

# Bridging the scale gap: enhancing point-scale rainfall estimates by post-processing ERA5

Fatima M. Pillosu<sup>a,b</sup> Timothy D. Hewson<sup>b</sup> Estibaliz Gascòn<sup>b</sup> Milana Vuckovic<sup>b</sup> Christel  
Prudhomme<sup>b,c,d</sup> Hannah Cloke<sup>a,e</sup>

<sup>a</sup> *Department of Geography and Environmental Science, University of Reading, Reading, UK*

<sup>b</sup> *Forecast and Services Department, European Centre for Medium-range Weather Forecasts,  
Reading, UK*

<sup>c</sup> *Department of Geography and Environment, University of Loughborough, Loughborough, UK*

<sup>d</sup> *UK Centre for Ecology and Hydrology, Wallingford, United Kingdom*

<sup>e</sup> *Department of Meteorology, University of Reading, Reading, UK*

<sup>11</sup> Corresponding author: Fatima Pillosu, tj835682@student.reading.ac.uk

12 ABSTRACT: Accurately estimating rainfall distributions, from small to extreme totals, is crucial  
13 for addressing various environmental challenges, including flood forecasting, water resource man-  
14 agement, and disaster preparedness. Global Numerical Weather Prediction (NWP) models can  
15 provide useful rainfall estimates; yet, they often misrepresent point-scale observations from rain  
16 gauges, underestimating the frequency of small rainfall totals and underestimating extreme val-  
17 ues. This study provides a systematic, global verification of four NWP-modelled rainfall datasets  
18 with different resolutions - ERA5's Ensemble Data Assimilation (62 km, probabilistic), ERA5's  
19 short-range forecasts (31 km, deterministic), short-range ECMWF reforecasts for cycle 46r1 (18  
20 km, control run), and ERA5-ecPoint (point-scale, probabilistic) - against 20 years of point-rainfall  
21 observations from rain gauges around the world. **The analysis focuses on liquid precipitation and**  
22 **is not stratified by season, with observations pooled across all months to characterise the overall**  
23 **climatological distribution at each station.** The models' ability to represent the entire rainfall dis-  
24 tribution, including extreme rainfall, was assessed. Overall, the higher spatial resolution of NWP  
25 models enables a more accurate representation of gauge-based climatologies. Nonetheless, ERA5-  
26 ecPoint provides the most accurate representation, capturing the frequency of zeros, the growth  
27 rates of rainfall totals, and the wet tails more accurately. Moreover, due to its probabilistic nature,  
28 ERA5-ecPoint can estimate long return periods (e.g., 1000 years and more), offering insights into  
29 extremely rare or unprecedented events at specific locations. The model significantly improves  
30 performance in flat, hilly/mountainous regions. In very mountainous areas (e.g., the Andes), it  
31 underestimates zero rainfall totals and overestimates the length of the wet tails. These findings  
32 underscore the importance of using post-processing to enhance the local-scale validity of global  
33 NWP models. Moreover, as climate change intensifies extreme rainfall events, these findings are  
34 crucial for estimating accurate long-period rainfall climatologies, as needed for effective mitigation  
35 and resilience building, particularly in areas lacking comprehensive and reliable rain gauge records.

36 SIGNIFICANCE STATEMENT: The purpose of this study is to better understand how modelled  
37 rainfall datasets represent the distribution of point-scale rainfall as measured by rain gauges,  
38 including extremes. We considered four modelled rainfall datasets, including ERA5 at 62 and  
39 31 km, ECMWF reforecasts at 18 km, and the ecPoint post-processed version of ERA5 (ERA5-  
40 ecPoint), which produces probabilistic point-scale rainfall estimates. All models were evaluated  
41 against twenty years of rain gauge measurements worldwide. Our findings demonstrate that ERA5-  
42 ecPoint largely improves the estimates of point-rainfall, correctly capturing both small and extreme  
43 rainfall totals. This advancement is crucial since many regions lack (appropriate) rain gauge  
44 coverage. Moreover, as climate change intensifies extreme rainfall globally, ERA5-ecPoint enables  
45 planners to quantify rare, unseen events, providing crucial information for infrastructure design  
46 and disaster preparedness.

## 47 1. Introduction

48 Accurately estimating the full range of past and future rainfall distributions, from light to extreme  
49 totals, is one of the biggest challenges in modern meteorology. Yet, it is essential to address a  
50 range of critical issues. In flood forecasting, accurately estimating the spatial distribution of small  
51 and extreme rainfall totals influences the catchment response to the rainfall event, impacting runoff  
52 generation and streamflow patterns (Cuo et al. 2011; Wang and Karimi 2022). In water resource  
53 management, understanding the full rainfall distribution informs the management of reservoirs for  
54 flood control, power generation, and irrigation purposes (Tie et al. 2023). It also helps in agricultural  
55 applications such as crop selection and planting schedules (Janmohammadi and Sabaghnia 2023;  
56 Maurya et al. 2024). It also supports urban planning by helping to design effective urban drainage  
57 systems and manage storm-water runoff (Hossain et al. 2024; Laouacheria et al. 2019). Analysing  
58 changes over time in the entire rainfall distribution provides insights into climate change impacts  
59 such as shifts in the frequency and intensity of extreme rainfall events (Tye et al. 2022), droughts'  
60 characteristics (Haile et al. 2020), biodiversity and ecosystems stability (Lamprecht et al. 2021),  
61 and food security (Balasundram et al. 2023). Extreme rainfall, in particular, has received significant  
62 attention in recent literature (Gimeno et al. 2022; Schumacher 2017) due to its catastrophic impacts  
63 for society, infrastructure, and the environment (IPCC 2023; WMO 2024). It not only reduces  
64 worldwide macroeconomic growth rates and slows global economic rise (Liang 2022), but also

can cause long-term anxiety and post-traumatic stress on affected communities, hindering recovery efforts (Doocy et al. 2013). With climate change expected to intensify both the frequency and severity of extreme rainfall, even in regions where average precipitation is decreasing (Asadieh and Krakauer 2015; Westra et al. 2014; Zittis et al. 2021), understanding its past and anticipating future trends is crucial to inform disaster preparedness and response efforts.

Precipitation time series can be obtained from various sources. Rain gauges are a primary source of ground truth. They provide highly accurate direct point-scale precipitation measurements when properly maintained and calibrated (Lanza and Stagi 2008). In regions with dense networks, rain gauges offer a good spatial representation of localised extremes (Haiden and Duffy 2016). Moreover, stations have been operating for decades in some locations, providing high-quality long-term historical records for trend analysis (Anand and Karunanidhi 2020; Tadeyo et al. 2020). Rain gauge coverage is notably spatially and temporally uneven, leaving many regions unmonitored (Kidd et al. 2017). In areas with complex topography or low-density networks, gauges may fail to represent the rainfall's spatial variability (Di Curzio et al. 2022). Inadequate rain gauge maintenance can also lead to data gaps or inaccuracies (Lanza and Stagi 2008). Furthermore, systematic quality control of gauge records remains challenging, as automated screening procedures may not reliably detect all forms of measurement error, including timing errors, blocked funnels, or spurious accumulations (Wang et al. 2023). Satellite- and radar-derived gridded datasets provide broader spatial and temporal coverage, particularly in ungauged regions (Herold et al. 2017). Their rainfall estimates may, however, differ from rain gauge measurements, especially extremes which might be severely underestimated and mislocated (Ensor and Robeson 2008; Gupta et al. 2020; Satgé et al. 2020). Numerical Weather Prediction (NWP) models, such as reanalyses and reforecasts, offer spatially and temporally consistent precipitation datasets with global, multi-decadal coverage. Reanalyses, like ERA5 and its Ensemble Data Assimilation (EDA) component (Hersbach et al. 2020) or NCEP/NCAR Reanalysis (Hamill et al. 2022; Kalnay et al. 1996), integrate historical weather observations with a state-of-the-art NWP model to produce high-resolution precipitation datasets. Reforecasts, such as NCEP's Global Ensemble Forecast System (Guan et al. 2022) and ECMWF's Integrated Forecast System (Richardson et al. 2014), provide 20-30 years of retrospective forecasts generated with current operational NWP models. Reanalyses capture rainfall's spatial patterns and temporal trends (Lavers et al. 2022) but tend to underestimate

extreme precipitation due to their coarse resolution of about 50 or 30 km (Alexandridis et al. 2023; Donat et al. 2016; Espinosa et al. 2024; Gomis-Cebolla et al. 2023)<sup>1</sup>. Reforecasts also capture rainfall's spatial patterns and temporal trends, but still underestimate extreme precipitation even though their resolution is half, i.e. 18 km (Hewson 2024)<sup>2</sup>. A fundamental challenge therefore persists: NWP grid-box values represent areal averages, whereas rain gauges measure precipitation at a single point, creating a scale gap that must be bridged (*bridging the scale gap*) before modelled and observed climatologies can be meaningfully compared or operationally combined. Statistical post-processing methods can enhance the local-scale representation of rainfall (Giorgos et al. 2024), but their effectiveness commonly depends on the availability of high-quality observations, leading to a patchy geographical coverage of post-processed reanalysis/reforecasts (Vannitsem et al. 2021). The post-processing method proposed by Hewson and Pillou (2021), called ecPoint, improves the local-scale representation of NWP model outputs globally, particularly for extremes, without requiring high-density observational networks, using a non-local calibration strategy. The ecPoint approach was applied to ERA5 for rainfall and temperature through the Highlander project (Hewson et al. 2023; Bottazzi et al. 2024).

The primary aim of this study is to assess the fitness-for-purpose of the ERA5-ecPoint dataset by comparing its representation of point rainfall climatologies around the world against the rain gauge-based equivalent. A secondary goal is to evaluate the impact of spatial resolution on the representation of point rainfall climatologies from three additional datasets: ERA5's Ensemble Data Assimilation (EDA, 62 km **effective grid spacing**), ERA5's short-range forecasts (31 km), and ECMWF 46r1 reforecasts (18 km). Two research questions are, therefore, examined. How do NWP models represent the overall distribution of point-rainfall observations (RQ1)? How do NWP models represent, in particular, extreme rainfall (RQ2)? With a reliable post-processed climatology in hand, one could analyse extreme rainfall trends over long periods (+80 years), place observed or forecast extremes into a climatological context, and assess whether a predicted event has historical precedent or lies beyond the range of physically plausible magnitudes.

---

<sup>1</sup>Note that the studies comparing both ERA5 and ERA5-Land against rain gauge observations are considered in this study only for their analysis of ERA5. These are somewhat flawed as ERA5-Land simply re-grids, without any statistical or dynamical downscaling, the precipitation in ERA5 onto ERA5-Land's grid (Muñoz-Sabater et al. 2021).

<sup>2</sup>Although ECMWF reforecasts are now produced at a horizontal resolution of approximately 9 km, Hewson (2024) demonstrated that the magnitude of precipitation extremes does not increase substantially when moving from the previous 18 km resolution to 9 km, suggesting limited added value at the tail of the distribution from this resolution increase alone.

121 The study is organised as follows. Section 2 describes the rain gauge observations and the  
122 NWP models used in this study. Section 3 describes the methods adopted to answer the research  
123 questions. Section 4 presents the results from the objective verification and a case study, while  
124 Section 5 discusses them. Final remarks are drawn in Section 6.

125 **2. Data**

126 *a. Point-scale rain gauge precipitation observations*

127 This study considered 24-hourly precipitation from surface synoptic observations (SYNOP)  
128 from the Global Telecommunication System (GTS) network and additional gauge data stored  
129 internally at ECMWF. SYNOP observations consist of standardised, historical and near-real-time  
130 meteorological reports that ensure data quality and format consistency across diverse regions. High-  
131 density national rain gauge networks (primarily from European countries and available internally  
132 at ECMWF) were also integrated into the analysis (Haiden and Duffy 2016).

133 Rain gauge networks are subject to well-documented issues, such as gauge undercatch and  
134 localised siting effects (Pollock et al. 2018; Kochendorfer et al. 2020), that cannot be fully eliminated  
135 through post-hoc quality control. The rain gauge observations used in this study underwent a multi-  
136 step quality control procedure prior to analysis. Rainfall frequency distributions were visually  
137 inspected across multiple intensity ranges to identify anomalous features, including suspicious  
138 peaks at regular intervals (e.g., 300, 310, 320 mm) suggestive of encoding or data transmission  
139 errors, artificial clustering at round integer values (e.g., 100, 200, 300 mm), and physically  
140 implausible totals exceeding known world records (1825 mm/24h, La Réunion). Flagged values  
141 were cross-checked against nearby stations (akin to buddy checking) and compared against the  
142 independent CPC Global Unified Gauge-Based Analysis<sup>3</sup>, a gridded product at 50 km resolution  
143 (akin to background checking). The CPC dataset was deliberately selected rather than ERA5 to  
144 preserve the independence of the verification framework. Comparison with the CPC data involved  
145 applying a scaling factor to account for representativeness differences between the gridded product  
146 and point-scale observations. After a sensitivity analysis, the scale factor of 20 was selected  
147 to clean erroneous rainfall observations. The procedure successfully removed the majority of  
148 anomalous features whilst yielding distributions more consistent with expected climatological

---

<sup>3</sup><https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html>

133 TABLE 1. Characteristics (columns 1 to 5) of the observational rainfall dataset (row 1) and the considered  
 134 NWP models (rows 2 to 5), as well as their derived climatologies (columns 6 to 8). It is worth noting that the  
 135 unavailable data for ERA5-EDA (row 2, column 6) is due to sporadic damage to the storage tapes on which the  
 136 original data were archived (MARS archive). These losses occurred at irregular intervals with no systematic  
 137 temporal or spatial pattern. There is therefore no reason to expect that the missing data would introduce bias into  
 138 the derived rainfall climatology.

No. Col	1	2	3	4	5	6	7	8
No. Row	Type of climatology	Dataset to compute climatology	Dataset description	Spatial coverage & resolution at equator	Temporal Coverage (20-year return period)	No. of independent realizations	Max return period that can be computed in the 20-year period	Max percentile (%) that can be computed $100 - \frac{100}{\text{no.total real.}}$
1	Observational, for points	SYNOP + regional datasets at ECMWF	Rain gauges	Global (patchy), point scale	01/01/2000 to 31/12/2019	1 daily realization (real.) - 1 real. X 365 days x 20 years = 7300 total real.	(7300 real. x 0.75*) / 365 days = <b>1 in a 15-year event</b> <i>* required at least 75% of valid obs.</i>	99.98174
2	NWP-modelled, gridded	ERA5-Ensemble Data Assimilation (ERA5-EDA)	Reanalysis, probabilistic (10 ensemble members)	Global, 62 km	01/01/2000 to 31/12/2019	10 daily real. - 10 real. X 365 days x 20 years = 73000* total real. <i>* we had only 66940 real. as some dates were not available</i>	66940 real. / 365 days = <b>1 in a 183-year event</b>	99.99851
3	NWP-modelled, gridded	ERA5	Reanalysis, deterministic	Global, 31 km	01/01/2000 to 31/12/2019	1 daily real. - 1 real. X 365 days x 20 years = 7300 total real.	7300 real. / 365 days = <b>1 in a 20-year event</b>	99.98630
4	NWP-modelled, gridded	46r1 ECMWF Rforecasts (reforecast_46r1)	Reforecasts, probabilistic (10 ensemble members, up to day 10). Used only control run	Global, 18 km	Past 20 years for period between 01/07/2019 and 30/06/2020 - Reforecasts run only on Mondays and Thursdays	10 daily real. (all lead times for 2 control runs a week were used as daily independent real.) - 10 real. X 2 runs X 52 weeks x 20 years = 20800 total real.	20800 real. / 365 days = <b>1 in a 56-year event</b>	99.99519
5	NWP-modelled, gridded	ERA5-ecPoint	Reanalysis, Probabilistic (99 ensemble members)	Global, at point scale, but provided on the ERA5's grid	01/01/2000 to 31/12/2019	99 daily real. (all ensemble members were used as daily independent real.) - 99 real. X 365 days x 20 years = 722700 total real.	722700 real. / 365 days = <b>1 in a 1980-year event</b>	99.99986

155 behavior. Notwithstanding these efforts, we acknowledge that some observational errors may  
156 remain undetected. However, the relative comparison between datasets is less sensitive to such  
157 biases, given the common observational reference used throughout.

