill values are reported in Table 1 and are used to contextualize inter-model spread in projections.

To evaluate whether the reduced future-projection subset is historically representative of the broader core ensemble, we compare the intended 17-model future subset with the full 21-model core ensemble using these historical $p99(T_{\max})$ skill metrics (Fig. S9). The subset is fully represented in the skill table and has similar historical performance to the full core ensemble: mean cRMSE, mean absolute bias, and mean spatial R^2 are 2.55 °C, 1.23 °C, and 0.851 for the 17-model subset, compared with 2.59 °C, 1.14 °C, and 0.842 for the full core ensemble.

2.3 GLM scaling of record-like exceedance likelihood

To quantify how the likelihood of a record-like hottest-land year changes with warming, we use a fixed-threshold exceedance definition based on the historical period. We define a historical benchmark

$$T^*_{\mathrm{hist}} = \max_{y \in [1980, 2014]} T^*(y),$$

where y indicates year, and an annual exceedance indicator

$$I(y) = \begin{cases} 1, & T^*(y) > T^*_{\mathrm{hist}}, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

This fixed threshold yields an impacts-style interpretation (how often future climate exceeds a late-20th-century record) and avoids the progressively tightening target that occurs under running-maximum "new record" definitions.

We relate exceedance probability to tropical land warming using a pooled logistic model fit over 2015–2100:

$$\operatorname{logit} P[I(y) = 1] = \alpha_0 + \alpha_1 \, \Delta\overline{T}_{TL}(y), \tag{2}$$

where $\Delta\overline{T}_{TL}(y)$ is tropical land-mean `tas` anomaly relative to 2005–2014. The slope α_1 is reported as an odds ratio $\mathrm{OR} = \exp(\alpha_1)$ per 1°C of tropical land warming. To stabilize the pooled fit given inter-model heterogeneity, we apply ridge regularization (L2 penalty) and estimate uncertainty by bootstrapping the model ensemble (resampling models with replacement and refitting the pooled model), reporting the 10th–90th percentile envelope of predicted probabilities.

For interpretability, we also summarize the expected number of exceedance-years over a future window (e.g., 2026–2100) by summing fitted annual probabilities, $E = \sum_y P[I(y) = 1]$, computed separately for each model's warming trajectory and then summarized across models.

3 Results

3.1 Model performance for extreme land surface temperatures

Table 1 reports these metrics for the 21-model core ensemble; the final two rows show the multi-model mean and multi-model median. We use the median as a more robust benchmark because a small number of outliers can skew mean-based summaries.

Model	**cRMSE (°C)**	**Mean bias (°C)**	**Spatial R^2**
ACCESS-CM2	2.44	-0.28	0.851
ACCESS-ESM1-5	3.63	1.35	0.686
AWI-CM-1-1-MR	2.55	0.53	0.850
CAMS-CSM1-0	2.54	1.80	0.849
CMCC-ESM2	2.30	0.67	0.863
CNRM-CM6-1	2.53	-0.78	0.874
CNRM-CM6-1-HR	2.97	-1.57	0.841
CNRM-ESM2-1	2.49	-0.05	0.878
CanESM5	3.76	3.19	0.785
EC-Earth3	2.31	0.24	0.882
EC-Earth3-CC	2.33	0.89	0.875
EC-Earth3-Veg	2.31	0.25	0.879
EC-Earth3-Veg-LR	2.29	0.09	0.877
GFDL-CM4	2.24	-1.14	0.903
GFDL-ESM4	2.26	-0.47	0.896
INM-CM4-8	3.26	0.63	0.766
INM-CM5-0	2.92	0.43	0.806
IPSL-CM6A-LR	2.76	-2.90	0.847
NorESM2-LM	2.49	0.52	0.809
NorESM2-MM	2.31	0.19	0.839
TaiESM1	2.30	-6.47	0.873
Multi-model mean	**2.62**	**-0.14**	**0.844**
Multi-model median	**2.49**	**0.24**	**0.851**

Table 1. Performance of 21 CMIP6 models in simulating the spatial pattern of extreme land heat. Centered RMSE (cRMSE; °C), mean bias (°C), and spatial R^2 are computed by comparing the historical $p99(T_{\max})$ field (1980–2014) with Berkeley Earth over land grid cells using latitude-area weights. The final two rows show the multi-model mean and multi-model median.

