

Technical Memo



909

Can global rainfall forecasts identify areas at flash flood risk? Proof of concept for Ecuador

Fatima M. Pillusu^{1,2}, Agathe Bucherie³, Andrew Kruczkievicz^{3,4,5}, Thomas Haiden², Calum Baugh², Carolynne Hultquist^{6,11}, Humberto Vergara⁷, Florian Pappenberger², Elisabeth Stephens^{8,4}, Christel Prudhomme^{2,9,10}, Hannah L. Cloke^{1,8}

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

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

³International Research Institute for Climate and Society (IRI), Columbia Climate School, New York, USA

⁴Red Cross Red Crescent Climate Centre, The Hague, The Netherlands

⁵Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands

⁶Center for International Earth Science Information Network (CIESIN), Columbia Climate School, New York, USA

⁷Civil and Environmental Engineering, IIHR - Hydroscience and Engineering, University of Iowa

⁸Department of Meteorology, University of Reading, Reading, UK

⁹Department of Geography and Environment, University of Loughborough, Loughborough, UK

¹⁰UK Centre for Ecology and Hydrology, Wallingford, United Kingdom

¹¹School of Earth and Environment, University of Canterbury, Christchurch, NZ

February 2024

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
<http://www.ecmwf.int/en/publications/>

Contact: library@ecmwf.int

© Copyright 2024

European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. The content of this document is available for use under a Creative Commons Attribution 4.0 International Public License.
See the terms at <https://creativecommons.org/licenses/by/4.0/>.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.

Abstract

Globally, flash floods are one of the costliest natural hazards for property damage and loss of life. The low accuracy of flash flood forecasts beyond a few hours limits their use in early warning systems. The aim of this study is twofold. First, this study aims to assess the performance of probabilistic rainfall forecasts from two different systems - the ECMWF ensemble (ENS) and the ecPoint post-processing system. The forecasts from these two systems are profoundly different: the first system provides forecasts at grid-scale. The second system provides forecasts at point-scale that mirror point observations such as rain gauge. ecPoint rainfall forecasts have been shown to provide better guidance than ENS in the prediction of extreme localized rainfall up to day 10. Hence, ecPoint should enhance ENS performance in the identification of areas at risk of flash floods up to medium-range leads. A one-year objective verification analysis was conducted for flash flood events in Ecuador, whose varied climate and the presence of a comprehensive flood report database made it an attractive site for this type of verification. To carry out such verification analysis, it is imperative to know what the magnitudes of flash-flood-triggering rainfall events are. Since this information was not known by the authors and there were no suitable rainfall observations to estimate such rainfall magnitudes, this study aims to propose a methodology that uses short-term ecPoint-Rainfall forecasts as a proxy for point rainfall observations to define the magnitude of flash-flood-triggering rainfall events to be used in the objective verification analysis. Due to the probabilistic nature of the ecPoint output, this approach allowed the simulation of a very high-density observational network in a de facto data-scarce region. The verification results suggest that ecPoint outperforms ENS in areas where rainfall originates from small-scale convective systems. Where rainfall originates from large-scale convective systems, ecPoint and ENS performances are comparable, except for the better identification by ecPoint of the actual magnitude of the rainfall events. This means that, under certain weather conditions, raw global NWP rainfall forecasts could be used to predict areas at risk of flash floods around the globe up to medium-range leads.

Plain language summary

Flash floods cause severe infrastructure damage and loss of life around the globe. Currently, it is possible to predict this hazard only up to one or two days ahead. Furthermore, many regions around the world have no access to flash flood predictions. Thus, protecting people and infrastructure against this hazard is one of the challenges of the century. This study has two objectives. The first objective relates to a verification analysis of two different global rainfall prediction systems, ECMWF ensemble (ENS) and ecPoint, that provide forecasts up to 10 days ahead. The verification analysis was carried over Ecuador because, besides having a very diverse climate, a detailed database has been developed to record historical flash flood events in the country. When the triggering rainfall event originates from small, localized storms, results show that ecPoint outperforms ENS in the identification of areas at risk of flash floods. For large-scale storms, ecPoint and the ECMWF ensemble performed similarly. However, ecPoint estimates the magnitude of extreme rainfall totals more accurately. The second objective of this study relates to the computation of thresholds to define flash-flood-triggering rainfall events for forecast verification. Here, short-range ecPoint rainfall forecasts were proposed as a proxy for rainfall observations. Results show that using this approach makes it possible to define rainfall thresholds closer to those calculated using rain gauge measurements, which would otherwise be the gold standard for this type of analysis. Furthermore, this methodology allows us to define the rainfall thresholds also in ungauged regions. These positive results give hope that, with further research, ecPoint could be used to predict areas at risk of flash floods up to 10 days ahead and over the whole globe.

1 Introduction

Flash floods cause significant societal, economic, and environmental impacts (Dordevic et al., 2020; Jonkman and Vrijling, 2008). They can occur due to prolonged moderate-to-extreme rainfall or extreme rainfall over a short-term period, leading to a sudden rise of river water levels or the inundation of low-lying areas (Speight et al., 2021; Zanchetta and Coulibaly, 2020). In Latin America, flash flood impacts are exacerbated by rapid, unregulated urbanization of floodplains, human-induced catchment degradation, high poverty levels, lack of preparedness plans, and inadequate infrastructure (Pinos and Quesada-Román, 2022). For example, flash floods in Ecuador are the deadliest form of flooding, and long-term impacts include infrastructure damage, agricultural losses, business and education interruptions, health service disruptions, and waterborne disease outbreaks (Galarza-Villamar et al., 2018). Approximately 60% of all floods in Ecuador are flash floods (Kruczakiewicz et al., 2021b), and their frequency is expected to increase due to climate change (Hirabayashi et al., 2021). This study aims to enhance preparedness and mitigation strategies against this escalating threat, particularly in regions with limited resources and data, by assessing the effectiveness of global rainfall forecasts in identifying areas at risk of flash floods. It is worth noting that, in this study, the term “identification of areas at risk of flash floods” refers to the prediction of polygons – or areas – that might experience flash flooding due to the expected rainfall.

Forecast-triggered mitigation strategies, such as early warning systems (de Perez et al., 2022; Šakić Troglić et al., 2022) and forecast-based financing protocols (Bischiniotis et al., 2019; De Perez et al., 2016), have been shown to improve resilience, decrease mortality, and lower recovery costs against riverine floods. Yet, these strategies hinge on accurate, timely predictions. In lower-income countries, accurate forecasts with even longer lead times are required to set cost-effective mitigation strategies (Bazo et al., 2019; Kiptum et al., 2023). Over the years, flash flood forecasting systems have been developed at local/regional (Corral et al., 2019; Ibarreche et al., 2020; Ramos Filho et al., 2021; Shuvo et al., 2021; Speight et al., 2018), national (Javelle et al., 2016; Liu et al., 2018), and continental scale (Gourley et al., 2017; Raynaud et al., 2015), with different degrees of model complexity and forecast accuracy. Due to large uncertainties in their forecasting chain, flash floods are more difficult to predict than riverine floods (Speight et al., 2021; Zanchetta and Coulibaly, 2020). Poor historical data on flash flood occurrence and impact (Lowrie et al., 2022), inaccurate predictions of extreme localised rainfall (Zeman et al., 2021), and challenging representation of detailed hydrological processes dependent on topography, soil conditions, and terrain coverage that modulate flash flood occurrence and severity (Xing et al., 2019) limit the predictability of flash floods. Thus, forecast-triggered mitigation strategies have been limited for flash floods (De Perez et al., 2016).

Several studies have found that low-complexity, index-based systems using key predictors like rainfall and soil moisture can offer more accurate flash flood predictions than more complex, physically-based models, especially in larger domain systems (Alfieri and Thielen, 2015; Hurford et al., 2012; Luong et al., 2021; Ma et al., 2021; Raynaud et al., 2015; Zanchetta et al., 2022). However, the prevailing belief that high-density observational networks (Javelle et al., 2010) or km-scale rainfall forecasts (Davolio et al., 2017; Song et al., 2019) are needed to represent the features of flash-flood-triggering rainfall events has limited the development of a flash flood forecasting system that provides medium-range forecasts over a continuous global domain. Radar-derived rainfall predictions extend only a few hours ahead (Imhoff et al., 2022) and the skill of km-scale rainfall forecasts decreases significantly after day 2 lead time Barrett et al. (2019). Furthermore, their spatial coverage is patchy. Due to the limitations outlined above, the Flash Flood Guidance System with Global Coverage is not developed over a continuous global domain but is built instead on a patchwork of systems over multiple locations (Georgakakos et al., 2021). Consequently, many areas of the world still remain without access to flash flood guidance, highlighting the need for alternative approaches to address the challenge of providing medium-range

flash flood predictions over a continuous global domain.

Global Numerical Weather Prediction (NWP) models, such as the ECMWF's Integrated Forecasting System, are increasingly seen as a viable solution to this challenge. These models provide daily global rainfall predictions up to medium-range leads but have historically struggled to accurately predict extreme localized rainfall events due to their coarse resolution and parametrization schemes (Emerton et al., 2016; Wen et al., 2021). Owing to recent improvements in global NWP forecast accuracy (Haiden et al., 2023; Lavers et al., 2021), the interest in using them to provide flash flood guidance in data-scarce regions and extend predictions' lead times has recently increased (Bucherie et al., 2022b). In addition, the development of state-of-the-art post-processing techniques can improve the quality of the raw forecasts and make them more suitable for flash flood prediction (Vannitsem et al., 2021). For example, the ecPoint statistical post-processing technique transforms global grid-based forecasts into probabilistic point-scale predictions and improves the reliability and discrimination ability of rainfall forecasts up to day 10, especially for extremes (Hewson and Pillou, 2021).

Since there remains a significant gap in leveraging global rainfall NWP forecasts for flash flood forecasting, this study aims to bridge this gap by evaluating and comparing the performance of ENS and ecPoint rainfall forecasts in identifying areas at risk of flash floods. Ecuador, with its extensive flash flood database and high susceptibility to flash flooding, serves as an ideal test bed for this research (Kruczakiewicz et al., 2021b). Two research questions are posed in this study. Can short-range ecPoint rainfall forecasts be used as proxies for in-situ rainfall observations in data-scarce regions to establish rainfall thresholds for an objective flash flood verification analysis? How well do ENS and ecPoint rainfall forecasts identify areas at risk of flash floods? The innovation of this research is threefold. First, it provides a complete objective verification analysis of forecasting systems, including estimates of forecast reliability and discrimination ability, to provide a nuanced assessment of the systems' performance. Second, rainfall forecasts are assessed against rainfall observations, and their suitability for flash flood prediction is simply assumed (Gascón et al., 2023; Hewson and Pillou, 2021). The objective verification in this study is conducted using historical flash flood reports in recognition of the non-linear relationship between flash floods and triggering rainfall events. Third, it proposes using short-range ecPoint rainfall forecasts as proxies for in-situ observations in data-scarce regions to calculate the flash-flood-triggering rainfall thresholds to be used in objective verification.

This paper is organized as follows. Section 2 provides background information on Ecuador's geography and rainfall/flood climatology. Sections 3 and 4 describe the data and methods used in the verification analysis. Section 5 presents the results of the objective verification analysis, while section 6 presents results from a case-study-based subjective verification analysis. Section 7 discusses the verification results, and section 8 draws concluding remarks for the study.

2 Background: geography, rainfall climatology and flooding in Ecuador

Located in north-western South America, Ecuador includes continental Ecuador and the Galápagos Islands in the Pacific Ocean, 1000 km from the mainland (insert in Figure 1a). The Andes run north to south through Ecuador (Figure 1a) and split it into three main regions (Vuille et al., 2000): "La Costa", which comprises the Andes' western slopes and the coastal plains along the Pacific Ocean; "El Oriente", which covers a plateau containing 2% of the Amazon basin and the eastern slopes; and "La Sierra", which contains the inter-Andean region between the western and eastern slopes of the Andes. This study considers only the continental landmass, hereafter referred to as "Ecuador". The domain of interest, including the borders with Colombia, Peru, and the Pacific Ocean, contains a total of 1090 grid boxes. This

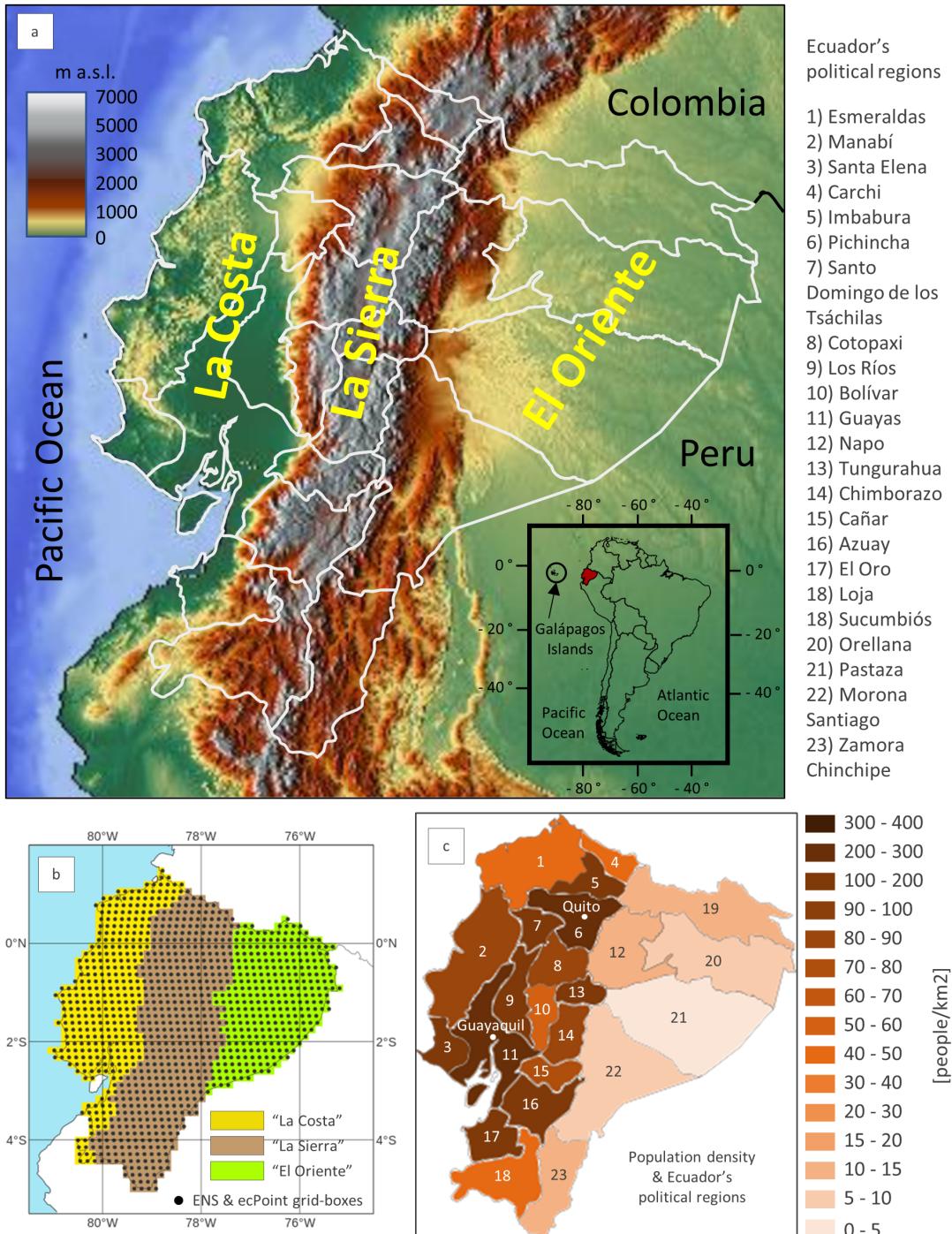


Figure 1: Panel (a) displays Ecuador's topography, outlines its political regions in white, and indicates the location of the country's three main geographical regions: the coast ("La Costa"), the Andean highlands ("La Sierra"), and the Amazon forest ("El Oriente"). The insert highlights Ecuador's location (in red) in South America. Panel (b) shows Ecuador's domain within the ENS and ecPoint grid, with grid boxes indicated with black dots. The regions are colour-coded in yellow for "La Costa", brown for "La Sierra", and green for "El Oriente". Panel (c) shows the population density from the 2020 census (people/km²) in each political region.

general domain is then split into the three main regions mentioned above using Ecuador's topography in the ENS grid (Figure 1b). Grid boxes below 600 m above sea level and to the west and the east of the longitude 78.2 °W belong to "La Costa" (i.e., 321 grid boxes) and "El Oriente" (299 grid boxes). Grid boxes above 600 m were assigned to "La Sierra" region (470 grid boxes).

