

Assessing the plausibility of unprecedented events: A process-based approach applied to month-long heatwaves in Western Europe

Florian E. Roemer¹, Erich M. Fischer¹, Robin Noyelle¹, and Reto Knutti¹

¹Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland

Correspondence: Florian E. Roemer (florian.roemer@env.ethz.ch)

Abstract. Climate model based storylines of individual climate and weather events are increasingly used to quantify impacts, vulnerability, or stress-test infrastructure to inform adaptation decisions. Here, we present an approach to test the plausibility of unprecedented climate storylines based on physical conformity, internal consistency, and historical precedent. We apply this approach to assess the plausibility of month-long heatwaves in Western Europe that would exceed existing record temperatures by more than 5 K. These heatwaves are based on model simulations using ensemble boosting, a computationally efficient method to simulate unprecedented events. We compare these unprecedented heatwaves with historical heatwaves in a reanalysis data set, using standardised anomalies relative to a time-evolving climatology of relevant physical variables. We show that these unprecedented heatwaves are associated with well-known physical drivers that are similar to historical heatwaves, including strong anticyclones and low soil moisture. The anomalies of these drivers are more intense in magnitude, but similar in their spatial patterns and in their relationships with each other. However, we also reveal substantial discrepancies between historical and unprecedented heatwaves, which are related to atmospheric blocking, land-atmosphere interactions, and their temporal persistence. Despite these discrepancies, we conclude that these long-lasting, unprecedented heatwaves should not be ruled out as implausible and thus highlight the need to anticipate such events when planning adaptation measures. Similar approaches can be used to assess the plausibility of other types of extreme events.

1 Introduction

Unprecedented events represent a particularly difficult challenge for extreme event research: It is essential to study them for disaster preparedness, but their absence from reanalysis records limits our ability to make robust statements about their potential magnitude and characteristics (Fischbacher-Smith, 2010; Kelder et al., 2025). Unprecedented events can be simulated in climate models, but the lack of comparable historical analogues leads to the fundamental question whether these simulated events are plausible to occur in the real world. The importance of anticipating future unprecedented events has been further highlighted by recent extreme events such as the 2021 Pacific North West heatwave, which broke previous temperature records by almost 5 K (Bartusek et al., 2022; White et al., 2023; Fleishman et al., 2025). Here we assess the plausibility of simulated month-long heatwaves in Western Europe that would exceed existing records by similar margins (Lüthi et al., 2024).

Future record-breaking events can be studied using climate storylines, which provide physically self-consistent unfoldings of plausible future events, making it possible to study their characteristics and underlying physical processes (Hazeleger et al.,

2015; Shepherd et al., 2018; Sillmann et al., 2021; Baulenas et al., 2023; van den Hurk et al., 2023; Baldissera Pacchetti et al., 2024). Among other approaches, such unfoldings can be selected from single-model initial-condition large ensembles of climate models (Bador et al., 2017; van der Wiel et al., 2021; Suarez-Gutierrez et al., 2020; Fischer et al., 2021) or from ensembles of weather prediction models (UNSEEN approach, van den Brink et al., 2005; Thompson et al., 2017). Furthermore, model ensembles can be optimised to simulate more frequent and intense extreme events by using rare event algorithms such as the GKTL algorithm (Giardina et al., 2011; Ragone et al., 2018; Ragone and Bouchet, 2021; Yiou et al., 2023; Noyelle et al., 2025a, b), ensemble boosting (Gessner et al., 2021; Fischer et al., 2023; Lüthi et al., 2024; Suarez-Gutierrez et al., 2025), or based on adaptive multilevel splitting (Finkel and O’Gorman, 2024, 2026). In recent years, similar approaches have been used to create extreme event storylines using machine learning (Mahesh et al., 2025a, b; Whittaker and Di Luca, 2026). However, how can we test whether simulated unprecedented events are plausible to occur in the real world?

Assessing the plausibility of unprecedented events is essential because it informs us about the potential intensity of events that society needs to prepare for (Nordmann, 2013). This makes plausibility an essential criterion for decision making, such as for natural disaster preparations (van der Helm, 2006; Shepherd et al., 2018; Sillmann et al., 2021) and is useful in a wide range of fields, ranging from epidemiology to economics (Derbyshire, 2022; Goodwin and Wright, 2010). This is because studying extreme events that are considered plausible — irrespective of their probability of occurrence — allows for the exploration of future scenarios while embracing inherent uncertainties (Selin and Guimarães Pereira, 2013). Therefore, studying plausible future extremes is essential for providing tangible what-if scenarios that are needed to inform disaster preparation.

But what does plausible mean? Located between possible (without contradiction) and probable (likely to occur), plausible is associated with being credible, trustworthy, and reasonable (Amara, 1991; Wilson, 1998; Selin, 2006; van der Helm, 2006; Nordmann, 2013; Urueña, 2019). However, the concept of plausibility is based on subjective judgement, and thus there are no universal criteria (Selin and Guimarães Pereira, 2013). In this work, we base our assessment of plausibility on tangible criteria suitable for climate storylines: conformity with physical principles, internal consistency, and historical precedent (Amara, 1991; Nordmann, 2013; Shepherd et al., 2018). These criteria can be used to either affirm or question the plausibility, depending on whether the underlying physical processes are internally consistent with process understanding and externally supported by historical events. Fundamentally, it is not possible to conclusively prove that a given event is plausible; the only feasible approach is to perform a series of tests in which the confidence increases with the number of distinct tests performed.

There are a number of different ways to translate plausibility criteria into concrete analysis steps. Most commonly, the plausibility of simulated extremes is statistically analysed through extreme value theory and model fidelity tests, for example, by comparing statistical distributions and moments of model output with observations or reanalysis (Thompson et al., 2017, 2019; Vautard et al., 2019; Philip et al., 2020; Kelder et al., 2022a, b). In contrast, only few studies assess the plausibility of simulated extreme events by analysing the physical processes that drive them, for example, by studying the temporal evolution of variables and the correlations between them (Vautard et al., 2019; Kelder et al., 2022b) or the dynamic and thermodynamic precursors of the events (Klif et al., 2026). An important part of this assessment involves comparisons with historical events; however, there are inherent limitations of this approach when analysing unprecedented extremes in a transient climate. This is particularly obvious for extreme events that are directly affected by global and regional warming trends; for example, heatwaves that reach

previously impossible temperatures are becoming possible in a warming climate (Fischer et al., 2025). To account for this, we propose a process-based approach centred around standardised anomalies with respect to a time-evolving climatology, that is, we quantify how extreme an event is relative to the background climate (mean and variability) in which it occurs. This allows us to directly and systematically compare anomalies in a wide range of physical drivers during simulated unprecedented events with anomalies during historical events and to evaluate their consistency with current process understanding. In this way, we expand on existing process-based studies and offer an additional layer of evidence to complement statistical approaches.

To demonstrate our approach, we focus on heatwaves in Western Europe. This is motivated by the fact that both mean summer temperatures and extreme heat in Western Europe are increasing faster than in almost any other region, caused by a combination of anthropogenic global warming, regional reductions in aerosol emissions, and changes in circulation (Singh et al., 2023; Vautard et al., 2023; Schumacher et al., 2024; World Meteorological Organization et al., 2025). As a consequence, Western Europe has been a hotspot for record-breaking heat extremes: Since the turn of the 21st century, Western Europe experienced record-breaking heat extremes in 2003, 2006, 2015, 2018, 2019, 2022, 2025, and most recently in 2026 (Rebetez et al., 2009; García-Herrera et al., 2010; Russo et al., 2015; Kornhuber et al., 2019; Sánchez-Benítez et al., 2022; Feser et al., 2024; Hotz et al., 2024; Copernicus Climate Change Service, 2025; World Meteorological Organization, 2026). Given current global warming trends, the likelihood of record-breaking heatwaves is expected to increase further in the future (Fischer et al., 2025), with parts of Western and Central Europe among the most at-risk regions for such heatwaves (Thompson et al., 2023).

The processes that govern heatwaves occur on a wide range of spatial and temporal scales and include interactions between the atmosphere, ocean, cryosphere, land surface, and biosphere (Barriopedro et al., 2023; Domeisen et al., 2023). In mid-latitudes, heatwaves are often caused by quasi-stationary Rossby wave packets that lead to persistent anticyclones (Röthlisberger et al., 2019). These anticyclones can disrupt the zonal atmospheric circulation leading to atmospheric blocking (Drouard and Woollings, 2018; Röthlisberger and Martius, 2019; Kautz et al., 2022). In summer, this can favour heat extremes through horizontal temperature advection, vertical adiabatic subsidence, and diabatic heating at the surface through solar radiation and turbulent fluxes (Zschenderlein et al., 2019; Barriopedro et al., 2023; Röthlisberger et al., 2026). These surface fluxes are often amplified by land-atmosphere feedbacks (Fischer et al., 2007; Seneviratne et al., 2010), particularly during persistent heatwaves that last multiple weeks (Tuel and Martius, 2023; Pappert et al., 2025). These generally well-understood physical drivers provide the foundation for our process-based approach to assess plausibility.

Heatwaves have severe impacts on human health, ecosystems, and infrastructure (Easterling et al., 2000; Zuo et al., 2015), and these impacts can be particularly severe during long-lasting heatwaves (D’Ippoliti et al., 2010; Anderson and Bell, 2011; Polt et al., 2023). Prominent examples of long-lasting heatwaves include the Eastern European heatwave of 2010 (Barriopedro et al., 2011; Dole et al., 2011) as well as the Western European heatwaves of 1976 (Green, 1977; Kendon et al., 2024) and 2003 (Black et al., 2004; García-Herrera et al., 2010), which resulted in thousands of excess deaths (Le Tertre et al., 2006; Robine et al., 2007). Recent model experiments based on ensemble boosting suggest that persistent, month-long heatwaves that would break reanalysis records by more than 5 K in Western Europe are possible in the near future, which would have severe effects on heat-related mortality (Lüthi et al., 2024). Given the unprecedented nature of these month-long heat extremes, here we address the pressing question whether they are plausible to occur in the real world.

2 Methods

2.1 Climate model and reanalysis data

We create unprecedented heatwave storylines using ensemble boosting simulations in CESM2 with daily model output (see Appendix A). As a climatological reference, we use daily output of the 100-member CESM2 large ensemble (LE) that follows the historical (1850–2014) and SSP3-7.0 (2015–2100) scenarios (Danabasoglu et al., 2020; Rodgers et al., 2021). Most of the atmospheric and surface variables that we use are available as model output (Tab. 1). Exceptions are the diurnal temperature range (DTR, the difference between daily maximum temperature TX and daily minimum temperature TN), total precipitation (PREC, the sum of convective and large-scale precipitation), and the evaporative fraction (EF, calculated from the surface fluxes of latent heat LHF and sensible heat SHF as $EF = LHF / (LHF + SHF)$).

To compare with historical heatwaves, we use daily output of the ERA5 reanalysis for 1950–2025 (Hersbach et al., 2020) for all variables in Tab. 1 except for SM, for which we use ERA5-Land instead (Muñoz-Sabater et al., 2021). In addition to these quality-checked final products, we use preliminary ERA5 and ERA5-Land data for the month-long heatwave in Western Europe from 2026-06-17 to 2026-07-16 (Hersbach et al., 2020; World Meteorological Organization, 2026).

For the temperature budget analysis, we additionally use daily model output on the 850 hPa level from the boosting simulations and from ERA5 for the temperature (T850), zonal wind (U850), meridional wind (V850), and vertical wind (ω 850), as well as the temperature tendency due to physical parametrisations. In CESM2, this is available as the PTTEND variable, in ERA5 we use the avg_ttpm variable which is only available from the deterministic forecast model.

For SM, the top-layer of the soil differs between CLM, CESM2’s land model (top 10 cm) and ERA5-Land (top 7 cm). To compare the absolute values of SM, the ERA5 values (volumetric soil water in m^3/m^3) are converted to CESM2 units (soil water mass in kg/m^2). Performing the same analysis for the top 1 m of the soil does not alter our conclusions (not shown).

We focus on a box in Western Europe (WE) where the boosting simulations detect heatwaves that exceed historical records by particularly large margins (0–10°E, 46–51°N). All CESM2 and ERA5 variables are averaged over this domain before we calculate standardised anomalies (see below). To study large-scale conditions, we further use the full horizontal fields of all variables analysed in Europe (12°W – 42° E, 35°N – 72°N). For this analysis, we remap the horizontal ERA5 fields to the CESM2 grid using first order conservative remapping (Schulzweida, 2023). Unless stated otherwise, all of our analysis is performed for the summer months of June, July, and August (JJA).

2.2 Standardised anomalies

To enable a meaningful comparison between the unprecedented heatwave storylines and historical heatwaves, we express every variable x as standardised anomaly z_x (in the following simply “anomaly”):

$$z_x(r, y, s) = \frac{x(r, y, s) - \bar{x}(r, y \pm 15 \text{ yr}, s \pm 2 \text{ d})}{x'(r, y \pm 15 \text{ yr}, s \pm 2 \text{ d})}. \quad (1)$$

Table 1. Heatwave characteristics and drivers analysed in this study.

TM	daily mean temperature
TX	daily maximum temperature
TN	daily minimum temperature
DTR	diurnal temperature range
Z500	500 hPa geopotential height
U500	500 hPa zonal wind (positive eastwards)
V500	500 hPa meridional wind (positive northwards)
ω 500	500 hPa vertical wind (positive downwards)
SSR	surface downwelling solar flux (positive downwards)
PREC	total precipitation
SM	top-layer soil moisture
LHF	surface latent heat flux (positive upwards)
SHF	surface sensible heat flux (positive upwards)
EF	surface evaporative fraction

Here, the region r refers to an individual grid cell of the CESM2 grid or the box-average over WE, y refers to the year, s refers to the day of the summer season (JJA), and $\bar{x}$ and x' refer to the mean and standard deviation of x , respectively, calculated over a centred 31-year window in time and a centred 5-day window over the seasonal cycle.

For the boosting simulations, $\bar{x}$ and x' are derived using the full 100-member CESM2-LE, except for SM, which is available for only 90 members. For CESM2-LE itself, the standardisation is performed for the period 2005–2035 (the same period as the 35-member CESM2 ensemble on which the boosting simulations are based). For ERA5, the standardisation is limited to the period 1965–2010, where a centred 31-year reference period can be constructed based on the data record available at the time of the analysis. When LHF and SHF are of opposite sign but similar magnitude, this can lead to very large absolute values in EF locally, particularly over ocean. Thus, only grid cells with $0 < EF < 1$ are used to calculate EF anomalies.

The standardisation is performed separately for daily values and a 30-day running mean. The 30-day running mean is assigned to the “central day”, the 15th day within a given 30-day period. We consider all 30-day periods whose central day lies in JJA, including periods that start in May or end in September. To calculate the 30-day running mean of the box-mean over WE, we require data points on all 30 days within a given window; for the running mean of individual grid points, this is relaxed to at least 25 data points within a 30-day window in order to reduce the effect of individual missing values in EF (see above).

2.3 Heatwave selection

Based on anomalies in the 30-day running mean of TM in WE, we select all 20 heatwaves in our boosting simulations that exceed the most extreme heatwave in ERA5 (“unprecedented” heatwaves U1–U20, Tab. B1). From ERA5, we select all 30-day periods whose TM anomalies in WE exceed the 95th percentile of all JJA 30-day means in the reanalysis record, yielding a

total of 11 “historical” heatwaves (Tab. B2). Our anomaly-based approach limits the reanalysis record to 1965–2010, which
 145 excludes many of the most extreme heatwaves that occurred in the last 15 years. Thus, we also perform the same analysis
 for the absolute TM values in WE in 2011–2026, which yields 10 “recent” historical 30-day heatwaves (in 2013, 2015, 2018,
 2019, 2020, 2022, 2023, 2024, 2025, and 2026).

To avoid selecting temporally overlapping heatwaves, we only consider the largest 30-day anomaly of each boosting run,
 and only consider the largest 30-day anomalies within ± 30 days in ERA5, allowing more than one heatwave per year to
 150 be selected. In this way, we select the 30-day periods that feature the most anomalously high temperatures with respect to
 the background climate and the timing within the summer season in which they occur. To directly compare the historical
 heatwaves to CESM2 heatwaves of similar magnitude, we select the 20 heatwaves in CESM2-LE with the most similar 30-day
 TM anomaly over WE for each historical heatwave, separately for two different periods (the historical period 1965–2010 and
 the period of the boosting simulations 2005–2035). In total, we thus select 220 “moderate” CESM2 heatwaves for each climate
 155 period, whose TM anomalies differ from the respective historical heatwave by at most $\pm 0.1 \sigma$ (Fig. 1a).

Of the 20 unprecedented heatwaves in our simulations, 15 originate from the same parent ensemble P1 (model year 2015),
 and the other five come from a second parent ensemble P2 (model year 2031). Unprecedented heatwaves from the same
 parent feature quite similar SM anomalies, but are still distinct in their atmospheric conditions (Fig. B1) and in the relative
 importance of relevant physical drivers (Fig. B2b). Nevertheless, to increase the independence and representativeness of the
 160 analysed heatwaves, we also include in our analysis the most extreme 30-day heatwaves of the three other parent ensembles
 P3–P5 (model years 2030, 2028, and 2028). However, because their 30-day TM anomalies are smaller than the most extreme
 historical heatwave, they are not considered “unprecedented” heatwaves and instead analysed separately.

2.4 Temperature budget analysis

Following Tyrllis et al. (2013) and Garfinkel et al. (2024), the Eulerian temperature budget can be written as

$$165 \quad \underbrace{\frac{\partial T}{\partial t}}_{\text{local}} = - \underbrace{\frac{u}{r \cos(\phi)} \frac{\partial T}{\partial \lambda}}_{\text{zonal}} - \underbrace{\frac{v}{r} \frac{\partial T}{\partial \phi}}_{\text{meridional}} - \underbrace{\frac{T}{\theta} \frac{\partial \theta}{\partial p} \omega}_{\text{adiabatic}} + \underbrace{\frac{J}{C_p}}_{\text{diabatic}} + \text{residual}, \quad (2)$$

where T is the temperature, t is the time, u , v , and ω are the zonal, meridional, and vertical wind components, θ is the potential
 temperature, p is pressure, ϕ is latitude, λ is longitude, r is the radius of Earth, J is the diabatic heating rate, and C_p is the
 heat capacity of air at constant p . In this way, the local temperature tendency is decomposed into four terms: zonal temperature
 advection, meridional temperature advection, adiabatic changes resulting from vertical temperature advection, and diabatic
 170 sources and sinks (such as radiation, evaporation/condensation, and surface exchanges), as well as a residual term.

For historical and recent heatwaves in ERA5 (except 2026), and for precededent and unprecedented heatwaves in CESM2,
 we evaluate Eq. (2). To reduce the effects of topography, we focus on the 850 hPa level (Tyrllis et al., 2013; Garfinkel et al.,
 2024). We calculate the zonal, meridional, and adiabatic terms from second order centred differences of daily model output on
 the native grids of ERA5 and CESM2, but the results are not sensitive to spatial remapping. For the diabatic term, we use the

175 daily mean temperature tendency due to physical parametrisations. The resulting values are averaged over WE, and separately averaged over days with positive and negative local T tendencies during each event.

2.5 Persistence

To analyse temporal persistence, we calculate the number of consecutive days on which daily anomalies are $\geq +1\sigma$ (for TM and Z500) or $\leq -1\sigma$ (for SM). For ERA5 and CESM2-LE, we count the occurrences of consecutive days whose end date lies
180 in JJA and divide this count by 92 to yield the frequency per summer season. For precededented and unprecedented heatwaves, we use the longest persistence duration that ends in the respective boosting run, including periods that start in the parent run. Thus, persistence durations from all sources can include days in months preceding JJA.

We analogously evaluate the persistence of atmospheric blocking using a simple blocking index (Tibaldi and Molteni, 1990), setting the northern and southern latitudes to the edges of our European domain ($35^\circ\text{N} - 72^\circ\text{N}$). A day counts as blocked, if
185 Z500 satisfies both blocking criteria of Tibaldi and Molteni (1990) for at least one longitude within our WE domain ($0-10^\circ\text{E}$). All calculations are performed on the CESM2 grid, but spatial mapping has only a negligible effect.

