Climate Services 44 (2026) 100742

Contents lists available at ScienceDirect

Climate Services

journal homepage: www.elsevier.com/locate/cliser

A transparent workflow for selecting heatwave stress-test years for energy system modelling

Johannes Schuhmacher *, Stephanie Gutmayer, Max Kleinebrahm, Armin Ardone, Wolf Fichtner

Institute for Industrial Production, Karlsruhe Institute of Technology (KIT), Germany

HIGHLIGHTS

- Climate data are translated into model-ready heatwave stress tests.
- Regional aggregation combines local HWMId severity with spatial coverage.
- Selected scenario years depend on climate-model chain and system perimeter.
- The 2003 and 2026 heatwaves are comparable in terms of temporal and spatial extent.

ARTICLE INFO

Keywords:
Heatwave
Heatwave magnitude index daily
HWMId
Energy-system modelling
Weather-year selection
Stress test
Climate scenario
Workflow
Reproducibility

ABSTRACT

Energy-system studies often rely on a small number of weather years to test system performance under extreme conditions. Translating observational and climate-projection datasets into traceable and model-ready stress-test scenarios is an important interface between climate information and energy-system modelling. However, the selection of heatwave stress-test years is difficult to reproduce because the link between climate data, heatwave definition, spatial aggregation and scenario year is not always documented. This study presents a transparent workflow for selecting heatwave years from gridded observations and climate-model projections. The workflow uses the Heat Wave Magnitude Index daily (HWMId) as an event-based metric that combines heatwave duration and locally normalized temperature anomaly and ranks years according to spatially aggregated heatwave severity. Applied to a Germany–France case study, the workflow identifies 2003 as the strongest completed historical heatwave year and shows that the June 2026 event, evaluated using ERA5 and NOAA CORe data through 1 July 2026, approaches but remains slightly below the 2003 benchmark. Future stress-test years vary across CORDEX-CMIP5 and CORDEX-CMIP6 projection chains, confirming that selected years should be reported together with the climate data product, model chain, spatial domain and ranking criterion defining them. Sensitivity analyses across temperature thresholds, reanalysis datasets, regional definitions, electricity-generation-capacity and population weightings show robust stress-test years per dataset and model chain, although their ordering can change. The workflow provides energy-system researchers and practitioners with a reproducible way to translate climate data into traceable heatwave stress-test scenarios while retaining information on the data product, climate-model chain, spatial domain and selection criterion.

Practical implications

Detailed energy-system models cannot always be run for every available historical and projected weather year while retaining high spatial and temporal resolution, detailed operational constraints and the desired range of technology and policy sensitivities. Reducing the number of weather-year runs can therefore enable the downstream model to retain a higher level of technological, operational and spatial detail. However, selecting only one or a few years informally can make resilience findings depend on scenarios whose meteorological basis is difficult to trace. The paper addresses this gap by providing an ex ante screening method for selecting a compact and documented set of heatwave stress tests. It is intended for targeted assessments of supply adequacy, heat-related electricity demand, thermal-generation constraints, storage, demand response and cross-border balancing. The appropriate number of weather years remains a study-design choice; the workflow is particularly useful when a limited and traceable scenario set allows the

* Corresponding author.
E-mail address: johannes.schuhmacher@kit.edu (J. Schuhmacher).

https://doi.org/10.1016/j.cliser.2026.100742
Received 29 July 2026; Received in revised form 11 September 2026; Accepted 26 September 2026
2405-8807/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Open access funding enabled and organized by Projekt DEAL.

downstream model to retain the detail required for the research question.

The workflow is intended for energy-system modellers, infrastructure planners, utilities, public agencies and researchers. It translates gridded observations, reanalysis and climate projections into concrete weather-year selections for downstream models. Heatwave severity is quantified using daily maximum temperature and the Heat Wave Magnitude Index daily, an event-based metric that combines event duration with the locally normalized temperature anomaly. Annual grid-cell values are then aggregated across the geographical perimeter of the energy-system study and ranked. Importantly, the service does not return an isolated calendar year: each selection remains linked to its climate data product, reference period, model chain, emissions pathway, spatial domain, aggregation rule and weighting choice.

In practice, users make four choices. First, they select an observational, reanalysis or projection dataset consistent with their time horizon. Second, they define a geographical domain matching the energy-system model. Third, they choose the baseline aggregation and, where relevant, test alternative weights representing exposure or infrastructure relevance, such as population or national electricity-generation capacity. Fourth, they select one or more high-ranking years. The selected calendar years can then be used to extract the complete and internally consistent weather time series required by the downstream model. Sensitivities across domains, weighting rules, ranking criteria and climate-model chains distinguish a primary stress test from additional cases. The open-source implementation and configuration files allow the procedure to be repeated for other regions and datasets.

The Germany–France application illustrates the practical interpretation. Both the *E*-OBS observational dataset and the ERA5 reanalysis identify 2003 as the strongest completed heatwave year over the study domain. Within ERA5, the June 2026 event, evaluated using data through 1 July 2026, ranks second and approaches, but remains slightly below, the 2003 benchmark in spatially aggregated heatwave magnitude. This provides a traceable answer to whether 2026 was comparable to 2003 under a stated metric, dataset and system boundary. For future conditions, selected years differ among regional climate-model chains. A future stress-test year should therefore not be communicated as a universal year for a particular emissions pathway. It is a model-chain-specific realization whose relevance must be assessed for the intended energy-system application. Alternative regional and capacity-based weightings may also change secondary years, helping users choose supplementary scenarios for specific stress mechanisms.

For applications in which cooling demand, human thermal stress or evaporative cooling is the dominant impact pathway, users can additionally test humidity-aware metrics. In the present case, this sensitivity changes 2026 from the second-ranked temperature-based year to the highest-ranked humid-heat year, illustrating why the meteorological metric should match the downstream application.

The selected years are meteorologically defined stress-test inputs, not forecasts, occurrence probabilities or return-period estimates. The index does not directly quantify electricity demand, plant derating, cooling-water restrictions, outages or cascading system effects. These impacts must be represented in the downstream energy-system model using appropriate sector-specific relationships and the complete temporal sequence of the selected weather year. A transparent application should report the selection metadata, simulate the primary and relevant sensitivity years, and keep conclusions conditional on the chosen climate product, model chain and system perimeter. Used in this way, the service connects climate information with detailed resilience modelling while preserving a clear audit trail from weather data to stress-test scenario.

1. Introduction

Recent heatwaves, like the one in June 2026 in Europe (Mohr et al., 2026; University of Freiburg, n.d.), are often discussed in comparison with well-known historical benchmark events, such as the 2003 European heatwave. However, whether a recent event is “stronger” or more relevant for energy-system stress testing depends on the heatwave definition, the spatial domain, the aggregation method and the meteorological dataset used. This creates a need for transparent workflows that can compare heatwave years consistently rather than relying on ad hoc event selection.

Heatwaves are relevant stress-test events for energy systems because they can increase cooling-related electricity demand while simultaneously constraining thermoelectric generation, hydropower availability, network operation and cross-border balancing (Breteau et al., 2026; Gonçalves et al., 2024; van Vliet et al., 2013). These effects are nonlinear and spatially heterogeneous, so stress-test years must preserve the temporal and spatial coherence of extreme conditions rather than relying on annual, seasonal or spatially averaged climate signals (Gøtske et al., 2024; Grochowicz et al., 2024). A selected heatwave scenario can then be used to assess operational resilience, including supply adequacy under peak-load conditions, regional bottlenecks, import dependence and the contribution of storage, demand response or other flexibility options.

Despite the growing recognition of heatwaves as relevant stress-test events, energy-system studies use a wide range of approaches to represent extreme heat. Some studies identify heatwaves directly from meteorological data using explicit metrics. For example, Kabir et al. (2026) study the correlation between heatwaves and outage occurrence using historical temperature and outage records. They use the heat index, which combines air temperature and humidity, as the heat exposure metric and aggregate it to the daily maximum heat index. Threshold values corresponding to the 80th – 95th percentiles are then derived from the empirical distribution of all historical warm-season observations (May – September). The final threshold is selected to ensure a sufficient number of heatwave days and concurrent outage observations for regression analysis, and heatwave events are defined as periods of at least two consecutive days above this threshold. Saki et al. (2025) similarly analyse the relationship between heatwaves, compound weather events and power outages using county-level outage data and meteorological observations for the United States. While they calculate the heat index, it is not used to define heatwaves. Instead, heat-related events are identified using official historical National Weather Service (NWS) heat advisories and warnings. Meteorological characteristics are then computed for these predefined event periods and combined with outage observations to investigate the relationship between weather conditions and power-system impacts. Molina et al. (2023) investigate the impact of heatwaves on electricity demand and wind power production in Southern Europe. Heatwaves are identified at the grid-cell level using *E*-OBS daily maximum temperature, where a heatwave is defined as a period of at least three consecutive days with temperatures exceeding the 90th percentile of a 31-day moving window during the 1971–2000 reference period. Since anomalies in electricity demand are the primary focus, the analysis is restricted to grid cells with high population density, and only heatwave events affecting at least 30% of these populated areas are considered. You et al. (2025) investigate the impact of heatwaves on global wind power availability. Heatwaves are identified using ERA5 daily maximum temperature, where a heatwave is defined as a period of at least three consecutive days exceeding the 90th-percentile threshold. The threshold is calculated separately for each grid cell and calendar day using a 15-day moving window over the 1980–2023 reference period. The identified heatwave periods are then combined with wind power density estimates to assess concurrent wind power shortages during heatwaves.

Other energy-system studies investigate heatwave impacts without explicitly defining heatwaves from meteorological data. Rübbelke and Vögele (2011), for example, select a future climate projection scenario that is described as comparable to the European heatwave of 2003. Abdin et al. (2019) model a sequence of projected years and discuss the impact of heatwave events, without applying an explicit heatwave

metric or defining a threshold. Another approach is the construction of synthetic scenarios, such as in Ke et al. (2016), where a five-day period in historical temperature data is increased by 10 °C to represent extreme conditions. Together, these studies demonstrate that heatwaves are commonly used in energy-system analysis, but also illustrate the lack of a consistent and reproducible methodology linking meteorological data, heatwave identification and scenario selection.

In contrast, the climate-science literature has developed a large number of formal heatwave definitions and metrics. Most definitions combine an intensity criterion with a minimum duration requirement, commonly three or more consecutive days above a temperature threshold (Bunting et al., 2024; Perkins and Alexander, 2013). Thresholds are often defined relative to the historical climatology at each location, commonly using the calendar-day 90th percentile of historical daily maximum or minimum temperature, estimated with a moving-window approach to account for the seasonal cycle. Perkins and Alexander (2013) note that the 90th percentile represents a compromise, providing a sufficient number of events for statistical analysis while still capturing conditions that are climatologically extreme.

A widely used index is the Heat Wave Magnitude Index daily (HWMId) introduced by Russo et al. (2014, 2015), which combines event duration and temperature exceedance into a single metric, enabling the ranking of individual heatwave events across regions and time periods. Subsequent studies have used HWMId in different ways to identify and characterize heatwaves. For Europe, Zschenderlein et al. (2019) adapt the HWMId event-identification procedure to represent spatially coherent regional heatwaves: in addition to the temporal HWMId criteria, at least 5% of a predefined European region must simultaneously satisfy the heatwave conditions, resulting in regional heatwave climatologies for six different European regions. Lin et al. (2022) instead calculate the annual maximum HWMId at each grid cell and analyse the resulting spatial fields and area-weighted regional HWMId across *E*-OBS observations and EURO-CORDEX GCM–RCM simulations. Their analysis focuses on the representation of historical heatwave magnitude, projected changes and the uncertainty introduced by different GCM–RCM combinations. Wang and Yan (2021) further extend HWMId to spatially coherent regional events by aggregating grid-cell magnitudes across the affected area. These studies demonstrate that HWMId can be used for grid-cell heatwave characterization, spatially coherent regional-event identification, regional aggregation and the comparison of historical and projected heatwave magnitude.

Other indices such as the Wet-Bulb Globe Temperature and Universal Thermal Climate Index are often motivated by impacts on human health, as they consider additional factors such as humidity (Becker et al., 2022). Comparative studies have shown that no single metric is universally superior, but rather that different metrics emphasize different aspects of heatwave risk and are suited to different applications. Recent methodological developments have also considered how multiple heatwave characteristics should be combined within a single magnitude measure. Zhang et al. (2026), for example, propose a Comprehensive Heatwave Magnitude Index (CHMI) that accounts explicitly for the balance among heatwave characteristics such as intensity, duration, frequency and spatial extent, addressing the possibility that compensatory aggregation may mask strongly imbalanced events. For energy-system applications, no single heat metric captures all relevant stress mechanisms. Daily maximum air temperature is directly relevant to several engineering impacts, including temperature-dependent electricity demand and constraints on thermal generation and cooling (Ke et al., 2016; Molina et al., 2023; van Vliet et al., 2013). Humidity can additionally affect human thermal stress, cooling demand and the performance of evaporative cooling, while full outdoor heat stress may also depend on radiation and wind. For the present application, HWMId provides a suitable common meteorological basis because it requires only daily maximum temperature, combines temperature exceedance and event duration at the grid-cell level, and can be applied consistently across the observational and climate-projection datasets considered here. Because humidity may nevertheless alter which years are identified as most severe, we additionally test the robustness of the ERA5 ranking using humidity-aware heat metrics.

Previous energy-system studies have shown that weather-year selection and interannual climate variability can materially affect system design, operation and robustness assessments (Antonini et al., 2024; Craig et al., 2022; Gøtske et al., 2024). Consequently, the procedure used to identify representative or extreme years can influence the resulting stress-test scenarios and the conclusions drawn from them. This need for transparent weather and climate data workflows is consistent with broader calls to better connect climate modelling and energy-system modelling, particularly when future climate conditions are translated into energy-system inputs (Craig et al., 2022).