The rainy season in "La Costa" spans from December to May (Ilbay-Yupa et al., 2021). Broad-scale atmospheric and oceanic phenomena modulate the intensity and spatial variability of rainfall. The extreme phases of El Niño Southern Oscillation, known as El Niño (i.e., above-average sea surface temperature in the Pacific Ocean) and La Niña (i.e., below-average), enhanced and decreased the average rainfall during the rainy season, respectively (Recalde-Coronel et al., 2014; Tobar and Wyseure, 2018). In addition, certain phases of the Madden-Julian Oscillation, 1 and 8 (i.e., when a convection centre is over the Western Hemisphere and Africa) and 4 and 5 (i.e., when a convection centre is over the Maritime continent), are associated with an enhancement and decrease in precipitation, respectively (Recalde-Coronel et al., 2020). "La Sierra" has two main rainy seasons (i.e., February-May and October-November). Precipitation spatial patterns in the inter-Andean valleys are more complex than in "La Costa" because rainfall is typically generated by smaller-scale convective systems (Vuille et al., 2000). Additionally, as air masses lose much of their humidity on both flanks of the Andes, precipitation amounts in "la Sierra" are relatively lower than those in the other two regions, varying between 800 and 1500 mm/year (Buytaert et al., 2006; Vuille et al., 2000). Several studies (Bendix et al., 2006; Buytaert et al., 2006; Junquas et al., 2022) have shown that in the Andean region, most of the rainfall occurs during the daytime, particularly in the afternoon hours, between 2 pm and 7 pm local time (LT). It rains throughout the year in "El Oriente", with the wettest (driest) months being April-July (September-October). Rainfall climatology in "El Oriente" is primarily influenced by strong convective activity across the Amazon Forest and water vapor variations from the sea surface temperature of the tropical Atlantic Ocean (Vuille et al., 2000).

Floods cause considerable material loss and deaths in "La Costa" and "La Sierra" because they are heavily populated and contain Ecuador's two most important industrial areas, Guayaquil in Guayas and Quito in Pichincha (Figure 1c). Prolonged rainfall events in "La Costa" can generate extensive, severe surface runoff far from rivers and cause rivers to flood vast plain areas (Galarza-Villamar et al., 2018). Intense shorter-lived rainfall events (i.e., less than one day) can also cause severe, sudden surface runoff (Galarza-Villamar et al., 2018). Rivers in "La Sierra" are susceptible to extreme localised rainfall events and, consequently, are prone to flash flooding (Laraque et al., 2009; Pinos and Timbe, 2020). The river flows in "El Oriente" show a much stronger response to seasonal rainfall than a single rainfall event because of the size and length of Amazonian rivers, floodplain storage, and shallow riverbeds (Trigg et al., 2009).

3 Data

This section introduces and provides statistical summaries for the datasets used in this verification analysis. This summary will be useful to interpret the verification results.

3.1 Flash flood reports

Disaggregation by flood type and specific documentation about historical flash flood events and their impacts are rare in many regions of the world, primarily due to a lack of commonly accepted flash flood definitions (Bucherie et al., 2022a; Kruczakiewicz et al., 2021a). Kruczakiewicz et al. (2021b) developed a method to assign an "Enhanced Flash Flood Confidence Index (EFFCI)" for flood events in historical

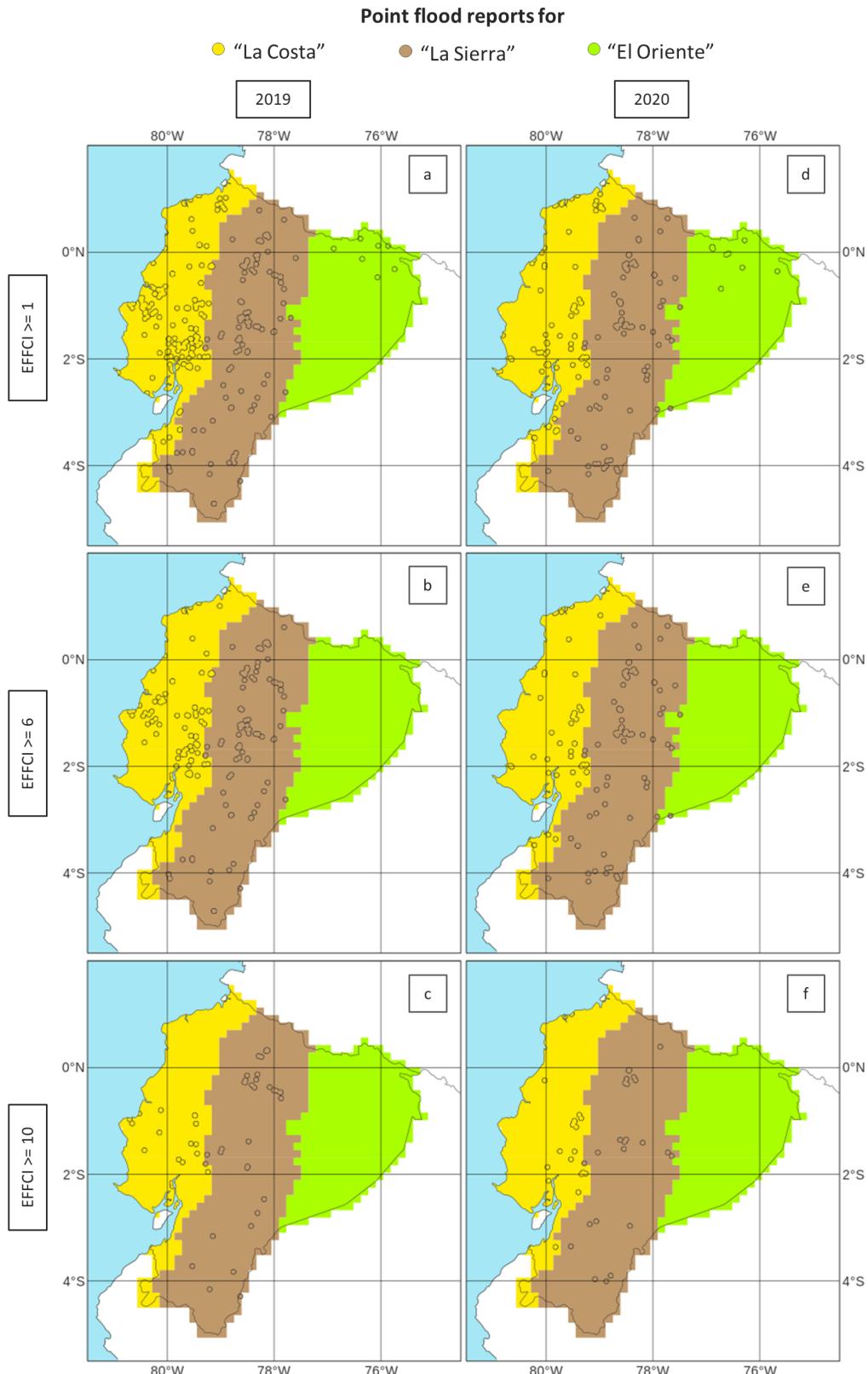


Figure 2: Panels (a), (b), and (c) show the spatial distribution of point flood reports in 2019 with EFFCIs, respectively, ≥ 1 , 6, and 10. Panels (d), (e), and (f) are the same but for flood reports in 2020.

Table 1: The first column indicates the years considered in this analysis. The second column shows the total number of flood reports in the database in each year. The third column shows the number of excluded reports from the study because they did not contain any reporting location (in lat/lon coordinates) and/or reporting time (with date and time). The fourth column shows the number of remaining flood reports. The remaining columns show the distribution of flood reports per region (C: Costa, S: Sierra, O: Oriente) and EFFCI threshold.

Year	N. of Raw FR	N. of deleted FR	N. of cleaned FR	N. of flood reports with								
				EFFCI>=1			EFFCI>=6			EFFCI>=10		
				C	S	O	C	S	O	C	S	O
2019	302	3	299	176	116	7	93	102	0	17	34	0
2020	190	0	190	96	93	88	48	79	0	22	26	0

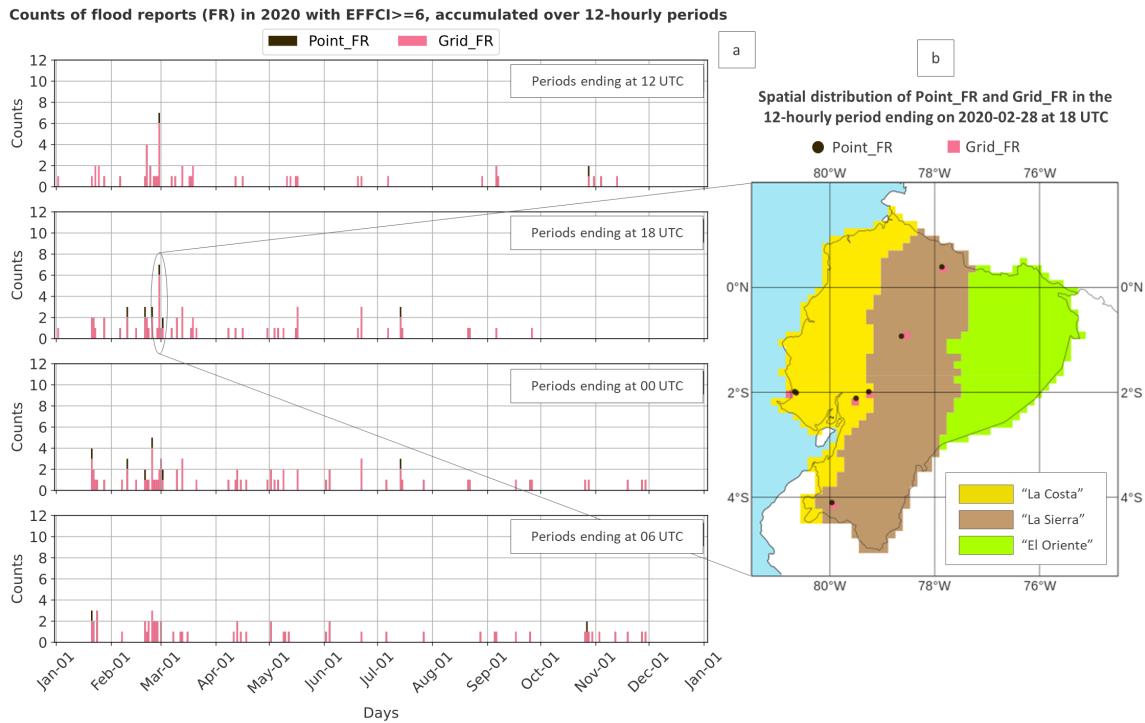


Figure 3: Panel (a) displays the timeseries of the counts of 2020's flood reports with $\text{EFFCI} \geq 6$, over four overlapping 12-hourly accumulation periods, ending at 12 (first row), 18 (second row), 00 (third row), and 06 UTC (fourth row). Figure 2e shows their collective spatial distribution. The overlapped black and pink bars represent point and grid flood reports, respectively. Where the black bars are visible, more than one point flood report was assigned to the grid box. Panel (b) shows the spatial distribution of the point and grid flood reports for the accumulation period that contains the largest number of flood reports, namely the one ending at 18 UTC on 2020-02-28.

flood datasets, based on text mining of disaster reports and a flash flood susceptibility index extracted from the geophysical properties of the location of the events. The EFFCI is an estimate of the likelihood of a flood event being a flash flood, ranging from 1 (not very likely) to 10 (extremely likely). The flash flood database in Ecuador was mainly compiled from two datasets, DesInventar ([UNDRR, 2021](#)) and the Ecuadorian Secretariat for Disaster Management (SNGRE), and it contains 4967 flood events from 2007 to 2020. In addition to the EFFCI index, most entries in the flash flood database contain information about the location (with latitude and longitude coordinates) and the day and time (in local time) of flood occurrence. As a result of applying this method to Ecuador, a historical dataset of flood occurrences and impacts is available, with specific information on the likelihood of events being flash floods ([Bucherie et al., 2021](#)). Although this dataset is the best attempt to address historical flash floods in Ecuador, it is essential to note that it is based on disaster reporting processes carried out on the ground and not systematically collected over time. Consequently, it can present gaps, inconsistent descriptions of flood processes over time, and uncertainty in geolocation.

This study considered flood reports from 2019 to define the climatology of rainfall events associated with flash floods. Events from 2020 were used to perform an objective verification analysis. Three EFFCI thresholds were considered to evaluate the impact of uncertainty around a flood report as a flash flood event: $\text{EFFCI} \geq 1$ (all flood reports), $\text{EFFCI} \geq 6$ (flood reports that are likely to be flash floods), and $\text{EFFCI} \geq 10$ (flood reports that are highly likely to be flash floods). Table 1 shows the total number of flood reports in 2019 and 2020, as well as their distribution between the three considered regions and EFFCIs. Figure 2 shows the spatial distribution of such flood reports, while Figure 3 shows the time distribution of flash flood reports with $\text{EFFCI} \geq 6$.

3.2 Rainfall observations

Rainfall observations from the SYNOP network transmitted by the Global Telecommunication System (GTS) were used in this study to represent Ecuador's rainfall climatology and provide a context for both the estimated rainfall totals associated with flash flood events and the objective verification results. In the ECMWF's internal database, only 12-hourly rainfall observations with accumulation periods ending at 00 and 12 UTC, between the 1st of January 2010 and the 31st of December 2020, are available for Ecuador. Figure 4a and Figure 4b show their spatial and temporal distribution, respectively. Not all days have observations. "El Oriente" has only one observation in a given day, while most of the days in "La Costa" and "La Sierra" have 4 to 8 observations in a given day, with peaks of 10 observations in "La Costa" and 14 in "La Sierra". Figure 4c shows the average rainfall over the entire study period for each accumulation period. The sinusoidal pattern in all three regions confirms what was reported in Section 2 regarding the marked rainfall diurnal cycle in Ecuador. From the available observations, "La Costa" is the region with the most marked rainfall diurnal cycle with peak at night-time (i.e., accumulation period ending at 12 UTC) of ~ 3.5 mm/12h and troughs at daytime (i.e., accumulation period ending at 00 UTC) of ~ 1 mm/12h, and an average rainfall overall of ~ 2.2 mm/12h. "La Sierra" shows instead a very small diurnal cycle with an amplitude of just ~ 0.1 mm/12h between night-time and daytime rain, and an average of ~ 1.6 mm/12h. It also appears that the times at which peaks and troughs occur are inverted to those in "La Costa". Although the spatial coverage of the rainfall observations in "El Oriente" is very poor, it is possible to observe that the average rainfall in the rain forest is substantially bigger than in the other two regions (i.e., ~ 4.8 mm/12h) and that the amplitude of the rainfall's diurnal cycle is ~ 1 mm/12h. Figure 4d shows the distribution of the rainfall totals on a given day for each accumulation period. More than 50% of the observations were less than 1 mm. "La Costa" is the region that has more frequent observations above 100 mm/12h in the accumulation period ending at 12 UTC, which corresponds to the local night-time. During daytime rain (i.e., accumulation period ending at 00 UTC)

Description of SYNOP observations of 12-hourly rainfall between 2010-01-01 and 2020-12-31

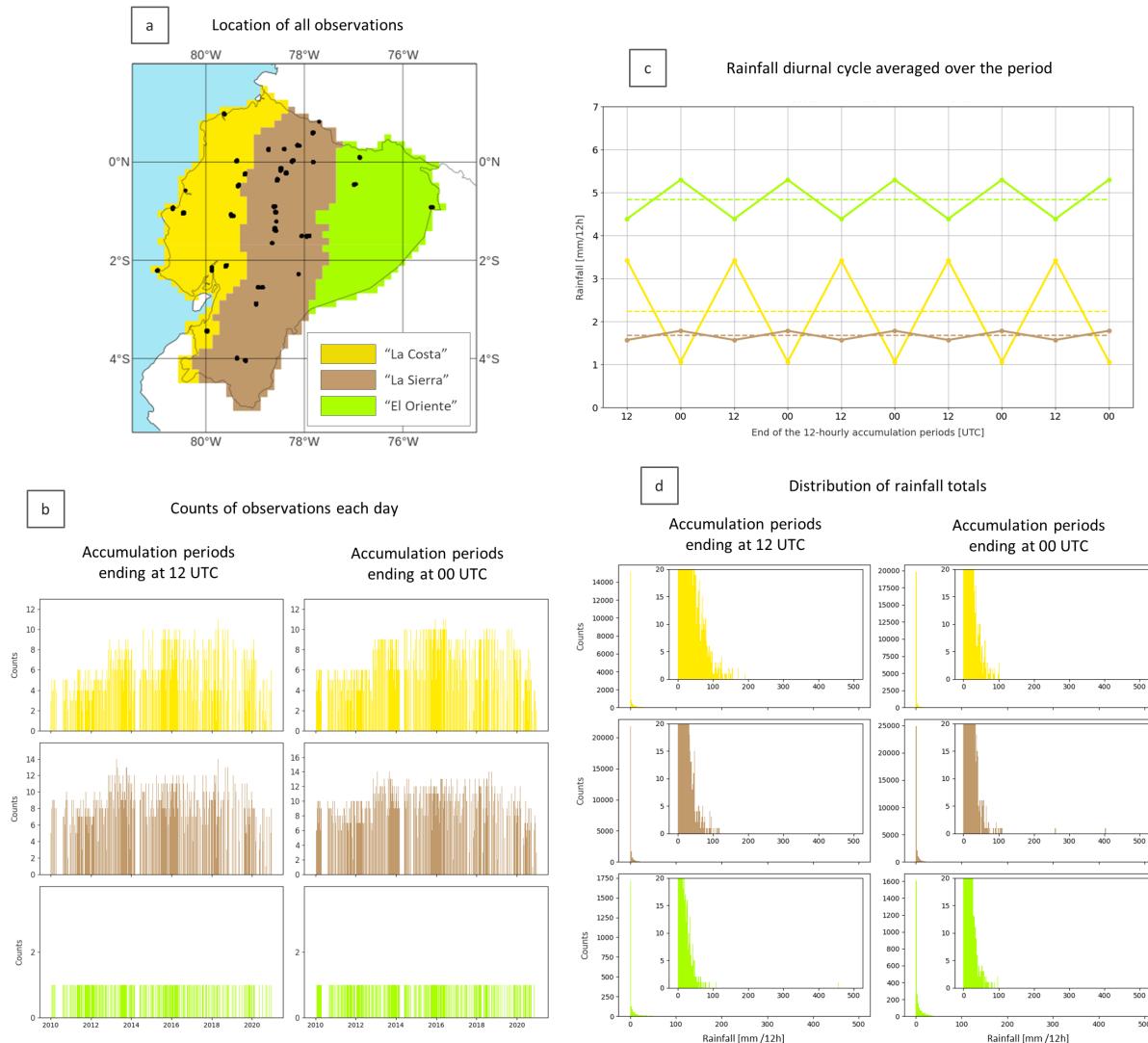


Figure 4: Description of 12-hourly SYNOP rainfall observations in Ecuador from 2010-01-01 to 2020-12-31. Only accumulation periods ending at 12 and 00 UTC were available. Each plot is colour-coded according to the region they refer to: yellow for “La Costa”, brown for “La Sierra”, and green for “El Oriente”. Panel (a) displays a map plot with the location of all SYNOP observations (some might have moved locations over time). Panel (b) shows the daily counts of observations available for the accumulation period ending at 12 UTC (first column) and at 00 UTC (second column). Panel (c) shows the average rainfall for all considered years, for the accumulation period ending at 12 and at 00 UTC (the two values are repeated five times in the plot), and the corresponding trend lines. In panel (d), the distributions of 12-hourly rainfall totals are shown for the accumulation period ending at 12 UTC (first column) and at 00 UTC (second column). Each distribution includes zoomed-in inserts to remove the focus over the very small rainfall totals.

