



# Rethinking design rainfall estimation under non-stationary extremes and imperfect reanalysis data

Dimos Touloumidis<sup>1,2\*</sup>, Efstathios K. Oikonomou<sup>3</sup>

<sup>1</sup> Information Systems and e-Business Laboratory, Department of Applied Informatics, School of Information Sciences, University of Macedonia, 546 36 Thessaloniki, Greece.

<sup>2</sup> Centre for Research and Technology Hellas, Hellenic Institute of Transport, 57001 Thessaloniki, Greece.

<sup>3</sup> School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece.

\* Corresponding author. E-mail: touldimos@uom.edu.gr

**Abstract:** Design rainfall return levels are critical inputs for engineering applications such as urban drainage design, flood mitigation, resilient transport infrastructure and climate-sensitive planning. However, their estimation is significantly challenged by changing rainfall regimes, which can weaken the stationarity assumptions traditionally used in design rainfall analysis. A second challenge is data availability: design estimates rely on long gauge records that are often spatially sparse, while reanalysis products such as ERA5-Land provide spatial continuity but may poorly represent point-scale extremes, especially in convectively active and topographically complex settings. This study develops a reproducible GEE workflow to jointly examine these challenges by assessing temporal changes in extreme precipitation return levels and evaluating the reliability of ERA5-Land for design-oriented rainfall screening. Annual maxima from ERA5-Land daily precipitation are modelled using Gumbel and GEV distributions, with the workflow demonstrated globally and evaluated locally in Thessaloniki, Greece, against a long in-situ gauge record for return periods of 5, 10, 50 and 100 years. Results show that ERA5-Land systematically underestimates gauge-based design rainfall, with discrepancies increasing toward rarer events; consequently, gauge-to-ERA5-Land correction factors range from 1.17 to 1.29 under Gumbel and from 1.13 to 1.46 under GEV. A reference-to-recent period comparison further indicates substantial gauge-inferred intensification, including an approximately 33% increase in the 100-year Gumbel return level, while ERA5-Land shows little corresponding change. The proposed framework is reproducible and transferable for screening applications, but correction factors remain site-specific and require local gauge validation before engineering use.

**Keywords:** Extreme precipitation; Design rainfall; Extreme value theory; Correction factors.

## 1 INTRODUCTION

The hydraulic reliability of urban drainage elements (e.g., inlets, culverts, channels and underpasses) depends strongly on the magnitude of extreme rainfall that drives peak runoff (Willems, 2013; Ochoa-Rodriguez et al., 2015). In urban catchments, pluvial flooding can occur when rainfall intensity and accumulation exceed local conveyance capacity, producing rapid surface inundation and disruption to critical transport assets, as demonstrated for road-network disruption under pluvial flooding (Iliadis et al., 2023; Pregnolato et al., 2017); consequently, design rainfall (typically expressed through return levels for specified return periods) remains a key input to engineering design (Touloumidis et al., 2025) and flood-risk mitigation (Willems, 2013).

Estimating design return levels is most robust when long, homogeneous gauge records are available to constrain Extreme Value Theory (EVT) inference (Coles, 2001; Katz et al., 2002; Koutsoyiannis, 2004), yet such records are often sparse, unevenly distributed or not representative of local micro-climatic conditions (Kidd and Huffman, 2011). This limitation has accelerated the use of Gridded Precipitation Products (GPPs), including reanalyses such as ERA5-Land (Muñoz-Sabater et al., 2021) and multi-source merged precipitation datasets (Beck et al., 2017), while cloud-based platforms such as Google Earth Engine (GEE) have further enabled scalable, reproducible analysis of such datasets (Gorelick et al., 2017).

However, in addition to well-known limitations in gridded precipitation estimation (Maggioni et al., 2016), including retrieval uncertainty, model-physics limitations, calibration uncertainty, bias-correction uncertainty, sampling constraints and land-surface or topographic influences, GPPs do not measure the same quantity as a point gauge since daily precipitation from ERA5-Land represents an areal mean over a grid cell (~0.1°), whereas gauges record point rainfall. For extreme events (particularly convective storms with small spatial footprints) this representativeness difference attenuates annual maxima in gridded data relative to gauges (Eggert et al., 2015). Studies report systematic underestimation of extreme rainfall by reanalysis products, with biases amplifying in topographically complex and convectively active areas where spatial averaging most smoothes localized peaks (de Leeuw et al., 2015; Hamouda and Pasquero, 2021; Gomis-Cebolla et al., 2023; Lavers et al., 2022); critically, agreement in mean climatology does not imply agreement in the upper tail that governs design, as discrepancies emerge precisely at long return periods (>50 years) where extrapolation is most sensitive (Coles, 2001) and underestimation translates into under-design risk (Hamouda and Pasquero, 2021; Gomis-Cebolla et al., 2023; Lavers et al., 2022; Rivoire et al., 2022).

Taken together, these findings indicate that the critical issue for engineering use is whether gridded products preserve the upper-tail structure required for design inference, not merely whether they reproduce mean precipitation or moderate rainfall

variability. In point-scale applications, spatial averaging, imperfect representation of convective processes and local orographic/coastal controls can jointly dampen long-return-period estimates; for example, Lavers et al. (2022) and Gomis-Cebolla et al. (2023) report larger ERA5/ERA5-Land precipitation errors for heavy/extreme precipitation metrics, particularly where convective activity, complex terrain or regional precipitation-regime characteristics challenge reanalysis representation.

A second challenge is that conventional frequency analysis assumes stationarity of extremes (i.e., that exceedance probabilities inferred from historical samples remain valid for current conditions) (Coles, 2001; Milly et al., 2008). While fully non-stationary EVT frameworks exist (Coles, 2001; Katz et al., 2002), they are not yet routine in operational design; therefore, a practical alternative is a period-stratified "stress test", estimating return levels separately for earlier and later periods to assess whether design-relevant extremes have shifted over time (Kysely et al., 2011; Erić et al., 2021).

Recent efforts have compared gridded or model-based precipitation extremes against gauge networks, quantified biases through diagnostic metrics and explored non-stationary EVT and correction approaches under changing climate (de Leeuw et al., 2015; Lavers et al., 2022; Gomis-Cebolla et al., 2023; Maraun et al., 2017; Onderka et al., 2024). However, much of this literature still emphasizes characterizing discrepancies, while fewer studies translate these diagnostics into correction frameworks that planners can readily apply (de Leeuw et al., 2015; Lavers et al., 2022; Gomis-Cebolla et al., 2023), as many evaluations report performance without deriving transformation factors that adjust gridded return levels for engineering implementation (Maraun et al., 2017). Furthermore, rigorous tail evaluations in topographically complex, convectively dominated environments, where mechanisms of spatial mismatch are most pronounced, remain limited and workflows often rely on specialized tooling, which can constrain accessibility in data-scarce regions (Onderka et al., 2024; Erić et al., 2021).

Three operational barriers therefore persist (Gorelick et al., 2017; Tamiminia et al., 2020; Hamouda and Pasquero, 2021; Ferguglia et al., 2024): cloud-based workflows enabling reproducible, at-scale analysis accessible to non-specialists remain limited; evaluations rarely derive return-period-specific correction factors that reconcile gridded estimates with gauge-consistent design levels for direct engineering implementation; and systematic tail evaluation in convectively active, orographically complex environments is sparse, leaving uncertainty about GPP performance despite long-standing regionalization/scaling approaches used for Intensity-Duration-Frequency (IDF) curves (Bara et al., 2010; Gaál et al., 2009).

This study addresses these gaps using Thessaloniki, Greece, as a case study and ERA5-Land as the gridded dataset; Thessaloniki represents a challenging test environment as a coastal Mediterranean city where extreme rainfall is driven by convective activity, sea-breeze convergence and orographic enhancement, producing spatially heterogeneous storms that stress grid-scale representativeness. These mechanisms can operate across sub-daily to daily scales, as localized convection and thermodynamic controls are especially relevant for short-duration intensities, whereas organized synoptic and mesoscale dynamics can shape daily precipitation totals. Therefore, the use of daily annual maxima is appropriate for 24-h design return levels, but it cannot resolve sub-daily process mixtures that may be important for IDF applications or for separating thermodynamic from circulation-driven intensification (Andria et al., 2025; Haslinger et al., 2025; Marra et al., 2026).

