Scientific Research Publishing

Atmospheric and Climate Sciences, 2025, 15(2), 426-450
https://www.scirp.org/journal/acs
ISSN Online: 2160-0422
ISSN Print: 2160-0414

Evaluating the Performance of the EcPoint Post-Processing Method for Ensemble Rainfall Forecasts over south China

Teddy Mwira[1,2], Sonum Stejik[1]

[1]School of Atmospheric Science, Nanjing University of Information Science and Technology, Nanjing, China
[2]Department of Meteorology, Ministry of Water and Environment, Kampala, Uganda
Email: tedbirungi@gmail.com

How to cite this paper: Mwira, T. and Stejik, S. (2025) Evaluating the Performance of the EcPoint Post-Processing Method for Ensemble Rainfall Forecasts over south China. *Atmospheric and Climate Sciences*, **15**, 426-450.
https://doi.org/10.4236/acs.2025.152022

Received: February 27, 2025
Accepted: April 18, 2025
Published: April 21, 2025

Abstract

Developing reliable weather prediction systems is still a challenging task due to the complexity of the Earth System and the chaotic behavior of its components. Small errors introduced by observations, their assimilation and the forecast model configuration escalate chaotically, leading to a significant loss in forecast skill with time. Traditionally, rainfall forecasts have been generated at grid-based spatial resolutions, providing valuable information on regional precipitation patterns. However, Weather varies markedly within a grid box and forecasts for specific sites have occasionally failed inevitably. The grid-based forecasts may not always meet the needs of decision-makers at specific points of interest. The major challenge is dealing with variations in sub-grid variability, that is to say, the variation seen amongst rainfall point values within a given model grid box, more especially in convective situations. While ensemble forecasts have shown promise in capturing the uncertainty inherent in average rainfall predictions of much larger grid boxes, their utility at point locations has not been extensively explored. Most evaluation studies focus on grid-based verification metrics, which may not accurately reflect forecast performance at individual points of interest. EcPoint, a post-processing approach developed at the European Centre for Medium Range Forecasts (ECMWF), is tailored to forecast rainfall at point locations. In this study, we evaluate the performance of the EcPoint post-processing method over the south China region. The analysis focuses on the reliability, accuracy and discrimination skill of this post-processing method over the three provinces in south China (Anhui, Zhejiang and Jiangsu). We examine performance versus lead time, seasons, and altitude. Through verifications, the study highlighted the added value of the post-processing method over Raw ensemble forecasts. One year of verification demonstrates that, between the Raw ensemble and post-pro-

cessed EcPoint forecasts, EcPoint is the more reliable and skillful system, adding significant value to most rainfall events occurring during the day and the seasonal associated events, as well as the topography-associated rainfall events. To complement the one-year verification analysis, a case study was conducted on an extremely heavy rainfall event observed on June 2, 2022 at 12 UTC. The analysis demonstrated EcPoint's ability to provide more localized and refined forecasts whereas Raw didn't provide any possibility of rainfall, particularly at short lead times. At longer lead times, EcPoint ensembles maintained relatively low probabilities, but offered improved performance in capturing rainfall variability, while Raw ensemble exhibited broader but less precise rainfall predictions with a tendency of over warning of some areas. Future work can extend the evaluation to more diverse climatic and topographic regions of China to enhance the general applicability of the method. Although based solely on the global ECMWF-IFS model, EcPoint performs well over the small domain of south China (three provinces). Besides verifying EcPoint, the study confirms that the post-processing method can significantly improve the forecast performance.

Keywords

Ensemble Forecasts, Post-Processing, Point Rainfall Forecasts, EcPoint, Verifications

1. Introduction

Accurate predictions of heavy and intense rainfall are crucial for impact-based forecasting, which plays a vital role in mitigating significant damages and losses worldwide. In regions with diverse topography and highly variable climatic conditions, precise rainfall forecasts are particularly critical for reducing the risks of weather-related hazards, such as floods and landslides [1]. However, existing ensemble rainfall forecasts in such areas often fail to adequately capture the spatial and temporal variability of rainfall. This limitation hinders effective decision-making and resource allocation, resulting in inefficiencies and less optimal outcomes [2]. General Circulation Models (GCMs), despite their capability to simulate large-scale atmospheric circulation, exhibit low skill in predicting local rainfall. This deficiency arises because local precipitation is heavily influenced by topography and land-sea contrasts, which are often poorly represented in coarse-scale models [3] [4]. Ensemble forecasts offer critical insights into the uncertainty surrounding future weather conditions, empowering decision-makers to act confidently in uncertain scenarios. By quantifying forecast uncertainty, ensemble systems enhance preparedness and planning. However, these ensemble forecasts are prone to biases, such as systematically underestimating or overestimating rainfall, or being overly confident or overly uncertain about future weather. Additionally, global ensemble forecasts typically represent "average weather" over predefined grid regions. This generalization often fails to capture the localized variability of

rainfall, especially in convective events like thunderstorms, where rainfall can differ significantly within the grid region [5]. One major challenge in rainfall forecasting is actually the significant variation observed in point values within a model grid box, particularly in convective situations. This variability underscores the pressing need to evaluate post-processing techniques that can enhance the accuracy and reliability of ensemble rainfall forecasts at point locations within a grid box. While ensemble forecasts have shown potential in addressing the inherent uncertainty of weather predictions, their utility at specific point locations remains underexplored. Conventional ensemble forecasts typically provide average rainfall predictions for much larger grid boxes rather than for specific sites of interest. This limitation highlights the necessity for tailored evaluation methodologies that account for the spatial variability of rainfall and the distinct characteristics of point-based forecasts. Correction algorithm, such as post-processed forecast ensembles (EcPoint), can be applied to ensemble forecasts to effectively address these limitations and challenges. The EcPoint post-processing method, developed by ECMWF, aims to enhance ensemble rainfall forecasts by addressing sub-grid variability and improving the reliability of probabilistic predictions. It has been shown to discriminate successfully between rainfall events and non-events, across both small and large rainfall amounts, while reliably quantifying forecast uncertainty [6]. The core premise of EcPoint lies in its ability to relate forecasted values to point observations based on physical considerations rather than specific locations. This approach accounts for sub-grid variability and grid-scale biases according to prevailing grid box weather situations [7]. The EcPoint post-processing method, designed specifically for point-based rather than region-based forecasts, has been shown to significantly enhance forecast quality considerably compared to ECMWF's global weather model forecasts (ECMWF-IFS), in a timely manner and at low cost [5]. This study therefore aims to verify EcPoint over the south China region, evaluating its accuracy, reliability and discrimination skill through rigorous verifications and case studies to validate its potential improvement and applicability in operational forecasting systems. The study evaluates the performance of EcPoint over south China, particularly in three provinces: Anhui, Zhejiang, and Jiangsu. While previous studies have demonstrated EcPoint's effectiveness, its performance in a broader range of topographical and climatic conditions remains an open research question.

EcPoint

[5] and [8] give a comprehensive introduction to EcPoint and ECMWF's associated point forecast product, respectively. Here, we just provide a summary of EcPoint, which closely follows theirs. EcPoint, a postprocessing approach developed at the European Centre for Medium-Range Weather Forecasts (ECMWF) is tailored for point forecasting of rainfall and has been shown to increase forecast skill considerably. It is a statistical post-processing method that accounts for the degree of variation within each grid box, bias on the grid box scale and weather dependence of each. The true utility of EcPoint lies in creating and using separate frequency distributions for weather types that differ from one another in signifi-

cant and physically realistic ways. It was specifically designed to improve the reliability and discrimination ability of the forecast, particularly for large totals.