158 Rain gauge observations stored at ECMWF have increased considerably since the 2000s. Thus,  
159 we consider a 20-year verification period between the 1<sup>st</sup> of January 2000 to the 31<sup>st</sup> of December  
160 2019. Since the verification in this study is conducted at the climatological level — comparing  
161 full distributions of 24-hourly accumulations rather than matching individual events on a timestep-  
162 by-timestep basis — the analysis was not restricted to a single accumulation window. For each  
163 station, 24-hourly rainfall climatologies were computed using the accumulation window consistent  
164 with the station's reporting time (e.g., 00–00, 01–01, 02–02 UTC, and so on). This approach  
165 maximises spatial coverage by retaining stations in regions where reporting times do not align with  
166 the 00 UTC convention, such as parts of South Asia and Australasia. The resulting station-level  
167 climatologies were then pooled into a single verification database. This aggregation is considered  
168 defensible because, over a 20-year record, the statistical distribution of 24-hourly totals at a given  
169 location is not expected to vary substantially with modest shifts in the accumulation window,  
170 although this assumption may be less robust in regions with strongly diurnal precipitation regimes.  
171 In total, the verification database comprises 7300 potential daily realisations per station within the  
172 20-year period (Table 1, row 1). Many rain gauge stations had missing data. To ensure that the  
173 timeseries were representative of the considered 20-year period, only sites with at least 75% of  
174 valid recordings were considered, which reduced the number of sites in the database from 28834  
175 to 4546.

176 *b. Gridded NWP-modelled precipitation estimates*

177 1) ERA5 REANALYSIS (ERA5) AND ERA5 ENSEMBLE DATA ASSIMILATION (ERA5-EDA)

178 ERA5 is the fifth generation of atmospheric reanalysis produced by the Copernicus Climate  
179 Change Service (C3S) run by ECMWF (Hersbach et al. 2020). Compared to its predecessor, ERA-  
180 Interim, ERA5 offers high spatial ( 31 km) and temporal (hourly) resolution and extended temporal  
181 coverage from 1940 to near-real time. ERA5 assimilates a diverse range of observational data from  
182 satellites, weather balloons, aircraft, and ground stations, employing a 4D-Var assimilation system.  
183 This system not only improves the accuracy of the data by adjusting it in four dimensions but also

184 enhances the continuity and stability of the climatological records. No precipitation observations  
185 are assimilated into ERA5 (Hersbach et al. 2020).

186 The ERA5 Ensemble Data Assimilation (EDA) system enhances the robustness of the ERA5  
187 reanalysis by generating multiple simulations with slightly varied initial conditions (Hersbach et al.  
188 2020). Each ensemble member in ERA5 EDA provides an equally probable realisation of the  
189 atmospheric state, quantifying the uncertainty associated with observational errors and limitations  
190 within the forecasting model itself. ERA5-EDA has 10 ensemble members, running at 62 km  
191 spatial resolution and 3-hour temporal resolution.

192 To match the rain gauge observations, ERA5 and ERA5-EDA data between the 1<sup>st</sup> of January 2000  
193 and the 31<sup>st</sup> of December 2019 were extracted, and only 24-hourly precipitation ending at 00 UTC  
194 was considered. Hence, ERA5 precipitation distribution was built with 7300 realisations, while  
195 ERA5-EDA distribution, considering the 10 ensemble members as equally probable precipitation  
196 realisations, was constructed with 73000 realisations (Table 1, rows 2 and 3).

## 197 2) ECMWF REFORECASTS

198 Reforecasts are retrospective weather forecasts generated with a fixed NWP model version. The  
199 reforecast uniformity (i.e., with no discrepancies caused by historical changes in model configu-  
200 rations) ensures that differences in climatological patterns are attributable to actual atmospheric  
201 variations rather than artefacts of evolving model technologies. To match the temporal span of the  
202 precipitation observations as closely as possible, reforecasts from the ECMWF's IFS 46r1 cycle  
203 were considered - since 46r1 run operationally from June 2019 to June 2020, the reforecasts span  
204 from the 1st of July 1999 to the 30th of June 2019. 46r1 reforecasts are provided at 18 km spatial  
205 resolution, and are produced only on Mondays and Thursdays. They consist of an ensemble of  
206 one control run and 10 perturbed members, produced at 00 UTC with a 6-hourly resolution up to  
207 t+1104 (day 46). The control and the perturbed members' model configurations (e.g., resolution,  
208 parametrisations) are the same. However, the control run uses the best estimate of the initial  
209 conditions (i.e., the operational analysis), and it has been shown to have a different precipitation  
210 climatology than the perturbed members. Hence, in this study, only the control run was used.  
211 Since reforecasts have fewer realisations per year (as they are produced only on Mondays and  
212 Thursdays), we increased the precipitation realisations by considering lead times up to day 10 as

213 equally probable precipitation realisations. This was possible as there was no drift in the forecasts  
214 up to day 10 (not shown). Hence, the precipitation distribution built with ECMWF reforecasts  
215 contains 20800 realisations (Table 1, row 4).

216 3) ERA5-ecPOINT

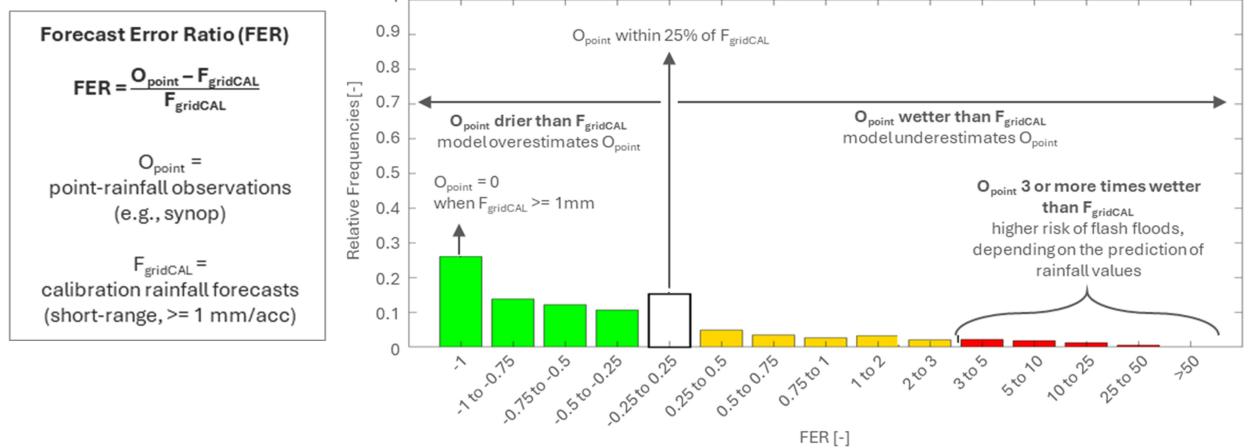
217 ERA5’s raw rainfall reanalysis exhibits known limitations in representing point-rainfall, rain-  
218 gauge-like estimates. As any other gridded model output, ERA5 provides a rainfall estimate that  
219 represents the average of point values over the grid-box area. ecPoint is a statistical post-processing  
220 technique that transforms global gridded NWP outputs into probabilistic point-scale estimates (?).  
221 The post-processing technique aims to provide post-processed estimates that mirror observations  
222 from rain gauges by addressing the two main factors affecting the performance of global NWP model  
223 outputs against point verification: systematic biases (Lavers et al. 2021) and lack of representation  
224 of sub-grid variability (Göber et al. 2008). The errors between global gridded rainfall forecasts  
225 (i.e., up to t+30, control run of ECMWF’s ENS) and point-rainfall observations (i.e., rain gauges)  
226 are computed for a one-year calibration period. The error computed for accumulated variables (like  
227 rainfall) is the Forecast Error Ratio (FER), whose formulation is shown in Figure 1a. The error  
228 distribution is named Mapping Function (MF), and its shape is linked to the degree of sub-grid  
229 variability and biases at grid scale in the raw forecasts. The MF for all data points in the calibration  
230 period is also shown in Figure 1a, and it shows that ECMWF’s ENS both overestimates (green  
231 bars) and underestimates (yellow and red bars) versus gage reports. The white bar indicates that  
232 only ~15% of the point-rainfall observations were correctly predicted.  
233

234 The MF is used to post-process the raw rainfall outputs from NWP models. Suppose all grid-  
235 boxes in the raw forecasts are post-processed by sampling only the MF in Figure 1a (also the  
236 same MF in the black circle in Figure 1b). In this case, the ecPoint post-processing would follow  
237 a univariate approach (U-ecPoint), and be thought of as a single-leaf decision tree (Figure 1b).  
238 U-ecPoint generally increases the number of zeros in the distribution of point-rainfall forecasts to  
239 correct ENS’s tendency to overpredict small rainfall totals (Haiden et al. 2025). It also increases  
240 the rainfall amounts in the wet tail of the rainfall distribution to correct for ENS underestimation  
241 of high rainfall values (Haiden et al. 2025). The MF shape can, however, change significantly  
242 according to different weather scenarios at grid-scale (called Grid-box Weather Type, G-WT). The  
243

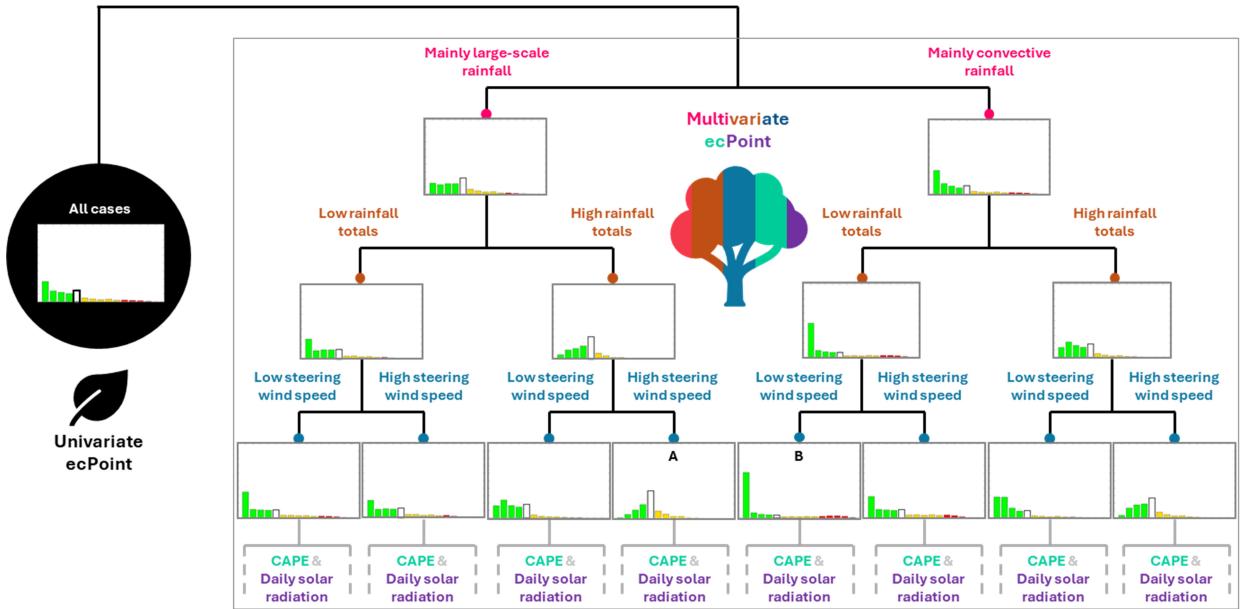
## Graphical representation of the ecPoint post-processing technique

Mapping function and decision tree

(a) Error formulation for accumulated variables (Forecast Error Ratio, FER), and errors' distribution for all cases in the training dataset (Mapping Function, MF)



(b) Univariate and multivariate ecPoint represented, respectively, as a "single-leaf" and "multiple-leaf" decision tree (DT)



217 FIG. 1. Panel (a) shows the error formulation for accumulated variables (Forecast Error Ratio, FER) and  
 218 the error distribution for all cases in the training dataset (Mapping Function, MF). The example pertains to  
 219 the calibration of 47r3 ECMWF ENS forecasts for 12-hourly rainfall forecasts. Panel (b) shows the univariate  
 220 approach for ecPoint (U-ecPoint) represented as a "single-leaf" decision tree (DT, within the black circle), while  
 221 the multivariate approach (M-ecPoint) is represented as a "multiple-leaf" DT (within the grey square).

multiple MFs can be visualised with a decision-tree-like representation, where each leaf of the decision tree corresponds to a G-WT and its corresponding MF (Figure 1b). When each grid-box in the raw forecast is post-processed differently by using the MF corresponding to the matching G-WT in the decision tree (within the grey square in Figure 1b), the ecPoint post-processing follows a multivariate approach (M-ecPoint). It corresponds to the original ecPoint system developed by ? . When for a grid-box, the raw ENS predicts high totals of mainly large-scale rainfall and strong steering wind speeds (case A in Figure 1b), the MF takes a Gaussian-like form. This means the raw model output is relatively representative of the point-scale rainfall totals. When the raw ENS predicts mainly convective rainfall with light steering wind speeds (case B in Figure 1b), the MF might take an exponential-like form. This means that the raw model output is not representative of the point-scale rainfall totals and that the expected degree of sub-grid variability is bigger than in case A. Each raw forecast is converted into a distribution of N point-scale forecasts using the MFs (for example, operationally, for each raw ensemble member, N=100 point-scale forecasts are created). Hence, while M-ecPoint increases overall the frequency of small and large rainfall totals in the post-processed forecasts, as U-ecPoint does, its adjustments are applied according to different G-WTs. M-ecPoint reduces the probabilities at certain locations more than U-ecPoint; this relates to the fact that corrections are applied differently across the ensemble members rather than uniformly, as done by U-ecPoint. Moreover, Hewson and Pillosu (2021) have shown that, due to the G-WT differentiation in the corrections, one of M-ecPoint's features is the ability to shift the location of areas at higher risk of extreme localised rainfall. This feature is lost in U-ecPoint as all grid-boxes are post-processed identically (Pilosu et al. 2025).

The ecPoint post-processing technique has been applied to ERA5 reanalysis within the Highlander project, co-financed by the EU and coordinated by Italy's Cineca supercomputing centre (Bottazzi et al. 2024). The deterministic realisations of ERA5 reanalysis are transformed to a distribution of 100 point-rainfall totals, and distilled in 99 percentiles, i.e. in percentiles 1,2,... 99. Currently, the ERA5-ecPoint dataset spans from 1950 to the near-present, providing a long-term, continuously updated record of 24-hourly point-scale rainfall estimates with an accumulation period ending at 00 UTC. They are provided in the same native grid of ERA5 (reduced Gaussian grid N320, 31 km). The 99 percentiles can be considered equally probable precipitation outcomes, at a gauge within

276 a grid-box, so that the precipitation distributions are built with 722700 daily realisations (Table 1,  
277 row 5).

