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

Fatima M. Pillosu[a,b] Timothy D. Hewson[b] Estíbaliz Gascòn[b] Milana Vučković[b] Christel Prudhomme[b,c,d] Hannah Cloke[a,e]

[a] *Department of Geography and Environmental Science, University of Reading, Reading, UK*

[b] *Forecast and Services Department, European Centre for Medium-range Weather Forecasts, Reading, UK*

[c] *Department of Geography and Environment, University of Loughborough, Loughborough, UK*

[d] *UK Centre for Ecology and Hydrology, Wallingford, United Kingdom*

[e] *Department of Meteorology, University of Reading, Reading, UK*

Corresponding author: Fatima Pillosu, tj835682@student.reading.ac.uk

Early Online Release: This preliminary version has been accepted for publication in *Bulletin of the American Meteorological Society*, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-25-0019.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT: Accurately estimating rainfall distributions, from-small-to-extreme totals, is crucial for addressing various environmental challenges (e.g., flood forecasting, water resource management, disaster preparedness). Global Numerical Weather Prediction (NWP) models can provide useful rainfall estimates; yet, they often misrepresent point-scale observations from rain gauges, underestimating the frequency of small rainfall totals and extreme values. In general, finer resolutions yield more accurate representation of gauge-based climatologies. Hence, this study provides a systematic, global verification of four NWP-modelled rainfall datasets of differing resolutions (with ”resolution” meaning ”horizontal grid spacing”) - ERA5’s Ensemble Data Assimilation (62 km, probabilistic), ERA5’s short-range forecasts (31 km, deterministic), short-range 46r1 ECMWF reforecasts (18 km, control run), and ERA5-ecPoint (point-scale, probabilistic) — against 20 years of global rain gauge observations, assessing each dataset’s ability to represent the entire rainfall distribution. Although in very mountainous areas (e.g., the Andes) ERA5-ecPoint underestimates zero-rainfall frequency and overestimates wet tail length, it dramatically improves upon raw NWP performance in many other regions by capturing more accurately the frequency of zeros, the ”growth rates” of rainfall totals, and the wet tails. Moreover, due to its probabilistic nature, ERA5-ecPoint can estimate long return periods (e.g., 1000 years) without using distribution fitting, thereby offering valuable insights into extremely rare or unprecedented events at specific locations. Such findings underscore the importance of using post-processing to enhance the local-scale validity of global NWP models. Moreover, as climate change intensifies extreme rainfall events, such post-processing becomes crucial for estimating accurate long-period rainfall climatologies, as needed for effective mitigation and resilience building, particularly in areas lacking comprehensive and reliable rain gauge records.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

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

1. Introduction

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

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.

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, within e.g., 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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

underestimate extreme precipitation due to their coarse grid spacing of e.g., 30-50 km (Alexandridis et al. 2023; Donat et al. 2016; Espinosa et al. 2024; Gomis-Cebolla et al. 2023)[1]. 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)[2]. Note that hereafter we generally adopt the commonly used term "resolution" to refer to grid spacing.

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 Pillosu (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.

---

[1]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 studies 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).

[2]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.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

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

2. Data

a. Point-scale rain gauge precipitation observations

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

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

[3]https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

TABLE 1. Characteristics (columns 1 to 5) of the observational rainfall dataset (row 1) and the considered NWP models (rows 2 to 5), as well as their derived climatologies (columns 6 to 8). It is worth noting that the unavailable data for ERA5-EDA (row 2, column 6) is due to sporadic damage to the storage tapes on which the original data were archived (MARS archive). These losses occurred at irregular intervals with no systematic temporal or spatial pattern. There is therefore no reason to expect that the missing data would introduce bias into 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 = 1 in a 15-year event ** required at least 75% of valid obs.*	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. ** we had only 66940 real. as some dates were not available*	66940 real. / 365 days = 1 in a 183-year event	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 = 1 in a 20-year event	99.98630
4	NWP-modelled, gridded	46r1 ECMWF Reforecasts (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 = 1 in a 56-year event	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 = 1 in a 1980-year event	99.99986

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

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

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

b. Gridded NWP-modelled precipitation estimates

1) ERA5 Reanalysis (ERA5) and ERA5 Ensemble Data Assimilation (ERA5-EDA)

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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

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

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

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

2) ECMWF Reforecasts

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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

(equiprobable) precipitation realisations by considering lead times up to day 10. This was viable as there was no drift in the forecasts up to day 10 (not shown). Hence, the precipitation distribution built with ECMWF reforecasts contains 20800 realisations (Table 1, row 4). Further work could look in more detail at whether perturbed member climatologies could provided a more accurate benchmark, but because of the perturbations this is not the expected behaviour and in any case the fundamental ”scale gap” would still be there, so this approach was considered out of scope for this study.

3) ERA5-ecPoint

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

The MFs are used to post-process the raw rainfall outputs from NWP models. Suppose all grid-boxes in the raw forecasts are post-processed by sampling only the MF in Figure 1a (also the same MF in the black circle in Figure 1b). In this case, the ecPoint post-processing would follow a univariate approach (U-ecPoint), and be thought of as a single-leaf decision tree (Figure 1b).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

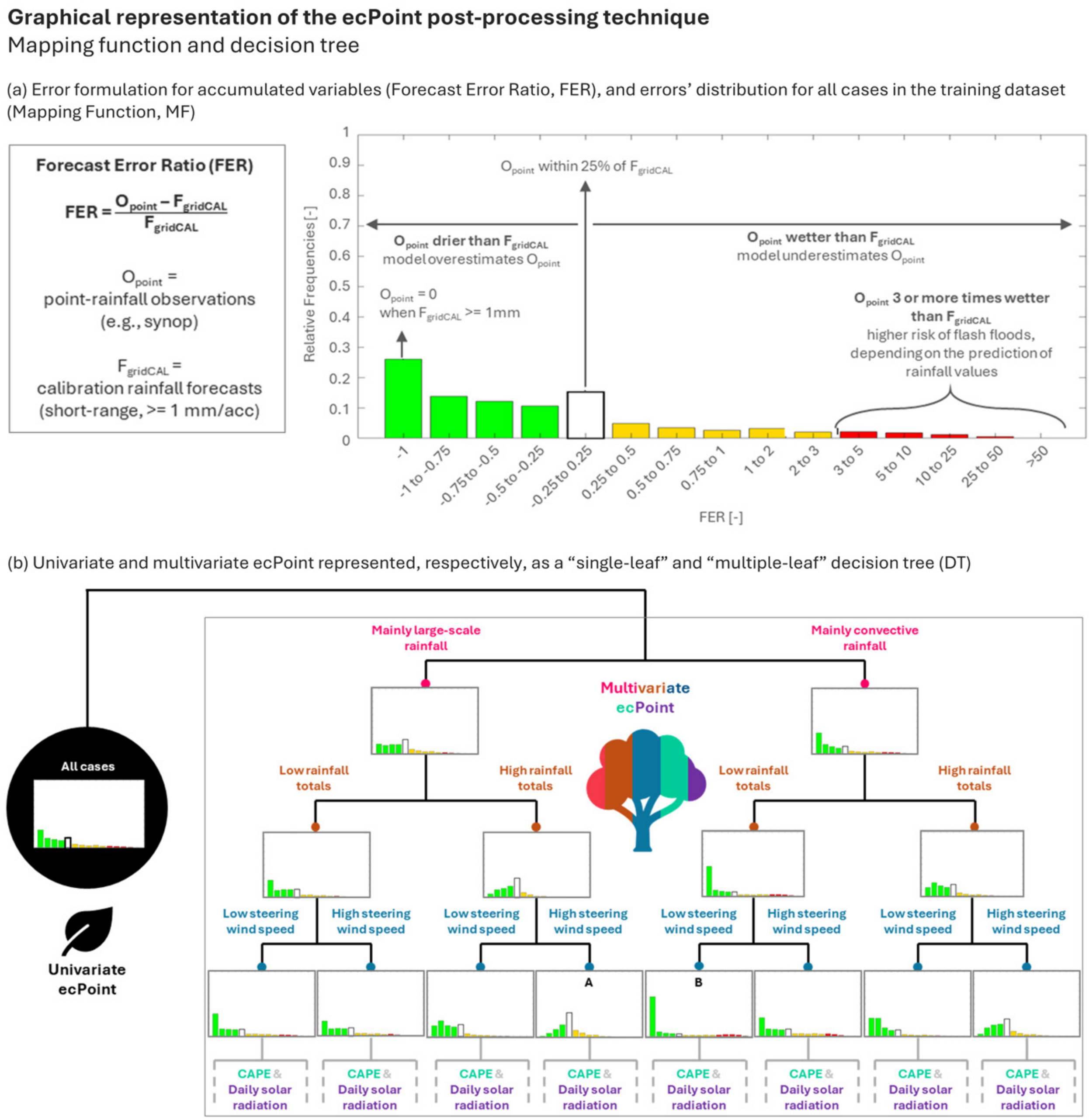


Fig. 1. Panel (a) shows the error formulation for accumulated variables (Forecast Error Ratio, FER) and the error distribution for all cases in the training dataset (Mapping Function, MF). The example pertains to the calibration of 47r3 ECMWF ENS forecasts for 12-hourly rainfall forecasts. Panel (b)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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

U-ecPoint generally increases the number of zeros in the distribution of point-rainfall forecasts to correct ENS's tendency to overpredict small rainfall totals (Haiden et al. 2025). It also increases the rainfall amounts in the wet tail of the rainfall distribution to correct for ENS underestimation of high rainfall values (Haiden et al. 2025).

The MF shape can, however, change significantly according to different weather scenarios at grid-scale (called Grid-box Weather Type, G-WT). The 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 Hewson and Pillosu (2021).

When for a grid-box, the raw ENS predicts, for example, 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. Conversely, when the raw ENS predicts, for example, 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 (Pillosu 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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

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 a grid-box, so that the 24h precipitation distributions are built with 722700 realisations (Table 1, row 5).