Taken together, these two strands of literature leave a specific methodological step less explicitly addressed: the translation from meteorological heatwave characterization to the selection of model-ready weather years. HWMId by itself does not specify how grid-cell heatwave information should be translated into a limited set of complete weather-year scenarios for a spatially resolved energy-system model. This translation requires additional choices concerning the geographical system perimeter, the spatial aggregation and weighting of heatwave severity, the criterion used to rank complete years, and the treatment of differences across observational products and climate-model chains. The methodological contribution of the present study is therefore not the use of HWMId itself, but a reproducible scenario-selection workflow that makes these downstream choices explicit, tests their influence on the resulting year selection, and retains a traceable link between each selected scenario and the climate information from which it was derived.

From a climate-services perspective, this creates a specific information need at the interface between climate science and energy-system modelling. Observational and climate-model datasets provide large ensembles of meteorological time series, whereas detailed energy-system studies often require a limited number of concrete weather scenarios. Energy-system researchers and practitioners conducting climate-resilience analyses therefore need climate information that identifies severe heat conditions, preserves the original temporal structure of the selected weather year, matches the geographical perimeter of the energy-system model, and retains information on the underlying dataset, model chain and selection criterion.

Against this background, the study addresses three questions. First, how can concrete heatwave stress-test years be selected reproducibly from gridded climate information for downstream energy-system models? Second, how sensitive is the resulting selection to the climate-model chain, spatial domain, aggregation criterion and energy-system weighting? Third, was the June 2026 European heatwave comparable to the 2003 benchmark when duration, local anomaly and spatial extent are evaluated consistently?

To answer these questions, we develop a reproducible and inspectable workflow that transforms gridded observational and climate-projection data into traceable, model-ready heatwave stress-test scenarios. The method is demonstrated for Germany and France and can support assessments of energy-system configurations and adaptation options under extreme heat. The workflow is released as open-source code, and its main methodological choices are configurable, allowing it to be reproduced and adapted to other country sets, regional definitions, weighting schemes and heatwave indices.

2. Design requirements for a stress-test year definition in energy systems

In this study, a heatwave stress-test year is defined as a concrete weather year selected for use as an input scenario in downstream energy-system optimization. It is not intended to estimate the probability or return period of a heatwave, and it is not a forecast of a specific future year. Instead, it provides a reproducible scenario year that

preserves the temporal structure of a heatwave event and can be communicated as part of a compact scenario set.

Simulating a large ensemble of historical or projected weather years is the most comprehensive approach when the research question is to quantify weather-year variability or to identify problematic years ex post. Such multi-year analyses are valuable, but they are not always compatible with the modelling tasks for which stress-test years are commonly used. Detailed energy-system studies often combine high temporal resolution, spatially resolved networks, brownfield capacity constraints, sector coupling, technology-specific operating limits, market design assumptions, or many policy and investment sensitivities (Hoffmann et al., 2024). In such settings, running all available weather years can require substantial simplifications, such as reduced spatial scope, temporal clustering, simplified operational constraints, or a more stylised greenfield formulation.

The purpose of an ex-ante heatwave-year selection is a transparent scenario-definition step for modelling applications in which a small number of documented stress-test years must be passed to a detailed downstream model. This is particularly useful when the question is not which year becomes problematic in a large ex post ensemble, but how a specified energy-system configuration performs under a meteorologically well-defined and reproducible heatwave stress event.

For this purpose, the selection rule should satisfy four requirements. It should identify heatwave events using a recognised meteorological index, aggregate the event severity over the spatial perimeter relevant to the energy-system model, remain reproducible from stated data products and model chains, and expose key sensitivities such as spatial domain, climate-model choice and weighting criterion.

The workflow was built around four design choices: daily maximum temperature as the meteorological input, HWMId as the event-intensity metric, a country-mask aggregation over Germany and France, and sensitivity checks that expose alternative choices. HWMId is used because it captures exceedance duration and magnitude relative to a local reference distribution while remaining directly computable from gridded maximum-temperature fields (Russo et al., 2014, 2015). Human-biometeorological indices such as WBGT or UTCI are valuable for health-impact questions (Błażejczyk et al., 2013; Budd, 2008), and the Excess Heat Factor is widely used for severity classification (Nairn and Fawcett, 2015), but the present purpose is a meteorological energy-system stress-test definition based on temperature extremes, with spatial coverage incorporated through the subsequent regional aggregation.

In this context, heatwave stress testing can support climate adaptation and energy-system resilience assessments by providing a consistent meteorological basis for evaluating alternative system configurations under extreme heat conditions. The workflow does not prescribe adaptation measures; rather, it provides traceable climate information against which such measures can be assessed.

3. Data and workflow

The workflow uses complementary input data sources. For historical rankings, we use *E*-OBS v33.0e daily maximum temperature (tx) for 1950–2025, a station-based gridded observational dataset for Europe that is widely used in European climate-extreme analyses (Cornes et al., 2018). ERA5 hourly 2 m temperature is used as a reanalysis-based comparison and for the 2026 event assessment, because complete E-OBS running-year data are not yet available for 2026 (C3S, 2018). As a second historical reanalysis comparison, we used the NOAA Conventional Observation Reanalysis (CORe) daily maximum 2 m temperature. CORe spans 1950 to near real time and primarily assimilates conventional observations (National Oceanic and Atmospheric Administration, 2026). For projected stress-test years, we use CORDEX-CMIP5/RCP bias-adjusted 3-hourly near-surface air temperature (Copernicus Climate Change Service, 2021) and CORDEX-CMIP6/SSP ICON-CLM hourly near-surface air temperature (Nikulin et al., 2026). In both projection ensembles, sub-daily temperature is converted to daily maximum temperature before HWMId is calculated. The resulting workflow from temperature input data to ranked heatwave stress-test years is summarized in Fig. 1. ERA5 hourly 2 m dewpoint temperature is additionally used with 2 m air temperature to derive the exploratory humidity-aware metrics reported in Appendix D.

3.1. HWMId calculation

We use the Heat Wave Magnitude Index daily (HWMId) following Russo et al. (2015) to identify and rank heatwave years from gridded temperature data. HWMId is suitable for this purpose because it requires only daily maximum temperature and can therefore be applied consistently across long historical and projected climate time series. At the same time, it combines three properties that are relevant for climate-impact and energy-system applications: exceedance above a locally defined heat threshold, event duration, and the magnitude of the temperature anomaly relative to local climatology.

For each grid cell, daily maximum temperature is needed. Heatwave detection is then based on a grid-cell-specific and calendar-day-specific threshold. This threshold is calculated as the 90th percentile of daily maximum temperature within a 31-day moving calendar window over the 1981–2010 reference period. The moving window accounts for the seasonal cycle while providing a sufficiently large sample for percentile estimation. As a result, each grid cell has its own local threshold for each day of the year.

A heatwave event is defined as a period of at least three consecutive days on which daily maximum temperature exceeds the local 90th-percentile threshold. This minimum duration follows the established HWMId definition and distinguishes persistent heatwave events from isolated hot days (Russo et al., 2015). For each event day d in each grid cell g, the temperature magnitude $M_{g,d}$ is normalized using the interquartile range of annual maximum temperatures during the reference period:

$$M_{g,d} = max\left(\frac{T_{max} - T_{25}}{T_{75} - T_{25}}, 0\right) \qquad \forall g, d \tag{1}$$

where T_{max} is the daily maximum temperature of the event day d, and T_{25} and T_{75} are the first and third quartiles of the annual T_{max} maxima in the 1981–2010 reference period. The daily magnitude therefore increases with the exceedance above T_{25}, scaled by the local spread of annual maximum temperatures. The $HWMId_{g,e}$ each grid cell g of an event e is calculated as the sum of daily magnitudes over all event days d. This formulation assigns higher values to events that are more intense, longer-lasting, or both.

$$HWMId_{g,e} = \sum_{d} M_{g,d}, \qquad \forall g, e \tag{2}$$

The annual value $H_{g,y}$ is the maximum event $HWMId_{g,e}$ of the strongest event in that grid cell g and year y.

$$H_{g,y} = \max_{e} HWMId_{g,e}, \qquad \forall g, y \tag{3}$$

For each grid cell and year, only the strongest heatwave event is retained. The result is an annual gridded HWMId field in which each grid cell represents the magnitude of the most severe heatwave event occurring at that location in that year. Years without a qualifying heatwave event receive no positive contribution at that grid cell. This annual field forms the basis for the spatial aggregation and year ranking described below. Because the strongest event is retained independently for each grid cell, the resulting annual field may combine local heatwave events that occur on different dates. The spatially aggregated score is therefore interpreted as a year-level measure of spatially distributed heatwave severity. It does not require the strongest local events retained in the individual grid cells to occur simultaneously. When the selected year is passed to a downstream energy-system model, the complete

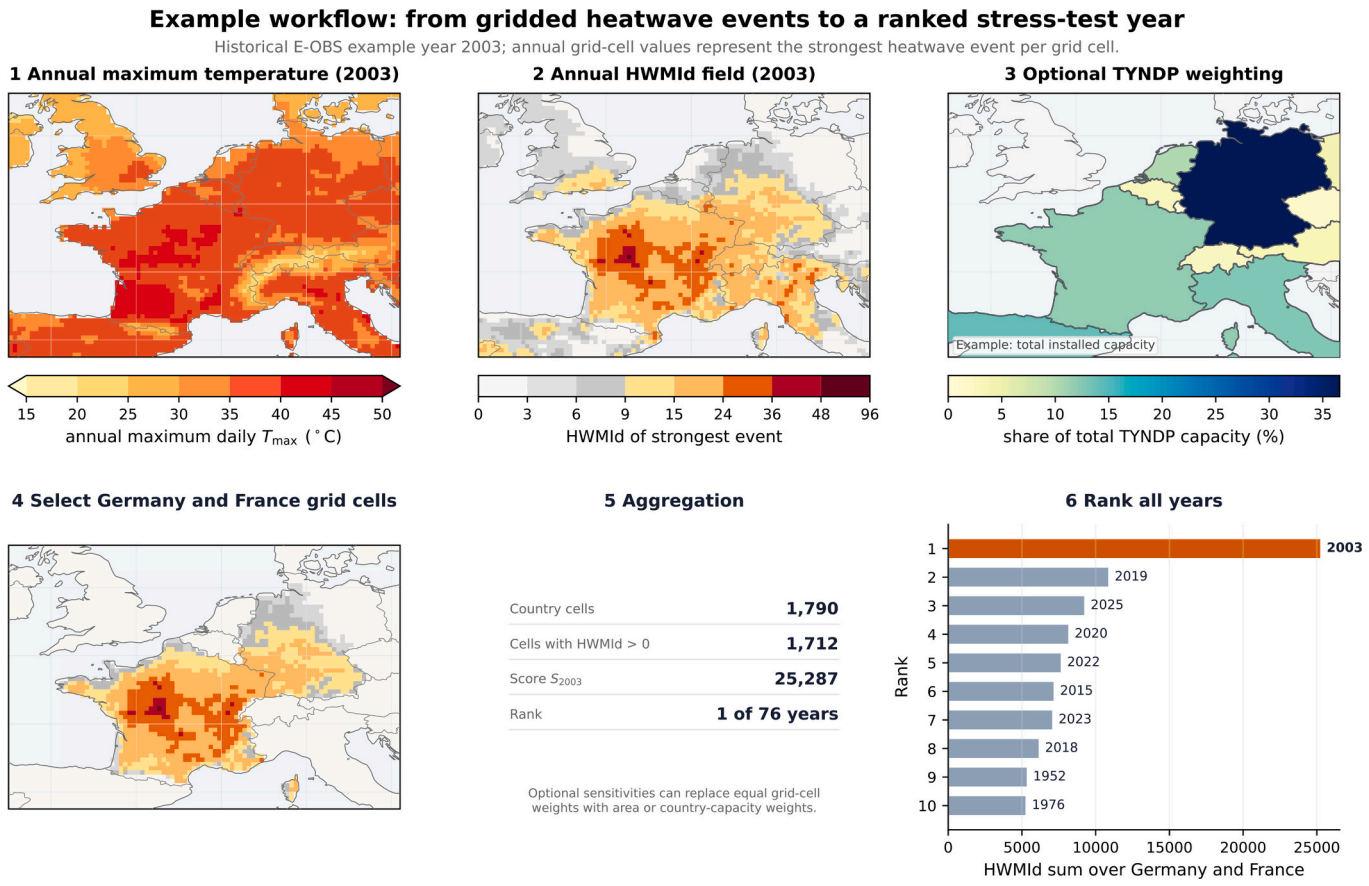


Fig. 1. Workflow for deriving heatwave stress-test years from gridded temperature data. Daily maximum temperature is converted into grid-cell HWMId values, the strongest heatwave event is retained for each grid cell and year, and the resulting annual HWMId fields are spatially aggregated over the selected country domain. Years are then ranked by the aggregated score, with sensitivity analyses varying the spatial domain, weighting scheme and ranking criterion.

original weather-year time series is retained, including the actual timing and sequence of all heatwave events.

Fig. 2 illustrates the temporal construction of HWMId for the E-OBS grid cell with the highest HWMId in the 2003 Germany-France ranking domain. Daily maximum temperature is compared with the local 90th-percentile threshold, the selected heatwave event is identified as the consecutive exceedance period, and the lower panel shows the daily magnitude contributions that are summed to the event HWMId.