278 **3. Methods**

279 *a. RQ1: assessment of the representation by NWP models of the overall distribution of point-  
280 rainfall observations*

289 The assessment of how well NWP models represent the overall distribution of point-rainfall  
290 observations is conducted by assessing the similarity between observed and the NWP-modelled  
291 rainfall distributions over the 20 years. The modelled estimates are extracted at the rainfall  
292 observation locations, considering the modelled value at the nearest grid-box. This approach is  
293 regarded as standard practice for rainfall, as interpolation may reduce extremes. This study adopts  
294 the method proposed by Gudmundsson et al. (2012), which assesses the similarity between the  
295 Empirical Cumulative Distribution Functions (ECDFs) of the observed and NWP-modelled rainfall  
296 estimates, constructed with the empirical percentiles (Boé et al. 2007). The similarity is assessed  
297 by averaging the Mean Absolute Errors (MAE) at corresponding  $x^{\text{th}}$  percentiles:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\text{tp}_{\text{OBS}}(x_i^{\text{th}}) - \text{tp}_{\text{NWP}}(x_i^{\text{th}})| \quad (1)$$

298  $n = 99$  percentiles

299 MAE values are expressed in mm and range from 0 (for perfect similarity) to  $+\infty$  (for poor  
300 similarity). Graphically, the MAE represents the areal difference between the two ECDFs (Figure  
301 2a)<sup>4</sup>. Ninety-nine percentiles (1st to 99th) were used to avoid variations related to sampling issues  
302 in the observational dataset. Moreover, more extreme rainfall events will be considered separately.

303 Gudmundsson et al. (2012) methodology was, however, adapted to compare ECDFs from different  
304 climatologies. MAE values were normalised ( $\text{MAE}_{\text{NORM}}$ ) to avoid having consistently bigger MAE  
305 values in wetter climates (see Figure 2b and Figure 2c). The normalisation consists of computing

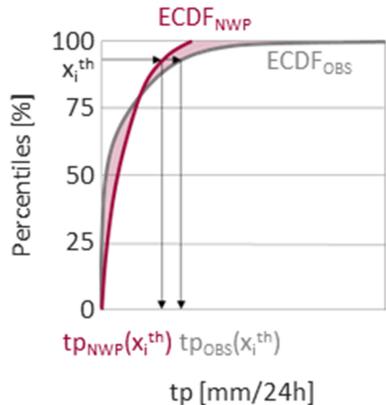
---

<sup>4</sup>The Mean Error (ME) or Bias could have also been used to measure the similarity between the ECDFs. These two measures, although complementary, can, however, provide very different pictures, as we could have big MAEs while the MEs could be very small if they cancel each other. Moreover, the ME has already been computed for ERA5 by Lavers et al. (2022) to assess its performance in climate monitoring. Results from both studies will, however, be compared in the discussion section

## Normalized Mean Absolute Error ( $MAE_{NORM}$ ) for 24-hourly total precipitation (tp)

Schematic on how to interpret  $MAE_{NORM}$

(a) Schematic representation and computation of MAE and  $MAE_{NORM}$  at a specific rain gauge



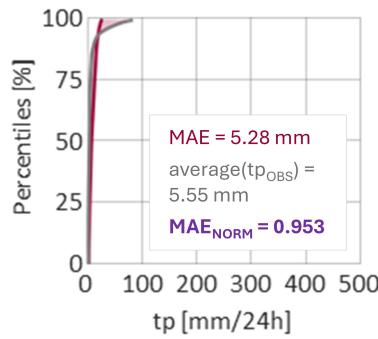
$$MAE = \frac{1}{n} \sum_{i=1}^n |tp_{OBS}(x_i^{\text{th}}) - tp_{NWP}(x_i^{\text{th}})|$$

$n = \text{tot n. of percentiles sampling ECDF}$

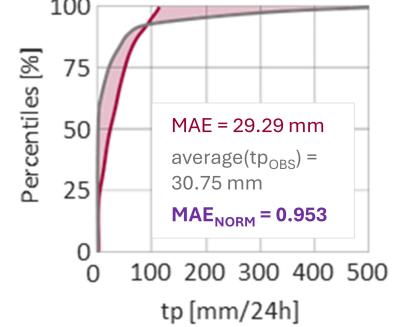
MAE is expressed in mm

$$MAE_{NORM} = \frac{MAE}{\text{average}(tp_{OBS})}$$

(b) Example of MAE and  $MAE_{NORM}$  in drier climates



(c) Example of MAE and  $MAE_{NORM}$  in wetter climates



$$MAE_{NORM} = \frac{MAE}{\text{average}(tp_{OBS})}$$

$$\text{average}(tp_{OBS}) = \frac{1}{m} \sum_{i=1}^m (tp_{OBS})_i$$

$m = \text{tot n. of observed records in the rain gauge}$

$MAE_{NORM}$  is adimensional

(d) Examples of  $MAE_{NORM}$  values for different degrees of similarity between ECDFs

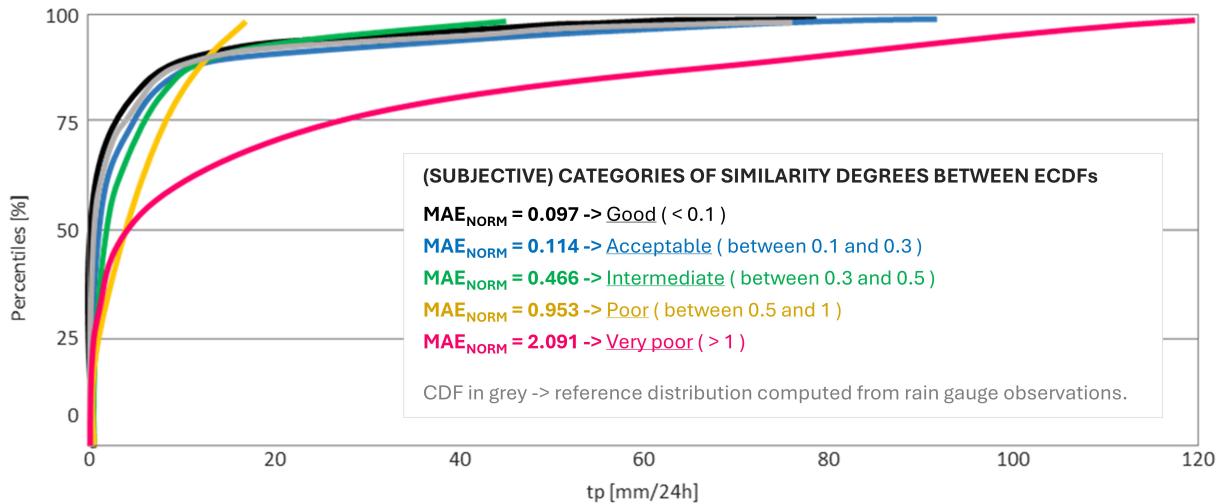


FIG. 2. Panel (a) shows a schematic representation and computation of the Mean Absolute Error (MAE) and the Normalised Mean Absolute Error ( $MAE_{NORM}$ ) for total precipitation (tp). Panels (b) and (c) show how the values of MAE change for drier and wetter climates, respectively, and how dividing them by the average of the observed precipitation helps to normalise the MAE values for different climatologies. Panel (d) shows examples of  $MAE_{NORM}$  values for different degrees of similarity between ECDFs: black, blue, green, yellow, and pink represent, respectively, a "good" (less than 0.1), "acceptable" (between 0.1 and 0.3), "intermediate" (between 0.3 and 0.5), "poor" (between 0.5 and 1), and "very poor" degree of similarity (greater than 1). The observational distribution is shown in grey.

306 dimensionless coefficients by dividing MAE by the corresponding station's average observed  
307 rainfall:

$$\text{MAE}_{\text{NORM}} = \frac{\text{MAE}}{\text{average}(\text{tp}_{\text{OBS}})} \quad (2)$$

308 To guide the reader on what was considered by the authors a better or worse similarity degree,  
309 Figure 2d shows a selection of  $\text{MAE}_{\text{NORM}}$  values for different similarity degrees between ECDFs.  
310 Four categories of similarity degrees were subjectively defined:  $\text{MAE}_{\text{NORM}}$  values below 0.1, in  
311 black, indicate a "good" degree of similarity,  $\text{MAE}_{\text{NORM}}$  values between 0.1-0.3, in blue, indicate  
312 an "acceptable" degree of similarity,  $\text{MAE}_{\text{NORM}}$  values between 0.3-0.5, in green, indicate an  
313 "intermediate" level of similarity degree,  $\text{MAE}_{\text{NORM}}$  values between 0.5-1, in yellow, indicate a  
314 "poor" degree of similarity, and  $\text{MAE}_{\text{NORM}}$  values greater than 1, in pink, indicate a "very poor"  
315 degree of similarity.

316 Formal statistical tests such as Kolmogorov-Smirnov, Cramer-von-Mises, and Anderson-Darling  
317 can also assess the similarity between two distributions (Stephens 1974). However, for large sample  
318 sizes (as in this study, see Table 1, row 6), the tests' statistical significance levels become extremely  
319 sensitive to minor differences between distributions that might not be practically significant (Eng-  
320 mann and Cousineau 2011; Janssen 2000). This is being referred to as "the problem of practical  
321 insignificance" (Kirk 1996), where the test flags differences that are statistically significant but not  
322 meaningful in practice, causing the rejection of the null hypothesis (i.e., the two samples come  
323 from the same population) when it is nonetheless practically valid. Gudmundsson's approach was  
324 tested to assess whether it was as sensitive to sample size as the formal statistical tests. The ECDFs  
325 were sampled with 99, 999, 9999, and 99999 percentiles to assess the sensitivity of  $\text{MAE}_{\text{NORM}}$  to  
326 the choice of sampling resolution. The results showed negligible differences across the range of  
327 percentiles tested (not shown). Moreover, Gudmundsson's approach assesses similarity between  
328 the observed and NWP-modelled rainfall distributions by comparing the whole ECDFs differently  
329 to other formal tests that assess similarity only for specific moments of the distribution, such as  
330 the mean, standard deviation, skewness, or specific percentiles (Anthanahalli Nanjegowda and  
331 Kulamulla Parambath 2022). Finally, the use of MAE concerning other commonly used scores,  
332 such as the Root Mean Squared Error (RMSE), is preferable as it is not unduly sensitive to outliers  
333 (e.g., caused by erroneous observations or atypical events), typically observed in the wet tails of

334 the distribution. Hence, MAE should be more representative of the distribution as a whole (Jolliffe  
335 and Stephenson 2011). Moreover, the RMSE is more appropriate when errors follow a normal  
336 distribution, which is very atypical for rainfall (Chai and Draxler 2014). Nonetheless, the RMSE  
337 was computed for the examples shown in Figure 2d, and its property of giving more weight to the  
338 larger errors (in the wetter part of the distribution) did not change the ranking obtained with the  
339 MAE<sub>NORM</sub> between the different CDFs in 2d (not shown), reassuring the reader that using MAE  
340 instead of RMSE should not change the final picture.

341 Maps plotting MAE<sub>NORM</sub> values at different rain gauge locations are shown to compare the  
342 performance of the four analysed NWP models. The maps are accompanied by pie charts that  
343 summarise, for specific regions, the percentage of locations falling in the five MAE<sub>NORM</sub> categories  
344 defined in Figure 2d (< 0.1, between 0.1 and 0.3, 0.3 and 0.5, 0.5 and 1, and > 1). The regions  
345 considered are North America, South America, Europe, the Mediterranean, Africa, the Arabian  
346 Peninsula, Asia, and Oceania.

347 Finally, a selection of representative ECDFs for all four models against their corresponding  
348 observed point-scale precipitation distributions is also shown, [to illustrate how the agreement](#)  
349 [between modelled and observed rainfall climatologies varies across contrasting physiographic set-](#)  
350 [tings. Each ECDF is computed at a single grid point, and sites were selected to represent four broad](#)  
351 [environment types: flat terrain, hilly to mountainous terrain with moderate orographic complexity,](#)  
352 [very mountainous terrain \(e.g., The Andes, The Rocky Mountains, The Alps, The Himalayas\), and](#)  
353 [desert. These descriptors are qualitative labels reflecting the dominant physiographic and climatic](#)  
354 [character of each site rather than categories defined by formal elevation or aridity thresholds. To](#)  
355 [ensure that the selected locations are not anomalous, ECDFs at numerous additional grid points](#)  
356 [within each environment type were visually inspected; the sites shown were retained because their](#)  
357 [distributional shapes are broadly representative of their respective categories.](#)

### 358 b. RQ2: assessment of the representation by NWP models of extreme rainfall

359 To assess how well each modelled dataset captures observed rainfall extremes, we compare large  
360 return periods derived from the observational and modelled climatologies. Such estimates are  
361 inherently uncertain because the events of interest, i.e., those in the far tail of the distribution, are  
362 poorly sampled. Regional pooling increases the effective sample size at a point by assuming a

common frequency distribution across nearby stations (Hosking and Wallis 1997). This technique may reduce the uncertainty in the computation of large return periods (Kim et al. 2025). However, rainfall climatologies may vary substantially over short distances in regions of complex orography, coastlines, or convective regimes. Hence, pooling risks smoothing the very local differences this study seeks to assess (Khaliq et al. 2006; Ahmed et al. 2023). Return periods can also be computed using extreme value theory, by fitting parametric distributions such as the generalised extreme value or the generalised Pareto (Coles 2001). Although commonly applied in the literature (Acero et al. 2018; Osei et al. 2021), assumptions on the form of the distribution's tail can substantially alter return period estimates, and parameter uncertainty may grow rapidly when fitted to short records (Scarrott and MacDonald 2012; Pan et al. 2022). This study adopts instead the non-parametric empirical approach of Zsoter et al. (2020), developed for reforecast-based flood thresholds in the Global Flood Awareness System, which avoids distributional assumptions. Since no regional pooling is applied, the uncertainty around the resulting estimates, particularly at the largest computable return periods, is quantified via bootstrapping with replacement over 1,000 repetitions.

For each grid point (in the modelled datasets) and each station location (in the observed dataset), we rank all available independent daily realisations in descending order. The return period associated with the  $m^{\text{th}}$  largest value in the sample of  $N$  independent realisations is estimated as:

$$T = \frac{N}{m \times 365} \quad [\text{years}]$$

where the factor 365 converts daily realisations to years.  
The percentile associated with rank  $m$  is:

$$P = 100 \times \left(1 - \frac{m}{N}\right) \quad [\%]$$

To compute the maximum return period for the considered dataset (rank  $m = 1$ , i.e., the largest value in the distribution), the equation b and b transforms to:

389

$$T = \frac{N}{1 \times 365} \quad [\text{years}] \quad (3)$$

390

391

$$P = 100 \times \left(1 - \frac{1}{N}\right) \quad [\%] \quad (4)$$

392

393 For each dataset, column 7 in Table 1 shows the worked examples of what are the maximum  
 394 return period (in years) that can be computed in the 20-year verification period considered in this  
 395 study (i.e., between 2000 and 2019); column 8 shows the corresponding maximum percentile (in  
 396 %). The 10-year return period (computed from the 99.9726<sub>th</sub> percentile of the rainfall realisations)  
 397 is the biggest common extreme event that can be computed for all datasets. The sensitivity of  
 398 return period estimates to sampling variability was examined using a bootstrap-based instability  
 399 analysis (not shown), which indicates that the 10-year return period is generally well constrained  
 400 across the majority (70%) of locations. Therefore, the 10-year return period provides a suitable  
 401 choice for analysing how NWP modelled rainfall estimates represent localised extreme rainfall  
 402 totals. Maps of the observed and modelled 10-year return period are presented for each dataset.  
 403 The percentage of locations where the modelled value exceeds the observed value is computed.  
 404 If a model is unbiased, this percentage should be approximately 50%, since overestimation and  
 405 underestimation are equally likely (Wilks 2020). Substantial departures from this proportion  
 406 indicate systematic bias in the representation of extremes: percentages smaller than 50% indicate  
 407 that modelled return periods underestimate the observed ones, whereas percentages larger than  
 408 50% indicate overestimation.

