A Diagnostic Framework for Hot Temperature Trends over Midlatitude Land

Yi Zhang,[a] William R. Boos,[b,c] Isla R. Simpson,[d]

[a] *Courant Institute of Mathematical Sciences, New York University, New York, New York*

[b] *Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, California*

[c] *Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California*

[d] *Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

Corresponding author: Yi Zhang, y.zhang@nyu.edu

ABSTRACT: Observations over the past four decades show that the annual hottest temperatures (TXx) over Northern Hemisphere midlatitude land have warmed at nearly twice the rate of global-mean surface temperature (GMST), whereas global climate models span a broad range of TXx-to-GMST sensitivities from approximately 1 to 4.5. To interpret model spread in terms of physical processes, we propose a diagnostic framework based on the near-convective neutrality assumption for hot days in Zhang and Boos (2023). This framework decomposes the sensitivity of TXx to GMST into two factors: the sensitivity of TXx to mid-tropospheric temperature T_{500}, and the sensitivity of T_{500} to GMST. For the first factor, ERA5 indicates stronger near-surface warming per unit midtropospheric warming than most CMIP6 members, which is consistent with its smaller increases in near-surface specific humidity on the hottest days than models. The second factor shows substantial inter-experiment spread, to which aerosol forcing makes an important contribution. The framework also extends to summer means and helps interpret regional behavior, including amplified TXx warming over western Europe and muted warming over the U.S. Midwest.

SIGNIFICANCE STATEMENT: This study develops a physically grounded diagnostic framework for hot extremes in near-surface air temperatures. It decomposes trends in the annual hottest temperatures into contributions from free-tropospheric warming and near-surface humidity changes, corresponding respectively to a top-down stability control and a near-surface moisture compensation. The framework can be applied at regional or even grid-point level and, unlike existing ones, does not require rerunning climate model simulations, making it readily applicable to a variety of datasets including observations, reanalysis, and model outputs. This then greatly facilitates identifying the physical causes of model biases in simulating the hot extremes of midlatitude temperatures.

1. Introduction

The intensification of hot summer temperatures since the 1950s is among the most robust effects of anthropogenic global warming (Seneviratne et al. 2021). Extreme heat events threaten ecosystems, human health, and infrastructure, and their negative effect on human health and well-being is especially prominent in mid-latitude land areas with limited air conditioning access. Accurately projecting the rate at which the hottest temperatures are warming is thus a key objective.

In this paper, we examine the model simulated changes in hot temperature extremes conditioned on the annual mean of the global mean surface temperature (GMST) warming based on theory and observational evidence. Specifically, we analyze the sensitivity of the annual hottest temperatures (TXx) to global mean surface temperature (GMST), quantified by the ratio $\Delta\mathrm{TXx}/\Delta\mathrm{GMST}$. In parallel, the absolute warming rate of TXx (in K/year) over recent decades serves as a useful benchmark for evaluating model performance against observations.

Interpreting model simulated extreme events ideally requires a robust diagnostic framework that enables the decomposition of the model output into contributions from relevant physical components. As an example, a widely used diagnostic framework for precipitation extremes (P_e) is based on the following scaling (O'Gorman and Schneider 2009):

$$P_e \sim -\epsilon \int \rho w \left.\frac{\mathrm{d}q_{\mathrm{sat}}}{\mathrm{d}z}\right|_{\theta_e^*} \mathrm{d}z. \tag{1}$$

Here, changes in P_e can be attributed to changes in three factors: thermodynamic effects ($\left.\frac{\mathrm{d}q_{\mathrm{sat}}}{\mathrm{d}z}\right|_{\theta_e^*}$, the vertical gradient in saturation specific humidity along a moist adiabat of constant saturation equivalent potential temperature θ_e^*), dynamic effects (ρw), and precipitation efficiency (ϵ). This decomposition has been adopted in numerous studies on extreme precipitation to understand individual model behavior (e.g. Sugiyama et al. 2010) and regional extreme precipitation changes (e.g. Pfahl et al. 2017). Such decompositions then help to focus research efforts on the factors that are not best represented in models.

For extreme heat, several studies have decomposed temperature anomalies into contributions from atmospheric circulation, thermodynamics, land surface conditions, and remote effect of sea surface temperature variability (e.g. Schumacher et al. 2022; Vautard et al. 2023; Faranda et al. 2023). The flow analog method identifies days with similar circulation patterns and analyzes extreme temperatures under comparable dynamical conditions (Horowitz et al. 2022; Faranda et al. 2023; Vautard et al. 2023). This approach separates circulation effects from other influences using only reanalysis data. Another, more commonly used method involves running a GCM in which selected components (winds, atmospheric temperature, soil moisture, or sea surface temperature) are prescribed with or nudged to observations (e.g. Schumacher et al. 2022; Duan et al. 2025). This allows the contributions of individual physical drivers to extreme temperature changes to be disentangled. Both methods are typically applied to specific regions and can be extended to others, but with significant effort.

Despite the value of existing approaches, a diagnostic framework for extreme temperatures analogous to Eq. (1) remains absent. Here, we aim to develop such a framework: a physics-based decomposition of extreme temperature anomalies that can be applied at the grid-point level to various global datasets, including observations, reanalyses, and model simulations. This framework is intended to address the need to diagnose model biases, allow process-based evaluation, and facilitate the much needed model-observation comparisons of regional temperature extremes (Shaw et al. 2024).

A scaling relationship that could serve as the counterpart to Eq. (1) for extreme temperatures is the upper bound theory introduced by Zhang and Boos (2023). This theory recognizes that extreme temperature events over most midlatitude land are close to convective neutrality, which imposes an effective upper bound on 2-meter air temperature as a function of the 500-hPa temperature (T_{500}), and the deviation of actual surface temperature from this upper bound is governed by the 2-meter specific humidity (q_s; see Figure S5b in Zhang and Boos (2023)). Previous studies (Fischer and Knutti 2013; Byrne 2021) have also emphasized the role of q_s in modulating hot temperature extremes, but this idea has not yet been formulated into a quantitative diagnostic framework. Here, we apply the near-convective neutrality condition to decomposing changes in extreme 2-meter temperature into contributions from T_{500} and q_s. The former can be loosely interpreted as the dynamic effect and the latter thermodynamic, drawing a parallel to the decomposition of extreme precipitation.

In the following, we describe this framework along with data and methods in Section 2. The results applying this framework to the diagnosis of GCMs are presented in Section 3, followed by discussion and conclusions in Section 4.

2. Methods

a. Data

We use observations, reanalysis, and global climate model output from the CMIP6 archive, all based on the r1i1p1f1 realization. For observations, we use the TXx variable from the HadEX3 dataset (Dunn et al. 2020) and the global and hemispheric mean surface temperatures from the HadCRUT5 dataset (Morice et al. 2021). For the ERA5 reanalysis, we use the daily maximum of hourly 2-m air temperature to compute TXx. Daily mean specific humidity is computed from the daily averages of hourly 2-m dew point temperature and surface pressure, given that the diurnal cycle of humidity (after applying the L_v/c_p factor for comparison) is not as pronounced as that of temperature. We use the daily mean 500-hPa temperature from ERA5 for T_{500}, as there the diurnal variability is weak at that level. For CMIP6, we analyze the Historical, AMIP, SSP3-7.0, greenhouse-gas-only historical experiment (Hist-GHG), and the idealized 1% per year CO_2 increase experiment (1pctCO2). TXx is computed from the daily maximum near-surface air temperature. We use the daily mean near-surface specific humidity and 500-hPa air temperature from each model simulation. All analysis focuses on land regions in the Northern Hemisphere between 35°N and 70°N starting from 1979.

b. Decomposition Framework

The main insight from Zhang and Boos (2023) is that very hot temperatures of the near surface air are under the constraint of convective neutrality. With this, the difference between the near-surface moist static energy (MSE) and the free-tropospheric saturation MSE is small. This condition can be expressed approximately as

$$\mathrm{MSE}_s \approx \mathrm{MSE}^*_a, \tag{2}$$

where the subscript s denotes the near-surface level, a the free-tropospheric level, and the superscript $*$ represents the MSE the air would have at saturation. MSE is defined as

$$\mathrm{MSE} = c_p T + L_v q + g z, \tag{3}$$

where c_p is the specific heat capacity of air at constant pressure, L_v is the latent heat of vaporization, q is the specific humidity, T is temperature, z is the geopotential height, and g is the gravitational acceleration.

Zhang and Boos (2023) show that the annual hottest temperatures over midlatitude land regions are, on average, at this critical condition described by Eq. (2) (Figure 2 in Zhang and Boos (2023)). This suggests that these events are nearly neutrally stratified in an average sense, although the equality remains approximate for individual heatwave events. To proceed, we assume that the annually hottest days may maintain a small $\mathrm{MSE}_s - \mathrm{MSE}^*_{500}$ at the grid-point level, but this violation does not change appreciably with warming. This difference can be attributed to entrainment (Duan et al. 2024) for tropical and subtropical regions and a low-level barrier to convection that allows the build-up of convective available potential energy (CAPE) in the midlatitudes (Li and Tamarin-Brodsky 2025). The assumption that it remains invariant can be revisited if compelling evidence emerges; although we do not see signs of such changes in the present analysis, it remains an open question for future study.

We now derive the framework within which changes in the annual hottest temperatures will be decomposed. We perturb Eq. (2) with respect to a reference climate, divide both sides by $c_p \Delta T_{500}$

and noting that this equation is evaluated on the hottest days, replace T_s with TXx:

$$\frac{\Delta \mathrm{TXx}}{\Delta T_{500}} = \underbrace{\frac{\mathrm{d}T_b(T_{500})}{\mathrm{d}T_{500}}}_{\substack{\text{change in}\\ \text{upper bound}}} - \underbrace{\frac{L_v}{c_p}\frac{\Delta q_s}{\Delta T_{500}}}_{\substack{\text{change in proximity}\\ \text{to bound}}}, \tag{4}$$

where $T_b(T_{500})$ is the upper bound for T_s proposed in Zhang and Boos (2023) (see Appendix). The first term on the right-hand side represents the contribution of free-tropospheric warming, which raises the upper bound for T_s. The second term accounts for the effect of near-surface humidity (q_s), which changes the proximity to the upper bound.

