Received: 31 May 2023 | Revised: 28 November 2023 | Accepted: 28 January 2024

DOI: 10.1002/joc.8393

RESEARCH ARTICLE

International Journal of Climatology RMetS

Climate change signals of extreme precipitation return levels for Germany in a transient convection-permitting simulation ensemble

Marie Hundhausen | **Hendrik Feldmann** | **Regina Kohlhepp** | **Joaquim G. Pinto**

Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Correspondence
Marie Hundhausen, Institute of Meteorology and Climate Research–Troposphere Research (IMK-TRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
Email: marie.hundhausen@kit.edu

Funding information
German Federal Ministry of Education and Research (BMBF), Grant/Award Numbers: 01LR2007B, 01LR2002B, 01LP1901A; AXA Research Fund

Abstract

The increase in extreme precipitation with global warming (GW) and associated uncertainties are major challenges for climate adaptation. To project future extreme precipitation on different time and intensity scales (return periods [RPs] from 1 to 100 a and durations from 1 h to 3 days), we use a novel convection-permitting (CP), multi-global climate model ensemble of COSMO-CLM regional simulations with a transient projection time (1971–2100) over Germany. We find an added value of the CP scale (2.8 km) with respect to the representation of hourly extreme precipitation intensities compared to the coarser scale with parametrized deep convection (7 km). In general, the return levels (RLs) calculated from the CP simulations are in better agreement with those of the conventional observation-based risk products for the region for short event durations than for longer durations, where an overestimation by the simulation-based results was found. A maximum climate change signal of 6–8.5% increase per degree of GW is projected within the CP ensemble, with the largest changes expected for short durations and long RPs. Analysis of the uncertainty in the climate change signal shows a substantial residual standard deviation of a linear approximation, highlighting the need for transient data sets instead of time-slice experiments to increase confidence in the estimates. Furthermore, the ensemble spread is found to be smallest for intensities of short duration, where changes are expected to be based mainly on thermodynamic contributions. The ensemble spread is larger for long, multi-day durations, where a stronger dependence on the dynamical component is ascribed. In addition, an increase in spatial variance of the RLs with GW implies a more variable future climate and points to an increasing importance of accounting for uncertainties.

KEYWORDS

convection-permitting, COSMO-CLM, extreme precipitation, multi-model ensemble, regional climate modelling

1 | INTRODUCTION

Record-breaking heavy precipitation events have become more frequent in Europe (Lehmann et al., 2015), exemplified by the severe flood in July 2021 with unprecedented impacts in Germany, Belgium and Luxembourg (Kreienkamp et al., 2021). With intensities above 150 mm/h, this event surpassed the 100-year return level (RL) provided by conventional rainfall risk products (Mohr et al., 2023), revealing the limitations of relying solely on past observations. To address these limitations and incorporate climate change, innovative heavy rainfall risk products are necessary. For example, the UK government has already integrated climate change allowances derived from climate projections into regulations, demonstrating the need for regional-scale climate information customized to meet specific user requirements (UK Environment Agency, 2012).

The change in heavy precipitation in a warming climate can be attributed to two main components: thermodynamic and dynamic processes. Thermodynamic changes alone lead to an intensification of precipitation extremes close to the Clausius–Clapeyron (CC) scaling, which describes the rate of increase in atmospheric moisture with warming, which increases by 7% per degree of warming, under the assumption that relative humidity remains constant. On the other hand, dynamic processes at various scales—from small (cloud-scale) to large (large-scale stability, humidity and large-scale circulation patterns)—impact extreme events (Fowler et al., 2021), potentially dampening or amplifying the change signal and causing regional variations (Fowler et al., 2021; Pfahl et al., 2017; Santos et al., 2016).

Scaling of the short-term (day-to-day) variability (e.g., in Berg & Haerter, 2013; Guerreiro et al., 2018; Lenderink & Van Meijgaard, 2008; Mishra et al., 2012; Westra et al., 2013), as well as pseudo global warming (GW) studies (e.g., Lenderink et al., 2019; Ludwig et al., 2023; Michaelis et al., 2017) mainly investigate the thermodynamic component. To consider the interplay between dynamic and thermodynamic processes, reliable climate projections are essential. While conventional regional climate simulations on a scale of tens of kilometres horizontal resolution have limited ability to represent extreme precipitation (e.g. Ban et al., 2014; Hohenegger et al., 2008), the new generation of higher resolution convection-permitting (CP) simulations with a grid spacing of ≤4 km that permit the explicit resolution of deep convection, appears to be a promising tool. The improvements comprise especially the intensities and frequencies of extreme precipitation, the diurnal cycle and orographically enhanced extreme precipitation (Ban et al., 2014, 2021; Hohenegger et al., 2008; Kendon et al., 2012; Prein et al., 2013; Sato et al., 2009). An added value of the temperature scaling of precipitation was found in Ban et al. (2014) and Chan et al. (2016). All in all, the improved skill in CP simulations provides confidence in their ability to reliably project future changes in extreme precipitation (Kendon et al., 2012; Lucas-Picher et al., 2021; Westra et al., 2014).

Due to high computational costs, CP studies are often limited to short simulation periods, specific domains and single ensemble members. First multi-decade simulations for the United Kingdom demonstrated increased short-duration precipitation beyond what coarser simulations captured (Kendon et al., 2014). Ban et al. (2015) could derive an intensification of extreme events with warming over an alpine domain that was asymptotically constrained by CC-scaling. Chan et al. (2022) derived change signals from a perturbed parameter ensemble comparing time slices from 1981–2000 to 2061–2080 of 30% for hourly RL intensities. However, due to limited simulation periods, these studies lacked representation of internal variability and uncertainty. Only recently, the publication by Kendon et al. (2023) showed for the first time results of hourly and daily precipitation from a transient ensemble over the United Kingdom that reveal 5%–15% increased intensity of hourly events with every degree of regional warming. As the study is based on an ensemble of CP simulations, estimation of uncertainty and internal variability was possible. However, the ensemble spread probably does not cover the whole uncertainty range, given the use of a perturbed parameter ensemble based on a single model family.

In our study, we use the KIT-KLIWA ensemble (Hundhausen et al., 2023) to track for the first time the evolution of future heavy rainfall in a CP, multi-global climate model (GCM) ensemble with four ensemble members and a transient simulation period from 1971 to 2100 (1971–2005: historical, 2006–2100: projection). The domain covers Southern and Central Germany. We investigate precipitation intensities of different scales: from hourly to multi-day (3 days) events and for different return periods (RPs) up to 100 a derived from extreme value statistics. Our goal is to cover the wide range of events relevant to various stakeholder applications, such as design drainage or regional planning. The transient simulation enhances the confidence in deriving climate change signals and allows to identify associated uncertainties and temporal variance in the projected climate change signals. The study addresses three research questions:

1. How do CP climate projections compare with observations or observation-based rainfall risk products? And are there patterns of the bias?
2. How do RLs evolve under GW as a function of event duration (ED) and RP?

3. What uncertainties are expected in the derived climate change signal?

The paper is structured as follows: Section 2 provides a description of the model and data. Section 3 presents the methods used for extreme value statistics and the climate change signal. Sections 4–6 present the results corresponding to the three research questions. The discussion and conclusion are provided in Section 7.

2 | DATA

2.1 | The CP ensemble

The CP simulations analysed in this study were conducted as part of the KLIWA project ('Climate change and consequences for water management', www.kliwa.de), extended in the ISAP project (Integrative city-regional adaptation strategies) within the German research program RegIKlim ('Regional information for action on climate change', www.regiklim.de), and joined to form the KIT-KLIWA ensemble.

The ensemble is a regional downscaling of four CMIP5 GCMs with the emission scenario RCP 8.5, namely MPI-ESM-LR (r1i1p1), EC-EARTH (r12i1p1), CNRM-CM5 (r1i1p1) and HadGEM2-ES (r1i1p1). The realization is given in brackets, using the standard *rNiMpL*, with *N* identifying the initial conditions, *M* the initialization method and *L* the perturbed physics version (Taylor et al., 2011). The four GCMs project different GW, due to different climate sensitivities—the equilibrium climate sensitivity ranges from 3.28 to 4.64K (cf. Supporting Information; Nijsse et al., 2020). Similar to Teichmann et al. (2018), we have chosen the GW within the period 1971–2000 as the reference period, where a GW of 0.46K compared to the pre-industrial conditions is assumed. Future GW is analysed over a 30-year running window. The GW since 1971–2000 is relatively similar for MPI-ESM-LR, EC-EARTH and CNRM-CM5 (Figure 1a). For HadGEM2-ES a stronger warming is found which is based on a higher climate sensitivity and in addition, HadGEM2-ES had a relatively cold period in 1971–2000. The 30-year periods centred around 2° and 3° of GW are hereafter referred to as GWL2 and GWL3 (details in Hundhausen et al., 2023).

The four GCMs are downscaled with the regional climate model (RCM) COSMO5.0-CLM9 (Sørland et al., 2021). The COSMO-CLM (CCLM) model is based on the former forecast model COSMO of the German weather service (Baldauf et al., 2011). CCLM combines a three-dimensional, non-hydrostatic numerical model of the atmosphere and a multi-layer soil-vegetation transfer model TERRA-ML. For additional information, the reader is referred to Rockel et al. (2008).

The downscaling strategy comprises three nesting steps: the first nest covers Europe with a grid resolution of 0.44°. The second nest with a resolution of 0.0625° is centred over Germany, extends into the neighbouring countries and completely encloses the Alps. The third, CP grid is centred over Southern Germany and has a resolution of 0.025° (approximately 2.8 km). While in the first two grids, convection is parametrized using the Tiedtke scheme (Tiedtke, 1989), in the third nest (2.8 km), only shallow convection is parametrized while deep convection is explicitly resolved. An overview of the nesting setup can be found in Hundhausen et al. (2023), where the first results of the KIT-KLIWA ensemble are used to study the changes in future heat wave characteristics. To eliminate boundary effects, a boundary region of 48 grid points was truncated. The resulting CP domain consists of 226 by 232 grid points (approximately 411,000 km^2) and is shown in Figure 1b. The hatched area in Figure 1b corresponds to the German part of the domain, which is analysed in the comparison of CP simulations with the rainfall hazard product KOSTRA (Section 4.2), and the climate change signals and their associated uncertainties (Sections 5 and 6) are derived for this area. The downscaled simulations driven by specific GCMs are referred to with an appended -C (GCM-Name-C).