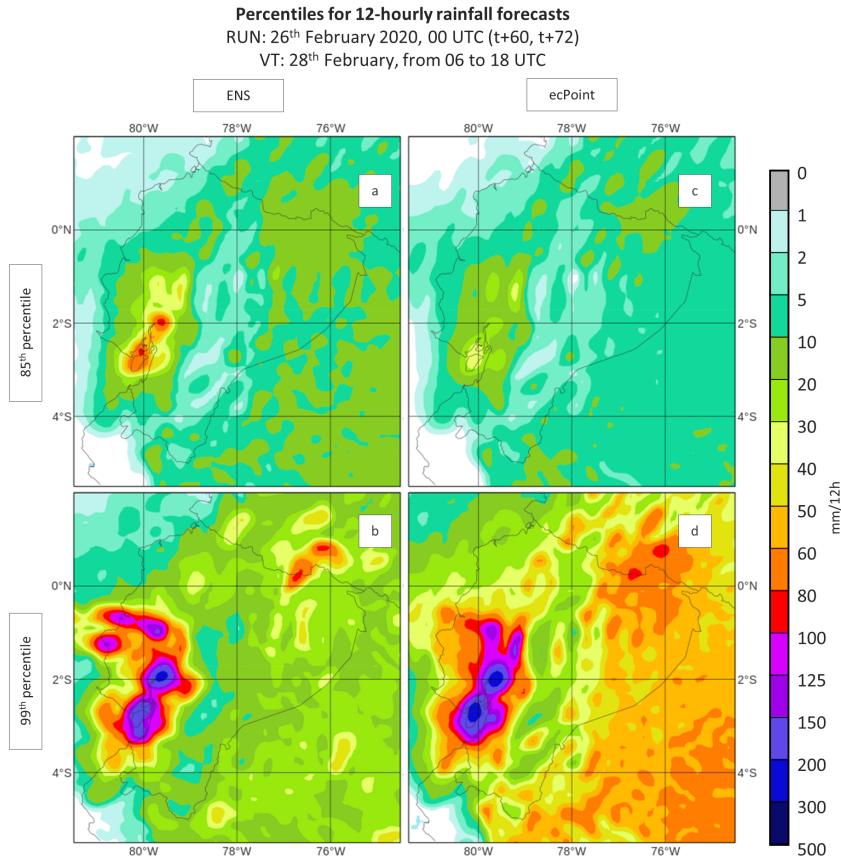


Figure 5: Example of a probabilistic ENS (first column) and ecPoint (second column) medium-range rainfall forecast (in mm/12h). The forecast is from the midnight run (00 UTC) on the 26th of February 2020, for the accumulation period ending at t+72, and valid for the 28th of February 2020 between 06 and 18 UTC (i.e., 00 and 12 LT). Panels (a) and (c) show examples of the 85th percentile for ENS and ecPoint, respectively, while panels (b) and (d) show examples of the 99th percentile.

in “La Costa”, and in both accumulation periods in “La Sierra” is rare to observe any rainfall ≥ 100 mm/12h. However, it is possible to observe two extreme rainfall observations in the accumulation period ending at 00 UTC in “La Sierra” of up to ~ 260 and ~ 400 mm/12h.

3.3 Rainfall forecasts: ECMWF ENS and ecPoint

The ECMWF ENS consists of one control run starting from the best possible representation of unperturbed initial conditions, and 50 perturbed members starting from perturbed initial conditions (using singular vectors and a data assimilation ensemble) and stochastic model uncertainties (Buizza, 2019). Up to day 15, ENS forecasts are saved in the native octahedral reduced-Gaussian with a resolution of 18 km at the equator (Owens and Hewson, 2018). Over the period used to compute the climatology of rainfall events associated with flash flood events (1st January to 31st December 2019) and the verification period (1st January to 31st December 2020), three different model versions were run operationally at ECMWF: 45r1¹ (for forecasts from 1st January to 10th June 2019), 46r1² (from 11th June 2019 to

¹www.ecmwf.int/en/forecasts/documentation-and-support/evolution-ifs/cycles/summary-cycle-45r1

²www.ecmwf.int/en/forecasts/about-our-forecasts/evolution-ifs/cycles/summary-cycle-46r1

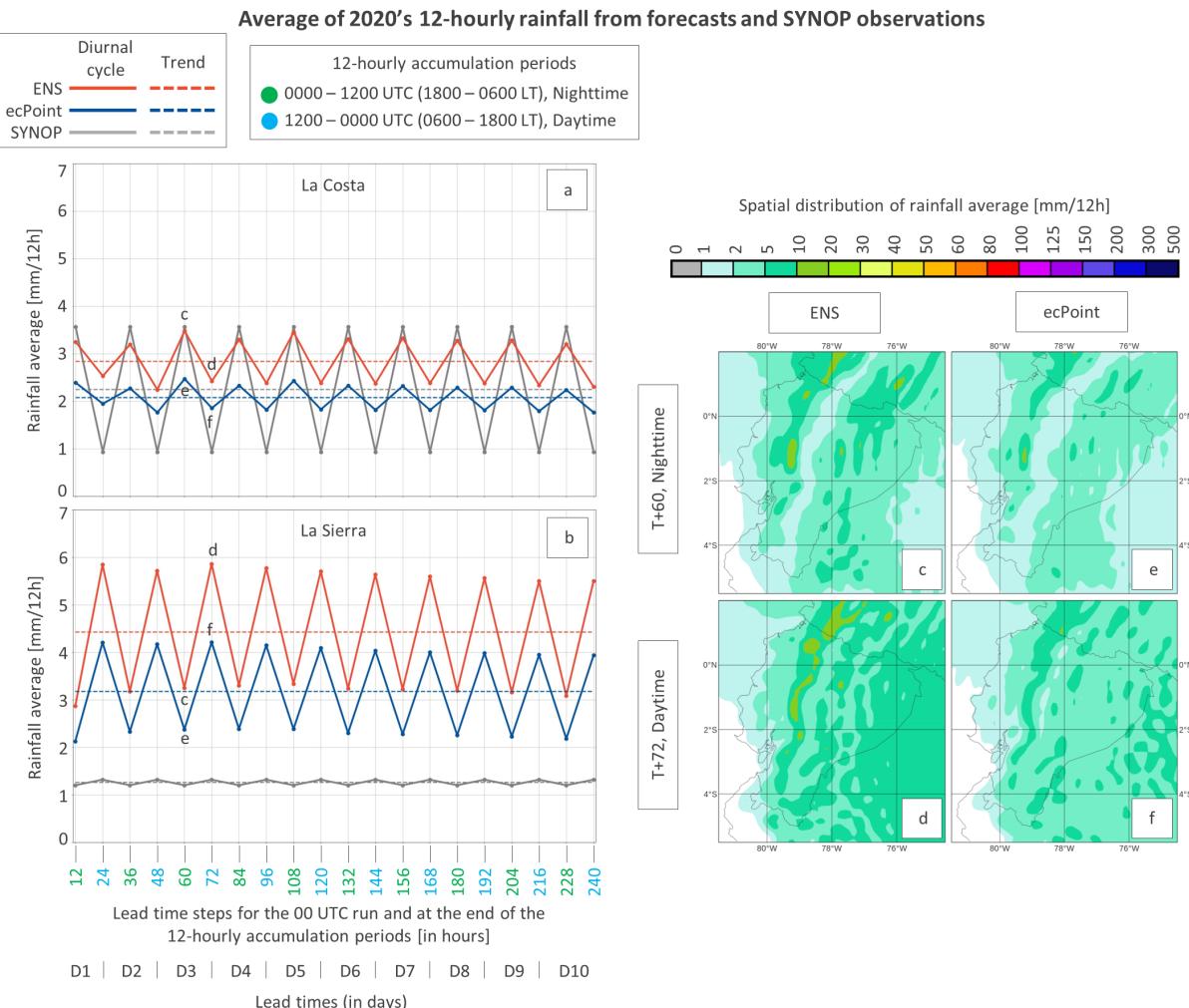


Figure 6: Panels (a) and (b) show, for “La Costa” and “La Sierra”, respectively, the 12-hourly rainfall average (continuous line) and their trend (dashed line) as a function of forecast lead time up to day 10. SYNOP, ENS, and ecPoint are shown in grey, red, and blue, respectively. The x-axis indicates the lead time steps (in hours) at the end of the 12-hourly accumulation period. Each step is colour-coded according to the valid accumulation period (in UTC and LT) based on the 00 UTC run: the steps representing the accumulation period between 0000-1200 UTC (or 1800-0600 LT, night-time) are shown in green, while the steps representing the accumulation period between 1200-0000 UTC (or 0600-1800 LT, daytime) are shown in cyan. Panels (c) and (d) illustrate examples of a typical spatial distribution of rainfall averages, respectively, at night-time and daytime for a day 3 forecast in ENS (accumulation periods ending at steps t+60 and t+72, respectively). Panels (e) and (f) show the same but for ecPoint.

12th July 2020), and 47r³ (from 13th July to 31st December 2020). This mismatch of model versions is unlikely to adversely affect the verification results because no significant changes were made in the physics of the rain generation mechanisms in the considered model versions.

ecPoint is a decision-tree-based statistical post-processing technique that transforms global grid-based forecasts into probabilistic point-scale forecasts (Hewson and Pillou, 2021). The post-processing technique aims to provide forecasts that mirror observations from rain gauges by addressing the two main factors affecting the performance of the global NWP model output against point verification: systematic biases (Lavers et al., 2021) and lack of information on forecast sub-grid variability (Göber et al., 2008). For each raw ENS member, ecPoint generates an ensemble of 100 point-rainfall values based on the error distributions between forecasts and observations that vary according to different weather scenarios at the grid-box level. For example, when on a grid box, the model mainly predicts large-scale rainfall with light winds, the raw model output tends to be representative of point rainfall totals within that grid box, and ecPoint generates an ensemble with a smaller spread compared to the case of mainly convective rainfall with light winds. In the latter case, many points within the grid-box are expected to show zero rainfall while very large rainfall amounts might be observed at a few points. From the current operational configuration of ENS forecasts (i.e., 51 ensemble members), the 5100 point-scale rainfall values were distilled in percentiles from the 1st to the 99th. ecPoint forecasts are provided in the same native grid of ENS forecasts up to day 10 lead times and in four overlapping 12-hourly accumulation periods with valid times starting at 0, 6, 12, and 18 UTC.

Figure 5 shows examples of the ENS and ecPoint rainfall forecasts from the 85th and 99th percentiles. Typically, percentiles from ecPoint lower than or equal to the 85th percentile (Figure 5c) have lower rainfall forecast values than the ENS (Figure 5a). This is because, generally, the number of zero rainfall totals is larger in ecPoint than in ENS. This is a bias correction applied to the rainfall forecasts by ecPoint, as ENS tends to overpredict small rainfall totals (Haiden et al., 2023). On the contrary, big percentiles (typically above the 90th percentile) tend to show larger rainfall totals on ecPoint than in ENS. This can be noticed in Figure 5d by the overall domination of the orange colour (i.e., rainfall totals between 50 and 80 mm/12h) compared to Figure 5b where the dominant colour is green (i.e., rainfall totals between 10 and 30 mm/12h). It is noteworthy that ecPoint does not always increase the amounts of the rainfall forecasts. In certain parts of “La Costa”, the rainfall totals in ecPoint are lower than those in ENS because the post-processing considered that the raw rainfall forecasts might be overpredicted under the predicted grid-box weather type. While Hewson and Pillou (2021) showed with a global objective verification analysis over a one-year period that, up to medium-range lead times (i.e., day 10 forecasts), ecPoint provides forecasts for point-scale rainfall with better reliability and discrimination ability than ENS, especially for extremes, it is interesting to compare how ENS and ecPoint perform in the prediction of rainfall in Ecuador. Figure 6a and 6b shows 2020’s 12-hourly rainfall average from SYNOP observations (in grey) and forecasts (ENS in red and ecPoint in blue), respectively, for “La Costa” and “La Sierra”. There is no degradation in performance with lead times up to day 10, and both forecasting systems reproduce the fact that Ecuador’s rainfall is affected by a diurnal cycle. However, neither ENS nor ecPoint can correctly represent the peaks and troughs of the diurnal cycle of rainfall. While in “La Costa” ENS represents the rainfall peaks during night-time exceptionally well, it significantly overestimates the rainfall over daytime (Figure 6a). With the aim of reducing the daytime bias, ecPoint unnecessarily reduces the night-time rainfall and reduces the overall amplitude of the rainfall’s diurnal cycle. This is because for 12-hourly rainfall, ecPoint makes no distinction between rainfall at different times of the day. Figure 6c shows how ENS predicts most of the rainfall during night-time on the western slopes of the Andes, while ecPoint tends to remove most of the rainfall, leaving the rainfall in the Pacific coast mostly unchanged (Figure 6e). During the daytime, the ENS still produces high amounts of rainfall, primarily on the

³www.ecmwf.int/en/forecasts/about-our-forecasts/evolution-ifs/cycles/summary-cycle-47r1

western slopes of the Andes (Figure 6d). ecPoint applies a general reduction in rainfall in “La Costa” (Figure 6f). It can be observed that ecPoint’s overall rainfall average (blue dashed line in Figure 6a) is closer than that of ENS (red dashed line) to the observed average (grey dashed line). In “La Sierra” (Figure 6b) although again ecPoint’s overall rainfall average is closer than ENS to the observed average, both forecasts significantly overestimate the absolute values of daytime and night-time rainfall and the amplitude of the observed rainfall diurnal cycle. In Figure 6c to Figure 6f, it can be observed that ecPoint applies a reduction in the rainfall forecasts generalized over the entire region.

4 Methods

The two main attributes of any probabilistic forecasting system are reliability and discrimination ability, which together determine the performance of the system (Jolliffe and Stephenson, 2011). Reliability and discrimination ability are defined for events exceeding a certain rainfall threshold (e.g., 50 mm/12h). Hereafter, this rainfall threshold will be referred to as verifying rainfall threshold (VRT). Reliability measures whether the chosen VRT is predicted with a probability that equals the average frequency at which such an event is observed. Discrimination measures the ability of the forecasting system to distinguish situations that lead to events exceeding the VRT or not. This study faced two main challenges. Since they were not known a priori, the first challenge was to define the magnitude (in mm) of flash-flood-triggering rainfall events to be used as the VRT values in the objective verification analysis. The pink area in the flowchart in Figure 7 provides a graphical representation of the steps described in section 4.1 to define the VRT values. The second challenge relates to the objective verification analysis. First, due to the intrinsic characteristics of the observational dataset, it was not possible to define univocally observational yes- and no-events. Second, since the probabilistic rainfall forecasts were converted into binary predictions (yes- or no-event), reliability in the usual probabilistic sense cannot be computed. The frequency bias was then calculated to determine the overall reliability of the system, indicating whether the system is, on average, under- or over-forecasting the occurrence of flash flood events. The green area in the flowchart in Figure 7 provides the graphical representation of the steps taken in this study to conduct the objective verification, whose details are described in section 4.2.

4.1 Definition of the verifying rainfall thresholds (VRT)

If not known a priori, VRT magnitudes can be defined in several ways (see Figure 7, pink area “Definition of VRT”). If point rainfall observations are available (e.g., rain gauges or radars), one can create the distribution of the observed flash-flood-triggering rainfall totals. VRT values would then correspond to the specific percentiles of the distribution. The higher the percentile, the higher the magnitude of the VRT, and the higher the severity level of flash flood events considered in the objective verification analysis would be. This approach requires high-density rainfall observations, in both space and time, to capture the extreme (localised) rainfall totals that trigger flash floods (Haiden and Duffy, 2016; Ramos Filho et al., 2021). In the absence of a suitable observational network (for example, in Ecuador, no high-density in-situ 12-hourly rainfall observations were available as shown in Figure 4a and Figure 4b), the VRT values can be defined only from gridded rainfall products such as reanalysis e.g., ERA5 (Hersbach et al., 2020), reforecasts (Hamill et al., 2006), or blended rainfall observations provided on a grid such as MSWEP (Beck et al., 2019) or GPCP (Adler et al., 2018). These datasets tend to, however, underestimate rainfall extremes because of their coarse resolution (Tapiador et al., 2019).