To link back to global mean warming, we additionally have

$$\frac{\Delta \mathrm{TXx}}{\Delta \mathrm{GMST}} = \left(\frac{\Delta \mathrm{TXx}}{\Delta T_{500}}\right)\left(\frac{\Delta T_{500}}{\Delta \mathrm{GMST}}\right), \tag{5}$$

where GMST denotes the global and annual mean surface temperature.

Equations (4) and (5) together comprise the framework within which TXx warming is decomposed.

c. Trends and Sensitivities

All sensitivity metrics of the form $\Delta y/\Delta x$ are computed as the ratio of linear trends in the time series $\{y_t\}$ and $\{x_t\}$. For each series we fit

$$y_t = b_y t + a_y + \epsilon_t^{(y)}, \qquad x_t = b_x t + a_x + \epsilon_t^{(x)},$$

where b_y and b_x are trends, a_y and a_x are intercepts, and $\varepsilon_t^{(y)}$ and $\varepsilon_t^{(x)}$ are regression residuals. The sensitivity is defined as

$$\frac{\Delta y}{\Delta x} = \frac{b_y}{b_x}. \tag{6}$$

The standard errors of the slopes, σ_{b_y} and σ_{b_x}, are also obtained from the least-squares fits. Assuming independence between b_y and b_x, the uncertainty in $\Delta y/\Delta x$ is estimated by standard error propagation:

$$\sigma_{\Delta y/\Delta x} = \left|\frac{b_y}{b_x}\right| \sqrt{\left(\frac{\sigma_{b_y}}{b_y}\right)^2 + \left(\frac{\sigma_{b_x}}{b_x}\right)^2}. \tag{7}$$

These uncertainties are shown as error bars in various figures.

This method may misrepresent changes when the trends in x and y are not linear, however, such nonlinearity is minimal over the historical period beginning in 1979 and in the future scenario SSP3-7.0 considered here. The advantage of this approach compared to regressing y directly against x is to avoid confounding influences from short-term covariability associated with ENSO or volcanic forcing and to emphasize the long-term warming signal.

3. Results

a. Overview of the decomposition

We focus on the average of TXx over land within 35°N–70°N where surface elevation is less than 1.5 km (denoted by $\overline{\mathrm{TXx}}$), broadly representative of the midlatitudes (we avoid elevated topography to reduce the likelihood that the boundary layer encompasses 500 hPa). The AMIP and Historical experiments exhibit multi-model mean trends in $\overline{\mathrm{TXx}}$ of 0.47 and 0.55 $\mathrm{K\,dec^{-1}}$, respectively, with substantial spread across the ~10 ensemble members analyzed. The observed trend over 1979–2014 is 0.33 $\mathrm{K\,dec^{-1}}$ in HadEX3 and 0.31 $\mathrm{K\,dec^{-1}}$ in ERA5, ranking near the lowest Historical member and between the 10th and 20th percentile of the AMIP ensemble. Extending the ERA5 record to 2023 gives a slightly higher trend of 0.37 $\mathrm{K\,dec^{-1}}$ (Figure 1a).

The sensitivity $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ facilitates comparison across experiments with differing warming levels (Figure 1b). The AMIP and Historical ensemble means of $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ are 3.1 and 2.2 respectively, both with notable inter-model spread. TXx from ERA5 paired with GMST from HadCRUT5 yields 1.8, and HadEX3/HadCRUT5 gives 2.0 for $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$. Both of these values approximately tie with the lowest AMIP member and lie near the 20th percentile in Historical. The SSP3-7.0 ensemble shows a lower mean of 1.5 with narrower spread, placing observations and reanalysis around its 70th percentile.

Decomposing $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ into two multiplicative components, as in Eq. (5), reveals that the forcing-pathway differences in $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ arise primarily from $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$, while $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ remains relatively consistent among experiments (Figure 2), where $\overline{T_{500}}$ is the spatial average of T_{500} on the annual hottest days. The ERA5 reanalysis estimate of $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ is notably higher than those from nearly all model members: it lies above the 90th percentile of AMIP and exceed all ten Historical and ten SSP3-7.0 models. The ERA5 estimate of $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$ ranks

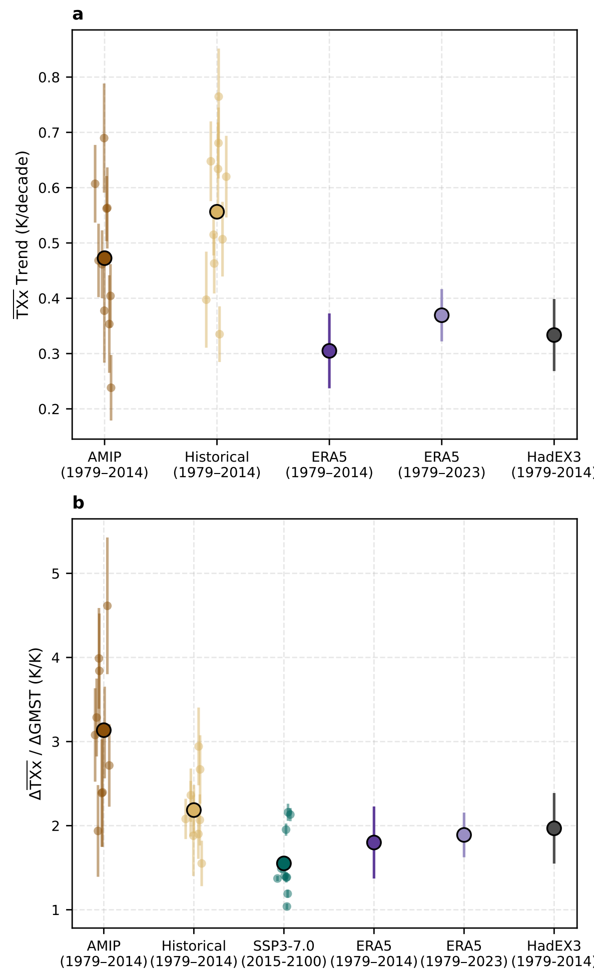

FIG. 1. Trends in $\overline{\text{TXx}}$ (a) and ratios of $\overline{\text{TXx}}$ to global-mean surface temperature (GMST) trends (b) across CMIP6 models, reanalysis, and observations. $\overline{\text{TXx}}$ refers to the annual maximum daily temperature averaged over land between 35°N–70°N. Small dots show individual model ensemble members; large markers indicate multi-model means (for CMIP6) or the observational estimate (for ERA5 and HadEX3). Vertical bars represent the standard error of the trend: in panel (a), this is the standard error of the linear fit; in panel (b), it is the standard error of the trend ratio (see Section c).

at the very bottom of the CMIP6 model ensembles, suggesting that models produce more mid-tropospheric warming on the hottest days than observed for a given GMST increase. These offsetting tendencies lead to partial compensation in the full ratio $\Delta\overline{\text{TXx}}/\Delta\text{GMST}$.

We are unable to apply the same decomposition to purely observational data, due to the lack of long records and sufficient spatial coverage of daily T_{500} values. However, the consistency of

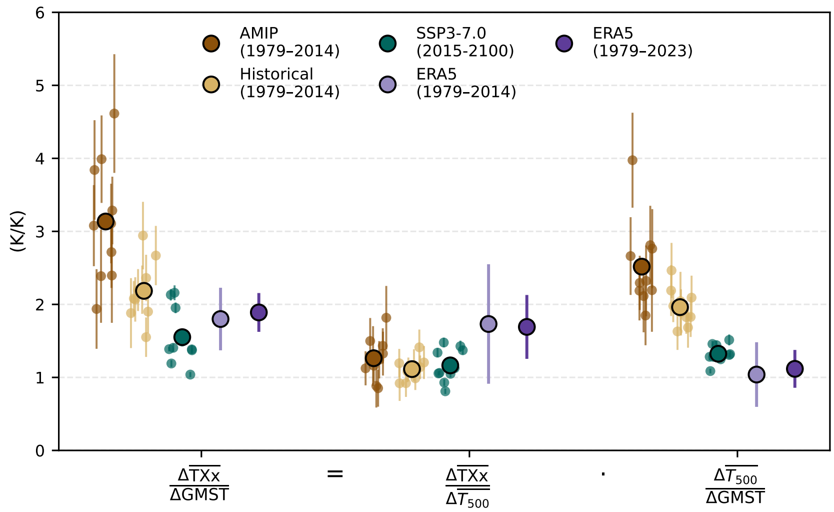

FIG. 2. Decomposition of TXx warming per Kelvin of GMST warming. Shown are the ratios $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$, $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$, and $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ for CMIP6 models and ERA5. For CMIP6 models, each small dot shows a model's central estimate with vertical bars denoting standard errors of the ratio of two trends (see Section c for details); large markers indicate multi-model means. For reanalysis and observations, markers show central estimates with standard error bars estimated the same way as individual models.

$\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ and $\overline{\mathrm{TXx}}$ trends between ERA5 reanalysis and HadEX3 observations in Figure 1 and the consistency of ERA5 mid-tropospheric temperature with satellite and radiosonde estimates (Zhang and Boos 2023) indicate that ERA5 might be a sufficiently good representation of the atmospheric state during hottest days.