Simulations for the two coarser grids were performed in a transient manner from 1971 to 2100. The CP simulations were initially computed for only three time slices, each including a 3-year spin-up period: 1968–2005, 2018–2050 and 2068–2100. In a second step, the CP simulations were extended by the two missing gaps, 2006–2020 and 2051–2070. A comparison of the overlapping years showed no significant differences in the mean precipitation or temperature.

2.2 | Observation data

We evaluate simulated rainfall data down to a temporal resolution of 1 h. Station-based observations of hourly precipitation are available from the German Weather Service (DWD) since 1995 (DWD, 2023; see Figure 1b). The DWD stations are operated in accordance with the WMO guidelines using 'Pluvio' pluviometers or 'Rain[e]' weighting precipitation sensors, which have an accuracy of ±0.1 mm or ±1% of the measured value. The initial stage of quality control involves a basic examination at the automatic stations (details in Kaspar et al., 2013). We restrict the analysis to stations with a large data coverage in the model historical period, defined as at least 9 years

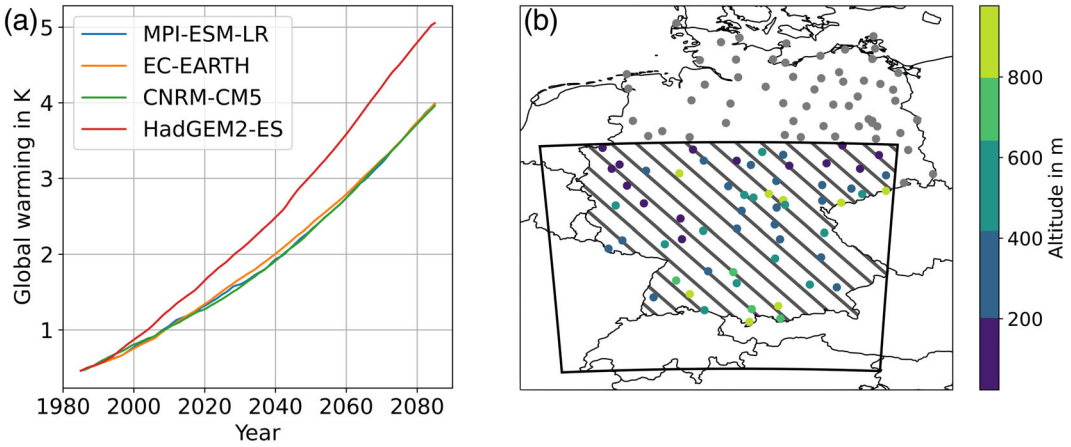

FIGURE 1 Figure (a) shows the global warming (GW) levels in the four driving GCMs over time. GW is averaged over a running window of 30 years and the reference of 0.46K for 1971–2000 is assumed similar to Teichmann et al. (2018). Figure (b) shows the study area with the simulation domain (thick black line). The dots locate the DWD stations that have been recording hourly precipitation since 1995 and have a data coverage of at least 9 years from 1995 to 2005—the coloured station locations within the simulation domain are evaluated in the study. The hatched area covers the German part of the domain and thus shows the part covered by both, the KOSTRA data set and the simulation. GCM, global climate model. [Colour figure can be viewed at wileyonlinelibrary.com]

of observations from 1995 to 2005 and stations that are located within the model domain.

As a second reference, we use the 'Coordinated heavy precipitation regionalization and evaluation' (KOSTRA) of the version KOSTRA-DWD-2010R (Junghänel et al., 2017). The data set, containing statistics on intensity, duration, and frequency, was derived from gauge observations in Germany in the period May to September 1951–2010. Based on 5-min resolution time series, a block rainfall method was applied, with adjustments to instantaneous events by multiplication by 1.14, 1.07, 1.04 and 1.03 for 5-, 10-, 15- and 20-min windows and no adjustment for longer durations. For durations below 4 h, a precipitation-free time period of at least 4 h was defined, while for longer durations, the required time period is equal to the duration itself. For daily and longer durations, station observations were first aggregated regionally in the REGNIE data set ('REGionalisierte NIEderschlagshöhen'; Rauthe et al. (2013)). A Peak Over Threshold (POT) method was applied to station data. The used threshold is determined by the length of the time series such that the threshold is restricted from including more data than the number of years times 2.718. Subsequently, an exponential distribution was fitted to the data. To avoid discontinuities in intensities, an adjustment across different EDs was applied. The results are regionalized using location as well as orographic information (orographically modified variogram analysis, detailed in Malitz and Ertel (2015)).

Uncertainties of the extreme value statistical evaluation, including regionalization due to the spatial and temporal variability of the heavy precipitation and the resulting limited representativeness of the station values, the long-term natural climate variability, the measurement error and the limitations of the station-based extreme value statistical approach, are expected. The DWD provides the following tolerance limits 10% for 1 a $<$ RP $<$ 5 a, 15% for 5 a $<$ RP $<$ 50 a and 20% for 5 a $<$ RP $<$ 100 a.

The data set is available from the German Weather Service (DWD) and provides RLs for RPs as of 1 a, 2 a, 3 a, 5 a, 10 a, 20 a, 30 a, 50 a and 100 a with an ED of 5, 10, 15, 20, 30, 45, 60, 90, 120, 180, 240, 360, 540, 720, 1080, 1440, 2880 and 4320 min over a 8.2 km times 8.2 km grid on German territory. KOSTRA was developed for practical applications such as urban drainage, but has also been used in the peer-reviewed literature to assess extreme events in climate models (Berg et al., 2019; Ettrichrätz et al., 2023; Poschlod et al., 2021).

3 | METHODS

3.1 | Determination of extreme values

In general, two steps are required to apply extreme value statistics: selecting the extremes and fitting an extreme value distribution. Various procedures for event selection and extreme value distributions are used in the literature. As elaborated by Pendergrass (2018), deciding which definition of extreme precipitation to use is critical, especially when communicating climate change information to other sectors. As our work relates to the KOSTRA data set and is intended to be close to the hydrological

application for Germany, the method is based on the German DWA-A 531 guideline (DWA, 2012), which is the basis for the KOSTRA-DWD-2010R data set:

1. For the selection of extremes, a partial series is generated. This is done over the selection of the $e=2.72$ times the number of analysed years' highest peak values of individual rain events. Rain events are separated from each other by a rain-free period that was adapted to 24 h to save computational costs.
2. Second, a function is fitted to the data to provide a relationship for the RL dependent on the RP. Therefore, plotting positions of the events of the partial series are derived using Equation (1), with L sample size, M length of the time series in a and k running index of the sorted samples ($k=1$ largest, $k=L$ smallest intensity). Then, an exponential function in the logarithmic form of Equation (2) is fitted to the data using a least square polynomial fit. The fit parameter slope w_p and offset u_p define the intensity distribution for different RP.

$$\mathrm{RP}(k)=\frac{L+0.2}{k-0.4}\frac{M}{L}. \tag{1}$$

$$\mathrm{RL}(\mathrm{RP})=u_p+w_p\cdot\ln\mathrm{RP}. \tag{2}$$

The method is applied to time series from station data and for the CP simulations for time series for each ensemble member and at each grid point individually. For the reference period, the analysis of the CP simulation is conducted for the years 1971–2005, which is the historical period of the driving GCMs. In addition, an evaluation at the nearest neighbour cells to the station location was performed on the CP simulation time slices for the available measurement period, usually 1995 to 2005. In the projection period, consecutive 30-year running time slices were evaluated, starting from 2006 to 2035, ending with 2071–2100.

The ED of 1, 6, 12, 24 and 72 h is analysed, which was evaluated by aggregating the hourly time series over running windows of the respective length prior to the event selection. In the reference period, the ED 2, 4, 9 and 48 h were also evaluated. In order to correct for systematic underestimation of precipitation intensities due to an equidistant sampling window for ED similar to the sampling frequency, the suggested correction factors from guideline DWA-A 531/KOSTRA were applied to increase precipitation depth for small ED: correction factor 1.14, 1.07, 1.04 and 1.03 for the number of 1, 2, 3 and 4 measurement points per ED. In our case, an hourly resolution of the simulation and the observation is available, which leads to corrections for ED $\leq$ 4 h.

3.2 | Climate change signal of extreme precipitation

According to the IPCC report, the climate change signal of extreme precipitation with GW is relatively independent of the forcing scenario (IPCC, 2021). Therefore, we evaluate changes in extreme precipitation with respect to GW instead of time. This has the advantage of allowing the comparison of models with different climate sensitivities, as in the present KIT-KLIWA ensemble. Moreover, the analysis with respect to GW will facilitate comparisons with other modelling studies based on different emission scenarios or considering different time periods.

To compare changes in precipitation intensity with GW over different RP and ED, we make use of the concept of relative climate change signal (or change factor [CF]/uplifts) similar to Chan et al. (2022) to become more independent of the absolute values that are subject to model bias (Ho et al., 2012). We evaluated the CF for every ensemble member individually according to Equation (3) over the absolute change in precipitation normalized by the simulation's value in the reference period 1971–2005 ($\mathrm{pr}_{\mathrm{hist}}$). To derive a robust change signal, the CF is evaluated over the slope of a linear regression to the respective trajectory of precipitation intensity over GW (Δpr in mm/(K global warming)). CFs are evaluated for the median of the grid points in the evaluation domain (hatched area in Figure 1b), except for the spatial analysis in Figure S7, where the procedure is applied to the model output at each grid point individually.

$$\mathrm{CF}(\mathrm{RP},\mathrm{ED})=1+\frac{\Delta\mathrm{pr}(\mathrm{RP},\mathrm{ED})}{\mathrm{pr}(\mathrm{RP},\mathrm{ED})_{\mathrm{hist}}}. \tag{3}$$

4 | COMPARISON OF SIMULATION AND OBSERVATION-BASED DATA

4.1 | Comparison with station data

The observation of hourly precipitation intensity from stations considered within this study ranges from 0 to 67 mm h^{-1}. The largest fractional contribution is expected for small intensities (I), with approximately equal contribution for $I\leq1$ mm h^{-1} (Figure 2a). With increasing I, the fractional contribution declines. For large I, scattering due to sparse data coverage is visible.