In the absence of more suitable point-scale datasets, this study proposes a methodology for creating a

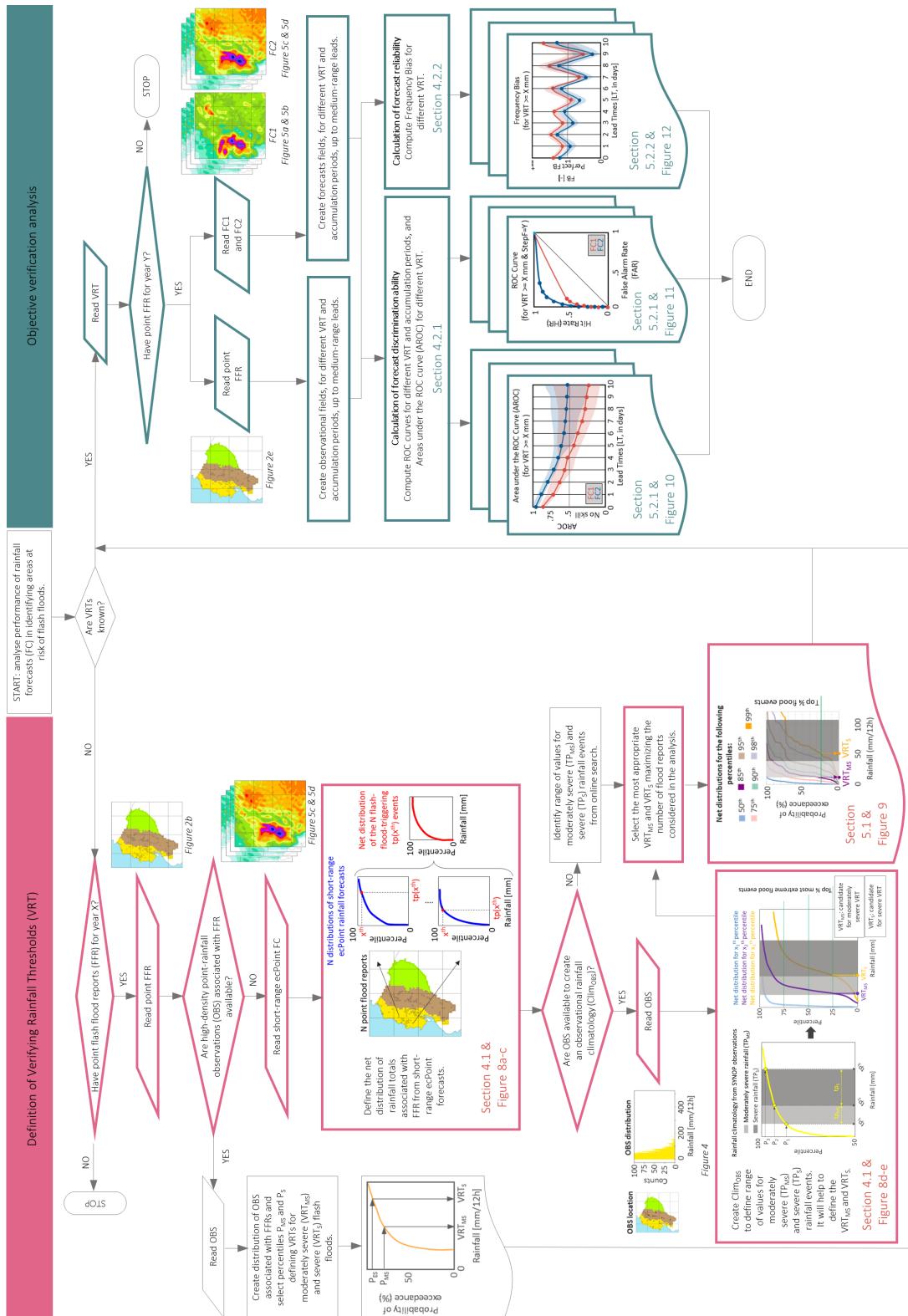


Figure 7: Flowchart of the general methodological steps to determine the verifying rainfall thresholds (pink area) and to assess the forecasts' performance in identifying areas at risk of flash floods (green area). The coloured boxes highlight the actual steps taken in this study.

Schematic representation of how short-range ecPoint rainfall forecasts are used to determine verifying rainfall thresholds (VRT)

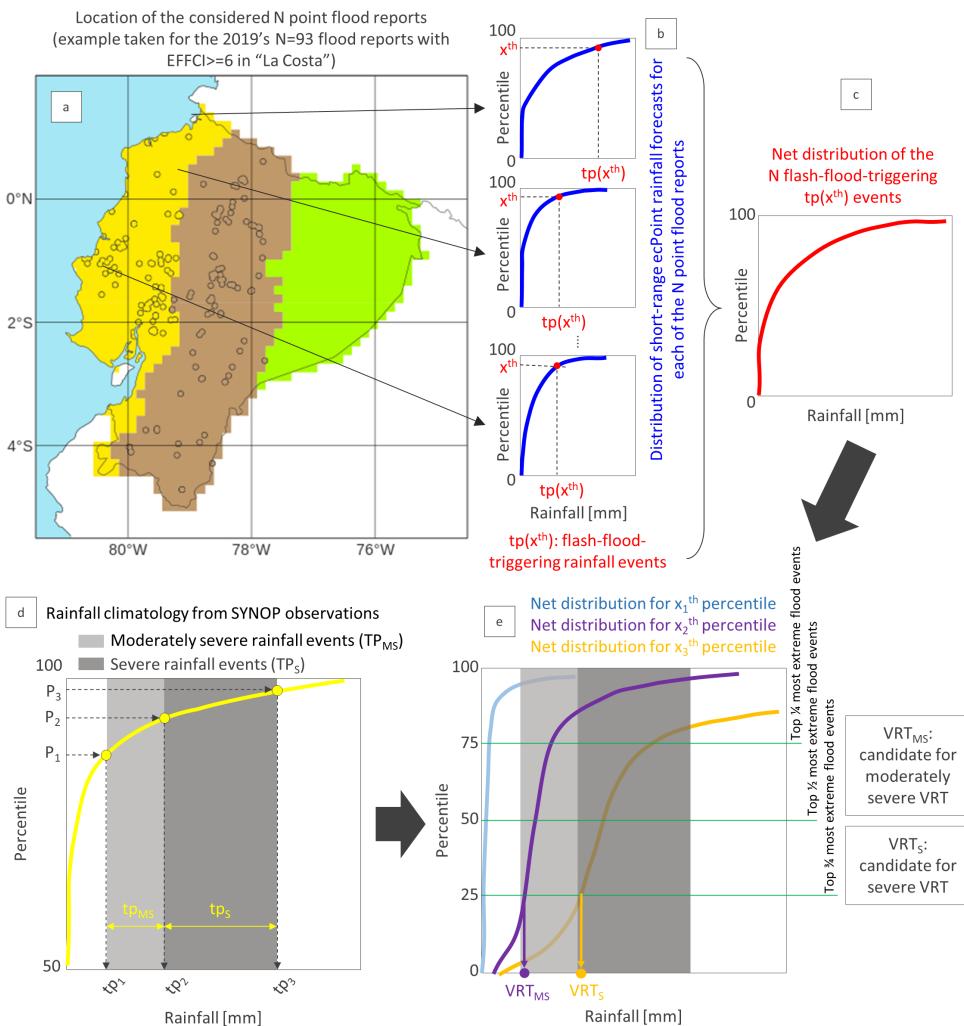


Figure 8: Schematic representation of how short-range ecPoint rainfall forecasts are used to determine verifying rainfall thresholds (VRT) using 2019's $N=93$ point flood reports with $\text{EFFCI} \geq 6$ in "La Costa" as an example (see Figure 2b and Table 1). Panel (a) shows the location of the N point flood reports. Panel (b) shows examples of possible distributions of short-range ecPoint rainfall forecasts associated with each of the N point flood reports (in blue). If a sufficiently large x^{th} percentile is considered in each distribution, the corresponding $tp(x^{\text{th}})$ values can be regarded as flash-flood-triggering rainfall events (in red). Panel (c) displays the net distribution (in red) of the N flash-flood-triggering rainfall events. In the example, only one distribution is shown. However, as many distributions as desired can be created by considering different percentiles. Panel (d) shows the rainfall climatology (in yellow, as this example refers to "La Costa") built from SYNOP observations. The percentiles P_1 and P_2 , and the corresponding rainfall values tp_1 and tp_2 , define the range of moderately severe rainfall values (tp_{MS}), whose width is visually highlighted by the light grey rectangle. The percentiles P_2 and P_3 , and the corresponding rainfall values tp_2 and tp_3 , define the range of severe rainfall values (tp_S), whose width is visually highlighted by the dark grey rectangle. Using the two rainfall ranges and a pre-established top fraction of the most extreme N flood reports, panel (e) illustrates how candidates for moderately severe VRT_{MS} and severe VRT_S are determined when a series of net distributions of flash-flood-triggering rainfall events are available.

Table 2: Day 1 ecPoint-Rainfall forecasts used to define the verifying rainfall thresholds (VRT).

Flood reports' reporting time (in UTC for day X)	Valid 12-hourly periods (in UTC) containing the flood reports' reporting time	Correspondent 12-hourly accumulation periods for day 1 ecPoint-Rainfall forecasts (forecast run date / forecast run time in UTC / lead time in hours)
Between 0 and 5:59	18 (on day X-1) to 5:59	Day(X-1) / 00 UTC / (t+18, t+30) Day(X-1) / 12 UTC / (t+6, t+18) Day(X) / 00 UTC / (t+0, t+12) Day(X-1) / 12 UTC / (t+12, t+24)
	0 to 11:59	Day(X) / 00 UTC / (t+0, t+12) Day(X-1) / 12 UTC / (t+12, t+24) Day(X) / 00 UTC / (t+6, t+18) Day(X-1) / 12 UTC / (t+18, t+30)
Between 6 and 11:59	0 to 11:59	Day(X) / 00 UTC / (t+0, t+12) Day(X-1) / 12 UTC / (t+12, t+24) Day(X) / 00 UTC / (t+6, t+18) Day(X-1) / 12 UTC / (t+18, t+30)
	6 to 17:59	Day(X) / 00 UTC / (t+6, t+18) Day(X-1) / 12 UTC / (t+18, t+30) Day(X) / 00 UTC / (t+12, t+24) Day(X) / 12 UTC / (t+0, t+12)
Between 12 and 17:59	6 to 17:59	Day(X) / 00 UTC / (t+6, t+18) Day(X-1) / 12 UTC / (t+18, t+30) Day(X) / 00 UTC / (t+12, t+24) Day(X) / 12 UTC / (t+0, t+12)
	12 to 23:59	Day(X) / 00 UTC / (t+12, t+24) Day(X) / 12 UTC / (t+0, t+12) Day(X) / 00 UTC / (t+18, t+30) Day(X) / 12 UTC / (t+6, t+18)
Between 18 and 23:59	12 to 23:59	Day(X) / 00 UTC / (t+12, t+24) Day(X) / 12 UTC / (t+0, t+12) Day(X) / 00 UTC / (t+18, t+30) Day(X) / 12 UTC / (t+6, t+18)
18 to 5:59 (on day X+1)		

synthetic distribution of flash-flood-triggering rainfall events using short-range ecPoint rainfall forecasts (Figure 7, pink area). The synthetic distribution will also be compared with the distribution obtained from the low-resolution point-rainfall observations available in Ecuador. At short-range leads, the ecPoint rainfall realisations can be considered proxies for point rainfall observations within the grid boxes (Hewson and Pillou, 2021). Each of the N flood reports is associated with ecPoint rainfall forecasts from the nearest grid box (Figure 8a). At each forecast run, two 12-hourly accumulation periods span each flood report's reporting time (see the first and second columns of 2, so a distribution of 396 ecPoint rainfall realisations (that is, 99 ecPoin rainfall values \times 2 accumulation periods \times 2 runs) can be built for each flood report (blue lines in Figure 8b). Owing to the high number of forecast realizations per flood report, only one year of short-range forecasts is sufficient to define the VRT values. The rainfall value (tp) associated with the X^{th} percentile of the distribution (red dots in Figure 8b) characterizes a certain level of severity for a flash-flood-triggering rainfall event. The net distribution of the N $tp(x^{\text{th}})$ values (red distribution in Figure 8c) represents the distribution of flash-flood-triggering rainfall events. Several reasonably high X^{th} percentiles, eg., the 50th, 75th, 85th, 90th, 95th, 98th, and 99th, were tested to exclude low rainfall totals from the analysis as they would have unlikely been the drivers of any flash flood event. To choose which X^{th} percentiles retain in the analysis, two severity categories were defined from the distribution of flash-flood-triggering rainfall events computed from a time series of 10 years SYNOP observations (Figure 8d): moderately severe (MS) between $P_1 = 95^{\text{th}}$ and $P_2 = 99^{\text{th}}$ percentiles, and severe (S) between $P_2 = 99^{\text{th}}$ and $P_3 = 99.99^{\text{th}}$ percentiles. The values of the percentiles P_1 , P_2 , and P_3 were chosen considering best practices for the definition of thresholds found in the literature (Zsoter et al., 2020; Mitheu et al., 2023) and Hewson and Pillou (2021) recommendations for defining extremes in ecPoint rainfall forecasts. Finally, the corresponding VRT_{MS} and VRT_{S} were defined by deciding how many flood reports would like to be retained in the analysis, for example, the top 1/4, 1/2, or 3/4 (Figure 8d). While such decision can be seen as arbitrary, it depends on the number of reports available at the beginning of the analysis, so that a reasonable number of events are maintained to produce robust statistics. In this study, the top 3/4 flood events were maintained due to the initial small number of events in

Table 3: Definition of the four quadrants in a contingency table.

FORECASTS (COLUMNS)/ OBSERVATIONS (ROWS)		YES	NO
		QUADRANT I Hits (H) The event was observed when it was predicted.	QUADRANT II False Alarms (FA) The event was not observed when it was predicted.
YES	QUADRANT III Misses (M) The event was observed when it was not predicted.	QUADRANT IV Correct Negatives (CN) The event was not observed when it was not predicted.	
	NO		

the database (Table 1). Separate VRT values were calculated for “La Costa” and “La Sierra” to capture their hydro-climatological regimes.

4.2 Objective verification to assess the rainfall forecasts’ performance in the identification of areas at flash flood risk

4.2.1 Assessment of forecasts’ discrimination ability: Receiver Operating Characteristic (ROC) curves and Area Under the ROC curve (AROC)

This study used the Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AROC) to estimate and compare the discrimination ability of ENS and ecPoint rainfall forecasts in the prediction of areas at risk of flash floods ([Jolliffe and Stephenson, 2011](#)). ROC curves were constructed using a 2×2 contingency table that quantifies the hits (H), misses (M), false alarms (FA), and correct negatives (CN) that occur when action is advised based on the VRT exceeding each sampled probability threshold (see Table 3 for the definition of the constituting elements of the contingency table). Correspondent hit rates (HR) and false alarm rates (FAR) can be then computed, respectively, from equations (1) and (2):

$$HR = \frac{H}{H+M} \quad [\text{values between 0 and 1}] \quad (1)$$

$$FAR = \frac{FA}{FA+CN} \quad [\text{values between 0 and 1}] \quad (2)$$

For each sampled probability threshold, the ROC curves map HRs on the Y-axis against FARs on the X-axis. The location of the ROC curve in the graph and the geometrical area under the ROC curve (AROC) determine the discrimination ability of the forecasting system. Perfect discrimination is obtained when only HRs grow, whereas the FARs always remain equal to zero. This is represented by an ROC curve that rises from the bottom left corner (0,0) along the Y-axis to the top-left corner (0,1) and moves straight to the top-right corner (1,1). In this case, the AROC was equal to 1. When the forecasting system has no discriminatory ability (i.e., it does not provide any additional information beyond climatological predictions), the HRs and FARs grow at the same rate. Therefore, the ROC curve lies along the diagonal of the graph, and the AROC is equal to 0.5. How ROC curves are built and AROCs are computed has a significant impact on the interpretation of the results. To ensure that the ROC curves are as complete as possible, probability thresholds are determined using the full discretization available in the ensemble, rather than using fixed percentage bins ([Bouallègue and Richardson, 2022](#)). Hence, ROC curves for ENS and ecPoint were constructed with 51 and 99 points, respectively. No curve fitting was used to build

or complete the ROC curves, and straight lines were drawn between consecutive points in the graph, as well as the last meaningful point of the ROC curve and the top-right corner ([Bouallègue and Richardson, 2022](#)). Moreover, AROCs are computed using the trapezoidal approximation, which simply sums the areas of single trapeziums formed by the straight lines between the ROC's consecutive points. As a result, ROC curves for high VRT values might cluster on the bottom left corner of the graph, and if built with fewer points, they might look incomplete, and AROCs might result smaller. However, this approach focuses on the analysis of “real” and not on the “potential” discrimination ability of rainfall forecasts. The percentile bootstrapping technique was applied to evaluate whether the differences between the AROC for ENS and ecPoint were significant ([DiCiccio and Efron, 1996](#)). Sampling with replacement with 10,000 replicates and 95% confidence intervals were considered.

