ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
24 March 2025

REVISED
2 June 2025

ACCEPTED FOR PUBLICATION
6 June 2025

PUBLISHED
17 June 2025

LETTER

A global perspective on the spatial representation of climate extremes from km-scale models

Lukas Brunner[1,*], **Benjamin Poschlod**[1], **Emanuel Dutra**[2,3], **Erich M Fischer**[4], **Olivia Martius**[5] **and Jana Sillmann**[1]

[1] Research Unit Sustainability and Climate Risk, Center for Earth System Research and Sustainability (CEN), University of Hamburg, Hamburg, Germany
[2] Instituto Portugués do Mar e da Atmosfera, 1749-049 Lisbon, Portugal
[3] Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal
[4] Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[5] Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland
[*] Author to whom any correspondence should be addressed.

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

Keywords: climate extremes, km-scale models, temperature, precipitation, CMIP6

Supplementary material for this article is available online

Abstract

Weather and climate extremes are rare manifestations of climate variability that can severely impact society and the environment. To investigate their properties and changes on a global scale, observational records are often complemented by climate models such as those from the latest Coupled Model Intercomparison Project (CMIP6). However, typical CMIP6 models have a grid spacing of about 100 km and, therefore, do not allow the representation of extremes at local scales important for impacts. Here, we provide a global view on the information lost at such resolutions, focusing on temperature and precipitation extremes. We draw on two next-generation, km-scale global climate models, run with a grid spacing of about 10 km, and regrid them to a range of coarser resolutions. From the regridded data, we then investigate the spatial *sub-grid variability* hidden at the CMIP6-like 100 km grid spacing to quantify the effect resolution has on the representation of extremes globally. We find clear patterns of such a resolution effect on a diverse set of temperature extremes, particularly in mountainous areas, at coastlines, and along large rivers. For the example of annual maximum temperature, the difference between the 10 km and 100 km grid spacings can exceed 10 °C. For precipitation, globally aggregated low and high extremes are shown to be underestimated at coarse resolution, with the strongest spatial signals emerging in regions with complex topography and in the tropics. Our results quantify existing knowledge and demonstrate the importance of spatial resolution for the representation of climate extremes, in particular for hotspot regions such as coastlines, which often coincide with densely populated areas. The advent of ever higher resolved global models, hence, allows improved estimates of local climate impacts and related risk assessments with global coverage.

1. Introduction

Climate extremes can develop on a wide range of spatial scales from continental to local. To cover these scales and investigate extreme event properties and their changes in different regions, global climate models (GCMs) are often used. However, established GCMs, such as those from the latest Coupled Model Intercomparison Project (CMIP6; Eyring *et al* (2016)) have a grid spacing of 100–200 km, limiting their applicability for the investigation of local extreme properties and impacts (e.g. Wehner *et al* 2010, Kopparla *et al* 2013, Torma *et al* 2015, Iles *et al* 2020, John *et al* 2024). Here, we use two new GCMs with a previously unprecedented resolution of about 10 km to provide a global quantification of what spatial information about climate extremes is lost at coarser, CMIP6-like resolutions.

For individual domains, regional climate models (RCMs) are already able to resolve local details. Some even explicitly resolve convection (in so-called convection-permitting RCMs; CPRCMs) which has been shown to lead to a better representation of extremes (Coppola *et al* 2020, 2021, Poschlod and Koh 2024, Soares *et al* 2024). However, they do not provide global coverage and show conceptual deficiencies, such as the potential inadequacy of the lateral boundary conditions and the lack of two-way interactions with the driving GCM. Further, their production often lags many years behind due to the additional processing steps needed (Davies 2014, Giorgi 2019, Sobolowski *et al* 2025).

Recently emerged GCMs, able to simulate the climate system with a grid spacing of about 10 km or even less (km-scale models), combine global coverage with the resolution of local details for the first time and aim to build towards digital twins of Earth to support adaptation decisions at a community level (Bauer *et al* 2021, Hohenegger *et al* 2023, Hazeleger *et al* 2024, Stevens *et al* 2024, Rackow *et al* 2025). First studies have already confirmed the potential of this new generation of models to better represent important climate variables and processes such as temperature, precipitation, and land-atmosphere feedbacks compared to observations (Kuma *et al* 2024, Lee and Hohenegger 2024, Li *et al* 2024, Spät *et al* 2024, Wille *et al* 2024).

Here, we take a complementary approach and investigate the added value of km-scale models for the spatial representation of various climate extremes purely within a given model and without reference to observations. We take advantage of the first 30 year integrations from the ICOsahedral Non-hydrostatic (ICON) and the Integrated Forecasting System (IFS) models. Their output is available on a 13 km equal-area grid, that we use to produce a range of coarser resolved datasets through regridding to then quantify the spatial variability lost at these coarser scales. This allows a globally consistent evaluation of the resolution effect on climate extremes without the need to consider the complex chain of observations, CPRCMs, RCMs, and GCMs. Our method, hence, isolates the effect of output resolution while not allowing direct conclusions about model performance or the effect of actually running models at different resolutions, which are covered in other works (e.g. Iles *et al* 2020, Ban *et al* 2021, Ha *et al* 2024, Lee and Hohenegger 2024, Soares *et al* 2024, Poujol *et al* 2025).

We apply our approach to a diverse set of temperature- and precipitation-based climate extreme indices to address the following questions: (1) what is the effect of resolution on the spatial and spatio-temporal distribution of extremes globally, (2) in which regions does output resolution matter for the representation of extremes, and (3) how do the results differ for different extremes represented by a range of indices?

2. Data and methods

2.1. Km-scale model data

We use data from two fully coupled GCMs: the ICON model in its Sapphire configuration (Hohenegger *et al* 2023) and the IFS model coupled to the Finite-volumE Sea ice-Ocean Model (FESOM2; Rackow *et al* (2025)). Here, we refer to these models simply as ICON and IFS.

Both were developed and run in the frame of the next Generation of Earth Modeling Systems (nextGEMS) project and we use their latest version, produced in the fourth development cycle (Segura *et al* 2025). ICON is run with a grid spacing of about 10 km on an icosahedral-triangular grid and IFS with about 9 km on a reduced Gaussian grid. For more detailed information about the models' setup and specification, we refer to the reference publications cited above.

Here, we briefly raise the representation of convection as a major difference between the two models. In ICON convection parametrization is turned off entirely (Hohenegger *et al* 2023) while IFS still parameterizes convection but reduces its influence (Rackow *et al* 2025). The effect of convection parametrization on different climate variables, including precipitation and temperature extremes, has already been described based on analysis of limited-area CPRCMs (Coppola *et al* 2020, Ban *et al* 2021, Ha *et al* 2024, Sangelantoni *et al* 2024, Soares *et al* 2024) and will be discussed in the context of ICON-IFS differences throughout this study.

We use all 29 full years available from both models, representing future projections from 2021 to 2049 using the high emission scenario SSP3-7.0 (Meinshausen *et al* 2020). ICON provides 15-minute mean temperature and accumulated precipitation, which we aggregate to hourly values by taking the mean and sum, respectively. IFS provides hourly instantaneous temperature and hourly accumulated precipitation. The hourly values of both models are processed to daily maximum temperature and daily accumulated precipitation.