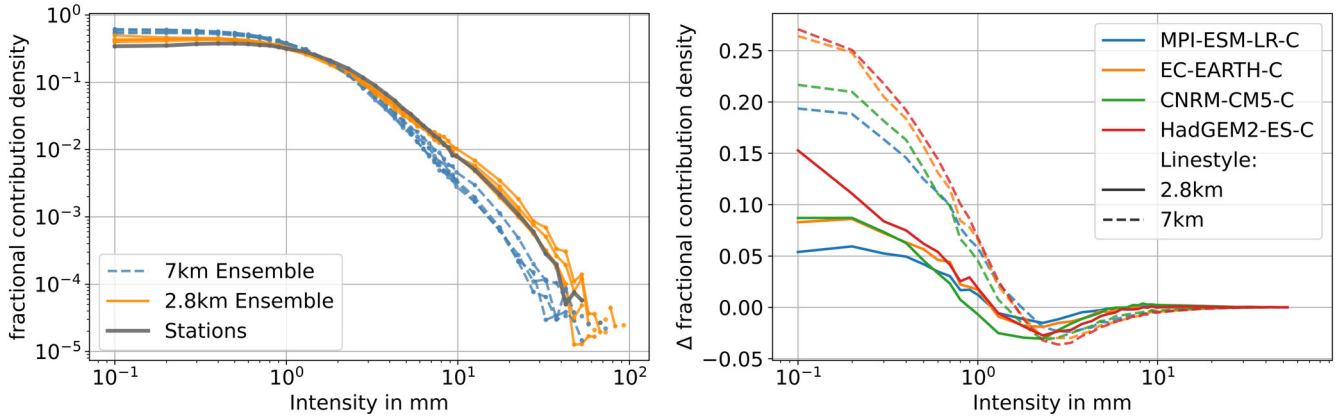

FIGURE 2 The density distribution of fractional contributions in the historical period (1971–2005) of hourly rainfall from station measurements and the according results in the nearest neighbour cell of 2.8 and 7 km simulation output is shown in (a). Coloured single lines correspond to the single ensemble members. Figure (b) provides the differences between observation and simulations for the four ensemble members. The 2.8 km resolution simulation is represented in thick lines and according to colour, the 7 km one is in dashed lines. The bin width for both panels is 0.1 mm for the intensity range from 0 to 1 mm, aligned with the resolution of the station measurement. Bin width is 0.5 mm for the intensity range from 1 to 10 mm and 5 mm for the intensity range from 10 to 100 mm. [Colour figure can be viewed at wileyonlinelibrary.com]

Comparing those observation results to the fractional contribution of intensities in the nearest grid point of the 2.8 km resolution CP simulation, the simulations replicate the shape of the curve: a plateau from 0.01 to 1 mm and a declining contribution for $I>1$ mm. However, there is a systematic bias between the ensemble and the observations (Figure 2b). All ensemble members overestimate the contribution of small intensities, with the largest overestimation seen in HadGEM2-ES-C (solid red line) and the most realistic representation for MPI-ESM-LR-C (solid blue line). Medium intensities exceeding 0.9 (CNRM-CM5-C) or 1.2 mm (MPI-ESM-LR-C, EC-EARTH-C and HadGEM2-ES-C) are systematically underestimated by the simulations. The maximum underestimation occurs around 2 mm precipitation, with the largest underestimation in CNRM-CM5-C and again the best representation in MPI-ESM-LR-C (Figure 2b). Observations and simulations converge for higher values, generally with a slight overestimation of extreme events exceeding 6–8 mm (the range corresponds to the ensemble spread).

The coarser, convection-parametrizing simulations on the 7 km grid show similar patterns of over- and underestimation of small to medium intensities as the 2.8 km simulations, but the differences with respect to the observation are amplified and the overestimation for small I is extended to larger intensities of 1.5–1.8 mm, above which the frequency of the respective intensity is underestimated. This underestimation at higher I is also amplified, with the maximum deviation occurring around $I=3$ mm. This underestimation persists even for the highest intensities, and there is no full convergence between the coarser simulations and the observations (Figure 2a).

Comparing the results, the CP simulations better represent the whole spectrum of precipitation intensities than the coarser 7 km resolution simulations. The features of the density distribution at high intensities are especially more realistically reproduced on the CP scale. However, a slight overestimation of the most extreme events has to be expected using CPM. This finding holds for a short temporal resolution of 1 h. For longer aggregation times, the CP simulations tend to systematically overestimate the contribution of high intensities and there is an increased bias compared to the coarser grid. This was tested with aggregated station data as well as using the gridded daily data set HYRAS ('Hydrometeorological raster data', cf. Rauthe et al., 2013, for HYRAS) (Supporting Information in Figure S1). In the analysis, data with different horizontal resolutions are compared. As precipitation statistics are related to horizontal resolution, and especially short duration (convective) precipitation extremes are shown to increase in intensity with increasing resolution (Eggert et al., 2015), the shown improvement of the high intensities from 7 to 2.8 km is partly attributed to the finer grid resolution in the CPM. As expected, the difference in the CPM compared to the 7 km simulation decreases when the comparison is made on a common grid (see Supporting Information in Figure S2 for a comparison on the common 7 km grid). However, even when comparing on a common grid, the improvement is reproduced for both, low and high intensities. Therefore, we conclude that a further improvement comes from the CPM and presumably a more realistic representation of precipitation due to the explicit resolution of deep convection.

The statistically derived RLs from the time series of the station data compared to the results of the CP simulations at the respective grid points show the following patterns (Figure 3):

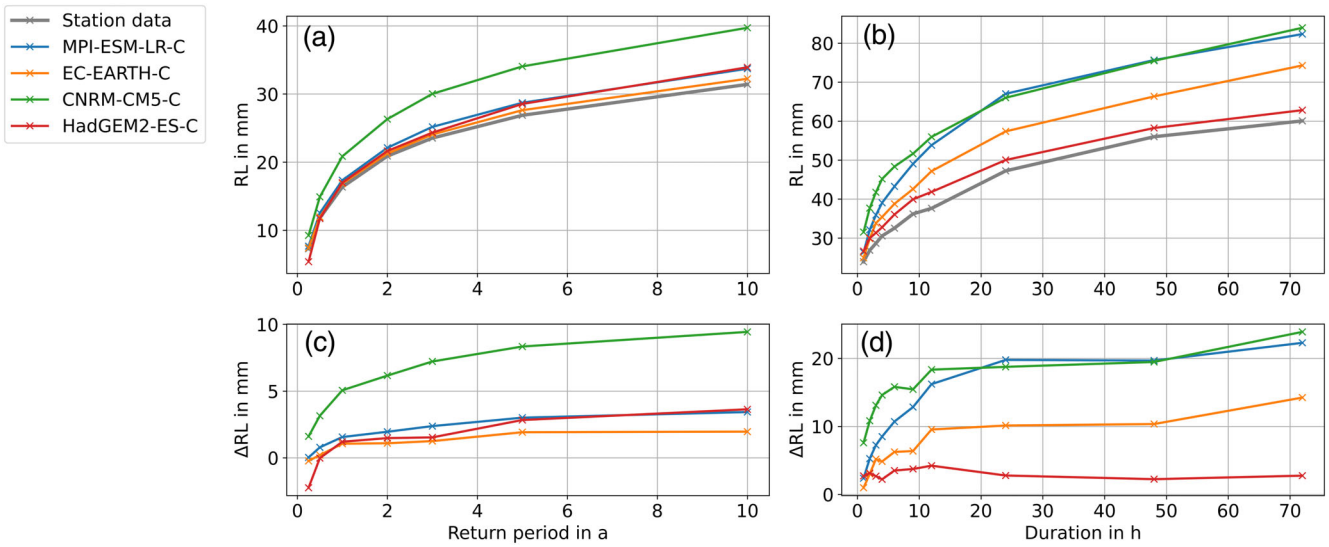

FIGURE 3 Median RLs in station observations and CP simulations at the nearest neighbour grid point for different RPs with fixed ED of 1 h (a) and for different ED with fixed RP of 5 a (b). In the second row, the differences between simulations and observations for the above are shown. CP, convection-permitting; ED, event duration; RL, return level; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

- For different RPs: the CP simulations generally reproduce the RLs derived from the station data within a deviation of 4 mm for the three ensemble members MPI-ESM-LR-C, EC-EARTH-C and HadGEM2-ES-C as shown for the duration of 1 h (Figure 3a). The function RL (RP) has a slightly steeper slope in all CP simulations leading to an increasing overestimation of RL with increasing RP. Thus, the differences are smallest for small RPs, and sub-yearly results tend to be underestimated by EC-EARTH-C and HadGEM2-ES-C (Figure 3c). The largest error is found for CNRM-CM5-C, with an overestimation of up to 10 mm for large RP.
- For different ED: taking the example of a 5-year RP and evaluating different ED (Figure 3b,d), the overestimation of RL is generally smallest for the shortest duration (1 h) and increases with accumulation time. The trend is most prominent for MPI-ESM-LR-C, where the overestimation ranges from 2 to 22 mm. Only in HadGEM2-ES-C, the described pattern is not visible. Here, the overestimation is similar across all EDs, with approximately 3 mm.

4.2 | Comparison with aggregated observations in KOSTRA

The aggregated rainfall hazard data set KOSTRA provides gridded RLs covering Germany, illustrated by the hourly 10-year RL (RL10) in Figure 4a. RL10 ranges from 27 to 48 mm. Higher values are visible in the Black Forest and the alpine foreland in the south, while flatter regions in the north generally show a smaller RL10. An overview of the geographical regions in Germany is provided in the Supporting Information (Figure S3).

The ensemble median of RL10 of the CP simulations during the reference period (1971–2005) ranges from 28 to 50 mm in the German part of the domain (Figure 4b). There is no clear dependence on orographical structures within this domain for the ensemble median. Same is true for the single ensemble members (additional information in Figure S4). The ensemble median shows a slight pattern of lower values in the centre of the domain, with RL slightly increasing towards the north and south. Strong spatial patterns in the CP results are detected only outside the evaluation domain in the Alps. For longer ED, spatial patterns of extreme precipitation associated with elevation become apparent in the CP ensemble. For example, all ensemble members reproduce higher daily RL10 in mountainous regions such as the Black Forest compared to the flat surroundings. This pattern is slightly more pronounced in the CP simulations than in observations, leading to an overestimation of RL10 in these mountainous areas (additional information in Figure S5).

