



A standardized rainfall extremity index for equitable flash-flood warning in South Korea

Misook Kwak¹, Hyunjun Ahn², and Sanghoo Yoon^{1,3}

¹Department of Statistics, Chonnam National University, Gwangju 61186, Republic of Korea

²Korea Hydro & Nuclear Power Co., Ltd., Gyeongju 38120, Republic of Korea

³G-LAMP Institute, Chonnam National University, Gwangju 61186, Republic of Korea

Correspondence: Sanghoo Yoon (statstar77@jnu.ac.kr)

Abstract. Conventional Intensity–Duration–Frequency (IDF) curves quantify rainfall depth at fixed durations. We develop their event-based dual—the Modified Inter-Amount Time (*M-IAT*), the duration for cumulative precipitation within an event to reach a critical depth—and derive a Standardized Torrential Rainfall Index (STRI) by mapping *M-IAT* to a normal-quantile scale through extreme-value theory. Hourly precipitation from 65 Automated Synoptic Observing System (ASOS) stations in South Korea (1990–2024) is analysed under stationary and non-stationary Generalized Extreme Value (GEV) and Generalized Pareto Distribution (GPD) models, with temporal dependence removed by a Ferro–Segers-verified cluster-minimum construction. Both shape parameters concentrate in the bounded reverse-Weibull domain ($\hat{\xi} < 0$); with the GEV shape Winsorized at the regularity boundary $\hat{\xi} = -1$, the GEV and GPD shapes agree closely (median $|\Delta\hat{\xi}| = 0.18$). The *magnitude* of *M-IAT* extremes is largely stationary—trends survive at only 3.1% (GEV) and 5.6% (GPD) of combinations, and only one GEV combination after multiplicity correction—whereas exceedance *frequency* trends positive at all nine critical values. This signal is not significant in the 35-year record but becomes significant across all critical values when extended to 1975–2024: intensification here manifests as more frequent, not larger, fast-accumulation events. Against precipitation-independent impact records from the national disaster yearbook, the event-level peak STRI is significantly rank-correlated with casualties (Spearman $\rho = 0.56$) and property losses ($\rho = 0.46$) over 27 events. STRI is proposed not as a hindsight competitor to precipitation totals but as a station-standardised diagnostic: its normal-quantile scale renders one threshold comparable across differing climatologies, and *M-IAT* is monitorable prospectively as rainfall accumulates. *Equity* here denotes between-station consistency of detection, not social or demographic fairness: a warning rule is equitable if, at a fixed alarm budget, its probability of capturing a damage day is independent of station wetness. At a matched damage-capture rate (recall = 0.60), recall under an absolute-depth threshold rises with wetness (Spearman $\rho = +0.32$, $p = 0.009$; the national advisory depth was met on only 53.9% of damage station–days); under a fixed STRI threshold the dependence vanishes ($\rho = 0.00$, $p = 0.99$) and the between-station spread of recall narrows from 0.18 to 0.13. Operational use requires local threshold calibration and a prospective, nowcasting-based demonstration of skill.



1 Introduction

Global climate change has altered not only mean air temperature but also precipitation patterns, driving a worldwide increase
 25 in the frequency and intensity of hydrological extremes—including short-duration torrential rainfall that triggers flash floods
 beyond the range of historical observations (Pendergrass et al., 2017; Fowler et al., 2021). Persistent increases in extreme
 rainfall have been reported across many regions, including Korea, Japan, India, and the United States (Karl and Knight, 1998;
 Krishnamurthy et al., 2009; Kamiguchi et al., 2010; Kunkel et al., 2013a; Melillo et al., 2014; Baek et al., 2017; Kim et al.,
 2025); in Korea, the number of days with torrential rainfall has increased by approximately 75% over the past 50 years (Do
 30 et al., 2023; Park and Suh, 2023). A consistent physical driver is that warming-induced increases in atmospheric water vapor
 amplify short-duration rainfall intensity at a much faster rate than total precipitation amounts (Min et al., 2011; Cheng et al.,
 2014; Stocker et al., 2013; Li et al., 2020; Fowler et al., 2021), simultaneously intensifying extreme events, raising their
 frequency, and extending their duration (Trenberth et al., 2005; Allan and Soden, 2008; Trenberth, 2011; Kunkel et al., 2013b);
 such events often exceed the physical infiltration capacity of the land surface and shorten the time of concentration, thereby
 35 triggering flash floods (Kim, 2007).

Conventional engineering practice quantifies rainfall extremes through Intensity–Duration–Frequency (IDF) curves: for a set
 of fixed durations $d \in \{1, 2, 3, 6, 12, 24\}$ h, IDF returns the rainfall depth $R(d, T)$ associated with return period T . IDF curves
 are therefore parameterised by *duration*; the depth is the random variable. A complementary, less-developed parameterisation
 fixes a critical depth and treats the duration as the random variable. This is precisely the structure of the Modified Inter-Amount
 40 Time ($M-IAT$) of Lee et al. (2021) and of related “inter-amount time” constructions (Schleiss and Smith, 2016): $M-IAT(c)$
 is the minimum time required, within a given rainfall event, for cumulative precipitation to reach a critical value c mm. We
 refer to this as the *event-based dual* to IDF curves throughout the paper.

This dual relationship, developed formally in Sec. 2.1, has two implications that motivate the present study. First, fixing
 the depth and observing the duration is operationally aligned with how flash-flood early warnings are triggered: the realised
 45 $M-IAT$ is precisely the lead time available for response, and it is observable *prospectively* as precipitation accumulates,
 whereas conventional severity proxies such as daily-maximum rainfall are known only once the event has ended. Second,
 although $M-IAT$ and IDF are statistically equivalent in information content, the EVT analysis of $M-IAT$ minima admits a
 clean reversed-Weibull (upper-bounded) interpretation that is hidden in standard upper-tail IDF analyses.

Any framework predicated on intensifying extremes must itself accommodate non-stationarity (Cheng et al., 2014; Lenderink
 50 et al., 2021): fitting a stationary GEV/GPD to time-trended extremes biases return-level estimates near the tail (Coles, 2001),
 so we embed a non-stationary extension within the methodology rather than treating it as a sensitivity check. Building on the
 $M-IAT$ formalism, the present study develops the Standardized Torrential Rainfall Index (STRI) and evaluates it against his-
 torical major torrential rainfall events. The paper contributes (i) a single normal-quantile index that replaces absolute, nationally
 uniform precipitation thresholds with a station-standardised measure of *relative* rarity, so that a common index value carries
 55 the same exceedance meaning across stations of differing climatology, critical values, and decades; (ii) two physically moti-
 vated sampling schemes—annual minima for the GEV and rainy-cluster minima for the GPD—replacing ad hoc inter-amount



duration cut-offs; (iii) a model-selection procedure combining declustered POT, stationary L-moments, and non-stationary MLE across 585 station–CV combinations; and (iv) external corroboration against a precipitation-independent impact register, which sidesteps the circularity present in most damage-validation studies of precipitation-based indices. The core contribution is thus not a new estimate of regional rainfall amount but a warning framework of *equal rarity*, under which a single operational threshold applies equitably across stations with contrasting rainfall regimes—a property demonstrated quantitatively in Sec. 4.8. *Equity* is used throughout in this between-station sense: at a fixed alarm budget, a warning rule is equitable to the extent that its per-station probability of capturing a damage day is independent of local climatology. In the natural-hazards literature the word more often carries a social or distributive meaning—the allocation of warnings or losses across population groups—which is not intended here; vulnerability- and exposure-weighted warning (Gall et al., 2009) is beyond the scope of this study. The analysis is confined to the South Korean network of Sect. 3; transferability to other climate regimes is discussed, but not demonstrated, in Sect. 5.

2 Research methodology

To quantify the hydrological risk of flash-flood-inducing torrential rainfall intensified by climate change, this study develops the Standardized Torrential Rainfall Index (STRI). The overall analytical framework consists of five primary stages: (1) data collection and construction, (2) extreme value distribution fitting and parameter estimation with explicit POT declustering, (3) standardized index calculation, (4) validation through historical damage cases, and (5) extension of (2)–(3) to a non-stationary setting.

2.1 *M-IAT* as the event-based dual of Intensity–Duration–Frequency (IDF) curves

Let $\{P(t)\}$ denote the hourly precipitation series at a given station. For an isolated rainfall event E with onset time t_0 , define the cumulative precipitation $S_E(\tau) = \int_{t_0}^{t_0+\tau} P(s) ds$. Conventional IDF curves operate on fixed durations: for d given, IDF returns the random variable $S_E(d)$. The *M-IAT* formalism of Schleiss and Smith (2016); Lee et al. (2021) reverses this construction. Given a critical precipitation value (CV) c mm, the event-wise *M-IAT* is

$$M-IAT_E(c) = \min\{\tau \geq 0 : S_E(\tau) \geq c\}. \quad (1)$$

Equation (1) is the natural event-based dual of an IDF curve: the depth axis is fixed, the duration axis is random. Smaller $M-IAT_E(c)$ values correspond to more intense, more concentrated rainfall and—other things equal—to larger flash-flood potential. The event-level quantity $M-IAT_E(c)$ is reduced to two extreme-value sampling schemes consistent with the EVT model used:

Annual minima (GEV model). For each calendar year y , the annual minimum across all events that reach depth c is

$$M-IAT_{\min}^{\text{GEV}}(c; y) = \min_{E \in \mathcal{E}_y} M-IAT_E(c), \quad (2)$$



where $\mathcal{E}_y$ is the set of events in year y . The annual block is the canonical block-maxima sampling unit (Coles, 2001); no sub-annual time window is required.

Rainy-cluster minima (GPD model). A rainy cluster $\mathcal{K}_k$ is defined as a maximal run of consecutive rainy days at the station (i.e. days with non-zero precipitation, separated by at least one dry day from neighbouring clusters), and its representative value is

$$M-IAT_{\min}^{\text{GPD}}(c; k) = \min_{E \in \mathcal{K}_k} M-IAT_E(c). \quad (3)$$

Taking the cluster minimum addresses the within-storm dependence that would otherwise inflate the apparent number of POT exceedances. Below, $M-IAT_{\min}$ denotes either sample, with the model made explicit where it matters.

This event-based approach ensures statistical stability by resolving the initial-condition dependence found in conventional methods. For analytical clarity, we distinguish between two key thresholds: (1) the Critical Rainfall Value (CV), the prescribed precipitation depth (mm) for calculating $M-IAT$, and (2) the GPD threshold (u), the statistical cutoff for fitting the extreme-value distribution.

The $M-IAT$ -based formulation exchanges the role of d and intensity. For a fixed critical depth c and any event E , let $R_d^E = \max_t S_E([t, t+d])$ denote the maximum d -hour cumulative rainfall within event E (the quantity targeted by conventional IDF analysis (Madsen et al., 1997; Koutsoyiannis et al., 1998)). Because $S_E(\tau)$ is non-decreasing, the events $\{R_d^E \geq c\}$ and $\{M-IAT_E(c) \leq d\}$ are identical, giving the deterministic identity

$$\mathbb{P}(M-IAT_E(c) \leq d) = \mathbb{P}(R_d^E \geq c), \quad (4)$$

from which the inverse relation between the IDF design depth $R_d(T)$ and the return-level $M-IAT$ follows: at the same return period T , the curve $\{R_d(T)\}_d$ in the depth–duration plane and the curve $\{M-IAT_{\min}(c; T)\}_c$ in the duration–depth plane are reflections of each other about the diagonal of equal exceedance probability. The two curves carry equivalent statistical content; they differ in which axis is treated as the random variable and therefore in which parameterisation the extreme-value model is fitted along.

This duality has immediate operational implications. The intensity axis is recovered as a derived quantity, $i_E(c) = c/M-IAT_E(c)$, and frequency analysis on the dual side operates on the physically bounded support $M-IAT_E(c) \geq 0$, so the natural maximum domain of attraction for the sign-reversed $X = C - M-IAT_{\min}$ is the bounded reverse-Weibull family ($\xi < 0$)—an interpretation hidden in standard upper-tail IDF analyses. Operationally, an IDF curve answers “what depth might fall in d hours, once every T years?”, whereas the dual answers “how rapidly might a critical depth c be reached?”—the lead time available to a flash-flood early-warning system. The Standardized Torrential Rainfall Index (STRI) of Sec. 2.5 converts this duration-based frequency to a common normal-quantile scale comparable across stations and CVs; Table 1 situates STRI among the related rainfall indices.



Table 1. Positioning of STRI among rainfall indices. Only the IDF curve is an exact dual (Eq. 4); the depth-based indices relate to STRI only empirically (Spearman ρ vs. event-peak GEV-STRI, $n = 27$; Fig. 12).

Index	Fixed $\rightarrow$ random variable	Relation to STRI
Hourly rainfall (P_{1h})	1-h window $\rightarrow$ depth	short-duration intensity; $\rho = 0.85$
Daily maximum ($P_{24, Rx1day}$)	24-h window $\rightarrow$ depth	daily volume; $\rho = 0.40$
Annual maximum rainfall	fixed duration, yearly $\rightarrow$ depth	climatological magnitude
IDF curve	duration $\rightarrow$ depth(T)	exact dual of $M-IAT$
STRI (this study)	critical depth $\rightarrow$ standardised time	—

2.2 Extreme value theory (EVT) framework

The annual minimum (2) and the rainy-cluster minimum (3) of $M-IAT$ are both rare extreme quantities, so probability estimation in the tail is unreliable for general distributions. We use the Generalized Extreme Value (GEV) distribution as a Block-Maxima model on the annual minima, and the Generalized Pareto Distribution (GPD) as a Peak-Over-Threshold (POT) model on the rainy-cluster minima (Coles, 2001). The GEV CDF is

$$G(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (5)$$

with location μ , scale $\sigma > 0$, and shape ξ . The GPD CDF is

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma} \right)^{-1/\xi}, \quad (6)$$

where $y = x - u$ is the exceedance above threshold u , $\sigma > 0$, and ξ is the shape parameter.

Because the GEV sample is built from annual minima of $M-IAT_E(c)$, no sub-annual time window or event-duration ceiling is required: only events that actually reach the critical depth c contribute to the minimum, and an annual block automatically excludes drought-dominated, drizzle-built, or otherwise non-extremal accumulation from the upper tail (Coles, 2001). Likewise, the GPD sample is built from minima over rainy-day clusters (3), so the cluster definition itself plays the role of a physical pre-filter for storm coherence (Restrepo-Posada and Eagleson, 1982; Dunkerley, 2008). To convert the minimum-value problem into a maximum-value problem, we apply the sign-reversal transformation $X = C - M-IAT_{\min}$ (Coles, 2001), with C an arbitrary constant exceeding the maximum observed $M-IAT_{\min}$. Under EVT this transformation is shape- and scale-invariant.

The GPD model additionally requires explicit treatment of temporal dependence in the POT exceedance series, applied as a three-stage procedure. *Stage 1*: the daily-rainfall record is segmented into rainy-day clusters $\{K_k\}$, each a maximal run of consecutive wet days bounded by at least one dry day, isolating meteorologically coherent storms (Restrepo-Posada and Dunkerley, 2008). *Stage 2*: each cluster is reduced to its representative minimum $M-IAT_E(c)$ as in (3), eliminating within-storm autocorrelation by construction. *Stage 3*: residual dependence between weather systems separated by short dry intervals is removed by the runs declustering of Ferro and Segers (2003), whose extremal index $\theta \in (0, 1]$ is estimated



from the rainy-cluster minimum sample as

$$\hat{\theta} = \frac{2 \left(\sum_{i=1}^{N-1} (T_i - 1) \right)^2}{(N-1) \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2)}, \quad (7)$$

140 where T_i are the inter-exceedance times above threshold u and N is the number of exceedances; the cluster maxima of the transformed series form the declustered POT sample on which the GPD is refitted. The empirical outcomes of this procedure—dominant decorrelation at Stages 1–2, extremal indices near unity at Stage 3 (so that $\hat{\theta}$ enters the return-level computation only through the annual cluster-exceedance rate $\lambda = \hat{\theta} n_{\text{excess}}/N_{\text{years}}$), and close GEV–GPD shape agreement—are reported in Sec. 4.3.