There is substantial inter-model spread across all three metrics. Several models show relatively high skill, with cRMSE ∼2–3 °C and high spatial correlations ($R^2 \geq 0.85$), indicating they capture the large-scale pattern of extreme heat reasonably well. Other models exhibit larger systematic biases (both warm and cold) and weaker spatial correlations. Where local biases are largest, they often occur in semi-arid and arid regions, suggesting that errors in land-surface processes and land–atmosphere coupling may contribute. Because these metrics are evaluated on native (typically coarse) model grids, they reflect grid-box mean extremes; local extremes could differ at higher resolution. To document the full range of modeled spatial patterns and biases, we provide a supplementary multi-panel "stamp" figure showing $p99(T_{\mathrm{max}})$ on each model's native grid and the corresponding bias relative to Berkeley Earth after regridding (Fig. S4). We also summarize overall performance with a composite skill score that combines cRMSE, |bias|, and spatial R^2 (Fig. S5).

Because observational uncertainty is an important limitation in sparsely observed hot regions, we also compare Berkeley Earth with ERA5 for historical $p99(T_{\mathrm{max}})$ over 1980–2014 (Fig. S7). After applying the Berkeley Earth climatology correction, the two products agree closely over land within 45°S–45°N, with area-weighted cRMSE of 0.20 °C, mean bias of 0.22 °C, and spatial $R^2 = 0.999$. These observational-product differences are much smaller than the inter-model spread in Table 1, supporting the use of Berkeley Earth as the primary benchmark while retaining observational uncertainty as an explicit caveat.

By contrast, several models exhibit larger systematic errors. CanESM5 and ACCESS-ESM1-5 show pronounced warm biases, whereas IPSL-CM6A-LR and especially TaiESM1 are strongly cold-biased relative to Berkeley Earth. These models also tend to have weaker spatial correlations.

This is also reflected in the composite model skill score (Fig. S5), which combines cRMSE, |bias|, and spatial R^2 into a single ranking. In that summary, GFDL-ESM4 and several EC-Earth3 variants cluster among the highest-skill models, whereas CanESM5 and ACCESS-ESM1-5 rank near the bottom, consistent with their larger mean biases and weaker spatial agreement.

3.2 Geography of Record-Setting Heat

Before quantifying trends, we examine where the most extreme land temperatures occur in observations and in model projections. Record temperatures may also reflect shifts in the *location* of the global land maximum if the hottest regions broaden or migrate under warming.

We diagnose hotspot locations using the hotspot-frequency maps in Fig. 1. For each year, we identify the single land grid cell with the global maximum of annual T_{max} and count how often each grid cell attains this maximum. In Berkeley Earth (1980–2014), hotspots cluster over the Middle East/Persian Gulf and adjacent northwestern South Asia, with fewer events over North Africa and Australia, indicating strong spatial persistence. Regions that are broadly hot in $p99(T_{\mathrm{max}})$ do not necessarily dominate hotspot frequency because this metric isolates the single hottest grid cell each year. Figure S2 shows hotspot-frequency maps for each core model over the historical period.

The middle and bottom right rows of Fig. 1 illustrate projected behavior for two models with contrasting historical skill. GFDL-ESM4 maintains a concentrated hotspot distribution dominated by the Middle East and nearby South Asia, whereas CanESM5 produces a more dispersed pattern, with frequent maxima over western Australia and increased occurrence over the western Sahara. Across the full ensemble, comparing Fig. S2 (historical) and Fig. S3 (SSP2-4.5) shows that dominant hotspot regions change little: the global land maximum remains anchored in a small set of subtropical arid and semi-arid regions, with differences mainly in how dispersed the hotspots become. Consistent with the historical evaluation (Table 1), higher-skill models tend to place hotspots closer to the observed Middle East–South Asia corridor, whereas lower-skill models more often introduce prominent maxima in regions that are less favored in the observations.

