ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
20 February 2026

REVISED
9 July 2026

ACCEPTED FOR PUBLICATION
20 August 2026

PUBLISHED
18 September 2026

The variability atlas: how internal climate variability affects the estimation of climate extreme indices

Lukas Brunner[1,*], **Vidur Mithal**[1], **Leonard F Borchert**[1], **Jana Sillmann**[1] and **Benjamin Poschlod**[1,2]

[1] Research Unit Sustainability and Climate Risk, Earth and Society Research Hub (ESRAH), University of Hamburg, Hamburg, Germany
[2] Institute for Global Water Security, Hamburg University of Technology, Hamburg, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: lukas.brunner@uni-hamburg.de

Keywords: internal climate variability, climate extremes, climate model

Supplementary material for this article is available online

Abstract

Robustly quantifying climate extremes is crucial for understanding their occurrence, reliable impact assessments, and efficient adaptation planning. To allow objective and comparable assessments of various climate extremes across regions, time periods, and datasets, standardized indices like those from the Expert Team on Climate Change Detection and Indices (ETCCDI) have been developed. These indices are routinely used for climate impact assessment. However, the occurrence of extremes depends not only on the climate state but also on the manifestation of internal climate variability, which is often not explicitly considered in such impact assessments. Here, we complement existing work on climate extremes by isolating the effect of internal climate variability on the estimation of extremes and raise awareness of potential implications for the community. We use a 50-member ensemble from the MPI-ESM global climate model to comprehensively map the effect of internal variability on various climate extremes as captured by the temperature- and precipitation-based ETCCDI indices in the period 1995–2014. We investigate the extent to which the assessment of extremes is affected by internal variability and discuss implications for users. These implications depend on many case-specific priorities, including the indices used, the region considered, and the focus on expected state versus worst-case scenario or single observed manifestation. To provide practical guidance, we introduce the *Variability Atlas*, an interactive online interface that allows for the investigation of the effect of internal variability for flexible user-defined cases and without coding experience.

1. Introduction

Climate extremes, such as heat or heavy precipitation, have devastating effects on human society and the environment alike and are expected to increase in frequency and severity due to continuing anthropogenic climate change. To objectively quantify and monitor their occurrence, standardized indices, like those from the Expert Team on Climate Change Detection and Indices (ETCCDI; Alexander *et al* 2006) have been established in the last decades. Such indices provide a widely used benchmark for the analysis of various climate extremes and inform many use cases and impact assessments. However, the effect of internal climate variability on their estimation is rarely considered in such impact assessments, limiting a comprehensive treatment of uncertainty. Here, we use a 50-member ensemble from the state-of-the-art global climate model MPI-ESM (Olonscheck *et al* 2023) to quantify the effect of internal variability on the frequency and severity of various climate extremes as estimated by the ETCCDI indices and to discuss implications for users.

Our work is based on the 26 core indices suggested by the ETCCDI and as described, e.g. in Zhang *et al* (2011). These indices aim to provide a comprehensive quantification of various temperature- and precipitation-based extremes using daily data. While the original indices have been developed for use on point measurements and many studies calculate them from station data (e.g. Keggenhoff *et al* 2014, Menang 2017, Dunn *et al* 2020), others also apply them directly to gridded observations (e.g. Rehana *et al* 2022, Nazrul *et al* 2025, Yate and Ren 2025), and climate model output (e.g. Sillmann *et al* 2013, Squintu *et al* 2021, Ebedi-Nding *et al* 2024). The indices have been used to investigate extremes at different scales from global (Wang *et al* 2017, Fan *et al* 2020, Kim *et al* 2020, Ajjur and Al-Ghamdi 2021, Ma *et al* 2025) and continental (Peterson *et al* 2008, Squintu *et al* 2021, Collazo *et al* 2022), to country level (Keggenhoff *et al* 2014, Menang 2017, Xu *et al* 2019, Rehana *et al* 2022, Getachew *et al* 2025, Nazrul *et al* 2025) and even for individual counties and catchments (Bohat *et al* 2025, Pyarali *et al* 2025, Sa'adi *et al* 2025).

For purely observation-based studies, the treatment of internal variability is fundamentally limited by data availability as (1) only one realization of past climate exists, (2) high-quality data are typically available only for the recent past, and (3) many regions do not provide dense observation networks. This limitation has been partially overcome by the use of climate models, though their representation of internal climate variability varies, depending on the spatiotemporal scale and climate variable, and might be biased against observations (Lee *et al* 2020, von Trentini *et al* 2020). Climate model ensembles can provide simulations spanning many thousands of years, with constant data quality, global coverage, and no missing values due to technical measurement faults. A typical approach in the literature is to draw on a range of different models provided in multi-model ensembles such as CMIP6 (Eyring *et al* 2016), or to use large ensembles of individual models (e.g. Maher *et al* 2025).

Yet, while large ensembles are frequently used to isolate the effect of internal variability and to sample extremes (e.g. Schaller *et al* 2018, Lehner *et al* 2020, Maher *et al* 2020, Milinski *et al* 2020, Gessner *et al* 2021, Blanusa *et al* 2023), their application in the more impact- and application-focused literature drawing on the ETCCDI indices is very limited (e.g. Iles *et al* 2024). Impact-focused studies, rather, draw on the CMIP6 multi-model ensemble and its predecessors (Seo *et al* 2014, Xu *et al* 2019, Fan *et al* 2020, Kim *et al* 2020, Lee *et al* 2020, Squintu *et al* 2021, Ebedi-Nding *et al* 2024, Mousavi *et al* 2024, Nazrul *et al* 2025), or bias-corrected and downscaled climate model output, like from the ISIMIP framework (Hempel *et al* 2013, Lange 2019). Crucially, such multi-model ensembles do not allow isolating the effect of internal variability, which is interwoven with structural model differences (Lehner *et al* 2020). The uncertainty ranges given in many of these studies are, hence, a combination of model uncertainty and internal variability.

To address this, we use a 50-member large ensemble from the CMIP6 version of the MPI-ESM model (Mauritsen *et al* 2019, Olonscheck *et al* 2023), to quantify the effect of internal variability on the climatology of the ETCCDI indices. While earlier work has already pointed to the dominant role internal variability can play for estimates of change in several ETCCDI indices (Tebaldi *et al* 2021, Blanusa *et al* 2023), we focus on the average over the historical 20 year period 1995–2014 here. We contrast the mean and spread across the 50 members for global, gridded maps, highlighting regions and indices that are particularly sensitive to internal variability. Using a single large ensemble ensures a robust estimation of the expected climate state and its uncertainty due to the remaining internal variability even in 20 year averages (Lehner *et al* 2020, Tebaldi *et al* 2021) and, at the same time, avoids convoluting model uncertainty and internal variability as detailed in section 3.1. We discuss selected results in this manuscript and provide the full set of indices in an easy-to-use web application, the *Variability Atlas* (https://lukasbrunner.github.io/variability_atlas). With this Atlas, we aim to raise awareness of the potential implications of internal variability on estimates of the ETCCDI indices in the community and to provide guidance to users of such extreme indices.

2. Data and methods

2.1. The MPI-ESM grand ensemble

To sample the effect of internal variability, we draw on the Max Planck Institute Earth System Model in its CMIP6 configuration (MPI-ESM1.2; Mauritsen *et al* 2019). We use daily minimum, mean, and maximum temperature as well as daily precipitation in the historical period 1961–2014 from the 50-member initial-condition large ensemble detailed in Olonscheck *et al* (2023).

We use MPI-ESM from a growing number of large ensembles (Maher *et al* 2025) as it provides a large number of members and can be considered fairly "middle of the road" in its representation of internal variability, compared to 15 other large ensembles. For precipitation, where the representation of variability is very heterogeneous in general, variability in MPI-ESM is mostly below average, and our results are hence rather a lower bound (see figures S2 and S3 in the supplement).