145 2.3 Parameter estimation and return-level calculation

For stationary parameter estimation we adopt L-moments, which are robust for extreme data sensitive to small sample sizes and outliers (Hosking, 1990) and remain valid across the bounded reverse-Weibull domain—including $\hat{\xi} \leq -1/2$, where the maximum-likelihood estimator becomes non-regular (Smith, 1985)—because they are linear combinations of order statistics rather than likelihood derivatives (Hosking and Wallis, 1997). The GEV shape is Winsorized at the regularity boundary $\hat{\xi} = -1$
 150 (re-fitting $\mu = \lambda_1$, $\sigma = 2\lambda_2$ there), so that the GEV is evaluated on the same shape domain as the GPD L-moments estimator ($\xi > -1$) and within the region where the GEV maximum likelihood exists; the empirical justification of this regularization is given in Sec. 4.1. Non-stationary models, for which L-moments are not defined, are estimated by maximum likelihood (Sec. 2.4); the declustered GPD uses L-moments analogously. Station–CV combinations in which the L-moments fit degenerates (block minima fewer than five or $\ell_2 \leq 0$) are flagged as non-regular and the GEV-STRI is reported as Not Available there;
 155 the GPD-STRI from the POT model (Davison and Smith, 1990) is unaffected and reported independently. The GEV-STRI and GPD-STRI are not numerically interchangeable at individual station–CV combinations: the Pickands–Balkema–de Haan theorem (Balkema and de Haan, 1974; Pickands, 1975) guarantees asymptotic equality of the shape parameters, not of the standardised normal quantiles, so the two indices are reported in parallel as distinct EVT readings of the same observation. The r -th order probability-weighted moment β_r is defined as

$$160 \quad \beta_r = E[X \{F(X)\}^r]. \quad (8)$$

To quantify statistical uncertainty, we perform parametric bootstrap with 1 000 simulated samples. Goodness-of-fit is assessed by the Cramér–von Mises (CvM) statistic

$$w^2 = \sum_{i=1}^n \left\{ F(x_{(i)}) - \frac{i-0.5}{n} \right\}^2 + \frac{1}{12n}. \quad (9)$$

Return levels follow standard EVT closed forms:

$$165 \quad z_T^{\text{GEV}} = \mu + \frac{\sigma}{\xi} \left[y_T^\xi - 1 \right], \quad y_T = -\log(1 - 1/T), \quad (10)$$

$$z_T^{\text{GPD}} = u + \frac{\sigma}{\xi} \left[(T \zeta_u)^\xi - 1 \right], \quad (11)$$



where ζ_u is the annual mean exceedance rate of u (after declustering, ζ_u is the annual mean rate of *cluster maxima* rather than raw exceedances).

2.4 Non-stationary extension

170 The stationary models of Sec. 2.3 assume time-invariant distribution parameters; coherence with the climate-change premise of Sec. 1 requires that the EVT framework accommodate potential temporal trends. Following Coles (2001) and Cheng et al. (2014), we let the location parameter of the GEV and the scale parameter of the GPD evolve linearly with time,

$$\mu(t) = \mu_0 + \mu_1 t, \quad \log \sigma(t) = \log \sigma_0 + \sigma_1 t, \quad \xi(t) = \xi, \quad (12)$$

$$\sigma_u(t) = \exp[\beta_0 + \beta_1 t], \quad \xi_u(t) = \xi_u, \quad (13)$$

175 where t is calendar year centred on the record midpoint (2007) for numerical stability (El Adlouni et al., 2007). Parameters are estimated by maximum likelihood with parametric bootstrap, and $H_0 : \mu_1 = 0$ (GEV) and $H_0 : \beta_1 = 0$ (GPD) are tested by likelihood-ratio tests (LRT) at the 5% level. The non-stationary fit is attempted only where the inter-annual standard deviation of the block extremes exceeds 0.2 h, below which a linear trend is not identifiable; the regularity constraint $\hat{\xi} \in (-1.2, 1)$ is applied to the non-stationary shape, and bootstrap standard errors use a robust median-absolute-deviation scale after discarding
 180 divergent re-fits. For every station–CV combination the stationary L-moments fit (Sec. 2.3) is obtained, and the non-stationary MLE fit is obtained independently wherever the trend is identifiable; the LRT then arbitrates between the two nested models. The non-stationary model is adopted only when it converges, its shape is regular, *and* the LRT rejects H_0 ; in all other cases—including non-convergence or a non-regular shape of the non-stationary fit—the stationary fit is retained (Cheng et al., 2014). The LRT p-values are further screened by the Benjamini–Hochberg false-discovery-rate procedure (Benjamini and Hochberg,
 185 1995) at $q = 0.05$ as an inferential diagnostic, separate from the adoption rule (Sec. 4.4). The non-stationary STRI is then computed from the time-conditional CDF $F(x | t, \hat{\theta}(t))$.

2.5 Calculation and application of STRI

The Standardized Torrential Rainfall Index transforms the EVT-modelled exceedance probability into a standard-normal z -score:

$$190 \text{ STRI} = Z = \Phi^{-1}\left(F(x | \hat{\theta})\right). \quad (14)$$

Under the non-stationary models (12)–(13), $F(\cdot)$ is replaced by the time-conditional CDF $F(x | t, \hat{\theta}(t))$; the resulting STRI(t) is the percentile of the event *within the climate of year t* , providing a fair comparison across decades.

Because the reference distribution F is fitted to the extreme-value population rather than to the full rainfall record, STRI=0 marks the *median* of a station’s extreme distribution—an event of average severity *among extremes*, and hence already
 195 damaging. The index thus ranks severity *within* an already-hazardous population rather than classifying hazard versus no hazard: a negative value denotes an event milder than the median extreme yet still within the damage-capable regime (see also Sec. 5).



Because $F(\cdot | \hat{\theta})$ is each station's *own* fitted distribution, a given STRI value encodes the same exceedance probability irrespective of absolute climatology—the flash-flood analogue of what standardised drought indices provide across climates.

200 To visualise the point-scale results, Inverse Distance Weighting (IDW) interpolation is applied to the 65 ASOS stations. All inference—EVT fits, STRI values, trend tests, and validation—is performed at the station points, so the interpolation is purely illustrative; geostatistical methods such as kriging would be the route to a future uncertainty-aware mapping were interpolated values and their uncertainty part of the inference. The full analytical workflow is summarised in Figure 1.

3 Research data

205 3.1 Hourly precipitation data and study area

The spatial scope of the analysis covers the entire territory of South Korea. The analytical data consist of hourly precipitation observed at 65 Automated Synoptic Observing System (ASOS) stations operated by the Korea Meteorological Administration (KMA), covering 1990–2024 (35 years). $M-IAT$ was computed from these data. Because the observations are hourly, the sub-hour time at which the critical depth is reached is obtained by linear (constant-rate) interpolation within the hour that crosses
 210 that depth; this convention is applied identically to the block extremes used for estimation and to the evaluated events, so its effect largely cancels in the percentile-based STRI and is further muted for the most intense (short- $M-IAT$) events, whose standardised values already lie near the saturation cap.

The 65 ASOS stations were selected based on the availability of continuous observations throughout 1990–2024 (Figure 2). Among currently operating stations, Dongducheon (ID 98), which began observations after 1990, was excluded for analytical
 215 consistency. The selected 65 stations representatively reflect mountainous, urban and coastal climates of South Korea.

To bracket South Korea's torrential-rain advisory criterion (60 mm in 3 h) and a range of inland/coastal intensities, nine CV values $c \in \{10, 20, \dots, 90\}$ mm were considered. The upper bound of 90 mm is dictated by estimation stability: it is the largest depth at which all 65 stations retain an adequate annual block-minimum sample (30–35 of the 35 blocks), the sample thinning at the drier inland gauges beyond it, so that larger critical values would compromise the GEV/GPD fits over part of the network.
 220 Because only hourly totals are recorded, the exact within-hour time at which the critical depth is crossed is recovered under the constant-rate interpolation described above; the residual timing uncertainty is bounded by the one-hour gauge interval and is largest in relative terms at low CV, where $M-IAT$ is itself of order one hour.

3.2 Sampling schemes for the GEV and GPD models

The two sampling schemes defined in Sec. 2.1 are applied as follows. The GEV model uses the annual minima (2), yielding
 225 35 block minima per station–CV combination over 1990–2024; no sub-annual window is imposed, the annual block being the canonical block-minima unit (Madsen et al., 1997; Coles, 2001). The GPD model uses the rainy-cluster minima (3): POT exceedances above the threshold u are extracted from the cluster-minimum series, and residual inter-storm dependence is removed by the Ferro–Segers declustering of Sec. 2.2, which provides the asymptotic-independence condition required

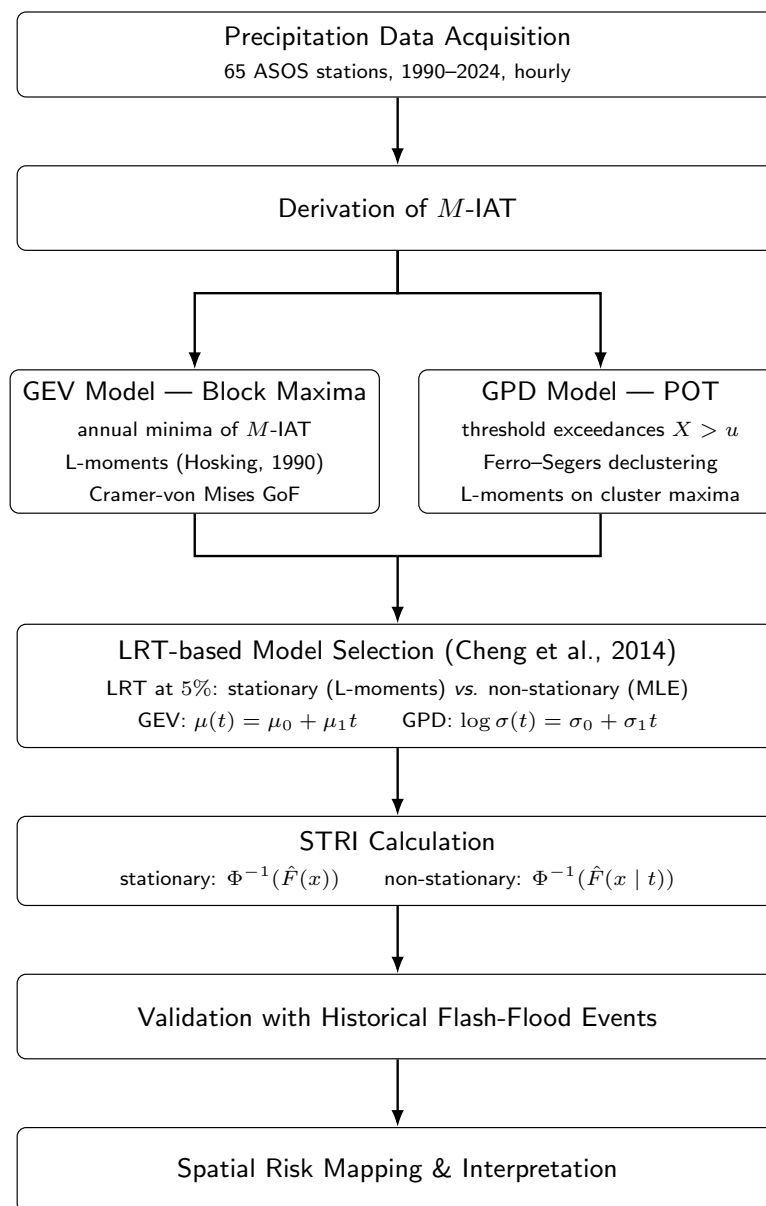


Figure 1. Research flowchart of the STRI methodology, showing the Ferro–Segers declustering step for the GPD model and the parallel stationary/non-stationary estimation paths.

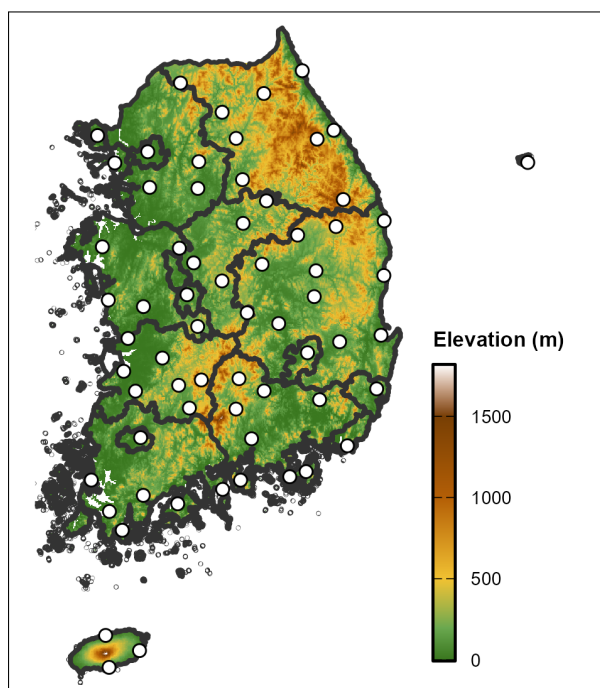


Figure 2. The 65 ASOS stations used in the study. Station identifiers and coordinates are listed in Table S1.

by the Pickands–Balkema–de Haan theorem (Coles, 2001, Ch. 3). The GEV scheme characterises the most extreme rainfall
 230 concentration of a typical year; the GPD scheme characterises the tail of within-storm concentration above an operational
 threshold. All results in Sec. 4 are produced under this two-model design.

The analysis is reported under both the stationary setting—the operative baseline, retained by the data-driven model selection
 at the great majority of combinations (Sec. 4.4)—and the non-stationary extension of Sec. 2.4, which ensures coherence with
 the climate-change premise wherever a trend is supported.

235 As anticipated in Sec. 2.2, inference is invariant to the sign-reversal constant C : comparing $C \in \{500, 1000, 2000\}$, μ shifted
 linearly with C while σ and ξ were unchanged (median relative difference below $10^{-10}\%$).

3.3 Historical torrential rainfall damage cases

For the application analysis of STRI, dates and regions associated with documented flood-damage cases over the past 35
 years were selected. A total of 36 flood damage cases—spanning typhoon and monsoon-front (Changma) heavy rainfall—
 240 were collected from the Water Disaster Status section of the Water Information Portal (<https://www.water.or.kr/>); complete
 documentation of all cases, with detailed station-level information, is provided in Supplementary Material 1.

The damage data are reported at the level of broad administrative zones rather than inundation coordinates, so stations not
 actually exposed to hydrological risk may fall inside damage zones owing to localised rainfall variability or topography. Guided
 by the national advisory level (3-h cumulative rainfall ≥ 60 mm or 12-h ≥ 110 mm), we adopt a daily-precipitation threshold of



110 mm as a practical *representativeness* criterion to identify, within each damage zone, the stations actually exposed to heavy rainfall, mitigating the spatial-aggregation bias of zone-level loss data (Gall et al., 2009). This daily threshold is an operational proxy for the sub-daily advisory; robustness to the choice (no filter; 80, 110, 150 mm) is assessed in Supplementary Material 1. In total, 886 station–damage cases were identified over the 35-year period, each with STRI computed for the nine CVs.

4 Results

This section presents stationary GEV/GPD fits, the corrected interpretation of the GEV/GPD shape comparison after declustering, the non-stationary extension, and validation against the historical damage cases. Throughout, $M-IAT_{\min}$ is referred to simply as $M-IAT$ for brevity.

4.1 GEV-based extreme value analysis

The annual minimum $M-IAT$ was treated as a maximum-value problem via $X = C - M-IAT$. L-moment estimation (Hosking, 1990) was used and 1 000 parametric bootstrap replications were applied to obtain standard errors (SE). Approximately 99.5% of all station–CV combinations satisfied the GEV assumption at the 5% significance level.