2.6 Multilinear model

We construct a multilinear statistical model based on a multivariate ordinary least-squares regression. For simplicity, we focus on a set of five predictors (Z500, U500, V500, SM, SSR) and apply it separately for three different prognostic variables (TM, TX, and TN). For each prognostic variable, the model is fitted to the maxima of 30-day z_{TM} of each summer season, with
190 separate versions trained on ERA5 and CESM2-LE. This gives six different model versions which we use to infer anomalies in the prognostic variable during the unprecedented heatwaves based on the simulated anomalies of the five predictors as

$$z_{x,\text{inf}}(d) = \sum_p \beta_p z_p. \quad (3)$$

Here, $x \in [\text{TM}, \text{TX}, \text{TN}]$ is the prognostic variable, $p \in [\text{Z500}, \text{U500}, \text{V500}, \text{SM}, \text{SSR}]$ are the predictors, $d \in [\text{CESM2-LE}, \text{ERA5}]$ are the data sources on which the model is trained, and β_p are the coefficients of the model.
195

The predictors were chosen as a compromise between the simplicity of the model and the inclusion of the most important processes that govern the unprecedented heatwaves. Z500 is chosen due to the importance of anticyclones for mid-latitude heatwaves. Because our model is based on WE-mean anomalies, it does not contain any information on the position of the anticyclone relative to WE, for which we include U500 and V500 as predictors. SM is included to capture the important role
200 of low soil moisture during heatwaves, and SSR is included to account for the diabatic contribution of radiation.

The model is meant to be a simplified representation of the relevant physical processes that drive month-long temperature extremes in Western Europe. Inherently, it cannot capture processes involving variables not included in the model, or non-linear relationships, including, but not limited to, land-atmosphere feedbacks between soil moisture and temperature. To further investigate this, we employ a modified version of the model to infer anomalies in SHF, LHF, and EF. As predictors, we chose
205 TM, SM and SSR, which are the variables most directly linked to the variables inferred in the absence of atmospheric humidity or near-surface wind in our analysis set of variables.

3 Results

To evaluate plausibility, we focus on storylines of unprecedented month-long heatwaves, defined as periods during which the 30-day running mean TM anomaly in WE surpasses the most extreme historical heatwave in June/July 1976 (3.6σ , the only historical event exceeding 3σ , see Tab. B2). In our CESM2 boosting simulations, we identify 20 such unprecedented heatwaves (labelled U1–U20), with anomalies between 3.6σ and 4.3σ (purple circles in Fig. 1a, Tab. B1, Sect. 2.3). The return period estimates of these unprecedented heatwaves range from around 300 years to more than 6000 years, with 95 % confidence intervals of 200–2000 years (lower bounds) to 500 years– ∞ (upper bounds), with infinite upper bounds for the three most extreme heatwaves (Fig. 1b; see Appendix A). By centring our assessment on the most extreme heatwaves in our simulations, we target the events that are most likely to be influenced by limitations in the applicability of CESM2. However, this approach also comes with some drawbacks. In addition to the limited sample size of unprecedented heatwaves in our simulations ($n = 20$), they originate from only two different parent ensembles, leading to similar preconditions in slow components such as the land surface (Sect. 2.3, Fig. B1c).

We address this challenge in two ways: (1) We increase the representativeness of our analysis by additionally including the most extreme heatwaves of the three other parent ensembles in our simulations for some of our assessments. Because their anomalies are smaller than the most extreme historical heatwave, we refer to them as “precedented” heatwaves (Sect. 2.3, pink circles in Fig. 1). (2) Using large-deviation theory, it has been shown that the typicality of heatwaves tends to increase with their magnitude, that is, more extreme heatwaves tend to become more similar in their physical pathways, as they require a “perfect storm” of strong contributions by most relevant heatwave drivers to occur (Lucarini et al., 2023; Noyelle et al., 2024). For this reason, we expect many of our findings to also apply to other unprecedented heatwaves as well as to less extreme heatwaves that might only exhibit a subset of these heatwave drivers. Furthermore, if even the most extreme heatwaves simulated by CESM2 are considered plausible, this ultimate test also improves confidence in CESM2 to simulate less extreme heatwaves.

Given that the central return period estimates for the unprecedented heatwaves analysed here are finite but much longer than a human lifetime, they appear statistically possible but not probable. In the following, we address the less straight-forward question of whether these unprecedented heatwaves are plausible by assessing their physical conformity, internal consistency, and historical precedent. In practice, our assessments of these characterisations naturally overlap, but generally fall into two categories: (1) By comparing the unprecedented heatwaves from CESM2 ensemble boosting simulations to less extreme heatwaves in CESM2-LE, we assess their internal consistency within the CESM2 model. This allows us to test whether the processes present during these unprecedented heatwaves are similar to those simulated during less extreme heatwaves in CESM2. (2) By comparing them to historical heatwaves, we assess whether — despite their by definition unprecedented magnitude — these heatwaves have external empirical support from historical precedents. This allows us to test whether the processes present during these unprecedented heatwaves are similar to heatwaves in the reanalysis record.

The conformity to physical principles represents the foundation of most of our assessment procedure; we test it by comparing the underlying physical processes in these unprecedented heatwaves both internally (within CESM2) and externally (with ERA5). In this way, we test the ability of CESM2 to realistically simulate extreme heatwaves with intensities beyond those

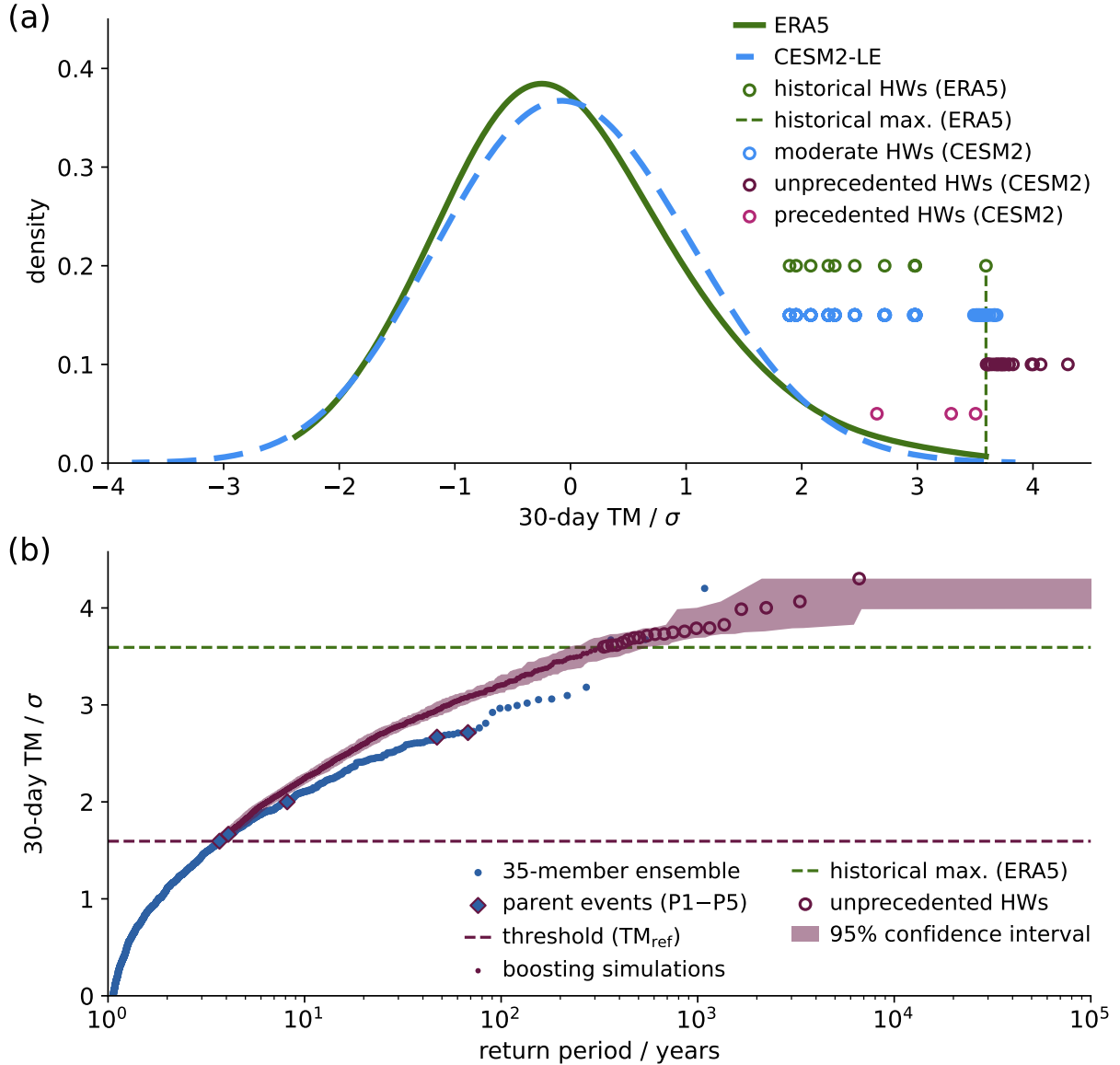


Figure 1. Unprecedented month-long heatwave storylines in Western Europe. (a) 30-day TM anomalies of the most extreme historical heatwaves (green circles), moderate heatwaves in CESM2-LE with the same TM anomalies (light blue circles), preceded (pink circles) and unprecedented heatwaves (purple circles) from ensemble boosting. Kernel density estimates of 30-day TM anomalies during JJA in ERA5 (1965–2010, green line) and CESM2-LE (2005–2035, light blue line), using a Gaussian kernel with a bandwidth of 0.5σ , truncated at the data limits. (b) Return period estimates of 30-day TM anomalies from boosting simulations (purple dots), including the selected unprecedented heatwaves (purple circles), and their 95 % confidence interval, as well as return period estimates of heatwaves in 35-member CESM2 ensemble (dark blue dots), including the selected parent heatwaves (dark blue diamonds). The thresholds for the return period calculation (TM_{ref} , dashed purple) and for the selection as unprecedented heatwave (historical maximum, dashed green) are also shown.

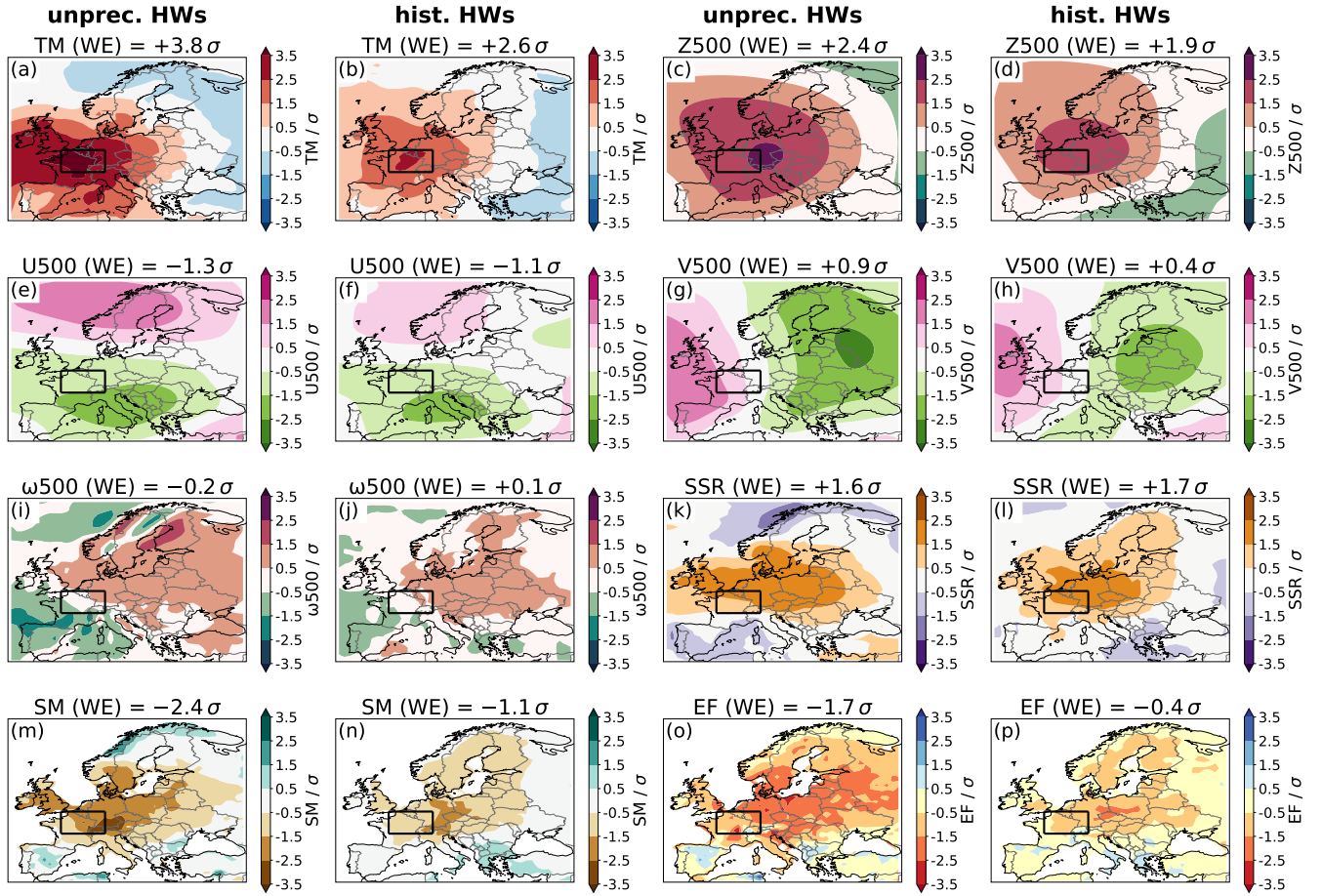


Figure 2. Spatial composites of anomalies in selected variables during unprecedented heatwaves (first and third columns) and during historical heatwaves (second and fourth columns). The surface variables SM and EF are only plotted over land.

found in the reanalysis record. The overall result of our assessment — whether the evidence supports or questions plausibility — depends on the combined results of several analysis criteria, which we present in the following.

3.1 Spatial fields

We start by exploring the large-scale spatial behaviour of unprecedented heatwaves over Europe. For this, we compare the composite spatial fields of 30-day anomalies during unprecedented and historical heatwaves for relevant heatwave variables (Fig. 2). In this context, it would support plausibility if TM anomalies arise from similar atmospheric and land-surface conditions over Europe in unprecedented and historical heatwaves. Conversely, it would not support plausibility if substantially different large-scale conditions happen to produce similar anomalies over WE.

First, the TM anomalies feature similar spatial distributions, extending slightly further west during unprecedented heatwaves (Fig. 2a,b). Similarly, the anomaly fields of the dynamical variables (Z500, U500, V500, and ω 500) are also very similar between unprecedented and historical heatwaves, with very similarly located anticyclones over Central Europe that disturb the zonal flow over Central and Southern Europe, and lead to Southerly wind anomalies over Western Europe and above-average vertical subsidence over much of Central and Eastern Europe (Fig. 2c–j). The magnitudes of the anomalies are greater during unprecedented heatwaves, which is discussed in more detail in Sect. 3.2. The largest SSR anomalies are found in areas of vertical subsidence and are very similar in magnitude during historical and unprecedented heatwaves (Fig. 2k,l).

The fields of SM and EF feature substantially more negative anomalies in large parts of Central and Eastern Europe during unprecedented heatwaves (Fig. 2m–p; see also Sect. 3.2). Despite these differences in magnitude, the spatial distributions are still similar, with the most negative values of both SM and EF occurring to the east of WE, consistent with previous findings (Noyelle et al., 2025b). Overall, the large-scale conditions during unprecedented heatwaves in WE thus appear to be remarkably consistent, both externally with historical heatwaves and internally by conforming to well-understood physical processes. However, this good agreement in standardised anomalies does not necessarily extend to the case of the underlying absolute values, which can be subject to model biases in the mean and variability of heatwave variables.

Therefore, to complement our anomaly-based analysis, we also analyse the spatial fields of absolute values during unprecedented heatwaves. For this, we compare them with “recent” heatwaves that have occurred during 2011–2026 and are not included in our anomaly time series due to their incomplete centred 31-year climatology. These heatwaves are among the most extreme heatwaves in terms of absolute TM and occur in a very similar background climate as most of our unprecedented heatwaves (Fig. 3; see Sect. 2.3). As mentioned above, the unprecedented heatwaves are on average almost 5 K warmer than most recent heatwaves (Fig. 3a,b). However, the 2026 heatwave exceeded the previous 30-day TM record by approximately 2 K, and was almost as extreme in 30-day TX as some unprecedented heatwaves (see Appendix B and Fig. B3).

Apart from more extreme TM, unprecedented heatwaves are also associated with stronger anticyclones than recent heatwaves, but they are located in very similar positions (Fig. 3c,d). During unprecedented heatwaves, anticyclones extend further northward, which can also be seen in the location of the strongest zonal winds (Fig. 3e,f), while the positions of the strongest meridional winds are more similar (Fig. 3g,h). Both historical and unprecedented heatwaves feature the strongest vertical subsidence over South-East Europe and the Mediterranean, with much smaller values over WE (Fig. 3i,j). The northward shift of the strongest westerlies in the unprecedented heatwaves relative to the historical heatwaves can also be seen in the SSR, where the unprecedented heatwaves feature much lower values over Northern Europe (Fig. 3k,l). Some of the largest discrepancies are again visible in the surface variables SM and EF, whose absolute values are substantially lower across most of Europe during unprecedented heatwaves and which (in the case of SM) also exhibit different spatial patterns (Fig. 3m–p). However, it should be noted that these surface variables are associated with substantial uncertainties in both climate models and reanalysis, and thus absolute values in particular should be compared with caution (see Sect. 4).

In summary, absolute values during unprecedented heatwaves are much more extreme than recent heatwaves, not only in TM, but also in Z500, SM, and EF. Our analysis based on absolute values thus reveals similar differences between unprecedented and (recent) historical heatwaves as our anomaly-based framework. In addition, it also highlights differences in the meridional

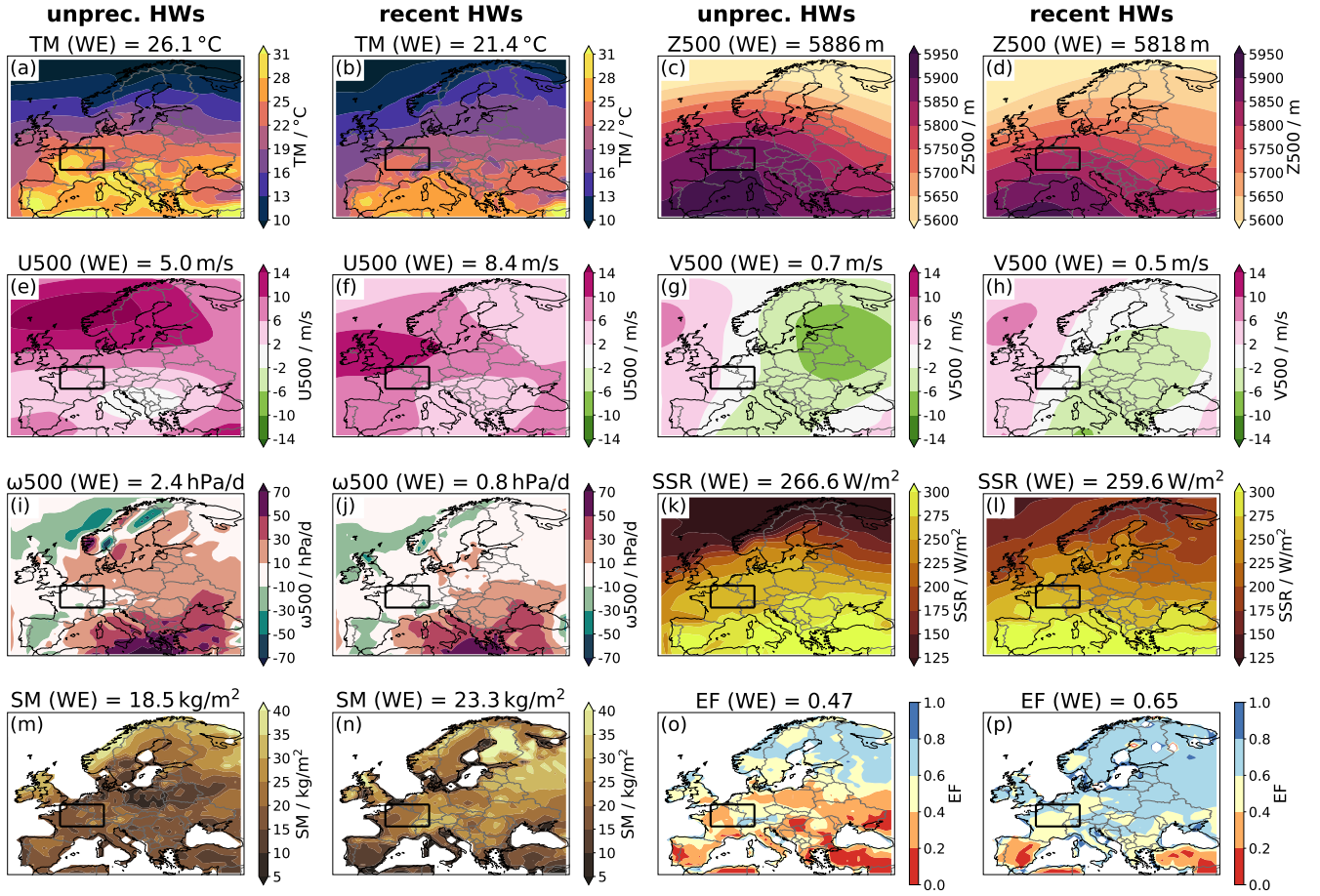


Figure 3. As Fig. 2, but for absolute values during unprecedented and recent heatwaves.

extent of the blocking anticyclone, which is greater in the unprecedented heatwaves. This indicates that CESM2 can simulate
 285 stronger blocking anticyclones than seen in the reanalysis record (see Sect. 3.4.3 for a more detailed discussion).