Extreme rainfall is characterized using an Annual Maxima Series (AMS) of daily totals and return levels are estimated under both Gumbel and Generalized Extreme Value (GEV) models, with all computations implemented in GEE to ensure reproducibility and scalability. The objectives are to:

- i. demonstrate a reproducible GEE workflow for screening annual-maxima-based Gumbel and GEV return levels from gridded daily precipitation products;
- ii. evaluate ERA5-Land return levels against a long in-situ record at Thessaloniki for  $T = 5, 10, 50$  and 100 years, quantifying bias structure across the return-period spectrum;
- iii. derive return-period specific correction factors for the Thessaloniki gauge-ERA5-Land comparison and assess how such factors can support engineering interpretation at locally evaluated sites;
- iv. quantify differences in inferred design rainfall between a reference period (1955–1980) and a recent period (1981–2022) as a pragmatic stress test for temporal change in extreme rainfall.

By combining tail-focused evaluation, site-specific correction factors and a scalable GEE workflow, this study provides diagnostic insight into the divergence of ERA5-Land from gauge-based extremes and demonstrates how gridded precipitation products can be linked with local calibration to support design-oriented screening. The remainder of the paper presents the study area and datasets, describes the EVT implementation and evaluation approach, reports results on return levels, biases and correction factors (including bootstrap uncertainty quantification) and discusses implications for the engineering use of gridded extremes under climate variability and change.

## 2 STUDY AREA AND DATA

Thessaloniki, Greece (40.63°N, 22.96°E), is selected as the case-study location (Fig. 1) as it combines a long, well-documented rainfall record from the Aristotle University of Thessaloniki (AUTH) campus/Department of Meteorology and Climatology station with a documented history of extreme-rainfall analysis and demonstrated engineering relevance for urban drainage performance under design storms. The station record has been utilized in peer-reviewed assessments of rainfall intensity and extremes (Philandras et al., 2010), including analyses of very long-period extreme precipitation behavior (Pakalidou and Karacosta, 2018) and the statistics of extreme rainfall events (Douka and Karacostas, 2018), while recent urban water-infrastructure work evaluates the response of the combined sewer network under storm events of different return periods, reinforcing the suitability of the city for research into design return levels and correction factors (Galiatsatou et al., 2024).

Daily gauge totals were screened prior to extreme-value analysis using year-level completeness criteria; a calendar year was retained when at least 95% of daily observations were available and when no consecutive missing-data gap longer than 7 days occurred during the main heavy-rainfall season. Years not satisfying these criteria were excluded before extracting annual maxima so that the annual maximum was not inferred from an incomplete record during the period when intense rainfall is most likely to occur and from this screened daily series, an AMS of 24-h precipitation totals was derived, yielding 73 valid annual maxima for 1950–2022.

The evaluated gridded dataset is ERA5-Land (ECMWF), which was selected instead of ERA5 because it provides a land-

focused reanalysis at finer spatial resolution ( $\sim 0.1^\circ \times 0.1^\circ$ , approximately 9 km) and is widely used for land-surface and hydrological applications, although previous evaluations suggest that ERA5 and ERA5-Land precipitation differences can be modest in some regions (Gomis-Cebolla et al., 2023). ERA5-Land total precipitation is provided as accumulated water-equivalent depth; these values were converted from metres to millimeters per day (mm/d) for consistency with the gauge record and sampled at the station coordinates using nearest-grid-cell extraction. The resulting ERA5-Land AMS comprises 76 years (1950–2025) and all analyses were performed on AMS derived from daily totals to ensure consistency between the point observations and the gridded product in both temporal aggregation and extreme-event definition.

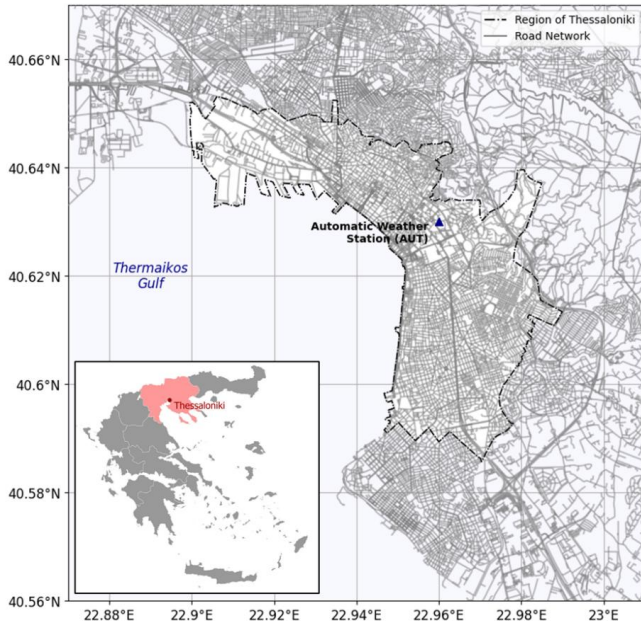


**Fig. 1.** Area of interest and location of in situ station.

The AUTH station is located at approximately 32 m a.s.l., while the corresponding ERA5-Land grid-cell elevation sampled at the station location is 38 m a.s.l.; this elevation information is reported because point-grid differences may reflect not only horizontal representativeness but also differences in effective terrain height. The gauge series used here represents daily total precipitation as a 24-h accumulation, while ERA5-Land daily totals were harmonized to daily water-equivalent precipitation totals before AMS extraction, ensuring the two datasets are consistent in accumulation duration and variable definition, although small differences in the exact daily accumulation window may remain and are treated as a minor source of timing uncertainty. Solid precipitation is not expected to dominate the annual daily maxima at this Mediterranean lowland site, but ERA5-Land total precipitation is expressed as water equivalent.

To provide a pragmatic stress test for potential temporal change, AMS and fitted return levels were estimated separately for a reference period (1955–1980) and a recent period (1981–2022), using 1980 as a practical breakpoint consistent with documented 1980s climate regime shifts across environmental systems and with common hydroclimate practice of contrasting pre-1980 and post-1980 precipitation statistics (Reid et al., 2016; North et al., 2013; Higgins et al., 2013); this cautious interpretation is particularly relevant for Mediterranean precipitation, where high temporal variability may dominate over monotonic trend signals (Vicente-Serrano et al., 2025),

while the split also roughly separates the pre-satellite and satellite-influenced reanalysis era around 1979, which may affect the quality and stability of ERA5/ERA5-Land precipitation estimates and further supports interpreting ERA5-Land period differences as descriptive diagnostics of temporal change.

The gauge and ERA5-Land records therefore have slightly different effective lengths because the gauge series is constrained by observational availability and quality screening, whereas ERA5-Land provides continuous gridded coverage; full-period return levels were estimated from the available AMS of each source, while direct event-space comparisons used only years with finite values in both series, and for the split-period analysis, the shorter samples within each time slice increase sampling uncertainty, especially for GEV estimates at long return periods, consequently making the reference–recent comparison function as a diagnostic stress test of temporal differences in return levels.

### 3 METHODOLOGY

#### 3.1 Extreme-value distribution fitting and return-level estimation

Extreme rainfall was characterized using a block maxima approach based on the AMS of daily precipitation totals (Kysely et al., 2011), wherein two extreme-value models were fitted to AMS: the GEV distribution (Coles, 2001) and the Gumbel distribution (Gumbel, 1962). The GEV cumulative distribution function (CDF) is defined as:

$$F(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad 1 + \xi \left( \frac{x - \mu}{\sigma} \right) > 0 \quad (1)$$

where  $\mu$ ,  $\sigma$ , and  $\xi$  denote the location, scale and shape parameters respectively. Parameters were estimated using the method of L-moments (probability-weighted moments, PWM) (Hosking, 1990) due to its robust behavior under moderate AMS sample sizes and suitability for stable server-side implementation; specifically, sample L-moments were computed from ordered AMS values using unbiased PWM weights (based on order index  $j$  and sample size  $n$ ), from which  $L_1$ ,  $L_2$  and the L-skewness  $\tau_3 = L_3/L_2$  were derived. For numerical stability,  $L_2$  was constrained to be positive (lower-bounded by a small  $\varepsilon$ ) and  $\tau_3$  was bounded to a finite interval prior to shape estimation using the standard Hosking-type approximation, which avoids iterative root solving and is well-suited to large-scale gridded processing. For a return period  $T$ , the non-exceedance probability is  $F_T = 1 - 1/T$  and the return level  $x_T$  is obtained by inverting the fitted CDF:

$$x_T = \begin{cases} \mu - \frac{\sigma}{\xi} \left[ 1 - (-\ln(F_T))^{-\xi} \right], & \xi \neq 0 \\ \mu - \sigma \ln[-\ln(F_T)], & \xi = 0 \end{cases} \quad (2)$$

In the implementation, the  $\xi = 0$  (Gumbel-limit) case was explicitly handled by switching to the closed-form Gumbel limit when  $|\xi|$  falls below a small threshold, ensuring the continuity of return levels as the estimated shape approaches zero.