Examining the CV-specific GEV parameter densities (Figure 3), the location parameter μ progressively shifts towards lower values (in transformed units) as CV increases, reflecting the physical fact that the time required to reach a higher CV lengthens. The scale parameter σ broadens with increasing CV, indicating that station-specific variability intensifies for more extreme events.

The shape parameter $\hat{\xi}_{\text{GEV}}$ moves *progressively more negative* as CV increases. This trajectory is the expected behaviour for annual minima of $M-IAT$: minima are more extreme than threshold exceedances and therefore converge faster to the upper-bounded reverse-Weibull domain. The increasingly negative $\hat{\xi}_{\text{GEV}}$ is therefore not a pathology, but a signature that the data lie in the bounded ($\xi < 0$) max-domain of attraction, consistent with the physical bound of $M-IAT_{\min}$ at the most concentrated rainfall achievable at the station. Parameter estimates for a representative east-coast station (Gangneung, ID 105) are given in Table 2; per-station estimates for all 65 stations are provided in Tables S8 and S9. Gangneung exemplifies this trajectory: its GEV shape deepens steadily from $\hat{\xi} = -0.22$ at CV=10 mm towards the regularity boundary $\hat{\xi} = -1$ at CV=90 mm, and it is also the station that recorded the single most extreme event in the archive (Typhoon Rusa, 2002; daily rainfall 870.5 mm), so its parameters anchor the most exposed tail of the network.

Of the 585 station–CV combinations, 295 had an unconstrained L-moments shape below -1 and are reported at the Winsorized boundary $\hat{\xi} = -1$ (Sec. 2.3); these lie in the non-regular regime where the maximum likelihood is undefined and the deep-tail shape is poorly identified. The Winsorization leaves goodness-of-fit essentially intact (97.4% of the GEV fits pass the Cramér–von Mises test at the 5% level, versus 99.5% unconstrained) and the damage-day index is rank-invariant to it (Spearman $\rho = 0.99$ against the unconstrained fit; per-CV median STRI shifts by at most 0.13; Supplementary Material 1, Secs. S1.10–S1.11). Because the affected combinations are predominantly stationary-adopted, excluding them outright would



leave roughly half of the station–CV grid with no index at all; Winsorizing at the estimator-validity boundary is therefore a principled regularization that preserves a complete, interpretable index.

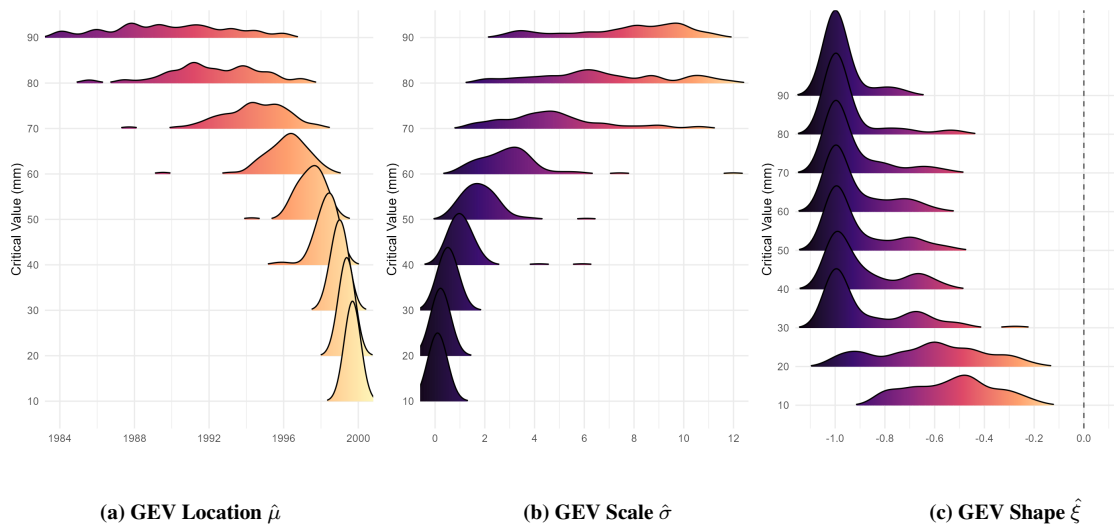


Figure 3. Distributions of the GEV parameter estimates (location, scale, shape) across the 65 stations and nine CVs, from stationary L-moments fits.

Table 2. Stationary GEV estimates for Gangneung (ID 105); SE from 1000 bootstrap replications. No shape SE is reported where the Winsorization at $\hat{\xi} = -1$ binds.

CV (mm)	$\hat{\mu}$		$\hat{\sigma}$		$\hat{\xi}$		CvM	
	estimate	SE	estimate	SE	estimate	SE	ω^2	p
10	1999.643	0.020	0.112	0.017	−0.219	0.138	0.029	0.979
20	1999.285	0.047	0.239	0.033	−0.277	0.132	0.028	0.983
30	1998.698	0.124	0.642	0.092	−0.488	0.146	0.054	0.858
40	1997.916	0.230	1.220	0.176	−0.561	0.156	0.061	0.809
50	1996.996	0.388	2.036	0.315	−0.691	0.179	0.025	0.990
60	1995.617	0.582	3.104	0.453	−0.678	0.159	0.052	0.865
70	1994.453	0.819	4.271	0.704	−0.806	0.174	0.037	0.950
80	1992.450	1.274	6.761	1.174	−0.949	0.210	0.030	0.978
90	1990.055	1.786	9.753	1.724	−1.000	—	0.037	0.951

Figure 4 shows the spatial distribution of the 50-year return-level M - IAT for CV from 30 to 90 mm: red zones below 0.5 h mark sensitive watersheds with high imperviousness, and under the most extreme CV=90 mm most basins converge to 1–3 h.
 280 Because the upper tail is bounded ($\hat{\xi} < 0$), the closed-form return level (10)–(11) reaches the physical accumulation floor in



61 (GEV) and 41 (GPD) of the 585 combinations; These cells indicate that the estimated 50-year return level lies beyond the physically meaningful range rather than representing a precisely resolved return time, and the corresponding STRI, which depends only on the exceedance probability of the observed event, is unaffected.

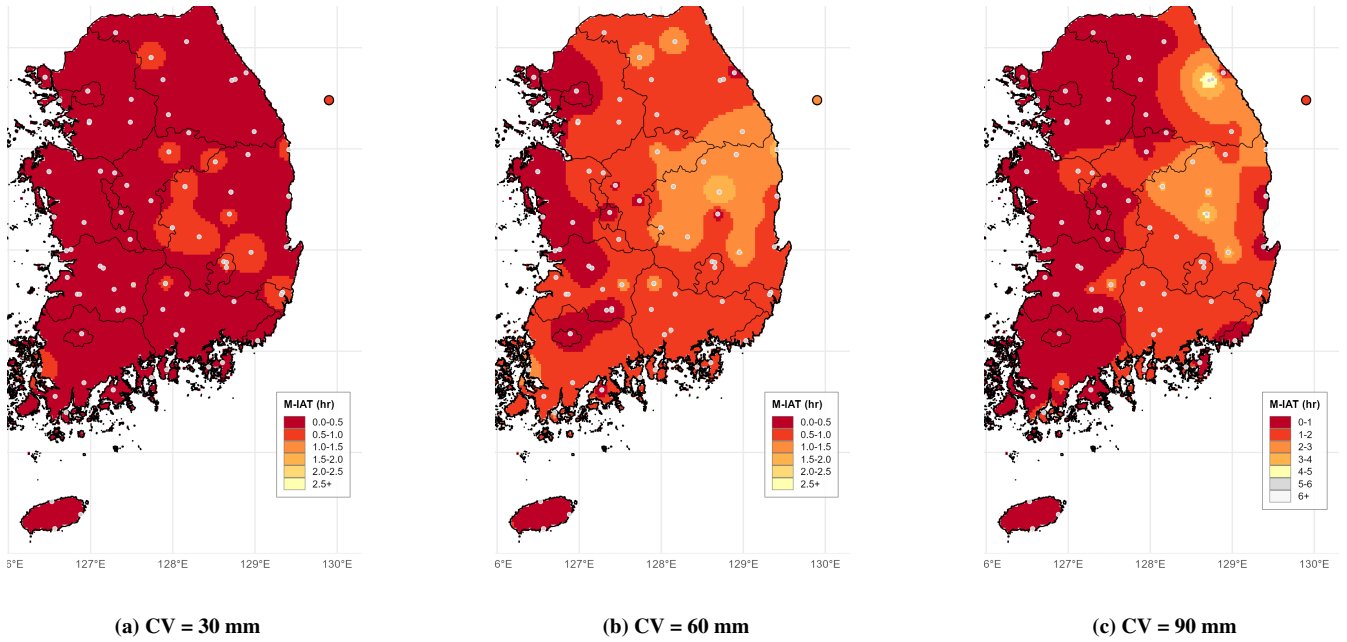


Figure 4. 50-year return-level $M-IAT$ from the stationary GEV model, at $CV = 30, 60, 90$ mm.

The 50-year return-level surface in Figure 4 can be read directly as the event-based dual of an IDF curve (Sec. 2.1). At
 285 $CV=60$ mm the 50-year return-level $M-IAT_{\min}(60; 50)$ is shortest (≤ 0.5 h) over the western and south-western lowlands—the Seoul metropolitan area, the Chungcheong region and Geum River basin, and interior Jeolla—where high imperviousness and intense frontal and typhoon rainfall drive the most rapid accumulation, and lengthens to 1–3 h over the eastern interior (upper Nakdong River basin). Where the return level approaches 3 h, it follows—via identity (4)—that the KMA heavy-rain advisory criterion (60 mm in 3 h) constitutes an approximately 30–50-year event at those locations, consistent with national
 290 rainfall-frequency analyses for Korea (Ministry of Land, Transport and Maritime Affairs, 2011; Lima et al., 2016), whereas over the western lowlands the same criterion is met considerably more often. Maps at low CV act as the duration-axis dual of high-intensity, short-duration IDF zones, and high- CV maps as the dual of low-intensity, long-duration zones; read together they trace the return-level surface without re-fitting the upper tail at each duration.

4.2 GPD-based extreme value analysis

295 The Peak-Over-Threshold methodology was applied, fitting the GPD to the transformed rainy-cluster minimum series $X = C - M-IAT_{\min}^{\text{GPD}}$ defined in (3). Within-storm autocorrelation is removed at the data-construction stage (Sec. 2.2); the Ferro–Segers extremal index applied to the cluster-minimum series returns $\hat{\theta} \approx 1$ throughout, confirming that the series is already



asymptotically independent and that $\hat{\theta}$ enters the return-level computation only through the rate factor $\lambda = \hat{\theta} n_{\text{excess}} / N_{\text{years}}$; bootstrap SE are computed conditionally on the cluster-minimum sample. The threshold u was selected as the lowest candidate
 300 quantile that passes the CvM test with at least 30 exceedances before declustering, scanning upward from the 0.90 level and relaxing to 0.85 where no candidate above 0.90 qualified; 95.4% of combinations were resolved above the 90th percentile and the remainder above the 85th. Declustering subsequently thins the retained sample, so the declustered count N_u reported in Table 3 falls below 30 at the highest critical values. Because the threshold is selected using the same CvM statistic subsequently reported as a goodness-of-fit measure, the reported p-values are conditionally optimistic, and threshold-choice uncertainty
 305 is not propagated to the return-level standard errors—a limitation shared by most automated POT procedures (Scarrott and MacDonald, 2012).

Figure 5 shows the GPD parameter densities. The threshold u concentrates in 1990–2000 across all CVs (equivalent to an $M-IAT$ threshold u_{M-IAT} of 0–10 h with $C = 2000$), the scale σ broadens with CV, and the shape $\hat{\xi}_{GPD}$ remains stable in a slightly negative band centred near -0.5 across CVs, in contrast with the more aggressively negative $\hat{\xi}_{GEV}$; its bootstrap
 310 standard error grows with CV, from 0.146 at 10 mm to 0.372 at 90 mm, the usual POT variance–threshold trade-off. After declustering this stability is preserved (Sec. 4.3), and the residual gap between $\hat{\xi}_{GPD}$ and $\hat{\xi}_{GEV}$ is consistent with the slower convergence of POT statistics to the bounded reverse-Weibull domain than block-minima statistics, as predicted by EVT (Coles, 2001, Ch. 4).

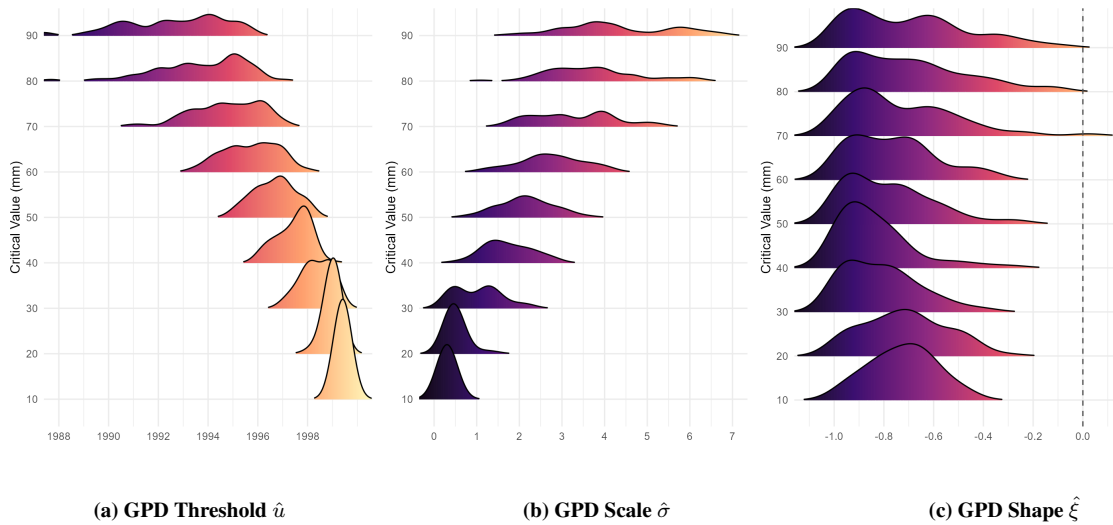


Figure 5. Distributions of the GPD threshold, scale and shape estimates across the 65 stations and nine CVs, from L-moments fits to declustered cluster maxima.

Figure 6 shows the spatial distribution of the GPD-based 50-year return-level $M-IAT$. Spatial tendencies are similar to the
 315 GEV result, but localised risk areas are more sharply identified.



Table 3. Declustered-GPD estimates for Gangneung (ID 105). N_u is the number of declustered exceedances; SE and CvM p from 1000 bootstrap replications.