2.2. The HEALPix grid

Both models are output on the Hierarchical Equal Area isoLatitude Pixelation of a sphere (HEALPix) grid (Górski *et al* 2005). HEALPix is an unstructured grid that allows representing geospatial data at different discrete resolutions or *zoom levels*. Higher zoom levels have higher resolution, with the highest available level used here being 9, corresponding to a horizontal grid spacing of 12.7 km. A defining feature of

Table 1. Summary of zoom levels used in this study and their respective grid spacing and number of grid cells globally. The nside parameter is also given for completeness as it is a frequently used alternative metric for the resolution. It indicates the number of pixels along a side of one of the 12 zoom level 0 pixels.

Zoom	nside	Grid spacing	Grid cells
9	512	12.7 km	3'145'728
8	256	25.5 km	786'432
7	128	50.9 km	196'608
6	64	101.9 km	49'152
5	32	203.7 km	12'288
4	16	407.5 km	3'072
3	8	815.0 km	768
2	4	1'629.9 km	192
1	2	3'259.8 km	48
0	1	6'519.6 km	12

HEAlpix is that the area of all grid cells at a given zoom level is identical everywhere on the globe.

A reduction of one zoom level corresponds to a doubling of the grid spacing or a reduction of the number of grid cells by a factor of four (see table 1). Computationally, this coarsening is very simple and only consists of averaging four neighboring grid cells.

2.3. Coarsening km-scale model output to lower resolutions

We draw on the zoom level 9 data from ICON and IFS to represent a range of other resolutions by coarsening them to different lower zoom levels (table 1). While this is clearly not equivalent to actually running a model at each of the different resolutions, the crucial advantage of this method is its simplicity: the comparison of resolutions is done purely within a single model run, eliminating all other factors.

2.4. Identifying regions with information loss at coarse resolutions

To identify regions with large information loss at a coarse grid spacing of about 102 km (zoom level 6), corresponding to the grid spacing of a typical CMIP6 model, we compare it to the 13 km (zoom level 9) reference resolutions. The resolution difference, hence, corresponds to three zoom levels or a change in the number of grid cells of $4^3 = 64$. While this approach uses the high-resolution model runs as a reference to quantify the effect of resolution, it is not reliant on them being completely unbiased compared to observation-based references as the comparison of a model to its own coarsened output provides an implicit bias-correction. We point out that this assumption only holds as long as the high-resolution model is not affected by spatially rapidly changing biases from one grid cell to the next and represents fundamental physics (land-sea contrasts, elevation dependence of temperature, and orographic lifting to give some examples) correctly.

We define the *sub-grid variability* of an extreme index as the standard deviation across the 64 zoom level 9 grid cells that make up a single grid cell at zoom level 6. To further quantify the effect of the resolution, we define the *sub-grid anomaly* as the difference between each grid cell at zoom level 9 minus the zoom level 6 grid cell within which it lies. The code to calculate and plot these metrics can be found in Brunner (2025b).

2.5. The ETCCDI extreme indices

To investigate the effect of resolution on climate extremes, we draw on a subset of 6 indices from the pool of 27 indices recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI; Zhang *et al* (2011)), focusing on hot and wet extremes. The used indices represent a diverse selection based on block maxima, absolute thresholds, and relative thresholds to cover different extreme properties. The calculated ETCCDI indices can be found in Brunner (2025a) and a database of time-mean figures in Brunner (2025c).

For temperature we use the annual maximum of daily maximum temperature, the annual number of summer days (days with maximum daily temperature exceeding 25 °C), and the warm spell duration index (annual sum of days with maximum daily temperature exceeding the local, seasonal 90th percentile for at least six consecutive days). For precipitation we use the annual maximum of daily precipitation, the annual number of heavy rain days (days with daily precipitation exceeding 10 mm), and the annual maximum length of consecutive days with daily precipitation exceeding 1 mm. Finally, the time average over all 29 years is taken for all indices.

We note that for the sub-grid anomaly (section 2.4) of extreme indices the order of calculation matters: calculating an extreme index at high resolution and then coarsening it is not the same as calculating it from a base variable at coarse resolution. To show the full potential of the difference in spatial resolutions we opt for the latter approach in the main paper: the first step is to remap the base variables (daily maximum temperature and daily precipitation) from zoom level 9 to zoom level 6. The next step is to calculate extreme indices at both zoom levels (6 and 9), and finally the calculation of the differences as described above. This means that the mean of all 64 anomalies within one coarse grid cell is not necessarily zero. The complementary approach is discussed in the supplement.

2.6. 10-year return level of hourly precipitation

In addition to the ETCCDI indices, which represent relatively moderate extremes on daily time scales, we also apply our approach to the 10-year return level of hourly precipitation. We sample the annual maxima of hourly precipitation and fit the generalized extreme value (GEV) distribution (Coles *et al* 2001):

$$\mathrm{GEV}(x;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\} & \text{otherwise} \end{cases} \quad (1)$$

The location parameter μ, the scale parameter σ, and the shape parameter ξ govern the center, spread, and tail behavior of the GEV distribution, respectively. Following Papalexiou and Koutsoyiannis (2013), we limit the shape parameter to values between 0 and 0.23, as GEV fits on small sample sizes tend to show a high estimation variance for the shape parameter. The assumption of a positive shape parameter indicating a heavy tail agrees with Ragulina and Reitan (2017), and Rivoire *et al* (2022). We assess the goodness of fit via the Anderson-Darling test and adjust the critical p-value for multiple testing (Wilks 2016, Poschlod 2021), as we conduct the test across more than 3 million grid cells for zoom level 9 (table 1). Less than 2.2% of all fits are rejected at the significance level of $\alpha = 0.05$.

3. Results

3.1. The effect of resolution on the spatio-temporal precipitation distribution

We start by visualizing the effect of output resolution in the new km-scale models and its implications for the example of daily precipitation in figure 1. The emergence of this new generation of models enables us to view global spatial patterns of precipitation with unprecedented local detail as shown for a selected day from ICON in figure 1(a). In fact, not every grid cell is represented in figure 1 due to limits in the graphical resolution. Still, the view allows us to identify several features such as fronts and extratropical cyclones, high-intensity convective rainfall in the tropics, and dry areas with precipitation rates of less than 0.01 mm d^{-1}.

The lower panels in figure 1 show the frequency of daily precipitation based on its pooled spatio-temporal distribution, which corresponds to more than 30 billion values for the highest zoom level. The values are grouped into 0.01 mm d^{-1} bins and normalized by the total number of values for each zoom level separately. The middle column shows the corresponding distributions for ICON and IFS, with the left and right columns allowing a more detailed view of the dry and wet extremes, respectively.

For precipitation values exceeding about 25 mm d^{-1} the different zoom levels start to diverge in figures 1(c) and (f) since cases with larger grid spacing represent averages over larger areas, while cases with smaller grid spacing are able to better resolve local high-intensity events. This becomes even more apparent in figures 1(d) and (g), where the frequency of exceedance is shown up to precipitation rates of 1 m d^{-1}. We note that these frequencies are based on the globally pooled precipitation distribution, which hides zonal and regional details. Yet, we choose this approach here as we are mainly interested in the effect of resolution on different extremes from a top-level view and with focus on the global picture. Maps of mean and extreme precipitation are shown in figures S5 and S6 in the supplement.

