PNAS RESEARCH ARTICLE | EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES OPEN ACCESS

The pace of change of summertime temperature extremes

Karen A. McKinnon[a,b,c,1], Isla R. Simpson[d], and A. Park Williams[e]

Affiliations are included on p. 9.

Edited by Kerry Emanuel, Massachusetts Institute of Technology, New Harbor, ME; received March 25, 2024; accepted August 30, 2024

Summer temperature extremes can have large impacts on humans and the biosphere, and an increase in heat extremes is one of the most visible symptoms of climate change. Multiple mechanisms have been proposed that would predict faster warming of heat extremes than typical summer days, but it is unclear whether this is occurring. Here, we show that, in both observations and historical climate model simulations, the hottest summer days have warmed at the same pace as the median globally, in each hemisphere, and in the tropics from 1959 to 2023. In contrast, the coldest summer days have warmed more slowly than the median in the global average, a signal that is not simulated in any of 262 simulations across 28 CMIP6 models. The observed stretching of the cold tail indicates that observed summertime temperatures have become more variable despite the lack of hot day amplification. The interannual variability and trend in the warming of both hot and cold extremes compared to the median can be explained from a surface energy balance perspective based on changes in net surface radiation and evaporative fraction. Tropical hot day amplification is projected to emerge in the future (2024–2099, SSP3-7.0 scenario), while Northern Hemisphere heat extremes are expected to continue to follow the median.

heat extremes | climate change | temperature variability

Significance

An increase in summertime temperature extremes is one of the most tangible components of climate change, but are they warming faster than typical temperatures at hemispheric and global scales? Here, we show that historical heat extremes warmed at the same pace as the median from 1959 to 2023, although the range of summertime temperatures widened due to slower warming of the cold tail. Trends in the extremes are consistent with trends in surface radiation and evaporative fraction, suggesting an important control of the land surface on changes in temperature distributions. Climate models agree on a lack of historical hot day amplification but miss the cold tail stretching, potentially indicating a misrepresentation of processes that control trends on cool summer days.

The question of whether temperature extremes have warmed faster than typical summer days, or should be expected to do so, gains new urgency each summer as record-breaking heat waves occur around the world (e.g., refs. 1–5). While an increase in heat extremes, including record-breaking events, is an expected result of increasing mean temperatures (6, 7), the magnitude and frequency of recent extremes raise questions about the nature of changes in summertime temperature distributions. In particular, can recent events be explained by a simple shift in the distribution combined with unpredictable sampling of weather variability, or are we now seeing evidence that extremes are warming faster than typical summer temperatures?

Multiple mechanisms have been proposed that would predict greater warming of hot days. In the midlatitudes, concurrent summertime heat extremes are commonly associated with amplified quasi-stationary planetary waves (8–11), and it has been hypothesized that the probability of these events is increasing, possibly due to a reduced equator-to-pole temperature gradient (12–15). These conclusions, however, remain debated in the literature, and observed trends are difficult to identify amid internal variability (16–21). In the tropics, climate models tend to project that the hottest summer days will warm faster than the average (22–25), which has been explained from both a dynamics and land surface perspective as a consequence of hot days being climatologically drier (25, 26). Finally, reductions in soil moisture on hot days, and the associated increases in sensible heating, have been used to explain amplified warming of heat extremes in climate model simulations in various regions around the globe (23, 27–31).

Despite this large body of work, it remains unclear whether we are observing amplified warming of hot extremes. We a priori expect the signal, if present, to be relatively small because the primary drivers of historical trends in observed (32, 33) and projected (34, 35) heat extremes are trends in the mean or the median, although the signal is nevertheless important to identify given the nonlinear impact of high temperatures (36). Prior work has primarily asked whether extremes are warming at a different rate from the middle of the distribution at the local level (22, 31, 32, 37–39), and these local trends will have a large contribution from unforced internal variability, even on multidecadal timescales. The combination of a small signal and large internal variability precludes clear conclusions about differential changes in temperature extremes from the current literature and motivates the spatially aggregated approach ultimately taken here.

In contrast to the focus on heat extremes, there has been little focus in the literature on mechanisms causing differential changes in the cold tail of summer temperatures, likely

Author contributions: K.A.M. designed research; K.A.M. and I.R.S. performed research; K.A.M., I.R.S., and A.P.W. analyzed data; I.R.S. contributed to data acquisition; and K.A.M., I.R.S., and A.P.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: kmckinnon@ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2406143121/-/DCSupplemental.

Published October 7, 2024.

due to their smaller human impacts. Nevertheless, quantifying changes in cold extremes compared to the median provides a more complete picture of changes in the spread of summertime temperature and is key to interpreting results that focus on trends in daily temperature variance (e.g., ref. 40), which reflects changes in both tails.

Here, we address the questions of whether large-scale changes in the shape of warm season continental temperature distributions are evident in the historical record, what causes differential changes in the distributions, and whether climate models are broadly consistent with observations.

A Framework for Quantifying Large-Scale Changes in Extremes

We are interested in understanding whether there are significant trends in the tails of daily maximum warm season temperature anomalies that cannot be explained by changes in the median. To do so, at each location, we calculate daily temperature deviations by first removing the climatological seasonal cycle, and then removing the sample median of each warm season's deseasonalized daily values. Long-term trends in the extreme values of the deviations then reflect a stretching or compression of the tails of the distribution compared to the middle. We take this nonparametric approach to avoid the need to specify the correct parametric form for the forced time evolution of the median at each location, which is not generally known. This statistical framework reflects the testing of a specific physical hypothesis: by definition, 50% of warm season days will be warmer than the median, and 50% will be cooler. Therefore, the median is unaffected by any change in the magnitude of temperature on hot or cold days, and a trend in differences between the tails and the median provides evidence of a mechanism that is preferentially affecting the extremes within a season.

Averaging information about extremes across space requires careful consideration, because the probability of a given extreme (e.g., an event measured in SD, or sigma, units) is strongly dependent on the underlying nonnormality of the distribution of daily temperature (2, 33, 41), and random sampling of these nonnormal distributions across space can give rise to larger or smaller trends in the spatial average of extremes than would be expected from sampling normal distributions (*SI Appendix*, Fig. S1). To address this issue, after defining the extremity of a given season as the maximum or minimum of the temperature deviations from the seasonal median, we transform this metric into a rank. For the 65-y record we consider in the main text, ranks will range between 1 and 65; the year with the largest difference between the seasonal extreme and median will have the largest rank. Thus, an increase in ranks over time indicates faster warming of the extremes compared to the median, and a decrease indicates slower warming. Analyses with synthetic datasets indicate that this method is effective at distinguishing between median and variance changes, and allows for detection of signals that are small locally, but consistent across many locations (*Materials and Methods* and *SI Appendix*, Fig. S2).

The primary results use the ERA5 reanalysis (42) given its spatial completeness and duration. The ERA5 results are also compared to those using high-quality station data from the Global Historical Climatology Network-Daily (43), as well as the shorter-duration CHIRTSdaily dataset, which focuses on providing improved estimates of climate variables in regions with few in situ observations (44). However, the three datasets are not independent of each other.

Little Evidence of Amplification of Hot Extremes Locally, Hemispherically, or Globally

We do not find evidence that hot extremes tended to warm faster than the median at any spatial scale over 1959–2023 (Fig. 1 *A*–*E*). At the local (gridbox) scale (Fig. 1*A*), there are few significant positive or negative trends in the ranks (0.7% of the land area for positive, 1.0% for negative) when accounting for the large number of hypothesis tests performed through controlling for a false discovery rate of $\alpha_{FDR} = 0.05$ (*Materials and Methods*). Notably, one of the few regions with a significant positive trend toward faster warming of heat extremes is northwest Europe where it has been suggested that dynamical changes, largely not simulated by climate models, have caused amplified warming of heat extremes (15, 39, 45). There is good agreement between ERA5 and available GHCND stations in terms of both the magnitude and spatial pattern of the trends (*SI Appendix*, Fig. S3), although the number and spatial distribution of GHCND stations is limited because complete temporal sampling is necessary to quantify the extreme values in each season. In many ways, the lack of clear signal locally is unsurprising: as discussed above, if amplification of extremes is present, it is likely a small signal that would be difficult to detect at the local scale. We are thus motivated to consider whether differential warming of hot extremes is present in aggregate over larger spatial scales.

Because the distribution of ranks at each location is identical, they can be straightforwardly averaged across gridboxes with appropriate area-weighting. The global average rank time series has been largely constant since 1959 (Fig. 1*B*), indicating that warm season heat extremes are, on average at the global scale, following the median. Because heat extremes are driven by distinct processes in the tropics and extratropics, we further divide the global domain into the extratropics in each hemisphere (poleward of 10°), and the deep tropics (equatorward of 10°, Fig. 1 *C*–*E*). In the Northern Hemisphere, there is no significant or visually apparent trend in the rank time series. The trends in both the tropics and the Southern Hemisphere are visibly positive; while neither are found to be significant when controlling for $\alpha_{FDR} = 0.05$, the Southern Hemisphere trend has an estimated P-value of 0.06, so is near the edge of what could be expected from random sampling of internal variability. The result that hot days are warming at the same rate as the median at the hemispheric, tropical, and global scale holds when using four other metrics of extremity, some of which account for heat wave duration (*SI Appendix*, Figs. S4–S7). Results are also consistent if the seasonal medians are first smoothed using a 5-y running average (*SI Appendix*, Fig. S8), if the analysis is limited to the satellite era (1980–2023, *SI Appendix*, Fig. S9), and when using the shorter CHIRTSdaily dataset (1983–2016, *SI Appendix*, Fig. S10).

