



Seasonal trends and spatial variability of precipitation extremes in Berlin-Brandenburg, Germany, from 1981 to 2025

Frederik Bart¹, Marco Otto¹, Daniel Fenner¹, Henning W. Rust², Fred Meier¹, Achim Holtmann¹, Benjamin Schmidt¹, and Dieter Scherer¹

¹Fachgebiet Klimatologie, Technische Universität Berlin, Rothenburgstr. 12, 12165 Berlin, Germany

²Arbeitsgruppe Statistische Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10 12165 Berlin, Germany

Correspondence: Frederik Bart (frederik.bart@tu-berlin.de)

Abstract. Precipitation extremes have caused substantial damage in the Berlin-Brandenburg region in recent decades. This study assesses seasonal trends and spatial variability in daily precipitation return levels for 2-, 5-, and 10-year return periods between 1981 and 2025. We use a peaks-over-threshold approach and a non-stationary Generalised Pareto Distribution with a time-dependent scale parameter. The analysis is based on observations from 223 rain-gauge stations operated by the German Meteorological Service (DWD) and the 2 km Central Europe Refined analysis version 2 (CER), a dynamically downscaled product based on the ERA5 atmospheric reanalysis. CER-based return levels show contrasting seasonal patterns: spring return levels decline across nearly the entire domain by up to approximately 20 %, while summer return levels show the largest absolute increases, averaging spatially approximately 5-10 mm d⁻¹ (+14-15 %), depending on the return period. Summer increases are also spatially more heterogeneous than in other seasons. Autumn return levels increase by up to 19 % across the analysed return periods, whereas winter changes remain weak. The DWD and CER datasets show consistent directions of change, although the CER underestimates the magnitude of summer increases relative to DWD. In some locations, the 2025 2-year summer return level approaches the magnitude of the 5-year return level estimated for 1981, highlighting the relevance of non-stationary return levels for regional risk assessment and future infrastructure planning.

1 Introduction

During recent decades, changes in the frequency and intensity of precipitation extremes have been observed across Central Europe. In Germany, previous studies have reported intensification of heavy precipitation, particularly for summer and winter extremes, alongside pronounced seasonal variability (Passow and Donner, 2019; Deumlich and Gericke, 2020; Gericke et al., 2019; Zeder and Fischer, 2020; Peter et al., 2024). In northeastern Germany, these changes occur alongside decreasing spring precipitation and associated soil-moisture deficits (Ionita et al., 2020). Together, more severe rainfall events and longer dry periods create growing challenges for infrastructure, agriculture, and urban flood management in the Berlin-Brandenburg region, located in northeastern Germany. The relevance of these hazards was demonstrated by severe rainfall events in Berlin in 2017 and 2019, which caused widespread urban pluvial flooding and substantial damage (Berghäuser et al., 2021). Statistical analyses suggest that the observed intensities exceeded the corresponding estimated 100-year return levels. Return levels provide



a useful way to describe the intensity of rare precipitation events and are widely used in infrastructure planning, engineering standards, and risk assessment (International Organization for Standardization, 2019, 2021; Nguyen et al., 2024). Within Extreme Value Theory (EVT), return levels can be estimated from either block maxima or threshold exceedances (Coles, 2001). However, precipitation extremes vary substantially between seasons and may change over time in response to climatic and environmental conditions. Assuming a stationary extreme-value distribution may therefore lead to biased estimates of current or future extremes (Gao et al., 2017). For the Berlin-Brandenburg region, projected changes in the water balance, including a potential decrease in summer precipitation of up to 30 % by the end of the twenty-first century, further motivate the consideration of time-dependent extreme-value characteristics (Cubasch and Kadow, 2019).

Non-stationary EVT models can represent temporal changes by allowing distributional parameters to depend on covariates such as time, temperature, or large-scale circulation patterns. Previous studies have already demonstrated the importance of such approaches for precipitation extremes in the Berlin-Brandenburg region. Using a Generalised Extreme Value model with seasonal and spatial covariates, Fischer et al. (2018) identified a pronounced seasonal cycle in extreme quantiles, including the 100-year return levels ($q_{0.99}$). Values were lower in winter and higher in summer, when daily return levels exceeded 120 mm. Other investigations, e.g. Fauer and Rust (2022), included large-scale covariates in their analysis such as the North Atlantic Oscillation (NAO) index and a blocking index, as well as surface temperature and time. Their findings suggest a general increase in precipitation extremes since 1950 and a positive association with the NAO during winter. Atmospheric blocking situations were found to have an enhancing effect on convective events during summer, while a suppression during winter was observed (Fauer and Rust, 2022). It is important to note that these previous investigations focused exclusively on rain-gauge measurements.

While evaluations based on station data typically allow for long time series, their spatial coverage is often limited and strongly depends on the density of the station network. Methods such as spatial interpolation of quantiles can allow for return-level estimations at locations without observations. However, these methods are highly dependent on the availability and density of station networks or the incorporation of auxiliary data. Small-scale precipitation events are often not captured by standard meteorological measurement networks (Terink et al., 2018; Rombeek et al., 2025). Even the relatively dense station network operated in Germany was shown to be often not dense enough to reliably capture intensive small-scale hourly rainfall events (Lengfeld et al., 2019). Accordingly, assessing the spatial variability of return levels based on station data alone can be subject to considerable uncertainties.

The inclusion of spatially high-resolution gridded data could allow for a more detailed analysis of long-term changes in precipitation extremes in a region. As such, weather radar data can provide high-resolution data, yet their temporal coverage is typically limited to time spans not longer than 20-30 years, and their quantitative accuracy depends on the availability of a sufficiently dense rain-gauge network for calibration (Saltikoff et al., 2019). In contrast, gridded output from numerical atmospheric models enables analyses over longer time spans and at variable spatial resolutions, supported by advances in computing power that have made convective-scale regional simulations feasible (Coppola et al., 2020). Reanalysis-based data products were found to be suitable to evaluate annual return levels in Bavaria, Germany (Poschlod, 2021). For the Berlin-Brandenburg region, the Central Europe Refined analysis version 2 (CER) is available as a gridded dataset covering the period 1981–2025



(Bart et al., 2025). This dataset stems from daily re-initialised, dynamically downscaled, and convection-permitting simulations of ERA5 reanalysis data (Hersbach et al., 2020). It provides up to hourly precipitation data (and other variables) with a maximum spatial resolution of 2 km for Berlin-Brandenburg and neighbouring regions. The CER has been evaluated for the time period 1990 to 2021 for monthly to annual precipitation rates, showing suitable accuracy for assessing spatial precipitation patterns and temporal changes. Most importantly, the dataset shows substantial improvements in terms of accuracy and reduction of mean deviations compared to the ERA5 forcing data (Bart et al., 2025).

In this study, we investigate temporal changes in seasonal precipitation extremes in the Berlin–Brandenburg region over the period 1981–2025, focusing on seasonal return levels and their spatial variability. Previous studies suggest that the expected changes should differ between seasons. Increasing atmospheric moisture and the thermodynamic scaling of precipitation may favour more intensive convective events during summer (Berg et al., 2019; Zeder and Fischer, 2020). At the same time, the strong spatial localisation of convective precipitation may produce heterogeneous trends, even across a relatively flat region such as Berlin-Brandenburg. Differences in storm tracks, atmospheric circulation, land use, and urban effects may all contribute to this spatial variability. Existing analyses based on radar observations and catalogues of heavy precipitation events have documented substantial spatial differences in the occurrence of heavy rainfall in and around Berlin-Brandenburg (Lengfeld et al., 2019; Lorenz et al., 2019). However, the spatial characteristics of seasonal return levels have not yet been quantified systematically across the region.

Here, we combine CER data with German Meteorological Service (DWD) rain-gauge observations to quantify seasonal return levels, their temporal evolution, and their spatial variability across Berlin-Brandenburg. The station data provide long-term local observations, whereas CER enables a spatially continuous analysis of the region. Comparing both datasets also allows us to assess the extent to which the identified temporal signals are consistent across observational and gridded data. The analysis addresses the following research questions:

R1 How have seasonal precipitation return levels changed across Berlin-Brandenburg between 1981 and 2025?

R2 How strongly do seasonal return levels and their temporal changes vary spatially across the region?

To address these questions, we use a Peaks-Over-Threshold (POT) framework with a Generalised Pareto Distribution (GPD) whose scale parameter varies with time. The model is first fitted and evaluated using DWD station observations and is then applied to CER to quantify spatial patterns in seasonal return-level changes. We estimate 2-, 5-, and 10-year return levels for the four meteorological seasons and compare the results between the two datasets. Sect. 2 describes the study area, datasets, statistical methods, and model evaluation. Sect. 3 presents the results, Sect. 4 discusses their implications and uncertainties, and Sect. 5 concludes the study.



2 Methods

2.1 Study design

90 The analysis workflow is illustrated in Fig. 1. After quality control and temporal aggregation, precipitation extremes are identified using a POT approach. Thresholds are selected using parameter-stability diagnostics, and dependent exceedances are reduced to independent cluster maxima through runs-based declustering. A GPD is then fitted separately for each season and location, with a scale parameter that varies linearly with time. Return levels and their temporal changes are estimated for the DWD stations and corresponding CER grid cells separately. Model fit and the statistical support for the time-dependent formulation are evaluated before the entire CER data are used to investigate spatial variability across the study region. Finally, the results from both datasets are compared.

2.2 Study area and period

The study area was defined to encompass the Berlin–Brandenburg region, which represents a climatically and socio-economically relevant transition zone between a densely built metropolitan area and its rural surroundings. The spatial extent of the innermost model domain of the CER dataset fully covers this region and ensures a consistent representation of precipitation extremes across the study area. This region encompasses the federal German states Berlin and Brandenburg (Fig. 2) as well as parts of the neighbouring states with the Berlin-Brandenburg metropolitan region located in the centre of the domain. Berlin’s administrative area covers approximately 900 km². The surroundings are sparsely populated and consist mainly of agricultural and forestry land. Located in the northern German lowlands, only minor plateaus and uplands are present in this region. The most noteworthy are the Barnim Plateau in the northern part of Berlin, the Teltow Plateau in the south, and the Nauen Plate in the southwest. The most southwestern and southern parts of the study region encompass regions of the Harz Mountains (lower Harz) and Lower Lusatia, respectively. The climatology of the region represents a transitional region between the more maritime climate in the west and north and more continental conditions in the east and south. Precipitation-wise the region is notably drier than the rest of Germany with a mean annual precipitation sum of 572 mm a⁻¹ (German Meteorological service station Berlin-Tempelhof, 1991-2020) and largest amounts of precipitation typically occur during summer (App. A1; Pagenkopf, 2011). Previous studies in this region indicate a high temporal and spatial variability of precipitation, which is characterised by long periods of drought and local heavy precipitation events (Berghäuser et al., 2021; Pohle et al., 2025). This study focuses on the period 1981-2025, covering 45 years that represent a period with accelerated changing climate conditions during the past century in the region (Cubasch and Kadow, 2019).

115 2.3 Data

2.3.1 Observational data

Daily quality-controlled measurement data from stations of the DWD were obtained for the study region for the years 1981-2025 (DWD, 2024b, a). The total number of stations with daily precipitation data in the study area is 970. To ensure sufficient

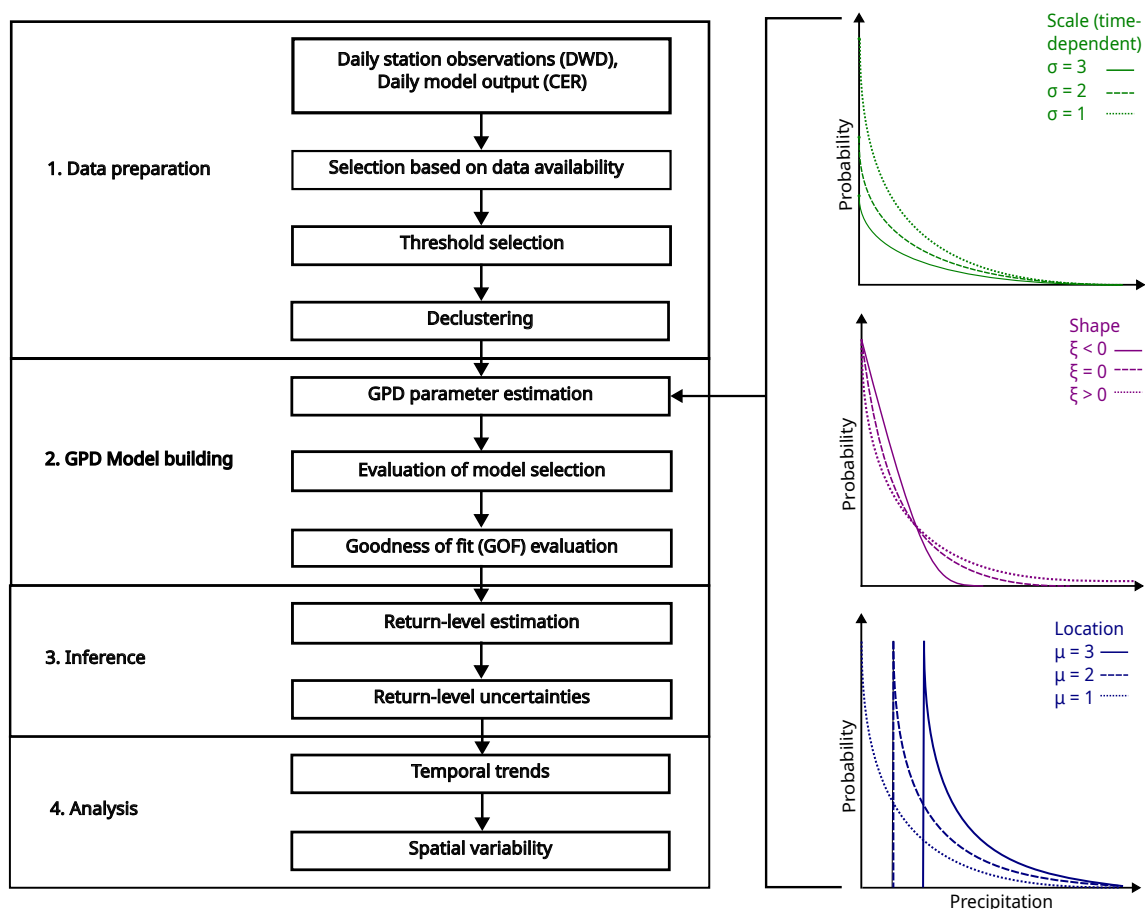


Figure 1. Schematic overview of the data analysis procedure. Curves in the right panel show exemplary values of the different parameters applied in the Generalised Pareto Distribution (GPD) to illustrate their influence on the shape of the distribution.

data availability, the time series were checked for large data gaps. Data completeness was assessed separately for each station, year, and season. A station-season-year record was excluded if fewer than 75 % of the expected daily values were available. Stations were then retained only if at least 75 % of the years within each season met this completeness criterion. After filtering, 223 stations remained for the analysis, which are relatively uniformly distributed in the study area (Fig. 2B). For the daily values provided by the DWD, it should be noted that the values refer to the period from 6:00 a.m. UTC of the current day to 6:00 a.m. UTC on the following day. The data were grouped into the meteorological seasons winter DJF (December-February), spring MAM (March-May), summer JJA (June-August), and autumn SON (September-November). December was assigned to the winter season of the following year.