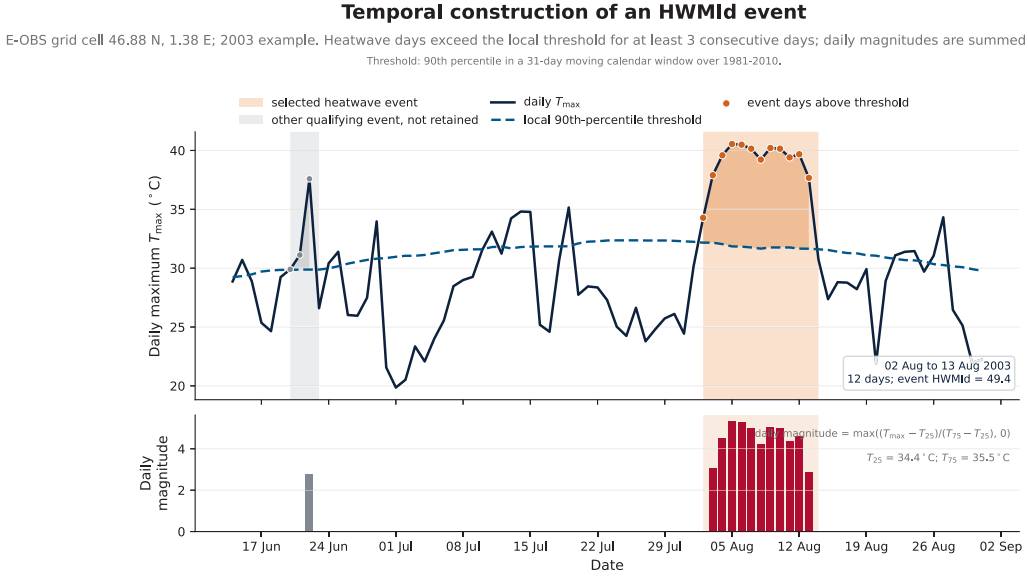


Fig. 2. Temporal construction of HWMId for an *E*-OBS grid cell during the 2003 heatwave. Daily maximum temperature is compared with the local 90th-percentile threshold estimated from a 31-day moving calendar window over the 1981–2010 reference period. The shaded period marks the selected heatwave event, defined by at least three consecutive exceedance days, and the bars show the daily magnitude contributions that are summed to the event HWMId.

The open-source code allows for operational choices, including the reference period, percentile threshold, minimum event duration, input temperature variable, spatial country mask, aggregation criterion and number of ranked years.

The reanalysis-dataset and threshold sensitivity analysis repeat the E-OBS, ERA5 and CORe calculations at the 90th, 95th and 99th percentiles while holding the reference period, three-day minimum-duration rule and spatial aggregation fixed (Appendix E). To test whether accounting for atmospheric humidity changes the historical event ranking, we additionally conduct an exploratory ERA5 sensitivity analysis using daily maximum Humidex, psychrometric wet-bulb temperature and a WBGT-inspired shade proxy. Humidex combines air temperature and humidity to represent perceived heat (Diaconescu et al., 2023). Psychrometric wet-bulb temperature characterizes the temperature attainable through evaporative cooling and therefore provides a humidity-sensitive measure of thermal conditions (Stull, 2011). The WBGT-inspired shade proxy combines air temperature and wet-bulb temperature using the weighting applied to radiation-free or shaded WBGT approximations, but does not represent full outdoor WBGT because direct solar radiation, wind effects and an explicitly calculated natural wet-bulb or globe temperature are not included (Buzan, 2024; Kong and Huber, 2022; Schwingshackl et al., 2024). For each alternative variable, local thresholds and reference-period quartiles are recalculated. The 1981–2010 reference period, 90th-percentile threshold, three-day minimum duration, annual strongest-event rule and Germany–France spatial aggregation are otherwise retained (Appendix D).

3.2. Spatial aggregation and ranking

The baseline ranking aggregates annual grid-cell HWMId values over the selected domain, here Germany and France. For each year, the heatwave score is calculated as the sum of annual HWMId values across all included grid cells. All years are then ranked in descending order of this score, and the highest-ranked years are retained as candidate heatwave scenario years.

The summed HWMId score is used as the baseline criterion because it combines local event severity with the geographical prevalence of high annual grid-cell values. A year with very high HWMId values in only a small part of the domain can therefore rank below a year with moderately high values distributed across a larger part of the domain. The criterion identifies years with spatially widespread annual heatwave severity, but it does not require the strongest local events to occur simultaneously. When the complete selected weather year is used in a downstream energy-system model, the original timing and sequence of all meteorological conditions are retained.

The baseline ranking treats all grid cells in the selected countries equally. Sensitivity analyses are used to test whether the selected years change when alternative spatial or energy-system assumptions are introduced. These include area-weighted and unweighted mean HWMId, maximum grid-cell HWMId, area-weighted heatwave duration, annual temperature anomaly, alternative regional masks and TYNDP 2024 capacity-based weights (ENTSO-E, 2024; ENTSO-E 2025). The Ten-Year Network Development Plan (TYNDP) provides European energy-system scenario data used in transmission-system planning, including country-level installed generation and storage capacities by technology. The capacities from the National Trends future energy system scenario for the year 2040 are used. Installed capacity describes the maximum rated power of assets such as solar PV, wind turbines, hydropower, thermal plants or storage units, and therefore provides a simple indicator of where energy-system infrastructure is located in the scenario data. We use these capacities as transparent country-level weights to test whether the meteorological heatwave ranking remains relevant when grid cells are weighted by the approximate spatial footprint of energy-system assets. In addition to generation and storage capacity weights, a demand-side weighting could also be considered. A simple option would be to weight grid cells by population density in order to approximate the spatial distribution of electricity demand. However, this would only imperfectly represent heatwave-related demand stress. During heatwaves, the additional electricity demand relevant for power-system stress tests is largely driven by cooling demand, which depends not only on population but also on air-conditioning penetration, building characteristics, income, behavioral responses and regional adaptation levels. Since harmonized spatial data on these determinants, in particular on air-conditioning ownership and use, are not available for the full study region, we treat population weighting as a separate robustness check rather than as an energy-demand weighting. The implementation and results are reported in Appendix B. The capacity-weighted sensitivity is therefore limited to generation and storage capacities, for which the TYNDP data provide a transparent and reproducible country-level proxy.

Because the TYNDP capacity data are available only at country level, we first calculate an area-weighted HWMId mean for each country and then weight these country-level values by the country's share of installed capacity in weighting scheme m. The resulting score S_y^m is a capacity-weighted mean of country-level HWMId values:

$$S_y^m = \sum_c \left(\frac{C_c^m}{\sum_{c'} C_{c'}^m} \cdot \sum_{g \in G_c} \frac{A_g}{\sum_{g' \in G_c} A_{g'}} H_{g,y} \right), \qquad \forall m, y \tag{4}$$

where $H_{g,y}$ is the annual HWMId value of grid cell g in year y, A_g is the grid-cell area, G_c is the set of grid cells in country c, and C_c^m is the TYNDP installed capacity of country c for weighting scheme m. The weighting schemes refer to different capacity categories, including total, renewables, PV, wind, storage, and thermal-plus-nuclear capacity. We use these capacities as transparent country-level weights in the broader Western/Central European sensitivity domain to test whether the meteorological heatwave ranking remains relevant when country-level energy-system infrastructure patterns are considered.

4. Scenario-year results

4.1. Results across historical data and climate-projection chains

Fig. 3 summarizes the top-ranked heatwave years across three historical data products and the CORDEX-CMIP5 and CORDEX-CMIP6 projection chains. The HWMId sums are used as ranking scores within each data set or model chain. They should not be interpreted as exact magnitude ratios across *E*-OBS, ERA5, NOAA CORe, CORDEX-CMIP5 and CORDEX-CMIP6, because the products differ in grid resolution, model formulation and data processing.

Across the common completed period 1950–2025, E-OBS, ERA5 and NOAA CORe all identify 2003 as the strongest heatwave year over Germany and France. E-OBS and NOAA CORe rank 2019 second, whereas ERA5 ranks 2025 second. When the available 2026 data through 1 July are included, ERA5 and NOAA CORe both place 2026 second behind 2003; E-OBS is available only through 2025. This agreement supports 2003 as the historical benchmark and 2026 as a closely matched recent stress-test candidate, while the ordering of lower-ranked years remains data-product dependent.

The CORDEX-CMIP5 results show that future stress-test years are strongly projection-chain specific. Under RCP4.5, the top-ranked years differ between the available chains: IPSL-WRF selects 2043 and 2070, MPI-CLM selects 2062 and 2030, and NCC-HIRHAM selects 2069 and 2070. Under RCP8.5, most selected years are located toward the end of the century, but the exact years again depend on the model chain, with 2091/2099 for CNRM-ALADIN, 2094/2093 for IPSL-WRF, 2092/2082 for MPI-CLM and 2091/2060 for NCC-HIRHAM.

The CORDEX-CMIP6 application shows that the same workflow can be applied to newer SSP-based projection data without changing the ranking workflow. The CNRM-driven ICON-CLM chain selects 2076/2059 under SSP1–2.6, 2078/2089 under SSP2–4.5, 2081/2080 under

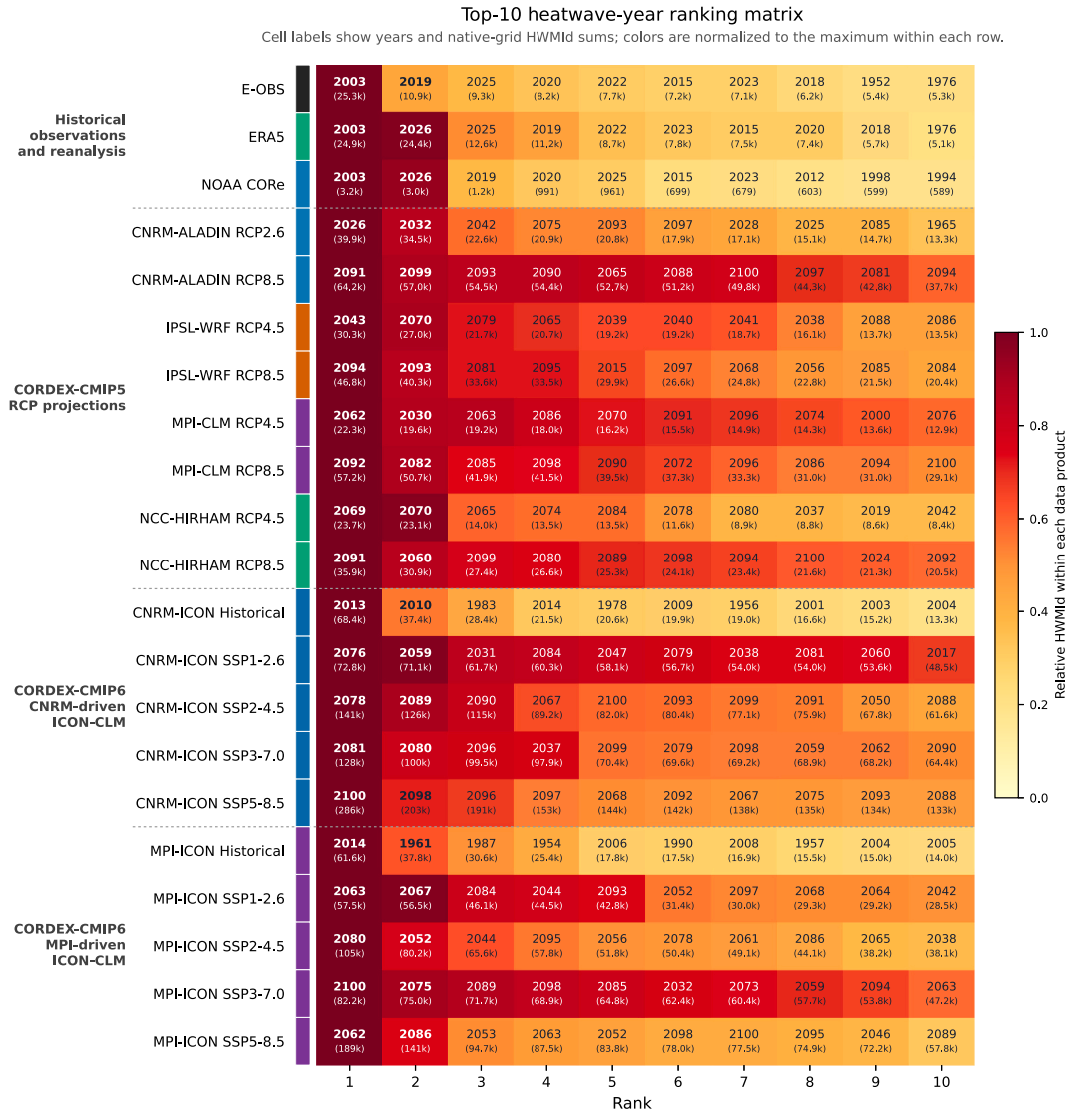


Fig. 3. Top-10 heatwave-year ranking matrix across historical data and CORDEX-CMIP5/CORDEX-CMIP6 projection chains. Cell labels show selected years and native-grid HWMId sum is in brackets; colour is normalized to the rank-1 value within each row. Values are intended for ranking within each data product or model chain. *E*-OBS includes complete years through 2025; ERA5 and NOAA CORe include the partial 2026 record through 1 July 2026.

SSP3–7.0 and 2100/2098 under SSP5–8.5. The MPI-driven ICON-CLM chain selects 2063/2067 under SSP1–2.6, 2080/2052 under SSP2–4.5, 2100/2075 under SSP3–7.0 and 2062/2086 under SSP5–8.5.

Fig. 4 presents the top-10 ranking curves for the main stress-test scenario set. The curves are grouped by data family to avoid over-interpreting absolute HWMId differences across products. In the historical panel, E-OBS and ERA5 use the left y-axis, while NOAA CORe is overlaid against the right y-axis because its coarser grid produces smaller raw sums. No normalization is applied. Within each data set or model chain, the vertical separation between ranked years indicates how clearly the selected years stand out from lower-ranked candidates. For the main energy-system stress-test package, the default historical benchmark is E-OBS 2003. ERA5 and NOAA CORe additionally identify 2026 as a relevant recent benchmark event. For projected applications, RCP4.5/IPSL-WRF 2043 remains the primary future heatwave year, with 2070 as a second-ranked sensitivity. RCP8.5/MPI-CLM 2092 remains a high-end late-century sensitivity and should always be reported together with its model chain.