The lack of clear signal in any domain is notable for two reasons. First, a growing body of literature has suggested that heat extremes will warm more than typical temperatures in the tropics, supported by climate model simulations (22, 23, 25, 26), atmospheric dynamics theory (25), and land-atmosphere processes (23, 26). While it is possible that the tropical signal of amplification will still emerge in the observations, it is notable that it is not yet present. Second, the rank values for recent Northern Hemisphere summers with multiple high-impact extremes are not particularly unusual. Thus, a different sampling of internal variability combined with the trends in the median could have led to even more extreme summers than were recently observed.

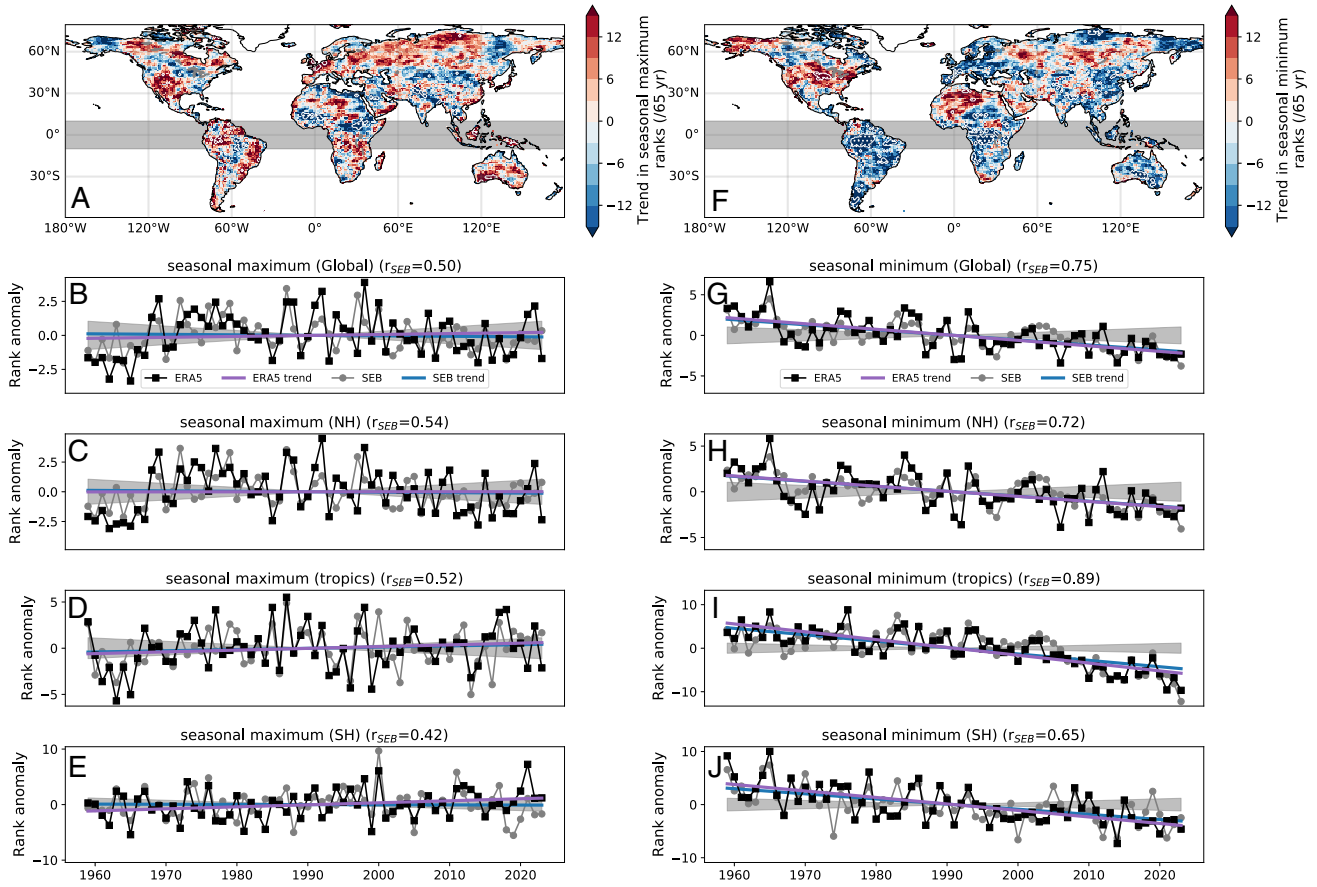

Fig. 1. Hot extremes have followed the median, and cold extremes have warmed more slowly. (*A*) The linear trend (1959–2023) in the ranks of the length of the upper tail (seasonal maximum minus median). Positive values indicate greater warming of hot extremes compared to the median. Gridboxes with significant trends are shown with white stippling outlined by white contours. The (*B*) global, (*C*) Northern Hemisphere, (*D*) tropical, and (*E*) Southern Hemisphere average rank anomaly by year (black squares and line) and the fitted linear trend (purple line). The 95% range of samples from the null hypothesis of no trend is shown in gray; *Materials and Methods*. The ranks predicted by the surface energy balance framework (*Large-Scale Behavior Explained through the Surface Energy Balance*; gray circles and line) and their fitted trend (blue line) are also shown. Comparable information for the cold tail in summertime temperature is shown in (*F–J*), based on the ranks of the seasonal minimum minus median.

Daily Summer Temperature Distributions Are Stretching Due to Damped Warming of the Cold Tail

While changes in hot summer days largely followed the median, trends in summertime cold extremes of daily maximum temperature did not keep pace, indicating a stretching of the cold tail of summertime temperatures. Based on our definition of ranks, a slower warming of the cold tail compared to the median is associated with a reduction in ranks over time. Similar to the analysis for hot extremes, only a small minority of gridboxes show significant trends for differential changes in the cold tail, but in this case the majority of the significant trends are negative, indicating slower warming of the cold days (0.6% of the land area for positive, 4.0% for negative). There is again good agreement with available GHCND stations in terms of the spatial pattern and magnitude of trends in ranks (*SI Appendix*, Fig. S3). The trends in the cold tail ranks are not explained by a change in the timing of the cold extremes within the warm season (*SI Appendix*, Fig. S11).

The signal of stretching of the cold tail becomes more obvious after averaging the ranks across larger domains, all of which show a clear and significant negative trend. The largest magnitudes of the negative trends are in the tropics and Southern Hemisphere. Given the relative sparsity of in situ measurements of temperature in these regions, we again compare to satellite-era ERA5 and CHIRTSdaily (*SI Appendix*, Figs. S9 and S10). While all trends remain visibly negative, and near the edge of what we would expect from the null hypothesis, only the tropical cold tail stretching is found to be significant for this shorter record. The change in estimated significance is due to both the larger range of trends that could be expected from sampling of internal variability over a shorter period, and the smaller value of the fitted trends outside of the tropics; in particular, the magnitude of the Southern Hemisphere trend using only satellite era data is half of its value during the full 1959–2023 period. The tropical signal, on the other hand, has the same magnitude during both periods. For all of the domains, nearly all of the possible trends based on at least 20 y of data during the satellite era are negative (*SI Appendix*, Fig. S12), suggesting a relatively consistent signal toward stretching of the cold tail over the last 44 y. Thus, the distribution of daily maximum temperature has generally been increasing in variance, although primarily due to cold day stretching rather than hot day amplification compared to the median.