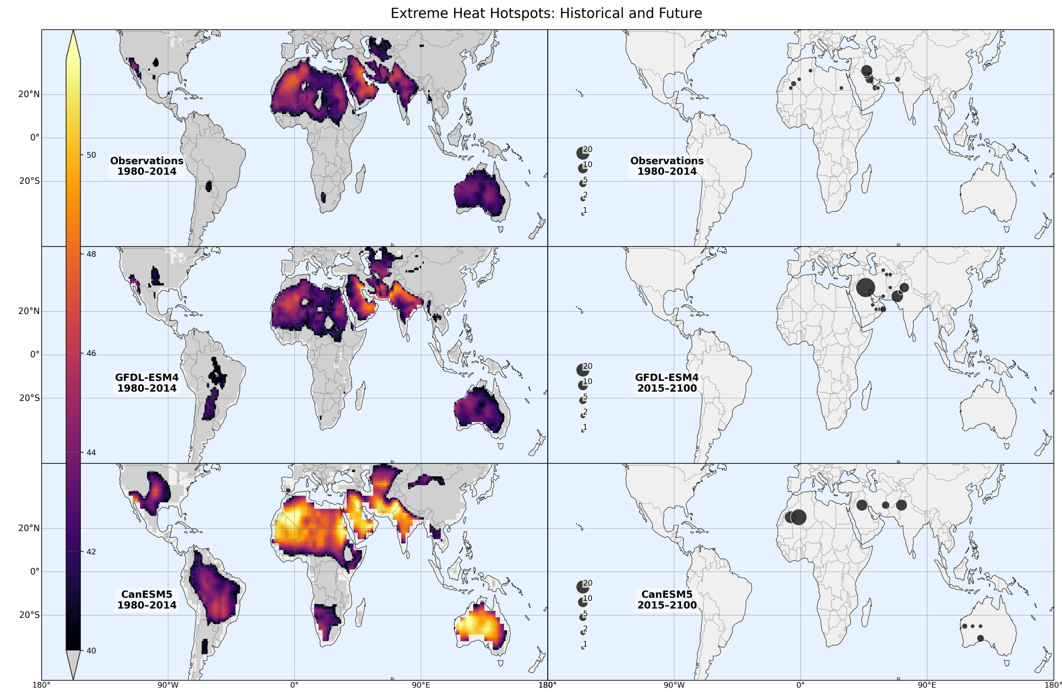

Figure 1. Observed and simulated spatial patterns of extreme land heat. Left column: $p99(T_{\max})$ (99th percentile of daily maximum temperature; shading) with the 40 °C isotherm contoured in black. Top row shows Berkeley Earth over 1980–2014; middle row shows GFDL-ESM4 (high skill in Table 1) and bottom row shows CanESM5 (strong warm bias) over 1980–2014. Right column: hotspot frequency, defined as the number of years in which a grid cell attains the *global land maximum* of annual $T_{\max}$ (marker size; 35 years for observations, 86 years for projections). Top row is for observations and bottom rows show SSP2-4.5 projections over 2015–2100 for each example model.

3.3 Transient and equilibrium trends of tropical land extremes

Figure 2 summarizes how tropical land temperature responds to warming in the mean state and in the far tail. We consider three annual indices: tropical land-mean temperature T_{TL}, the daily-mean temperature on the hottest day each year T_{DM}, and the annual maximum daily-maximum temperature T_{DX} (Methods). Panel (a) shows SSP2-4.5 projections (2015–2100) as anomalies relative to 2005–2014. For additional context on inter-model spread, Fig. S1 shows the corresponding per-model trajectories and model-by-model trend estimates for a common-model set.

For the abrupt-4×CO_2 panel, we interpret the final-30-year difference from piControl as a practical late-simulation response estimate rather than proof that every model has reached complete equilibrium.

All three indices warm substantially over the 21st century. In the multi-model mean, T_{TL} increases by about ∼2.5 °C by 2100, whereas the hottest-day metrics increase by ∼3–3.5 °C. Interannual variability is larger for T_{DM} and T_{DX} than for T_{TL}, but the forced warming signal is clear.