The spatial HWMId fields of the selected first- and second-ranked years are shown in Fig. 5. This spatial check is important because the selected stress-test years should not only contain high local HWMId values, but should also represent spatially coherent heat stress across the Germany-France energy-system domain.

4.2. *Comparison of the 2026 and 2003 heatwaves*

The European June 2026 heatwave provides a useful test case for the workflow because it raises a practical question that cannot be answered by temperature maps or peak values alone: how does the event compare with the 2003 benchmark heatwave when duration, relative local intensity and spatial extent are evaluated consistently? Since the E-OBS current-year file is not yet available for 2026, the comparison is carried out within ERA5, where both 2003 and 2026 are assessed with the same data product and the same HWMId workflow.

The event-window totals in Fig. 6 sum daily contributions from all qualifying grid-cell events within the selected regional window. They therefore differ slightly from the annual ranking scores in Figs. 3 and 4, which retain only the strongest event in each grid cell. For 2003, the workflow-selected regional event window extends from 1 August to 31 August and reaches an aggregated HWMId of 25,336 over Germany and France. The spatially coherent regional core lasts from 2 August to 13 August and accounts for an HWMId of 25,030. For 2026, the selected

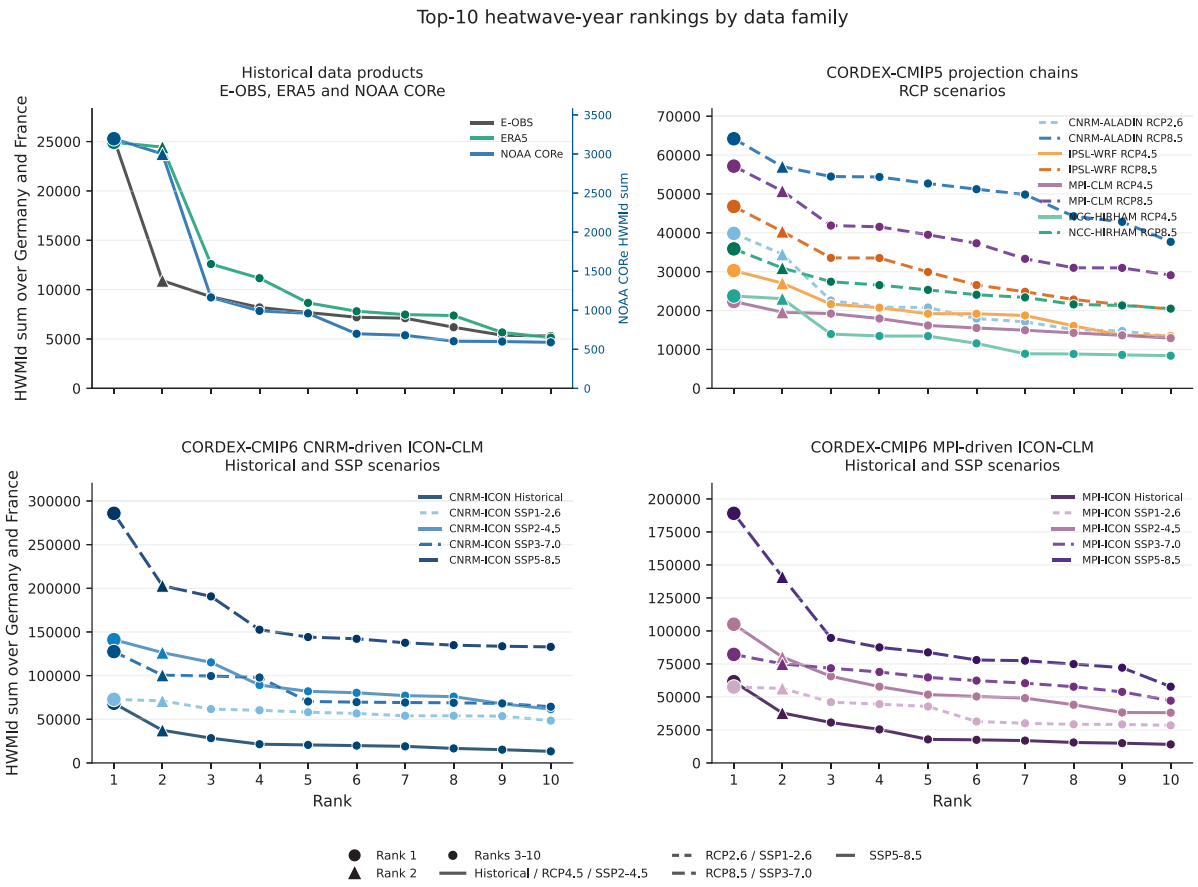


Fig. 4. Top-10 heatwave-year ranking curves by data family. Lines show ranked native-grid HWMId sums over Germany and France for historical products, CORDEX-CMIP5 projection chains and CORDEX-CMIP6 projection chains. In the historical panel, *E*-OBS and ERA5 use the left y-axis and NOAA CORe the right y-axis because its coarser native grid produces smaller raw sums; no normalization is applied. Curves are interpreted within each data product or model chain rather than as direct cross-product magnitude comparisons. Circles indicate rank 1, triangles indicate rank 2 and smaller circles indicate ranks 3–10.

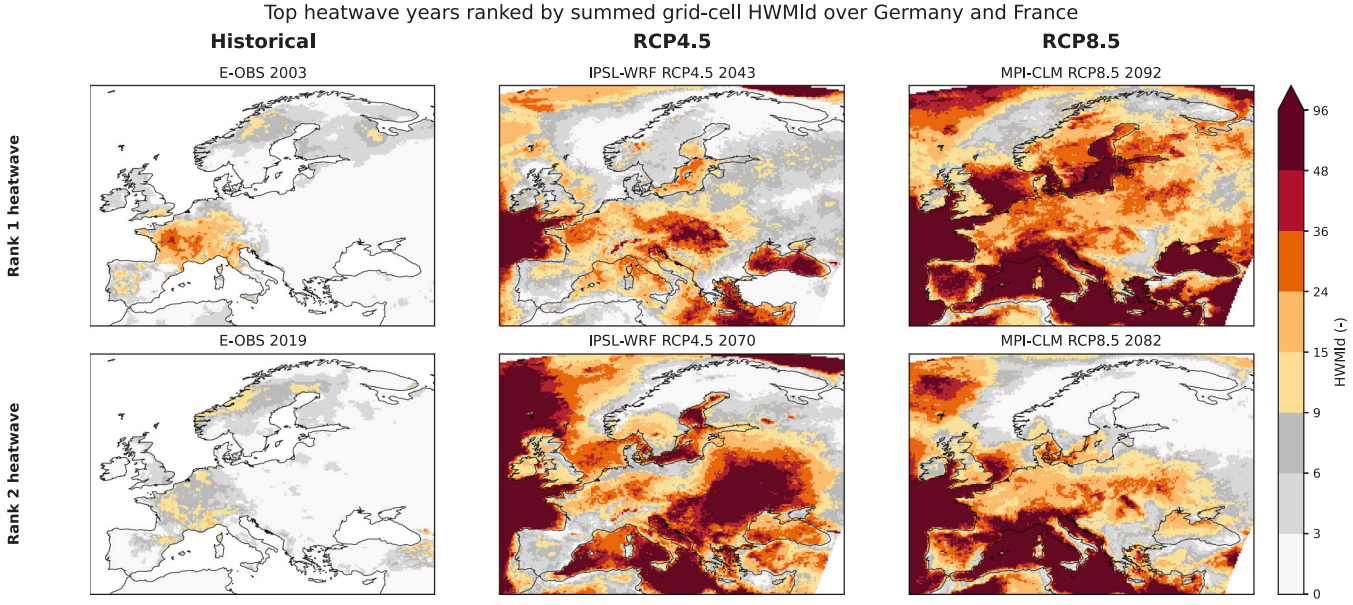


Fig. 5. Spatial distribution of annual HWMId values for the strongest and second-strongest heatwave years in the selected stress-test scenario set. Each grid cell shows the HWMId of the most severe local heatwave event occurring in that year. The maps indicate where the selected years produce spatially coherent heatwave severity across Europe and across the Germany-France ranking domain.

event period extends from 13 June to 1 July and reaches an aggregated HWMId of 24,594. Its regional core lasts from 17 June to 28 June and

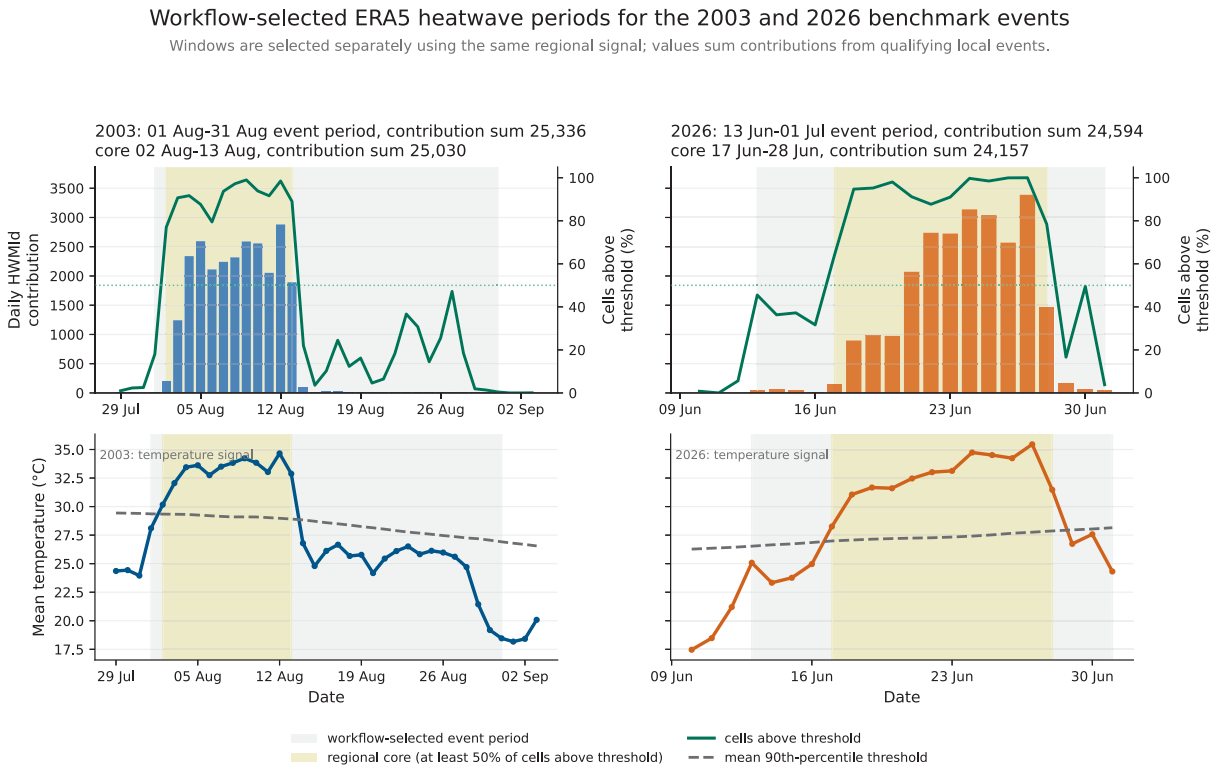


Fig. 6. ERA5 benchmark comparison of the 2003 and 2026 heatwaves over Germany and France. Grey shading indicates the selected event period: contiguous days with positive regional HWMId contribution that include a spatially coherent regional core. Yellow shading indicates the regional core, defined as contiguous days with at least 50% of grid cells exceeding their local 90th-percentile threshold. Bars show daily HWMId contributions; green lines show the share of cells above threshold, and dashed grey lines show the spatial mean threshold. Lower panels compare spatial mean daily maximum temperature with the mean threshold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reaches an HWMId of 24,157. The 2026 event is therefore very close to the 2003 benchmark in the ERA5 assessment, but remains slightly lower in aggregated HWMId over the Germany-France domain.

The comparison in Fig. 6 also illustrates the added value of the HWMId-based workflow. The maximum grid-cell temperature during the 2026 core is slightly higher than during the 2003 core in the ERA5 extraction, but the aggregated HWMId remains lower. This is because the metric also accounts for the duration of threshold exceedance, the local reference-period normalization and the spatial contribution of affected grid cells. In this sense, the workflow distinguishes a compact, intense recent event from a longer benchmark event without reducing the comparison to a single peak-temperature value.

The exploratory humidity-aware sensitivity in Appendix D further shows that the comparison depends on the impact pathway represented by the meteorological variable. With ERA5 data through 1 July 2026, all three humidity-aware variants rank 2026 first. The 2003 benchmark ranks second for Humidex and the shade-WBGT proxy and fifth for wet-bulb temperature. Rank correlations with the temperature-based HWMId ordering are approximately 0.91 for Humidex and the shade-WBGT proxy and 0.69 for wet-bulb temperature. Thus, 2026 appears more pronounced as a humid-heat event, whereas 2003 remains the slightly stronger temperature-based HWMId benchmark.

5. Robustness checks

5.1. Projection-chain sensitivity across CORDEX-CMIP5 and CORDEX-CMIP6

Applying the same workflow to the available CORDEX-CMIP5 and CORDEX-CMIP6 projection chains shows that selected stress-test years depend on the GCM/RCM chain and scenario framework. A GCM/RCM model chain denotes the combination of a global climate model, which provides the large-scale climate simulation, and a regional climate model, which dynamically downscales that simulation for Europe.

We do not expect the same calendar year to be selected across independent climate-model chains, because simulated years are not synchronized realizations of the same future weather. The purpose of this sensitivity check is to document how strongly the selected stress-test year, event magnitude and timing depend on the chosen GCM/RCM chain. The IPSL-WRF RCP4.5 chain reproduces 2043/2070 as the top two years. In contrast, alternative RCP4.5 chains select 2062/2030 for MPI-CLM and 2069/2070 for NCC-HIRHAM. RCP8.5 top years also vary across chains, ranging from 2091/2099 for CNRM-ALADIN to 2094/2093 for IPSL-WRF and 2092/2082 for MPI-CLM. This confirms that the scenario year is a model-chain-specific stress-test year rather than a generic scenario label. Fig. 7 visualizes this model-chain dependence by comparing the timing and magnitude of the ten highest-ranked years across the available Copernicus/CORDEX chains.