In the following, we address the spread across model ensemble members and observational datasets in both components of the decomposition: $\Delta\mathrm{TXx}/\Delta T_{500}$ in Section 3b and $\Delta T_{500}/\Delta\mathrm{GMST}$ in Section 3c. We also extend the framework to a broader range of summer days beyond annual extremes (Section 3d) and examine the spatial patterns with focus on two representative regions, namely U.S. Midwest and western Europe (Section 3e).

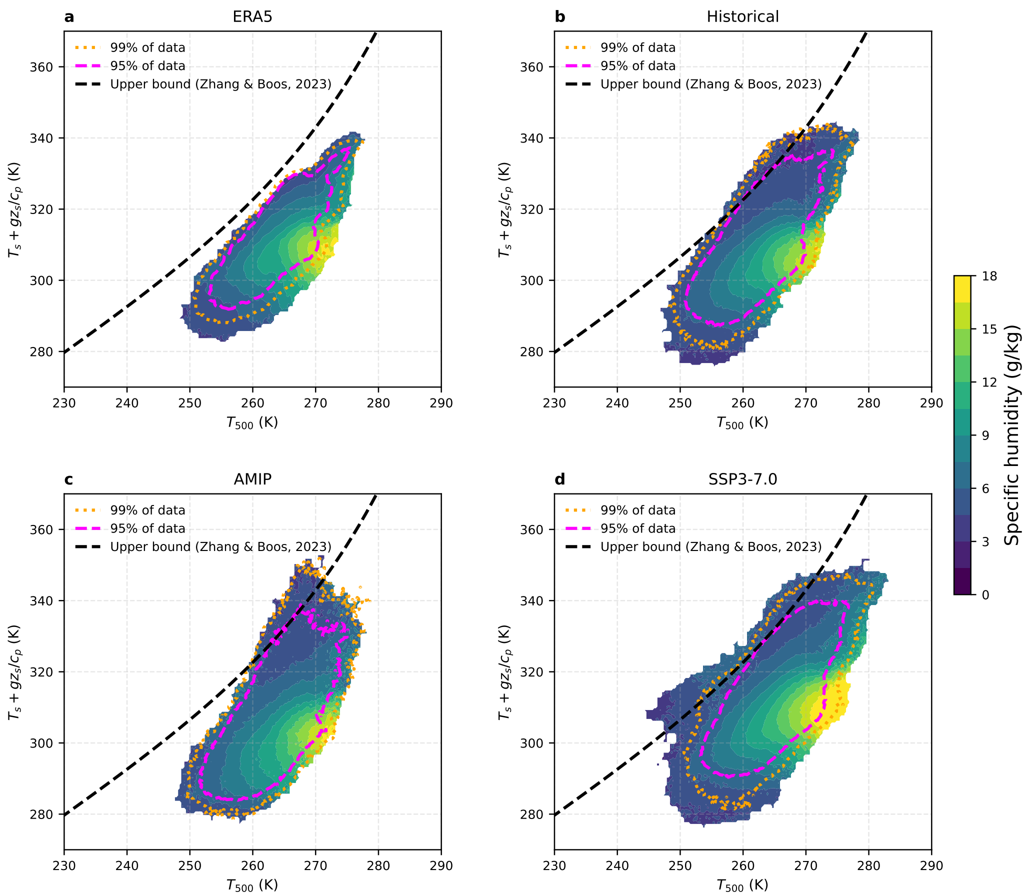

FIG. 3. The upper bound theory largely constrains the annual hottest temperatures in reanalysis and models. Joint distributions of free-tropospheric temperature (T_{500}) and surface temperature (T_s) with an offset from surface elevation (z_s) on TXx days are shown for (a) ERA5, (b) CMIP6 Historical simulations, (c) AMIP simulations, and (d) SSP3-7.0 projections. Shading indicates the mean specific humidity (g/kg) conditioned on TXx days in each (T_{500}, T_s) bin. Contours enclose the regions containing 99% (dotted orange) and 95% (dashed magenta) of the probability density (data points weighted by the cosine of latitudes). The black dashed line denotes the upper bound from Zhang and Boos (2023). A Gaussian filter is applied to smooth the plotted fields for clarity.

b. Model-reanalysis discrepancies in surface–midtroposphere coupling

1) TXx DAYS IN THE T_s-T_{500} PHASE SPACE

We now examine the model-reanalysis discrepancy in $\Delta\text{TXx}/\Delta T_{500}$ using Eq. (4). As a first step, we verify that the annual hottest days in ERA5 and models are constrained by the upper bound defined in Eq. (A6). This upper bound depends primarily on physical constants but includes the prefactor $\overline{z_{500}}/\overline{T_{500}}$ in the geopotential term (see Appendix), which could differ across models and between models and reanalysis. We find that this factor differs by less than 1% across models and ERA5, producing nearly identical upper bounds, and therefore use the ERA5 estimate of $\overline{z_{500}}/\overline{T_{500}}$ for plotting.

Figure 3 shows that surface temperatures adjusted for orography ($T_s + \frac{g}{c_p} z_s$) generally remain below the upper bound for TXx days, although some apparent violations occur in the CMIP6 dataset, likely due to model-specific factors such as convective parameterization schemes which we do not analyze further. The near surface specific humidity (q_s) at a height of 2 m increases largely with the difference between the actual $T_s + \frac{g}{c_p} z_s$ and the upper bound throughout the full range of T_{500} values, consistent with Eq. (A5), in which T_s and q_s act as compensatory terms on the left side. At lower T_s, the isolines of q_s no longer run parallel to the upper bound (Figure 3), but instead reflect the Clausius–Clapeyron relationship that the capacity of the near-surface air to hold moisture decreases at cooler temperatures.

We then examine the temporal shift of the joint distribution of T_s and T_{500} in time under anthropogenic forcing by dividing each dataset into four consecutive chunks in time. In ERA5, the shift in TXx is almost parallel to the upper bound (red line in Figure 4a), which implies that the first term on the right-hand side of Eq. (4), $\mathrm{d}T_b(T_{500})/\mathrm{d}T_{500}$, dominates over the second term, $L_v/c_p \cdot \Delta q_s/\Delta T_{500}$. This is consistent with the findings of Figure 4 in Zhang and Boos (2023) showing that both the upper bound and the actual TXx have changed by similar magnitudes, even at the 0.25° grid-point scale. CMIP6 models collectively show a more pronounced downward deviation from the upper bound (Figure 4b-d) not only in the multi-model mean (slopes of thick red lines) but also for individual members (slopes of thin red lines). If Eq. (4) holds, this downward departure from the upper bound suggests that the TXx warming of most CMIP6 models might be accompanied by stronger increases in the specific humidity of the near surface air than in ERA5, thus reducing surface warming per degree of T_{500} warming, which we will next verify by examining the surface specific humidity trends on the annual hottest days.

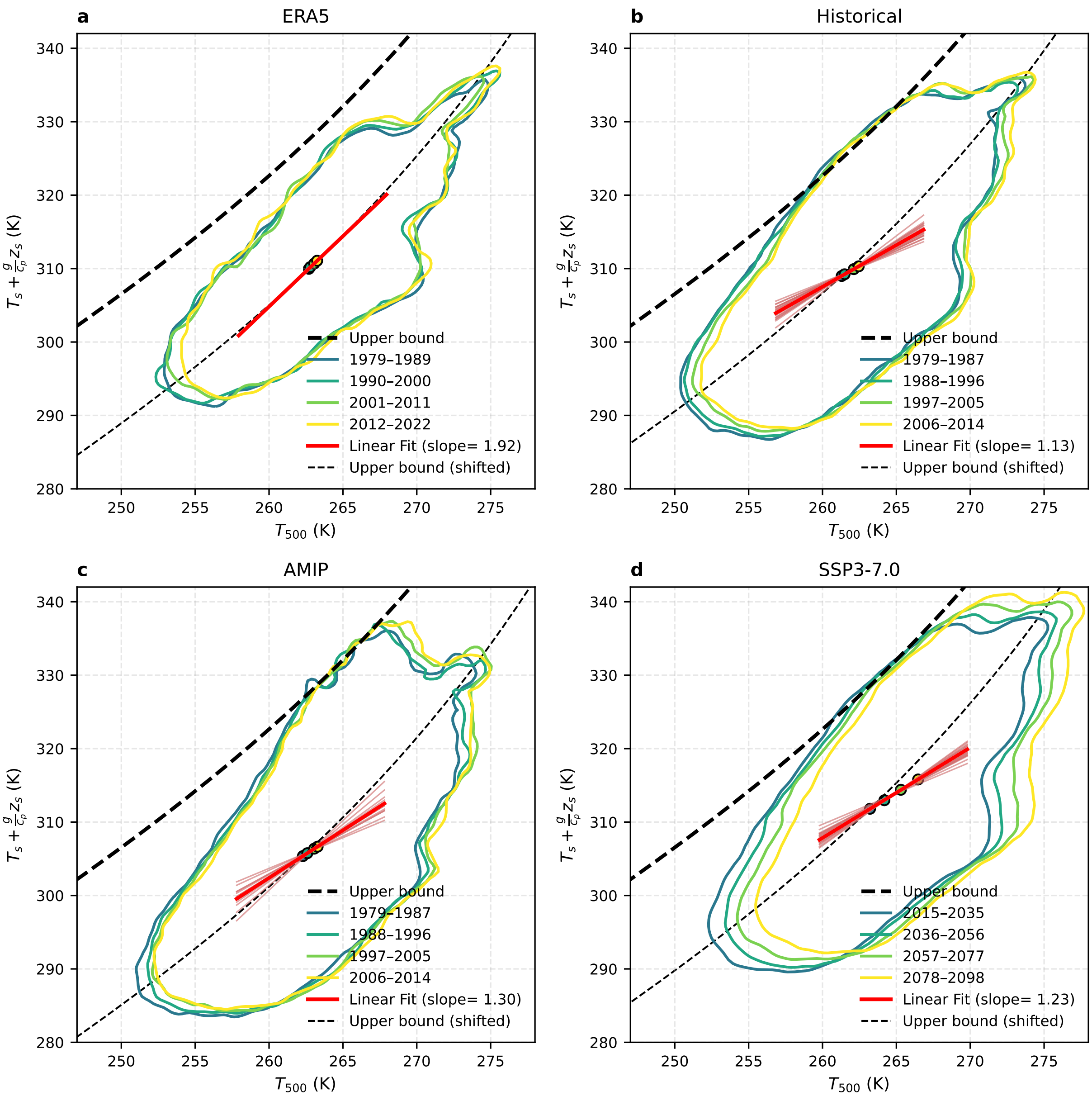

Fig. 4. Shift of TXx days in the $T_s - T_{500}$ space for (a) ERA5, (b) Historical, (c) AMIP, and (d) SSP3-7.0. Each contour encloses 95% of data (weighted by area) in each joint distribution. Colored dots are the centroids of each distribution, reflecting the shift of the distribution with warming. Thick red lines are linear fits of the multi-model centroids highlighting the shift and thin red lines are the same but for individual models. Dotted lines are parallel to the upper bound and go through the centroid of the initial distribution.