To conclude this section, the spatial fields of unprecedented heatwaves are very similar to those of historical heatwaves
 in anomaly space but, unsurprisingly, exhibit more discrepancies compared to recent heatwaves in their absolute values. This
 shows that unprecedented heatwaves appear to exhibit physically plausible large-scale conditions of heatwave driver anomalies,
 while the absolute values of these drivers should be interpreted with caution. In other words, plausibility in anomalies does not
 290 imply plausibility in absolute values, highlighting the importance of performing bias correction when using model output for
 impact assessment (Lüthi et al., 2024).

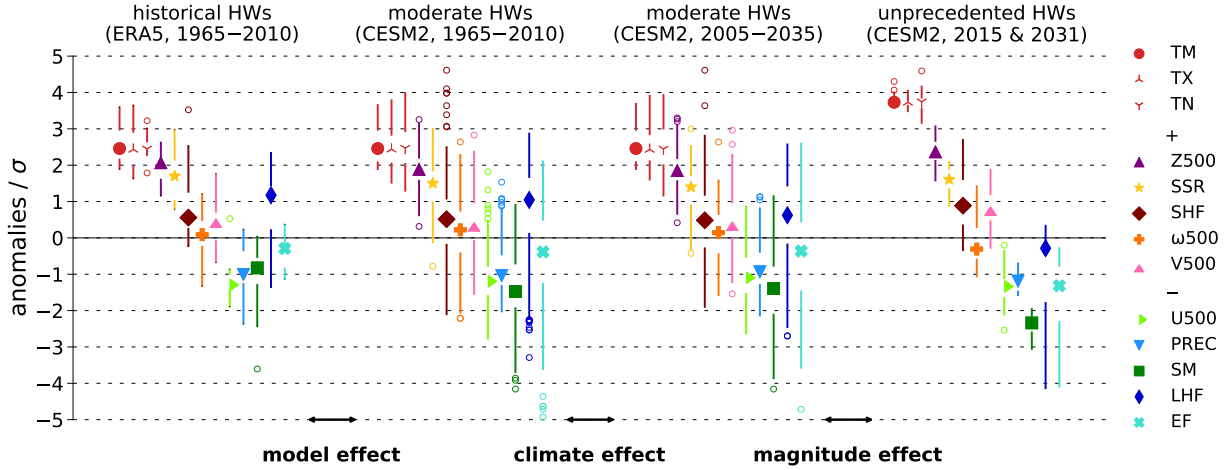


Figure 4. Anomalies of monthly mean heatwave variables in WE. Box plots of anomalies during historical heatwaves (HWs), moderate heatwaves from two different background climates, and unprecedented heatwaves: Median (filled symbols), inter-quartile range (IQR, empty space), whiskers (1.5 IQR, solid lines) and outliers (empty circles). We further indicate below the isolated effects of varying model, climate, and magnitude of heatwaves, respectively. For clarity of visualization, we broadly distinguish temperature variables (red), and variables with expected positive (“+”) and negative (“−”) anomalies during heatwaves.

3.2 Regional anomalies

As a next step, we analyse the monthly-mean anomalies in relevant heatwave variables during unprecedented heatwaves and compare them with the most extreme historical heatwaves (Tab. 1, Fig. 4). For this, we focus on the 30-day anomalies of these variables averaged over WE. In this analysis, results that would indicate plausibility include (1) similar dominant drivers during historical and unprecedented heatwaves and (2) temperature anomalies whose magnitudes are consistent with the magnitudes of relevant drivers (for example, large temperature anomalies being associated with large anomalies in key drivers). In contrast, findings that would question the plausibility of these unprecedented heatwaves include (1) systematically different dominant drivers in historical and unprecedented heatwaves and (2) temperature anomalies whose magnitudes are decoupled from the magnitudes of relevant drivers (for example, large temperature anomalies being associated with only small anomalies in key drivers).

Overall, historical and unprecedented heatwaves are governed by quite similar dominant drivers, with large positive anomalies in Z500 and SSR and substantial negative anomalies in U500, PREC, and SM. In contrast, anomalies in SHF, $\omega 500$ (which we use as a proxy for adiabatic warming) and V500 (which we use as a proxy for mid-tropospheric temperature advection) appear to play a secondary role in both historical and unprecedented heatwaves. This relatively minor role for advective and adiabatic processes is partly caused by our focus on monthly-mean regional anomalies in this section, which can obscure short-term anomalies in these drivers (see Sect. 3.4.2). To disentangle and quantify the effects of these processes more explicitly, we also perform a temperature budget analysis (Sect. 3.4.1).

The main differences between historical and unprecedented heatwaves can be seen in variables related to land-atmosphere exchanges. In relative terms, SM and EF appear to play a substantially more important role during the unprecedented heatwaves than during the historical heatwaves. Furthermore, the anomalies in LHF are generally positive in historical heatwaves but mostly negative in unprecedented heatwaves. So what exactly explains these differences?

The unprecedented heatwaves structurally differ from the historical heatwaves in three main ways: They originate from a climate model rather than reanalysis, occur in a different background climate, and are more extreme in magnitude. To disentangle these effects on the anomalies of heatwave drivers, we additionally analyse “moderate” heatwaves simulated by CESM2 that have the same magnitude as historical heatwaves and are selected separately for two different climate periods (“historical” 1965–2010 and “present” 2005–2035; see Sect. 2.3). This allows us to distinguish between the three main effects at play:

(1) The “model effect”: Most physical drivers have similar anomalies during ERA5 and CESM2 heatwaves from the same climate. The CESM2 heatwaves have slightly less positive anomalies in Z500, SSR, SHF, and LHF and more negative anomalies in SM, but only the latter difference exceeds $\pm 0.5 \sigma$. This suggests that, when comparing heatwaves of the same magnitude from the same background climate, CESM2 generally captures the physical drivers present during heatwaves in ERA5 well. However, it also shows that CESM2 overestimates the role of drying soils during heatwaves. This finding is consistent with the literature (Vogel et al., 2018; Röthlisberger et al., 2026) and will be a theme throughout this work (see Sect. 3.4).

(2) The “climate effect”: Almost all physical drivers feature very similar anomalies during moderate CESM2 heatwaves from the historical and present climates, respectively. The only noticeable difference concerns LHF which exhibits anomalies of both signs in both background climates, but is shifted towards fewer positive and more negative anomalies in the present climate. Notably, this shift is not accompanied by a shift towards more negative SM anomalies, which might be expected from physical process understanding. However, the median of absolute SM values during moderate heatwaves is around 10 % lower in the present climate than in the historical climate (not shown). This is consistent with the projected long-term decrease in SM in Western Europe (Seneviratne et al., 2010), which is not visible in the anomalies due to our use of a time-evolving climatology. This means that the same SM anomaly can imply that land-atmosphere fluxes are energy-limited during historical heatwaves (causing more positive LHF anomalies), but transitional or even moisture-limited in present or future heatwaves (causing less positive LHF anomalies; Seneviratne et al., 2010). Between the historical and present climates compared here, this effect is still small, but it might become more substantial for future climates that see even stronger soil drying.

(3) The “magnitude effect”: The TM anomalies are $> 1 \sigma$ more extreme in the unprecedented heatwaves compared to the moderate heatwaves. As expected, this is accompanied by more extreme anomalies in most relevant drivers (Z500, SHF, V500, U500, PREC, SM, and EF); only SSR and $\omega 500$ exhibit unchanged and slightly less extreme anomalies, respectively. For LHF, the median anomalies in the unprecedented heatwaves are approximately 1σ lower than during the moderate heatwaves, causing a switch from positive to negative median anomalies. This magnitude effect on LHF is noticeably greater than the combined model and climate effects, suggesting that the more negative SM anomalies are primarily responsible for the difference in LHF between historical and unprecedented heatwaves.

However, the shift from energy-limited to moisture-limited conditions also appears to have an effect among the unprecedented heatwaves themselves. The absolute values of SM are on average more than 15 % lower in the unprecedented heatwaves occurring in model year 2031 (parent ensemble P2) than in the other unprecedented heatwaves occurring in model year 2015 (parent ensemble P1, see Fig. B1c). This is presumably why their LHF anomalies are much more negative despite only slightly more negative SM anomalies (see, for example, U4 and U8 in Fig. B2b). This highlights an important limitation of our anomaly-based framework when comparing SM anomalies in different background climates, as this approach cannot capture non-linear behaviour such as changes in the regime of land-atmosphere interactions. At the same time, it also highlights the value of explicitly analysing anomalies in SHF, LHF, and EF, because these variables contain valuable information about the land-atmosphere interaction regime.

What conclusions does this analysis allow us to draw regarding the plausibility of unprecedented heatwaves? During historical, moderate, and unprecedented heatwaves, the largest anomalies consistently occur in the same physical drivers, most notably Z500, SSR, U500, and SM, and the unprecedented heatwaves feature more extreme anomalies in these drivers than the less extreme historical heatwaves (Fig. 4c). This implies that these unprecedented heatwaves are caused by “extreme anomalies of common drivers” (Fischer et al., 2021). This indicates that these heatwaves are internally consistent within CESM2 and also have external empirical support from historical precedent in ERA5. We also find that physically plausible, non-linear regime shifts in land-atmosphere interactions can limit the comparability of SM anomalies for future climate states.

3.3 Relationships between variables

The previous analysis has demonstrated that the relevant drivers of unprecedented heatwaves are similar to those during extreme heatwaves in both CESM2 and ERA5. To further assess the internal consistency and physical conformity of these heatwaves, we now study the bivariate and multivariate relationships between relevant heatwaves variables. This allows us to test whether these variables follow well-known physical principles that govern their intervariate relationships and whether anomalies in different drivers are internally consistent.

3.3.1 Bivariate relationships

We first investigate the bivariate relationships between different relevant physical variables. For this, we select the annual JJA maxima in 30-day z_{TM} in ERA5 (1965–2010, 45 heatwaves) and CESM2-LE (2005–2035, 3100 heatwaves). We then compare the bivariate distributions of the 30-day anomalies in heatwave variables during unprecedented heatwaves with those during heatwaves in ERA5 and CESM2-LE (Figs. 5 and 6). Unprecedented heatwaves for which knowledge of the bivariate distributions in ERA5 and CESM2-LE and the anomaly in a given variable provides a good estimate of the anomaly in a second variable are considered plausible from a bivariate perspective. Conversely, obvious visual outliers from the mentioned bivariate distributions can count against plausibility of a given unprecedented heatwave if they cannot be explained by the behaviour of a third variable.

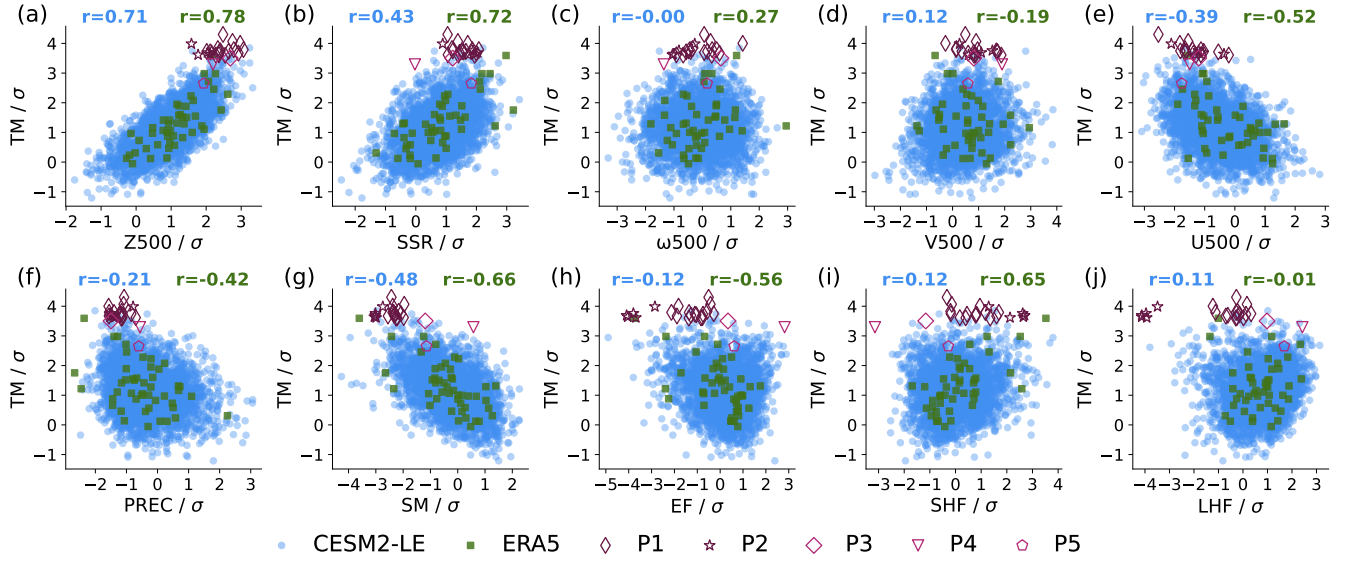


Figure 5. Bivariate distributions between TM and analysed heatwave drivers. Shown are the annual JJA maxima of 30-day temperature anomalies in CESM2-LE (2005–2035, blue circles) and ERA5 (1965–2010, green squares), as well as the 30-day anomalies during unprecedented heatwaves (purple) and preceded heatwaves (pink). The respective correlation coefficient r is noted at the top of the panel for CESM2-LE heatwaves (blue) and the ERA5 heatwaves (green).

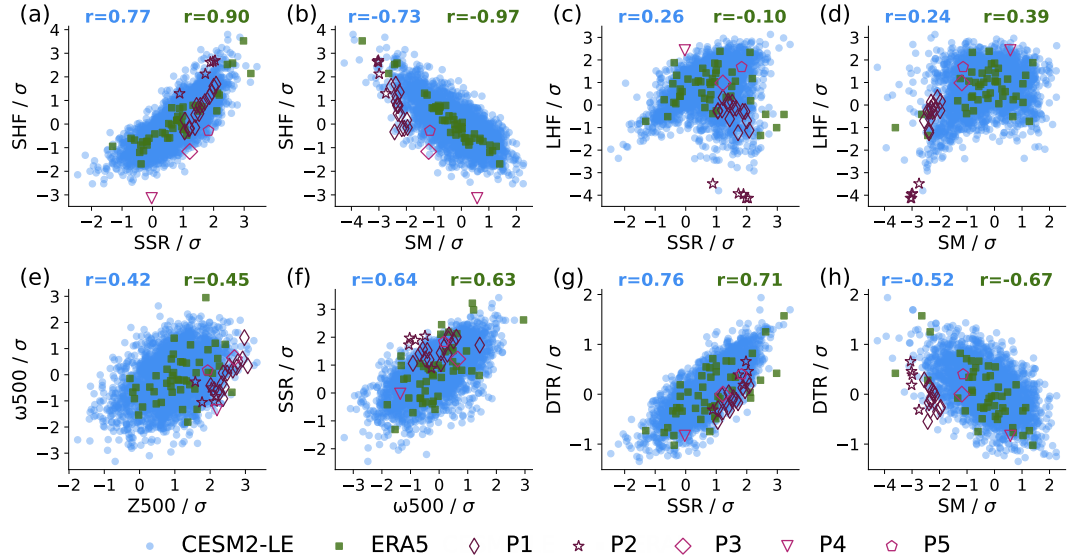


Figure 6. Same as Fig. 5, but for selected cross-relationships between different heatwave drivers and characteristics.

By definition, all unprecedented heatwaves have more extreme TM anomalies than heatwaves in ERA5 and almost all
375 heatwaves in CESM2-LE, placing them at the upper edge of all bivariate distributions involving TM (Fig. 5). The different
physical drivers can be broadly grouped into three different categories.

(1) Several drivers exhibit rather high correlations with TM, and their anomalies during preceded and unprecedented heat-
waves are generally at the extreme ends of the CESM2-LE and ERA5 distributions. This is the case for Z500 (Fig. 5a), U500
(Fig. 5e), SM (Fig. 5h) and, to a lesser extent, PREC (Fig. 5f). These correlations align with physical process understanding,
380 highlighting the importance of blocking anticyclones and drought conditions during heatwaves. Furthermore, the anomalies of
these drivers are consistent with other heatwaves in CESM2-LE, but also with historical heatwaves in ERA5.

(2) Anomalies in ω 500 (Fig. 5c) and V500 (Fig. 5d) have weak correlations with TM. For V500 this is not surprising given
that temperature advection, by definition, mostly occurs non-locally. For ω 500, this is contrary to the expectation that more
extreme heatwaves would be associated with stronger sinking motion. This is presumably again because the 30-day mean can
385 average out short periods of vertical subsidence or southerly wind (see Sect. 3.4.2). Furthermore, the regional anomalies on
a single vertical level do not capture the full effect of these processes; for example, the signal can be more pronounced on
other vertical levels (Hotz et al., 2024). Indeed, our temperature budget analysis at the 850 hPa level reveals substantial adia-
batic sinking during unprecedented heatwaves (Sect. 3.4.1). Nevertheless, the near-zero ω 500 anomalies of many preceded
and unprecedented heatwaves are quite similar to the behaviour of heatwaves in CESM2-LE and ERA5 and consistent with
390 anomalies in Z500 and SSR (Fig. 6e,f).

(3) The relationships of the surface heat flux variables with TM differ between CESM2 and ERA5. For EF and SHF,
CESM2 shows a much weaker correlation with T2M than ERA5 (Fig. 5h,i). During almost all preceded and unprecedented
heatwaves, SHF follows the strongly linear relationships with SSR and SM seen in ERA5 and CESM2-LE (Figs. 6a,b). The
only exception is P4, presumably due to the combined occurrence of low SSR anomalies and high SM anomalies, both of
395 which directly affect SHF, but whose individual contributions alone are not enough to explain the large negative SHF anomaly.
In contrast, both CESM2 and ERA5 show very weak correlations between LHF and T2M, but tend to show heteroscedastic
behaviour: Average TM anomalies feature a smaller spread in LHF than the most extreme TM anomalies, consistent with the
fact that LHF anomalies vary strongly among the unprecedented heatwaves (Fig. 5j). This behaviour is not surprising given
that surface fluxes strongly differ between moisture-limited and energy-limited conditions; during extreme conditions, this can
400 be amplified through land-atmosphere feedbacks. LHF also features strong non-linearities in its relationships with SSR and
SM in CESM2 and ERA5 (Figs. 6c,d). Most preceded and unprecedented heatwaves fall within the phase space covered by
CESM2-LE and ERA5, but heatwaves from the parent ensemble P2 (model year 2031) fall outside the range covered by ERA5
and are only similar to a single heatwave in CESM2-LE, presumably due to their lower absolute values in SM (Sect. 3.2).

The only variable that does not nicely fall into any of the three mentioned categories is SSR, which shows a substantial pos-
405 itive correlation with TM, particularly in ERA5 (Fig. 5b). The SSR anomalies during the unprecedented and most preceded
heatwaves align with the most extreme heatwaves in CESM2-LE; however, they are noticeably less extreme than expected
from the bivariate distribution in ERA5. This could be caused by differences in cloud fraction or aerosol concentration (see
Sects. 3.4.1 and 3.4.2). Nevertheless, the behaviours of other variables appear to be mostly consistent with relatively weak SSR

anomalies, most notably the anomalies in SHF (Fig. 6a), $\omega 500$ (Fig. 6f) and DTR (Fig. 6g). Finally, the small SSR anomalies
410 during unprecedented heatwaves are also presumably the reason why their DTR anomalies are smaller than expected from the
SM anomaly alone (Fig. 6h).

