Research Square

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

Predicting Unprecedented Hot Temperature Extreme Occurrences Up to a Year in Advance

Paolo De Luca

87paolo11@gmail.com

Barcelona Supercomputing Center (BSC) https://orcid.org/0000-0002-0416-4622

Juan C. Acosta-Navarro

European Commission Joint Research Centre https://orcid.org/0000-0001-5375-0639

Erich Fischer

ETH Zurich https://orcid.org/0000-0003-1931-6737

Markus Donat

Barcelona Supercomputing Center https://orcid.org/0000-0002-0608-7288

Article

Keywords:

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9860091/v1

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

Additional Declarations: There is **NO** Competing Interest.

Predicting Unprecedented Hot Temperature Extreme Occurrences Up to a Year in Advance

P. De Luca[1], J.C. Acosta Navarro[2], E.M. Fischer[3], and M.G. Donat[1,4]

[1]Barcelona Supercomputing Center (BSC), Barcelona, Spain
[2]European Commission, Joint Research Centre, Ispra, Italy
[3]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Corresponding author: Paolo De Luca, paolo.deluca@bsc.es

Abstract

Unprecedented hot temperature extremes (UTEx) severely stress human and natural systems by exceeding historical climate ranges and regionally even breaching potential adaptability limits, making reliable predictions crucial. Using initialised decadal hindcasts and historical forcing simulations from the Coupled Model Intercomparison Project Phase 6, we explore to what extent UTEx can be predicted on interannual time scales. We find that the occurrence of UTEx in the following year can be skilfully predicted in some predominantly tropical regions using initialised climate predictions, while historical climate simulations show little skill. This suggests that to predict the annual occurrences of UTEx, correctly capturing the phase of climate variability is key. UTEx in these regions occur predominantly during El Niño years, grounding the prediction of such extremes in tropical Pacific predictability. Our findings establish the feasibility of predicting annual occurrences of unprecedented climate extremes, providing a foundation for future early warning systems.

Main

Hot temperature extremes play a critical role in shaping societal, economic and environmental risks, particularly under a warming climate[1–3]. These events can cause significant impacts on human health[1], biodiversity[4,5], agriculture[6], energy systems[7], and infrastructure[8], often exacerbating existing inequalities[9,10]. In particular, hot extremes that surpass previous historical records tend to result in more severe impacts, as they expose human and natural systems to conditions outside the range they have previously experienced[11,12], which is often related to the level of adaptation. Such record-breaking extremes, hereafter defined as unprecedented hot temperature extremes (UTEx), are affected by global warming, so that their occurrence has intensified along with rising global temperatures driven by anthropogenic influence[11,13,14]. The probability of UTEx occurrences is projected to further increase across the globe under future warming scenarios[15] and is directly related to the forced warming rate and potentially amplified by increasing year-to-year temperature variability[12,16]. A notable UTEx was the 2021 Pacific Northwest summer heatwave, which caused many locations to break their historical maximum temperature records by more than 5°C, with Canada recording its historical national record of 49.6°C[17]. The impacts of this UTEx were enormous, with hundreds of deaths, damage to marine life and crops, river flooding from rapid snow and glacier melt and wildfires, making the heatwave one of the most extreme ever recorded globally[17,18].

In a stationary climate, the probability of observing an UTEx inherently decays inversely with the length of the observational record (1/n). However, anthropogenic forcing fundamentally disrupts this statistical baseline. Recent observations demonstrate that UTEx are occurring far more frequently than stationary theory predicts. Driven by rapid regional warming rates, in recent years UTEx have not only been declining slower than expected in a stationary climate but have regionally even become more frequent[11,12]. Yet, in addition to the long-term tendency there are also large year-to-year variations in the frequency of UTEx, with a tendency to cluster in particular years[12,16]—highlighting the potential for both the global warming trend and interannual variations to provide predictability. Previous studies have assessed the predictability of temperature extremes across various time scales. At time scales of numerical weather prediction of 10-15 days, the predictability

of UTEx is relatively high even though they are by definition outside the range of observations[17]. Multi-annual predictions have demonstrated good skill for summer temperature extremes defined with moderate thresholds over most global land areas[19,20], and skill has been found even in multi-decadal predictions[21]. Furthermore, the predictability of summer temperature extremes in the Northern Hemisphere has been linked to North Atlantic sea surface temperatures[22]. Interannual to decadal predictions have also been evaluated using a perfect-model setup, with results indicating skill across much of the globe[23]. These studies evaluating the predictability of temperature extremes have typically focused on indicators of relatively moderate extremes that occur at least once a year (such as the number of days above a moderate threshold, or the temperature of the hottest day of the year[24]). UTEx, however, are much rarer and often associated with more severe impacts, and the extent to which these are predictable on inter-annual time scales remains unknown.

To assess UTEx predictability, we use two multi-model ensembles (MMEs) of initialized predictions from the Decadal Climate Prediction Project (DCPP)[25] of the 6th phase of the Coupled Model Intercomparison Project (CMIP6), and un-initialized transient climate simulations (HIST)[26] in response to historical forcings (Tabs. S1–S2). DCPP contains 76 ensemble members from 6 different models, providing retrospective decadal predictions initialized from observed ocean and atmospheric conditions starting each year from 1961 to 2024. Model initialization can play a strong role in obtaining skillful climate predictions, with internal variability considered a fundamental aspect for these simulations[27]. On the other hand, HIST is composed of a total of 300 ensemble members from 33 models, which are simulations designed to simulate the response to external forcings from the mid-19th century to the near present, using observed forcings such as greenhouse gases, aerosols, solar variability, and volcanic activity. These simulations are uninitialized—meaning they do not start from observed ocean or atmospheric states, but from a preindustrial climate state consistent with the model's climate attractor—and therefore provide insight into the component of predictability arising solely from external forcings[26,27]. We use an ensemble of different reanalysis datasets, from which we identify UTEx in the real-world climate and use this to evaluate how well DCPP and HIST simulations can predict UTEx.

Here we explore the inter-annual predictability of UTEx occurrences. To this end, we consider the first calendar year after each yearly initialization from DCPP. We compute UTEx from annual maximum anomalies of 7-day moving averages of near-surface maximum temperatures (tasmax, K). The daily climatology was removed to calculate maximum anomalies, which are insensitive to seasonal variations. We also show results for UTEx computed on absolute values (i.e. with seasonality), occurring during the hottest period of the year, in the Supporting Information. We use a 7-day window to capture temporally persistent hot days, often representative of heatwaves, in contrast to daily maximum temperatures, which may represent more short-lived events. From these annual maxima, we compare against the highest temperatures that occurred in all previous years back to 1961. If the current year's annual maximum exceeds the previous annual maxima record, we flag it as an UTEx event (see Online Methods). To assess how well the DCPP and HIST MMEs can predict the observed UTEx occurrences, we use two complementary skill metrics specifically developed for handling binary datasets of extreme events, and these are the Symmetric Extremal Dependence Index (SEDI)[28] and the Brier Skill Score (BSS)[29], which assess deterministic and probabilistic skill, respectively (see Online Methods). After a global skill evaluation, we then shift our analysis to selected regions that show the highest annual prediction skill, namely Mainland Southeast Asia (MSEA), Northern South America (NSAm) and Western Africa (WAfr) (Fig. S1), and discuss the skillful UTEx predictions in the context of El Niño Southern Oscillation (ENSO) predictability.