3. Methods

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

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

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\mathrm{tp}_{\mathrm{OBS}}(x_i^{\mathrm{th}}) - \mathrm{tp}_{\mathrm{NWP}}(x_i^{\mathrm{th}})\right| \tag{1}$$

$$n = 99 \text{ percentiles}$$

MAE values are expressed in mm and range from 0 (for perfect similarity) to $+\infty$ (for poor similarity). Graphically, the MAE represents the areal difference between the two ECDFs (Figure 2a)[4]. Ninety-nine percentiles (1st to 99th) were used to avoid variations related to wet tail sampling

[4]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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

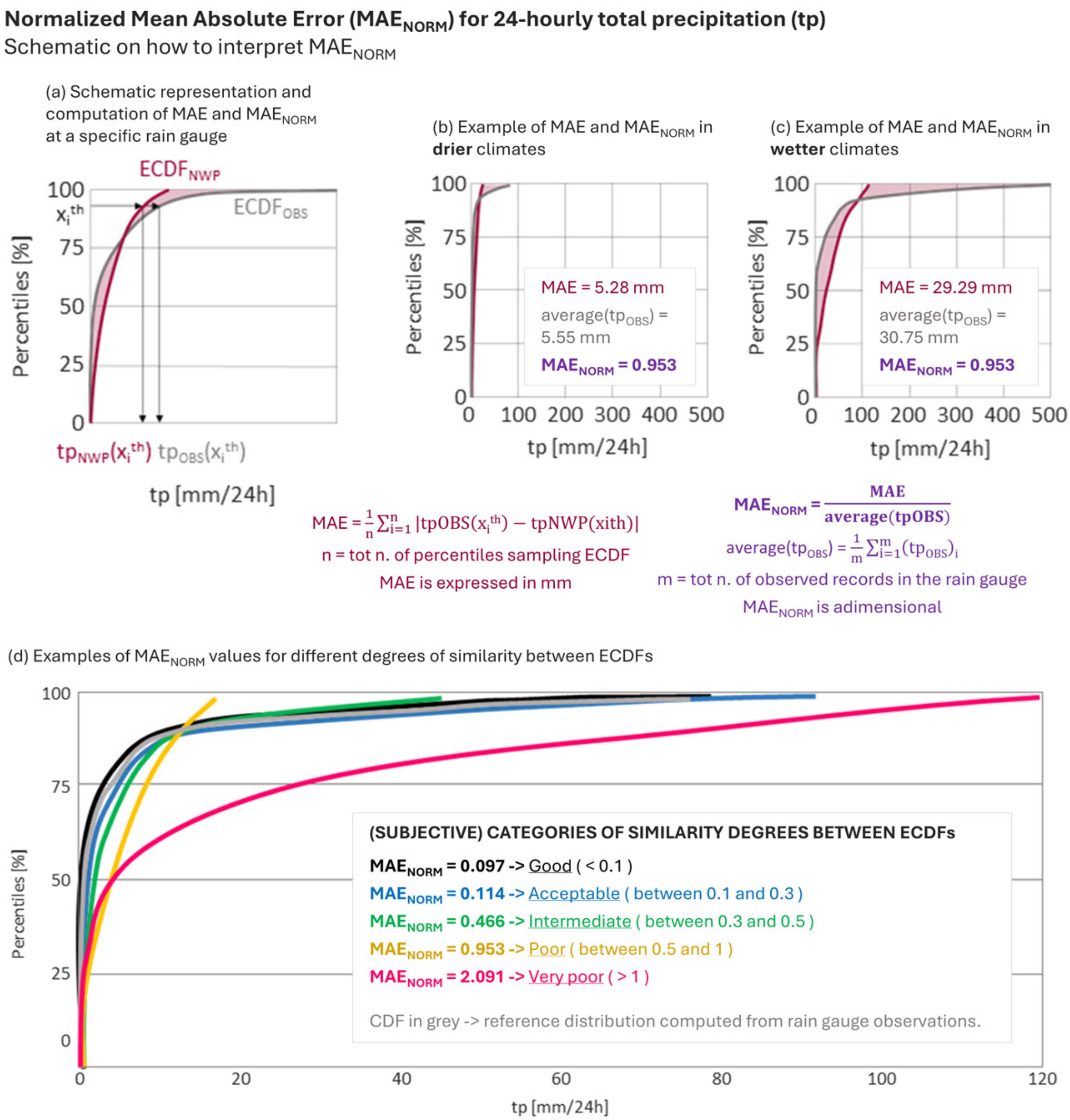


FIG. 2. Panel (a) shows a schematic representation and computation of the Mean Absolute Error (MAE) and the Normalised Mean Absolute Error (MAE_{NORM}) for total precipitation (tp). Panels (b) and (c) show how the values of MAE change for drier and wetter climates, respectively, and how dividing them 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.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

issues in the observational dataset. Moreover, more extreme rainfall events will be considered separately.

Gudmundsson et al. (2012) methodology was, however, adapted to compare ECDFs from different climatologies. MAE values were normalised ($\mathrm{MAE}_{\mathrm{NORM}}$) to avoid having consistently bigger MAE values in wetter climates (see Figure 2b and Figure 2c). The normalisation consists of computing dimensionless coefficients by dividing MAE by the corresponding station's average observed rainfall:

$$\mathrm{MAE}_{\mathrm{NORM}} = \frac{\mathrm{MAE}}{\mathrm{average}(\mathrm{tp}_{\mathrm{OBS}})} \tag{2}$$

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

Formal statistical tests such as Kolmogorov-Smirnov, Cramer-von-Mises, and Anderson-Darling can also assess the similarity between two distributions (Stephens 1974). However, for large sample sizes (as in this study, see Table 1, row 6), the tests' statistical significance levels become extremely sensitive to minor differences between distributions that might not be practically significant (Engmann and Cousineau 2011; Janssen 2000). This is being referred to as ”the problem of practical insignificance” (Kirk 1996), where the test flags differences that are statistically significant but not meaningful in practice, causing the rejection of the null hypothesis (i.e., the two samples come from the same population) when it is nonetheless practically valid. Gudmundsson's approach was tested to assess whether it was as sensitive to sample size as the formal statistical tests. The ECDFs were sampled with 99, 999, 9999, and 99999 percentiles to assess the sensitivity of $\mathrm{MAE}_{\mathrm{NORM}}$ to the choice of sampling resolution. The results showed negligible differences across the range of percentiles tested (not shown). Moreover, Gudmundsson's approach assesses similarity between the observed and NWP-modelled rainfall distributions by comparing the whole ECDFs differently

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

to other formal tests that assess similarity only for specific moments of the distribution, such as the mean, standard deviation, skewness, or specific percentiles (Anthanahalli Nanjegowda and Kulamulla Parambath 2022). Finally, the use of MAE compared to other commonly used scores, such as the Root Mean Squared Error (RMSE), is preferable as it is not so sensitive to outliers (e.g., caused by erroneous observations or atypical events), typically found in the wet tails of the distribution. Nonetheless, the RMSE was computed for the examples shown in Figure 2d, and its property of giving more weight to the larger errors (in the wetter part of the distribution) did not change the ranking obtained with the MAE_{NORM} between the different CDFs in 2d (not shown), reassuring the reader that using MAE instead of RMSE should not change the final picture.

Maps plotting MAE_{NORM} values at different rain gauge locations are shown to compare the performance of the four analysed NWP models. The maps are accompanied by pie charts that summarise, for specific regions, the percentage of locations falling in the five MAE_{NORM} categories 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 considered are North America, South America, Europe, the Mediterranean, Africa, the Arabian Peninsula, Asia, and Oceania.

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

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

To assess how well each modelled dataset captures observed rainfall extremes, we compare large return periods derived from the observational and modelled climatologies. Such estimates are

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

inherently uncertain because the events of interest, i.e., those in the far tail of the distribution, are 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). So no pooling is used here and all observation percentiles are station specific.

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). Return-period estimates are, however, sensitive to the choice of threshold and to the uncertainty in the estimated tail (shape) parameter (Schneider et al. 2021), particularly when extrapolated beyond the range supported by the sample (Scarrott and MacDonald 2012; Pan et al. 2022). Supported by the unprecedentedly large sample size provided by ERA5-ecPoint, 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_{th} largest value in the sample of N independent realisations is estimated as:

$$T = \frac{N}{m \times 365} \quad \text{[years]} \tag{3}$$

where the factor 365 converts daily realisations to years.

The percentile associated with rank m is:

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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

To compute the maximum return period for the considered dataset (rank m = 1, i.e., the largest value in the distribution), the equation 3 and 4 transform to:

$$T = \frac{N}{1 \times 365} \quad \text{[years]} \tag{5}$$

$$P = 100 \times \left(1 - \frac{1}{N}\right) \quad [\%] \tag{6}$$

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

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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

