|  |
| --- |
| **RESEARCH ARTICLE** |
| words (full document) |
| 7637 words (excluding intro, abstract, tables and figures) |
| 81 min to read the full document |

**Guidance on how to use ecPoint forecasts to predict flash floods in Ecuador**

Fatima M. Pillosu1,2, Calum Baugh2, Agathe Bucherie3, Andrew Kruczkiewicz3,4, Carolynne Hultquist5, Humberto Vergara6,7, Florian Pappenberger2, Elisabeth Stephens1,4,8, Christel Prudhomme2,9,10, Hannah L. Cloke1,8,11,12

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

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

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

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

5 Center for International Earth Science Information Network (CIESIN), Columbia Climate School and its Earth Institute, New York, USA

6 Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), The University of Oklahoma, Norman, OK

7 NOAA National Severe Storms Laboratory (NSSL), Norman, OK

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

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

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

11 Department of Earth Sciences, Air, Water and Landscape Science, Uppsala University, Sweden

12 Centre of Natural Hazards and Disaster Science, CNDS, Sweden

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

**Abstract.**

**Keywords.** Flash flood forecasting, flash flood observations, ecPoint.

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

# Introduction

Flash floods in Ecuador constitute a significant risk due to the hydro-climatological characteristics of the country. On the world scene, Ecuador is not an exception. Globally, flash floods account for ~85% of all types of floods and have the highest mortality rate, with more than 5000 fatalities per year (Dordevic *et al.*, 2020). For example, riverine floods are rarely associated with fatalities in Europe, while flash floods often result in loss of life (Gaume *et al.*, 2009). Flash floods also have severe social, economic, and environmental impacts due to the increased vulnerability of people living and having economic activities in flood-prone areas and cities (Dordevic *et al.*, 2020). Such impacts are exacerbated in developing countries that often have few resources to recover from the impacts of extreme natural hazards (Winsemius *et al.*, 2018). In Latin America, flood impacts are more severe than in developed regions due to exponential, unregulated urbanization of the floodplains, human-triggered catchment degradation, lack of preparedness and resilience for emergency response, the persistence of poverty, inefficient public policies, and infrastructural problems (Pinos & Quesada-Román, 2022). Hence, flash floods in Ecuador can have severe, long-term impacts such as damages to infrastructure and agriculture, interruptions to business and education, disruption of healthcare services, and outbursts of waterborne diseases (Galarza-Villamar *et al.*, 2018). Ecuador’s vulnerability to flash flooding is also worrying in the context of climate change since the frequency of floods is expected to increase in South America in the coming decades (Hirabayashi *et al.*, 2021).

Forecast-triggered flood risk reduction strategies, such as “early warning systems” (EWS, Bouwer *et al.*, 2014) or “forecast-based financing” protocols (FbF, Coughlan De Perez *et al.*, 2015), can increase resilience and reduce mortality. Success is dependent on being able to provide reliable predictions of the location, magnitude, and time of occurrence of floods with sufficiently long lead time to allow risk managers to implement risk reduction measures, e.g. evacuations (UNICEF & WFP, 2015). The value of forecast-triggered flood risk reduction strategies is enhanced in low-income countries, which tend to have more limited budgets than more developed regions to build also other types of flood risk reduction strategies such as hard flood protection infrastructures (Golnaraghi, 2012).

Forecasts must also be issued with sufficiently long lead time to allow risk managers to implement such measures (UNICEF & WFP, 2015), especially in developing countries where longer preparation times are needed (Bazo *et al.*, 2018). Flash flood forecasting systems have been developed at regional (Speight *et al.*, 2018; Corral *et al.*, 2019; Ibarreche *et al.*, 2020; Georgakakos *et al.*, 2021; Ramos Filho *et al.*, 2021; Shuvo *et al.*, 2021), national (Javelle *et al.*, 2016; Liu *et al.*, 2018), and continental scale (Raynaud *et al.*, 2015; Park *et al.*, 2019), but nonetheless flash floods remain one of the most difficult type of flood to predict (Zanchetta & Coulibaly, 2020). Flash floods are rapidly occurring events, typically within minutes or few hours, after a torrential triggering-rainfall event. Their limited predictability is linked to the challenge of predicting extreme localized rainfall accurately (Golding *et al.*, 2016), as well as representing in detail hydrological factors such as topography, soil conditions, and terrain coverage that can modulate the occurrence and severity of flash floods (Xing *et al.*, 2019; Kastridis & Stathis, 2020). In a recent review, Zanchetta & Coulibaly (2020) show that progress has been made in the field of high-resolution rainfall forecasts, but they also highlight that issues related to extending rainfall lead times and reducing computational costs for flash flood forecasting are not fully resolved.

ecPoint is a statistical post-processing technique that transform global, grid-based forecasts into probabilistic point-scale forecasts (Hewson & Pillosu, 2021). For rainfall, Hewson & Pillosu (2021) have shown that ecPoint provides more reliable and skilful forecasts against point verification than ECMWF’s ensemble (ENS) up to day 10 forecasts, in particular for localized extremes.

This study’s research questions are: can ecPoint-Rainfall provide added value in capturing flash floods in Ecuador compared to ENS, in particular in terms of forecast accuracy and lead-time extension? If so, why? ~~The metric given by the area under the ROC curve, and ROC curve themselves was used to answer the first question. The scores were computed for a one-year verification period comparing the time and location of extreme rainfall events with the time and location of flash floods reports for Ecuador (Kruczkiewicz~~ *~~et al.~~*~~, 2021).~~ ecPoint and ENS performance in the prediction of day- and night-time rainfall were analyzed to answer the second question, in view of the fact that Ecuador has strong rainfall diurnal cycles (Kikuchi & Wang, 2008), and such cycles tend to not be well represented in models with parametrized convections (Stephens *et al.*, 2010), including ECMWF ENS (Haiden *et al.*, 2021). .

Depending on the available datasets, computational power, and area to cover, different approaches can be adopted in flash flood forecasting. Some approaches consist in flash flood susceptibility assessments, rainfall observations or forecasts comparison with considered/neglected surface conditions, and flow comparison (Zanchetta & Coulibaly, 2020). Although there is a limited number of existing flash flood forecasting systems properly and publicly documented (**reference**), their comparison under operational constraints suggests that systems relying only on rainfall-exceedance criteria can deliver forecasts, whose quality can be comparable to physically-based systems (**reference**). Since rainfall-based systems also significantly reduce computational times and running costs, they can potentially be run for large domains (i.e., continental, or global), leaving the application of rainfall-runoff and routing models for smaller regions or urban environments (**reference**). Methods for obtaining the required high-resolution rainfall forecasts are the temporal extrapolation of distributed radar observations, the (dynamical or statistical) downscaling of coarser numerical weather prediction (NWP) model outputs, or the integration of both approaches (Zanchetta & Coulibaly, 2020). While radar-derived rainfall fields are still considered the most accurate family of approaches, they also provide very short-range forecasts, typically under 6 hours (**reference**). Dynamically downscaled NWP models (i.e., km-scale, limited area models) can increase forecasts lead time, typically up to 5 days due the exponential increase in the running costs beyond day 2/5 forecasts. While they provide a good representation of the rainfall totals distributions, the prediction of the time and location of the rainfall peaks might not be accurate (**reference**). One of the main advantages of using a statistical downscaling approach is the extremely low computational cost when compared to dynamical downscaling. Vannitsem *et al.* (2021) review a series of statistical post-processing for rainfall forecasts.

