**RESEARCH ARTICLE**

Post-processed modelled rainfall climatology represents better observational point-rainfall climatologies

Fatima M. Pillosu1,2, Tim Hewson2, Christel Prudhomme2,3,4, Elisabeth Stephens1,5,6, Hannah L. Cloke1,5

1 Department of Geography and Environmental Science, University of Reading, Reading, UK

2 Forecast and Services Department, European Centre for Medium-range Weather Forecasts, Reading, UK

3 Department of Geography and Environment, University of Loughborough, Loughborough, UK

4 UK Centre for Ecology and Hydrology, Wallingford, United Kingdom

5 Department of Meteorology, University of Reading, Reading, UK

6 Red Cross Red Crescent Climate Centre, The Hague, The Netherlands

**Correspondence:** Fatima M. Pillosu(fatima.pillosu@ecmwf.int)

**Abstract.**

**Keywords.**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Introduction

Among extreme weather events, extreme rainfall has drawn particular attention in recent scientific literature (Gimeno et al., 2022; Schumacher, 2017) Its impacts are increasingly recognised to be catastrophic for society, infrastructure, and the environment (IPCC, 2023; WMO, 2024). Extreme rainfall can lead to catastrophic flooding, which puts life/property/food security at risk, disrupts transportation networks, and damages critical infrastructure. Among the most recent devastating flood events induced by extreme rainfall are Western Europe, Afghanistan, and China in 2021, including hurricane Ida affecting primarily the North-East coast of the USA (WMO, 2021); Pakistan, Brazil in 2022, including hurricane Ian in Florida (USA) and Fiona in Dominican Republic and Puerto Rico (WMO, 2022); in 2023, storm Daniel in the Mediterranean, Cyclone Freddy in Madagascar, Mozambique, and Malawi, Hurricane Otis in Mexico, and extensive flooding in the Horn of Africa (WMO, 2024). At the time of writing, one of the most recent devastating events in 2024 include severe flooding due to persistent rainfall in Central Europe and Hurricane Helene in Florida (USA). Extreme rainfall events also have significant environmental consequences, leading to soil erosion, water quality degradation, and habitat destruction (Lugo, 2018). Extreme rainfall can also have severe psychological effects on severely affected communities, leading to long-term anxiety and post-traumatic stress, hindering recovery efforts (Doocy et al., 2013). Extreme rainfall also reduces worldwide macroeconomic growth rates and slows global economic rise (Liang, 2022). Extreme rainfall is expected to become more intense and frequent due to climate change (Fowler et al., 2021; Min et al., 2011; Tabari, 2020), even in regions where the average rainfall has decreased (Asadieh & Krakauer, 2015). Hence, understanding extreme rainfall and analysing its past and future trends more accurately can inform better decision-making in disaster preparedness, agricultural management, and urban planning.

There are different sources to obtain rainfall timeseries. Rain gauges provide direct rainfall measurements of at specific locations, making them a primary source of ground truth. When properly maintained and calibrated, rain gauges provide highly accurate point-scale rainfall measurements (Lanza & Stagi, 2008). Where high-density networks exist, they also offer a good spatial representation of localized extremes (Haiden & Duffy, 2016). Moreover, many rain gauge stations have been in operation for decades, providing long-term historical records for trend analysis (Anand & Karunanidhi, 2020; Tadeyo et al., 2020). Rain gauge coverage is spatially and temporally uneven, leaving many regions uncovered (Kidd et al., 2017). In areas with complex topography or low-density networks, rain gauges may not adequately represent rainfall’s spatial variability (Di Curzio et al., 2022). Inadequate rain gauge maintenance can also cause instrument malfunctioning, leading to errors and missing data (Lanza & Stagi, 2008). Satellite- and radar-derived rainfall gridded datasets offer broader spatial and temporal coverage, especially in ungauged areas. The statistical characteristics of the gridded rainfall distribution can, however, be extremely different to the observational one, and observed extremes can be severely underestimated (Ensor & Robeson, 2008; Gupta et al., 2020; Satgé et al., 2020). Rainfall estimates from Numerical Weather Prediction (NWP) models, such as reanalysis and reforecasts, offer spatially and temporally consistent datasets with global and multi-decadal coverage. Reanalyses, such as ERA5 (Hersbach et al., 2020) and NCEP/NCAR Reanalysis (Kalnay et al., 1996), integrate historical weather observations with state-of-the-art NWP models to produce a high-resolution, and spatially and temporally continuous rainfall dataset that spans several decades. Reforecasts, such as those generated by the Global Ensemble Forecast System (Hammil et al., 2006) and ECMWF (Richardson et al., 2014), run current operational NWP models on historical data (typically, 20 years in the past) to bias-correct operational forecasts and develop compatible climatologies. Modelling precipitation and detecting extreme events is challenging primarily due to the use of parametrization schemes to represent convection and sub-grid scale processes (Madakumbura et al., 2021). Since they relate to current operational forecasting systems, reforecasts tend to have more up-to-date physical representation of rainfall-generation mechanisms and higher spatial resolution than reanalysis. Hence, they should represent better extreme (localised) rainfall events. Reanalysis generally captures rainfall’s spatial patterns and temporal trends (Lavers et al., 2022). However, they tend to underestimate the intensity of extreme rainfall events, especially in regions with complex topography or sparse rain gauge measurements (Alexandridis et al., 2023; Espinosa et al., 2024; Gomis-Cebolla et al., 2023)[[1]](#footnote-2). Similar results are obtained when analysing extreme rainfall patterns and trends using reforecasts (T. Hewson, 2024). Statistical downscaling of reanalysis and/or reforecast datasets can improve the representation of extreme rainfall patterns and trends and offer more accurate local-scale information (Giorgos t al., 2024). The good performance of any downscaling depends on a good spatial and temporal coverage of rainfall observations (Vannitsem et al., 2021). The ecPoint post-processing have shown to improve the spatial and temporal representation of NWP-modelled rainfall, especially for extremes and including areas with scarce or no rain gauge observations (T. Hewson et al., 2023; T. D. Hewson & Pillosu, 2021). Hence, ecPoint has been used to post-process ERA5 (T. Hewson et al., 2023).