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

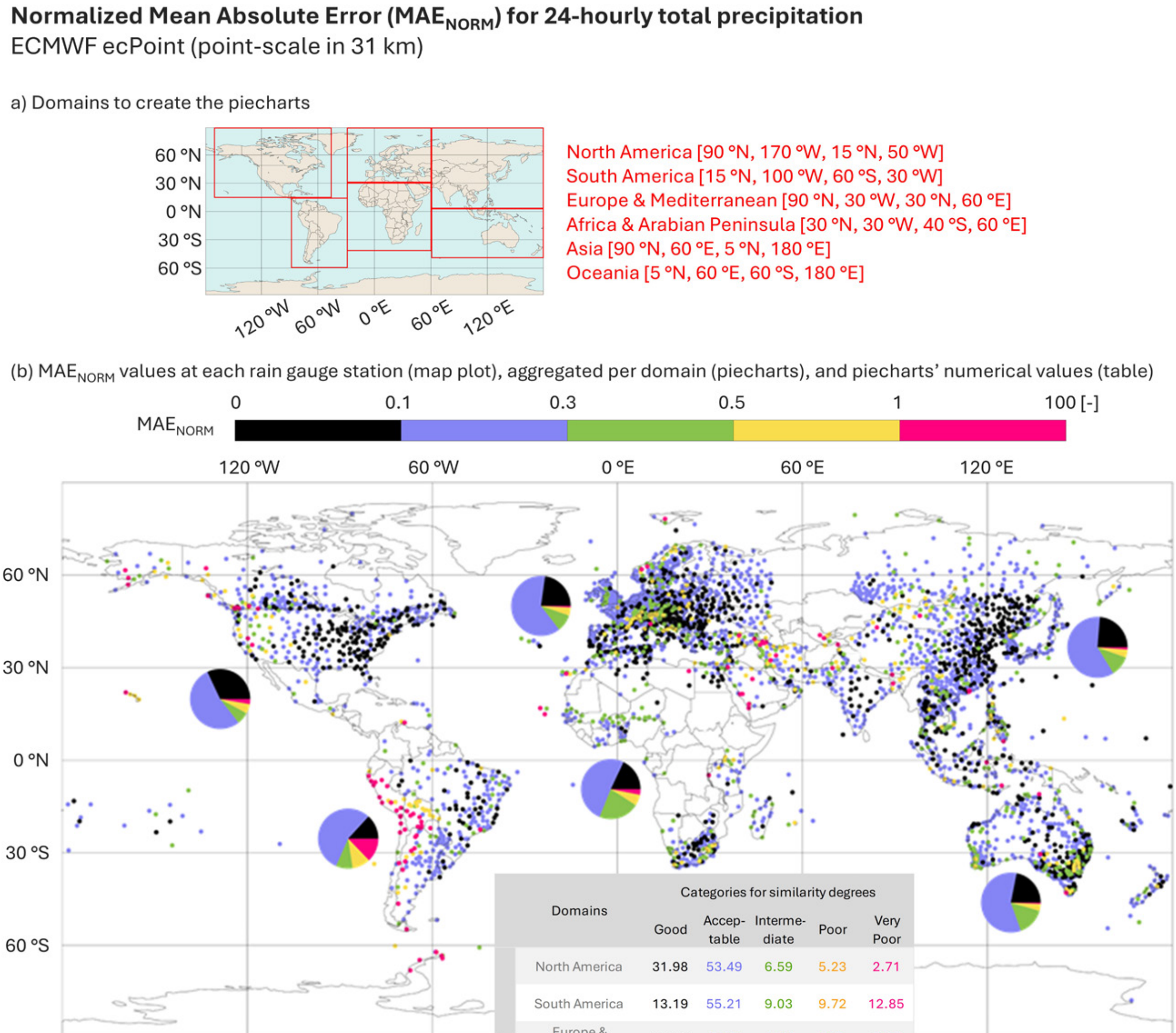
4. Results

a. RQ1: comparison of rainfall distribution climatologies

Out of all NWP-modelled precipitation estimates, ERA5-ecPoint reproduces observed point-precipitation distributions best. This can be seen by the larger percentage of small MAE_{NORM} values (black dots) in Figure 3b compared to Figures 4b (for ERA5-EDA), 5b (for ERA5), and 6b (for ECMWF Reforecasts for 46r1), where there are bigger percentages of larger MAE_{NORM} values (depicted by the coloured dots). ERA-ecPoint increases the number of MAE_{NORM} values in the category of "good similarity" (dots in black) by a factor of 10, 30, and 27 compared to the reforecasts, ERA5, and ERA5-EDA, respectively (see piecharts and inserted tables in Figures 3b to 6b). In the baseline NWP models (ERA5-EDA, ERA5, and ECMWF reforecasts), the proportion of grid points with the very high similarity between observed and modelled precipitation ("black" dots) remains consistently low, below 2% in most regions, except in North America, where reforecasts reach about 4%. In South America, the raw NWP models do not yield any such high-similarity points, whereas applying ERA5-ecPoint boosts this proportion to 13%, with representation along Brazil's eastern coast looking particularly good, in relative terms.

At the opposite extreme, points with poor similarity ("pink" dots) are substantially reduced when using ERA5-ecPoint. Compared to ERA5-EDA, the number of these poorly performing points declines by about 60%, and relative to ERA5 and reforecasts, by about 50%. These improvements are most pronounced in the Arabian Peninsula, Asia, and North America. Although reforecasts also have a lower count of poorly performing points in these areas, they exhibit slightly worse performance in parts of South America, especially the Bolivian Amazon, increasing the proportion of poor-similarity points by 2% and 5% relative to ERA5-EDA and ERA5, respectively. In contrast, ERA5-ecPoint markedly improves this situation in South America, reducing poor-similarity points

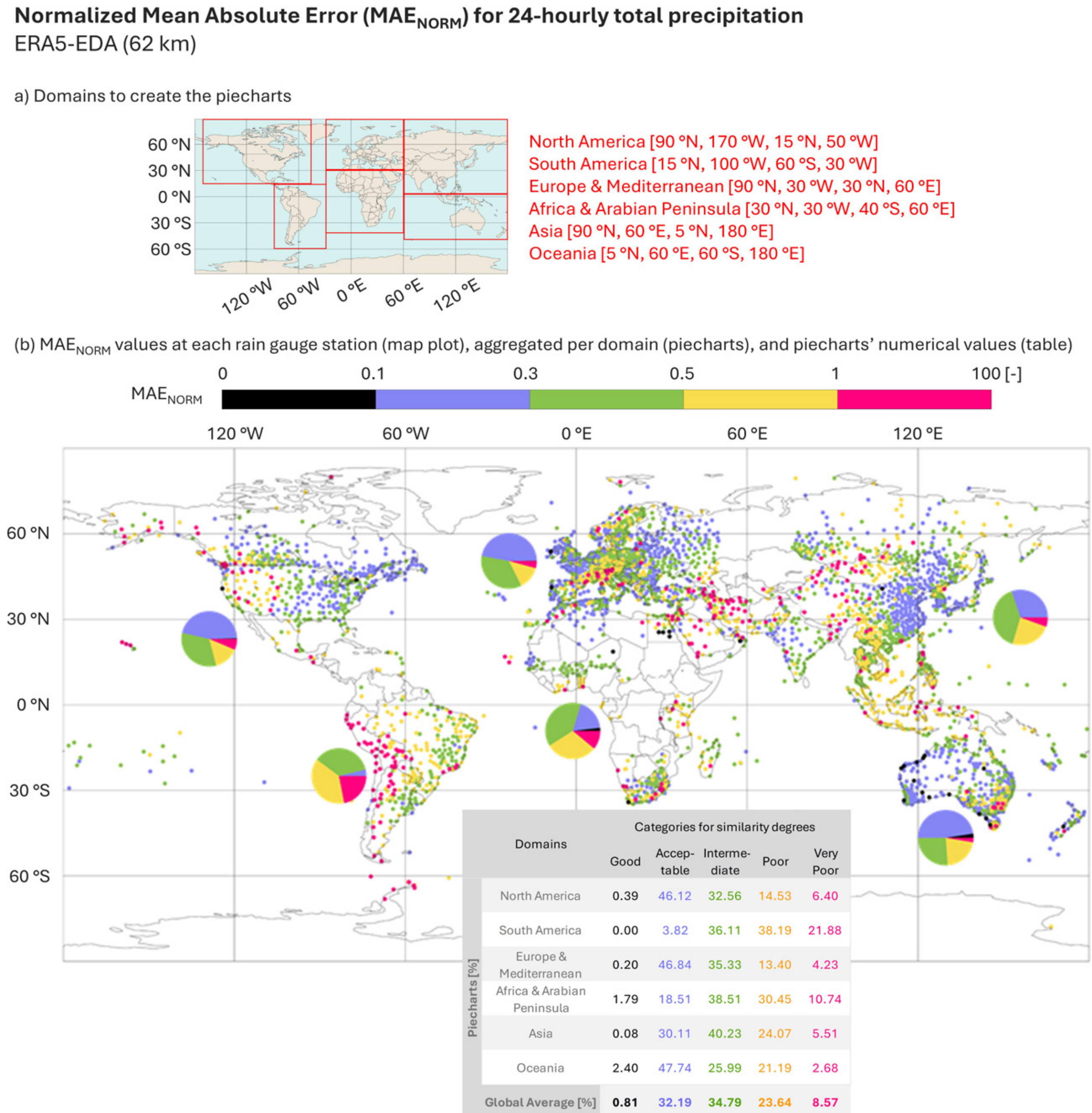
Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.



Piecharts [%] Domains	Good	Acceptable	Intermediate	Poor	Very Poor
	Categories for similarity degrees				
North America	31.98	53.49	6.59	5.23	2.71
South America	13.19	55.21	9.03	9.72	12.85
Europe & Mediterranean	22.77	62.78	9.24	4.36	0.85
Africa & Arabian Peninsula	17.91	50.75	22.09	5.97	2.99
Asia	23.71	60.21	10.66	4.26	1.16
Oceania	21.89	58.62	14.97	3.81	0.71
Global Average [%]	**21.91**	**56.84**	**12.10**	**5.56**	**3.55**

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

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.