2) Moistening vs. warming compensation

To test the above hypothesis, we directly examine the relationship between $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ and $L_v\Delta\overline{q_s}/(c_p\Delta\overline{T_{500}})$, where $\overline{q_s}$ is the spatial average of q_s on the annual hottest days. According to

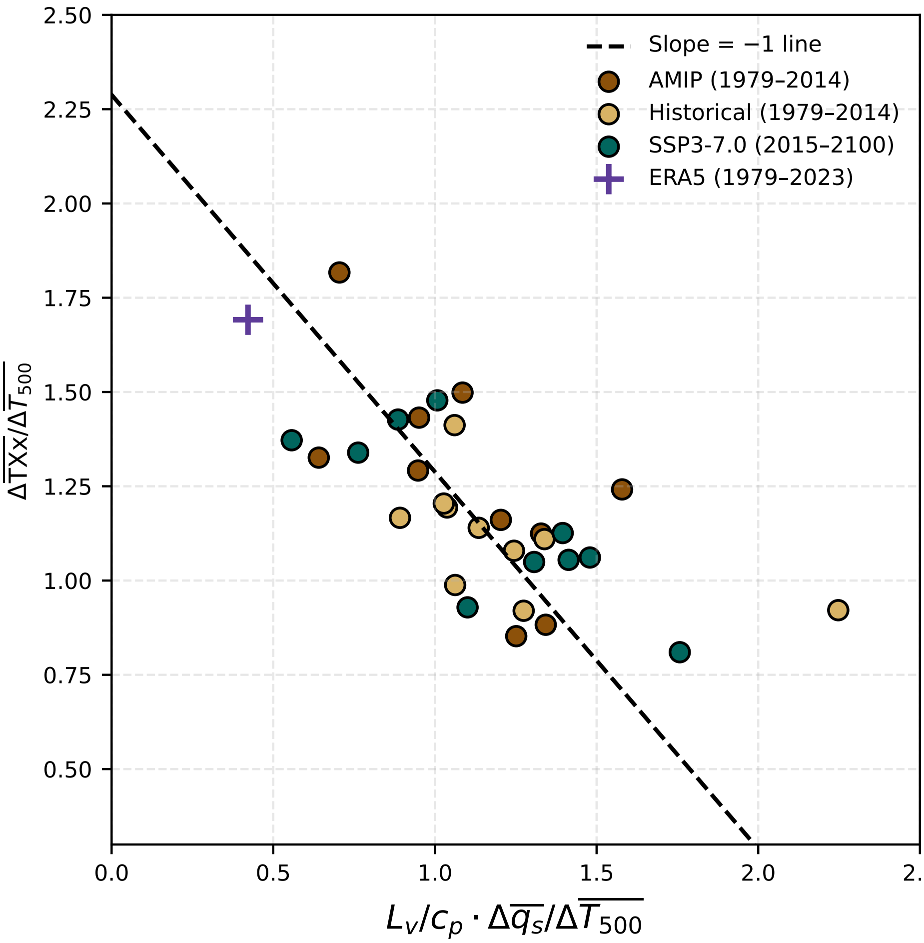

FIG. 5. Compensating relationship between surface warming and moistening on the annual hottest days. Scatterplot of $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ versus $L_v\Delta\overline{q_s}/(c_p\Delta\overline{T_{500}})$ from CMIP6 model simulations and ERA5. $\overline{T_{500}}$ and $\overline{q_s}$ are averages over land within 35°N-70°N on the annual hottest days, then the ratios are computed following the procedure in Section 2c.

Eq. (4), these two terms should sum to $\mathrm{d}T_b/\mathrm{d}T_{500}$, the slope of the upper bound. This expectation is partially supported by a negative correlation between the two terms among CMIP6 members, regardless of experiment (Figure 5). The ERA5 data point lies near the low-moistening, high-warming end of the model cluster, which helps explain the higher $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ diagnosed in ERA5 compared to CMIP6 models (Figure 3) and equivalently the steeper slopes in Figure 4. Among the 30 model members examined for three experiments in Figure 5, all produce stronger $L_v\Delta\overline{q_s}/(c_p\Delta\overline{T_{500}})$ than ERA5. These q_s trend patterns are broadly consistent with previous findings (e.g., Dunn et al. 2017; Douville and Plazzotta 2017; Simpson et al. 2024) that observations tend to show weaker increases in q_s than the model mean, particularly over arid regions such as the southwestern United States (Simpson et al. 2024), and over midlatitude land during summer (Douville and Plazzotta 2017). Our results suggest that the model–reanalysis discrepancies in humidity trends may contribute to differences in simulated and observed extreme temperature responses.

Eq. (4) motivates fitting a linear relationship with slope -1 to the CMIP6 and ERA5 data points, because $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ can be interpreted as the dependent variable, $\mathrm{d}T_b/\mathrm{d}T_{500}$ the intercept, and $L_v\Delta\overline{q_s}/(c_p\Delta\overline{T_{500}})$ the independent variable. Fitting such a linear relationship with fixed slope -1 minimizing the mean squared errors results in the following formula for the estimated intercept:

$$a = \frac{1}{N}\sum_i (y_i + x_i), \tag{8}$$

where x_i and y_i are the $L_v/c_p \cdot \Delta\overline{q_s}/\Delta\overline{T_{500}}$ and $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$, respectively. The intercept a is estimated to be 2.3, which is broadly consistent with theoretical expectations for the sensitivity of the upper bound to T_{500}, though it aligns more closely with the warmer T_{500} values seen in later decades of the SSP3-7.0 scenario (for the current climate the slope is around 1.9). This suggests that the effective upper-bound slope in models may be steeper than predicted by theory. In Figure 3b–d, the envelope formed by the hottest simulated days, which appears as the uppermost boundary of the T_s–T_{500} scatter, tilts more steeply than the theoretical line and also indicates a higher effective slope in the models. Possible explanations to this higher slope include enhanced entrainment associated with lower-tropospheric drying (Duan et al. 2024) and warming-related changes in convective available potential energy (CAPE) (Singh et al. 2017; Chen et al. 2020; Li and Tamarin-Brodsky 2025).

This anticorrelation is also evident at the grid-point level across model experiments, as shown in Figure 6, which shows the correlation coefficient between $\Delta\mathrm{TXx}/\Delta T_{500}$ and $L_v\Delta q_s/c_p\Delta T_{500}$ across CMIP6 models. The widespread negative correlation over most land areas indicates that stronger near-surface moistening is associated with weaker TXx warming, consistent with Eq. (4). This anticorrelation is particularly strong over northern midlatitude land, including Europe, North America, and parts of Asia. Correlations are weaker or even positive in dry subtropical regions and

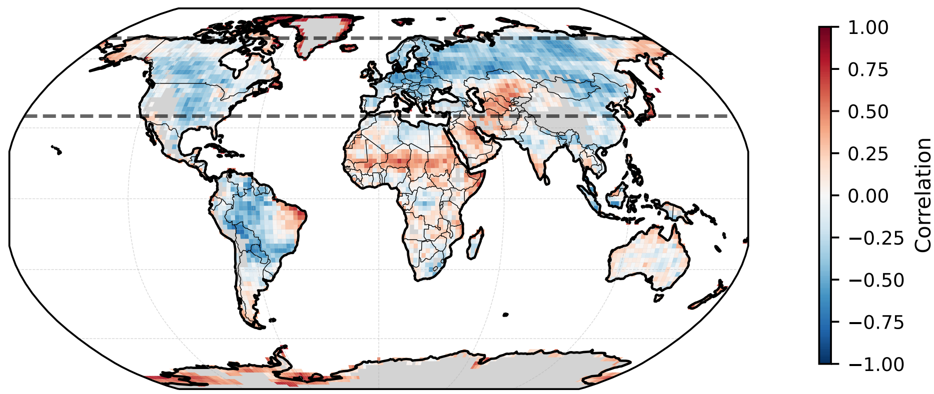

FIG. 6. Correlation across CMIP6 models between $\Delta q_s/\Delta T_{500}$ and $\Delta\mathrm{TXx}/\Delta T_{500}$ both evaluated on the annual hottest days. Correlations are computed at each grid point across models, using simulations from AMIP, historical, and SSP3-7.0 experiments. Dashed horizontal lines denote 35°N and 70°N, the Northern Hemispheric midlatitude band emphasized in the text. All model data are regridded to $2^\circ \times 2^\circ$ resolution before analysis. Locations of surface elevation over 1.5 km are masked in gray.

