**RESEARCH ARTICLE**

Guidance on how to use ecPoint-Rainfall 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.**

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

# Introduction

# Background: geography, rainfall climatology and flooding in Ecuador

Within 1.60 °N - 3.50 °S and 81.10°W - 75.28 °W, continental Ecuador is located in north-western South-America and intersected by the equator (small box in Figure 1a), with Peru to the south/south-east of the country, Colombia to the north/north-east and the Pacific Ocean to the west. Galapagos islands in the Pacific Ocean, 1000 km to the west of the mainland (small box in Figure 1a), are part of Ecuador but will not be included in this study. Thus, “continental Ecuador” will be referred only to as “Ecuador” from now on. The Andes cordillera, that runs from north to south across Ecuador, splits the country in three main geographical regions (Recalde-Coronel, Barnston and Muñoz, 2014, see Figure 1a): "La Costa", which includes the coastal plains next to the Pacific Ocean and the western slopes of the Andes cordillera; "El Oriente", which includes the Ecuadorian region of the Amazon basin and the eastern slopes of the Andes cordillera; and "La Sierra", the inter-Andean region between the western slopes of the Andean cordillera in "La Costa" region and the eastern slopes in the "El Oriente" region.

The rainy season in "La Costa" spans from December to May. Several studies (Recalde-Coronel, Barnston and Muñoz, 2014; Tobar and 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" and well established for a long time. It experiences 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 region show a similar behaviour to the coastal lowlands with respect to ENSO during its peak phase in December-February (Vuille, Bradley and Keimig, 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 why the increased precipitation associated with El Niño, otherwise limited to lower coastal areas, can sometimes be observed in this part of the Andes. However, the relationship is weak and does not hold true during all ENSO events. Similar observation have been made in coastal areas of Northern Peru (Goldberg, Tisnado and Scofield, 1987) and in more recent studies of the rainfall spatial patterns in Ecuador (Tobar and Wyseure, 2018).“La Sierra” region has a complex spatial precipitation pattern. Vuille, Bradley and Keimig (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, Bradley and Keimig (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, Bradley and Keimig (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 dipolike 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 in- fluence of this tropical Atlantic SSTA mode and coupled ITCZ displacements upon precipitation anomalies ex- tends even as far west as the eastern Andes of Ecuador. The Pacific influence on this mode, however, cannot be completely ruled out, since Pacific 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 valleys and basis are rather low, varying between 800 and 1500 mm/year (Vuille, Bradley and Keimig, 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). Vuille, Bradley and Keimig (2000) acknowledges a deficit in rainfall during El Niño events while no significant effects was identified by Tobar and Wyseure (2018), concluding 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, whereas La Niña can result in droughts which has slower acting but also potentially disastrous consequences. The rainfall-runoff relationship in “El Oriente”, as in other parts of the Amazonian basin, displays a large lag between peaks in rainfall and peaks in river discharge, with river flows showing a 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 interactions, and rivers generally having a shallow bed and topographical slopes, with relative slow moving waters. However, rivers located in upstream catchments in the Andean region are prone to flash flooding and are highly sensitive to extreme localized rainfall events (Laraque *et al.*, 2009).

# Data

## Flash flood reports for Ecuador

**Table 1** – 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** | **N. of flood reports (EFFCI>=1)** | **N. of flood reports eliminated** | **N. of flood reports retained** |
| 2019 | 300 | 3 | 297 |
| 2020 | 190 | 0 | 190 |

**Table 2** – 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 1).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **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 %) |