In conclusion, most of the bivariate relationships analysed are consistent with the bivariate distributions seen in heatwaves
in both CESM2-LE and ERA5, indicating that they are plausible from a bivariate perspective. In cases where this is not the
case, the discrepancies propagate to other variables in a physically consistent way. This primarily concerns variables related
415 to land-atmosphere interactions, which should be interpreted with caution given their substantial uncertainties in both CESM2
and ERA5/ERA5-Land. Furthermore, these discrepancies in part simply reflect changes in the land-atmosphere regime under
climate change rather than physical inconsistencies. However, given the fact that many variables depend on more than one
other variable, the bivariate approach is inherently limited in its ability to reliably detect multivariate outliers. This is therefore
studied more thoroughly below.

420 3.3.2 Multivariate relationships

We now investigate how consistent the TM anomalies during the unprecedented (and preceded) heatwaves are given the
combinations of anomalies in multiple different drivers. To this end, we employ a simple multilinear statistical model that
infers the 30-day temperature anomaly based on the 30-day anomalies in five physical drivers: Z500, U500, V500, SM, and
SSR. This selection is to some extent inherently arbitrary; it represents a compromise between model simplicity, predictor
425 independence, and explained variance (see Sect. 2.6). We use six different versions of the multilinear model, separately trained
to infer TM, TX, and TN based on heatwaves in CESM2-LE and ERA5, respectively. These different model versions are used
to infer 30-day anomalies in TM, TX, and TN during preceded and unprecedented heatwaves based on anomalies in the
five predictors (Fig. 7). The CESM2-LE trained model is primarily used to assess internal consistency, while the ERA5 trained
model evaluates whether the processes during these heatwaves are in line with those during historical heatwaves in reanalysis.

430 If both the CESM2-LE and ERA5 trained models are able to infer the temperature anomalies of unprecedented heatwaves
based only on anomalies of the five drivers, this suggests that the magnitudes of these heatwaves are explained by well-
established physical processes that behave linearly even outside the training data and across different models, supporting plau-
sibility. Conversely, non-linearities of land-atmosphere interactions discussed above could cause some discrepancies between
actual and inferred temperature anomalies. However, the model should still be able to infer a large fraction of the temperature
435 anomaly from the other drivers (and the linear component of SM), particularly given the high explained variance within the
training data. Therefore, substantial overestimations or underestimations of temperature anomalies by a model would question
the plausibility of the heatwave from a multivariate perspective. In this way, the multilinear model serves as a more quantitative
and holistic assessment of plausibility from a multivariate perspective compared to the more qualitative, bivariate assessment
in the previous section.

440 Our multilinear models capture 80–85 % of the variance in TX, 75 % in TM, but only 60–65 % of the variance in TN.
The lowest explained variance in TN is not surprising, because it is more strongly affected by processes not directly captured
in our model, such as the downwelling longwave radiation from water vapour and clouds. Although our models explain the

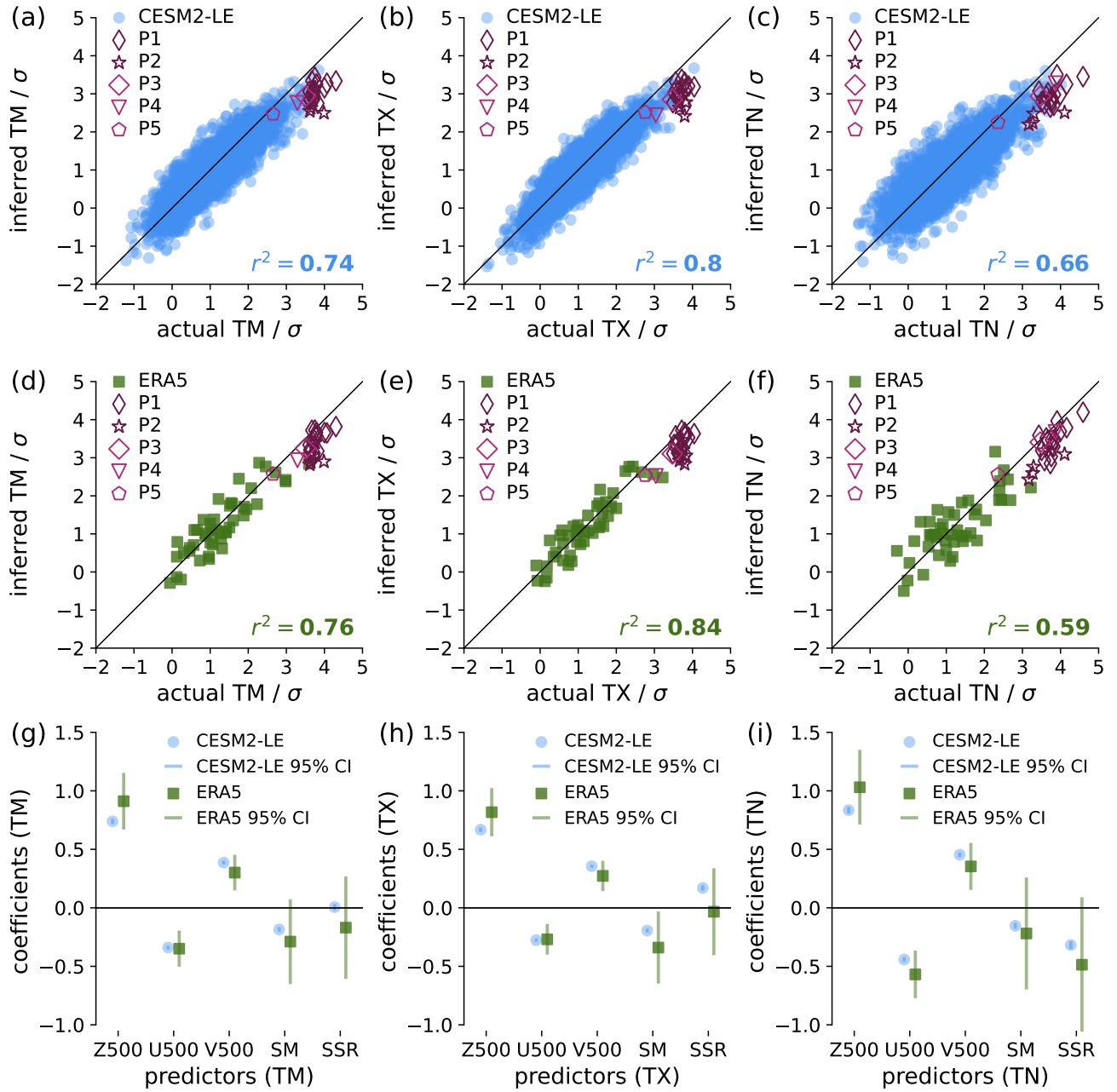


Figure 7. Actual anomalies of TM (a,d), TX (b,e), and TN (c,f) and anomalies inferred using our multilinear statistical model. Shown are the training data, namely the annual maxima of 30-day TM anomalies in WE during the summer season (JJA) in CESM2-LE (a–c, blue circles) and in ERA5 (d–f, green squares), respectively. The model is applied to 30-day anomalies of TM, TX, and TN in WE during unprecedented heatwaves (purple) and precededented heatwaves (pink). For reference, the 1:1 line and the explained variance r^2 of each model are also shown. The regression coefficients of each predictor and their 95 % confidence intervals (CI) are compared for TM (g), TX (h), and TN (i).

majority of the variance in the training data and thus perform well on average, they tend to underestimate the magnitude of the most extreme heatwaves, both for the training data and the (un)precedented heatwaves to which they are applied. This is not surprising given that (1) the models are trained to perform well in the most extreme heatwave in an average summer and are thus affected by regression to the mean, and (2) our simple least-squares approach neglects uncertainty in the independent variable and is thus sensitive to regression dilution (Pitkänen et al., 2016). This behaviour is further favoured by the strongly non-linear nature of land-atmosphere interactions and our omission of several relevant variables to achieve model simplicity. However, the main goal here is not to perfectly capture all processes affecting TM anomalies, but rather to assess whether the unprecedented heatwaves behave similarly in a multivariate phase space compared to other heatwaves in climate models and reanalysis. In Fig. 7, this corresponds to the question whether the data points of the unprecedented heatwaves behave consistently with the most extreme heatwaves in the training data.

In general, preceded and unprecedented heatwaves behave remarkably similarly to heatwaves in both CESM2-LE and ERA5. Although their magnitudes are underestimated by our models, this underestimate is similar to the underestimate of the most extreme heatwaves in the training data. For some heatwaves from parent P2, the magnitudes are underestimated more strongly than comparable heatwaves in the training data; however, the magnitude of the deviation from the 1-1 line is still comparable to the differences found for less extreme heatwaves in the training data. This underestimate is presumably related to non-linear shifts in the land-atmosphere interaction regime, as discussed in Sect. 3.2. Five of the six models include training data heatwaves of comparable magnitude to the unprecedented heatwaves, that is, the anomalies during the unprecedented heatwaves only slightly exceed the most extreme heatwave in the training data. However, the TN anomalies of many unprecedented heatwaves are far more extreme than any heatwave in ERA5 (Fig. 7f). Despite this fact, the ERA5 trained model only slightly underestimates the TN anomalies of the unprecedented heatwaves. This shows that these unprecedented heatwaves are governed by processes very similar to those of historical heatwaves.

Another way to assess plausibility from a multivariate perspective is through the model coefficients (Fig. 7g–i). In general, the differences in coefficients between the different prognostic variables exhibit physically reasonable behaviour. For example, the SSR coefficient is most positive for TX and most negative for TN, presumably due to the associated lower nighttime cloud cover. These two effects appear to largely cancel, which is why the effect of SSR on TM is much smaller than that of the other variables analysed. Furthermore, SM seems to play a larger role for TX than for TN, presumably as a result of surface heat fluxes induced by solar heating during the day. In contrast, the coefficients of the adiabatic and advective variables (Z500, U500, V500) are much stronger for TN than for TX. The coefficients for the models trained on CESM2-LE and ERA5 are generally similar and lie within their respective 95 % confidence intervals. Compared to ERA5, CESM2 has less positive Z500 coefficients, more positive V500 coefficients, less negative SM coefficients, and systematically higher SSR coefficients (more positive for TX, less negative for TM and TN). However, these differences should not be overinterpreted, given the large confidence intervals for the ERA5-trained model, caused by the small sample size and the high collinearity between SSR and SM in ERA5 (not shown).

To further investigate the role of non-linearities in land-atmosphere fluxes, we use the same model setup to predict SHF, LHF, and EF based on TM, SM, and SSR (see Sect. 2.6, Fig. B4). For SHF, the multilinear models capture 80–90 % of the

variance in the training data, a surprisingly high value given that the models do not contain information on the near-surface wind (Fig. B4a,d). The CESM2-LE trained model performs decently in inferring SHF during (un)precedented heatwaves, while the
480 ERA5 trained model systematically overestimates SHF during these heatwaves. For LHF, the multilinear model only explains 40–60 % of the variance (Fig. B4b,e). This is presumably due to missing information about near-surface wind and humidity and due to non-linearities between SM and LHF, visible from the strong deviations from the 1:1 line at large negative LHF anomalies for the CESM2-LE model in particular. These effects also affect EF, but to a somewhat lesser extent (Fig. B4c,f).

These multivariate models confirm that the surface fluxes during the precededent heatwave P4, which lie outside the re-
485 spectve bivariate distributions, are predicted reasonably well by the CESM2-LE trained model, but not by the ERA5 trained model — not surprising given different parametrisations of surface fluxes in CESM2 and ERA5. This is also apparent from the very different model coefficients of our multilinear model (Fig. B4g–i), demonstrating that, even in anomaly space, one should be careful in directly comparing them. Given the comparatively poor performance of our multilinear models in inferring surface fluxes during precededent and unprecedent heatwaves, it is remarkable that our models perform as well as they do in
490 inferring temperature anomalies during the same heatwaves. This is particularly notable for the ERA5 trained model, given the demonstrated discrepancies in the treatment of surface fluxes compared to CESM2. This demonstrates that, despite the presence of non-linearities in land-atmosphere interactions, temperatures during unprecedent heatwaves can be well explained by a linear combination of five predictors. This solidifies that it is in fact “extreme anomalies of common drivers” acting in an approximately linear way to cause these heatwaves.

495 In summary, all of our model versions can infer temperature anomalies during the unprecedent heatwaves with accuracies similar to those in the training data. This suggests that the linear combination of five drivers can explain most of the magnitude of the unprecedent heatwaves, and that, even very far in the tail, non-linearities appear to play a secondary role. Within CESM2, this is not entirely surprising, but, given substantial tuning to represent the mean climate state, it is still an important finding that internal consistency of physical processes still holds for extremes far in the tail. Remarkably, this consistency even
500 applies to model versions trained only on historical heatwaves in ERA5, showing that the multilinear model is robust across training data from different physical models. We can thus conclude that, although these unprecedent heatwaves are far more extreme than the most extreme historical heatwaves, they appear to be internally consistent and have external empirical support from a multivariate perspective.

3.4 Temporal substructure

505 The analyses above have exclusively focused on temporal averages over the entire 30-day durations of the unprecedent heatwaves. While this contains valuable information about the mean state during these heatwaves, it can average out variations on shorter time scales, and is thus inherently limited in fully assessing the relevant underlying processes. In the following, we therefore focus on the temporal substructure of unprecedent heatwaves on the daily timescale.

3.4.1 Temperature budget analysis

510 First, we perform a Eulerian temperature budget analysis at the 850 hPa level, disentangling the advective, diabatic, and adiabatic contributions to historical and recent heatwaves in ERA5 and to precededent and unprecedented heatwaves in CESM2 (Fig. 8, see Sect 2.4). In this analysis, the plausibility of unprecedented heatwaves would be confirmed by local temperature tendencies that occur due to similar physical processes, whereas it would be questioned for substantially different underlying processes. We average the tendencies separately for days with increasing local temperature (“warming days”, Fig. 8a) and decreasing local temperature (“cooling days”, Fig. 8b). Because the analysis is based on absolute values rather than anomalies, it is subject to systematic model biases. However, the local tendencies during the heatwaves analysed are generally quite similar, allowing for a comparison of the relative importance of the underlying processes.

During warming days, unprecedented heatwaves feature weaker adiabatic warming than ERA5 heatwaves. This difference is small for heatwaves from the P1 ensemble, but exceeds 50 % for the five heatwaves from the P2 ensemble. During the 520 three precededent heatwaves P3–P5, the adiabatic contribution is very similar to the ERA5 heatwaves. Conversely, there is net diabatic cooling during ERA5 heatwaves, presumably dominated by longwave radiation and evaporation, but net diabatic warming during unprecedented heatwaves. This is particularly pronounced for P2 heatwaves, consistent with their extreme anomalies in SM and SHF, while the precededent heatwaves P3–P5 feature net diabatic cooling, consistent with their much less extreme anomalies in SM and SHF (Fig. 5i). The zonal and meridional advective tendencies are smaller in magnitude, with 525 ERA5 heatwaves featuring net meridional warming rather than cooling, but also non-negligible positive residuals.

During cooling days, temperature tendencies are dominated by advective cooling, but while ERA5 heatwaves feature strong zonal cooling and weak meridional warming, unprecedented heatwaves feature weak zonal and meridional cooling. Again, precededent heatwaves are more similar to ERA5 heatwaves in this respect. Furthermore, CESM2 heatwaves feature close to zero adiabatic contributions, in contrast to substantial adiabatic cooling during ERA5 heatwaves. The diabatic tendencies are of 530 smaller magnitude than during warming days, but feature the same discrepancies regarding their signs. However, the residual terms are substantial and of opposite sign, and thus these results should be interpreted with caution.

Overall, we find that the unprecedented heatwaves in CESM2 are caused by stronger diabatic and weaker adiabatic tendencies than historical heatwaves in ERA5. This is in line with the findings of Röthlisberger et al. (2026) who found biases of CESM2 relative to ERA5 based on Lagrangian backward trajectories that are consistent with the differences found here. 535 However, the less extreme precededent heatwaves feature much more similar contributions compared with ERA5, indicating that these results depend on the exact heatwaves analysed. Nevertheless, these discrepancies in the processes that cause unprecedented heatwaves in CESM2 casts some doubt on their physical conformity. Because the discrepancies mainly concern the relative contributions of diabatic vs. adiabatic terms, some of this can compensate to still yield similar behaviour in temperature, as discussed below.

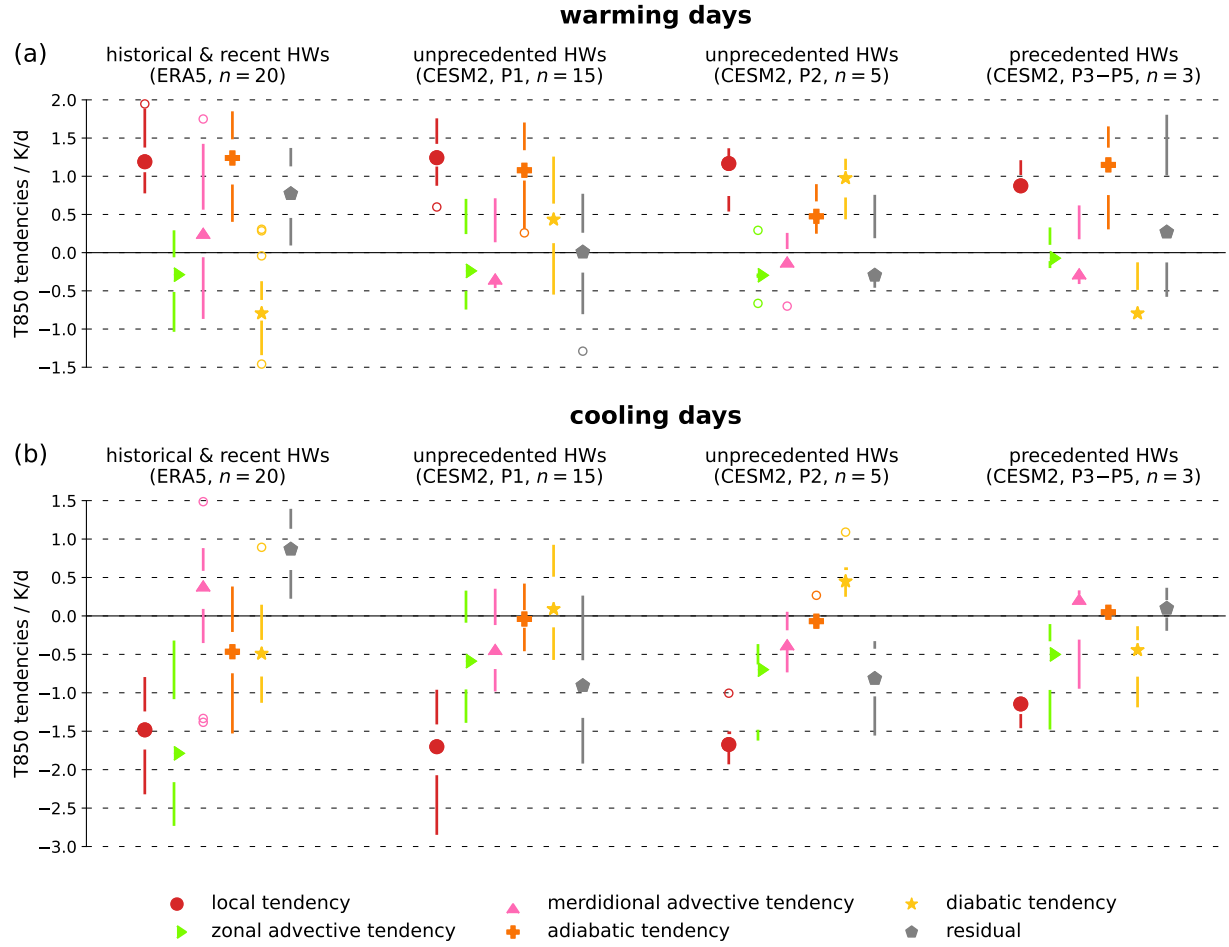


Figure 8. Tendencies of 850 hPa temperature (T850) during historical and recent heatwaves in ERA5, as well as preceded and unpreceded heatwaves in CESM2, shown separately for different parent ensembles. Shown are the terms of Eq. (2), averaged over each event during days with positive (a) and negative local T850 tendency (b): Median (filled symbols), inter-quartile range (IQR, empty space), whiskers (1.5 IQR, solid lines) and outliers (empty circles).