It is noted that for all thresholds and all lead times, EcPoint-Rainfall has better discrimination ability than the corresponding Ensemble Raw and post-processed COSMO forecasts [9]. The methodology is based on physically relevant statistical relationships between the larger-scale weather features well represented by ECMWF forecasts and local realizations represented by point observations. These relationships make it possible to compute statistically based probabilities for point rainfall rather than Raw-ensemble-based [10] [11]. This includes extremes, which can be used to infer the likelihood of flash floods for use on platforms such as the European and Global Flood Awareness Systems [12]. EcPoint is the name given to the post-processing philosophy, whilst the companion calibration software is called "EcPoint-calibrate" [5] [13]. The EcPoint-Calibrate performs a physical-statistical analysis of meteorological data (model data and observation data) that provides users with a tree-like graphical support tool to investigate biases and errors in the ensembles that is the conditional verification and post-process the ensembles to produce probabilistic ensemble forecasts for points.

EcPoint Calibration Process

EcPoint calibration follows a physically relevant statistical approach, classifying weather types using key governing variables such as convective precipitation, 700 hPa wind speed, solar radiation, CAPE, and total precipitation. These parameters help model sub-grid variability, making EcPoint adaptable to different locations without requiring site-specific calibration. Each forecast grid box from an ensemble member is assigned a weather type, and forecast errors are estimated based on the distribution of point rainfall outcomes within the grid box. This process accounts for both sub-grid variability and systematic biases, generating a large calibration dataset using just one year of observational data. EcPoint is based on the concept of conditional verification, which involves identifying and correcting errors in model rainfall forecasts based on diagnosed grid-box weather types. Calibration helps detect systematic forecast errors, while post-processing adjusts forecasts accordingly. Since weather types vary across grid boxes, lead times, and ensemble members, EcPoint uses a decision tree approach where each level represents a governing variable, each leaf corresponds to a distinct weather type, and all leaves collectively account for all possible atmospheric conditions [9]. Each weather type is assigned a mapping function, which defines typical forecast errors observed worldwide under similar conditions. This function accounts for both sub-grid variability and grid-scale bias, ensuring better representation of localized rainfall patterns. During forecast construction, EcPoint assigns a weather type to each grid box at each time step for all ensemble members. The system then converts forecasts into 100 equiprobable point rainfall values using the respective Forecast Error Representation (FER) mapping function. These values are blended across the ensemble, producing 5100 virtual ensemble members per grid point at each forecast time step. As a result, EcPoint post-processing transforms the raw

ensemble forecast into an "ensemble of ensembles", yielding 99 calibrated ensemble members that better capture rainfall variability and uncertainty.

A key advantage of EcPoint is that its calibration is independent of location. Instead of relying on local observations, EcPoint calibrates forecasts based on universal weather-type classifications, assuming that the underlying physical processes governing rainfall are globally consistent. This allows forecasts to be generated even in regions without direct observational data, making EcPoint applicable to areas where traditional calibration methods would fail. By blending information from different locations with similar rainfall generation mechanisms, EcPoint ensures that forecast improvements extend beyond observation sites, offering reliable probabilistic predictions across diverse geographic and climatic conditions [14]. The post-processing system has been fully automated and requires minimal computing resources to run compared to high-resolution numerical models [5]. The final standard EcPoint output does not provide a set of spatial-temporal forecast scenarios, but rather a single calibrated, probabilistic point forecast distribution for any point in space for each 12 h accumulation period for several lead times. It is depicted in percentile or probability format, with the user able to define, according to purpose [5] and [8]. The product aims to bridge the gap between the relatively coarse resolution of today's global forecasting models and the higher resolution limited-area models needed to describe localized heavy rainfall. Interestingly, it is believed that the point-rainfall product could support the prediction of flash floods across the globe [13]. The point rainfall focuses on estimating the range of totals likely within the grid box, and indeed delivers probabilities for different point values within that grid box. For example, like the range of values presented in the approximate grid box selected in **Figure 1**.

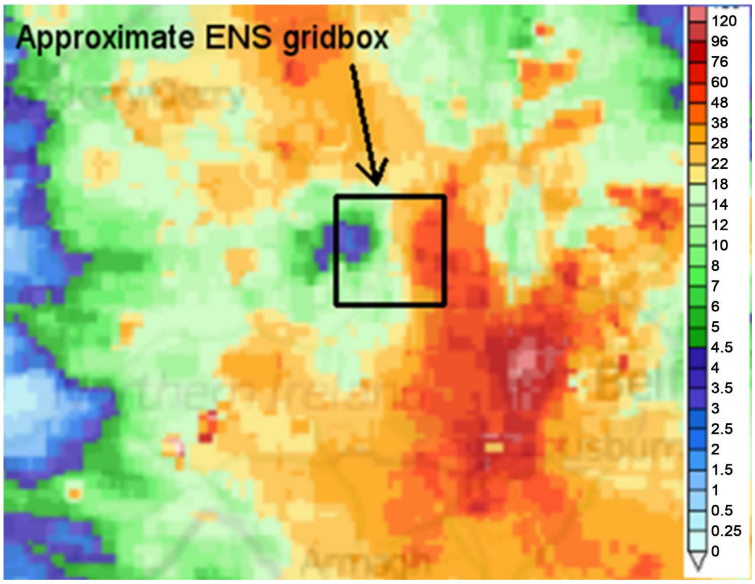

Figure 1. Radar-derived rainfall totals showing how rainfall can significantly vary within a grid box for example in the approximate grid box, it ranges from about 4 mm to about 60 mm.

2. Data and Methodology

2.1. Datasets

Observations database: Meteorological observation data (rain gauge data) for three provinces were used to represent real rainfall observation points. Observed Rainfall (12 hours accumulation) of these provinces for one year (2022) was used to verify our post processed forecast product, using a standard "nearest neighbor" technique to match station location with model grid forecast data. Figure 2 displays the study area for the three provinces in south China (Anhui, Zhejiang, and Jiangsu) and the array of rain gauge observation points that were used for verification. The observation points show altitude variation and the annual rainfall distribution over the region for the year 2022. Previous studies indicated that the summertime south China precipitation is affected by a number of climate systems, making its prediction very challenging. The seasonal mean precipitation over the region is affected by the western north Pacific summer monsoon activity as suggested by [15] and [16]. It is shown that the dominant patterns of interannual variation of early summer South China rainfall themselves depend on the phase of the Pacific decadal oscillation (PDO) [17]-[20]. The interannual variations of seasonal precipitations are quite consistent in winter and spring, indicating that anomalous precipitation tends to prolong in winter and spring [21]. Active warm-rain microphysical processes can also play an important role in some extreme rainfall events, although the relative contributions of warm-rain, riming, and ice-phase microphysical processes remain unclear [22]. Most rainfall anomalies on the whole regional scale of South China are well in phase during winter and spring, and the frequency of persistent drought is higher than that of persistent flood.

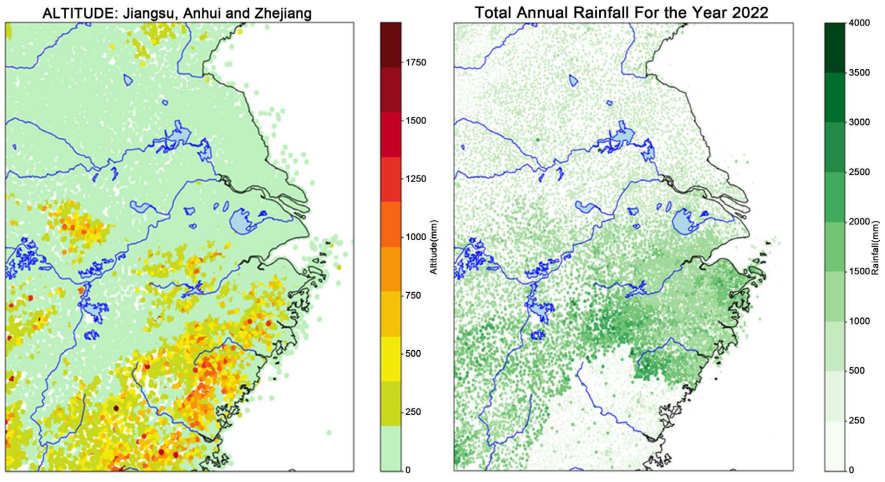

Figure 2. Observation points of the study area; on the left, it's the altitude data of the points and on the right, it's the total annual rainfall of the points (2022).