Large-Scale Behavior Consistent with the Surface Energy Balance

What is the cause of the observed stretching of the cold tail, as well as the lack of stretching of the warm tail? Given the importance of land-atmosphere fluxes for summertime daily maximum temperatures (e.g., refs. 46–50), we combine the surface energy balance with the definition of evaporative fraction (*EF*; the

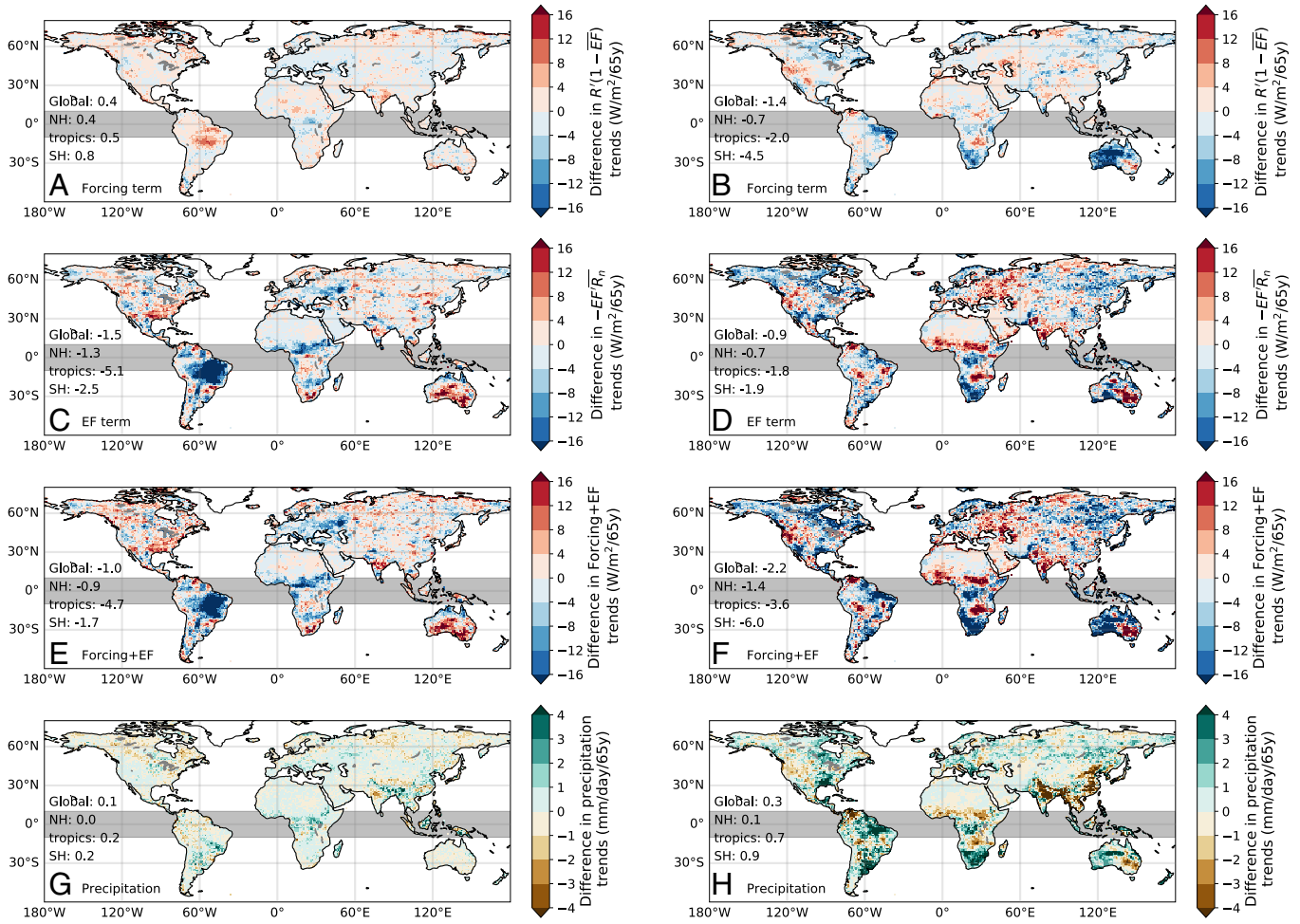

Fig. 2. Relative trends in surface energy balance terms and precipitation on hot and cold days versus median days. The difference between trends in (*A* and *B*) the forcing anomaly term, $(1-\overline{EF})R'$ (Wm^{-2}/65 y), (*C* and *D*) the evaporative fraction anomaly term, $-\overline{R_n}EF'$ (Wm^{-2}/65 y), (*E* and *F*) their sum, and (*G* and *H*) total daily precipitation (mm/day/65 y) in ERA5 on hot [left column, panels (*A*, *C*, *E*, and *G*)] or cold [right column, panels (*B*, *D*, *F*, and *H*)] minus median days. Hot days are defined as days with temperature anomalies exceeding the 90th percentile, cold days below the 10th percentile, and median days between the 45th and 55th percentiles. The global and subdomain averages are shown in each panel. A positive sign in panels (*A*–*F*) would be expected to be associated with greater warming of the extreme than the median.

fraction of incoming energy partitioned into evapotranspiration), linearize, and assume a linear scaling between temperature anomalies and sensible, ground, and longwave heat anomalies (*Materials and Methods*) to propose a proportionality for temperature anomalies,

$$T' \propto (1-\overline{EF})R' - \overline{R_n}EF'. \quad [1]$$

In the equation, EF is the evaporative fraction, R_n is the downward net surface radiative forcing, and R' is the downward net surface radiative forcing anomaly after removing the component of longwave that is collinear with temperature (*Materials and Methods*). Overbars indicate time means, and primes indicate anomalies from the climatological seasonal cycle. The evaporative fraction generally increases with increasing soil moisture (48, 51, 52), but can also be influenced by radiation and meteorological conditions including temperature, humidity, and wind speed (53, 54). The first ("forcing") term predicts that temperature increases with greater heating, but that the influence of the heating is scaled by the climatological EF, with drier regions having a greater sensitivity to the forcing. The second ("evaporative fraction") term predicts that temperature increases with decreases in evaporative fraction (drying), but that the influence of the drying is scaled by the climatological net radiation at the surface. Using daily ERA5 data to estimate each term on the right-hand side, we produce estimates of temperature from the surface energy balance that are then converted into ranks in the same manner as we did for the actual temperatures.

The trends in ranks predicted by the surface energy balance-based equation are shown in gray in the time series in Fig. 1, and closely align with what has been observed: no significant trends in hot days beyond the median, and a significant stretching of the cold tail in all domains. The interannual variability in ranks predicted by the surface fluxes is similar to the ranks based on actual temperatures for the cold tail, with correlations of 0.75, 0.72, 0.89, and 0.65 for the globe, Northern Hemisphere, tropics, and Southern Hemisphere respectively. The interannual correlations for the hot tail are more moderate (0.50, 0.54, 0.52, and 0.42 for the same regions), indicative of the importance of other processes such as advection (39, 55, 56) and/or limitations of the assumptions of the surface energy balance framework.

To identify the relative importance of each term in Eq. **1**, we show the difference in linear trends (1959–2023) on hot and cold days, minus those on median days (Fig. 2 *A*–*F*). Hot, cold, and median days are defined as those days that exceed the 90th percentile, are below the 10th percentile, and are between the 45th and 55th percentile of detrended

daily temperature anomalies, respectively, for each location. We detrend temperature before calculating the percentiles in order to more evenly sample days across the period of study, and percentiles are calculated individually for each day of the year within the warm season to sample evenly across the season.

The difference in trends between hot and median days is relatively small for the forcing term, although the values are, on average, positive in each domain considered (Fig. 2*A*). There are larger differences in the trends of the evaporative fraction term between hot and median days, with substantial regional variability in sign and magnitude; the average value in each domain, however, is negative, and thus counteracts the positive tendency in the forcing term at large scales (Fig. 2*C*). However, the maps of the relative trends in the forcing and evaporative fraction terms are only weakly negatively correlated across space (correlation of −0.14), indicating that they do not tend to counteract each other locally, and the sign of the average values of the relative trends in each domain often switches when instead considering the shorter satellite era period (*SI Appendix*, Fig. S13). Collectively, this suggests a lack of a clear surface energy balance signal in the trends of hot versus median day warming, as expected due to the lack of significant trends in the rank of the seasonal maximum minus median.

In contrast, there are large differences in trends in both the forcing and evaporative fraction term between median and cold days, with both the evaporative fraction and forcing trends causing, on average globally and in each subdomain, slower warming of the cold tail than the median (Fig. 2 *B* and *D*). Further, the two components exhibit similarities in their spatial patterns (the spatial correlation between the two maps is 0.40), and thereby reinforce the same types of changes in the temperature distribution. The similar spatial structure suggests the possibility of a common driver of changes in both net surface radiation and evaporative fraction. A likely candidate is changes in precipitation: cold days in the summer tend to be rainy (*SI Appendix*, Fig. S14), and a relative increase in precipitation on cold days would be associated with both a relative decrease in net radiation at the surface and increase in evaporative fraction due to increased soil moisture. Indeed, the spatial structure of the relative trends in precipitation in ERA5 during cold minus median days (Fig. 2*H* and *SI Appendix*, Fig. S15) mirrors the changes in the forcing and evaporative fraction (spatial correlations of −0.45 and −0.46, respectively), and their sum (Fig. 2*F*, correlation of −0.53), and all domains show greater increases in precipitation on cold than median days, with the greatest relative increases in the tropics and Southern Hemisphere. The relative trends in the surface energy balance terms are qualitatively similar when considering only the satellite era (*SI Appendix*, Fig. S13), as are the relative trends in precipitation using satellite-era ERA5 and five additional different precipitation datasets that span all or most of the satellite era and variously incorporate gauges, satellites, and reanalyses (*SI Appendix*, Fig. S16).

CMIP6 Models Also Do Not Amplify Heat Extremes in the Historic Period, but Miss Stretching of Cold Tail

Despite large differences in climate sensitivity (57) and spatial patterns of warming (58) across CMIP6 models, they largely agree with each other and the observations regarding a lack of hot day amplification globally and in each subdomain over 1959–2023 (Fig. 3 *A*–*D*). The multimodel ensemble-mean trend in ranks is surprisingly consistent with the observations in all domains, given that the latter includes sampling of internal variability that would be averaged out in the former.