The Gumbel model was fitted via linear least squares regression on Gumbel probability paper. Ordered maxima were assigned rank-based plotting positions consistent with the implementation ( $p_m = m/(n+2)$ ,  $q_m = 1 - p_m$ ) where  $m=1$  corresponds to the largest annual maximum and reduced variates were computed as  $y_m = -\ln[-\ln(q_m)]$  and the linear model  $x = a + by$  used to derive return levels for the specified design periods.

### 3.2 Return levels and period-change ratio mapping

For the global analysis, AMS were constructed per pixel from daily precipitation totals by extracting the maximum daily value within each calendar year; subsequently, GEV and Gumbel models were fitted at each pixel to compute return levels for periods spanning 5 to 100 years, thereby producing global return-level surfaces (Rivoire et al., 2022). A global period comparison was conducted exclusively for ERA5-Land by utilizing the extensive temporal coverage of the dataset, wherein two time windows were defined (reference vs recent), and models were fitted independently within each period. A period-change ratio was consequently expressed as the ratio of return levels between the recent and reference periods:

$$PCR_T = \frac{x_{T,Recent}}{x_{T,Reference}} \quad (3)$$

Similar relative-change calculations between earlier and later-period quantiles are commonly employed in climate-impact contexts (Onderka et al., 2024), here, the period-change ratio is used as a descriptive screening indicator of temporal differences in design extremes, without assigning those differences to a specific forcing mechanism.

### 3.3 Google Earth Engine implementation

Cloud-native workflows in GEE are widely recognized for enabling scalable Earth-observation analytics (Tamiminia et al., 2020; Gorelick et al., 2017) and have been successfully applied to diverse hydrological tasks ranging from global surface-water mapping (Pekel et al., 2016) to event-focused flood monitoring (DeVries et al., 2020). Leveraging this framework, all AMS extraction and EVT computations were operationalized within GEE to ensure reproducibility and scalable processing; specifically, the ERA5-Land daily precipitation product was ingested, filtered by analysis period and harmonized to consistent daily totals (including unit conversions where necessary). AMS were derived by extracting the maximum daily precipitation for each calendar year (computed per pixel for gridded mapping and at specific coordinates for the in-situ station), while statistical modeling relied on custom server-side functions developed to implement (i) probability-paper regression for the Gumbel distribution and (ii) L-moment/PWM estimation for the GEV distribution, thereby supporting a consistent period-stratified analysis across both the global and the local cases.

### 3.4 Product evaluation and derivation of design correction factors

For the Thessaloniki use case, return levels were computed for the in-situ gauge and each gridded precipitation product sampled at the station location; product performance was subsequently assessed by comparing return levels at matched return periods. Relative Bias (RB) was calculated as:

$$RB_T = \frac{x_{T,ERA5} - x_{T,Gauge}}{x_{T,Gauge}} \quad (4)$$

To translate discrepancies into implementable engineering guidance, multiplicative correction factors were derived per product and return period:

$$CF_T = \frac{x_{T,Gauge}}{x_{T,ERA5}} \quad (5)$$

In the present case study, the Correction Factor (CF) expresses the gauge–ERA5-Land discrepancy in engineering-relevant terms and provides a locally calibrated adjustment factor.

### 3.5 Model evaluation and uncertainty quantification

Uncertainty in return levels for the Thessaloniki evaluation (in-situ and ERA5-Land) was quantified using non-parametric bootstrap resampling of AMS; specifically, for each dataset,  $B=500$  bootstrap samples were generated by resampling annual maxima with replacement using a deterministic pseudo-random generator with fixed seeds to ensure reproducibility. For each replicate, Gumbel and GEV parameters were re-estimated and return levels recomputed, while the resulting distribution of  $x_T$  was summarized using 95% Confidence Intervals (CI):

$$CI_{95\%}(x_T) = [Q_{0.025}(x_T), Q_{0.975}(x_T)] \quad (6)$$

where  $Q_p$  denotes the  $p^{\text{th}}$  quantile of the bootstrap distribution, with the median reported as the central estimate.

## 5 RESULTS AND DISCUSSION

### 5.1 Global-scale return levels and period-change ratio maps

The GEE workflow was first applied globally to ERA5-Land annual maxima to demonstrate scalability and provide a climatological plausibility check before the station-based evaluation, wherein global return levels were derived for a range of design-relevant return periods (5, 10, 50 and 100 years) under both distributional assumptions to produce global Gumbel and GEV return-level surfaces (Fig. 2).

The Gumbel maps exhibit coherent climatological structures and highest extremes align with deep convection and monsoon regimes while dry subtropical belts and cold high-latitude interiors show lower values. Growth from shorter to longer return periods is spatially heterogeneous and tends to concentrate in regions where annual daily maxima are already large, consistent with known spatial patterns of heavy daily precipitation and supporting the internal consistency of the AMS extraction and fitting workflow (Westra et al., 2014). At the same time, this global plausibility does not imply local representativeness at the point scale, where spatial averaging, storm intermittency and localized convective structure can materially shift upper-tail behavior (Eggert et al., 2015; Hamouda and Pasquero, 2021; Lavers et al., 2022).

To examine whether allowing upper-tail flexibility changes global design levels, the same annual maxima were also modeled using GEV (Fig. 3). The large-scale spatial pattern remains similar to Gumbel, yet hotspot regions show stronger amplification at 50 and 100 years, reflecting the influence of the GEV shape parameter on long-return-period growth; this confirms that distributional choice becomes consequential precisely in the high-return-period range where engineering sensitivity is greatest and tail damping (if present) will be expressed most clearly (Pfahl et al., 2017). Comparable multi-dataset assessments show that continental-scale patterns of precipitation extremes are usually robust, but magnitudes and tails can diverge materially across gridded in-situ products and reanalyses, with the largest inter-dataset spread typically occurring for rare extremes and in regions where observational constraints or representativeness are challenging (Donat et al., 2014; Dunn et al., 2014). These supports treating the global maps here primarily as a scalability check while reserving engineering

interpretation of the upper tail for location-specific evaluation and local calibration.

For Europe, the broad spatial organization is consistent with the rationale of Rivoire et al. (2022), who also emphasized that high return-level estimates from ERA5 precipitation require

careful treatment of spatial dependence and tail behavior; however, the present global maps are used only as screening and scalability outputs, as engineering interpretation remains tied to the local gauge-based evaluation.

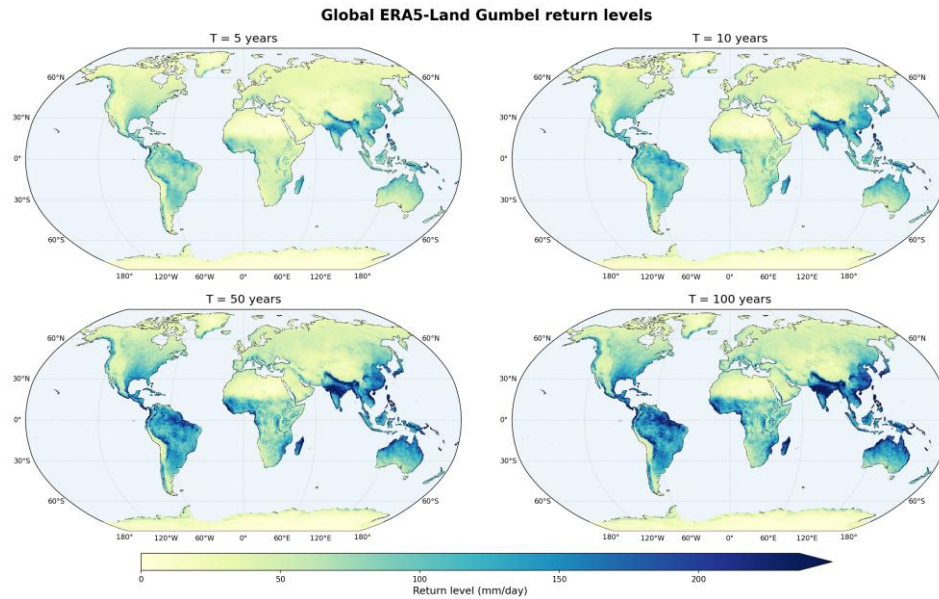


**Fig. 2.** Global Gumbel return-level maps from ERA5-Land daily precipitation for  $T = 5, 10, 50$  and  $100$  years [mm/day].