Results

Global Skill Patterns

Global analysis of prediction skill over land grid cells (excluding Antarctica) for UTEx using DCPP, as measured by SEDI, demonstrate significant skill (p-value < 0.05) across a range of global land areas, including the broad regions of Northern South America, Eastern Europe, Western Africa, India, Mainland Southeast Asia, and Northern Australia (Fig. 1a). HIST also exhibit significant SEDI skill in a few areas—notably parts of North America, South and East Africa, and Western to Central Russia (Fig. 1b). This indicates that external forcing alone can impart a baseline degree of predictability in

some regions. Crucially, however, DCPP skill substantially exceeds HIST skill in most of these overlapping regions and provides skill in larger areas (27.4% vs 18.5% of global land with positive significant skill, Fig. 1a,b), confirming that initialization contributes a distinct predictive signal well beyond the underlying forced response. We complement the analysis with the BSS, considering also probabilistic aspects of the predictions. The BSS shows spatially more limited skill than SEDI, with DCPP retaining significant skill primarily over Equatorial, Western, and Southern Africa, Northern South America, and localized sectors of central and eastern North America (Fig. 1c). In stark contrast, the HIST ensemble fails to yield any significantly skillful regions globally under the stricter BSS metric (10.0% of global land area with significantly positive BSS in DCPP vs 0.0% in HIST, Fig. 1c,d). This indicates that the warming trend in response to forcing is in most regions not sufficient to predict the annual occurrence of UTEx, and that initialization to align the phasing of climate variability is needed to predict these extremes in most regions where UTEx are predictable (Fig. S2). Overall, similar patterns of skill are found for UTEx computed using absolute values, which capture the occurrence of UTEx during the hottest period of the year (primarily summer), as opposed to UTEx based on anomalies, where seasonality is removed (Figs. S3–S4). Note, however, that fewer areas exhibit significant SEDI skill in the northern hemisphere extra-tropics (Fig. S3a), compared to UTEx computed on anomalies (Fig. 1a). Skill patterns also look similar when evaluating against individual observations-based datasets, indicating the robustness of the results with respect to dataset choice (Figs. S5–S6). However, when compared to JRA3Q, we note that ERA5 contains high temperature values in the early part of the record (i.e. the 1960s) over the Amazon and equatorial Africa, which prevents the detection of UTEx and, consequently, the assessment of their prediction skill in these regions.

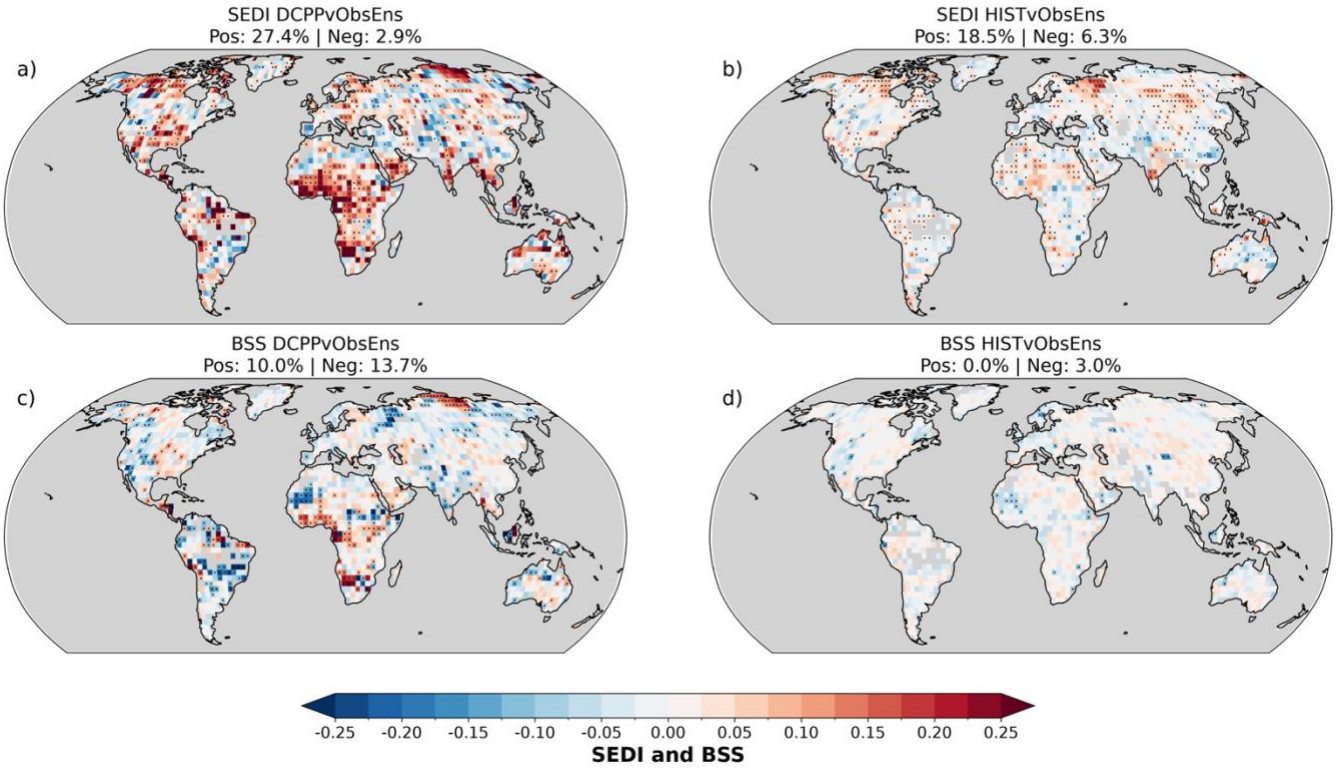


Figure 1. Annual prediction skill of unprecedented temperature extremes (UTEx). (a)-(b) Symmetric Extremal Dependence Index (SEDI) and (c)-(d) Brier Skill Score (BSS) of Coupled Model Intercomparison Project Phase 6 (CMIP6) Decadal Climate Prediction Project (DCPP) and Historical (HIST) simulations. All panels use the mean of the observations-based datasets (ObsEns) as observational dataset and cover the period 1971-2024. The reference forecast is the theoretical expectation of UTEx under stationary conditions (see Online Methods). DCPP uses the first calendar year after initialization, and stippling indicates grid-points where skill is statistically significant (p-value <0.05, two-sided bootstrap with n=1000 samples). Pos and Neg indicate the percentage (%) of total global land area (excluding Antarctica) with significant positive and negative skill.

Global Land Area Affected by Unprecedented Heat

Next, we quantify the percentage of global land area affected by annual UTEx across the DCPP, HIST, and the observations-based reanalysis datasets (Fig. 2). The time series exhibit generally elevated values—up to ~12% of the global land area —during the early 1970s, which is due to the UTEx identification method: the shorter observational record starting in 1961 leads to higher likelihood of setting UTEx in the first decades. A decreasing trend follows, and from the early 1990s onward, the time series stabilize, indicating a relatively stationary phase that signifies that the extreme records keep being broken but in a smaller fraction of the land area. Among the datasets, the observations exhibit the highest interannual variability, with notable peaks in 1973, 1983, 1998, 2010 and 2024—

all years that began with El Niño events[30–34]. This interannual variability is reasonably well captured by DCPP (Spearman's ρ = 0.51, p-value <0.01, from 1971 to 2024), although the magnitude of the peaks is often underestimated in the ensemble mean compared to the reanalyses; note that the observed UTEx area lays for most years well within the DCPP ensemble spread (Fig. S7). In contrast, HIST fails to reproduce the timing of the variability and exhibits a much smoother time series with limited year-to-year fluctuations (ρ = 0.26, p-value <0.10), highlighting again the superior skill of DCPP in predicting UTEx and the added value of initialization over uninitialized simulations. Comparable results are found when UTEx is computed using absolute values (Fig. S8), and again, the findings are robust when using individual observations-based datasets to evaluate the skill of the hindcasts (Fig. S9).