Figures 1(b) and (e) zoom into the low end of the precipitation distribution, showing the frequency of dry days based on three thresholds. The highest threshold with a daily precipitation rate of 1 mm is often used as definition for dry days in climate models to avoid what has been termed 'drizzle bias' and to allow a fair comparison to observations (Trenberth and Zhang 2018, Chen *et al* 2021). This bias emerges due to the coarse model resolution and resulting parametrizations that result in too few days with no precipitation and too many days with small precipitation rates compared to the point measurements provided by stations. Here, we use two additional smaller thresholds (0.1 mm d^{-1} and 0.01 mm d^{-1}), showing the rapid decrease in dry days with coarsening resolutions. The 1 mm d^{-1} threshold, in turn, reduces most of the resolution difference between zoom levels 9 and 6 (13 km and 102 km, respectively) in IFS, indicating that the threshold works as intended for this case. In ICON, however, the mitigating effect is less pronounced due to its explicit representation of convection leading to smaller precipitation cells (Takasuka *et al* 2025).

Finally, we highlight a notable difference between the precipitation distributions of ICON and IFS, which warrants future investigation: ICON is more sensible to resolution, resulting in a larger spread of frequencies at almost all precipitation rates (figures 1(b) to (f)). As a result ICON has a higher number of dry days at the highest zoom level than IFS, and vice versa for the lowest zoom level (see also Wille *et al* 2024). At the same time, the IFS distributions have heavier tails at all resolutions. Understanding the underlying drivers of these differences is crucial for reliably modeling precipitation, downstream impact assessments, as well as further model development. Yet, such an investigation requires a dedicated study and is outside of the scope of this work.

3.2. Spatial sub-grid variability

Next, we analyze the spatial information lost at resolutions typical for current GCMs from CMIP6. As coarse grid spacing, we conservatively choose zoom level 6 or 102 km gird spacing, which contains exactly 64 grid cells at zoom level 9 (13 km; see methods for details). Zoom level 6 grid cells with low variability, hence, indicate that the higher output resolution does not add much information, while grid cells with

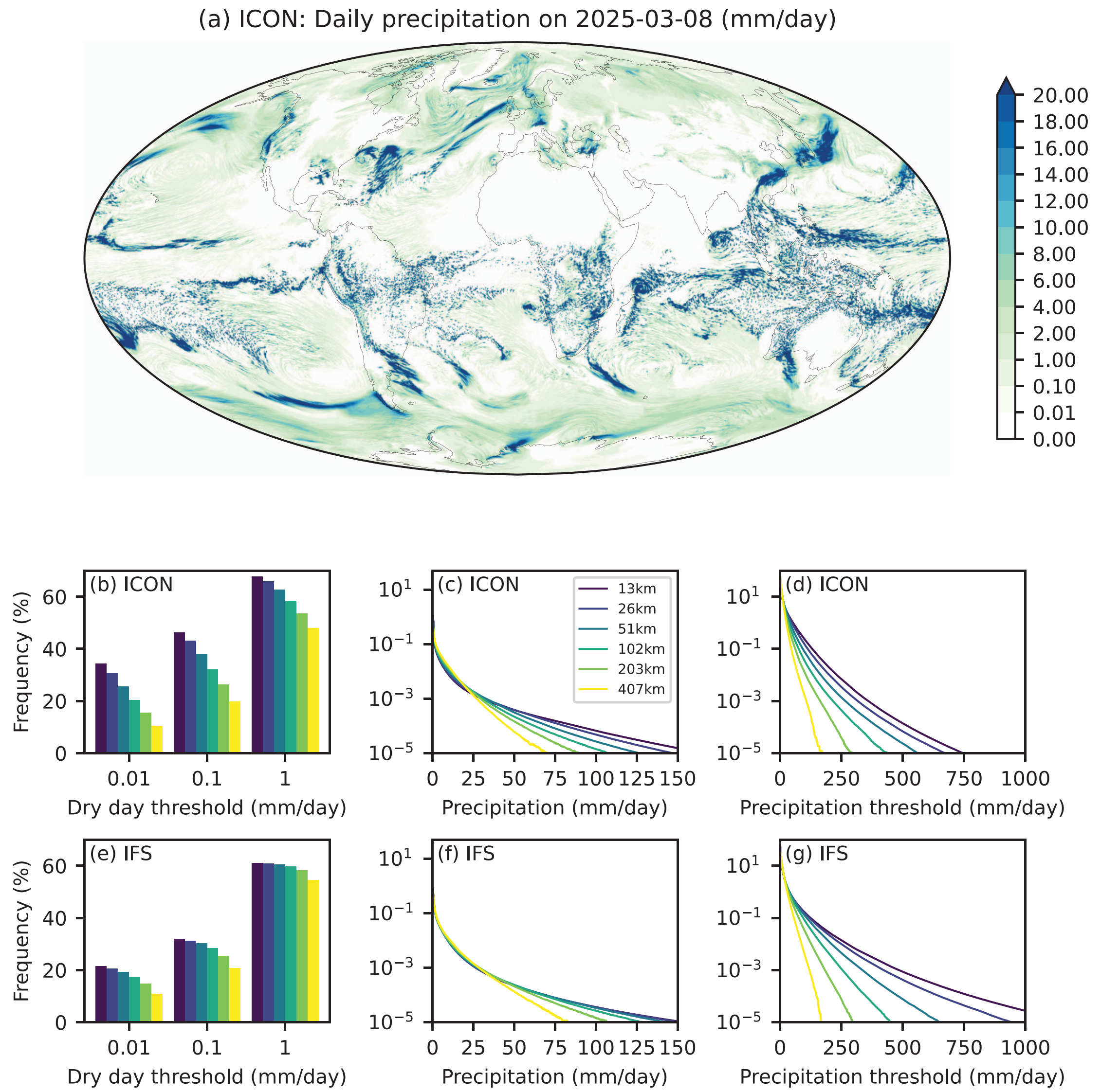

Figure 1. (a) Spatial distribution of daily precipitation in ICON on a selected day (March 8th 2025) for the highest zoom level 9. (b)–(g) Different views on the precipitation frequency for 0.01 mm d^{-1} bins sampled in the period 2021–2049 and over all grid cells (i.e. the total number of samples depends on the resolution). (b), (e) Dry days using different thresholds; (c), (f) precipitation frequency smoothed using a running average of size 100; (d), (g) precipitation exceedance frequency.

high values highlight regions where the resolution of a typical CMIP6 model is not sufficient to represent the spatial variability of extremes. We apply this methodology to a set of three extreme indices for temperature (figure 2) and precipitation (figure 3), respectively.