The use of a rather extensive 50-member ensemble further ensures that we comprehensively capture internal variability even for climate extremes at a grid-cell level (Milinski *et al* 2020, Tebaldi *et al* 2021). Using a subset of six extreme indices also used in this study, Tebaldi *et al* (2021) conclude that 20 members are sufficient to ensure accurate sampling of internal variability (less for temperature-based indices). Yet, their work does not consider the full set of available ETCCDI indices and approaches the issue from a model center perspective, while we, here, focus on a user-based view.

2.2. ETCCDI indices

In the main manuscript, we present a subset of the 26 core indices recommended by the ETCCDI, while the *Variability Atlas* covers the full set of indices. The ETCCDI indices are based on daily minimum, mean, and maximum temperature as well as on daily precipitation and mostly capture two types of extremes: block extrema and threshold exceedances Tebaldi *et al* (2021), see also. For the latter, an additional separation can be made into the sub-types of absolute and relative threshold-based indices, using a single fixed threshold and a spatially (and in some cases also seasonally) varying threshold, respectively. Each index type has specific properties that also determine how it is affected by internal variability (Sillmann *et al* 2014), as discussed in the results section. The indices presented in the manuscript are, hence, selected to cover both base variables and all three index types and their specific behavior in connection with internal variability:

- TXx (°C) – annual maximum of daily maximum temperature (block maximum)
- TX90p (%) – annual percentage of days where daily maximum temperature exceeds the local and daily 90th percentile temperature (relative threshold)
- WSDI (days) – annual count of days where daily maximum temperature exceeds the daily 90th percentile for at least six consecutive days (relative threshold)
- SU (days) – annual count of days where daily maximum temperature exceeds 25 °C (absolute threshold)
- Rx1day (mm) – annual maximum of daily precipitation (block maximum)
- R95p (mm) – annual sum of precipitation where daily precipitation exceeds the local 95th percentile (relative threshold)
- CWD (days) – maximum number of consecutive wet days (daily precipitation >1 mm) in the entire period (absolute threshold)
- R10mm (days) – annual sum of days where daily precipitation exceeds 10 mm (absolute threshold)

The relative threshold-based indices use the default 30 year base-period of 1961–1990. A full list of all indices can be found, e.g. in table 1 of Sillmann *et al* (2013). To practically calculate the ETCCDI indices, a range of options exists depending on specific user needs. These include software packages such as Climdex (Campalani *et al* (2025); https://www.climdex.org/), xclim ((Bourgault *et al* 2023); https://xclim.readthedocs.io/), or ESMValTool ((Weigel *et al* 2021); https://docs.esmvaltool.org/) as well as the use of pre-calculated indices from the Climate Data Store (2022) or the World Data Center for Climate (Brunner 2025).

Here we calculate the ETCCDI indices using the climate data operators (CDO) command line tool (Schulzweida 2023). CDO is a domain-specific language specialized for Earth system-related calculations, and provides very fast processing and easy input-output operations. This allows efficient and straightforward processing of 4 variables (daily minimum, mean, and maximum temperature as well as daily precipitation), the 50-member ensemble, and the 26 ETCCDI indices. A generalized collection of the corresponding scripts can be found at https://github.com/lukasbrunner/etccdi_cdo.

2.3. Estimating the contribution of internal variability

We focus our analysis on the 20 year out-of-base period 1995–2014 to completely avoid any potential in-base / out-of-base inhomogeneities (Zhang *et al* 2005) and differences in treatment of relative threshold-based indices and indices that do not require a base period. The end of the analysis period matches the end of the historical forcing period of CMIP6. For each index and member, the 20 annual values are averaged to represent a possible climate state.

Then, mean, standard deviation, and coefficient of variation (CV) are calculated across the 50 ensemble members for each index. The CV is defined as the ratio of standard deviation to mean and is, hence, dimensionless. Here, we provide it in the unit of a fraction rather than in percent to avoid any confusion with frequency indices, such as TX90p, which are given in percent. We note that for temperature-based block extrema we provide the mean and standard deviation in °C rather than in K as it is the more intuitive unit, while we exclusively use K as the base unit for the CV (see section 3.2 for a discussion).

3. Results

3.1. Showcasing the effect of internal variability on the example of TX90p

We start by showcasing possible effects of internal variability for the example of TX90p in Europe (figure 1). As the simplest possible case, figure 1(a) shows the TX90p frequency from a single realization of the MPI model (selected to be the member with median area mean frequency for demonstration purposes). This case still contains considerable (but *a priori* unknown) internal variability that is statistically similar to the single manifestation

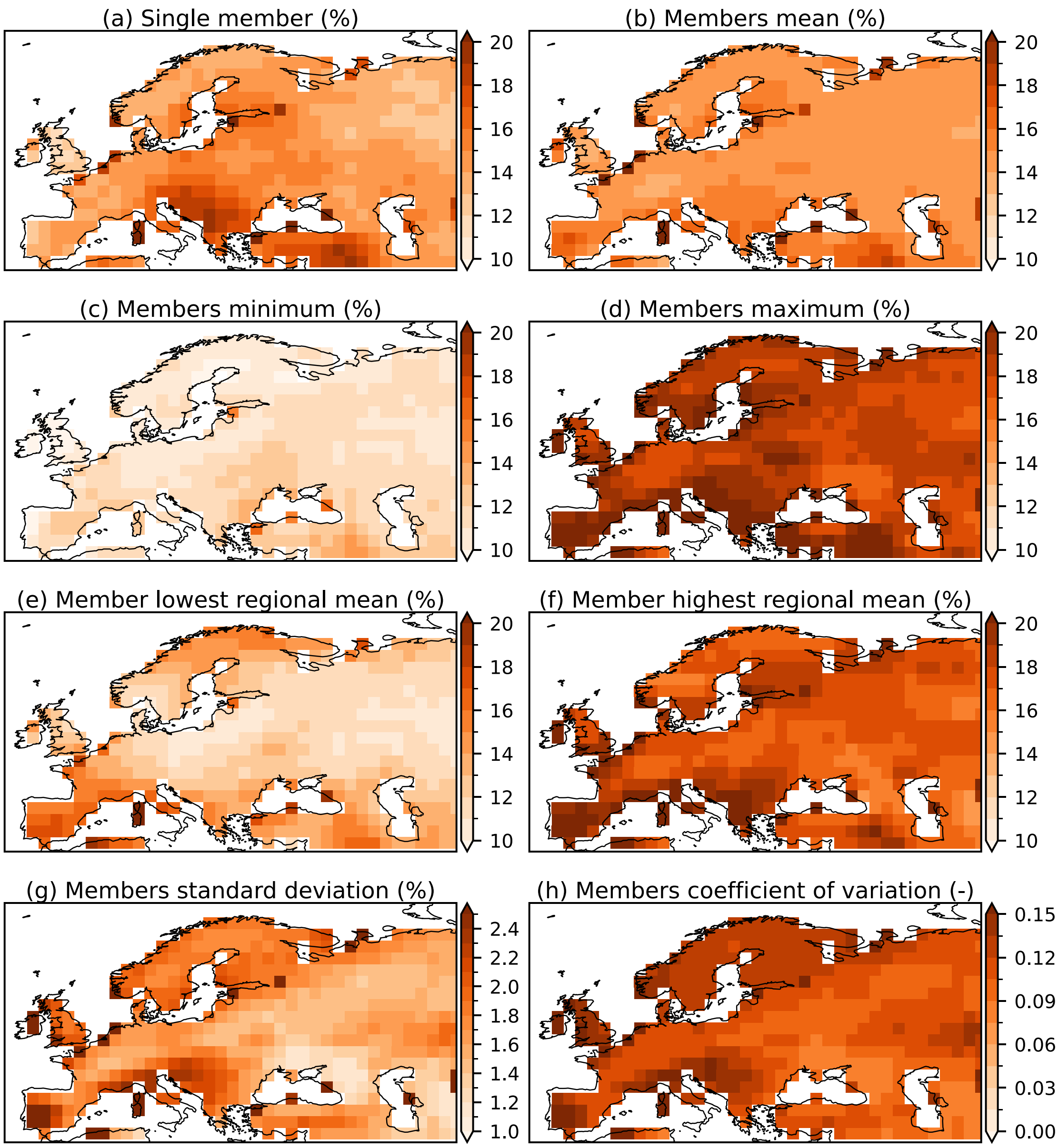

Figure 1. Effect of internal variability on TX90p for the example of Europe. TX90p from (a) a single member (picked here as the member with median region mean frequency for demonstration purposes) and the 50-member (b) mean, (c) minimum, and (d) maximum as well as from the member with (e) lowest and (f) highest region mean frequency. 50-member (g) standard deviation and (h) coefficient of variation (CV).

of observed climate, notwithstanding model biases (see supplement figure S1). In contrast to the single ensemble member case, figure 1(b) shows the mean TX90p frequency over all 50 members. This can be considered a robust estimate of the climatological mean of TX90p where internal variability has been averaged out (Milinski *et al* 2020, Tebaldi *et al* 2021). Different from the single realization, it is spatially very homogeneous at about 15%, with slightly higher values in southern Europe. At the same time, single grid cells still show higher peaks due to a high ocean fraction modulating the temperature.