Sample sizes differ across indices because complete processed time series are not available for all models (Table S1): T_{TL} uses $N = 24$ models, T_{DM} uses $N = 21$, and T_{DX} uses $N = 17$. These differences primarily reflect data availability requirements (daily-

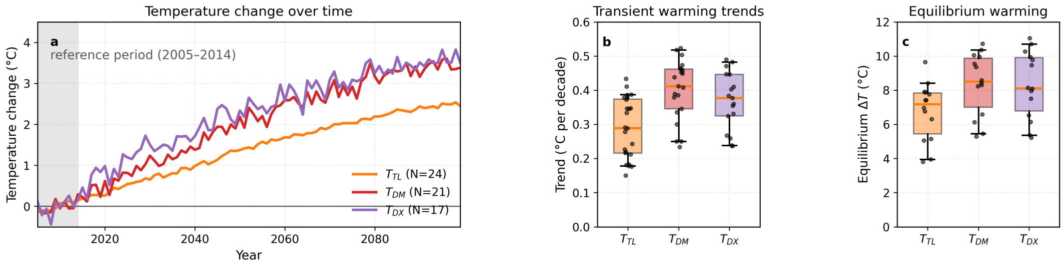

Figure 2. Multi-model response of tropical land temperatures in CMIP6 (all changes in °C). (a) Annual evolution of tropical land-mean temperature T_{TL}, the daily-mean temperature on the hottest day each year T_{DM}, and the annual maximum daily-maximum temperature T_{DX} under SSP2-4.5, shown as anomalies relative to 2005–2014. Sample sizes differ across indices due to data availability and completeness (Table S1): $N = 24$ for T_{TL}, $N = 21$ for T_{DM}, and $N = 17$ for T_{DX}. (b) Transient warming trends for each index over 2015–2100 in °C per decade. (c) Equilibrium warming in abrupt-4×CO_2 experiments, defined as the difference between the late-time mean of abrupt-4×CO_2 (last 30 years) and the corresponding piControl (°C); this panel uses the subset of models with both experiments available for the relevant index.

mean temperature for T_{TL} and T_{DM} versus daily-maximum temperature for T_{DX}), as well as completeness of the SSP2-4.5 record. As a multi-day heat sensitivity test, we also compute T_{X5D}, the annual maximum 5-day mean daily maximum temperature. In the 17-model common set, T_{X5D} and T_{DX} have nearly identical projected trends: 0.372 and 0.369 °C decade^{-1}, respectively, with a mean difference of only 0.002 °C decade^{-1} (Fig. S8). All calculations are performed on native model grids, so the plotted values represent grid-box mean extremes; local extremes could differ at higher spatial resolution.

Panels (b) and (c) compare transient and equilibrium responses. Panel (b) shows per-model warming trends over 2015–2100, while panel (c) shows the equilibrium response in abrupt-4×CO_2, defined as the difference between the late-time mean of abrupt-4×CO_2 and the corresponding piControl (Methods). In both frameworks, extremes warm faster than the mean: T_{DM} and T_{DX} exhibit larger trends and larger equilibrium warming than T_{TL}, while T_{DM} and T_{DX} are similar to each other, suggesting broadly consistent mechanisms controlling the hottest-day mean and the annual maximum of daily maxima.

The ratio of extreme to mean warming implies that the hottest-day indices increase roughly 30–40% faster than T_{TL}, consistent with theoretical expectations for land extremes under warming associated with moisture limitation and land–atmosphere coupling (Byrne, 2021). This scaling provides a useful baseline for interpreting record-breaking behavior: if record temperatures track the far tail, then record magnitudes should increase faster than the tropical land mean by a comparable factor. We therefore interpret the projected record-like exceedance behavior in the context of the historical pattern-skill diagnostics in Table 1.

3.4 Scaling of record-like extreme heat with tropical land warming

A key implication of percentile-dependent warming over land is that exceedances of a *fixed* historical extreme should become increasingly common as warming accumulates. In the tropical band, our CMIP6 ensemble shows clear tail amplification over land consistent with the theoretical expectation summarized by Byrne (2021): the scaling factor $S_x = \Delta T^x / \Delta \overline{T}$ rises above unity in the upper tail (Fig. S6), reaching $S_{99} \approx 1.27$ over land compared with $S_{99} \approx 1.08$ over ocean. This amplified warming of high percentiles motivates an impacts-style "record-like" metric that asks how often future climate surpasses a familiar late-20th-century benchmark.