semi-arid tropical regions like the Sahel. Although the signal-to-noise ratio is lower at regional scales and the number of ensemble members is too small to support robust linear fits like that in Figure 5, the spatial pattern of the correlation still confirms the broader conclusion that lower warming of TXx in models is linked to excessive trends in near-surface humidity over the Northern Hemispheric midlatitude in those models. The physical origin of q_s trends on the annual hottest days, however, is difficult to determine and lies beyond the scope of this study. Future work is needed to examine potential contributing factors, such as evapotranspiration, precipitation, and vertical and horizontal moisture advection.

c. Midtropospheric amplification and the potential role of aerosols

The previous section shows that different experiments produce a similar range of $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ values in CMIP6, despite being underestimated relative to ERA5. Here, we focus on the second

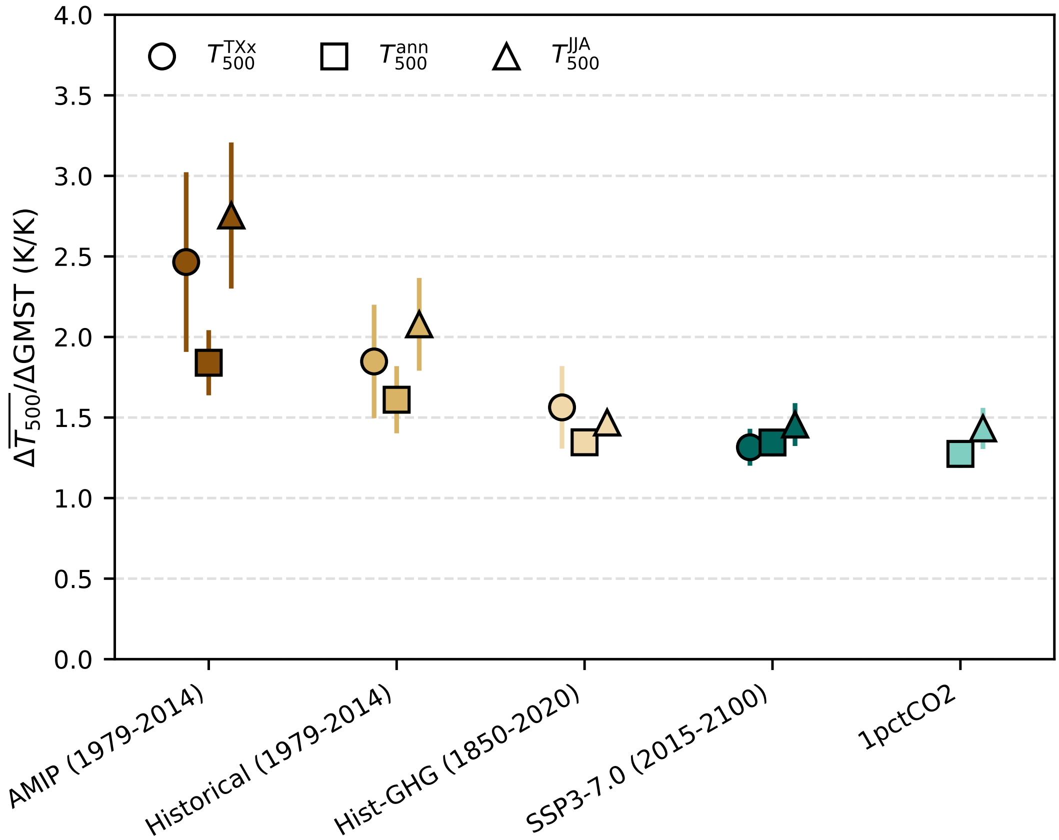

FIG. 7. Midtropospheric warming relative to the annual mean of the global mean surface temperature (GMST) warming, expressed as the ratio $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$. Circles indicate trends in 500-hPa temperature (T_{500}) computed on TXx days, squares indicate the annual mean T_{500}, and triangles indicate the June–July–August (JJA) mean T_{500}.

factor on the right-hand side of Eq. (5), $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$. As shown in Section 3a and Figure 2, this term explains much of the experiment dependence of in $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$.

We first ask whether the spread in $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$ among experiments is restricted to the annual hottest days or reflects a general pattern, and the results indicate the latter. When using the annual mean or the June-July-August mean T_{500} to compute $\Delta\overline{T_{500}}/\Delta\mathrm{GMST}$, AMIP still produces the highest ratio, followed by Historical and SSP3-7.0 (Figure 7). This consistency suggests that the divergent behavior is a general feature of midtropospheric warming, rather than a special case on the hottest days, with summer conditions providing the closest match to the hottest-day sensitivity.

The experiment-dependent behavior of $\Delta\overline{T_{500}}/\Delta$GMST suggests that differences in the forcing agents matter. Both AMIP and Historical experiments are influenced by substantial reductions in aerosol emissions in the NH mid-latitudes over recent decades, while SSP3-7.0 is dominated by greenhouse gas increases and weaker changes in aerosol forcing over this region. This raises the hypothesis that the pronounced aerosol reductions in AMIP and Historical are at least partly responsible for their higher $\Delta\overline{T_{500}}/\Delta$GMST values. To test this, we apply the same analysis to the historical simulation with greenhouse gas forcing only (Hist-GHG) and find that the resulting $\Delta\overline{T_{500}}/\Delta$GMST is reduced relative to Historical and comes close to that of SSP3-7.0. The idealized 1% per year CO_2 increase experiment (1pctCO2), another experiment that does not include aerosol forcing and therefore exhibits no changes in aerosol forcing, also yields similar $\Delta\overline{T_{500}}/\Delta$GMST to SSP3-7.0 (Figure 7). These results support the interpretation that the rapid decline in aerosol forcing over the NH mid-latitudes is a primary driver of the elevated $\Delta\overline{T_{500}}/\Delta$GMST values in AMIP and Historical relative to SSP3-7.0. The higher $\Delta\overline{T_{500}}/\Delta$GMST in AMIP compared to Historical likely comes from sea surface temperature patterns.

The higher $\Delta T_{500}/\Delta$GMST in global climate models is qualitatively consistent with prior work finding a faster warming trend in the free troposphere in the models than derived from satellite observations (Santer et al. 2017a), which is most pronounced in the tropics but persists into near-global averages (Santer et al. 2017b).

d. Extending the framework beyond annual hottest days

The preceding analysis focuses on the annual hottest days (TXx). Here, we examine whether the same physical mechanisms also govern temperature trends across a broader range of hot summer days. We show linear trends in surface temperature (T_s) and other relevant variables as a function of percentile thresholds that progressively include more days—from the single hottest day (1.1%) to the full June–July–August average (100%) (Figure 8).

The warming trend of average summer days (June–July–August mean) is comparable to that of the hottest days in both ERA5 (Fig. 8a) and the CMIP6 multi-model mean (Fig. 8b). This finding is consistent with previous observations that the most extreme summer temperatures have not warmed substantially faster than the seasonal average (McKinnon et al. 2024).

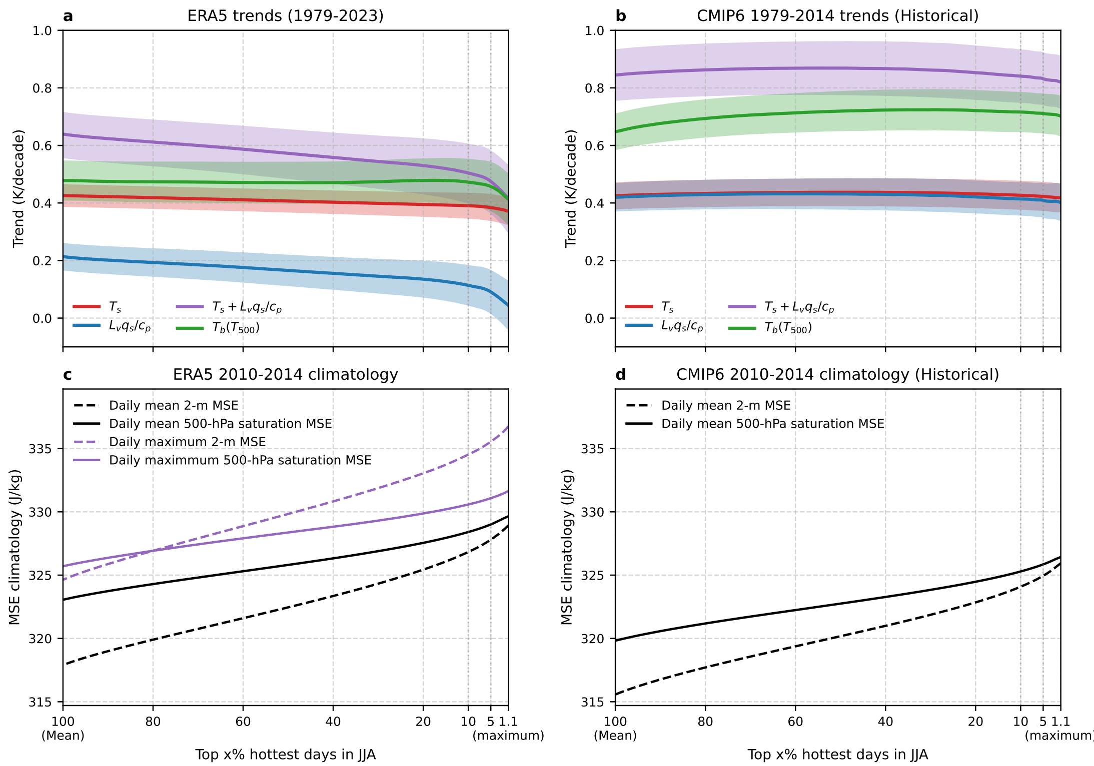

FIG. 8. (a) Linear trends in surface temperature (T_s), scaled near-surface specific humidity ($L_v q_s/c_p$), their sum ($T_s + L_v q_s/c_p$), and the upper-bound temperature ($T_b(T_{500})$), computed from ERA5 reanalysis over 1979–2023 for Northern Hemisphere midlatitude land (35°–70°N). Trends are calculated separately for subsets of summer (JJA) days, ranging from the hottest extremes (1.1%) to the full summer mean (100%). Shaded areas represent the standard error of the linear regression slope. (b) Same as (a) but for the CMIP6 Historical experiment. Shaded areas represent the standard error of the intermodel spread of the trends, rather than the standard error of the regression slopes. (c) Climatologies of moist static energy (MSE) at 2-m level (MSE_s) and saturation MSE at 500-hPa level (MSE^*_{500}) across percentiles of air temperature at 2 m in ERA5. (d) Same as (c) but for multi-model mean of the CMIP6 Historical experiment.