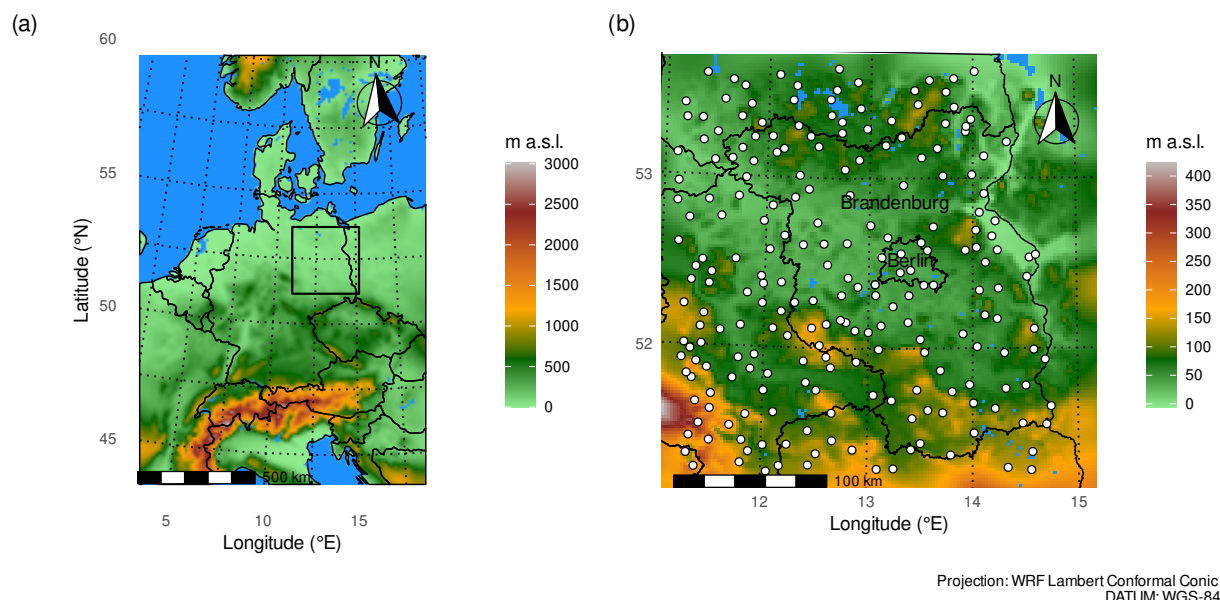


Figure 2. (a) Location of the study region Berlin-Brandenburg and Central Europe Refined analysis version 2 (CER) domain (black rectangle) with terrain height. (b) CER domain with terrain height and location of German Meteorological Service (DWD) rain gauges used in this study (white dots). Black lines indicate country (a) and federal state (b) borders. Note different colour bar scaling.

2.3.2 Reanalysis-based data

Gridded precipitation data of the Central Europe Refined analysis (CER; Bart et al., 2025) at 2 km spatial resolution was analysed alongside the station-based data. The CER represents a WRF-based (Skamarock et al., 2021) dynamical downscaling of ERA5 (Hersbach et al., 2020). The simulation strategy of the CER was adopted from Maussion et al. (2011, 2014) and employs a daily re-initialization approach. Each day is simulated separately with a 36-hour run, where the first 12 hours are discarded as spin-up and the remaining 24 hours are retained for analyses. The domain setup of the CER features two two-way nested domains. The outermost domain covers all of Germany with a horizontal grid spacing of 10 km. The inner domain with a grid spacing of 2 km covers Berlin and Brandenburg (Fig. 2). Further details of the dataset as well as the evaluation of monthly, seasonal and annual precipitation products are given in Bart et al. (2025). The dynamical downscaling approach of the CER reduces the wet bias found in the ERA5 forcing data for the study region significantly and provides several improvements of precipitation products over the previous version of the dataset (Jänicke et al., 2017). For a comparative analysis with the DWD station data, hourly precipitation data of the CER was extracted at the nearest grid cell to the DWD station coordinates. To minimize the impact of border effects, the outermost 10 km of CER data were removed. These hourly values were aggregated to daily precipitation sums as in the DWD station data, i.e., from 6:00 am UTC to 6:00 am UTC of the following day.



2.4 Extreme-value modelling of precipitation extremes

2.4.1 POT and GPD formulation

According to the Pickands-Balkema-de Haan theorem, the exceedances over a sufficiently large threshold u can be approximated by a GPD for a wide class of underlying probability distributions. Based on that assumption, it can be expected that the GPD is able to model the probability of extremes for the majority of precipitation regimes (Balkema and de Haan, 1974; Pickands, 1975). Extreme events are selected using the POT approach, in which all daily rainfall data x above a sufficiently large threshold u is extracted. After the temporal declustering described in subsect. 2.4.3 these are used to fit a distribution function. The selection procedure for these thresholds is described in sect. 2.4.2 and visualised in Fig. 3. The main advantage of this approach over alternatives such as the block-maxima method is the larger sample size, which has the potential to improve the statistical efficiency of parameter estimation. A detailed description of the GPD and POT is given in Coles (2001). The cumulative distribution function (CDF) of the GPD can be expressed either for the original observations above u , in which case the lower endpoint is represented by a location parameter $\mu = u$, or for the excesses $y = x - u$, in which case the location parameter is zero (x denotes the daily precipitation totals). We use the excess formulation below.

$$F_{(\sigma, \xi)}(x) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{for } \xi = 0. \end{cases} \quad (1)$$

The location parameter μ defines the shift or the lower bound of the support of the distribution, see Fig. 1, bottom right. The scale parameter σ describes the dispersion or scale of the GPD, see Fig. 1, middle right. The shape parameter ξ defines the tail behaviour ("heaviness" of the tail), see Fig. 1, top right. For $\xi = 0$ the distribution corresponds to the exponential distribution while for $\xi < 0$ the tail is bound by an upper endpoint. For $\xi > 0$ the distribution becomes heavy tailed. The GPD parameters were estimated using Maximum Likelihood Estimation (MLE). Since the main focus of this work is on the temporal variability of return level estimates, we employed a time-dependent σ parameter as described in Coles (2001). A constant shape parameter was used to avoid excessive instabilities in the estimation of the distribution.

$$\sigma(t) = \sigma_0 + \sigma_1 t \quad (2)$$

The time covariate was centred and scaled to the study period, and parameter estimates were checked to ensure that $\sigma(t)$ remained positive throughout 1981-2025. Using this method results in an intercept term of the scale parameter σ_0 which sets the baseline of σ and a slope (trend) coefficient σ_1 which describes how σ changes with per unit increase t (years). We employed functions from the R package `extRemes` (Gilleland and Katz, 2016) for modelling of the GPD on a seasonal scale, including the declustering, as well as to estimate MLE-parameters and return levels.



2.4.2 Threshold selection

The selection of an adequate threshold u is essential for the estimation of the GPD and the subsequent estimation of the return levels. On the one hand, selecting a too low u could lead to the inclusion of non-extreme events and introduce a bias to the distribution model. On the other hand, selecting a too high u may result in too few exceedances, leading to high variance in the estimated return levels. Standard selection methods for adequate u values rely on parameter stability, commonly assessed through quantile-quantile-plots (Q-Q plots) (Coles, 2001). These offer an easy-to-interpret standard tool for the identification of the highest u values at which no abrupt changes in the GPD parameters can be observed. The behaviours of σ_0 , σ_1 and ξ for a range of percentile-based u between the 85th and 98th percentile (1 % steps) are visualised in Fig. 3. u were estimated based on percentiles of daily precipitation for each station individually for each season. When considering the parameter stability of the entire station collective, the 90th percentile was assessed as the most suitable. Higher u show significant fluctuations in σ_1 , particularly in JJA. This behaviour is also observed for ξ for JJA and MAM extremes. The resulting u were highest during JJA and showed relatively similar values during the other seasons with a wider spread in DJF and SON.

2.4.3 Temporal declustering

An important aspect for estimating parameters of the GPD is the independence of extremes in the observational data. Contrary to the block-maxima approach, where maxima stem from predefined disjunct time intervals (e.g. years), the POT approach risks the sampling of two or more consecutive daily precipitation extremes. The latter are likely to be dependent, i.e., they stem from the same event, and challenge the assumption of independence of the GPD. A typical approach is to "decluster" these events and select only the one largest threshold exceedance from a cluster. We employed a "runs"-based declustering method following Coles (2001). First, we grouped consecutive threshold exceedances into clusters. Specifically, we considered consecutive daily threshold exceedances to be part of the same cluster, i.e. they were not separated by at least one non-exceedance day. Next, we reduced each cluster to a single data point by keeping only the highest daily maximum value and removing all other values in that cluster. The resulting cluster maxima were treated as approximately independent observations for the GPD fit. We performed this declustering procedure separately to the data at each station and CER grid point to maintain data independence for our analysis.

2.4.4 Return-level estimation

Return levels are expressed as quantiles $q_p(t)$, where $p = 1 - \frac{1}{N}$ denotes the non-exceedance probability (i.e. the probability that a threshold will not be exceeded this year) and N the corresponding return period in years. Our analysis focuses on precipitation extremes with return periods of $N = 2$, $N = 5$ and $N = 10$, which correspond to the quantiles of $q_{0.5}$, $q_{0.8}$, and $q_{0.9}$, respectively. We estimate that $q_p(t)$ for longer N would be associated with substantially increased uncertainty given the 45-year length of the available time series. These return levels are defined with respect to the relevant season. Due to the established naming convention of N -year return levels the quantiles will be referenced as such from this point onwards.

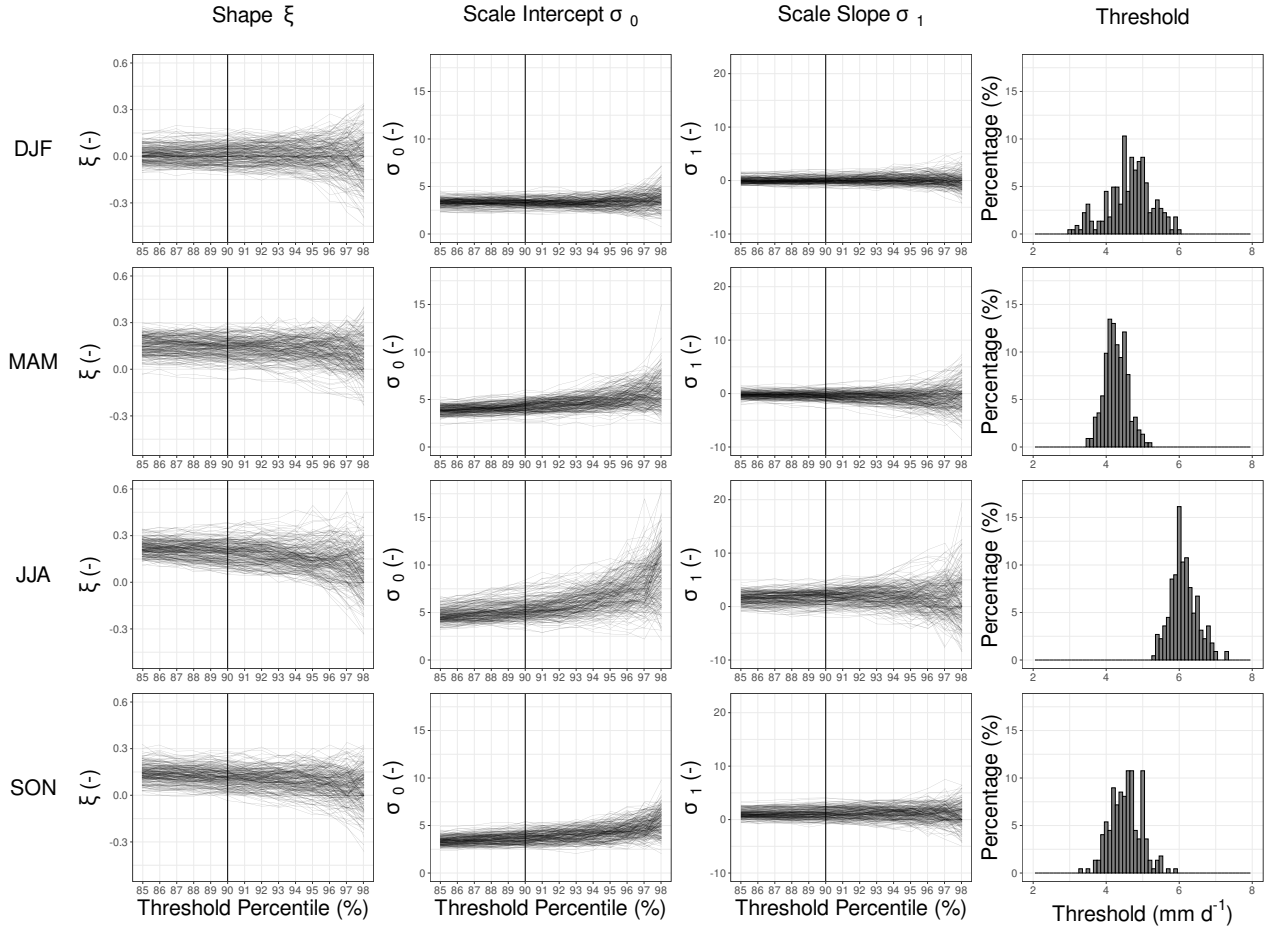


Figure 3. Estimated values of ξ (shape), σ_0 (scale intercept) and σ_1 (scale slope) for percentile-based thresholds u from the 85th to the 98th percentile for all stations and seasons (first three columns from the left). The selected 90th percentile u is indicated as a black vertical line. The distributions of resulting u values (90th percentile) across DWD stations are visualised as histogram plots for each season in the right column.