# Background: geography, rainfall climatology and flooding in Ecuador

Continental Ecuador is located in north-western South-America within 1.60 °N - 3.50 °S and 81.10°W - 75.28 °W (small box in **Figure 1a**). On its north/north-east side and south/south-east sides it borders, respectively, with Colombia and Peru. It borders with the Pacific Ocean at the west, and it is intersected by the equator. Galápagos islands are located in the Pacific Ocean, 1000 km to the west of the mainland (small box in **Figure 1a**). They are part of Ecuador but are not included in this study. Thus, “continental Ecuador” will be referred only to as “Ecuador” from now on. The Andes cordillera runs from north to south across Ecuador (**Figure 1**). Several authors (Vuille *et al.*, 2000; Recalde-Coronel *et al.*, 2014; Tobar & Wyseure, 2018; Towner *et al.*, 2021) agree that the Andes split the country in three main geographical regions. The first region is called "La Costa", and includes the coastal plains next to the Pacific Ocean (with hills that, on average, do not exceed 300 m above sea level) and the western slopes of the Andes cordillera. The second region is called "El Oriente". It is a plateau that contains 2% of the whole Amazon basin and the eastern slopes of the Andes cordillera. The third region is called "La Sierra", which contains the inter-Andean region between the western and the eastern slopes of the Andean cordillera. While floods are common throughout many regions of Ecuador, climate and hydrology vary dramatically within regions.

The rainy season in "La Costa" spans from December to May. Several studies (Recalde-Coronel *et al.*, 2014; Tobar & Wyseure, 2018) have shown that the extreme phases of El Niño Southern Oscillation (ENSO), known as El Niño (i.e., above-average SST in the Pacific Ocean) and La Niña (i.e., below-average SST), strongly modulate precipitation and temperature in "La Costa". “La Costa” would experience higher (lower) than average rainfall events between February and April during El Niño (La Niña) events, strengthening the normal rainfall seasonality in the region. The western Andean slopes is the only Andean region that shows a similar behaviour to the coastal lowlands with respect to ENSO (Vuille *et al.*, 2000). The close proximity of this region to the Pacific Ocean and the exposure of its slopes to air masses penetrating from the ocean may explain the increased precipitation associated with El Niño, which would otherwise be limited to the lower coastal area. However, the relationship with ENSO in the western Andean slopes is weak and does not hold true during all ENSO events. This behaviour has been confirmed in more recent studies for Ecuador (Tobar & Wyseure, 2018) and it is similar to the behaviour of coastal areas in Northern Peru (Goldberg *et al.*, 1987). “La Sierra” region has a complex spatial precipitation pattern. Vuille *et al.* (2000) found that rainfall in the western Andean region are influenced by air masses originating in the Pacific Ocean. One of the most significant results from Vuille *et al.* (2000) is the observation of a strong relationship between precipitation and the ENSO phases. El Niño events are associated with below-average precipitation since an anomalous Hadley cell subdues and inhibits convection and precipitation in tropical South America mainly between December-February. The opposite is true during La Niña events due to a northward displaced ITCZ, weakened easterly trade winds over the Caribbean, and accelerated south-westerly cross-equatorial flow in the eastern Pacific. Vuille *et al.*, (2000) also show that the eastern Andean Cordillera is another region that shows similar negative precipitation anomalies associated with El Niño events. Even though the eastern Andean range is not directly connected to a Pacific forcing, it is affected by it through atmospheric circulation anomalies over the interior of the continent. However, during most of the year, the eastern Andean are more closely related to SST anomalies in the tropical Atlantic Ocean. Increased precipitation is associated with a dipole-like correlation structure in the tropical Atlantic featuring warmer than normal waters to the south of the ITCZ, while below average SSTs occur to the north. This pattern affects precipitation variability in the eastern Andes from March to November. On the other hand, a strengthened South Atlantic trade flow, cold SSTA in the South Atlantic, and an early withdrawal of the ITCZ toward the warm SSTA over the tropical North Atlantic are associated with below-average precipitation in this region. Therefore, the influence of this tropical Atlantic SSTA mode and coupled ITCZ displacements upon precipitation anomalies extends even as far west as the eastern Andes of Ecuador. The Pacific influence on this mode, however, cannot be completely ruled out since El Niño events trigger tropical North Atlantic warm events and thereby could lead to a reduction in Andean precipitation. The inter-Andean valleys experience a varying influence form oceanic and continental air masses with two main rainy seasons (February to May and October-November). As air masses lose much of their humidity on both flanks of the Andes, precipitation amounts in the inter-Andean region are rather low, varying between 800 and 1500 mm/year (Vuille *et al.*, 2000). It rains throughout the year in "El Oriente" with the wettest (driest) months being April-July (September-October). The rainfall climatology in “El Oriente” is primarily influenced by the strong convective activity across the Amazon Forest and the water vapour variations from the SST of the tropical Atlantic Ocean. The association between rainfall and ENSO in "El Oriente" is much less well understood owing to its remoteness and sparse population (Laraque *et al.*, 2007), and as a consequence, evidence is much conflicting between studies (Towner *et al.*, 2020). For example, (Vuille *et al.*, 2000) acknowledges a deficit in rainfall during El Niño events while no significant effects was identified by (Tobar & Wyseure, 2018), who conclude that the Andes cordillera acts as an eastward barrier for the impacts of ENSO in "La Sierra" and in "El Oriente" regions.

Flooding in Ecuador is also complex. Torrential rains, high river runoff and flooding are experienced during El Nino events. The flooding in “La Costa” during El Niño events can cause considerable material loss and deaths, as prolonged rainfall events (from 1 to multiple days of continuous rainfall) can make rivers experience overbank flows that consequently cause flooding of the alluvial plain and can cause severe surface runoff in areas far from rivers due to the saturation of the ground. The rainfall-runoff relationship in “El Oriente”, as in other parts of the Amazonian basin, displays a large lag between rainfall peaks and river discharge peaks. River flows show instead a much stronger response to seasonal rainfall patterns as opposed to single rainfall events (Trigg *et al.*, 2009). Trigg *et al.* (2009) found that this lag is related to the size and length of Amazonian rivers, flood plain storage and relative interactions, and rivers generally having a shallow bed and topographical slopes, with relative slow moving waters. Instead, rivers located in the Andean region are highly sensitive to extreme localized rainfall events, and consequently are prone to flash flooding (Laraque *et al.*, 2009). In the Amazon basin, the influence of climate variables on flood risk remains understudied (Towner et al., 2020) as a result of the nonlinear relationship between precipitation and streamflow (Stephens et al., 2015).