Forecast data: The forecast data consists of Raw Ensemble forecast (ECMWF-IFS model forecast) comprising 50 perturbed ensemble members with a horizontal resolution of 18 km. The EcPoint forecast output is made available with 99

members for overlapping periods of up to day 10, namely T + 240 h. Both the Raw ensemble forecast and post processed EcPoint ensemble forecasts were obtained from ECMWF for the same one-year period (2022) and 12-hour accumulation.

2.2. Verification Methods

Relative Operating Characteristics (ROC) diagram: To evaluate forecast discrimination ability for different thresholds, we used receiver operating characteristic (ROC) curves, a curve for a particular event, e.g., exceedance of a threshold generated by plotting the true positive rate against the false positive rate. A ROC shows how well the forecast discriminates between two events, and it measures the ability of the forecast system to discriminate between events and non-events (e.g., rain or no rain). It evaluates the hit rate (how often the event is correctly forecasted) against the false alarm rate (how often a non-event is incorrectly forecasted as an event). In our study, probabilistic forecasts were transformed into categorical yes/no forecasts (Observed/not observed event). For different thresholds, corresponding Hit rate (true positive rate) and false alarm rate (false positive rate) were computed and displayed on Relative Operating Characteristics (ROC) diagram. The closer the ROC curve is to the upper left corner of the plot, the better the performance. The area under the ROC curve (AUC) was also computed to quantify the overall performance of the forecast. Overall, AUC provides a single scalar value that summarizes the performance of a classification model across all possible classification thresholds. AUC values between 0.5 and 1 indicate better performance. The higher the AUC, the better the performance in distinguishing between events and non-events.

Continuous Rank Probability Scores (CRPS): We used the continuous ranked probability score (CRPS) [23], which depends on both forecast spread and bias, in order to assess the overall forecast performance. It assesses the entire cumulative distribution function (CDF) of the forecast compared to the CDF of the observed outcomes. By comparing the entire forecast CDF with the observed outcome, CRPS assesses not just the accuracy of the forecast's central value, but also reliability. Therefore, the CRPS value is a comprehensive metric that tells us how well the forecast performs both in predicting the actual rainfall amounts which is accuracy and in ensuring the forecast probabilities are in alignment which is reliability. It is a single metric value that penalizes both misplaced forecasts and mis-calibrated probabilities of the probabilistic forecast. A low CRPS value indicates a forecast that is both accurate (close to the observed value) and reliable (probabilities align well with reality). High values suggest deficiencies in either accuracy or reliability or both. Unlike other scoring rules, like Mean Squared Error (MSE), which only evaluate deterministic forecasts, CRPS evaluates the entire probabilistic forecast by taking into account the spread or range of the predicted values. Continuous ranked probability score is unbounded and can go to infinity except for cases in which the forecast target itself has bounded support. It is the integral of the square of the difference between the cumulative distribution function (CDF) of the prob-

abilistic predictions F and the ground truth y.

$$\mathrm{CRPS}(F,y)=\int_{-\infty}^{+\infty}\left[F(x)-1(x\geq y)\right]^2 \mathrm{d}x$$

where:

- $F(x)$ is the cumulative distribution function of the forecast.
- $1(x\geq y)$ is the indicator function, which equals 1 if $x\geq y$ and 0 otherwise.
- y is the observed value.

While CRPS provide an aggregate measure of forecast accuracy and reliability across the full range of potential outcomes, it does not explicitly prioritize performance for specific thresholds (specific amount of rainfall). This limitation highlights the need for the complementary metric, such as the ROC curve and AUC described earlier, which evaluate the forecasts for events defined by thresholds.

Different verification methods vary differently, and they often give conflicting indications. They vary with region, seasons, and lead times among other factors. What looks good might not be as good as under other circumstances. Therefore, to have a diverse assessment, the verification analysis is considered verse lead time over the whole year, then across seasonal changes and also considered for topography variation. A Case study on a specified date was also considered to verify the forests across different stations with different physical characteristics to analyze the spatial distribution and how the forecast systems capture those observed rainfall distributions.

3. Results and Discussions

Verifications

Verification for one year of 12 h-rainfall forecasts was carried out using truth rain gauge observations from specialized high-density datasets of Anhui, Zhejiang and Jiangsu provinces (see observation point coverage in **Figure 2**). Two probabilistic ensemble forecasts were assessed: the ECMWF Raw Ensemble forecast and the post processed EcPoint forecast. In this framework of study, the fundamental aspects of assessing were the capacity to discriminate events, as well as accuracy and reliability. To evaluate the capacity to discriminate events, we used Relative Operating Characteristic curves and also considered the Area under the Relative Operating Characteristic curve (AROC) [24]. Three different precipitation thresholds were initially considered, given the importance of issuing skillful predictions for thresholds of rain/no rain or alert levels for heavy precipitation or flooding possibilities. We present results for three 12 h accumulation thresholds of 0.2 mm for rain or no rain event, 10 mm for a wet event or moderate rainfall event and 30 mm for high rainfall event. **Figures 3-5** display the plots of ROC analysis for one-year verification across lead time together with their associated area under the curves (AUC). There is a diurnal cycle reflected in AUC values with lead time across all the thresholds for both forecasting systems. The AUC values for EcPoint forecasts remain consistently higher across all lead times, with a diurnal pattern more noticeably observed for the threshold of 0.2 mm, reflecting the system's ca-

pability to better capture the subtle diurnal influences on the no rainfall events. Conversely, the Raw ensemble forecasts exhibit lower AUC values, with a gradual decline as lead time increases, indicating a reduced ability to maintain skill for longer forecasts. Both the Raw Ensemble Forecasts (ECMWF) and EcPoint forecasts provide skillful predictions for all thresholds across all lead times, with all AUC values consistently above 0.5. Although EcPoint does better at capturing localized rainfall events, as reflected in its consistently higher AUC values compared to the raw ensemble throughout. EcPoint demonstrates consistent performance across all thresholds, with the same AUC value particularly for the first lead time, including the 10 mm and 30 mm thresholds. This highlights its ability to maintain skill in event discrimination regardless of rainfall intensity. The added value of the EcPoint system, as observed in the improvement of the ROC area, is attributed to its enhanced ability to capture the wet tails of sub-grid point rainfall distribution, as was noted by [5]. By effectively addressing the variability of rainfall within grid cells, it delivers more refined predictions for high rainfall events than the Raw ensemble forecast.