The return level $q_p(t)$ is defined as

$$q_p(t) = \begin{cases} u + \sigma(t) \ln(Nn\zeta_u), & \text{if } \xi = 0, \\ u + \frac{\sigma(t)}{\xi} [(Nn\zeta_u)^\xi - 1], & \text{if } \xi \neq 0, \end{cases} \quad (3)$$

where n is the number of daily observations per season and $\zeta_u = P(X > u)$ is the threshold exceedance rate. Since a time-dependent scale parameter was used, $q_p(t)$ estimates were recalculated for each year, resulting in 45 values per season and station. The change in these estimates is mainly governed by σ_1 (Fig. 4). To evaluate changes in $q_p(t)$ over the selected time



frame, we estimated the difference ($\Delta q_p(t)$) between the final $q_p(t = 2025)$ and initial time step $q_p(t = 1981)$ for each season.

205 Relative changes were calculated as the percentage:

$$\Delta q_p = 100 \cdot \frac{q_p(2025) - q_p(1981)}{q_p(1981)} \quad (4)$$

2.4.5 Uncertainty estimation

Uncertainties in the estimated $q_p(t)$ were evaluated using a parametric bootstrapping approach. This approach is less dependent on the sample size than asymptotic or profile likelihood methods. Parametric bootstrapping is generally considered as a reliable and robust method for the evaluation of model uncertainties, being able to handle even smaller sample sizes and higher degrees of model complexity well (Davison and Hinkley, 1997). Using the MLE-estimates for each station and season, random GPD-distributed data was generated for each of the 223 stations and four seasons, resulting in a total of 908 bootstrapping procedures. Employing the newly generated sample data, the MLE-parameters were re-estimated. This process was repeated for each station and season with 100 replications. Finally, confidence intervals for each MLE-parameter were estimated using $q_p(t)$ quantiles, with confidence intervals ranging from $q_{0.025}$ to $q_{0.975}$ of the bootstrapping samples yielding a coverage probability of $1 - \alpha = 0.95$ (Fig. 4).

2.5 Model evaluation

2.5.1 Goodness-of-fit

The inclusion of a time-varying scale parameter, as chosen in this work, increases the model complexity compared to a stationary GPD. However, it is unclear if and how much a higher degree of model complexity can influence the Goodness of Fit (GOF). The assessment of the GOF is essential when fitting a distribution model like the GPD. Generally, the GOF evaluates how well the selected distribution model matches the actual data. Fig. 5 shows the visual evaluation of the standardised empirical and theoretical quantiles for each station and season (Coles, 2001). Observations are transformed to standard exponential quantiles via the probability integral transform (PIT) and the resulting empirical quantiles are compared to the 1:1 line. Across both datasets and all seasons, we find close agreement in the lower to mid-range tail. In the upper tail, a systematic positive deviation is visible across all seasons and both datasets. Here, empirical quantiles consistently exceed modelled quantiles and indicate that the GPD underestimates the most extreme observations. This upper-tail underestimation varies in degree across seasons rather than being categorically absent in any of them, with DJF and MAM exhibiting the most pronounced deviations, while JJA and SON show a diffuse but equally systematic positive bias. Upper-tail quantiles also indicate a vertical banding of points near standardised modelled quantile values of approximately 5-6, reflecting a discrete clustering of extreme events across many stations at similar modelled quantile values. The upper tail is constrained similarly across stations, likely governed by the estimated shape parameter ξ , while the actual observed extremes vary more continuously. The systematic overestimation of observed extremes relative to GPD-modelled quantiles implies that return level estimates derived from the fitted model are likely conservative with a slight tendency to underestimate rare events.

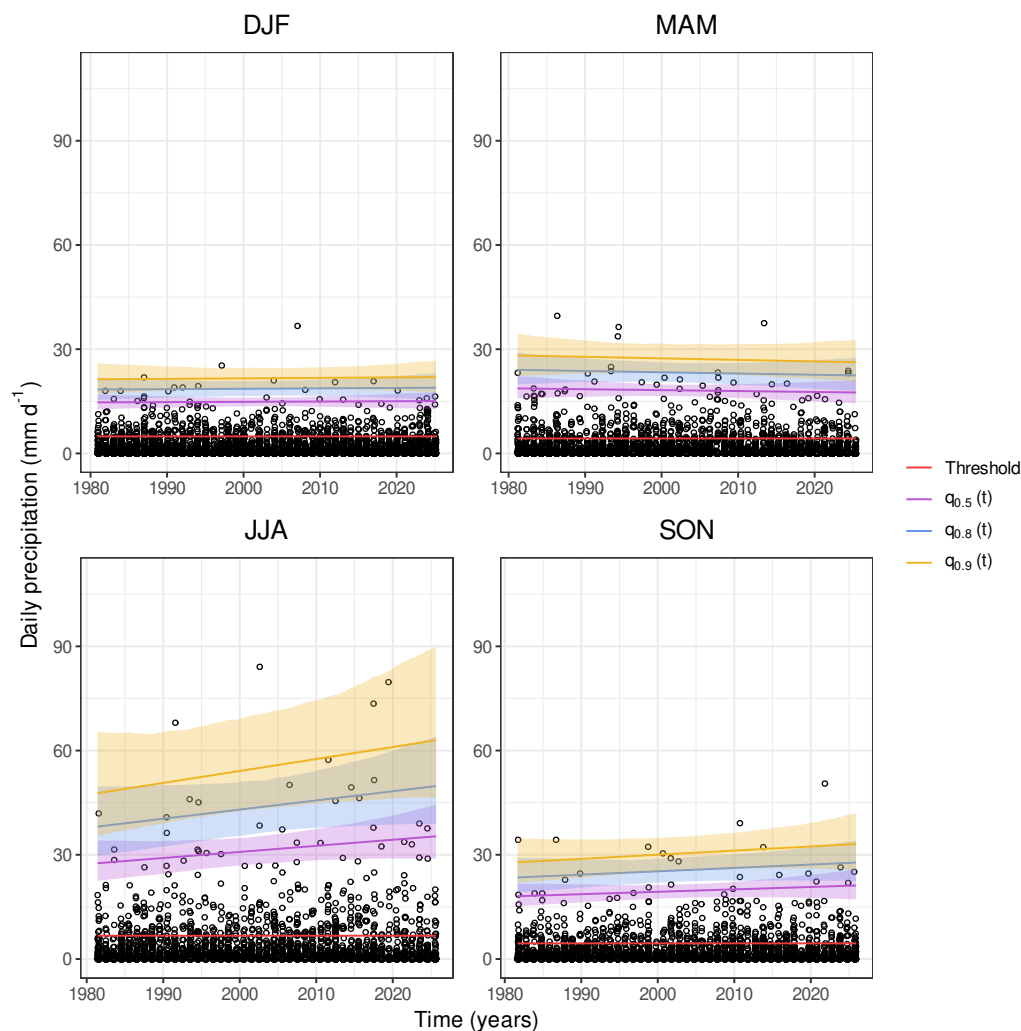


Figure 4. Example of the daily precipitation data (black points) at the German Meteorological Service station Potsdam per season with the estimated return levels $q_{0.5}$ (2-year return period), $q_{0.8}$ (5-year return period) and $q_{0.9}$ (10-year return period) based on the 90th percentile threshold value (red line) from 1981 to 2025. Shading indicates 95 % confidence intervals based on parametric bootstrapping.

235 2.5.2 Model Selection

We evaluated GOF using complementary likelihood-based criteria that differ in how strongly they penalise model complexity. Since the comparison here was between nested models only, we employed a Likelihood Ratio Test (LRT), which assesses whether the improvement in Negative Log Likelihood is statistically significant given the additional parameters (Wilks, 1938). Furthermore, the Akaike Information Criterion (AIC) was used for the comparison, which introduces a penalty on the number
240 of parameters and is commonly used to assess predictive performance (Akaike, 1974). The Bayesian Information Criterion

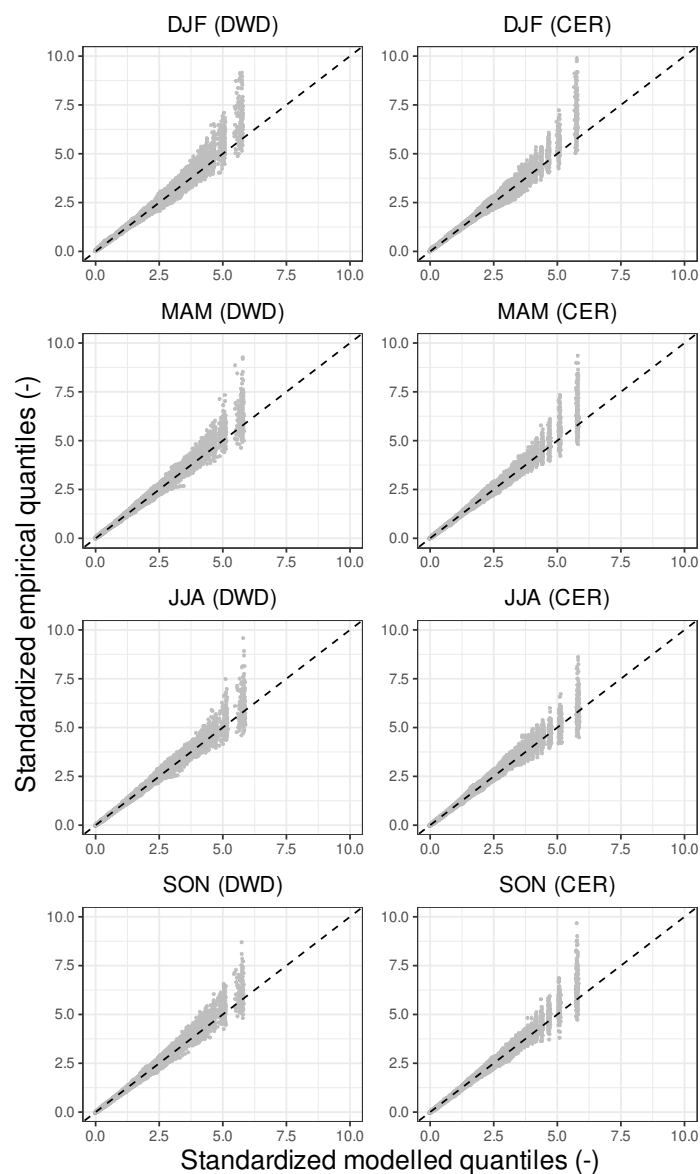


Figure 5. Quantile-quantile (Q-Q) plots for the GPD fit based on German Meteorological Service (DWD) data (left) and Central Europe Refined analysis version 2 (CER) data (right) for each season. Standardised empirical quantiles represent the observed/modelled data transformed by the fitted cumulative distribution function. The modelled quantiles or theoretical quantiles result from the reference standard exponential distribution. Both empirical and modelled quantiles have been standardised using a probability integral transformation of the residuals of the Generalised Pareto Distribution.

(BIC) applies an even stronger penalty that additionally depends on sample size, favouring more parsimonious models and potentially becoming conservative for large datasets (Schwarz, 1978). For AIC and BIC, lower values indicate better GOF; for

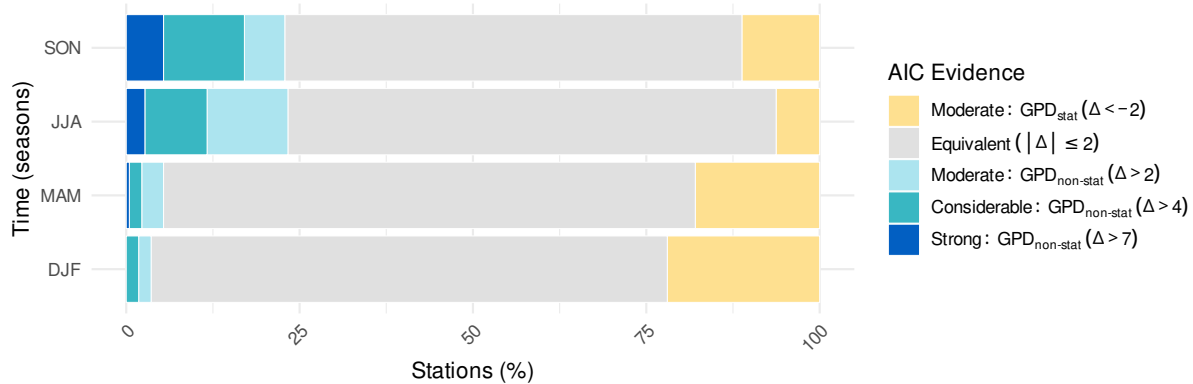


Figure 6. Model comparison between the stationary Generalised Pareto Distribution (GPD_{stat}) and the non-stationary model ($GPD_{non-stat}$) based on the Akaike Information Criterion (AIC) at each German Meteorological Service (DWD) station, aggregated by season. The x-axis shows the percentage of stations falling into each evidence category, defined by the AIC difference $\Delta = AIC_{stat} - AIC_{non-stat}$. Positive Δ indicates support for $GPD_{non-stat}$ (blue shading), negative Δ for GPD_{stat} (yellow shading). Evidence strength follows established thresholds described in Burnham and Anderson (2004).

the LRT, a p-value below $\alpha = 0.05$ indicates a statistically significant improvement of the more complex model. The model selection in this study focused on the comparison between a stationary model with a constant scale parameter ("less complex model") and the non-stationary model with a time-varying scale parameter as described in section 2.4.1 ("more complex model"). For each station and season, the AIC and BIC were compared directly. The overall increase or decrease in GOF for the non-stationary model was then evaluated based on the recommendations in Kass and Raftery (1995) for the BIC and as in Burnham and Anderson (2004) for the AIC.