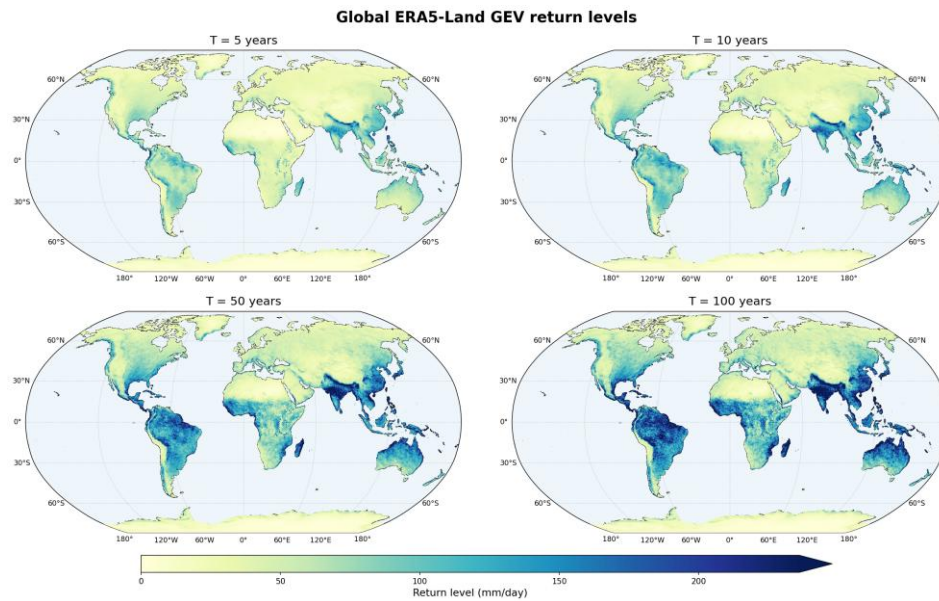


**Fig. 3.** Global GEV return-level maps from ERA5-Land daily precipitation for  $T = 5, 10, 50$  and  $100$  years [mm/day].

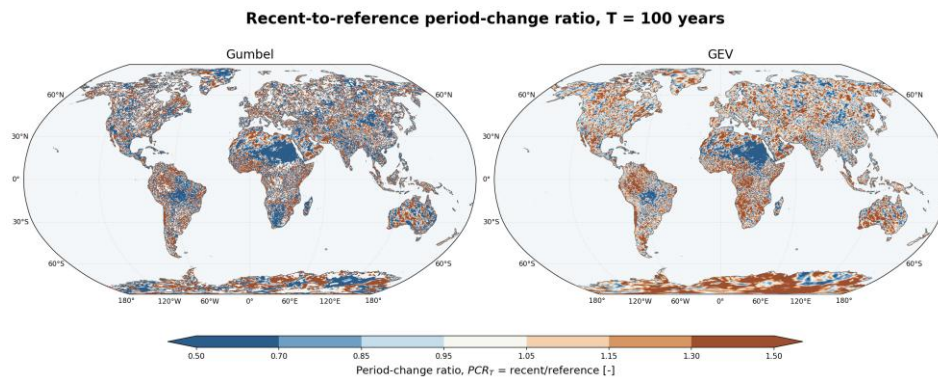


**Fig. 4.** Global period-change ratio  $PCR_{100}$  (recent/reference) for ERA5-Land: (a) Gumbel and (b) GEV.

A split-period analysis was used as a screening indicator of temporal differences in ERA5-Land design extremes. Fig. 4 presents the recent-to-reference period-change ratio of the 100-year return level ( $PCR_{100}$ ) under Gumbel and GEV; both show pronounced spatial heterogeneity rather than a uniform global signal. Differences between panels reflect distributional choice, as GEV can amplify contrasts where tail shape influences long-return-period growth, while Gumbel-based ratios can display localized extreme values where the reference-period return level is very small and estimation becomes numerically unstable. Extreme ratio values should therefore be interpreted cautiously, especially in regions where reference-period return levels are small or where ERA5 precipitation errors are known to be large; this is particularly relevant in tropical regions where reanalysis precipitation uncertainty can be substantial (Lavers et al., 2022).  $PCR_{100}$  therefore serves as a screening layer to flag regions where ERA5-Land implies temporal differences in design extremes and where more detailed regional or site-specific evaluation may be valuable and to motivate site-specific evaluation rather than as a standalone adjustment for design (O’Gorman, 2015). The strong spatial heterogeneity in  $PCR_{100}$  is also consistent with regional evidence that changes in heavy precipitation can be non-uniform and regime-dependent, including elevation-conditioned behavior in reanalyses and projected changes in extremes that become more pronounced toward the upper tail (Ferguglia et al., 2024; Kysely et al., 2011).

An additional numerical qualification concerns the few non-physical values that may appear in the Gumbel-based ratio map in extremely dry regions (Fig. 4); although precipitation return levels are physically non-negative, the unconstrained Gumbel fit can occasionally produce near-zero or slightly negative fitted reference-period return levels in hyper-arid grid cells where annual maxima are very small and parameter estimation is unstable, consequently requiring that the resulting ratio values be interpreted as fitting artefacts rather than meaningful precipitation-change signals and excluded from any engineering interpretation of  $PCR_{100}$ .

## 5.2 Annual-maxima behaviour at Thessaloniki and correspondence in event space

The station-based evaluation focuses on Thessaloniki (AUTH gauge) and the co-located ERA5-Land grid cell sampled at the station location, where annual maxima were extracted for both datasets to yield 73 valid gauge years and 76 ERA5-Land years;

these sample sizes define the basis for the subsequent inference of extreme values.

Paired annual maxima were compared only for years with finite values in both datasets to assess year-paired correspondence between the gauge and the co-located ERA5-Land grid cell (Fig. 5). The comparison shows substantial dispersion and increasing departure from parity at higher magnitudes, with Mean Absolute Error (MAE) = 12.1 mm, Root Mean Square Error (RMSE) = 16.5 mm, Mean Absolute Percentage Error (MAPE) = 26.7% and coefficient of determination ( $R^2$ ) = -0.10. These metrics should not be interpreted as strict storm-event matching performance, because the annual maximum recorded at a point gauge and the annual maximum represented by a grid-cell mean may occur on different dates within the same year; rather, they indicate that the two series do not preserve the same magnitude structure in event space. A date-lag or storm-event matching analysis would provide a useful complementary diagnostic of timing correspondence.

The reduced-variate comparison further suggests that the central part of the distribution is more consistent than the upper tail, where ERA5-Land increasingly falls below the gauge, consistent with documented underestimation of heavier daily precipitation categories in ERA5/ERA5-Land evaluations (Gomis-Cebolla et al., 2023; Lavers et al., 2022). The separation becomes visually evident for annual maxima above roughly 40 mm/day, and this threshold should be interpreted as an empirical feature of the Thessaloniki sample rather than as a universal physical cutoff, as a plausible explanation is that higher daily totals are increasingly influenced by localized convective or coastal–orographic events whose spatial footprint is poorly represented by a ~9 km grid-cell mean; this structured mismatch aligns with spatial averaging, storm intermittency and localization of convection acting together with coastal and topographic modulation of Mediterranean extreme precipitation (Mastrantonas et al., 2021; Dayan et al., 2015).

## 5.3 Full-period design return levels and relative bias

Return levels were estimated from the full annual-maxima samples using both models described in the Methods and evaluated across the standard design spectrum (5–100 years); the resulting return levels for the gauge and ERA5-Land, together with the corresponding relative bias ( $RB_T$ ), are reported in Table 1.