This study evaluates the strengths and weaknesses of four gridded rainfall datasets with different spatial resolutions against rain gauge observations: the Ensemble Data Assimilation (EDA) for ERA5 at 62 km, the short range ERA5 forecasts at 31 km, ECMWF reforecasts for 46r1 at 18km resolution, and ERA5-ecPoint at point-scale. Two research questions are examined in this study. First, how does the distribution of NWP-modelled rainfall realizations compare to that built with rain gauge observations? Second, how is the tail of the NWP-modelled rainfall distribution different from that of the observed rainfall distribution? To answer the first question, an empirical quantile analysis is carried out. To examine the tail differences, NWP-modelled and observed rainfall for extreme return periods are compared and a case-study-based analysis for extreme rainfall events is carried out.

The study is organized as follows. Section two describes the observations and the NWP models used in this study. Section three describes the methods used for the comparison of the observational and NWP-modelled rainfall realizations. Section 4 presents the results while Section 5 discusses them. Finally, Section 6 draw the final conclusions for this analysis.

References

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

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. (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>

Emrich, C. T., Alvarez, S., Knox, C. C., Sadiq, A. A., & Zhou, Y. (2019). Hurricane Irma and Cascading impacts. In C. B. Rubin & S. L. Cutter (Eds.), U.S. emergency management in the 21st century from disaster to catastrophe (1st ed., p. 290). Routledge. https://doi.org/10.4324/ 9780429424670

Gimeno, L., Sorí, R., Vázquez, M., Stojanović, M., Algarra, I., Eiras‐Barca, J., … & Nieto, R. (2022). Extreme precipitation events. Wiley Interdisciplinary Reviews Water, 9(6). <https://doi.org/10.1002/wat2.1611>

Gimeno, L., Dominguez, F., Nieto, R., Trigo, R., Drumond, A., Reason, C. J. C., Taschetto, A. S., Ramos, A. M., Kumar, R., & Marengo, J. (2016). Major mechanisms of atmospheric moisture transport and their role in extreme precipitation events. Annual Review of Environment and Resources, 41(1), 117–141. https://doi.org/10.1146/annurev-environ-110615-085558

Giorgos, Ntagkounakis, Panagiotis T. Nastos, and Yiannis Kapsomenakis. 2024. "Creating High-Resolution Precipitation and Extreme Precipitation Indices Datasets by Downscaling and Improving on the ERA5 Reanalysis Data over Greece" *Eng* 5, no. 3: 1885-1904. <https://doi.org/10.3390/eng5030101>

Hamill, T. M., Whitaker, J. S. & Mullen, S. L. Reforecasts: An important dataset for improving weather predictions. *Bulletin of the American Meteorological Society* **87**, 33–46 (2006).

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

Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020).

Hewson (2024). ECMWF Newsletter n. 178.

Kalnay et al.,The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 437-470, 1996

Kharin, V. V., Zwiers, F. W., Zhang, X., & Wehner, M. (2013). Changes in the temperature and precipitation extremes in the CMIP5 ensemble. Climatic Change, 119, 345–357. https://doi.org/10.1007/s10584-013-0705-8

Kidd, C. *et al.* So, how much of the Earth’s surface is covered by rain gauges? *Bulletin of the American Meteorological Society* **98**, 69–78 (2017).

Liang, X.-Z. (2022). Extreme rainfall slows the global economy. Nature, 601, 193–194.

Lugo, A.E. Characterization of ten extreme disturbance events in the context of social and ecological systems. *Biogeochemistry* **141**, 385–400 (2018). https://doi.org/10.1007/s10533-018-0453-y

Madakumbura, G. D., Thackeray, C. W., Norris, J., Goldenson, N., & Hall, A. (2021). Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. Nature Communications, 12, 3944. <https://doi.org/10.1038/s41467-021-24262-x>

McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., Kominoski, J. S., McPhearson, T., Méndez-Lazaro, P., Rosi, E. J., & Shafiei Shiva, J. (2018). Defining extreme events: A cross-disciplinary review. Earth's Future, 6, 441–455. [https://doi.org/10. 1002/2017EF000686](https://doi.org/10.%201002/2017EF000686)

Min, S. K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. Nature, 470, 378– 381. https://doi.org/10.1038/nature09763

Richardson et al., (2014). Calibration of ECMWF forecasts. ECMWF Newsletter.

Rosenzweig, B., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., … & Davidson, C. (2018). Pluvial flood risk and opportunities for resilience. Wiley Interdisciplinary Reviews Water, 5(6). <https://doi.org/10.1002/wat2.1302>

Schumacher, R. S. Heavy Rainfall and Flash Flooding. *Oxford Research Encyclopedia of Natural Hazard Science* **1**, (2017).

Suppasri, A., Maly, E., Kitamura, M., Syamsidik, Pescaroli, G., Alexander, D., & Imamura, F. (2021). Cascading disasters triggered by tsunami hazards: A perspective for critical infrastructure resilience and disaster risk reduction. International Journal of Disaster Risk Reduction, 66, 102597. <https://doi.org/10.1016/j.ijdrr.2021.102597>

Tabari, H., Madani, K., & Willem, P. (2020). The contribution of anthropogenic influence to more anomalous extreme precipitation in Europe. Environmental Research Letters, 15(10), 104077. <https://doi.org/10.1088/1748-9326/abb268>

Ward, P., Jongman, B., Weiland, F., Bouwman, A., Beek, L., Bierkens, M., … & Winsemius, H. (2013). Assessing flood risk at the global scale: model setup, results, and sensitivity. Environmental Research Letters, 8(4), 044019. https://doi.org/10.1088/1748-9326/8/4/044019

Westra, S., H. J. Fowler, J. P. Evans, L. V. Alexander, P. Berg, F. Johnson, E. J. Kendon, G. Lenderink, and N. M. Roberts (2014), Future changes to the intensity and frequency of short-duration extreme rainfall, *Rev. Geophys.*, 52, 522–555, doi:[10.1002/2014RG000464](https://doi.org/10.1002/2014RG000464).

Zittis, G., Bruggeman, A., & Lelieveld, J. (2021). Revisiting future extreme precipitation trends in the Mediterranean. Weather and Climate Extremes., 34, 100380. <https://doi.org/10.1016/j.wace.2021.100380>

# Data

## Point-scale observational climatology

The point-scale observational rainfall climatologies were computed using 24-hourly rainfall from surface synoptic observations (SYNOP) from the Global Telecommunication System (GTS) network and stored internally at ECMWF. SYNOP data consists of standardized meteorological reports, historical and near-real time, that ensure consistency in data quality and format across diverse regions. To enhance the spatial coverage of rain gauges used to compute the rainfall climatologies, high-density national rain gauge networks available internally at ECMWF, primarily from European countries (Haiden & Duffy, 2016), were also integrated into the analysis. The observations were manually quality-controlled to remove erroneous values. Focus was given to eliminating erroneous high totals, which, if not corrected, could have significantly distorted the upper tails of the climatological distribution, leading to inaccurate results. Given the increased number of SYNOP observations at ECMWF from the 2000s, the analysis focused on the 20-year period between 1 January 2000 to 31 December 2019. For each location, the rainfall climatology was computed when at least 50% of the days within this period had valid observations, ensuring a reliable and comprehensive representation of the rainfall distribution.