Model selection between the stationary (GPD_{stat}) and non-stationary ($GPD_{non-stat}$) GPD was assessed using AIC (Fig. 6), BIC (App. A2), and the Likelihood Ratio Test (LRT) (App. A3) across all four seasons. In DJF and MAM, the majority of stations showed equivalent model performance ($|\Delta| \leq 2$) or moderate evidence favoring GPD_{stat} ($\Delta < -2$). In JJA and SON, support for $GPD_{non-stat}$ increased markedly, with combined moderate ($\Delta > 2$), considerable ($\Delta > 4$), and strong ($\Delta > 7$) evidence accounting for approximately 25 % and 20 % of stations, respectively. The BIC results were substantially more conservative, attributing moderate evidence in favor of GPD_{stat} to the large majority of stations in DJF and MAM. Even in JJA and SON, only a small fraction (~ 5 %) of stations reached moderate or considerable BIC support for $GPD_{non-stat}$, reflecting the stronger complexity penalty of BIC relative to AIC. In DJF and MAM, the overwhelming majority of stations showed no significant improvement from $GPD_{non-stat}$ ($p \geq 0.1$). In JJA, approximately 25 % of stations reached at least marginal significance ($p < 0.1$), with a visible portion achieving $p < 0.05$ or $p < 0.01$. SON exhibited the largest highly significant fraction, with roughly 12 % of stations at $p < 0.01$. All three criteria consistently indicate that support for $GPD_{non-stat}$ is weakest in DJF and MAM, and strongest in JJA and SON. However, $GPD_{non-stat}$ does not substantially underperform GPD_{stat} at any station, while statistically significant improvement is limited to a minority of stations.



3 Results

3.1 Temporal changes in return levels

The non-stationary GPD fitted to DWD station observations and corresponding CER data grid points yields time-varying return
265 levels $q_p(t)$ with distinct seasonal trends during 1981-2025. JJA return levels are substantially higher than those of all other
seasons and increase over time, while DJF return levels remain low and in most cases near-stationary (Fig. 7). MAM exhibits a
progressive decline in return levels, while SON shows a moderate upward trend. The decline in $q_p(t)$ during MAM is observed
across the entire study area and shows little spatial variation (e.g. compared with JJA).

In JJA, and to a lesser extent in SON, the upward trend in $q_{0.5}(t)$ over the study period is large enough to approach the
270 return-level magnitudes of $q_{0.8}(t)$ in 1981, indicating an effective increase in the frequency of such precipitation events over
the past 45 years. A similar, although weaker, relationship is found between the $q_{0.8}(t)$ and $q_{0.9}(t)$ estimates. This indicates that
some precipitation intensities associated with relatively rare events at the beginning of the study period correspond to shorter
effective return periods by 2025.

The changes in return levels Δq_p are directionally consistent between DWD and CER across all seasons and quantile levels,
275 though systematic magnitude differences are present (Table 1). The agreement is closest for DJF and SON results, when mean
deviation (MD) between the two datasets remains small. In MAM, CER results overestimate the negative trend compared to
DWD at all quantile levels, with the discrepancy growing larger toward lower exceedance probabilities. The opposite is observ-
able in JJA, when DWD q_p -estimates show consistently larger positive changes than CER, again with increasing divergence
from $q_{0.5}$ to $q_{0.9}$. The spread between stations/grid points is similar between datasets during all seasons except JJA, when a
280 higher variability in $q_p(t)$ estimates can be observed (Fig. 8). The decline in MAM return levels and increase in SON return
levels from 1981 to 2025 lead to a swap between the two seasons, i.e., while in 1981 return levels in MAM were generally
higher than in SON, the opposite is true in 2025. This trend in seasonal return levels is also observed for higher quantiles (or
events with higher return periods), with the rainfall amounts for the corresponding events being substantially higher (App. B1,
B2). Despite these magnitude differences, the sign of Δq_p remains consistent between the two datasets in all seasons. Thus,
285 the comparison indicates stronger agreement in the direction of the temporal changes than in their magnitude.

3.2 Spatial variations in return levels

Changes in return levels based on the CER between 1981 and 2025 exhibit pronounced differences in both direction and
spatial coherence across seasons (Fig. 8). DJF shows a weakly positive domain mean, but the near-equal share of positive and
negative grid cells indicates limited spatial coherence (Table 2, Fig. 9). MAM stands in sharp contrast, with 94.2% of grid cells
290 displaying a negative change and a comparatively low standard deviation, confirming a near-domain-wide decline in return
levels that is the most spatially consistent of all seasons. This decline intensifies toward higher return levels, with the mean
difference at $\Delta q_{0.9}$ approximately 1.7 times larger in absolute terms than at $\Delta q_{0.5}$.

Seasons JJA and SON both show predominantly positive changes, but differ markedly in spatial coherence. In JJA, the ma-
jority of grid cells exhibit positive changes, yet the standard deviation nearly doubles from $q_{0.5}$ to $q_{0.9}$, reflecting the spatially

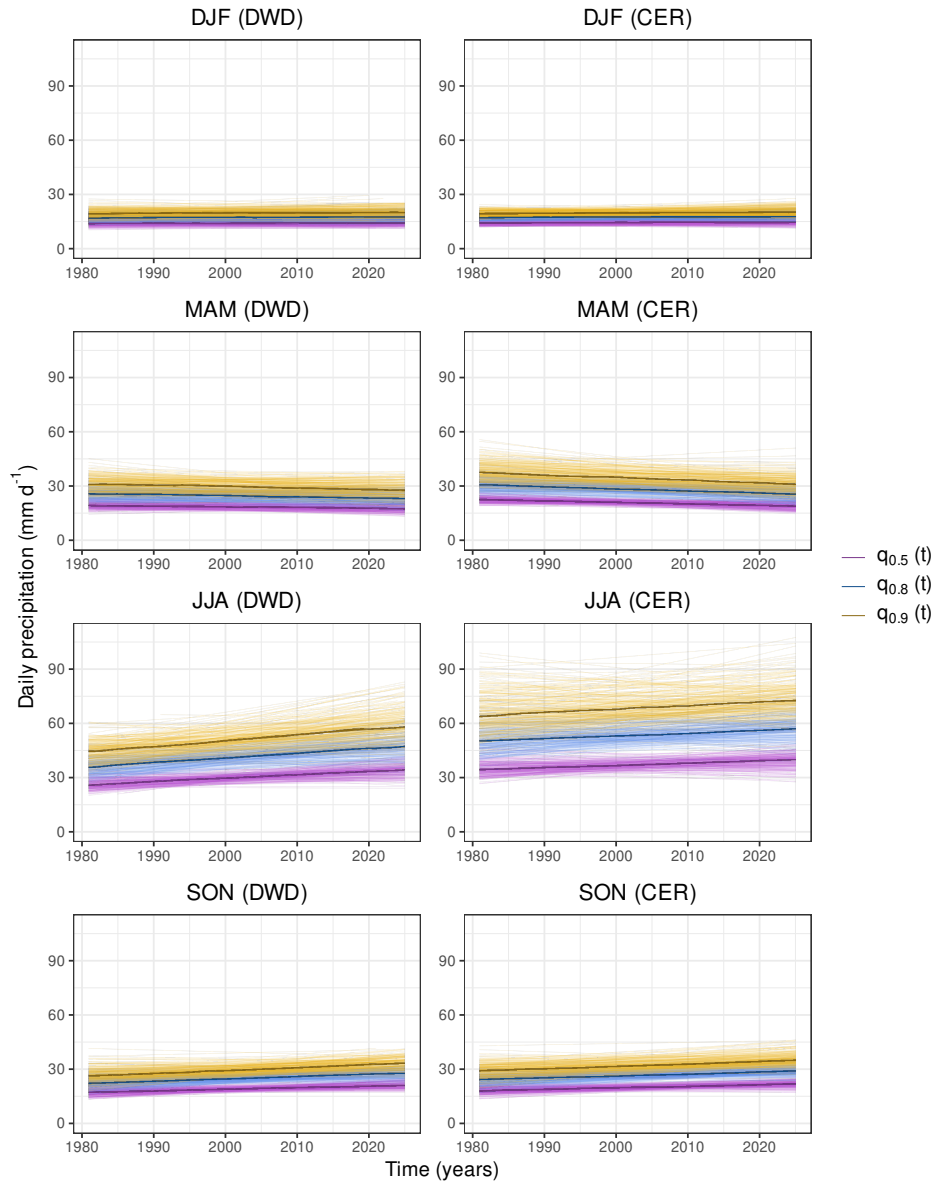


Figure 7. Comparison of the estimated time-dependent seasonal return levels $q_p(t)$ using a non-stationary Generalised Pareto Distribution based on 223 stations of the German Meteorological Service (DWD, left column) and corresponding grid points in the Central Europe Refined analysis version 2 (CER, right column) from 1981 to 2025 during the different seasons. Each line corresponds to one station or grid point, the dark line shows the median. Results are given for events with a non-exceedance probability of 50 % ($q_{0.5}(t)$), 80 % ($q_{0.8}(t)$) and 90 % ($q_{0.9}(t)$), i.e., 2-year, 5-year and 10-year return levels.



Table 1. Statistical metrics calculated for the difference in seasonal return levels $q_{0.5}$ (2-year return period), $q_{0.8}$ (5-year return period) and $q_{0.9}$ (10-year return period) in 1981 compared to 2025 ($\Delta q_p = q_p(t = 2025) - q_p(t = 1981)$), based on German Meteorological Service (DWD) station data and the Central Europe Refined analysis version 2 (CER) data at the corresponding locations for different return levels. Mean values and standard deviation are provided for both datasets. Mean Deviation (MD) and Mean Absolute Deviation (MAD) are calculated between Δq_p for CER and DWD.

Season	Δq_p	Mean		Std. Dev.		MD	MAD
		CER	DWD	CER	DWD		
		(mm d ⁻¹)	(mm d ⁻¹)	(mm d ⁻¹)	(mm d ⁻¹)	(mm d ⁻¹)	(mm d ⁻¹)
DJF	$\Delta q_{0.5}$	0.69	0.46	1.79	1.69	0.23	1.49
	$\Delta q_{0.8}$	0.89	0.60	2.34	2.24	0.30	1.97
	$\Delta q_{0.9}$	1.03	0.69	2.73	2.65	0.34	2.33
MAM	$\Delta q_{0.5}$	-3.73	-1.70	2.77	2.71	-2.03	3.28
	$\Delta q_{0.8}$	-5.40	-2.40	4.10	3.87	-3.00	4.76
	$\Delta q_{0.9}$	-6.86	-2.98	5.30	4.84	-3.87	6.06
JJA	$\Delta q_{0.5}$	4.34	7.66	6.34	5.08	-3.33	6.24
	$\Delta q_{0.8}$	6.51	11.31	9.66	7.61	-4.80	9.33
	$\Delta q_{0.9}$	8.49	14.55	12.77	9.94	-6.05	12.16
SON	$\Delta q_{0.5}$	3.60	3.85	2.63	3.07	-0.25	3.35
	$\Delta q_{0.8}$	5.08	5.36	3.71	4.25	-0.28	4.68
	$\Delta q_{0.9}$	6.30	6.58	4.61	5.20	-0.27	5.76

heterogeneous nature of convective precipitation trends (Fig. 9, App. B3, B4). The spatial character of MAM and JJA changes is fundamentally opposed: MAM exhibits a coherent, domain-wide decrease, while JJA is dominated by spatially heterogeneous and locally divergent trends. Particularly in the northern part of the study area, several regions show a substantially lower increase and, in some cases, even negative trends. SON combines a strong and spatially extensive positive signal with comparatively low variability, remaining consistent across the entire domain. However, increases are less pronounced and in some cases negative in the southern and southwestern parts of the study region. Also notable are areas along the northwestern edges where there is a decrease in return levels during JJA and SON, which could indicate border effects within the inner model domain of the CER. Across all seasons, the magnitude of Δq_p increases systematically from $q_{0.5}$ to $q_{0.9}$, with the most spatially coherent trend in MAM and SON, and the greatest spatial heterogeneity in JJA. Overall, the two datasets show con-

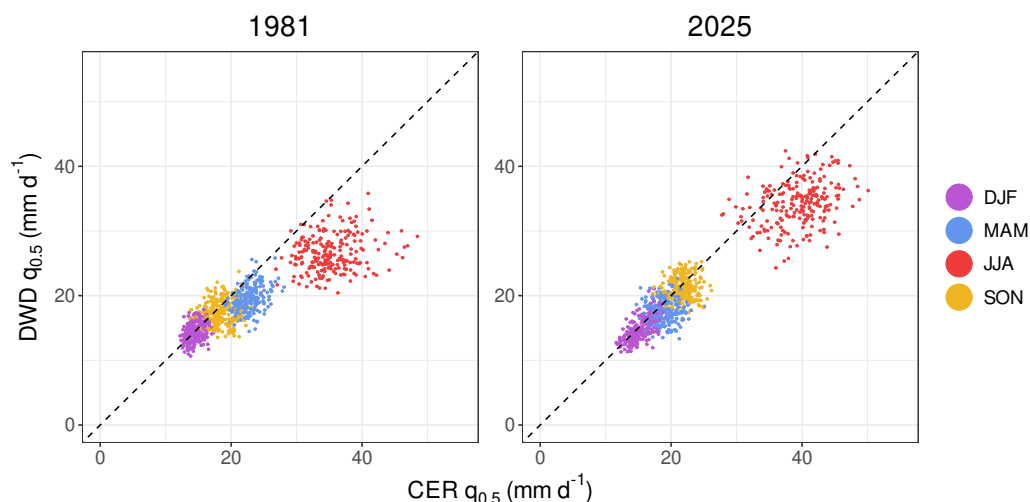


Figure 8. Comparison of time-dependent return levels ($q_{0.5}$, corresponding to a 2-year return period) derived from the German Meteorological Service (DWD) station observations and Central Europe Refined analysis version 2 (CER) data at the corresponding station locations, shown for the beginning ($t = 1981$, left) and end ($t = 2025$, right) of the investigation period. Each point represents one station-season combination, coloured by season. The dashed black line indicates the 1:1 reference line.

305 sistent seasonal directions of change: return levels decrease during MAM, increase most strongly during JJA, increase more coherently during SON, and change only weakly during DJF. The spatial analysis confirms that the spring decline is the most coherent regional signal, whereas summer changes are the most heterogeneous.