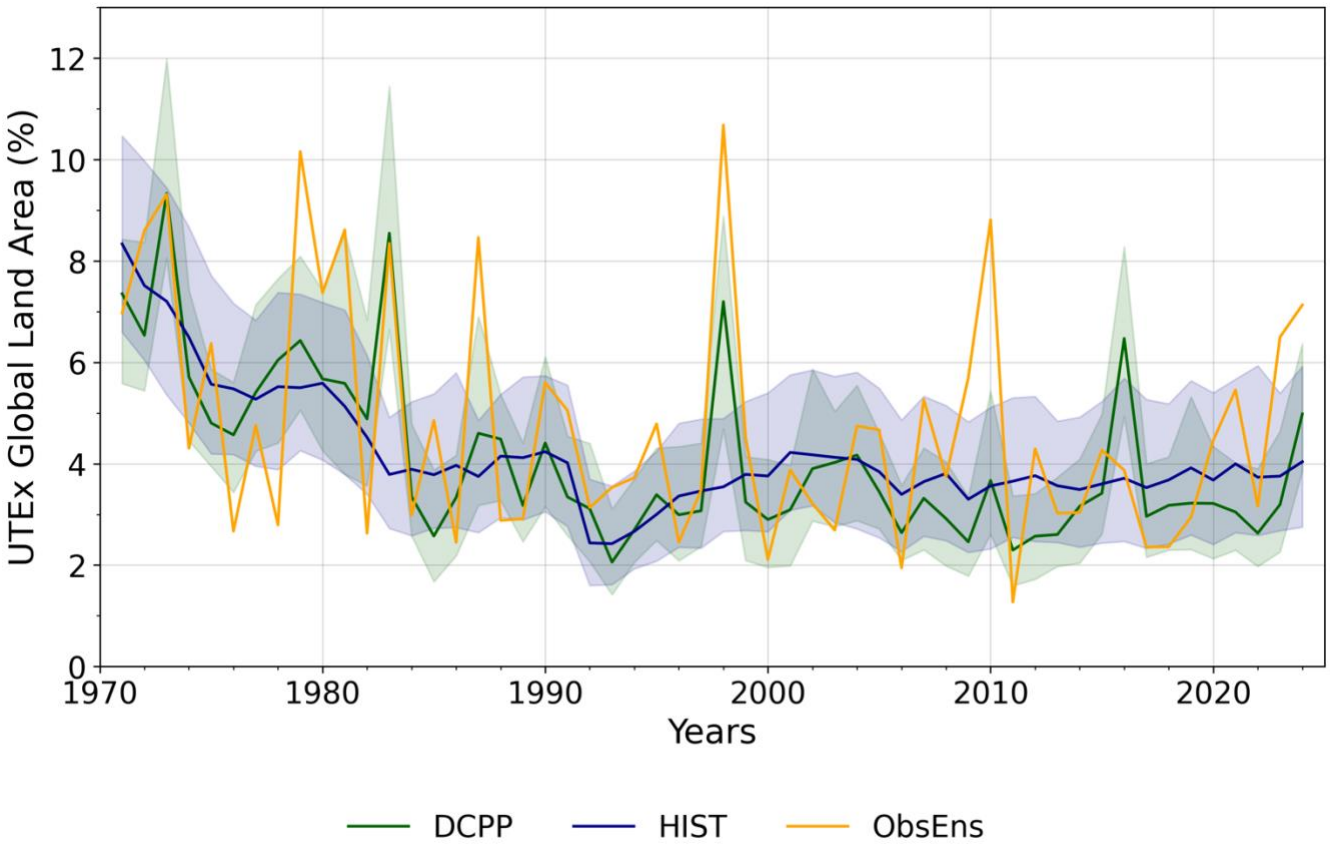


Figure 2. Global land area (%) with UTEx occurrences from 1971 to 2024. Dark green line shows DCPP multi-model ensemble (MME) median, dark blue HIST MME median and orange the observational ensemble (ObsEns). The dark green and dark blue shaded areas represent the interquartile range (i.e. 25th and 75th percentiles) of respectively DCPP and HIST. Global land area does not include Antarctica.

Regional Interannual Variability

After providing a global perspective, we now focus on three regions, such as Mainland Southeast Asia (MSEA), Northern South America (NSAm) and Western Africa (WAfr), which exhibited high prediction skill for annual UTEx (Fig. S1). For these regions, we compute annual tasmax anomalies relative to the full climatological period (1961–2024) using DCPP and HIST single members, as well as the observational ensemble (Fig. 3). Consistent with the global analysis (Fig. 2), DCPP (Fig. 3a,c,e) more accurately captures the interannual variability of observations compared to HIST (Fig. 3b,d,f). The DCPP ensemble means generally follow the observed values, although in some peak years the observations fall outside the interquartile range of the ensemble. This agreement is corroborated by Spearman's ρ correlation coefficients, which are higher for DCPP versus Observations ($0.57 \leq \rho \leq 0.73$, p-value < 0.01) than for HIST versus Observations ($0.36 \leq \rho \leq 0.56$, p-value < 0.05). Among the three regions, DCPP shows the strongest correlation with Observations in NSAm ($\rho = 0.73$), while HIST performs best in WAfr ($\rho = 0.56$). HIST also shows much smaller interannual variations in tasmax, along with a smoother time-series with a gradual increase in temperatures. Similar patterns emerge when UTEx is calculated using absolute values (Fig. S10), with the main difference being that WAfr yields the highest DCPP–Observations correlation ($\rho = 0.79$, Fig. S10e). These regional patterns are also reproduced when comparing against the individual observations-based datasets (Figs. S11–S12, Tabs. S3-S4).