## Rainfall forecasts: ecPoint and ECMWF ensemble

ecPoint (Hewson and Pillosu, 2021) is a statistical post-processing technique that addresses the two main factors that affect the utility of global NWP model outputs, especially for guidance on extreme local events: lack of information on forecast sub-grid variability (Göber, Zsótér and Richardson, 2008) and systematic biases (see Lavers, Harrigan and Prudhomme (2021) for a description of biases in the rainfall forecasts of the ECMWF ensemble prediction system). Systematic biases are model errors which lead to forecast under-/overestimations at grid scale, whilst the lack of information on forecast sub-grid variability is a characteristic of NWP models. NWP models forecast only grid-box averages and do not enable forecasters to assess when/where smaller/higher local values than the model average might occur. Statistical post-processing can improve the forecast accuracy and reliability by correcting biases and converting forecasts from grid-box to point scale (Buizza, 2018). A plethora of statistical post-processing techniques have been developed over the last 50 years (see Vannitsem *et al.* (2021) for a review). However, the operational creation of statistically post-processed forecasts remains problematic to this day, especially in the case of global NWP models and extreme 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, and how ecPoint has addressed them by applying a “global remote calibration” approach. The concept behind this approach is 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, **Fig. 1a**) that are used to convert each single raw grid-box forecast into a distribution of equally probable bias-corrected point-scale realizations (**Fig. 1b**).

Rainfall forecasts from the ECMWF ENSemble prediction system (ENS) are the benchmark dataset in this study as they provide good quality rainfall forecasts (Haiden *et al.*, 2018). Furthermore, since the focus of the study is analysing primarily medium range forecast (from day 3), the ENS provides forecasts with long enough lead times for a comparison with ecPoint-Rainfall which will not be possible with km-scale models which forecasts typically stop at ~day2. ENS consists of a control run and 50 perturbed ensemble members, both with a spatial resolution of ~18km. The control run is started from the best possible unperturbed initial conditions, while the 50 members are started from perturbed initial conditions and are subject to a stochastic representation of model uncertainties (Leutbecher and Palmer, 2008).

# Methods

## Creation of Ecuador’s mask

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.

## Management of the flood reports

* We extracted the flood reports for 2019 to compute the rainfall thresholds, and for 2020 to compute the verification statistics for ecPoint and ECMWF ENS. See **Table 2** for the distribution of the flood reports in the different regions and for different values of EFFCI. See **Figure 3** for the spatial distribution of the flood reports.
* We created 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. The grid-boxes in the Ecuador mask that contained at least a flash flood report on a specific accumulation period were assigned the value 1; otherwise, they were assigned the value 0. This approach allowed us to compute all the elements of the contingency table, i.e., non only hits and misses, but also false alarms and correct negative. This allows us to present the full picture of the verification.
* We are considering four overlapping accumulation periods for each day in the verification period (2020-01-01 to 2020-12-31) to have the best chance to capture the rainfall event that generated the flash flood event considered. This approach is considered as we do not have more precise information on when the rainfall that caused the flash flood event happened. However, considering this approach means that one flash flood report is always associated to two observational fields, and the same flood report could contribute to the hits or the misses elements in the contingency table. The implications of this experiment design will be shown in the Results section and discussed in the Discussion section.

## Verification of ecPoint-Rainfall forecasts

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. The reliability component of the Brier score (RBS) has been considered in this study to test the reliability of the ecPoint and ENS forecasts to predict flash floods. 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 RBS and ROC curves are created for different rainfall thresholds. Hamill and 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. Seven relative thresholds (i.e., 75th, 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 the following section.

Regarding the use of ROC curves for the verification of extreme events, Bouallegue and 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:

* 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.
* ROC curves are closed by drawing a straight line between the last meaningful point of the ROC curve and the top-right corner.
* 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 so the as a sum of trapeziums.

Bouallegue and 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 of the two competing forecasting systems. Otherwise comparing the ROC curves for the two competing forecasts could lead to misleading conclusions. 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 10000 replicates was adopted. 95% of the observational dataset is resampled 10000 times with no replacement to provide a bootstrap distribution for which the statistics are recomputed every time. The differences in the ecPoint/ENS statistics is calculated using a 95% confidence interval.

The plots show the step at the end of the period. The plots were created considering two model runs, the 00 UTC and the 12 UTC run. This means that, in order to consider the same validation period, two different lead times had to be considered for the two different runs. In the plots, the average lead time is shown as representative of the two accumulation periods. For example, for the validation period 2020-02-01 between 00 UTC and 12 UTC, the following forecasts were considered: 2020-02-01, 00 UTC between (t+0,t+12), and 2020-01-31, 12 UTC between (t+12,t+24). In the plots, the steps considered for this particular accumulation period is t+18 (i.e., (12+24)/2 = 18).