We define $T^*(y)$ as the annual hottest-land temperature in year y, computed as the annual maximum of the daily land maximum of `tasmax` (land within 45°S–45°N). We then define a fixed historical benchmark from the 1980–2014 period and an exceedance indicator,

$$I(y) = \begin{cases} 1, & T^*(y) > T^*_{\mathrm{hist}}, \\ 0, & \text{otherwise}, \end{cases} \qquad T^*_{\mathrm{hist}} = \max_{y \in [1980, 2014]} T^*(y). \tag{3}$$

Unlike a "new record" (running-maximum) definition, this fixed-threshold formulation does not tighten the target over time; instead it directly measures the increasing frequency of exceeding a historical record-like extreme.

To quantify how exceedance likelihood scales with warming, we relate $I(y)$ to tropical land-mean warming $\Delta\overline{T}_{\mathrm{TL}}(y)$ (45°S–45°N land mean `tas`, baseline 2005–2014) using a pooled ridge-regularized logistic regression over 2015–2100. The pooled fit yields an odds ratio of OR ≈ 1.35 per $1\,°\mathrm{C}_{\mathrm{TL}}$, implying that each additional degree of tropical land warming increases the odds of exceeding the 1980–2014 hottest-land benchmark by $\sim$35%. In other words, tail amplification documented in Supporting Information Fig. S6 translates into a rapid increase in the frequency with which the climate exceeds a fixed historical extreme. Summing fitted annual exceedance probabilities over 2026–2100 gives an expected $\sim$42 exceedance-years in the multi-model mean (out of 75), indicating that surpassing the late-20th-century hottest-land record becomes a common outcome later in the century under SSP2-4.5.

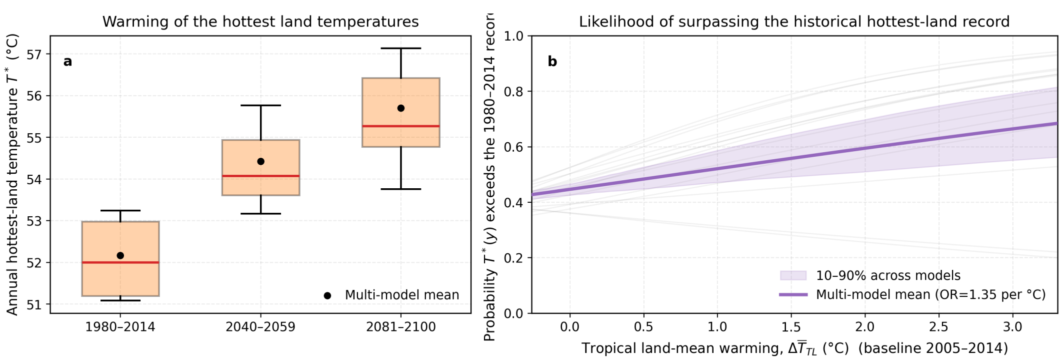

Figure 3. Fixed-threshold exceedance of the historical hottest-land benchmark. (a) Model spread of the annual hottest-land temperature T^* across time slices. (b) Probability that $T^*(y)$ exceeds the historical benchmark $T^*_{\mathrm{hist}} = \max_{1980\text{–}2014} T^*(y)$ (Eq. 1) as a function of tropical land warming $\Delta\overline{T}_{\mathrm{TL}}$ relative to 2005–2014. Thin curves show individual-model ridge-logit fits; the thick curve shows the pooled multi-model ridge-logit fit (2015–2100). Shading indicates the 10th–90th percentile range across bootstrap resamples of the model ensemble.

4 Conclusion

CMIP6 models project that the hottest land temperatures intensify faster than mean tropical land warming, but the magnitude and spatial fidelity of this amplification vary across models (Table 1). In several cases, strong spatial agreement coexists with a sizable mean bias, so both pattern skill and bias magnitude matter when interpreting projections.