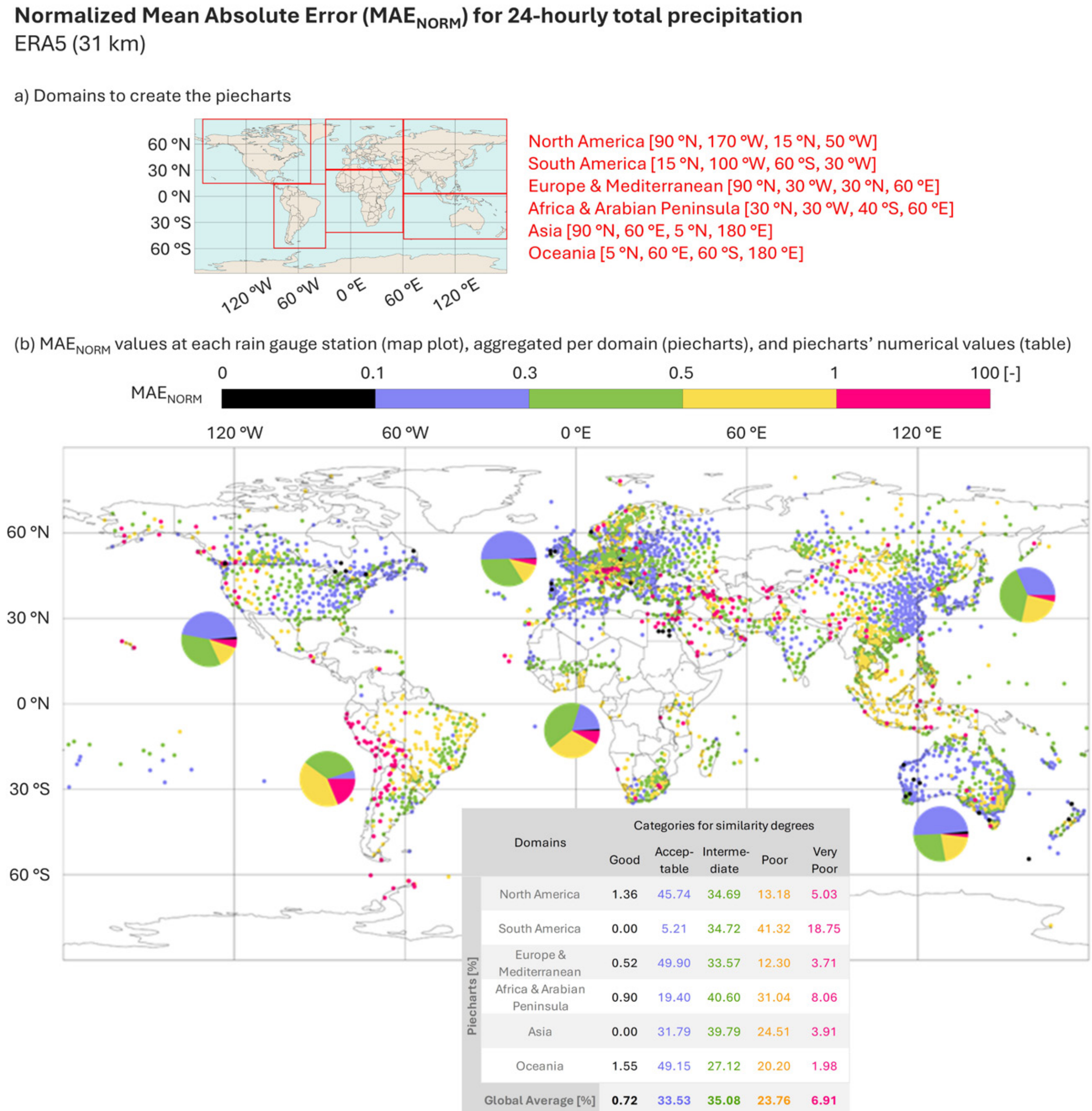
Domains	Categories for similarity degrees: Good	Accep-table	Interme-diate	Poor	Very Poor
North America	0.39	46.12	32.56	14.53	6.40
South America	0.00	3.82	36.11	38.19	21.88
Europe & Mediterranean	0.20	46.84	35.33	13.40	4.23
Africa & Arabian Peninsula	1.79	18.51	38.51	30.45	10.74
Asia	0.08	30.11	40.23	24.07	5.51
Oceania	2.40	47.74	25.99	21.19	2.68
Global Average [%]	0.81	32.19	34.79	23.64	8.57

FIG. 4. Like Figure 3 but for ERA5-EDA (at 62 km spatial resolution).

by 47%, 23%, and 10% compared to reforecasts, ERA5, and ERA5-EDA, respectively. Much of this improvement occurs 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 consis-

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

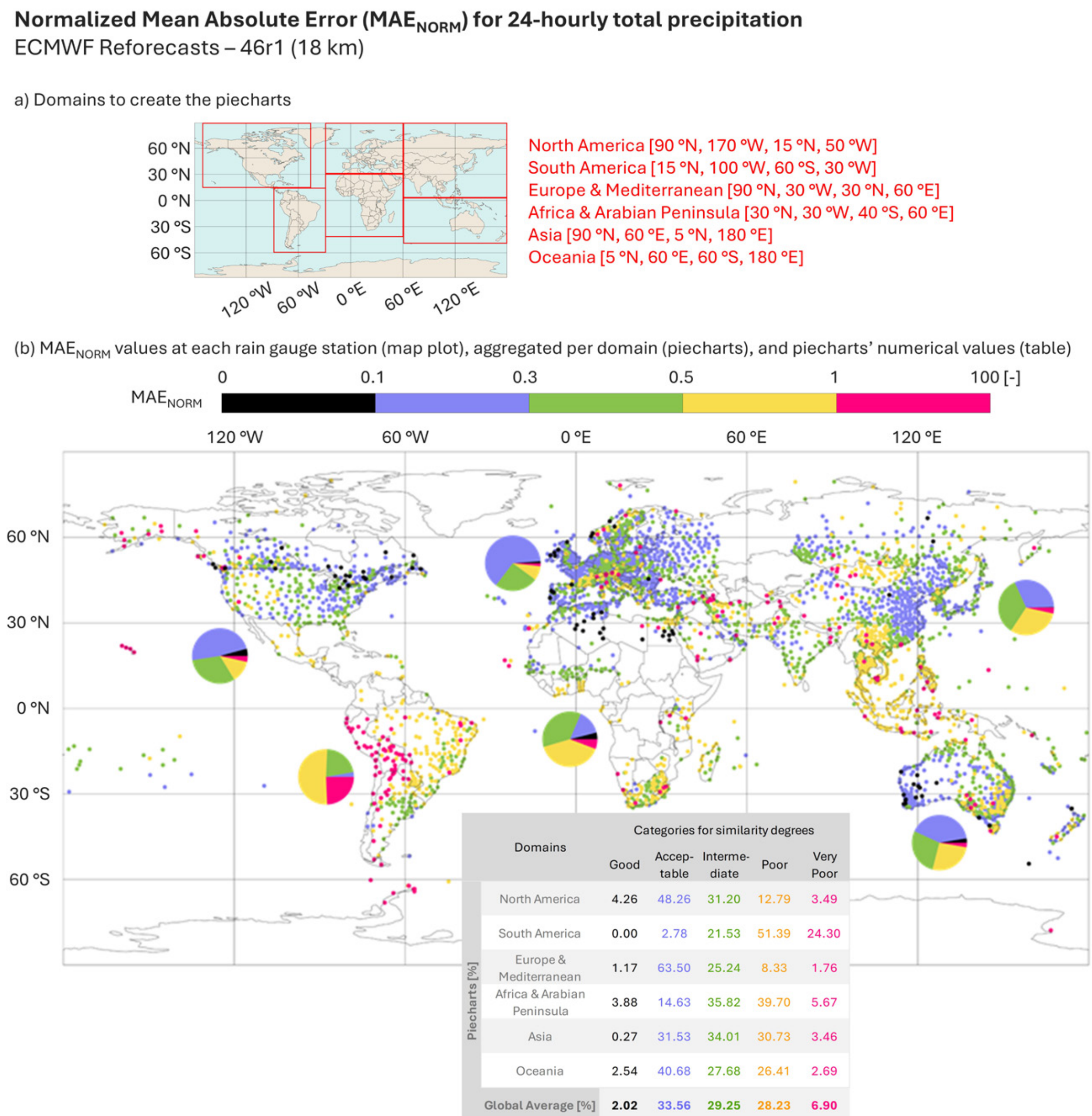


Domains	Good	Accep-table	Interme-diate	Poor	Very Poor
	Categories for similarity degrees				
Piecharts [%]					
North America	1.36	45.74	34.69	13.18	5.03
South America	0.00	5.21	34.72	41.32	18.75
Europe & Mediterranean	0.52	49.90	33.57	12.30	3.71
Africa & Arabian Peninsula	0.90	19.40	40.60	31.04	8.06
Asia	0.00	31.79	39.79	24.51	3.91
Oceania	1.55	49.15	27.12	20.20	1.98
Global Average [%]	0.72	33.53	35.08	23.76	6.91

FIG. 5. Like Figure 3 but for ERA5 (at 31 km spatial resolution).

tently shifts conditions toward a higher level of agreement across all domains. This results in fewer points showing poor similarity 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,

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.



Domains	Categories for similarity degrees: Good	Accep-table	Interme-diate	Poor	Very Poor
Piecharts [%] North America	4.26	48.26	31.20	12.79	3.49
South America	0.00	2.78	21.53	51.39	24.30
Europe & Mediterranean	1.17	63.50	25.24	8.33	1.76
Africa & Arabian Peninsula	3.88	14.63	35.82	39.70	5.67
Asia	0.27	31.53	34.01	30.73	3.46
Oceania	2.54	40.68	27.68	26.41	2.69
Global Average [%]	2.02	33.56	29.25	28.23	6.90

FIG. 6. Like Figure 3 but for ECMWF Reforecasts for 46r1 (at 18 km spatial resolution).

thereby offering a notably better representation of precipitation patterns than the baseline NWP models.

It is worth comparing the observed and the NWP-modelled ECDFs to gain insights into how the distributions differ (Figure 7). Each ECDF (in linear scale) has an insert with the ECDF's

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.