We considered overlapping periods to have the best chance of capturing the rainfall event that caused that flash flood event, as we don’t have any other better information to pair rainfall-flash flood events. Therefore, it could be that a flash flood event recorded at 13 UTC might be a hit if a rainfall forecasts is > RT in the period 6-18 UTC but it might be a missed in the period 12-24 UTC because the event happened in the previous six hours.

## Determination of rainfall thresholds for verification

In the absence of a dense network of rainfall observations needed to assess the magnitude of localized rainfall extremes in small regions, i.e. one country, and in a limited period of time, i.e. only one year verification period (Haiden and Duffy, 2016), the rainfall values associated with flash flood events in Ecuador were determined from day 1 ecPoint-Rainfall forecasts. Day one ecPoint-Rainfall forecasts are considered the nearest equivalent to rainfall observations, and their 99 percentiles are considered representative of the rainfall variability within the grid-box that contains the flash flood reports (Hewson and Pillosu, 2021). This last aspect provides an important advantage over the use of “real” observations that, by chance, might or not capture the extreme rainfall value that generated the flash flood (Hewson and Pillosu, 2021).

Only flood reports for “La Costa” and “La Sierra” were considered as in “El Oriente” there are not many reports (8 in 2019 and 11 in 2020 for EFFCI>=1, and there are no reports for EFFCI>=6 and >= 10. See **Table 2**).

How ecPoint-Rainfall is capable to identify flash floods was analysed by creating Definition of rainfall thresholds

* The rainfall thresholds were computed for “La Costa” and “La Sierra” separately to capture possible differences in the distribution of rainfall values associated with flash flood events in those regions.
* For each flash flood report, extract all the percentiles of day1 ecPoint-Rainfall forecasts. Such forecasts are considered to represent the rainfall sub-grid variability within the grid-box.
* The 00 UTC and 12 UTC runs for ecPoint-Rainfall were considered. This means that up to four accumulation periods can include each flash flood report. Thus, each flash flood report can have associated 99, 198, 297 or 396 day1 ecPoint-Rainfall realizations. The time associated to each flood report was used to associate it with the forecast period that included that report. The flood reports provided the times in local time. They were converted into UTC time for a direct comparison with the forecasts that are provided in UTC time.
* We pull together all the 198 X n realizations (where n corresponds to the number of flash flood reports in the calibration dataset) and compute percentiles that will function as rainfall thresholds for the verification.

# Results

## Area under the ROC curves (AURCs)

Figure 4 shows the AURC up to day 3 for all considered relative rainfall thresholds (75th, 85th, 90th, 95th, 98th, 99th percentiles). Only the AURCs for flood reports with an EFFCI>=6 are shown. The AURCs for flood reports with and EFFCI>=1 and >=10 can be found in the supplemental material.

The AURCs are relatively flat for both ecPoint and ENS, and for both “La Costa” and “La Sierra” regions. Therefore, no significant decrease in skill with lead time is observed.

The AURC for ecPoint tend to be always higher than the AURC for ENS at all lead times, all relative rainfall thresholds, and in both “La Costa” and “La Sierra” regions. The only exceptions are the AURC for the relative rainfall thresholds >= 75th, 85th and 90th percentile in “La Sierra” region. However, in both “La Costa” and “La Sierra” regions, the difference between ecPoint’s and ENS’s AURCs appears to be not significant for such relative rainfall thresholds. The same behaviour is observed in “La Costa” also for higher relative rainfall thresholds (95th, 98th and 99th percentiles), while in “La Sierra” the difference between ecPoint’s and ENS’s AURCs tends to increase and become more significant. For the relative rainfall thresholds >=98th and 99th percentile, the ENS’s AURC is almost equal to 0.5 showing no discrimination ability for the ENS forecasts in forecasting flash floods due to the fact that no ENS forecast exceeded the rainfall thresholds (equal to 45.9 mm/12h for the 98th percentile, and 65.1 mm/12h for the 99th percentile).