In contrast, the observed stretching of the cold tail is entirely outside the envelope of all model simulations (262 simulations

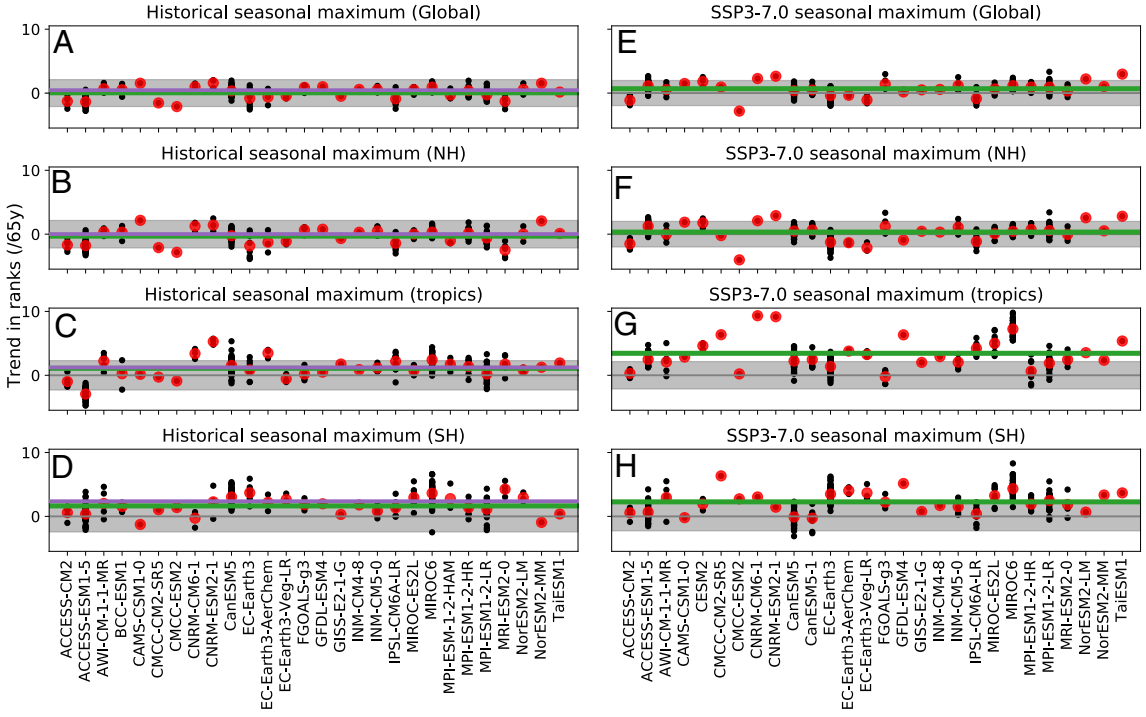

Fig. 3. Historical and future trends in the length of the hot tail of temperature in CMIP6 simulations. The trend in ranks of the difference between the seasonal maximum and median for the historical (1959–2023, *A*–*D*) and future (2024–2099 using SSP3-7.0, *E*–*H*) periods for the globe, Northern Hemisphere, tropics, and Southern Hemisphere (top to bottom rows, respectively) in CMIP6 models with available daily maximum temperature output. Individual model simulations (black dots), the single-model ensemble mean (red dots), the multimodel ensemble mean (green line, calculated as average across red dots), the trend in ERA5 (purple line, historical period only), and the 95% range of samples from the null hypothesis of no trend (gray shading) are shown.

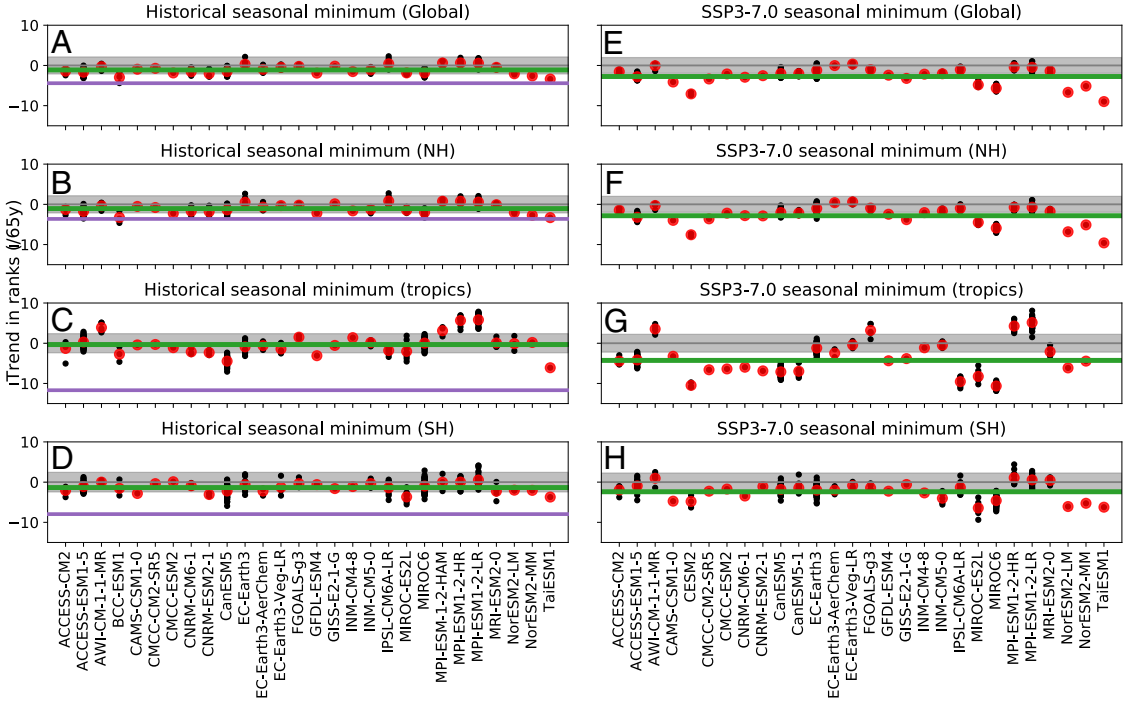

Fig. 4. Historical and future trends in the length of the cold tail of temperature in CMIP6 simulations. As in Fig. 3, but for the trend in ranks of the difference between the seasonal minimum and median for the historical (*A–D*) and future (*E–H*) periods. In panel (*G*), the single outlier for TaiESM1 is not shown; it has a trend value of −18.54/65 y (the ranks are unitless).

across 28 models) for the global, tropical, and Southern Hemisphere averages, and only one simulation has more cold day stretching than the observations in the Northern Hemisphere (Fig. 4 *A–D*). The multimodel ensemble-mean in all cases is consistent with the null hypothesis of the coldest temperatures warming at the same pace as the median. While this may seem less important than hot day amplification given that cold summer days do not tend to have adverse human impacts, cold temperatures during the warm season can constrain net primary productivity across colder regions (59) and provide a window for wildfire suppression given the strong, nonlinear relationship between vapor pressure deficit (largely a function of temperature) and wildfire spread (60). In addition, it suggests that the processes controlling changes in the shape of summertime temperature distributions differ between the models and observed world. Specifically, the prior observational analyses suggest that changes in the distribution of precipitation across days, with a more positive trend in precipitation on the coldest days compared to median days, may be a dominant cause of the observed stretching via reductions in net surface radiation and increases in the evaporative fraction through increased soil moisture on cold days. Is this behavior simulated by the climate models?

For each of the CMIP6 simulations spanning 1959–2023, we similarly identify cold and median days, calculate the difference in precipitation trends conditional on each type of day, and compare to the trends in cold tail ranks (*SI Appendix*, Fig. S17). The across-simulation spread in precipitation trends on cold minus median days has a significant negative relationship with the trends in cold tail ranks for the global, Northern Hemisphere, and tropical averages: simulations with greater precipitation increases on cold days tend to have greater stretching of the cold tail, as expected. However, outside of the tropics, the relationship is weak, and not significant in the Southern Hemisphere, indicating that additional factors are needed to explain the cross-model spread in cold tail stretching in the extratropics. We next compare the observed relative trends in cold day precipitation to the model trends. The observed trends are at the edge of the distribution of climate model simulations for the tropics and Southern Hemisphere, but near the multimodel ensemble mean for the global and Northern Hemisphere averages. Combined, the evidence suggests that the difference between models and observations for cold tail stretching in the tropics can be partially explained by model-observational differences in precipitation trends on cold versus median days. However, more work is needed to identify the processes controlling the cross-model spread in cold tail stretching, and whether the same processes can explain the mismatch between the models and observations. Other factors not explored here such as model-observational differences in sea surface temperature trends (61), and the warm season temperature response to ozone loss and recovery (62, 63) or aerosols (64) may be playing a role.