# Data

## Flash flood reports for Ecuador

## Rainfall forecasts: ECMWF ensemble and ecPoint

The ECMWF ENSemble prediction system (ENS) consists of one control run and 50 perturbed members. The control run is started from the best possible representation of unperturbed initial conditions. The 50 members are instead started from perturbed initial conditions and model uncertainties. Initial uncertainties are simulated using singular vectors and ensemble of data assimilations, while a stochastic representation simulates model uncertainties (Buizza, 2019). Up to medium-range forecasts (i.e., forecasts 15 days ahead), ENS native resolution is provided by an octahedral reduced-Gaussian grid with 640 latitude lines between the pole and the equator (O640), which corresponds to ~18 km spatial resolution at the equator (Owens & Hewson, 2018).

ecPoint (Hewson & Pillosu, 2021) is a statistical post-processing technique that helps to address the two main factors that affect the performance of global NWP models forecasting extreme localized rainfall: systematic biases (Lavers *et al.*, 2021) and lack of information on forecast sub-grid variability (Göber *et al.*, 2008). Systematic biases are model errors that lead to forecasts under-/overestimations at grid scale. The lack of information on sub-grid variability is instead a characteristic of NWP models as they forecast only grid-box averages. Statistical post-processing can address this deficiencies (Buizza, 2018). Thus, a plethora of statistical post-processing techniques have been developed over the last 50 years (Vannitsem *et al.*, 2021). However, the operational creation of statistically post-processed forecasts remains problematic to this day, especially for global NWP models and extreme localized rainfall events (Table 1 in (Hewson & Pillosu, 2021) presents a list of thirteen challenges faced by well-established and state-of-the-art statistical post-processing techniques). The ecPoint post-processing technique has addressed those challenges by applying a “global remote calibration” approach, which considers that sub-grid variability and systematic biases are weather-dependent, and such dependency is generally not location-related. Thus, observations from everywhere in the world are gathered to define a set of grid-box weather types (G\_WT, **Figure 3a**) that are used to convert each single raw grid-box forecast into a distribution of equally probable bias-corrected point-scale realizations (**Figure 3b**). Hewson & Pillosu (2021) have shown with a one-year period global verification that ecPoint-Rainfall (i.e., the branch of ecPoint-family products that post-processes rainfall forecasts) can produce more reliable and skilful rainfall forecasts than the raw ENS for point verification, especially in case of extreme events. Currently, ecPoint-Rainfall generates 100 new ensemble members for each raw ENS member. This produces a total of 5100 post-processed members, which are finally distilled in percentiles from 1st to 99th.

# Methods

## Quantification of the four quadrants of the contingency table

The first challenge faced in this study was to develop a verification methodology that quantifies the four quadrants of the contingency table despite having report-based flood observations. When verifying the performance of a forecasting system in the prediction of a parameter (e.g., rainfall, temperature, wind speed, river discharge), observations from stationary instruments (e.g., rain gauges, thermometers, anemometers, or discharge gauges) are typically used to compare the forecast values at the nearest grid-box. Verification is therefore carried out at the observations' sites. Stationary observational instruments report both events and non-events, allowing one to populate the four quadrants of the contingency table, i.e., H, FA, M, and CN (see **Table 1** for the definition of these terms), and fully assess the performance of the forecasting system under consideration.

When observations are provided as reports, one will typically receive reports only at the events' locations, and there will be no reports for non-events. Therefore, verification is carried out at the single event site. One could assume that non-events are represented in the dataset by the absence of reports at a specific location and time. This would be true if report-based observational datasets were not incomplete in nature. Several studies (Gaume *et al.*, 2009; Robbins & Titley, 2018) explain why these types of reports cannot be considered complete. Several studies have been conducted on this topic for many years up to today, indicating no solution has been found. Several studies (Robbins & Titley, 2018) that verify forecasts performance using report-based observations tend to verify the forecasts only for yes-events with the caveat that only quadrants I (i.e., hits) and III (i.e., misses) of the contingency table can be populated. However, it is argued here that it is equally valuable to provide information on false alarms. Risk-tolerant users are mainly interested in the forecasting system's hit rates (computed from misses and hits), whereas risk-averse users might look at false alarms and misses more closely because if they allocate resources and act based on a forecast, and the event does not happen, such resources would be wasted. Therefore, the methodology to verify forecasts using report-based observations must change to quantify false alarms. The verification must be carried out at every grid-box covering the interest domain, and the observational reports must be re-gridded into the domain's grid. Section 4.1.1 explains how the domain used in this study was developed, and section 4.1.2 illustrates the management of the flood reports.

### Definition of the verification domain

The mask built for this study from the ECMWF ENS and ecPoint grid-box covers all continental Ecuador, and fully covers Ecuador’s borders with Colombia and Peru and the Pacific coast. The mask contains a total of 1090 grid-boxes. The three main regions in Ecuador, “La Costa”, “La Sierra” and “La Selva” were selected on the basis of the topography of Ecuador as represented in the ECMWF ENS. Points below 600 m and before and after the longitude 78.2 °W were considered belonging to “La Costa” and “El Oriente”, respectively. The grid-boxes above 600 m were considered belonging to “La Sierra”. Following this criteria, 321, 470, and 299 grid-boxes belong to “La Costa”, “La Sierra”, and “El Oriente” respectively.

### Flood reports management

One must create observational fields that overlap in space and time with the forecasting fields. Flood observational fields for four 12-hourly accumulation periods each day, i.e., 0-12 UTC, 6-18 UTC, 12-00 (for the following day) UTC, and 18-6 (for the following day) UTC were created. Four overlapping accumulation periods for each day are considered in order to have the best chance to capture the rainfall event that generated the flash flood events. This approach was adopted in this study as the authors did not have a more precise information on when the rainfall that caused the flash flood events happened. Considering this approach means, however, that one flash flood report is always associated to two accumulated observational fields. Therefore, the same flood report can contribute to the hits on one accumulation period and to the misses on another accumulation period. The implications of this experiment design will be shown in the Results section and discussed in the Discussion section.

For each observational field, one will count how many observations are present in each grid-box. 0s and 1s will be allocated to the grid-boxes that, respectively, do not contain any report or contain at least one report. Each forecasting accumulation period will be examined with the correspondent observational field, and false alarms and correct negatives will be assigned to those grid-boxes that did not contain any reports, but the system forecasted or not forecasted, respectively, the event.

Only flood reports with an EFFCI>=1 (i.e., flood reports that very likely to be reporting different types of floods, including flash floods), 6 (i.e., flood reports that likely to be reporting flash floods), and 10 (i.e., flood reports that are very likely to be reporting flash floods) were considered.

## Selection of verification scores

The second challenge faced in this study was to use appropriate scores that will allow to evaluate the performance of rainfall forecasts as proxies to identify areas at risk of flash floods.

The two main attributes of any probabilistic forecast are reliability and discrimination, and together determine the usefulness of a probabilistic forecasting system (Candille *et al.*, 2003). A probabilistic forecasting system is called reliable if provides unbiased estimates of the observed frequencies associated with different forecasts probability values. However, reliability alone is not sufficient for a probabilistic forecast to be useful as it might not provide any forecast information beyond climatology. A useful forecast system should be able to discriminate in advance between situations that lead to a different verifying observed events, being able to distinguish among situations under which an event occurs with lower or higher than climatological frequency values. This ability is called discrimination.

In this study, it does not make sense to examine reliability because we are not examining rainfall observations. We are using binary events, flash flood yes-events and non-events. Therefore, the only attribute that could be examined here is the discrimination ability of the forecasting system to predict which area might be at risk of flash flooding. The relative operating characteristic (ROC) curve and the area under the ROC curve (AURC) are used as a summary measure of the forecast discrimination abilities. In particular, the AURC shows perfect discrimination when equal to 1, and shows no skill for values <=0.5.