Then, three separate days were considered for analysis, specifically for night lead times that are at 024 for day one (short lead time), 120 for day 5 (medium lead time) and high lead time for day 10 at 240 lead time. Given that the performance

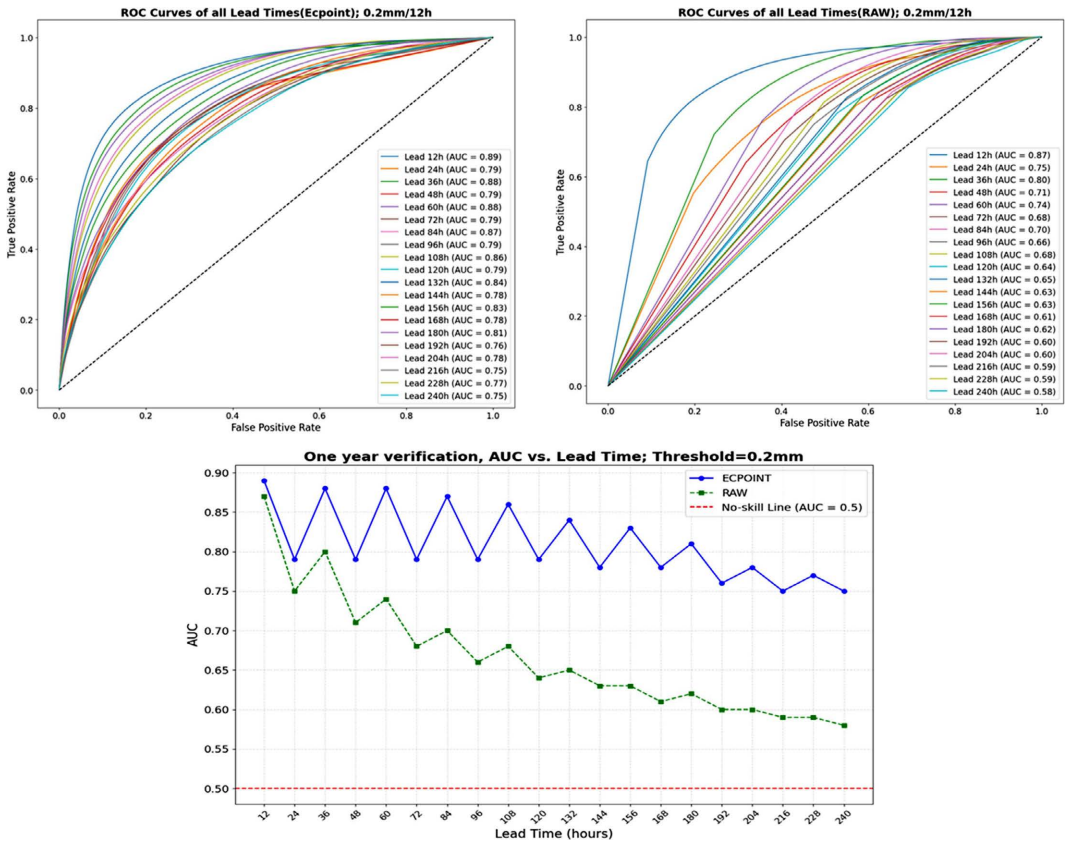

Figure 3. Comparison of ROC curves and AUC values for different lead times at a 0.2 mm threshold with the no-skill reference line at AUC = 0.5.

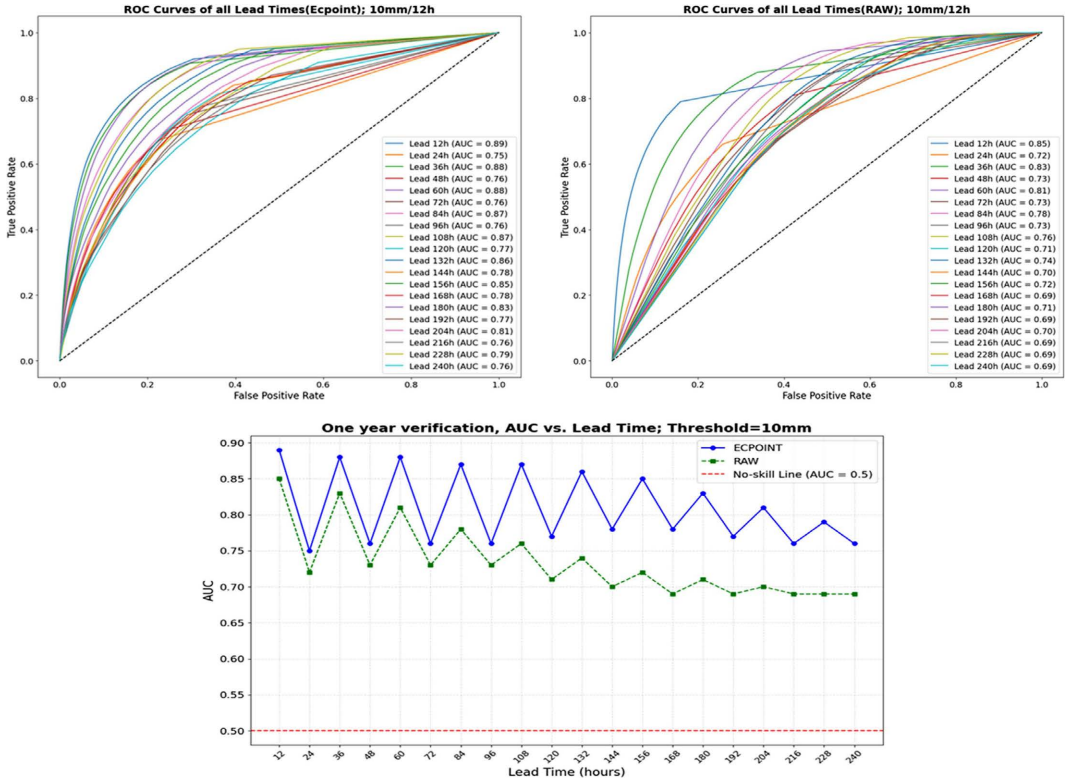

Figure 4. Comparison of ROC curves and AUC values for different lead times at a 10mm threshold with the no-skill reference line at AUC = 0.5.

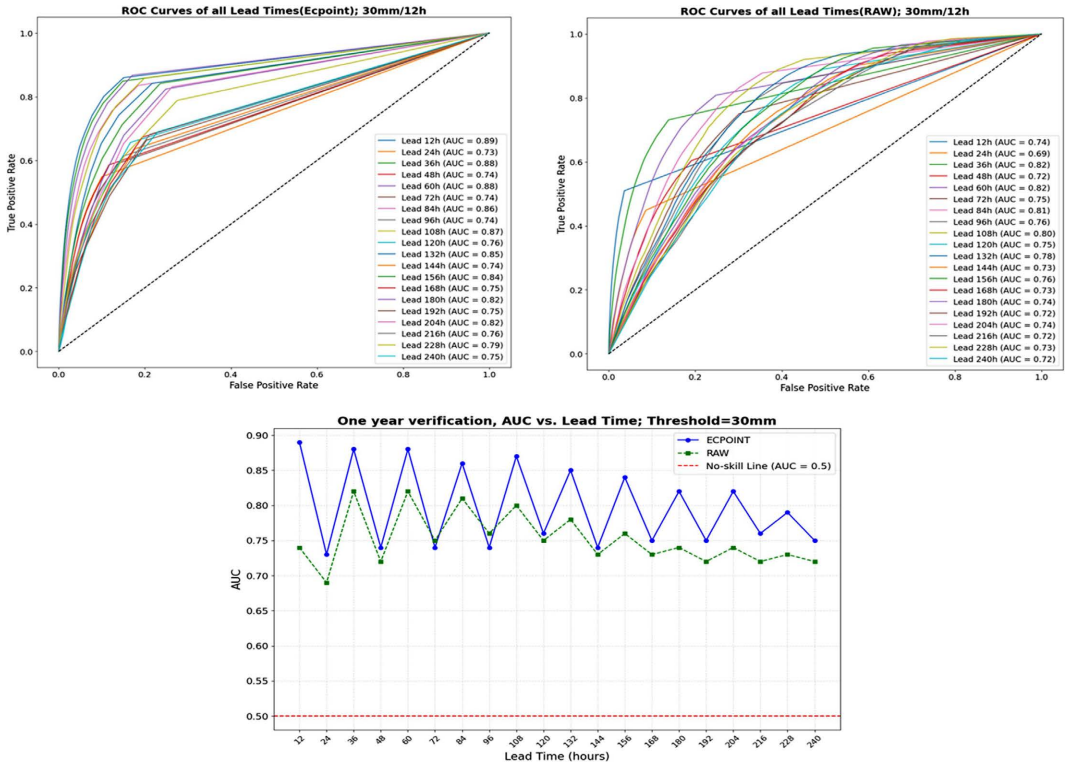

Figure 5. Comparison of ROC curves and AUC values for different lead times at a 30 mm threshold with the no-skill reference line at AUC = 0.5.