## NWP-modelled climatologies

### ERA5 Reanalysis (ERA5) and ERA5 Ensemble Data Assimilation (ERA5-EDA):

ERA5 is the fifth generation of atmospheric reanalysis produced at ECMWF (Hersbach et al., 2020). Compared to its predecessor ERA-Interim, ERA5 offers higher spatial (~31 km) and temporal resolution (hourly), providing more detailed insights into global atmospheric processes. The ERA5 dataset 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 (three spatial and one temporal) but also enhances the continuity and stability of the climatological records. The data spans from 1940 to the present, allowing for long-term climate variability and trend analysis. It also allows to validate the performance of weather prediction models against observed atmospheric conditions. Providing a single realization per day, the maximum rainfall totals that can be computed in the ERA5-derived climatology corresponds to the 20-year return period.

The ERA5 Ensemble Data Assimilation (EDA) system enhances the robustness of the ERA5 reanalysis by incorporating ensemble forecasting techniques, which generate multiple simulations with slightly varied initial conditions. This approach not only estimates the most likely state of the atmosphere but also quantifies the uncertainty associated with observational errors and limitations within the forecasting model itself. Each ensemble member of the ERA5 EDA provides a possible realization of the atmospheric state, which, when aggregated, offers a probabilistic forecast that is crucial for risk assessment and decision-making processes in climate studies. ERA5-EDA has 10 ensemble members, running at 62 km spatial resolution and at 3 hour temporal resolution. Figure 3b shows the rainfall values (in mm/24h) corresponding to the 10-year return period in ERA5. Since the 10 ensemble members of ERA5-EDA were used to compute independent realizations of 24-hourly rainfall totals per day, , the maximum rainfall total that can be computed corresponds to the 20-year return period (Table 1, row 3).

### ECMWF Reforecasts

Reforecasts are created by retrospectively applying the latest operational model settings to past weather events, creating a historical dataset with no discrepancies caused by historical changes in model configurations. Hence, reforecasts provide a consistent basis for evaluating accuracy and bias of weather prediction models, and developing adjustments that correct systematic model biases. The uniformity of reforecasts is crucial for this analysis, as it ensures that differences in climatological patterns are attributable to actual atmospheric variations rather than artifacts of evolving model technologies. In this study, the 46r1 model cycle was considered as it was produced for over a year and, hence, we can use only one model cycle, and it runs over a similar period of interest (1st July 1999 to 30th June 2019).

### ERA5-ecPoint

Observation-based climatologies help detect trends and patterns in the climate over a long period of time and can contextualise extreme, high-impact weather events. However, observations can be inaccurate and are unevenly distributed in space and time. The ERA5 reanalysis, which is produced by the Copernicus Climate Change Service (C3S) run by ECMWF and is based on state‑of‑the‑art data assimilation and numerical weather prediction (NWP), provides an alternative: an accurate, temporally consistent, gridded estimate of the past state of the Earth system worldwide. However, this does not satisfy all needs due to ERA5’s relatively coarse model resolution, which precludes representation of localised extremes, and because of some intrinsic biases. Within the Highlander project, co‑financed by the EU and coordinated by Italy’s Cineca computing centre, ECMWF’s ecPoint post-processing technique was applied to raw ERA5 ‘deterministic’ fields to address ERA5 limitations (T. Hewson et al., 2023). ecPoint is formulated for independent application to any single model realisation and aims to infer sub-grid variability and to correct biases (both according to ongoing weather and geographical scenarios). So, in this way we create much more reliable (probabilistic) point-scale climatologies. ERA5\_ecPoint has so far been created for 1950 to near-real time, for 24- and 12‑hourly rainfall, and 24 h minimum, maximum, and mean 2m temperature. Two product classes were developed: grid-scale bias-corrected (‘deterministic’) and point-scale (probabilistic; percentiles 1 to 99). Currently, ERA5-ecPoint is provided on its native (reduced Gaussian) grid, with a spatial resolution of approximately 31 km (TL639).

# Methods

## Computation of the rainfall climatologies

The 24-hourly rainfall realizations ending at 00 UTC were computed for each day in the analysis period (01/01/2000 to 31/12/2019) from the rain gauge observations and the reanalysis/reforecast datasets. According to the considered dataset, a different number of daily realizations can be computed (Table 1, column 6). The total number of realizations in the 20-year analysis period (Table 1, column 6) were aggregated and ranked to compute a percentile-based climatology. Since the 1 in 10 year event (= 99.97260th percentile) was the maximum rainfall event that could be computed in the observational climatologies (Table 1, column 7 and 8, row 1), the comparison between the latter and the NWP-modelled climatologies used events up to a 10-year return period even though larger events could have been computed (Table 1, column 7 and 8, rows 2 to 5). Figure 3 shows the rainfall values (in mm/24h) corresponding to the 10-year return period for the observational (Figure 3) and NWP-modelled climatologies (Figure 3b-e).

## Statistical test to assess representativeness of NWP-modelled climatologies of the observational ones: k-sample Anderson-Darling test

In the k-sample Anderson-Darling (AD) test, the null hypothesis states that all samples are drawn from the same underlying distribution. Rejecting the null hypothesis indicates that there is sufficient evidence to conclude that at least one of the samples comes from a different distribution, suggesting significant differences between the samples. The critical values in this test increase with higher confidence levels, as more stringent thresholds are required to reject the null hypothesis when higher confidence is sought. For instance, at a 99% confidence level, the critical value is larger, meaning stronger evidence (i.e., a larger test statistic) is necessary to reject the null hypothesis compared to a 90% confidence level. This adjustment reduces the likelihood of making a Type I error—incorrectly rejecting the null hypothesis—at higher confidence levels. As a result, the test becomes more conservative, demanding more extreme evidence to detect distributional differences between the samples as statistical significance is evaluated with increasing rigor.