5.2. Regional-mask and weighting sensitivity

Regional-domain sensitivity tests whether the selected years are specific to the Germany-France perimeter or whether they also emerge when the spatial domain is narrowed or broadened. For RCP4.5/IPSL-WRF, the baseline Germany-France selection of 2043/2070 is retained when the domain is expanded to Germany, France, Benelux, Alpine countries and the broader Western/Central European domain (Fig. 8). The N-1 sensitivity in Appendix C shows that 2043 remains the top-ranked year in all Western/Central European leave-one-country-out variants except when France is omitted, where the top-ranked year shifts to 2039. This indicates that France contributes materially to the system-wide ranking of the 2043 event.

The RCP8.5/MPI-CLM case is more sensitive to the regional perimeter. Germany-France selects 2092/2082, whereas the broader Western/

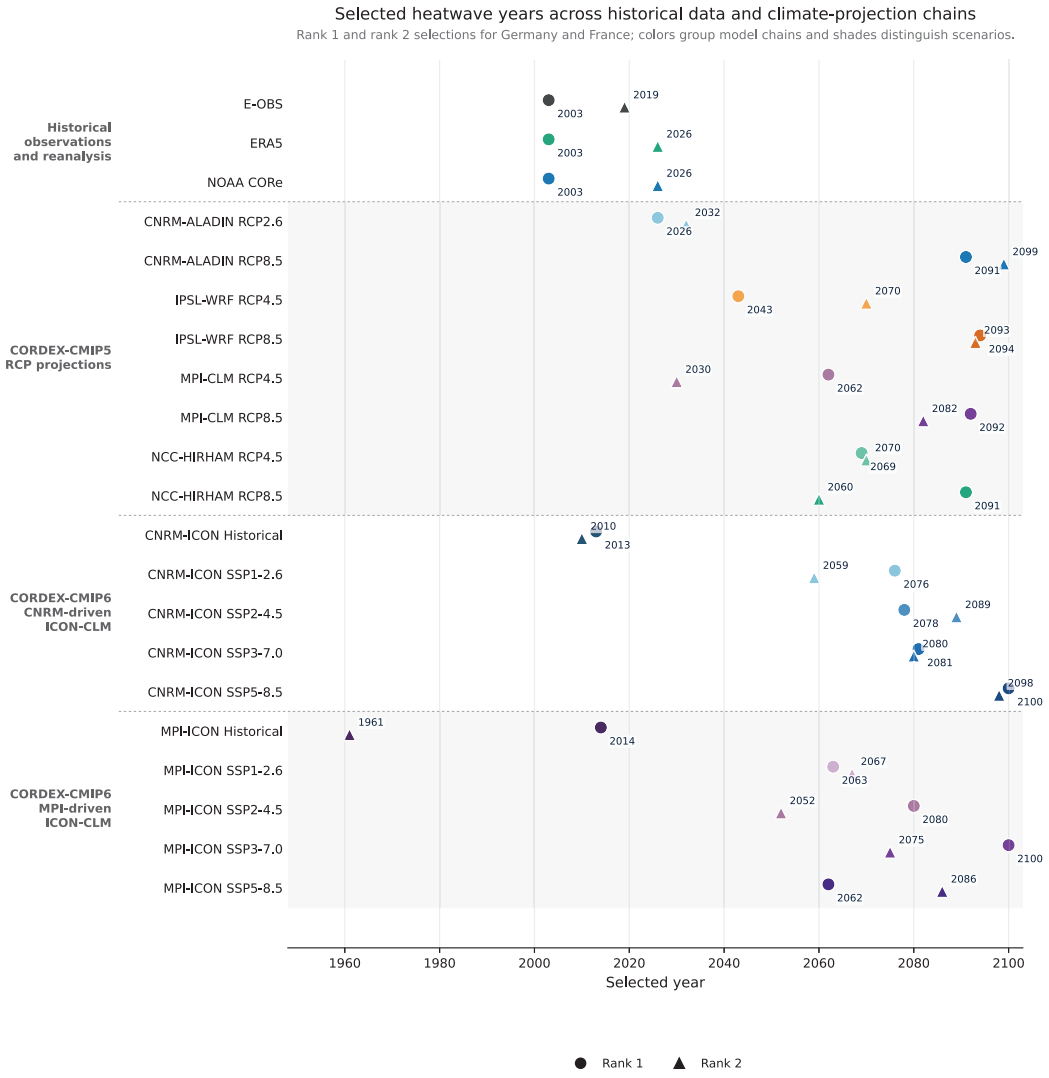


Fig. 7. Selected heatwave years across historical data and CORDEX-CMIP5/CORDEX-CMIP6 projection chains. Points show the first- and second-ranked years selected by the same Germany-France HWMId ranking workflow. Circles indicate rank 1 and triangles indicate rank 2. Colours group model chains and shades distinguish scenarios. The figure illustrates that the workflow is transferable across data products and projection ensembles, while selected calendar years remain projection-chain and scenario specific.

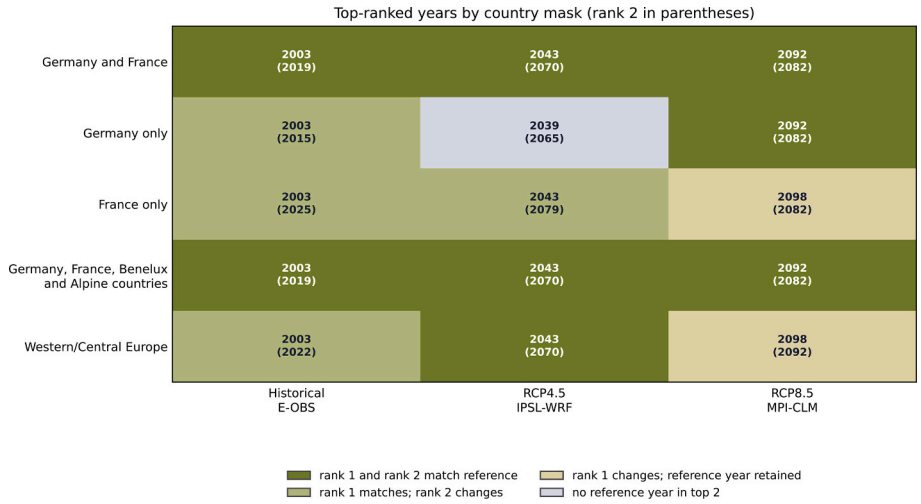


Fig. 8. Sensitivity of the top-ranked heatwave years to the selected regional domain. Cells show the highest-ranked year for each country mask and data basis, with the second-ranked year in parentheses. Colours indicate whether both selected years match the Germany-France reference, only rank 1 matches, rank 1 changes but one reference year remains among the top two, or neither reference year remains among the top two.

Central European domain selects 2098/2092. In the N-1 sensitivity, 2098 remains rank 1 in all leave-one-country-out variants except when Spain is omitted, where 2092 becomes rank 1. The high-end scenario should therefore be interpreted as model-chain and system-perimeter dependent.

After testing sensitivity to the regional domain, we assess whether the ranking changes when country-level heatwave severity is weighted by installed electricity-generation and storage capacity. For the historical *E*-OBS case, 2003 remains the highest-ranked year under all representative TYNDP 2024 weightings; 2019 remains second except under the thermal-plus-nuclear weighting, which selects 2022. For RCP4.5/IPSL-WRF, total capacity, renewables, photovoltaics and wind select 2070/2043, battery-plus-pumped-hydro capacity selects 2070/2039, and thermal-plus-nuclear capacity selects 2043/2041. For RCP8.5/MPI-CLM, all weightings retain 2092 as the highest-ranked year, while the second-ranked year is 2098 for total capacity, renewables, photovoltaics and wind, and 2090 for thermal-plus-nuclear and battery-plus-pumped-hydro capacity. The results therefore indicate that capacity weighting generally retains the dominant historical and high-forcing stress years but can change the ordering and the selection of secondary years, particularly for the RCP4.5 case (Fig. 9). These technology-weighted results are not used to redefine the baseline scenario because they are country-level capacity proxies from one TYNDP scenario year. They are nevertheless useful for downstream energy-system interpretation: a heatwave year that is most stressful for the spatial footprint of solar or wind capacity need not be identical to the year selected by an unweighted Germany-France HWMId sum.

As an additional exposure-oriented sensitivity, we also rerank the Germany-France years with a population-weighted HWMId mean. The historical WorldPop weighting retains E-OBS 2003/2019. For future projections, every fixed-2040 ISIMIP2b population distribution from SSP1 to SSP5 retains 2043/2070 for RCP4.5/IPSL-WRF and 2092/2082 for RCP8.5/MPI-CLM (Appendix B).

5.3. Ranking-criterion sensitivity

For the selected Germany-France case, the HWMId-based main selection is stable under three closely related aggregation criteria: HWMId sum, area-weighted mean HWMId and unweighted mean HWMId select the same top two years. This confirms the robustness of the selected example, but not of the method across all possible masks or datasets. Criteria that target different physical properties produce different sensitivity years: maximum grid-cell HWMId emphasizes localized extremes, duration emphasizes persistence, and annual T_{max} anomaly highlights warm years that do not necessarily contain the strongest HWMId event. The baseline HWMId sum is therefore retained because it balances intensity and spatial extent and is aligned with the energy-system stress-test objective. Fig. 10 compares the resulting top-two years across the alternative ranking criteria.

5.4. Historical reanalysis data and threshold sensitivity

E-OBS, ERA5 and NOAA CORe all rank 2003 first for the completed 1950–2025 period. With data through 1 July 2026, ERA5 and CORe both rank 2026 second behind 2003. For completed years, pairwise Spearman rank correlations are 0.99 for E-OBS/ERA5, 0.96 for E-OBS/CORe and 0.95 for ERA5/CORe; their top-ten sets share nine, six and seven years, respectively (Fig. E.1).

The E-OBS top two remain 2003/2019 at the 90th, 95th and 99th percentiles. ERA5 retains 2003/2026 at the 90th and 95th percentiles, with score gaps of 1.94% and 0.34%, but reverses the order at the 99th percentile, where 2003 is 41.39% below 2026 (Fig. E.2). Because these scores are deterministic transformations of dependent gridded time series and no stochastic null model is specified, a conventional *p*-value is not directly applicable; separation is assessed through score gaps and sensitivity across definitions and data products.

6. Discussion

The results demonstrate that the proposed workflow provides a transparent scenario-definition method for heatwave stress tests in energy-system modelling. The motivating comparison between 2003 and 2026 shows why such a workflow is needed. Within ERA5, the 2026 heatwave is almost as strong as the 2003 benchmark over Germany and France, but the ranking combines event-based HWMId information on duration and local anomaly with subsequent spatial aggregation rather than on maximum temperature alone. The small ERA5 score gaps at the 90th and 95th percentiles and the reversal at the 99th percentile show that 2003 and 2026 are best treated as closely matched candidates rather than as universally ordered events.

The humidity-aware sensitivity also demonstrates that the term comparable is impact-pathway dependent. Metrics that include

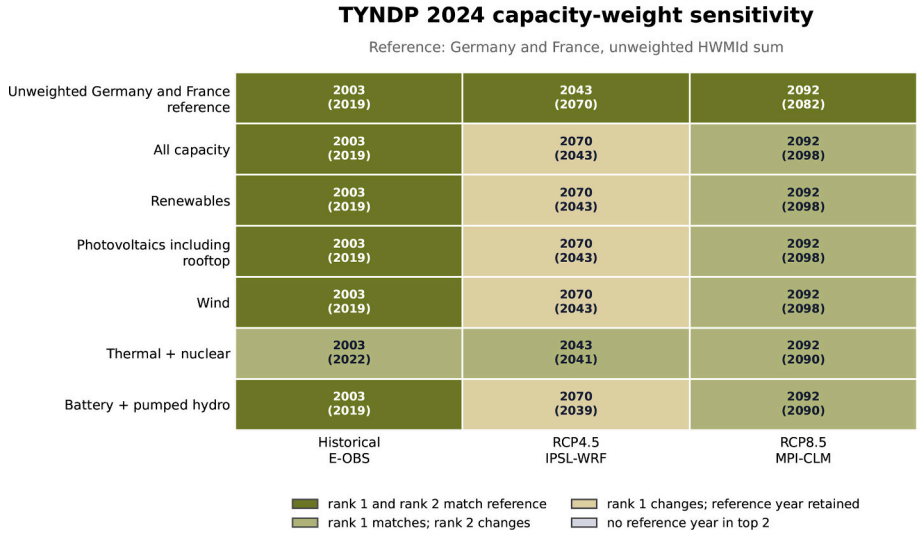


Fig. 9. Representative TYNDP 2024 capacity-weighted heatwave-year sensitivity. Rows compare the unweighted Germany-France reference with representative country-level capacity-weighted sensitivities for the broader Western/Central European domain, based on installed electricity-generation capacities from the TYNDP 2024 National Trends 2040 dataset. For each weighting scheme, country-level capacity shares weight the area-weighted country mean HWMId values. Cells show the highest-ranked year, with the second-ranked year in parentheses; colours indicate whether the top-two years match the unweighted reference ranking.