4 Discussion

4.1 Seasonal changes in precipitation extremes

Our analysis of trends in daily precipitation return levels reveals distinct temporal and spatial patterns over the Berlin-
310 Brandenburg region. The change in return levels during MAM shows the clearest and most spatially robust decrease in precipitation extremes. All return levels decrease almost uniformly across the entire domain, indicated by ~ 94 % of CER grid cells with negative sign. This decrease in precipitation is also observable in the domain-wide seasonal sums for the selected time period (App. A1). The average decrease of 18-20 % across all CER grid cells (depending on the return period) agrees well with independent evidence for increasing spring dryness and drought risk in Central Europe (Ionita et al., 2020; Haslinger and
315 Mayer, 2023). Our findings indicate that this decreasing return level trend extends to the upper tail of the daily precipitation distribution. Results for both DWD and CER data show that not only mean and moderate rainfall, but also the more severe spring events have become substantially less intense during the last 45 years.



Table 2. Spatial statistics of temporal changes in return level $\Delta q_p = q_p(t = 2025) - q_p(t = 1981)$ across the CER model domain, aggregated by season and non-exceedance probability p , where return levels $q_{0.5}$, $q_{0.8}$, and $q_{0.9}$ correspond to return periods of $N = 2, 5$, and 10 years, respectively. The domain-wide mean and median is given in absolute (mm d^{-1}) terms, the relative change expressed as a percentage change from 1981-2025 for the domain-wide mean. The standard deviation (mm d^{-1}) quantifies spatial variability across all grid cells. Furthermore, the percentage of grid cells with a positive (increasing return levels) or negative sign (decreasing return levels) of the temporal change is also included.

Season	Δq_p	Mean		Median	Std. Deviation	Positive Sign	Negative Sign
		(mm d^{-1})	(%)	(mm d^{-1})	(mm d^{-1})	(%)	(%)
DJF	$\Delta q_{0.5}$	0.5	3.1	0.4	1.7		
	$\Delta q_{0.8}$	0.6	3.3	0.5	2.2	60.1	39.9
	$\Delta q_{0.9}$	0.7	3.4	0.6	2.6		
MAM	$\Delta q_{0.5}$	-4.1	-17.8	-4.1	2.7		
	$\Delta q_{0.8}$	-6.0	-18.9	-5.9	3.9	5.8	94.2
	$\Delta q_{0.9}$	-7.5	-19.5	-7.4	5.0		
JJA	$\Delta q_{0.5}$	5.0	14.1	5.2	6.6		
	$\Delta q_{0.8}$	7.5	14.7	7.9	10.0	77.4	22.6
	$\Delta q_{0.9}$	9.7	14.9	10.3	13.2		
SON	$\Delta q_{0.5}$	3.2	18.0	3.4	2.8		
	$\Delta q_{0.8}$	4.6	18.7	4.8	3.9	86.6	13.4
	$\Delta q_{0.9}$	5.7	19.3	6.0	4.9		

In contrast, an overall positive trend in return levels is seen during JJA and SON. Here, JJA shows the strongest overall increase in return levels with domain averages from 5-10 mm d^{-1} based on CER data. Domain-mean changes are positive for all quantiles, yet trends differ substantially between subregions. This is particularly noticeable in the differences between CER-based estimates in the north and south of the study region (e.g. Fig. 9). The overall increase in precipitation agrees with previous investigations for Central Europe, where an intensification in summer precipitation extremes was observed, scaling positively with the Northern Hemisphere temperature anomalies (Zeder and Fischer, 2020). The amplification of trend magnitude from $q_{0.5}$ to $q_{0.9}$ further supports the notion that the most intense summer events are changing most rapidly. However, it must be noted that CER data also show a systematic underestimation of the positive trend in summer extreme events compared to DWD observational data. This indicates some limitations of the CER in representing convective processes at daily time scales, which has also been indicated in previous evaluations (Bart et al., 2025).

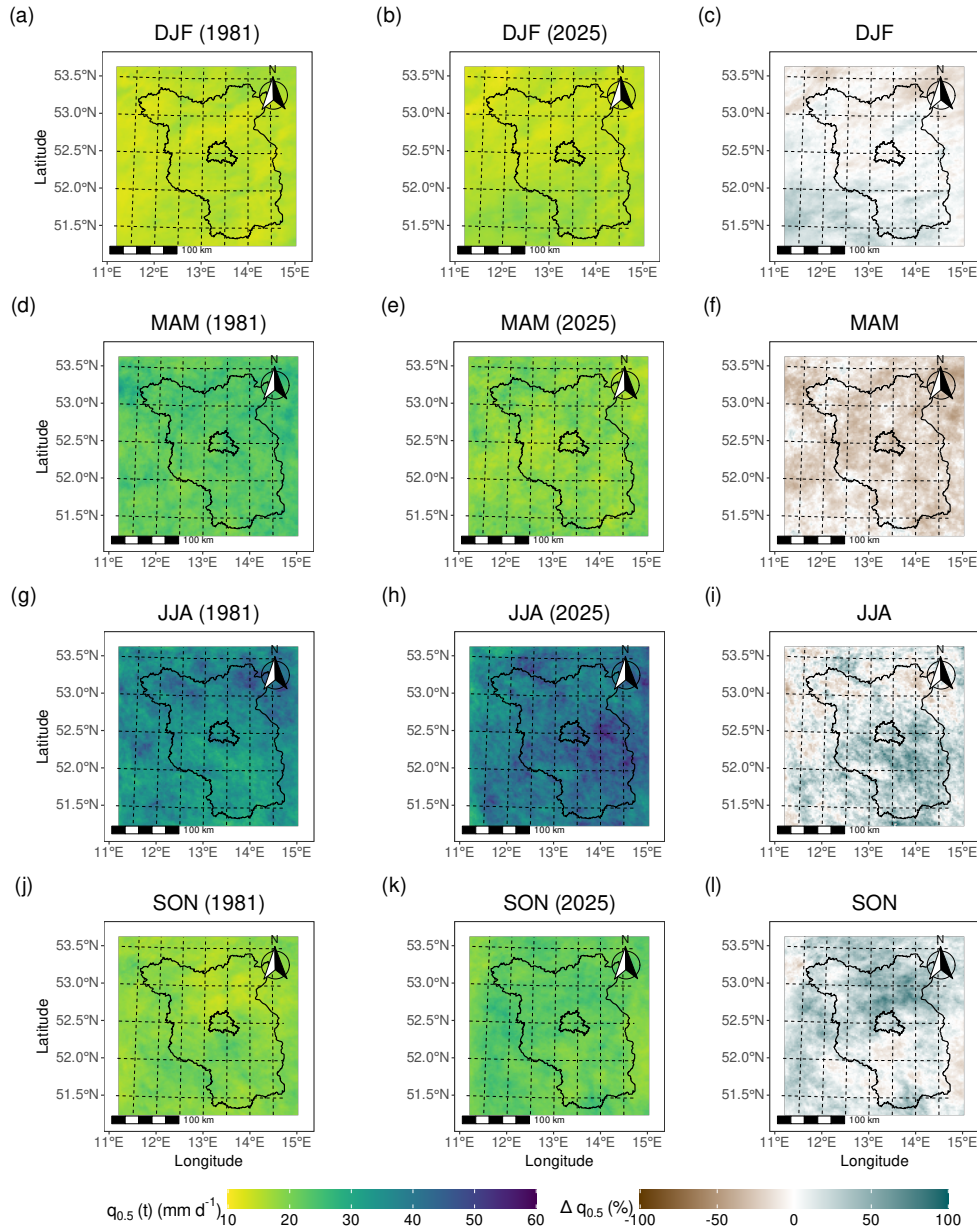


Figure 9. Seasonal return levels $q_{0.5}(t)$ (corresponding to a 2-year return period) and their relative changes $\Delta q_{0.5}$ from 1981 to 2025 derived from the non-stationary Generalised Pareto Distribution fitted to the data of the Central Europe Refined analysis version 2 (CER). The left and centre columns show the spatial distribution of $q_{0.5}(t)$ at $t = 1981$ and $t = 2025$, respectively, for winter (a, b), spring (d, e), summer (g, h), and autumn (j, k). The right column (c, f, i, l) shows the relative temporal change at each grid cell. The black outline denotes the boundaries of Berlin and Brandenburg.



An increase in return levels is also observable in both CER and DWD data during SON with a higher spatial coherence. Here, most CER grid cells exhibit positive trends and relative changes in q_p from 1981 to 2025 are comparable to JJA in some regions. This increase in return levels is especially noticeable in the north of the study region, where a seasonal shift in precipitation extremes between summer (decrease) to autumn (increase) becomes apparent (Fig. 9). Previous analyses of seasonal precipitation and drought indices over Central Europe have similarly reported wetting tendencies in autumn and winter, in contrast to spring drying (Spinoni et al., 2017; Stagge et al., 2017). Our findings extend this picture to daily extremes and suggest that enhanced cool-season precipitation in the region is accompanied by a systematic increase in the intensity of severe autumn events. This is consistent with previous findings of an overall increase in autumn precipitation extremes (Casanueva et al., 2014). In terms of spatial variability during autumn, we found that trends are typically less positive in the south and in some regions even negative, which was also observed in the western region of Poland (Lupikasza, 2010). Conversely, another investigation for the Czech Republic found increases in extreme precipitation indices, suggesting a higher spatial variability of changes in autumn extreme precipitation in Central Europe (Rulfová and Kyselý, 2014).

Return level estimates for DJF exhibit only modest increases/decreases, with spatially incoherent patterns. This is broadly consistent with previous studies over Germany and Central Europe, which report limited or weakly positive trends in winter heavy precipitation, despite a general increase in atmospheric moisture under warming conditions (Zeder and Fischer, 2020; Fauer and Rust, 2022). Furthermore, the intensity of return levels across all quantiles were also generally lower during winter than during any other season, which agrees well with previous investigations on the seasonality of precipitation extremes in the region (Fischer et al., 2019).

4.2 Possible physical causes

The seasonal contrast in return-level trends is consistent with the influence of different physical processes, although these mechanisms were not tested directly in the present analysis and go beyond the scope of our study. The following interpretations should therefore be regarded as plausible explanations rather than causal attribution. Particularly noteworthy in that context are the uniform decreases in spring return levels, the strong but spatially heterogeneous intensification of summer extremes, and the overall increase during autumn. The widespread reduction in return levels during spring is consistent with increasingly frequent dry springs and early-season droughts observed in Central Europe (Ionita et al., 2020). These conditions are driven by a persistent high-pressure system over the North Sea and northern Germany that, under positive NAO index conditions, diverts Atlantic storm tracks northward and suppresses large-scale frontal precipitation across the region (Hurrell and Deser, 2009; Ionita et al., 2020). Under such conditions, soil-moisture deficits can accumulate already in early spring and limit the potential for intense precipitation events later in the season (Ionita et al., 2020).

The summer increase and high spatial variability in return levels is consistent with a combination of greater atmospheric moisture availability and changes in the atmospheric circulation and organisation of precipitation-producing systems. The warming of the lower troposphere during recent decades and increases in near-surface humidity raise the moisture content that can be converted into precipitation during convective storms. The resulting precipitation events are usually short in duration and more intense, according to the Clausius–Clapeyron scaling and consistent with convection-permitting modelling studies



over Germany (Purr et al., 2021). The pronounced subregional differences in our results are consistent with spatial variations in convective organization, moisture supply, and mesoscale forcing, as also suggested by recent works on German convective environments and extreme short-term rainfall (Ulrich, 2022). However, this does not explain the differences in return level changes between the north and south of the domain that we found. One possible explanation could be related to recent findings that indicate an increase in summer NAO-index regarding variability and frequency of both positive and negative NAO extremes, amplifying their impacts on surface conditions over northwestern Europe (Liu et al., 2025). Furthermore, an overall positive trend in the summer NAO mean state might contribute to the decrease in precipitation extremes in northern parts of Berlin-Brandenburg, consistent with earlier work linking summer extremes to blocking situations (Fauer and Rust, 2022). Accordingly, the summer results of our analysis show some consistency with previous studies. The potential role of circulation changes should be tested using explicit NAO, blocking, or other large-scale covariates in future work.

A possible explanation for the intensification of autumn return levels is atmospheric warming, which could increase the air's moisture-holding capacity and thereby enhance precipitation intensity in large-scale weather systems (Gericke et al., 2019; Zeder and Fischer, 2020). An additional contribution may arise from the extension of the convective season into early autumn due to warmer conditions. Observational evidence from Central Europe shows that autumn is the only season in which both convective and stratiform precipitation simultaneously exhibit increasing trends, with stratiform extremes showing particularly pronounced increases likely linked to Mediterranean cyclone activity (Rulfová and Kysely, 2014). This dual intensification in both precipitation types is consistent with both thermodynamic amplification of individual weather systems and a lengthening of the connectively active season into September and October (Rulfová and Kysely, 2014). The spatial and temporal variability of the SON signal in our domain is further consistent with findings from adjacent regions, where autumn extreme precipitation trends are characterised by considerable temporal instability across different 30-year analysis windows (Lupikasza, 2010).

4.3 Sources of uncertainties

When interpreting return-level estimates and their trends, several sources of uncertainty should be considered. The DWD station records could be subject to gauge undercatch, instrumentation changes, and undocumented relocations that may introduce residual inhomogeneities despite quality control procedures (Dunn et al., 2025). The CER reanalysis is additionally affected by model and assimilation errors, which are, e.g., reflected in the systematic underestimation of return-level changes during summer relative to DWD, consistent with known limitations of reanalyses in representing convective extremes at daily timescales (Hu and Franzke, 2020). Changes in the observing system over time may further influence apparent trends in the reanalysis (Hersbach et al., 2020).

Notably, the northeastern and western border regions of the CER domain show systematic negative deviations that previous evaluations have attributed to border effects (Bart et al., 2025), which appear to extend far enough into the domain that removing the outermost 10 km does not fully compensate for them. Comparative investigations of extreme precipitation performance in the EURO-CORDEX initiative found that only few climate simulations can adequately represent temporal clustering of observed extreme precipitation in both summer and winter (Yang et al., 2023). Therefore, we estimate that reanalysis-downscaling data like CER should be used in conjunction with high-quality station observations when quantifying convective conditions.