In the first case (figure 1(a)), uncertainty from internal variability is implicitly not considered, as only a single (arbitrary) realization is used. As such, this case can be seen as analogous to an observation-based analysis without explicit consideration of internal variability. The second case, in contrast, explicitly excludes internal variability uncertainty (figure 1(b)). Yet, internal variability can have a considerable effect on the estimation of the TX90p frequency, even in the 20 year mean that we examine here, with the difference between the member minimum (figure 1(c)) and maximum (figure 1(d)) frequency exceeding 10%. This amounts to up to half of the single-member and mean frequencies, illustrating the potential influence that internal variability can play in shaping climate extreme index occurrence. Taking the minimum or maximum over 50 members at each grid cell serves illustration purposes here, but breaks spatial consistency and may therefore be inappropriate for some use cases: no single ensemble member shows a 20 year mean TX90p frequency pattern for Europe comparable to figures 1(c) and (d) as neighboring grid cells are not physically independent from each other in a single ensemble member. Still, such an approach might still be useful for an investigation of the full range of possibilities in a grid cell-by-grid cell view.

To address a context where the spatial pattern is relevant, we also show the two members with the lowest and highest regional mean frequency in figures 1(e) and (f), respectively. For the region of Europe shown here, their differences are smaller but comparable to the grid cell minimum and maximum in figures 1(c) and (d). While this approach is somewhat more targeted and depends strongly on the region (size), its clear advantage is the preservation of spatial patterns in the estimate of uncertainty. This is particularly relevant in contexts where multiple regions are analyzed, and physical coherence between these regions must be considered.

In the rest of the manuscript, we focus on the standard deviation rather than the minima and maxima, as it provides a simple, robust, and well-known metric to quantify variability. Yet, a range of other metrics, including the grid-cell and regional maxima and minima, is provided in the interactive online *Variability Atlas* as discussed in section 4. Figure 1(g) shows the standard deviation pattern of TX90p in Europe exemplarily. This view has several desirable properties as it (1) is very straightforward to calculate, (2) quantifies member differences in the unit of the index itself, hence making it easily comparable to the mean, and (3) shows an undistorted pattern, which highlights regions of low and high effect. Yet, there are contexts in which there are also several disadvantages of this approach as (1) understanding the relative importance of variability for the index calculation requires manual comparison to the mean, (2) the effect of internal variability between indices with different units is not directly comparable, and (3) for indices with large regional differences, the value of the standard deviation is dominated by the member mean rather than member differences.

Addressing these shortcomings, we also discuss the CV as a second metric to summarize internal variability uncertainty attached to ETCCDI indices (figure 1(h). The CV more directly quantifies the relative contribution of internal variability, accounting for regional differences in the climatological mean and for differences in the index units. For the example of TX90p, the resulting pattern looks very similar to the standard deviation, while providing a more direct comparison to the mean. For other indices, however, the CV has specific properties, which require careful interpretation, as we will discuss in the next section.

3.2. The effect of internal variability on selected temperature-based ETCCDI indices

Overall, the importance of internal variability strongly depends on the specific extreme index and region considered. Here, we show a subset of four temperature-based extreme indices and discuss the effect of internal variability based on the member standard deviation (figure 2) and CV (figure 3). Our aim is to highlight relevant examples, not to provide an extensive description of internal variability for all regions and indices. We cover the latter in the interactive *Variability Atlas* that allows visualization of flexible user-defined cases and is detailed in section 4.

For TXx, standard deviations are mostly below 5 °C (figure 2(a)), which could be considered small compared to the highest means of over 40 °C (figure 2(b)). However, from an impact viewpoint, such rather small values might still be crucial as the effects of heat can be highly non-linear (e.g. Dunne *et al* 2013, Matthews *et al* 2025). In particular, for high absolute temperature (potentially in combination with humidity), an additional degree can have huge impacts on human well-being. Basing the estimation of the TXx amplitude purely on the member mean (or, conceptually similar, only a multi-model mean), hence ignores the roughly 15% and 2.5% probabilities of exceeding the mean by more than one or two standard deviations, respectively, just due to internal variability.

From a climate-system perspective, the variability maps, such as the standard deviation map of TXx presented in figure 2, can also provide a first-order check on whether regions might be sensitive to processes that drive downstream variability in extremes and hence warrant closer investigation. For the specific example of TXx, strong standard deviation signals emerge in the Northern Hemisphere over the Great Plains in the US and in a band ranging from the Mediterranean Sea eastward through central Asia. These regions of high TXx variability closely coincide with those highlighted in recent work by Maraun *et al* (2025) as exhibiting particularly strong event-specific soil moisture-temperature coupling (measured as the correlation of latent heat flux and temperature on the days of TXx occurrence). The complex interplay of soil moisture availability (which, among other things, depends on the exact timing of TXx as soil moisture typically depletes during a heatwave) and coupling strength (which, in turn, depends on the climatological soil moisture state) might provide an explanation for the high variability values in these regions. While a more detailed analysis of this phenomenon is outside the scope of our work here, it still serves as a tangible example of the type of analysis that might be prompted and/or supported by the information the *Variability Atlas* can provide.