CV (mm)	u	u_{M-IAT}	N_u	$\hat{\sigma}$		$\hat{\xi}$		CvM	
				estimate	SE	estimate	SE	ω^2	p
10	1999.286	0.714	166	0.355	0.044	−0.622	0.133	0.070	0.750
20	1998.965	1.035	78	0.428	0.074	−0.434	0.170	0.023	0.993
30	1997.167	2.833	98	2.055	0.338	−0.781	0.199	0.049	0.886
40	1996.191	3.809	78	2.311	0.423	−0.518	0.188	0.072	0.739
50	1996.064	3.936	48	3.373	0.791	−0.945	0.312	0.031	0.973
60	1997.167	2.833	26	1.684	0.556	−0.641	0.359	0.031	0.975
70	1996.143	3.857	27	2.764	0.894	−0.850	0.405	0.057	0.839
80	1994.715	5.285	28	3.845	1.218	−0.768	0.382	0.038	0.945
90	1992.719	7.281	28	6.639	2.277	−0.995	0.461	0.037	0.952

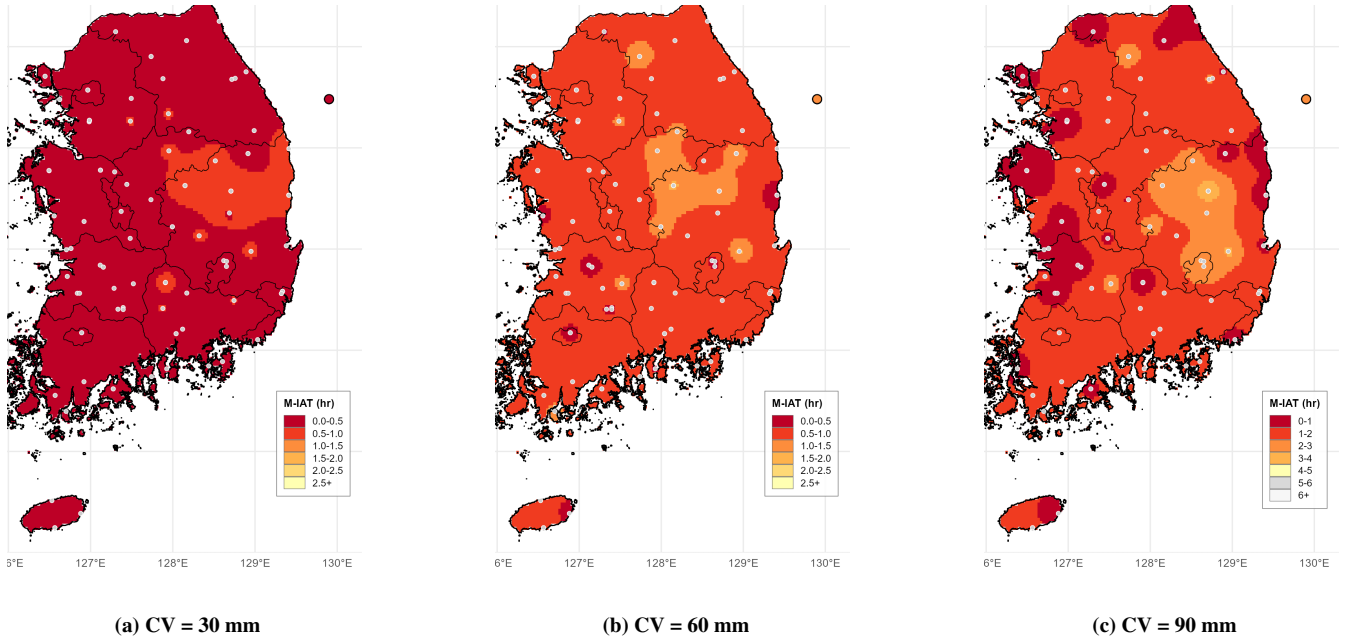


Figure 6. 50-year return-level $M-IAT$ from the declustered GPD model, at CV = 30, 60, 90 mm.

4.3 Reconciling the GEV and GPD shape parameters

EVT requires the GEV and GPD shape parameters to coincide asymptotically (Coles, 2001, Ch. 3–4); observed divergence in finite samples is attributable to (i) inappropriate threshold selection, (ii) insufficient declustering of POT exceedances, or (iii)



the slower convergence of upper-tail POT statistics to the reverse-Weibull domain compared with block-minima statistics. The unadjusted comparison of $\hat{\xi}_{\text{GEV}}$ and $\hat{\xi}_{\text{GPD}}$ in Figures 3 and 5 therefore conflates a finite-sample artefact with substantive methodological information. We test cause (ii) directly: the Ferro–Segers extremal index of the exceedances is close to unity (median $\hat{\theta} \approx 1.0$; below 0.8 in only 1.2% of station–CV combinations), so the POT are already near-independent and autocorrelation is not a material source of the discrepancy. Declustering is nevertheless applied to guarantee the independence assumption.

We re-fit the GPD on the declustered sample of Section 2.2 and compare the resulting shape parameters with the GEV shape over the common set of station–CV combinations. Across the 583 combinations for which both converged, the median absolute gap $|\hat{\xi}_{\text{GEV}} - \hat{\xi}_{\text{GPD}}|$ is 0.18 ($|\Delta\hat{\xi}| < 0.5$ in 91%, < 0.3 in 73% of combinations), confirming close agreement of the two Winsorized-domain shape estimates; the gap is smallest at low CV and widens only modestly toward CV=90 mm. Matched at the same threshold quantile, declustering shifts the GPD shape by a median of only 0.10 and narrows the GEV–GPD gap only modestly—from a median $|\Delta\hat{\xi}|$ of 0.22 (86% of combinations within 0.5) before declustering to 0.18 (91%) after—so the close agreement is robust to the declustering choice rather than an artefact of autocorrelated exceedances or of a particular analyst decision (declustering-path sensitivity in Supplementary Material 1, Sec. S1.5).

The residual gap reflects (i) the slower convergence of POT statistics to the reverse-Weibull domain (Davison and Smith, 1990; Coles, 2001, Ch. 4); (ii) finite-sample bias of the L-moments estimator at small declustered samples (Hosking and Wallis, 1997); and (iii) the Winsorization of the stationary GEV shape at $\hat{\xi} = -1$, which pins the GEV at the regularity boundary for the deeply-bounded high-CV combinations while the declustered-GPD shape floats above -1 , accounting for the modest widening of the gap toward CV=90 mm. Of these, only (iii) involves the Winsorization, and it *narrows*—rather than creates—the discrepancy by pulling the bounded GEV shape estimates toward the GPD estimates at high CV; the residual disagreement is therefore an intrinsic consequence of the block-minima versus POT sampling schemes, not an artefact of the shape regularization. Figure 7 confirms this: panel (a) pairs the two estimates at each station–CV combination (90.7% within $|\Delta\hat{\xi}| < 0.5$ of the 1:1 line; median $|\Delta\hat{\xi}| = 0.18$), and panel (b) shows both shapes remaining in the bounded reverse-Weibull domain ($\hat{\xi} < 0$) at every CV, with the GEV median pinned at $\hat{\xi} = -1$ for $\text{CV} \geq 30$ mm and the GPD median near -0.75 .

4.4 Non-stationary GEV/GPD analysis

We refit the GEV with a linear-in-time location (12) at each station and CV, and refit the declustered GPD with a linear-in-time log-scale (13). Likelihood-ratio tests of $H_0 : \mu_1 = 0$ (GEV) and $H_0 : \beta_1 = 0$ (GPD) are performed at the 5% level (Wilks, 1938; Coles, 2001). Following Cheng et al. (2014), combinations in which the non-stationary estimator failed to converge, returned a non-regular shape ($\hat{\xi} \leq -1.2$), or was not retained by the LRT are assigned the stationary model as a fallback, preserving a unified data-driven mixed-model framework across all 65 stations and nine CVs. The stationary model uses L-moments throughout (Sec. 2.3). The regularity filter $\hat{\xi} \in (-1.2, 1)$ is still applied to the non-stationary MLE model, where it is required for asymptotic identifiability.

For the GPD, the non-stationary fit converged at $269/585 = 46.0\%$ of combinations; among these, the non-stationary model was retained by the LRT in 33 combinations (5.6% of all combinations), with the retention rate decreasing monotonically in CV from 21.5% at CV=20 mm to 0% at CV ≥ 60 mm (Table 4). At CV=90 mm, two stations (Jeju, ID 184; Namwon,

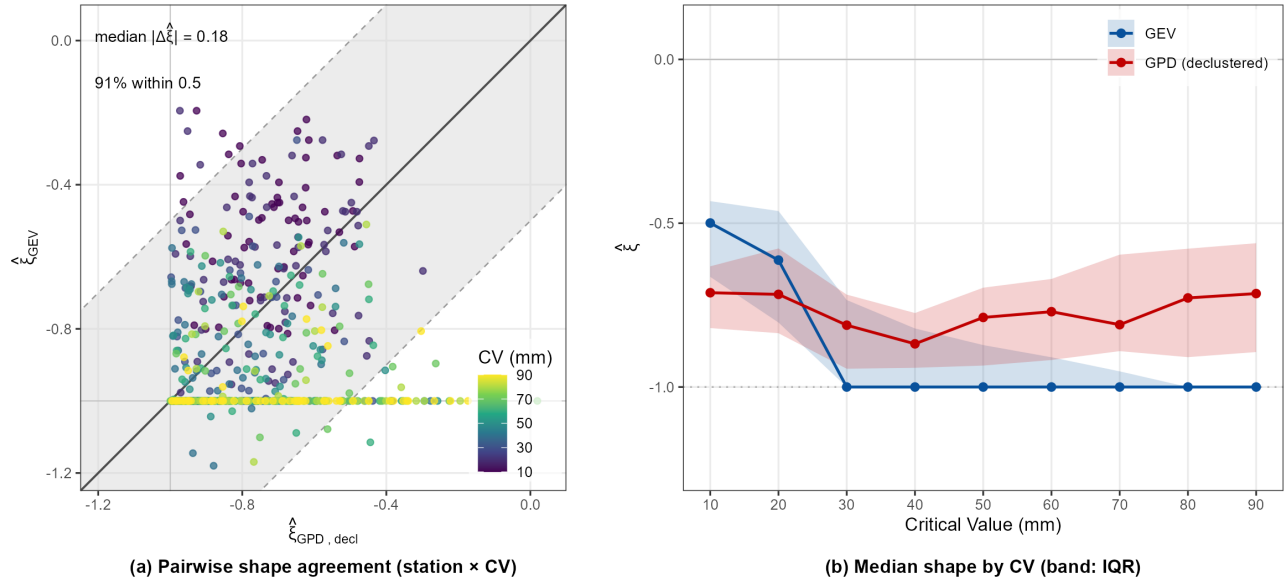


Figure 7. GEV versus declustered-GPD shape over the 65 stations and nine CVs. (a) Pairwise estimates coloured by CV; line is the 1:1 identity, band $|\Delta\hat{\xi}| < 0.5$. (b) Median shape by CV (band: IQR).

ID 247) yield no regular declustered-GPD fit, reducing the adopted total to 583/585; adoption rates are nevertheless expressed over all 585 attempted combinations, so that these two count as non-adoptions rather than being dropped. The GEV adoption sequence is analogous: the non-stationary MLE is attempted only where the inter-annual standard deviation of block minima exceeds 0.2 h and is adopted when its shape passes the regularity filter *and* the LRT rejects $H_0 : \mu_1 = 0$ at the 5% level; the non-stationary GEV is retained at 18 of the 585 combinations (3.1%; per-CV breakdown in Table S4), the remainder adopting the stationary L-moments fit, so the GEV-STRI is available at every station–CV combination. The lower filter value -1.2 is a noise margin below the maximum-likelihood existence boundary $\hat{\xi} = -1$ (Smith, 1985), so that combinations whose non-stationary shape dips just below -1 through sampling noise are arbitrated by the LRT rather than rejected outright; tightening the filter to $\hat{\xi} \in (-1, 1)$ leaves the adoption counts unchanged (Supplementary Material 4). In both families the non-stationary adoption rate decreases with CV, consistent with the “stationarity is undead” interpretation of Serinaldi and Kilsby (2015).

Overall, stationarity is not generally rejected for $M-IAT_{\min}$ -derived extremes in the Korean record (Serinaldi and Kilsby, 2015). Where non-stationarity is supported it concentrates at low CV (≤ 20 mm) and dissipates at extreme CV, although LRT power itself varies strongly with CV—the small declustered samples at high CV can hide real trends, while the larger low-CV samples inflate sensitivity. Multiplicity is the sharper caution: a Benjamini–Hochberg false-discovery-rate adjustment (Benjamini and Hochberg, 1995) leaves none of the 33 GPD rejections significant at $q = 0.05$ (smallest adjusted $p = 0.12$), and in the GEV model a single regular combination survives (Yeongdeok, ID 277, CV= 40 mm; adjusted $p = 0.002$), so individual-combination evidence for non-stationarity is essentially absent once multiplicity is accounted for (full adjusted p-



values in Supplementary Material 4). Two aggregate observations nevertheless argue against dismissing the retained set as pure noise: the asymmetric sign distribution (25/33 NS-GPD combinations with $\hat{\beta}_1 > 0$; two-sided binomial $p = 0.005$) and a physically meaningful typical effect size (median $|\hat{\beta}_1|$ corresponding to a $\sim 14\%$ cumulative change in σ_u over the 35-year record). Per-combination Wald statistics remain below 2 throughout, however, so per-station trend attribution should be regarded as indicative rather than confirmatory.

Accordingly, we report a single *hybrid* stationary/non-stationary GPD model with the combination-specific selection summarised in Table 4. The 1990–2024 trajectories of the time-varying parameters at four representative CVs (20, 40, 60, 80 mm) are shown in Supplementary Material 4 (Figure S1).

The two models detect complementary aspects of change—GEV adoption peaks at intermediate CV (40–60 mm), where the location of the annual extreme shifts, whereas GPD adoption concentrates at the lowest CVs, where the POT log-scale broadens—and both signs coexist at every CV, indicating spatially heterogeneous change rather than Korea-wide intensification. Of the 48 combinations in which either model adopts non-stationarity, three are detected by both models at the same CV—Sokcho (CV=30 mm), Seosan (CV=20 mm), and Yeosu (CV=50 mm)—all with agreeing positive trends and a shortening of the 50-year fastest-accumulation time (per-station ΔRL_{50} annotated in Fig. 8).

The preceding trend analysis concerns the *magnitude* parameters (location and log-scale) and holds the occurrence rate constant; it cannot detect a change in how *often* extreme fast-accumulation events occur—the quantity national assessments report as rising (Do et al., 2023; Park and Suh, 2023). We therefore counted, per station, CV, and year, the rainy-cluster events whose *M-IAT* falls below the fitted GPD threshold (the network is fixed at 65 stations throughout 1990–2024) and fitted a Poisson trend to the annual counts. The pooled national frequency trends positive at every CV (+3.5 to +7.7% per decade) and per-station trend signs are overwhelmingly positive (399/583 = 68%), but no individual trend survives multiplicity correction and the pooled trends lose significance under the substantial interannual overdispersion of a 35-year record ($\hat{\phi} \approx 9$ –19; smallest adjusted $p = 0.07$). The frequency channel thus carries a coherent positive *direction*—reconciling the reported national increase in heavy-rainfall days with the stationarity of the magnitude distribution—but has not yet emerged above the noise of a record this short (Serinaldi and Kilsby, 2015); we accordingly retain the stationary model as the operative choice (per-CV values in Supplementary Material 4).

To distinguish limited power from a genuine absence of change, the exceedance-frequency analysis was extended to the longer 1975–2024 record available at 60 of the 65 stations. With the same fixed thresholds and the same overdispersion-adjusted Poisson model, the frequency of fast-accumulation events increases significantly at *all* nine CVs (+7.7 to +10.7% per decade; $p < 0.02$ throughout)—a cumulative rise of roughly +45 to +66% over the fifty years, consistent with the national increase in heavy-rainfall days (Do et al., 2023; Park and Suh, 2023). The 35-year non-detection is therefore attributable to limited power rather than to a stationary occurrence process. The extended series were screened for non-climatic inhomogeneity using the Pettitt and Standard Normal Homogeneity tests with Benjamini–Hochberg control (Pettitt, 1979; Alexandersson, 1986; Benjamini and Hochberg, 1995); no retained gauge carries an identifiable instrumental discontinuity, and the increase is gradual and spread across critical values and stations rather than concentrated at a break date or a single gauge, so the 50-year signal is unlikely to be a non-climatic artefact (screening details in Supplementary Material 4).



Table 4. Model selection for the declustered GPD. NS: non-stationary retained by the LRT (5%); S (LRT): stationary retained by the LRT; S (NS failed): stationary adopted after non-convergence or a non-regular shape. Rates are over all 585 attempted combinations.

CV (mm)	NS	S (LRT)	S (NS failed)	Total	NS rate (%)
10	9	56	0	65	13.8
20	14	50	1	65	21.5
30	5	35	25	65	7.7
40	2	17	46	65	3.1
50	1	21	43	65	1.5
60	0	19	46	65	0.0
70	1	14	50	65	1.5
80	1	13	51	65	1.5
90	0	11	52	63	0.0
Total	33	236	314	583	5.6 (33/585)

405 To probe for a coherent spatial gradient, Figure 8 maps each station by the net sign of its non-stationary detections in each model. The GEV signal is sparse and spatially diffuse (10 stations, no regional cluster), consistent with its low adoption rate; the GPD signal is denser (20 stations), with positive-dominant stations concentrating in the mid-latitude band (35–37°N) plus Jeju, and the inland Gangwon mountain belt the most stationarity-dominated zone. Negative-trend stations are few and weak (four in the GEV, six in the GPD) and show no systematic topographic correspondence. Read with the limited-power caveat
 410 above, the strongest GPD-detected intensification signals lie on the inland mid-latitude belt and Jeju rather than on the south coast.