3.2.1. Temperature extreme indices

For the annual maximum of daily maximum temperature and summer days (figures 2(a)–(d)) three distinct regions can be identified where information is hidden at coarse resolutions: coasts, topographically complex regions, and larger water bodies over land such as lakes and even larger rivers. From an impact perspective, this finding alone is crucial as coastal areas and large river basins are often densely populated (Kummu *et al* 2011, Cosby *et al* 2024). As a consequence, these results suggest that risk to human society and infrastructure from heat extremes might not be adequately captured when estimated from a coarse resolution model; in fact, this misrepresentation mostly manifests as an underestimation of risk following from an underestimation of the amplitude of hot extremes on land as we will discuss in detail in section 3.3.

Summer days, as an index with an absolute threshold, have the additional property of their heterogenous climatology. At high latitudes and high elevations, their climatological occurrence is zero, while at some low-latitude regions every day exceeds the 25 °C threshold (figure S7). Consequently, the summer day sub-grid variability is zero in these regions while their edges are coined by sharp geographical gradients in the summer day climatology, leading to high sub-grid variability. Apparent examples are the Andes in South America and the Himalayas in Asia: both are framed by values of high

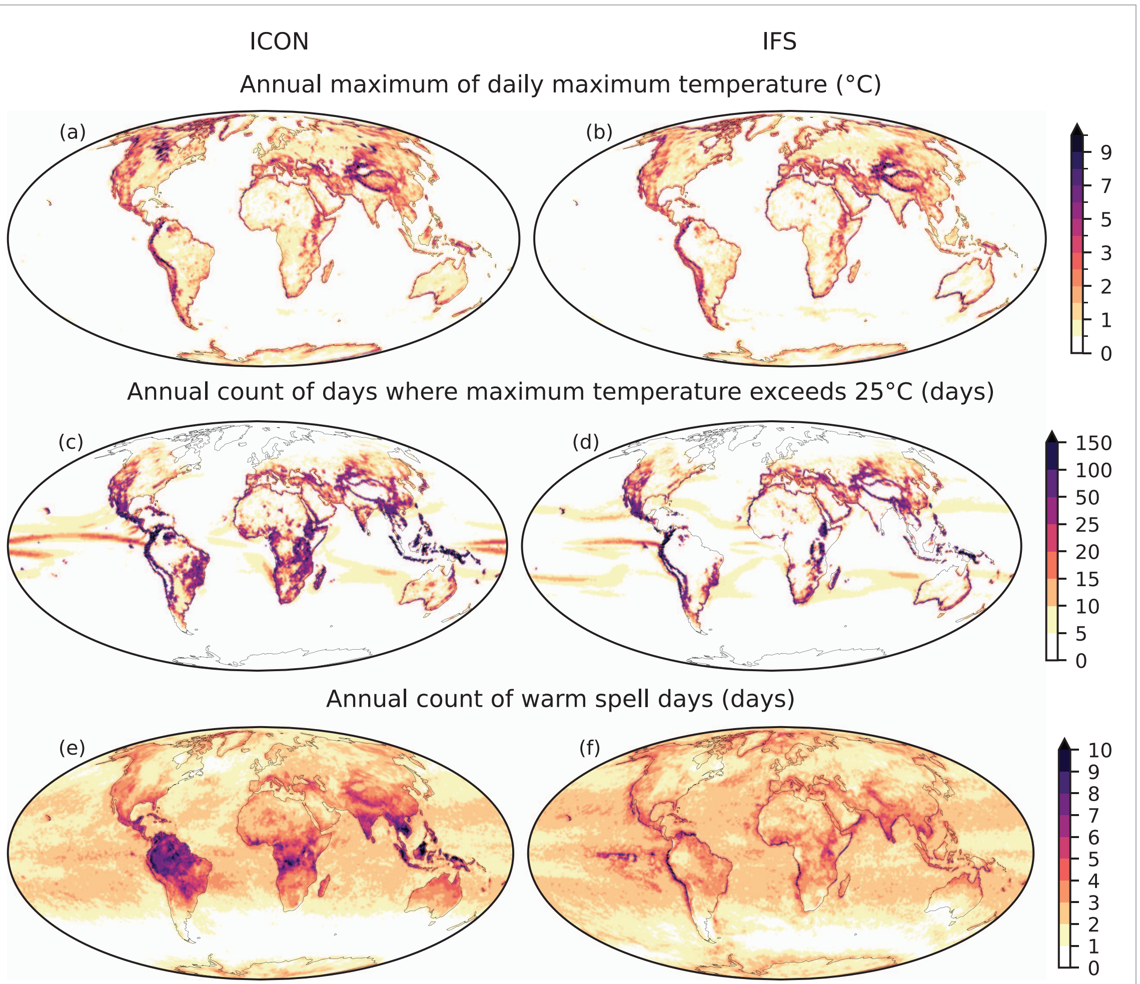

Figure 2. Standard deviation across 64 high-resolution grid cells falling into each coarse-resolution grid cell for (a), (b) annual maximum of daily maximum temperature, (c), (d) number of summer days, and (e), (f) warm spell duration. Shown is the 2021–2049 mean of annual values for (left) ICON and (right) IFS. Note that ocean grid cells are not masked, but appear white in many cases due to their low variability.

variability in the transition zone where the finer grid spacing resolves topographical gradients, while the highest mountain ranges have no variability as temperatures never exceed 25 °C (see also model topography and its sub-grid variability in figure S14). A similar consideration holds, for example for the topical Pacific, where the number of summer days can sharply transition from region saturated at 365 days to regions with very low summer day counts (figure S7).

The strongest impact of km-scale resolution for the representation of summer days can be found at the northern end of the Andes in Ecuador and Colombia, in Central America, and across the islands of Indonesia, the Philippines, and Papa New Guinea. Here, the effects of land-sea contrast and topography are combined in regions with a high climatological count of summer days due to their low latitude location.

The warm spell duration index (figures 2(e) and (f)) is distinct from the other two temperature-based indices mainly because it uses a percentile-based threshold. Coastal and topographical effects are less pronounced with the relative threshold accounting for them to a degree. In addition, the effect of climatological differences (mainly between low and high latitudes) is eliminated for percentile-based indices. This is showcased in figure S4 for the example of two conceptually similar indices based on an absolute and relative threshold (summer days and hot days, respectively). While the absolute threshold-based summer days show a very clear imprint of climatology, topography, and land-sea contrast, all these effects are weakened for the relative threshold hot days. Yet, some imprint of topography and land-sea contrasts remains even for hot days, as mountain ranges can lead to strongly varying conditions even at neighboring grid cells and oceans dampen variability compared to land in general (see section S5 and figure S4 for additional discussion). Finally, there are notable differences in the sub-grid variability between ICON and IFS in the equatorial region. We will discuss these in the next section together with the complementary differences in precipitation.