A second source of uncertainty arises from the modeling framework. The linear time-varying scale parameter is a simplification that may not capture non-linear or low-frequency variability in the precipitation tails. The inclusion of additional large-scale covariates could further improve model accuracy. Model selection diagnostics indicate that the non-stationary GPD is most justified in JJA and SON, where the AIC identifies moderate to strong evidence at $\sim 25\%$ and $\sim 20\%$ of stations respectively, while the more conservative BIC reduces this to $\sim 5\%$. The LRT corroborates this, with $\sim 25\%$ of JJA stations reaching marginal significance ($p < 0.1$) and $\sim 12\%$ of SON stations achieving $p < 0.01$. That the non-stationary model does not underperform the stationary model at any station supports it as a suitable modelling choice, with limited formal significance being an expected consequence of signal-to-noise constraints in a 45-year record (Fig. 6, App. A2). Bootstrapped confidence intervals are narrowest in MAM and DJF, reflecting a better GPD fit for large-scale frontal extremes, and widest in JJA, where the irregular nature of convective rainfall increases uncertainty in the estimated return levels. The 90th seasonal percentile threshold balances bias and variance effectively, as confirmed by the parameter stability evaluation across the 85th–98th percentile range. Finally, the 45-year record limits the precision with which long return periods and temporal changes in the upper tail can be estimated. This is why the analysis focuses on 2-, 5-, and 10-year return levels. The results should not be extrapolated directly to longer return periods without additional data or a more strongly constrained model.

4.4 Added value of reanalysis-based data

Despite the discussed limitations, the implementation of reanalysis-based data products such as the CER has the potential to complement station-based investigations by identifying changes in precipitation extremes on a high spatial resolution. This may be especially useful for regions with only sparse ground-based observation networks. The CER fields provide a dense grid of spatially coherent time series, which enables the estimation of non-stationary GPD parameters and return-level trends at every grid point, and have the potential to allow for a more detailed assessment of regional differences. The overall agreement in the sign and approximate magnitude of return level changes between CER and DWD at station locations indicates that CER captures the main seasonal signals, even though the amplitude of change in summer extremes is underestimated compared to DWD station-based estimates.

The usage of convection-permitting, reanalysis-based data products provides added value wherever spatially coherent information on precipitation extremes is required and enables different applications. Examples include developing engineering design values for locations that lack extensive long-term observational records, the assessment of changing hazard patterns across river or lake basins, and the coupling with hydrological or impact models that require spatially distributed forcing data (Alexopoulos et al., 2023; Somogyvári et al., 2024). Overall, the combination of DWD station data and CER reanalysis exploits the strengths of both data sources: the local accuracy of gauges and the spatial completeness and physical consistency of the reanalysis, thereby enabling a more robust characterization of seasonal and spatial changes in precipitation extremes than would be possible from either dataset alone.



5 Conclusions

This study examines changes in seasonal daily precipitation return levels (2-, 5-, 10-year) across Berlin-Brandenburg between 1981 and 2025 using DWD station observations and the gridded high-resolution CER dataset within a non-stationary POT-GPD framework. The results reveal pronounced seasonal contrasts in the evolution of precipitation extremes. Spring return levels decline across nearly the entire CER domain, with relative decreases of up to approximately 19% across the analysed return period. Summer return levels increase most strongly, and in some locations the 2025 2-year return level approaches the magnitude of the 5-year return level estimated for 1981. Autumn exhibits a relatively coherent positive trend, whereas winter changes are weak and spatially inconsistent.

The spatial variability of return levels and their temporal changes strongly depend on the season. Spring and autumn show greatest spatial coherence, while summer exhibits largest spatial variability, particularly for the higher return periods. The DWD and CER datasets show consistent directions of change across seasons and return periods, supporting the robustness of the seasonal pattern. However, the magnitude of the changes differ between the datasets, with CER generally underestimating the positive summer changes relative to DWD. CER should therefore be regarded as a complement to, rather than a replacement for, station observations. Its spatial completeness enables regional analyses at locations without long-term gauge records, while the DWD data provide an essential reference for local changes.

Our findings are relevant to infrastructure planning and regional hazard assessment, as precipitation return levels vary with both season and time. They suggest that design approaches based on stationary return-level assumptions may not fully represent changing seasonal hazards. However, the estimates should not be interpreted as direct replacements for engineering design values without further assessment of model, dataset, and sampling uncertainties. Future work could investigate dynamical drivers of the observed spatial variability by incorporating large-scale climate covariates, such as the North Atlantic Oscillation and blocking frequency, into the extreme-value framework. Extending the analysis to sub-daily precipitation could be particularly relevant for assessing urban pluvial flood risk and for evaluating the applicability of the framework in regions with different precipitation regimes and observational network densities. As climate change and meteorological hazards continue to evolve, our findings can support more informed and seasonally differentiated approaches to risk assessment and adaptation planning in urban regions and beyond, facing the combined challenges of more severe wet extremes and prolonged dry periods.

Code availability. The R-code used for the analyses in this study is available on the GitLab server of the Chair of Climatology at the TU Berlin at <https://gitlab.klima.tu-berlin.de/bart/gpd>.

Data availability. The CER data set is available for download on the web server of the Chair of Climatology at the Technische Universität Berlin at <https://www.tu.berlin/en/klima/research/regional-climatology/central-europe/cer>. This data set is continuously updated with the release of new ERA5 data.



Appendix A: Introduction & Methods

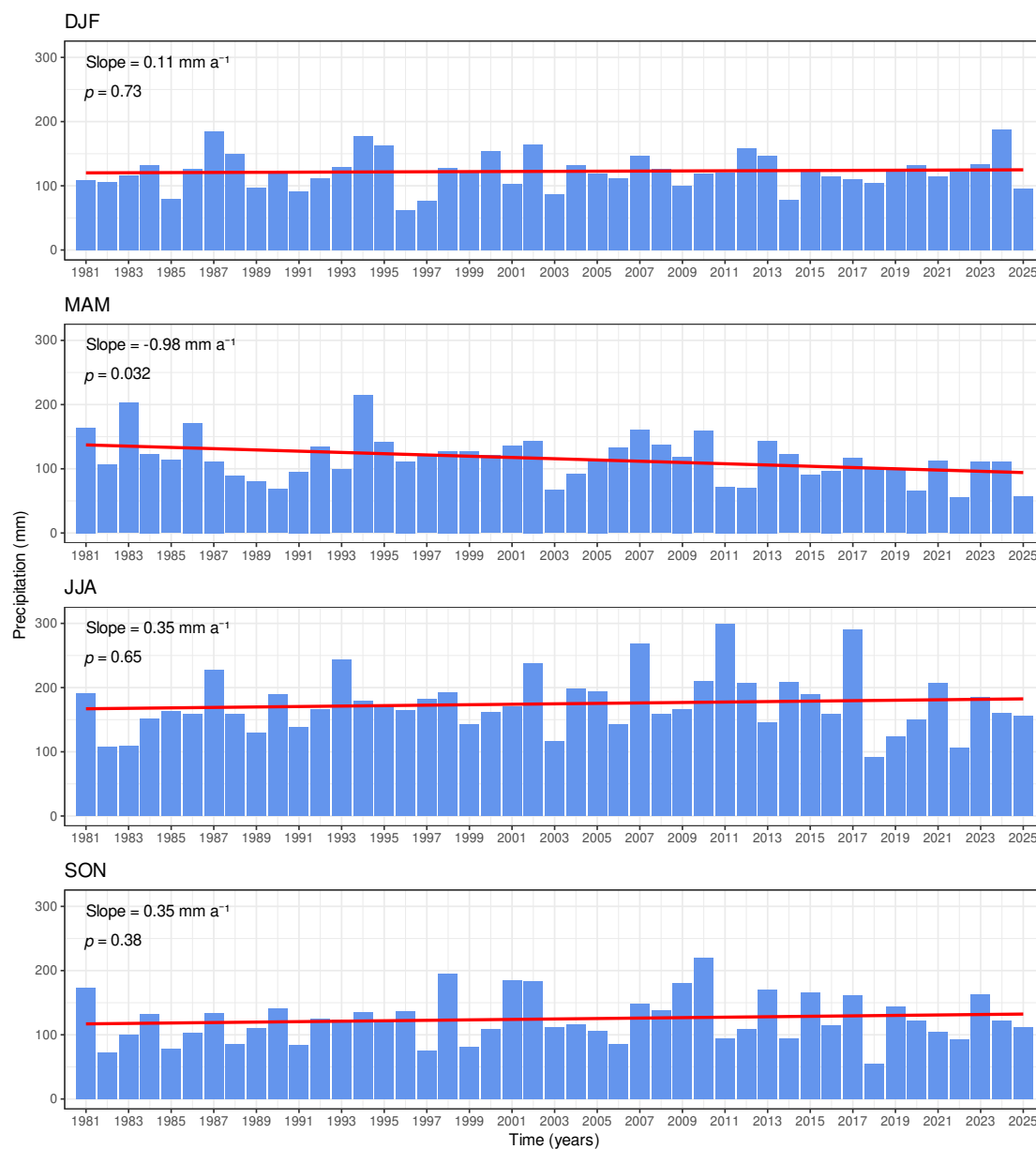


Figure A1. Average monthly precipitation sums across the selected 233 stations of the German Meteorological Service for each season and year. The trend was estimated using a linear regression (red line). The resulting change of precipitation over time (slope) is given in the top left corner with the corresponding p -value of the regression.

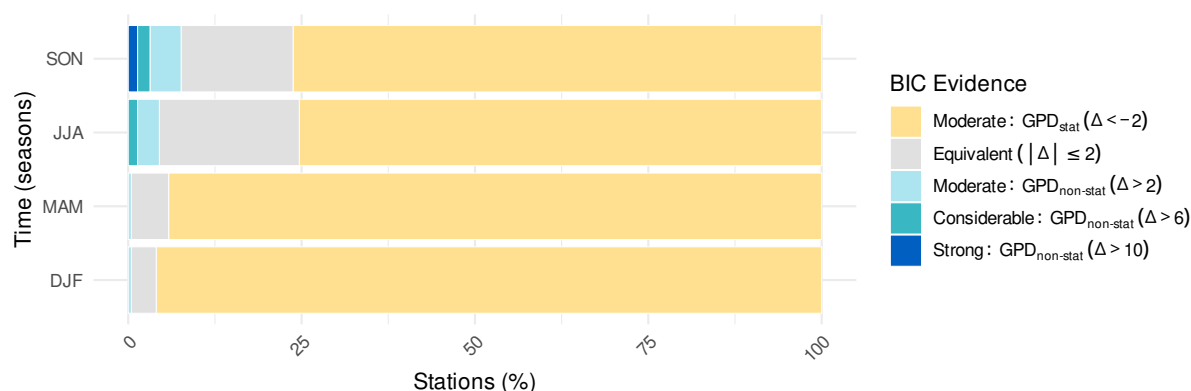


Figure A2. Bayesian Information Criterion (BIC)-based model comparison between the stationary (GPD_{stat}) and the non-stationary Generalised Pareto Distribution ($GPD_{non-stat}$) at each station of the German Meteorological Service, aggregated by season. The x-axis shows the percentage of stations falling into each evidence category, defined by the BIC difference $\Delta = BIC_{stat} - BIC_{non-stat}$. Positive Δ indicates support for $GPD_{non-stat}$, negative Δ for GPD_{stat} . Evidence thresholds were selected based on Kass and Raftery (1995).

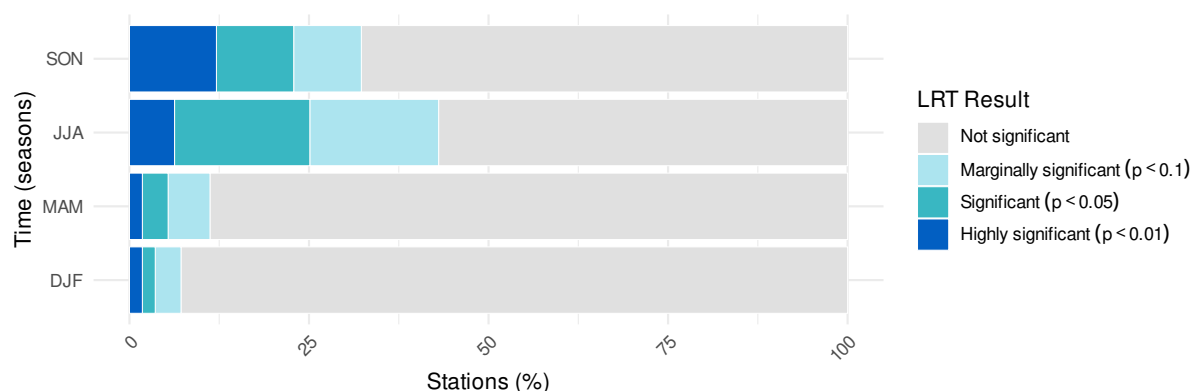


Figure A3. Likelihood Ratio Test (LRT) results comparing the stationary (GPD_{stat}) and the non-stationary Generalised Pareto Distribution ($GPD_{non-stat}$) at each station of the German Meteorological Service (aggregated by season). The LRT tests the null hypothesis that the additional time-dependent scale parameter of $GPD_{non-stat}$ does not significantly improve the model fit relative to GPD_{stat} . The x-axis shows the percentage of stations falling into each significance category: not significant ($p \geq 0.1$, grey), marginally significant ($p < 0.1$, light blue), significant ($p < 0.05$, medium blue), and highly significant ($p < 0.01$, dark blue).

Appendix B: Results

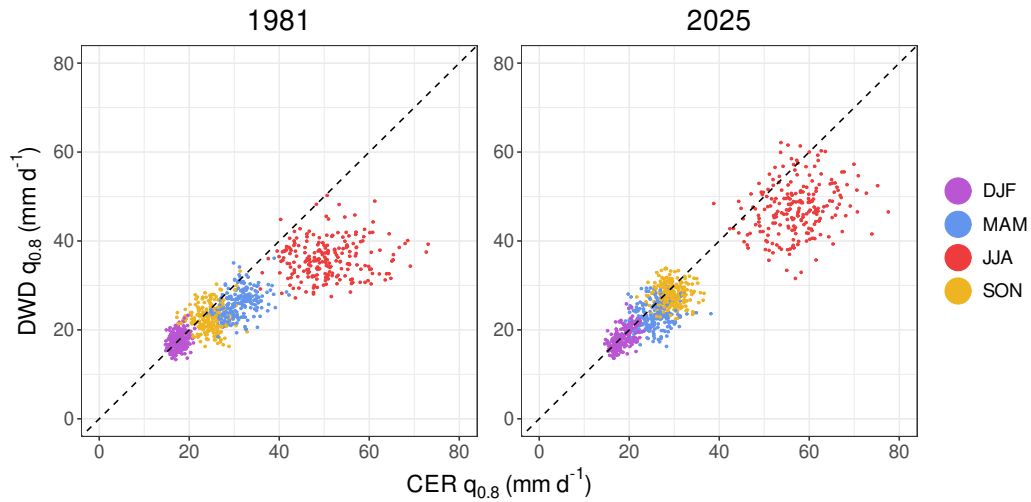


Figure B1. Comparison of time-dependent return levels ($q_{0.8}$, corresponding to a 5-year return period) derived from station observations of the German Meteorological Service (DWD) and the Central Europe Refined analysis version 2 (CER) data at the corresponding station locations, shown for $t = 1981$ (left) and $t = 2025$ (right). Each point represents one station-season combination, coloured by season. The dashed line indicates the 1:1 reference line of perfect agreement. Points above the line indicate higher DWD values relative to CER, and vice versa.