Populating the contingency tables (Table 3) is one of the challenges of this objective verification analysis. Stationary observations (i.e., provided by instruments installed at a specific location, e.g., rain gauges) provide time series that record both yes- and non-events at the location where the instrument was installed. Thus, all four elements in the contingency table can be quantified. Non-stationary observations record only yes-events at the location where the events occurred. As a result, it is impossible to answer the question “if there are no reports in an area, is it because an event happened, but nobody reported it, or because there was no event to report?”. Some studies verify only yes-events with the caveat that only quadrants I (i.e., hits) and III (i.e., misses) of the contingency table can be populated ([Robbins and Titley, 2018](#)). Instead, this study followed the method proposed by [Tsonevsky et al. \(2018\)](#), which allows to populate all quadrants of the contingency table. This method assumes that a non-report corresponds to a non-event. Because of the high-impact nature of the event and the care used to create the observational flood database, this assumption was also considered valid for this study. Observational fields are built by assigning 1 to grid-boxes containing at least one flood report (i.e., observational yes-event); otherwise, a value of 0 is assigned (i.e., observational non-event). Forecast fields are built by assigning a value of 1 to those grid-boxes where the considered VRT is exceeded with a considered probability threshold (i.e., forecast yes-event); otherwise, the grid boxes are assigned a value of 0 (i.e., forecast non-event). The 2X2 contingency tables are built by examining overlapping grid boxes in correspondent observational and forecast fields; when both grid boxes are assigned a value of 1 or 0, they count as H and CN, respectively. When a grid box in the observational field is assigned a value of 1, and the corresponding grid box in the forecast field is assigned a value of 0, it counts as M. It counts as a false alarm (FA) if the opposite occurs.

4.2.2 Assessment of forecasts’ calibration: Frequency Bias (FB)

The frequency bias was used to evaluate the overall reliability of ENS and ecPoint rainfall forecasts in the prediction of areas at risk of flash floods. For each lead time, the frequency bias was determined by dividing the total number of ensemble members exceeding the considered VRT by the product of the number of ensemble members and the total number of instances when a flash flood was observed. Equation (3) was used to compute the FB:

$$FB = \frac{\sum_{i=1}^M \sum_{j=1}^N n_i \text{ ensemble members exceeding VRT}}{(n_i \text{ ensemble members}) \times (\sum_{i=1}^M \sum_{j=1}^N n_i \text{ instances when a flash flood was observed})} \quad (3)$$

FB values range from 0 to $+\infty$. FB=1 indicates perfect calibration. Scores greater and smaller than 1 indicate, respectively, that the forecasting system over- and under-predicts the observed events. It is worth noting that FB measures the overall ratio of forecast events to observed events and is not a measure

of forecast skill. As such, it can provide a score of 1 when there are compensating errors. It can also be observed that a forecasting system appears to overestimate the observed number of events if the observed events are heavily underreported.

5 Results

5.1 Verifying rainfall thresholds

The rainfall climatologies computed from the SYNOP observations show in “La Costa” (Figure 9a) ranges for tp_{MS} (between 10 and 45 mm/12h) and tp_S (between 45 and 108 mm/12h) that are much bigger than in “La Sierra” (Figure 9c), where tp_{MS} ranges between 10 and 25 mm/12h and tp_S ranges between 25 and 64 mm/12h. Considering these ranges, the net distributions of flash-flood-triggering rainfall events corresponding to the 50th and 75th percentiles provide rainfall values that are deemed too small (Figure 9b and Figure 9d). In both regions, the net distributions for the 85th, 90th, 95th, and 98th percentiles provide good candidates to define the VRT_{MS} , while the net distribution for the 90th, 95th, and 98th, and 99th percentiles provide good candidates to define the VRT_S . To increase the count of predicted moderately severe and severe events (and so produce robust statistics), it was decided to define the VRT_{MS} and VRT_S from the first net distribution (within the ranges provided) intersected by the horizontal line corresponding to retain the top 3/4 of the observed flood reports. Therefore, VRT_{MS} and VRT_S are defined from the net distribution corresponding to the 85th and 99th percentile. The VRT_{MS} and VRT_S are indicated by purple and orange dots in Figure 9c for “La Costa” and in Figure 9f for “La Sierra.” The rounded values in mm/12h, used in the objective verification analysis are listed in Table 4.

5.2 Objective verification

5.2.1 Forecast discrimination ability

Figure 10 shows the evolution of the AROC with the lead time for VRT_{MS} and VRT_S for “La Costa” and “La Sierra.” No degradation of the AROC values with lead time was observed in “La Costa” (Figure 10a and Figure 10b), while there is some degradation observed in “La Sierra” (Figure 10c and Figure 10d). The AROC values for ENS and ecPoint diminished at the same rate for both VRT. Overall, the AROC values were larger for VRT_{MS} (Figure 10a and Figure 10c) than for VRT_S (Figure 10b and Figure 10d), and for “La Costa” (10a and Figure 10b) compared to “La Sierra” (10c and Figure 10d). A feature that stands out in all panels in Figure 10, but especially in “La Costa,” is the sinusoidal pattern shown by the AROC in correspondence with the different accumulation periods throughout the day. In “La Costa” (Figure 10a and Figure 10c), the peaks in the AROC are observed between 0000-1200 LT (i.e., lead time steps labelled in purple) and 0600-1800 LT (i.e., lead time steps labelled in cyan, that correspond to rainfall occurred mainly during daytime), while troughs are mostly observed between 1200-0000 LT (i.e., lead time steps labelled in pink) and 1800-0600 LT (i.e., lead time steps labelled in green, that correspond to rainfall occurred mainly during night-time). Overall, the discrimination ability for ecPoint in “La Costa” tended to be better than the discrimination ability for ENS, especially for VRT_S (Figure 10c). However, it is worth noting that ecPoint and ENS’s AROC values are closer in the peaks and further apart in the troughs, meaning that the improvements that ecPoint brings in the identification of areas at risk of flash floods are mainly over rainfall events occurring in the daytime. A similar sinusoidal pattern, although noisier, was observed for the AROC values in “La Sierra” (Figure 10c and Figure 10d). Unlike in “La Costa”, in “La Sierra” there are no specific times of the day where ecPoint adds value to the

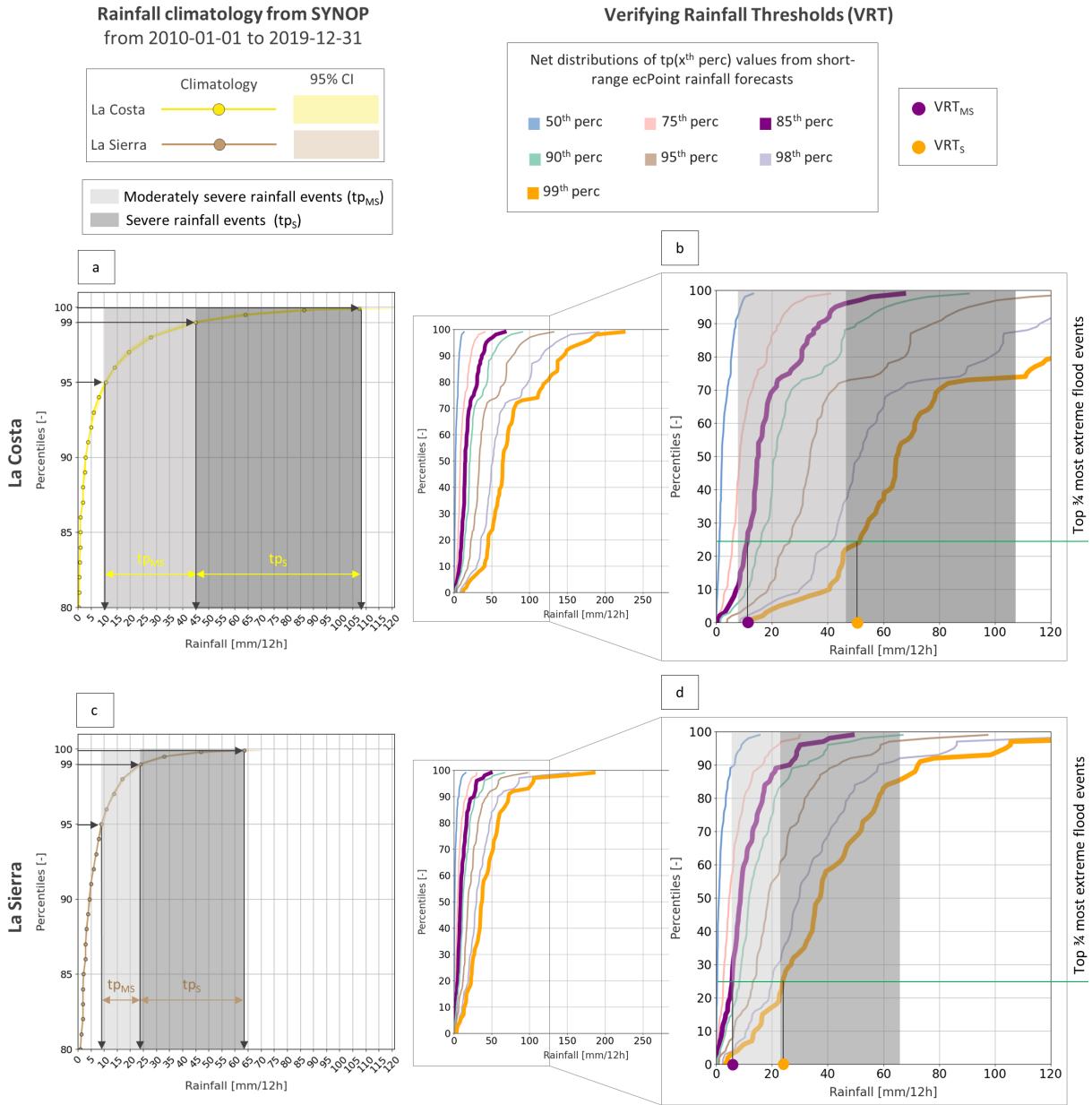


Figure 9: Panels (a) and (c) show the rainfall climatologies from SYNOP observations, respectively, for “La Costa” and “La Sierra”, with 95% confidence intervals. The range of moderately severe (tp_{MS}) and severe rainfall totals (tp_S) are indicated, respectively, with a light and dark grey rectangle. Panels (b) and (d) show, respectively, the net distributions of flash-flood-triggering rainfall events (i.e., distribution in red in Figure 8c) for “La Costa” and “La Sierra”, built from the rainfall totals corresponding to the 50th (in pale blue), 75th (in pale pink), 85th (in purple), 90th (in pale green), 95th (in pale brown), 98th (in grey), and 99th percentiles (in orange) in the distributions of short-range ecPoint rainfall forecasts (i.e., the x th percentile indicated in red in Figure 8b). The VRT_{MS} (purple circle) and VRT_S (orange circle) are defined using the net distribution that contains the cross point between the top 3/4 most extreme flood events and the lower threshold of, respectively, the range of moderately severe and severe rainfall events.

Table 4: Verifying rainfall thresholds (in mm/12h).

Region	VRT85	VRT99
La Costa	9.865 mm/12h (rounded to 10 mm/12h)	50.452 mm/12h (rounded to 50 mm/12h)
La Sierra	5.885 mm/12h (rounded to 6 mm/12h)	25.551 mm/12h (rounded to 26 mm/12h)

performance of ENS. For VRT_{MS} (Figure 10c), ENS shows an overall better discrimination ability than ecPoint, while the latter shows an overall better performance than ENS in the AROC values for VRT_{S} (Figure 10d).

The analysis of the ROC curves provides further information on the discrimination ability of forecasts. Figure 11 shows the ROC curves for ENS (in red) and ecPoint (in blue) for the 12-hourly accumulation period ending at $t+72$ (i.e., day 3 forecast), which corresponds to a mainly daytime rainfall whose valid accumulation period ends at 1800 LT. For both regions, the AROC is larger. However, substantial differences were observed in the shapes of ROC curves. For “La Costa” (continuous line), the ROC curves are mostly overlapping, and only the points corresponding to the two top percentiles in ecPoint (98th and 99th) are above the ROC curve for ENS. This means that the underlying curve for ENS and ecPoint is the same, and the greater AROC from ecPoint is due to the capability to identify extremes better thanks to the bigger number of ensemble members. In this case, it can be concluded that ENS contains the information required to predict whether an extreme flash-flood-triggering rainfall event might occur, although it might not be possible to predict the actual magnitude of the event. To achieve this goal, forecasts from ecPoint are better suited because they can double the hit rate from the ENS with only a very small increase in the false alarm rate. On the contrary, although noisy, it can be observed that the ROC curves in “La Sierra” (dashed lines) do not overlap. This implies that the underlying curves for ENS and ecPoint are different. Thus, ecPoint can identify events for which ENS may not provide any signal.

5.2.2 Forecast frequency bias

Figure 12 shows the evolution of the FB with lead time for VRT_{MS} and VRT_{S} , for “La Costa” and “La Sierra.” No significant degradation in FB was observed with the lead time in either region. Overall, FB is larger for VRT_{MS} (Figure 12a and Figure 12c) than for VRT_{S} (Figure 12b and Figure 12d), and for “La Sierra” (Figure 12c and Figure 12d) than for “La Costa” (Figure 12a and Figure 12b). In “La Costa,” ecPoint diminishes the FB at all lead times, but especially for the accumulation period that ends at 0000 LT (steps indicated in pink) in the case of VRT_{MS} and for the accumulation period that corresponds to mainly night-time rainfall and ends at 0600 LT (steps indicated in green) in the case of VRT_{S} . For the latter, only accumulation periods that correspond mainly to daytime rainfall and end at 1800 LT (steps indicated in cyan) show a slightly worse FB for ecPoint compared to ENS. ecPoint’s FB for VRT_{MS} in “La Sierra” (Figure 12c) is significantly better than ENS’s FB, while for VRT_{S} (Figure 12d) it is worse, with the exception of the accumulation periods that correspond mainly to night-time rainfall and end at 0600 LT (steps indicated in green). It is worth noting that opposite to what happens in “La Costa” where the performance is qualitatively the same for both forecasting systems (i.e., the peaks and troughs are observed for the same accumulation periods in both ENS and ecPoint), in “La Sierra” peaks and troughs happen in different accumulation periods for different forecasting systems. For VRT_{MS} , peaks occurred for daytime rainfall in the ENS and for night-time rainfall in ecPoint. The opposite is true for the VRT_{S} .

A feature that stands out in all panels in Figure 12 is that the FB values are significantly larger than 1 (i.e., the perfect value for FB), indicating that ENS and ecPoint significantly overestimate the frequency

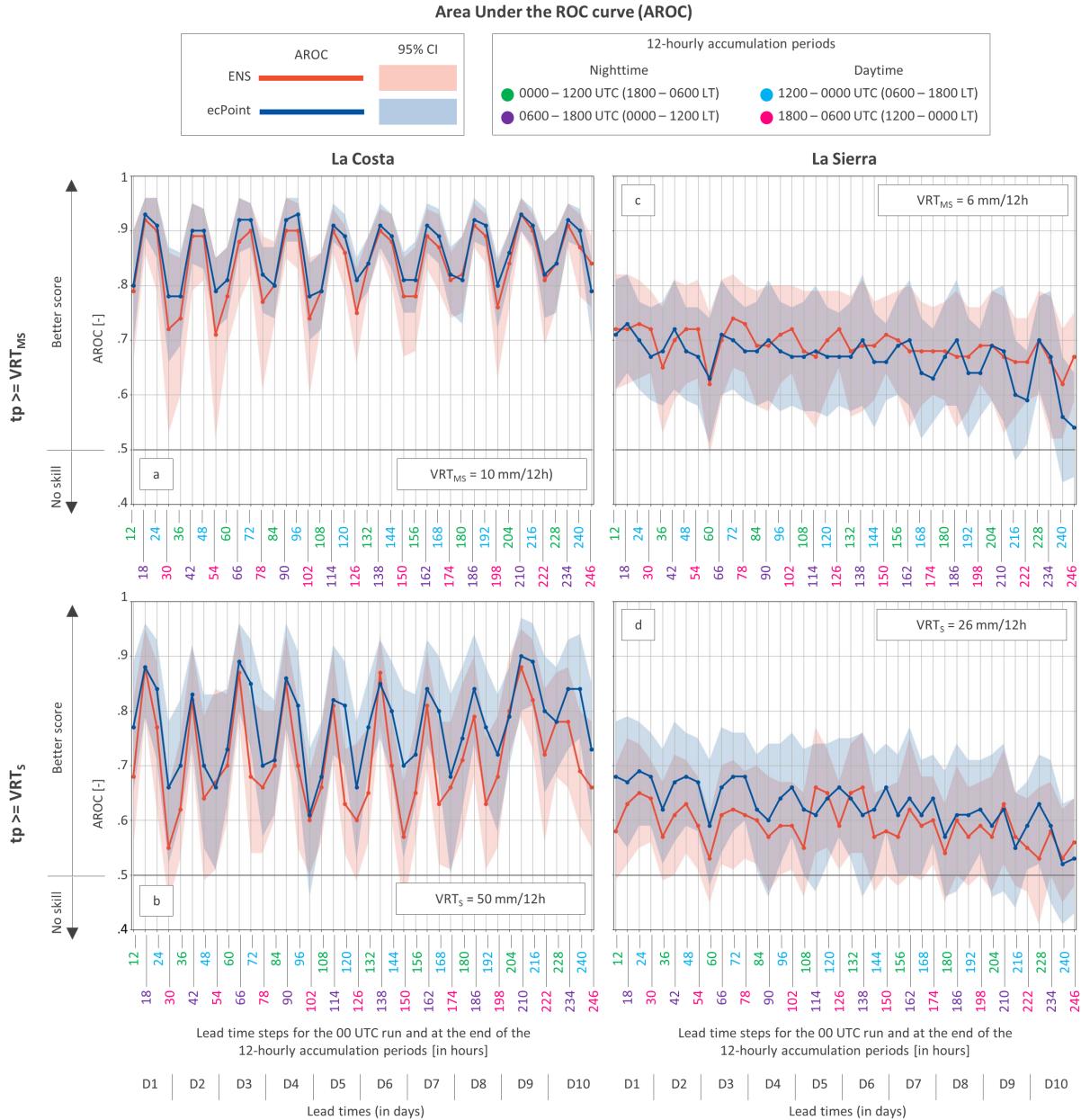


Figure 10: Area under the ROC curve (AROC) for flood reports with $EFFCI \geq 6$. Panels (a) and (b) show the AROC, respectively, for $VRT_{MS} \geq 10 \text{ mm/12h}$ and $VRT_S \geq 50 \text{ mm/12h}$ in “La Costa”. Panels (c) and (d) show the AROC, respectively, for $VRT_{MS} \geq 6 \text{ mm/12h}$ and $VRT_S \geq 26 \text{ mm/12h}$ in “La Sierra”. The lines and the shaded areas represent, respectively, the values of the AROC and the confidence interval (CI) at 95% for ENS (in red) and ecPoint (in blue). The x-axis indicates the lead time steps for the 00 UTC run at the end of the 12-hourly accumulation period, expressed in hours. The colours associated with each step indicate the valid 12-hourly accumulation periods in UTC and local time (LT). The steps in green for 0000-1200 UTC (or 1800-0600 LT) and in purple for 0600-1800 UTC (or 0000-1200 LT) represent the 12-hourly accumulation periods during the night-time. The steps in cyan for 1200-0000 UTC (or 0600-1800 LT) and in fuchsia for 1800-0600 UTC (or 1200-0000 LT) represent the 12-hourly accumulation periods during the daytime. Lead times are also expressed in days (from 1 to 10).