The non-parametric k-sample Anderson-Darling test is used to assess whether NWP-modelled rainfall climatologies are representative of observational climatologies. This non-parametric test is a good choice because it assesses whether the empirical cumulative distribution functions (ECDFs) of the two considered datasets have the same underlying distribution without having to specify the distribution function of that population. The test gives emphasis not only to differences in the centre of the distributions (like other tests such as Kolmogorov-Smirnov), but it also gives emphasis to differences in the tails. This makes the Anderson-Darling test highly effective for detecting departures in the extreme values in distributions with significant skew or kurtosis, e.g. rainfall distributions (Scholz & Stephens, 1987).

The test statistic Ak is calculated by integrating the squared difference between the two ECDFs, weighted by the variance function of the combined sample. A higher value of the Anderson-Darling statistic indicates a greater discrepancy between the two samples, suggesting that they may come from different distributions. The Ak values are compared to critical values of the test statistic, Acrit, to make inferential judgments about the homogeneity of the populations under study. Acrit is determined, at a certain confidence level (CL), via Monte Carlo simulations due to the complexity of the test’s distribution under the null hypothesis:

* **Ak <= Acrit(99.99% CL)**, the NWP-modelled rainfall climatology **is representative** of the observational climatology.
* **Ak > Acrit(99.99% CL)**, the NWP-modelled rainfall climatology **is not representative** of the observational climatology.

Such a binary decision was considered not appropriate for our case because the critical values were the same for all the NWP models.

The Anderson-Darling test is designed to compare distributions using all available data points, regardless of the difference in sample sizes, allowing for a more precise comparison of differences in the tails of the distribution (Baumgartner & Kolassa, 2021). This is particularly useful in our case where the NWP-modelled distributions can have 10 times more data points than the distribution from observations (see Table 1, column 6). However, it's important to note that very large differences in sample sizes (such as in the case of the distribution from ERA5-ecPoint that is 99 times bigger than the one from observations) may affect the test's sensitivity and introduce biases (Makarov & Simonova, 2017). Specifically, the larger sample may dominate the test statistic, potentially masking meaningful differences in the smaller sample or exaggerating minor deviations in the larger one. This imbalance can lead to either Type II errors (failing to detect true differences) or Type I errors (detecting differences that are not meaningful). Resampling techniques, such as bootstrapping, can help reduce this bias, but caution is needed if the focus is on differences in the tails of the distributions (as it is in this study), as resampling can sometimes obscure tail behaviour due to random sampling variability. Permutation methods, on the other hand, are a robust alternative because they preserve the structure of the original data, making them particularly effective when analysing differences in the tails. The permutation method tests the null hypothesis by randomly shuffling the group labels between samples, simulating a scenario where there is no true difference between the groups. By recalculating the test after each shuffle and repeating this process many times, the method generates a distribution of test statistics under the null hypothesis. If the original test statistic, calculated from the unshuffled data, falls in the extreme tail of this distribution, it suggests that the observed difference between the groups is unlikely to have occurred by chance, providing strong evidence to reject the null hypothesis. This approach ensures that any detected differences are meaningful and not simply due to random variation. Differences in the computed p-values reduced drastically after 100 permutations (not shown), so such value was considered as the best compromise between accuracy in the computation of the p-values and reducing code runtime.

As the sample size of your data increases, your chances of discovering non-normality increase. Small sample sizes may give you a false reading of normality. If you are using a probability plot, don’t be deceived by the impact of the sample size. Let your decision be guided by the p-value. The p-value of your AD test will indicate, with your desired level of risk, whether you can reject your null hypothesis.

# Results

Out of all NWP-modelled rainfall climatologies, ERA5-ecPoint provides the best representation of observational point-rainfall climatologies. This overall better representation can be seen by a much wider extension of blue dots (i.e., locations where NWP-modelled climatologies are representative of observational point-rainfall climatologies) in Figure 2e compared to Figure2b-d, where there is a much more significant number of pink dots (i.e., locations where NWP-modelled climatologies are not representative of observational point-rainfall climatologies). ERA5-ecPoint represents observational climatologies in more than 90% of locations in all six considered domains (whose definitions are provided in Figure 2a), except for South America, where it reaches only 74.3%. The best representation is reached in the Oceania and Europe & Mediterranean region, where the representativeness reaches 98.3% and 98.8% of locations. The lower resolution NWP-modelled climatologies, i.e. ERA5\_EDA (Figure 2a), ERA5 (Figure 2b), and reforecast\_46r1 (Figure 2c) are representative of observational climatologies in a much smaller number of locations, with performances that, on average, increase with the resolution (ERA5\_EDA < ERA5 < Reforecast\_46r1). The performance of ERA5-EDA and ERA5 climatologies is mostly comparable, with the worst representation in South America (less than 1% of locations with representative climatologies) and relatively poor representation over North America, Europe & Mediterranean, and Asia (less than 6%). Both NWP-modelled climatologies in Africa & Arabian Peninsula and Oceania reach at least 10% of locations that represent observational climatologies. However, it is worth noting that those points (blue dots in Figure 2b-c) are mainly located over the Sahara Desert in Africa and the arid regions of Australia. Overall, the climatologies from reforecasts provide a larger number of locations where they are presentative of observational climatologies, with improvements compared to ERA5-EDA and ERA5 that exceed 15% in Oceania, Asia, Africa & Arabian Peninsula, and 10% in Europe & Mediterranean and North America. Climatologies from reforecasts do not improve the representation of observational climatologies in South America, where they remain below 1%.

It is worth comparing the observational and the NWP-modelled climatologies to gain insights into where and why the climatological distributions differ. The climatologies are displayed with the x-axis in a logarithmic scale to compress the small rainfall totals and expand the higher ones. Compared to a graph with x-axis in linear scale, the logarithmic scale makes it easier to compare the climatologies on a single graph. In flat (Figure 2a) and orographic areas (Figure 2b), all NWP-model climatologies, with the exception of ERA5-ecPoint, are not able to represent the growth rate that observational climatologies have between 0.1 to 10, the frequency of zero rainfall totals, and the long wet tails in the observational climatology. In the flat areas, there is a progressive better representation of the observational climatology with increasing spatial resolutions. In orographic areas, ERA5-EDA and ERA5 perform very similarly for small rainfall totals, with ERA5 performing better in the representation of the wet tail. Reforecasts perform better than ERA5-EDA and ERA5. In very complex orographic areas such as the Andean region (Figure 2c), also ERA5-ecPoint fails to represent the growth rate that observational climatologies have between 0.1 to 10 and the frequency of zero rainfall totals, while still providing a better representation of the wet tail compared to the other NWP-modelled climatologies. As highlighted above, the AD test highlights that ERA5-ecPoint is indeed not representative of the observational climatology in the Andean region (Figure 1e). In desertic areas (Figure 2d), all NWP-modelled climatologies represent well the observational climatology, a part from the tail of the observational climatology.