The ROC curves and the AURC are created for different rainfall thresholds. (Hamill & Juras, 2006) recommend the use of thresholds expressed in relative terms (e.g., quantiles of a climatology) in order to avoid over-interpretation of the AURC results by mixing the forecast ability to distinguish between wet and dry regions and genuine predictive skill. Six relative thresholds (i.e., 85th, 90th, 95th, 98th, 99th percentiles) have been considered in this study. The description of the creation of a rainfall climatology for rainfall totals associated is described in section 4.3.

Regarding the use of ROC curves for the verification of extreme events, (Bouallegue & Richardson, 2021) suggest that special care is taken in the description of the construction and interpretation of the verification results. For the creation of the ROC curves and the AUC, the following procedure was followed. First, the ROC curves for ecPoint and ENS are created using their maximum available discretization (i.e., each member exceeding the rainfall threshold rather than fixed percentage bins). This ensures the ROC curves are as complete as possible. This means the ROC curves are computed using 99 members for ecPoint and 51 members for ENS. Second, ROC curves are closed by drawing a straight line between the last meaningful point of the ROC curve and the top-right corner. Finally, the AUC is computed using the trapezoidal approximation, which means the area under the ROC curve is estimated considering straight lines between two consecutive points of the plot, and the area is equal to the sum of the areas of the single trapeziums.

Bouallegue & Richardson (2021) advocate that a fair comparison for the underlying discrimination ability of different systems would rely on using the same discretization for the ROC curves for the two competing forecasting systems, otherwise the resulting conclusions could be misleading. However, the present study does not focus on determining the underlying discrimination ability of the two different systems. Instead, it focuses on the actual discrimination abilities given their two different configurations and how to use them to forecast flash floods. For this reason, it is considered that the configuration adopted for the ROC curves is the most appropriate to answer the posed research question.

To test the significance of the differences between ecPoint and ENS statistics, a non-parametric bootstrapping technique with replacement and with 1000 replicates was adopted. The differences in the ecPoint/ENS statistics are calculated using a 95% confidence interval.

The ROC curves and the AURC were created considering both available model runs, the 00 and 12 UTC run, in order to maximize the available data.

## Definition of verifying rainfall events

The third challenge faced in this study was to develop a point-based rainfall climatology to determine the verifying rainfall events.

Examining the rainfall observations near the location of flood event reports is common practice to determine the average extreme rainfall values that are likely to generate flash floods in a given area. Two scenarios may arise. Scenario 1 shows a flash flood generated by a widespread rainfall event with relatively low local variability. In this scenario, most point rainfall totals represent the average rainfall event that caused the flash flood. Scenario 2 shows a flash flood generated by a rainfall event with relatively high local variability, with some small point rainfall values and some localized peaks. In this case, only the most extreme part of the distribution of rainfall values would represent the average rainfall event that caused the flash flood because, physically, the smallest rainfall values would unlikely be the driver of any flood event. The knowledge gained by the analysis of rainfall observations near a given flood event could then be used to define the rainfall events for flash flood verification in such an area.

This methodology requires a dense network of rainfall observations to have a higher chance to capture the location and the time of the highest localized rainfall totals that caused the flash flood event (Haiden and Duffy, 2016). This is even more true when the analysis is carried out for small regions (from city to national scale) or limited analysis periods (from a few days to a few months).

The authors of this study did not have any 12-hourly observations to develop the rainfall thresholds for the flash flood verification in Ecuador. Therefore, day 1 ecPoint-Rainfall forecasts were considered instead. The ecPoint's 99 percentiles at such a short-range forecast can indeed be considered representative of the rainfall sub-grid variability (Hewson and Pillosu, 2021) and be used as a proxy for rainfall observations for the definition of rainfall thresholds. The use of day 1 ecPoint-Rainfall forecasts also provides the advantage of having information on the extreme rainfall that in the observations could be missing if, by chance, the observations did not capture the extreme rainfall that generated the flash flood because there were no rain gauges where the highest rainfall totals occurred (Hewson and Pillosu, 2021). A set of n ecPoint-Rainfall forecast instances was built for each flood report in the database. Table 3 shows which forecasts were used depending on the reporting time of the flood event. Four overlapping accumulation periods were used for each report. Therefore, the distribution of rainfall values associated with each report is constituted by 4 times 99 ecPoint-Rainfall percentiles, which equals 396 instances per report. Only a representative extreme rainfall value, provided as an Xth percentile of such distribution, was considered for each of those rainfall distributions. The 50th, 75th, 85th, 90th, 95th, 98th, and 99th percentiles were considered in this study. Finally, the distribution of extreme rainfall values for all flood reports was created to understand which rainfall values might cause flash floods in Ecuador (Figure 5). The analysis was conducted separately for "La Costa" and "La Sierra" regions, respectively, to capture possible differences in the distribution of rainfall values associated with flash flood events. Similar distributions were provided for ENS (also in Figure 5) to draw a comparison between the two forecasting systems. In this case, the distribution of rainfall values associated with each report was constituted by 4 times 51 ENS ensemble members, which equals 204 instances per report.

The verification rainfall events were chosen based on how many flash flood reports were retained in the analysis. The percentiles of the representative extreme rainfall values distribution depict the number of retained flood reports. This means that if the 25th, 50th or 75th percentiles are considered, the definition of the verification rainfall events is carried out using the top 75%, 50% or 25% of flood reports in the database, respectively. The 25th percentile was considered in this study to ensure that the flash floods events used do not represent only the top extreme flash flood events in Ecuador. Such percentile should also ensure that the value of the verification rainfall events is robustly computed from a not too small number of retained flood reports. This is especially true for the database of flood reports with EFFCI>=10, which contains a small number of flood reports (Table 2). For example, if the 75th percentile had been considered, the verification rainfall events for "La Costa" and "La Sierra" would have been defined only by 4.5 and 8.25 flood reports, respectively, instead of 13.5 and 24.75 if the 25th percentile would have been considered.

**Figure 5** shows the distributions for the extreme rainfall associated with flood reports in “La Costa” and “La Sierra” regions of Ecuador, used to determine the verifying rainfall events. Considering a 75% percent of retained flood reports (i.e., 25th percentile of such distributions), the distribution of the 50th and 75th percentiles provided small rainfall values (see **Table 6**) that are unlikely to generate any flash flood events. Only percentiles which translated in around 10 mm/12h were considered in the analysis. This rainfall threshold was lowered for the “La Sierra” region to 5 mm/12h. Therefore, the flash flood verification was run considering only the verification rainfall events from the 85th, 90th, 95th, 98th and 99th percentiles (highlighted in red in **Table 6**) for the category that retains in the analysis at least 75% of the flood reports present in the database.

## The selection of different EFFCI thresholds

This study verifies different EFFCI thresholds. The application context may lead indeed to selection of different critical levels of flash flood confidence (Kruczkiewicz *et al.*, 2021b). If the results of this study are used for FbF, the tolerance to uncertainty may be lower than other use cases, and as such may need necessitate a lower critical threshold to be selected. Alternatively, if the tolerance to uncertainty is higher for a particular use case, perhaps for applications beyond the humanitarian context, the critical threshold could be raised.