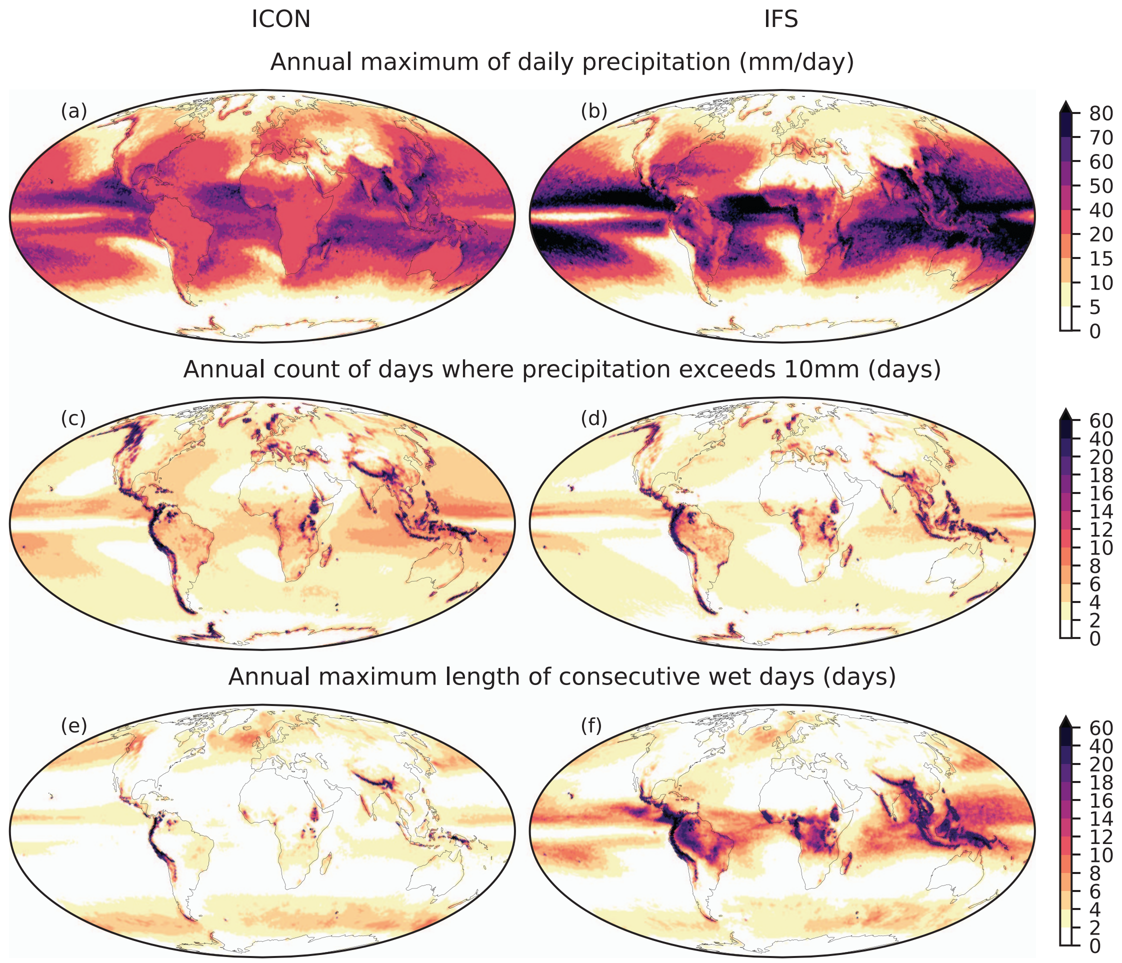

Figure 3. Standard deviation across 64 high-resolution grid cells falling into each coarse-resolution grid cell for (a), (b) the annual maximum of daily precipitation, (c), (d) number of heavy rain days, and (e), (f) maximum annual length of consecutive wet days. Shown is the 2021–2049 mean of annual values for (left) ICON and (right) IFS.

3.2.2. Precipitation extreme indices

The spatial patterns of sub-grid variability of extreme precipitation (shown in figure 3) differ distinctively from their temperature counterparts. Our assessment of lost information at coarse scales shows high variability for the annual maximum of daily precipitation at low latitudes that also have the highest absolute daily precipitation (see supplement figure S9). In addition, several hotspots coincide with large population centers such as the Gulf of Mexico, the Mediterranean, and India. While agreeing on the spatial patterns of this index, major differences between ICON and IFS emerge in the amplitude of variability. IFS shows larger maxima of variability at low latitudes consistent with the overall higher frequency of precipitation extremes (figure 1).

For the number of days where daily precipitation exceeds 10 mm, the tropics show relatively high variability based on a combination of high climatological values and spatially heterogenous convective cells dominating precipitation. Further, we detect a high variability on the western sides of the mountain ranges, where advective precipitation is enhanced by orographic lifting on the windward sides. The Himalayan mountain range is a particularly apparent example showing high variability for all three precipitation extreme metrics. In addition to the orographic effect, it marks the transition between regions with high precipitation in the south to regions with hardly any precipitation in the north (figures S2 and S9) and is, hence, characterized by sharp gradients that are better resolved at high resolution.

For the variability in the maximum number of consecutive wet days, ICON and IFS also agree on the general global pattern, but IFS simulates larger sub-grid variability widely across the tropics, while ICON only selectively shows high values in complex topography. These differences are based on the considerably lower absolute consecutive wet day count in ICON compared to IFS (figure S9), which leads to lower standard deviations (while the coefficient of variation is generally higher in ICON; figure S12). The differing behavior can be traced to the treatment of convection, with work by Spät *et al* (2024) finding that parametrizing convection leads to temporally more persistent precipitation patterns in IFS (see also lag 1 day auto-correlation in figure S11). ICON, in turn, features more local precipitation events driven

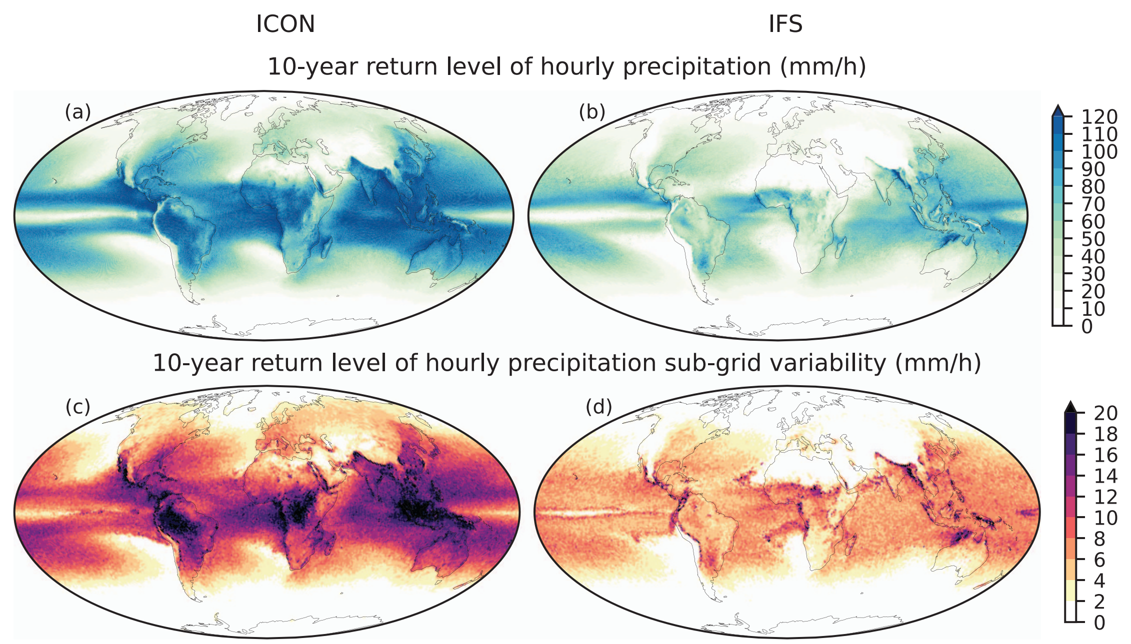

Figure 4. Climatology and sub-grid variability for the 10-year return level of hourly precipitation. (a), (b) Climatology at high resolution and (c), (d) standard deviation across 64 high-resolution grid cells falling into each coarse-resolution grid cell. Shown is the 2021–2049 mean of annual values for (left) ICON and (right) IFS.