# Discussions

The results of our study suggest that by post-processing ERA5 to better align with observed climatology, especially through enhancements in spatial resolution and region-specific model adjustments, we can significantly improve the representativeness of the climatological outputs. This alignment is crucial for the accurate prediction of localized weather patterns and can greatly enhance the capability of these models to issue timely and accurate warnings for extreme localized events. The analysis of the representativeness of various NWP-modelled climatologies against observed climatology, as demonstrated through the Anderson-Darling test at a 99.99% confidence level, reveals significant regional and model-specific variations. The ERA5-EDA (62 km) model shows high representativeness in North and South America, with over 95% of locations aligning with observed data, but less so in the Africa & Arabian Peninsula where only 87.2% of locations are representative. In contrast, the finer resolution ERA5 (31 km) model indicates improvements in most regions except Oceania, where the representativeness drops to 89.7%. The ECMWF Reforecasts (18 km) model further enhances fidelity, particularly in Asia and the Europe & Mediterranean regions, achieving over 99% representativeness in South America. However, challenges persist in the Africa & Arabian Peninsula, where only 66.6% of locations are representative in this model setup. Lastly, the ERA5-ePoint model, offering point-scale granularity, shows a notable improvement in North America, with a representativeness of 95.5%, but still struggles in Oceania and the Africa & Arabian Peninsula, with only 74.3% and 92.2% of locations found representative, respectively. This variability underscores the complexities in model performance across different spatial resolutions and geographic domains, highlighting the need for region-specific adjustments in NWP models to enhance their predictive accuracy globally.

The ability to establish a threshold for extreme rainfall predictions globally, independent of direct observational data, marks a significant advancement in meteorological science. This methodology leverages improved computational models that integrate historical data and predictive analytics, enabling the prediction of extreme weather events even in regions lacking physical measurement infrastructure. As a result, warnings for extreme rainfall can now be disseminated across ungauged areas, ensuring that all regions, regardless of their monitoring capabilities, are forewarned and can prepare adequately. This approach not only expands the coverage of weather warnings to a global scale but also enhances the resilience of vulnerable communities to the impacts of severe weather events.

The significant improvements in global weather models have facilitated a more accurate representation of climatic conditions worldwide, including extreme weather events. This enhanced modelling is pivotal for predicting phenomena such as severe storms, heatwaves, and intense precipitation with greater precision. By integrating advanced algorithms and extensive meteorological data, these models can now more reliably forecast the intensity and frequency of extreme conditions, providing crucial information for disaster preparedness and risk management. Such advancements not only aid in safeguarding lives and property by enabling more effective early warning systems but also support adaptive strategies in agriculture, infrastructure, and emergency planning across diverse geographic landscapes.

Utilizing ecPoint ERA5 for creating weather event thresholds requires careful calibration, particularly for Europe, to avoid setting thresholds that are too high, which could lead to underestimating or missing significant weather events. This model's high resolution can sometimes produce data that skew towards more extreme values, potentially setting the bar too high for triggering alerts. To prevent this, it is essential to rigorously validate the model against historical weather events across Europe, ensuring that the thresholds are adjusted to accurately reflect the frequency and intensity of occurrences. This approach will help maintain a balance, ensuring both the sensitivity and specificity of the model in predicting genuine weather threats without overlooking milder but still impactful events.

# Conclusions

The analysis conducted in this study highlights the effectiveness of refining NWP models to more closely align with observed climatology, particularly by improving spatial resolutions and making region-specific model adjustments. Such enhancements have significantly bolstered the representativeness of climatological outputs across various global regions, critical for the accurate forecasting of localized weather patterns and extreme events. Our findings, validated through the Anderson-Darling test at a 99.99% confidence level, demonstrate considerable variations in model performance by region and resolution. For instance, the ERA5-EDA (62 km) model exhibits high accuracy in North and South America, while the ERA5 (31 km) and ECMWF Reforecasts (18 km) models show substantial improvements in Europe, Asia, and South America. However, challenges remain in regions like Africa & the Arabian Peninsula and Oceania, underscoring the need for targeted adjustments to enhance model accuracy. Furthermore, our study underscores a significant advancement in meteorology with the ability to set thresholds for extreme weather predictions globally, independent of direct observational inputs. This advancement is crucial for issuing timely warnings, particularly in ungauged areas, thus expanding protective measures worldwide and enhancing community resilience against severe weather impacts. The integration of sophisticated algorithms and comprehensive data has also facilitated a more precise prediction of extreme weather phenomena, crucial for effective disaster preparedness and risk management. However, caution must be exercised, particularly with high-resolution models like ecPoint ERA5, to avoid setting overly high thresholds that may miss significant weather events in regions such as Europe. Rigorous validation against historical data is essential to ensure these models maintain a balance between sensitivity and specificity, providing reliable and actionable weather warnings. This holistic approach to improving NWP models not only helps in safeguarding lives and properties but also supports adaptive strategies across various sectors, thereby enhancing global readiness and response to climatic extremes.

# Tables

Table 1 – Summary of the characteristics of the rainfall observations and NWP models considered in this study (rows 1 to 5). Description of the computed climatologies (row 6 to 8).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.**  **Column** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **No.**  **Row** | **Type of climatology** | **Dataset for computation of climatology** | **Dataset description** | **Spatial coverage & resolution at equator** | **Temporal coverage** | **No. of independent daily & total no. of realizations in the 20-year period (between 2000 and 2019)** | **Max return period that can be computed in the 20-year period**  **(Max return period computed)** | **Max percentile (%) computed**  100 – ( \* 100 ) |
| 1 | Observational,  for points | SYNOP observations + higher resolution national databases | Rain gauges | Global (patchy),  Point-scale | From 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 / 2  (required a minimum of 50% of valid obs) =  1 in 10 year event  (1 in 10 year event) | 99.97260 |
| 2 | NWP-modelled, gridded | ERA5 - Ensemble Data Assimilation  (ERA5-EDA) | Analysis, probabilistic (10 ensemble members) | Global,  62 km | From 01/01/2000 to 31/12/2019 | 10 daily real.  (all ensemble members were used as daily independent realizations)  -  10 real. x 365 days x 20 years = 73000 total real. | 73000 real. / 365 days =  1 in 200 year event  (1 in 100 year event)\*  \*Some dates were not available, so only 66940 real. were available) | 99.99726 |
| 3 | NWP-modelled, gridded | ERA5 | Analysis, deterministic | Global,  31 km | From 01/01/2000 to 31/12/2019 | 1 daily real.  -  1 real. x 365 days x 20 years = 7300 total real. | 7300 real. / 365 day  1 in 20 year event  (1 in 20 year event) | 99.98630 |
| 4 | NWP-modelled, gridded | ECMWF Reforecasts for 46r1 | Forecasts, 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; forecast runs only on Mondays and Thursdays | 10 daily real.  (all lead times for 2 control runs a week were used as daily independent realizations)  -  10 real. x 2 runs x 52 weeks x 20 years = 20800 total real. | 20800 real. / 365 days =  1 in 56 year event  (1 in 50 year event) | 99.99452 |
| 5 | NWP-modelled, gridded | ERA5-ecPoint | Analysis, probabilistic (99 ensemble members) | Global,  point-scale but provided on ERA5’s grid (31 km) | From 01/01/2000 to 31/12/2019 | 99 daily real.  (all ensemble members were used as daily independent realizations)  -  99 real. x 365 days x 20 years = 722700 total real. | 722700 real. / 365 days =  1 in 1980 year event  (1 in 1000 year event) | 99.99972 |