The lack of spatial structures for short EDs is consistent with the analysis of observational data, for example, radar data in Lengfeld et al. (2019), who concluded that the intensity of precipitation intensities of short compared to longer durations is less dependent on orography. Therefore, we attribute the lack of spatial patterns mainly to uncertainty due to the relatively short time series length compared to the RP of interest. We expect that longer time series are needed or that using a regression model, such as that used to derive the KOSTRA data set,

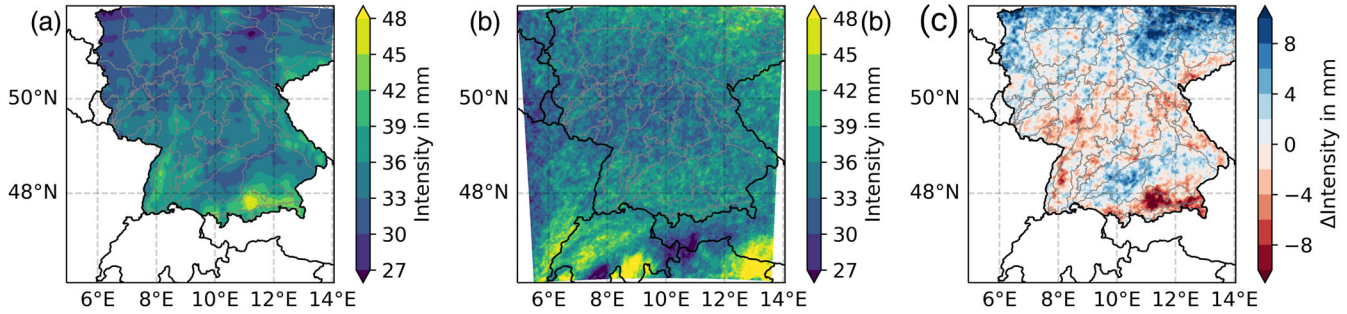

FIGURE 4 Spatial distribution of the return level with RP = 10 a, ED = 1 h in KOSTRA (a), the median of the ensemble results (b) and the differences (KOSTRA subtracted from the simulations) (c). ED, event duration; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

could improve the representation of the expected patterns associated with orography, which likely explains the differences with KOSTRA.

The differences between hourly RL10 in KOSTRA and the CP ensemble lead to an underestimation of RL10 for areas with high KOSTRA values (Alpine foreland and low mountain ridges in the southwest of the domain) and a general overestimation in the north of the domain (Figure 4c). Seemingly random spatial deviations between CP and KOSTRA appear superimposed on this pattern. Moreover, exceptionally high RL10 in KOSTRA in the Alpine foreland stand out (lon ≈ 11°E, lat ≈ 48°N). At this location, there is no orographic feature that could explain a mechanism for such a strong local maximum, but it is assumed that an anomalous observed local event has biased the statistics. This is thus an example of how the method can depend on individual events even for comparatively short RP (as in the example of 10 years) and shows that the available, finite observation time series do not always permit a robust estimate.

To complement the comparison, further extreme intensities with different RPs and EDs are analysed in KOSTRA and the reference period (1971–2005) of the CP ensemble. Based on the finding of random spatial deviations in the CP results for short ED, we use the median of the grid point results for the comparison. Both data sets are evaluated over the shared part of the domain (hatched area in Figure 1). The following patterns emerged from the comparison (Figure 5), the values of the relative differences are provided in Table S1 in the Supporting Information.

- For different RP, constant ED: two different regimes are apparent. For ED < 24 h (small-scale or predominantly convective events), the relative difference between CP simulations and KOSTRA decreases (becomes more negative) with increasing RP (Figure 5c). This behaviour is determined by the underestimation of the slope of the fitted function RL (RP), which tends to be too small for the ensemble members MPI-ESM-LR-C and EC-EARTH-C (Figure 5a). For HadGEM2-ES-C and CNRM-CM5-C, the slope agrees well with KOSTRA, resulting in a roughly constant absolute overestimation in the simulations. Apart from ED = 1 h, the absolute differences almost always increase with RP, except for further few sub-daily events in EC-EARTH-C (see Supporting Information: Figure S6).
- In the range of ED ≥ 24 h (large-scale events), the pattern in the relative differences is reversed for all GCM: the overestimation increases with RP.
- For different ED, constant RP: Generally, the difference between KOSTRA and CPM is smallest for ED = 1 h, with negative values for the realizations driven by MPI-ESM-LR and EC-EARTH. For ED < 24 h, the difference becomes larger with increasing ED, leading to an increasing overestimation of RL projected by MPI-ESM-LR-C, EC-EARTH-C and CNRM-CM5-C. No continuous increase is visible in the HadGEM2-ES-C realization and partially in CNRM-CM5-C for low RP.
- For ED ≥ 24 h, a reversal is again apparent. Overestimation is largest for ED = 24 h and decreases with higher aggregation times. For ED = 72 h, negative bias is found in HadGEM2-ES-C for low RP. Exceptionally large deviations between simulation and KOSTRA are found for daily ED.

The deviation between the spatial median in KOSTRA and CPM ranges from −7% to 10% for hourly duration, excluding the outlier CNRM-CM5-C with significantly higher deviations (up to 27%). For sub-daily events with ED > 1 h, a positive bias/overestimation is expected for the model (2%–23%, excluding CNRM-CM5-C). Longer events (ED ≥ 24 h) show the largest overestimation and variance with −2% to 27%

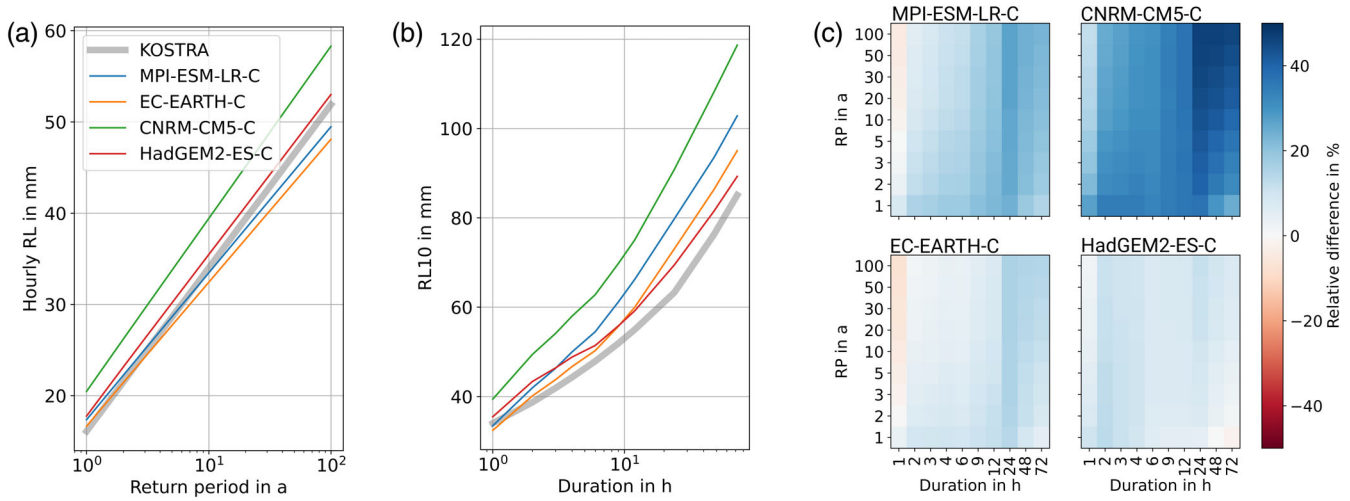

FIGURE 5 RLs in KOSTRA and CPM are shown in (a) for different RPs keeping ED constant with 1 h and in (b) for different EDs with constant RP of 10 a. The relative differences in RLs for all analysed RPs and EDs are shown in (c) for the four ensemble members separately—top left: MPI-ESM-LR-C, top right: CNRM-CM5-C, bottom left: EC-EARTH-C and bottom right: HadGEM2-ES-C. The values of the relative differences are provided in the Supporting Information (Table S1). ED, event duration; RL, return level; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

deviation—excluding CNRM-CM5-C with up to 47% overestimation.

The best agreement and no systematic over-/underestimation compared to KOSTRA was found in the ensemble for hourly ED and medium RP, which is in this case around an RP of 10 a, which corresponds to one third of the analysed time series length. Therefore, hourly RL10 serves as an example for more detailed analysis of the distribution in Figure 6. Both KOSTRA and the four ensemble members exhibit a positive skew, with the CP ensemble displaying more pronounced tailing and larger values for the most extreme events. The distribution of CP simulation results is generally smoother than in KOSTRA. RLs derived from station data are generally lower than KOSTRA or CP ensemble results, with larger scattering, which is mainly attributed to shorter time series which increase uncertainty in extreme value statistics.

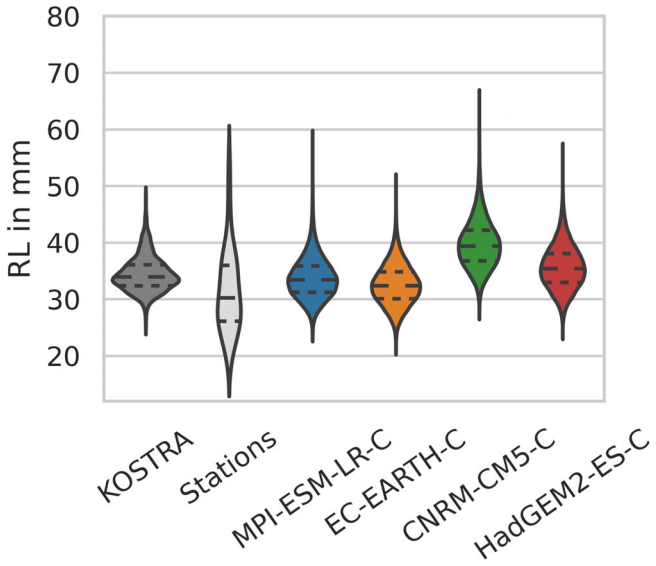

FIGURE 6 RLs for ED = 1 h and RP = 10 a in KOSTRA, derived from the available station data and from the four ensemble members in the CP ensemble for 1971–2005. The distribution is evaluated over the German part of the model domain for KOSTRA and CP simulations, and over all available stations within this domain for station data. CP, convection-permitting; ED, event duration; RL, return level; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | Implications

We conclude that there is adequate agreement between CP ensemble and station measurement for hourly rainfall intensities, with an added value in the CP simulations beyond the mere higher resolution compared to coarser simulations. Confidence in RLs in the CP simulations is higher for short ED (hourly to sub-daily), as they are in better agreement with the observations (station and KOSTRA). The mean bias is not necessarily smaller with short RP. In general, there is an overestimation by the CP simulations, except for ED = 1 h. The positive skew of the distribution of hourly RL10 is represented by the simulations, but there is a greater magnitude of the largest possible extreme events. We find higher confidence in the ensemble members driven by MPI-ESM-LR, EC-EARTH and HadGEM2-ES. CNRM-CM5-C is likely to overestimate extreme precipitation.