## ROC curves

Figure 5 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.

# Case studies

## Flash flood event in “La Costa”

## Flash flood event in “La Sierra”

# Discussions

## Area under the ROC curves (AURCs)

## On the overlapping ROC curves for ecPoint and ENS in “La Costa”

Bouallegue and Richardson (2021) argue that if the ROC curves for a post-processed and raw forecasts overlap with the exception of only a couple of points, the post-processing might not be altering the underlying information content in the raw forecasting system. Instead, the improvement seen on few points in the ROC curve for the post-processed forecasts might be only attributed to a change in the frequency of the events (i.e., due to a bigger number of ensemble members, and therefore a bigger spread, the post-processed forecasts exceed the event-threshold more often than the raw forecasts). Bouallegue and Richardson (2021) arrive to this conclusion by analysing the ROC curves for raw and post-processed forecasts that add information on sub-grid variability to the raw forecasts and increase the number of raw ensemble members. They compared ROC curves and AURC calculated with the trapezium and the bi-normal technique. The ROC curves computed with the latter technique do not differ much (concluding that the post-processing technique does not alter the underlying discrimination ability of the raw forecasting system) while the ROC curves computed with the former technique overlap perfectly with the exception of few points (concluding that the ROC curve computed with the trapezium technique provides the false illusion of an improved underlying discrimination ability of the post-processed forecasts over the raw ones).

The ecPoint post-processing technique not only increases the number of ensemble members available to users (i.e., ecPoint provides a bigger spread than the raw forecasts). It also adds information on the possible rainfall sub-grid variability to produce forecasts for a point within the grid-box. If the information added by the post-processing technique regards also sub-grid variability (and not only, for example, bias correction), the raw forecasts will still not show a similar spread by simply increasing its ensemble members because, even a bigger ensemble, would still represent average forecasts over the grid-box which will not coincide with point-based rainfall observations unless the sub-grid variability is small. The two forecasting systems will still produce forecasts with different resolutions (i.e., grid-box scale and point-base scale) even when the spread of the raw ensemble is increased. This difference between the resolutions of ecPoint and raw NWP models has been shown in Pillosu et al. (2021) for rainfall case studies in Costa Rica and Hungary. It is important to notice that this distinction is valid for ecPoint and for whatever other post-processing technique that adds information on sub-grid variability. Therefore, it is valid also for the post-processing technique used by Bouallegue and Richardson (2021), and described in Ben Bouallegue *et al.* (2020).

Another important consideration regards the fact that the information provided by the bigger ensemble could be useful for a certain type of users, and they will not have access to that type of information if they use only the raw ensemble. Bouallegue and Richardson (2021) show that the potential economic value of the post-processed forecasts is higher than the one for the raw forecasts only for users with small cost-loss ratios (typically less than 2%) in case of overlapping ROC curves with the exception of few points closer to the top right corner of the diagram. This means that typically users who would incur very large losses when a rare event happens and not protection was taken would benefit of the post-processed forecasts (Richardson D., 2000). Even in the case that the resolution of the two forecasting systems (raw and post-processed) were similar and the ROC curves were proved to not be different with the exception of few points since the post-processed has a bigger number of members, when more points can be added to a ROC curve by having a greater number of ensemble members, the area will increase and so will the value for users (Richardson D., 2000). This is because the users will be actually able to use the extra information provided by the post-processed forecasting system that is not available in the raw ensemble. One could parametrize the ROC curves (for example using the bi-normal technique) to demonstrate the potential value that would be achieved if all possible probability thresholds could be used (i.e., an infinite ensemble of forecasts) to estimate the benefit that could be achieve by the raw forecast if it had a larger number of ensembles. Even in the case the two ROC curves (trapezium and bi-normal) would be the same, the reality stays that the raw forecast has a smaller number of ensembles that might not provide to users the full picture of what could happen (e.g., rainfall event exceeding 50 mm/12h) if the uncertainty of the event is high (typically for rare events). Furthermore, to have only one more point in the ROC curve like in the case of the raw ENS and ecPoint (that goes from the maximum 98th percentile for ENS to the 99th percentile for ecPoint), the number of ensemble members should double which comes to a high computational cost which might not be affordable. Here the focus is to show what are the practical benefits on using two forecasting systems with different configurations to forecast flash floods. The focus is not to compare their underlying discrimination abilities. The results in Figure 5 show that the raw ENS does a good job in forecasting rainfall that might cause flash floods, and would probably fully overlap with the ROC curve for ecPoint if it had 100 members. However, increasing ENS to that number of ensemble members would come with a very high cost that cannot be afford with the current computational resources available.