# Figures

A graph of a graph of a graph

Description automatically generated with medium confidence

Figure 1 – Schematic representation of the calculation of the average areal difference (ECDFdiff) between two ECDFs.

A screenshot of a computer screen

Description automatically generated

Figure 2 – Average areal differences between Empirical Cumulative Distribution Functions (ECDFs) for observed and NWP-modelled 24-hourly rainfall. The differences, expressed as a percentage (in %) of the correspondent station’s mean observed rainfall, are summarized in piecharts for the domains defined in panel (a). Panels (b) to (e) show the average ECDFs’ areal differences for ERA5-EDA, ERA5, reforecasts, and ERA5-ecPoint, respectively. The tables on the right offer a numerical representation (in %) of the piecharts.

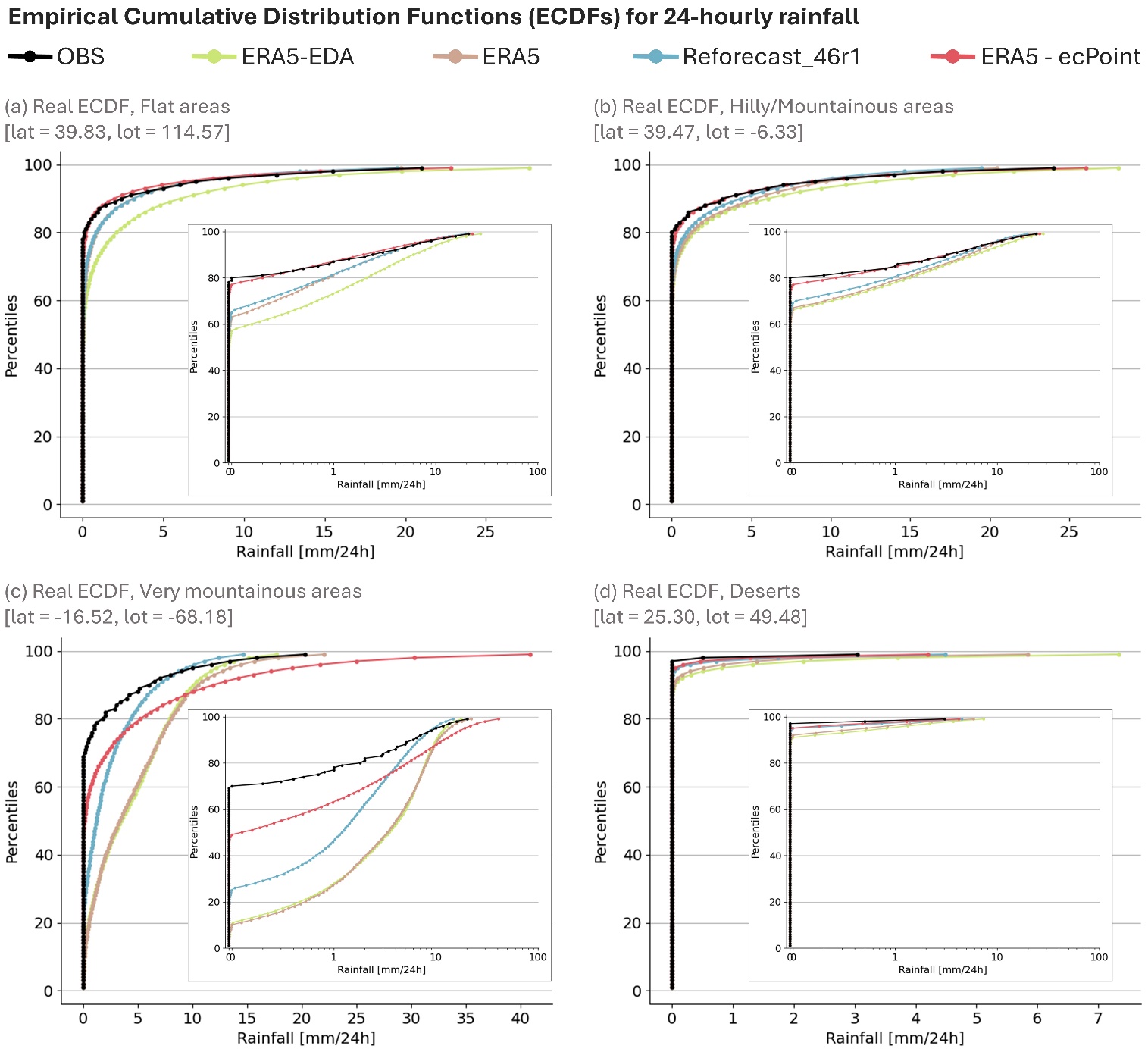


Figure 3 – Empirical Cumulative Distribution Functions (ECDFs) for 24-hourly rainfall from rain gauge observations (OBS, in black) and the NWP models ERA5-EDA (green), ERA5 (brown), reforecasts (Reforecasts\_46r1, light blue), and ERA5-ecPoint (coral). Panels (a) to (d) show representative ECDFs, respectively, 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 (Log ECDF).