Hotspot locations are relatively stable: the annual global land maximum most often occurs within a subtropical corridor spanning North Africa, the Arabian Peninsula/Middle East, and South Asia (Fig. 1; Figs. S2–S3). Across the ensemble, extremes warm more than the mean: T_{DM} and T_{DX} increase faster than T_{TL} in both transient and equilibrium frameworks (Fig. 2). Consistent with this amplification, exceedances of a fixed 1980–2014 "record-like" benchmark become increasingly common under SSP2-4.5, with pooled fits indicating ∼35% higher odds per 1 °C of tropical land warming (Fig. 3). The sensitivity analyses indicate that this interpretation is robust to using a multi-day TX5D heat metric and that the intended 17-model future subset has historical $p99(T_{\max})$ skill comparable to the full 21-model core ensemble. We focus on daytime maximum-temperature extremes; nighttime minimum temperature, humidity, and heat-stress metrics involve additional boundary-layer, cloud, and moisture processes and remain important directions for future work.

Acknowledgments

We thank the editor and anonymous referee for constructive comments that helped improve the manuscript, including the suggestion to examine a multi-day heat sensitivity analysis. We thank Berkeley Earth, ECMWF (ERA5), and the CMIP6 modeling groups for datasets. AKW and AJM were partly supported by National Science Foundation grants OCE-2022868 and OCE-2224726. This study forms part of the Ph.D. dissertation of AKW at SIO–UCSD.

Conflict of Interest Statement

The authors have no conflicts of interest to disclose.

Open Research

CMIP6 model output used in this study is available through the Earth System Grid Federation (ESGF) (Earth System Grid Federation, 2026). Berkeley Earth surface temperature data are available from Berkeley Earth (Berkeley Earth, 2026). ERA5 reanalysis data were obtained from the Copernicus Climate Data Store (Copernicus Climate Change Service, 2023). Code and analysis workflows and derived materials used in this study are archived permanently on Zenodo (Wilson et al., 2026a). The GitHub repository used during peer review is retained as a development copy (Wilson et al., 2026b). Figures and analyses were produced using Python software including `xarray` and `Matplotlib` (Hoyer & Hamman, 2017; Hunter, 2007).

References

Berkeley Earth. (2026). *[Dataset] berkeley earth surface temperature data.* `http://berkeleyearth.org/data/`. Berkeley Earth. (Accessed 17 June 2026)

Byrne, M. P. (2021). Amplified warming of extreme temperatures over tropical land. *Nature Geoscience*, *14*(11), 837–841. doi: 10.1038/s41561-021-00828-8

Chen, Y.-H., Huang, X. L., Chen, X.-H., & Flanner, M. G. (2019). The effects of surface longwave spectral emissivity on atmospheric circulation and convection over the sahara and sahel. *Journal of Climate*, *32*(15), 4873–4890. doi: 10.1175/JCLI-D-18-0615.1

Coffel, E. D., Horton, R. M., & de Sherbinin, A. (2018). Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century. *Environmental Research Letters*, *13*(1), 014001. doi: 10.1088/1748-9326/aaa00e

Copernicus Climate Change Service. (2023). *[Dataset] era5 hourly data on single lev-*

els from 1940 to present. https://doi.org/10.24381/cds.adbb2d47. Copernicus Climate Data Store. doi: 10.24381/cds.adbb2d47

Earth System Grid Federation. (2026). *[Dataset] cmip6 model output.* https://esgf-node.llnl.gov/search/cmip6/. Earth System Grid Federation. (Accessed 17 June 2026)

El Fadli, K. I., Cerveny, R. S., Burt, C. C., Eden, P., Parker, D. E., Brunet, M., & Peterson, T. C. (2013). World meteorological organization assessment of the purported world record 58 °C temperature extreme at El Azizia, Libya (13 september 1922). *Bulletin of the American Meteorological Society*, *94*(2), 199–204. doi: 10.1175/BAMS-D-12-00093.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., & Schär, C. (2007). Soil moisture–atmosphere interactions during the 2003 european summer heat wave. *Journal of Climate*, *20*(20), 5081–5099. doi: 10.1175/JCLI4288.1

Hoyer, S., & Hamman, J. (2017). [Software] xarray: N-d labeled arrays and datasets in python. *Journal of Open Research Software*, *5*(1), 10. doi: 10.5334/jors.148