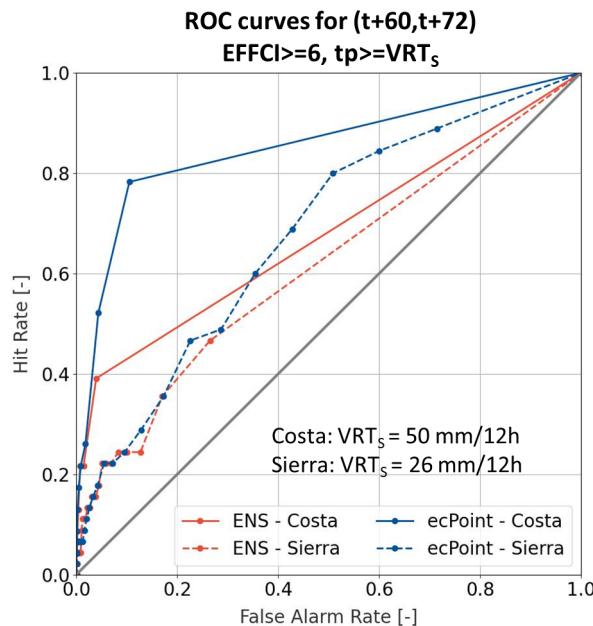


Figure 11: ROC curves built with flood reports with $\text{EFFCI} \geq 6$, for $\text{tp} \geq \text{VRT}_S$ and for the accumulation period between $(t+60, t+72)$, i.e., for 0600 to 1800 local time (LT). The red and the blue lines denote, respectively, the ROC curves for ENS and ecPoint. The continuous and the dashed lines correspond to the ROC curves for “La Costa” and “La Sierra”.

of flash flood events. Figure 13 shows the counts of yes-events in forecasts and observations for a day 3 forecast (i.e., accumulation period ending at $t+72$). In “La Costa” (Figure 13a and Figure 13b) there is a good overall correspondence between days with yes- and non-flash-flood events, especially for VRT_S . The larger FB is mainly due to the slightly larger count of grid-boxes corresponding to yes events in the forecasts. In “La Sierra” (Figure 13c and Figure 13d), there is also a good representation of the periods with yes- and non-flash-flood events, especially for VRT_S . However, the count of grid-boxes corresponding to yes-events in the forecasts is much larger than in the observations, contributing to FB values that are well above the perfect value of 1.

6 Case study: intense rainfall and flash floods on the 8th of March 2021

This subjective verification analysis is presented to further support the outcomes of the objective verification analysis. March is one of the wettest months in 2021 in Ecuador. As a result of numerous heavy rainfall events, rivers such as Guayas, Los Ríos, Esmeraldas, and Manabí burst their banks, with landslides being observed in many different regions. The 8th of March was one of the wettest days (Figure 14). Significant impacts⁴ were mainly observed in the highly populated city of Guayaquil, where very heavy rainfall was reported to occur in the afternoon after 4 pm (Local Time, LT), with rainfall totals exceeding 100 mm/24h⁵ in the city centre (zoomed red area in Figure 14a). Around the 8th of March, the MJO was reported by various centres to be in phase 8⁶, which tends to be conducive to, or at least

⁴<https://www.eluniverso.com/guayaquil/comunidad/la-mayor-lluvia-del-2021-en-guayaquil-provoco-afectaciones-en-64-zonas-entre-inundaciones-arboles-caidos-canales-rebosados-y-otros-nota/>

⁵<https://www.wunderground.com/history/daily/SEGU/date/2021-3-8>

⁶https://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/ARCHIVE/PDF/mjo_evol-status-fcsts-20210315.pdf

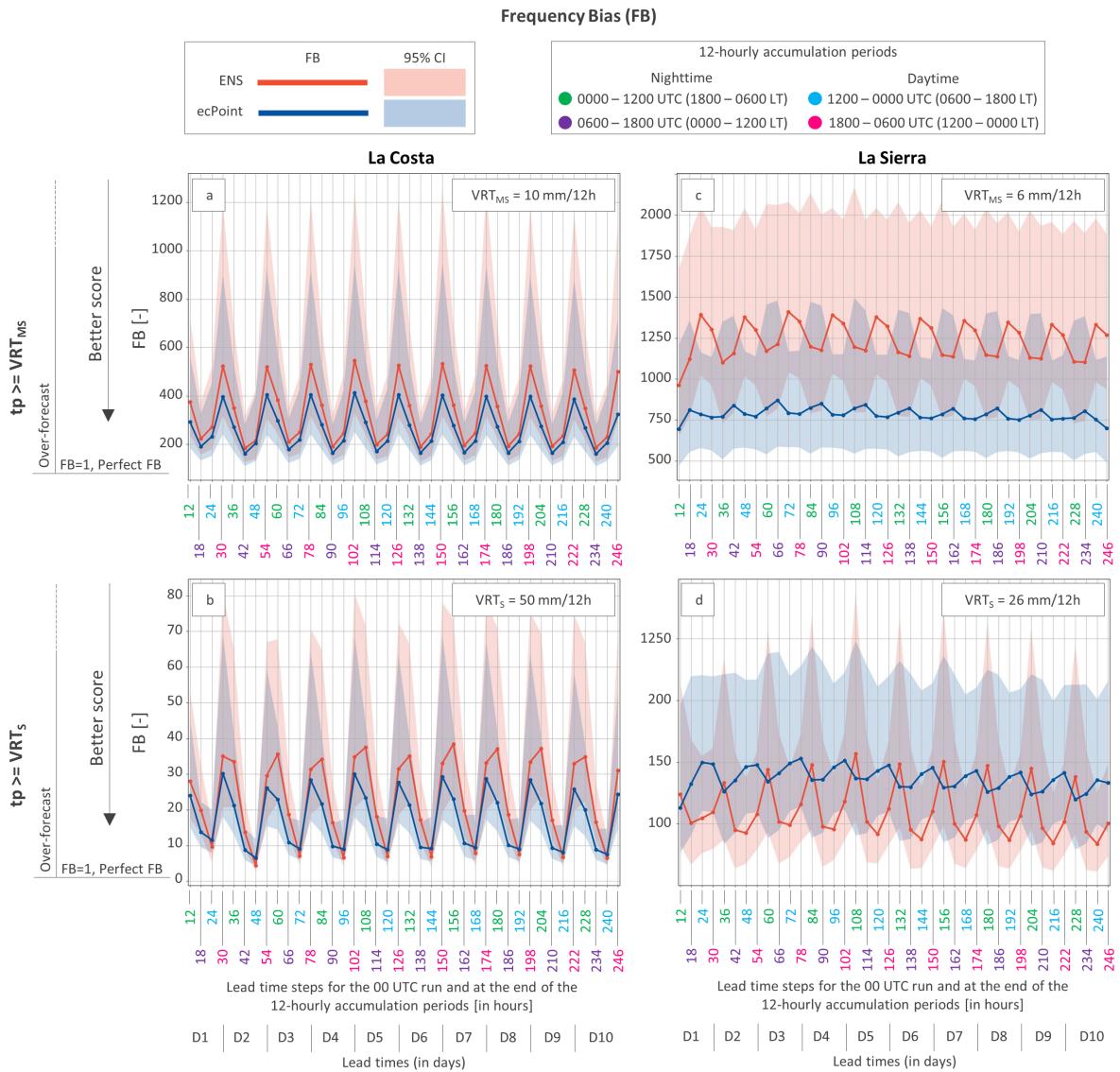


Figure 12: Frequency bias (FB) for flood reports with $\text{EFFCI} \geq 6$. Panels (a) and (b) show the FB, respectively, for $\text{VRT}_{\text{MS}} \geq 10 \text{ mm}/12\text{h}$ and $\text{VRT}_S \geq 50 \text{ mm}/12\text{h}$ in “La Costa”. Panels (c) and (d) show the FB, respectively, for $\text{VRT}_{\text{MS}} \geq 6 \text{ mm}/12\text{h}$ and $\text{VRT}_S \geq 26 \text{ mm}/12\text{h}$ in “La Sierra”. The lines and the shaded areas represent, respectively, the values of the FB and the confidence interval (CI) at 95% for ENS (in red) and ecPoint (in blue). The x-axis indicates the lead times steps for the 00 UTC run at the end of the 12-hourly accumulation period, expressed in hours. The colours associated with each step indicate the valid 12-hourly accumulation periods in UTC and local time (LT). The steps in green for 0000-1200 UTC (or 1800-0600 LT) and in purple for 0600-1800 UTC (or 0000-1200 LT) represent the 12-hourly accumulation periods during the night-time. The steps in cyan for 1200-0000 UTC (or 0600-1800 LT), and in fuchsia for 1800-0600 UTC (or 1200-0000 LT) represent the 12-hourly accumulation periods during the daytime. The lead times are also expressed in days (from 1 to 10).

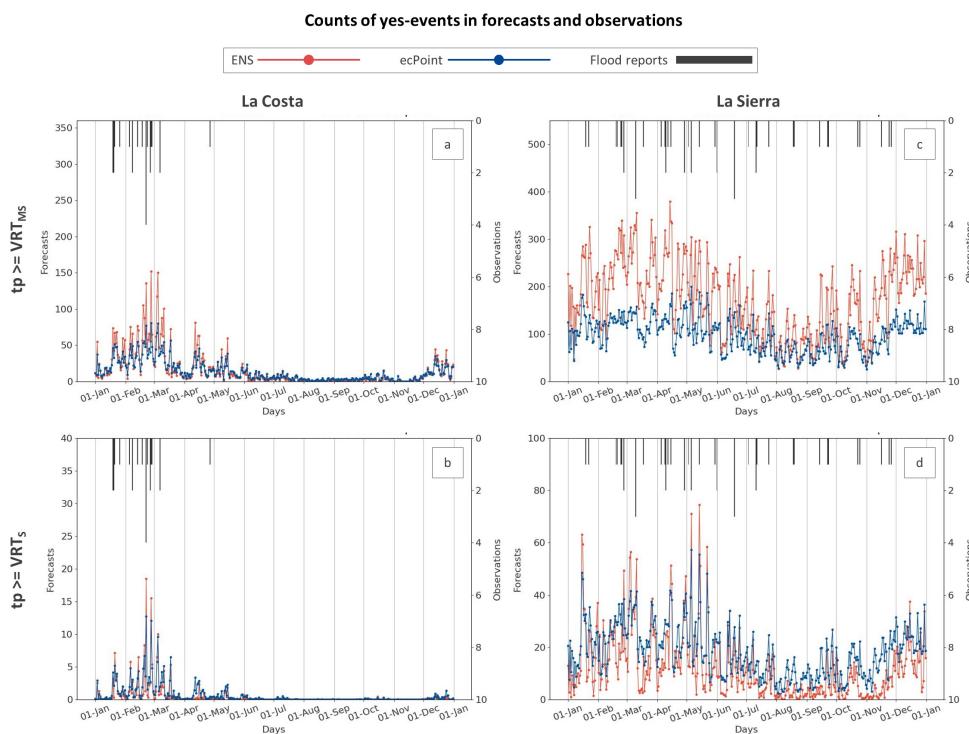


Figure 13: Counts of yes-events (flash floods) in forecasts and observations. The yes-events in the forecasts are displayed on the bottom x-axis, in red for ENS and in blue for ecPoint. The example considered is for the accumulation period ending at $t+72$ (day 3 forecast, daytime rainfall). The yes events in the observations are displayed in black on the top x-axis. Panels (a) and (b) show the counts for “La Costa” for $tp \geq VRT_{MS}$ and VRT_S , respectively. Panels (c) and (d) show the same but for “La Sierra”.

correlated with, onshore lower tropospheric westerly wind anomalies near the equatorial west-facing coasts of South America ([Wheeler and Hendon, 2004](#)). In conjunction, analysts from the NOAA have highlighted the likelihood of enhanced convective activity in the region in routine bulletins. ECMWF's numerical model sounding (Figure 14b) valid for on the 8th of March at 6 am (LT) appeared particularly conducive to flash-flood-triggering rainfall activity. For instance, the very high Convective Available Potential Energy (CAPE) shows that there is potential for sufficiently high dew point depression insolation-based triggering that might not be impeded by thick clouds. It also shows the potential for very high-altitude convective cloud tops, very strong wind shear that favours prolonged convective cell life cycles (as down-draughts would not interfere with up-draughts), and relatively light steering winds (favouring slow movement of convective cells). This description is supported by SYNOP and METAR observations and satellite imagery (not shown), suggesting that the cause of this rainfall event was organized convective cells, whose development was triggered by insolation.

Figure 14c shows ENS and ecPoint rainfall forecasts from the 00 UTC run for day 1 (first row), day 3 (second row), and day 7 (third row) lead times. The forecasts are valid for the 12-hourly accumulation period between the 8th of March 2021, at 12 am and the 9th of March 2021, at 0 am (LT), that is, the fraction within the 24-hourly period of the observations reported in Figure 14a when most of the rainfall fell. Forecasts for the 50th (first and second columns), 95th (third and fourth columns), and 99th percentiles (fifth and sixth columns) are shown. The median (i.e., 50th percentile) represents the dividing line for the equi-probable observation categories. By comparing the rainfall observations (Figure 14a) and the forecast for the 50th percentile (first and second columns in Figure 14c), it can be seen that, overall, the ENS overestimates the mean rainfall. On the contrary, owing primarily to its bias correction for rainfall overprediction at the grid-scale, ecPoint's 50th percentile is systematically smaller than in ENS, showing a better fit with the observations. At the same time, ecPoint's 95th (third and fourth columns in Figure 14c) and 99th percentiles (fifth and sixth columns in Figure 14c) highlight a higher potential than the ENS for having higher local rainfall totals in certain areas (e.g., Guayaquil). While far more observations would be needed to analyse robustly the performance of ENS and ecPoint forecasts for such a high percentile, it appears that both ENS and ecPoint predicted well the local rainfall extremes in "La Costa." For example, there is a signal of extreme rainfall in the ENS for Guayaquil at the 95th percentile). In contrast, it appears that ecPoint adds the most value to the prediction of extreme rainfall in "La Sierra." For example, ecPoint's 99th percentile shows a 1% chance of having up to 60 mm/12h at some location in "La Sierra", and one location observed such an amount.