409 A case study of widespread flash floods in Italy complements the general global comparison be-  
 410 tween observed and NWP-modelled extreme precipitation, using 24-hourly precipitation estimates.  
 411 Italy offers the highest-spatial-resolution rain gauge network of all countries in the database, a prop-  
 412 erty vital for case-study-based analysis of extreme rainfall events, as it increases the likelihood of  
 413 capturing extreme localised totals. For ERA5, only deterministic rainfall estimates are available.  
 414 The ERA5-EDA and ECMWF Reforecast ensembles each comprise 10 members, meaning the 90<sub>th</sub>  
 415 is the highest computable percentile. It corresponds to an event occurring on average 37 days per

416 year, which does not constitute an extreme. The Control run is therefore preferred as it constitutes  
417 the member with the best model representation. For ERA5-ecPoint, the 99<sub>th</sub> percentile is selected.  
418 Given its probabilistic characteristic (i.e., the plotted rainfall totals correspond to an event with a  
419 1% chance of being exceeded at any given location), an accompanying map indicating the locations  
420 where observed rainfall exceeded the ecPoint 99th percentile is also shown.

## 421 4. Results

### 422 a. RQ1: comparison of rainfall distribution climatologies

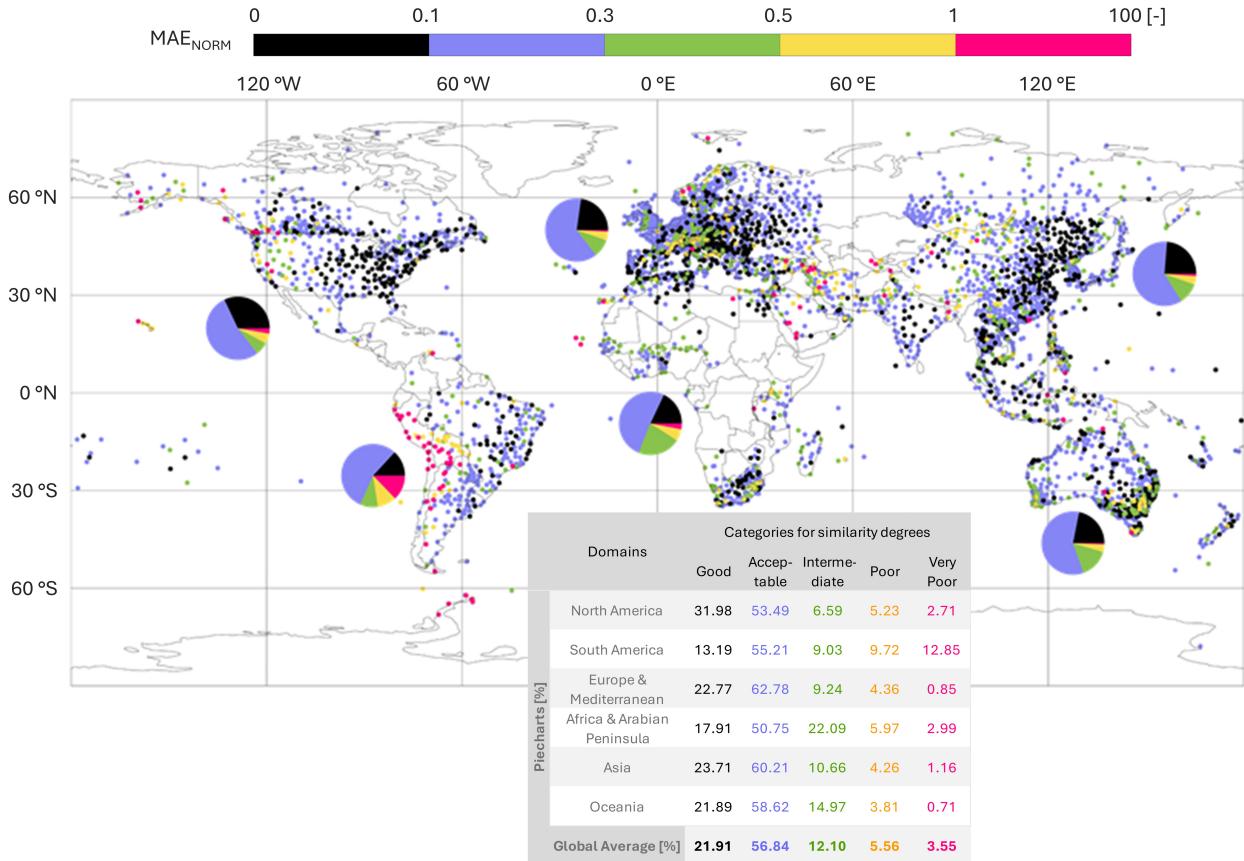
432 Out of all NWP-modelled precipitation estimates, ERA5-ecPoint reproduces observed point-  
433 precipitation distributions best. This can be seen by the larger percentage of small MAE<sub>NORM</sub>  
434 values (black dots) in Figure 3b compared to Figures 4b (for ERA5-EDA), ??b (for ERA5), and  
435 6b (for ECMWF Reforecasts for 46r1), where there are bigger percentages of larger MAE<sub>NORM</sub>  
436 values (depicted by the coloured dots). ERA-ecPoint increases the number of MAE<sub>NORM</sub> values  
437 in the category of "good similarity" (dots in black) by a factor of 10, 30, and 27 compared to the  
438 reforecasts, ERA5, and ERA5-EDA, respectively (see piecharts and inserted tables in Figures 3b to  
439 6b). In the baseline NWP models (ERA5-EDA, ERA5, and ECMWF reforecasts), the proportion of  
440 grid points with the very high similarity between observed and modelled precipitation ("black" dots)  
441 remains consistently low, below 2% in most regions, except in North America, where reforecasts  
442 reach about 4%. In South America, the raw NWP models do not yield any such high-similarity  
443 points, whereas applying ERA5-ecPoint boosts this proportion to 13%, with representation along  
444 Brazil's eastern coast looking particularly good, in relative terms. At the opposite extreme,  
445 points with poor similarity ("pink" dots) are substantially reduced when using ERA5-ecPoint.  
446 Compared to ERA5-EDA, the number of these poorly performing points declines by about 60%,  
447 and relative to ERA5 and reforecasts, by about 50%. These improvements are most pronounced in  
448 the Arabian Peninsula, Asia, and North America. Although reforecasts also have a lower count of  
449 poorly performing points in these areas, they exhibit slightly worse performance in parts of South  
450 America, especially the Bolivian Amazon, increasing the proportion of poor-similarity points by  
451 2% and 5% relative to ERA5-EDA and ERA5, respectively. In contrast, ERA5-ecPoint markedly  
452 improves this situation in South America, reducing poor-similarity points by 47%, 23%, and 10%  
453 compared to reforecasts, ERA5, and ERA5-EDA, respectively. Much of this improvement occurs

**Normalized Mean Absolute Error (MAE<sub>NORM</sub>) for 24-hourly total precipitation  
ECMWF ecPoint (point-scale in 31 km)**

a) Domains to create the piecharts



(b) MAE<sub>NORM</sub> values at each rain gauge station (map plot), aggregated per domain (piecharts), and piecharts' numerical values (table)



423 FIG. 3. Panel (a) defines the geographical domains (North America, South America, Europe and Mediterranean,  
 424 African and Arabian Peninsula, Asia, and Oceania) used to build the piecharts in the following panel. Panel  
 425 (b) shows the Normalised Mean Absolute Error (MAE<sub>NORM</sub>) for 24-hourly total precipitation at each rain gauge  
 426 location for ERA5-ecPoint (point-scale in the ERA5 grid at 31 km spatial resolution). Dots in black, blue, green,  
 427 yellow, and pink represent, respectively, a "good", "acceptable", "intermediate", "poor", and "very poor" degree  
 428 of similarity to the corresponding point observed climatology. The pie charts indicate the frequency (in %) of  
 429 MAE<sub>NORM</sub> values in the domains defined in panel (a). The inserted table offers the numerical representation of  
 430 the pie charts in each geographical domain and in each similarity category. The numbers in bold in the last row  
 431 represent the global average for each similarity category.

## Normalized Mean Absolute Error (MAE<sub>NORM</sub>) for 24-hourly total precipitation

ERA5-EDA (62 km)

a) Domains to create the piecharts



(b) MAE<sub>NORM</sub> values at each rain gauge station (map plot), aggregated per domain (piecharts), and piecharts' numerical values (table)

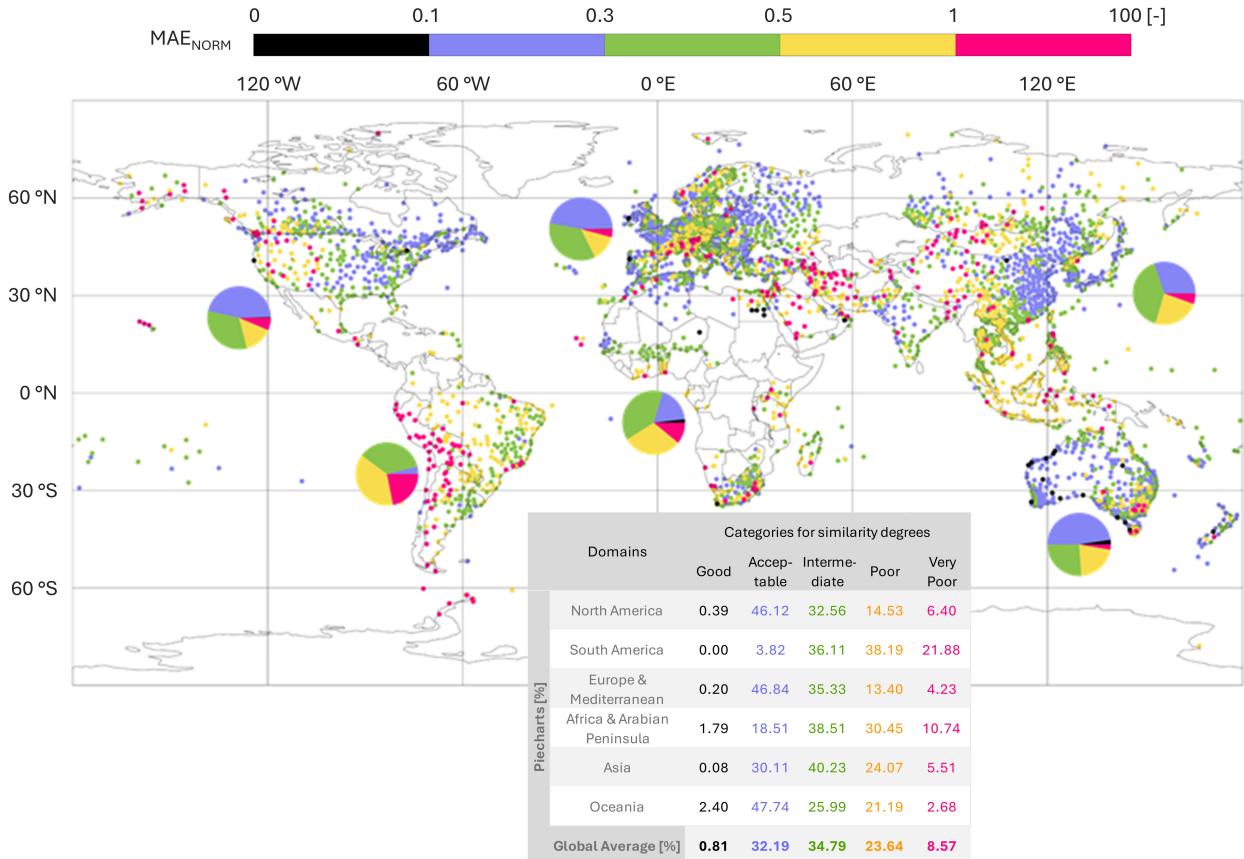


FIG. 4. Like Figure 4 but for ERA5-EDA (at 62 km spatial resolution).

in the flatter Amazonian regions east of the Andean highlands. Still, even with ERA5-ecPoint, some challenging areas remain, such as the Andean slopes and the narrow desert-like coastlines of Peru and Chile. For intermediate similarity levels (previously represented by "blue", "green", and "yellow" categories), the application of ERA5-ecPoint consistently shifts conditions toward a higher level of agreement across all domains. This results in fewer points showing poor similarity

## Normalized Mean Absolute Error (MAE<sub>NORM</sub>) for 24-hourly total precipitation

ERA5 (31 km)

a) Domains to create the piecharts



(b) MAE<sub>NORM</sub> values at each rain gauge station (map plot), aggregated per domain (piecharts), and piecharts' numerical values (table)

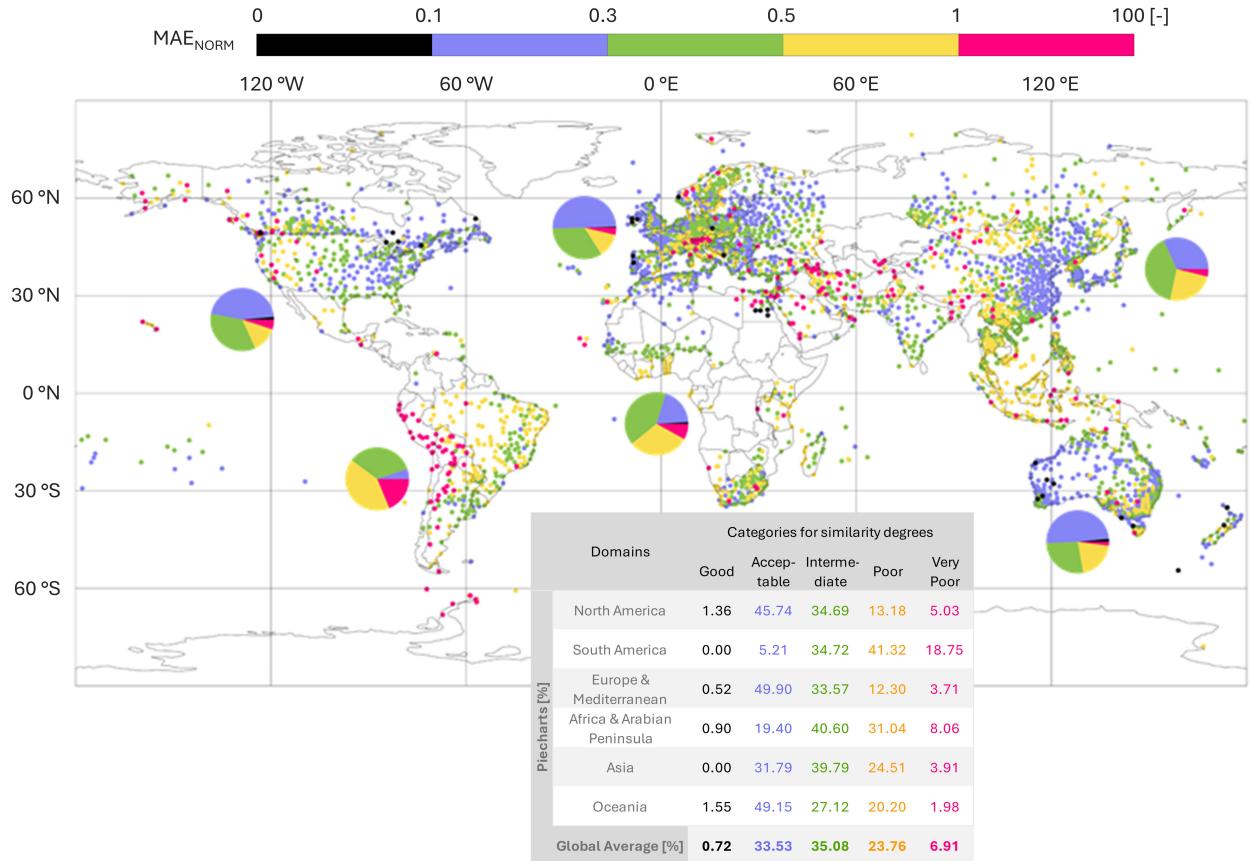


FIG. 5. Like Figure 4 but for ERA5 (at 31 km spatial resolution).

and more points reaching acceptable or good similarity levels. The improvements are especially apparent in South America, Africa, Asia, and Oceania, where ERA5-ecPoint generally transitions more points into categories reflecting moderate to good agreement, thereby offering a notably better representation of precipitation patterns than the baseline NWP models.