We then examine the trends in near-surface moist static energy (MSE_s), specific humidity (q_s), and T_{500} across percentiles. All these trends are relatively insensitive to percentile in both ERA5 and CMIP6. The key difference between the datasets is that the MSE_s trend in ERA5 is predominantly driven by T_s, whereas in CMIP6, T_s and q_s contribute roughly equally.

Near-convective neutrality implies that MSE_s and MSE^*_{500} should be approximately equal. Using daily averaged fields from CMIP6, we find that models reproduce the ERA5 climatology of these quantities: on the annual hottest days, MSE_s and MSE^*_{500} are nearly equal, and their difference increases to $\sim 5\ \mathrm{J\,g^{-1}}$ for the summer mean. Analysis of daily maxima in ERA5 shows a similar pattern, with MSE_s exceeding MSE^*_{500} by about $5\ \mathrm{J\,g^{-1}}$ on the hottest days and nearly matching it for the summer mean. Overall, the MSE differences (MSE_s-MSE^*_{500}) remain within about $\pm 5\ \mathrm{J\,g^{-1}}$, indicating that both the hottest days and mean summer days remain close to convective neutrality.

These results suggest that the convective upper-bound framework can be extended beyond annual extremes to help diagnose and interpret trends in summer mean temperatures. This may help explain the observed lack of amplification in extreme temperatures relative to the seasonal average (McKinnon et al. 2024).

e. Spatial patterns and regional signals

We next examine the spatial patterns of $\Delta\mathrm{TXx}/\Delta\mathrm{GMST}$ and assess the potential of this framework for interpreting regional trends. Both HadEX3 and ERA5 exhibit strongly nonuniform sensitivities over the past few decades, with the most notable signals emerging over Europe and North America (Figure 9a,b). Over Europe, TXx has increased faster than over other regions, consistent with prior findings (Vautard et al. 2023). In contrast, parts of North America show a negative trend in TXx, with the cooling signal more widespread in HadEX3 than in ERA5. This "warming hole" in TXx could be linked to irrigation-related effects (Thiery et al. 2020) and circulation changes (Singh et al. 2023) over that region. In the following, we focus on these two regions.

The near-zero observed trend of TXx (Figure 10a) and the sensitivity $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ (Figure 10b) over parts of North America lie outside the model ensemble in which all members predict positive trends (this is true for the central estimates, although some of the standard errors overlap). Decomposition reveals that the term $\Delta\overline{\mathrm{TXx}}/\Delta T_{500}$ in CMIP6 models is larger than in ERA5 (Figure 11a), contributing to their more positive $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ (there is less disagreement when the ERA5 trend is extended through 2023, but this goes beyond the end date of the Historical and AMIP runs). This is consistent with the fact that the models simulate insufficient increase in q_s during hot days per degree of T_{500} warming (Figure 11b). This discrepancy may reflect enhanced irrigation and soil moisture storage of that region over recent decades and these processes may not be well represented in CMIP6 simulations. The expected anti-correlation between the TXx

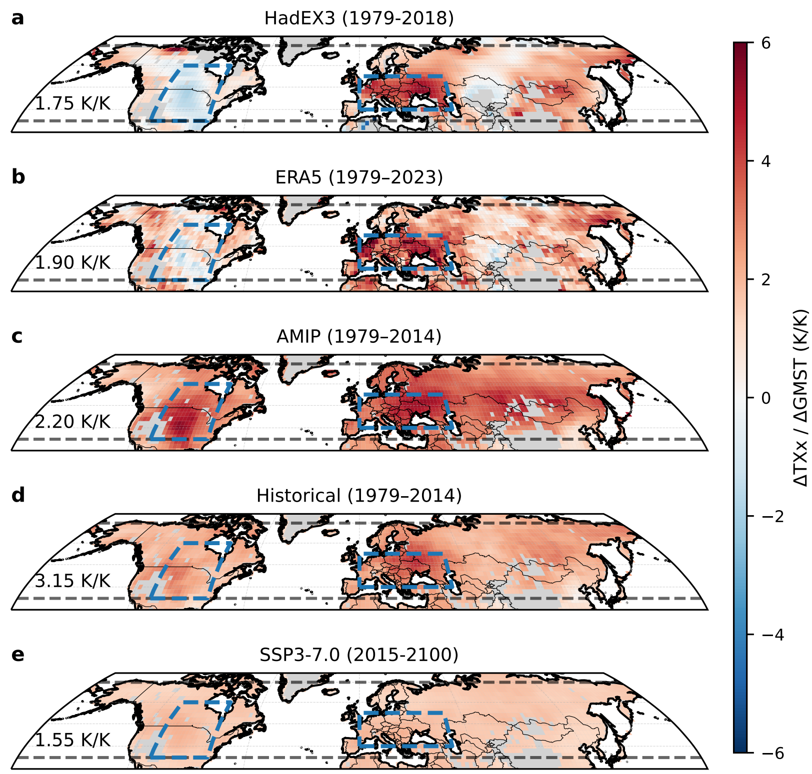

Fig. 9. Spatial patterns of ΔTXx$/\Delta$GMST for various datasets as labeled. All datasets are interpolated to a common 2°×2° grid prior to computation. For CMIP6 experiments, the fields shown represent the multi-model mean of ΔTXx$/\Delta$GMST. Values of $\Delta\overline{\mathrm{TXx}}/\Delta$GMST, where $\overline{\mathrm{TXx}}$ is the land average between 35°–70°N, are indicated in the lower-left of each panel. Regions with surface elevation above 1.5 km are masked in gray. Black dashed lines denote the 35° and 70°N latitudes. Blue dashed boxes highlight two regions in North America (35°–60°N, 80°–110°W) and in Europe (40°–55°N, 0°–55°E).

warming term and the surface air moistening term holds; the slope is shallower than the -1 predicted by Equation (4), but a fair amount of scatter exists.

European TXx has been warming at a rate of around 4 K with each Kelvin of global mean surface warming according to both HadEX3 and ERA5 (Figure 10 c), which is above most models in the Historical experiment but similar to the average AMIP ensemble. The absolute TXx trends over this region fall within the model spread of both AMIP and Historical experiments (Figure 10d). Decomposition of the sensitivity $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ shows that ERA5 exhibits higher values of $\Delta\overline{\mathrm{TXx}}/\Delta T_{500}$ than CMIP6 multi-model means, while the $\Delta T_{500}/\Delta\mathrm{GMST}$ term remains relatively consistent across ERA5 and the Historical ensemble (Figure 11c). The lower $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$ seen in the Historical simulations is therefore primarily driven by a muted $\Delta\overline{\mathrm{TXx}}/\Delta T_{500}$ response which is mechanistically linked to a stronger increase in surface air specific humidity q_s per unit T_{500} warming on hot days (Figure 11d). The same anti-correlation between $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ and $L_v/c_p\cdot\Delta\overline{q_s}/\Delta\overline{T_{500}}$ observed in the midlatitude land mean (Figure 2) also holds when the analysis is restricted to Europe (Figure 11d).

These two regional cases illustrate contrasting directions of model–observation disagreement. In both, trends in $L_v/c_p\cdot\Delta q_s/\Delta T_{500}$ on the hottest days emerge as a key factor shaping differences in $\Delta\overline{\mathrm{TXx}}/\Delta T_{500}$. This result thus echoes previous studies that emphasize the importance of land–atmosphere feedbacks and soil moisture control on European heat extremes (e.g., Vogel et al. 2017), as well as the influence of irrigation on temperature extremes over the US Midwest (Nocco et al. 2019).

4. Summary and Conclusions

This study presents a physics-based framework for diagnosing hot temperature trends over midlatitude land. Building on the upper bound theory, we formulate a decomposition that separates the trends of the annual hottest temperatures (TXx) into two components: the sensitivity of TXx to midtropospheric warming ($\Delta\mathrm{TXx}/\Delta T_{500}$) and the amplification of midtropospheric warming relative to global mean surface temperature ($\Delta T_{500}/\Delta\mathrm{GMST}$). This framework enables interpretation of modeled TXx trends in terms of free-tropospheric dynamics and local-scale land-atmosphere coupling, and facilitates comparison across models, experiments, and observations.