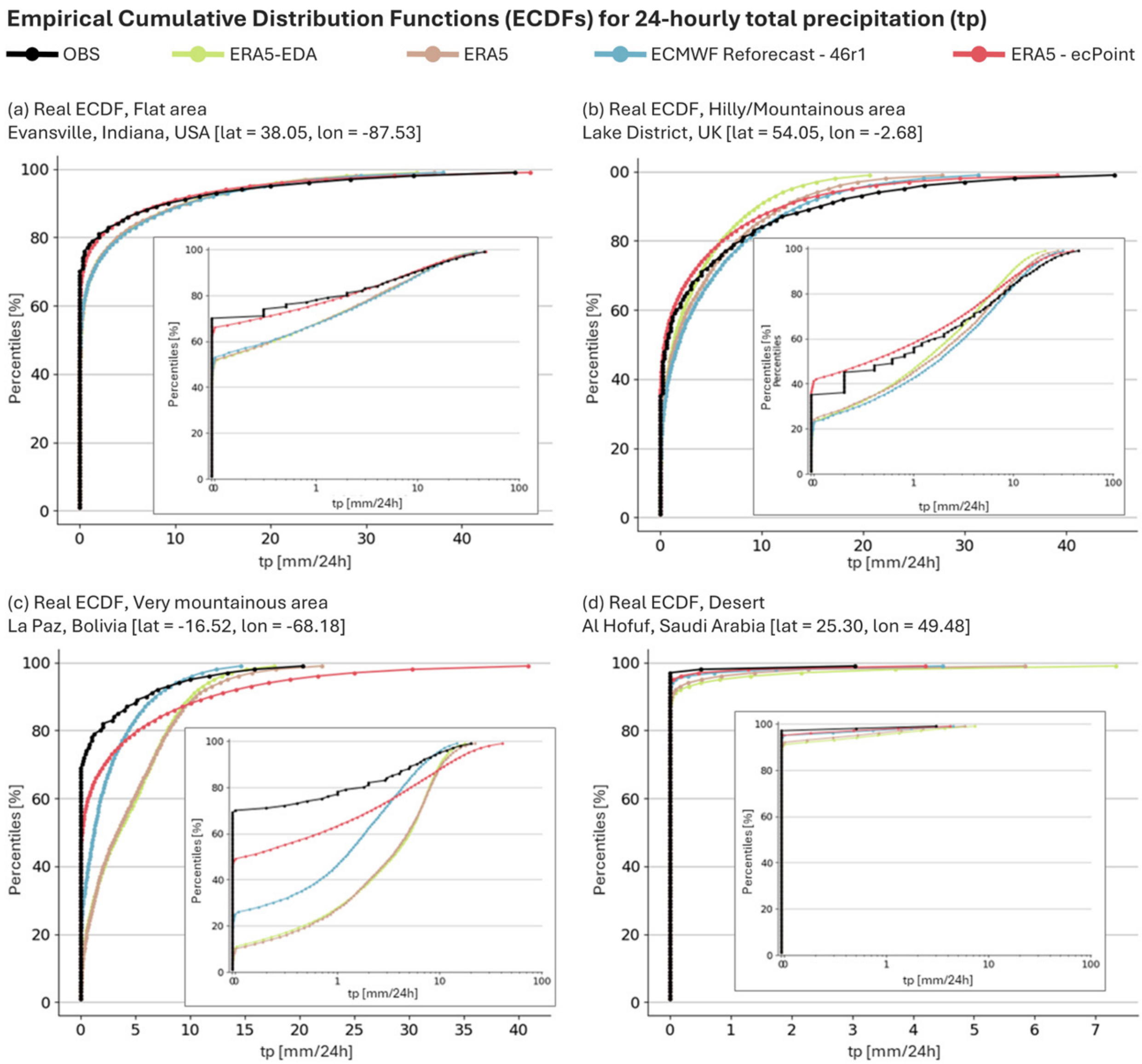


FIG. 7. Empirical Cumulative Distribution Functions (ECDFs) for 24-hourly total precipitation (tp) from rain gauge observations (OBS, in black) and the NWP models ERA5-EDA (in green), ERA5 (in brown), ECMWF Reforecasts-46r1 (in light blue), and ERA5-ecPoint (in coral). Panels (a) to (d) show examples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

x-axis in logarithmic scale to compress/expand the small/high rainfall totals, and see more clearly differences in the distributions.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

In flat areas (Figure 7a), ERA5-ecPoint (in coral) represents the distribution of point-scale precipitation observations better than the baseline raw NWP-models: it captures well the frequency of observed zero precipitation totals (see ECDF in log scale), the growth rate of the precipitation observations[5] (see ECDF in log scale), and the length of the wet tail (going up to the 99th percentile, see ECDFs in linear scale). There are no notable differences between the distributions from ERA5-EDA (in green), ERA5 (in brown), and reforecasts (in blue): they all underestimate, although to different degrees, the frequency of observed zero precipitation totals, and they have similar growth rates, which are greater than that in the observed distribution. They all underestimate the length of the wet tail but to different degrees: in general, ERA5-EDA shows the biggest underestimation, reforecasts show the smallest, and ERA5 falls in between the two.

In hilly/mountainous areas (Figure 7b), ERA5-ecPoint behaves similarly to flat areas. It represents the frequency of zero precipitation totals observed and the growth rate of the precipitation observations well. However, ERA5-ecPoint tends to slightly overestimate the distribution's wet tail of the observed ECDF (in black). Compared to point-rainfall observations, raw NWP models show behaviour similar to that observed in flat areas. The main difference lies in a progression in a better representation of the observed ECDF for NWP models with increasing spatial resolution, i.e., ERA5-EDA at 62 km (in green, in Figure 7b), which shows a worse representation of the observed ECDF compared to ERA5 at 31 km (in brown), and ERA5 shows a worse representation than reforecasts (in blue). This behaviour is seen in other sites too (not shown).

In very mountainous areas (Figure 7c), all NWP models fail to represent the observed ECDFs. This should not be that surprising as the observations used to train ERA5-ecPoint, and indeed to validate all representations, come primarily from valleys and hilly areas. First, all the NWP model versions underestimate the frequency of observed zero precipitation totals. ERA5-ecPoint tends to double such a frequency, but it does not reach the values in the observed ECDFs. The ECDFs from raw NWP models show a growth rate that is too large compared to the observed ECDFs, while ERA5-ecPoint also improves on that. Finally, while the raw NWP models tend to slightly underestimate the tail length of the observed ECDFs (with ERA5 providing the best representation out of the three models), ERA5-ecPoint tends to overestimate it.

In desert areas (Figure 7d), all NWP models represent the observed ECDFs well, apart from the wet tails that tend to all be overestimated. The overestimation is reduced with the increase in the

[5]Growth rate here is intended as the rate of change of the logarithm of precipitation totals.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

spatial resolution of the NWP models, with ERA5-ecPoint representing the actual length of the wet tail best.

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

1) Comparison of the 10-year period

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

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

In Asia, there is a varied picture. The raw NWP models highlight the wetter climatologies of India (especially the Northeast regions), East China, Japan, Southeast Asia, and the Malay Archipelago. However, they do not reach the peaks of 300-500 mm/24h seen in the observations. ERA5-ecPoint

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

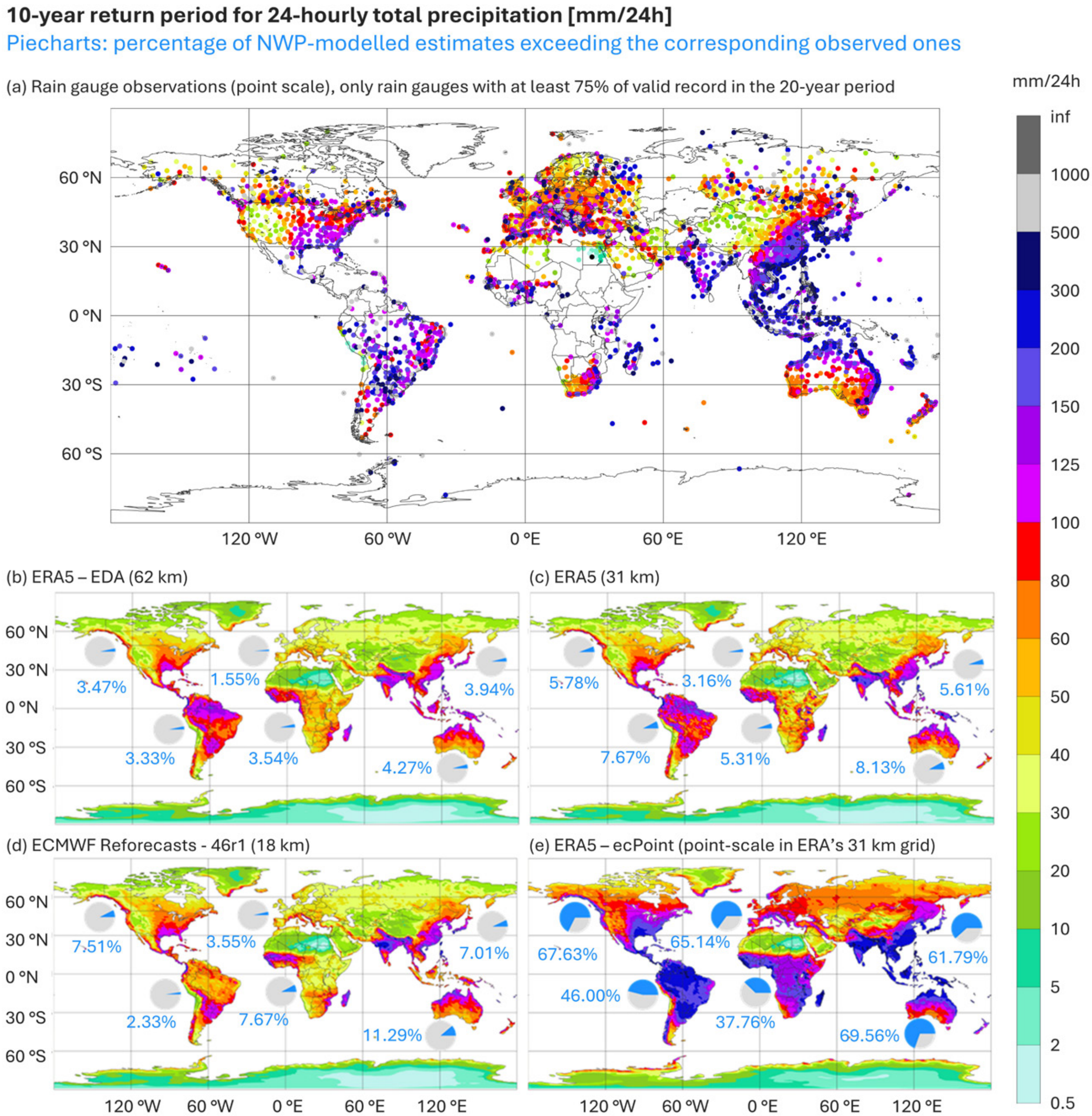


Fig. 8. Panel (a) displays the 10-year return period (computed from the 99.9726$_{th}$ percentile) for 24-hourly total precipitation from rain gauge observations, calculated over the 20-year period between 2000 and 2019, and using only rain gauges with at least 75% of valid records. Panels (b) to (e) show the 10-year return period for NWP-modelled 24-hourly total precipitation: ERA5-EDA (62 km), ERA5 (31 km), ECMWF Reforecasts-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.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.