drops at night, these separate days were analyzed to further highlight the capabilities of the two models under such atmospheric conditions. These results would highlight the strengths and challenges of the forecasting systems, particularly when factoring in the distinct performance patterns observed during nighttime hours. Four thresholds were considered for this analysis: 0.2 mm, 10 mm, 30 mm and 50 mm (considered for an extremely high event). EcPoint continues to consistently outperform the Raw ensemble across all thresholds and lead times, highlighting the value of statistical post-processing in addressing systematic biases and improving probabilistic forecast skills. In Figure 6, for the 0.2 mm threshold representing rain or no rain event, EcPoint significantly reduces false positives, achieving higher AUC scores compared to the raw ensemble. For instance, at the 24 h lead time, the AUC is 0.79 for EcPoint compared to 0.75 for the Raw ensemble, with much better trends observed at 120 h and 240 h lead times. However, forecasting rain or no rain still remains challenging, particularly at longer lead times for both forecast systems where AUC values drop gradually degrading accuracy. Although EcPoint's adjustments (AUC = 0.79 at 120 h and AUC = 0.75 at 240 h) compared to Raw Ensembles (AUC = 0.64 at 120 h and AUC = 0.58 at 240 h) show that EcPoint can still perform better at high lead times making it a more reliable option for applications sensitive to light precipitation. The sharp drop of AUC values for Raw ensembles suggests a tendency to overpredict the light rainfall, given that the curves are more to the false positive rate. For the 10 mm threshold moderate rainfall EcPoint achieves 0.75 at 24 h compared to 0.72 for the Raw ensemble and 0.77 at 120 h compared to 0.71. The benefits of EcPoint's post-processing are particularly evident here for a higher threshold and at longer lead times, where it demonstrates a better performance with a higher lead time and with curves more shifted to the true positive rate (reduced false alarms).

At the 30 mm threshold, a similar trend is observed. EcPoint still enhances forecast skill by improving discrimination between event and non-event cases particularly with a higher threshold and longer lead times, with AUC values of 0.73 (24 h), 0.76 (120 h), and 0.75 (240 h), compared to the Raw Ensemble's 0.69 (24 h), 0.75 (120 h), and 0.72 (240 h). For the 50 mm threshold, the overall AUC values are lower than those for the 30 mm threshold, reflecting the increasing difficulty in predicting more extreme rainfall events, especially for EcPoint which exhibits a uniform performance for all the lead times while Raw ensembles improve comparatively. The AUC values for EcPoint are 0.70 (24 h), 0.70 (120 h), and 0.70 (240 h), while the Raw Ensemble values are 0.64 (24 h), 0.76 (120 h), and 0.75 (240 h). Notably, the Raw ensemble curve has higher AUC values than EcPoint particularly for higher thresholds, meaning that the Raw ensemble has comparable skill at medium-range forecasts, although it still lacks the added precision that EcPoint provides in the tails of the distribution. Despite having the higher AUC values, the shape of the ROC curves reveals important differences in their behavior. EcPoint demonstrates lower false positive rates, which means it is more cautious and precise in identifying rainfall events, which could be attributed to its post-processing

system, which accounts for sub-grid variability and reduces unnecessary event predictions, while the Raw ensemble tends to demonstrate more false positive rates implying a tendency of overpredicting the event. While the benefits of EcPoint diminish slightly at the highest threshold, it still remains the dependable forecasting system, particularly for these nighttime scenarios studied at the short, medium and longer lead times. Generally, it is observed that both EcPoint and the Raw ensemble do achieve high true positive rate, but the Raw ensemble does so at the cost of a high false positive rate too. The findings underscore.

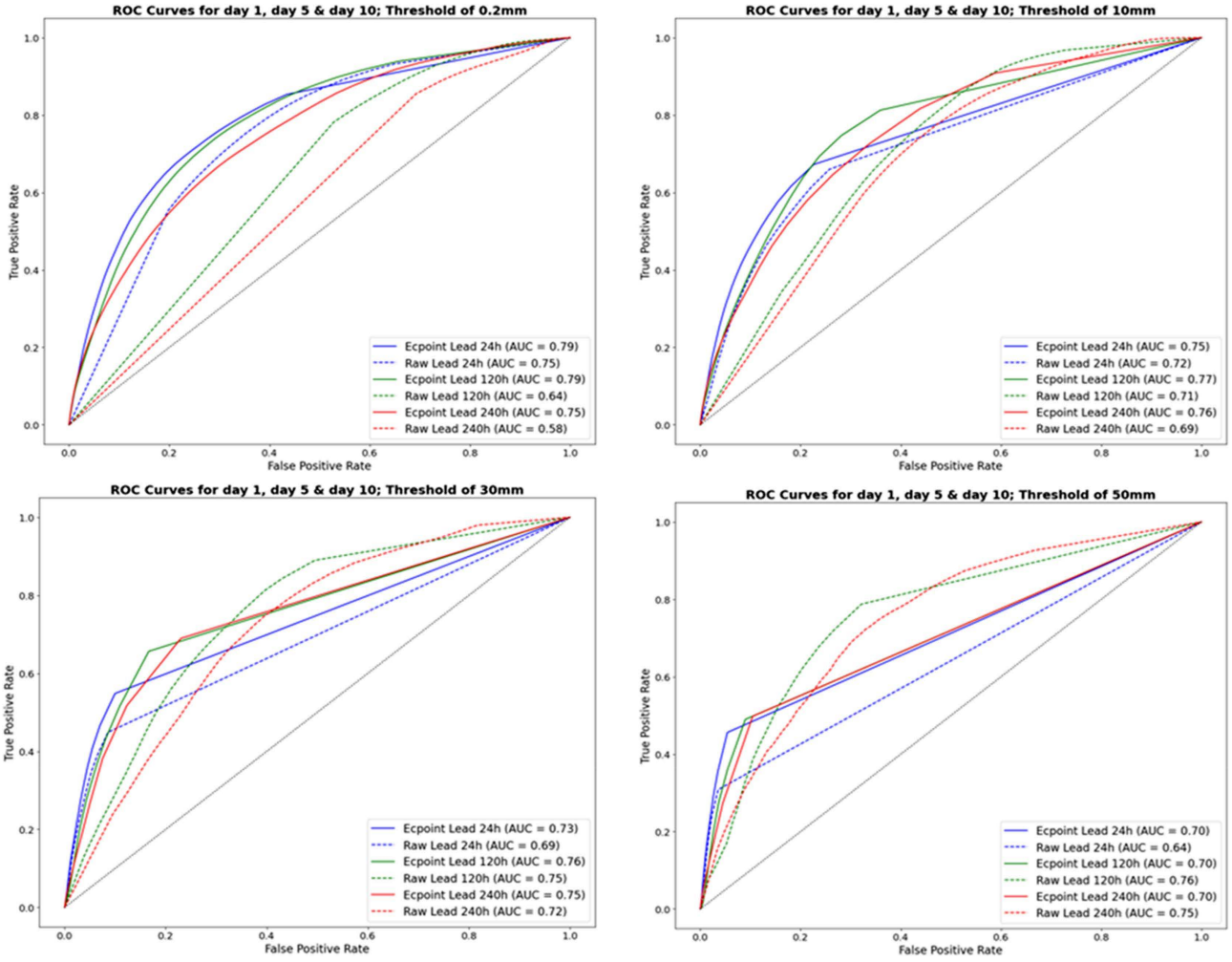

Figure 6. ROC curves considered at three specific days for the four different thresholds.