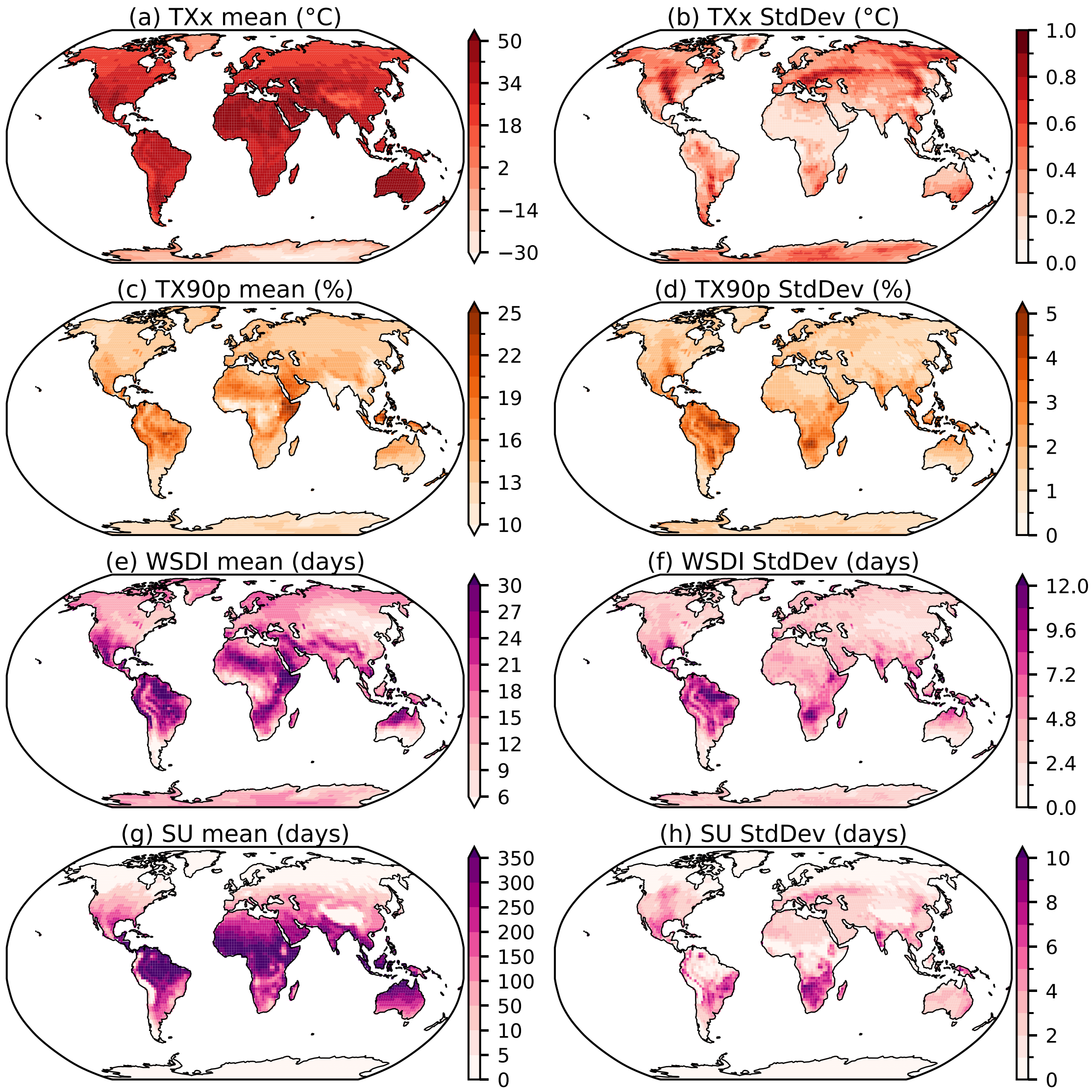

Figure 2. 50-member (left) mean and (right) standard deviation for four selected temperature-based ETCCDI indices: (a), (b) TXx, (c), (d) TX90p, (e), (f) WSDI, and (g), (h) SU. Indices with different units are indicated by different colormaps.

For the percentile-based TX90p index, the potential contribution of internal variability is about 3%–5% across the globe (figure 2(d)), which is considerable compared to the mean frequency of about 15%–20% (figure 2(c)). For the WSDI, which is a more complex metric derived from TX90p, the effect of internal variability becomes even more important with member standard deviations of up to 12 d (figure 2(f)). TX90p hotspot regions such as South America, parts of South Africa, and South-East Asia, but also parts of Louisiana in the US, are further amplified for the WSDI (figures 2(d) and (f)). While the mean WSDI values also tend to be higher than elsewhere for these hotspot regions, there is no one-to-one correspondence between mean and standard deviation, leading to regionally very high CVs for the WSDI (figure 3(c)).

The summer day index (SU), finally, has quite a different behavior compared to the other three indices, as shown in figure 2. This is due to its property as a fixed threshold-based index: at high latitudes in both hemispheres, the threshold is never crossed, leading to zero summer days in both mean (figure 2(g)) and standard deviation (figure 2(h)). Similarly, there are no summer days at high elevations, such as in the Himalayas or the Andes. At low latitudes, in turn, all days of the year exceed 25 °C, also leading to low variability.

To also provide the effect of the standard deviation relative to the mean severity or frequency of extremes, we show the CV in figure 3. While this might be considered a more comprehensive metric than the standard deviation, we warn against using it as a measure for internal variability in isolation in several cases, as detailed in the following.

For severity-based temperature indices such as TXx (i.e. indices based on the total amplitude of temperature), the issue of units arises: if the index is used

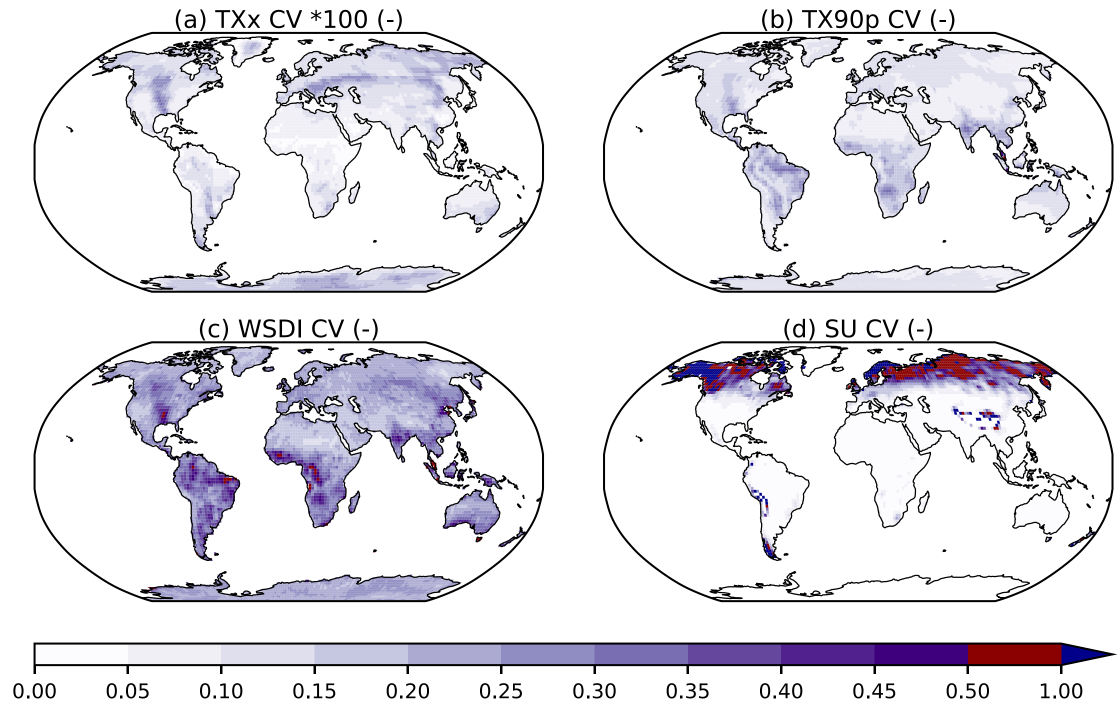

Figure 3. 50-member coefficient of variation (CV) for four selected temperature-based ETCCDI indices: (a) TXx, (b) TX90p, (c) WSDI, and (d) SU. Regions where the standard deviation accounts for more than half of the mean value are highlighted in red and blue. Note that for TXx the CV value is multiplied by 100 to be of the same order of magnitude as the other indices (see discussion in the text for details).

in units of °C (which is often more intuitive for communication and corresponds to many of the impact thresholds), the arbitrary zero-point of the Celsius scale leads to largely inflated CV values for regions with mean temperatures close to 0 °C. To circumvent this problem, we, here, converted TXx from °C to K before calculating the CV in figure 3(a). However, this leads to very small CV values as the standard deviation is in the order of magnitude of 1 K while the mean is of magnitude 100 K. As a result, the Celsius- as well as the Kelvin-based CV for severity-based temperature indices hide some of the potential impact of internal variability, and we show it here to raise awareness of this fact.

For fixed threshold indices such as the summer days, the CV strongly highlights transition regions (figure 3(d)), where the mean is close to zero (figure 2(g)) but internal variability can still lead to some occurrences (figure 2(h)). In Scandinavia and Alaska, for example, this combination leads to CV values exceeding 1, indicating that the standard deviation is higher than the mean, yet in absolute terms, both are well below 1 d per year. In such cases, the use of standard deviation or CV as a metric to provide the most informative assessment of internal variability strongly depends on the application. For cases where the physical extreme threshold is directly connected to physiological or economic impacts, the CV might be more appropriate as it highlights the risk for extreme occurrence, even in a climate where we would not expect it based purely on the mean. An example of such a case might be the potentially detrimental impact of even a single frost day (FD; minimum temperature below 0 °C) on apple tree blossom (Unterberger *et al* 2018). For other cases, such as the summer days shown in figures 2(g), (h) and 3(d), the threshold is less directly connected to impacts and small fluctuations around a near-zero mean might be less important, making the standard deviation the preferable variability metric.