4.5 STRI calculation and historical damage validation

To evaluate STRI's consistency in representing actual torrential rainfall damage, the 36 documented damage cases were analysed: matching each damage date with observation stations, $7974 = 886 \times 9$ STRI values were computed under both the GEV
 415 and the declustered-GPD models for the nine CVs (complete station-wise values in Supplementary Material 2). Because the case stations are selected conditional on observed heavy rainfall (Sec. 3), this validation is, strictly, a verification on a conditioned sample (Jolliffe and Stephenson, 2012) and should be read as a consistency check rather than an unconditional skill estimate; sensitivity to the conditioning threshold is quantified in Supplementary Material 1.

In 1701 of the 7974 damage-day cases (21.3%), GPD-STRI was *Not Available*: 1673 observations fell below the GPD
 420 threshold u and 28 lay in combinations with no regular GPD fit (Sec. 4.3). The below-threshold cases are not detection failures but events the GPD legitimately classifies as non-extreme on the POT scale; their GEV-STRI is correspondingly low (median -1.06), so the two models agree in classifying these events as non-extreme where their scales overlap, while supplying complementary information in the moderate-to-extreme region.

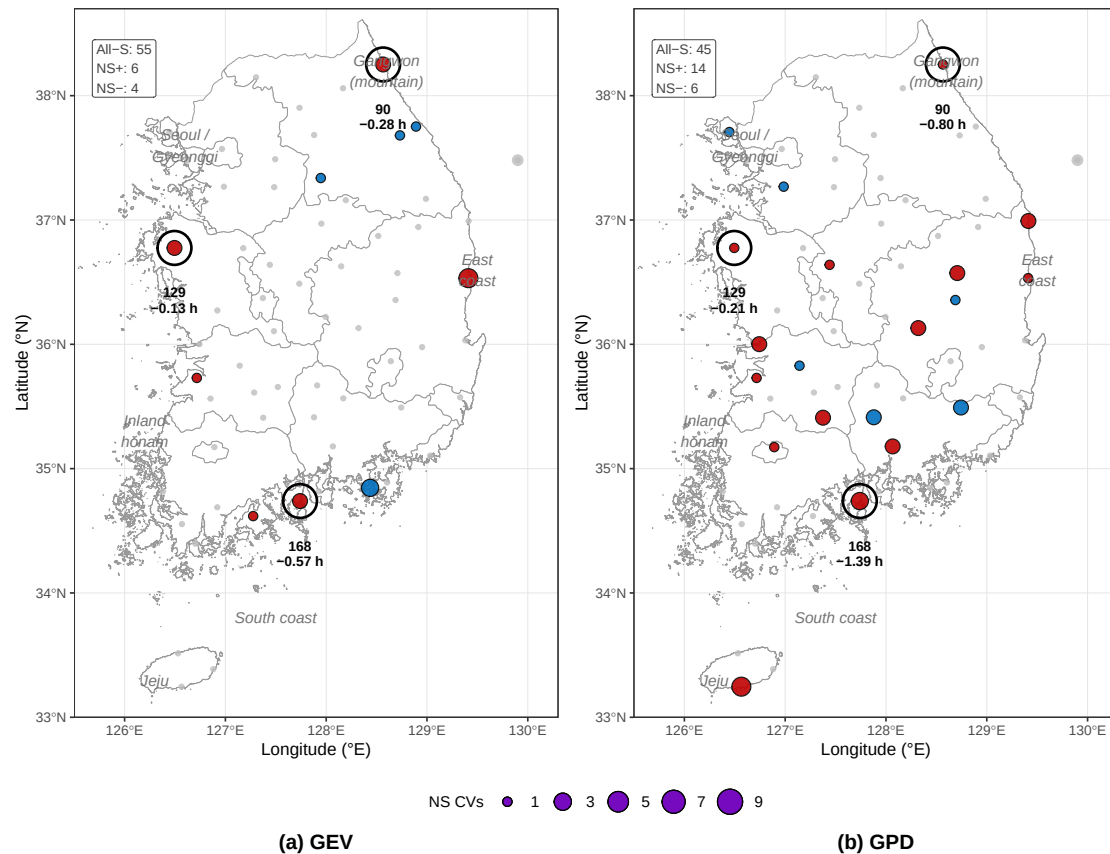


Figure 8. Stations by net sign of non-stationary adoption across the nine CVs: (a) GEV, (b) declustered GPD. Grey: stationary at all CVs; red/blue: positive/negative trends dominant; marker size scales with the number of NS-adopted CVs. Open rings mark the three stations at which both models adopt NS at the same CV, labelled with station ID and the implied 50-year ΔRL_{50} (h).

After declustering and shape regularization (Table 5; full station-wise values in Supplementary Material 2), the GPD-STRI
 425 does not saturate at the 99.99th percentile: damage-day medians lie in a moderate range (0.25–0.57 across CVs), and even
 for the most catastrophic event in the record (Typhoon Rusa, 2002; daily rainfall 870.5 mm at Gangneung) the GPD-STRI at
 the two most exposed stations attains a median of only $z \approx 2$, with only about 2% of combinations reaching the $z = 3.72$ cap.
 The paired GPD–GEV mean difference—computed over the cases for which both indices are defined, and therefore not equal
 to the difference of the two marginal means in Table 5—shifts monotonically with CV, positive (GPD more alarming) at CV
 430 ≤ 50 mm and negative at CV ≥ 60 mm, so the two models carry complementary rather than redundant operational information,
 and the GPD exceedance rate declines from 0.91 at CV=10 mm to 0.68 at CV=90 mm, reflecting the increasing rarity of events
 that reach high critical depths quickly.



Table 5. Descriptive statistics of STRI by CV. “Exceedance Rate (GPD)” is the fraction of damage-day events above the GPD threshold, i.e. with a defined GPD-STRI; “Mean Diff” is the paired GPD–GEV difference.

CV (mm)	Mean		Median		SD		Mean Diff (GPD-GEV)	Exceedance Rate (GPD)
	GPD	GEV	GPD	GEV	GPD	GEV		
10	0.59	−0.59	0.57	−0.48	1.19	1.35	0.93	0.912
20	0.34	−0.57	0.28	−0.47	1.19	1.36	0.49	0.836
30	0.29	−0.34	0.33	−0.27	1.08	1.21	0.25	0.783
40	0.30	−0.21	0.38	−0.22	0.96	1.02	0.21	0.828
50	0.31	−0.10	0.33	−0.10	0.93	0.92	0.12	0.799
60	0.29	0.01	0.30	0.00	0.95	0.86	−0.02	0.773
70	0.25	0.10	0.33	0.11	1.03	0.78	−0.15	0.748
80	0.19	0.18	0.27	0.22	1.04	0.74	−0.28	0.725
90	0.16	0.24	0.25	0.23	1.02	0.74	−0.39	0.675

To verify robustness to the 110 mm station-selection criterion (Sec. 3), the analysis was repeated at daily-precipitation cutoffs of *none*, 80, 110, and 150 mm: both indices shift upward monotonically with the cutoff while the CV-dependence and the relative GEV/GPD ordering are preserved, so the principal conclusions are insensitive to this choice (Supplementary Material 1).

Beyond the threshold-robustness check, the operational utility of STRI as a flash-flood trigger is assessed through probability-of-detection (POD) profiles as a function of the warning threshold z^* , an alarm being issued whenever $\text{STRI} \geq z^*$ (Fig. 9). A POD-vs- z^* representation is used instead of an ROC display (Jolliffe and Stephenson, 2012) because the GPD model returns STRI=NA for the 21.3% of damage-day events below the POT threshold (Sec. 4.5), imposing a structural POD ceiling—the detection rate attainable as the warning threshold is relaxed without bound—that declines with CV (from 0.89 at CV=10 mm to 0.61 at CV=90 mm; Fig. 9b), whereas the GEV-STRI is defined at every station–CV combination and reaches POD=1 in the same limit (Fig. 9a). Because damage events are rare in the candidate alarm pool (positive base rate $\approx 0.4\%$), any single fixed threshold operates at a high false-alarm ratio—a generic consequence of severe class imbalance in rare-event detection (Jolliffe and Stephenson, 2012)—so the index would be deployed not at a single global threshold but as one component of a multi-criterion alarm logic, with per-CV thresholds tuned to the local positive base rate. Per-CV POD ceilings, Youden-optimal thresholds, and false-alarm ratios at conventional thresholds are tabulated in Supplementary Material 3.

4.6 Spatial validation against the historical damage record

To verify the complementarity of the two models spatially, the distributions of the statistical risk index (STRI) and of the physical arrival time ($M\text{-}IAT$) across the 36 historical damage cases are compared.

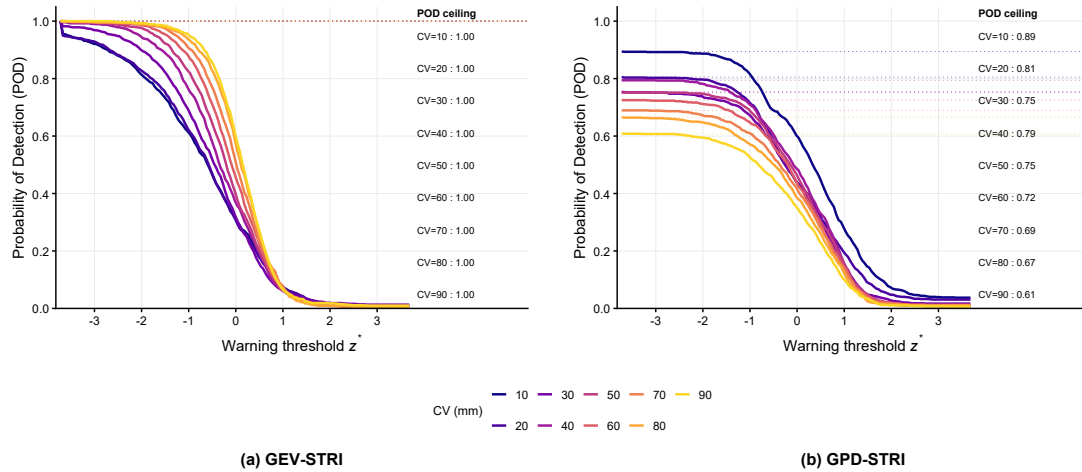


Figure 9. Probability-of-detection curves for STRI as a flash-flood trigger across the nine CVs: (a) GEV-STRI, (b) GPD-STRI. Dotted lines mark the POD ceiling reached as the threshold z^* is relaxed without bound.

The spatial distribution of the minimum STRI across damage stations is summarised in Figure 10, which compares GEV- and GPD-based indices. In the GEV maps the extreme-STRI area (redder bins) broadens with CV, as reaching a higher critical depth already selects severe storms; the declustered GPD, standardised within the POT tail, instead marks localised hotspots, preserving spatial sensitivity without saturation.

455 Because STRI normalises each observation against its station- and model-specific distribution, the z -scale in Figure 10 is commensurate across stations but not between models. Stations where both models show high z experienced events extreme under both sampling frameworks; disagreement most commonly marks a damage-day M - IAT that falls in the upper tail of the GEV block-minima distribution yet does not exceed the GPD threshold, and is therefore classified as non-extreme on the POT scale.

460 The corresponding spatial distribution of that same least-extreme (minimum-STRI) case's M - IAT is shown in Figure 11, in which the GEV and GPD models are placed side by side to expose how each model projects disaster risk onto the physical arrival-time scale.

The two models emphasise different parts of the same distribution: the GEV-based maps, built on annual block minima, portray the typical annual worst case relative to the 35-year climatology (longer arrival times than the POT tail, since block minima also include quieter years), whereas the GPD-based maps resolve the most extreme events more sharply (arrival times ~ 0.5 – 1.0 h)—two sampling views of one hazard rather than two kinds of risk. A CV-dependent gradient is evident in both: the shortest M - IAT occurs at CV=30 mm and lengthens progressively towards CV=90 mm, indicating intensity-dominated flash-flood characteristics around 30 mm and volume-dominated river-overflow characteristics above 90 mm; stations near Daejeon are an exception, retaining a comparatively short M - IAT even at CV=90 mm, where most stations lengthen to several hours:

470 the effect is sharpest in the GEV maps, whose return level shortens with CV and reaches the physical floor by CV=90 mm,

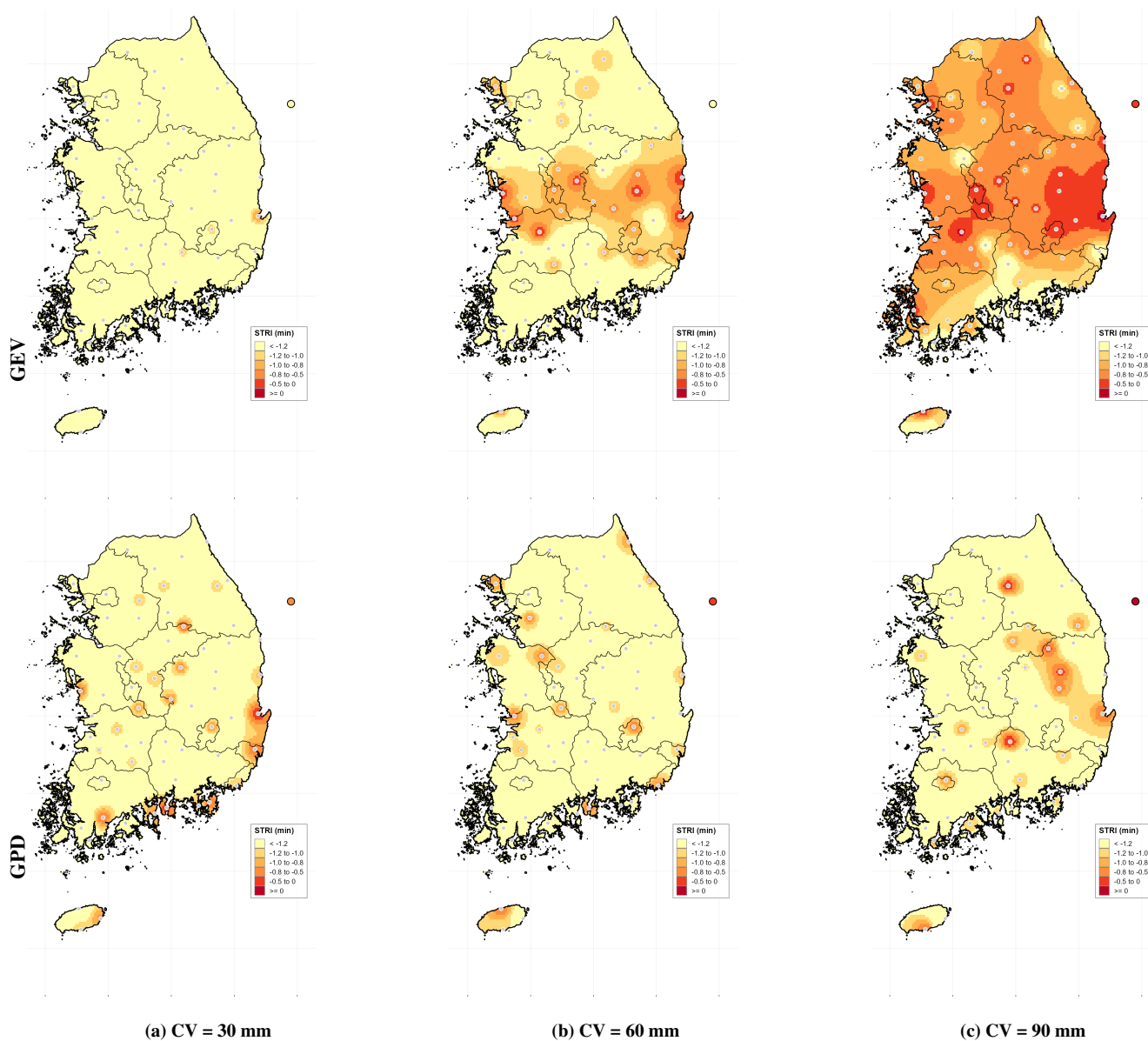


Figure 10. Minimum STRI across the historical damage cases; rows GEV/GPD, columns CV = 30, 60, 90 mm. Common six-bin z-score legend; bins skew negative because the minimum is taken over several damage events per station.