## Normalized Mean Absolute Error (MAE<sub>NORM</sub>) for 24-hourly total precipitation

ECMWF Reforecasts – 46r1 (18 km)

a) Domains to create the piecharts



(b) MAE<sub>NORM</sub> values at each rain gauge station (map plot), aggregated per domain (piecharts), and piecharts' numerical values (table)

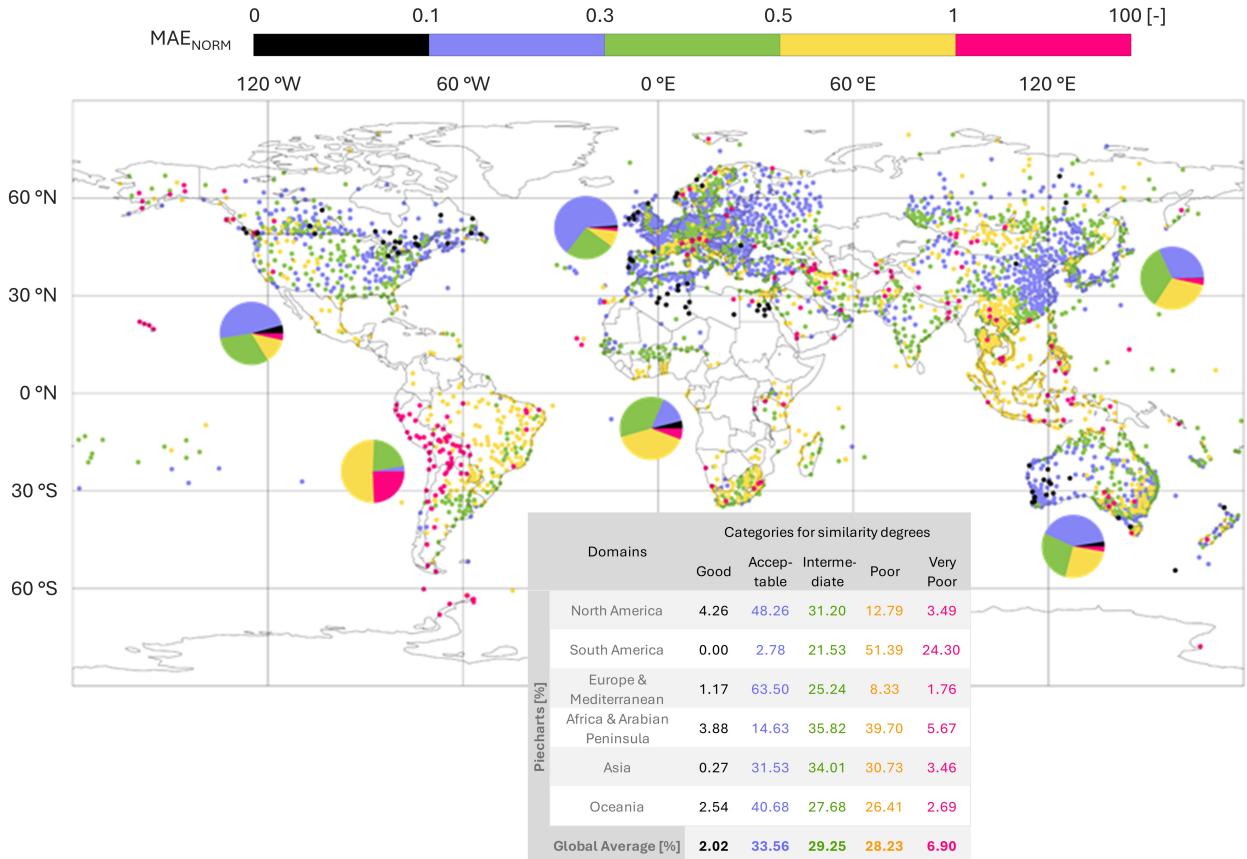
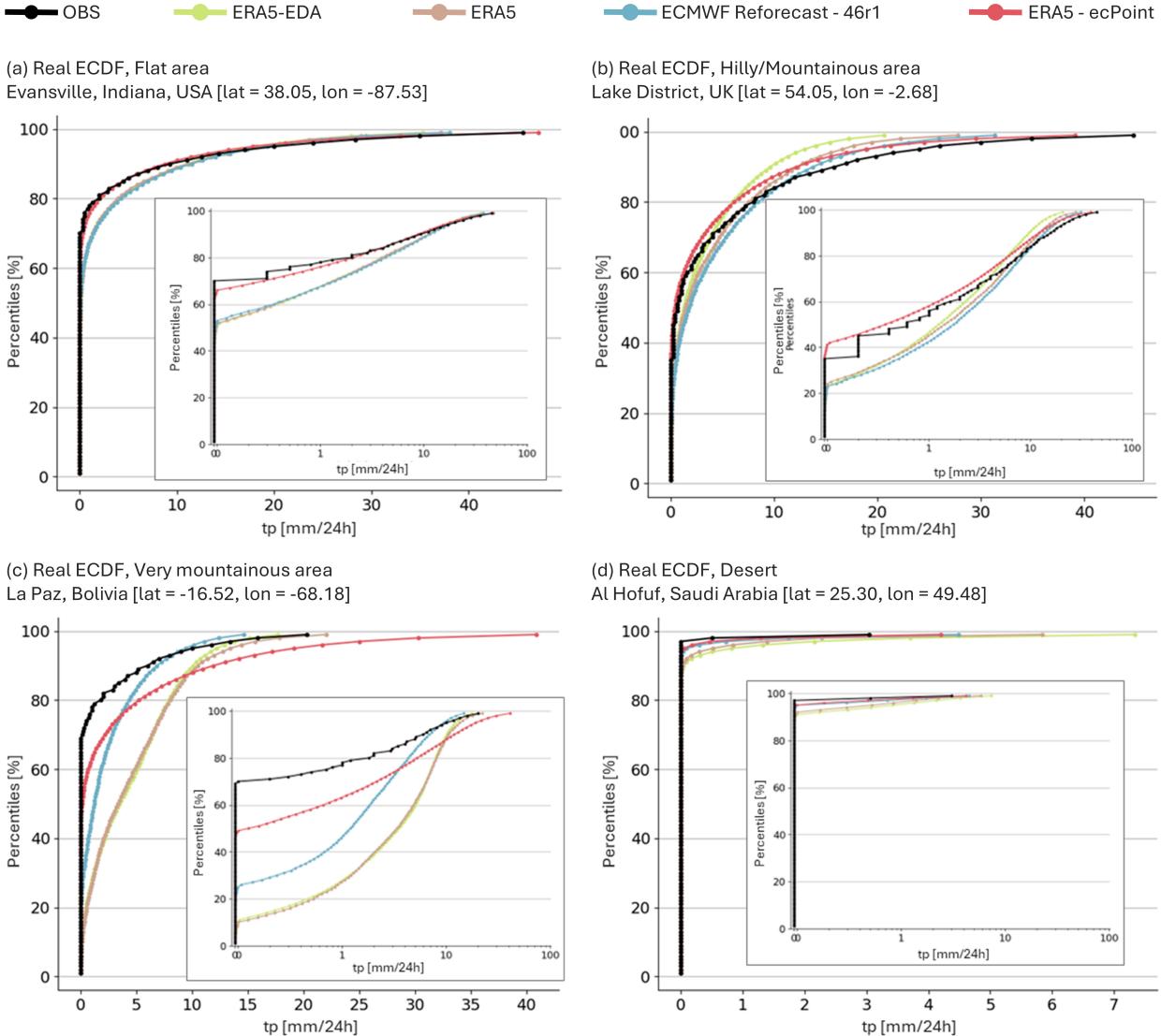


FIG. 6. Like Figure 4 but for ECMWF Reforecasts for 46r1 (at 18 km spatial resolution).

468 It is worth comparing the observed and the NWP-modelled ECDFs to gain insights into how  
 469 the distributions differ (Figure 7). Each ECDF (in linear scale) has an insert with the ECDF's  
 470 x-axis in logarithmic scale to compress/expand the small/high rainfall totals, and see more clearly  
 471 differences in the distributions. In flat areas (Figure 7a), ERA5-ecPoint (in coral) represents the  
 472 distribution of point-scale precipitation observations better than the baseline raw NWP-models:

## Empirical Cumulative Distribution Functions (ECDFs) for 24-hourly total precipitation (tp)



463 FIG. 7. Empirical Cumulative Distribution Functions (ECDFs) for 24-hourly total precipitation (tp) from rain  
 464 gauge observations (OBS, in black) and the NWP models ERA5-EDA (in green), ERA5 (in brown), ECMWF  
 465 Reforecasts-46r1 (in light blue), and ERA5-ecPoint (in coral). Panels (a) to (d) show examples of ECDFs,  
 466 respectively, for flat areas, hilly/mountainous areas, very mountainous areas, and deserts. The inserts represent  
 467 the same ECDFs but with the x-axis on a logarithmic scale.

473 it captures well the frequency of observed zero precipitation totals (see ECDF in log scale), the  
 474 growth rate of the precipitation observations<sup>5</sup> (see ECDF in log scale), and the length of the wet  
 475 tail (going up to the 99th percentile, see ECDFs in linear scale). There are no notable differences

<sup>5</sup>(Growth rate here is intended as the rate of change of the logarithm of precipitation totals)

476 between the distributions from ERA5-EDA (in green), ERA5 (in brown), and reforecasts (in blue):  
477 they all underestimate, although to different degrees, the frequency of observed zero precipitation  
478 totals, and they have similar growth rates, which are greater than that in the observed distribution.  
479 They all underestimate the length of the wet tail but to different degrees: in general, ERA5-EDA  
480 shows the biggest underestimation, reforecasts show the smallest, and ERA5 falls in between  
481 the two. In hilly/mountainous areas (Figure 7b), ERA5-ecPoint behaves similarly to flat areas. It  
482 represents the frequency of zero precipitation totals observed and the growth rate of the precipitation  
483 observations well. However, ERA5-ecPoint tends to slightly overestimate the distribution’s wet  
484 tail of the observed ECDF (in black). Compared to point-rainfall observations, raw NWP models  
485 show behaviour similar to that observed in flat areas. The main difference lies in a progression in a  
486 better representation of the observed ECDF for NWP models with increasing spatial resolution, i.e.,  
487 ERA5-EDA at 62 km (in green, in Figure 7b), which shows a worse representation of the observed  
488 ECDF compared to ERA5 at 31 km (in brown), and ERA5 shows a worse representation than  
489 reforecasts (in blue). This behaviour is seen in other sites too (not shown). In very mountainous  
490 areas (Figure 7c), all NWP models fail to represent the observed ECDFs. It is worth noting that  
491 this is not surprising as the observations used to train ERA5-ecPoint, and indeed to validate all  
492 representations, come primarily from valleys and hilly areas. First, all the NWP model versions  
493 underestimate the frequency of observed zero precipitation totals. ERA5-ecPoint tends to double  
494 such a frequency, but it does not reach the values in the observed ECDFs. The ECDFs from raw  
495 NWP models show a growth rate that is too large compared to the observed ECDFs, while ERA5-  
496 ecPoint also improves on that. Finally, while the raw NWP models tend to slightly underestimate  
497 the length of the observed ECDFs (with ERA5 providing the best representation out of the three  
498 models), ERA5-ecPoint tends to overestimate it. In desert areas (Figure 7d), all NWP models  
499 represent the observed ECDFs well, apart from the wet tails that tend to all be overestimated. The  
500 overestimation is reduced with the increase in the spatial resolution of the NWP models, with  
501 ERA5-ecPoint representing the actual length of the wet tail best.

### 10-year return period for 24-hourly total precipitation [mm/24h]

Piecharts: percentage of NWP-modelled estimates exceeding the corresponding observed ones

(a) Rain gauge observations (point scale), only rain gauges with at least 75% of valid record in the 20-year period

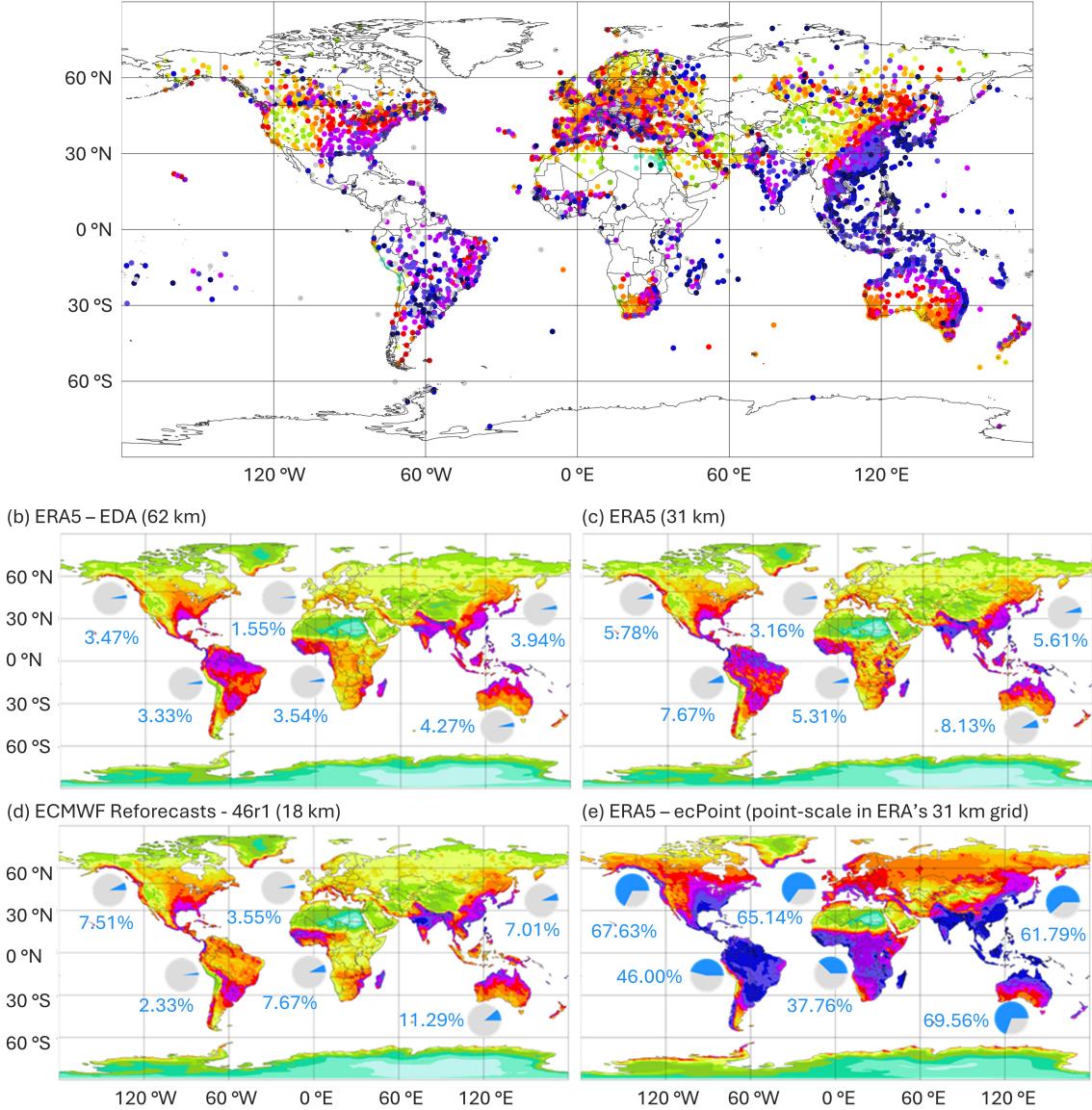
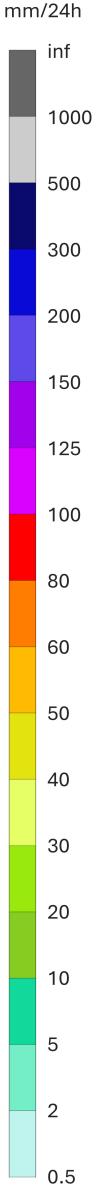


FIG. 8. Panel (a) displays the 10-year return period (computed from the 99.9726<sup>th</sup> percentile) for 24-hourly total precipitation from rain gauge observations, calculated over the 20-year period between 2000 and 2019, and using only rain gauges with at least 75% of valid records. Panels (b) to (e) show the 10-year return period for NWP-modelled 24-hourly total precipitation: ERA5-EDA (62 km), ERA5 (31 km), ECMWF Rerforecasts-46r1 (18 km) and ERA5-ecPoint (point-scale, provided on ERA5 grid). The pie charts represent the percentage (in %) of modelled climatologies exceeding the observed climatologies over the domains defined in Figures 3 to 6.