by the explicit representation of convection. The pattern in precipitation also corresponds relatively well to an inverse pattern in the warm spell duration index (figures 2(e) and (f)). Here, ICON features more local precipitation events driven by the explicit representation of convection, driving the high variability in warm spells.

3.2.3. Precipitation return levels

For the return levels, we start by focusing on annual maximum hourly precipitation itself rather than its sub-grid variability (figures 4(a) and (b)). For this hourly metric ICON features higher intensities than IFS, in contrast to the daily values we have investigated so far (figures 1 and 2). We suggest that this relates to the differing representations of convection, which become even more pronounced at sub-daily time-scales. This difference in the amplitude then propagates to the assessment of sub-grid variability with values up to over 20 mm h^{-1} in the tropics and over 10 mm h^{-1} in the Mediterranean and Central Europe within the storm-resolving ICON simulation.

As sub-daily heavy rainfall is the main trigger for urban flooding, the additional information contained in km-scale models, compared to typical CMIP6-like resolutions is of utmost importance for local adaptation planning in cities. Locally, the higher resolution has the potential to improve the representation of extreme precipitation at individual locations instead of large areal averages on the one hand (figure 1) and to better represent spatial variability on the other hand (figure 4).

3.3. Spatial sub-grid anomaly

Finally, we point to the implications of resolution for the assessment of climate impacts and associated risks on the example of the hottest annual maximum. We choose this index as the impacts of extreme heat are very much non-linear (Gasparrini *et al* 2015, Vicedo-Cabrera *et al* 2021) and a quantification of differences between resolutions is therefore crucial for robust risk assessments.

In figure 5(a), we show a global view of the sub-grid anomaly for the hottest annual maximum, which reveals the patterns driving the corresponding sub-grid variability (figure 2(a)). For coasts and land water, high variability arises from a dipole pattern emerging from the moderating effect of water bodies on air temperature that is better represented at the km-scale, while the coarse grid spacing of 100 km × 100 km averages both domains (see e.g. the dipoles along the coast of India in figure 5(b) which correspond to high sub-grid variability in the ocean fraction; figure S13). For topography, the variability arises from the better representation of elevation and its gradients at higher resolutions, often leading to slightly more complex structures (see, e.g. the strong anomalies at the Himalayan mountain range in figure 5(b) corresponding to high sub-gird variability in the topography; figure S14).

The zoom-in to the Indian subcontinent in figure 5(b) reveals differences between the resolutions exceeding 5 °C on most of the coastal land, where many of the largest population centers of the region are located (figure 5(c)). For the example of Karachi (Pakistan), with more than 20 million inhabitants,

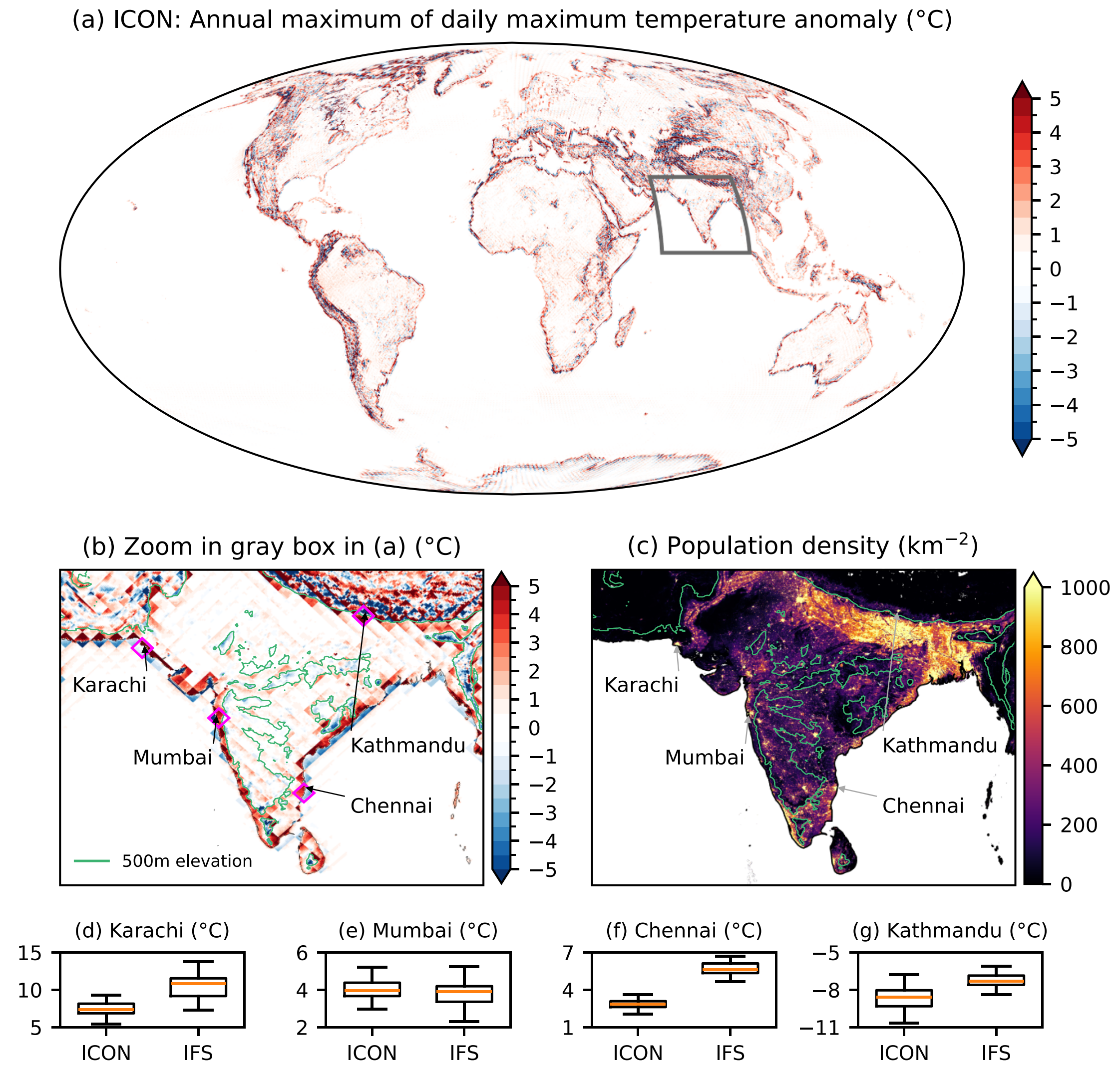

Figure 5. High resolution minus low resolution difference for (a), (b) the annual maximum of daily maximum temperature. Four major cities are marked by an arrow in (b) pointing to the high-resolution grid cell and a pink square highlighting the low-resolution grid cell. (c) Population density from the worldpop (https://hub.worldpop.org/) dataset. The green contour-lines indicate 500 m elevation. (d)–(g) Distribution of anomalies from all 29 years at the four cities marked in (b) for ICON and IFS. Each anomaly is the maximum of daily maximum temperature in a given year from the single high-resolution grid cell closest to the city location minus the low-resolution grid cell within which it lies.