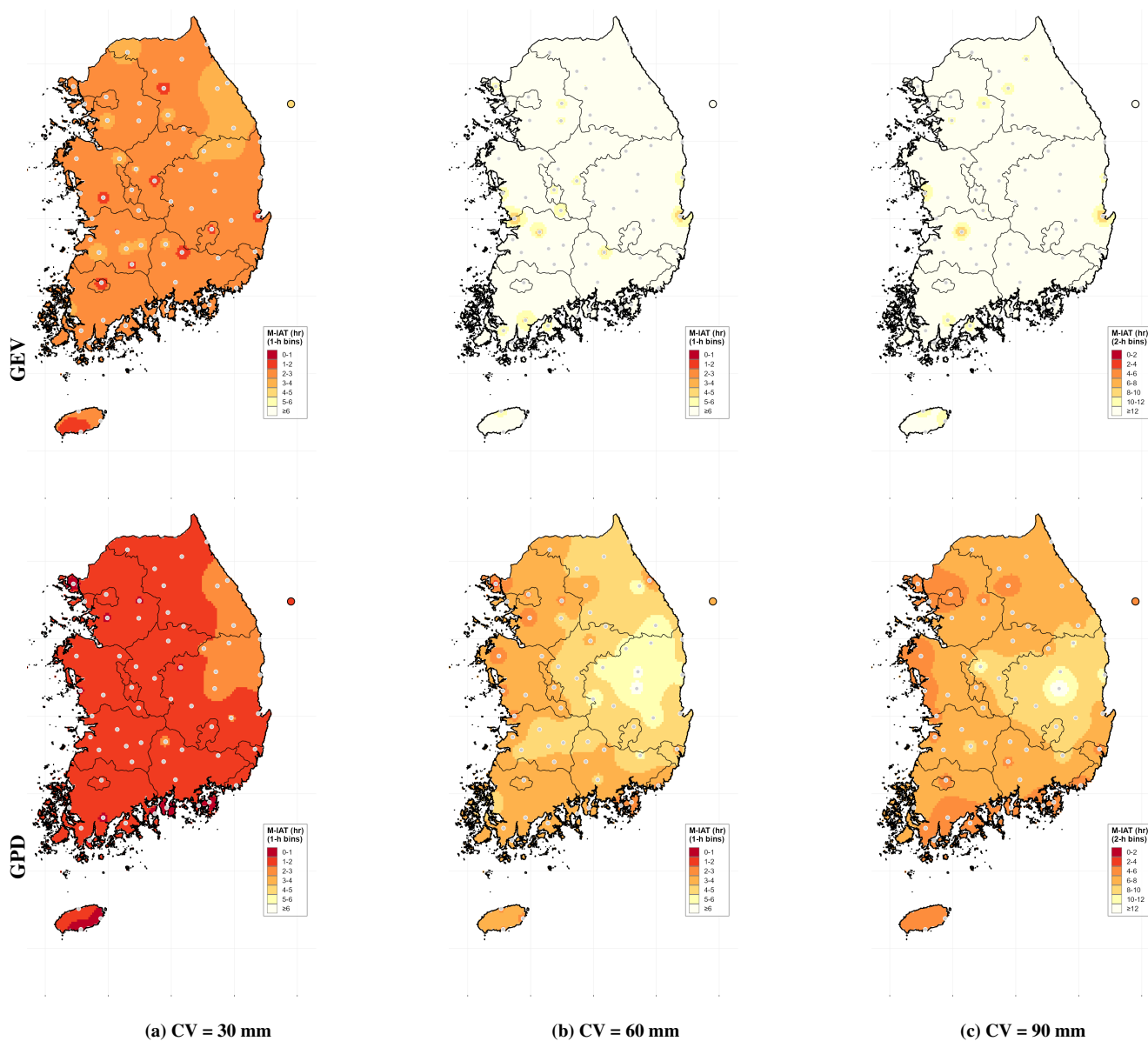


Figure 11. $M-IAT$ of the same least-extreme damage case mapped in Figure 10; rows GEV/GPD, columns CV = 30, 60, 90 mm. Redder tones mark shorter $M-IAT$; bin widths are CV-specific, so each panel carries its own legend.



whereas the GPD return level—though it still lengthens with CV—remains near 1 h, well below the network norm. This relative persistence suggests an abrupt weakening of the local hydrological defence margin above a station-specific critical amount.

The spatial arrangement of these values follows the principal rain-producing mechanisms over the peninsula (Lee et al., 2017): the shortest fast-accumulation times concentrate along the typhoon-exposed south coast and Jeju Island (median $M-IAT$ at CV = 30 mm of 0.88 and 0.84 h across the damage record, against 1.00 h at interior stations), interior stations occupy an intermediate range consistent with convective and Changma-front mesoscale systems, and the east-coast Yeongdong stations behind the Taebaek range show the *longest* values (median 1.27 h), reflecting prolonged, topographically forced easterly-flow rainfall. These contrasts are hydrologically informative precisely because $M-IAT$ isolates accumulation *rate*; urban imperviousness does not enter the index, so the elevated damage in metropolitan watersheds reflects exposure and accelerated runoff superimposed on the meteorological signal—a vulnerability multiplier rather than a rainfall-climatological feature.

4.7 Event-level validation against independent damage severity

The damage validations above are conditioned on stations selected by a daily-precipitation criterion; because STRI is itself a function of rainfall extremity, they are best read as internal consistency checks exposed to a circularity between predictor and label (Jolliffe and Stephenson, 2012). As an external, precipitation-*independent* criterion, we therefore additionally validate STRI against the casualties and property damage recorded for each event in the national disaster yearbook (Ministry of the Interior and Safety, 2025): these are determined by the realised impact of an event and are not derived from any rainfall threshold, so a positive association with STRI cannot be an artefact of the labelling rule.

For each documented event, STRI is reduced to a single event-level severity proxy: the median STRI across the nine CVs for every affected station–day, maximised over the event’s affected stations (the event “peak” STRI)—the peak rather than the mean, because realised loss is driven by the most extreme local concentration within an event, whereas the mean is diluted by the many marginally exposed stations that large events necessarily contain. The peak STRI is related to recorded losses by the Spearman rank correlation (Spearman, 1904), which is robust to the heavy-tailed, non-inflation-adjusted loss figures; multi-year bundled records and events without a yearbook entry were excluded, leaving $n = 27$ events.

Table 6. Spearman rank correlation between the event-level peak STRI and independently recorded damage severity, $n = 27$ events. The baselines are the event-peak daily and running 3-hour maximum precipitation.

Event peak index	Casualties		Property damage	
	ρ	p	ρ	p
GPD-STRI (max)	+0.56	0.004	+0.46	0.018
GEV-STRI (max)	+0.47	0.018	+0.35	0.074
Daily maximum precipitation (max)	+0.72	< 0.001	+0.43	0.027
3-h maximum precipitation (max)	+0.53	0.005	+0.36	0.066



The event peak STRI is positively and significantly associated with both impact measures (Table 6, Fig. 13): GPD-STRI
 495 attains $\rho = +0.56$ with casualties and $+0.46$ with property damage, the GEV-based index slightly lower. As the loss figures
 are independent of the precipitation criterion defining damage days, this is external corroboration that the events STRI ranks
 most extreme also produced the largest realised impacts. The GPD index discriminates slightly better than the GEV, consistent
 with GEV-STRI saturation at the upper cap for the most extreme events (Sec. 4.5).

A leave-one-out jackknife confirms the correlations are not driven by influential events: removing each of the 27 events in
 500 turn reduces ρ by less than 0.05, except one *concordant* low-severity event (Typhoon Soulik) whose removal *lowers* ρ by ~ 0.06 ,
 so the reported correlations are conservative; the visually discordant high-STRI/low-damage points reflect genuine exposure-
 and-vulnerability heterogeneity (Gall et al., 2009) rather than measurement error (per-event jackknife values in Supplementary
 Material 4).

To quantify the incremental value of STRI over simpler precipitation summaries, the identical event-level validation was
 505 repeated for two baselines derived from the same hourly record: the event-peak daily and 3-hour maximum precipitation
 (Table 6). The daily-maximum baseline correlates more strongly with casualties ($\rho = +0.72$) than GPD-STRI ($+0.56$), whereas
 for property damage GPD-STRI is the best-correlated index ($+0.46$, exceeding all baselines); none of the pairwise differences
 is statistically significant at $n = 27$ (paired bootstrap over events; Supplementary Material 4). At the event level, therefore,
 STRI is statistically indistinguishable from raw precipitation maxima as a *retrospective* severity correlate—an expected parity,
 510 since the value of the index lies in its prospective, station-standardised design rather than in additional correlation with realised
 loss (Secs. 4.8, 5). A station-day POD/AUC comparison against these baselines is structurally uninformative, because the
 damage-day labels embed a precipitation criterion that guarantees near-perfect AUC for raw-precipitation scores irrespective of
 hydrological merit; only the impact-based comparison escapes this circularity. The limitation of an absolute, nationally uniform
 criterion is nevertheless visible in the present record: only 53.9% of damage station-days reached the national torrential-rain
 515 advisory (60 mm in 3 h) on the day of damage—precisely the location-dependence that a station-standardised index is designed
 to absorb (Sec. 4.8).

A complementary question is *which* component of the rainfall STRI carries. Across the same 27 events, the event-peak GEV-
 STRI is only weakly rank-correlated with the event-peak daily total ($\rho = 0.40$) but strongly correlated with the event-peak
 hourly rainfall ($\rho = 0.85$; Fig. 12); the two maxima are taken independently over each event's affected stations. STRI therefore
 520 preferentially encodes short-duration *intensity*—the driver of flash-flood generation once infiltration and drainage capacity are
 exceeded (Doswell et al., 1996)—rather than accumulated volume. This ordering is robust to both the station pairing and the
 sample: over all 36 documented events, with the precipitation taken at the station where STRI itself peaks, the correlations
 are 0.83 (hourly) and 0.54 (daily) (Supplementary Material 4). STRI is accordingly best read not as a competitor to existing
 indices but as a *complementary* perspective: IDF curves are a design-rainfall tool and daily-maximum indices summarise event
 525 magnitude, whereas STRI standardises the *temporal concentration* of an event, separating storms of comparable return period
 or daily total into distinct risk characters when their speed of accumulation differs.

Finally, because all preceding validations are in-sample, a temporal-split cross-validation was performed: stationary fits re-
 estimated from 1990–2010 alone were used to recompute STRI for the 17 documented events of 2011–2024. The out-of-sample

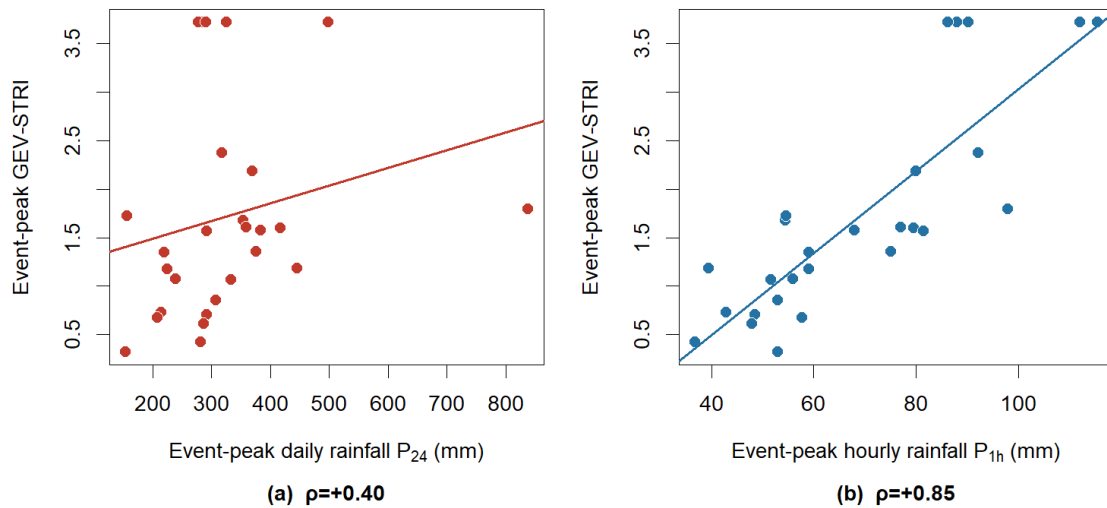


Figure 12. Event-peak GEV-STRI against (a) daily-maximum rainfall P_{24} and (b) hourly-maximum rainfall P_{1h} for the 27 documented events. STRI tracks short-duration intensity ($\rho = 0.85$) far more closely than daily volume ($\rho = 0.40$).

index closely tracks its in-sample counterpart (station-day correlation $r = 0.95$ for GEV-STRI and 0.89 for GPD-STRI),
 and the external validation survives the split: out-of-sample GPD-STRI remains significantly rank-correlated with casual-
 ties ($\rho = +0.64$, $p = 0.005$; GEV-STRI $+0.58$, $p = 0.016$), while property-damage correlations are positive but not significant
 in this reduced sample. Cell-level shape estimates are considerably noisier across the split, so the out-of-sample stability of
 STRI derives from the fitted CDF as a whole rather than from any individual parameter: a climatology frozen in 2010 would
 have ranked the 2011–2024 impacts essentially as the full-record index does (protocol and per-event values in Supplementary
 Material 4).

The moderate rather than near-perfect correlation is expected from a hazard index: realised loss is governed jointly by hazard
 intensity and by exposure and vulnerability (Gall et al., 2009), so the results are best read as corroborative external evidence for
 the operational meaning of STRI rather than as a calibrated loss model; the casualty count is dimensionally stable over time,
 whereas the property-damage figures span three decades without inflation adjustment.

4.8 Warning equity across the station network

A single event illustrates the construction and added value of the index. On 21 September 2024 two gauges recorded almost
 identical 24-hour totals—Seosan (ID 129), 123.1 mm, and Ulleungdo (ID 115), 124.4 mm—yet at Seosan the first 30 mm
 accumulated in $M-IAT = 0.41$ h, a fast, flash-flood-prone burst, whereas at Ulleungdo the same depth took 4.16 h (Fig. 14).
 Passing each station's own $M-IAT$ through its fitted GEV (CV= 30 mm) returns STRI= $+1.13$ at Seosan and -2.54 at
 Ulleungdo: two events that a rainfall total cannot tell apart are separated into a locally rare, dangerous accumulation and a
 locally ordinary one.

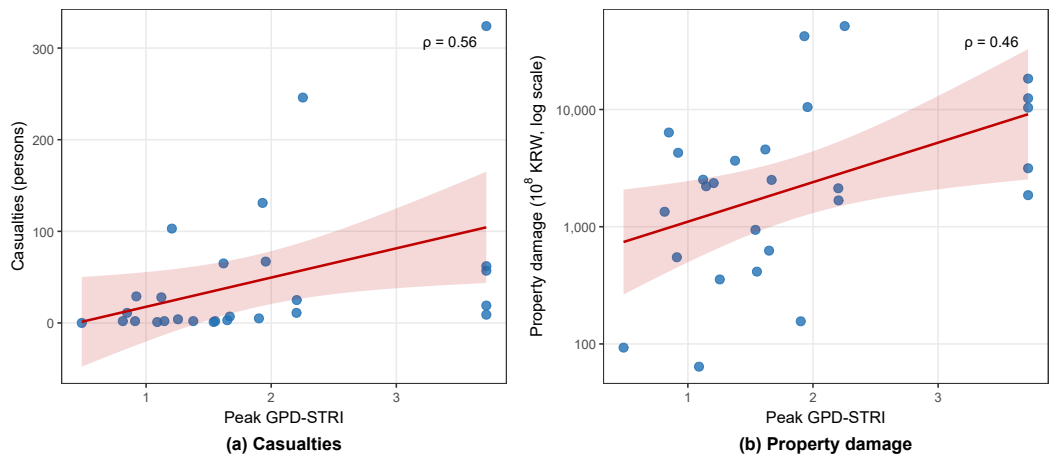


Figure 13. Event-level peak GPD-STRI versus (a) recorded casualties and (b) recorded property damage (log scale) for the 27 documented events, with Spearman correlations annotated; GEV-STRI results are in Table 6.