Finally, for TX90p and the WSDI, the two percentile-based indices (figures 3(b) and (c)), the CV is well-defined and reveals relative contributions of internal variability of about 10%–20% and 20%–40%, respectively. The effect of variability on the WSDI is highest at low latitudes, with hotspot regions exceeding CV values of 0.5.

3.3. The effect of internal variability on selected precipitation-based ETCCDI indices

For precipitation-based indices, the amplitude of the standard deviation relatively closely follows the climatology for all three index types: block maxima and relative and absolute threshold-based indices (figure 4). This leads to rather uniform CV values globally (figure 5), except for desert regions, where the low climatological mean precipitation can lead to very high values exceeding 1.

For the two indices capturing precipitation severity (in mm), Rx1day and R95p, the standard deviations show substantial spatial variations, driven by regional topographical features such as the

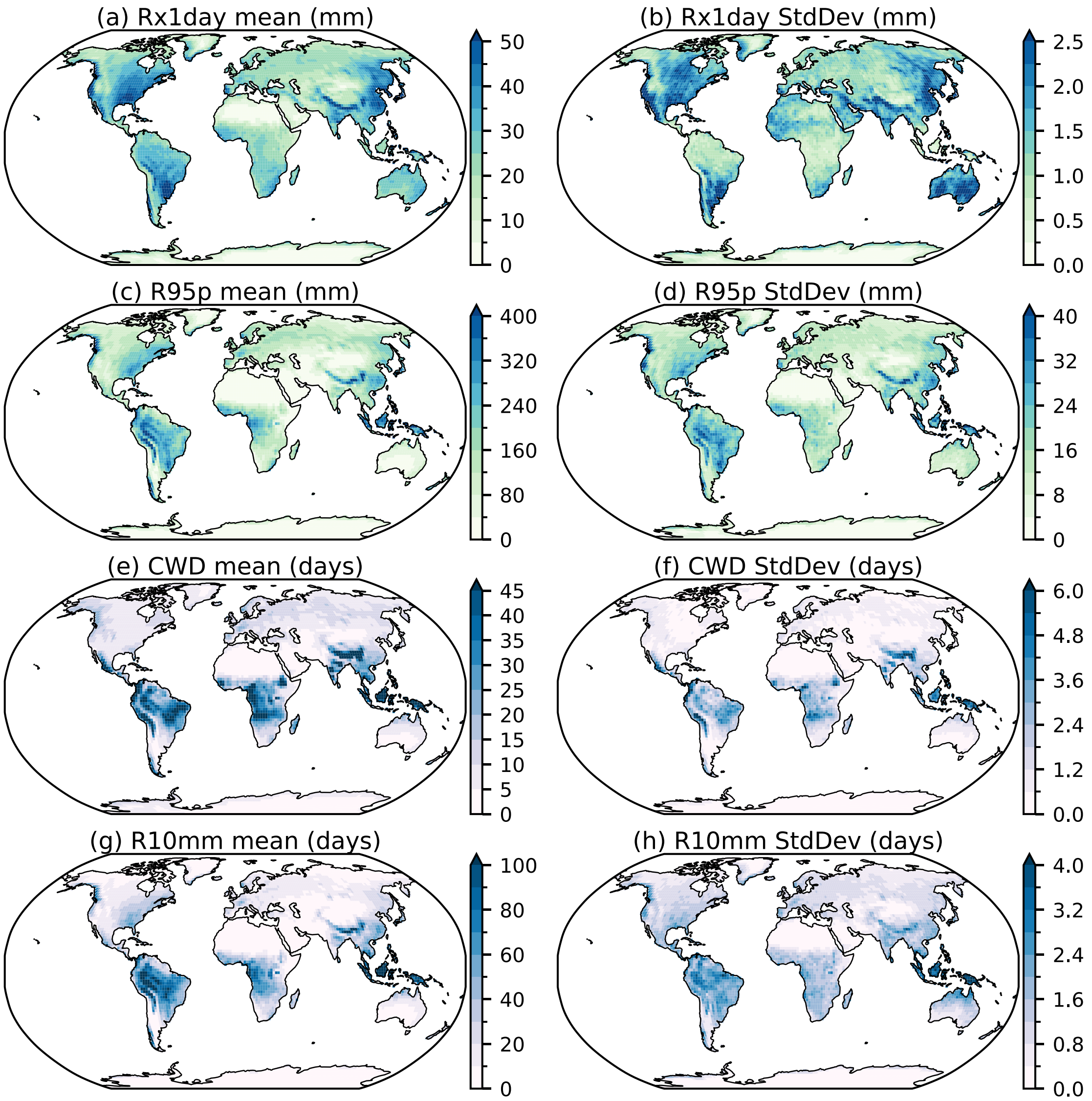

Figure 4. 50-member (left) mean and (right) standard deviation for four selected precipitation-based ETCCDI indices: (a), (b) Rx1day, (c), (d) R95p, (e), (f) CWD, and (g), (h) R10mm. Indices with different units are indicated by different colormaps.

Himalayan mountain range, which leads to high internal variability as well as high climatological values (figures 4(a)–(d)). The *Variability Atlas* allows the user to zoom into their regions of interest, revealing more tangible examples of the effect of internal variability on precipitation extremes. For instance, the R95p shows a standard deviation of 20 mm over Germany, equal to a CV of 0.1 (figure 6(a)) and substantial differences between the wettest and driest member (figure 6(b)), illustrating the notable contribution of climate variability for this case.

Both absolute threshold indices (CWD, R10mm) are, in turn, dominated by latitudinal differences, with the highest values in standard deviation and mean being found in the tropics. The CWD index, in particular, measures whether sequences of days consistently cross the threshold and is, therefore, quite sensitive to individual days dropping below the threshold and interrupting an otherwise connected wet-spell. This, in turn, makes it somewhat sensitive to model resolution, as coarser-resolved models (like the MPI-ESM used here) tend to have a higher 'drizzle bias' that leads to lower but more persistent precipitation (Trenberth and Zhang 2018). In contrast, recent work has shown that novel km-scale global models, in particular those allowing explicit convection, have a considerably lower precipitation persistence, more in line with observations (Spät *et al* 2024, Wille *et al* 2025, Takasuka *et al* 2026). For large regions of the low latitudes, (Brunner *et al* 2025) have shown vast differences in CWD between explicit and parameterized convection schemes, hinting at the limitation of the model used here and coarser CMIP6-like models in general. At the same time, the imprint of large-scale topographical features such as the Andes on CWD duration is consistent between coarse and km-scale models and across convection treatments, so that the *Variability Atlas* can still be informative for the investigation of internal variability of precipitation depending on the case.