CRPS Analysis

The CRPS (Continuous Ranked Probability Score) analysis for one-year verification reveals distinct differences between the performance of EcPoint and the Raw ensemble forecasts across lead times (**Figure 7**). The CRPS values for the Raw ensemble consistently increase with lead time, while EcPoint demonstrates a nearly flat CRPS trend across all lead times, with values consistently very low as lead

times extend even up to day 10. The ability to maintain low CRPS values throughout the forecast period underscores the robustness of EcPoint's post-processing system in delivering reliable precipitation forecasts. However, the higher CRPS values observed for the Raw ensemble forecasts could be partly attributed to their reliance on grid-averaged rainfall estimates. This grid-averaging approach inherently smooths localized variations, which are often critical for accurate and reliable forecasts of extreme weather events. Consequently, the Raw ensembles may appear more penalized in terms of CRPS, not necessarily because they lack accuracy or reliability, but because the metric evaluates discrepancies over the full distribution. EcPoint, on the other hand, employs a post-processing methodology that disaggregates grid-averaged forecasts to better represent localized point rainfall events. This approach aligns more closely with observations at specific locations, leading to consistently lower CRPS values. While this suggests better performance, it also reflects EcPoint's ability to resolve fine-scale features that are diluted in grid-averaged forecasts. By recognizing this dynamic, the analysis underscores the value of combining grid-based and post-processed forecast systems. While it does not necessarily provide insight into the exact amount of rainfall or specific thresholds for its accuracy, CRPS is actually indicating that EcPoint excels at forecasting rainfall totals at observation point locations.

The CRPS analysis was thereafter applied to the daily CRPS trends for the 12-hour lead time and the 24-hour lead time, which shows noticeable variability in performance between EcPoint and the Raw ensemble forecasts throughout the year (**Figure 8** & **Figure 9**). Both models generally exhibit low CRPS values on most days, indicating relatively accurate short-term predictions. However, there are significant spikes in CRPS values on certain days, suggesting episodes where forecast errors are higher, likely associated with challenging weather conditions. Notably, the Raw ensemble displays frequent spikes in CRPS throughout the year, reflecting its sensitivity to forecast uncertainties and error propagation, while EcPoint exhibits a trend with fewer and less pronounced spikes. This pattern is particularly prominent during certain clusters of days for both lead times. The days with the highest peak of CRPS (common to both systems) are observed to mostly fall in the seasons with pronounced days of heavy rainfall events. Therefore, High CRPS values seem to align with days featuring heavy rainfall with associated severe weather phenomena or challenging weather conditions, which often involve small-scale or rapidly evolving processes that are difficult for models to capture accurately. Particularly on these days of the highest peak seen in both forecast systems for both lead times, EcPoint generally shows lower values compared to Raw forecasts, indicating its ability to mitigate the uncertainties, though it cannot fully eliminate errors under the most challenging conditions like the ones observed in the days in question.

The histograms in **Figure 10** illustrate the distribution of CRPS values for both Raw and EcPoint forecasts at lead times of 12 hours and 24 hours. At both lead times for both Raw and EcPoint, all histograms show a highly skewed distribution

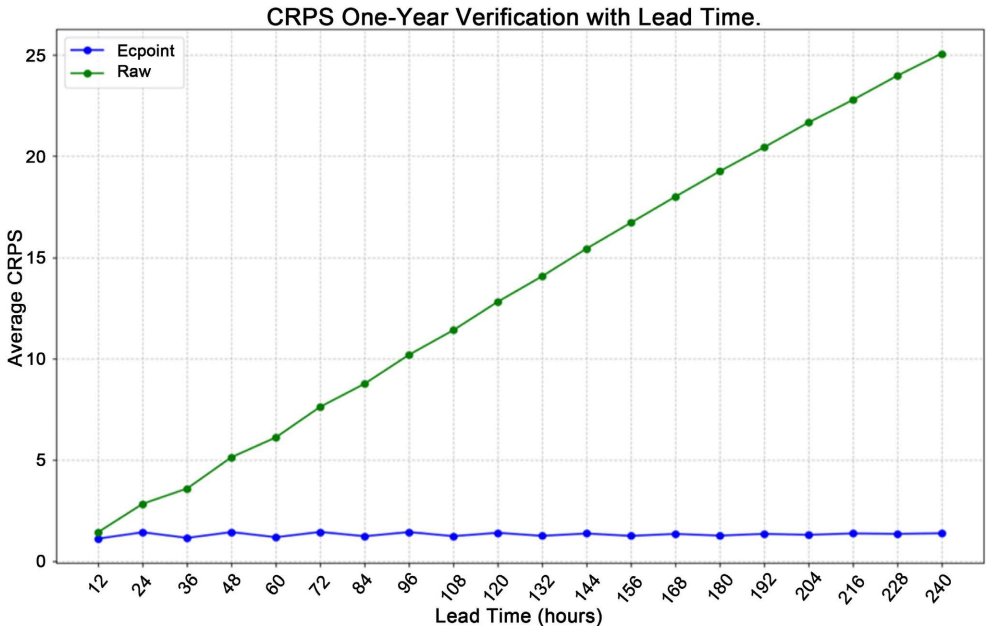

Figure 7. Average CRPS against lead time aggregated over all stations and days for the whole one year.

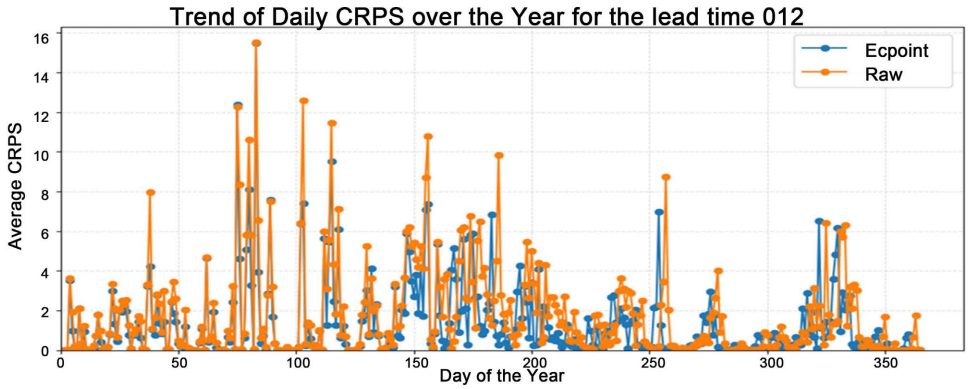

Figure 8. The trend of daily average CRPS aggregated for all stations for the lead time of 12 h at daytime (12UTC).

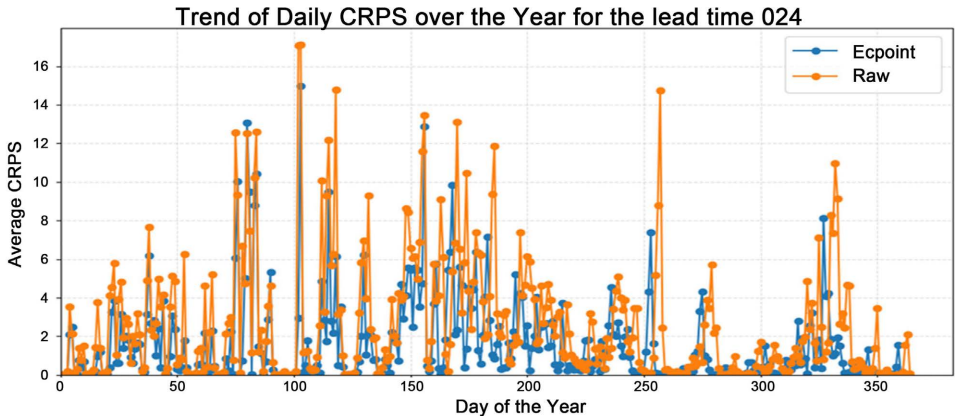

Figure 9. The trend of daily average CRPS aggregated for all stations for the lead time of 24 h at nighttime (00 UTC).