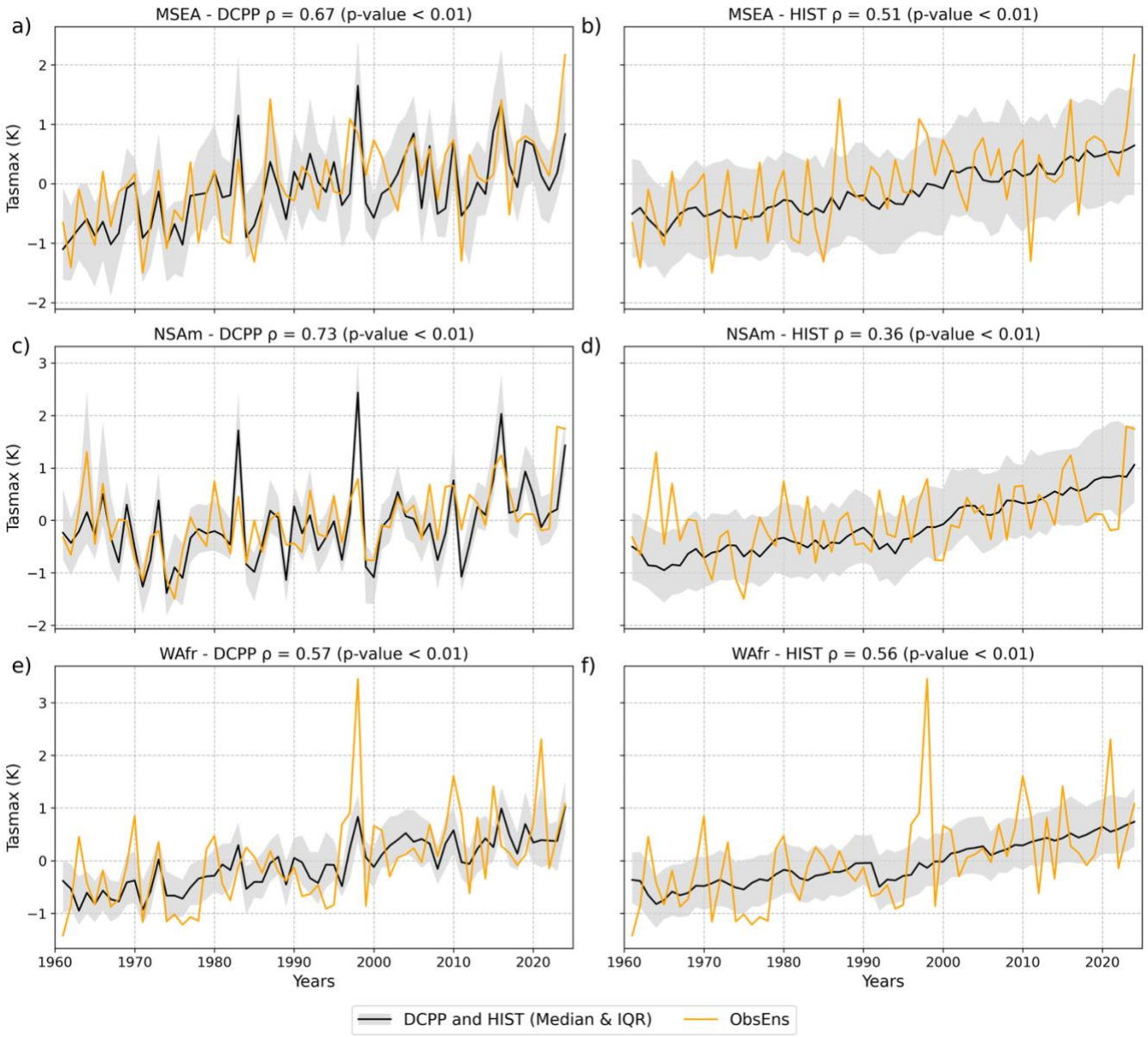


Figure 3. Annual maxima of 7-day average regional near-surface maximum temperature anomalies (tasmax, K). (a)-(b) Mainland Southeast Asia; (c)-(d) Northern South America (NSAm); and (e)-(f) Western Africa (WAfr). (a),(c),(e) DCPP MME and (b),(d),(f) HIST MME. All data span from 1961 to 2024, with the full period mean removed to compute the anomalies. In all panels the ObsEns median is shown as orange lines. All panels show the Spearman's ρ correlation coefficient and corresponding p-value computed between the DCPP, HIST MME medians and the ObsEns median.

For the same three regions (MSEA, NSAm and WAfr), we further quantify the percentage of ensemble members in DCPP and HIST that predict UTEx events in the different years and compare these with the binary UTEx record from observations (i.e. event/no event, Fig. 4). Notable examples

include the years 1973, 1987, 1998, 2010, 2016 and 2024 for MSEA (Fig. 4a); 1983, 1998, 2010, 2016 and 2024 for NSAm (Fig. 4b); and 1973, 1983 and 1998 for WAfr (Fig. 4c). In contrast, HIST shows relatively uniform UTEx occurrence probabilities across all years, close to the curve of the theoretically expected rate of UTEx in a stationary climate but showing slightly higher rates in recent decades, indicative of a warming (i.e. non-stationary) climate. Importantly, HIST also lacks the pronounced clustered UTEx occurrence in few specific years seen in observations and therefore fails to predict the timing of observed events. This discrepancy underscores the added value of DCPP in skillfully predicting annual UTEx events, as the regional interannual variability is captured through initialization. Additionally, DCPP predicts a UTEx event over NSAm in 1998 that is not captured by any of the reanalysis datasets (Fig. 4b). This discrepancy represents either a false positive in the model or a missed event due to uncertainties in the observational data. We further evaluate the regional annual UTEx prediction skill for these regional averages and find, consistent with Figure 1, that DCPP exhibits statistically significant SEDI and BSS scores (p-value <0.05), while HIST shows lower skill with scores that are not statistically significant. Additionally, DCPP shows a clear BSS-added value over HIST for MSEA and WAfr (Tab. S5). Between 20% and 80% of DCPP members typically predict UTEx events in years when these also occurred based on observations. Similar findings are obtained when UTEx are computed using absolute thresholds (Fig. S13, Tab. S5) and across the individual observations-based datasets (Fig. 4 and Tabs. S6-S8).