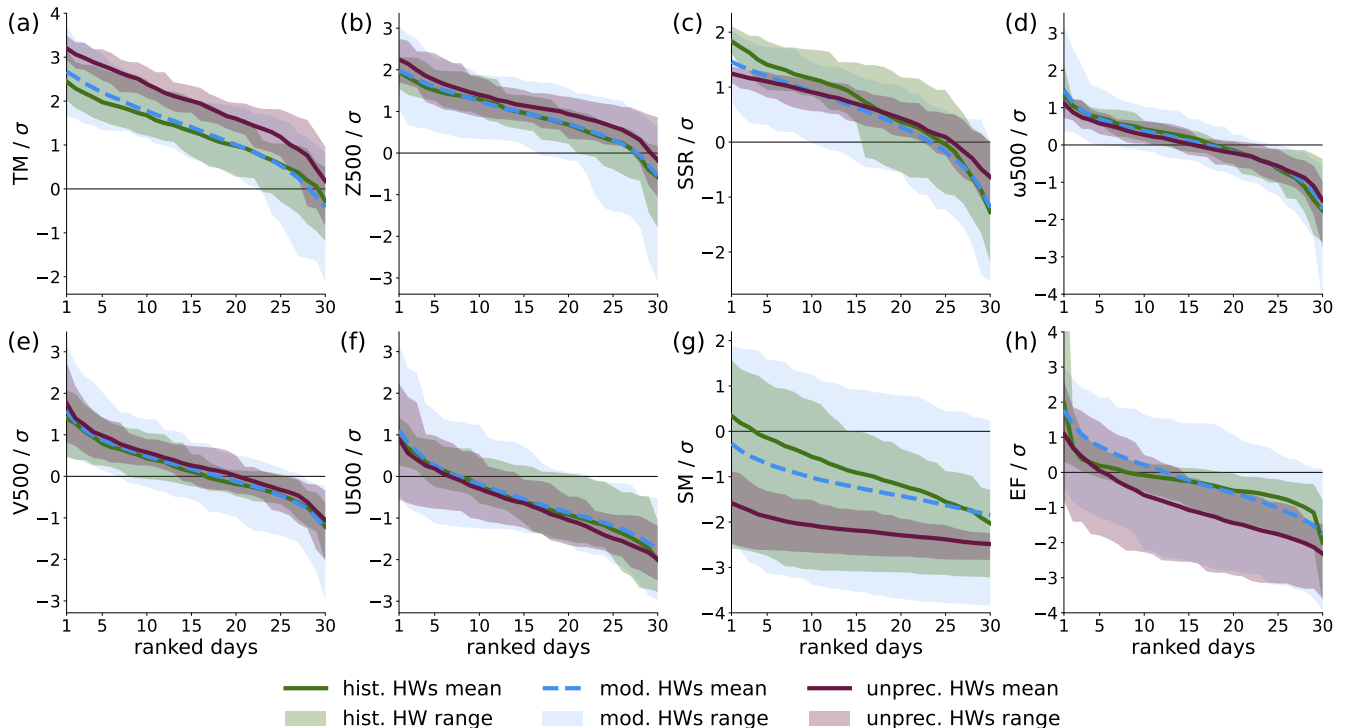


Figure 9. Ranked daily anomalies of heatwave variables. Shown are the mean (lines) and range (shading) during historical heatwaves (green), moderate heatwaves (blue), and unprecedented heatwaves (purple). Variables are sorted from highest to lowest anomalies, meaning that the most favourable conditions for heatwaves are found at low-ranked days in panels (a)–(e), but at high-ranked days in panels (f)–(h).

540 3.4.2 Rank anomalies

Next, we rank the daily anomalies of heatwave variables from highest to lowest within the 30-day historical, moderate, and unprecedented heatwaves (Fig. 9), similar to the approach of Röthlisberger et al. (2020). Because we rank the anomalies separately for each variable, the ranked days of the different variables do not correspond to each other. To be plausible, unprecedented heatwaves should feature similar relationships between the most extreme and least extreme daily anomalies as historical and moderate heatwaves. In contrast, unprecedented heatwaves dominated by only a few days of outliers would be considered less plausible. For this analysis, we only consider the model effect (ERA5 vs. CESM2) and the magnitude effect (moderate vs. unprecedented heatwaves) but not the much smaller climate effect (historical vs. present climate; see Sect. 3.2). Thus, we only include moderate heatwaves from the present climate (2005–2035). The two effects are discussed below.

(1) The model effect: The temporal substructures of most variables are very similar during historical and moderate heatwaves of the same magnitude (TM, Z500, ω 500, V500, and U500; Figs. 9,a,b,d,e,f), with three notable exceptions. (a) SSR anomalies on the sunniest days are less positive in moderate heatwaves (ranked days 1–15 in Fig. 9c). As discussed above, this is presumably related to differences in cloud fraction, indicating that even on the sunniest days, CESM2 might still simu-

late some cloud cover. It might also be affected by different aerosol concentrations, as aerosol-radiation interactions are most important on days with low cloud cover. (b) SM anomalies on less dry days are more negative in the moderate heatwaves, suggesting discrepancies in the persistence of agricultural droughts in CESM2 (ranked days 1–20 in Fig. 9g; see Sect. 3.4.3). (c) EF anomalies exhibit more strongly positive and more strongly negative daily anomalies in moderate heatwaves, while most daily anomalies during historical heatwaves are much closer to zero (Fig. 9h). This discrepancy presumably reflects differences in the parametrisation of surface heat fluxes between ERA5 and CESM2.

(2) The magnitude effect: Features (a)–(c) appear to be model specific, as they can also be detected for the unprecedented heatwaves. In addition, unprecedented heatwaves unsurprisingly feature substantially greater daily TM anomalies throughout the 30-day period than moderate heatwaves, with a weaker shift on the coolest days (ranked days 25–30 in Fig. 9a). Most other variables also exhibit very similar daily anomaly distributions between moderate and unprecedented heatwaves (Z500, SSR, ω 500, V500, and U500; Figs. 9b,c,d,e,f). For Z500, unprecedented heatwaves exhibit a shift of the ranked anomalies towards more positive values for days of weaker Z500, indicating more persistent anticyclones (ranked days 20–30 in Fig. 9b). For V500 and ω 500, the temporal substructures reveal that their small 30-day mean anomalies are partially due to the long averaging period; short-term periods of southerly wind or vertical subsidence are largely cancelled out by periods of negative anomalies as the anticyclone itself moves (Fig. 9d,e). SM anomalies are more negative during unprecedented heatwaves throughout the 30-day periods, but particularly for less dry days (ranked days 1–20 in Fig. 9g). This suggests that it is the persistence of low SM, rather than the driest SM days, that makes the unprecedented heatwaves so exceptional. Interestingly, this behaviour is not mirrored by EF, whose shift towards more negative anomalies is strongest for days with intermediate and low EF anomalies (ranked days 10–25 in Fig. 9h), which might be partially explained by the smaller SSR anomalies on sunny days.

To assess how representative these results are, we repeat this analysis for the three preceded heatwaves from three different parent ensembles (Fig. B5). Although the small sample size means that the results should not be overinterpreted, we find that most variables again feature similar daily anomaly distributions (TM, Z500, ω 500, V500, U500) and that the daily anomalies in SM and EF show the opposite behaviour of the unprecedented heatwaves, with uniform shifts to substantially less negative (or even positive) anomalies throughout their duration. This suggests that the much less dry soil is primarily responsible for the less extreme TM anomalies compared to the unprecedented heatwaves.

Overall, the temporal substructures of most variables are internally consistent within CESM2 and have external empirical support from historical precedent in ERA5. The most notable exception is the apparent overestimation of persistence in SM anomalies by CESM2, which will be analysed in more detail next.

3.4.3 Persistence

Finally, we evaluate the temporal persistence of relevant heatwaves variables, comparing the occurrence of periods in which the anomalies of TM and Z500 exceed $+1\sigma$, and in which the anomalies of SM remain below -1σ , respectively. For this, we compare preceded and unprecedented heatwaves with historical heatwaves, as well as with the JJA distributions of ERA5 and CESM2-LE (see Sect. 2.5). With this analysis, we assess the systematic differences between ERA5 and CESM2 and the position of the unprecedented heatwaves relative to the CESM2 distribution and the historical heatwaves. Findings that would

support their plausibility include: (1) ERA5 and CESM2-LE exhibiting similar distributions, (2) the unprecedented heatwaves consistently remaining within CESM2-LE distribution, and (3) the persistence of the unprecedented heatwaves being higher than that of historical heatwaves, but still of roughly comparable magnitude. If one or more of these conditions are not met, this would argue against the plausibility of these heatwaves. For the persistence of TM, a systematic overestimation by CESM2 would count against plausibility, as it would indicate that a simulated heatwave occurs only due to model biases. In contrast, a systematic underestimation of TM persistence by CESM2 would not necessarily suggest the same, but might instead indicate that the persistence of real-world heatwaves could even exceed that of unprecedented heatwaves.

For Z500, CESM2 captures the frequencies of short persistence durations (<10 days) quite well, but it substantially underestimates the frequency of longer persistence durations (Fig. 10a). The discrepancy is even more pronounced when we compare the persistence of atmospheric blocking (Fig. 10b, Sect. 2.5). CESM2 strongly underestimates the occurrence of consecutive blocked days for short persistence durations and even more strongly for long persistence durations. This systematic underestimate of persistent atmospheric blocking regimes in CESM2 also seems to affect the unprecedented and preceded heatwaves. Although they tend to exhibit longer persistence durations with respect to 1σ -threshold exceedances of Z500 anomalies than historical heatwaves, the picture is almost exactly reversed for consecutive blocked days, except for only the preceded P3 heatwave. This is consistent with previous findings that global circulation models (GCMs) struggle to reproduce persistent blocking regimes (Tibaldi and Molteni, 1990; Davini and D’Andrea, 2016; Liu et al., 2022).

The opposite phenomenon can be seen for SM, where CESM2 simulates a higher frequency of long persistence durations (>15 days). Furthermore, while preceded heatwaves feature persistence durations comparable to historical heatwaves, unprecedented heatwaves feature persistence durations up to three times longer (Fig. 10c). Presumably, this is at least partially because CESM2 simulates many more negative SM anomalies relative to ERA5-Land (Fig. B6). It is also consistent with the findings of Vogel et al. (2018) that climate models tend to overestimate soil drying due to land-atmosphere feedbacks. Furthermore, although monthly soil moisture memory is generally well represented in GCMs, there are substantial regional and intermodel differences (Seneviratne et al., 2006).

The net effect of these two processes can be seen in the persistence of TM, where CESM2 and ERA5 simulate very similar occurrences of short persistence durations (Fig. 10d). CESM2 underestimates the occurrence of longer persistence durations in TM, but the differences relative to ERA5 are smaller than for Z500 and SM. Despite this systematic underestimate, preceded and unprecedented heatwaves on average feature longer persistence durations in TM than historical heatwaves.

Overall, CESM2 struggles to accurately capture persistence durations relative to ERA5, with substantial discrepancies between the roles of diabatic contributions (related to persistence of SM) and adiabatic contributions (related to persistence of Z500), as already shown above. Thus, unprecedented heatwaves simulated by CESM2 occur for “partly the wrong reasons”, in line with previous findings (Röthlisberger et al., 2026). These limitations of CESM2 in simulating the physical processes causing long persistence durations consistently with historical precedent raise concerns regarding the physical conformity — and thus the overall plausibility — of the unprecedented heatwaves. Nevertheless, despite these discrepancies in the underlying processes, the discrepancies in the persistence of TM itself are comparatively smaller. Given that CESM2 underestimates the persistence of TM and only a few unprecedented heatwaves feature longer TM persistence durations than seen in the short

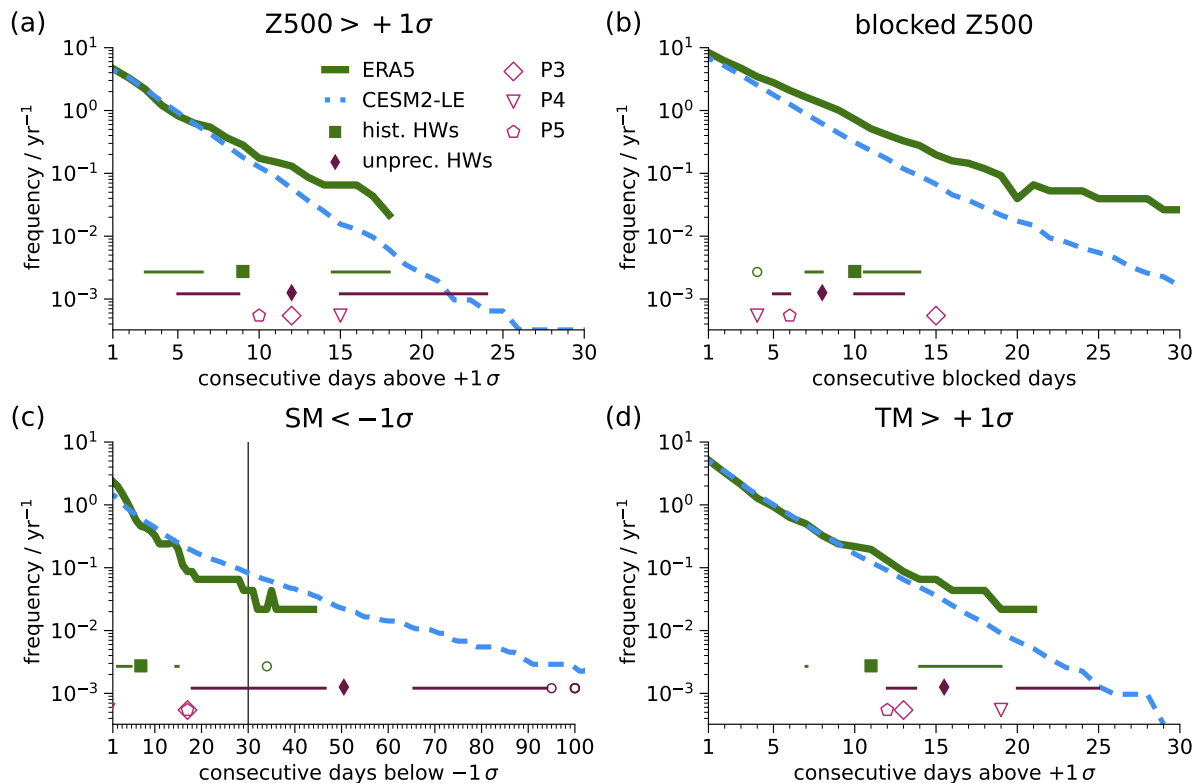


Figure 10. Persistence of selected heatwave variables. (a,c,d) Number of consecutive days in JJA with daily anomalies $\geq +1\sigma$ (TM and Z500) or $\leq -1\sigma$ (SM). Shown are the relative frequencies in ERA5 (green) and CESM2-LE (blue), and box plots of values during historical heatwaves (green) and unprecedented heatwaves (purple): Median (filled symbols), inter-quartile range (IQR, empty space), whiskers (1.5 IQR, solid lines) and outliers (empty circles). The values during the precededent heatwaves from other parent heatwaves are also shown (pink symbols). Note the different x-axis for SM in panel (c). (b) As panel (a), but showing the number of consecutive blocked days (Sect. 2.5.)

reanalysis record, this suggests that heatwaves with comparable or even longer persistence of TM are plausible to occur in the real world, although with different roles for diabatic and adiabatic processes. In fact, we argue that these heatwaves might even represent a conservative estimate for worst-case long-lasting heatwaves in terms of their persistence in TM.

625 4 Discussion

Our assessment of plausibility is mostly based on standardised anomalies, which is essential to meaningfully compare extremes occurring in different background climates. By construction, this approach removes systematic differences in mean state and variability between CESM2 and ERA5. Thus, our assessment that extreme heatwaves simulated in CESM2 are plausible to occur in the real world refers primarily to standardised anomalies rather than to the absolute physical state. This means that

630 even for plausible heatwaves, absolute values should be treated with caution, an important caveat given that impact studies are

generally based on absolute thresholds rather than anomalies. Thus, it remains essential to perform a bias correction of the raw model output when model-based storylines are used for stress testing and disaster preparedness (Lüthi et al., 2024).

Although standardised anomalies correct for many model biases in mean state and variability, some systematic differences remain. Compared to observations, GCMs tend to overestimate soil drying during heatwaves due to differences in the representation of land-atmosphere feedbacks (Vogel et al., 2018). Furthermore, the atmospheric circulation in GCMs is too zonal, that is, GCMs systematically underestimate the occurrence and persistence of atmospheric blocking, a major driver of heatwaves (Tibaldi and Molteni, 1990; Davini and D’Andrea, 2016; Liu et al., 2022). These biases can partially compensate, leading to simulated heatwaves that occur for “partly the wrong reasons” (Sects. 3.4.1 and 3.4.3; Röthlisberger et al., 2026).

Furthermore, although we often refer to the non-temperature variables analysed as heatwave “drivers”, it is not always possible to cleanly separate cause and effect. For example, soil moisture can cause temperature anomalies, but can also be amplified by them through land-atmosphere feedbacks (Seneviratne et al., 2010). Therefore, one must be careful in asserting causality, especially at a monthly time scale. For this reason, we further analyse the partitioning of latent and sensible heat flux in the evaporative fraction and focus on internal consistency rather than causal relationships.

Another caveat concerns the use of monthly regional anomalies in V500, ω 500, and SHF as proxies for temperature advection, adiabatic subsidence, and diabatic warming in some of our assessments. This neglects upstream and non-local processes, and vertical variations in the atmosphere (Hotz et al., 2024). For further studies, our approach could therefore be complemented by a more comprehensive analysis of the underlying atmospheric dynamics, for example, using Lagrangian backward trajectories or circulation-regime diagnostics (Röthlisberger and Papritz, 2023; Hochman et al., 2021). Our Eulerian approach is motivated by two main arguments: (1) It is much easier and less computationally expensive to implement and could therefore realistically be included in an operational assessment of plausibility (Kelder et al., 2022a). (2) The stationarity of the weather systems required to produce heatwaves on a monthly time scale makes the Eulerian approach much more suitable than for heatwaves or other extreme events occurring on shorter time scales.

In order to still gain more insights into the underlying processes, we perform a Eulerian temperature budget analysis, but this also comes with a few caveats: Our focus on the 850 hPa level reduces the effects of topography, but does not remove them. This leads to large residuals over mountain ranges, which is why we exclusively evaluate the temperature budget for the regional WE mean. Furthermore, focusing on the 850 hPa level presumably underestimates the diabatic contributions of land-atmosphere exchanges, which are strongest close to the surface. Furthermore, the systematically opposite signs of both the diabatic contribution and the residuals between CESM2 and ERA5 might reflect differences in the parametrisations underlying the diabatic tendencies. Finally, part of the non-negligible residual terms are presumably explained by our use of daily time steps, whereas the spatial resolution of the model output seems to play only a secondary role for the regional WE mean budget.

Evaluating historical precedent for heatwaves in a rapidly warming climate is inherently challenging, but is made possible by our approach of considering standardised anomalies relative to a time-evolving climatology. While useful, this view can in principle obscure the fact that the same anomaly might be plausible in past climates, but not necessarily in future climates. This is particularly true for non-negative variables that can exhibit substantial non-Gaussianity, such as PREC or SM, particularly when combined with a decreasing trend over time. But it can also complicate the interpretation of negative anomalies in U500,

as it is not obvious whether they imply weaker-than-average westerlies or easterly winds. The time-evolving climatology itself is also subject to uncertainties; for example, a 30-year climatology is difficult to precisely constrain in regions with unusually warm or cold periods caused by internal variability. Nevertheless, it offers clear improvements in assessing the extremeness of an event relative to the climate in which it occurred compared to a climatology based on a fixed time period.

670 Another limitation concerns the inherently limited sample size of historical heatwaves in the reanalysis record. Some of this might be overcome by choosing a longer data set for reanalysis, such as the Twentieth Century Reanalysis (Compo et al., 2011). However, by their very nature, unprecedented heatwaves do not have precedents of comparable magnitude in reanalysis data sets, regardless of the covered time period. Furthermore, due to much sparser observations in the first half of the twentieth century, such reanalyses are much less constrained by observations and instead rely much more heavily on the underlying
675 physical model, leading to additional uncertainties and potential biases.