502 b. RQ2: comparison of the wet tail in the distributions built with NWP-modelled precipitation  
503 estimates and rain gauge observations

504 1) COMPARISON OF THE 10-YEAR PERIOD

511 The precipitation maps for the 10-year return period show that ERA5-ecPoint provides a better  
512 representation than the raw NWP models ([Figure 8a](#) and [Figure 8e](#)). In North America, the  
513 extremes in 24-hourly precipitation over the west coast of Alaska, Canada and North-West USA,  
514 which reach peaks up to 500 mm, are better represented in ERA5-ecPoint than in ERA5-EDA  
515 ([Figure 8b](#)), ERA5 ([Figure 8c](#)), and reforecasts ([Figure 8d](#)) that tend not to exceed 125 mm. The  
516 peaks around the Gulf of Mexico, the USA's East Coast and the border between Canada and the  
517 USA are also better represented in ERA5-ecPoint. However, in the latter case, there seems to  
518 also be sampling-related noise in the observations. The raw NWP models better represent the  
519 extremes over the Rocky Mountains since ERA5-ecPoint overestimates them. However, the latter  
520 shows an overall closer representation of the observed ECDFs apart from the tail. ERA5-ecPoint  
521 greatly improves the precipitation peaks over Mexico and South America over the other three NWP  
522 models, apart from the Andean region and the desert on the west coast of Peru and Chile, where  
523 ERA5-ecPoint overestimates the wet tails (as shown in section 4.a). It is worth noting that the  
524 ECMWF reforecasts from 46r1 halved the precipitation extremes over the Amazon compared to  
525 ERA5-EDA and ERA5.

526 The extremes over Europe also verify better on ERA5-ecPoint than the three raw NWP models.  
527 The wetter climatology with peaks up to 300-500 mm around the Mediterranean catchment (in-  
528 cluding the African part), the Alps, the Atlantic coast of Spain and the UK, and the Norwegian  
529 Fiords is better captured in ERA5-ecPoint than in the three raw NWP models. The higher spatial  
530 resolution in the reforecasts helps to increase the extremes compared to both reanalyses, but they  
531 still do not exceed 100 mm in 24-hours.

532 In Asia, there is a varied picture. The raw NWP models highlight the wetter climatologies of India  
533 (especially the Northeast regions), East China, Japan, Southeast Asia, and the Malay Archipelago.  
534 However, they do not reach the peaks of 300-500 mm/24h seen in the observations. ERA5-ecPoint  
535 represents such peaks. However, the peaks greater than 500 mm/24h observed in the Malay  
536 Archipelago remain underestimated, also in the post-processed ERA5. The overall overestimation

537 in the mountainous regions of Western China has a similar flavour to the ones discussed over  
538 the Rocky Mountains in the USA: ERA5-ecPoint shows the best overall representation of the full  
539 observed ECDFs, but tends to overestimate the wet tails. In the Arabian Peninsula, all models  
540 represent the overall observed ECDF tails quite well. As discussed in section 4.a for desert  
541 areas, such good representation originates from the high frequency of zero precipitation totals  
542 well estimated by all NWP models. The only exception is on the peninsula’s south coast, where  
543 precipitation peaks can reach 200 mm/24h, and raw NWP models estimate a maximum peak of only  
544 up to 80 mm/24h. ERA5-ecPoint increases them up to 150 mm/24h. In Oceania, all NWP models  
545 show a good overall representation of the observed ECDFs with slight underestimations of the  
546 wet tails. The added value of ERA5-ecPoint in this region mainly provides a better representation  
547 of the precipitation peaks. There are a few observations in Africa, and nothing can be said  
548 about the model representation of precipitation extremes in the numerous ungauged areas of this  
549 continent. All NWP models represent the wet climatology of West Africa, including its Atlantic  
550 coast. However, ERA5-ecPoint best represents the observed local peaks that vary between 100  
551 and 500 mm/24h. It is worth noting that ECMWF 46r1 reforecasts degrade the representation  
552 of the extreme precipitation around the Gulf of Guinea by producing maximum peaks only up to  
553 80-100 mm/24h. Similarly, out of all NWP models, ERA5-ecPoint somewhat better represents the  
554 varied precipitation peaks, between 80 and 500 mm/24h, in South Africa, where raw NWP models  
555 suggest extreme precipitation might not exceed 80 mm/24h. Also, in East Africa, ERA5-ecPoint  
556 provides a more realistic representation of the extreme precipitation peaks (up to 500 mm/24h)  
557 than raw NWP models. The reforecasts considerably reduce the precipitation in this area. The wet  
558 climatology of Madagascar is well represented in all NWP models, but ecPoint can increase the  
559 wet tail of ERA5 and provide extreme precipitation totals that are closer to those observed. Finally,  
560 all NWP models seem to represent quite well the observed precipitation distribution in the Sahara,  
561 with the caveat that data coverage there is poor. In any case, good performance likely connects to  
562 the prevalence of dry weather.

## 563 2) CASE STUDY: STORM VAIA IN ITALY (28TH OF OCTOBER 2018)

564 We now examine the case of widespread (flash) flooding in Italy on the 28th of October 2018  
565 (Figure 9). This event is part of a weather system that persisted over different parts of Italy

566 between the end of October and the beginning of November 2018. It is called Storm Vaia. In the  
567 observations (Figure 9a), one can see extreme precipitation amounts between 300-400 mm/24h  
568 over Veneto (north-east), up to 200 mm/24h over Lombardi (North) and Liguria (North-East),  
569 up to 240 mm/24h in Lazio (centre), up to 130 mm/24h in Puglia (Southwest), and up to 260  
570 mm/24h in Calabria (Southeast). ERA5-EDA (Figure 9b), ERA5 (Figure 9c), and reforecasts  
571 (Figure 9d) provide a good signal on which might be the wetter areas in Italy for that day, apart  
572 from the south of Italy, that does not stand out as a possible area at risk of extreme precipitation.  
573 The precipitation peaks over the Italian Peninsula increase with the increasing spatial resolution  
574 of the NWP models, but they do not reach the observed extreme precipitation totals. ERA5-EDA  
575 estimated a maximum total of 100 mm/24h, and ERA5 pushed the estimated peaks to 150 mm/24h  
576 over Veneto. Reforecasts increased the precipitation peaks in Veneto and Lazio up to 200 mm/24h,  
577 but precipitation in Liguria, Puglia, and Calabria remains highly underestimated. ERA5-ecPoint  
578 (Figure 9e) represents better the areas where the precipitation peaks were observed. In the north  
579 (Figure 9f, Northern Italy), where the storm created the biggest impacts, roughly 1% of the rainfall  
580 observations exceeded the 99th percentile of ERA5-ecPoint (red dots), indicating reliable point-  
581 scale rainfall estimates. In the rest of the peninsula (Figure 9f, Central and Southern Italy), 6% of  
582 the observations exceed the ERA5-ecPoint estimates, indicating an under-prediction of point-scale  
583 rainfall over Le Marche, Puglia, and Calabria. In this specific case, the location of the red dots  
584 along coastlines indicates underestimation primarily due to the known issue of convective cells  
585 generated over the sea not moving onto land ([Bechtold et al. 2014](#)).

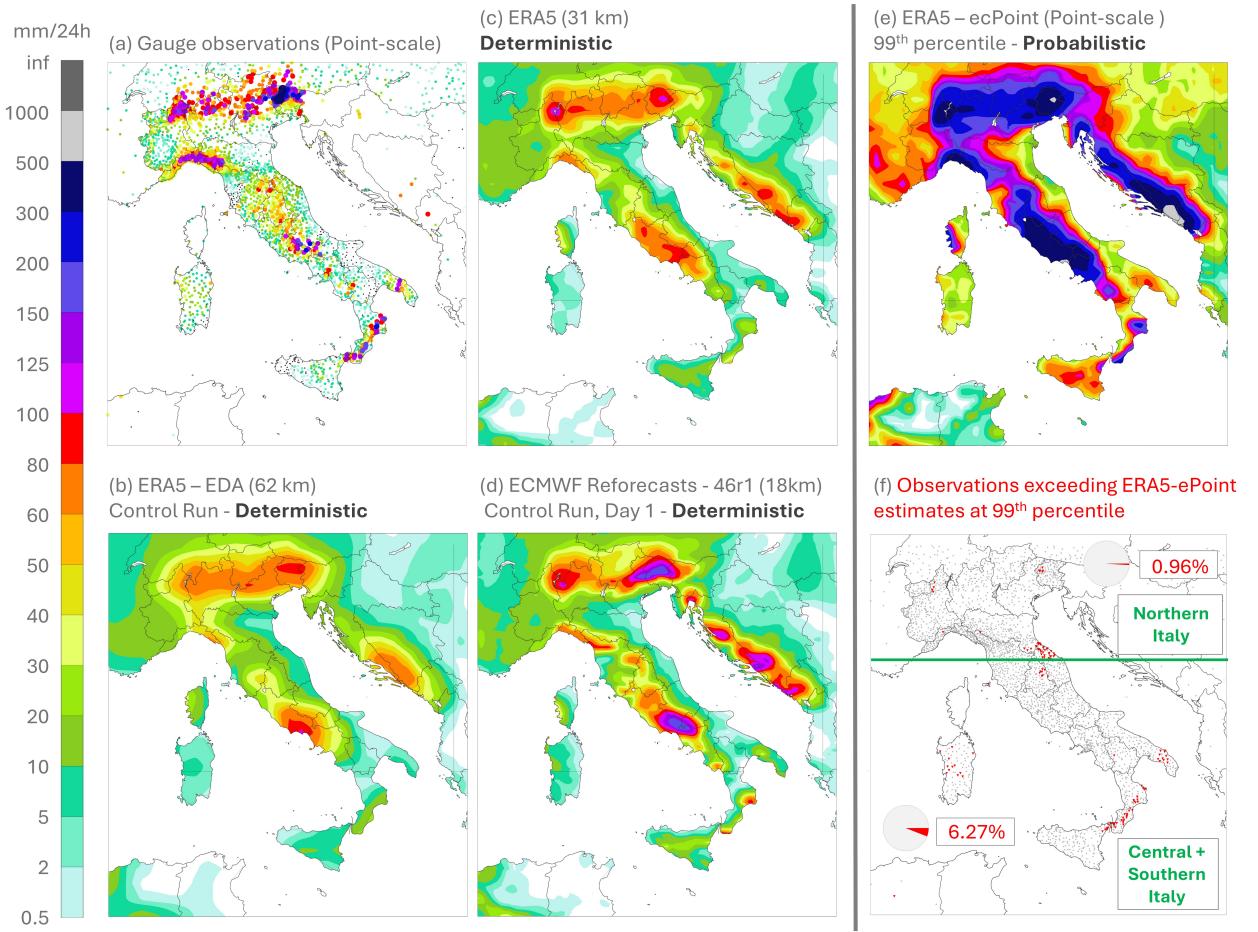
593

## 594 5. Discussion

595 The results from this study show that ERA5-ecPoint provides, overall, the best representation  
596 of point-rainfall distributions out of all the NWP models tested. Specifically, ERA5-ecPoint  
597 captures better the frequency of the observed zero rainfall totals, the growth rate within the rainfall  
598 observation CDFs, and the longer wet tails. The bigger improvements are particularly evident  
599 in flat and hilly/mountainous regions. However, in very mountainous areas such as the Andes,  
600 ERA5-ecPoint underestimates the frequency of zero rainfall totals and overestimates the length of  
601 the wet tails, raising some questions about its effectiveness over very complex orography. [This](#)

## 24-hourly total precipitation [mm/24h] for widespread flash floods in Italy (Storm Vaia)

Valued valid between 28<sup>th</sup> October 2018 at 00 UTC and 29<sup>th</sup> October 2018 at 00 UTC



586 FIG. 9. Widespread (flash) flooding in Italy on the 28th of October 2018 due to Storm Vaia. Panel (a)  
 587 represents the rain gauge observations. Panels (b) to (d) show the deterministic rainfall estimates, respectively,  
 588 for ERA5-EDA at 62 km (from control run), ERA at 31 km (single realisation), and ECMWF Reforecasts from  
 589 46r1 at 18 km (from control run, day 1 lead time). Panel (e) shows the probabilistic rainfall estimates from  
 590 ERA5-ecPoint (99th percentile). Panel (f) shows the locations of the rain gauge observations exceeding the  
 591 ERA5-ecPoint estimates at the 99th percentile. The representation is split into two geographical areas: Northern  
 592 and Southern Italy, with pie charts denoting total counts for these two areas.

602 **should not be surprising**, as ERA5-ecPoint is post-processed with observations primarily coming  
 603 from valleys and hilly areas, although, on the other hand, verifying data comes from such sites too.  
 604 **There likely exists a complex interplay, whereby** data from non-mountainous regions is sometimes  
 605 used to train for mountainous areas, despite the inclusion of a sub-grid orography variable in the

606 ERA5-ecPoint decision tree. The growth rate of the ERA5-ecPoint rainfall estimates closely aligns  
607 with that of the observations, indicating that the post-processing system is making meaningful  
608 adjustments to the rainfall estimates. Additional observational data from regions at high altitudes  
609 are necessary to refine the corrections, particularly to increase the accuracy in representing the  
610 frequency of zero rainfall totals and to reduce the overestimation observed in the wet tail.

611 Overall, the raw NWP models (i.e., ERA5-EDA, ERA5, and ECMWF Reforecasts – 46r1)  
612 consistently show an underestimation of the zero rainfall totals and the wet tails, and the growth  
613 rate of the modelled rainfall estimates is consistently bigger than that observed. This means that  
614 the raw NWP models overestimate the frequency of small rainfall totals and underestimate the  
615 frequency of extreme rainfall events, as one might expect from representivity considerations, and  
616 as has been reported previously by National Meteorological and Hydrological Services around  
617 Europe (Hewson and Chevallier 2024). ERA5 (at 31 km) improves the overall representation of  
618 point-rainfall distributions compared to ERA5-EDA (at 62 km), especially in mountainous regions  
619 such as the Rocky Mountains, the Alps, and the Norwegian Fjords. However, the improvements  
620 in these regions remain modest in proportion, despite the twofold increase in spatial resolution.  
621 The ECMWF Reforecasts provide general improvements due to the increased spatial resolution  
622 (18 km) and a more up-to-date model version (46r1 rather than 41r2 of ERA5-EDA and ERA5).  
623 The observed degradations over Australasia and Africa in 46r1 (see pie charts on Figure 4) are  
624 counterintuitive and may be symptomatic of a physics issue that manifests in those areas. Compared  
625 to ERA5, the 46r1 improvements are focused again on mountainous areas and extend to most of  
626 Europe, the arid regions of Northern Africa, and the Arabian Peninsula.

627 Focusing on extreme rainfall events, there is a general increase in the values with the increase  
628 of the raw NWP models' spatial resolution, which better agrees with the observed wetter tails.  
629 The major difference is observed between ERA5-EDA and ERA5, while the differences between  
630 the latter and the ECMWF reforecasts are less prominent. Indeed, for the rainfall in the Amazon  
631 region, Equatorial Africa, and Indonesia, the reforecasts show rainfall estimates that do not exceed  
632 100 mm/24h. In contrast, both reanalysis, ERA5-EDA and ERA5, show rainfall estimates up  
633 to 300 mm/24h, which better represent the observed rainfall totals in the region. These results  
634 similarly contradict expectations and may indicate regional limitations in cycle 46r1 employed for  
635 the reforecast dataset.