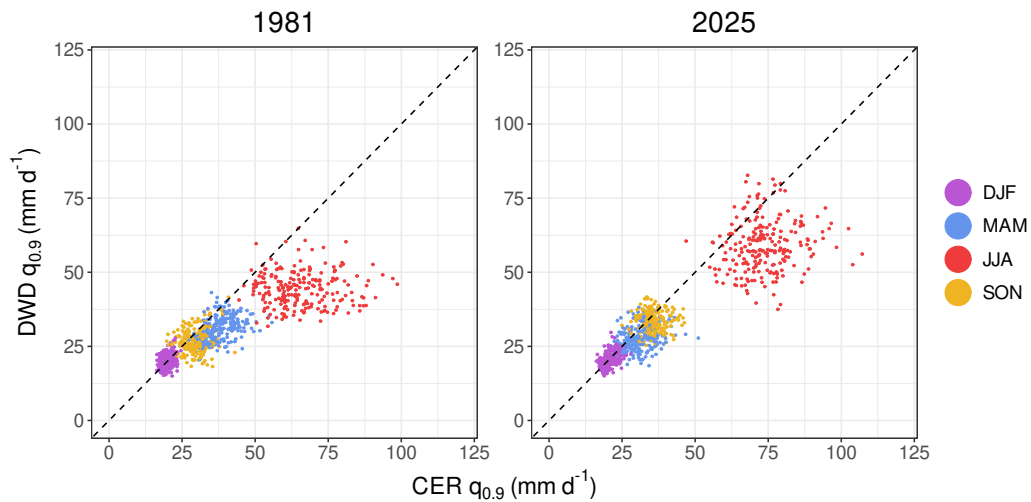


Figure B2. Comparison of time-dependent return levels ($q_{0.9}$, corresponding to a 10-year return period) derived from station observations of the German Meteorological Service (DWD) and the Central Europe Refined analysis version 2 (CER) dataset at the corresponding station locations, shown for $t = 1981$ (left) and $t = 2025$ (right). Each point represents one station-season combination, coloured by season. The dashed line indicates the 1:1 reference line of perfect agreement. Points above the line indicate higher DWD values relative to CER, and vice versa.

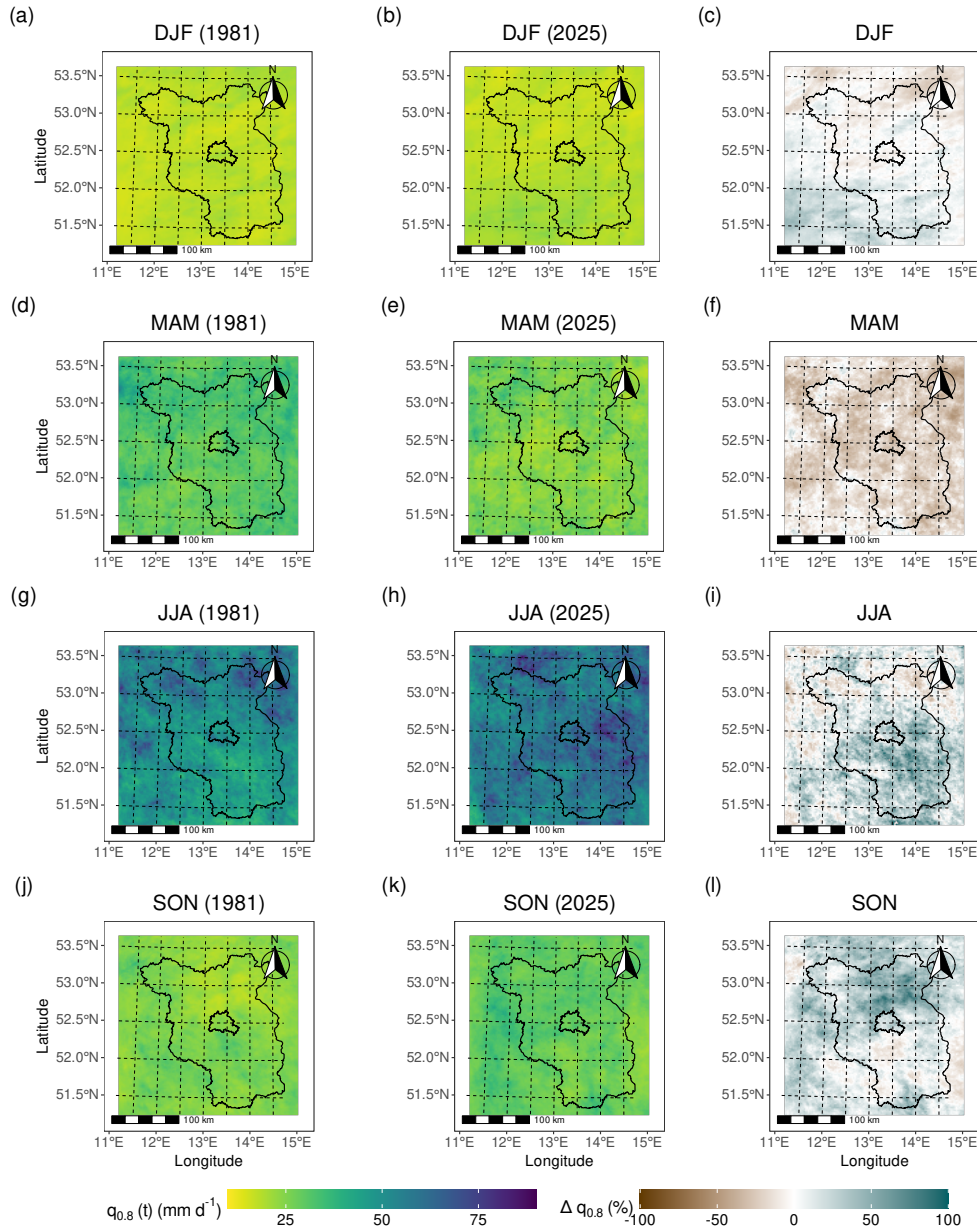


Figure B3. Seasonal return levels $q_{0.8}(t)$ (corresponding to a 5-year return period) and their relative changes $\Delta q_{0.8}$ between 1981 and 2025 derived from the non-stationary Generalised Pareto Distribution fitted to the data of the Central Europe Refined analysis version 2 (CER). The left and centre columns show the spatial distribution of $q_{0.8}(t)$ at $t = 1981$ and $t = 2025$, respectively, for winter (a, b), spring (d, e), summer (g, h), and autumn (j, k). The right column (c, f, i, l) shows the relative change at each grid cell. Positive values (blue) indicate an increase in the return level, negative values (brown) a decrease. The black outline denotes the boundary of the study region. Return levels are expressed in mm d^{-1} and the relative change in percent.

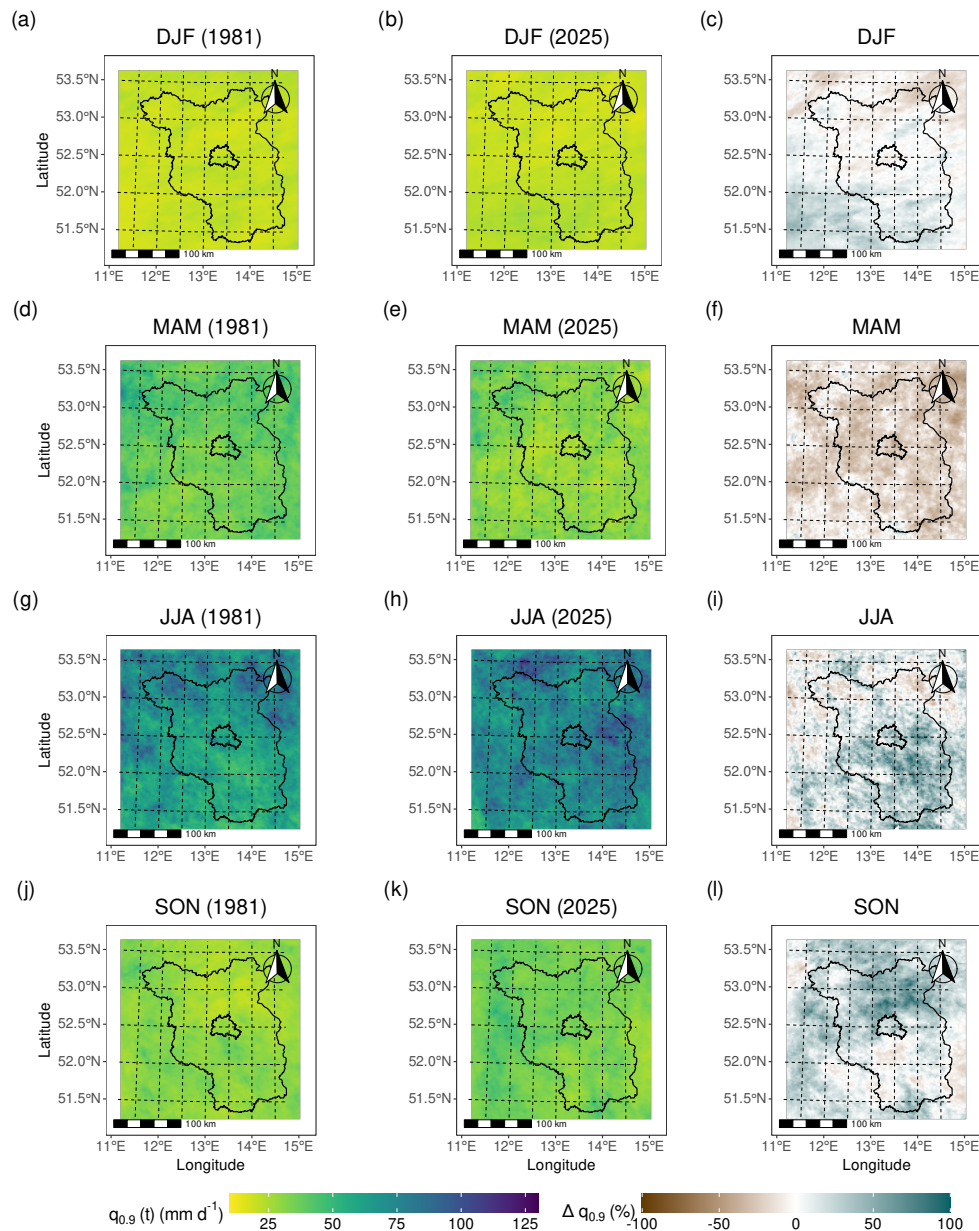


Figure B4. Seasonal return levels $q_{0.9}(t)$ (corresponding to a 10-year return period) and their relative changes $\Delta q_{0.9}$ between 1981 and 2025 derived from the non-stationary Generalised Pareto Distribution fitted to the data of the Central Europe Refined analysis version 2 (CER). The left and centre columns show the spatial distribution of $q_{0.9}(t)$ at $t = 1981$ and $t = 2025$, respectively, for DJF (a, b), MAM (d, e), JJA (g, h), and SON (j, k). The right column (c, f, i, l) shows the relative change at each grid cell. Positive values (blue) indicate an increase in the return level, negative values (brown) a decrease. The black outline denotes the boundary of the study region. Return levels are expressed in mm d^{-1} and the relative change in percent.



460 *Author contributions.* FB, MO, DF, HR, FM, AH and DS worked on the conceptualisation and methodology of the research activities. FB did the data analysis and visualisation. FB, MO, DF, HR, FM, AH and BS contributed to the writing, review and editing of the manuscript draft. BS and MO provided or helped with the usage of computing resources and other analysis tools. DF and DS served as supervisors for the research activities.

Competing interests. One of the co-authors is a member of the editorial board of NHESS. Aside from that, the authors declare that they have no conflict of interest.

465 *Acknowledgements.* We thank the German Meteorological Service (DWD) for the provision of observational data. This study was funded by the Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) in the context of the CELTIC-NEXT project SUSTAINET-Advance (Sustainable Technologies for Advanced Resilient and Energy-Efficient Networks – Advance) (16KIS2277). Generative artificial intelligence tools were used to assist with language editing and parts of the R code development. The authors reviewed and verified all AI-assisted outputs and take full responsibility for the content of the manuscript, the analysis, and the code.