For related reasons, it is important to be careful when comparing surface energy variables between CESM2 and ERA5. Although temperature and circulation variables in reanalysis are strongly constrained by observations, this constraint is much weaker for variables related to the surface energy budget, such as SHF, LHF, and SM. Instead, their values are strongly affected by the atmospheric and land surface models and parametrisations of land-atmosphere interactions used; furthermore, ERA5
680 fails to close the surface energy budget (Hersbach et al., 2020; Muñoz-Sabater et al., 2021). Therefore, reanalysis values of these variables should not be considered as “ground truth” but instead be interpreted cautiously. Our anomaly-based approach is designed to correct for many of these discrepancies, including biases in the mean and variance of surface variables, but not for differences in higher-order moments of the distributions.

We demonstrate the feasibility of our process-based plausibility assessment approach for heatwaves, because these are the
685 extreme events with the most solid understanding of the underlying physical processes. In principle, our approach can also be applied to other extreme events, but it should be noted that the physical drivers of these events are less well understood and in some cases less straight-forward to implement in our framework. For example, it should be quite straight-forward to apply our approach to large-scale events that can be realistically simulated in general circulation models, such as cold spells or droughts. For smaller-scale events, such as heavy convective precipitation or wind storms, high-resolution modelling is required.

690 Finally, all unprecedented heatwaves analysed in this study are based on a single model. Given well-documented intermodel differences among GCMs, this inherently limits the generalisability of our results. Our approach presented here can serve as a blueprint for performing similar analyses with other models, which would be very valuable to allow conclusions to be made less dependent on the specific climate model. In particular, models that have already been used to create climate storylines should be evaluated in this way, which could provide a better picture of how plausible extreme heatwaves they simulate are.

695 5 Conclusions

We present a process-based approach to assess the plausibility of unprecedented climate storylines and apply it to month-long heatwaves in Western Europe that would exceed the most extreme historical heatwaves by more than 5 K. We compare 20

unprecedented month-long heatwaves in Western Europe from CESM2 ensemble boosting simulations with the most extreme historical heatwaves in ERA5 using standardised anomalies relative to a time-evolving climatology.

700 The unprecedented heatwaves are associated with physical drivers well-known from less extreme heatwaves in climate models and reanalysis, including strong anticyclones, low levels of soil moisture, and high values of surface solar radiation, demonstrating the physical conformity of these heatwaves. Compared to historical heatwaves, the anomalies in these drivers are unusually extreme, leading to the conclusion that these unprecedented heatwaves are caused by “extreme anomalies of common drivers” (Fischer et al., 2021). Furthermore, the studied physical drivers are very similar in their spatial patterns, and
705 in their relationships with each other, demonstrating that they are internally consistent.

However, CESM2 systematically underestimates adiabatic contributions, including persistent blocking anticyclones, and overestimates diabatic contributions, primarily land-atmosphere interactions, and it is similarly biased in their respective temporal persistence. This indicates that unprecedented heatwaves occur for “partly the wrong reasons”, a known weakness of climate models (Röthlisberger et al., 2026; Vogel et al., 2018). These discrepancies raise some doubts about the physical
710 conformity of CESM2 to simulate unprecedented long-lasting heatwaves. However, the fact that CESM2 systematically underestimates the frequency of long temperature persistence suggests that, if anything, the simulated temperature persistence during these unprecedented heatwaves might represent a conservative estimate of what worst-case heatwaves in the real world might look like, which is an important consideration for disaster preparedness. We therefore conclude that such unprecedented heatwaves should not be ruled out as implausible, but rather should be included in risk assessments when preparing adaptation
715 measures, ideally using bias-corrected model output.

There are many ways to assess the plausibility of extreme weather and climate events. Our process-based approach complements existing methods and builds on understanding the underlying physical processes. Future studies could expand on our approach by using Lagrangian methods to study the atmospheric dynamics during unprecedented heatwaves. Furthermore, our methodology would also have to be adjusted before it can be applied to other types of extreme events. Nevertheless, our
720 approach contains several important ingredients that can provide powerful insights into the plausibility of different extreme events. This includes the use of a time-evolving climatology to meaningfully compare anomalies across different background climates, as well as a multivariate perspective that provides a holistic picture of the underlying dynamics of unprecedented events.

Code and data availability. The code used to perform the analysis is available at <https://doi.org/10.5281/zenodo.19492889> (Roemer et al.,
725 2026a). Pre-processed data to reproduce the figures is available at <https://doi.org/10.5281/zenodo.19494785> (Roemer et al., 2026b). CESM2-LE data are publicly available from the NSF NCAR Geoscience Data Exchange at <https://gdex.ucar.edu/datasets/d651056/> (Danabasoglu et al., 2020; Rodgers et al., 2021). Reanalysis data is publicly available from the Copernicus Climate Change Service Climate Data Store: ERA5 data on pressure levels at <https://doi.org/10.24381/cds.bd0915c6> (Hersbach et al., 2023a), ERA5 data on single levels at <https://doi.org/10.24381/cds.adbb2d47> (Hersbach et al., 2023b), and ERA5-Land data at <https://doi.org/10.24381/cds.e9c9c792> (Muñoz-Sabater, 2019).

730 **Appendix A: Ensemble boosting simulations**

Ensemble boosting is a computationally efficient method to simulate record-breaking extremes in the CESM2 climate model (Gessner et al., 2021). It consists of re-initialising a model run a few weeks before an extreme event and creating new ensembles based on round-off sized perturbations in the humidity field. The simulations used here are based on a 35-member CESM2 ensemble running from 2005 to 2035 under the historical (2005–2014) and SSP3-7.0 (2015–2035) scenarios (Fischer et al., 2023; Lüthi et al., 2024). From this ensemble, the five most extreme 14-day heat extremes in Switzerland (“parents”) were selected, occurring in the model years 2015, 2028, 2028, 2030, and 2031 in different ensemble members. For each parent, 100 ensemble members per day start 14–25 days before the peak of the parent event and subsequently run for 50 days, yielding 1100 alternative realisations of each parent heatwave. These simulations were already used and described by Lüthi et al. (2024).

To estimate the return periods of the unprecedented heatwaves from the ensemble boosting simulations, we use a method that was specifically developed for this purpose and is described in detail by Bloin-Wibe et al. (2025). In short, this method uses the shared antecedent conditions between a boosted simulation and its parents to derive conditional probabilities of exceeding a threshold TM_{ref} and uses bootstrapping to derive the 95 % confidence interval of the estimated return period. It provides an unbiased estimate of return periods and offers clear improvements over approaches based on fitting extreme value distributions.

When applying this method to our simulations, it should be noted that the parent heatwaves were selected based on the 14-day absolute TM in Switzerland, while our study considers the 30-day TM anomaly in WE. As a consequence, the five selected parent heatwaves are not among the most extreme heatwaves in the ensemble; rather, they are ranked 16th, 23rd, 133rd, 265th, and 294th in terms of their 30-day TM anomaly in WE (out of the 35 ensemble members · 31 ensemble years = 1085 simulated years; see Fig. 1b). Thus, setting TM_{ref} equal to the least extreme parent run, the antecedent conditions are sampled less efficiently than is theoretically possible, but, importantly, this does not affect the quality of the return period estimator. Return periods of heatwaves in the 35-member ensemble (including parent heatwaves) are estimated using an empirical frequency estimator based on the occurrence within the ensemble (Bloin-Wibe et al., 2025).

The ten selected unprecedented heatwaves are from ensemble members that are re-initialised 11–20 days before the peak of the 30-day TM anomaly in WE in the parent run and thus we use all ensemble members with lead times of 11–20 days to estimate conditional probabilities. Note that these lead times do not correspond directly to the lead times defined with respect to the 14-day TM in Switzerland in Sect. 2.1. All days of the unprecedented 30-day heatwaves occur after the re-initialisation and thus do not include any simulated days from the respective parent simulation.

Appendix B: Selected heatwaves

This appendix provides more detailed information on precededented and unprecedented heatwaves (Tab. B1), historical heatwaves (Tab. B2) and recent heatwaves analysed (Tab. B3). For unprecedented heatwaves, further details are given on daily time series of TM Z500 and SM (Fig. B1). Furthermore, 30-day anomalies of heatwave drivers during individual unprecedented, precededented, and historical heatwaves are shown (Fig. B2). We also include the temporal substructure of precededented heatwaves

(Fig. B5), climatological distributions of SM in ERA5-Land and CESM2-LE (Fig. B6), and the results of our multilinear model for SHF, LHF, and EF (Fig. B4).

Furthermore, we compare unprecedented heatwaves with recent heatwaves in terms of 30-day mean absolute values and daily time series of relevant heatwave variables (Fig. B3). On average, unprecedented heatwaves are around 5 K warmer than recent heatwaves, but the 2026 heatwave exhibits an almost as high 30-day TX as some unprecedented heatwaves (Fig. B3a-d). On average, the 500 hPa geopotential height is also substantially higher during unprecedented heatwaves, while the surface solar radiation (SSR) is very similar but shows substantial variability (Fig. B3e-h). Furthermore, unprecedented heatwaves occur under drier conditions with less evaporative cooling of the surface (Fig. B3i-l). Overall, although the unprecedented heatwaves are much more extreme than recent heatwaves, several variables are similar to the 2026 heatwave.

Comparing the time series of the 2026 heatwave with selected unprecedented heatwaves, we find that during the U1 heatwave, TX is strikingly similar to that of 2026 in the first and last 10 days of the respective 30-day periods, only differing noticeably between (Fig. B3a). The 2026 heatwave features systematically lower values in TN and Z500 and slightly higher values in SSR compared to the unprecedented heatwaves (Fig. B3c,e,g). The 2026 heatwave starts out with a much higher SM in the beginning, but dries out strongly and exhibits SM values similar to the unprecedented heatwaves at the end of the 30-day period (Fig. B3i).

Author contributions. FR, EF, and RK conceived the idea for this study. FR carried out the analysis and drafted the manuscript. EF and RN helped with developing and refining the methodology. All authors contributed to discussing the results and revising the manuscript.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Weather and Climate Dynamics.

Acknowledgements. We thank U. Beyerle for technical assistance with the running and data acquisition of CESM2. We also thank L. Pitz-Bloin for help in calculating return periods, G. Gyuleva for useful input on the statistical model, A. Prein for help with acquiring additional CESM2 data, and P. Stott for encouraging feedback on an early version of this work. RN has received funding from the Swiss National Science Foundation (SNSF) through grant No. 216710. RK is part of SPEED2ZERO, a Joint Initiative co-financed by the ETH Board. This research was supported by the NCCR CLIM+, funded by the Swiss National Science Foundation (grant 51AU-0_229306).

Table B1. Unprecedented (U1–U20) and precededented 30-day heatwaves in WE. Listed are the parent heatwaves, the date of re-initialisation, the selected ensemble member, the beginning and end of the 30-day period, and the 30-day TM anomaly.

heatwave	parent	re-initialisation	member	day 1	day 30	TM / σ
U1	P1	2015-07-09	3	2015-07-25	2015-08-23	4.3
U2	P1	2015-07-11	37	2015-07-27	2015-08-25	4.1
U3	P1	2015-07-16	7	2015-07-30	2015-08-28	4.0
U4	P2	2031-07-12	29	2031-07-23	2031-08-21	4.0
U5	P1	2015-07-17	50	2015-07-26	2015-08-24	3.8
U6	P1	2015-07-12	62	2015-07-30	2015-08-28	3.8
U7	P1	2015-07-17	63	2015-07-23	2015-08-21	3.8
U8	P2	2031-07-13	76	2031-07-31	2031-08-29	3.8
U9	P1	2015-07-16	47	2015-07-22	2015-08-20	3.7
U10	P1	2015-07-13	6	2015-07-24	2015-08-22	3.7
U11	P1	2015-07-10	56	2015-07-21	2015-08-19	3.7
U12	P1	2015-07-11	92	2015-07-27	2015-08-25	3.7
U13	P1	2015-07-10	74	2015-07-25	2015-08-23	3.7
U14	P2	2031-07-18	83	2031-07-31	2031-08-29	3.7
U15	P1	2015-07-11	51	2015-07-22	2015-08-20	3.7
U16	P2	2031-07-18	33	2031-07-30	2031-08-28	3.6
U17	P1	2015-07-15	47	2015-07-29	2015-08-27	3.6
U18	P1	2031-07-18	40	2031-07-30	2031-08-28	3.6
U19	P1	2015-07-13	53	2015-07-27	2015-08-25	3.6
U20	P1	2015-07-12	19	2015-07-25	2015-08-23	3.6
	P3	2030-08-06	75	2030-08-20	2030-09-18	3.5
	P4	2028-07-17	78	2028-07-05	2028-08-03	3.3
	P5	2028-07-18	100	2028-07-31	2028-08-29	2.7

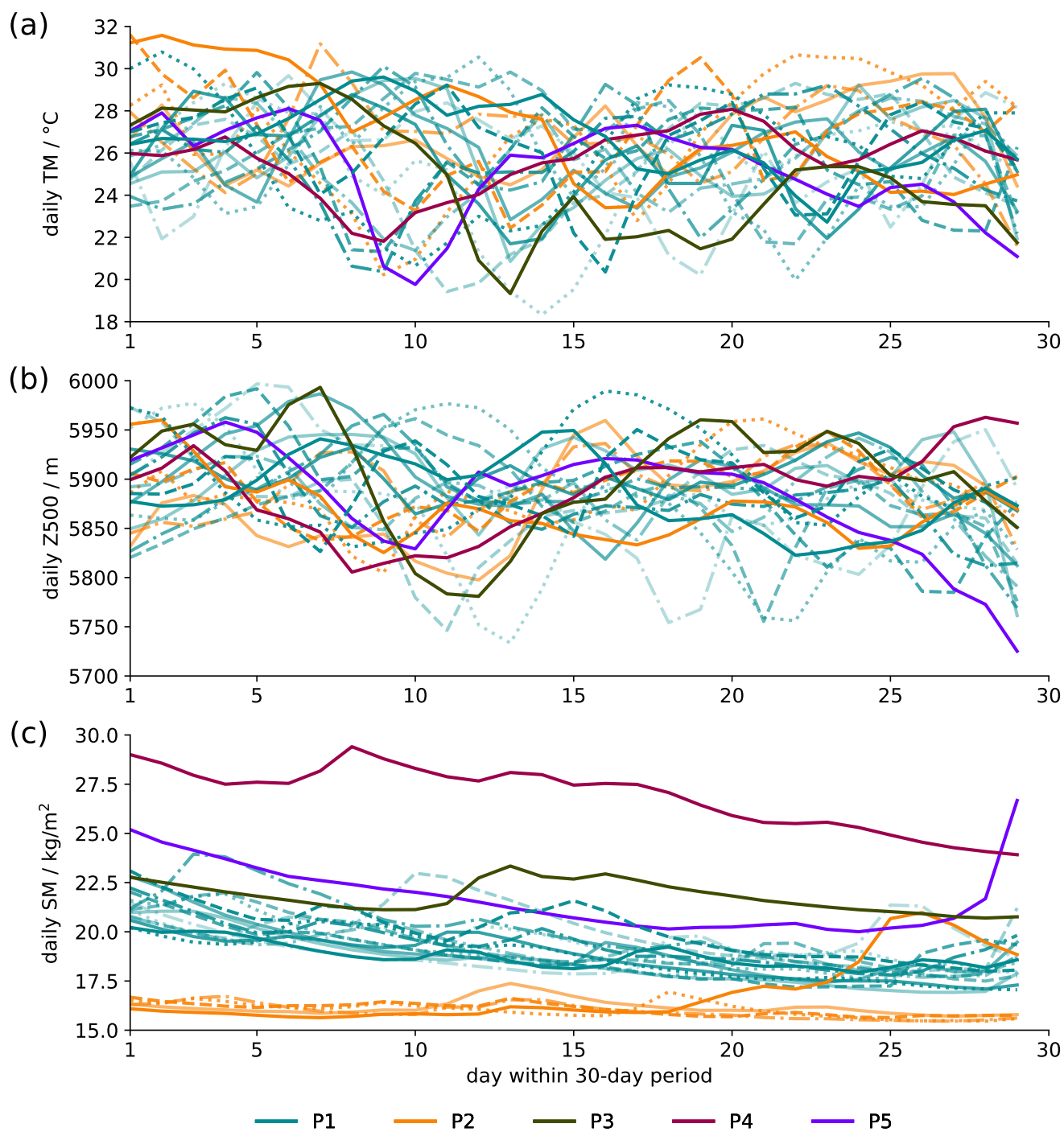


Figure B1. Daily values of (a) TM, (b) Z500, (c) SM during the unprecedented heatwaves (P1, P2) and preceded heatwaves (P3–P5).

Table B2. Historical 30-day heatwaves in WE selected from ERA5 (beginning and end of the selected 30-day period), sorted in descending order by their 30-day mean TM anomaly.

day 1	day 30	TM / σ
1976-06-07	1976-07-06	3.6
2003-08-01	2003-08-30	3.0
1973-08-11	1973-09-09	3.0
2003-05-29	2003-06-27	3.0
1983-07-03	1983-08-01	2.7
2006-07-01	2006-07-30	2.5
1969-07-16	1969-08-14	2.3
1997-08-08	1997-09-06	2.2
1994-07-11	1994-08-09	2.1
1975-07-30	1975-08-28	2.0
1995-07-07	1995-08-05	1.9

Table B3. Recent 30-day heatwaves in WE selected from ERA5 (beginning and end of the selected 30-day period) in chronological order, with their absolute 30-day mean TM.

day 1	day 30	TM / °C
2013-07-07	2013-08-05	20.9
2015-06-25	2015-07-24	21.2
2018-07-12	2018-08-10	21.9
2019-06-26	2019-07-25	20.9
2020-07-23	2020-08-21	21.4
2022-07-16	2022-08-14	21.6
2023-08-13	2023-09-11	20.3
2024-07-18	2024-08-16	20.8
2025-06-12	2025-07-11	20.6
2026-06-17	2026-07-16	24.0

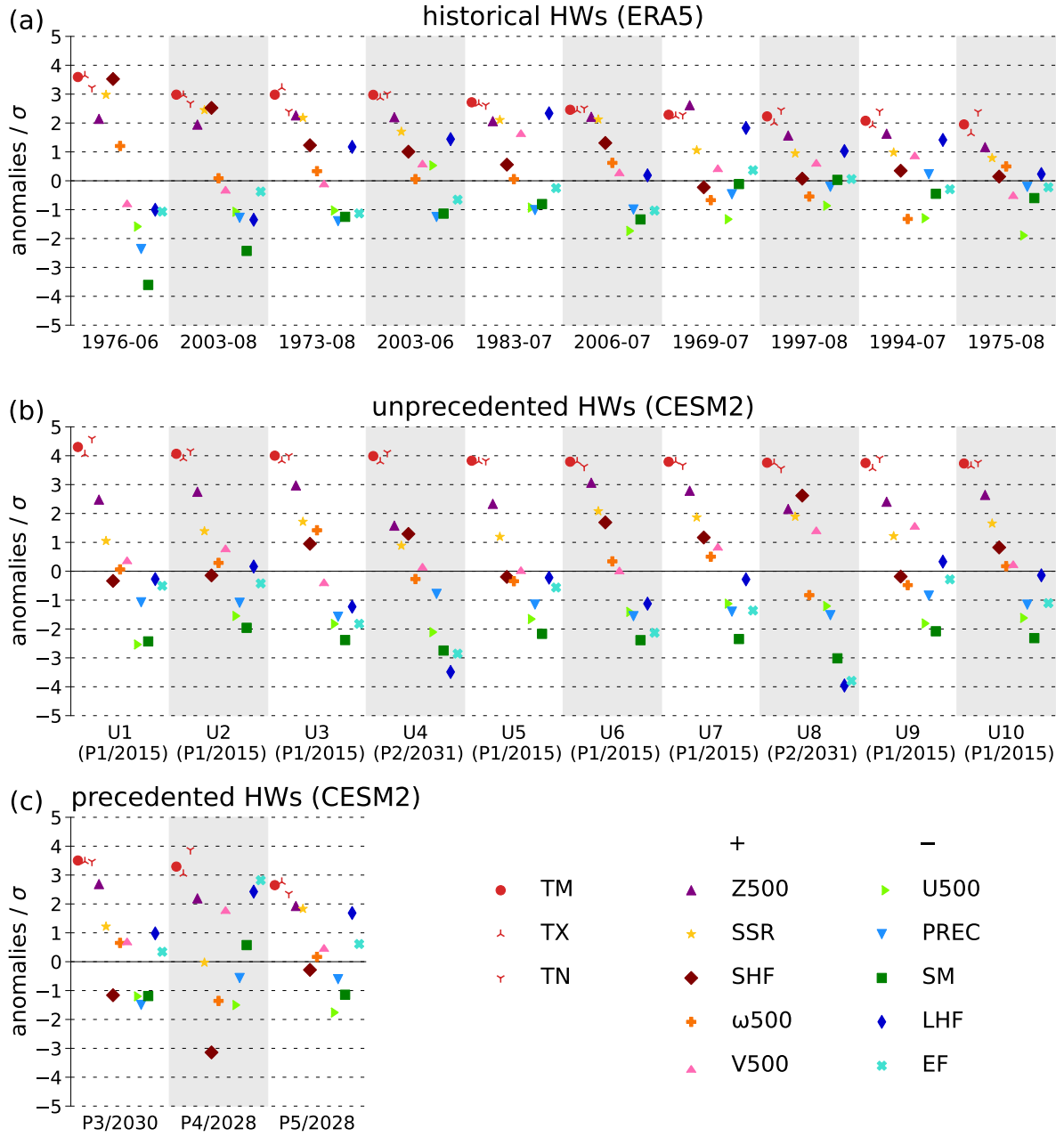


Figure B2. Anomalies of heatwave characteristics and drivers in WE of selected heatwaves (Tab. 1). (a) Ten most extreme historical month-long heatwaves in ERA5 (Tab. B2) (b) Ten most extreme unprecedented month-long heatwaves in CESM2 ensemble boosting simulations (Tab. B1). (c) The most extreme heatwaves from the parent ensembles P3–P5 (Tab. B1). Shown are the temperature variables (red), and variables with expected positive (“+”) and negative (“–”) anomalies during heatwaves. This separation is a simplification primarily adopted for ease of visualization; several variables feature substantial anomalies of both signs during heatwaves.