Future simulations (2024–2099) using the SSP3-7.0 scenario suggest that the signal of hot day amplification will emerge in the tropics, as has been noted in prior studies (23, 25): the multimodel ensemble mean exhibits a significant trend in the areal-averaged ranks, and 299 out of 316 simulations show a trend in the same direction (Fig. 3*G*), although there is substantial intermodel spread in the magnitude of the amplification. While the multimodel ensemble mean in the Southern Hemisphere is at the edge of significance, over 80% of simulations also suggest hot day amplification (Fig. 3*H*). In contrast, the multimodel behavior for the globe and Northern Hemisphere remains consistent with the hottest days shifting with the median (Fig. 3 *E–F*). The stretching of the lower tail also emerges in the future simulations (Fig. 4 *E–H*), with a significant negative trend in the ensemble mean for the global, Northern Hemisphere, and tropical averages.

It is notable that the observed cold day stretching is at the edge of or entirely exceeds historical CMIP6 simulations in all

domains, and is also greater in magnitude than nearly all of the trends from the future simulations in the tropics and Southern Hemisphere. There are three plausible explanations for the outlier behavior of the observed cold tail amplification. The first is that the observed trends are erroneous. The greatest differences between the models and observations are in the tropics and Southern Hemisphere, where long-term in situ measurements are limited, so it is difficult to both constrain and validate a reanalysis product. We do find agreement between the spatial structure of trends in ERA5 and GHCND in Australia (*SI Appendix*, Fig. S3), but ground observations are assimilated into ERA5 so this does not offer independent validation. When considering only satellite era data, the magnitudes of the extratropical trends are reduced, which may place them in the envelope of the models; however, the magnitude of the tropical trend is unchanged, and would likely remain outside the envelope of the models. Nevertheless, given a lack of ground truth, the trends being an artifact of the data cannot be excluded. The second is that the observed trends are primarily due to sampling of low-frequency internal variability that is not simulated by the climate models; indeed, studies comparing climate model simulations and proxy records have suggested that climate models tend to underestimate longer-than-decadal temperature variability, especially at subglobal scales (65, 66). While it is difficult to definitively conclude that the observed trends are not due to sampling of internal variability without relying on the model-simulated variability, we note that the sign of the trend in the domain-average ranks across starting and ending dates is very consistent (*SI Appendix*, Fig. S12 *E–H*). As such, very low-frequency (centennial-scale) internal variability would be required to explain the observed trends. The third explanation is that the observations are indicating a forced trend that is underestimated in models and therefore emerges later in the models than observations. There are two pieces of evidence in support of this explanation, although neither excludes the prior two explanations: 1) The observed trend far exceeds the range of trends that would be expected from sampling internal variability as estimated through either averaging across a random sample of ranks (gray shading in Figs. 3 and 4; *Materials and Methods*) or the spread across climate model large ensembles, and 2) most climate model simulations exhibit a trend in the same direction as the observations under stronger, future forcing. It is also possible that multiple or all of the above explanations are true, as has often been found for other model-observational discrepancies (e.g., the tropical tropospheric warming discrepancy; 67, 68).

Discussion

An increase in summertime heat extremes is one of the most obvious and high-impact consequences of global warming. But are extremes around the globe warming faster than typical summertime temperatures? Despite the intensity of recent heatwaves, we suggest the answer is "not yet." We have shown that changes in historical warm season heat extremes can be explained by changes in the median when aggregating across global and hemispheric scales, indicating that proposed processes for amplification of heat extremes are not yet dominant at large scales. Nevertheless, the spread of summertime temperatures has increased in the global average due to the slower warming of cold summer days. Trends in both warm and cold extremes compared to the median can largely be explained by changes in net surface radiation and evaporative fraction. We do not diagnose the origin of the relative changes in radiation and evaporative fraction for the warm tail, which could be responses to local conditions like surface moisture availability (48, 51, 52, 69, 70), changes in the large-scale circulation (71), and/or changes in forcing, including aerosols (72). For the observed stretching of the cold tail, the changes in radiation and evaporative fraction generally reinforce each other, which we interpret as suggesting a common driver. In particular, there have been greater increases in precipitation on cold compared to median days that would be expected to reduce net radiation through increased cloudiness and increase the evaporative fraction through increases in soil moisture, although the changes could be moderated by other, associated changes in radiation and meteorology.

The behavior of the extremes was benchmarked with respect to the seasonal median, rather than mean, because the former is unaffected by changes in the tails of a distribution. In particular, the slower warming of the cold tail relative to the median (combined with the hot tail keeping pace with the median) would be expected to induce slower warming of the mean compared to the median. As a direct result of this asymmetry, repeating our analysis of the amplified warming of hot days but benchmarking against the mean, instead of the median, leads to a conclusion that there has been significantly faster warming of hot days in the tropics and Southern Hemisphere where there has been the greatest stretching of the cold tail (*SI Appendix*, Fig. S18). The use of the median allows for a clearer separation of the trend in each tail, and we suggest that the amplification of hot day warming compared to the mean can be viewed as an artifact of the cold tail stretching, which has rather different implications in terms of impacts than if the variability increase were due to stretching of the warm tail.

There is general agreement between CMIP6 models and observations that hot days have warmed at the same pace as the median at large scales. The good agreement historically provides some confidence that the projected hot day amplification in the tropics, and perhaps in the Southern Hemisphere, may indeed emerge, and motivates future work to reduce the substantial intermodel spread in the magnitude of the amplification. This stands in contrast to the Northern Hemisphere where models agree that future heat extremes will follow changes in the median. The agreement between models and observations is diminished for the cold tail, where the observed signal generally falls outside the envelope of 28 CMIP6 models. However, it should be noted that the disagreement is largest in the tropical and Southern Hemisphere domains where our observing systems, particularly in situ observations, are more limited, and the possibility of data artifacts cannot be excluded.

Extreme heat is the most deadly type of extreme weather (73). The results found here suggest that historical changes in heat extremes at large spatial scales can be relatively simply explained by a spatially variable shift in median temperatures. The extent to which this continues to hold in the future should be monitored to provide relevant information about the expected future probabilities of extreme temperatures.

Materials and Methods

Period of Analysis. The main analysis focuses on the years 1959–2023. The year 1958 is included for the Southern Hemisphere because we assign any Southern Hemisphere warm season that begins in year X to the year X + 1. The starting year of the analysis was chosen because of data quality concerns in ERA5 temperature in prior years (*SI Appendix*, Fig. S19 as an example), as well as the fact that many GHCND stations come online throughout the late 1940s and 1950s; prior to 1950, very few stations are available outside of North America (43).

ERA5 Data. The primary source of data is the ERA5 global reanalysis (42) due to its duration and quality. Daily maximum values of 2 m air temperature (ERA5 variable name "2m_temperature") are calculated as the maximum hourly value in each day. This will, in general, lead to underestimated values of the maxima since they are not recorded instantaneously, but will not induce spurious trends in the data. Net surface shortwave ("surface_net_solar_radiation"), net surface longwave ("surface_net_thermal_radiation"), latent heat flux ("surface_latent_heat_flux"), and sensible heat flux ("surface_sensible_heat_flux") are also used. For shortwave, longwave, latent, and sensible heat fluxes, the daily value is the average across the hourly values. The hourly data are downloaded at the native $0.25° \times 0.25°$ latitude-longitude resolution. After calculating the daily maximum or average, the data are regridded to a $1° \times 1°$ latitude-longitude grid. The evaporative fraction, EF, is calculated as the fraction of net radiation at the surface that is partitioned into latent heating: $EF = \frac{LH_{\uparrow}}{SW_{\downarrow} - LW_{\uparrow}}$, where the direction of the subscripted arrow indicates our sign convention for the fluxes. Note that the raw data from ERA5 always have the sign convention of positive downward.

GHCND Data. To provide some validation of the ERA5 temperature trends, we compare to in situ measurements contained within the Global Historical Climatology Network-Daily (GHCND; 43) database. Stations are first selected if their first year is at least as early as 1959 for the Northern Hemisphere and 1958 for the Southern Hemisphere, and their last year is 2023 or later. Measurements that fail any quality control checks or are lagged with respect to the reported time of observation (mflag of "L") are set to missing. Because our analysis relies on the seasonal maxima and minima of temperature, seasons that do not have 100% data coverage within the season are removed. We then retain only the stations that have data for at least 75% of the years (at least 48 out of 65 y), with the additional requirement that they must also have the same fraction of coverage during the first and last 10 y of the study period (at least eight out of 10 y). The remaining stations are primarily in the United States, Europe, and Australia.

Additional Gridded Observational Datasets. We compare the ERA5 results to other gridded, daily datasets for temperature and precipitation that are available for various subsets of the satellite era. For temperature, we use the CHIRTSdaily dataset (1983–2016) that merges remote sensing, in situ data, and ERA5 reanalysis, and is specifically focused on creating a high-quality product in regions with low in situ data coverage (44). For precipitation, we consider five additional datasets beyond ERA5: 1) CHIRPS (1981–2023), based on in situ and remote sensing measurements, with a focus on quality in regions with sparse in situ measurements (74); 2) MSWEP (1979–2023), which merges gauge, satellite, and reanalysis (ERA5) data (75); 3) PERSIANN-CDR (1983–2023), which uses an artificial neural network to predict precipitation based on satellite imagery, which is trained using National Centers for Environmental Prediction (NCEP) Stage IV hourly precipitation and bias-corrected with Global Precipitation Climatology Project (GPCP) monthly precipitation data (76); 4) CPC Global Unified Gauge-Based Analysis of Daily Precipitation (77); and 5) GPCC Full Data Daily Version 2022 (1982–2020) (78). The latter two datasets are based solely on in situ gauge measurements that are statistically interpolated; GPCC generally draws on a greater number of in situ observations than CPC (79). All datasets are bilinearly interpolated to the same $1° \times 1°$ latitude-longitude grid used for ERA5.