470 References

- Akaike, H.: A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19, 716–723, <https://doi.org/10.1109/TAC.1974.1100705>, 1974.
- Alexopoulos, M. J., Müller-Thomy, H., Nistahl, P., Šraj, M., and Bezak, N.: Validation of Precipitation Reanalysis Products for Rainfall-Runoff Modelling in Slovenia, *Hydrology and Earth System Sciences*, 27, 2559–2578, <https://doi.org/10.5194/hess-27-2559-2023>, 2023.
- 475 Balkema, A. A. and de Haan, L.: Residual Life Time at Great Age, *The Annals of Probability*, 2, 792–804, <https://doi.org/10.1214/aop/1176996548>, 1974.
- Bart, F., Schmidt, B., Wang, X., Holtmann, A., Meier, F., Otto, M., and Scherer, D.: The Central Europe Refined Analysis Version 2 (CER v2): Evaluating Three Decades of High-Resolution Precipitation Data for the Berlin-Brandenburg Metropolitan Region, *Meteorologische Zeitschrift*, pp. 339–363, <https://doi.org/10.1127/metz/2024/1233>, 2025.
- 480 Berg, P., Christensen, O. B., Klehmet, K., Lenderink, G., Olsson, J., Teichmann, C., and Yang, W.: Summertime Precipitation Extremes in a EURO-CORDEX 0.11° Ensemble at an Hourly Resolution, *Natural Hazards and Earth System Sciences*, 19, 957–971, <https://doi.org/10.5194/nhess-19-957-2019>, 2019.
- Berghäuser, L., Schoppa, L., Ulrich, J., Dillenardt, L., Jurado, O. E., Passow, C., Samprogna Mohor, G., Seleem, O., Petrow, T., and Thieken, A. H.: Starkregen in Berlin : Meteorologische Ereignisrekonstruktion und Betroffenenbefragung, Tech. rep., Universität Potsdam, <https://doi.org/10.25932/publishup-50056>, 2021.
- 485 Burnham, K. P. and Anderson, D. R., eds.: *Model Selection and Multimodel Inference*, Springer, New York, NY, ISBN 978-0-387-95364-9, <https://doi.org/10.1007/b97636>, 2004.
- Casanueva, A., Rodríguez-Puebla, C., Frías, M. D., and González-Reviriego, N.: Variability of Extreme Precipitation over Europe and Its Relationships with Teleconnection Patterns, *Hydrology and Earth System Sciences*, 18, 709–725, [https://doi.org/10.5194/hess-18-709-](https://doi.org/10.5194/hess-18-709-2014)
- 490 2014, 2014.
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, Springer, London, ISBN 978-1-84996-874-4 978-1-4471-3675-0, <https://doi.org/10.1007/978-1-4471-3675-0>, 2001.
- Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., Ban, N., Bastin, S., Belda, M., Belusic, D., Caldas-Alvarez, A., Cardoso, R. M., Davolio, S., Dobler, A., Fernandez, J., Fita, L., Fumiere, Q., Giorgi, F., Goergen, K., Güttler, I., Halenka, T., Heinzeller, D., Hodnebrog, Ø., Jacob, D., Kartsios, S., Katragkou, E., Kendon, E., Khodayar, S., Kunstmann, H., Knist, S., Lavín-Gullón, A., Lind, P., Lorenz, T., Maraun, D., Marelle, L., van Meijgaard, E., Milovac, J., Myhre, G., Panitz, H.-J., Piazza, M., Raffa, M., Raub, T., Rockel, B., Schär, C., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Stocchi, P., Tölle, M. H., Truhetz, H., Vautard, R., de Vries, H., and Warrach-Sagi, K.: A First-of-Its-Kind Multi-Model Convection Permitting Ensemble for Investigating Convective Phenomena over Europe and the Mediterranean, *Climate Dynamics*, 55, 3–34, <https://doi.org/10.1007/s00382-018-4521-8>, 2020.
- 500 Cubasch, U. and Kadow, C.: Global Climate Change and Aspects of Regional Climate Change in the Berlin-Brandenburg Region, *Die Erde*, 142 (2011), 3–20, <https://doi.org/10.17169/refubium-2278>, 2019.
- Davison, A. C. and Hinkley, D. V.: *Bootstrap Methods and Their Application*, Cambridge University Press, ISBN 978-0-521-57471-6, 1997.
- Deumlich, D. and Gericke, A.: Frequency Trend Analysis of Heavy Rainfall Days for Germany, *Water*, 12, 1950, <https://doi.org/10.3390/w12071950>, 2020.
- 505 Dunn, R. E., Fowler, H. J., Green, A. C., and Lewis, E.: Tipping-Bucket Rain Gauges: A Review of the Undercatch Phenomenon, and Methods for Its Reduction and Correction, *Weather*, 80, 196–205, <https://doi.org/10.1002/wea.7736>, 2025.



- DWD: Daily Precipitation Observations for Germany, Version V24.3, 2024a.
- DWD: Historical Daily Station Observations (Temperature, Pressure, Precipitation, Sunshine Duration, Etc.) for Germany, Version V24.3, 2024b.
- 510 Fauer, F. S. and Rust, H.: Non-Stationary Large-Scale Statistics of Precipitation Extremes in Central Europe, *Stochastic Environmental Research and Risk Assessment*, 37, 4417–4429, <https://doi.org/10.1007/s00477-023-02515-z>, 2022.
- Fischer, M., Rust, H. W., and Ulbrich, U.: Seasonal Cycle in German Daily Precipitation Extremes, *Meteorologische Zeitschrift*, pp. 3–13, <https://doi.org/10.1127/metz/2017/0845>, 2018.
- Fischer, M., Rust, H. W., and Ulbrich, U.: A Spatial and Seasonal Climatology of Extreme Precipitation Return-Levels: A Case Study, *Spatial Statistics*, 34, 100 275, <https://doi.org/10.1016/j.spasta.2017.11.007>, 2019.
- 515 Gao, T., Wang, H. J., and Zhou, T.: Changes of Extreme Precipitation and Nonlinear Influence of Climate Variables over Monsoon Region in China, *Atmospheric Research*, 197, 379–389, <https://doi.org/10.1016/j.atmosres.2017.07.017>, 2017.
- Gericke, A., Kiesel, J., Deumlich, D., and Venohr, M.: Recent and Future Changes in Rainfall Erosivity and Implications for the Soil Erosion Risk in Brandenburg, NE Germany, *Water*, 11, 904, <https://doi.org/10.3390/w11050904>, 2019.
- 520 Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in R, *Journal of Statistical Software*, 72, 1–39, <https://doi.org/10.18637/jss.v072.i08>, 2016.
- Haslinger, K. and Mayer, K.: Early Spring Droughts in Central Europe: Indications for Atmospheric and Oceanic Drivers, *Atmospheric Science Letters*, 24, e1136, <https://doi.org/10.1002/asl.1136>, 2023.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- 525 Hu, G. and Franzke, C. L. E.: Evaluation of Daily Precipitation Extremes in Reanalysis and Gridded Observation-Based Data Sets Over Germany, *Geophysical Research Letters*, 47, e2020GL089 624, <https://doi.org/10.1029/2020GL089624>, 2020.
- Hurrell, J. W. and Deser, C.: North Atlantic Climate Variability: The Role of the North Atlantic Oscillation, *Journal of Marine Systems*, 78, 28–41, <https://doi.org/10.1016/j.jmarsys.2008.11.026>, 2009.
- International Organization for Standardization: ISO 14090:2019, <https://www.iso.org/standard/68507.html>, 2019.
- 535 International Organization for Standardization: ISO 14091:2021, <https://www.iso.org/standard/68508.html>, 2021.
- Ionita, M., Nagavciuc, V., Kumar, R., and Rakovec, O.: On the Curious Case of the Recent Decade, Mid-Spring Precipitation Deficit in Central Europe, *npj Climate and Atmospheric Science*, 3, 49, <https://doi.org/10.1038/s41612-020-00153-8>, 2020.
- Jänicke, B., Meier, F., Fenner, D., Fehrenbach, U., Holtmann, A., and Scherer, D.: Urban–Rural Differences in near-Surface Air Temperature as Resolved by the Central Europe Refined Analysis (CER): Sensitivity to Planetary Boundary Layer Schemes and Urban Canopy Models, *International Journal of Climatology*, 37, 2063–2079, <https://doi.org/10.1002/joc.4835>, 2017.
- 540 Kass, R. E. and Raftery, A. E.: Bayes Factors, *Journal of the American Statistical Association*, 90, 773–795, <https://doi.org/10.2307/2291091>, 1995.



- Lengfeld, K., Winterrath, T., Junghänel, T., Hafer, M., and Becker, A.: Characteristic Spatial Extent of Hourly and Daily Precipitation Events in Germany Derived from 16 Years of Radar Data, *Meteorologische Zeitschrift*, 28, 363–378, <https://doi.org/10.1127/metz/2019/0964>, 2019.
- Liu, Q., Bader, J., Jungclaus, J. H., and Matei, D.: More Extreme Summertime North Atlantic Oscillation under Climate Change, *Communications Earth & Environment*, 6, 474, <https://doi.org/10.1038/s43247-025-02422-x>, 2025.
- Lorenz, J. M., Kronenberg, R., Bernhofer, C., and Niyogi, D.: Urban Rainfall Modification: Observational Climatology Over Berlin, Germany, *Journal of Geophysical Research: Atmospheres*, 124, 731–746, <https://doi.org/10.1029/2018JD028858>, 2019.
- Lupikasza, E.: Spatial and Temporal Variability of Extreme Precipitation in Poland in the Period 1951–2006, *International Journal of Climatology*, 30, 991–1007, <https://doi.org/10.1002/joc.1950>, 2010.
- Maussion, F., Scherer, D., Finkelnburg, R., Richters, J., Yang, W., and Yao, T.: WRF Simulation of a Precipitation Event over the Tibetan Plateau, China – an Assessment Using Remote Sensing and Ground Observations, *Hydrology and Earth System Sciences*, 15, 1795–1817, <https://doi.org/10.5194/hess-15-1795-2011>, 2011.
- Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., and Finkelnburg, R.: Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis, *Journal of Climate*, 27, 1910–1927, <https://doi.org/10.1175/JCLI-D-13-00282.1>, 2014.
- Nguyen, V. D., Aerts, J., Tesselaar, M., Botzen, W., Kreibich, H., Alfieri, L., and Merz, B.: Exploring the Use of Seasonal Forecasts to Adapt Flood Insurance Premiums, *Natural Hazards and Earth System Sciences*, 24, 2923–2937, <https://doi.org/10.5194/nhess-24-2923-2024>, 2024.
- Pagenkopf, A.: Urbane Niederschlagsbeeinflussung, Doctoral Thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, <https://doi.org/10.18452/16421>, 2011.
- Passow, C. and Donner, R. V.: A Rigorous Statistical Assessment of Recent Trends in Intensity of Heavy Precipitation Over Germany, *Frontiers in Environmental Science*, 7, 143, <https://doi.org/10.3389/fenvs.2019.00143>, 2019.
- Peter, M., Rust, H. W., and Ulbrich, U.: Interannual Variations in the Seasonal Cycle of Extreme Precipitation in Germany and the Response to Climate Change, *Natural Hazards and Earth System Sciences*, 24, 1261–1285, <https://doi.org/10.5194/nhess-24-1261-2024>, 2024.
- Pickands, J.: Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics*, 3, 119–131, 1975.
- Pohle, I., Zeilfelder, S., Birner, J., and Creutzfeldt, B.: The 2018–2023 Drought in Berlin: Impacts and Analysis of the Perspective of Water Resources Management, *Natural Hazards and Earth System Sciences*, 25, 1293–1313, <https://doi.org/10.5194/nhess-25-1293-2025>, 2025.
- Poschlod, B.: Using High-Resolution Regional Climate Models to Estimate Return Levels of Daily Extreme Precipitation over Bavaria, *Natural Hazards and Earth System Sciences*, 21, 3573–3598, <https://doi.org/10.5194/nhess-21-3573-2021>, 2021.
- Purr, C., Brisson, E., and Ahrens, B.: Convective Rain Cell Characteristics and Scaling in Climate Projections for Germany, *International Journal of Climatology*, 41, 3174–3185, <https://doi.org/10.1002/joc.7012>, 2021.
- Rombeek, N., Hrachowitz, M., Droste, A., and Uijlenhoet, R.: Evaluation of High-Intensity Rainfall Observations from Personal Weather Stations in the Netherlands, *Hydrology and Earth System Sciences*, 29, 4585–4606, <https://doi.org/10.5194/hess-29-4585-2025>, 2025.
- Rulfová, Z. and Kysely, J.: Trends of Convective and Stratiform Precipitation in the Czech Republic, 1982–2010, *Advances in Meteorology*, 2014, 1–11, <https://doi.org/10.1155/2014/647938>, 2014.
- Saltikoff, E., Friedrich, K., Soderholm, J., Lengfeld, K., Nelson, B., Becker, A., Hollmann, R., Urban, B., Heistermann, M., and Tassone, C.: An Overview of Using Weather Radar for Climatological Studies: Successes, Challenges, and Potential, *Bulletin of the American Meteorological Society*, 100, 1739–1752, <https://doi.org/10.1175/BAMS-D-18-0166.1>, 2019.
- Schwarz, G.: Estimating the Dimension of a Model, *The Annals of Statistics*, 6, 461–464, 1978.



- Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., and Huang, X.-y.: A Description of the Advanced Research WRF Model Version 4.3, NCAR Technical Notes, <https://doi.org/10.5065/1dfh-6p97>, 2021.
- Somogyvári, M., Scherer, D., Bart, F., Fehrenbach, U., Okujeni, A., and Krueger, T.: A Hybrid Data-Driven Approach to Analyze the Drivers of Lake Level Dynamics, *Hydrology and Earth System Sciences*, 28, 4331–4348, <https://doi.org/10.5194/hess-28-4331-2024>, 2024.
- 585 Spinoni, J., Naumann, G., and Vogt, J. V.: Pan-European Seasonal Trends and Recent Changes of Drought Frequency and Severity, *Global and Planetary Change*, 148, 113–130, <https://doi.org/10.1016/j.gloplacha.2016.11.013>, 2017.
- Stagge, J. H., Kingston, D. G., Tallaksen, L. M., and Hannah, D. M.: Observed Drought Indices Show Increasing Divergence across Europe, *Scientific Reports*, 7, 14 045, <https://doi.org/10.1038/s41598-017-14283-2>, 2017.
- Terink, W., Leijnse, H., van den Eertwegh, G., and Uijlenhoet, R.: Spatial Resolutions in Areal Rainfall Estimation and Their Impact on
- 590 Hydrological Simulations of a Lowland Catchment, *Journal of Hydrology*, 563, 319–335, <https://doi.org/10.1016/j.jhydrol.2018.05.045>, 2018.
- Ulrich, J.: Statistical Modeling of Extreme Precipitation on Different Timescales: Towards Consistent and More Efficient Use of Available Data, Doctoral Thesis, Freie Universität Berlin, Berlin, <https://doi.org/10.17169/refubium-35793>, 2022.
- Wilks, S. S.: The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, *The Annals of Mathematical*
- 595 *Statistics*, 9, 60–62, 1938.
- Yang, L., Franzke, C. L. E., and Duan, W.: Evaluation and Projections of Extreme Precipitation Using a Spatial Extremes Framework, *International Journal of Climatology*, 43, 3453–3475, <https://doi.org/10.1002/joc.8038>, 2023.
- Zeder, J. and Fischer, E. M.: Observed Extreme Precipitation Trends and Scaling in Central Europe, *Weather and Climate Extremes*, 29, 100 266, <https://doi.org/10.1016/j.wace.2020.100266>, 2020.