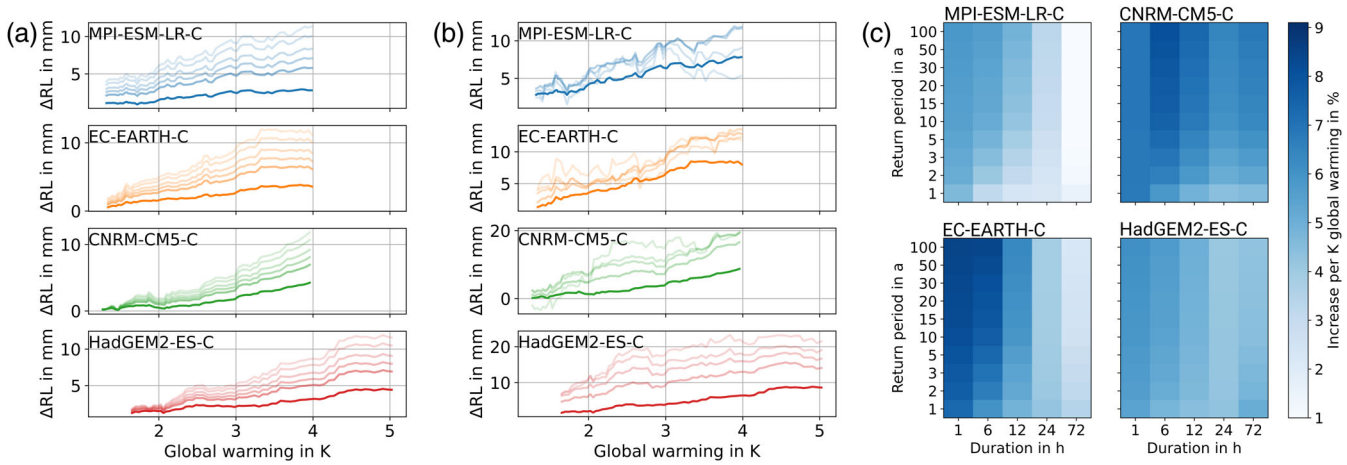

FIGURE 7 (a) Projected change compared to the reference period (1971–2005) for ED = 1 h and variable RP = 1, 5, 10, 20, 50 and 100 a (with increasing transparency) for MPI-ESM-LR-C, EC-EARTH-C, CNRM-CM5-C and HadGEM2-ES-C (from top to bottom), (b) for RP = 10 a and variable ED = 1, 6, 12, 24 and 72 h (with increasing transparency). (c) Relative percentage increase of RL, normalized to the corresponding value in the reference period 1971–2000. The increase is evaluated over a linear regression in the projection period—top left: MPI, top right: CNRM-CM5-C, bottom left: EC-EARTH-C and bottom left: HadGEM2-ES-C. ED, event duration; RL, return level; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

Our analysis implies that we have more confidence in the simulated values that are similar to those in KOSTRA. However, it is important to consider the limitations of this comparison. KOSTRA and the CP ensemble consider different time periods. Moreover, also observational data sets come with limitations in accurate, high-resolution rainfall measurements, especially for intense events accompanied by strong winds (e.g., Sieck et al., 2007). Data retrieval and interpolation in complex terrain is challenging due to sparse coverage of ground-based measurements and radar limitations (e.g., Henn et al., 2018). Systematic errors can be expected in the observations, especially for short events and complex terrain. Moreover, the large bias between the model and KOSTRA associated with daily ED raises the question whether this effect could originate from a discontinuity in the data source in KOSTRA or is an impact of applied corrections in KOSTRA. The bias compared to the simulations is therefore treated here as a reference rather than the ultimate truth. Based on these considerations, observation data was therefore not used for bias correction but to account for systematic model errors; change signals of percentage increase are derived in the following similar to Chan et al. (2022).

5 | FUTURE DEVELOPMENT OF HEAVY PRECIPITATION

The projection period 2006 to 2100 of the CP ensemble covers GW levels from 1 to 5K, depending on the GCM (Figure 1). The following patterns are derived from the change signal in Figure 7:

- If ED is kept constant and RP is varied, the main change is an increased slope with greater RP. The shape of the curve (local maxima and minima) remains independent of RP (Figure 7a), as expected for the displayed values from the same underlying extreme value fits. For the relative increase with GW, approximated by a linear regression over the normalized RL-GW curves, three out of four ensemble members project an enhanced increase of RL with RP for ED < 24 (Figure 7c). Only CNRM-CM5-C shows an approximately constant CF for all RPs for ED = 1 h. For large ED = 72 h, a reverse pattern is observed for the majority of the ensemble members (MPI-ESM-LR-C, EC-EARTH-C and HadGEM2-ES-C), where the relative increase is larger for small RPs than for large RPs. The apparent increase is small and ranges from 0.8 (MPI-ESM-LR and HadGEM2-ES) to 1.2 PPT (EC-EARTH-C). There is no agreement on a common trend in the projections at ED = 24 h and the change signal is approximately constant, especially for EC-EARTH-C and HadGEM2-ES-C.
- If RP is kept constant and ED is varied, the projected increase is substantially affected, altering the shape of the intensity trajectory with GW (Figure 7b). Different local maxima and minima emerge. However, similarities in the curves indicate that the same event appears to be partially sampled at several EDs. The effect of ED on the shape of the curve depends crucially on the

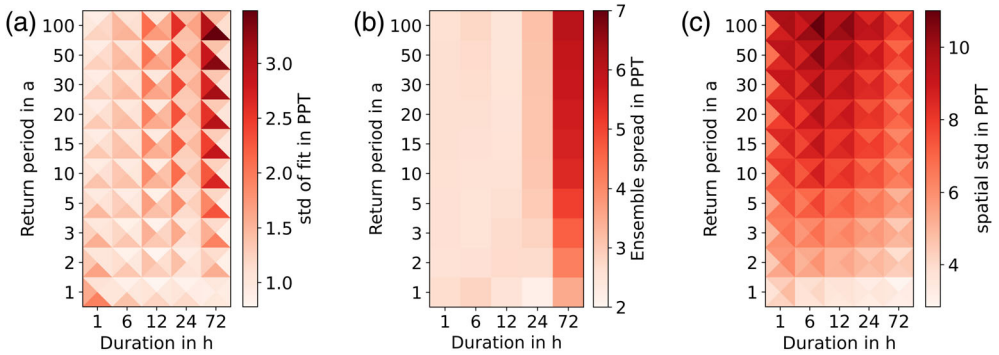

FIGURE 8 Different types of uncertainty in the climate change signal of extreme precipitation for different RP and ED. (a) The residual standard deviation is calculated for the linear regression that was used to approximate the change signal of extreme precipitation with global warming. (b) The ensemble spread corresponds to the range between the minimum and maximum ensemble member projections of the relative change signal. (c) The spatial standard deviation is calculated over the increase at each grid point in the domain. For (a) and (c), the sections in the graph correspond to the ensemble members, top: MPI-ESM-LR-C, right: EC-EARTH-C, bottom: CNRM-CM5-C and left: HadGEM2-ES-C. ED, event duration; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

ensemble member. In terms of the relative change signal (Figure 7c), increasing ED generally leads to decreasing CF for DS < 24 h. Only CNRM-CM5-C shows lower values for ED = 1 and thus no steady decrease. For ED = 72 h, there is no common trend in the ensemble. While in EC-EARTH-C and MPI-ESM-LR-C, the pattern is continued and the smallest increases are for ED = 72 h, CNRM-CM5-C and HadGEM2-ES-C show similar or slightly higher magnitudes for ED = 72 h compared to ED = 12 h.

The climate change signal is therefore expected to be largest for short ED and long RP. The projected increase in extreme precipitation per degree warming in this case reaches up to 8.5% (EC-EARTH-C) or up to 5.9% and 6.1% (MPI-ESM-LR-C and HadGEM2-ES-C). CNRM-CM5-C shows a maximum change signal of 8.1% for sub-daily events, which is well within the range of other models assigned higher confidence due to better agreement with observations and similar future behaviour. The change signal is in the range of the CC scaling. Super CC scaling is projected by EC-EARTH-C and CNRM-CM5-C. Events with longer EDs also increase with GW but to a lesser extent from 0.8% (MPI-ESM-LR-C) over 2.3% (EC-EARTH-C) to 4.3% (HadGEM2-ES-C) for 3-day ED and large RP. CNRM-CM5-C indicates that also for those long ED larger change signals may be expected (up to 6.8%). Please refer to the Supporting Information for an overview of all extracted change signals (Table S2).

In summary, for sub-daily events, the change signal increases with RP, whereas for large-scale multi-day events (ED = 72 h), the change signal decreases with RP. The highest climate change signals, ranging from 6% to 9%, are associated with the shortest ED, falling within the range of CC-scaling and partly super CC-scaling.