# Results

## Is ecPoint-Rainfall improving upon ENS with regard to flash flood prediction? Areas Under the ROC Curve (AURC)

**Figure 6** shows the AURC for verifying rainfall events with relative thresholds >= 85th and 99th percentile, and for flood reports with EFFCI>=6. The AURC for more verifying rainfall events and EFFCI thresholds are included in the supplemental material. The AURC for both regions ("La Costa" and "La Sierra") and both rainfall forecasts (ENS and ecPoint) show three similar features. First, the AURC gradually decline as the relative threshold for the verifying rainfall event rises. Such decrease is a feature of ROC curves when the value of the verifying event is increased, which does not necessarily imply a decrease in the forecasts’ discrimination ability to predict extreme events (Gneiting & Vogel, 2021). Second, the AURC get noisier as the EFFCI thresholds increase. This is likely due to the progressive decrease in flood reports for higher EFFCI thresholds as shown in **Table 4**. Third, AURC differences between ENS and ecPoint tend to be null or very small in both regions for small verifying rainfall events and gradually increase with the verifying rainfall events, with ecPoint showing higher AURC than ENS. This means that ecPoint adds value when forecasting flash flood events generated by extreme rainfall. Notwithstanding such similarities, four significant differences in the AURC for “La Costa” and “La Sierra” can be observed. First, “La Costa” shows overall higher AURC than “La Sierra”. For verifying rainfall events with relative thresholds >=85th and 99th percentile, the AURC for “La Costa” are between 0.8-0.9 and 0.6-0.85, respectively. In “La Sierra”, AURC values are between 0.65-0.75 and 0.55-0.7. Second, there is almost not reduction in skill with lead time for both, ENS and ecPoint, in “La Costa”, while in “La Sierra” such decrease is more marked for both forecasts. Third, the behaviour of the AURC is overall nosier in “La Sierra” than in “La Costa”. Such noisy behaviour cannot be explained by a lack of flood reports since the two regions have a similar number of reports. Fourth, the differences between the best and the worst AURC values in “La Costa” are bigger than the differences in “La Sierra”. Fifth, while the AURC values diminish with the increasing of the verifying rainfall event, in “La Sierra”, the ENS’s AURC for verifying rainfall events >= 85th percentile is bigger than the ecPoint’s AURC for verifying rainfall events >= 99th percentile.

Let us consider the AURC for verifying rainfall events with relative thresholds >= 85th percentile for “La Costa” (top left panel in **Figure 6**).The highest AURC values, for both ENS and ecPoint, are observed in the accumulation periods between 0000-1200 LST, and 0600-1800 LST, which correspond to the troughs of the rainfall’s diurnal cycle in “La Costa” (**Figure 2**). The end of such accumulation periods are highlighted in the x-axis of **Figure 6** in purple and cyan, respectively. The AURC values for the accumulation periods between 0000-1200 LST stay constant at all lead times, while the AURC values between 0600-1800 LST degrade from day 5. The smallest AURC values, for both ENS and ecPoint, are observed in the accumulation periods between 1200-0000 LST, and 1800-0600 LST, which correspond to the peaks of the rainfall’s diurnal cycle in “La Costa” (**Figure 2**). The end of such accumulation periods are highlighted in the x-axis of **Figure 6** in pink and green, respectively. Let us consider the AURC for verifying rainfall event with relative threshold >= 99th percentile (bottom left panel in **Figure 6**).For ENS, the maximum AURC values are observed only between 0000-1200 LST, while the values between 0600-1800 LST are so low since day 1 to be comparable to the minimum AURC values between 1200-0000 LST and 1800-0600 LST. ecPoint seems to not correct the rainfall totals for the accumulation period between 0000-1200 LST since the AURC values are equal to those for ENS. Instead, ecPoint seems to apply significant corrections to the rainfall forecasts in the other three accumulation periods since the AURC are much higher than those for ENS, with correction peaks of 0.16 for the accumulation period between 0600-1800 LST from day 3, and 0.1 for the accumulation period between 1200-0000 LST from day 1.

Let us now consider the AURC for verifying rainfall events with relative thresholds >= 85th percentile for “La Sierra” (top right panel in **Figure 6**). Up to day 2, the highest AURC values for ENS are observed between 0000-1200 LST, 0600-1800 LST, 1200 – 0000 LST. The end of such accumulation periods are highlighted in the x-axis of **Figure 6**, respectively, in purple, cyan and pink. The first accumulation period corresponds to one of the troughs of the rainfall’s diurnal cycle in “La Sierra”, while the last two correspond to the peaks (**Figure 2**). From day 2, the highest AURC values are mainly observed only between 1200 – 0000 LST. ecPoint shows a minor degradation in the AURC values compared to ENS. In addition, ecPoint has nosier AURC values that do not indicate specific accumulation periods where the skill in identifying areas at flash flood risk is better or worst. Let us consider the AURC for verifying rainfall events with relative thresholds >= 99th percentile (bottom right panel in **Figure 6**). As opposed to the AURC for verifying rainfall events with relative thresholds >= 85th percentile, ecPoint shows in this case overall better AURC than ENS. AURC values for ecPoint range mainly between 0.6 and 0.7 through all lead times, while AURC values for ENS range mainly between 0.5 and 0.65. Up to day 4, ecPoint and ENS shows higher AURC values for accumulation periods between 0600-1800 LST and 1200-0000 LST, which correspond to the peaks of the rainfall’s diurnal cycle in “La Sierra” (**Figure 2**). The end of such accumulation periods are highlighted in the x-axis of **Figure 6**, respectively, in cyan and pink. ecPoint and ENS show instead smaller AURC values mainly for the accumulation period between 18000-0600 LST and 0000-1200 LST, which correspond to the troughs of the rainfall’s diurnal cycle in “La Sierra” (**Figure 2**). From day 4, ENS and ecPoint show an asynchronized pattern instead. While ecPoint continues showing highest AURC values between 1200-0000 LST, ENS shows the lowest AURC values for the same accumulation period up to day 10. The period between 0600-1800 LST deteriorates after day 3 in ecPoint, but it still maintains a relatively high AURC compared to the other two accumulation periods (i.e., 1800-0600 LST and 0000-1200 LST). Therefore, in “La Sierra” short range forecasts show in both ENS and ecPoint high AURC for the peak of rainfall’s diurnal cycle. However, from day 4 ENS skill diminishes while ecPoint maintains the trend through all lead times.

## Where is the added ecPoint’s value coming from?

### ROC curves

In “La Sierra”, the AURC values for ENS at the 85th percentile are similar to the AURC values for ecPoint at the 99th percentile. This means that the ability of ecPoint to discriminate between flash flood events and non-events when considering high rainfall events is similar to the discrimination ability of ENS when considering high rainfall events. **Figure 8** shows the ROC curves from day 1 to day 3 (rows from top to bottom) and for the 95th, 98th, and 99th percentile (columns from left to right). Only the ROC curves for flood reports with an EFFCI>=6 are shown. The ROC curves for flood reports with and EFFCI>=1 and >=10 can be found in the supplemental material. For “La Costa” region, the ecPoint’s and ENS’s ROC curves are overlapping with the exception of the last one or two top forecast percentiles, at all lead times ad all relative rainfall thresholds. Therefore, the ROC curves for ecPoint and ENS belong to the same underlying curve. For “La Sierra” region, the ecPoint’s and ENS’s ROC curves differ much more significantly.

### Distribution of Weather Types (WTs)

Flash floods typically occur during night-time (Maddox *et al.*, 1979).