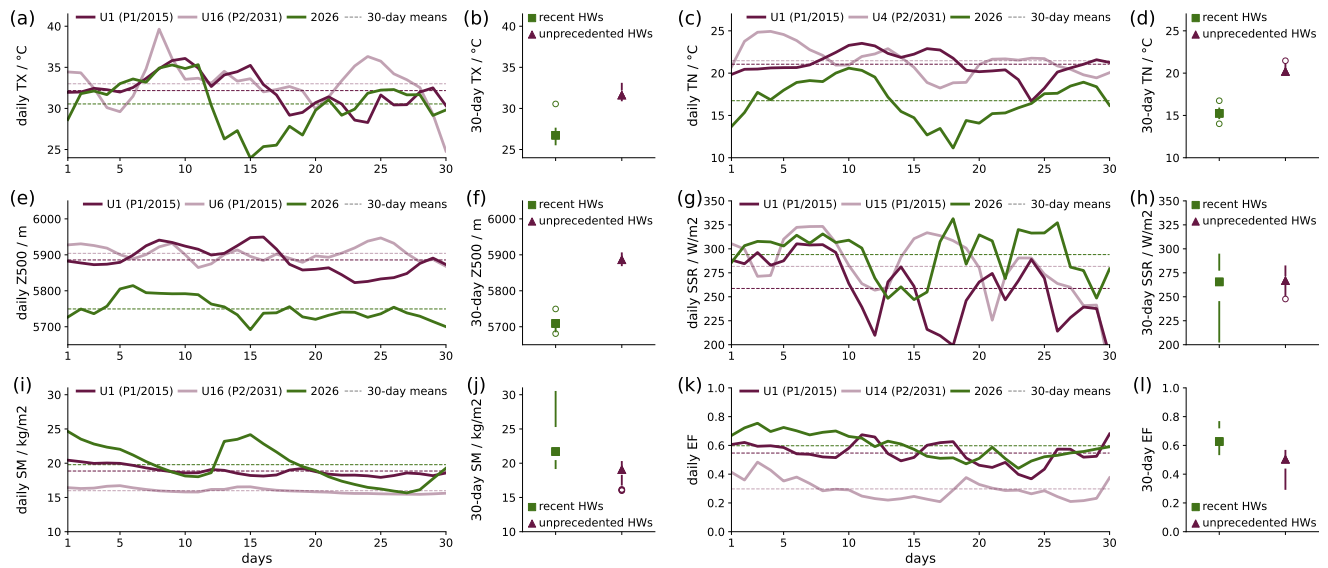


Figure B3. Absolute values in selected variables of unprecedented heatwaves and recent heatwaves. Time series for the 2026 heatwave (green) and for two unprecedented heatwaves: the most extreme in terms of TM anomaly (U1), and the most extreme in terms of absolute value of the respective variable (first and third column). 30-day means of unprecedented and recent heatwaves: Median (filled symbols), inter-quartile range (IQR, empty space), whiskers (1.5 IQR, solid lines) and outliers (empty circles) (second and fourth column).

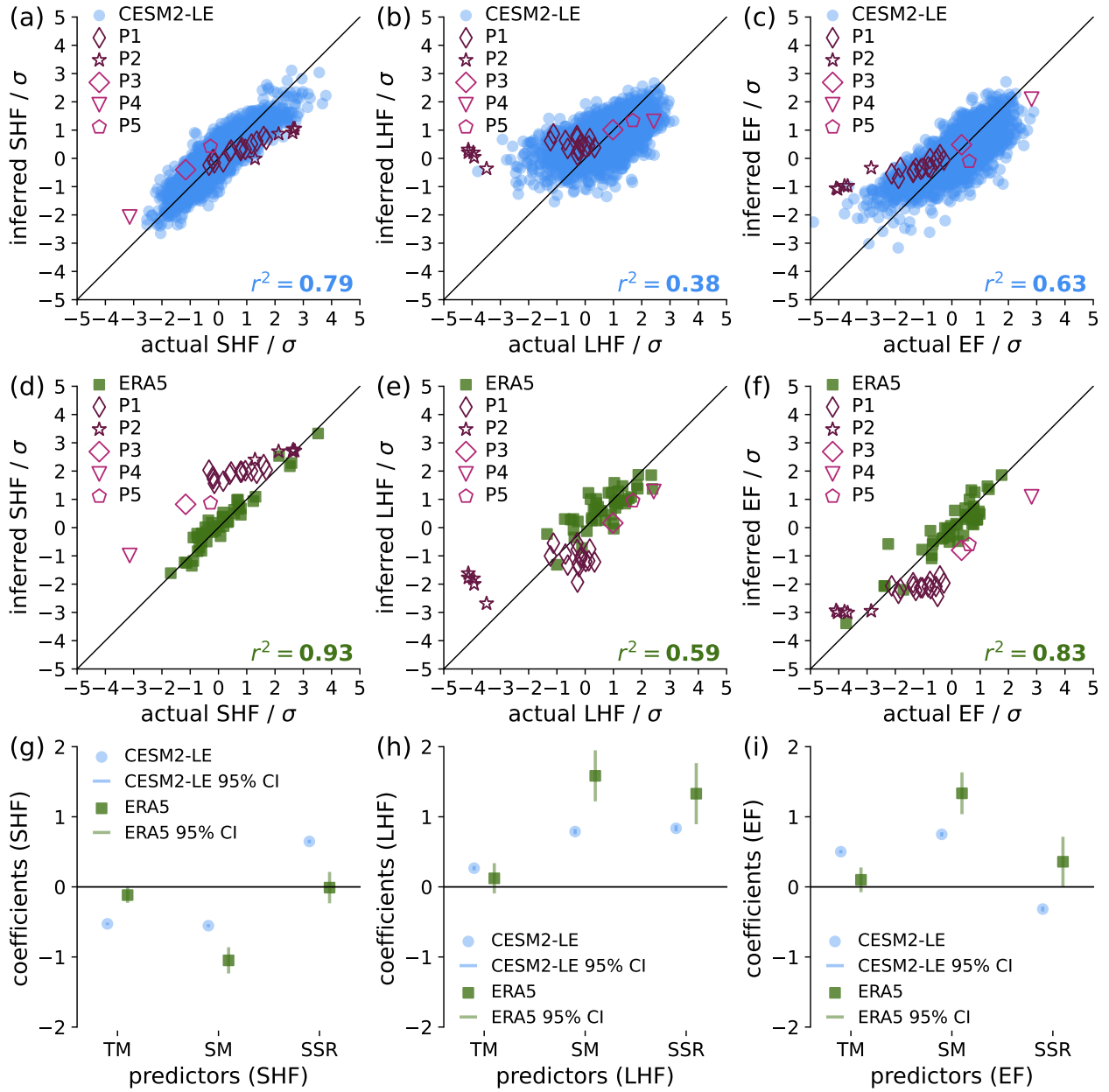


Figure B4. Actual anomalies of SHF (a,d), LHF (b,e), and EF (c,f) and anomalies inferred using our multilinear statistical model. Shown are the training data, namely the annual maxima of 30-day TM anomalies in WE during the summer season (JJA) in CESM2-LE (a–c, blue circles) and in ERA5 (d–f, green squares), respectively. The model is applied to 30-day anomalies of SHF, LHF, and EF in WE during the unprecedented heatwaves (purple) and three less extreme heatwaves from CESM2 boosting simulations (pink). For reference, the 1:1 line and the explained variance r^2 of each model are also shown. The regression coefficients of each predictor and their 95 % confidence intervals (CI) are compared for LHF (g), SHF (h), and EF (i).

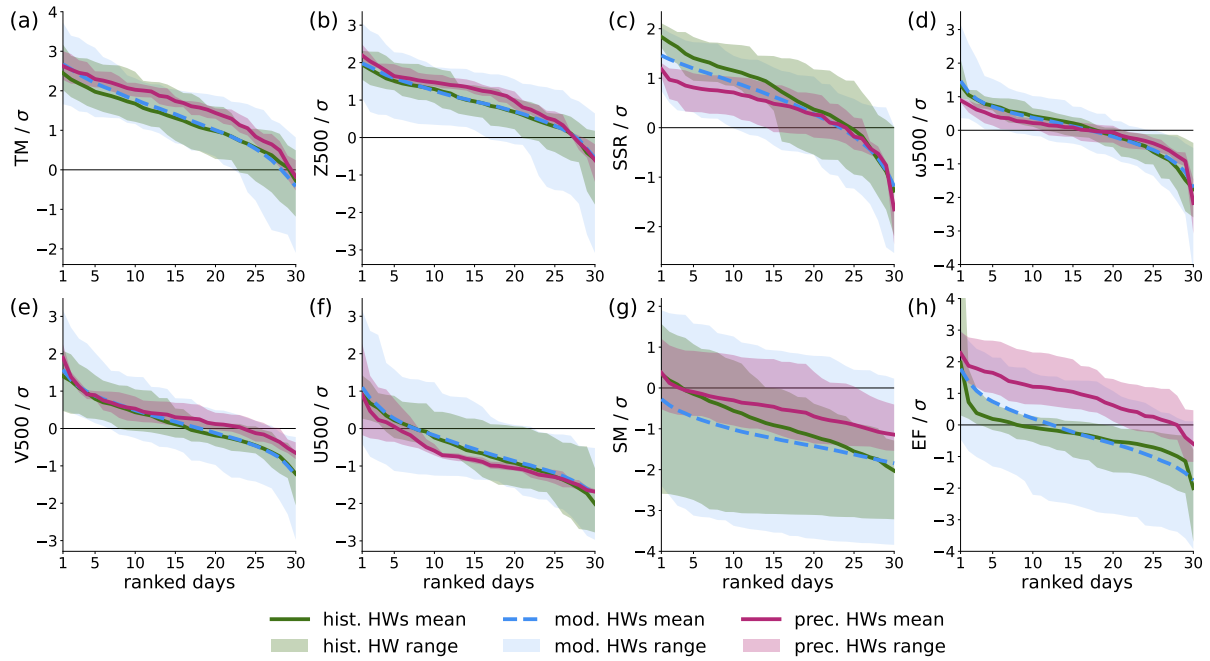


Figure B5. Ranked daily anomalies of heatwave characteristics and drivers during historical heatwaves (green), moderate heatwaves (blue), and preceded heatwaves from CESM2 boosting simulations (pin). Variables are sorted from highest to lowest anomalies, meaning that the most favourable conditions for heatwaves can be found at low ranked days panels (a)–(e), but at high ranked days for panels (f)–(h).

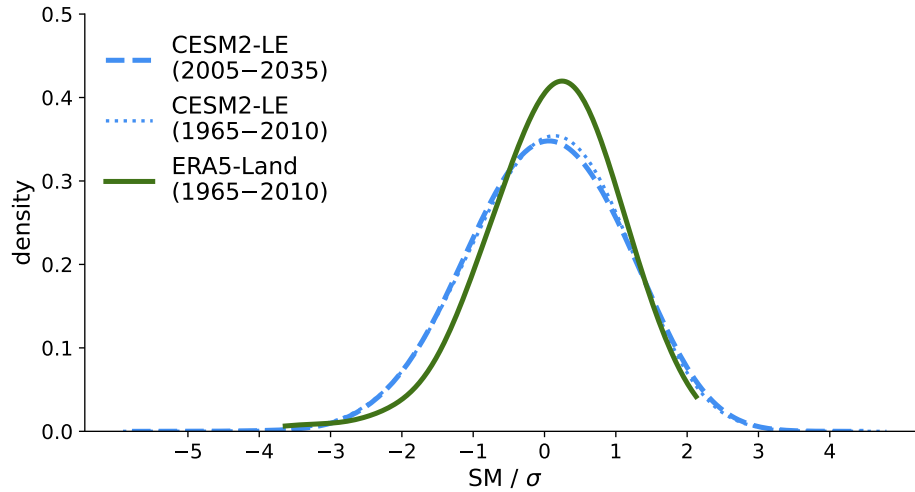


Figure B6. Kernel density estimates of 30-day SM anomalies during the summer season (JJA) in ERA5-Land (1965–2010, green) and in CESM2-LE (1965–2010 and 2005–2035, blue), using a Gaussian kernel with a bandwidth of 0.5σ , truncated at the data limits.

- Amara, R.: Views on futures research methodology, *Futures*, 23, 645–649, [https://doi.org/10.1016/0016-3287\(91\)90085-G](https://doi.org/10.1016/0016-3287(91)90085-G), 1991.
- Anderson, G. B. and Bell, M. L.: Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities, *Environmental Health Perspectives*, 119, 210–218, <https://doi.org/10.1289/ehp.1002313>, 2011.
- Bador, M., Terray, L., Boé, J., Somot, S., Alias, A., Gibelin, A.-L., and Dubuisson, B.: Future summer mega-heatwave and record-breaking temperatures in a warmer France climate, *Environmental Research Letters*, 12, 074 025, <https://doi.org/10.1088/1748-9326/aa751c>, 2017.
- Baldissera Pacchetti, M., Coulter, L., Dessai, S., Shepherd, T. G., Sillmann, J., and Van Den Hurk, B.: Varieties of approaches to constructing physical climate storylines: A review, *WIREs Climate Change*, 15, e869, <https://doi.org/10.1002/wcc.869>, 2024.
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, *Science*, 332, 220–224, <https://doi.org/10.1126/science.1201224>, 2011.
- Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D. G., and Salcedo-Sanz, S.: Heat Waves: Physical Understanding and Scientific Challenges, *Reviews of Geophysics*, 61, e2022RG000 780, <https://doi.org/10.1029/2022RG000780>, 2023.
- Bartusek, S., Kornhuber, K., and Ting, M.: 2021 North American heatwave amplified by climate change-driven nonlinear interactions, *Nature Climate Change*, 12, 1143–1150, <https://doi.org/10.1038/s41558-022-01520-4>, 2022.
- Baulenas, E., Versteeg, G., Terrado, M., Mindlin, J., and Bojovic, D.: Assembling the climate story: use of storyline approaches in climate-related science, *Global Challenges*, 7, 2200 183, <https://doi.org/10.1002/gch2.202200183>, 2023.
- Black, E., Blackburn, M., Harrison, G., Hoskins, B., and Methven, J.: Factors contributing to the summer 2003 European heatwave, *Weather*, 59, 217–223, <https://doi.org/10.1256/wea.74.04>, 2004.
- Bloin-Wibe, L., Noyelle, R., Humphrey, V., Beyerle, U., Knutti, R., and Fischer, E.: Estimating return periods for extreme events in climate models through Ensemble Boosting, *Weather and Climate Dynamics*, 6, 1147–1177, <https://doi.org/10.5194/wcd-6-1147-2025>, 2025.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, O., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Quarterly Journal of the Royal Meteorological Society*, 137, 1–28, <https://doi.org/10.1002/qj.776>, 2011.
- Copernicus Climate Change Service: Heatwaves contribute to the warmest June on record in western Europe | Copernicus, <https://climate.copernicus.eu/heatwaves-contribute-warmest-june-record-western-europe>, 2025.
- Danabasoglu, G., Deser, C., Rodgers, K., and Timmermann, A.: CESM2 Large Ensemble, <https://doi.org/10.26024/kgmp-c556>, 2020.
- Davini, P. and D’Andrea, F.: Northern Hemisphere Atmospheric Blocking Representation in Global Climate Models: Twenty Years of Improvements?, *Journal of Climate*, 29, 8823–8840, <https://doi.org/10.1175/JCLI-D-16-0242.1>, 2016.
- Derbyshire, J.: Increasing Preparedness for Extreme Events using Plausibility-Based Scenario Planning: Lessons from COVID-19, *Risk Analysis*, 42, 97–104, <https://doi.org/10.1111/risa.13827>, 2022.
- D’Ippoliti, D., Michelozzi, P., Marino, C., de’Donato, F., Menne, B., Katsouyanni, K., Kirchmayer, U., Analitis, A., Medina-Ramón, M., Paldy, A., Atkinson, R., Kovats, S., Bisanti, L., Schneider, A., Lefranc, A., Iñiguez, C., and Perucci, C. A.: The impact of heat waves on mortality in 9 European cities: results from the EuroHEAT project, *Environmental Health*, 9, 37, <https://doi.org/10.1186/1476-069X-9-37>, 2010.
- Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.-W., Xu, T., and Murray, D.: Was there a basis for anticipating the 2010 Russian heat wave?, *Geophysical Research Letters*, 38, <https://doi.org/10.1029/2010GL046582>, 2011.