CMIP6 data. For comparison to the observed trends in temperature, we use all CMIP6 models (80) that provide daily data in the Earth System Grid Federation database for daily maximum temperature ("tasmax") for both the historical (up to 2014) and SSP3-7.0 scenarios (2015 and onward) (81), leading to 262 simulations across 28 models, four of which are large ensembles (at least 30 members; ACCESS-ESM1-5, CanESM5, MIROC6, and MPI-ESM1-2-LR). For the future trends, only the SSP3-7.0 simulations are required, and 316 simulations across 28 models are available, five of which are large ensembles (ACCESS-ESM1-5, CanESM5, EC-Earth3, MIROC6, and MPI-ESM1-2-LR). To create *SI Appendix*, Fig. S17, we additionally use daily precipitation ("pr") for the historical (up to 2014) and SSP3-7.0 scenarios (2015–2023); there are slightly fewer (255) simulations available for this analysis because a small number of models do not provide daily precipitation data for both the historical and SSP3-7.0 scenarios.

Geographic Domain and Definition of Land. We consider 60°S to 80°N for the analysis. Both Greenland and Antarctica are excluded given the different processes that govern temperature over persistently ice-covered surfaces. Gridboxes marked as having at least 50% land fraction are included in the analysis. For ERA5, the native landmask is regridded to $1° \times 1°$ latitude-longitude resolution before masking; the same land mask is used for all other gridded observational products. For CMIP6, the land mask for each model is used.

Anomaly Calculation. The climatological seasonal cycle in temperature is removed before calculation of the seasonal median and extremes. The seasonal cycle is estimated by fitting the data to the first five Fourier basis functions with annual frequency (e.g., a sine and cosine pairs with frequencies 1/ y, 2/ y,..., 5/ y). This model explains >99% of the variance in the empirical seasonal cycle (estimated by averaging by day of year) on average across gridboxes in ERA5, and comparable amounts for the other temperature datasets. The terms in the surface energy balance in Eq. **1** are also considered as anomalies from the seasonal cycle. In these cases, the seasonal cycle is estimated through applying a forward-backward third-order lowpass Butterworth filter with a cutoff frequency of 1/30 d to the empirical seasonal cycle. This approach is taken for the budget terms because their seasonality (particularly for latent and sensible heat fluxes) is harder to capture with a small number of Fourier basis functions.

Defining the Warm Season. In order to perform a global analysis of changes in warm season temperature distributions, we consider a latitudinally varying warm season. At each latitude, the fitted seasonal cycle (from the first five Fourier bases, discussed above) over land is averaged across longitude, the hottest day of year from the averaged seasonal cycles is identified, and then the warm season is defined as a 91-d window around that day (*SI Appendix*, Fig. S20). Warm seasons south of the equator are shifted forward by half a year, such that, e.g., the austral summer that spans 1958–1959 is assigned the year 1959.

Alternative Metrics of Extremes. The primary analysis defines the extremity of temperature deviations (after removing the seasonal median) of a given year by the seasonal maximum and minimum. We also consider three other metrics of extremity to provide a more holistic perspective. For the hot tail, the metrics are (1a) the cumulative degree-days over the 95th percentile of the temperature deviations from the median (i.e., the sum of temperature deviations across the season that exceed the 95th percentile), (2a) the number of days per season that exceed the 95th percentile of the temperature deviations, and (3a) the average exceedance over the 95th percentile of the temperature deviations, i.e. (1a) normalized by (2a). The comparable metrics for the cold tail are (1b) the cumulative degree-days less than the 5th percentile of the temperature deviations, (2b) the number of days per season that are below the 5th percentile of the temperature deviations, and (3b) the average exceedance less than the 5th percentile of the temperature deviations, i.e. (1b) normalized by (2b). The equivalent of Fig. 1 for these alternative definitions of extremes are in *SI Appendix*, Figs. S4–S6; the spatial patterns and time series are very similar to the main results. An advantage of performing the full analysis in rank space is that the ranks from the different definitions of extremes can be averaged together, providing a single metric that incorporates all of the characteristics of each metric. The trends in the average metric are shown in *SI Appendix*, Fig. S7, and are again very similar to the others.

Null Hypothesis for Rank Trends and Spatial Degrees of Freedom. The null hypothesis for the rank trends is that there is no change in the ranks over time. To quantify this null hypothesis for the domain-average values, we perform the following procedure 1,000,000 times: N random time series of ranks (e.g., a random draw without replacement of values between 1 and 65 for a 65-y record) are created, and averaged, where N is an estimate of the number of spatial degrees of freedom in the field (the global land masses, or each hemisphere or tropics). The 1,000,000 resulting time series (each an average across N time series) are treated as samples from the null hypothesis. The value of N is estimated using the eigenvalue formula method of Bretherton et al. (82) and is found to be 57, 53, 48, and 43 for the seasonal maximum ranks for the

globe, Northern Hemisphere, tropics, and Southern Hemisphere, respectively, and 57, 54, 45, and 43 for seasonal minimum ranks in the same domains. Note that the degrees of freedom estimated with this method are limited by the number of time samples (65, in our case), and may be biased low in cases when the estimated number of degrees of freedom is close to 65, such as for the global and Northern Hemisphere averages. Increasing the degrees of freedom (e.g., ref. 33 estimated ≈100 degrees of freedom in the Northern Hemisphere for similar data) would slightly shrink the null intervals. However, none of the general conclusions would change given the general lack of signal of hot day amplification in the global and Northern Hemisphere averages and strong signal of cold tail stretching.

Multiple Hypothesis Testing and Significance. To account for multiple hypothesis testing (on average, 5% of null hypotheses will be rejected if performing individual tests at the 0.05 level), we control for a false discovery rate (83) of 0.05. Through this control, in expectation, no more than 5% of rejected nulls will be false discoveries (i.e., true nulls). The false discovery control is performed independently to test the null hypothesis of no trend at each gridbox in each map in Fig. 1, for each of the eight time series in Fig. 1, and for each of the 16 multimodel ensemble means in Figs. 3 and 4. For the domain-average trends, a two-sided *P*-value for each trend is estimated as the percentile of the observed trend within the trends estimated from the 1,000,000 samples of the null hypothesis produced from the procedure discussed above. For the gridbox trends, the two-sided p-value is estimated using the scipy.stats.linregress function, which uses the Wald Test with t-distribution of the test statistic. For Figs. 1 *B*–*E* and *G*–*J*, 3, and 4, each null envelope (gray shading) shows the 95% range of the samples from the null hypothesis, while the cutoff p-value for significance will vary (and generally be less than 0.05) depending on the distribution of *P*-values and number of hypotheses being tested.

Illustration of Trends in Ranks Approach Using Synthetic Data. In order to test and demonstrate the power of the spatially aggregated rank-based approach to detecting changes in extremes, we apply the method to synthetic data. The synthetic data are generated from a normal distribution to straightforwardly simulate separate changes in the median (equal to the mean for a normal distribution) and the tails (as a function of the variance). We explore three different scenarios: 1) increase in median and no change in variance, 2) increase in variance and no change in median, and 3) increase in both median and variance. In scenario (1), all changes in extremes should be accounted for by the change in the median, and we use it to confirm that our proposed method is not leading to false positives. In scenarios (2) and (3), the extremes are changing beyond what is explained by a median shift, and we use them to test the power of the proposed method. We also consider a case with no change in median or variance to provide the null hypothesis. For each experiment, we generate data for 50 "locations" (analogous to the aforementioned ≈50 spatial degrees of freedom estimated for each domain), each with 65 "years" of data containing 91 d, analogous to the observational records used in the analysis. We consider median changes over 65 y between 0.25 °C and 2 °C in steps of 0.25 °C, and fractional variance changes between 5% and 50% in steps of 5%. For each case of median-only, variance-only, or median-and-variance changes, we assess the fraction of tests with a significant trend in ranks for both the individual "locations" and for the average across locations. In the case of individual locations, we employ a false discovery rate control, analogous to our actual analysis of individual gridboxes (although with a smaller number of tests here). *SI Appendix*, Fig. S2 shows the fraction of N = 100 hypothesis tests that were rejected for each median and variance combination, before and after averaging. Even in cases of large variance changes (35% increase over 65 y), only ≈30% of individual locations were found to have a significant trend in ranks. In contrast, after averaging the rank time series, variance changes as small as 5% over 65 y could be detected ≈65% of the time, and a variance change of 10% was detected over 99% of the time. Importantly, when only the median was changing over time, a "significant" trend in the spatially averaged ranks was found, on average, 5% of the time, consistent with our choice of significance level.