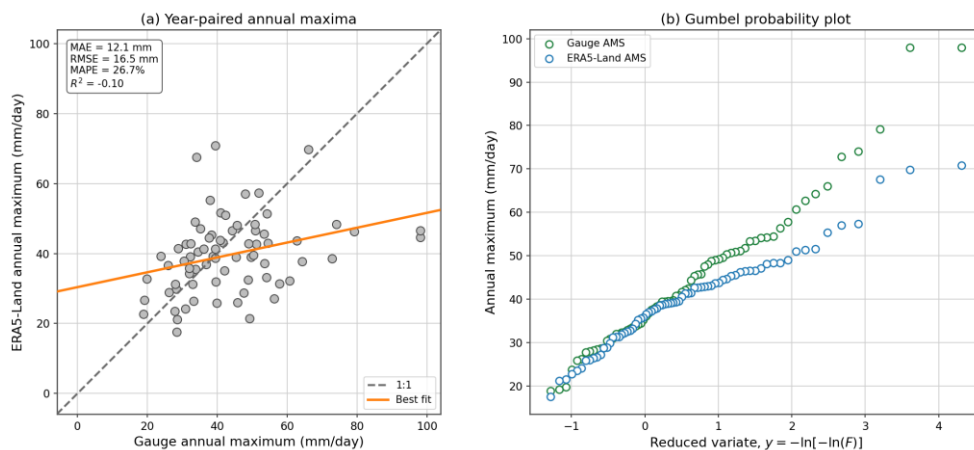


Fig. 5. Annual maxima correspondence scatter and Gumbel probability plot for the Thessaloniki gauge and ERA5-Land.

**Table 1.** Full-period design return levels and relative bias at Thessaloniki.

| Return period<br>(years) | $X_{T,Gauge}$<br>(Gumbel)<br>[mm/day] | $X_{T,ERA5}$ (Gumbel)<br>[mm/day] | $RB_T$<br>(Gumbel) | $X_{T,Gauge}$<br>(GEV)<br>[mm/day] | $X_{T,ERA5}$<br>(GEV)<br>[mm/day] | $RB_T$<br>(GEV) |
|--------------------------|---------------------------------------|-----------------------------------|--------------------|------------------------------------|-----------------------------------|-----------------|
| 5                        | 55.7                                  | 47.7                              | -0.15              | 54.4                               | 48.3                              | -0.11           |
| 10                       | 65.7                                  | 54.3                              | -0.17              | 64.0                               | 53.5                              | -0.16           |
| 50                       | 87.6                                  | 68.9                              | -0.21              | 86.3                               | 62.8                              | -0.27           |
| 100                      | 96.9                                  | 75.1                              | -0.22              | 96.3                               | 66.0                              | -0.32           |

Across return periods and under both models, ERA5-Land produces lower design return levels than the gauge, and the separation increases toward rarer events (Gomis-Cebolla et al., 2023; Lavers et al., 2022; Rivoire et al., 2022). Specifically, under GEV, the gauge increases from 54.4 mm/day (5 years) to 96.3 mm/day (100 years) whereas ERA5-Land increases from 48.3 to 66.0 mm/day, yielding a widening gap that exceeds 30 mm/day at 100 years; under Gumbel, the gap is smaller but remains substantial. The return-period dependence indicates that the mismatch is not a constant offset but a damped upper tail in the gridded product, consistent with known limitations of gridded representations of point-scale extremes (Peleg et al., 2024; Hamouda and Pasquero, 2021; Eggert et al., 2015; Ochoa-Rodriguez et al., 2015).

An exploratory comparison across the fitted design spectrum indicates that ERA5-Land underestimation is already present at the lower end of the annual-maxima distribution but becomes more pronounced toward the upper tail. For example, at  $T = 5$  years, ERA5-Land is lower than the gauge by approximately 8.0 mm/day under Gumbel and 6.1 mm/day under GEV, whereas at  $T = 100$  years the corresponding differences increase to approximately 21.8 and 30.3 mm/day, respectively; this suggests that the bias is not limited to the most extreme extrapolated quantiles, but its practical importance increases strongly for rare design levels.

The present analysis is intentionally focused on annual maxima and therefore does not diagnose the full precipitation distribution or annual accumulation bias; consequently, the upper-tail underestimation reported here characterizes the design-relevant extremes and does not by itself describe mean daily precipitation or annual-total bias. In Mediterranean climates, annual accumulation may depend strongly on a small number of storms, but compensating errors across weak, moderate and heavy daily precipitation classes can allow a product to reproduce mean totals while still damping design-relevant extremes.

A further interpretation concerns the event-generating mechanisms represented within the AMS; in a Mediterranean coastal environment such as Thessaloniki, annual maxima may arise from different precipitation regimes, including localized convective storms, organized frontal systems, sea-breeze-related convergence and orographically enhanced rainfall. Consequently, the AMS may combine events generated by distinct seasonal or dynamical mechanisms; although this does not invalidate the block-maxima approach, it means that the classical EVT assumption of annual maxima as independent and identically distributed draws from a stationary homogeneous parent distribution may be only approximately satisfied, and in this context, ERA5-Land underestimation may reflect not only point-grid smoothing of intense convective cores but also imperfect representation of the relative frequency, intensity and spatial organization of the regimes generating station-scale annual maxima, particularly given the regime-dependent nature of Mediterranean extreme precipitation and its links to large-

scale atmospheric patterns, convection, coastal effects and orographic modulation (Dayan et al., 2015; Mastrantonas et al., 2021).

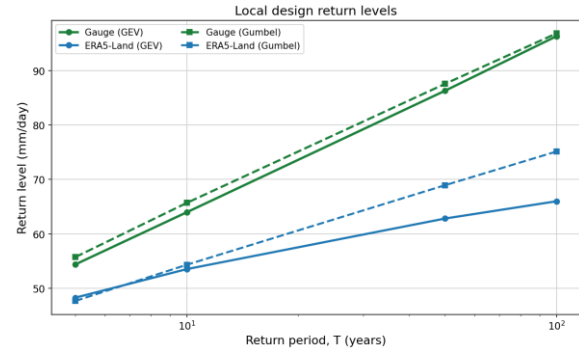
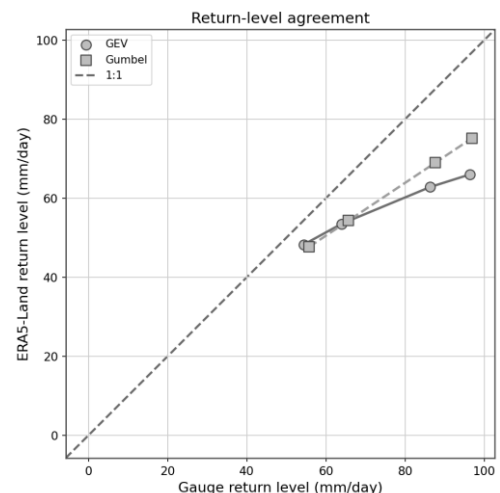
**Fig. 6.** Design return-level curves for return periods of 5, 10, 50, and 100 years.

Fig. 6 visualizes the widening curve separation (which is more pronounced under GEV), reflecting the interaction between distributional flexibility and tail behavior where GEV increases sensitivity to rare-event structure in the gauge record while ERA5-Land remains comparatively damped; consequently, for design interpretation, this supports presenting Gumbel and GEV as complementary envelopes rather than treating one as a redundant sensitivity case.

It is also notable that Gumbel and GEV estimates are relatively close for the gauge over most of the evaluated range, whereas the ERA5-Land curves show stronger model dependence at longer return periods; this supports retaining both models because Gumbel provides a parsimonious and more stable benchmark while GEV exposes sensitivity to tail-shape estimation, allowing the two models to be interpreted as complementary sensitivity bounds rather than as competing claims of a single definitive design value.

**Fig. 7.** Agreement in return-level space.

Agreement in return-level space (Fig. 7) shows systematic departure from the one-to-one relationship that strengthens with increasing return levels, reinforcing that the discrepancy is structured and concentrated in the design-relevant upper tail; in engineering terms, this implies that uncorrected gridded return levels can embed systematic under-design of drainage and flood-mitigation measures, with the largest deficits occurring where design checks commonly anchor (50–100 years) and where rainfall–runoff responses and capacity thresholds can make consequences disproportionate (Ochoa-Rodriguez et al., 2015; Willems, 2013).