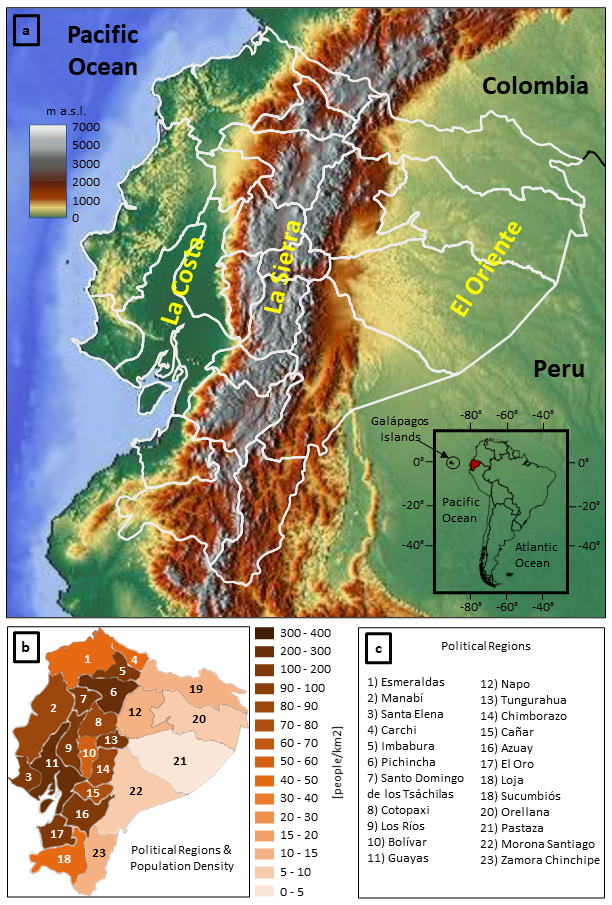
## On the effect in the verification scores of the spatial/temporal uncertainty in the flood reports

* The verification method needs to consider that the flash flood reports are not collected everywhere.