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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

2) Case study: Storm Vaia in Italy (28th of October 2018)

We now examine the case of widespread (flash) flooding in Italy on the 28th of October 2018 (Figure 9). This event is part of a weather system that persisted over different parts of Italy between the end of October and the beginning of November 2018. It is called Storm Vaia. In the observations (Figure 9a), one can see extreme precipitation amounts between 300-400 mm/24h over Veneto (north-east), up to 200 mm/24h over Lombardi (North) and Liguria (North-East), up to 240 mm/24h in Lazio (centre), up to 130 mm/24h in Puglia (Southwest), and up to 260 mm/24h in Calabria (Southeast). ERA5-EDA (Figure 9b), ERA5 (Figure 9c), and reforecasts (Figure 9d) provide a good signal on which might be the wetter areas in Italy for that day, apart from the south of Italy, that does not stand out as a possible area at risk of extreme precipitation. The precipitation peaks over the Italian Peninsula increase with the increasing spatial resolution of the NWP models, but they do not reach the observed extreme precipitation totals. ERA5-EDA estimated a maximum total of 100 mm/24h, and ERA5 pushed the estimated peaks to 150 mm/24h over Veneto. Reforecasts increased the precipitation peaks in Veneto and Lazio up to 200 mm/24h, but precipitation in Liguria, Puglia, and Calabria remains highly underestimated. ERA5-ecPoint (Figure 9e) represents better the areas where the precipitation peaks were observed. In the north (Figure 9f, Northern Italy), where the storm created the biggest impacts, roughly 1% of the rainfall observations exceeded the 99th percentile of ERA5-ecPoint (red dots), indicating reliable point-scale rainfall estimates. In the rest of the peninsula (Figure 9f, Central and Southern Italy), 6% of the observations exceed the ERA5-ecPoint estimates, indicating an under-prediction of point-scale rainfall over Le Marche, Puglia, and Calabria. In this specific case, the location of the red dots along coastlines indicates underestimation primarily due to the known issue of convective cells generated over the sea not moving onto land (Bechtold et al. 2014).

5. Concluding Remarks

Characterising the full distribution of rainfall totals, from light to extreme, is a long-standing challenge in meteorology. Yet, it is essential for various fields of environmental research that require a more accurate representation of point rainfall, such as flood forecasting, water management, and disaster preparedness under climate change. Rainfall's high spatial variability makes such characterisation reliant on point-scale measurements, traditionally supplied by rain gauges. Yet,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

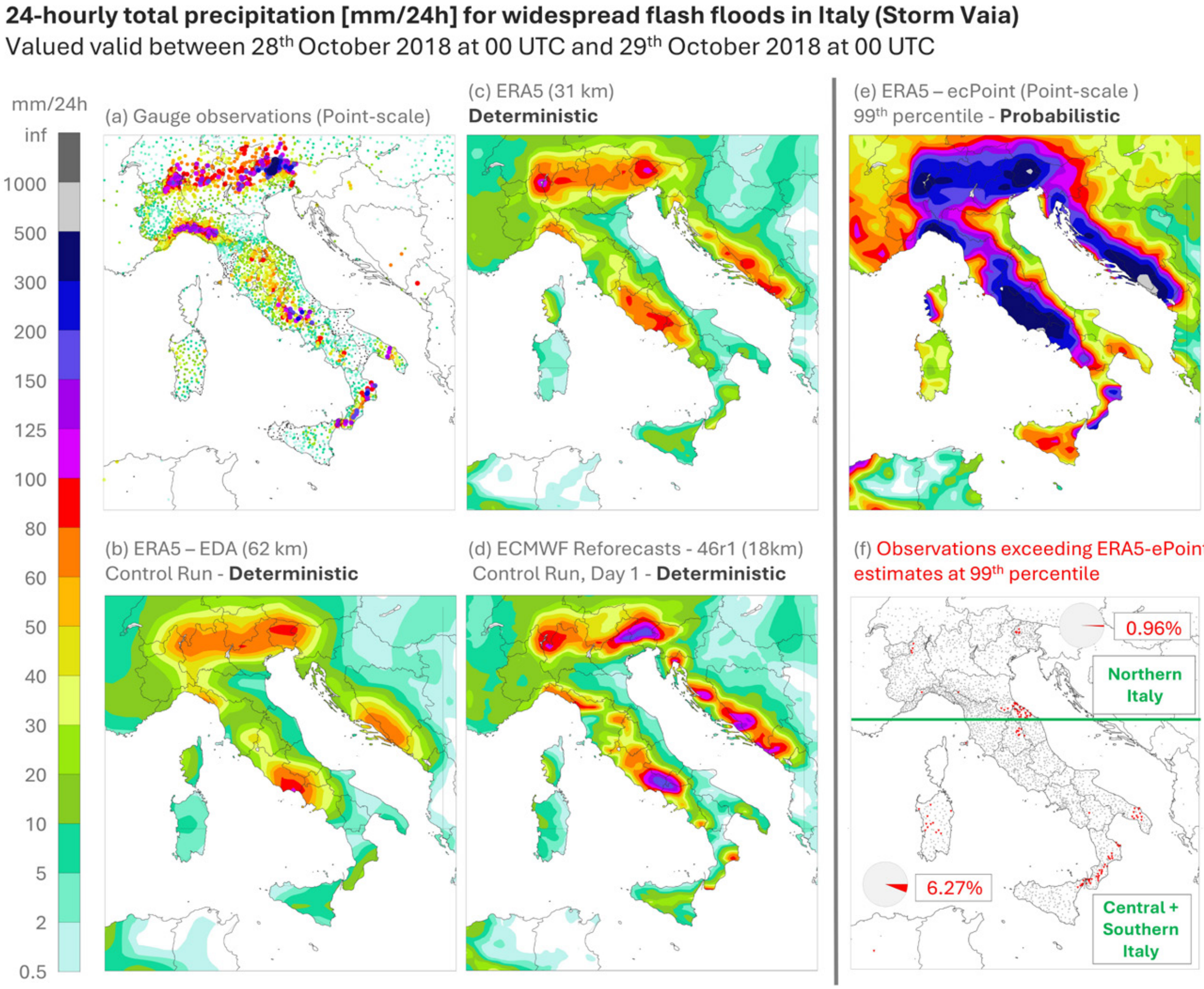


FIG. 9. Widespread (flash) flooding in Italy on the 28th of October 2018 due to Storm Vaia. Panel (a) represents the rain gauge observations. Panels (b) to (d) show the deterministic rainfall estimates, respectively, for ERA5-EDA at 62 km (from control run), ERA at 31 km (single realisation), and ECMWF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

the uneven spatial distribution of rain gauges and the lack of long, multi-decadal records prevent seamless characterisation of rainfall totals over space and time, motivating the use of gridded, modelled datasets such as the ERA5 reanalysis. However, due to model parametrisations and

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

spatial resolution, raw gridded, modelled datasets do not provide a good representation of point-scale rainfall distributions.

This study provides a systematic, global verification of *how well* NWP models represent the distribution of point-rainfall observations. It considered point-rainfall observations over 20 years and four different modelled, gridded datasets with distinct spatial resolutions: ERA5-EDA (62 km), ERA5 (31 km), ECMWF Reforecasts for model cycle 46r1 (18 km), and ERA5-ecPoint (point-scale but provided over ERA5's grid at 31 km). Among the tested models, this study shows that ERA5-ecPoint, a statistically post-processed version of ERA5, is instead capable of reproducing much more faithfully the climatological distributions of point-rainfall.

We would advocate adoption of the ERA5-ecPoint dataset for multiple applications. Firstly, the embedded climatological information can enhance long-term strategic planning, to deliver more resilient stormwater infrastructure and urban drainage systems, which are also facing increasing pressure from the extreme rainfall intensification due to climate change. Secondly, better rainfall representation can also facilitate more efficient water resource management, for reservoirs and irrigation, and enhanced distribution for agriculture, power generation, and urban water supply systems. And thirdly, flash flood warning systems, that aim to safeguard communities through a targetted emergency response, can clearly benefit: point-scale return period thresholds provide a form of flash flood benchmark against which real-time point-scale rainfall forecasts can be compared. Hydrologically speaking, more accurate rainfall estimates at local scale are crucial for predicting runoff and streamflow dynamics, particularly in catchments prone to flash floods.

In all the above applications the ERA5-ecPoint characteristics which will be key for users are its ability to capture the full spectrum of climatological rainfall values, including zeros and extremes, and its applicability over a continuous global domain that includes areas short of or devoid of raingauge data.

We would also highlight that ERA5-ecPoint facilitates the estimation of rainfall thresholds with significantly longer return periods than have been presented in this study (say up to a 1000-year return period, as noted in Table 1, row 5, column 7). Hewson et al. (2024) have shown, using 2023 Storm Daniel in Libya as an example, that ERA5-ecPoint data 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 truly exceptional rainfall,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

facilitating preparedness for the unseen (Heinrich et al. 2024; Ommer et al. 2024) or for levels so infrequent that they have faded from collective memory (Ludwig et al. 2023; Merz et al. 2024).

Nonetheless, some caution is needed when generalising. While ERA5-ecPoint demonstrates strong performance in estimating point-rainfall totals overall, one should again note that the verification dataset contains large regions with little or no raingauge data, and whilst providing estimates for those regions is rightly cited as a strength one cannot truly know how good those representations actually are. In this regard, the verification for complex mountainous terrains (e.g., the Andes) stands out. The study showed that for such areas the post-processed reanalysis is probably underestimating the frequency of zero totals, and overestimating wet tail length. This outcome should not surprise, as ERA5-ecPoint is post-processed with observations preferentially situated, for practical historical reasons, at lower altitudes in such regions. Nevertheless, the ERA5-ecPoint rainfall distribution growth rates do still closely align with those observations we do have, indicating that the post-processing system can make meaningful adjustments even when the training dataset for a given category (in this case, complex mountainous terrain) is small. This it does, in practice, by pooling with data from other topographic classes; whilst a sub-grid orography variable is included in the ERA5-ecPoint decision tree, this could not be used in as comprehensive a way as one might have liked, due to the sample size issues. So further refinement of post-processed forecasts could be desirable, ideally by incorporating more rain gauge data if/when available. One could also strive to clarify how the system is drawing from other categories when training data for a specific category of interest is scarce.