6 | UNCERTAINTY IN THE CLIMATE CHANGE SIGNAL

Apart from identifying the major patterns of change in extreme precipitation, the continuous assessment over GW using a running 30-year window reveals the inherent variance in the increase in extreme events, visible as local minima and maxima in the trajectories of RLs over GW (Figure 7a,b). This variance is approximated by the residual standard deviation of the linear fit, which is a measure of how well a linear relationship describes the future development of RLs with GW. The residual standard deviation ranges from 0.8 and 2 percentage points (PPT) (10th to 90th percentile) for most RP-ED combinations (Figure 8a). The highest residual standard deviations are found for large EDs in MPI-ESM-LR-C, HadGEM2-ES-C and CNRM-CM5-C. However, the magnitude differs for the individual ensemble members. CNRM-CM5-C and HadGEM2-ES-C reach maximum residual standard deviation of 3.5 and 3.0 PPT for long ED and large RP, while MPI-ESM-LR-C and EC-EARTH-C show maximum values of 1.4 and 1.7 PPT for all tested configurations. For ED > 6 h, all GCMs show an increasing residual standard deviation with increasing RP. However, this pattern is reversed for most ensemble members (MPI-ESM-LR-C, CNRM-CM5-C and HadGEM2-ES-C) for ED = 1 h and no consistent result is found for ED = 6 h. The derived trends at large ED, as the apparent decreasing change signal with RP for ED = 72 h, are thus associated with

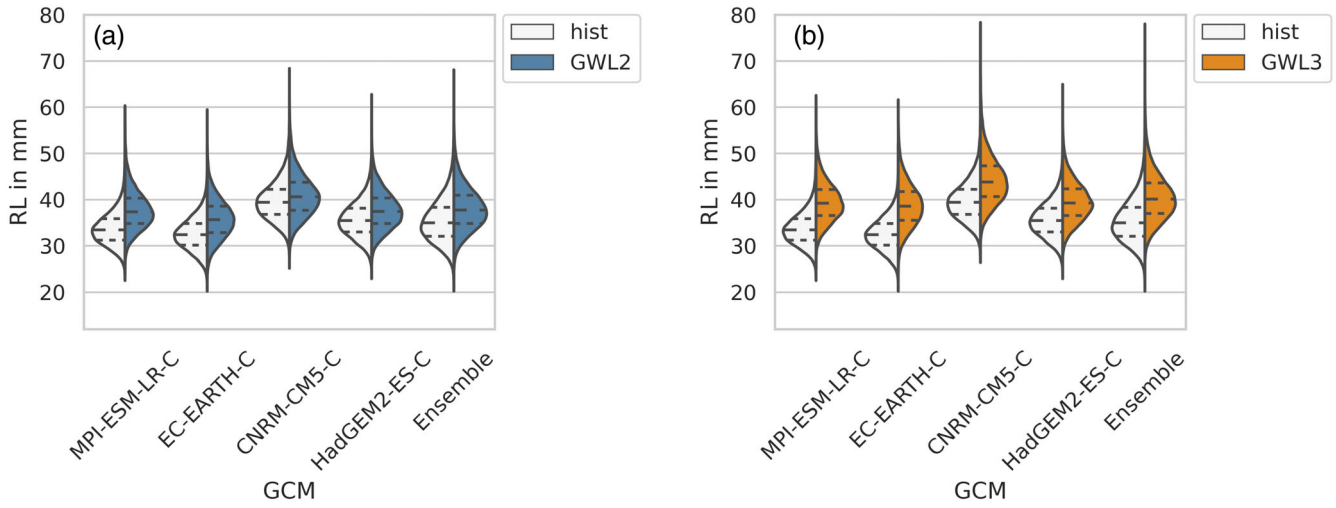

FIGURE 9 Distribution of 1 h RP10 over the evaluation area for GWL2 (a) and GWL3 (b). The distribution in white represents the results from the reference period 1971–2005. RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

greater uncertainty. These findings suggest that a linear regression describes the trend better for short EDs and less accurately for large ED and large RP. The availability of a transient simulation allowed local minima and maxima in the climate change signal to be smoothed. The analysis suggests that caution should be taken when applying a linear assumption to the analysis of time slice experiments, as significant stochastic errors can be expected, in contrast to transient data series.

Further uncertainties arising from differences in GCM formulation are expected and can be estimated from the ensemble spread (Figure 8b). The ensemble spread in the relative increase with GW depends mainly on the ED and is between 2 and 3 PPT for the majority of the EDs studied up to 24 h and is relatively independent of RP for ED < 24 h. The spread is significantly higher for ED = 72 h up to 6 PPT, mainly due to the exceptionally low RL changes projected by MPI-ESM-LR-C. The spread increases with RP for ED = 24 and 72 h. The relationship found indicates that the spread is particularly pronounced for long events (72 h), which are thought to depend on large-scale circulation patterns predetermined by the GCM and propagated into the RCM domain. Events of shorter duration, for which little spread is found here, are likely to be mainly controlled by the RCM, which is identical for all ensemble members in our study.

In contrast, the spatial variance depends mainly on the RP (Figure 8c). For all ED, it increases with RP up to 9 PPT (HadGEM2-ES-C), 10 PPT (MPI-ESM-LR-C, CNRM-CM5-C) and 11 PPT (EC-EARTH-C) for RP = 100 a. MPI-ESM-LR-C tends to show higher variance for shorter durations, while HadGEM2-ES-C shows the highest variance for longer durations of 12 and 24 h. For CNRM-CM5-C and EC-EARTH-C, the maximum spatial variance is centred around ED = 6 and 12 h. The magnitude of the spatial variance is similar for all ensemble members. For large RP (approximately RP > 15 a), the spatial variance is in the order of magnitude of the change signal itself. The results highlight uncertainties in estimating RLs from time series with a length close to or smaller than the RP. This implies that the robustness of a single grid point result is not given and that information can be derived from the consideration of multiple points only.

Analysis of the spatial patterns with respect to the climate change signal shows no dependence on the location or orographic features (see Supporting Information: Figure S7). Because of this stochastic nature, we treat the spatial distribution of the individual grid point results as an estimate of uncertainty. On exemplary examination of the evolution of RL10 with ED = 1 h and RP = 10 a at GWL2 and GWL3 in Figure 9, the main feature is a shift of the distribution. All ensemble members project a positive shift of the median from historical, over GWL2 to GWL3 (Table 1). The distribution is right-skewed—with longer tailing towards higher RLs, indicating the possibility of local exceptionally high RLs. The skewness does not show a clear climate change signal, remaining relatively constant for the ensemble members EC-EARTH-C, and HadGEM2-ES-C, slightly increasing for CNRM-CM5-C, and decreasing for MPI-ESM-LR-C (Table 1). The distribution widens for all ensemble members from historical to GWL2 to GWL3, as indicated by the increase in the interquartile range (Table 1). This effect is strongest for CNRM-CM5-C (increase of 1.5 mm) and EC-EARTH-C (increase of 1.2 mm) and less pronounced for MPI-ESM-LR-C and HadGEM2-ES-C with 0.9 and

TABLE 1 Statistics over the distribution of hourly 10-year RL sampled all grid points in the domain.

	Median in mm			Interquartile range in mm			Skewness		
	Hist	GWL2	GWL3	Hist	GWL2	GWL3	Hist	GWL2	GWL3
MPI-ESM-LR-C	33.5	37.4	39.2	4.7	5.5	5.6	0.68	0.52	0.38
EC-EARTH-C	32.4	35.7	38.6	4.7	5.7	6.2	0.32	0.29	0.33
CNRM-CM5-C	39.4	40.6	43.8	5.4	6.1	6.7	0.5	0.55	0.63
HadGEM2-ES-C	35.5	37.5	39.3	5.1	5.6	5.8	0.37	0.43	0.4

Abbreviation: RL, return level.

0.7 mm. The broadening of the distribution indicates that extreme precipitation will become more variable in the future. These patterns derived from RL10 are largely consistent for events of further RPs and duration, and are provided in the Supporting Information for RPs between 5 and 30 years in both hourly and daily EDs (Table S3). The median increases consistently for all configurations. The interquartile range shows an increase except for the ensemble member MPI-ESM-LR-C for the daily ED in GWL2. In general, the change in RP alone has little effect on the observed patterns within the ensemble, but amplifies the changes for the median and interquartile range. The skewness shows no common projected patterns even in further configurations.

In summary, sensitivity analysis provides evidence that there is a substantial residual standard deviation in the linear fit (5% confidence interval from 0.8 to 2 PPT) and that scatter in the climate change signal is expected, especially for large-scale events. The ensemble spread in the multi-GCM, single-RCM ensemble is significantly smaller for small-scale events (2–3 PPT) than for large-scale events (up to 6 PPT). Analysis of the spatial distribution of the climate change signal revealed a standard deviation in the order of magnitude of the climate change signal itself for large RP. Further analysis of the spatial distribution shows that the change in mean RL is accompanied by an increase in variance.

7 | DISCUSSION AND CONCLUSION

Based on the analysis of extreme precipitation in a transient CP climate ensemble over Germany, we compare the results to observation-based extreme values, derive climate change signals, and assess the uncertainties. Three main conclusions were drawn:

- The CP simulations show a good coverage of the intensity range observed in the station measurements in the study area. Compared to the observations and the observation-based data set KOSTRA, a better agreement with the CP simulations is found for RLs with short EDs. In general, the simulations overestimate the RL for EDs longer than 1 h.
- The CP simulations project climate change signals of extreme precipitation intensities of up to 6 or 8.5% increase per K GW depending on the ensemble member. Events with short duration and long RPs are expected to change the most.
- Analysis of the uncertainty in the climate change signal revealed a substantial residual standard deviation of the linear approximation. The model spread is significantly smaller for short ED. The apparent spatial uncertainty implies that estimates for long RP are only possible for pooled spatial information. Furthermore, the spatial variance is expected to increase with climate change.

The comparison of the CP ensemble with station data shows the agreement of the hourly intensity range in CPM and an added value in the fractional contribution of precipitation intensities compared to conventional, convection parameterized simulations (here 7 km). This more realistic representation of rain events, especially of short duration, is consistent with previous research (Ban et al., 2014; Ban et al., 2020; Chan et al., 2014; Fosser et al., 2015). However, also the tendency to overestimate heavy rainfall in the present CPM is known from previous analyses (Fosser et al., 2015; Kendon et al., 2012) and is supported by the data presented.