When focusing on extreme precipitation events, ERA5-ecPoint consistently demonstrates a superior ability to replicate observed extremes compared to raw NWP models. For example, the 10-year return period precipitation maps show that ERA5-ecPoint provides a much closer representation of observed extreme rainfall events in regions like North America, Europe, and parts of Asia. The Italian case study on Storm Vaia further underscores this finding. While raw models captured the general distribution of wetter areas, they underestimated the magnitude of precipitation peaks across multiple regions, including Veneto, Lazio, and Liguria. ERA5-ecPoint, on the other hand, was able to capture these extremes better, providing a more realistic forecast of the potential for flash floods. Some underestimation of rainfall along coastlines is highlighted due to non-moving convective cells generated over the sea that fail to generate rain over the land. Convective cell drift is something that has been explored in the ecPoint framework, but not implemented yet. Applying it should bring intrinsic improvements in the areas of triggering, via the bias correction aspect (as shown in the Hewson and Pillosu (2021), Norway example).

Furthermore, ERA5-ecPoint enables one to estimate rainfall events with significantly longer return periods than those presented in this study (e.g., up to a 1000-year return period, as noted in Table 1, row 5, column 8). Hewson et al. (2024) have shown for 2023 Storm Daniel in Libya that applying the ecPoint post-processing technique to ERA5 can deliver usable estimates of an  $n$ -year return period rainfall from  $m$  years of data, where  $n \gg m$ . Consequently, datasets like ERA5-ecPoint offer valuable insights into the potential magnitude of extreme rainfall events, improving our preparedness for unseen events (Heinrich et al. 2024; Ommer et al. 2024) or ones so infrequent that they have faded from collective memory (Ludwig et al. 2023; Merz et al. 2024).

One area where ERA5-ecPoint did not seem to provide significant benefits, and where extremes were overestimated, was in the high-altitude and relatively dry western parts of the USA, where mountain barriers can block external moisture sources. Parts of inland northern China fall into the same class. Dry boundary layers often characterise such areas. We know from experimenting with ecPoint and considering physics that low-level rainfall under-evaporation in such situations can lead to large net positive raw model rainfall biases at the grid scale, particularly in convective situations. Although ERA5-ecPoint includes a low-level humidity parameter within its decision tree, which can combat such biases, it is probably not activated in enough weather-type scenarios to be fully effective. Hence, there is some cross-contamination in the calibration from data in

666 areas with much moister boundary layers. This could thus be a focal point for future work. While  
667 ecPoint’s remote calibration approach has shown significant benefits compared to a purely local  
668 approach, as in this paper and in Hewson and Pillou (2021), there can evidently be some local  
669 downsides.

670 The climatological distributions evaluated here pool observations across all months of the year  
671 without seasonal stratification, and focus exclusively on liquid precipitation as recorded by rain  
672 gauges. This design choice was deliberate: the primary objective is to assess each dataset’s  
673 ability to represent the overall statistical distribution of 24-hourly rainfall accumulations across a  
674 wide range of climatic regimes, rather than to diagnose season-dependent performance. However,  
675 pooling across seasons may mask compensating biases (for instance, a model that overestimates  
676 light frontal rainfall in winter but underestimates intense convective rainfall in summer could  
677 produce an aggregate distribution that appears well calibrated despite meaningful seasonal errors).  
678 Similarly, in regions where a substantial fraction of annual precipitation falls as snow, the rain  
679 gauge observations inherently exclude that component, meaning the verification characterises only  
680 the liquid-phase subset of the precipitation climate. Decomposing the verification by season, and  
681 potentially by dominant precipitation type, would provide a more complete picture of each dataset’s  
682 strengths and weaknesses across different meteorological regimes. This could represent a natural  
683 extension of this work.

684 The general improvements provided by ERA5-ecPoint open up significant opportunities across  
685 various fields of environmental research that require a more accurate representation of point rainfall  
686 estimates. We advocate that such improvements would enhance both long-term strategic planning  
687 (e.g., using this dataset for climatological studies) and short-term emergency response (e.g., this  
688 dataset to create point-scale rainfall thresholds that are compatible with ecPoint rainfall medium-  
689 range forecasts to determine areas at risk of extreme localised rainfall), thereby contributing to  
690 developing more resilient societies in the face of climate change. In the realm of flood forecasting,  
691 more accurate rainfall estimates at local scales are crucial for predicting runoff and streamflow  
692 dynamics, particularly in catchments prone to flash floods. Precise point-scale rainfall data is  
693 pivotal in enhancing early warning systems, which are essential for safeguarding communities from  
694 the severe impacts of extreme rainfall and flooding. Better rainfall representation could facilitate  
695 more efficient management of reservoirs and irrigation planning in water resource management,

696 optimising water storage and distribution for agriculture, power generation, and urban water supply  
697 systems. Furthermore, enhanced point-scale precipitation estimates are crucial for designing more  
698 resilient stormwater infrastructure and urban drainage systems, which are facing increasing pressure  
699 from the intensification of extreme rainfall events due to climate change. In the context of disaster  
700 preparedness, ERA5-ecPoint's ability to capture the full spectrum of rainfall values, including zeros  
701 and extremes, provides valuable insights into the risks posed by changing precipitation patterns.

## 702 **6. Conclusions**

703 Modern-day NWP systems and reanalysis products do not provide a good representation of 24h  
704 climatological rainfall distributions for gauged sites around the world, whilst ecPoint, in its ERA5  
705 variant form, though not perfect everywhere, **performs much better**.

706 This study provides a systematic, global verification of how well NWP models represent the  
707 distribution of point-rainfall observations. It considered point-rainfall observations over 20 years  
708 and four different modelled, gridded datasets with distinct spatial resolutions: ERA5-EDA (62  
709 km), ERA5 (31 km), ECMWF Reforecasts for 46r1 (18 km), and ERA5-ecPoint (point-scale but  
710 provided over ERA5's grid at 31 km). Among the tested models, this study shows that ERA5-  
711 ecPoint most accurately captures both the frequency of zero rainfall totals and the wet tails of the  
712 observed point-rainfall distributions.

713 Since ERA5-ecPoint provides rainfall totals over a continuous global domain, the post-processed  
714 reanalysis could be used to provide seamless point-rainfall estimates, including over regions with  
715 sparse or no rain gauge observational data. While increasing NWP ensemble size addresses  
716 only grid-scale forecast uncertainty and convection-permitting resolutions (1 km) that could  
717 explicitly resolve sub-grid rainfall variability remain computationally prohibitive for global multi-  
718 decadal applications, statistical post-processing approaches such as ecPoint offer a practical and  
719 operationally viable means of bridging the scale gap between NWP grid-box averages and point-  
720 scale observations.

721 However, caution is needed when generalising the verification results. While ERA5-ecPoint  
722 demonstrates strong performance in estimating point-rainfall totals overall, it is essential to note  
723 that the verification dataset contains large regions with sparse or no rain gauge observations.  
724 Furthermore, ERA5-ecPoint has shown some limitations in very complex mountainous terrains  
725 (e.g. the Andes), where the post-processed reanalysis remains short in representing the frequency

726 of zero rainfall totals and overestimates the wet tails. Hence, this finding highlights the need for  
727 further refinement of the post-processed forecasts in these regions by incorporating, when available,  
728 more rain gauge observations in the calibration process.

729 The improved performance of ERA5-ecPoint over raw NWP models in representing point-scale  
730 rainfall totals, whether small or large, emphasises post-processing's critical role in addressing the  
731 inherent limitations of gridded rainfall estimates in guiding point-scale rainfall. **The effectiveness**  
732 **of ecPoint, however, remains contingent** on the quality of the underlying NWP models it post-  
733 processes. Without accurate raw NWP estimates at a grid-scale, the skill demonstrated by the  
734 ERA5-ecPoint rainfall estimates would be diminished. Moving forward, the authors advocate  
735 enhancing the spatial resolution and the skill of raw NWP models alongside ongoing improvements  
736 of post-processing techniques such as ecPoint to reduce further errors in estimating the whole  
737 distribution of point-rainfall totals. Such improvements will be particularly significant as climate  
738 change intensifies the frequency and severity of extreme rainfall, making accurate and reliable point-  
739 rainfall estimates indispensable for effective mitigation and response efforts related to droughts,  
740 extreme rainfall, flooding, food security, and urban resilience.

741 *Acknowledgments.* The authors gratefully acknowledge Hans Hersbach, Paul Berrisford, and  
742 Adrian Simmons for their invaluable insights on the characteristics of ERA5 and ERA5-EDA  
743 datasets and their implications in the outcomes shown in this paper.

744 *Data availability statement.* Data and/or the codes used to generate the figures that are incorpo-  
745 rated into this manuscript will be made available upon reasonable request.

## 746 References

- 747 Acero, F. J., S. Parey, J. A. García, and D. Dacunha-Castelle, 2018: Return Level Estimation of Ex-  
748 treme Rainfall over the Iberian Peninsula: Comparison of Methods. *Water*, **10** (2), <https://doi.org/10.3390/w10020179>.
- 750 Ahmed, A., G. Yildirim, K. Haddad, and A. Rahman, 2023: Regional Flood Frequency Analysis:  
751 A Bibliometric Overview. *Water*, **15** (9), <https://doi.org/10.3390/w15091658>.

- 752 Alexandridis, V., S. Stefanidis, and S. Dafis, 2023: Evaluation of era5 and era5-land reanalysis  
753 precipitation data with rain gauge observations in greece. *Environmental Sciences Proceedings*,  
754 **26 (1)**, 104, <https://doi.org/10.3390/ENVIRONSCIPROC2023026104>.
- 755 Anand, B., and D. Karunanidhi, 2020: Long-term spatial and temporal rainfall trend analysis  
756 using gis and statistical methods in lower bhavani basin, tamil nadu, india. *Indian Journal of*  
757 *Geo-Marine Sciences*, **49 (3)**, 419–427.
- 758 Anthanahalli Nanjegowda, R., and S. Kulamulla Parambath, 2022: A novel bias correction method  
759 for extreme rainfall events based on l-moments. *International Journal of Climatology*, **42 (1)**,  
760 250–264, <https://doi.org/10.1002/JOC.7242>.
- 761 Asadieh, B., and N. Y. Krakauer, 2015: Global trends in extreme precipitation: Climate models  
762 versus observations. *Hydrology and Earth System Sciences*, **19 (2)**, 877–891, <https://doi.org/10.5194/HESS-19-877-2015>.
- 764 Balasundram, S. K., R. R. Shamshiri, S. Sridhara, and N. Rizan, 2023: The role of digital agriculture  
765 in mitigating climate change and ensuring food security: An overview. *Sustainability*, **15 (6)**,  
766 <https://doi.org/10.3390/SU15065325>.
- 767 Bechtold, P., N. Semane, P. Lopez, J. P. Chaboureau, A. Beljaars, and N. Bormann, 2014: Rep-  
768 resenting equilibrium and nonequilibrium convection in large-scale models. *Journal of the*  
769 *Atmospheric Sciences*, **71 (2)**, 734–753, <https://doi.org/10.1175/JAS-D-13-0163.1>.
- 770 Bottazzi, M., and Coauthors, 2024: High performance computing to support land, climate, and  
771 user-oriented services: The highlander data portal. *Meteorological Applications*, **31 (2)**, e2166,  
772 <https://doi.org/10.1002/met.2166>.
- 773 Boé, J., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of  
774 the seine basin climate for hydro-meteorological studies. *International Journal of Climatology*,  
775 **27 (12)**, 1643–1655, <https://doi.org/10.1002/joc.1602>.
- 776 Chai, T., and R. R. Draxler, 2014: Root mean square error (rmse) or mean absolute error (mae)?  
777 -arguments against avoiding rmse in the literature. *Geoscientific Model Development*, **7 (3)**,  
778 1247–1250, <https://doi.org/10.5194/GMD-7-1247-2014>.

- 779 Coles, S., 2001: Threshold Models. *An Introduction to Statistical Modeling of Extreme Values*,  
780 Springer, 74–91, [https://doi.org/10.1007/978-1-4471-3675-0\\_4](https://doi.org/10.1007/978-1-4471-3675-0_4).
- 781 Cuo, L., T. C. Pagano, and Q. J. Wang, 2011: A review of quantitative precipitation forecasts  
782 and their use in short- to medium-range streamflow forecasting. *Journal of Hydrometeorology*,  
783 **12** (5), 713–728, <https://doi.org/10.1175/2011JHM1347.1>.
- 784 Di Curzio, D., A. Di Giovanni, R. Lidori, M. Montopoli, and S. Rusi, 2022: Comparing rain  
785 gauge and weather radar data in the estimation of the pluviometric inflow from the apennine  
786 ridge to the adriatic coast (abruzzo region, central italy). *Hydrology*, **9** (12), 225, <https://doi.org/10.3390/HYDROLOGY9120225>.
- 788 Donat, M. G., L. V. Alexander, N. Herold, and A. J. Dittus, 2016: Temperature and precip-  
789 itation extremes in century-long gridded observations, reanalyses, and atmospheric model  
790 simulations. *Journal of Geophysical Research*, **121** (19), 11 174–11 189, <https://doi.org/10.1002/2016JD025480>.
- 792 Doocy, S., A. Daniels, S. Murray, and T. D. Kirsch, 2013: The human impact of floods: A historical  
793 review of events 1980-2009 and systematic literature review. *PLoS Currents*, <https://doi.org/10.1371/CURRENTS.DIS.F4DEB457904936B07C09DAA98EE8171A>.
- 795 Engmann, S., and D. Cousineau, 2011: Comparing distributions: The two-sample anderson-  
796 darling test as an alternative to the kolmogorov-smirnoff test. *Journal of Applied Quantitative  
797 Methods*, **6** (3), URL [https://jaqm.ro/issues/volume-6,issue-3/1\\_engmann\\_cousineau.php](https://jaqm.ro/issues/volume-6,issue-3/1_engmann_cousineau.php).
- 798 Ensor, L. A., and S. M. Robeson, 2008: Statistical characteristics of daily precipitation: Compar-  
799 isons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, **47** (9),  
800 2468–2476, <https://doi.org/10.1175/2008JAMC1757.1>.
- 801 Espinosa, L. A., M. M. Portela, and S. Gharbia, 2024: Assessing changes in exceptional rain-  
802 fall in portugal using era5-land reanalysis data (1981/1982–2022/2023). *Water*, **16** (5), 628,  
803 <https://doi.org/10.3390/W16050628>.
- 804 Gimeno, L., R. Sorí, M. Vázquez, M. Stojanovic, I. Algarra, J. Eiras-Barca, L. Gimeno-Sotelo, and  
805 R. Nieto, 2022: Extreme precipitation events. *Wiley Interdisciplinary Reviews: Water*, **9** (6),  
806 <https://doi.org/10.1002/WAT2.1611>.