7 Discussion

7.1 Short-range ecPoint rainfall forecasts can be used as proxy for in-situ observations in data scarce regions to calculate VRT

The literature shows that the definition of flash-flood-triggering rainfall thresholds is a still widely used way to predict areas at risk of flash floods ([Ma et al., 2021](#); [Papagiannaki et al., 2015](#); [Ramos Filho et al., 2021](#)). Central to these studies is the need for high-density rainfall observations to establish thresholds without excessive over- or underprediction biases. The lack of high-density, in situ 12-hourly rainfall observations in Ecuador would have limited the ability of defining the flash-flood-triggering rainfall thresholds, used in this study as VRT values. The method for calculating VRT values using ecPoint rainfall forecasts as a proxy for in-situ observations addresses this data gap by simulating a very high-density observational system in the proximity of the flash flood report. However, the probabilistic nature of the ecPoint output raises the following question: of all ecPoint's rainfall realizations associated to a flash

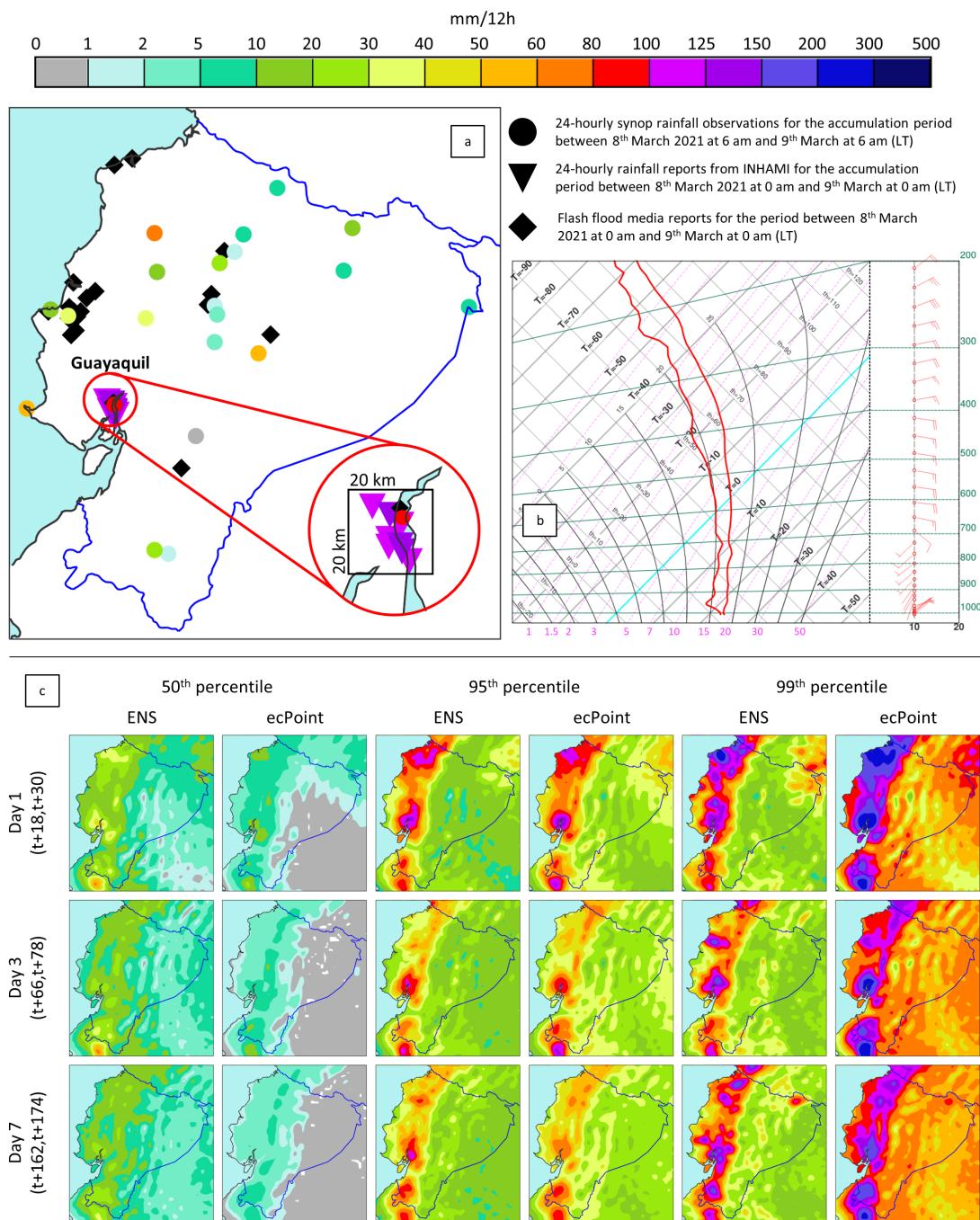


Figure 14: Flash floods in Ecuador on 8th March 2021. Panel (a) shows 24-hourly synop rainfall observations between 8th March at 6 am and 9th March at 6 am (coloured dots), 24-hourly rainfall reports from INAMHI for Guayaquil between 8th March at 0 am and 9th March at 0 am (coloured triangles), and flash flood reports in different regions between 8th March at 0 am and 9th March at 0 am (black diamonds). Panel (b) shows the sounding for Guayaquil (lat: -2.2; lon: -79.9) valid for 8th March 2021 at 6 am. Panel (c) shows day 1, 3, and 7 forecasts from 00 UTC runs for ENS and ecPoint, valid for the accumulation period between 8th March at 12 am and 9th March at 0 am (when the rainfall event was at its peak). All reported times are meant to be in LT.

flood report, which amount represents better the one that triggered the corresponding flash flood event? The comparison of net distributions of potential flash-flood-triggering rainfall events with observational rainfall climatologies (even though computed using low-density point-rainfall observations) is essential to distinguish between net distributions that represent “moderately severe” rainfall events (i.e., events that might or might not generate flash floods), from net distributions that represent “severe” rainfall events, that are very likely to generate flash floods most of the times.

Several studies have found that flash flood forecasts are very sensitive to rainfall’s spatial-temporal variability ([Borga et al., 2014](#); [Demissie et al., 2021](#); [Douinot et al., 2016](#); [Norbiato et al., 2008](#); [Song et al., 2019](#)). While the average rainfall does not seem to vary significantly in “La Sierra,” the coastal areas in “La Costa” appear drier than on the western slopes of the Andes, suggesting that more than one VRT may be required to capture the variability in the magnitude of flash-flood-triggering rainfall events. Furthermore, the differences between daytime and night-time rainfall amounts suggest that a distinction between VRT values for daytime and night-time events could be beneficial to reduce biases in the verification analysis. Due to the limited availability of observational data, it was decided to aggregate all these conditions to provide more robust results, and only two VRT values were computed, one for the entire “La Costa” and one for the entire “La Sierra”. However, this data aggregation comes at the expense of a greater granularity that might be beneficial and required to predict better areas at risk of flash floods. Finally, focusing solely on rainfall when defining VRT values may overlook other important factors. [Dinis et al. \(2021\)](#) highlighted how terrain and urban planning influence flash flood vulnerability in Benguela, Angola, even under moderate rainfall. Therefore, a more comprehensive approach incorporating geomorphological, socio-economic, and environmental factors alongside rainfall data is imperative for refining further the flash flood predictions and reducing false alarms.

7.2 Comparison of ENS and ecPoint’s performance in the prediction of areas at risk of flash floods

In Section 2, we learned that the meteorological dynamics influencing rainfall in “La Costa” and “La Sierra” are significantly different. This contextual backdrop is essential for interpreting the ENS and ecPoint performance in the identification of areas at flash flood risk in the two regions. In “La Costa”, flash-flood-triggering rainfall events are generated predominantly by large-scale convective systems such as El Niño and the MJO ([Recalde-Coronel et al., 2020](#); [Tobar and Wyseure, 2018](#)). ENS can predict with a reasonable degree of accuracy up to six weeks ahead of these systems ([Haiden et al., 2021](#)). Ergo, ENS can proficiently identify areas at risk of heavy rain (and flash floods), albeit it might still misforecast the absolute rainfall amounts. Herein, ecPoint’s contribution is primarily confined to rectifying the rainfall estimates. This type of correction is influential in continuous rainfall amount verification, but it becomes less relevant in binary event verification. Namely, it is not significant if a yes-event is obtained from exceeding the VRT by 1 mm or 20 mm as far as it is exceeded, and the correction of the absolute rainfall totals will matter only when it tips the balance between having or not having a yes-event. This behaviour can be observed for corrections applied by ecPoint to daytime and night-time rain in “La Costa”. The needed reduction of daytime rain in “La Costa” applied by ecPoint to ENS forecasts is rewarded by both AROC and FB metrics. However, no significant changes in performance are observed when unnecessary corrections are applied to night-time rainfall (because the post-processing does not differentiate bias corrections depending on LT): ecPoint’s AROC values are the same as those for ENS, and the FB values are only slightly inferior. Conversely, flash-flood-triggering rainfall events in “La Sierra” are generated predominantly by small-scale convective systems ([Recalde-Coronel et al., 2014](#)). ENS struggles to capture extreme rainfall from such systems even on day 1, and when it does, its predictive skill decreases significantly after a few days ([Haiden et al., 2023](#)). [Hewson and Pillosu](#)

(2021) have shown that ecPoint significantly enhances the detection of small-scale convective systems and increases the lead time at which these systems can be identified in forecasts. As such, ecPoint has shown to perform better in identifying areas at risk of flash floods, especially for the most extreme flash flood events. The case-study-based analysis in Section 6 aligns with the discussion above. It indeed shows that ecPoint reduces the overprediction bias present in ENS while, at the same time, it improves ENS's capability of predicting very extreme localised extreme rainfall events, especially when such an event is originated from a small-scale convective system.

These findings underscore the importance of considering the scale and nature of the weather systems that generate extreme rainfall in the region of interest when evaluating rainfall forecast performance in the identification of areas at risk of flash floods. When considering small-scale convective systems, ecPoint enhances structurally the performance of ENS, while whenever the main cause of extreme rainfall is a large-scale system, ENS can do a respectable job, and ecPoint would mainly help in predicting more accurately the entity (in mm) of the flash-flood-triggering rainfall events. This is consistent with the findings in Bucherie et al. (2022b), who showed that larger-scale patterns linked to the occurrence of flash floods can be discerned in coarser global-scale models. This result is extremely important. So far, the research and the forecasting community have believed that km-scale rainfall forecasts were needed to identify areas at flash flood risk. However, this study establishes that raw global rainfall forecasts can be successfully used in the identification of areas at risk of flash floods under certain weather conditions. The two main benefits of being able to use global raw and post-processed NWP rainfall forecasts are (1) the availability of rainfall forecasts over a continuous domain (no more patchy coverage of km-scale limited-area models or radar-derived products) and (2) the availability of predictions up to medium-range lead times (instead of just a few hours or a couple of days).

7.3 Adequacy of the impact observations for verifying flash floods

This study demonstrates that enhanced flood report databases, such as the one developed in Ecuador by Krucziewicz et al. (2021b), are instrumental in setting flash-flood-triggering rainfall thresholds and assessing rainfall forecast performance in flash flood prediction. The better and more reliable spatial/temporal coverage of the flood reports in Ecuador's database enabled an in-depth, long-term assessment that would have been unfeasible with publicly available flood reports due to their inevitably poorer spatial/temporal coverage and poorer identification of flood features (e.g., type of flood). For these reasons, flash flood verification in the past was primarily based on case studies as more detailed information was available for single events (Gaume et al., 2009). While a case-study-based verification approach is invaluable to understanding how forecasts predict flash flood events, provided enough observations are available, the results might not hold to other events due to the focused nature of the analysis. Alternatively, taking advantage of the better quality and spatial coverage of rainfall observations, researchers might use them to infer the performance of rainfall forecasts in predicting flash floods. However, as seen in this study, the results of these two verification analyses are different. ecPoint almost always performs better than ENS in predicting extreme (localized) rainfall (Gascón et al., 2023; Hemri et al., 2022; Hewson and Pillou, 2021). However, in the binary prediction of whether an area was affected by a flash food or not (yes- or no-event), the verification results are more nuanced. Not considering these two cases separately would do a disservice to the assessment of how well raw ENS forecasts can predict areas at risk of flash floods. Thus, this study underscores the importance of enhancing flash flood report databases to assess more in detail the performance of rainfall forecasts for flash flood prediction, contributing significantly to a more effective disaster preparedness and risk management strategies.

Although this dataset is the best attempt to build a comprehensive historical record of flash flood events

in Ecuador, event underreporting still needs to be addressed to provide a more comprehensive assessment of forecast performance. In alignment with previous studies attempting to use impact-based observations to estimate forecast performance ([Hitchens et al., 2013](#); [Mitheu et al., 2023](#); [Robbins and Titley, 2018](#)), the verification of the rainfall forecasts against underreported flash flood events can lead to an underestimation of forecast skill and undermine the confidence in the forecasts, causing the dismissal of valuable predictions crucial for preparedness actions. It is, therefore, required to read beyond the actual numbers of verification results, and read them in a critical, although subjective, way. For example, although the FB is far larger than 1 (that would mean the forecasts overestimate substantially the areas at risk of flash floods), the counts of yes-events in the forecast and the observations show there is a good correspondence between wet/dry conditions (meaning the forecasts can identify fairly well which areas are at risk of flash floods) and that the high FB might be due mainly to the low spatial coverage of the reports. Techniques such as assigning 1s (i.e., yes-events) to adjacent grid-boxes to those already containing flood reports could help to reduce significantly the FB. We argue, however, that this method is more appropriate for forecasts provided on higher resolution grids because rainfall might not be predicted at the right location and/or the flash flood event might extend to adjacent grid-boxes. Since forecasts are here provided on a grid at 18 km, if a flash-flood-triggering rainfall event is predicted within a grid-box, it is reasonable to think that there is a good chance to see the flash flood within that grid-box (unless the event happens on the grid-box boundary, but this is an exception whose handling goes beyond the scope of this analysis). It is, therefore, argue here that, when verifying global NWP forecasts, improvements in the observational spatial coverage are the way to reduce the number of “false” false alarms, which for this type of analysis (using non-standard observations) is the biggest and most difficult problem to address ([Marsigli et al., 2021](#)).

8 Conclusions

This study defined a new method to calculate VRT values in data-scarce regions. By using short-range ecPoint rainfall forecasts as proxies for in-situ observations to calculate them, this method addresses significant data gaps that limit the verification and creation of flash flood forecasting systems in data-scarce regions.

The performance of two global rainfall forecasts, ENS and ecPoint, is critically examined in identifying areas at risk of flash floods in Ecuador. It highlighted the nuances in the performance of the two forecasting systems in different meteorological contexts. While ecPoint outperforms ENS when flash-flood-triggering rainfall events are mainly originated from small-scale convective systems, in the case of large-scale systems ENS and ecPoint show comparable performance in the identification of areas at risk of flash floods, with ecPoint enhancing primarily only the accuracy of the magnitude prediction of the flash-flood-triggering rainfall event.

Finally, while the quality of the flash flood report database built for Ecuador is much higher than other public databases, this study has also discussed the impacts in objective flash flood forecast verification of a still large event underreporting. While the enhanced quality of the database used in this analysis allowed the authors to conduct an in-depth, long-term verification analysis, there is still work to do better assess forecast performance to support preparedness and action during decision-making processes.

The authors of this study suggest that future research should be focused in two main areas. Since the definition of flash-flood-triggering rainfall thresholds is linked to the availability of flash flood reports and those are not uniformly available around the globe, it would not be possible to develop rainfall-base flash flood forecasting systems over a continuous global domain. It would be interesting to develop an

ecPoint-compatible gridded rainfall climatology, for example using the newly developed ERA5-ecPoint (Hewson et al., 2023), to create flash-flood-triggering rainfall thresholds over a continuous global domain and support the development of a rainfall-based flash flood forecasting system over the same domain. The authors also suggest that more resources are spent in the development of more flood databases like the one presented in this study as they incorporate invaluable details on the type of flood that can be used to target verification and forecast development efforts. At the same time, research efforts should be focused in addressing the uncertainties around the spatial coverage of this type of observations to provide clearer guidance on forecast performance.

References

- Adler, R. F., Sapiano, M. R., Huffman, G. J., Wang, J. J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R. and Shin, D. B. (2018) The global precipitation climatology project (gpcp) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138.
- Alfieri, L. and Thielen, J. (2015) A european precipitation index for extreme rain-storm and flash flood early warning. *Meteorological Applications*, **22**, 3–13.
- Barrett, A. I., Wellmann, C., Seifert, A., Hoose, C., Vogel, B. and Kunz, M. (2019) One step at a time: How model time step significantly affects convection-permitting simulations. *Journal of Advances in Modeling Earth Systems*, **11**, 641–658.
- Bazo, J., Singh, R., Destrooper, M. and Perez, E. C. D. (2019) Pilot experiences in using seamless forecasts for early action: The “ready-set-go!” approach in the red cross, 387–398.
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Dijk, A. I. V., McVicar, T. R. and Adler, R. F. (2019) Mswev v2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, **100**, 473–500.
- Bendix, J., Rollenbeck, R. and Reudenbach, C. (2006) Diurnal patterns of rainfall in a tropical andean valley of southern ecuador as seen by a vertically pointing k-band doppler radar. *International Journal of Climatology*, **26**, 829–846.
- Bischiniotis, K., van den Hurk, B., de Perez, E. C., Veldkamp, T., Nobre, G. G. and Aerts, J. (2019) Assessing time, cost and quality trade-offs in forecast-based action for floods. *International Journal of Disaster Risk Reduction*, **40**, 101252.
- Borga, M., Stoffel, M., Marchi, L., Marra, F. and Jakob, M. (2014) Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows. *Journal of Hydrology*, **518**, 194–205.
- Bouallègue, Z. B. and Richardson, D. S. (2022) On the roc area of ensemble forecasts for rare events. *Weather and Forecasting*, **37**, 787–796.
- Bucherie, A., Ayala, F. and Krucziewicz, A. (2021) Ecuador historical flood occurrences and impacts dataset with flash flood confidence index (2007-2020). *Zenodo*.
- Bucherie, A., Hultquist, C., Adamo, S., Neely, C., Ayala, F., Bazo, J. and Krucziewicz, A. (2022a) A comparison of social vulnerability indices specific to flooding in ecuador: principal component analysis (pca) and expert knowledge. *International Journal of Disaster Risk Reduction*, **73**, 102897.