# Conclusions

# Tables

# Figures

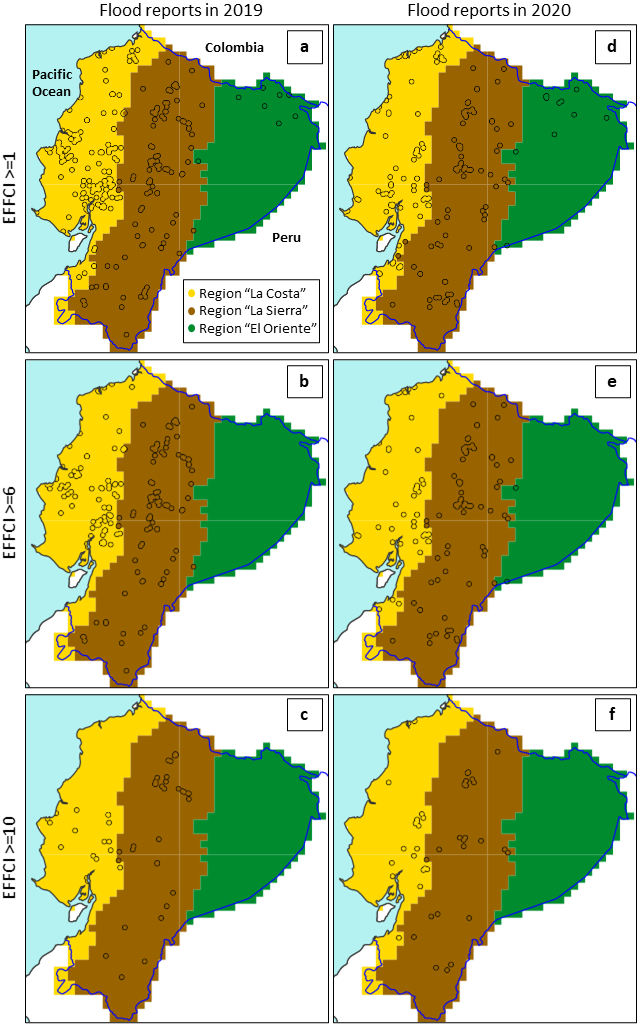


**Figure 1** - Panel (a) shows Ecuador’s orography, its political regions, and the location of Ecuador’s three main geographical regions: the coast (“La Costa”), the highlands (“La Sierra”), and the Amazon (“EL Oriente”). The small box shows Ecuador’s location in South America. Panel (b) shows the population density (in people/km2) for each region from 2020 census (source: <https://es.wikipedia.org/wiki/Provincias_de_Ecuador>). Panel (c) lists the names of Ecuador’s political regions following the numbers indicated in panel (b).