#### 5.4 Correction factors for design adjustment

To express the discrepancy in a compact form suitable for engineering applications, return-period-specific correction factors  $CF_T$  (calculated as the ratio of the gauge to ERA5-Land) were computed at each return period for both models (Table 2).

**Table 2.** Return-period-specific correction factors  $CF_T$  across return periods.

| Return period<br>(years) | $CF_T$<br>(Gumbel) | $CF_T$<br>(GEV) |
|--------------------------|--------------------|-----------------|
| 5                        | 1.17               | 1.13            |
| 10                       | 1.21               | 1.20            |
| 50                       | 1.27               | 1.37            |
| 100                      | 1.29               | 1.46            |

The correction factors increase with return period under both models, consistent with a tail-dependent discrepancy rather than a constant multiplicative bias; specifically, under GEV the required adjustment reaches 1.46 at 100 years whereas under Gumbel it reaches 1.29. Operationally, this provides an implementable bridge wherein ERA5-Land can be treated as a spatially complete signal aligned to gauge-consistent tail behavior through return-period-specific calibration rather than as a substitute for local extremes. While related work has addressed the same design-translation problem via explicit bias-correction of extreme quantiles (e.g., Onderka et al. (2024) apply a time-varying GEV framework to correct upper-tail rainfall quantiles from regional climate model outputs), the approach presented here provides a site-specific, return-period-indexed multiplicative correction for ERA5-Land anchored to the observed Thessaloniki gauge–reanalysis discrepancy. Their use at other locations, periods or return periods would require additional gauge-based validation or regionalization.

The monotonic trend (Fig. 8) demonstrates that a single constant adjustment would not preserve return-period dependence. For hydraulic sizing and resilience of transport infrastructure, this distinction is critical because design rainfall

enters nonlinearly into surcharge, overtopping and surface-flow concentration (Ochoa-Rodriguez et al., 2015; Willems, 2013); consequently, a damped tail can shift events into more frequent exceedance regimes and increase the likelihood of closures and cascading disruption during intense storms (Pregolato et al., 2017), while the dependence of correction magnitude on distributional choice further motivates reporting both models as bounds for uncertainty-aware design interpretation (Coles, 2001; Maraun et al., 2017).

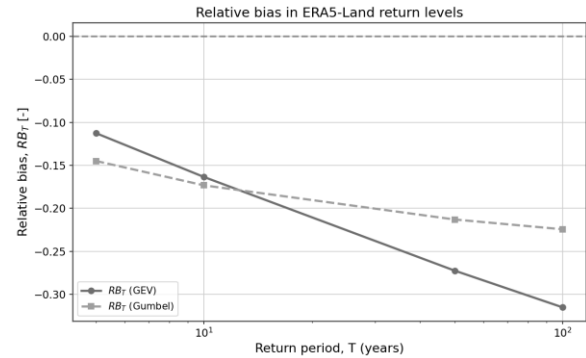


**Fig. 8.** Relative bias  $RB_T$  of ERA5-Land return levels with respect to the gauge across return periods  $T$  (GEV and Gumbel).

#### 5.5 Reference and recent period stratification: changes in return levels

Return levels were estimated separately for a reference period (1955–1980) and a recent period (1981–2022) to provide a test of whether design-relevant extremes differ between time slices; consequently, the slice-specific return levels and period-change ratios ( $PCR_T$ ) are reported for Gumbel in Table 3 and for GEV in Table 4. The gauge indicates substantial reference-to-recent increases (especially for rarer events) whereas ERA5-Land shows muted changes and under Gumbel, slight decreases across all evaluated return periods. This split-period comparison provides a diagnostic stress test of temporal differences and highlights that ERA5-Land alone may understate both the magnitude of the reference-period tail and the change implied by local observations (Bara et al., 2010).

Under Gumbel (Table 3), the gauge indicates a clear reference-to-recent increase, with  $PCR_T$  rising from 1.16 at  $T = 5$  years to 1.33 at  $T = 100$  years, corresponding to an approximate 33% increase in the 100-year daily return level, whereas ERA5-Land shows  $PCR_T$  values close to or slightly below 1.00 across all evaluated return periods; this contrast suggests that the inferred period-change signal depends on both the source of the precipitation data and the tail model, reinforcing the value of aligning gridded period-change diagnostics with local evidence where such evidence is available.

**Table 3.** Reference and recent period return levels and period-change ratios  $PCR_T$  under Gumbel.

| Return pe-<br>riod (years) | Gauge refer-<br>ence | Gauge recent | $PCR_T$ (Gauge,<br>Gumbel) | ERA5 reference | ERA5 recent | $PCR_T$ (ERA5,<br>Gumbel) |
|----------------------------|----------------------|--------------|----------------------------|----------------|-------------|---------------------------|
| 5                          | 50.6                 | 58.8         | 1.16                       | 48.4           | 48.0        | 0.99                      |
| 10                         | 58.0                 | 70.9         | 1.22                       | 55.8           | 55.2        | 0.99                      |
| 50                         | 74.5                 | 97.4         | 1.31                       | 72.1           | 70.9        | 0.98                      |
| 100                        | 81.4                 | 108.6        | 1.33                       | 79.0           | 77.6        | 0.98                      |

**Table 4.** Reference and recent period return levels and period-change ratios  $PCR_T$  under GEV.

| Return period (years) | Gauge reference | Gauge recent | $PCR_T$ (Gauge, GEV) | ERA5 reference | ERA5 recent | $PCR_T$ (ERA5, GEV) |
|-----------------------|-----------------|--------------|----------------------|----------------|-------------|---------------------|
| 5                     | 51.8            | 55.2         | 1.07                 | 49.0           | 48.0        | 0.98                |
| 10                    | 56.6            | 67.4         | 1.19                 | 54.1           | 53.9        | 1.00                |
| 50                    | 63.8            | 100.5        | 1.57                 | 62.6           | 65.6        | 1.05                |
| 100                   | 65.9            | 117.6        | 1.78                 | 65.3           | 70.1        | 1.07                |

**Table 5.** Bootstrap CI for return levels (GEV and Gumbel).

| T (years) | Gauge median (Gumbel) | Gauge $CI_{95}$ (Gumbel) | ERA5 median (Gumbel) | ERA5 $CI_{95}$ (Gumbel) |
|-----------|-----------------------|--------------------------|----------------------|-------------------------|
| 5         | 54.1 (−5.0/+5.1)      | 48.1 (−2.9/+2.7)         | 55.4 (−5.9/+5.3)     | 47.5 (−3.3/+2.9)        |
| 10        | 63.4 (−7.4/+7.0)      | 53.2 (−4.2/+3.9)         | 65.1 (−7.9/+7.6)     | 53.9 (−4.4/+4.1)        |
| 50        | 85.4 (−16.3/+15.7)    | 62.0 (−8.2/+9.0)         | 86.4 (−13.0/+12.0)   | 68.3 (−7.6/+6.2)        |
| 100       | 95.0 (−22.9/+21.4)    | 65.0 (−10.1/+11.6)       | 95.4 (−15.0/+13.7)   | 74.3 (−8.8/+7.1)        |

Under GEV (Table 4), the gauge shows stronger amplification toward rarer events, with  $PCR_T$  reaching 1.57 at  $T = 50$  years and 1.78 at  $T = 100$  years, whereas ERA5-Land exhibits only modest increases at longer return periods and values close to 1.00 at shorter return periods, indicating that the product does not reproduce the magnitude of upper-tail amplification implied by the gauge; this divergence is important for design-oriented screening since local extremes controlled by localized or mixed precipitation mechanisms may produce stronger upper-tail changes than those represented by the ERA5-Land grid cell.

The divergence between the gauge-based and ERA5-Land split-period signals should also be interpreted in light of possible regime dependence; if the recent-period intensification in the gauge record is partly driven by localized convective or coastal–orographic events, ERA5-Land may reproduce only a muted signal because the grid-cell representation smooths point-scale maxima and may not fully capture shifts in the relative contribution of the generating regimes.