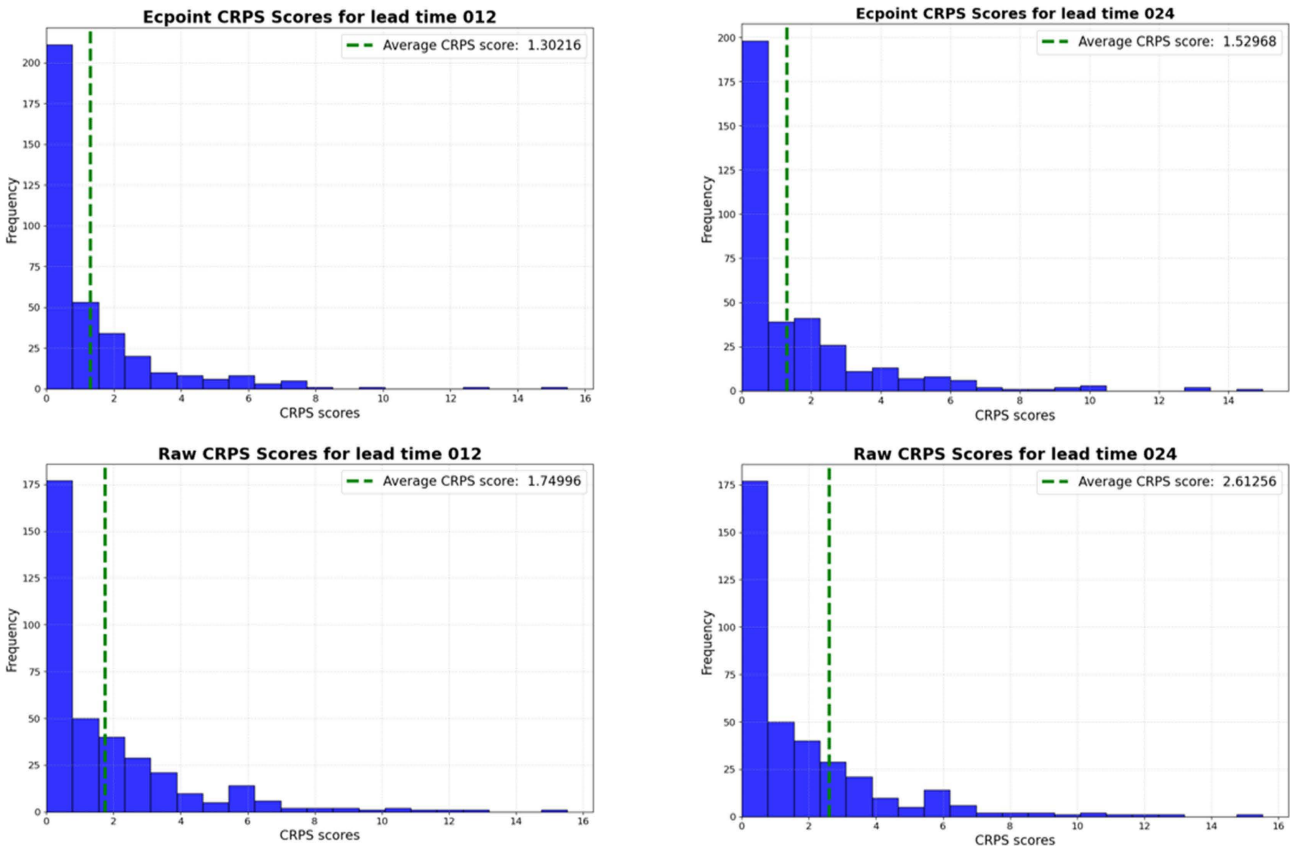

Figure 10. The frequency distribution of the daily CRPS scores over the year, with their annual CRPS average calculated for both ensembles forecast systems (EcPoint and Raw) at the two lead times.

with a significant concentration of CRPS scores near zero (mostly clustered below 2), indicating better forecast performance for the majority of the days. The distributions also exhibit long tails with some scores extending towards higher CRPS values, which corresponds to outliers or challenging forecast conditions. The average CRPS for Raw forecasts increases from 1.75 at 12 hours to 2.61 at 24 hours, indicating a degradation in forecast skill with increasing lead time, whereas EcPoint forecasts increase slightly from 1.30 at 12 hours to 1.52 at 24 hours, demonstrating better performance compared to Raw forecasts as lead time increases. More so for EcPoint, the frequency of lower CRPS scores is higher, with fewer cases in the tail region than in the raw ensembles.

Seasonal Breakdown

The seasonal ROC analysis was conducted using a threshold of 20 mm/12 hours, acknowledging that the event count naturally diminishes when subdividing data into seasonal subsets. The seasonal evaluation of AUC values against lead time provided key insights into the comparative performance of EcPoint and Raw Ensemble forecasts. Overall, EcPoint consistently demonstrates superior skill, particularly at shorter lead times. The ability of EcPoint to refine predictive accuracy across different seasons suggests its effectiveness in capturing rainfall variability, offering improved discrimination compared to Raw Ensembles.

In Figure 11, the summer (JJA) season reflects the peak of diurnally-driven convective activity, leading to greater variability in forecast performance. Both EcPoint and Raw Ensembles exhibit fluctuations in AUC values, which can be attributed to the chaotic nature of convective precipitation. Notably, a sharp diurnal cycle in AUC with lead time is observed for both forecasting systems, highlighting the influence of convection-driven precipitation patterns. Also, the higher number of observed rainfall cases during summer could contribute to the increased variability in performance with more events available for evaluation. Strong performance of EcPoint also holds for discrimination ability particularly for Autumn (SON) and Spring seasons (MAM). This is attributed to the fact that rainfall systems are generally less challenging to predict in these seasons due to their well-defined dynamics, which evolve over larger spatial and temporal scales. On the other hand, summer is influenced by strong diurnal heating characterized by localized instability, short duration heavy rainfall convective in nature thus introducing significant unpredictability. Also, for Autumn (SON), the gap between EcPoint and Raw Ensemble becomes more evident at longer lead times, reinforcing the value of EcPoint in extended forecasts. This season shows relatively stable AUC values, indicating consistent forecast skill over time. For Winter (DJF), an interesting pattern emerges where Raw Ensemble performs comparably to EcPoint at longer lead times, with AUC values converging after 150 hours. However, at shorter lead times, EcPoint still holds an advantage. A key reason for the performance in winter could be the lower frequency of the rainfall events (≥20 mm/12 h) during winter. With fewer heavy rainfall cases, the ability to statistically discriminate between event and non-event occurrences decreases, leading to greater variability in AUC values. Additionally, the scarcity of such events reduces the opportunity for EcPoint's post-processing to make significant corrections, resulting in a smaller performance gap between the two forecasting systems. The dominance of light-to-moderate precipitation in winter further limits the impact of the chosen threshold on ROC analysis, as there might not be as many cases of 20 mm as there may be for other seasons. This rarity introduces uncertainty in AUC calculations, causing fluctuations and potentially masking EcPoint's advantages over Raw Ensembles. Consequently, while EcPoint still offers improvements, its benefits appear less pronounced in winter due to the limited number of high-intensity events available for evaluation.