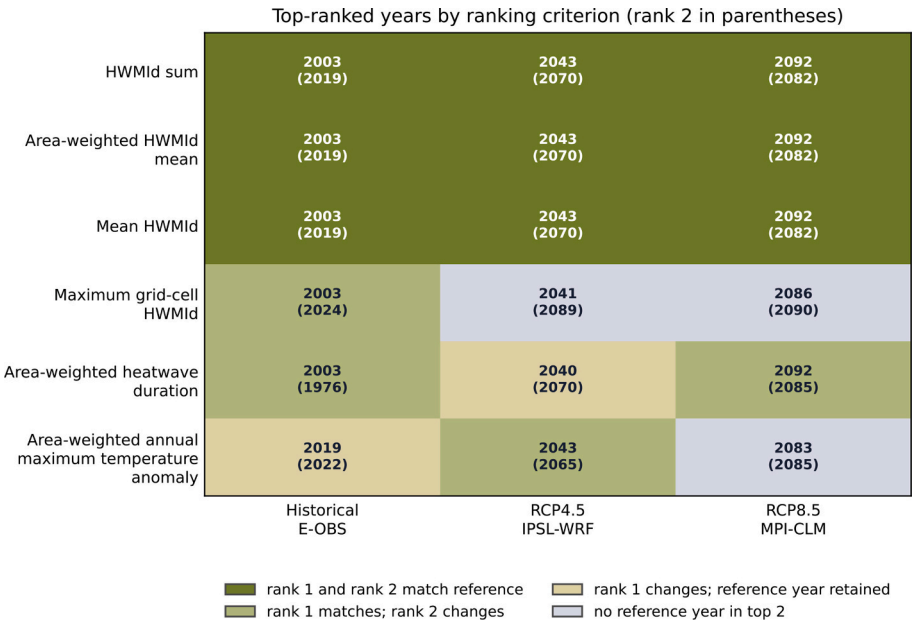


Fig. 10. Sensitivity of selected heatwave years to alternative ranking criteria for Germany and France. Rows compare the baseline summed-HWMId score with area-weighted mean HWMId, unweighted mean HWMId, maximum grid-cell HWMId, area-weighted heatwave duration and area-weighted annual maximum-temperature anomaly. Cells show the highest-ranked year, with the second-ranked year in parentheses. Colours indicate whether both selected years match the baseline HWMId-sum ranking, only rank 1 matches, rank 1 changes but one baseline year remains among the top two, or neither baseline year remains among the top two.

atmospheric moisture may be informative for cooling-demand, human-exposure or evaporative-cooling applications, whereas dry-bulb maximum temperature remains more directly connected to other technical stressors. The alternative metrics therefore complement the HWMId selection for specific applications rather than providing a universally superior ranking (Cvijanovic et al., 2023).

The historical and projection results also show that a heatwave year should not be reported as a generic "RCP4.5", "RCP8.5" or "SSP" year. It is always the result of a specific data product, model chain, spatial domain, reference period and ranking criterion. This is especially important for downstream energy-system studies, where a selected weather year may influence electricity demand, renewable generation, thermal plant availability, storage operation and market outcomes.

The comparison between CORDEX-CMIP5 and CORDEX-CMIP6 should therefore be interpreted as a demonstration of transferability and projection-chain sensitivity, not as a direct attribution of differences between the two ensemble generations. The workflow can be applied to both data families, but the selected calendar years remain chain-specific. This is visible in the high-forcing cases: CNRM-ICON SSP5–8.5 selects end-century years, whereas MPI-ICON SSP5–8.5 selects 2062 and 2086. The selected year reflects both the climate-change signal and internal variability within the respective model chain.

For practical energy-system applications, the main value of the workflow is not to replace simulations of many weather years where those are feasible. Its value is to define a small, traceable set of stress-test years before running computationally expensive or highly detailed model studies. This is useful when full multi-decadal weather-year ensembles would require strong model simplifications, repeated greenfield optimizations or large scenario matrices. The selected years should therefore be understood as transparent stress-test inputs, not as probabilistic estimates of future heatwave occurrence.

The sensitivity analyses indicate how the workflow can be adapted to different modelling questions. A purely meteorological Germany-France stress test supports 2043 as the primary RCP4.5/IPSL-WRF year. A broader or capacity-weighted energy-system perspective may shift the emphasis toward 2070 or other secondary years. This is not a weakness of the method; it makes explicit that the "most relevant" heatwave year depends on the system boundary and the stress mechanism being tested. The recommended practice is therefore to report the baseline meteorological selection together with a small number of documented sensitivity years.

7. Limitations and appropriate use

The workflow is designed to define internally consistent heatwave stress-test years. The selected years therefore depend on the specific climate data product, emission pathway, bias correction and GCM/RCM model chain used in the analysis. Accordingly, a reported scenario year should not be interpreted as a generic RCP4.5, RCP8.5 or SSP heatwave year, but as the outcome of a defined ranking procedure applied to a specific model chain and spatial domain. HWMId values should be interpreted in the same way: they are relative ranking scores within the chosen dataset and reference-period definition.

The heatwaves can be compared directly for the 2003 and 2026 ERA5 event-period comparison because the data product and aggregation setup are identical, but they should not be used as exact physical magnitude ratios across *E*-OBS, ERA5, NOAA CORe, CORDEX-CMIP5 and CORDEX-CMIP6. The humidity-aware extension is exploratory. Wet-bulb temperature is estimated with the standard-sea-level-pressure approximation of Stull (2011), and the shade-WBGT proxy assumes no solar load and uses psychrometric wet-bulb temperature in place of a measured natural wet-bulb temperature. It is therefore not a full outdoor WBGT calculation. Applying the HWMId event and ranking logic to these alternative daily variables produces study-specific sensitivity scores, not standardized heatwave indices or direct estimates of electricity demand, health outcomes or technical derating.

The Germany-France domain reflects the motivating cross-border energy-system use case of this study. Applying the method to another model perimeter may change the ranking and its stability, especially if the spatial extent of the heatwave differs across countries. The same applies to the capacity-weighted sensitivity checks. These weights approximate the spatial relevance of future energy-system assets using one TYNDP 2024 National Trends 2040 dataset, but they do not represent detailed plant locations, hourly generation profiles or weather-dependent availability. They should therefore be understood as diagnostic checks on the meteorological ranking rather than as a general

energy-system weighting scheme.

8. Conclusions

This study presents a transparent HWMId-based workflow for selecting concrete heatwave years for energy-system stress-test scenarios and for comparing recent events with historical benchmarks. Applied to Germany and France, the workflow identifies 2003 as the historical reference year and, at the primary threshold, places June 2026 closely behind it in ERA5 and CORe; alternative thresholds and humidity-aware metrics can reverse the exact ordering. For projected climate conditions, it identifies 2043 under RCP4.5/IPSL-WRF as the primary future stress-test year, 2070 under the same chain as an additional future sensitivity case, and 2092 under RCP8.5/MPI-CLM as a high-impact late-century sensitivity case. The CORDEX-CMIP5 and CORDEX-CMIP6 comparison demonstrates that selected future years remain projection-chain specific. We therefore recommend that heatwave stress-test years are always reported together with the data product, climate-model chain, spatial domain, ranking criterion, weighting choice and HWMId-score interpretation used to select them.

The workflow is released as open-source code, with configurable methodological choices including the heatwave index, regional definition and weighting scheme, so that it can be reproduced and adapted to other applications. Future research could extend the approach by testing additional heatwave and heat-stress metrics, developing alternatives for explicitly spatially coherent event identification, and assessing how different weather-year selections propagate into downstream energy-system outcomes such as demand, generation availability, storage operation, network flows and supply adequacy.

CRediT authorship contribution statement

Johannes Schuhmacher: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stephanie Gutmayer:** Writing – review & editing, Writing – original draft, Conceptualization. **Max Kleinebrahm:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Data curation. **Armin Ardone:** Writing – review & editing, Supervision, Funding acquisition. **Wolf Fichtner:** Writing – review & editing, Supervision, Funding acquisition.

Code availability

The accompanying Python workflow has been prepared as a small package with configuration-driven runs, tests and publication checks. The public repository is released under the MIT License at https://doi.org/10.5281/zenodo.23081776 and https://github.com/johannesschuhmacher/heatwave_definition.git.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the E-OBS dataset from the EU-FP6 project UERRA, the Copernicus Climate Change Service, and the data providers in the ECA&D project. The authors also acknowledge the Copernicus Climate Change Service climate data products and the ENTSO-E and ENTSOG TYNDP Scenarios data used in the scenario workflow.

This work was supported by the Helmholtz Association as part of the project “Helmholtz platform for the design of robust energy systems and their supply chains” (RESUR).

This work was also supported within the scope of energy research by the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag under grant number 03EI1123A.

Appendix A. Data used

Table A.1 summarizes the data basis used in this study.

Table A.1
Data products and their role in the scenario-definition workflow.

Data product	Period used	Role in this study
E-OBS v33.0e daily Tmax (tx)	1950–2025	Historical ranking and plausibility check against known European heatwaves.
ERA5 hourly 2 m temperature (t2m) and dewpoint temperature (d2m)	1950–2026	Reanalysis-based comparison, 2003/2026 benchmark assessment and exploratory humidity-aware sensitivity.
NOAA CORe daily maximum 2 m temperature	1950–2026	Independent historical reanalysis sensitivity and 2003/2026 ranking check.
CORDEX-CMIP5/RCP tasAdjust	to 2100	Primary future scenario selection and raw-data ensemble sensitivity.
CORDEX-CMIP6/SSP ICON-CLM tas	historical 1950–2014; SSP periods from 2015 onward	Workflow transferability check and additional projection-chain sensitivity.
TYNDP 2024 National Trends 2040 capacities	2040	Country-level electricity-generation-capacity weighting sensitivity.
WorldPoP 2020 and ISIMIP2b SSP1-SSP5 population	2020; 2040	Cell-level historical and future population-weighting sensitivities, respectively.

Appendix B. Population-weighted sensitivity

To test population exposure sensitivity, we apply cell-level population weights. Historical *E*-OBS uses WorldPop, 2020 counts, whereas future projections use ISIMIP2b population for a fixed 2040 planning year under SSP1-SSP5 (Jones and O'Neill, 2016; Piontek and Geiger, 2017; WorldPop, 2020). The population year is independent of the selected weather-year label. Population is remapped to the HWMId grid and normalized before the annual grid-cell HWMId values are aggregated. The resulting score is

$$S_y^{pop} = \frac{\sum_g P_g * H_{g,y}}{\sum_g P_g}, \qquad \forall y \tag{B.1}$$

where P_g is the population assigned to grid cell g and $H_{g,y}$ is the annual HWMId value of that grid cell in year y.

The historical WorldPop weighting retains E-OBS 2003/2019. All five fixed-2040 SSP population distributions retain 2043/2070 for RCP4.5/IPSL-WRF and 2092/2082 for RCP8.5/MPI-CLM. The scores vary slightly, but the selected years do not change (Fig. B.1). Population weighting is treated as a separate exposure sensitivity, because cooling demand also depends on air-conditioning, buildings and adaptation.

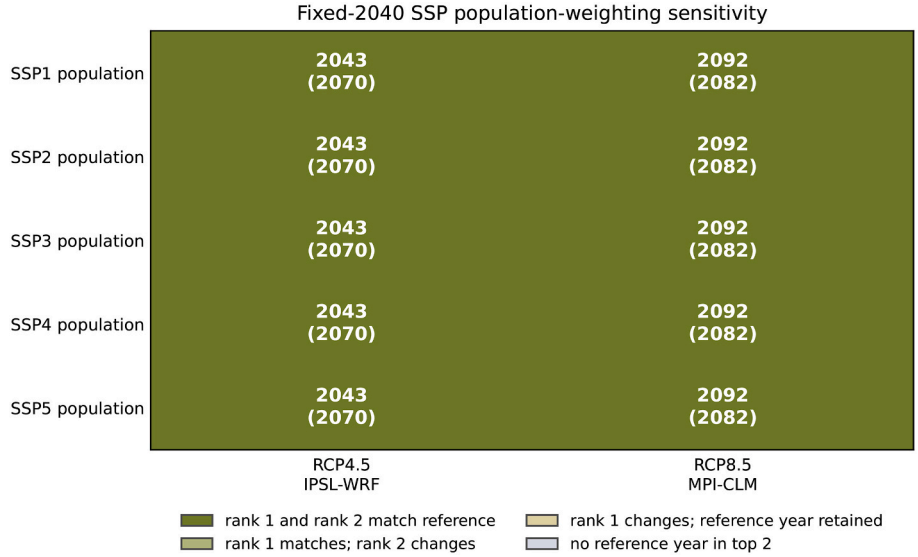


Fig. B.1. Fixed-2040 SSP population-weighting sensitivity for the two primary future projection chains. Rows apply cell-level ISIMIP2b population weights for SSP1-SSP5. Cells show the highest-ranked year, with the second-ranked year in parentheses. Green indicates that both years match the unweighted reference. The population year is fixed at 2040 and is independent of the selected weather-year label. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix C. N-1 country sensitivity