The differences in the AURC values for the different times of the day can be explained by looking at the distribution of WTs in **Figure 7**. In “La Costa” there is not much variability of WTs. The most observed WTs (green crosses in **Figure 7**) are mainly in the 1???? and 2???? categories. However, the peaks correspond to small rainfall totals, low wind speeds, small CAPE, and medium solar radiation. This type of WT is observed across all four categories of convective precipitation ratio (CPR, defined as the ratio between convective rainfall and total precipitation), with the category 1???? (i.e., CPR<0.25, mainly large-scale rainfall in the ENS model) contributing more often than the other three (2????, 3????, and 4????), and the category 4???? (i.e., CPR>0.75, mainly convective rainfall in the ENS model) The one contributing less often. This is shown by the percentages in red in **Figure 7**. In the category 1???? There is also a big contribution by the WT that represents small rainfall totals, high wind speed, small cape, and high solar radiation. This two categories of WTs tend to correct mainly biases in the ENS model as opposed to correct mainly sub-grid variability. The bias correction is mainly for overestimation of rainfall, with a correction factor of around 0.85 in both cases. In “La Sierra”, there is much more variability of WTs across all CPR categories. This can be seen from a more uniform distribution of the percentages in red in **Figure 7**, with the exception of the accumulation period between 1800 to 0600 LST where the category 1???? holds more than the half of the WTs during wet days. There are still the same than those in “La Costa” region (i.e., those indicated by black and blue crosses). However, in “Sierra”, the crosses are indicated with a different colour (magenta) as they represent mainly large solar radiation. This is probably due to the fact that the cases are in the in the heights of the Andean region, and the optical depth of the climatological aerosol distribution is smaller in the mountains, as used in the computation of the clear sky solar radiation.

The different proportion in the WTs distribution during wet days can be linked to the different types of corrections applied by ecPoint, and shown in **Figure 6**. ecPoint does not apply any corrections to the accumulation periods that correspond to the accumulation periods between 0000 and 1200 LST (the final step of these accumulation periods are indicated in green in **Figure 6**). This accumulation period corresponds to one of the lows in the diurnal cycle in “La Costa”. Instead, ecPoint applies the biggest corrections in the rainfall accumulation periods between 1200 and 0000 LST (the final step of these accumulation periods are indicated in black in **Figure 6**). This accumulation period corresponds to one of the peaks of the diurnal cycle in “La Costa”. Over the accumulation periods between 1800-0600 LST and between 0600-1800 LST ecPoint-Rainfall forecasts also apply significant corrections to the ENS forecasts, tending to improve the AURC scores. The characteristic here to highlight is that when ecPoint is not applying much correction is when the ENS model comes more often with weather scenarios that lie in the 1???? category of WTs. If we examine the WTs in such category that might contribute to have rainfall events > 50 mm / 12h, one can see that the WTs tend to be mainly gaussian, with peaks around the white bar that indicates that the ENS forecasts were representative of the point rainfall observations on the ground.

# Case Studies

# Discussions

## Overall benefits of using ecPoint post-processing technique to develop a flash flood forecasting with continuous global domain

## Development of a flash flood forecasting system with continuous global domain without relying on local knowledge by incorporating hydro-meteorological climatologies

## Improvement of observational datasets to better verify flash flood forecasting systems

# Conclusions

* This study analyses the performance of ecPoint at forecasting flash flood events in Ecuador. ECMWF ENS forecasts were used as a benchmark.
* Both ecPoint and ENS provide good flash flood guidance in Ecuador, especially in “La Costa” where it is not observed any substantial forecast skill degradation up to day 10. In “La Sierra”, both forecasts have less skill; however, ecPoint outperforms ENS for extreme rainfall events, showing its added value where flash floods are generated mainly by deep, localized convective systems.
* However, while ENS shows their best performance for small rainfall totals with high probabilities of occurrence, ecPoint shows its added value for large rainfall totals with low probabilities of occurrence. The last type of rainfall event is more often associated with flash flood events than the former by forecasters, so they would connect better with ecPoint-based flash flood forecasts. Furthermore, such forecasts would typically verified better against point rainfall observations than ECMWF ENS, increasing forecasters’ confidence in the forecasting system outputs.
* While the results shown in this study are valid only for Ecuador, it is argued that ecPoint would add value to forecasts based on raw ECMWF ENS if a flash flood forecasting system with a continuous global domain was developed.
* Further research is however needed to develop such system. This study bases the definition of warning-triggering events on the ability of having detailed flash flood observations. It is difficult, if not impossible, to find in other countries such high quality flash flood databases. The authors of this study recommend the incorporation of a rainfall climatology compatible with ecPoint resolution to define the rainfall warning thresholds allowing the definition of warnings without needing detailed local knowledge. Detailed flash flood observational databases would then be required only for verification.
* The development of a flash flood database as detailed as the one developed for Ecuador remains nonetheless fundamental for the development of flash flood forecasting system. Their use in flash flood verification would increase the confidence in the system, and would encourage end-users to adopt it in programmes such as FbA.

# Tables

**Table 1** – Definition of the four quadrants in a contingency table.

|  |  |  |
| --- | --- | --- |
| **FORECASTS (COLUMNS) /**  **OBSERVATIONS (ROWS)** | **YES** | **NO** |
| **YES** | 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*. |
| **NO** | 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*. |

**Table 2** – Number of all flood reports in the database in 2019 and 2020 (i.e., flood reports with EFFCI>=1). The second column of the table shows the number of reports eliminated from the database because the reports did not contain the location of the report in lat/lon coordinates, or the reports did not contain the date and time in which the flood event happened.

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Total n. of flood reports** | **N. of eliminated flood reports** | **N. of retained flood reports** |
| 2019 | 300 | 3 | 297 |
| 2020 | 190 | 0 | 190 |

**Table 3** – Number of flood reports for 2019 and 2020, in each region, and for EFFCI>=1 (first group), EFFCI>=6 (second group), and EFFCI>=10 (third group). The percentages within parenthesis for the reports with EFFCI>=6 and EFFCI>=10 represent the reduction in percent of the number of flood reports compared to the number of flood reports with EFFCI>=1, which correspond to the total number of flood reports in the correspondent year (see **Table 2**).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **N. of flood reports with EFFCI>=1** | | | | **N. of flood reports with EFFCI>=6** | | | | **N. of flood reports with EFFCI>=10** | | | |
| Region/  Year | Costa | Sierra | Oriente | All | Costa | Sierra | Oriente | All | Costa | Sierra | Oriente | All |
| 2019 | 175 | 114 | 8 | 297 | 92  (-47%) | 101  (-12%) | 0  (-100%) | 193  (-35%) | 18  (-90 %) | 33  (-71 %) | 0  (-100 %) | 51  (-83 %) |
| 2020 | 91 | 88 | 11 | 190 | 48  (-47%) | 77  (-12 %) | 2  (-82%) | 127  (-33%) | 22  (-76 %) | 26  (-70 %) | 0  (-100 %) | 48  (-75 %) |

**Table 4** – Accumulation periods for day 1 ecPoint and ENS forecasts (i.e., from t+0, up to t+30) considered for each flood report in order to define the values of the verification rainfall events.