- Bucherie, A., Werner, M., Homberg, M. V. D. and Tembo, S. (2022b) Flash flood warnings in context: Combining local knowledge and large-scale hydro-meteorological patterns. *Natural Hazards and Earth System Sciences*, **22**, 461–480.
- Buizza, R. (2019) Introduction to the special issue on “25 years of ensemble forecasting”. *Quarterly Journal of the Royal Meteorological Society*, **145**, 1–11.
- Buytaert, W., Celleri, R., Willem, P., Bièvre, B. D. and Wyseure, G. (2006) Spatial and temporal rainfall variability in mountainous areas: A case study from the south ecuadorian andes. *Journal of Hydrology*, **329**, 413–421.
- Corral, C., Berenguer, M., Sempere-Torres, D., Poletti, L., Silvestro, F. and Rebora, N. (2019) Comparison of two early warning systems for regional flash flood hazard forecasting. *Journal of Hydrology*, **572**, 603–619.
- Davolio, S., Silvestro, F. and Gastaldo, T. (2017) Impact of rainfall assimilation on high-resolution hydrometeorological forecasts over liguria, italy. *Journal of Hydrometeorology*, **18**, 2659–2680.
- De Perez, E. C., Van Den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongman, B., Lopez, A., Mason, S., Mendler De Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J. and Zsoter, E. (2016) Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences*, **20**, 3549–3560.
- Demissie, B., Teklemariam, D., Haile, M., Meaza, H., Nyssen, J., Billi, P., Abera, W., Gebrehiwot, M., Haug, R. and Eetvelde, V. V. (2021) Flood hazard in a semi-closed basin in northern ethiopia: Impact and resilience. *Geo: Geography and Environment*, **8**.
- DiCiccio, T. J. and Efron, B. (1996) Bootstrap confidence intervals. *Statistical Science*, **11**, 189–212.
- Dinis, P. A., Huvi, J., Pinto, M. C. and Carvalho, J. (2021) Disastrous flash floods triggered by moderate to minor rainfall events. recent cases in coastal benguela (angola). *Hydrology 2021, Vol. 8, Page 73*, **8**.
- Dordevic, M., Mutic, P. and Kim, H. (2020) Flash flood guidance system: Response to one of the deadliest hazards. *WMO Bulletin*, **69**, 29–33.
- Douinot, A., Roux, H., Garambois, P. A., Larnier, K., Labat, D. and Dartus, D. (2016) Accounting for rainfall systematic spatial variability in flash flood forecasting. *Journal of Hydrology*, **541**, 359–370.
- Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A. and Cloke, H. L. (2016) Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water*, **3**, 391–418.
- Galarza-Villamar, J. A., Leeuwis, C., Pila-Quinga, G. M., Cecchi, F. and Párraga-Lema, C. M. (2018) Local understanding of disaster risk and livelihood resilience: The case of rice smallholders and floods in ecuador. *International Journal of Disaster Risk Reduction*, **31**, 1107–1120.
- Gascón, E., Montani, A. and Hewson, T. D. (2023) Post-processing output from ensembles with and without parametrised convection, to create accurate, blended, high-fidelity rainfall forecasts.
- Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutoulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D. and Viglione, A. (2009) A compilation of data on european flash floods. *Journal of Hydrology*, **367**, 70–78.

- Georgakakos, K. P., Modrick, T. M., Shamir, E., Campbell, R., Cheng, Z., Jubach, R., Sperfslage, J. A., Spencer, C. R. and Banks, R. (2021) The flash flood guidance system implementation worldwide: A successful multidecadal research-to-operations effort. *Bulletin of the American Meteorological Society*, **-1**, 1–35.
- Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P. E., Clark, R. A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., Hong, Y. and Howard, K. W. (2017) The flash project - improving the tools for flash flood monitoring and prediction across the united states. *Bulletin of the American Meteorological Society*, **98**, 361–372.
- Göber, M., Zsótér, E. and Richardson, D. S. (2008) Could a perfect model ever satisfy a naïve forecaster? on grid box mean versus point verification. *Meteorological Applications*, **15**, 359–365.
- Haiden, T. and Duffy, S. (2016) Use of high-density observations in precipitation verification. *ECMWF Newsletter*, 20–25.
- Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, F. and Richardson, D. (2021) Evaluation of ecmwf forecast, including 2021 upgrade. *ECMWF Technical Memoranda*, 56. URL: <https://www.ecmwf.int/node/20142>.
- Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z. and Prates, F. (2023) Evaluation of ecmwf forecasts, including the 2023 upgrade. *ECMWF Technical Memoranda*, **911**, 1–60.
- Hamill, T. M., Whitaker, J. S. and Mullen, S. L. (2006) Reforecasts: An important dataset for improving weather predictions. *Bulletin of the American Meteorological Society*, **87**, 33–46.
- Hemri, S., Hewson, T., Gascón, E., Rajczak, J., Bhend, J., Spirig, C., Moret, L. and Liniger, M. A. (2022) How do ecpoint precipitation forecasts compare with postprocessed multi-model ensemble predictions over switzerland? *ECMWF Technical Memoranda*, **901**.
- 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., Abellán, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J. N. (2020) The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049.
- Hewson, T. and Pillosu, F. (2021) A low-cost post-processing technique improves weather forecasts around the world. *Communications Earth and Environment*, **2**.
- Hewson, T., Pillosu, F., Gascòn, E. and Vučović, M. (2023) Post-processing era5 output with ecpoint. *ECMWF Newsletter*.
- Hirabayashi, Y., Tanoue, M., Sasaki, O., Zhou, X. and Yamazaki, D. (2021) Global exposure to flooding from the new cmip6 climate model projections. *Scientific Reports*, **11**, 1–7.
- Hitchens, N. M., Brooks, H. E. and Kay, M. P. (2013) Objective limits on forecasting skill of rare events. *Weather and Forecasting*, **28**, 525–534.
- Hurford, A. P., Priest, S. J., Parker, D. J. and Lumbroso, D. M. (2012) The effectiveness of extreme rainfall alerts in predicting surface water flooding in england and wales. *International Journal of Climatology*, **32**, 1768–1774.

- Ibarreche, J., Aquino, R., Edwards, R. M., Rangel, V., Pérez, I., Martínez, M., Castellanos, E., Álvarez, E., Jimenez, S., Rentería, R., Edwards, A. and Álvarez, O. (2020) Flash flood early warning system in colima, mexico. *Sensors (Switzerland)*, **20**, 1–26.
- Ilbay-Yupa, M., Lavado-Casimiro, W., Rau, P., Zubierta, R. and Castillón, F. (2021) Updating regionalization of precipitation in ecuador. *Theoretical and Applied Climatology*, **143**, 1513–1528.
- Imhoff, R. O., Brauer, C. C., van Heeringen, K. J., Uijlenhoet, R. and Weerts, A. H. (2022) Large-sample evaluation of radar rainfall nowcasting for flood early warning. *Water Resources Research*, **58**.
- Javelle, P., Fouchier, C., Arnaud, P. and Lavabre, J. (2010) Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations. *Journal of Hydrology*, **394**, 267–274.
- Javelle, P., Organde, D., Demargne, J., Saint-Martin, C., Saint-Aubin, C. D., Garandéau, L. and Janet, B. (2016) Setting up a french national flash flood warning system for ungauged catchments based on the aiga method. *E3S Web of Conferences*, **7**, 718010–18010.
- Jolliffe, I. T. and Stephenson, D. B. (2011) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley Sons, Ltd, 2nd editio edn.
- Jonkman, S. and Vrijling, J. (2008) Loss of life due to floods. *Journal of Flood Risk Management*, **1**, 43–56.
- Junquas, C., Heredia, M. B., Condom, T., Ruiz-Hernández, J. C., Campozano, L., Dudhia, J., Espinoza, J. C., Menegoz, M., Rabatel, A. and Sicart, J. E. (2022) Regional climate modeling of the diurnal cycle of precipitation and associated atmospheric circulation patterns over an andean glacier region (antisana, ecuador). *Climate Dynamics*, **58**, 3075–3104.
- Kiptum, A., Mwangi, E., Otieno, G., Njogu, A., Kilavi, M., Mwai, Z., MacLeod, D., Neal, J., Hawker, L., O'Shea, T., Saado, H., Visman, E., Majani, B. and Todd, M. C. (2023) Advancing operational flood forecasting, early warning and risk management with new emerging science: Gaps, opportunities and barriers in kenya. *Journal of Flood Risk Management*, e12884.
- Kruczkiewicz, A., Bucherie, A., Ayala, F. and Bazo, J. (2021a) Forecast-based financing for flash floods: a flash flood confidence index to improve flood reporting. *Anticipation Hub*. URL: <https://www.anticipation-hub.org/news/forecast-based-financing-for-flash-floods-a-flash-flood-confidence-index>.
- Kruczkiewicz, A., Bucherie, A., Ayala, F., Hultquist, C., Vergara, H., Mason, S., Bazo, J. and de Sherbinin, A. (2021b) Development of a flash flood confidence index from disaster reports and geophysical susceptibility. *Remote Sensing*, **13**, 2764.
- Laraque, A., Bernal, C., Bourrel, L., Darrozes, J., Christophoul, F., Armijos, E., Fraizy, P., Pombosa, R. and Guyot, J. L. (2009) Sediment budget of the napo river, amazon basin, ecuador and peru. *Hydrological Processes*, **23**, 3509–3524.
- Lavers, D. A., Harrigan, S. and Prudhomme, C. (2021) Precipitation biases in the ecmwf integrated forecasting system. *Journal of Hydrometeorology*.
- Liu, C., Guo, L., Ye, L., Zhang, S., Zhao, Y. and Song, T. (2018) A review of advances in china's flash flood early-warning system. *Natural Hazards*, **92**, 619–634.
- Lowrie, C., Kruczkiewicz, A., McClain, S. N., Nielsen, M. and Mason, S. J. (2022) Evaluating the usefulness of vgi from waze for the reporting of flash floods. *Scientific Reports 2022 12:1*, **12**, 1–13.

- Luong, T. T., Pöschmann, J., Kronenberg, R. and Bernhofer, C. (2021) Rainfall threshold for flash flood warning based on model output of soil moisture: Case study wernersbach, germany. *Water (Switzerland)*, **13**, 1–15.
- Ma, M., Wang, H., Yang, Y., Zhao, G., Tang, G., Hong, Z., Clark, R. A., Chen, Y., Xu, H. and Hong, Y. (2021) Development of a new rainfall-triggering index of flash flood warning-case study in yunnan province, china. *Journal of Flood Risk Management*, **14**, 1–9.
- Marsigli, C., Ebert, E., Ashrit, R., Casati, B., Chen, J., Coelho, C. A. S., Dorninger, M., Gilleland, E., Haiden, T., Landman, S. and Mittermaier, M. (2021) Review article: Observations for high-impact weather and their use in verification. *Natural Hazards and Earth System Sciences*, **21**, 1297–1312.
- Mitheu, F., Tarnavsky, E., Ficchi, A., Stephens, E., Cornforth, R. and Petty, C. (2023) The utility of impact data in flood forecast verification for anticipatory actions: Case studies from uganda and kenya. *Journal of Flood Risk Management*, e12911.
- Norbiato, D., Borga, M., Esposti, S. D., Gaume, E. and Anquetin, S. (2008) Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. *Journal of Hydrology*, **362**, 274–290.
- Owens, R. and Hewson, T. (2018) Ecmwf forecast user guide. URL: <https://www.ecmwf.int/node/16559>.
- Papagiannaki, K., Lagouvardos, K., Kotroni, V. and Bezes, A. (2015) Flash flood occurrence and relation to the rainfall hazard in a highly urbanized area. *Natural Hazards and Earth System Sciences*, **15**, 1859–1871.
- de Perez, E. C., Harrison, L., Berse, K., Easton-Calabria, E., Marunye, J., Marake, M., Murshed, S. B., Shampa and Zuisomue, E. H. (2022) Adapting to climate change through anticipatory action: The potential use of weather-based early warnings. *Weather and Climate Extremes*, **38**, 100508.
- Pinos, J. and Quesada-Román, A. (2022) Flood risk-related research trends in latin america and the caribbean. *Water (Switzerland)*, **14**, 10.
- Pinos, J. and Timbe, L. (2020) Mountain riverine floods in ecuador: Issues, challenges, and opportunities. *Frontiers in Water*, **2**, 36.
- Ramos Filho, G. M., Rabelo Coelho, V. H., da Silva Freitas, E., Xuan, Y. and das Neves Almeida, C. (2021) An improved rainfall-threshold approach for robust prediction and warning of flood and flash flood hazards. *Natural Hazards*, **105**, 2409–2429.
- Raynaud, D., Thielen, J., Salamon, P., Burek, P., Anquetin, S. and Alfieri, L. (2015) A dynamic runoff co-efficient to improve flash flood early warning in europe: Evaluation on the 2013 central european floods in germany. *Meteorological Applications*, **22**, 410–418.
- Recalde-Coronel, C. G., Barnston, A. G. and Ángel G. Muñoz (2014) Predictability of december-april rainfall in coastal and andean ecuador. *Journal of Applied Meteorology and Climatology*, **53**, 1471–1493.
- Recalde-Coronel, G. C., Zaitchik, B. and Pan, W. K. (2020) Madden–julian oscillation influence on sub-seasonal rainfall variability on the west of south america. *Climate Dynamics*, **54**, 2167–2185.
- Robbins, J. C. and Titley, H. A. (2018) Evaluating high-impact precipitation forecasts from the met office global hazard map (ghm) using a global impact database. *Meteorological Applications*, **25**, 548–560.

- Shuvo, S. D., Rashid, T., Panda, S. K., Das, S. and Quadir, D. A. (2021) Forecasting of pre-monsoon flash flood events in the northeastern bangladesh using coupled hydrometeorological nwp modelling system. *Meteorology and Atmospheric Physics*, 1–23.
- Song, H. J., Lim, B. and Joo, S. (2019) Evaluation of rainfall forecasts with heavy rain types in the high-resolution unified model over south korea. *Weather and Forecasting*, **34**, 1277–1293.
- Speight, L., Cole, S. J., Moore, R. J., Pierce, C., Wright, B., Golding, B., Cranston, M., Tavendale, A., Dhondia, J. and Ghimire, S. (2018) Developing surface water flood forecasting capabilities in scotland: an operational pilot for the 2014 commonwealth games in glasgow. *Journal of Flood Risk Management*, **11**, S884–S901.
- Speight, L. J., Cranston, M. D., White, C. J. and Kelly, L. (2021) Operational and emerging capabilities for surface water flood forecasting. *Wiley Interdisciplinary Reviews: Water*, **8**.
- Tapiador, F. J., Roca, R., Genio, A. D., Dewitt, B., Petersen, W. and Zhang, F. (2019) Is precipitation a good metric for model performance? *Bulletin of the American Meteorological Society*, **100**, 223–233.
- Tobar, V. and Wyseure, G. (2018) Seasonal rainfall patterns classification, relationship to enso and rainfall trends in ecuador. *International Journal of Climatology*, **38**, 1808–1819.
- Trigg, M. A., Wilson, M. D., Bates, P. D., Horritt, M. S., Alsdorf, D. E., Forsberg, B. R. and Vega, M. C. (2009) Amazon flood wave hydraulics. *Journal of Hydrology*, **374**, 92–105.
- Šakić Trogrlić, R., van den Homberg, M., Budimir, M., McQuistan, C., Sneddon, A. and Golding, B. (2022) *Early Warning Systems and Their Role in Disaster Risk Reduction*, 11–46.
- Tsonevsky, I., Doswell, C. A. and Brooks, H. E. (2018) Early warnings of severe convection using the ecmwf extreme forecast index. *Weather and Forecasting*, **33**, 857–871.
- UNDRR (2021) Desinventar project. *Plataforma Desinventar Sendai*. URL: <https://db.desinventar.org/>.
- Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., Bouallègue, Z. B., Bhend, J., Dabernig, M., de Cruz, L., Hieta, L., Mestre, O., Moret, L., Plenković, I. O., Schmeits, M., Taillardat, M., van den Bergh, J., van Schaeybroeck, B., Whan, K. and Ylhaisi, J. (2021) Statistical postprocessing for weather forecasts review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*, **102**, E681–E699.
- Vuille, M., Bradley, R. S. and Keimig, F. (2000) Climate variability in the andes of ecuador and its relation to tropical pacific and atlantic sea surface temperature anomalies. *Journal of Climate*, **13**, 2520–2535.
- Wen, Y., Schuur, T., Vergara, H. and Kuster, C. (2021) Effect of precipitation sampling error on flash flood monitoring and prediction: Anticipating operational rapid-update polarimetric weather radars. *Journal of Hydrometeorology*, **22**, 1913–1929.
- Wheeler, M. C. and Hendon, H. H. (2004) An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. *Monthly Weather Review*, **132**, 1917–1932.
- Xing, Y., Liang, Q., Wang, G., Ming, X. and Xia, X. (2019) City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution. *Natural Hazards*, **96**, 473–496.

- Zanchetta, A. D. and Coulibaly, P. (2020) Recent advances in real-time pluvial flash flood forecasting. *Water (Switzerland)*, **12**, 570.
- Zanchetta, A. D., Coulibaly, P. and Fortin, V. (2022) Forecasting high-flow discharges in a flashy catchment using multiple precipitation estimates as predictors in machine learning models. *Hydrology*, **9**, 216.
- Zeman, C., Wedi, N. P., Dueben, P. D., Ban, N. and Schär, C. (2021) Model intercomparison of cosmo 5.0 and ifs 45r1 at kilometer-scale grid spacing. *Geoscientific Model Development [preprint]*, **2021**, 1–35.
- Zsoter, E., Prudhomme, C., Stephens, E., Pappenberger, F. and Cloke, H. (2020) Using ensemble re-forecasts to generate flood thresholds for improved global flood forecasting. *Journal of Flood Risk Management*, **13**.