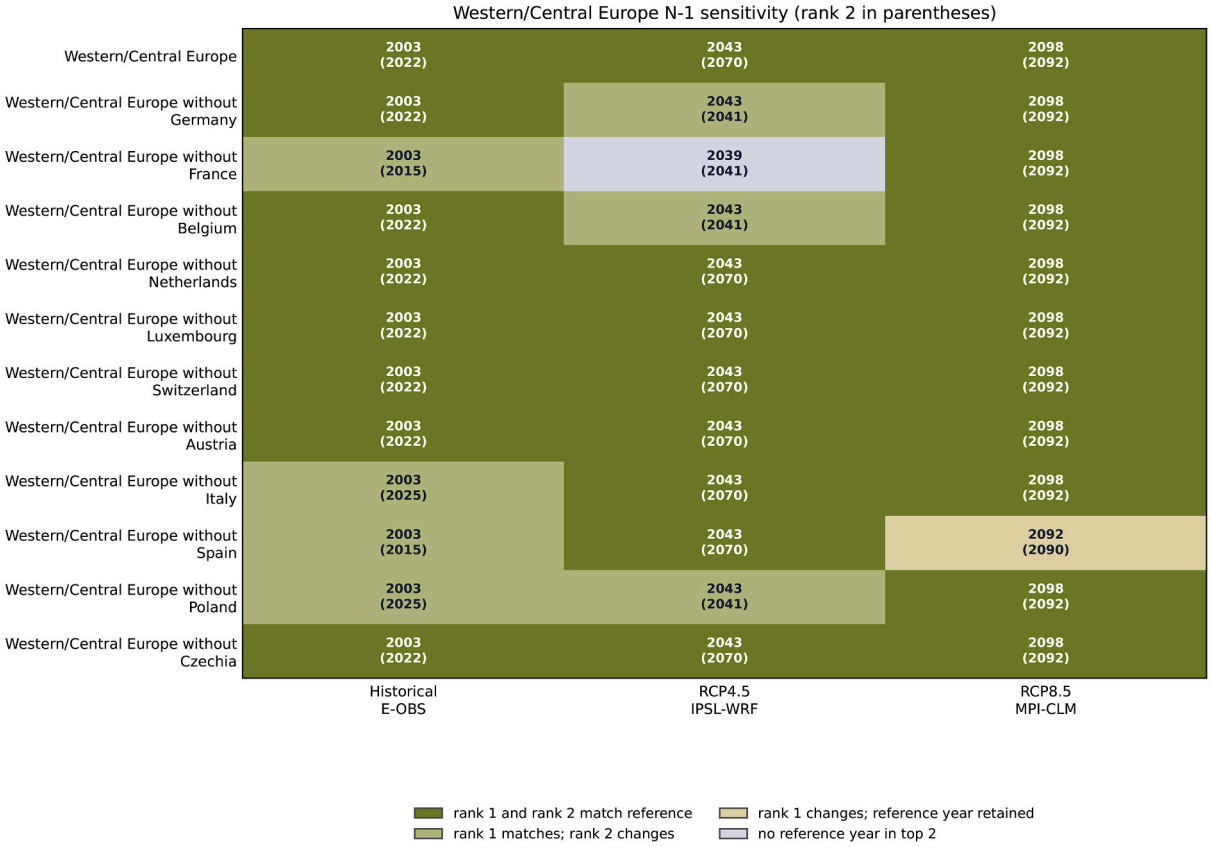


Fig. C.1. Western/Central European N-1 sensitivity of the heatwave-year ranking. Each row omits one country from the full Western/Central European domain before aggregating annual HWMId values. Cells show the highest-ranked year, with the second-ranked year in parentheses. Colours indicate whether both selected years match the full-domain reference, only rank 1 matches, rank 1 changes but one reference year remains among the top two, or neither reference year remains among the top two. This sensitivity test identifies whether the regional ranking is driven by any single country.

Appendix D. Humidity-aware metric sensitivity

Humidity can influence human thermal stress, cooling demand and the performance of evaporative cooling. We therefore test whether the ERA5 Germany–France ranking changes when daily maximum air temperature is replaced by three humidity-aware variables: Humidex, psychrometric wet-bulb temperature and a WBGT-inspired shade proxy.

Hourly 2-m air temperature T and 2-m dewpoint temperature T_d are taken from ERA5. Relative humidity is derived from T and T_d using the Magnus-form approximation for saturation vapour pressure with the coefficients recommended by Alduchov and Eskridge (1996):

$$RH = 100 * \frac{\exp\left(\frac{17.625\ T_d}{243.04+T_d}\right)}{\exp\left(\frac{17.625\ T}{243.04+T}\right)} \tag{D.1}$$

where RH is expressed in percent and T and T_d are in degrees Celsius.

Humidex is calculated from air temperature and dewpoint-derived vapour pressure following the formulation introduced by Masterton and Richardson (1979) and used in subsequent climate applications (e.g. Chakraborty et al., 2022; Diaconescu et al., 2023):

$$Hu = T + 0.5555(e - 10)\ with\ e = 6.11\ exp\left(5417.753\left(\frac{1}{273.15} - \frac{1}{T_{d,k}}\right)\right) \tag{D.2}$$

where e is vapour pressure in hPa and $T_{d,k} = 273.15 + T_d$ is dewpoint temperature in Kelvin.

Psychrometric wet-bulb temperature is estimated from air temperature and relative humidity using the empirical approximation of Stull (2011):

$$T_w = T \cdot arctan\left[0.151977(RH + 8.313659)^{\frac{1}{2}}\right] + \arctan(T + RH) - \arctan(RH - 1.676331) + 0.00391838\ RH^{3/2}\arctan(0.023101\ RH) - 4.686035 \tag{D.3}$$

The approximation was derived for standard sea-level pressure and is applicable for air temperatures from −20 to 50 °C and relative humidity from 5% to 99%, except for combinations of low temperature and low humidity. Across its stated valid range, Stull (2011) reports errors between approximately −1 and +0.65 °C and a mean absolute error below 0.3 °C.

As a third humidity-aware sensitivity metric, we calculate a WBGT-inspired shade proxy as

$$WBGT_{shade} = 0.7\ T_w + 0.3\ T \tag{D.4}$$

The same weighting is commonly used for simplified WBGT calculations under radiation-free conditions, for which globe temperature is assumed to equal air temperature (e.g. Schwingshackl et al., 2024). Here, however, T_w is the psychrometric wet-bulb temperature estimated with Eq. (D.3), rather than the natural wet-bulb temperature required for a full outdoor WBGT calculation. $WBGT_{shade}$ is therefore used only as a humidity-aware proxy and should not be interpreted as full outdoor WBGT, which additionally depends on radiative and wind conditions and requires explicit natural wet-bulb and globe-temperature calculations (Budd, 2008; Kong and Huber, 2022).

For each variable, hourly values are reduced to daily maxima. The local 90th-percentile thresholds and reference quartiles are then recalculated separately for each metric over the 1981–2010 reference period. The same event-detection and ranking procedure as in the temperature-based analysis is subsequently applied: a minimum event duration of three consecutive days, selection of the strongest annual event at each grid cell, and spatial aggregation over the Germany-France domain. Because each metric has its own physical scale, threshold distribution and normalization, absolute event scores are compared only within the same metric; comparisons between metrics are based on ranks, rank correlations and top-ten overlap.

The completed-year comparison retains 2003 at rank 1 for temperature-based HWMId, Humidex and the shade-WBGT proxy, but places it at rank 4 for wet-bulb temperature. When ERA5 data through 1 July 2026 are included, 2026 ranks second under HWMId, reaching 98.1% of the 2003 score, but ranks first under all three humidity-aware variants. Within the respective metric, the 2026 score is 146.8% of the 2003 Humidex score, 221.0% of the 2003 wet-bulb score and 151.0% of the 2003 shade-WBGT score. Spearman rank correlations with the temperature-based ordering are 0.91 for Humidex, 0.69 for wet-bulb temperature and 0.91 for the shade-WBGT proxy. Fig. D.1 therefore indicates that the June 2026 event was substantially more exceptional when humidity was included than when it was characterized by dry-bulb temperature alone.

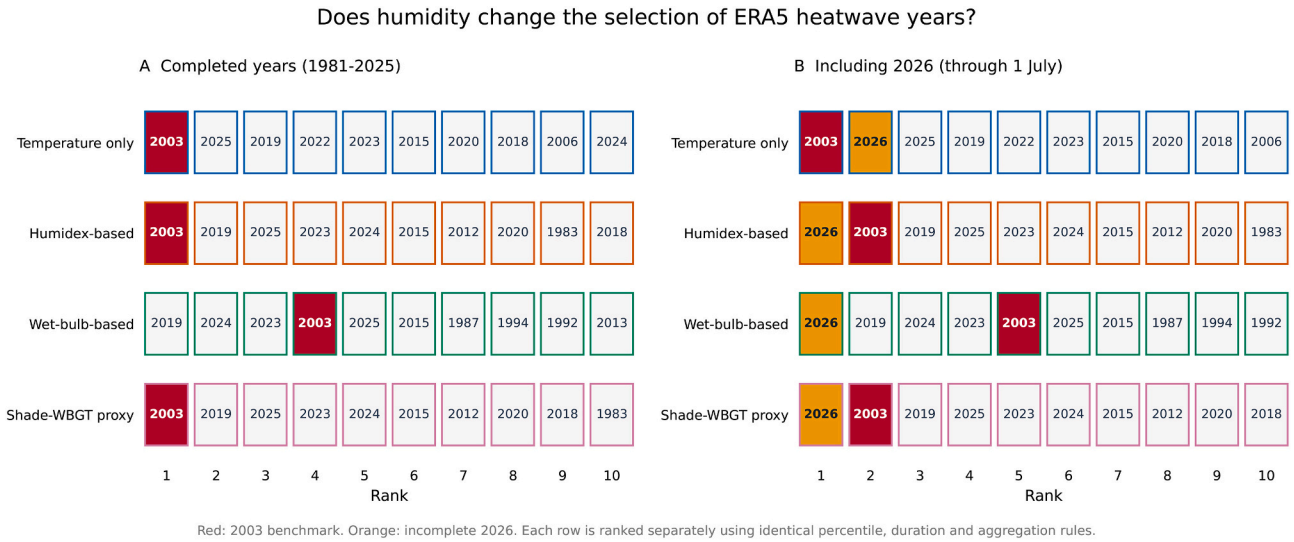


Fig. D.1. Humidity-aware sensitivity of the ERA5 Germany-France heatwave-year ranking. Rows show the ten highest-ranked years for temperature-only HWMId, Humidex, wet-bulb temperature and a shade-WBGT proxy. Panel A includes completed years from 1981 to 2025; Panel B additionally includes 2026 through 1 July. Each row uses locally recalculated thresholds and reference quartiles while retaining the same 1981–2010 reference period, 90th-percentile threshold, minimum three-day duration, strongest-event rule and spatial aggregation. Red identifies 2003 and orange identifies the incomplete 2026 year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix E. Reanalysis data and threshold sensitivity

The historical-product comparison retains each product on its native grid and therefore compares ranks rather than absolute HWMId sums.

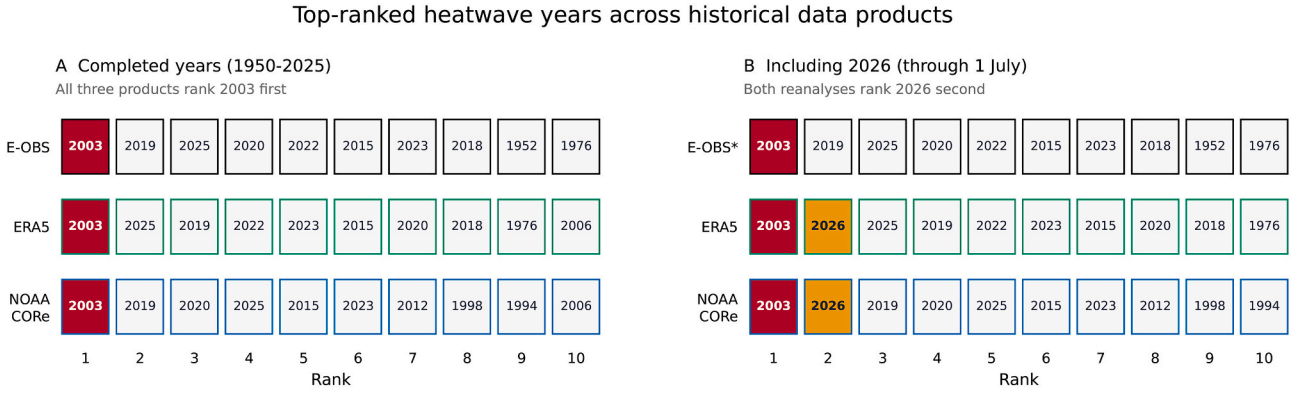


Fig. E.1. Top-ten heatwave-year rankings over Germany and France for *E*-OBS, ERA5 and NOAA CORe. Panel A contains completed years from 1950 to 2025. Panel B additionally includes ERA5 and CORe data through 1 July 2026; E-OBS remains available through 2025. CORe lacks 19 May 2026, which is outside the selected June event. Red identifies 2003 and orange the incomplete 2026 year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

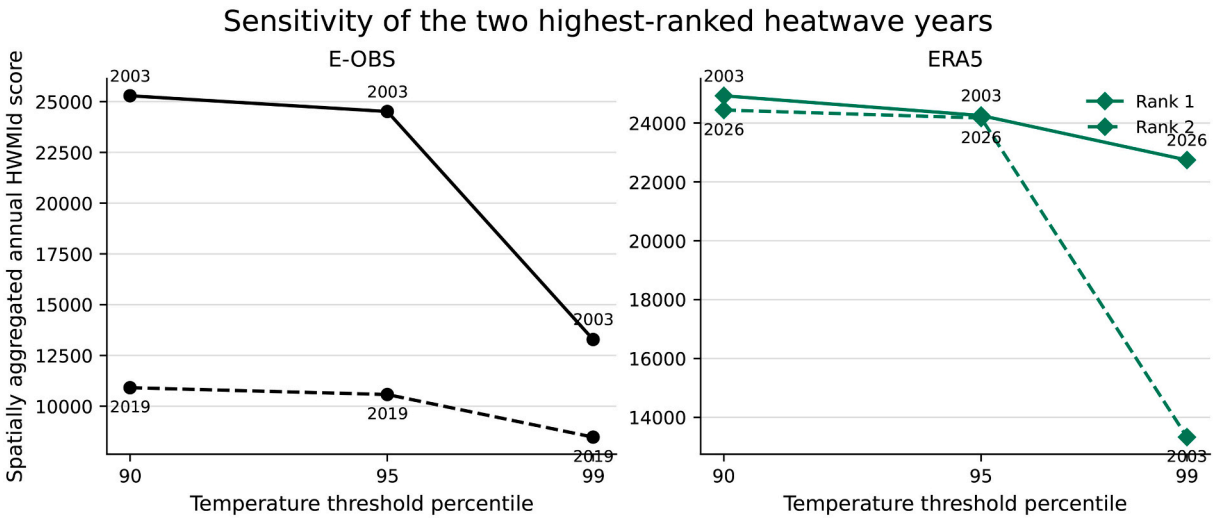


Fig. E.2. Threshold sensitivity of the two highest-ranked Germany-France heatwave years. The E-OBS and ERA5 calculations use 90th-, 95th- and 99th-percentile thresholds while retaining the 1981–2010 reference period, minimum three-day duration and spatial aggregation. Solid lines denote rank 1 and dashed lines rank 2; labels identify the selected years. ERA5 includes data through 1 July 2026, whereas E-OBS contains completed years through 2025.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT by OpenAI and Codex by OpenAI to support manuscript drafting, linguistic refinement, grammar checking, and code development. All AI-assisted outputs were subsequently subjected to critical review, validation, and editorial revision by the authors. The authors take full responsibility for the accuracy, integrity, originality, and final content of the published article.