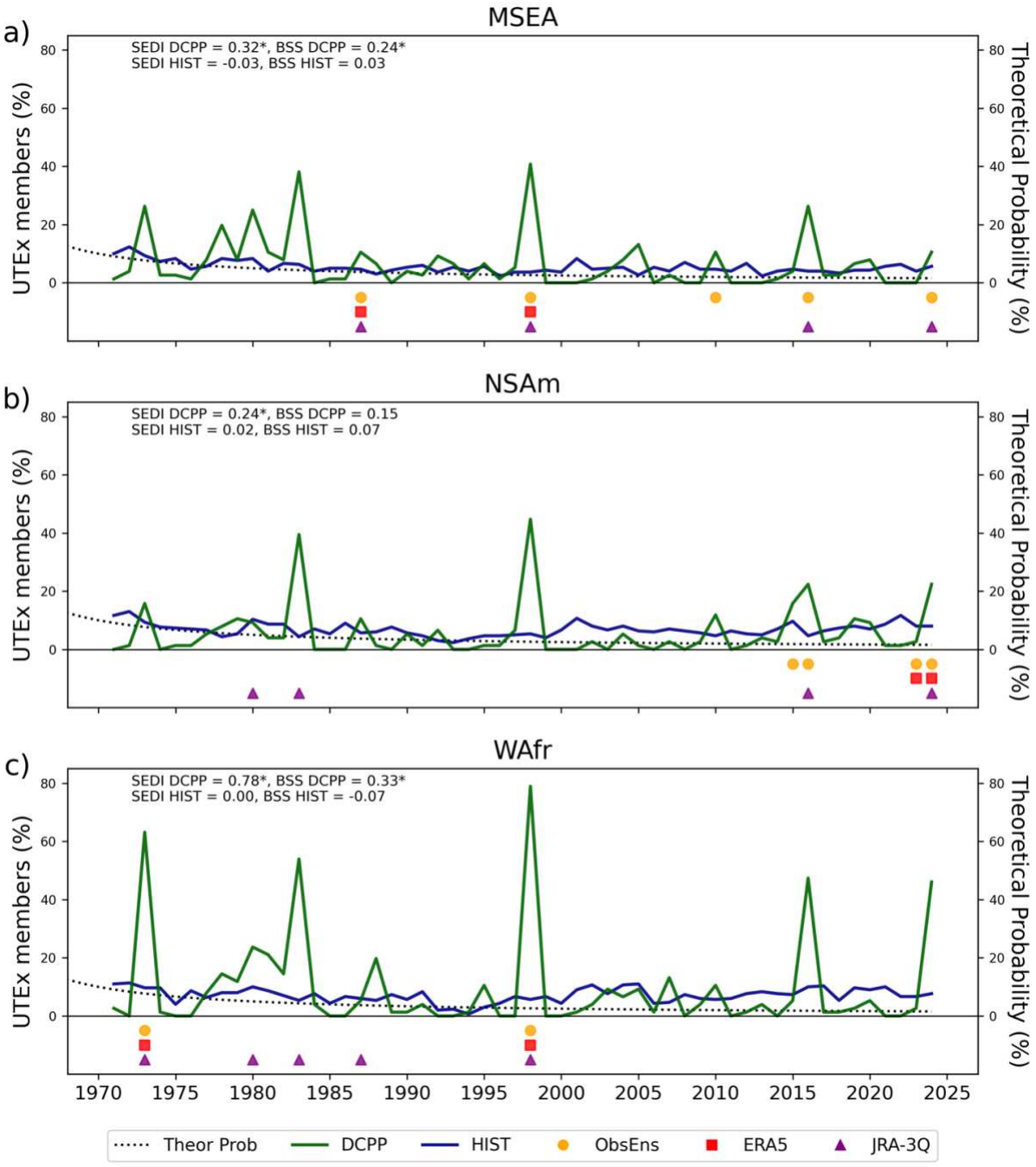


Figure 4. Percentage (%) of DCPP and HIST members with regional UTEx from 1971 to 2024. The yellow symbols represent the occurrence of UTEx for ObsEns, red for ERA5 and purple for JRA-3Q reanalyses. The dashed line represents the UTEx theoretical probability of occurrence (%) in a stationary climate (see Online Methods). DCPP and HIST regional annual prediction skill are shown at the top of each panel as SEDI and BSS scores. Skill metrics evaluate the hindcasts in capturing events in ObsEns, and the reference forecast for BSS is the theoretical expectation of UTEx occurrences. Asterisks indicate statistically significant regional prediction skill (p-value <0.05).

The Role of El Niño in the Predictability of UTEx

El Niño events are known to exacerbate global and regional temperature extremes, including in the MSEA, NSAm and WAfr regions[35–38], where we obtained the highest annual prediction skill for UTEx. Therefore, we next investigate the links between UTEx and El Niño by quantifying their joint occurrences. We find that a high fraction (>0.9) of UTEx events is observed to occur during years that start with El Niño conditions in some regions, including our three focus regions of high skill (Fig. 5a). Similarly, although with smoother spatial variability, DCPP also points to a higher fraction of UTEx during El Niño years in some regions including these three skillful regions (fractions ranging from 0.5 to 0.8, Fig. 5b). In contrast, HIST shows relatively low values of about 0.1 to 0.3 UTEx events occurring during observed El Niño years in all regions, indicating that without initialization the climate simulations do not reproduce the link between UTEx and El Niño events. These findings suggest that the capability of the DCPP hindcasts in predicting ENSO (e.g. the Oceanic Niño Index (ONI), over the Niño3.4 region, is skillfully predicted at interannual time scales; Pearson's *r* correlation=0.61, p-value <0.01) also provides skill for predicting UTEx occurrences in regions where these events occur in relation to El Niño events. This association of unprecedented events with El Niño years is also robust when computing UTEx based on absolute temperature values (Fig. S14). Teleconnections can differ between canonical Eastern Pacific and Central Pacific El Niño events[39]. Indeed, we also find some differences in the global composite maps when separated by El Niño type (Figs. S15-S16). For example, WAfr shows the largest frequency of UTEx events associated with Eastern Pacific events (Figs. S15b-S16b). This is also reflected in Figure 4c, which shows the UTEx events in WAfr consistent across observational datasets in 1973 and 1998 (both Eastern Pacific El Niño events; Tab. S9)[40], and highest DCPP forecast probabilities in these same years.

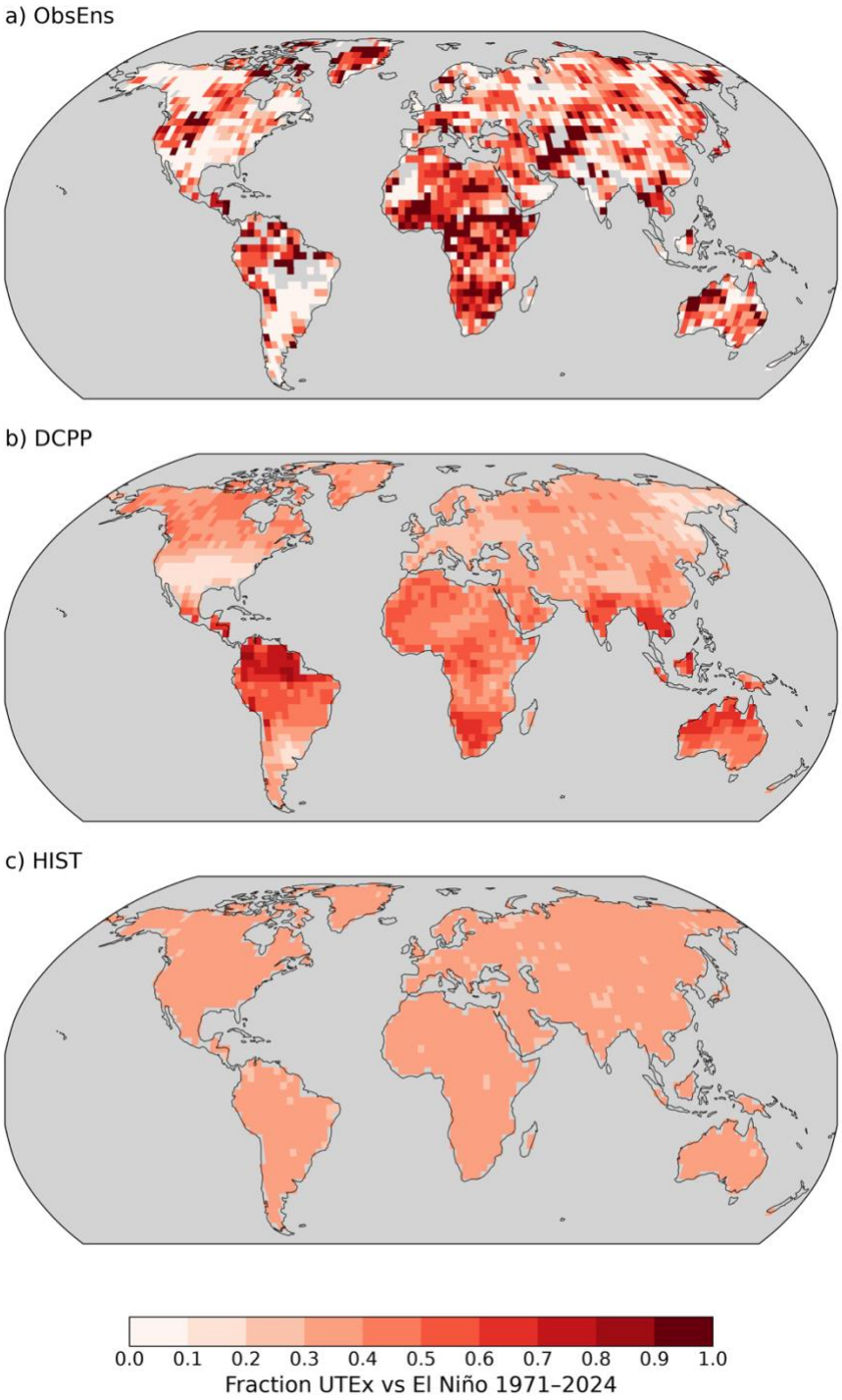


Figure 5. Fraction of UTEx occurring during El Niño years from 1971 to 2024. El Niño years are obtained from the Oceanic Niño Index (ONI), which is computed with ERSST.v5[41] observational data over the Niño3.4 region (5°N-5°S, 170°-120°W, https://www.ncei.noaa.gov/access/monitoring/enso/sst).