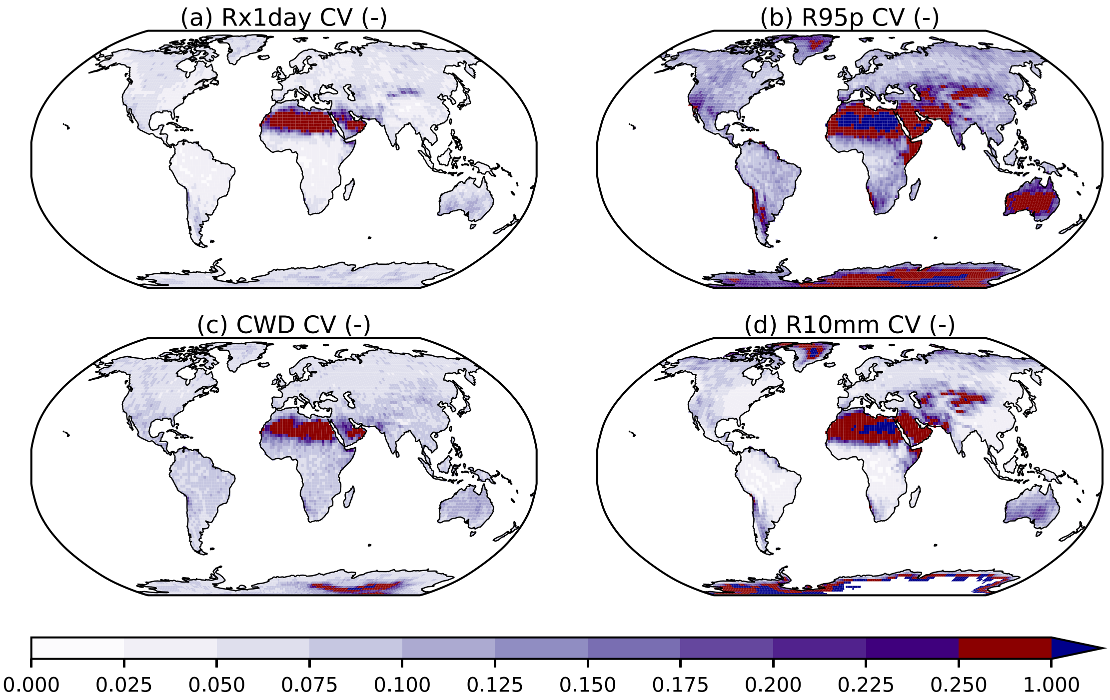

Figure 5. 50-member coefficient of variation (CV) for four selected precipitation-based ETCCDI indices: (a) Rx1day, (b) R95p, (c) CWD, and (d) R10mm. Regions where the standard deviation accounts for more than a quarter of the mean value are highlighted in red and blue.

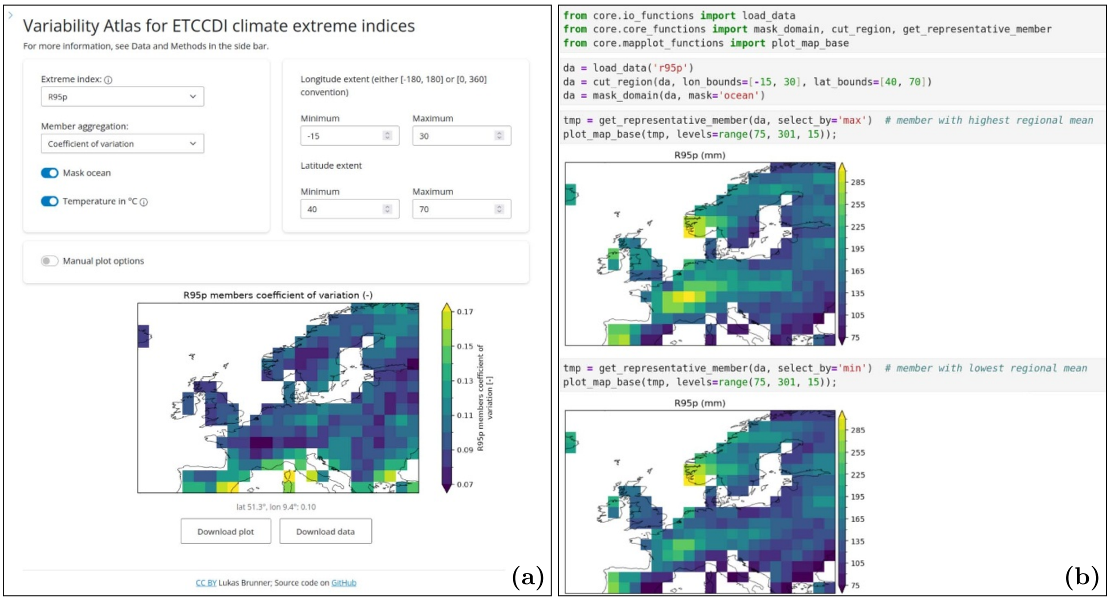

Figure 6. Example usage of the interactive tools provided to analyze the effect of internal variability by the *Variability Atlas* (https://lukasbrunner.github.io/variability_atlas/). (a) Shiny app and (b) example script as Jupyter notebook.

4. Introducing the interactive variability atlas

In the previous sections, we provided an overview of the effect of internal variability on selected ETCCDI indices for an example region and on a global scale. Yet, to estimate the importance of internal variability for a given case, a multitude of case-specific choices need to be considered, including the region and indices, as well as the potential emphasis on rather central estimates versus outliers, and whether spatial consistency is important. It is, therefore, impossible to provide the best possible information for all of these cases based on static plots.

To address this issue, we provide a layered *Variability Atlas* (https://lukasbrunner.github.io/variability_atlas) with increasing complexity in each layer to account for different user needs and levels of expertise. The first, most accessible, high-level overview is provided by a fully browser-based app with a graphical user interface. As shown in the screenshot in figure 6(a), variables and metrics can be selected via simple drop-down menus, and rectangular regions can be specified. As the resulting calculations are fairly simple, they can be done on the fly in the background, and the resulting visualizations can be inspected online or downloaded for further comparison.

This approach is necessarily limited to fairly simple cases, as restrictions apply due to the complexity of the interface and the available computing time. Therefore, the second option is to clone the git repository containing the high-level Python scripts used for the analysis in this study from https://github.com/lukasbrunner/etccdi_internal_variability. This repository also contains all pre-calculated 20 year mean ETCCDI indices in the period 1995–2014, allowing easy data handling while also enabling the full flexibility provided by a general-purpose language. Simple Python functions and Jupyter notebook examples are provided, allowing users to easily cut regions, mask ocean grid cells, and calculate member statistics, including selecting individual members representing extreme realizations of variability in a region, as shown in figures 1(e) and (f).

Finally, the full extreme index time-series 1850–2014 is also provided to users who might require specific time periods or seasonal results and can handle larger datasets: https://doi.org/10.26050/WDCC/ETCCDI-MPIGE (Brunner *et al* 2026b).

As a tangible example, bringing all pieces together, consider TX90p over Europe (figures 1 and S1). A practitioner interested in extreme heat in the Mediterranean might use observations (e.g. from ERA5; figure S1(a)) to assess recent changes and plan adaptation measures. Doing so would provide no information on the uncertainty stemming from internal climate variability in the extreme heat estimates based on which adaptation measures might be planned. Further, the single realization of observations might fail to capture plausible extreme cases, which could be of particular importance for decision-makers. A quick check of the *Variability Atlas* app might convince them that internal variability can play an important role for such analyses, particularly in the Mediterranean (compared to other regions in Europe). To follow up on the initial results provided by the app, they could analyze the entire ensemble or individual members depending on their specific use case by cloning the Git repository. Still, the 1995–2014 period we used for our pre-calculated climatologies might not quite fit their needs, so, in a final step, they could download the full time series and adjust it to match their case exactly for a detailed assessment.

5. Discussion and conclusions

We have quantified the effect of internal variability on the estimation of various climate extreme indices recommended by the ETCCDI with the aim of informing impact studies and users about internal variability uncertainty attached to those indices. Our main analysis focused on global patterns of mean, standard deviation, and CV across a 50-member large ensemble from a state-of-the-art coupled climate model. In addition, we also provide an interactive online *Variability Atlas* that allows users to estimate a range of internal variability measures on a case-by-case basis (https://lukasbrunner.github.io/variability_atlas).

As we show in detail in figure 1, this focus on the application is crucial as the specific user context helps to set priorities for the investigation of the variability contribution: For a simple documentation of the status quo based on historical observations, using the single manifestation of observed climate (see figure S1; equivalent to the single model run in figure 1(a) might be sufficient. Similarly, for a user only interested in the expected value from a climate perspective, the mean over a large ensemble (as in figure 1(b) or, equivalently, a multi-model ensemble, is adequate. Yet, in the face of the substantial relative contributions of internal variability to different ETCCDI indices showcased in this paper, we argue that even for such cases, it is important to clearly acknowledge the potential contribution of internal variability and to discuss its implications.