|  |  |  |
| --- | --- | --- |
| **Interval periods (in UTC time) for reporting times of floods on day X** | **12-hourly forecast accumulation periods (in UTC time) containing the interval periods of flood event reporting times on day X** | **Forecast accumulation periods** |
| From 0 to 6 UTC | From 18 (on day X-1) to 6 UTC | Day(X-1), 00 UTC (t+18,t+30) |
| Day(X-1), 12 UTC (t+6,t+18) |
| From 0 to 12 UTC | Day(X), 00 UTC (t+0,t+12) |
| Day(X-1), 12 UTC (t+12,t+24) |
| From 6 to 12 UTC | From 0 to 12 UTC | Day(X), 00 UTC (t+0,t+12) |
| Day(X-1), 12 UTC (t+12,t+24) |
| From 6 to 18 UTC | Day(X), 00 UTC (t+6,t+18) |
| Day(X-1), 12 UTC (t+18,t+30) |
| From 12 to 18 UTC | From 6 to 18 UTC | Day(X), 00 UTC (t+6,t+18) |
| Day(X-1), 12 UTC (t+18,t+30) |
| From 12 to 0 (on day X+1) UTC | Day(X), 00 UTC (t+12,t+24) |
| Day(X), 12 UTC (t+0,t+12) |
| From 18 to 0 (on day X+1) UTC | From 12 to 0 (on day X+1) UTC | Day(X), 00 UTC (t+12,t+24) |
| Day(X), 12 UTC (t+0,t+12) |
| From 18 to 6 (on day X+1) UTC | Day(X), 00 UTC (t+18,t+30) |
| Day(X), 12 UTC (t+6,t+18) |

**Table 5** – Verifying rainfall events (in mm/12h) provided by the 85th and 99th percentile of the distribution of day 1 ecPoint-Rainfall forecasts associated with 2019 flash flood events. 75% of flood report retention is considered. Verifying rainfall events are provided for “La Costa” and “La Sierra” regions, and for EFFCI>=6.

|  |  |  |
| --- | --- | --- |
|  | **85th percentile** | **99th percentile** |
| **La Costa** | 9.865 mm/12h | 50.452 mm/12h |
| **La Sierra** | 5.885 mm/12h | 25.551 mm/12h |

**Table 6** – Relationship between the four overlapping 12-hourly accumulation periods in a day and the distribution of WTs.

|  |  |  |  |
| --- | --- | --- | --- |
| 12-hourly accumulation period | “La Costa” | “La Sierra” | Example of WTs |
| 0000 - 1200 UTC  (1800 – 0600 LST) |  |  |  |
| 0600 – 1800 UTC  (0000 – 1200 LST) |  |  |  |
| 1200 – 2400 UTC  (0600 – 1800 LST) |  |  |  |
| 1800 – 0600 UTC  (1200 – 2400 LST) |  |  |  |

# Figures

Map

Description automatically generated

**Figure 1** - Panel (a) shows Ecuador’s orography, its political regions, and the location of Ecuador’s three main geographical regions: the coast (“La Costa”), the highlands (“La Sierra”), and the Amazon (“EL Oriente”). The small box shows Ecuador’s location (in red) within South America. Panel (b) shows the definition of the Ecuador’s three main geographical regions using the ENS and ecPoint grid (“La Costa”, “La Sierra” and “El Oriente” are represented in yellow, brown, and green, respectively) and the location of the ENS and ecPoint grid-boxes (black dots) within Ecuador’s domain. Panel (c) shows the population density (in people/km2) for each region from the 2020 census (source: <https://es.wikipedia.org/wiki/Provincias_de_Ecuador>). Panel (d) lists the names of Ecuador’s political regions following the numbers indicated in panel (c).

Chart, line chart

Description automatically generated

**Figure 2** – ENS (red lines) and ecPoint (blue lines) diurnal cycle for 12-hourly rainfall annual mean in “La Costa” (solid lines) and “La Sierra” (dashed lines). Only forecasts for the 00 UTC run and up to day 2 lead time are shown. The x-axis labels in purple and brown indicate, respectively, the 12-hourly accumulation periods in UTC and local time.

Diagram

Description automatically generated

**Figure 3** - Panels (a) shows the workflow for the calibration of rainfall forecasts based on the ecPoint technique. Panel (b) shows the workflow for the forecasts generation.

Map

Description automatically generated

**Figure 4** – Panels (a), (b), and (c) show the location of point flood reports with an EFFCI>=1, EFFCI>=6, and EFFCI>=10, respectively, for events occurred in 2019. Panels (d), (e), and (f) show the same but for point flood reports of events occurred in 2020.

Chart, diagram

Description automatically generated

**Figure 5** – Distribution of the extreme rainfall values associated with flood reports in 2019. The first and second column show the rainfall distributions for “La Costa” and “La Sierra”, respectively. The first, second and third row show the rainfall distributions for flood reports with an EFFCI>=1, >=6 and >=10, respectively. How “extreme” the rainfall values are is defined considering an Xth of the distribution of forecasts instances associated to each flood report. The 50th (in yellow), 75th (in green), 85th (in purple), 90th (in cyan), 95th (in blue), 98th (in magenta), and 99th (in orange) percentiles define the most extreme rainfall values for each flood report. The continuous and dashed lines correspond to distributions for ENS and ecPoint, respectively. The 25th, 50th and 75th percentiles of such distributions (marked with horizontal grey lines indicate the percentage of retained flood reports (FRs) that would be considered in the definition of the verification rainfall events.

**Chart, histogram

Description automatically generated**

**Figure 6** – Areas under the ROC curve (AURC) up to day 10 (t+246), for flood reports with EFFCI>=6 in “La Costa” (first column) and in “La Sierra” (second column). AURCs for rainfall events greater than the relative threshold of 85th (first row) and 99th percentile (second row) are shown. Red and blue lines represent the AURC for ENS and ecPoint, respectively. The lead times of the forecasts are indicated in hours, as the steps at the end of the accumulation period. The equivalent lead time in days is indicated below.

Graphical user interface, application, table, Excel

Description automatically generated

**Figure 7** – Distribution of WTs for 2020 for “La Costa” (left column) and “La Sierra” (right column) for all cases in which ENS forecasted rainfall greater than zero. The first, second, third, and fourth row represent, respectively, the WTs distribution for the accumulation period from 0000 to 1200 UTC (1800 to 0600 LST), 0600 to 1800 UTC (0000 to 1200 LST), 1200 to 0000 UTC (0600 to 1800 LST), and 1800 to 0600 UTC (1200 to 0000 LST). WT codes 1???? and 2???? correspond to those cases in which ENS forecasted rainfall with a convective precipitation ratio (CPR) < 0.25 and between 0.25 and 0.5 (i.e., mainly large-scale rainfall). WT codes 3???? and 4???? correspond to those cases in which ENS forecasted rainfall with a convective precipitation ratio (CPR) between 0.5 and 0.75, and >=0.75 (i.e., mainly convective rainfall). The percentages in grey indicate the frequency of WTs within each CPR category. The peaks in the WTs distribution marked by green crosses correspond to WTs for small rainfall totals, light wind speeds, small CAPE, and medium solar radiation across all ranges of CPR. The peaks marked by fuchsia crosses correspond to the same type of WTs but with high solar radiation across all ranges of CPR. The peaks marked by cyan crosses correspond to WTs for mainly large-scale rainfall, small rainfall totals, strong wind speeds, small CAPE, and high solar radiation.