Spatial patterns of extreme precipitation associated with orography are represented in the CP ensemble for longer, daily durations. However, likely due to the relatively short time series length, shorter (hourly) extremes do not reflect spatial patterns without the application of regionalization techniques. Our confidence in RLs from CPM is higher for short (hourly to sub-daily) events than for daily to multi-day events due to better agreement with station data and KOSTRA. The mean absolute bias is usually—but not necessarily—smaller for short RPs. The

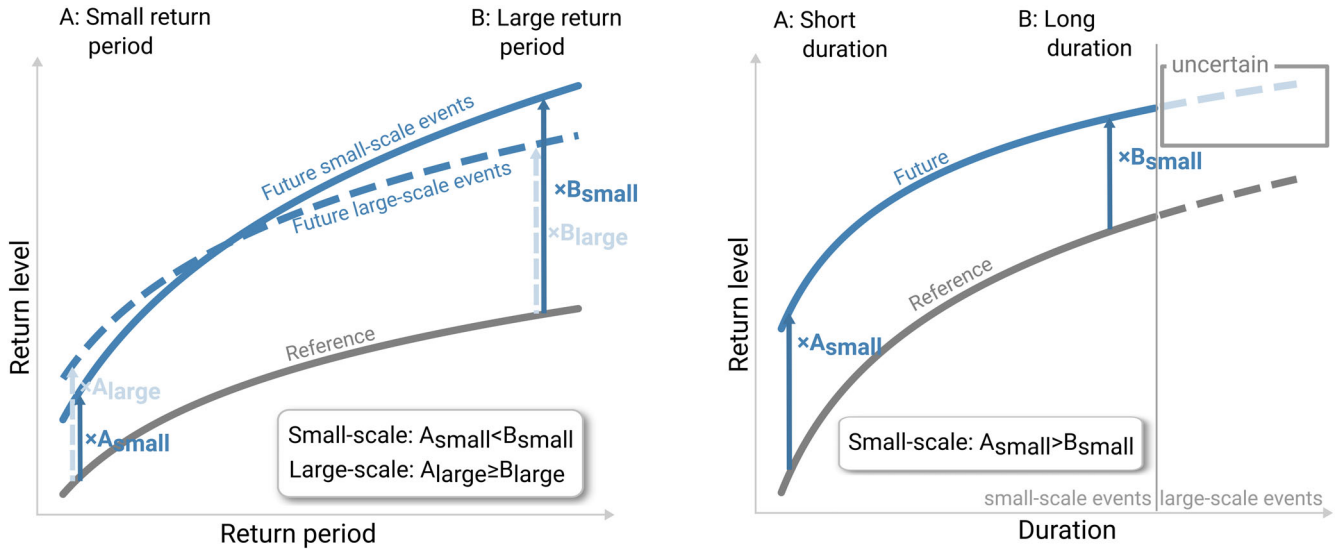

FIGURE 10 Schematic for the development of future intensity and duration curves. Changes are expressed in terms of relative CFs. The dependency of those CFs $A_{small/large}$ and $B_{small/large}$ on RP and duration is illustrated. The absolute values are not the subject of the diagram. CF, change factor; RP, return period. [Colour figure can be viewed at wileyonlinelibrary.com]

relative mean bias generally decreases with RP for sub-daily (except hourly) ED and increases for daily to multi-day events. In general, RLs are overestimated by CPM. An underestimation of the RL was only found for the shortest evaluated ED of 1 h and high RP. We have higher confidence in the three ensemble members driven by MPI-ESM-LR, EC-EARTH and HadGEM2-ES, due to a larger overestimation for the simulation driven by CNRM-CM5. We expect a mean bias in the order of ±10% for hourly ED, up to 23% for sub-daily events with ED > 1 h and up to 27% for daily and longer events in the CP ensemble, excluding CNRM-CM5-C. The findings complement previous comparisons of simulations with KOSTRA or observations by Berg et al. (2019), Poschlod et al. (2021), and Ban et al. (2020), who described increasing overestimation with ED by the simulations but for coarser resolution. As we assume a systematic bias in the simulations and shortcomings in the observations, we support the concept of climate CFs or uplifts (Chan et al., 2022) for a more robust assessment of future extreme precipitation.

There is agreement among the ensemble members that extreme precipitation increases with GW on all time scales considered. Figure 10 summarizes the main patterns in the ensemble regarding the relative climate change signal: For sub-daily events, the change signal increases with RP. In contrast, for large-scale events (ED = 72 h), the change signal is relatively constant or decreases with RP. The strongest change signals are associated with the shortest ED and tend to decrease with larger ED for sub-daily events. For daily to multi-day events, no consistent pattern could be derived. The highest change signals are therefore associated with the shortest ED (hourly) and large RP, where an increase of 6%–8.5% is projected, which is in the range of CC-scaling to slightly super CC-scaling. This strongest increase of extreme precipitation for short rainfall events is in agreement with previous research (Ban et al., 2015; Ban et al., 2020; Berg et al., 2019; Hodnebrog et al., 2019; Lenderink et al., 2021; Rajczak & Schär, 2017).

The finding of a larger increase for more rare events from hourly to daily duration is consistent with previous findings based on RCMs (e.g., Ban et al., 2015; Helsen et al., 2020; Li et al., 2019; Rajczak & Schär, 2017). For very long, multi-day events, we could not identify an increase in the change signal with RP. Ban et al. (2020) observed a similar behaviour of 5-day events in winter over the Alpine region. This behaviour of long events is likely influenced by shifts in the large-scale circulation patterns.

The sensitivity analysis showed that there is a substantial residual standard deviation for the linear fit of the intensity trajectories over GW, especially for long EDs. The ensemble spread is significantly smaller for small-scale, sub-daily events (2–3 PPT) than for large-scale, multi-day events (up to 6 PPT). This difference is attributed to the influence of large-scale circulation patterns on long events (Kautz et al., 2022) carried by the GCM, while small-scale events depend more on the thermodynamic representation in the CP model. The spatial distribution of the climate change signal was shown to have a standard deviation in the range of the magnitude of the change signal itself for large RP. This suggests that robust estimates can only be derived from the

distribution of results or pooled spatial information. Moreover, the width of this spatial distribution is found to increase with GW, indicating an increasing variability of extreme events.

Limitations of the study are that our conclusion is based on CMIP5 simulations and only four ensemble members and thus results potentially differ for other CMIP-based ensembles. Comparing daily extremes of annual RPs, Seneviratne and Hauser (2020) did not find different climate change signals in CMIP6 compared to CMIP5 in Central Europe. However, there are indications of an added value of CMIP6 simulations, including a better representation of blocking for some models (Schiemann et al., 2020). Furthermore, our analysis is limited to one RCM, which is suggested to largely determine the representation of short events. The uncertainty in the change signal for small-scale events may not be fully captured in the multi-GCM and single-RCM ensemble used. A matrix of CMIP6 GCMs and RCMs, such as is currently being targeted in the project NUKLEUS (Actionable local climate information for Germany), could be key to clarifying the robustness of trends in future ensembles.

The study also highlighted uncertainties in the methods used to estimate future extreme precipitation statistics. The variability in the results of individual grid points was discussed and is probably determined by short time series and therefore limited occurrence of rare events. Future investigations making use of the increasing number of CP simulations available will show whether the uncertainty can be reduced with additional model data. In addition, spatial pooling approaches could reduce uncertainty. For example, Chan et al. (2022) were able to resolve spatial patterns of short events using a regression model.

Overall, the study emphasizes that new products for rainfall risk management are needed today for sustainable (future/long-term) planning and management of climate adaptation and gives an indication of which thresholds are particularly sensitive to climate change, namely short-duration and long RPs. Moreover, we found an increased variability of future extreme precipitation. This is expected to pose a challenge for the application as the incorporation of uncertainty, also in impact modelling, may become more important. To avoid underestimating the risk, more conservative approaches that consider higher percentiles than the median may be reasonable.

In addition, the study emphasizes the considerable standard error associated with evaluating individual time slices as it is common practice in CP downscaling. By utilizing a transient projection and considering inherent variance, we obtained a more robust CF. This highlights the benefits of conducting long transient ensemble simulations and underscores the necessity for such simulations.

AUTHOR CONTRIBUTIONS

Marie Hundhausen: Writing – original draft; visualization; formal analysis; methodology; conceptualization. **Hendrik Feldmann:** Conceptualization; writing – review and editing; methodology; funding acquisition. **Regina Kohlhepp:** Writing – review and editing; methodology. **Joaquim G. Pinto:** Conceptualization; writing – review and editing; funding acquisition.

ACKNOWLEDGEMENTS

We thank Hans-Jürgen Panitz (KIT-IMK-TRO) for his great efforts in the RCM simulations. The simulations were performed on the national supercomputers Cray XC40 Hazel Hen and HPE Apollo Hawk at the High Performance Computing Centre Stuttgart (grant number HRCM ID-12801). Parts of the processing were performed at DKRZ. We thank DWD for providing observational data. Finally, we thank Elizabeth Kendon (UK Met Office), Steven Chan and Hayley Fowler (Newcastle University) for valuable discussions. This study is funded by the German Federal Ministry of Education and Research (BMBF) research programs, RegIKlim: projects ISAP and NUKLEUS (Grant numbers: 01LR2007B and 01LR2002B) and ClimXtreme: A1 SEVERE (Grant number: 01LP1901A), and AXA Research Fund. Open Access funding enabled and organized by Projekt.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Station data and the KOSTRA data set are available from Deutscher Wetterdienst (https://opendata.dwd.de/climate_environment/CDC/, last accessed November 2023). Simulation data from the KIT-KLIWA ensemble are available on request from the authors.

ORCID

Marie Hundhausen https://orcid.org/0000-0001-5400-3088
Hendrik Feldmann https://orcid.org/0000-0001-6987-7351
Joaquim G. Pinto https://orcid.org/0000-0002-8865-1769

REFERENCES

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and

sensitivities. *Monthly Weather Review*, 139, 3887–3905. Available from: https://journals.ametsoc.org/view/journals/mwre/139/12/mwr-d-10-05013.1.xml.

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M. et al. (2021) The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Climate Dynamics*, 57, 275–302.

Ban, N., Rajczak, J., Schmidli, J. & Schär, C. (2020) Analysis of alpine precipitation extremes using generalized extreme value theory in convection-resolving climate simulations. *Climate Dynamics*, 55, 61–75.

Ban, N., Schmidli, J. & Schär, C. (2014) Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, 119, 7889–7907.

Ban, N., Schmidli, J. & Schär, C. (2015) Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, 42, 1165–1172.

Berg, P., Christensen, O.B., Klehmet, K., Lenderink, G., Olsson, J., Teichmann, C. et al. (2019) Summertime precipitation extremes in a EUR-OCORDEX 0.11° ensemble at an hourly resolution. *Natural Hazards and Earth System Sciences*, 19, 957–971.

Berg, P. & Haerter, J. (2013) Unexpected increase in precipitation intensity with temperature—A result of mixing of precipitation types? *Atmospheric Research*, 119, 56–61.

Chan, S., Kendon, E., Fowler, H.J., Youngman, B.D., Dale, M. & Short, C. (2022) *New extreme rainfall projections for improved climate resilience of urban drainage systems*. Available at SSRN 4303999.

Chan, S.C., Kendon, E.J., Fowler, H.J., Blenkinsop, S., Roberts, N.M. & Ferro, C.A. (2014) The value of high-resolution met office regional climate models in the simulation of multihourly precipitation extremes. *Journal of Climate*, 27, 6155–6174.

Chan, S.C., Kendon, E.J., Roberts, N.M., Fowler, H.J. & Blenkinsop, S. (2016) Downturn in scaling of UK extreme rainfall with temperature for future hottest days. *Nature Geoscience*, 9, 24–28.

DWA (2012) *Starkregen in Abhängigkeit von Wiederkehrzeit und Dauer, DWA-Regelwerk 531*. Dt. Vereinigung für Wasserwirtschaft, Abwasser u. Abfall e.V, Hennef.