Discussion

Our findings demonstrate that annual UTEx occurrences can be skillfully predicted on interannual time scales in some regions where their occurrence is controlled by ENSO. These predictions require initialized simulations (i.e. DCPP) to capture the relevant interannual variations. In contrast, transient historical simulations (i.e. HIST) largely fail to capture the observed annual UTEx occurrences. This also highlights that external forcing alone does not sufficiently predict the occurrence of UTEx in specific years, whereas the combination of forcing and initialization significantly enhances predictive skill—particularly over Mainland Southeast Asia, Northern South America, and Western Africa, where DCPP exhibits statistically significant skill across multiple verification metrics. The HIST simulations did show isolated areas of positive and statistically significant skill for SEDI—such as in parts of North America, South and East Africa, and Western to Central Russia—meaning that forcing alone can, to some extent, predict annual occurrences of UTEx in these regions. However, initialization in the DCPP typically leads to further enhanced skill also in these regions. While we demonstrate predictive skill for UTEx in 7-day average time scales, we expect it to extend also to somewhat shorter and longer aggregation periods or to ocean temperatures that are directly related to ENSO variability.

While our study focuses on predicting UTEx over the next calendar year, the occurrence and predictability of these extremes vary throughout that period. DCPP predictive skill for UTEx is notably higher during the first half of the year (Fig. S17) compared to both the second half (Fig. S18) and the full calendar year (Fig. 1). Nevertheless, the second half of the year still exhibits significant SEDI skill in regions such as Northern South America and South Asia (Fig. S18). Furthermore, time series of global area fractions affected by UTEx demonstrate significant correlations between DCPP predictions and observations in both the first and second half of the year following initialization, consistently outperforming the uninitialized HIST ensemble (Fig. S19). Beyond the closer proximity to initialization in the first six months, this seasonal disparity in skill also likely reflects the seasonal characteristics of ENSO evolution[42] and UTEx occurrence. Specifically, larger global land areas are affected by UTEx during El Niño events in the first half of the year than in the second half (Fig. S19), indicating a preferential occurrence during the decaying phase of El Niño. Conversely, skill during

the second half of the year in regions like Northern South America may be linked to UTEx preferentially occurring during the build-up phase of El Niño (Figs. S20-S21).

In our study we aimed at analyzing the year-to-year predictability of UTEx over a long hindcast period (1961-2024), also including recent years characterized by UTEx in some regions. While a larger number of models provide hindcasts with initializations until 2014 (within the dcppA experiment of CMIP6), the extension of the analysis period until 2024 implies using a reduced ensemble of hindcasts (see Online Methods). Previous studies have shown a benefit of large decadal prediction ensembles, and reduced skill of multi-annual predictions when using smaller ensembles[43–45]. We therefore also assess UTEx prediction skill up to 2014 using a larger DCPP hindcast (158 members from 12 models) and HIST (525 members from 45 models) MMEs (Figs. S22-S23). While the spatial patterns of positive and statistically significant skill (p-value <0.05) are generally preserved between the predictions up to 2024 and those limited to 2014, the performance of the larger MMEs until 2014 is slightly improved compared to their smaller counterpart. We find slightly increased DCPP skill over north America and Europe (Fig. S22) along with improved added skill over Northern South America (Fig. S23). Nonetheless, as operational forecast products rely on the most up-to-date predictions (i.e. including data beyond 2014), broader participation from modeling centers providing forecasts could therefore help enable more skillful UTEx predictions and early warnings.

Previous studies have shown that as a result of the high anthropogenic rate of warming, the probability of UTEx has declined slower than expected in a stationary climate and has even increased in recent years given the high rate of warming[11,46,47], a trend that is projected to continue into the future. While uninitialized projections capture the overall tendency in UTEx probabilities, they largely fail to predict their year-to-year variations. The occurrence probability of UTEx is high during the first decades of the evaluation period, as the availability of DCPP hindcasts limits UTEx calculations to the period after 1960 and the theoretical probability of UTEx decreases with time. To confirm that these initially high rates do not bias our results, we also performed the evaluation only over the later decades from 1990 onwards when UTEx rates stabilized (Figs. 2 and 4). Our findings

remain robust, and computing the skill over the shorter 1990–2024 period yields results which are comparable to the full 1971–2024 baseline (Figs. S24–S25).

We here show for the first time that initialized prediction systems can skillfully predict the UTEx probability of the coming year, with largest skill in some low-latitude regions. We show that the predictability of UTEx is related to the predictive skill of ENSO variability, that directly relates to the annual occurrences of such extremes in some regions. The relation of ENSO variability to UTEx is non-linear and the area affected by UTEx is particularly high in El Niño years relative to neutral years, whereas the difference is less pronounced between La Niña and neutral years (Fig. S26). Therefore, our results collectively suggest that DCPP currently represent a suitable tool for predicting annual UTEx events in tropical regions affected by El Niño. Given the high potential impacts of UTEx, which have never been seen in the observational record, it is highly relevant not only to understand the overall long-term tendency but also the probability of such events in the coming year. Thus, our results provide a foundation for developing early warning systems, thereby enhancing preparedness for anticipated UTEx events and mitigating their most severe impacts.

Online Methods

CMIP6 and observational ensemble data

We analyze DCPP[25] and HIST[26] simulations. In total we use 76 members from 6 different models for DCPP and 300 members and 33 models for HIST, for the hindcast period 1961-2024 (Tabs. S1–S2). DCPP are initialized each year and simulations of each initialization extend for ten years, allowing assessment of near-term climate predictability originated for example from modes of climate variability in addition to the external forcings[25]. From the DCPP data, we use the first calendar year after each initialization. For computing the UTEx, we extract from these data daily tasmax (K), so that for DCPP we include all dcppA-hindcast (1961-2014) concatenated with dcppB-forecast (2015-2024) experiments, which are fewer than the dcppA-hindcast alone. On the other hand, for HIST, we use Historical (1961-2014) and Historical concatenated with Shared Socio-economic Pathway 2-4.5 (2015-2024) experiments. To estimate UTEx in observations-based data we also use the European Centre for Medium-Range Weather Forecasts Reanalysis 5th generation (ERA5)[48] and

Japanese Reanalysis for Three Quarters of a Century (JRA3Q)[49], which form an ensemble of renalysis datasets that cover the hindcast and forecast period provided by DCPP. For the CMIP6 MMEs and the two reanalysis datasets we first remap the tasmax data to a common 2.8° x 2.8° grid, which represents the horizontal resolution of the coarsest model available (i.e. CanESM5) and then compute the UTEx. After obtaining the UTEx, we also apply a land-sea mask so that only grid-points over land are retained. In the main text we show results only for the ensemble mean of observations-based datasets (i.e. ObsEns) to compensate the observational uncertainties and show results for the individual observations-based datasets in the Supplementary Information. As part of the analysis, we also investigated the interannual variability of tasmax and UTEx in three regions that yield high annual prediction skill. These regions are Mainland Southeast Asia (MSEA, 91°E–112°E, 8°N–23°N); Northern South America (NSAm, 80°W–50°W, 5°S–12°N); and Western Africa (WAfr, 20°W–20°E, 0°N–20°N) (Fig. S1).