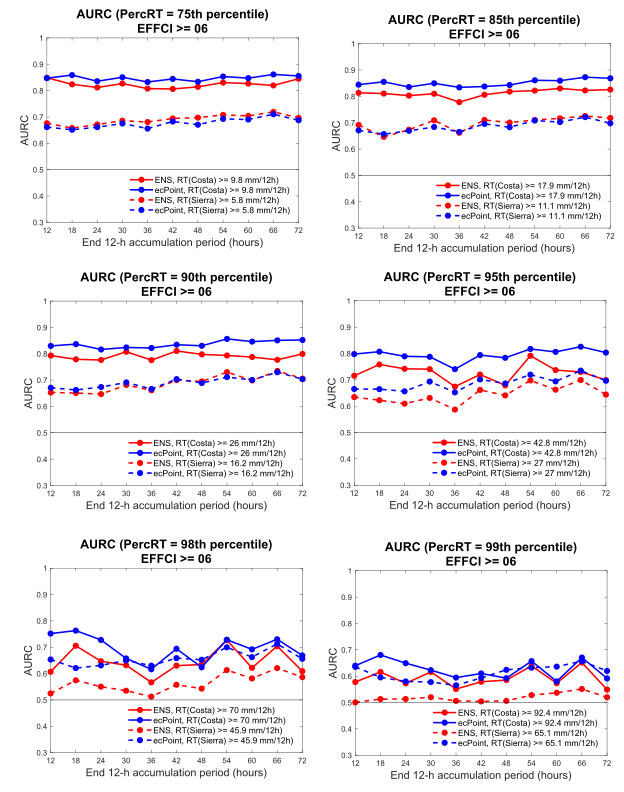
A picture containing diagram

Description automatically generated

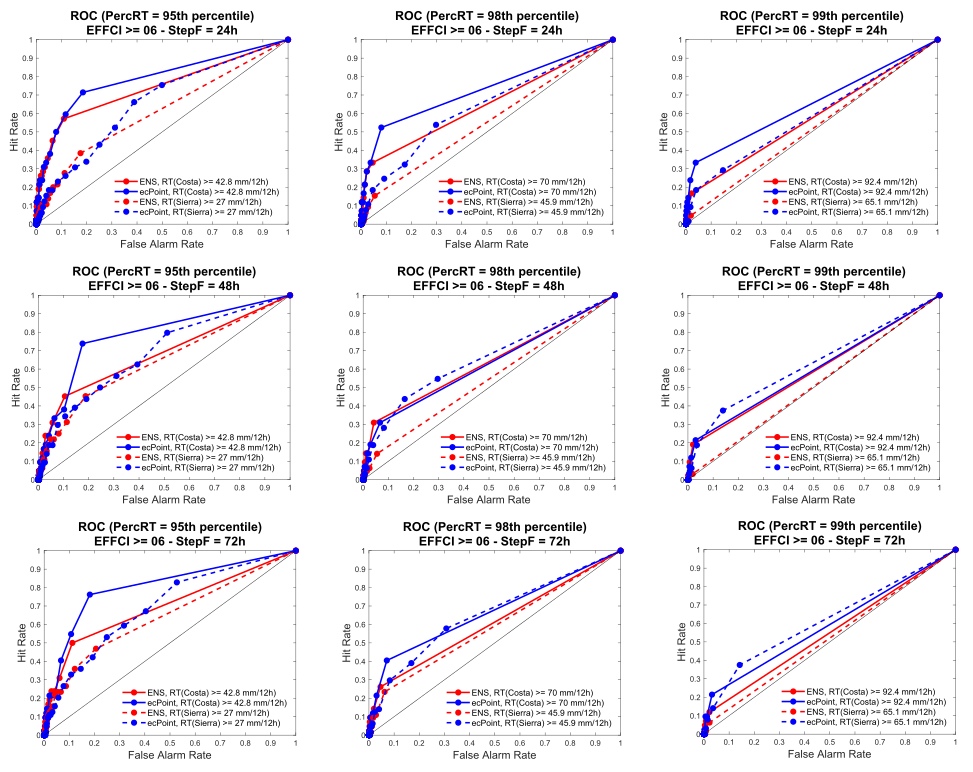
**Figure 2** - Panels (a) and (b) show, respectively, the workflow for the calibration and the forecasts generation using the ecPoint methodology. Panel (c) shows an example of the two ecPoint-Rainfall products for 12-hourly rainfall, available on ecCharts (https://www.ecmwf.int/en/forecasts/ ecCharts), i.e., a map plot for the 99th percentile (first row) and a map for the probabilities of exceeding 10 mm/12h (second row).



**Figure 3** - Flood reports for 2019 (first column) and 2020 (second column) indicated by the black circles. The maps show flood reports with EFFCI>=1 (first row), an EFFCI>=6 (second row) and an EFFCI>=10 (third row).



**Figure 4** – Area under the ROC curve (AURC) for relative rainfall thresholds greater than the 75th, 85th, 90th, 95th, 98th and 99th percentile. The legends in each diagram indicate the correspondent rainfall amounts (in mm/12h) for each relative rainfall threshold and for each region. All AURC refer to flood reports with an EFFCI>=6 with forecast up to day 3. The red and blue line correspond to ENS and ecPoint forecasts, respectively. The solid and the dash lines correspond to the AURC for “La Costa” and “La Sierra” regions, respectively.



**Figure 5** – Each row (from top to bottom) correspond to ROC curves for day 1 (12-h accumulation period between t+12 and t+24), day 2 (t+36,t+48), and day 3 (t+60,t+72). Each column (from left to right) correspond to ROC curves for the rainfall threshold of 95th, 98th and 99th percentile. The legends in each diagram indicate the correspondent rainfall amounts (in mm/12h) for each relative rainfall threshold and for each region. All ROC curves refer to flood reports with an EFFCI>=6. The red and blue line correspond to ENS and ecPoint forecasts, respectively. The solid and the dash lines correspond to the AURC for “La Costa” and “La Sierra” regions, respectively.

# References

Ben Bouallegue, Z. *et al.* (2020) ‘Accounting for Representativeness in the Verification of Ensemble Precipitation Forecasts’, *Monthly Weather Review*, 148(5), pp. 2049–2062. doi: 10.1175/mwr-d-19-0323.1.

Bouallegue, Z. Ben and Richardson, D. S. (2021) ‘On the ROC Area of Ensemble Forecasts for Rare Events’. doi: 10.20944/PREPRINTS202111.0535.V1.