Inferring Temperature Anomalies from the Surface Energy Balance. The evaporative fraction is generally defined as,

$$EF = \frac{LH}{R_n} = \frac{LH}{SW - LW}, \qquad [2]$$

where LH is latent heat (positive upward) and R_n is the net radiation at the surface, which can be decomposed into net shortwave (SW, positive downward) and net longwave (LW, positive upward) components. Each component of the equation can be written in terms of a time-mean state (shown as $\bar{\cdot}$) and an anomaly (shown as $\cdot'$). Through assuming that any product of anomalies is small compared to terms that include the mean state, and noting that Eq. **2** holds for the mean state,

$$\overline{EF}(SW' - LW') + \overline{R_n}EF' = LH' = SW' - LW' - SH' - G', \qquad [3]$$

where the second equality draws on the surface energy budget ($SW - LW = LH + SH + G$; G is the ground heat flux and SH is the sensible heat flux, both are defined as positive upward). The anomalies are calculated from the climatological seasonal cycle, although the mean states are chosen to be time-constant for simplicity, and because the seasonality in the mean is much smaller than the time-mean (*SI Appendix*, Fig. S21). In order to move from the surface energy budget to temperature, it is necessary to assume some relationships between the terms and temperature anomalies. We first assume that the longwave heat flux anomalies can be decomposed as $LW' = aT' - LW'_f$. The first term varies linearly with temperature, and is assumed to capture both upward and downward longwave radiation caused by surface and atmospheric temperature, respectively. The linearity in upward longwave is from the linearization of the nonlinear Stefan–Boltzmann law. The second term is interpreted as the forcing related to increased greenhouse gas concentrations. In addition, we assume $SH' = bT'$ and $G' = cT'$. Note that a, b, and c can be spatially variable, are all expected to be positive and largely unchanged between climate states (see p. 1664 in ref. 69 for a discussion of these assumptions). Using these assumptions, we can rearrange to get an expression for temperature,

$$T' = \frac{1}{a(1 - \overline{EF}) + b + c}\left((1 - \overline{EF})R' - \overline{R_n}EF'\right), \qquad [4]$$

where $R' = SW' + LW'_f$, i.e., the forcing anomaly is the sum of shortwave and longwave anomalies. We assume that a, b, and c do not vary between hot, cold, and median days, yielding Eq. **1**, which is used to calculate the gray lines in Fig. 1.

Data, Materials, and Software Availability. All codes (in Python) to analyze the data and make the figures are available at https://zenodo.org/doi/10.5281/zenodo.12640136 All datasets used in this study are freely available at the following websites: ERA5: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5 (42); GHCND: https://www.ncei.noaa.gov/data/global-historical-climatology-network-daily/ (43); CHIRTSdaily: http://data.chc.ucsb.edu/products/CHIRTSdaily/ (44); CHIRPS: https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_daily/ (74); MSWEP: https://www.gloh2o.org/mswep/ (75); PERSIANN-CDR: https://www.ncei.noaa.gov/data/precipitation-persiann/access/ (76); CPC: https://downloads.psl.noaa.gov/Datasets/cpc_global_precip/ (77); GPCC: https://opendata.dwd.de/climate_environment/GPCC/full_data_daily_v2022/ (78); and CMIP6: https://esgf-node.llnl.gov/search/cmip6/ (80).

ACKNOWLEDGMENTS. K.A.M. acknowledges support from the NSF (Awards No. 1939988 and 2338237) and the David and Lucile Packard Foundation. I.R.S. acknowledges funding from the NSF National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. A.P.W. acknowledges support from the MacArthur Foundation.

Author affiliations: [a]Department of Statistics and Data Science, University of California, Los Angeles, CA 90095; [b]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095; [c]Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095; [d]Climate and Global Dynamics Laboratory, NSF National Center for Atmospheric Research, Boulder, CO 80305; and [e]Department of Geography, University of California, Los Angeles, CA 90095