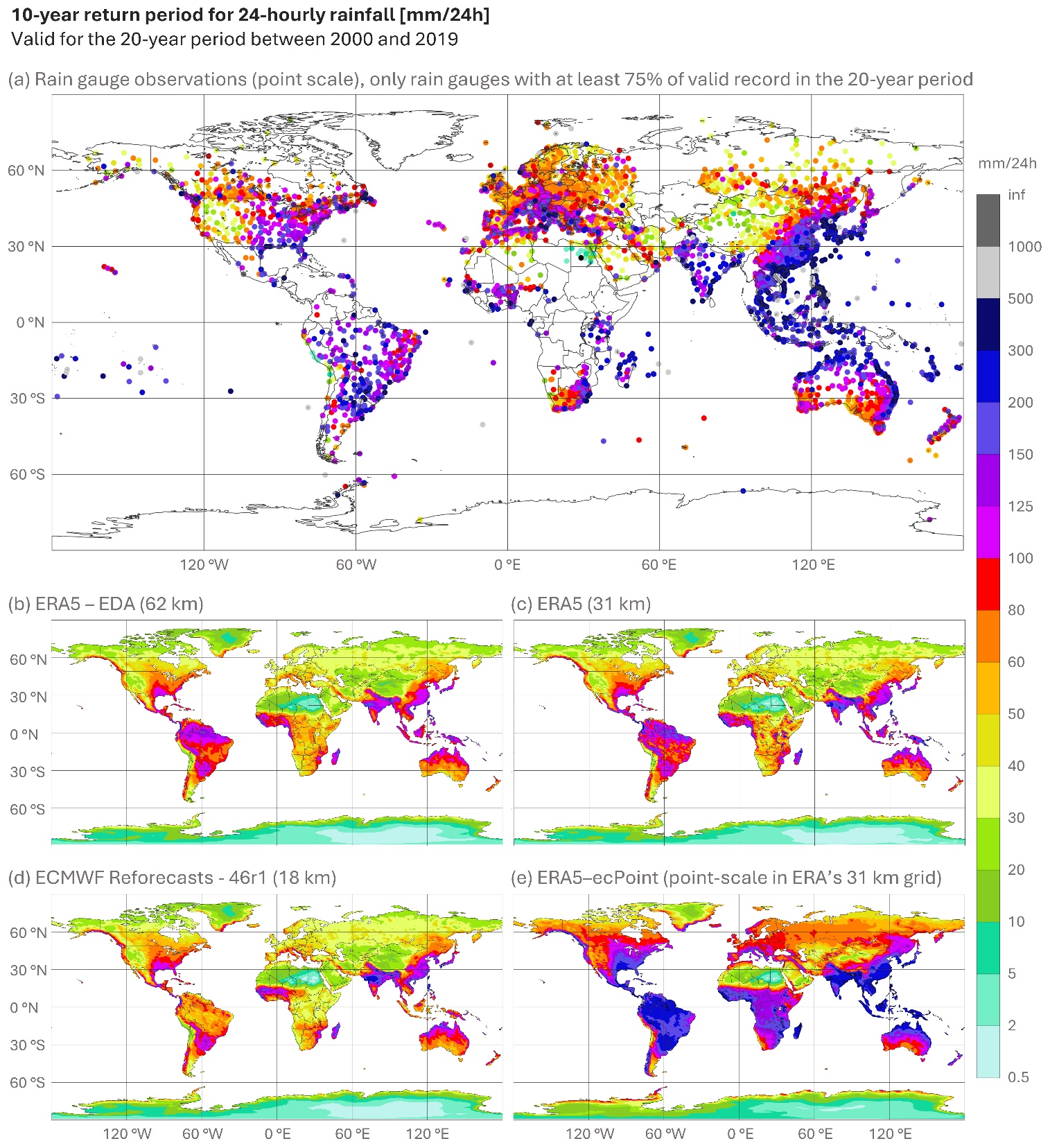


Figure 4 – Panel (a) displays the 10-year return period for 24-hourly rainfall from rain gauge observations, calculated over the 20-year period between 2000 and 2019, and using only rain gauges with at least 75% of valid records. Panels (b) to (e) show the 10-year return period for NWP-modelled 24-hourly rainfall: ERA5-EDA (62 km), ERA5 (31 km), reforecasts (Reforecasts\_46r1, 18 km) and ERA5-ecPoint (point-scale, provided on ERA5 grid).