Buizza, R. (2018) ‘Ensemble Forecasting and the Need for Calibration’, in *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, pp. 15–48. doi: 10.1016/b978-0-12-812372-0.00002-9.

Buizza, R., Milleer, M. and Palmer, T. N. (1999) ‘Stochastic representation of model uncertainties in the ECMWF ensemble prediction system’, *Quarterly Journal of the Royal Meteorological Society*, 125(560), pp. 2887–2908. doi: 10.1002/qj.49712556006.

Candille, G. *et al.* (2003) *Forecast Verification. A practicioner’s Guide in Atmospheric Science*. Edited by Ian T. Joliffe and David B. Stephenson. Wiley.

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(3), pp. 359–365. doi: 10.1002/met.78.

Goldberg, R. A., Tisnado, G. M. and Scofield, R. A. (1987) ‘Characteristics of extreme rainfall events in Northwestern Peru during the 1982-1983 E1 Nifio Period’, *Journal of Geophysical Research: Oceans*, pp. 14225–14241. doi: 10.1029/JC092iC13p14225.

Haiden, T. *et al.* (2018) ‘Evaluation of ECMWF forecasts, including the 2018 upgrade’, *ECMWF Technical Memoranda*, (831). doi: 10.21957/ldw15ckqi.

Haiden, T. and Duffy, S. (2016) ‘Use of high-density observations in precipitation verification’, *ECMWF Newsletter*, (147), pp. 20–25. doi: 10.21957/hsacrdem.

Hamill, T. M. and Juras, J. (2006) ‘Measuring forecast skill: Is it real skill or is it the varying climatology?’, *Quarterly Journal of the Royal Meteorological Society*, 132(621 C), pp. 2905–2923. doi: 10.1256/qj.06.25.

Hewson, T. D. and Pillosu, F. M. (2021) ‘A new low-cost technique improves weather forecasts across the world’, *Communications Earth & Environment*, 2(1), p. 132. doi: 10.1038/s43247-021-00185-9.

Laraque, A. *et al.* (2007) ‘Heterogeneous distribution of rainfall and discharge regimes in the Ecuadorian Amazon basin’, *Journal of Hydrometeorology*, 8(6), pp. 1364–1381. doi: 10.1175/2007JHM784.1.

Laraque, A. *et al.* (2009) ‘Sediment budget of the Napo River, Amazon Basin, Ecuador and Peru’, *Hydrological Processes*, 23(25), pp. 3509–3524. doi: 10.1002/hyp.7463.

Lavers, D. A., Harrigan, S. and Prudhomme, C. (2021) ‘Precipitation Biases in the ECMWF Integrated Forecasting System’, *Journal of Hydrometeorology*. doi: 10.1175/jhm-d-20-0308.1.

Leutbecher, M. and Palmer, T. N. (2008) ‘Ensemble forecasting’, *Journal of Computational Physics*, 227(7), pp. 3515–3539. doi: 10.1016/j.jcp.2007.02.014.

Recalde-Coronel, C. G., Barnston, A. G. and Muñoz, Á. G. (2014) ‘Predictability of december-april rainfall in coastal and Andean Ecuador’, *Journal of Applied Meteorology and Climatology*, 53(6), pp. 1471–1493. doi: 10.1175/JAMC-D-13-0133.1.

Tobar, V. and Wyseure, G. (2018) ‘Seasonal rainfall patterns classification, relationship to ENSO and rainfall trends in Ecuador’, *International Journal of Climatology*, 38(4), pp. 1808–1819. doi: 10.1002/joc.5297.

Towner, J. *et al.* (2020) ‘Attribution of Amazon floods to modes of climate variability: A review’, *Meteorological Applications*, 27(5). doi: 10.1002/met.1949.

Trigg, M. A. *et al.* (2009) ‘Amazon flood wave hydraulics’, *Journal of Hydrology*, 374(1–2), pp. 92–105. doi: 10.1016/j.jhydrol.2009.06.004.

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

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(14), pp. 2520–2535. doi: 10.1175/1520-0442(2000)013<2520:CVITAO>2.0.CO;2.