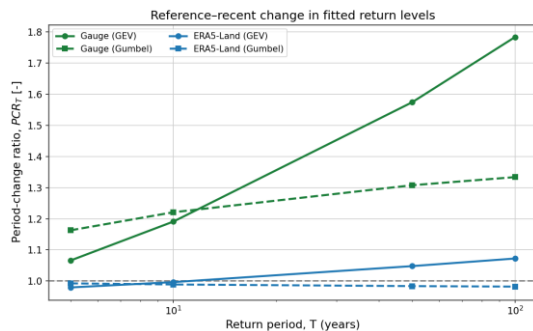
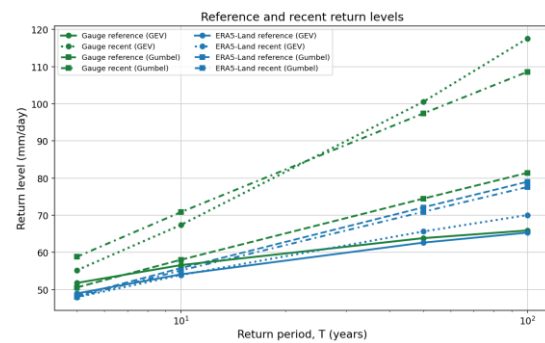
**Fig. 9.** Reference-to-recent percentage change in return levels as a function of  $T$ .

Fig. 9 highlights the strong return-period dependence of change for the gauge (especially under GEV) versus the comparatively flat response in ERA5-Land.

Fig. 10 provides the corresponding slice-based return-level curves and demonstrates that the strongest divergence emerges in the high-return-period range where design and risk screening are most sensitive; these results strengthen the conclusion that uncorrected gridded extremes can diverge materially from gauge-based design inference regarding both baseline magnitude and inferred change between periods.

**Fig. 10.** Reference and recent return-level curves.

## 5.6 Uncertainty quantification

Sampling uncertainty in design return levels was quantified through non-parametric bootstrap resampling of the annual maxima series (500 resamples) by propagating variability through the full fitting and return-level estimation workflow under both GEV and Gumbel; the resulting medians and 95% confidence intervals are reported in Table 5.

Under GEV, uncertainty widens markedly with return period and is slightly asymmetric, reflecting the sensitivity of upper-tail inference under finite samples, whereas intervals remain tighter under Gumbel due to the fixed-tail constraint. Across return periods, ERA5-Land medians remain below gauge medians and the separation increases toward rarer events, indicating that the discrepancy is systematic rather than explained by sampling variability alone. Fig. 11 displays the corresponding uncertainty bands around median return-level curves, demonstrating that even with tighter Gumbel bands the upper-tail separation persists, thereby supporting return-period-specific calibration for engineering use. The results of the uncertainty analysis also bound the interpretation of long-return-period extrapolation and of the reference–recent split, which reduces sample sizes within each period and can increase sensitivity, particularly for GEV estimates at long return periods; this motivates a careful interpretation of period-based differences, especially at longer return periods.

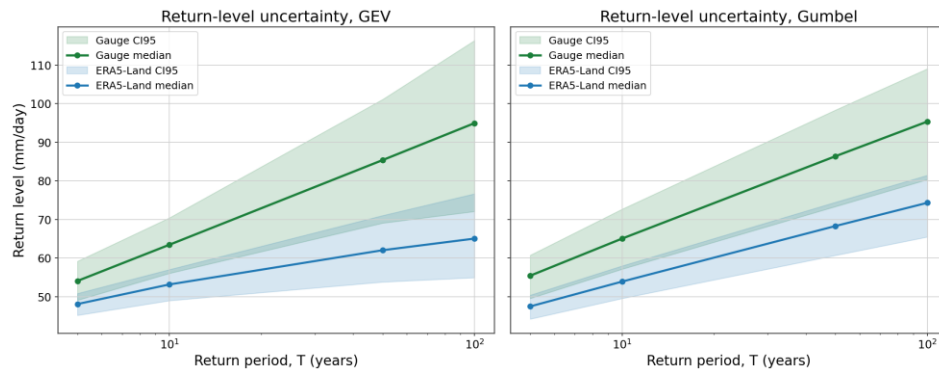


**Fig. 11.** Return levels with 95% bootstrap uncertainty bands (GEV and Gumbel).

The transferability of correction factors depends on the stability of the point-to-grid relationship, which may vary with convection, orography, coastal gradients, elevation mismatch and the seasonality of event-generating processes, and although the present analysis does not quantify date-level event matching, seasonality of annual maxima or biases across low, moderate and high daily precipitation classes and annual accumulations, these diagnostics would support future extension of the correction-factor approach across space and time.

## 6 CONCLUSIONS

A cloud-based, reproducible workflow for annual-maxima extreme-value analysis in GEE was developed and demonstrated by fitting both Gumbel and GEV models to estimate design rainfall return levels from gridded daily precipitation archives; the workflow was applied at two complementary scales comprising global mapping as a first-order screening exercise and local evaluation and calibration against a long in-situ record in Thessaloniki, Greece, where convective organization and coastal–orographic effects can amplify point-to-grid representativeness gaps. The contribution is primarily diagnostic as it evaluates how ERA5-Land represents upper-tail design rainfall at a Mediterranean coastal site and illustrates how local point–grid differences can be quantified for design-oriented screening. Key findings are:

- Global design screening from gridded archives is operationally feasible in GEE. Automated AMS extraction and EVT fitting generate physically interpretable global return-level surfaces for  $T = 5$ –100 years and provide a transparent pathway to derive screening-level design-relevant extremes from widely available gridded products.
- Period-stratified fitting provides a pragmatic stress test for temporal change in gridded extremes. Separate estimation for reference (1955–1980) and recent (1981–2022) periods and mapping recent-to-reference ratios yields a spatially heterogeneous screening layer indicating where gridded archives imply intensification or weakening. At Thessaloniki, this comparison indicates substantial gauge-inferred intensification at longer return periods, including an approximately 33% increase in the gauge-based 100-year Gumbel return level, which is weakly reproduced by ERA5-Land; this provides a basis for targeted local follow-up where design-relevant changes are indicated.
- At Thessaloniki, ERA5-Land systematically underestimates gauge-consistent design return levels, with tail divergence increasing for rarer events. Across the design range, ERA5-Land return levels are below

gauge-based estimates, with the gap widening toward long return periods and being more pronounced under GEV, indicating damping of the upper tail that governs design even when large-scale precipitation fields are climatologically coherent.

- Return-period-specific correction factors provide locally calibrated adjustment factors for design-oriented interpretation. Multiplicative gauge-to-ERA5-Land factors increase with return period, spanning approximately 1.17–1.29 under Gumbel and 1.13–1.46 under GEV for  $T = 5$ –100 years. This shows that a single constant correction is inadequate and their wider use would require validation across additional gauges.

Beyond the Thessaloniki case study, the broader contribution is a transferable GEE-based EVT workflow that produces global return-level and period-change screening layers from gridded archives and provides a framework for site-specific calibration where gauges exist; future work should extend evaluation across additional gauges and climatic regimes, test duration dependence and sub-daily consistency with IDF practice, examine the seasonality and timing of annual maxima, evaluate biases across low, moderate and high precipitation classes and annual accumulations and apply moving-window or covariate-based non-stationary EVT formulations to determine whether reference–recent differences emerge gradually, abruptly or as a function of circulation and local thermodynamic conditions.

**Acknowledgement:** This research was financially supported by the Research Committee (ELKE) of the University of Macedonia. The authors gratefully acknowledge this support.

**Supplementary Materials.** The following link contains the Google Earth Engine script used in this study: <https://code.earthengine.google.com/a0ede94f4534b06de3a4cd10926e9e44>.