For applications where covering a possible 'worst case' due to specific realizations of internal variability is important, considering the full range (figures 1(c) and (d)) could be appropriate. Still, while the minimum-maximum differences across 50 members allow to assess the full range of possibilities, it is also sensitive to outliers. Therefore, more robust estimates of variability may be preferable in most cases. In our manuscript we base our estimate on the member standard deviation (e.g. figures 1(g) and (h)), but other options might be the likely (66%) or very likely (90%) ranges as defined by the Intergovernmental Panel on Climate Change (IPCC), which are possible metrics in the expert version of the *Variability Atlas* provided as part of this study as detailed in section 4 (Brunner *et al* 2026a) .

The toolbox provided by the *Variability Atlas*, hence, allows for investigating variability effects on the ETCCDI indices for specific, user-defined cases. These cases could include, for example, a simple quantification of uncertainty in the estimation of climate extreme indices attributable to internal climate variability. Other applications could include investigating the effect of internal variability for evaluating the representation of extremes in models against observations, as e.g. streamlined in the ESMValTool (Weigel *et al* 2021), or building spatially consistent regional extreme storylines based on selected individual members (see figures 1(e) and (f)).

At the same time, it is important to keep in mind that our model-based approach can only estimate real-world variability to the degree it is well-represented in the MPI-ESM model we use. To build confidence in our results based on MPI-ESM, we compare them to 15 other large ensembles in the supplement, finding that, at least in model space, its representation of variability is quite reasonable. The high TXx standard deviations in the Great Plains and across parts of Asia discussed above, for example, are reproduced by the ACCESS and GFDL ensembles and, to some degree, also by CESM1, E3SM, EC-Earth3, and MIROC6, lending confidence for temperature-based results. For precipitation, (model) uncertainty in general is higher, and MPI-ESM falls on the low end of variability compared to the other models, in particular in the tropics. Information from the *Variability Atlas* should, hence, be treated (1) as a first guess, with case-specific analysis and interpretation required for a reliable and holistic assessment and (2) as a lower bound, since higher-resolution datasets generally show greater variability as discussed in Brunner *et al* (2025). Due to its global scope, the *Variability Atlas* allows targeted analysis across different scales from global to regional, while the investigation of local details is limited by the rather coarse model resolution of about 150 km. At such scales, processes like deep convection cannot be resolved and, hence, are represented by statistical parameterizations, which introduce biases into the model (e.g. Trenberth and Zhang 2018, Spät *et al* 2024; section 3.3). Similarly, the resolution also limits the representation of topographical features such as mountain ranges and coastlines both within the models and in the output fields (Iles *et al* 2020, Poschlod and Koh 2024, Brunner *et al* 2025).

The *Variability Atlas* is, hence, well-suited for comparing variability across different extreme indices or regions and for providing a first estimate of internal variability while being mindful of these caveats. It highlights the effect of internal climate variability on extreme occurrence as estimated by a range of temperature- and precipitation-based indices, which is not considered in many studies, through a unique user lens. Depending on the extreme index and region, internal variability can account for a considerable part of the severity, frequency, or duration of extreme indices with implications for impact-relevant threshold exceedances and their interpretation for derived climate risk assessments and planning of adaptation measures.

Acknowledgments

The authors thank the World Data Center for Climate (WDCC) for hosting the ETCCDI indices calcualted for this paper and Eileen Hertwig (WDCC) for her help in the publication process. This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2037 'CLICCS-Climate, Climatic Change, and Society' - Project No. 390683824, a contribution to the Center for Earth System Research and Sustainability (CEN) of University of Hamburg.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/lukasbrunner/etccdi_internal_variability (Brunner *et al* 2026a).

Supplementary figures available at https://doi.org/10.1088/1748-9326/ae9c57/data1.

Author contributions

Lukas Brunner 0000-0001-5760-4524
Conceptualization (supporting), Data curation (supporting), Formal analysis (lead), Investigation (lead), Methodology (lead), Project administration (lead), Software (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (equal)

Vidur Mithal 0000-0003-2662-6936
Data curation (lead), Writing – review & editing (equal)

Leonard F Borchert 0000-0002-3232-7409
Conceptualization (supporting), Writing – review & editing (equal)

Jana Sillmann 0000-0002-0219-5345
Funding acquisition (lead), Writing – review & editing (equal)

Benjamin Poschlod 0000-0003-0247-2514
Conceptualization (lead), Writing – review & editing (equal)

References

Ajjur S B and Al-Ghamdi S G 2021 Global hotspots for future absolute temperature extremes from CMIP6 models *Earth Space Sci.* **8** e2021EA001817

Alexander L V *et al* 2006 Global observed changes in daily climate extremes of temperature and precipitation *J. Geophys. Res. Atmos.* **111** 1–22