- 807 Giorgos, N., P. T. Nastos, and Y. Kapsomenakis, 2024: Creating high-resolution precipitation and  
808 extreme precipitation indices datasets by downscaling and improving on the era5 reanalysis data  
809 over greece. *Eng*, **5** (3), 1885–1904, <https://doi.org/10.3390/ENG5030101>.
- 810 Göber, M., E. Zsótér, and D. S. Richardson, 2008: Could a perfect model ever satisfy a naïve  
811 forecaster? On grid box mean versus point verification. *Meteorological Applications*, **15** (3),  
812 359–365, <https://doi.org/10.1002/met.78>.
- 813 Gomis-Cebolla, J., V. Rattayova, S. Salazar-Galán, and F. Francés, 2023: Evaluation of era5 and  
814 era5-land reanalysis precipitation datasets over spain (1951–2020). *Atmospheric Research*, **284**,  
815 <https://doi.org/10.1016/J.ATMOSRES.2023.106606>.
- 816 Guan, H., and Coauthors, 2022: GEFSv12 Reforecast Dataset for Supporting Subseasonal and  
817 Hydrometeorological Applications. *Monthly Weather Review*, **150** (3), 647–665, <https://doi.org/10.1175/MWR-D-21-0245.1>.
- 819 Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical note:  
820 Downscaling rcm precipitation to the station scale using statistical transformations - a comparison  
821 of methods. *Hydrology and Earth System Sciences*, **16** (9), 3383–3390, <https://doi.org/10.5194/hess-16-3383-2012>.
- 823 Gupta, V., M. K. Jain, P. K. Singh, and V. Singh, 2020: An assessment of global satellite-based  
824 precipitation datasets in capturing precipitation extremes: A comparison with observed precipi-  
825 tation dataset in india. *International Journal of Climatology*, **40** (8), 3667–3688, <https://doi.org/10.1002/JOC.6419>.
- 827 Haiden, T., and S. Duffy, 2016: Use of high-density observations in precipitation verification.  
828 *ECMWF Newsletter*, (147), 20–25, <https://doi.org/10.21957/hsacrdem>.
- 829 Haiden, T., M. Janousek, F. Vitart, F. Prates, M. Maier-Gerber, C. W. Y. Li, and M. Chevallier,  
830 2025: Evaluation of ECMWF forecasts. *ECMWF Technical Memoranda*, **931**, <https://doi.org/10.21957/51e665e5c8>.
- 832 Haile, G. G., and Coauthors, 2020: Projected impacts of climate change on drought patterns over  
833 east africa. *Earth's Future*, **8** (7), <https://doi.org/10.1029/2020EF001502>.

- 834 Hamill, T. M., and Coauthors, 2022: The reanalysis for the global ensemble forecast system, version  
835 12. *Monthly Weather Review*, **150** (1), 59–79, <https://doi.org/10.1175/MWR-D-21-0023.1>.
- 836 Heinrich, D., E. Stephens, and E. Coughlan de Perez, 2024: More than magnitude: Towards a mul-  
837 tidimensional understanding of unprecedented weather to better support disaster management.  
838 *Water Security*, **23**, 100 181, <https://doi.org/10.1016/J.WASEC.2024.100181>.
- 839 Herold, N., A. Behrangi, and L. V. Alexander, 2017: Large uncertainties in observed daily  
840 precipitation extremes over land. *Journal of Geophysical Research: Atmospheres*, **122** (2),  
841 668–681, <https://doi.org/10.1002/2016JD025842>.
- 842 Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal  
843 Meteorological Society*, **146** (730), 1999–2049, <https://doi.org/10.1002/qj.3803>.
- 844 Hewson, T., 2024: Capturing extreme rainfall events. *ECMWF Newsletter*, (178).
- 845 Hewson, T., and M. Chevallier, 2024: Use and verification of ecmwf products. *ECMWF Technical  
846 Memoranda*, (919), <https://doi.org/10.21957/734d5d4d39>.
- 847 Hewson, T., and F. Pillusu, 2021: A low-cost post-processing technique improves weather  
848 forecasts around the world. *Communications Earth and Environment*, **2** (1), <https://doi.org/10.1038/s43247-021-00185-9>.
- 850 Hewson, T., F. Pillusu, E. Gascòn, and M. Vučković, 2023: Post-processing era5 output with  
851 ecpoint. *ECMWF Newsletter*, (176).
- 852 Hewson, T., and Coauthors, 2024: Medicane daniel: an extraordinary cyclone with devastating  
853 impacts. *ECMWF Newsletter*, (179), <https://doi.org/10.21957/th3wxk861d>.
- 854 Hosking, J. R. M., and J. R. Wallis, 1997: *Regional Frequency Analysis: An Approach  
855 Based on L-Moments*. Cambridge University Press, Cambridge, <https://doi.org/10.1017/CBO9780511529443>.
- 857 Hossain, I., S. Gato-Trinidad, M. Imteaz, and S. Rayburg, 2024: Future scenarios of design rainfall  
858 due to upcoming climate changes in nsw, australia. *Atmosphere*, **15** (9), 1101, <https://doi.org/10.3390/ATMOS15091101>.

- 860 IPCC, 2023: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and*  
861 *III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* IPCC,  
862 Geneva, Switzerland, 35–115 pp., <https://doi.org/10.59327/IPCC/AR6-9789291691647>.
- 863 Janmohammadi, M., and N. Sabaghnia, 2023: Strategies to alleviate the unusual effects of climate  
864 change on crop production: a thirsty and warm future, low crop quality. a review. *Biologija*,  
865 **69 (2)**, 121–133, <https://doi.org/10.6001/BIOLOGIJA.2023.69.2.1>.
- 866 Janssen, A., 2000: Global power functions of goodness of fit tests. *The Annals of Statistics*, **28 (1)**,  
867 239–253.
- 868 Jolliffe, I. T., and D. B. Stephenson, 2011: *Forecast Verification: A Practitioner's Guide in Atmo-*  
869 *spheric Science.* 2nd ed., John Wiley & Sons, Ltd., <https://doi.org/10.1002/9781119960003>.
- 870 Kalnay, E., and Coauthors, 1996: The ncep/ncar 40-year reanalysis project. *Bulletin of the Amer-*  
871 *ican Meteorological Society*, **77**, 437–472, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- 872 Khaliq, M. N., T. B. M. J. Ouarda, J. C. Ondo, P. Gachon, and B. Bobée, 2006: Frequency analysis  
873 of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review.  
874 *Journal of Hydrology*, **329 (3)**, 534–552, <https://doi.org/10.1016/j.jhydrol.2006.03.004>.
- 875 Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B.  
876 Kirschbaum, 2017: So, how much of the earth's surface is covered by rain gauges? *Bulletin of the*  
877 *American Meteorological Society*, **98 (1)**, 69–78, <https://doi.org/10.1175/BAMS-D-14-00283.1>.
- 878 Kim, S., K. Sung, J.-Y. Shin, and J.-H. Heo, 2025: At-Site Versus Regional Frequency Analysis of  
879 Sub-Hourly Rainfall for Urban Hydrology Applications During Recent Extreme Events. *Water*,  
880 **17 (15)**, <https://doi.org/10.3390/w17152213>.
- 881 Kirk, R. E., 1996: Practical significance: A concept whose time has come. *Educational and*  
882 *Psychological Measurement*, **56 (5)**, 746–759, <https://doi.org/10.1177/0013164496056005002>.
- 883 Kochendorfer, J., and Coauthors, 2020: Undercatch Adjustments for Tipping-Bucket Gauge  
884 Measurements of Solid Precipitation. *Journal of Hydrometeorology*, **21 (6)**, 1193–1205,  
885 <https://doi.org/10.1175/JHM-D-19-0256.1>.

- 887 Lamprecht, A., and Coauthors, 2021: Disentangling anthropogenic drivers of climate change  
888 impacts on alpine plant species: Alps vs. mediterranean mountains. final report. *Earth System*  
889 *Sciences (ESS)*, 1–72, <https://doi.org/10.1553/ESS-MEDIALPSS1>.
- 890 Lanza, L. G., and L. Stagi, 2008: Certified accuracy of rainfall data as a standard re-  
891 quirement in scientific investigations. *Advances in Geosciences*, **16**, 43–48, <https://doi.org/10.5194/ADGEO-16-43-2008>.
- 893 Laouacheria, F., S. Kechida, and M. Chabi, 2019: Modelling the impact of design rainfall on  
894 the urban drainage system by storm water management model. *Journal of Water and Land*  
895 *Development*, **40 (1)**, 119–125, <https://doi.org/10.2478/JWLD-2019-0013>.
- 896 Lavers, D. A., S. Harrigan, and C. Prudhomme, 2021: Precipitation biases in the ECMWF  
897 integrated forecasting system. *Journal of Hydrometeorology*, **22 (5)**, 1315–1334, <https://doi.org/10.1175/jhm-d-20-0308.1>.
- 899 Lavers, D. A., A. Simmons, F. Vamborg, and M. J. Rodwell, 2022: An evaluation of era5  
900 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*,  
901 **148 (748)**, 3152–3165, <https://doi.org/10.1002/QJ.4351>.
- 902 Liang, X. Z., 2022: Extreme rainfall slows the global economy. *Nature*, **601 (7892)**, 193–194,  
903 <https://doi.org/10.1038/d41586-021-03783-x>.
- 904 Ludwig, P., and Coauthors, 2023: A multi-disciplinary analysis of the exceptional flood event of july  
905 2021 in central europe - part 2: Historical context and relation to climate change. *Natural Hazards*  
906 and *Earth System Sciences*, **23 (4)**, 1287–1311, <https://doi.org/10.5194/NHESS-23-1287-2023>.
- 907 Maurya, S. K., A. Kalhapure, V. K. Verma, A. Tiwari, C. Chaubey, D. K. Maurya, and M. Kumar,  
908 2024: Irrigation scheduling and cultivar management for increasing water productivity under  
909 dryland condition: A review. *International Journal of Environment and Climate Change*, **14 (1)**,  
910 461–470, <https://doi.org/10.9734/IJECC/2024/V14I13856>.
- 911 Merz, B., and Coauthors, 2024: Spatial counterfactuals to explore disastrous flooding. *Environ-  
912 mental Research Letters*, **19 (4)**, <https://doi.org/10.1088/1748-9326/AD22B9>.

- 913 Muñoz-Sabater, J., and Coauthors, 2021: Era5-land: A state-of-the-art global reanalysis dataset  
914 for land applications. *Earth System Science Data*, **13** (9), 4349–4383, [https://doi.org/10.5194/  
915 ESSD-13-4349-2021](https://doi.org/10.5194/ESSD-13-4349-2021).
- 916 Ommer, J., J. Neumann, M. Kalas, S. Blackburn, and H. L. Cloke, 2024: Surprise floods: The  
917 role of our imagination in preparing for disasters. *Natural Hazards and Earth System Sciences*,  
918 **24** (8), 2633–2646, <https://doi.org/10.5194/NHESS-24-2633-2024>.
- 919 Osei, M. A., L. K. Amekudzi, A. Y. Omari-Sasu, E. I. Yamba, E. Quansah, J. N. A. Aryee,  
920 and K. Preko, 2021: Estimation of the return periods of maxima rainfall and floods at the Pra  
921 River Catchment, Ghana, West Africa using the Gumbel extreme value theory. *Heliyon*, **7** (5),  
922 <https://doi.org/10.1016/j.heliyon.2021.e06980>.
- 923 Pan, X., A. Rahman, K. Haddad, and T. B. M. J. Ouarda, 2022: Peaks-over-threshold model in flood  
924 frequency analysis: A scoping review. *Stochastic Environmental Research and Risk Assessment*,  
925 **36** (9), 2419–2435, <https://doi.org/10.1007/s00477-022-02174-6>.
- 926 Pillusu, F. M., T. Hewson, E. Stephens, C. Prudhomme, and H. Cloke, 2025: Does a multivariate  
927 approach enhance univariate grid-to-point post-processed rainfall forecasts? A comparative  
928 analysis. *ECMWF Technical Memoranda*, **924**, 1–21, <https://doi.org/10.21957/eb6c62b7af>.
- 929 Pollock, M. D., and Coauthors, 2018: Quantifying and Mitigating Wind-Induced Undercatch in  
930 Rainfall Measurements. *Water Resources Research*, **54** (6), 3863–3875, [https://doi.org/10.1029/  
931 2017WR022421](https://doi.org/10.1029/2017WR022421).
- 932 Richardson, D., S. Hemri, K. Bogner, T. Gneiting, T. Haiden, F. Pappenberger, and M. Scheuerer,  
933 2014: Calibration of ecmwf forecasts. *ECMWF Newsletter*, **(142)**, URL <https://doi.org/10.21957/45t3o8fj>.
- 935 Satgé, F., D. Defrance, B. Sultan, M. P. Bonnet, F. Seyler, N. Rouché, F. Pierron, and J. E. Paturel,  
936 2020: Evaluation of 23 gridded precipitation datasets across west africa. *Journal of Hydrology*,  
937 **581**, 124412, <https://doi.org/10.1016/J.JHYDROL.2019.124412>.
- 938 Scarrott, C., and A. MacDonald, 2012: A Review of Extreme Value Threshold Estimation  
939 and Uncertainty Quantification. *REVSTAT-Statistical Journal*, **10** (1), 33–60, [https://doi.org/  
940 10.57805/revstat.v10i1.110](https://doi.org/10.57805/revstat.v10i1.110).

- 941 Schumacher, R. S., 2017: Heavy rainfall and flash flooding. *Oxford Research Encyclopedia of*  
942 *Natural Hazard Science*, **1**, <https://doi.org/10.1093/acrefore/9780199389407.013.132>.
- 943 Stephens, M. A., 1974: Edf statistics for goodness of fit and some comparisons. *Journal of the*  
944 *American Statistical Association*, **69 (347)**, 730, <https://doi.org/10.2307/2286009>.
- 945 Tadeyo, E., D. Chen, B. Ayugi, and C. Yao, 2020: Characterization of spatio-temporal trends and  
946 periodicity of precipitation over malawi during 1979-2015. *Atmosphere*, **11 (9)**, <https://doi.org/10.3390/ATMOS11090891>.
- 948 Tie, J., S. Kim, and A. Sharma, 2023: How does increasing temperature affect the sub-annual  
949 distribution of monthly rainfall? *Environmental Research: Climate*, **2 (1)**, <https://doi.org/10.1088/2752-5295/ACB5B9>.
- 951 Tye, M. R., K. Dagon, M. J. Molina, J. H. Richter, D. Visioni, B. Kravitz, and S. Tilmes, 2022:  
952 Indices of extremes: geographic patterns of change in extremes and associated vegetation  
953 impacts under climate intervention. *Earth System Dynamics*, **13 (3)**, 1233–1257, <https://doi.org/10.5194/ESD-13-1233-2022>.
- 955 Vannitsem, S., and Coauthors, 2021: Statistical postprocessing for weather forecasts review,  
956 challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*,  
957 **102 (3)**, E681–E699, <https://doi.org/10.1175/BAMS-D-19-0308.1>.
- 958 Wang, X., S. Shi, L. Zhu, Y. Nie, and G. Lai, 2023: Traditional and Novel Methods of Rainfall  
959 Observation and Measurement: A Review. *Journal of Hydrometeorology*, **24 (12)**, 2153–2176,  
960 <https://doi.org/10.1175/JHM-D-22-0122.1>.
- 961 Wang, Y., and H. A. Karimi, 2022: Impact of spatial distribution information of rainfall in runoff  
962 simulation using deep learning method. *Hydrology and Earth System Sciences*, **26 (9)**, 2387–  
963 2403, <https://doi.org/10.5194/HESS-26-2387-2022>.
- 964 Westra, S., and Coauthors, 2014: Future changes to the intensity and frequency of short-duration ex-  
965 treme rainfall. *Reviews of Geophysics*, **52 (3)**, 522–555, <https://doi.org/10.1002/2014RG000464>.
- 966 Wilks, D. S., 2020: *Statistical Methods in the Atmospheric Sciences*. 4th ed., Elsevier.
- 967 WMO, 2024: State of the global climate 2023. *WMO-No.1347*.

968 Zittis, G., A. Bruggeman, and J. Lelieveld, 2021: Revisiting future extreme precipitation trends in  
969 the mediterranean. *Weather and Climate Extremes*, **34**, <https://doi.org/10.1016/J.WACE.2021.100380>.

971 Zsoter, E., C. Prudhomme, E. Stephens, F. Pappenberger, and H. Cloke, 2020: Using ensemble  
972 reforecasts to generate flood thresholds for improved global flood forecasting. *Journal of Flood  
973 Risk Management*, **13 (4)**, <https://doi.org/10.1111/jfr3.12658>.