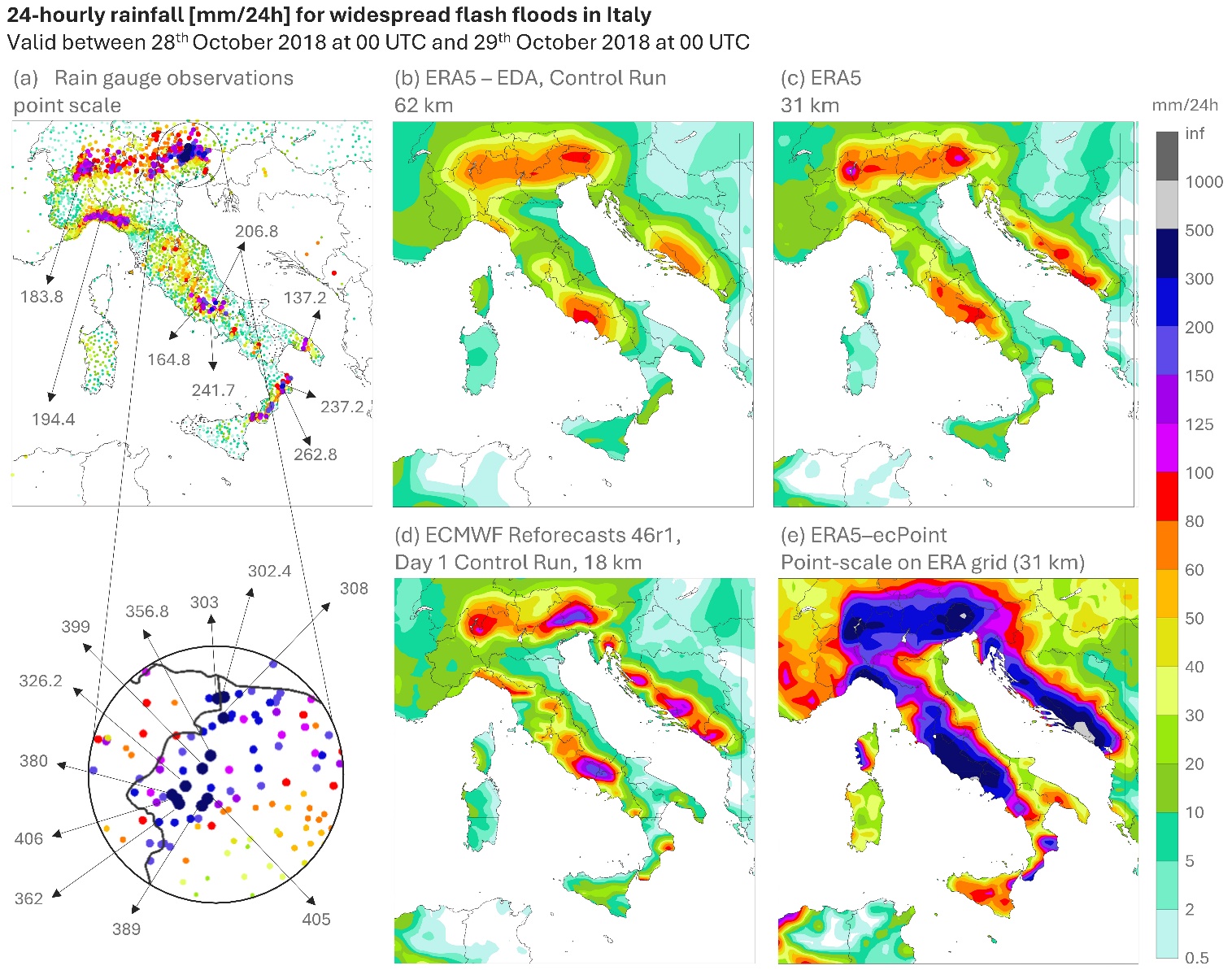


Figure 5 – 24-hourly rainfall [mm/24h] for widespread flash floods in Italy on 28th October 2018. Panel (a) shows observations from rain gauges. Numbers in grey indicate peak rainfall totals in mm/24h. Panel (b) to (e) indicate rainfall totals from, respectively, control run for ERA\_EDA (62 km), ERA5 (31 km), day 1 control run ECMWF Reforecasts 46r1 (18 km), and 99th percentile for ERA5-ecPoint (point-scale, provided on ERA5 grid).

# References

Alexandridis, V., Stefanidis, S., & Dafis, S. (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., & Karunanidhi, D. (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*(03), 419–427.

Asadieh, B., & Krakauer, N. Y. (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

Baumgartner, D., & Kolassa, J. (2021). Power considerations for Kolmogorov–Smirnov and Anderson–Darling two-sample tests. *Communications in Statistics. Simulation and Computation*, *52*(7), 3137–3145. https://doi.org/10.1080/03610918.2021.1928193

Di Curzio, D., Di Giovanni, A., Lidori, R., Montopoli, M., & Rusi, S. (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

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The Human Impact of Floods: A Historical Review of Events 1980-2009 and Systematic Literature Review. *PLoS Currents*, *APR 2013*. https://doi.org/10.1371/CURRENTS.DIS.F4DEB457904936B07C09DAA98EE8171A

Ensor, L. A., & Robeson, S. M. (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., Portela, M. M., & Gharbia, S. (2024). Assessing Changes in Exceptional Rainfall in Portugal Using ERA5-Land Reanalysis Data (1981/1982–2022/2023). *Water 2024, Vol. 16, Page 628*, *16*(5), 628. https://doi.org/10.3390/W16050628

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., Guerreiro, S., Haerter, J. O., Kendon, E. J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., & Zhang, X. (2021). Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth & Environment*, *2*(2), 107–122. https://doi.org/10.1038/s43017-020-00128-6

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

Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., & Francés, F. (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

Gupta, V., Jain, M. K., Singh, P. K., & Singh, V. (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., & Duffy, S. (2016). Use of high-density observations in precipitation verification. *ECMWF Newsletter*, *147*, 20–25. https://doi.org/10.21957/hsacrdem

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (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. D., & Pillosu, F. M. (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., Pillosu, F., Gascòn, E., & Vučković, M. (2023). Post-processing ERA5 output with ecPoint. *ECMWF Newsletter*, *176*.

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 [Core Writing Team, H. Lee and J. Romero (eds.)]. *IPCC, Geneva, Switzerland*, 35–115. https://doi.org/10.59327/IPCC/AR6-9789291691647

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (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

Lanza, L. G., & Stagi, L. (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

Lavers, D. A., Simmons, A., Vamborg, F., & Rodwell, M. J. (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

Lugo, A. E. (2018). Characterization of ten extreme disturbance events in the context of social and ecological systems. *Biogeochemistry*, *141*(3), 385–400. https://doi.org/10.1007/S10533-018-0453-Y/FIGURES/11

Makarov, A. A., & Simonova, G. I. (2017). Studying the Power of the Two-Sample Anderson–Darling Test in the Case of Contamination of One Sample. *Journal of Mathematical Sciences*, *221*(4), 580–587. https://doi.org/10.1007/S10958-017-3251-5

Min, S. K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378–381. https://doi.org/10.1038/nature09763

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J. N. (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

Satgé, F., Defrance, D., Sultan, B., Bonnet, M. P., Seyler, F., Rouché, N., Pierron, F., & Paturel, J. E. (2020). Evaluation of 23 gridded precipitation datasets across West Africa. *Journal of Hydrology*, *581*, 124412. https://doi.org/10.1016/J.JHYDROL.2019.124412

Scholz, F. W., & Stephens, M. A. (1987). K-Sample Anderson-Darling Tests. *Journal of the American Statistical Association*, *82*(399), 918. https://doi.org/10.2307/2288805

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

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-70816-2

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

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., Bouallègue, Z. Ben, Bhend, J., Dabernig, M., de Cruz, L., Hieta, L., Mestre, O., Moret, L., Plenković, I. O., Schmeits, M., … Ylhaisi, J. (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

WMO. (2021). *State of the Global Climate 2021*. *WMO-No. 1290*.

WMO. (2022). *State of the Global Climate 2022*. *WMO-No. 1316*.

WMO. (2024). *State of the Global Climate 2023*. *WMO-No.1347*.

1. The above-mentioned studies compare ERA5 and ERA5-Land. This is a faulty comparison because rainfall in ERA5-Land simply re-grids the rainfall in ERA5 onto ERA5-Land’s grid (Muñoz-Sabater et al., 2021). Such re-grid cannot be considered as a statistical or dynamical downscaling which would have added value to the higher-resolution dataset (i.e., ERA5-Land) compared to the lower-resolution one (i.e., ERA5). However, the results in these studies regarding ERA5 remain valid. Thus, we consider only that aspect of such studies. [↑](#footnote-ref-2)