Blanusa M L, López-Zurita C J and Rasp S 2023 Internal variability plays a dominant role in global climate projections of temperature and precipitation extremes *Clim. Dyn.* **61** 1931–45
Bohat N, Rajora K, Joshi V and Sarma K 2025 Assessment of climatic trends using Rclimdex indices in the Mandakini Watershed, Uttarakhand, India *Discov. Environ.* **3** 145
Bourgault P *et al* 2023 Xclim: Xarray-based climate data analytics *J. Open Source Softw.* **8** 5415
Brunner L 2025 Annual ETCCDI extreme indices for ICON-sapphire and IFS-FESOM (nextGEMS cycle 4) (https://doi.org/10.26050/WDCC/nextGEMS_ExtrInd)
Brunner L*et al* 2026a lukasbrunner/etccdi_internal_variability: Brunner_etal_2026_ERL (Version v1.0) [Computer software] *Zenodo* (https://doi.org/10.5281/zenodo.22205382)
Brunner L, Mithal V, Borchert L, Sillmann J and Poschlod B 2026b ETCCDI Indices for MPI Grand Ensemble *World Data Center for Climate (WDCC) at DKRZ* (https://doi.org/10.26050/WDCC/ETCCDI-MPIGE)
Brunner L, Poschlod B, Dutra E, Fischer E M, Martius O and Sillmann J 2025 A global perspective on the spatial representation of climate extremes from km-scale models *Environ. Res. Lett.* **20** 074054
Campalani P, Crespi A, Pittore M and Zebisch M 2025 Climdex-Kit: an open software for climate index calculation, sharing and analysis towards tailored climate services *Environ. Model. Softw.* **190** 106442
Collazo S, Barrucand M and Rusticucci M 2022 Evaluation of CMIP6 models in the representation of observed extreme temperature indices trends in South America *Clim. Change* **172** 21
Dunn R J H *et al* 2020 Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3 *J. Geophys. Res.: Atmos.* **125** e2019JD032263
Dunne J P, Stouffer R J and John J G 2013 Reductions in labour capacity from heat stress under climate warming *Nat. Clim. Change* **3** 563–6
Ebedi-Nding D D, Tamoffo A T and Mouassom F L 2024 Extremes events and socio-economic impacts in central Africa: a CMIP6-based analysis of projections *Model. Earth Syst. Environ.* **10** 6575–98
Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58
Fan X, Miao C, Duan Q, Shen C and Wu Y 2020 The performance of CMIP6 versus CMIP5 in simulating temperature extremes over the global land surface *J. Geophys. Res.: Atmos.* **125** e2020JD033031
Gessner C, Fischer E M, Beyerle U and Knutti R 2021 Very rare heat extremes: quantifying and understanding using ensemble re-initialization *J. Clim.* **34** 1–46
Getachew F, Bayabil H, Hoogenboom G, Kiker G A, Yu Z and Li Y 2025 Exploring the dynamics of past and future climate extremes under CMIP6: implications for rainfed agriculture *Theor. Appl. Climatol.* **156** 618
Hempel S, Frieler K, Warszawski L, Schewe J and Piontek F 2013 A trend-preserving bias correction & the ISI-MIP approach *Earth Syst. Dyn.* **4** 219–36
Iles C E, Samset B H, Sandstad M, Schuhen N, Wilcox L J and Lund M T 2024 Strong regional trends in extreme weather over the next two decades under high- and low-emissions pathways *Nat. Geosci.* **17** 845–50
Iles C E, Vautard R, Strachan J, Joussaume S, Eggen B R and Hewitt C D 2020 The benefits of increasing resolution in global and regional climate simulations for European climate extremes *Geosci. Model Dev.* **13** 5583–607
Keggenhoff I, Elizbarashvili M, Amiri-Farahani A and King L 2014 Trends in daily temperature and precipitation extremes over Georgia *Weather Clim. Extrem.* **4** 75–85
Kim Y-H, Min S-K, Zhang X, Sillmann J and Sandstad M 2020 Evaluation of the CMIP6 multi-model ensemble for climate extreme indices *Weather Clim. Extrem.* **29** 100269
Lange S 2019 Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0) *Geosci. Model Dev.* **12** 3055–70
Lee Y, Paek J, Park J-S and Boo K-O 2020 Changes in temperature and rainfall extremes across East Asia in the CMIP5 ensemble *Theor. Appl. Climatol.* **141** 143–55
Lehner F, Deser C, Maher N, Marotzke J, Fischer E M, Brunner L, Knutti R and Hawkins E 2020 Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6 *Earth Syst. Dyn.* **11** 491–508
Ma X, Zhang H, Zhao S and Xu K 2025 Assessing climate extremes indices over global drylands under real world warming beyond 1.5 °C: spatial distribution and temporal trends *Int. J. Climatol.* **45** e70020
Maher N, Lehner F and Marotzke J 2020 Quantifying the role of internal variability in the temperature we expect to observe in the coming decades *Environ. Res. Lett.* **15** 054014
Maher N, Phillips A S, Deser C, Wills R C J, Lehner F, Fasullo J, Caron J M, Brunner L, Beyerle U and Jeffree J 2025 The updated multi-model large ensemble archive and the climate variability diagnostics package: new tools for the study of climate variability and change *Geosci. Model Dev.* **18** 6341–65
Maraun D, Schiemann R, Ossó A and Jury M 2025 Changes in event soil moisture-temperature coupling can intensify very extreme heat beyond expectations *Nat. Commun.* **16** 734
Matthews T, Raymond C, Foster J, Baldwin J W, Ivanovich C, Kong Q, Kinney P and Horton R M 2025 Mortality impacts of the most extreme heat events *Nat. Rev. Earth Environ.* **6** 193–210
Mauritsen T *et al* 2019 Developments in the MPI-M earth system model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2 *J. Adv. Model. Earth Syst.* **11** 998–1038
Menang K P 2017 Climate extreme indices derived from observed daily precipitation and temperature data over Cameroon: the need for further assessments *Meteorol. Appl.* **24** 167–71
Milinski S, Maher N and Olonscheck D 2020 How large does a large ensemble need to be? *Earth Syst. Dyn.* **11** 885–901
Mousavi R, Johnson D L, Byrne J M and Kroebel R 2024 Exploring climate change impacts on temperature extremes in the south Saskatchewan river Watershed, Alberta, Canada *Atmosphere* **15** 1293
Nazrul F B, Karim M R, Pavan W and Kamruzzaman M 2025 Future climate extremes in Bangladesh: insights from a Multi-Model ensemble of CMIP6 simulations *Earth Syst. Environ.* **10** 1369–88
Olonscheck D *et al* 2023 The new max Planck institute grand ensemble with CMIP6 forcing and high-frequency model output *J. Adv. Model. Earth Syst.* **15** e2023MS003790
Peterson T C, Zhang X, Brunet-India M and Vazquez-Aguirre J L 2008 Changes in North American extremes derived from daily weather data *J. Geophys. Res.: Atmos.* **113** D7
Poschlod B and Koh J 2024 Convection-Permitting climate models can support observations to generate rainfall return levels *Water Resour. Res.* **60** e2023WR035159
Pyarali K, Zhang L and Forkel M 2025 Assessment of selected climate indicators across the Elbe river basin to analyse changes in climate extremes and their effects *Earth Syst. Environ.* **10** 4219–38
Rehana S, Yeleswarapu P, Basha G and Munoz-Arriola F 2022 Precipitation and temperature extremes and association with large-scale climate indices: an observational evidence over India *J. Earth Syst. Sci.* **131** 170
Sa'adi Z *et al* 2025 Examining spatiotemporal patterns of temperature extremes in Johor: a link to public perception *Next Res.* **2** 100485
Schaller N, Sillmann J, Anstey J, Fischer E M, Grams C M and Russo S 2018 Influence of blocking on Northern European

and Western Russian heatwaves in large climate model ensembles *Environ. Res. Lett.* **13** 054015

Schulzweida U 2023 CDO user guide Version 2.3.0. (https://doi.org/10.5281/zenodo.10020800)

Seo Y-W, Kim H, Yun K-S, Lee J-Y, Ha K-J and Moon J-Y 2014 Future change of extreme temperature climate indices over East Asia with uncertainties estimation in the CMIP5 *Asia-Pac. J. Atmos. Sci.* **50** 609–24

Service, Copernicus Climate Change 2022 Climate extreme indices and heat stress indicators derived from CMIP6 global climate projections (https://doi.org/10.24381/cds.776e08bd7)

Sillmann J, Kharin V V, Zhang X, Zwiers F W and Bronaugh D 2013 Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. model evaluation in the present climate *J. Geophys. Res. Atmos.* **118** 1716–33

Sillmann J, Kharin V V, Zwiers F W, Zhang X, Bronaugh D and Donat M G 2014 Evaluating model-simulated variability in temperature extremes using modified percentile indices *Int. J. Climatol.* **34** 3304–11

Spät D, Biasutti M, Schuhbauer D and Voigt A 2024 Autocorrelation-a simple diagnostic for tropical precipitation variability in global kilometer-scale climate models *Geophys. Res. Lett.* **51** e2024GL108856

Squintu A A, van der Schrier G, van den Besselaar E, van der Linden E, Putrasahan D, Roberts C, Roberts M, Scoccimarro E, Senan R and Tank A K 2021 Evaluation of trends in extreme temperatures simulated by HighResMIP models across Europe *Clim. Dyn.* **56** 2389–412

Takasuka D, Becker T and Bao J 2026 Precipitation characteristics and thermodynamic-convection coupling in global kilometer-scale simulations *J. Adv. Model. Earth Syst.* **18** e2025MS005343

Tebaldi C *et al* 2021 Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6 *Earth Syst. Dyn.* **12** 253–93

Trenberth K E and Zhang Y 2018 How often does it really rain? *Bull. Am. Meteorol. Soc.* **99** 289–98

Unterberger C, Brunner L, Nabernegg S, Steininger K W, Steiner A K, Stabentheiner E, Monschein S and Truhetz H 2018 Spring frost risk for regional apple production under a warmer climate *PLoS One* **13** 1–18

von Trentini F, Aalbers E E, Fischer E M and Ludwig R 2020 Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe *Earth Syst. Dyn.* **11** 1013–31

Wang X, Jiang D and Lang X 2017 Future extreme climate changes linked to global warming intensity *Sci. Bull.* **62** 1673–80

Weigel K *et al* 2021 Earth system model evaluation tool (ESMValTool) v2.0- diagnostics for extreme events, regional and impact evaluation and analysis of Earth system models in CMIP *Geosci. Model Dev.* **14** 3159–84

Wille J D, Koch R, Becker T and Fischer E 2025 Extreme precipitation depiction in convection-permitting earth system models within the nextGEMS project *J. Adv. Model. Earth Syst.* **17** e2024MS004840

Xu K, Xu B, Ju J, Wu C, Dai H and Hu B X 2019 Projection and uncertainty of precipitation extremes in the CMIP5 multimodel ensembles over nine major basins in China *Atmos. Res.* **226** 122–37

Yate T A and Ren G 2025 The opposite trends in precipitation total and extremes during two rain-seasons across Ethiopia, the Water Tower of Africa *Weather Clim. Extrem.* **50** 100813

Zhang X, Alexander L, Hegerl G C, Jones P, Tank A K, Peterson T C, Trewin B and Zwiers F W 2011 Indices for monitoring changes in extremes based on daily temperature and precipitation data *WIREs Clim. Change* **2** 851–70

Zhang X, Hegerl G, Zwiers F W and Kenyon J 2005 Avoiding inhomogeneity in percentile-based indices of temperature extremes *J. Clim.* **18** 1641–51