- Domeisen, D. I. V., Eltahir, E. A. B., Fischer, E. M., Knutti, R., Perkins-Kirkpatrick, S. E., Schär, C., Seneviratne, S. I., Weisheimer, A., and Wernli, H.: Prediction and projection of heatwaves, *Nature Reviews Earth & Environment*, 4, 36–50, <https://doi.org/10.1038/s43017-022-00371-z>, 2023.
- 825 Drouard, M. and Woollings, T.: Contrasting Mechanisms of Summer Blocking Over Western Eurasia, *Geophysical Research Letters*, 45, 12,040–12,048, <https://doi.org/10.1029/2018GL079894>, 2018.
- Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O.: Climate Extremes: Observations, Modeling, and Impacts, *Science*, 289, 2068–2074, <https://doi.org/10.1126/science.289.5487.2068>, 2000.
- Feser, F., Van Garderen, L., and Hansen, F.: The Summer Heatwave 2022 over Western Europe: An Attribution to Anthropogenic Climate Change, *Bulletin of the American Meteorological Society*, 105, E2175–E2179, <https://doi.org/10.1175/BAMS-D-24-0017.1>, 2024.
- 830 Finkel, J. and O’Gorman, P. A.: Bringing Statistics to Storylines: Rare Event Sampling for Sudden, Transient Extreme Events, *Journal of Advances in Modeling Earth Systems*, 16, e2024MS004 264, <https://doi.org/10.1029/2024MS004264>, 2024.
- Finkel, J. and O’Gorman, P. A.: Rare Event Sampling for Moving Targets: Extremes of Temperature and Daily Precipitation in a General Circulation Model, *Journal of Advances in Modeling Earth Systems*, 18, e2025MS005 456, <https://doi.org/10.1029/2025MS005456>, 2026.
- 835 Fischbacher-Smith, D.: Beyond the worst case scenario: ‘Managing’ the risks of extreme events, *Risk Management*, 12, 1–8, <https://doi.org/10.1057/rm.2009.17>, 2010.
- Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., and Schär, C.: Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave, *Journal of Climate*, 20, 5081–5099, <https://doi.org/10.1175/JCLI4288.1>, 2007.
- Fischer, E. M., Sippel, S., and Knutti, R.: Increasing probability of record-shattering climate extremes, *Nature Climate Change*, 11, 689–695, <https://doi.org/10.1038/s41558-021-01092-9>, 2021.
- 840 Fischer, E. M., Beyerle, U., Bloin-Wibe, L., Gessner, C., Humphrey, V., Lehner, F., Pendergrass, A. G., Sippel, S., Zeder, J., and Knutti, R.: Storylines for unprecedented heatwaves based on ensemble boosting, *Nature Communications*, 14, 4643, <https://doi.org/10.1038/s41467-023-40112-4>, 2023.
- Fischer, E. M., Bador, M., Huser, R., Kendon, E. J., Robinson, A., and Sippel, S.: Record-breaking extremes in a warming climate, *Nature Reviews Earth & Environment*, pp. 1–15, <https://doi.org/10.1038/s43017-025-00681-y>, 2025.
- 845 Fleishman, E., Rupp, D. E., Loikith, P. C., Bumbaco, K. A., and O’Neill, L. W.: Synthesis of Publications on the Anomalous June 2021 Heat Wave in the Pacific Northwest of the United States and Canada, *Bulletin of the American Meteorological Society*, 106, E1155–E1174, <https://doi.org/10.1175/BAMS-D-24-0188.1>, 2025.
- García-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J., and Fischer, E. M.: A Review of the European Summer Heat Wave of 2003, *Critical Reviews in Environmental Science and Technology*, 40, 267–306, <https://doi.org/10.1080/10643380802238137>, 2010.
- 850 Garfinkel, C. I., Rostkier-Edelstein, D., Morin, E., Hochman, A., Schwartz, C., and Nirel, R.: Precursors of summer heat waves in the Eastern Mediterranean, *Quarterly Journal of the Royal Meteorological Society*, 150, 3757–3773, <https://doi.org/10.1002/qj.4795>, 2024.
- Gessner, C., Fischer, E. M., Beyerle, U., and Knutti, R.: Very Rare Heat Extremes: Quantifying and Understanding Using Ensemble Reinitialization, *Journal of Climate*, 34, 6619–6634, <https://doi.org/10.1175/JCLI-D-20-0916.1>, 2021.
- 855 Giardina, C., Kurchan, J., Lecomte, V., and Tailleur, J.: Simulating Rare Events in Dynamical Processes, *Journal of Statistical Physics*, 145, 787–811, <https://doi.org/10.1007/s10955-011-0350-4>, 2011.
- Goodwin, P. and Wright, G.: The limits of forecasting methods in anticipating rare events, *Technological Forecasting and Social Change*, 77, 355–368, <https://doi.org/10.1016/j.techfore.2009.10.008>, 2010.

- Green, J. S. A.: The Weather During July 1976: Some Dynamical Considerations of the Drought, *Weather*, 32, 120–126, <https://doi.org/10.1002/j.1477-8696.1977.tb04532.x>, 1977.
- Hazeleger, W., van den Hurk, B. J. J. M., Min, E., van Oldenborgh, G. J., Petersen, A. C., Stainforth, D. A., Vasileiadou, E., and Smith, L. A.: Tales of future weather, *Nature Climate Change*, 5, 107–113, <https://doi.org/10.1038/nclimate2450>, 2015.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1940 to present, <https://doi.org/10.24381/cds.bd0915c6>, 2023a.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1940 to present, <https://doi.org/10.24381/cds.adbb2d47>, 2023b.
- Hochman, A., Messori, G., Quinting, J. F., Pinto, J. G., and Grams, C. M.: Do Atlantic-European Weather Regimes Physically Exist?, *Geophysical Research Letters*, 48, e2021GL095574, <https://doi.org/10.1029/2021GL095574>, 2021.
- Hotz, B., Papritz, L., and Röthlisberger, M.: Understanding the vertical temperature structure of recent record-shattering heatwaves, *Weather and Climate Dynamics*, 5, 323–343, <https://doi.org/10.5194/wcd-5-323-2024>, 2024.
- Kautz, L.-A., Martius, O., Pfahl, S., Pinto, J. G., Ramos, A. M., Sousa, P. M., and Woollings, T.: Atmospheric blocking and weather extremes over the Euro-Atlantic sector – a review, *Weather and Climate Dynamics*, 3, 305–336, <https://doi.org/10.5194/wcd-3-305-2022>, 2022.
- Kelder, T., Marjoribanks, T. I., Slater, L. J., Prudhomme, C., Wilby, R. L., Wagemann, J., and Dunstone, N.: An open workflow to gain insights about low-likelihood high-impact weather events from initialized predictions, *Meteorological Applications*, 29, e2065, <https://doi.org/10.1002/met.2065>, 2022a.
- Kelder, T., Wanders, N., van der Wiel, K., Marjoribanks, T. I., Slater, L. J., Wilby, R. L., and Prudhomme, C.: Interpreting extreme climate impacts from large ensemble simulations—are they unseen or unrealistic?, *Environmental Research Letters*, 17, 044052, <https://doi.org/10.1088/1748-9326/ac5cf4>, 2022b.
- Kelder, T., Heinrich, D., Klok, L., Thompson, V., Goulart, H. M. D., Hawkins, E., Slater, L. J., Suarez-Gutierrez, L., Wilby, R. L., Coughlan de Perez, E., Stephens, E. M., Burt, S., van den Hurk, B., de Vries, H., van der Wiel, K., Schipper, E. L. F., Carmona Baéz, A., van Bueren, E., and Fischer, E. M.: How to stop being surprised by unprecedented weather, *Nature Communications*, 16, 1–15, <https://doi.org/10.1038/s41467-025-57450-0>, 2025.
- Kendon, E. J., Ciavarella, A., McCarthy, M., Brown, S., Christidis, N., Kay, G., Dunstone, N., Fereday, D., and Pope, J. O.: Multiperspective view of the 1976 drought–heatwave event and its changing likelihood, *Quarterly Journal of the Royal Meteorological Society*, 150, 232–261, <https://doi.org/10.1002/qj.4594>, 2024.
- Klif, A., Garfinkel, C. I., Rostkier-Edelstein, D., and Hochman, A.: Process-based evaluation of Eastern Mediterranean heatwave development in the CMIP6 models, *Weather and Climate Extremes*, 53, 100918, <https://doi.org/10.1016/j.wace.2026.100918>, 2026.

- Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S., and Gray, L.: Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern, *Environmental Research Letters*, 14, 054 002, <https://doi.org/10.1088/1748-9326/ab13bf>, 2019.
- Le Tertre, A., Lefranc, A., Eilstein, D., Declercq, C., Medina, S., Blanchard, M., Chardon, B., Fabre, P., Filleul, L., Jusot, J.-F., Pascal, L., Prouvost, H., Cassadou, S., and Ledrans, M.: Impact of the 2003 Heatwave on All-Cause Mortality in 9 French Cities, *Epidemiology*, 17, 75, <https://doi.org/10.1097/01.ede.0000187650.36636.1f>, 2006.
- Liu, P., Reed, K. A., Garner, S. T., Zhao, M., and Zhu, Y.: Blocking Simulations in GFDL GCMs for CMIP5 and CMIP6, *Journal of Climate*, 35, 5053–5070, <https://doi.org/10.1175/JCLI-D-21-0456.1>, 2022.
- Lucarini, V., Melinda Galfi, V., Riboldi, J., and Messori, G.: Typicality of the 2021 Western North America summer heatwave, *Environmental Research Letters*, 18, 015 004, <https://doi.org/10.1088/1748-9326/acab77>, 2023.
- Lüthi, S., Huber, V., Pascal, M., Beyerle, U., Pyrina, M., Domeisen, D., Vicedo-Cabrera, A. M., and Fischer, E.: Storylines for month-long heatwaves and associated heat-related mortality impacts over Western Europe, <https://doi.org/10.21203/rs.3.rs-5356341/v1>, 2024.
- Mahesh, A., Collins, W., Bonev, B., Brenowitz, N., Cohen, Y., Elms, J., Harrington, P., Kashinath, K., Kurth, T., North, J., OBrien, T., Pritchard, M., Pruitt, D., Risser, M., Subramanian, S., and Willard, J.: Huge Ensembles Part I: Design of Ensemble Weather Forecasts using Spherical Fourier Neural Operators, <https://doi.org/10.48550/arXiv.2408.03100>, arXiv:2408.03100 [physics], 2025a.
- Mahesh, A., Collins, W., Bonev, B., Brenowitz, N., Cohen, Y., Harrington, P., Kashinath, K., Kurth, T., North, J., OBrien, T., Pritchard, M., Pruitt, D., Risser, M., Subramanian, S., and Willard, J.: Huge Ensembles Part II: Properties of a Huge Ensemble of Hindcasts Generated with Spherical Fourier Neural Operators, <https://doi.org/10.48550/arXiv.2408.01581>, arXiv:2408.01581 [cs], 2025b.
- Muñoz-Sabater, J.: ERA5-Land monthly averaged data from 1950 to present., <https://doi.org/10.24381/cds.68d2bb30>, 2019.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>, 2021.
- Nordmann, A.: (Im)Plausibility2, *International Journal of Foresight and Innovation Policy*, 9, 125–132, <https://doi.org/10.1504/IJFIP.2013.058612>, 2013.
- Noyelle, R., Yiou, P., and Faranda, D.: Investigating the typicality of the dynamics leading to extreme temperatures in the IPSL-CM6A-LR model, *Climate Dynamics*, 62, 1329–1357, <https://doi.org/10.1007/s00382-023-06967-5>, 2024.
- Noyelle, R., Caubel, A., Meurdesoif, Y., Faranda, D., and Yiou, P.: Evolution of the Dynamics of Centennial Hot Summers in Western Europe With Climate Change, *Geophysical Research Letters*, 52, e2025GL115 552, <https://doi.org/10.1029/2025GL115552>, 2025a.
- Noyelle, R., Caubel, A., Meurdesoif, Y., Yiou, P., and Faranda, D.: Statistical and Dynamical Aspects of Extremely Hot Summers in Western Europe Sampled with a Rare Event Algorithm, *Journal of Climate*, 38, 4763–4787, <https://doi.org/10.1175/JCLI-D-24-0635.1>, 2025b.
- Pappert, D., Tuel, A., Coumou, D., Vrac, M., and Martius, O.: Long vs. short: understanding the dynamics of persistent summer hot spells in Europe, *Weather and Climate Dynamics*, 6, 769–788, <https://doi.org/10.5194/wcd-6-769-2025>, 2025.
- Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R., and van Aalst, M.: A protocol for probabilistic extreme event attribution analyses, *Advances in Statistical Climatology, Meteorology and Oceanography*, 6, 177–203, <https://doi.org/10.5194/ascmo-6-177-2020>, 2020.
- Pitkänen, M. R. A., Mikkonen, S., Lehtinen, K. E. J., Lipponen, A., and Arola, A.: Artificial bias typically neglected in comparisons of uncertain atmospheric data, *Geophysical Research Letters*, 43, 10,003–10,011, <https://doi.org/10.1002/2016GL070852>, 2016.

- Polt, K. D., Ward, P. J., De Ruiter, M., Bogdanovich, E., Reichstein, M., Frank, D., and Orth, R.: Quantifying impact-relevant heatwave durations, *Environmental Research Letters*, 18, 104 005, <https://doi.org/10.1088/1748-9326/acf05e>, 2023.
- Ragone, F. and Bouchet, F.: Rare Event Algorithm Study of Extreme Warm Summers and Heatwaves Over Europe, *Geophysical Research Letters*, 48, e2020GL091 197, <https://doi.org/10.1029/2020GL091197>, 2021.
- Ragone, F., Wouters, J., and Bouchet, F.: Computation of extreme heat waves in climate models using a large deviation algorithm, *Proceedings of the National Academy of Sciences*, 115, 24–29, <https://doi.org/10.1073/pnas.1712645115>, 2018.
- Rebetez, M., Dupont, O., and Giroud, M.: An analysis of the July 2006 heatwave extent in Europe compared to the record year of 2003, *Theoretical and Applied Climatology*, 95, 1–7, <https://doi.org/10.1007/s00704-007-0370-9>, 2009.
- Robine, J.-M., Cheung, S. L. K., Roy, S. L., Oyen, H. V., Griffiths, C., Michel, J.-P., and Herrmann, F. R.: Death toll exceeded 70,000 in Europe during the summer of 2003, *Comptes Rendus. Biologies*, 331, 171–178, <https://doi.org/10.1016/j.crv.2007.12.001>, 2007.
- Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J.-E., Simpson, I. R., Stein, K., Stuecker, M. F., Yamaguchi, R., Bódai, T., Chung, E.-S., Huang, L., Kim, W. M., Lamarque, J.-F., Lombardozzi, D. L., Wieder, W. R., and Yeager, S. G.: Ubiquity of human-induced changes in climate variability, *Earth System Dynamics*, 12, 1393–1411, <https://doi.org/10.5194/esd-12-1393-2021>, 2021.
- Roemer, F. E., Fischer, E., Noyelle, R., and Knutti, R.: Supplementary code for "Assessing the plausibility of unprecedented events: A process-based approach applied to month-long heatwaves in Western Europe", <https://doi.org/10.5281/zenodo.19492889>, 2026a.
- Roemer, F. E., Fischer, E., Noyelle, R., and Knutti, R.: Supplementary data for "Assessing the plausibility of unprecedented events: A process-based approach applied to month-long heatwaves in Western Europe", <https://doi.org/10.5281/zenodo.19494785>, 2026b.
- Russo, S., Sillmann, J., and Fischer, E. M.: Top ten European heatwaves since 1950 and their occurrence in the coming decades, *Environmental Research Letters*, 10, 124 003, <https://doi.org/10.1088/1748-9326/10/12/124003>, 2015.
- Röthlisberger, M. and Martius, O.: Quantifying the Local Effect of Northern Hemisphere Atmospheric Blocks on the Persistence of Summer Hot and Dry Spells, *Geophysical Research Letters*, 46, 10 101–10 111, <https://doi.org/10.1029/2019GL083745>, 2019.
- Röthlisberger, M. and Papritz, L.: Quantifying the physical processes leading to atmospheric hot extremes at a global scale, *Nature Geoscience*, 16, 210–216, <https://doi.org/10.1038/s41561-023-01126-1>, 2023.
- Röthlisberger, M., Frossard, L., Bosart, L. F., Keyser, D., and Martius, O.: Recurrent Synoptic-Scale Rossby Wave Patterns and Their Effect on the Persistence of Cold and Hot Spells, *Journal of Climate*, 32, 3207–3226, <https://doi.org/10.1175/JCLI-D-18-0664.1>, 2019.
- Röthlisberger, M., Sprenger, M., Flaounas, E., Beyerle, U., and Wernli, H.: The substructure of extremely hot summers in the Northern Hemisphere, *Weather and Climate Dynamics*, 1, 45–62, <https://doi.org/10.5194/wcd-1-45-2020>, 2020.
- Röthlisberger, M., Sprenger, M., Beyerle, U., Fischer, E. M., and Wernli, H.: Advective, adiabatic and diabatic contributions to heat extremes simulated with the Community Earth System Model version 2, *Weather and Climate Dynamics*, 7, 1363–1384, <https://doi.org/10.5194/wcd-7-1363-2026>, 2026.
- Schulzweida, U.: CDO User Guide, <https://doi.org/10.5281/zenodo.10020800>, 2023.
- Schumacher, D. L., Singh, J., Hauser, M., Fischer, E. M., Wild, M., and Seneviratne, S. I.: Exacerbated summer European warming not captured by climate models neglecting long-term aerosol changes, *Communications Earth & Environment*, 5, 182, <https://doi.org/10.1038/s43247-024-01332-8>, 2024.
- Selin, C.: Trust and the illusive force of scenarios, *Futures*, 38, 1–14, <https://doi.org/10.1016/j.futures.2005.04.001>, 2006.
- Selin, C. and Guimarães Pereira, A.: Pursuing plausibility, *International Journal of Foresight and Innovation Policy*, 9, 93–109, <https://doi.org/10.1504/IJFIP.2013.058616>, 2013.

- Seneviratne, S. I., Koster, R. D., Guo, Z., Dirmeyer, P. A., Kowalczyk, E., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Oleson, K. W., and Verseghy, D.: Soil Moisture Memory in AGCM Simulations: Analysis of Global Land–Atmosphere Coupling Experiment (GLACE) Data, *Journal of Hydrometeorology*, 7, 1090–1112, <https://doi.org/10.1175/JHM533.1>, 2006.
- 975 Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, *Earth-Science Reviews*, 99, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004>, 2010.
- Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., Fowler, H. J., James, R., Maraun, D., Martius, O., Senior, C. A., Sobel, A. H., Stainforth, D. A., Tett, S. F. B., Trenberth, K. E., van den Hurk, B. J. J. M., Watkins, N. W., Wilby, R. L., and
- 980 Zenghelis, D. A.: Storylines: an alternative approach to representing uncertainty in physical aspects of climate change, *Climatic Change*, 151, 555–571, <https://doi.org/10.1007/s10584-018-2317-9>, 2018.
- Sillmann, J., Shepherd, T. G., van den Hurk, B., Hazeleger, W., Martius, O., Slingo, J., and Zscheischler, J.: Event-Based Storylines to Address Climate Risk, *Earth’s Future*, 9, e2020EF001 783, <https://doi.org/10.1029/2020EF001783>, 2021.
- Singh, J., Sippel, S., and Fischer, E. M.: Circulation dampened heat extremes intensification over the Midwest USA and amplified over
- 985 Western Europe, *Communications Earth & Environment*, 4, 432, <https://doi.org/10.1038/s43247-023-01096-7>, 2023.
- Suarez-Gutierrez, L., Müller, W. A., Li, C., and Marotzke, J.: Dynamical and thermodynamical drivers of variability in European summer heat extremes, *Climate Dynamics*, 54, 4351–4366, <https://doi.org/10.1007/s00382-020-05233-2>, 2020.
- Suarez-Gutierrez, L., Beyerle, U., Mittermeier, M., Vautard, R., and Fischer, E.: Worst-Case European Heat and Drought Storylines generated using Ensemble Boosting, <https://doi.org/10.21203/rs.3.rs-7113183/v1>, iSSN: 2693-5015, 2025.
- 990 Sánchez-Benítez, A., Goessling, H., Pithan, F., Semmler, T., and Jung, T.: The July 2019 European Heat Wave in a Warmer Climate: Storyline Scenarios with a Coupled Model Using Spectral Nudging, *Journal of Climate*, 35, 2373–2390, <https://doi.org/10.1175/JCLI-D-21-0573.1>, 2022.
- Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., and Belcher, S. E.: High risk of unprecedented UK rainfall in the current climate, *Nature Communications*, 8, 107, <https://doi.org/10.1038/s41467-017-00275-3>, 2017.
- 995 Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Hardiman, S. C., Ren, H.-L., Lu, B., and Belcher, S. E.: Risk and dynamics of unprecedented hot months in South East China, *Climate Dynamics*, 52, 2585–2596, <https://doi.org/10.1007/s00382-018-4281-5>, 2019.
- Thompson, V., Mitchell, D., Hegerl, G. C., Collins, M., Leach, N. J., and Slingo, J. M.: The most at-risk regions in the world for high-impact heatwaves, *Nature Communications*, 14, 2152, <https://doi.org/10.1038/s41467-023-37554-1>, 2023.
- Tibaldi, S. and Molteni, F.: On the operational predictability of blocking, *Tellus A*, 42, 343–365, <https://doi.org/10.1034/j.1600-0870.1990.t01-2-00003.x>, 1990.
- 1000 Tuel, A. and Martius, O.: Weather persistence on sub-seasonal to seasonal timescales: a methodological review, *Earth System Dynamics*, 14, 955–987, <https://doi.org/10.5194/esd-14-955-2023>, 2023.
- Tyrlis, E., Lelieveld, J., and Steil, B.: The summer circulation over the eastern Mediterranean and the Middle East: influence of the South Asian monsoon, *Climate Dynamics*, 40, 1103–1123, <https://doi.org/10.1007/s00382-012-1528-4>, num Pages: 1103-1123, 2013.
- 1005 Urueña, S.: Understanding “plausibility”: A relational approach to the anticipatory heuristics of future scenarios, *Futures*, 111, 15–25, <https://doi.org/10.1016/j.futures.2019.05.002>, 2019.
- van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., and Burgers, G.: Estimating return periods of extreme events from ECMWF seasonal forecast ensembles, *International Journal of Climatology*, 25, 1345–1354, <https://doi.org/10.1002/joc.1155>, 2005.