Computation of UTEx

We compute yearly UTEx for all the datasets using tasmax 7-day moving averages. Our method starts quantifying UTEx eleven years after the start date of the hindcast period (i.e. in 1971) and for each year it compares against the maximum value of all previous years since 1961. We begin the quantification of UTEx in 1971 to ensure a sufficiently long historical period for establishing a reliable baseline. Records are inevitably set in the first year of a time series, and the likelihood of breaking these early records remains high during the initial years. We do not consider the first 10 years (1961–1970) of the UTEx calculation in the analysis and for the skill evaluation. Accordingly, for UTEx events in 2024, we use a 62-year reference window spanning 1961–2023. We also confirm the robustness of our results when evaluating the hindcast skill only over the shorter period 1990-2024, when theoretical UTEx probabilities have stabilized after still somewhat increased rates in the 1970s and 1980s (Figs. S24-S25).

The details of our approach can be described as follows: i) for each grid-point and daily tasmax value remove the seasonal cycle, by subtracting from the temperature at each calendar day its daily climatology computed from the entire period 1961-2024; ii) apply a 7-day running mean to these

daily anomalies; iii) compute the annual maximum of the 7-day running means; iv) for each year, starting from 1971 and up to 2024, assess whether its annual maximum exceeds the ones found in the previous years; v) if they exceed the previous records they are labelled as UTEx. This approach diagnoses unprecedented temperature anomalies that can occur at any time of the year.

In addition to UTEx computed from anomalies, we also compute, using the same datasets, UTEx based on absolute temperature values. This approach targets the hottest days of the year, typically in summer, and it is the same as described above with the only difference being that here we do not remove the daily climatology and proceed in computing the 7-day running mean directly from the daily tasmax data. We note that results for the UTEx based on absolute values are qualitatively similar to the ones derived from anomalies and we show them in the Supplementary Information.

Skill metrics

For assessing how well the DCPP and HIST MMEs predict observational UTEx we use two complementary skill metrics that are specifically tailored for handling binary climate data, in our case 1 being a yearly record of a UTEx event and 0 a non-record UTEx event.

The first skill metric is the Symmetric Extremal Dependence Index (SEDI)[28], which is insensitive to event frequency or is not affected by how rare the event is, is symmetric and bounded as $-1 < SEDI < 1$. The SEDI is computed from the contingency table using the hit rate (H) and false alarm rate (F) as follows:

$$SEDI = \frac{\ln(F) - \ln(H) + \ln(1-H) - \ln(1-F)}{\ln(F) + \ln(H) + \ln(1-H) + \ln(1-F)}$$

where $H = \frac{hits}{hits+misses}$ and $F = \frac{false\ alarms}{false\ alarms+correct\ negatives}$. $SEDI = 1$ represents the perfect forecast, $SEDI = 0$ an unskillful random forecast and $SEDI < 0$ a forecast worse than a random one.

The second skill metric is the Brier Skill Score (BSS)[29] which is a standardized metric that evaluates how much a probabilistic forecast improves upon a baseline prediction. It is defined as:

$$BSS = 1 - \frac{BS_{forecast}}{BS_{reference}}$$

where $BS_{forecast}$ is the Brier Score (BS)[50] of the prediction model, expressed as:

$$BS_{forecast} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

where N is the number of the forecasts, f_i is the forecast probability for event i calculated from the fraction of ensemble members predicting an event and o_i is the observed binary outcome, in our case represented by the DCPP or HIST ensemble members and Observations respectively. $0 < BS_{forecast} < 1$ where 0 represents the perfect forecast and 1 the worst one. On the other hand, the $BS_{reference}$ serves as the benchmark against which $BS_{forecast}$ is evaluated and can be expressed as:

$$BS_{reference} = \frac{1}{N}\sum_{i=1}^{N}(p - o_i)^2$$

Where in our analysis we define p as the chance of setting a record at the nth year, which equals the probability that all previous values are smaller. For a stationary, independently drawn random variable, this likelihood decreases steadily with time. The first sample is always unprecedented. The probability that the second sample surpasses the first is 1/2, for the third to exceed the first two is 1/3, etc[51]. Thus, the probability of exceedance in the nth year is $p(n)$ =1/n[15]. The first ten years are excluded from the skill evaluations, to avoid focusing on unprecedented events that occur primarily due to the short record during the first years.

Therefore, the BSS represents the percentage improvement in squared error relative to the baseline. $\infty < BSS < 1$, with $BSS = 1$ being the perfect forecast, $BSS = 0$ a forecast with no improvement over the reference and $BSS < 0$ a forecast worse than the reference.

We employ both SEDI and BSS because they evaluate different but complementary aspects of how the DCPP and HIST ensembles predict observed UTEx. SEDI provides a deterministic assessment of whether a model correctly captures the binary occurrence of UTEx events, whereas the BSS provides a probabilistic evaluation of the model's reliability in simulating the likelihood of such extremes.

To assess the statistical significance of both SEDI and BSS, we perform a bootstrapping test at grid-point level using n=1000 random resampling (with repetition) from the original timeseries. The resampled timeseries is of the same size as the original series and is then used to compute the SEDI and BSS for each random selection, and to compute the 95% confidence intervals, so that only grid-points with p-value < 0.05 are stippled.

UTEx during El Niño years

To quantify the fraction of UTEx occurring during El Niño years we use the National Oceanic and Atmospheric Administration (NOAA) Oceanic Niño Index (ONI, https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php), which is composed by 3-month running means of ERSST.v5 sea surface temperature anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W), with El Niño events identified as five consecutive anomalies of minimum +0.5°C. From this dataset we extract data from 1971 to 2024 and find that the highest fraction of UTEx occur in years starting with an El Niño month, which are: 1973, 1977, 1978, 1980, 1983, 1987, 1988, 1992, 1995, 1998, 2003, 2005, 2007, 2010, 2015, 2016, 2019 and 2024. Then, for each grid-point, we assess the fraction of UTEx which occur within these 18 El Niño years.

Acknowledgments

This study has been funded by the Horizon Europe projects ASPECT (Grant no. 101081460) and EXPECT (Grant no. 101137656). The authors are grateful to Margarida Samso Cabre for downloading the datasets used in the analysis.

Data availability

All CMIP6 data used in the analysis (i.e. DCPP and HIST MMEs) are available in the ESGF website (https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/). ERA5 renalysis data are available in the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets) and JRA3Q reanalysis data in the JMA website (https://www.data.jma.go.jp/jra/html/JRA-3Q/index_en.html). Source Data are provided with this paper.

Code availability

All the codes used in the analysis are available as supplementary files.

References

1. Ebi, K. L. *et al.* Hot weather and heat extremes: health risks. *Lancet* **398**, 698–708 (2021).
2. Sun, Q. *et al.* Global heat stress on health, wildfires, and agricultural crops under different levels of climate warming. *Environ. Int.* **128**, 125–136 (2019).
3. Sun, Y. *et al.* Global supply chains amplify economic costs of future extreme heat risk. *Nature* **627**, 797–804 (2024).
4. Trisos, C. H., Merow, C. & Pigot, A. L. The projected timing of abrupt ecological disruption from climate change. *Nature* **580**, 496–501 (2020).
5. Meyer, A. S. *et al.* Temporal dynamics of climate change exposure and opportunities for global marine biodiversity. *Nat. Commun.* **15**, 5836 (2024).
6. Tran, B.-L., Tseng, W.-C. & Chen, C.-C. Climate change impacts on crop yields across temperature rise thresholds and climate zones. *Sci. Rep.* **15**, 23424 (2025).