Another geographical region where ERA5-ecPoint representation was sub-optimal, with extremes overestimated, was the high-altitude and relatively dry western parts of the USA, where mountain barriers can block ingress from 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 from physics scheme considerations, 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 scenarios (see also Gascón et al. (2025)). 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 moister boundary layer scenarios, found else-

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

where. This could thus be another focal point for future work. While ecPoint's remote calibration approach has shown some very significant benefits compared to a purely local approach, as here and in Hewson and Pillosu (2021), there can evidently be some local downsides.

The climatological distributions evaluated here pool observations across all months of the year, so without seasonal stratification. This design choice was deliberate: the primary objective was to assess each dataset's ability to represent the overall statistical distribution of 24-hourly rainfall accumulations across a wide range of climatic regimes, rather than to diagnose season-dependent performance. However, pooling across seasons may mask compensating biases (for instance, a model that underestimates in winter but overestimates in summer could produce an aggregate distribution that appears well-calibrated despite meaningful seasonal errors). Similarly, in regions where a substantial fraction of annual precipitation falls as snow, the rain gauge observations may represent that component (i.e., its liquid water equivalent) less well, lending more focus to the liquid phase in our analysis. Decomposing the verification by season, and potentially by dominant precipitation type, would provide a more complete picture of each dataset's strengths and weaknesses across different meteorological regimes. This could represent a natural extension of this work, albeit with the downside of reduced sample sizes.

Despite the above limitations, this study shows that statistical post-processing approaches such as ecPoint offer the practical and operationally viable means of bridging the scale gap between NWP grid-box averages and point-scale observations that large ensembles and convection-permitting resolution models (~1 km) cannot. The first of these addresses only grid-scale forecast uncertainty, while the second, although able to explicitly resolve sub-grid rainfall variability, remains computationally prohibitive for global, multi-decadal applications. The improved performance of ERA5-ecPoint in representing point-scale rainfall totals over raw gridded, modelled estimates emphasises post-processing's critical role. The effectiveness of ecPoint, however, remains contingent on the quality of the underlying NWP models it post-processes. Without reasonably accurate raw NWP estimates at grid-scale, the skill of the post-processed fields would be diminished. Moving forward, the authors advocate enhancing the spatial resolution and the skill of raw NWP models alongside continual improvement of post-processing techniques such as ecPoint to further reduce errors in estimates of the full distribution of point-rainfall totals worldwide. Such improvements

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

will be of particular practical significance as climate change intensifies the severity of extreme rainfall, and its attendant impacts in many spheres.

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

Data availability statement. Data and/or the codes used to generate the figures that are incorporated into this manuscript will be made available upon reasonable request.

References

Ahmed, A., G. Yildirim, K. Haddad, and A. Rahman, 2023: Regional Flood Frequency Analysis: A Bibliometric Overview. *Water*, **15 (9)**, https://doi.org/10.3390/w15091658.

Alexandridis, V., S. Stefanidis, and S. Dafis, 2023: Evaluation of ERA5 and ERA5-land reanalysis precipitation data with rain gauge observations in Greece. *Environmental Sciences Proceedings*, **26 (1)**, 104, https://doi.org/10.3390/ENVIRONSCIPROC2023026104.

Anand, B., and D. Karunanidhi, 2020: Long-term spatial and temporal rainfall trend analysis using gis and statistical methods in lower Bhavani basin, Tamil Nadu, India. *Indian Journal of Geo-Marine Sciences*, **49 (3)**, 419–427.

Anthanahalli Nanjegowda, R., and S. Kulamulla Parambath, 2022: A novel bias correction method for extreme rainfall events based on l-moments. *International Journal of Climatology*, **42 (1)**, 250–264, https://doi.org/10.1002/JOC.7242.

Asadieh, B., and N. Y. Krakauer, 2015: Global trends in extreme precipitation: Climate models versus observations. *Hydrology and Earth System Sciences*, **19 (2)**, 877–891, https://doi.org/10.5194/HESS-19-877-2015.

Balasundram, S. K., R. R. Shamshiri, S. Sridhara, and N. Rizan, 2023: The role of digital agriculture in mitigating climate change and ensuring food security: An overview. *Sustainability*, **15 (6)**, https://doi.org/10.3390/SU15065325.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

Bechtold, P., N. Semane, P. Lopez, J. P. Chaboureau, A. Beljaars, and N. Bormann, 2014: Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, **71 (2)**, 734–753, https://doi.org/10.1175/JAS-D-13-0163.1.

Bottazzi, M., and Coauthors, 2024: High performance computing to support land, climate, and user-oriented services: The HIGHLANDER Data Portal. *Meteorological Applications*, **31 (2)**, e2166, https://doi.org/10.1002/met.2166.

Boé, J., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, **27 (12)**, 1643–1655, https://doi.org/10.1002/joc.1602.

Coles, S., 2001: Threshold Models. *An Introduction to Statistical Modeling of Extreme Values*, Springer, 74–91, https://doi.org/10.1007/978-1-4471-3675-0_4.

Cuo, L., T. C. Pagano, and Q. J. Wang, 2011: A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. *Journal of Hydrometeorology*, **12 (5)**, 713–728, https://doi.org/10.1175/2011JHM1347.1.

Di Curzio, D., A. Di Giovanni, R. Lidori, M. Montopoli, and S. Rusi, 2022: Comparing rain gauge and weather radar data in the estimation of the pluviometric inflow from the Apennine ridge to the Adriatic coast (Abruzzo region, central Italy). *Hydrology*, **9 (12)**, 225, https://doi.org/10.3390/HYDROLOGY9120225.

Donat, M. G., L. V. Alexander, N. Herold, and A. J. Dittus, 2016: Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. *Journal of Geophysical Research*, **121 (19)**, 11 174–11 189, https://doi.org/10.1002/2016JD025480.

Doocy, S., A. Daniels, S. Murray, and T. D. Kirsch, 2013: The human impact of floods: A historical review of events 1980-2009 and systematic literature review. *PLoS Currents*, https://doi.org/10.1371/CURRENTS.DIS.F4DEB457904936B07C09DAA98EE8171A.

Engmann, S., and D. Cousineau, 2011: Comparing distributions: The two-sample anderson-darling test as an alternative to the kolmogorov-smirnoff test. *Journal of Applied Quantitative Methods*, **6 (3)**, URL https://jaqm.ro/issues/volume-6,issue-3/1_engmann_cousineau.php.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

Ensor, L. A., and S. M. Robeson, 2008: Statistical characteristics of daily precipitation: Comparisons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, **47 (9)**, 2468–2476, https://doi.org/10.1175/2008JAMC1757.1.

Espinosa, L. A., M. M. Portela, and S. Gharbia, 2024: Assessing changes in exceptional rainfall in Portugal using ERA5-land reanalysis data (1981/1982–2022/2023). *Water*, **16 (5)**, 628, https://doi.org/10.3390/W16050628.

Gascón, E., L. Magnusson, T. Hewson, J. Rey, and J. Rodriguez, 2025: Extreme precipitation in Spain's Valencia region. *ECMWF Newsletter*, **(183)**.

Gimeno, L., R. Sorí, M. Vázquez, M. Stojanovic, I. Algarra, J. Eiras-Barca, L. Gimeno-Sotelo, and R. Nieto, 2022: Extreme precipitation events. *Wiley Interdisciplinary Reviews: Water*, **9 (6)**, https://doi.org/10.1002/WAT2.1611.

Giorgos, N., P. T. Nastos, and Y. Kapsomenakis, 2024: Creating high-resolution precipitation and extreme precipitation indices datasets by downscaling and improving on the ERA5 reanalysis data over greece. *Eng*, **5 (3)**, 1885–1904, https://doi.org/10.3390/ENG5030101.

Göber, M., E. Zsótér, and D. S. Richardson, 2008: Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. *Meteorological Applications*, **15 (3)**, 359–365, https://doi.org/10.1002/met.78.

Gomis-Cebolla, J., V. Rattayova, S. Salazar-Galán, and F. Francés, 2023: Evaluation of ERA5 and ERA5-land reanalysis precipitation datasets over spain (1951–2020). *Atmospheric Research*, **284**, https://doi.org/10.1016/J.ATMOSRES.2023.106606.

Guan, H., and Coauthors, 2022: GEFSv12 Reforecast Dataset for Supporting Subseasonal and Hydrometeorological Applications. *Monthly Weather Review*, **150 (3)**, 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrology and Earth System Sciences*, **16 (9)**, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.

Gupta, V., M. K. Jain, P. K. Singh, and V. Singh, 2020: An assessment of global satellite-based precipitation datasets in capturing precipitation extremes: A comparison with observed precipitation dataset in India. *International Journal of Climatology*, **40 (8)**, 3667–3688, https://doi.org/10.1002/JOC.6419.

Haiden, T., and S. Duffy, 2016: Use of high-density observations in precipitation verification. *ECMWF Newsletter*, **(147)**, 20–25, https://doi.org/10.21957/hsacrdem.

Haiden, T., M. Janousek, F. Vitart, F. Prates, M. Maier-Gerber, C. W. Y. Li, and M. Chevallier, 2025: Evaluation of ECMWF forecasts. *ECMWF Technical Memoranda*, **931**, https://doi.org/10.21957/51e665e5c8.

Haile, G. G., and Coauthors, 2020: Projected impacts of climate change on drought patterns over east africa. *Earth's Future*, **8 (7)**, https://doi.org/10.1029/2020EF001502.

Hamill, T. M., and Coauthors, 2022: The reanalysis for the global ensemble forecast system, version 12. *Monthly Weather Review*, **150 (1)**, 59–79, https://doi.org/10.1175/MWR-D-21-0023.1.

Heinrich, D., E. Stephens, and E. Coughlan de Perez, 2024: More than magnitude: Towards a multidimensional understanding of unprecedented weather to better support disaster management. *Water Security*, **23**, 100 181, https://doi.org/10.1016/J.WASEC.2024.100181.