the annual maximum of daily maximum temperatures can even be more than 10 °C hotter at high resolution in IFS (figure 5(d)). In Mumbai (India) the effect is weaker but can still reach 5 °C in both models. For Chennai on the Indian east coast the two models show some diversity with differences in ICON being limited to less than 4 °C, while IFS reaches 6 °C. For the city of Kathmandu (Nepal), the resolution effect is reversed and temperatures are higher at coarse resolution (figure 5(g)). This is due to its location at about 1.5 km elevation and on a mountain slope in combination with the geometry of the coarse grid. As can be seen in figure 5(b), Kathmandu is located at the northern (and hence high) end of the coarse grid cell, and therefore the average representation overestimates maximum temperatures there, while it underestimates them in the lowlands to the south.

Sub-grid anomalies for the other extreme indices discussed in the paper can be found in the supplement (figures S8 and S10). We also discuss the annual maximum of daily precipitation as an example for a precipitation-based extreme index in more detail in the supplement (section S2 and figure S2).

4. Summary and discussion

Drawing on ICON and IFS, two kilometer-scale, fully-coupled GCMs, we have provided a global picture of the importance of model output resolution for the representation of climate extremes in many regions of the Earth. Our study is the first to quantify the resolution effect at kilometer-scale consistently in a global model and confirms earlier regional work and process understanding. Our global approach is, hence, complementary to regional climate modeling efforts and allows us to compare different regions as well as to suggest areas for prioritization, where dynamical downscaling or regional grid refinement are most beneficial.

We calculated coarser resolutions from the high-resolution (about 13 km grid spacing) ICON and IFS data to allow a fully consistent comparison, isolating only the effect of output resolution. While this approach is clearly not equivalent to actually running a climate model at coarser resolutions we argue that it provides an upper limit for model fidelity at a given resolution: running a model at high resolution and then remapping it to a coarser resolution will lead to a better representation of reality than running the model directly at a coarser resolution (given no bugs emerge at higher resolutions and proper tuning; Proske *et al* (2024)). In addition, the latter approach necessarily introduces additional sources of differences only indirectly connected to resolution. ICON, for example, is run with convection parametrization disabled at its 10 km grid spacing, and running it at considerably coarser resolutions would require activating this the parametrization, leading to considerable changes in the model setup as a second-order resolution effect.

Our approach, hence, does not allow any conclusions about absolute model performance nor do our results necessarily indicate that higher resolution is always better. Both points depend on a multitude of factors and trade-offs. For example, a high-resolution model is typically able to resolve more processes and hence better from a physical perspective but empiric parametrizations in a lower-resolution model might still lead to better model performance compared to observations.

Our results highlight the potential of kilometer-scale models, such as those from nextGEMS (Segura *et al* 2025) and DestinE (John *et al* 2024), for impact assessments in global intercomparison projects such as ISIMIP and GGCMI (Rosenzweig *et al* 2017, Franke *et al* 2020) as well as in impact studies (e.g. Lüthi *et al* 2023, Orlov *et al* 2024)). Here, we have focused on a set of relatively simple standard indices based on temperature and precipitation. The next steps could be the combination of several climate variables into more impact-relevant indices, such as the wet-bulb temperature, as well as an analysis of the effects of climate change. In particular, with continuously increasing temperature records and the likely exceedance of 1.5 °C global warming in the next decades (Bevacqua *et al* 2025), an accurate assessment of climate extremes is crucial for robust risk assessments (Sillmann *et al* 2024).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: www.wdc-climate.de/ui/entry?acronym=nextGEMS_ExtrInd (Brunner 2025a).

Acknowledgment

The authors thank Benjamin Blanz and Johanna Baehr (University of Hamburg), Lukas Kluft (MPI for Meteorology, Hamburg), and Dorian Spät (University of Vienna) for helpful discussions on the manuscript. We thank two anonymous reviewers for their helpful comments.

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. ED, EF, and OM acknowledge support by the Horizon 2020 project nextGEMS under Grant Agreement Number 101003470.

This work was conceptualized by LB with contributions from ED and BP. Data analysis and writing (original draft) by LB with contributions from BP. Methodology by LB with contributions from all authors. Writing (review and editing) by all authors. Project administration, data curation, software, and visualization by LB.

Code availability

The code to calculate and plot the extreme metrics from this study can be found in Brunner (2025b).

ORCID iDs

Lukas Brunner https://orcid.org/0000-0001-5760-4524
Benjamin Poschlod https://orcid.org/0000-0003-0247-2514
Emanuel Dutra https://orcid.org/0000-0002-0643-2643
Erich M Fischer https://orcid.org/0000-0001-8722-9876
Olivia Martius https://orcid.org/0000-0002-8645-4702
Jana Sillmann https://orcid.org/0000-0002-0219-5345

References

Ban N *et al* 2021 The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation *Clim. Dyn.* **57** 275–302
Bauer P, Stevens B and Hazeleger W 2021 A digital twin of Earth for the green transition *Nat. Clim. Change* **11** 80–83
Bevacqua E, Schleussner C-F and Zscheischler J 2025 A year above 1.5 °C signals that Earth is most probably within the 20-year period that will reach the Paris Agreement limit *Nat. Clim. Change* **15** 1–4