A key finding is that ERA5 estimates of $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ lie at the upper end of the CMIP6 model ensemble distribution examined, which coexists with excessive moistening in models on the annual hottest days. This factor, $\Delta\mathrm{TXx}/\Delta T_{500}$, quantifies the shift of annual hottest days in the T_s-T_{500} phase space. While $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ remains relatively consistent across CMIP6 experiments for average annual hottest days over land between 35°N and 70°N, the ERA5 value exceeds nearly all 30 individual model members across AMIP, Historical, and SSP3-7.0 ensembles (Figures 2, 4),

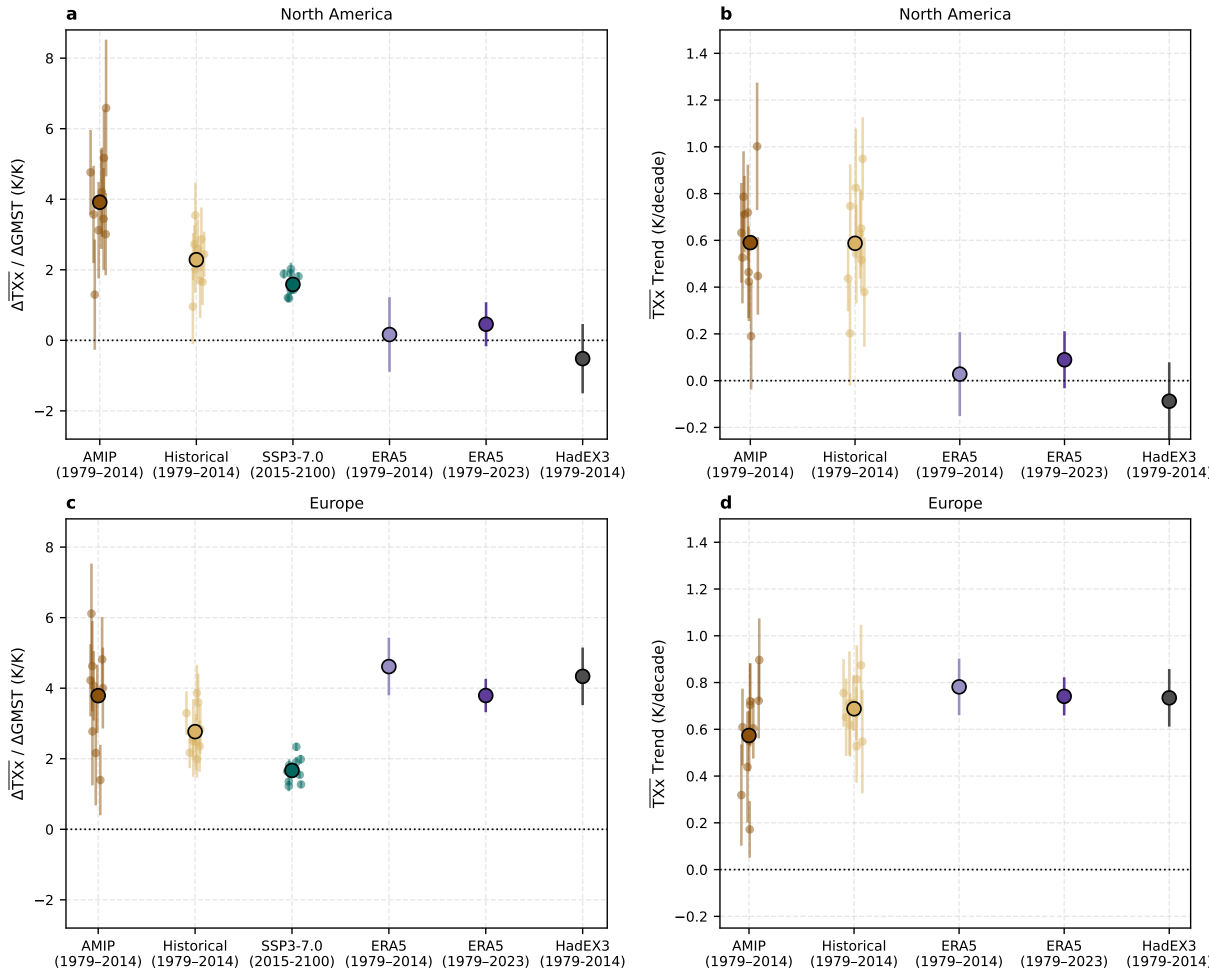

FIG. 10. Warming of $\overline{\mathrm{TXx}}$ averaged over North America and Europe boxes across models, reanalysis, and observations. Panels (a) and (b) are the same as Figure 1a and b, but for North America (35°–60°N, 80°–110°W), and panels (c) and (d) are for Europe (40°–55°N, 0°–5°E).

although the standard error of the ERA5 estimate does encompass most models. Across models and ERA5, $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T_{500}}$ is negatively correlated with $L_v\Delta\overline{q_s}/(c_p\Delta\overline{T_{500}})$ (Figure 5), consistent with the thermodynamic constraint imposed by moist convective neutrality that stronger near-surface moistening reduces the allowable surface warming under the same moist static stability. Further research is warranted on the causes of excessive near-surface moistening during midlatitude summers in CMIP6 models.

The diagnosis of the warming–moistening relation for regions also reveals useful information. Parts of North America and Europe are known for exhibiting TXx trends that diverge from the

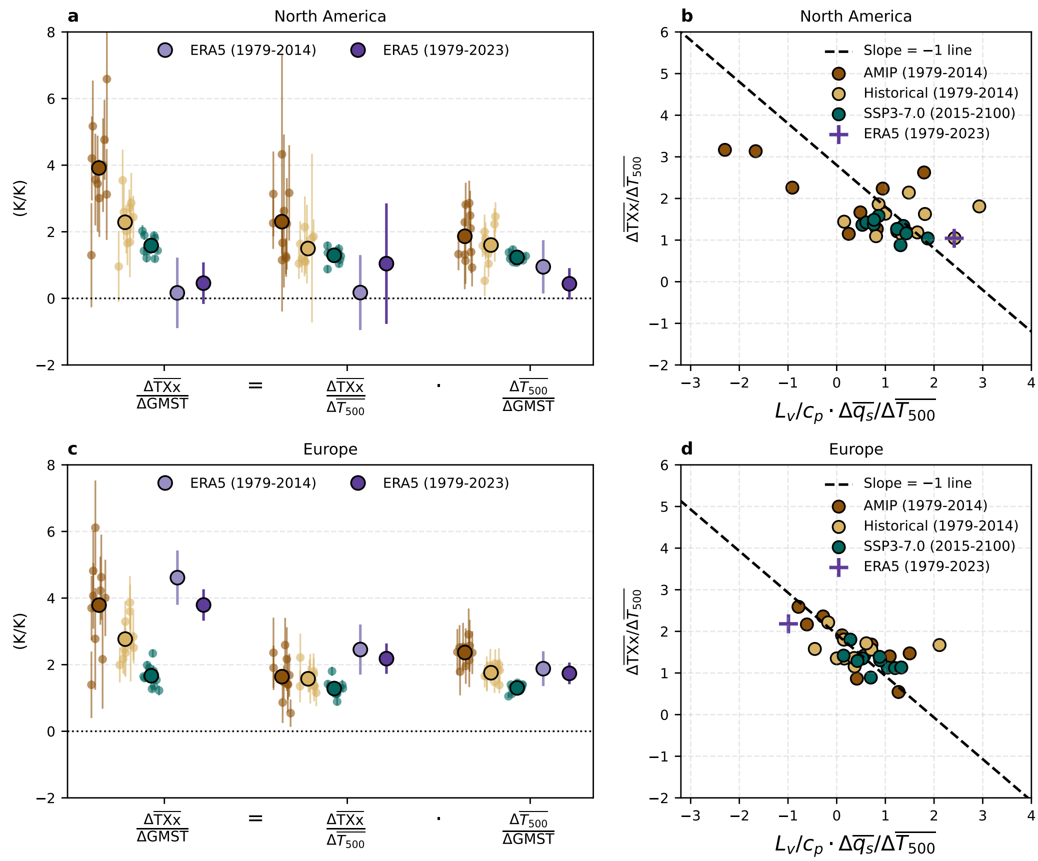

FIG. 11. Decomposition of TXx warming per Kelvin of GMST warming for land grid points in the North America box (a) and the Europe box (c), and the compensationg relationship between surface warming and moistening on the annual hottest days for these two regions (b,d)

broader midlatitude land patterns, in opposite directions (Figure 9). The anticorrelation between $\Delta\overline{\mathrm{TXx}}/\Delta\overline{T}_{500}$ and $L_v\,\Delta\overline{q_s}/(c_p\,\Delta\overline{T}_{500})$ for these smaller regions indicates that Eq. (4) remains a useful constraint at regional scales. ERA5 lying at the high-warming, low-moistening end of the model distribution for Europe, and at the low-warming, high-moistening end for North America (Figure 11b,d) suggests that land-surface processes contribute to the anomalous TXx trends in these two regions.

$\Delta\overline{T}_{500}/\Delta\mathrm{GMST}$ varies strongly across experiments and explains most of the experiment dependence of $\Delta\overline{\mathrm{TXx}}/\Delta\mathrm{GMST}$. This variation is potentially linked to the changes in aerosol forcing. Experiments with large aerosol reductions (e.g., AMIP, Historical) show exaggerated $\Delta\overline{T_{500}}/\Delta$GMST, while greenhouse gas–dominated experiments such as SSP3-7.0 show moderate $\Delta\overline{T_{500}}/\Delta$GMST. Other factors such as SST warming patterns and arctic amplification could also contribute, but are not investigated here.

We further show that this diagnostic framework based on near-convective neutrality applies similarly well to the annual hottest days as to the June-July-August mean. This is broadly consistent with prior findings that midlatitude land regions in summer exhibit behaviors considered typical of the tropics, including the near-moist adiabatic vertical temperature profiles and a net divergence of moist static energy (Korty and Schneider 2007; Miyawaki et al. 2022).

Finally, the decomposition framework and the anti-correlation between warming and moistening on the hottest days shown in Figure 5 may serve as a basis of emergent constraints on future TXx change. Extensions to other climate zones, variables, and heat stress metrics could also be explored in future work. Although our analysis is primarily based on ERA5, the close agreement between ERA5 TXx trends and observational datasets, as well as the consistency of ERA5 mid-tropospheric temperatures with satellite and radiosonde estimates (Zhang and Boos 2023), supports the credibility of the results. Future work incorporating more regional observations will help further evaluate the robustness of these findings.