1. R. Vautard *et al*., Human contribution to the record-breaking June and July 2019 heatwaves in Western Europe. *Environ. Res. Lett.* **15**, 094077 (2020).
2. K. A. McKinnon, I. R. Simpson, How unexpected was the 2021 Pacific Northwest heatwave? *Geophys. Res. Lett.* **49**, e2022GL100380 (2022).
3. R. H. White *et al*., The unprecedented Pacific Northwest heatwave of June 2021. *Nat. Commun.* **14**, 727 (2023).
4. T. Zhang, Y. Deng, J. Chen, S. Yang, Y. Dai, An energetics tale of the 2022 mega-heatwave over central-eastern China. *npj Clim. Atmos. Sci.* **6**, 162 (2023).
5. C. Qian *et al*., Rapid attribution of the record-breaking heatwave event in North China in June 2023 and future risks. *Environ. Res. Lett.* **19**, 014028 (2024).
6. S. Rahmstorf, D. Coumou, Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17905–17909 (2011).
7. E. Fischer, S. Sippel, R. Knutti, Increasing probability of record-shattering climate extremes. *Nat. Clim. Change* **11**, 689–695 (2021).
8. V. Petoukhov, S. Rahmstorf, S. Petri, H. J. Schellnhuber, Quasiresonant amplification of planetary waves and recent northern hemisphere weather extremes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5336–5341 (2013).
9. J. A. Screen, I. Simmonds, Amplified mid-latitude planetary waves favour particular regional weather extremes. *Nat. Clim. Change* **4**, 704–709 (2014).
10. M. Röthlisberger, L. Frossard, L. F. Bosart, D. Keyser, O. Martius, Recurrent synoptic-scale Rossby wave patterns and their effect on the persistence of cold and hot spells. *J. Clim.* **32**, 3207–3226 (2019).
11. K. Kornhuber *et al*., Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nat. Clim. Change* **10**, 48–53 (2020).
12. D. Coumou, K. Kornhuber, J. Lehmann, V. Petoukhov, "Weakened flow, persistent circulation, and prolonged weather extremes in boreal summer" in *Climate Extremes: Patterns and Mechanisms*, S. -Y. Simon Wang, J. -H. Yoon, C. C. Funk, R. R. Gillies, Eds. (Wiley, 2017), pp. 61–73.
13. M. E. Mann *et al*., Projected changes in persistent extreme summer weather events: The role of quasi-resonant amplification. *Sci. Adv.* **4**, eaat3272 (2018).
14. K. Kornhuber *et al*., Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environ. Res. Lett.* **14**, 054002 (2019).
15. E. Rousi, K. Kornhuber, G. Beobide-Arsuaga, F. Luo, D. Coumou, Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia. *Nat. Commun.* **13**, 3851 (2022).
16. J. A. Screen, I. Simmonds, Caution needed when linking weather extremes to amplified planetary waves. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2327 (2013).
17. E. A. Barnes, J. A. Screen, The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? *Wiley Interdiscip. Rev. Clim. Change* **6**, 277–286 (2015).
18. T. Woollings *et al*., Blocking and its response to climate change. *Curr. Clim. Change Rep.* **4**, 287–300 (2018).
19. J. Riboldi, F. Lott, F. d'Andrea, G. Rivière, On the linkage between Rossby wave phase speed, atmospheric blocking, and Arctic amplification. *Geophys. Res. Lett.* **47**, e2020GL087796 (2020).
20. R. Blackport, J. A. Screen, Insignificant effect of Arctic amplification on the amplitude of midlatitude atmospheric waves. *Sci. Adv.* **6**, eaay2880 (2020).
21. M. Stendel, J. Francis, R. White, P. D. Williams, T. Woollings, "The jet stream and climate change" in *Climate Change*, T. M. Letcher, Ed. (Elsevier, 2021), pp. 327–357.
22. M. H. Gross, M. G. Donat, L. V. Alexander, Changes in daily temperature extremes relative to the mean in Coupled Model Intercomparison Project Phase 5 models and observations. *Int. J. Climatol.* **39**, 5273–5291 (2019).
23. S. Q. Duan, K. L. Findell, J. S. Wright, Three regimes of temperature distribution change over dry land, moist land, and oceanic surfaces. *Geophys. Res. Lett.* **47**, e2020GL090997 (2020).
24. C. Simolo, S. Corti, Quantifying the role of variability in future intensification of heat extremes. *Nat. Commun.* **13**, 7930 (2022).
25. M. P. Byrne, Amplified warming of extreme temperatures over tropical land. *Nat. Geosci.* **14**, 837–841 (2021).
26. S. Q. Duan, K. A. McKinnon, I. R. Simpson, Two perspectives on amplified warming over tropical land. *J. Clim.* **37**, 4743–4760 (2024).
27. N. S. Diffenbaugh, J. S. Pal, F. Giorgi, X. Gao, Heat stress intensification in the Mediterranean climate change hotspot. *Geophys. Res. Lett.* **34**, L11706 (2007).
28. M. G. Donat, A. J. Pitman, S. I. Seneviratne, Regional warming of hot extremes accelerated by surface energy fluxes. *Geophys. Res. Lett.* **44**, 7011–7019 (2017).
29. M. M. Vogel *et al*., Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks. *Geophys. Res. Lett.* **44**, 1511–1519 (2017).
30. C. Schwingshackl, M. Hirschi, S. I. Seneviratne, A theoretical approach to assess soil moisture-climate coupling across CMIP5 and GLACE-CMIP5 experiments. *Earth Syst. Dyn.* **9**, 1217–1234 (2018).
31. A. K. Srivastava, M. Wehner, C. Bonfils, P. A. Ullrich, M. Risser, Local hydroclimate drives differential warming rates between regular summer days and extreme hot days in the Northern Hemisphere. *Weather Clim. Extremes* **45**, 100709 (2024).
32. K. A. McKinnon, A. Rhines, M. P. Tingley, P. Huybers, The changing shape of Northern Hemisphere summer temperature distributions. *J. Geophys. Res. Atmos.* **121**, 8849–8868 (2016).
33. S. Van Loon, D. W. Thompson, Comparing local versus hemispheric perspectives of extreme heat events. *Geophys. Res. Lett.* **50**, e2023GL105246 (2023).
34. D. Argüeso, A. Di Luca, S. E. Perkins-Kirkpatrick, J. P. Evans, Seasonal mean temperature changes control future heat waves. *Geophys. Res. Lett.* **43**, 7653–7660 (2016).
35. K. van der Wiel, R. Bintanja, Contribution of climatic changes in mean and variability to monthly temperature and precipitation extremes. *Commun. Earth Environ.* **2**, 1 (2021).
36. T. A. Carleton, S. M. Hsiang, Social and economic impacts of climate. *Science* **353**, aad9837 (2016).
37. N. R. Cavanaugh, S. S. Shen, Northern Hemisphere climatology and trends of statistical moments documented from GHCN-daily surface air temperature station data from 1950 to 2010. *J. Clim.* **27**, 5396–5410 (2014).
38. D. Coumou, G. Di Capua, S. Vavrus, L. Wang, S. Wang, The influence of Arctic amplification on mid-latitude summer circulation. *Nat. Commun.* **9**, 2959 (2018).
39. M. Patterson, North-West Europe hottest days are warming twice as fast as mean summer days. *Geophys. Res. Lett.* **50**, e2023GL102757 (2023).
40. M. Kotz, L. Wenz, A. Levermann, Footprint of greenhouse forcing in daily temperature variability. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103294118 (2021).
41. P. D. Sardeshmukh, G. P. Compo, C. Penland, Need for caution in interpreting extreme weather statistics. *J. Clim.* **28**, 9166–9187 (2015).
42. H. Hersbach *et al*., The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
43. M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, T. G. Houston, An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* **29**, 897–910 (2012).
44. A. Verdin *et al*., Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Sci. Data* **7**, 303 (2020).
45. R. Vautard *et al*., Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends. *Nat. Commun.* **14**, 1–9 (2023).
46. G. Lenderink, A. Van Ulden, B. Van den Hurk, E. Van Meijgaard, Summertime inter-annual temperature variability in an ensemble of regional model simulations: Analysis of the surface energy budget. *Clim. Change* **81**, 233–247 (2007).
47. A. J. Teuling *et al*., Contrasting response of European forest and grassland energy exchange to heatwaves. *Nat. Geosci.* **3**, 722–727 (2010).
48. S. I. Seneviratne *et al*., Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Rev.* **99**, 125–161 (2010).
49. D. O. Benson, P. A. Dirmeyer, Characterizing the relationship between temperature and soil moisture extremes and their role in the exacerbation of heat waves over the contiguous United States. *J. Clim.* **34**, 2175–2187 (2021).
50. J. Zhou, A. J. Teuling, S. I. Seneviratne, A. L. Hirsch, Soil moisture-temperature coupling increases population exposure to future heatwaves. *Earth's Future* **12**, e2024EF004697 (2024).
51. M. Budyko, *Climate and Life* (Academic Press, 1974), p. 508.
52. P. S. Eagleson, Climate, soil, and vegetation: 4. The expected value of annual evapotranspiration. *Water Resour. Res.* **14**, 731–739 (1978).
53. T. W. Ford, C. O. Wulff, S. M. Quiring, Assessment of observed and model-derived soil moisture-evaporative fraction relationships over the United States Southern Great Plains. *J. Geophys. Res. Atmos.* **119**, 6279–6291 (2014).
54. E. Haghighi, D. J. Short Gianotti, R. Akbar, G. D. Salvucci, D. Entekhabi, Soil and atmospheric controls on the land surface energy balance: A generalized framework for distinguishing moisture-limited and energy-limited evaporation regimes. *Water Resour. Res.* **54**, 1831–1851 (2018).
55. C. R. Holmes, T. Woollings, E. Hawkins, H. De Vries, Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *J. Clim.* **29**, 2221–2236 (2016).
56. T. Tamarin-Brodsky, K. Hodges, B. J. Hoskins, T. G. Shepherd, Changes in Northern Hemisphere temperature variability shaped by regional warming patterns. *Nat. Geosci.* **13**, 414–421 (2020).
57. M. D. Zelinka *et al*., Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
58. S. I. Seneviratne, M. Hauser, Regional climate sensitivity of climate extremes in CMIP6 versus CMIP5 multimodel ensembles. *Earth's Future* **8**, e2019EF001474 (2020).
59. R. R. Nemani *et al*., Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
60. J. K. Balch *et al*., Warming weakens the night-time barrier to global fire. *Nature* **602**, 442–448 (2022).
61. R. C. Wills, Y. Dong, C. Proistosecu, K. C. Armour, D. S. Battisti, Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophys. Res. Lett.* **49**, e2022GL100011 (2022).
62. L. M. Polvani, D. W. Waugh, G. J. Correa, S. W. Son, Stratospheric ozone depletion: The main driver of twentieth-century atmospheric circulation changes in the Southern Hemisphere. *J. Clim.* **24**, 795–812 (2011).
63. J. Bandoro, S. Solomon, A. Donohoe, D. W. Thompson, B. D. Santer, Influences of the Antarctic ozone hole on Southern Hemispheric summer climate change. *J. Clim.* **27**, 6245–6264 (2014).
64. Y. Xu, J. F. Lamarque, B. M. Sanderson, The importance of aerosol scenarios in projections of future heat extremes. *Clim. Change* **146**, 393–406 (2018).
65. T. Laepple, P. Huybers, Ocean surface temperature variability: Large model-data differences at decadal and longer periods. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16682–16687 (2014).
66. T. Laepple *et al*., Regional but not global temperature variability underestimated by climate models at supradecadal timescales. *Nat. Geosci.* **16**, 958–966 (2023).
67. S. Po-Chedley, T. J. Thorsen, Q. Fu, Removing diurnal cycle contamination in satellite-derived tropospheric temperatures: Understanding tropical tropospheric trend discrepancies. *J. Clim.* **28**, 2274–2290 (2015).
68. S. Po-Chedley *et al*., Internal variability and forcing influence model-satellite differences in the rate of tropical tropospheric warming. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2209431119 (2022).
69. W. Kong, K. A. McKinnon, I. R. Simpson, M. M. Laguë, Understanding responses of summer continental daily temperature variance to perturbations in the land surface evaporative resistance. *J. Clim.* **36**, 1653–1678 (2023).
70. Q. Kong, M. Huber, Regimes of soil moisture-wet-bulb temperature coupling with relevance to moist heat stress. *J. Clim.* **36**, 7925–7942 (2023).
71. M. Chiacchio, M. Wild, Influence of NAO and clouds on long-term seasonal variations of surface solar radiation in Europe. *J. Geophys. Res. Atmos.* **115**, D00D22 (2010).
72. R. M. Horton, J. S. Mankin, C. Lesk, E. Coffel, C. Raymond, A review of recent advances in research on extreme heat events. *Curr. Clim. Change Rep.* **2**, 242–259 (2016).
73. W. M. Organization, 2023 state of climate services: Health (Tech. Rep., 2023).
74. C. Funk *et al*., The climate hazards infrared precipitation with stations–A new environmental record for monitoring extremes. *Sci. Data* **2**, 1–21 (2015).
75. H. E. Beck *et al*., MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.* **21**, 589–615 (2017).
76. H. Ashouri *et al*., PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Am. Meteorol. Soc.* **96**, 69–83 (2015).
77. P. Xie, M. Chen, W. Shi, "CPC unified gauge-based analysis of global daily precipitation" in *Preprints, 24th Conference on Hydrology* (American Meteorological Society, Atlanta, GA, 2010), vol. 2.
78. U. Schneider *et al*., GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* **115**, 15–40 (2014).

79. Q. Sun *et al*., A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.* **56**, 79–107 (2018).
80. V. Eyring *et al*., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model. Dev.* **9**, 1937–1958 (2016).
81. B. C. O'Neill *et al*., The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci. Model. Dev.* **9**, 3461–3482 (2016).
82. C. S. Bretherton, M. Widmann, V. P. Dymnikov, J. M. Wallace, I. Bladé, The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* **12**, 1990–2009 (1999).
83. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Ser. B (Methodological)* **57**, 289–300 (1995).