# References

Bazo J., Singh R., Destrooper M., De Perez EC. 2018. Pilot experiences in using seamless forecasts for early action: The “ready-set-go!" approach in the red cross. In: *Sub-seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*. 387–398.

Bouwer LM., Papyrakis E., Poussin J., Pfurtscheller C., Thieken AH. 2014. The Costing of Measures for Natural Hazard Mitigation in Europe. *Nat. Hazards Rev.* **15**:04014010.

Corral C., Berenguer M., Sempere-Torres D., Poletti L., Silvestro F., Rebora N. 2019. Comparison of two early warning systems for regional flash flood hazard forecasting. *J. Hydrol.* **572**:603–619.

Coughlan De Perez E., Van Den Hurk B., Van Aalst MK., Jongman B., Klose T., Suarez P. 2015. Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Nat. Hazards Earth Syst. Sci.* **15**:895–904.

Dordevic M., Mutic P., Kim H. 2020. Flash Flood Guidance System: Response to one of the deadliest hazards. *WMO Bull.* **69**:29–33.

Galarza-Villamar JA., Leeuwis C., Pila-Quinga GM., Cecchi F., Párraga-Lema CM. 2018. Local understanding of disaster risk and livelihood resilience: The case of rice smallholders and floods in Ecuador. *Int. J. Disaster Risk Reduct.* **31**:1107–1120.

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., Koutroulis A., Marchi L., Matreata S., Medina V., Preciso E., Sempere-Torres D., Stancalie G., Szolgay J., Tsanis I., Velasco D., Viglione A. 2009. A compilation of data on European flash floods. *J. Hydrol.* **367**:70–78.

Georgakakos KP., Modrick TM., Shamir E., Campbell R., Cheng Z., Jubach R., Sperfslage JA., Spencer CR., Banks R. 2021. The Flash Flood Guidance System Implementation Worldwide: A Successful Multidecadal Research-To-Operations Effort. *Bull. Am. Meteorol. Soc.* **1**:1–35.

Golding B., Roberts N., Leoncini G., Mylne K., Swinbank R. 2016. MOGREPS-UK convection-permitting ensemble products for surface water flood forecasting: Rationale and first results. *J. Hydrometeorol.* **17**:1383–1406.

Golnaraghi M. 2012. *Institutional partnerships in multi-hazard early warning systems: a compilation of seven national good practices and guiding principles*. Springer.

Haiden T., Jannousek M., Vitart F., Ben-Bouallegue Z., Ferranti L., Prates C., Richardson D. 2021. Evaluation of ECMWF forecasts, including the 2020 upgrade. *ECMWF Tech. Memo.* **880**.

Hewson TD., Pillosu FM. 2021. A new low-cost technique improves weather forecasts across the world. *Commun. Earth Environ.* **2**:132.

Hirabayashi Y., Tanoue M., Sasaki O., Zhou X., Yamazaki D. 2021. Global exposure to flooding from the new CMIP6 climate model projections. *Sci. Rep.* **11**:1–7.

Ibarreche J., Aquino R., Edwards RM., Rangel V., Pérez I., Martínez M., Castellanos E., Álvarez E., Jimenez S., Rentería R., Edwards A., Álvarez O. 2020. Flash flood early warning system in Colima, Mexico. *Sensors (Switzerland)* **20**:1–26.

Javelle P., Organde D., Demargne J., Saint-Martin C., De Saint-Aubin C., Garandeau L., Janet B. 2016. Setting up a French national flash flood warning system for ungauged catchments based on the AIGA method. In: *E3S Web of Conferences*. 718010–18010.

Kastridis A., Stathis D. 2020. Evaluation of hydrological and hydraulic models applied in typical mediterranean ungauged watersheds using post-flash-flood measurements. *Hydrology* **7**:12.

Kikuchi K., Wang B. 2008. Diurnal precipitation regimes in the global tropics. *J. Clim.* **21**:2680–2696.

Kruczkiewicz A., Bucherie A., Ayala F., Hultquist C., Vergara H., Mason S., Bazo J., de Sherbinin A. 2021. Development of a flash flood confidence index from disaster reports and geophysical susceptibility. *Remote Sens.* **13**:2764.

Liu C., Guo L., Ye L., Zhang S., Zhao Y., Song T. 2018. A review of advances in China’s flash flood early-warning system. *Nat. Hazards* **92**:619–634.

Park S., Berenguer M., Sempere-Torres D. 2019. Long-term analysis of gauge-adjusted radar rainfall accumulations at European scale. *J. Hydrol.* **573**:768–777.

Pinos J., Quesada-Román A. 2022. Flood risk-related research trends in Latin America and the Caribbean. *Water (Switzerland)* **14**:10.

Ramos Filho GM., Coelho VHR., Freitas E da S., Xuan Y., Almeida C das N. 2021. An improved rainfall-threshold approach for robust prediction and warning of flood and flash flood hazards. *Nat. Hazards* **105**:2409–2429.

Raynaud D., Thielen J., Salamon P., Burek P., Anquetin S., 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. *Meteorol. Appl.* **22**:410–418.

Shuvo SD., Rashid T., Panda SK., Das S., Quadir DA. 2021. Forecasting of pre-monsoon flash flood events in the northeastern Bangladesh using coupled hydrometeorological NWP modelling system. *Meteorol. Atmos. Phys.*:1–23.

Speight L., Cole SJ., Moore RJ., Pierce C., Wright B., Golding B., Cranston M., Tavendale A., Dhondia J., Ghimire S. 2018. Developing surface water flood forecasting capabilities in Scotland: an operational pilot for the 2014 Commonwealth Games in Glasgow. *J. Flood Risk Manag.* **11**:S884–S901.

Stephens GL., L’Ecuyer T., Forbes R., Gettlemen A., Golaz JC., Bodas-Salcedo A., Suzuki K., Gabriel P., Haynes J. 2010. Dreary state of precipitation in global models. *J. Geophys. Res. Atmos.* **115**:24211.

UNICEF., WFP. 2015. UNICEF/WFP Return on Investment for Emergency Preparedness Study. :34.

UNISDR. 2002. Guidelines for Reducing Flood Losses. *United Nations Int. Strateg. Disaster Reduct.*:79.

Vannitsem S., Bremnes JB., Demaeyer J., Evans GR., Flowerdew J., Hemri S., Lerch S., Roberts N., Theis S., Atencia A., Bouallègue Z Ben., Bhend J., Dabernig M., de Cruz L., Hieta L., Mestre O., Moret L., Plenković IO., Schmeits M., Taillardat M., van den Bergh J., van Schaeybroeck B., Whan K., Ylhaisi J. 2021. Statistical postprocessing for weather forecasts review, challenges, and avenues in a big data world. *Bull. Am. Meteorol. Soc.* **102**:E681–E699.

Winsemius HC., Jongman B., Veldkamp TIE., Hallegatte S., Bangalore M., Ward PJ. 2018. Disaster risk, climate change, and poverty: Assessing the global exposure of poor people to floods and droughts. *Environ. Dev. Econ.* **23**:328–348.

Xing Y., Liang Q., Wang G., Ming X., Xia X. 2019. City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution. *Nat. Hazards* **96**:473–496.

Zanchetta ADL., Coulibaly P. 2020. Recent advances in real-time pluvial flash flood forecasting. *Water (Switzerland)* **12**:570.