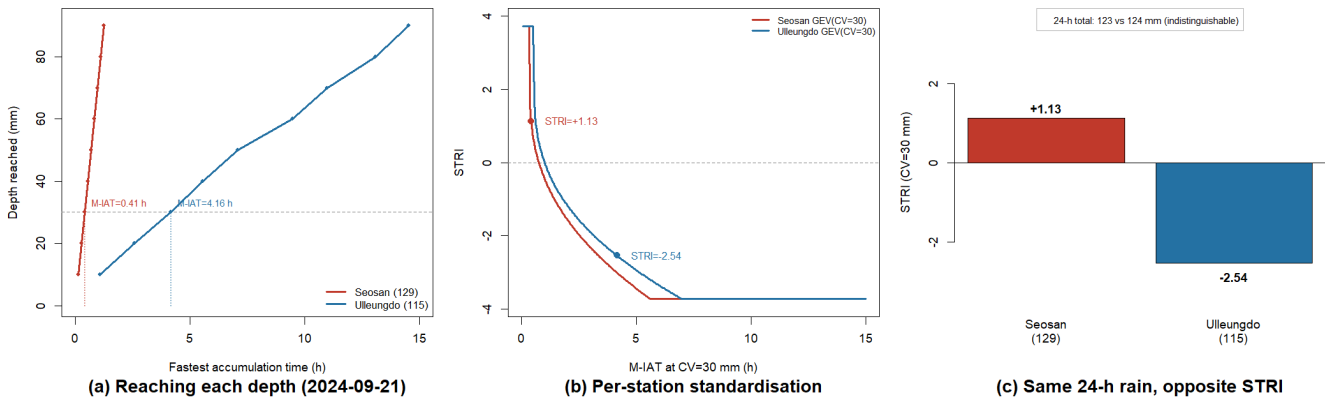


Figure 14. Two gauges on 21 September 2024 with near-identical 24-hour totals but very different accumulation rates. (a) Time to reach each depth; (b) the same *M-IAT* standardised through each station’s own GEV (*CV* = 30 mm); (c) identical totals, opposite STRI.



Because flood damage is generated by locally—not nationally—extreme rainfall, a single nationwide depth threshold is expected to over-serve climatologically wet gauges and under-serve dry ones, whereas a station-standardised threshold fires at a constant exceedance probability everywhere. We test this directly on the damage station–days, operationalising the between-station equity defined in Sect. 1. For each CV, an absolute-depth rule and an STRI rule are calibrated to the *same* overall damage-capture rate (recall = 0.60), so the two rules expend an identical alarm budget and differ only in *which* events they capture; per-station recall is then related to the station’s climatological wetness, taken as the 95th percentile of its daily-precipitation record.

At CV= 30 mm (Fig. 15; all 65 stations resolved), per-station recall under the absolute-depth rule is significantly correlated with station wetness (Spearman $\rho = +0.32$, $p = 0.009$): at matched overall recall, climatologically wet gauges capture systematically more of their damage days than dry gauges. Under the STRI rule this dependence vanishes ($\rho = +0.00$, $p = 0.99$) and the cross-station dispersion of recall falls (standard deviation $0.18 \rightarrow 0.13$); the contrast holds at every CV (Supplementary Material 4). A concrete instance is the inland gauge at Andong (ID 136): none of its six damage days (101–117 mm daily precipitation) reached the 128 mm cutoff at which the absolute rule operates at this budget, so that rule would have issued no warning on any of them, whereas the standardised index flags three of the six as locally extreme.

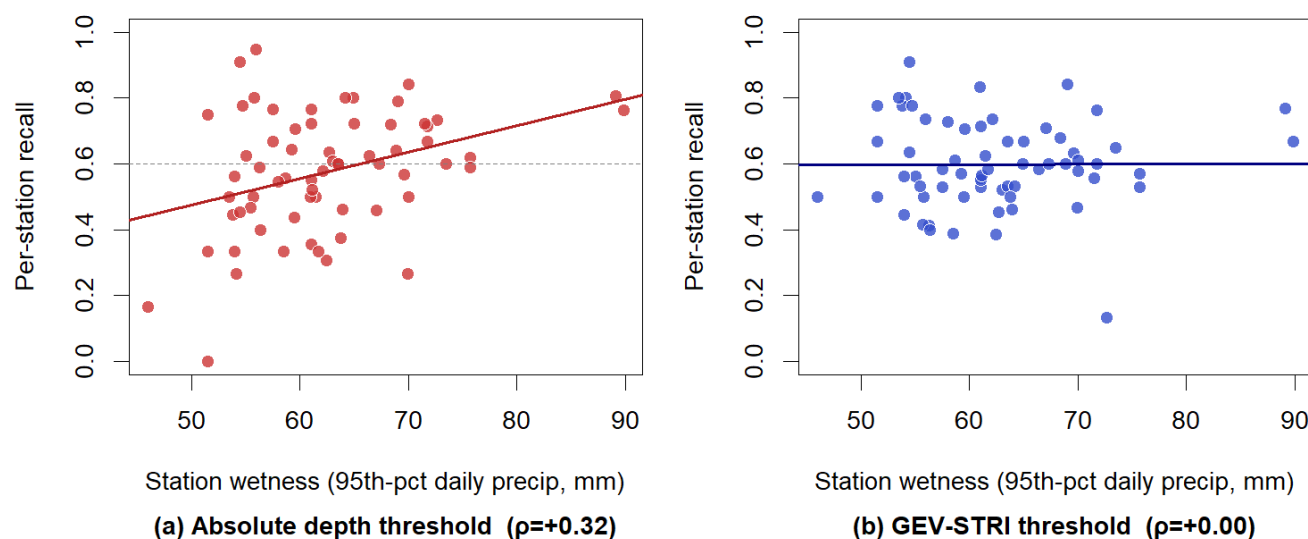


Figure 15. Warning equity at a matched damage-capture rate (recall = 0.60), CV= 30 mm. Each point is a station: recall against climatological wetness. (a) Fixed national depth threshold; (b) fixed GEV-STRI threshold. Dashed line marks recall = 0.60.



5 Discussion

$M-IAT$ is the event-based dual to IDF curves: the same rainfall record is parameterised by fixed depth and random duration rather than by fixed duration and random depth (Eq. 4). The dual does not, by itself, supply additional information about the rainfall process; what it changes is which quantity is observable at decision time. The realised $M-IAT$ is the lead time
 565 available to a depth-triggered warning system (“how long until the basin sees c mm?”) and can be monitored prospectively as rainfall accumulates, whereas conventional severity proxies—daily or running multi-hour maxima—are hindsight quantities, known only once the event has ended (Doswell et al., 1996). The analytical investment of this study accordingly lies not in a new data source but in the rigorous EVT treatment of the dual side: sampling schemes consistent with the block-minima and POT paradigms, explicit declustering, estimation on a common shape domain, and a non-stationary extension coherent with
 570 the climate-change premise.

The GEV and GPD shape parameters, which EVT requires to coincide asymptotically (Pickands, 1975; Balkema and de Haan, 1974; Coles, 2001), agree closely (median $|\hat{\xi}_{\text{GEV}} - \hat{\xi}_{\text{GPD}}| = 0.18$) once within-storm dependence is removed by the cluster-minimum construction—verified, not produced, by the Ferro–Segers index $\hat{\theta} \approx 1$ (Ferro and Segers, 2003)—and the two estimators are placed on a common shape domain by Winsorizing the GEV shape at the regularity boundary $\hat{\xi} = -1$
 575 (Smith, 1985; Hosking and Wallis, 1997). The residual difference reflects finite-sample bias and the slower convergence of POT statistics to the bounded domain (Davison and Smith, 1990); full numerical agreement should not be expected, or claimed, at these sample sizes. More informative is that both shapes concentrate in the bounded reverse-Weibull domain ($\hat{\xi} < 0$): the fastest accumulation achievable at a station imposes a finite endpoint on the sign-reversed variable, so this boundedness is the expected—not anomalous—signature of $M-IAT$ extremes, hidden in standard upper-tail IDF analyses (Sec. 2.1).

Coherence with the climate-change premise requires a non-stationary treatment, yet the data do not support a blanket switch. Applied symmetrically to the GEV and GPD models via the mixed-model framework of Cheng et al. (2014), an LRT-supported non-stationary fit is retained at only 3.1% (GEV) and 5.6% (GPD) of the 585 station–CV combinations, and Benjamini–Hochberg correction leaves only a single regular GEV combination significant (Benjamini and Hochberg, 1995); the fitted *magnitude* of $M-IAT$ extremes is thus effectively stationary over this record, in line with the “stationarity is undead” position
 585 of Serinaldi and Kilsby (2015). The *frequency* channel differs: annual counts of fast-accumulation events trend positive at all nine CVs, significantly so once the record is extended to the homogeneity-screened 1975–2024 series (+7.7 to +10.7% per decade). Together the two channels resolve the apparent tension with the documented national rise in heavy-rainfall days (Do et al., 2023; Park and Suh, 2023): over this record, intensification manifests as *more frequent*, not *faster*, critical-depth accumulations. Operationally, a rising occurrence rate raises the alarm load at a fixed threshold without invalidating the percentile scale
 590 on which STRI is computed, whereas a magnitude trend would change the threshold’s meaning; only the former is currently supported.

The originality of the STRI lies in its interpretive framework. Unlike conventional indices such as the SPI (McKee et al., 1993), which measure deviations from the mean of a general precipitation distribution, STRI is a standardised normal index defined within an extreme-value population. Whereas meteorological drought has long had such a standardised, distribution-



595 referenced benchmark, no equivalent index has existed for sub-daily torrential rainfall, where operational practice still relies on absolute depth thresholds and IDF return periods; STRI is intended to fill this gap, and is accordingly benchmarked here against those existing references—precipitation maxima and absolute criteria—rather than against a pre-existing standardised index. Its zero point is the median past extreme—an event of average severity among extremes, and already damaging—so that a positive value marks an event more severe than the typical past extreme. A zero or negative STRI value does not imply
 600 absence of risk; rather, it signifies an event that is relatively common within the extreme distribution yet still capable of causing recurrent damage. The spatial distributions in Figures 10 and 11 show that flood damages have occurred repeatedly even where the underlying *M-IAT* values do not reach theoretical extreme levels; this residual mismatch reflects the exposure-and-vulnerability component of realised damage, which lies outside any rainfall metric (Gall et al., 2009), and motivates the detection and false-alarm analysis of Sec. 4.5.

605 STRI should therefore be understood not as a direct predictor of damage but as a standardised diagnostic of rainfall extremity with an explicit lead-time interpretation through *M-IAT*. The external validation delimits this honestly: the event-peak STRI is significantly rank-correlated with independently recorded casualties and property losses (Spearman $\rho = 0.56$ and 0.46 ; Sec. 4.7), yet at $n = 27$ events it is statistically indistinguishable from raw precipitation maxima as a *retrospective* severity correlate (Table 6)—the expected signature of an index that isolates the meteorological component of flood risk and leaves
 610 the socioeconomic component to exposure-and-vulnerability modelling (Gall et al., 2009). The case for STRI is therefore operational, and twofold. First, *warning equity*: at a matched warning budget an absolute-depth rule captures systematically more damage at climatologically wet gauges than at dry ones (wetness–recall correlation $+0.32$), a dependence that vanishes under the standardised trigger ($\rho = 0.00$; Sec. 4.8), so a single national threshold no longer under-warns the drier interior. Second, *prospective monitorability*: because *M-IAT* is observable as rainfall accumulates, the index carries a lead-time meaning
 615 that hindsight totals cannot provide. The false-alarm analysis tempers, without overturning, this case: at the $\approx 0.4\%$ positive base rate any single-threshold trigger operates at a high false-alarm ratio (Jolliffe and Stephenson, 2012), so STRI is best deployed as one component of a multi-criterion alarm logic with per-CV thresholds calibrated to the local base rate (Sec. 4.5). Of the two readings we recommend the GPD-STRI as the primary operational trigger—it resolves localised extremes more sharply and avoids the upper-cap saturation to which the GEV-STRI is prone for the rarest events (Fig. 6)—with the GEV-
 620 STRI retained as a complementary block-minima reading defined at every station–CV combination. Finally, the temporal-split cross-validation confirms that the index is estimable from a truncated record without material degradation; the decisive next step is a nowcasting-based real-time replay demonstrating improved *prospective* warning skill over absolute thresholds.

6 Conclusions

This study developed and validated the Standardized Torrential Rainfall Index (STRI), an extreme-value standardisation of
 625 the Modified Inter-Amount Time (*M-IAT*)—the time for cumulative precipitation within an event to reach a critical depth—positioned as the event-based dual to conventional IDF curves. The contributions are (i) a single normal-quantile index that replaces absolute, nationally uniform precipitation thresholds with a station-standardised measure of *relative* rarity, so that



one operational threshold carries the same exceedance meaning across stations of differing climatology, critical values, and decades; (ii) two physically motivated sampling schemes—annual minima for the GEV model and rainy-cluster minima for the GPD model—that supersede ad hoc inter-amount duration cut-offs; (iii) a model-selection procedure combining declustered POT with Ferro–Segers verification of asymptotic independence (median extremal index ≈ 1), stationary L-moments Winsorized at the regularity boundary $\hat{\xi} = -1$, and non-stationary MLE across the 585 station–CV combinations, which places both models on a common shape domain and yields close, threshold-robust GEV–GPD shape agreement (median $|\Delta\hat{\xi}| = 0.18$) in the bounded reverse-Weibull domain; and (iv) external corroboration against a precipitation-independent impact register—complemented by a temporal-split cross-validation and a warning-equity analysis at a matched alarm budget—which sidesteps the circularity present in most damage-validation studies of precipitation-based indices.

Three empirical findings stand out. First, the *magnitude* of *M-IAT* extremes in the Korean record is largely stationary—non-stationary fits are retained at only 3.1% (GEV) and 5.6% (GPD) of station–CV combinations, none surviving false-discovery-rate correction except a single GEV cell—whereas the *frequency* of fast-accumulation events trends positive at all nine critical values and becomes significant at every one of them over the extended 1975–2024 record: intensification over this record manifests as more frequent, not faster, critical-depth accumulations, reconciling the stationarity of the fitted extreme-value parameters with the documented national increase in heavy-rainfall days (Do et al., 2023; Park and Suh, 2023). Second, the event-level peak STRI is significantly rank-correlated with the casualties ($\rho = 0.56$) and property losses ($\rho = 0.46$) recorded in the national disaster yearbook (Ministry of the Interior and Safety, 2025), a criterion independent of any precipitation threshold, which addresses the circularity that arises when a precipitation-derived index is validated on precipitation-thresholded labels; at $n = 27$ events, however, STRI is statistically indistinguishable from raw precipitation maxima as a retrospective severity correlate. Third—and constituting the operational case for the index—STRI’s contribution is not a new estimate of regional rainfall amount but a warning framework of *equal rarity*: a fixed STRI trigger distributes a matched warning budget equitably across the wet–dry gradient of the network, removing the wetness bias of an absolute national threshold (Sec. 4.8).