Acknowledgments. This research used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

Data availability statement. The ERA5 hourly data on pressure levels and single levels from 1979 to present are downloaded from the Copernicus Climate Change Service Climate Data Store (`https://cds.climate.copernicus.eu`). The Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST) is downloaded from `https://www.metoffice.gov.uk/hadobs/hadisst/`. The CMIP6 model output data are downloaded from the Lawrence Livermore National Laboratory portal (https://aims2.llnl.gov/search/cmip6/).

APPENDIX

Derivation of the diagnostic framework

Expanding Eq. (2) using the definition of MSE in Eq. (3) leads to

$$c_p T_s + L_v q_s + g z_s \approx c_p T_{500} + L_v q_{\mathrm{sat}}(T_{500}) + g z_{500}, \tag{A1}$$

where the approximate equality holds when all terms are evaluated on the annual hottest days. T_{500}, $q_{\mathrm{sat}}(T_{500})$, and z_{500} represent the temperature, saturation specific humidity, and geopotential height at the 500-hPa pressure level, respectively.

The saturation specific humidity q_{sat} is computed from the saturation vapor pressure and ambient pressure using the following expression:

$$q_{\mathrm{sat}}(T, p) = \frac{\epsilon\, e_{\mathrm{sat}}(T)}{p - (1-\epsilon)\, e_{\mathrm{sat}}(T)}, \tag{A2}$$

where $\epsilon = R_d/R_v$ is the ratio of the gas constants for dry air and water vapor, and p is the ambient pressure. The saturation vapor pressure $e_{\mathrm{sat}}(T)$ is evaluated using the Clausius–Clapeyron relation in the form of the Tetens approximation:

$$e_{\mathrm{sat}}(T) = a_1 \exp\left(\frac{a_2(T - T_0)}{T - a_3}\right), \tag{A3}$$

where $a_1 = 611.21\,\mathrm{Pa}$, $a_2 = 17.502$, $a_3 = 32.19\,\mathrm{K}$, and $T_0 = 273.15\,\mathrm{K}$.

To further simplify Eq. (A1), the approximation for z_{500} from Zhang and Boos (2023) is adopted, assuming a linear relationship between z_{500} and T_{500},

$$z_{500} \approx \frac{\overline{z_{500}}}{\overline{T_{500}}} T_{500}, \tag{A4}$$

where $\overline{z_{500}}$ and $\overline{T_{500}}$ are June-July-August climatological geopotential height and temperature at 500 hPa, taking the values of 5682 m and 258.8 K, respectively. Substituting this approximation into Eq. (A1) gives

$$c_p T_s + g z_s + L_v q_s \approx c_p T_{500} + L_v q_{\mathrm{sat}}(T_{500}) + \frac{g\overline{z_{500}}}{\overline{T_{500}}} T_{500} = c_p T_b(T_{500}). \tag{A5}$$

T_b is the same as $T_{s,\max}$ in Zhang and Boos (2023) (switched to avoid confusion with the annual maximum temperature here) and is a function of T_{500} alone:

$$T_b(T_{500}) = T_{500} + \frac{L_v}{c_p} q_{\mathrm{sat}}(T_{500}) + \frac{g\overline{z_{500}}}{c_p\overline{T_{500}}} T_{500}. \tag{A6}$$

Perturbing Eq. (2) with respect to a reference climate thus gives

$$c_p \Delta T_s + L_v \Delta q_s \approx c_p \frac{\mathrm{d}T_b(T_{500})}{\mathrm{d}T_{500}} \Delta T_{500}. \tag{A7}$$

Dividing Eq. (A7) by $c_p \Delta T_{500}$ and replacing T_s with TXx gives Eq. (4).

References

Byrne, M. P., 2021: Amplified warming of extreme temperatures over tropical land. *Nature Geoscience*, **14 (11)**, 837–841.

Chen, J., A. Dai, Y. Zhang, and K. L. Rasmussen, 2020: Changes in convective available potential energy and convective inhibition under global warming. *Journal of Climate*, **33 (6)**, 2025–2050.

Douville, H., and M. Plazzotta, 2017: Midlatitude summer drying: An underestimated threat in cmip5 models? *Geophysical Research Letters*, **44 (19)**, 9967–9975.

Duan, S. Q., F. Ahmed, and J. D. Neelin, 2024: Moist heatwaves intensified by entrainment of dry air that limits deep convection. *Nature Geoscience*, **17 (9)**, 837–844.

Duan, S. Q., K. A. McKinnon, and I. R. Simpson, 2025: The Impact of Soil Preconditioning on the Evolution of Heatwaves Under Constrained Circulation: A Case Study of the 2021 Pacific Northwest Heatwave. *Earth's Future*, **13**, e2025EF006 216, doi:10.1029/2025EF006216.

Dunn, R. J., K. M. Willett, A. Ciavarella, and P. A. Stott, 2017: Comparison of land surface humidity between observations and cmip5 models. *Earth System Dynamics*, **8 (3)**, 719–747.

Dunn, R. J., and Coauthors, 2020: Development of an updated global land in situ-based data set of temperature and precipitation extremes: Hadex3. *Journal of Geophysical Research: Atmospheres*, **125 (16)**, e2019JD032 263.

Faranda, D., G. Messori, A. Jezequel, M. Vrac, and P. Yiou, 2023: Atmospheric circulation compounds anthropogenic warming and impacts of climate extremes in europe. *Proceedings of the National Academy of Sciences*, **120 (13)**, e2214525 120.

Fischer, E. M., and R. Knutti, 2013: Robust projections of combined humidity and temperature extremes. *Nature Climate Change*, **3 (2)**, 126–130.

Horowitz, R. L., K. A. McKinnon, and I. R. Simpson, 2022: Circulation and soil moisture contributions to heatwaves in the united states. *Journal of Climate*, **35 (24)**, 8031–8048.

Korty, R. L., and T. Schneider, 2007: A climatology of the tropospheric thermal stratification using saturation potential vorticity. *Journal of Climate*, **20 (24)**, 5977–5991.

Li, F., and T. Tamarin-Brodsky, 2025: Atmospheric stability sets maximum moist heat and convection in the midlatitudes. *arXiv preprint arXiv:2501.05351*.

McKinnon, K. A., I. R. Simpson, and A. P. Williams, 2024: The pace of change of summertime temperature extremes. *Proceedings of the National Academy of Sciences*, **121 (42)**, e2406143 121.

Miyawaki, O., T. A. Shaw, and M. F. Jansen, 2022: Quantifying energy balance regimes in the modern climate, their link to lapse rate regimes, and their response to warming. *Journal of Climate*, **35 (3)**, 1045–1061.

Morice, C. P., and Coauthors, 2021: An updated assessment of near-surface temperature change from 1850: The hadcrut5 data set. *Journal of Geophysical Research: Atmospheres*, **126 (3)**, e2019JD032 361.

Nocco, M. A., R. A. Smail, and C. J. Kucharik, 2019: Observation of irrigation-induced climate change in the midwest united states. *Global change biology*, **25 (10)**, 3472–3484.

O'Gorman, P. A., and T. Schneider, 2009: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, **106 (35)**, 14 773–14 777.

Pfahl, S., P. A. O'Gorman, and E. M. Fischer, 2017: Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, **7 (6)**, 423–427.

Santer, B. D., and Coauthors, 2017a: Causes of differences in model and satellite tropospheric warming rates. *Nature Geoscience*, **10 (7)**, 478–485.

Santer, B. D., and Coauthors, 2017b: Comparing tropospheric warming in climate models and satellite data. *Journal of Climate*, **30 (1)**, 373–392.

Schumacher, D., M. Hauser, and S. I. Seneviratne, 2022: Drivers and mechanisms of the 2021 pacific northwest heatwave. *Earth's Future*, **10 (12)**, e2022EF002 967.

Seneviratne, S., and Coauthors, 2021: Chapter 11: Weather and climate extreme events in a changing climate. 2021.

Shaw, T. A., and Coauthors, 2024: Regional climate change: consensus, discrepancies, and ways forward. *Frontiers in Climate*, **6**, 1391 634.

Simpson, I. R., K. A. McKinnon, D. Kennedy, D. M. Lawrence, F. Lehner, and R. Seager, 2024: Observed humidity trends in dry regions contradict climate models. *Proceedings of the National Academy of Sciences*, **121 (1)**, e2302480 120.

Singh, J., S. Sippel, and E. M. Fischer, 2023: Circulation dampened heat extremes intensification over the midwest usa and amplified over western europe. *Communications Earth & Environment*, **4 (1)**, 432.

Singh, M. S., Z. Kuang, E. D. Maloney, W. M. Hannah, and B. O. Wolding, 2017: Increasing potential for intense tropical and subtropical thunderstorms under global warming. *Proceedings of the National Academy of Sciences*, **114 (44)**, 11 657–11 662.

Sugiyama, M., H. Shiogama, and S. Emori, 2010: Precipitation extreme changes exceeding moisture content increases in miroc and ipcc climate models. *Proceedings of the National Academy of Sciences*, **107 (2)**, 571–575.

Thiery, W., and Coauthors, 2020: Warming of hot extremes alleviated by expanding irrigation. *Nature communications*, **11 (1)**, 290.

Vautard, R., and Coauthors, 2023: Heat extremes in western europe increasing faster than simulated due to atmospheric circulation trends. *Nature Communications*, **14 (1)**, 6803.

Vogel, M. M., R. Orth, F. Cheruy, S. Hagemann, R. Lorenz, B. J. van den Hurk, and S. I. Seneviratne, 2017: Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks. *Geophysical Research Letters*, **44 (3)**, 1511–1519.

Zhang, Y., and W. R. Boos, 2023: An upper bound for extreme temperatures over midlatitude land. *Proceedings of the National Academy of Sciences*, **120 (12)**, e2215278 120.