7. Liu, L. *et al.* Climate change impacts on planned supply–demand match in global wind and solar energy systems. *Nat. Energy* **8**, 870–880 (2023).
8. Rotta Loria, A. F. The silent impact of underground climate change on civil infrastructure. *Commun. Eng.* **2**, 44 (2023).
9. Schöngart, S., Nicholls, Z., Hoffmann, R., Pelz, S. & Schleussner, C.-F. High-income groups disproportionately contribute to climate extremes worldwide. *Nat. Clim. Chang.* **15**, 627–633 (2025).
10. Callahan, C. W. & Mankin, J. S. Globally unequal effect of extreme heat on economic growth. *Sci. Adv.* **8**, eadd3726 (2025).
11. Fischer, E. M. *et al.* Record-breaking extremes in a warming climate. *Nat. Rev. Earth Environ.* **6**, 456–470 (2025).
12. Fischer, E. M., Sippel, S. & Knutti, R. Increasing probability of record-shattering climate extremes. *Nat. Clim. Chang.* **11**, 689–695 (2021).
13. King, A. D. Attributing Changing Rates of Temperature Record Breaking to Anthropogenic Influences. *Earth's Futur.* **5**, 1156–1168 (2017).
14. Diffenbaugh, N. S. *et al.* Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci.* **114**, 4881–4886 (2017).
15. Benestad, R. E., Lussana, C. & Dobler, A. Global record-breaking recurrence rates indicate more widespread and intense surface air temperature and precipitation extremes. *Sci. Adv.* **10**, eado3712 (2025).
16. Coumou, D., Robinson, A. & Rahmstorf, S. Global increase in record-breaking monthly-mean temperatures. *Clim. Chang.* **118**, 771–782 (2013).
17. White, R. H. *et al.* The unprecedented Pacific Northwest heatwave of June 2021. *Nat. Commun.* **14**, 727 (2023).
18. Thompson, V. *et al.* The 2021 western North America heat wave among the most extreme events ever recorded globally. *Sci. Adv.* **8**, eabm6860 (2025).
19. Delgado-Torres, C. *et al.* Multi-annual predictions of the frequency and intensity of daily temperature and precipitation extremes. *Environ. Res. Lett.* **18**, 34031 (2023).
20. Aranyossy, A. *et al.* Multi-annual predictions of hot, dry and hot-dry compound extremes.

Earth Syst. Dyn. **16**, 2225–2251 (2025).

21. De Luca, P., Delgado-Torres, C., Mahmood, R., Samso-Cabre, M. & Donat, M. Constraining decadal variability regionally improves near-term projections of hot, cold and dry extremes. *Environ. Res. Lett.* (2023).
22. Borchert, L. F. *et al.* Decadal Predictions of the Probability of Occurrence for Warm Summer Temperature Extremes. *Geophys. Res. Lett.* **46**, 14042–14051 (2019).
23. Liu, Y., Donat, M. G., Rust, H. W., Alexander, L. V & England, M. H. Decadal predictability of temperature and precipitation means and extremes in a perfect-model experiment. *Clim. Dyn.* **53**, 3711–3729 (2019).
24. Zhang, X. *et al.* Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Clim. Chang.* **2**, 851–870 (2011).
25. Boer, G. J. *et al.* The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* **9**, 3751–3777 (2016).
26. Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
27. Meehl, G. A. *et al.* Initialized Earth System prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.* **2**, 340–357 (2021).
28. Ferro, C. A. T. & Stephenson, D. B. Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events. *Weather Forecast.* **26**, 699–713 (2011).
29. Wilks, D. S. *Statistical Methods in the Atmospheric Sciences*. (Elsevier, Amsterdam, the Netherlands, 2019).
30. Cane, M. A. Oceanographic Events During El Niño. *Science (80-.).* **222**, 1189–1195 (1983).
31. McPhaden, M. J. Genesis and Evolution of the 1997-98 El Niño. *Science (80-.).* **283**, 950–954 (1999).
32. Kim, W., Yeh, S.-W., Kim, J.-H., Kug, J.-S. & Kwon, M. The unique 2009–2010 El Niño event: A fast phase transition of warm pool El Niño to La Niña. *Geophys. Res. Lett.* **38**, (2011).
33. Trenberth, K. E. The Definition of El Niño. *Bull. Am. Meteorol. Soc.* **78**, 2771–2778 (1997).
34. Jiang, N. *et al.* Enhanced risk of record-breaking regional temperatures during the 2023–24

El Niño. *Sci. Rep.* **14**, 2521 (2024).
35. Thirumalai, K., DiNezio, P. N., Okumura, Y. & Deser, C. Extreme temperatures in Southeast Asia caused by El Niño and worsened by global warming. *Nat. Commun.* **8**, 15531 (2017).
36. Lin, L., Chen, C. & Luo, M. Impacts of El Niño–Southern Oscillation on heat waves in the Indochina peninsula. *Atmos. Sci. Lett.* **19**, e856 (2018).
37. Cai, W. *et al.* Climate impacts of the El Niño–Southern Oscillation on South America. *Nat. Rev. Earth Environ.* **1**, 215–231 (2020).
38. Arblaster, J. M. & Alexander, L. V. The impact of the El Niño-Southern Oscillation on maximum temperature extremes. *Geophys. Res. Lett.* **39**, (2012).
39. Capotondi, A. *et al.* Understanding ENSO Diversity. *Bull. Am. Meteorol. Soc.* **96**, 921–938 (2015).
40. Kao, H.-Y. & Yu, J.-Y. Contrasting Eastern-Pacific and Central-Pacific Types of ENSO. *J. Clim.* **22**, 615–632 (2009).
41. Huang, B. *et al.* Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).
42. Chen, H.-C. & Jin, F.-F. Fundamental Behavior of ENSO Phase Locking. *J. Clim.* **33**, 1953–1968 (2020).
43. Delgado-Torres, C. *et al.* Multi-Model Forecast Quality Assessment of CMIP6 Decadal Predictions. *J. Clim.* **35**, 4363–4382 (2022).
44. Smith, D. M. *et al.* Robust skill of decadal climate predictions. *npj Clim. Atmos. Sci.* **2**, 13 (2019).
45. Athanasiadis, P. J. *et al.* Decadal predictability of North Atlantic blocking and the NAO. *npj Clim. Atmos. Sci.* **3**, 20 (2020).
46. Bador, M., Terray, L. & Boé, J. Detection of anthropogenic influence on the evolution of record-breaking temperatures over Europe. *Clim. Dyn.* **46**, 2717–2735 (2016).
47. Meehl, G. A., Tebaldi, C., Walton, G., Easterling, D. & McDaniel, L. Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophys. Res. Lett.* **36**, (2009).
48. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049

(2020).

49. KOSAKA, Y. *et al*. The JRA-3Q Reanalysis. *J. Meteorol. Soc. Japan. Ser. II* **102**, 49–109 (2024).

50. BRIER, G. W. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Mon. Weather Rev.* **78**, 1–3 (1950).

51. Sena, E. T., Koren, I., Altaratz, O. & Kostinski, A. B. Record-breaking statistics detect islands of cooling in a sea of warming. *Atmos. Chem. Phys.* **22**, 16111–16122 (2022).

Supplementary Files

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

- UTExscripts.zip
- DeLucaetalSuppInfo.pdf
- DeLucaetalSuppInfo.pdf