DWD (2023) *Hourly station observations of precipitation for Germany*. Available from: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/precipitation/historical/.

Eggert, B., Berg, P., Haerter, J., Jacob, D. & Moseley, C. (2015) Temporal and spatial scaling impacts on extreme precipitation. *Atmospheric Chemistry and Physics*, 15, 5957–5971.

Ettrichrätz, V., Beier, C., Keuler, K. and Trachte, K. (2023) Identification of regions with a robust increase of heavy precipitation events. *EGUsphere*, 2023, 1–34. Available from: https://egusphere.copernicus.org/preprints/2023/egusphere-2023-552/.

Fosser, G., Khodayar, S. & Berg, P. (2015) Benefit of convection permitting climate model simulations in the representation of convective precipitation. *Climate Dynamics*, 44, 45–60.

Fowler, H.J., Lenderink, G., Prein, A.F., Westra, S., Allan, R.P., Ban, N. et al. (2021) Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth & Environment*, 2, 107–122.

Guerreiro, S.B., Fowler, H.J., Barbero, R., Westra, S., Lenderink, G., Blenkinsop, S. et al. (2018) Detection of continental-scale intensification of hourly rainfall extremes. *Nature Climate Change*, 8, 803–807.

Helsen, S., van Lipzig, N.P., Demuzere, M., Vanden Broucke, S., Caluwaerts, S., De Cruz, L. et al. (2020) Consistent scale-dependency of future increases in hourly extreme precipitation in two convection-permitting climate models. *Climate Dynamics*, 54, 1267–1280.

Henn, B., Newman, A.J., Livneh, B., Daly, C. & Lundquist, J.D. (2018) An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, 556, 1205–1219.

Ho, C.K., Stephenson, D.B., Collins, M., Ferro, C.A. & Brown, S.J. (2012) Calibration strategies: a source of additional uncertainty in climate change projections. *Bulletin of the American Meteorological Society*, 93, 21–26.

Hodnebrog, Ø., Marelle, L., Alterskjær, K., Wood, R.R., Ludwig, R., Fischer, E.M. et al. (2019) Intensification of summer precipitation with shorter time-scales in Europe. *Environmental Research Letters*, 14, 124050.

Hohenegger, C., Brockhaus, P. & Schar, C. (2008) Towards climate simulations at cloud-resolving scales. *Meteorologische Zeitschrift*, 17, 383–394.

Hundhausen, M., Feldmann, H., Laube, N. & Pinto, J.G. (2023) Future heat extremes and impacts in a convection-permitting climate ensemble over Germany. *Natural Hazards and Earth System Sciences*, 23, 2873–2893.

IPCC. (2021) *Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge and New York, NY: Cambridge University Press.

Junghänel, T., Ertel, H. & Deutschländer, T. (2017) *Bericht zur revision der koordinierten starkregenregionalisierung und -auswertung des deutschen wetterdienstes in der version 2010*. Offenbach am Main: German Meteorological Service, Department of Hydrometeorology.

Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A. et al. (2013) Monitoring of climate change in Germany—data, products and services of Germany's national climate data centre. *Advances in Science and Research*, 10, 99–106.

Kautz, L.-A., Martius, O., Pfahl, S., Pinto, J.G., Ramos, A.M., Sousa, P.M. et al. (2022) Atmospheric blocking and weather extremes over the euro-atlantic sector—a review. *Weather and Climate Dynamics*, 3, 305–336.

Kendon, E.J., Fischer, E.M. & Short, C.J. (2023) Variability conceals emerging trend in 100yr projections of UK local hourly rainfall extremes. *Nature Communications*, 14, 1133.

Kendon, E.J., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C. & Senior, C.A. (2014) Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4, 570–576.

Kendon, E.J., Roberts, N.M., Senior, C.A. & Roberts, M.J. (2012) Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, 25, 5791–5806.

Kreienkamp, F., Philip, S.Y., Tradowsky, J.S., Kew, S.F., Lorenz, P., Arrighi, J. et al. (2021) *Rapid attribution of heavy rainfall events leading to the severe flooding in western Europe during July 2021*. World Weather Atribution.

Lehmann, J., Coumou, D. & Frieler, K. (2015) Increased record-breaking precipitation events under global warming. *Climatic Change*, 132, 501–515.

Lenderink, G., Belušić, D., Fowler, H.J., Kjellström, E., Lind, P., van Meijgaard, E. et al. (2019) Systematic increases in the thermodynamic response of hourly precipitation extremes in an idealized warming experiment with a convection-permitting climate model. *Environmental Research Letters*, 14, 074012.

Lenderink, G., de Vries, H., Fowler, H.J., Barbero, R., van Ulft, B. & van Meijgaard, E. (2021) Scaling and responses of extreme hourly precipitation in three climate experiments with a convection-permitting model. *Philosophical Transactions of the Royal Society A*, 379, 20190544.

Lenderink, G. & Van Meijgaard, E. (2008) Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, 1, 511–514.

Lengfeld, K., Winterrath, T., Junghänel, T., Hafer, M. & Becker, A. (2019) Characteristic spatial extent of hourly and daily precipitation events in Germany derived from 16 years of radar data. *Meteorologische Zeitschrift*, 28, 363–378.

Li, C., Zwiers, F., Zhang, X., Chen, G., Lu, J., Li, G. et al. (2019) Larger increases in more extreme local precipitation events as climate warms. *Geophysical Research Letters*, 46, 6885–6891.

Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A. et al. (2021) Convection-permitting modeling with regional climate models: latest developments and next steps. *Wiley Interdisciplinary Reviews: Climate Change*, 12, e731.

Ludwig, P., Ehmele, F., Franca, M.J., Mohr, S., Caldas-Alvarez, A., Daniell, J.E. et al. (2023) A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe—part 2: historical context and relation to climate change. *Natural Hazards and Earth System Sciences*, 23, 1287–1311.

Malitz, G. & Ertel, H. (2015) Kostra dwd 2010: Starkniederschlagshöhen für Deutschland (Bezugszeitraum 1951 bis 2010), Abschlussbericht, Deutscher Wetterdienst.

Michaelis, A.C., Willison, J., Lackmann, G.M. & Robinson, W.A. (2017) Changes in winter North Atlantic extratropical cyclones in high-resolution regional pseudo–global warming simulations. *Journal of Climate*, 30, 6905–6925.

Mishra, V., Wallace, J.M. & Lettenmaier, D.P. (2012) Relationship between hourly extreme precipitation and local air temperature in the United States. *Geophysical Research Letters*, 39, L16403.

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J.E. et al. (2023) A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe—part 1: event description and analysis. *Natural Hazards and Earth System Sciences*, 23, 525–551.

Nijsse, F.J., Cox, P.M. & Williamson, M.S. (2020) Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, 11, 737–750.

Pendergrass, A.G. (2018) What precipitation is extreme? *Science*, 360, 1072–1073.

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

Poschlod, B., Ludwig, R. & Sillmann, J. (2021) Ten-year return levels of sub-daily extreme precipitation over Europe. *Earth System Science Data*, 13, 983–1003.

Prein, A., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N., Keuler, K. et al. (2013) Added value of convection permitting seasonal simulations. *Climate Dynamics*, 41, 2655–2677.

Rajczak, J. & Schär, C. (2017) Projections of future precipitation extremes over Europe: a multimodel assessment of climate simulations. *Journal of Geophysical Research: Atmospheres*, 122, 10–773.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A. & Gratzki, A. (2013) A central European precipitation climatology—part I: generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, 22, 235–256.

Rockel, B., Will, A. & Hense, A. (2008) The regional climate model COSMO-CLM (CCLM). *Meteorologische Zeitschrift*, 17, 347–348.

Santos, J.A., Belo-Pereira, M., Fraga, H. & Pinto, J.G. (2016) Understanding climate change projections for precipitation over western Europe with a weather typing approach. *Journal of Geophysical Research: Atmospheres*, 121, 1170–1189.

Sato, T., Miura, H., Satoh, M., Takayabu, Y.N. & Wang, Y. (2009) Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model. *Journal of Climate*, 22, 4809–4826.

Schiemann, R., Athanasiadis, P., Barriopedro, D., Doblas-Reyes, F., Lohmann, K., Roberts, M.J. et al. (2020) Northern Hemisphere blocking simulation in current climate models: evaluating progress from the climate model intercomparison project phase 5 to 6 and sensitivity to resolution. *Weather and Climate Dynamics*, 1, 277–292.

Seneviratne, S.I. & Hauser, M. (2020) Regional climate sensitivity of climate extremes in CMIP6 versus CMIP5 multimodel ensembles. *Earth's Future*, 8, e2019EF001474.

Sieck, L.C., Burges, S.J. & Steiner, M. (2007) Challenges in obtaining reliable measurements of point rainfall. *Water Resources Research*, 43, W01420.

Sørland, S.L., Brogli, R., Pothapakula, P.K., Russo, E., Van de Walle, J., Ahrens, B. et al. (2021) COSMO-CLM regional climate simulations in the coordinated regional climate downscaling experiment (CORDEX) framework: a review. *Geoscientific Model Development*, 14, 5125–5154.

Taylor, K. E., Balaji, V., Hankin, S., Juckes, M., Lawrence, B. & Pascoe, S. (2011) *CMIP5 data reference syntax (DRS) and controlled vocabularies*. Program for Climate Model Diagnosis and Intercomparison. Available from: http://pcmdi.github.io/mips/cmip5/docs/cmip5_data_reference_syntax.pdf [Accessed 1st November 2023].

Teichmann, C., Bülow, K., Otto, J., Pfeifer, S., Rechid, D., Sieck, K. et al. (2018) Avoiding extremes: benefits of staying below +1.5 °C compared to +2.0 °C and +3.0 °C global warming. *Atmosphere*, 9, 115.

Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117, 1779–1800.

UK Environment Agency (2012) *Guidance flood risk assessments: climate change allowances*. Available from: www.gov.uk/guidance/flood-risk-assessments-climate-change-allowances.

Westra, S., Alexander, L.V. & Zwiers, F.W. (2013) Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, 26, 3904–3918.

Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F. et al. (2014) Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, 52, 522–555.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hundhausen, M., Feldmann, H., Kohlhepp, R., & Pinto, J. G. (2024). Climate change signals of extreme precipitation return levels for Germany in a transient convection-permitting simulation ensemble. *International Journal of Climatology*, 1–18. https://doi.org/10.1002/joc.8393