Brunner L 2025a Annual ETCCDI extreme indices for ICON-Sapphire and IFS-FESOM (nextGEMS cycle 4) (https://doi.org/10.26050/WDCC/nextGEMS_ExtrInd)
Brunner L 2025b Scripts for Brunner et al. 2025 A global perspective on the spatial representation of climate extremes from km-scale models (https://doi.org/10.5281/zenodo.15619824)
Brunner L 2025c Supplementary figures for Brunner et al. 2025 A global perspective on the spatial representation of climate extremes from km-scale models (https://doi.org/10.5281/zenodo.15613611)
Chen D, Dai A and Hall A 2021 The convective-to-total precipitation ratio and the 'drizzling' bias in climate models *J. Geophys. Res.: Atmos.* **126** e2020JD034198
Coles S, Bawa J, Trenner L and Dorazio P 2001 *An Introduction to Statistical Modeling of Extreme Values.* vol 208 (Springer)
Coppola E *et al* 2020 A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean *Clim. Dyn.* **55** 3–34
Coppola E *et al* 2021 Climate hazard indices projections based on CORDEX-CORE, CMIP5 and CMIP6 ensemble *Clim. Dyn.* **57** 1293–383
Cosby A, Lebakula V, Smith C, Wanik D, Bergene K, Rose A, Swanson D and Bloom D 2024 Accelerating growth of human coastal populations at the global and continent levels: 2000–2018 *Sci. Rep.* **14** 22489
Takasuka D, Becker T and Bao J 2025 Precipitation Characteristics and Thermodynamic-Convection Coupling in Global Kilometer-Scale Simulations (https://doi.org/10.22541/essoar.174834999.93788721/v1)
Davies T 2014 Lateral boundary conditions for limited area models *Q. J. R. Meteorol. Soc.* **140** 185–96
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
Franke J A *et al* 2020 The GGCMI Phase 2 experiment: global gridded crop model simulations under uniform changes in CO_2, temperature, water and nitrogen levels (protocol version 1.0) *Geosci. Model Dev.* **13** 2315–36
Gasparrini A *et al* 2015 Mortality risk attributable to high and low ambient temperature: a multicountry observational study *The Lancet* **386** 369–75
Giorgi F 2019 Thirty years of regional climate modeling: where are we and where are we going next? *J. Geophys. Res.: Atmos.* **124** 5696–723
Górski K M, Hivon E, Banday A J, Wandelt B D, Hansen F K, Reinecke M and Bartelmann M 2005 HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere *Astrophys. J.* **622** 759
Ha M T *et al* 2024 Precipitation frequency in Med-CORDEX and EURO-CORDEX ensembles from 0.44° to convection-permitting resolution: impact of model resolution and convection representation *Clim. Dyn.* **62** 4515–40
Hazeleger W *et al* 2024 Digital twins of the Earth with and for humans *Commun. Earth Environ.* **5** 1–11
Hohenegger C *et al* 2023 ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales *Geosci. Model Dev.* **16** 779–811
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
John A *et al* 2024 Global storyline simulations at the kilometre-scale (available at: www.authorea.com/users/852952/articles/1239319-global-storyline-simulations-at-the-kilometre-scale?commit=12de5e231724accee75467b638937fafc3886f21)
Kopparla P, Fischer E M, Hannay C and Knutti R 2013 Improved simulation of extreme precipitation in a high-resolution atmosphere model *Geophys. Res. Lett.* **40** 5803–8
Kuma P *et al* 2024 Ship-based lidar evaluation of Southern Ocean clouds in the storm-resolving general circulation model ICON and the ERA5 and MERRA-2 reanalyses (https://doi.org/10.22541/essoar.173655440.05475709/v1)
Kummu M, De Moel H, Ward P J and Varis O 2011 How close do we live to water? a global analysis of population distance to freshwater bodies *PLoS One* **6** e20578
Lee J and Hohenegger C 2024 Weaker land-atmosphere coupling in global storm-resolving simulation *Proc. Natl Acad. Sci.* **121** e2314265121
Li L, Bisht G, Hao D and Leung L R 2024 Global 1 km land surface parameters for kilometer-scale Earth system modeling *Earth Syst. Sci. Data* **16** 2007–32
Lüthi S *et al* 2023 Rapid increase in the risk of heat-related mortality *Nat. Commun.* **14** 4894
Meinshausen M *et al* 2020 The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500 *Geosci. Model Dev.* **13** 3571–605
Orlov A *et al* 2024 Human heat stress could offset potential economic benefits of Co_2 fertilization in crop production under a high-emissions scenario *One Earth* **7** 1250–65
Papalexiou S M and Koutsoyiannis D 2013 Battle of extreme value distributions: a global survey on extreme daily rainfall *Water Resour. Res.* **49** 187–201
Poschlod B 2021 Using high-resolution regional climate models to estimate return levels of daily extreme precipitation over Bavaria *Nat. Hazards Earth Syst. Sci.* **21** 3573–98
Poschlod B and Koh J 2024 Convection-permitting climate models can support observations to generate rainfall return levels *Water Resour. Res.* **60** e2023WR035159
Poujol B, Lee J, Rackow T, Rotach M W and Ban N 2025 Are the largest benefits of kilometer-scale climate models over mountains or over flatland? *Geophys. Res. Lett.* **52** e2024GL113937
Proske U, Brüggemann N, Gärtner J P, Gutjahr O, Haak H, Putrasahan D and Wieners K-H 2024 A case for open communication of bugs in climate models, made with ICON version 2024.01 *EGUsphere* **1** 1–30
Rackow T *et al* 2025 Multi-year simulations at kilometre scale with the integrated forecasting system coupled to FESOM2.5 and NEMOv3.4 *Geosci. Model Dev.* **18** 33–69
Ragulina G and Reitan T 2017 Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach *Hydrol. Sci. J.* **62** 863–79
Rivoire P, Le Gall P, Favre A-C, Naveau P and Martius O 2022 High return level estimates of daily era-5 precipitation in europe estimated using regionalized extreme value distributions *Weather Clim. Extremes* **38** 100500
Rosenzweig C *et al* 2017 Assessing inter-sectoral climate change risks: The role of ISIMIP *Environ. Res. Lett.* **12** 010301
Sangelantoni L *et al* 2024 Investigating the representation of heatwaves from an ensemble of km-scale regional climate simulations within CORDEX-FPS convection *Clim. Dyn.* **62** 4635–71
Segura H *et al* 2025 nextGEMS: entering the era of kilometer-scale Earth system modeling *EGUsphere* 1–39
Sillmann J *et al* 2024 Climate extremes and risks: Links between climate science and decision-making *Front. Clim.* **6** 1499765
Soares P M M *et al* 2024 The added value of km-scale simulations to describe temperature over complex orography: the CORDEX FPS-Convection multi-model ensemble runs over the Alps *Clim. Dyn.* **62** 4491–514
Sobolowski S *et al* 2025 GCM Selection & Ensemble Design: Best Practices and Recommendations from the EURO-CORDEX Community (Bulletin of the American Meteorological Society) (https://doi.org/10.1175/BAMS-D-23-0189.1)
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
Stevens B *et al* 2024 Earth Virtualization Engines (EVE) *Earth Syst. Sci. Data* **16** 2113–22
Torma C, Giorgi F and Coppola E 2015 Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps *J. Geophys. Res.: Atmos.* **120** 3957–72
Trenberth K E and Zhang Y 2018 How Often Does It Really Rain? *Bull. Am. Meteorol. Soc.* **99** 289–98
Vicedo-Cabrera A M *et al* 2021 The burden of heat-related mortality attributable to recent human-induced climate change *Nat. Clim. Change* **11** 492–500
Wehner M F, Smith R L, Bala G and Duffy P 2010 The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model *Clim. Dyn.* **34** 241–7
Wilks D 2016 The stippling shows statistically significant grid points: how research results are routinely overstated and overinterpreted and what to do about it *Bull. Am. Meteorol. Soc.* **97** 2263–73
Wille J D, Koch R, Becker T and Fischer E M 2024 Extreme precipitation depiction in convection-permitting Earth System Models within the nextGEMS project accepted (https://doi.org/10.22541/essoar.173204240.08166674/v1)
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