Huang, X., Chen, X., Zhou, D. K., & Liu, X. (2016). An observationally based global band-by-band surface emissivity dataset for climate and weather simulations. *Journal of the Atmospheric Sciences*, *73*(9), 3541–3555. doi: 10.1175/JAS-D-15-0355.1

Hunter, J. D. (2007). [Software] matplotlib: A 2d graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

Jin, M., & Liang, S. (2006). An improved land surface emissivity parameter for land surface models using global remote sensing observations. *Journal of Climate*, *19*(12), 2867–2881. doi: 10.1175/JCLI3720.1

Merlone, A., Al-Dashti, H., Faisal, N., Cerveny, R. S., et al. (2019). Temperature extreme records: World meteorological organization metrology and verification of the 54.0 °C observations in Kuwait and Pakistan. *International Journal of Climatology*, *39*(14), 5154–5169. doi: 10.1002/joc.6132

Mildrexler, D. J., Zhao, M., & Running, S. W. (2011). Satellite finds highest land skin temperatures on earth. *Bulletin of the American Meteorological Society*, *92*(7), 855–860. doi: 10.1175/2011BAMS3067.1

Miralles, D. G., Gentine, P., Seneviratne, S. I., & Teuling, A. J. (2019). Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Annals of the New York Academy of Sciences*, *1436*(1), 19–35. doi: 10.1111/nyas.13912

Ogawa, K., & Schmugge, T. (2004). Mapping surface broadband emissivity of the sahara desert using ASTER and MODIS data. *Earth Interactions*, *8*(7), 1–14. doi: 10.1175/1087-3562(2004)008⟨0001:MSBEOT⟩2.0.CO;2

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*, 3357. doi: 10.1038/s41467-020-16970-7

Raymond, C., Matthews, T., & Horton, R. M. (2020). The emergence of heat and humidity too severe for human tolerance. *Science Advances*, *6*(19), eaaw1838. doi: 10.1126/sciadv.aaw1838

Rohde, R. A., & Hausfather, Z. (2020). The berkeley earth land/ocean temperature record. *Earth System Science Data*, *12*, 3469–3479. doi: 10.5194/essd-12-3469-2020

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., . . . Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3-4), 125–161. doi: 10.1016/j.earscirev.2010.02.004

Seneviratne, S. I., Zhang, X., Adnan, M., et al. (2021). Weather and climate extreme events in a changing climate. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change.* Cambridge University Press. doi: 10.1017/9781009157896.013

Sherwood, S. C., & Huber, M. (2010). An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, *107*(21), 9552–9555. doi: 10.1073/pnas.0913352107

Spencer, R. W., Christy, J. R., & Reid, W. T. (2026). Death valley illusion: Evidence against the 134 °F world record. *Bulletin of the American Meteorological Society*, *107*(1). doi: 10.1175/BAMS-D-24-0313.1

Stachelski, S. (2013). *A century of weather in death valley national park: 1911–2011* (Western Region Technical Memorandum No. NWS WR-289). NOAA National Weather Service, Western Region.

Wilson, A., Lutsko, N., Faggiani Dias, D., & Miller, A. J. (2026a). *[Dataset] projections of earth's hottest surface temperatures in cmip6.* `https://doi.org/10.5281/zenodo.20726359`. Zenodo. doi: 10.5281/zenodo.20726359

Wilson, A., Lutsko, N., Faggiani Dias, D., & Miller, A. J. (2026b). *[Software] extreme-heat-cmip6 development repository.* `https://github.com/a4wilson/extreme-heat-cmip6`. GitHub. (Permanent archive available from Zenodo at doi:10.5281/zenodo.20726359)

World Meteorological Organization. (2019, June). *Wmo verifies 3rd and 4th hottest temperature recorded on earth.* Press release. Retrieved from `https://wmo.int/news/media-centre/wmo-verifies-3rd-and-4th-hottest-temperature-recorded-earth` (18 June 2019)

Zhao, Y., Norouzi, H., Azarderakhsh, M., & AghaKouchak, A. (2021). Global patterns of hottest, coldest, and extreme diurnal variability on earth. *Bulletin of the American Meteorological Society*, *102*(9), E1672–E1681. doi: 10.1175/BAMS-D-20-0325.1