## REFERENCES

- Andria, S., Borga, M., Marani, M., 2025. Thermodynamic versus large-scale controls on extreme precipitation: Temporal scale dependence and Clausius-Clapeyron scaling redefined. *Geophysical Research Letters*, 52(21), e2025GL115204.
- Bara, M., Gaál, L., Kohnová, S., Szolgay, J., Hlavčová, K., 2010. On the use of the simple scaling of heavy rainfall in a regional estimation of IDF curves in Slovakia. *J. Hydrol. Hydromech.*, 58, 49–63.
- Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by

- merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, 21, 589–615.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 208 p.
- de Leeuw, J., Methven, J., Blackburn, M., 2015. Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations. *Q. J. R. Meteorol. Soc.*, 141, 798–806.
- Dayan, U., Nissen, K., Ulbrich, U., 2015. Review Article: Atmospheric conditions inducing extreme precipitation over the eastern and western Mediterranean. *Nat. Hazards Earth Syst. Sci.*, 15, 2525–2544.
- DeVries, B., Huang, C., Armston, J., Huang, W., Jones, J.W., Lang, M.W., 2020. Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. *Remote Sens. Environ.*, 240, 111664.
- Donat, M.G., Sillmann, J., Wild, S., Alexander, L.V., Lippmann, T., Zwiers, F.W., 2014. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, 27, 5019–5035.
- Douka, M., Karacostas, T., 2018. Statistical analyses of extreme rainfall events in Thessaloniki, Greece. *Atmos. Res.*, 208, 60–77.
- Dunn, R.J.H., Donat, M.G., Alexander, L.V., 2014. Investigating uncertainties in global gridded datasets of climate extremes. *Clim. Past*, 10, 2171–2199.
- Ferguglia, O., Palazzi, E., Arnone, E., 2024. Elevation dependent change in ERA5 precipitation and its extremes. *Clim. Dyn.*, 62, 8137–8153.
- Eggert, B., Berg, P., Haerter, J.O., Jacob, D., Moseley, C., 2015. Temporal and spatial scaling impacts on extreme precipitation. *Atmos. Chem. Phys.*, 15, 5957–5971.
- Erić, R., Kadović, R., Đurđević, V., Đukić, V., 2021. Future changes in extreme precipitation in central Serbia. *J. Hydrol. Hydromech.*, 69, 196–208.
- Gaál, L., Szolgay, J., Lapin, M., Fasko, P., 2009. Hybrid approach to delineation of homogeneous regions for regional precipitation frequency analysis. *J. Hydrol. Hydromech.*, 57, 226–249.
- Galiatsatou, P., Zafeirakou, A., Nikoletos, I., Gkatzoura, A., Kapouniari, M., Katsoulea, A., Malamataris, D., Kavouras, I., 2024. Capacity assessment of a combined sewer network under different weather conditions: Using nature-based solutions to increase resilience. *Water*, 16, 2862.
- Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., Francés, F., 2023. Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmos. Res.*, 284, 106606.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.*, 202, 18–27.
- Gumbel, E., 1962. Statistical theory of extreme values. *JORBEL – Belgian Journal of Operations Research, Statistics, and Computer Science*, 3(2), 3–11.
- Hamouda, M.E., Pasquero, C., 2021. European extreme precipitation: The effects of spatio-temporal resolution of the data. *Weather Clim. Extrem.*, 33, 100337.
- Haslinger, K., Breinl, K., Pavlin, L., Pistotnik, G., Bertola, M., Olefs, M., ..., Blöschl, G., 2025. Increasing hourly heavy rainfall in Austria reflected in flood changes. *Nature*, 639(8055), 667–672.
- Higgins, R.W., Kousky, V.E., 2013. Changes in observed daily precipitation over the United States between 1950–79 and 1980–2009. *J. Hydrometeorol.*, 14(1), 105–121.
- Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B (Methodol.)*, 52(1), 105–124.
- Iliadis, C., Galiatsatou, P., Glenis, V., Prinos, P., Kilsby, C., 2023. Urban Flood Modelling under Extreme Rainfall Conditions for Building-Level Flood Exposure Analysis. *Hydrology*, 10(8), 172.
- Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. *Adv. Water Resour.*, 25(8–12), 1287–1304.
- Kidd, C., Huffman, G., 2011. Global precipitation measurement. *Meteorol. Appl.*, 18(3), 334–353.
- Koutsoyiannis, D., 2004. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrol. Sci. J.*, 49(4), 591–610.
- Kysely, J., Gaál, L., Beranová, R., 2011. Projected changes in flood-generating precipitation extremes over the Czech Republic in high-resolution regional climate models. *J. Hydrol. Hydromech.*, 59(4), 217–227.
- Lavers, D.A., Simmons, A., Vamborg, F., Rodwell, M.J., 2022. An evaluation of ERA5 precipitation for climate monitoring. *Q. J. R. Meteorol. Soc.*, 148(748), 3152–3165.
- Maggioni, V., Meyers, P.C., Robinson, M.D., 2016. A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *J. Hydrometeorol.*, 17(4), 1101–1117.
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L.O., 2017. Towards process-informed bias correction of climate change simulations. *Nat. Clim. Change*, 7, 764–773.
- Marra, F., Dallan, E., Canale, A., Prosdoci, I., Papacharalampous, G., Borga, M., Papalexio, S.M., 2026. Apparent heavy tails of sub-daily precipitation explained by the coexistence of lighter-tailed processes. *Geophysical Research Letters*, 53(2), e2025GL119705.
- Mastrantonas, N., Herrera-Lormendez, P., Magnusson, L., Pappenberger, F., Matschullat, J., 2021. Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. *Int. J. Climatol.*, 41, 2710–2728.
- Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: Whither water management? *Science*, 319, 573–574.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... and Thépaut, J.-N., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, 13, 4349–4383.
- North, R.P., Livingstone, D.M., Hari, R.E., Köster, O., Niederhauser, P., Kipfer, R., 2013. The physical impact of the late 1980s climate regime shift on Swiss rivers and lakes. *Inland Waters*, 3, 341–350.
- O’Gorman, P.A., 2015. Precipitation extremes under climate change. *Curr. Clim. Change Rep.*, 1, 49–59.
- Ochoa-Rodriguez, S., Wang, L.P., Gires, A., Pina, R.D., Reinoso-Rondinel, R., Bruni, G., ... and ten Veldhuis, M.-C., 2015. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *J. Hydrol.*, 531, 389–407.
- Onderka, M., Pecho, J., Szolgay, J., Kohnová, S., Garaj, M., Mikulová, K., Varšová, S., Lukasová, V., Výleta, R.,

- Rutkowska, A., 2024. Applying a time-varying GEV distribution to correct bias in rainfall quantiles derived from regional climate models. *J. Hydrol. Hydromech.*, 72(4), 499–512.
- Pakalidou, N., Karacosta, P., 2018. Study of very long-period extreme precipitation records in Thessaloniki, Greece. *Atmos. Res.*, 208, 106–115.
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540, 418–422.
- Peleg, N., Wright, D.B., Fowler, H.J., Leitão, J.P., Sharma, A., Marra, F., 2024. A simple and robust approach for adapting design storms to assess climate-induced changes in flash flood hazard. *Adv. Water Resour.*, 193, 104823.
- Pfahl, S., O’Gorman, P.A., Fischer, E.M., 2017. Understanding the regional pattern of projected future changes in extreme precipitation. *Nat. Clim. Change*, 7, 423–427.
- Philandras, C.M., Nastos, P.T., Paliatsos, A.G., Repapis, C.C., 2010. Study of the rain intensity in Athens and Thessaloniki, Greece. *Adv. Geosci.*, 23, 37–45.
- Pregolato, M., Ford, A., Wilkinson, S.M., Dawson, R.J., 2017. The impact of flooding on road transport: A depth-disruption function. *Transp. Res. Part D: Transp. Environ.*, 55, 67–81.
- Reid, P. C., Hari, R. E., Beaugrand, G., Livingstone, D. M., Marty, C., Straile, D., ... and Zhu, Z., 2016. Global impacts of the 1980s regime shift. *Glob. Change Biol.*, 22(2), 682–703.
- Rivoire, P., Le Gall, P., Favre, A.-C., Naveau, P., Martius, O., 2022. High return level estimates of daily ERA-5 precipitation in Europe estimated using regionalized extreme value distributions. *Weather Clim. Extremes*, 38, 100500.
- Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J. Photogramm. Remote Sens.*, 164, 152–170.
- Touloumidis, D., Madas, M., Zeimpekis, V., Ayfantopoulou, G., 2025. Weather-Related Disruptions in Transportation and Logistics: A Systematic Literature Review and a Policy Implementation Roadmap. *Logistics*, 9(1), 32.
- Vicente-Serrano, S.M., Trambly, Y., Reig, F., González-Hidalgo, J.C., Beguería, S., Brunetti, M., ..., Potopová, V., 2025. High temporal variability not trend dominates Mediterranean precipitation. *Nature*, 639(8055), 658–666.
- Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.*, 52(3), 522–555.
- Willems, P., 2013. Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium. *J. Hydrol.*, 496, 166–177.

Received 13 March 2026

Accepted 5 June 2026