Data availability

The underlying E-OBS, ERA5, NOAA CORe, CORDEX-CMIP5, CORDEX-CMIP6, WorldPop, ISIMIP2b and TYNDP 2024 data are available from their official providers and are not redistributed in the public repository.

References

Abdin, A.F., Fang, Y.-P., Zio, E., 2019. A modeling and optimization framework for power systems design with operational flexibility and resilience against extreme heat waves and drought events. Renew. Sust. Energ. Rev. 112, 706–719. https://doi.org/10.1016/j.rser.2019.06.006.

Alduchov, O.A., Eskridge, R.E., 1996. Improved Magnus form approximation of saturation vapor pressure. J. Appl. Meteorol. Climatol. 35 (4), 601–609. https://doi.org/10.1175/1520-0450(1996)035%3C0601:IMFAOS%3E2.0.CO;2.

Antonini, E.G.A., Di Bella, A., Savelli, I., Drouet, L., Tavoni, M., 2024. Weather- and climate-driven power supply and demand time series for power and energy system analyses. Sci. Data 11 (1), 1324. https://doi.org/10.1038/s41597-024-04129-8.

Becker, F.N., Fink, A.H., Bissolli, P., Pinto, J.G., 2022. Towards a more comprehensive assessment of the intensity of historical European heat waves (1979–2019). Atmos. Sci. Lett. 23 (11), e1120. https://doi.org/10.1002/asl.1120.

Błażejczyk, K., Jendritzky, G., Bröde, P., Fiala, D., Havenith, G., Epstein, Y., Psikuta, A., Kampmann, B., 2013. An introduction to the universal thermal climate index (UTCI). Geogr. Pol. 86 (1), 5–10. https://doi.org/10.7163/GPol.2013.1.

Breteau, P., Garric, A., Imbach, R., & Prati, L. (2026, July 3). Four graphs to visualize the unprecedented scale of France's June 2026 heatwave. Le Monde. https://www.lemonde.fr/en/les-decodeurs/article/2026/07/03/four-graphs-to-visualize-the-unprecedented-scale-of-france-s-june-2026-heatwave_6755104_8.html.

Budd, G.M., 2008. Wet-bulb globe temperature (WBGT)—its history and its limitations. J. Sci. Med. Sport 11 (1), 20–32. https://doi.org/10.1016/j.jsams.2007.07.003.

Bunting, E.L., Tolmanov, V., Keellings, D., 2024. What is a heat wave: a survey and literature synthesis of heat wave definitions across the United States. PLOS Climate 3 (9), e0000468. https://doi.org/10.1371/journal.pclm.0000468.

Buzan, J.R., 2024. Implementation and evaluation of wet bulb globe temperature within non-urban environments in the community land model version 5. J. Adv. Model. Earth Syst. 16 (2), e2023MS003704. https://doi.org/10.1029/2023MS003704.

Chakraborty, T., Venter, Z.S., Qian, Y., Lee, X., 2022. Lower urban humidity moderates outdoor heat stress. AGU Adv. 3 (5), e2022AV000729. https://doi.org/10.1029/2022AV000729.

Copernicus Climate Change Service, 2021. Climate and energy indicators for Europe from 2005 to 2100 derived from climate projections [data set].. ECMWF. https://doi.org/10.24381/CDS.F6951A62.

C3S. (2018). *ERA5 hourly data on single levels from 1940 to present* [Data set]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Doi: 10.24381/CDS.ADBB2D47.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D., 2018. An ensemble version of the E-OBS temperature and precipitation data sets. J. Geophys. Res. Atmos. 123 (17), 9391–9409. https://doi.org/10.1029/2017JD028200.

Craig, M., Wohland, J., Stoop, L.P., Kies, A., Pickering, B., Bloomfield, H.C., Browell, J., De Felice, M., Dent, C.J., Deroubaix, A., Frischmuth, F., Gonzalez, P.L.M., Grochowicz, A., Gruber, K., Hartel, P., Kittel, M., Kotzur, L., Labuhn, I., Lundquist, J. K., Brayshaw, D.J., 2022. Overcoming the disconnect between energy system and climate modeling. Joule 6 (7), 1405–1417. https://doi.org/10.1016/j.joule.2022.05.010.

Cvijanovic, I., Mistry, M.N., Begg, J.D., Gasparrini, A., Rodó, X., 2023. Importance of humidity for characterization and communication of dangerous heatwave conditions. Npj Clim. Atmos. Sci. 6 (1), 33. https://doi.org/10.1038/s41612-023-00346-x.

Diaconescu, E., Sankare, H., Chow, K., Murdock, T.Q., Cannon, A.J., 2023. A short note on the use of daily climate data to calculate humidex heat-stress indices. Int. J. Climatol. 43 (2), 837–849. https://doi.org/10.1002/joc.7833.

ENTSO-E. (2024, October 30). *ENTSO-E Transparency Platform* [Data Platform]. https://transparency.entsoe.eu/.

ENTSO-E, 2025. *TYNDP* 2024 *Scenarios final package: PEMMDB 2.5 and supply inputs* [Data set]. TYNDP 2024 Scenarios. https://2024.entsos-tyndp-scenarios.eu/download/.

Gonçalves, A.C.R., Costoya, X., Nieto, R., Liberato, M.L.R., 2024. Extreme weather events on energy systems: a comprehensive review on impacts, mitigation, and adaptation measures. Sustain. Energy Res. 11 (1), 4. https://doi.org/10.1186/s40807-023-00097-6.

Gøtske, E.K., Andresen, G.B., Neumann, F., Victoria, M., 2024. Designing a sector-coupled European energy system robust to 60 years of historical weather data. Nat. Commun. 15 (1), 10680. https://doi.org/10.1038/s41467-024-54853-3.

Grochowicz, A., van Greevenbroek, K., Bloomfield, H.C., 2024. Using power system modelling outputs to identify weather-induced extreme events in highly renewable systems. Environ. Res. Lett. 19 (5), 054038. https://doi.org/10.1088/1748-9326/ad374a.

Hoffmann, M., Schyska, B.U., Bartels, J., Pelser, T., Behrens, J., Wetzel, M., Gils, H.C., Tang, C.-F., Tillmanns, M., Stock, J., Xhonneux, A., Kotzur, L., Praktiknjo, A., Vogt, T., Jochem, P., Linßen, J., Weinand, J.M., Stolten, D., 2024. A review of mixed-integer linear formulations for framework-based energy system models. Adv. Appl. Energy 16, 100190. https://doi.org/10.1016/j.adapen.2024.100190.

Jones, B., O'Neill, B.C., 2016. Spatially explicit global population scenarios consistent with the shared socioeconomic pathways. Environ. Res. Lett. 11 (8), 084003. https://doi.org/10.1088/1748-9326/11/8/084003.
Kabir, S.M.R., Rahman, M., Zisha, F.K., Meng, L., 2026. Quantifying event-based heatwave-induced power outage risk: a multi-year spatiotemporal analysis in Texas. Sustainability 18 (12). https://doi.org/10.3390/su18126205.
Ke, X., Wu, D., Rice, J., Kintner-Meyer, M., Lu, N., 2016. Quantifying impacts of heat waves on power grid operation. Appl. Energy 183, 504–512. https://doi.org/10.1016/j.apenergy.2016.08.188.
Kong, Q., Huber, M., 2022. Explicit calculations of wet-bulb globe temperature compared with approximations and why it matters for labor productivity. *Earth's* Future 10 (3), e2021EF002334. https://doi.org/10.1029/2021EF002334.
Lin, C., Kjellström, E., Wilcke, R.A.I., Chen, D., 2022. Present and future European heat wave magnitudes: Climatologies, trends, and their associated uncertainties in GCM-RCM model chains. Earth Syst. Dynam. 13 (3), 1197–1214. https://doi.org/10.5194/esd-13-1197-2022.
Masterton, J.M., Richardson, F.A., 1979. In: Environment Canada, Atmos. Environ (Ed.), Humidex: a method of quantifying human discomfort due to excessive heat and humidity. https://publications.gc.ca/site/eng/9.865813/publication.html.
Mohr, S., Benz, S.A., Böhnke, D., Cakmak, H., Cermak, J., Cheng, H., Dillerup, I.M., Ehret, U., Fink, G., Fischer, J., Ho, S., Hoffmann, J., Holmberg, E., Klimiuk, T., Lemburg, A., Ludwig, P., Mühr, B., Pinto, J.G., Rosenberg, S., Kunz, M., 2026. CEDIM forensic disaster analysis (FDA) group: heat wave June 2026: Western and Central Europe. https://doi.org/10.5445/IR/1000195086.
Molina, M.O., Gutiérrez, C., Ortega, M., Sánchez, E., 2023. Summer heatwaves, wind production and electricity demand in southern Europe: climatic conditions and impacts. Environ. Res. Commun. 5 (8), 085005. https://doi.org/10.1088/2515-7620/acec37.
Nairn, J.R., Fawcett, R.J.B., 2015. The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity. Int. J. Environ. Res. Public Health 12 (1), 227–253. https://doi.org/10.3390/ijerph120100227.
National Oceanic & Atmospheric Administration, 2026. Climate Prediction Center. NOAA - Conventional Observation Reanalysis (CORe). . NSF National Center for Atmospheric Research. https://gdex.ucar.edu/datasets/d090004.
Nikulin, G., Buntemeyer, L., Fernández, J., McGinnis, S., Evans, J.P., 2026. CORDEX-CMIP6 archiving specifications for dynamical downscaling. https://doi.org/10.5281/ZENODO.20274448.
Perkins, S.E., Alexander, L.V., 2013. On the measurement of heat waves. J. Clim. 26 (13), 4500–4517. https://doi.org/10.1175/JCLI-D-12-00383.1.
Piontek, F., Geiger, T., 2017. *ISIMIP2b secondary population input data* (version 1.0) [data set]. . ISIMIP repository.. https://doi.org/10.48364/ISIMIP.432399.
Rübbelke, D., Vögele, S., 2011. Impacts of climate change on European critical infrastructures: the case of the power sector. Environ. Sci. Policy 14 (1), 53–63. https://doi.org/10.1016/j.envsci.2010.10.007.
Russo, S., Dosio, A., Graversen, R.G., Sillmann, J., Carrao, H., Dunbar, M.B., Singleton, A., Montagna, P., Barbola, P., Vogt, J.V., 2014. Magnitude of extreme heat waves in present climate and their projection in a warming world. J. Geophys. Res. Atmos. 119 (22). https://doi.org/10.1002/2014JD022098.
Russo, S., Sillmann, J., Fischer, E.M., 2015. Top ten European heatwaves since 1950 and their occurrence in the coming decades. Environ. Res. Lett. 10 (12), 124003. https://doi.org/10.1088/1748-9326/10/12/124003.
Saki, S., Sofia, G., Kar, B., Anagnostou, E., 2025. A multi-year analysis of the impact of heatwaves and compound weather events on power outages. Sci. Rep. 15 (1), 30846. https://doi.org/10.1038/s41598-025-15065-x.
Schwingshackl, C., Daloz, A.S., Iles, C., Aunan, K., Sillmann, J., 2024. High-resolution projections of ambient heat for major European cities using different heat metrics. Nat. Hazards Earth Syst. Sci. 24 (1), 331–354. https://doi.org/10.5194/nhess-24-331-2024.
Stull, R., 2011. Wet-bulb temperature from relative humidity and air temperature. J. Appl. Meteorol. Climatol. 50 (11), 2267–2269. https://doi.org/10.1175/JAMC-D-11-0143.1.
University of Freiburg. (n.d.). Weather stations measure record-breaking temperatures in Freiburg City. Retrieved 10 July 2026, from https://uni-freiburg.de/en/weather-stations-measure-record-breaking-temperatures-in-freiburg-city/.
van Vliet, M.T.H., Vögele, S., Rübbelke, D., 2013. Water constraints on European power supply under climate change: impacts on electricity prices. Environ. Res. Lett. 8 (3), 035010. https://doi.org/10.1088/1748-9326/8/3/035010.
Wang, J., Yan, Z., 2021. Rapid rises in the magnitude and risk of extreme regional heat wave events in China. Weather Clim. Extrem. 34, 100379. https://doi.org/10.1016/j.wace.2021.100379.
WorldPop, 2020. *Global 1km population total adjusted to match the corresponding UNPD estimate* [data set]. University of Southampton. https://doi.org/10.5258/SOTON/WP00671.
You, J., Yin, F., Gao, L., 2025. Escalating wind power shortages during heatwaves. Commun. Earth Environ. 6 (1), 245. https://doi.org/10.1038/s43247-025-02239-8.
Zhang, T., Deng, G., Liu, X., He, Y., Chen, Q., 2026. Comprehensive heatwave magnitude index considering the effect of indicator equilibrium. Atmos. Res. 330, 108591. https://doi.org/10.1016/j.atmosres.2025.108591.
Zschenderlein, P., Fink, A.H., Pfahl, S., Wernli, H., 2019. Processes determining heat waves across different European climates. Q. J. R. Meteorol. Soc. 145 (724), 2973–2989. https://doi.org/10.1002/qj.3599.