Herold, N., A. Behrangi, and L. V. Alexander, 2017: Large uncertainties in observed daily precipitation extremes over land. *Journal of Geophysical Research: Atmospheres*, **122 (2)**, 668–681, https://doi.org/10.1002/2016JD025842.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hewson, T., 2024: Capturing extreme rainfall events. *ECMWF Newsletter*, **(178)**.

Hewson, T., and F. Pillosu, 2021: A low-cost post-processing technique improves weather forecasts around the world. *Communications Earth and Environment*, **2 (1)**, https://doi.org/10.1038/s43247-021-00185-9.

Hewson, T., F. Pillosu, E. Gascòn, and M. Vučković, 2023: Post-processing ERA5 output with ecPoint. *ECMWF Newsletter*, **(176)**.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

Hewson, T., and Coauthors, 2024: Medicane Daniel: an extraordinary cyclone with devastating impacts. *ECMWF Newsletter*, **(179)**, https://doi.org/10.21957/th3wxk861d.

Hosking, J. R. M., and J. R. Wallis, 1997: *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511529443.

Hossain, I., S. Gato-Trinidad, M. Imteaz, and S. Rayburg, 2024: Future scenarios of design rainfall due to upcoming climate changes in NSW, Australia. *Atmosphere*, **15 (9)**, 1101, https://doi.org/10.3390/ATMOS15091101.

IPCC, 2023: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 35–115 pp., https://doi.org/10.59327/IPCC/AR6-9789291691647.

Janmohammadi, M., and N. Sabaghnia, 2023: Strategies to alleviate the unusual effects of climate change on crop production: a thirsty and warm future, low crop quality. a review. *Biologija*, **69 (2)**, 121–133, https://doi.org/10.6001/BIOLOGIJA.2023.69.2.1.

Janssen, A., 2000: Global power functions of goodness of fit tests. *The Annals of Statistics*, **28 (1)**, 239–253.

Kalnay, E., and Coauthors, 1996: The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–472, https://doi.org/10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2.

Khaliq, M. N., T. B. M. J. Ouarda, J. C. Ondo, P. Gachon, and B. Bobée, 2006: Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, **329 (3)**, 534–552, https://doi.org/10.1016/j.jhydrol.2006.03.004.

Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, **98 (1)**, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.

Kim, S., K. Sung, J.-Y. Shin, and J.-H. Heo, 2025: At-Site Versus Regional Frequency Analysis of Sub-Hourly Rainfall for Urban Hydrology Applications During Recent Extreme Events. *Water*, **17 (15)**, https://doi.org/10.3390/w17152213.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.

Kirk, R. E., 1996: Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, **56 (5)**, 746–759, https://doi.org/10.1177/0013164496056005002.

Kochendorfer, J., and Coauthors, 2020: Undercatch Adjustments for Tipping-Bucket Gauge Measurements of Solid Precipitation. *Journal of Hydrometeorology*, **21 (6)**, 1193–1205, https://doi.org/10.1175/JHM-D-19-0256.1.

Lamprecht, A., and Coauthors, 2021: Disentangling anthropogenic drivers of climate change impacts on alpine plant species: Alps vs. mediterranean mountains. final report. *Earth System Sciences (ESS)*, 1–72, https://doi.org/10.1553/ESS-MEDIALPSS1.

Lanza, L. G., and L. Stagi, 2008: Certified accuracy of rainfall data as a standard requirement in scientific investigations. *Advances in Geosciences*, **16**, 43–48, https://doi.org/10.5194/ADGEO-16-43-2008.

Laouacheria, F., S. Kechida, and M. Chabi, 2019: Modelling the impact of design rainfall on the urban drainage system by storm water management model. *Journal of Water and Land Development*, **40 (1)**, 119–125, https://doi.org/10.2478/JWLD-2019-0013.

Lavers, D. A., S. Harrigan, and C. Prudhomme, 2021: Precipitation biases in the ECMWF integrated forecasting system. *Journal of Hydrometeorology*, **22 (5)**, 1315–1334, https://doi.org/10.1175/jhm-d-20-0308.1.

Lavers, D. A., A. Simmons, F. Vamborg, and M. J. Rodwell, 2022: An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, **148 (748)**, 3152–3165, https://doi.org/10.1002/QJ.4351.

Liang, X. Z., 2022: Extreme rainfall slows the global economy. *Nature*, **601 (7892)**, 193–194, https://doi.org/10.1038/d41586-021-03783-x.

Ludwig, P., and Coauthors, 2023: A multi-disciplinary analysis of the exceptional flood event of july 2021 in central Europe - part 2: Historical context and relation to climate change. *Natural Hazards and Earth System Sciences*, **23 (4)**, 1287–1311, https://doi.org/10.5194/NHESS-23-1287-2023.

Maurya, S. K., A. Kalhapure, V. K. Verma, A. Tiwari, C. Chaubey, D. K. Maurya, and M. Kumar, 2024: Irrigation scheduling and cultivar management for increasing water productivity under

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-25-0019.1.

dryland condition: A review. *International Journal of Environment and Climate Change*, **14 (1)**, 461–470, https://doi.org/10.9734/IJECC/2024/V14I13856.

Merz, B., and Coauthors, 2024: Spatial counterfactuals to explore disastrous flooding. *Environmental Research Letters*, **19 (4)**, https://doi.org/10.1088/1748-9326/AD22B9.

Muñoz-Sabater, J., and Coauthors, 2021: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, **13 (9)**, 4349–4383, https://doi.org/10.5194/ESSD-13-4349-2021.

Ommer, J., J. Neumann, M. Kalas, S. Blackburn, and H. L. Cloke, 2024: Surprise floods: The role of our imagination in preparing for disasters. *Natural Hazards and Earth System Sciences*, **24 (8)**, 2633–2646, https://doi.org/10.5194/NHESS-24-2633-2024.

Pan, X., A. Rahman, K. Haddad, and T. B. M. J. Ouarda, 2022: Peaks-over-threshold model in flood frequency analysis: A scoping review. *Stochastic Environmental Research and Risk Assessment*, **36 (9)**, 2419–2435, https://doi.org/10.1007/s00477-022-02174-6.

Pillosu, F. M., T. Hewson, E. Stephens, C. Prudhomme, and H. Cloke, 2025: Does a multivariate approach enhance univariate grid-to-point post-processed rainfall forecasts? A comparative analysis. *ECMWF Technical Memoranda*, **924**, 1–21, https://doi.org/10.21957/eb6c62b7af.

Pollock, M. D., and Coauthors, 2018: Quantifying and Mitigating Wind-Induced Undercatch in Rainfall Measurements. *Water Resources Research*, **54 (6)**, 3863–3875, https://doi.org/10.1029/2017WR022421.

Richardson, D., S. Hemri, K. Bogner, T. Gneiting, T. Haiden, F. Pappenberger, and M. Scheuerer, 2014: Calibration of ECMWF forecasts. *ECMWF Newsletter*, **(142)**, URL https://doi.org/10.21957/45t3o8fj.

Satgé, F., D. Defrance, B. Sultan, M. P. Bonnet, F. Seyler, N. Rouché, F. Pierron, and J. E. Paturel, 2020: Evaluation of 23 gridded precipitation datasets across West Africa. *Journal of Hydrology*, **581**, 124 412, https://doi.org/10.1016/J.JHYDROL.2019.124412.

Scarrott, C., and A. MacDonald, 2012: A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. *REVSTAT-Statistical Journal*, **10 (1)**, 33–60, https://doi.org/10.57805/revstat.v10i1.110.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

Schneider, L. F., A. Krajina, and T. Krivobokova, 2021: Threshold selection in univariate extreme value analysis. *Extremes*, **24 (4)**, 881–913, https://doi.org/10.1007/s10687-021-00405-7.

Schumacher, R. S., 2017: Heavy rainfall and flash flooding. *Oxford Research Encyclopedia of Natural Hazard Science*, **1**, https://doi.org/10.1093/acrefore/9780199389407.013.132.

Stephens, M. A., 1974: EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, **69 (347)**, 730, https://doi.org/10.2307/2286009.

Tadeyo, E., D. Chen, B. Ayugi, and C. Yao, 2020: Characterization of spatio-temporal trends and periodicity of precipitation over Malawi during 1979-2015. *Atmosphere*, **11 (9)**, https://doi.org/10.3390/ATMOS11090891.

Tie, J., S. Kim, and A. Sharma, 2023: How does increasing temperature affect the sub-annual distribution of monthly rainfall? *Environmental Research: Climate*, **2 (1)**, https://doi.org/10.1088/2752-5295/ACB5B9.

Tye, M. R., K. Dagon, M. J. Molina, J. H. Richter, D. Visioni, B. Kravitz, and S. Tilmes, 2022: Indices of extremes: geographic patterns of change in extremes and associated vegetation impacts under climate intervention. *Earth System Dynamics*, **13 (3)**, 1233–1257, https://doi.org/10.5194/ESD-13-1233-2022.

Vannitsem, S., and Coauthors, 2021: Statistical postprocessing for weather forecasts review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*, **102 (3)**, E681–E699, https://doi.org/10.1175/BAMS-D-19-0308.1.

Wang, X., S. Shi, L. Zhu, Y. Nie, and G. Lai, 2023: Traditional and Novel Methods of Rainfall Observation and Measurement: A Review. *Journal of Hydrometeorology*, **24 (12)**, 2153–2176, https://doi.org/10.1175/JHM-D-22-0122.1.

Wang, Y., and H. A. Karimi, 2022: Impact of spatial distribution information of rainfall in runoff simulation using deep learning method. *Hydrology and Earth System Sciences*, **26 (9)**, 2387–2403, https://doi.org/10.5194/HESS-26-2387-2022.

Westra, S., and Coauthors, 2014: Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, **52 (3)**, 522–555, https://doi.org/10.1002/2014RG000464.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.

Wilks, D. S., 2020: *Statistical Methods in the Atmospheric Sciences*. 4th ed., Elsevier.

WMO, 2024: State of the global climate 2023. *WMO-No.1347*.

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

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

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-25-0019.1.