Because the framework requires only a station’s own precipitation record, it is in principle transferable to other climate regimes—monsoon, mid-latitude, and tropical—although thresholds and any trend covariate would require local recalibration. Several limitations bear mention. The linear-in-time non-stationarity assumption (12)–(13) is a first-order approximation that should eventually be replaced by smoothing-spline or large-scale-covariate parameterisations such as global mean temperature or ENSO indices (Cheng et al., 2014). The 35-year record limits statistical power in both trend channels: the non-detection of magnitude trends and the non-significance of the 35-year frequency trends are compatible with real but undetected change, as the significant 50-year frequency extension indicates (Serinaldi and Kilsby, 2015). The external validation rests on $n = 27$ events drawn from administrative damage records biased towards disasters that crossed official reporting thresholds (Gall et al., 2009), and the heterogeneous composition of the register—typhoon-driven and stagnant-front events with differing temporal structures—may warrant stratified validation in future work. The automated POT threshold selection reuses the Cramér–von Mises statistic subsequently reported as a goodness-of-fit measure, so threshold-choice uncertainty is not propagated to the return-level standard errors (Scarrott and MacDonald, 2012). The conditional-independence assumption across the 65 stations could be relaxed by spatially correlated EVT, yielding event-level rather than station–day-level return estimates, and the IDW



maps are illustrative only, the station-scale estimates being the primary inference. Operationally, the incremental *trigger* skill of STRI—its timeliness and false-alarm behaviour at locally calibrated thresholds—remains to be demonstrated prospectively, through nowcasting-based replay of events, before adoption; the temporal-split cross-validation (Sec. 4.7) establishes the out-of-sample stability of the index but is not a substitute for that demonstration. The *M-IAT* construction also extends to drought: torrential-rainfall analysis uses the minimum time to reach a precipitation threshold, whereas the same inter-amount time—when long—quantifies water-supply-delay risk, so that floods and droughts can be treated within one framework.

Code availability. The R code implementing the *M-IAT* construction, the declustering procedure, the stationary and non-stationary GEV/GPD fits, and the STRI transform will be deposited in a public repository with a DOI upon acceptance, and is available from the corresponding author in the interim.

Data availability. The hourly precipitation records for the 65 ASOS stations are openly available, free of charge, from the Korea Meteorological Administration Open Data Portal (<https://data.kma.go.kr>) under the KMA open-data licence; station identifiers and coordinates needed to reproduce the retrieval are listed in Table S1. The documented flood-damage cases were compiled from the Water Information Portal (<https://www.water.or.kr/>) and the national disaster yearbook (Ministry of the Interior and Safety, 2025); the case list with station-level detail is given in Supplementary Material 1. The derived *M-IAT* event samples, the complete stationary and non-stationary GEV/GPD parameter estimates for all 65 stations, and the station-wise STRI values for every damage event are provided in Supplementary Materials 1–4 and will be deposited alongside the code upon acceptance.

Author contributions. Misook Kwak: methodology, validation, formal analysis, data curation, writing—original draft, visualisation. Hyunjun Ahn: methodology, validation, writing—review and editing. Sanghoo Yoon: conceptualisation, supervision, funding acquisition, writing—review and editing.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. Hourly precipitation data were provided by the Korea Meteorological Administration, and the flood-damage records by the Ministry of the Interior and Safety.

Financial support. This work was supported by the G-LAMP Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education of the Republic of Korea (No. RS-2024-00442775).



References

- Alexandersson, H.: A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661–675, <https://doi.org/10.1002/joc.3370060607>, 1986.
- Allan, R. P. and Soden, B. J.: Atmospheric warming and the amplification of precipitation extremes, *Science*, 321, 1481–1484, <https://doi.org/10.1126/science.1160787>, 2008.
- Baek, H.-J., Kim, M.-K., and Kwon, W.-T.: Observed short- and long-term changes in summer precipitation over South Korea and their links to large-scale circulation anomalies, *Int. J. Climatol.*, 37, 972–986, <https://doi.org/10.1002/joc.4753>, 2017.
- Balkema, A. A. and de Haan, L.: Residual life time at great age, *Ann. Probab.*, 2, 792–804, <https://doi.org/10.1214/aop/1176996548>, 1974.
- Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J. Roy. Stat. Soc. B*, 57, 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>, 1995.
- Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. W.: Non-stationary extreme value analysis in a changing climate, *Climatic Change*, 127, 353–369, <https://doi.org/10.1007/s10584-014-1254-5>, 2014.
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, UK, <https://doi.org/10.1007/978-1-4471-3675-0>, 2001.
- Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds, *J. Roy. Stat. Soc. B*, 52, 393–425, <https://doi.org/10.1111/j.2517-6161.1990.tb01796.x>, 1990.
- Do, H.-S., Kim, J., Cha, E.-J., Chang, E.-C., Son, S.-W., and Lee, G.: Long-term change of summer mean and extreme precipitations in Korea and East Asia, *Int. J. Climatol.*, 43, 3476–3492, <https://doi.org/10.1002/joc.8039>, 2023.
- Doswell III, C. A., Brooks, H. E., and Maddox, R. A.: Flash flood forecasting: An ingredients-based methodology, *Weather Forecast.*, 11, 560–581, [https://doi.org/10.1175/1520-0434\(1996\)011<0560:FFFAIB>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2), 1996.
- Dunkerley, D.: Rain event properties in nature and in rainfall simulation experiments: a comparative review with recommendations for increasingly systematic study and reporting, *Hydrol. Process.*, 22, 4415–4435, <https://doi.org/10.1002/hyp.7045>, 2008.
- El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., and Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43, W03410, <https://doi.org/10.1029/2005WR004545>, 2007.
- Ferro, C. A. T. and Segers, J.: Inference for clusters of extreme values, *J. Roy. Stat. Soc. B*, 65, 545–556, <https://doi.org/10.1111/1467-9868.00401>, 2003.
- Fowler, H. J., Ali, H., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Cabi, N. S., Chan, S., Dale, M., Dunn, R. J. H., Ekström, M., Evans, J. P., Fossier, G., Golding, B., Guerreiro, S. B., Hegerl, G. C., Kahraman, A., Kendon, E. J., Lenderink, G., Lewis, E., Li, X., O’Gorman, P. A., Orr, H. G., Peat, K. L., Prein, A. F., Pritchard, D., Schär, C., Sharma, A., Stott, P. A., Villalobos-Herrera, R., Villarini, G., Wasko, C., Wehner, M. F., Westra, S., and Whitford, A.: Towards advancing scientific knowledge of climate change impacts on short-duration rainfall extremes, *Philos. T. R. Soc. A*, 379, 20190542, <https://doi.org/10.1098/rsta.2019.0542>, 2021.
- Gall, M., Borden, K. A., and Cutter, S. L.: When do losses count? Six fallacies of natural hazards loss data, *B. Am. Meteorol. Soc.*, 90, 799–810, <https://doi.org/10.1175/2008BAMS2721.1>, 2009.
- Hosking, J. R. M.: L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. Roy. Stat. Soc. B*, 52, 105–124, <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>, 1990.
- Hosking, J. R. M. and Wallis, J. R.: *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK, <https://doi.org/10.1017/CBO9780511529443>, 1997.



- Jolliffe, I. T. and Stephenson, D. B. (Eds.): *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd edn., Wiley-Blackwell, Chichester, UK, <https://doi.org/10.1002/9781119960003>, 2012.
- 725 Kamiguchi, K., Arakawa, O., Kitoh, A., Yatagai, A., Hamada, A., and Yasutomi, N.: Development of APHRO_JP, the first Japanese high-resolution daily precipitation product for more than 100 years, *Hydrol. Res. Lett.*, 4, 60–64, <https://doi.org/10.3178/hrl.4.60>, 2010.
- Karl, T. R. and Knight, R. W.: Secular trends of precipitation amount, frequency, and intensity in the United States, *B. Am. Meteorol. Soc.*, 79, 231–241, [https://doi.org/10.1175/1520-0477\(1998\)079<0231:STOPAF>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0231:STOPAF>2.0.CO;2), 1998.
- Kim, J. H.: Development of Korean Flash Flood Guidance Model, PhD thesis, Sejong University, Seoul, South Korea, https://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=1062b9f8f2dc9d49ffe0bdc3ef48d419, 2007 (in Korean with English abstract).
- 730 Kim, Y.-J., Yeh, S.-W., Choi, Y.-S., Son, S.-W., Oh, S.-G., Yang, Y.-M., and Kim, J.: The need for climate risk assessment in response to the increasing occurrence of compound drought-heatwave events in Republic of Korea, *Atmosphere-Korea*, 35, 13–27, <https://doi.org/10.14191/Atmos.2025.35.1.013>, 2025 (in Korean with English abstract).
- 735 Koutsoyiannis, D., Kozonis, D., and Manetas, A.: A mathematical framework for studying rainfall intensity–duration–frequency relationships, *J. Hydrol.*, 206, 118–135, [https://doi.org/10.1016/S0022-1694\(98\)00097-3](https://doi.org/10.1016/S0022-1694(98)00097-3), 1998.
- Krishnamurthy, C. K. B., Lall, U., and Kwon, H.-H.: Changing frequency and intensity of rainfall extremes over India from 1951 to 2003, *J. Climate*, 22, 4737–4746, <https://doi.org/10.1175/2009JCLI2896.1>, 2009.
- Kunkel, K. E., Karl, T. R., Brooks, H., Kossin, J., Lawrimore, J. H., Arndt, D., Bosart, L., Changnon, D., Cutter, S. L., Doesken, N., Emanuel, K., Groisman, P. Ya., Katz, R. W., Knutson, T., O'Brien, J., Paciorek, C. J., Peterson, T. C., Redmond, K., Robinson, D., Trapp, J., Vose, R., Weaver, S., Wehner, M., Wolter, K., and Wuebbles, D.: Monitoring and understanding trends in extreme storms: State of knowledge, *B. Am. Meteorol. Soc.*, 94, 499–514, <https://doi.org/10.1175/BAMS-D-11-00262.1>, 2013a.
- 740 Kunkel, K. E., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X., and Hennon, P.: Probable maximum precipitation and climate change, *Geophys. Res. Lett.*, 40, 1402–1408, <https://doi.org/10.1002/grl.50334>, 2013b.
- 745 Lee, J.-Y., Kwon, M., Yun, K.-S., Min, S.-K., Park, I.-H., Ham, Y.-G., Jin, E. K., Kim, J.-H., Seo, K.-H., Kim, W., Yim, S.-Y., and Yoon, J.-H.: The long-term variability of Changma in the East Asian summer monsoon system: A review and revisit, *Asia-Pac. J. Atmos. Sci.*, 53, 257–272, <https://doi.org/10.1007/s13143-017-0032-5>, 2017.
- Lee, J., Kwon, T., and Yoon, S.: Analysis of the generalized extreme value of drought on the Korean Peninsula using inter-amount times, *J. Korean Data Inf. Sci. Soc.*, 32, 1023–1034, <https://doi.org/10.7465/jkdi.2021.32.5.1023>, 2021.
- 750 Lenderink, G., de Vries, H., Fowler, H. J., Barbero, R., van Ulft, B., and van Meijgaard, E.: Scaling and responses of extreme hourly precipitation in three climate experiments with a convection-permitting model, *Philos. T. R. Soc. A*, 379, 20190544, <https://doi.org/10.1098/rsta.2019.0544>, 2021.
- Li, X.-F., Blenkinsop, S., Barbero, R., Yu, J., Lewis, E., Lenderink, G., Guerreiro, S., Chan, S., Li, Y., Ali, H., Villalobos Herrera, R., Kendon, E., and Fowler, H. J.: Global distribution of the intensity and frequency of hourly precipitation and their responses to ENSO, *Clim. Dynam.*, 54, 4823–4839, <https://doi.org/10.1007/s00382-020-05258-7>, 2020.
- 755 Lima, C. H. R., Kwon, H.-H., and Kim, J.-Y.: A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate, *J. Hydrol.*, 540, 744–756, <https://doi.org/10.1016/j.jhydrol.2016.06.062>, 2016.
- Madsen, H., Rasmussen, P. F., and Rosbjerg, D.: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling, *Water Resour. Res.*, 33, 747–757, <https://doi.org/10.1029/96WR03848>, 1997.



- 760 McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993, American Meteorological Society, Boston, MA, USA, 179–184, https://www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf, 1993.
- Melillo, J. M., Richmond, T. C., and Yohe, G. W. (Eds.): Climate Change Impacts in the United States: The Third National Climate Assessment, U.S. Global Change Research Program, Washington, DC, USA, <https://doi.org/10.7930/J0Z31WJ2>, 2014.
- 765 Min, S.-K., Zhang, X., Zwiers, F. W., and Hegerl, G. C.: Human contribution to more-intense precipitation extremes, *Nature*, 470, 378–381, <https://doi.org/10.1038/nature09763>, 2011.
- Ministry of Land, Transport and Maritime Affairs: Improvement and complementation of the probability rainfall depth maps of Korea, Ministry of Land, Transport and Maritime Affairs, Republic of Korea, <https://www.riss.kr/link?id=M12721452>, 2011 (in Korean).
- 770 Ministry of the Interior and Safety: 2024 Disaster Yearbook, Ministry of the Interior and Safety, Republic of Korea, https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000014&nttId=122722, 2025 (in Korean; flood damage records by event accessed via the MyWater portal (K-water), <https://www.water.or.kr/>).
- Park, I.-G. and Suh, M.-S.: Recent (30, 50 years) extreme precipitation characteristics and changing trends in Korea, *J. Clim. Change Res.*, 14, 737–754, <https://doi.org/10.15531/KSCCR.2023.14.6.737>, 2023.
- 775 Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., and Sanderson, B. M.: Precipitation variability increases in a warmer climate, *Sci. Rep.*, 7, 17966, <https://doi.org/10.1038/s41598-017-17966-y>, 2017.
- Pettitt, A. N.: A non-parametric approach to the change-point problem, *J. Roy. Stat. Soc. C-Appl.*, 28, 126–135, <https://doi.org/10.2307/2346729>, 1979.
- Pickands III, J.: Statistical inference using extreme order statistics, *Ann. Stat.*, 3, 119–131, <https://doi.org/10.1214/aos/1176343003>, 1975.
- 780 Restrepo-Posada, P. J. and Eagleson, P. S.: Identification of independent rainstorms, *J. Hydrol.*, 55, 303–319, [https://doi.org/10.1016/0022-1694\(82\)90136-6](https://doi.org/10.1016/0022-1694(82)90136-6), 1982.
- Scarrott, C. and MacDonald, A.: A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT-Stat. J.*, 10, 33–60, <https://doi.org/10.57805/revstat.v10i1.110>, 2012.
- Schleiss, M. and Smith, J. A.: Two simple metrics for quantifying rainfall intermittency: The burstiness and memory of interamount times, *J. Hydrometeorol.*, 17, 421–436, <https://doi.org/10.1175/JHM-D-15-0078.1>, 2016.
- 785 Serinaldi, F. and Kilsby, C. G.: Stationarity is undead: Uncertainty dominates the distribution of extremes, *Adv. Water Resour.*, 77, 17–36, <https://doi.org/10.1016/j.advwatres.2014.12.013>, 2015.
- Smith, R. L.: Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, 72, 67–90, <https://doi.org/10.1093/biomet/72.1.67>, 1985.
- 790 Spearman, C.: The proof and measurement of association between two things, *Am. J. Psychol.*, 15, 72–101, <https://doi.org/10.2307/1412159>, 1904.
- Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (Eds.): Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324>, 2013.
- 795 Trenberth, K. E.: Changes in precipitation with climate change, *Clim. Res.*, 47, 123–138, <https://doi.org/10.3354/cr00953>, 2011.



Trenberth, K. E., Fasullo, J., and Smith, L.: Trends and variability in column-integrated atmospheric water vapor, *Clim. Dynam.*, 24, 741–758, <https://doi.org/10.1007/s00382-005-0017-4>, 2005.

Wilks, S. S.: The large-sample distribution of the likelihood ratio for testing composite hypotheses, *Ann. Math. Stat.*, 9, 60–62, <https://doi.org/10.1214/aoms/1177732360>, 1938.

800