The CRPS seasonal verification analysis reveals that Generally, the Raw Ensemble forecasts exhibit increasing CRPS values with lead time, indicating a degradation in forecast skill over time, while EcPoint demonstrates a significant reduction in CRPS across all seasons highlighting the benefit of post-processing in improving forecast accuracy and enhances forecast reliability. Among the seasons, summer (JJA) shows the highest CRPS values, suggesting greater forecast uncertainty, likely due to the convective and localized nature of summer rainfall events. In contrast, autumn (SON) and winter (DJF) exhibit lower CRPS values, reflecting the more predictable nature of large-scale weather systems characteristic of these seasons (Figure 12).

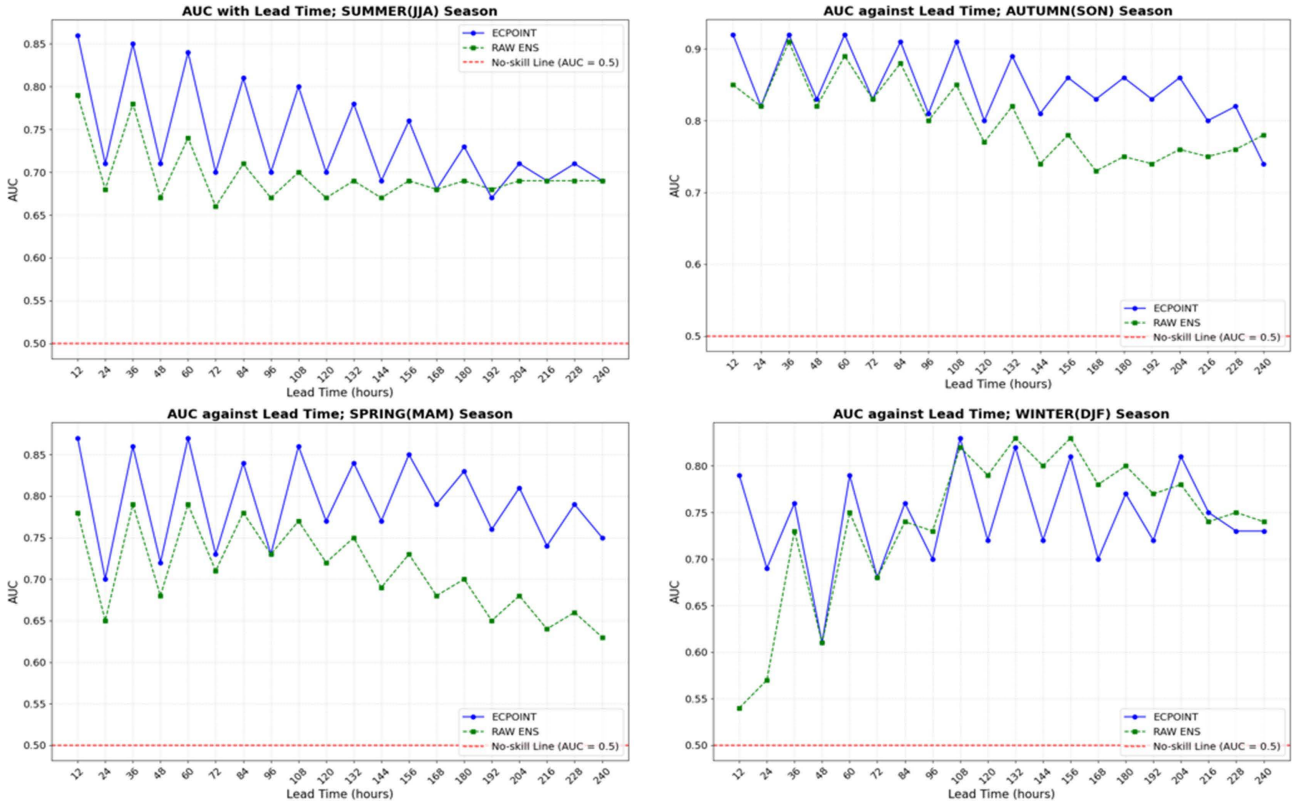

Figure 11. The variation of area under the curve (AUC) with lead time of up to 10 days for the four different seasons of the year.

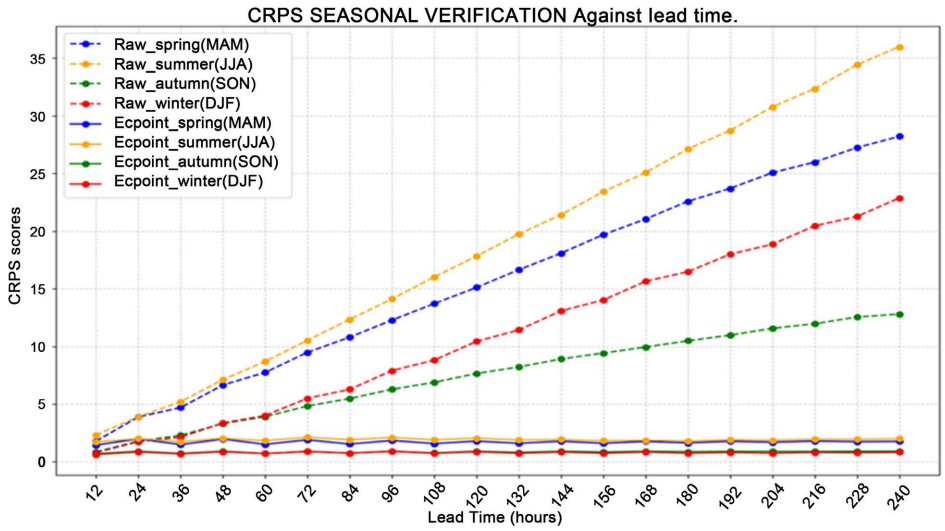

Figure 12. CRPS analysis for the seasons of both Raw ensembles (dashed lines) and EcPoint (solid lines) versus lead time of up to 10 days.

The superior performance of EcPoint compared to the raw ECMWF ensemble can be attributed to its ability to correct systematic biases, better represent sub-grid variability, and categorize rainfall forecasts based on distinct weather regimes. While raw ensemble forecasts provide grid-scale averages, EcPoint refines these estimates by associating each forecast grid box with a probabilistic distribu-

tion based on observed errors in similar meteorological conditions. This is particularly advantageous for convective and orographic rainfall, where raw ensembles often struggle due to resolution limitations. Additionally, EcPoint's calibration process is physics-based rather than location-dependent, allowing it to improve forecast skills even in regions with limited observational data. By enhancing probabilistic forecasts for both typical and extreme rainfall events, EcPoint significantly increases forecast reliability, as has been demonstrated by the CRPS and skill by ROC.

Topographic Analysis

In analyzing topography, altitude ranges were categorized into three groups: plains/low hills (altitude < 50 m), very hilly (50 m < altitude < 100 m), and mountainous (altitude > 100 m), as shown in **Figure 13**. These categories were not arbitrarily chosen; rather, they were based on prior research and relevant studies. The selection follows the approach of [9], who aligned their classification with breakpoints identified using decision tree calibration, with some rounding and normalization applied. The topography assessment evaluates the ability to discriminate rainfall events using ROC and AUC, focusing on a relatively high threshold of 20 mm/12 h. Additionally, it examines the reliability and accuracy of forecast systems through the CRPS analysis. Both CRPS and AUC results indicate that EcPoint consistently outperforms Raw Ensembles across various lead times for all altitude categories, providing more reliable forecasts with lower CRPS scores. For stations at lower altitudes (<50 m, orange lines), as shown in **Figure 14**, both Raw and EcPoint systems perform similarly well at the shortest lead times. However, over time, EcPoint demonstrates superior performance as it maintains a flat curve, while the Raw system exhibits a continuous gradual increase of CRPS scores. Conversely, CRPS scores increase for both Raw and EcPoint systems at higher altitudes, with particularly high CRPS values observed for stations above 100 m (green lines).

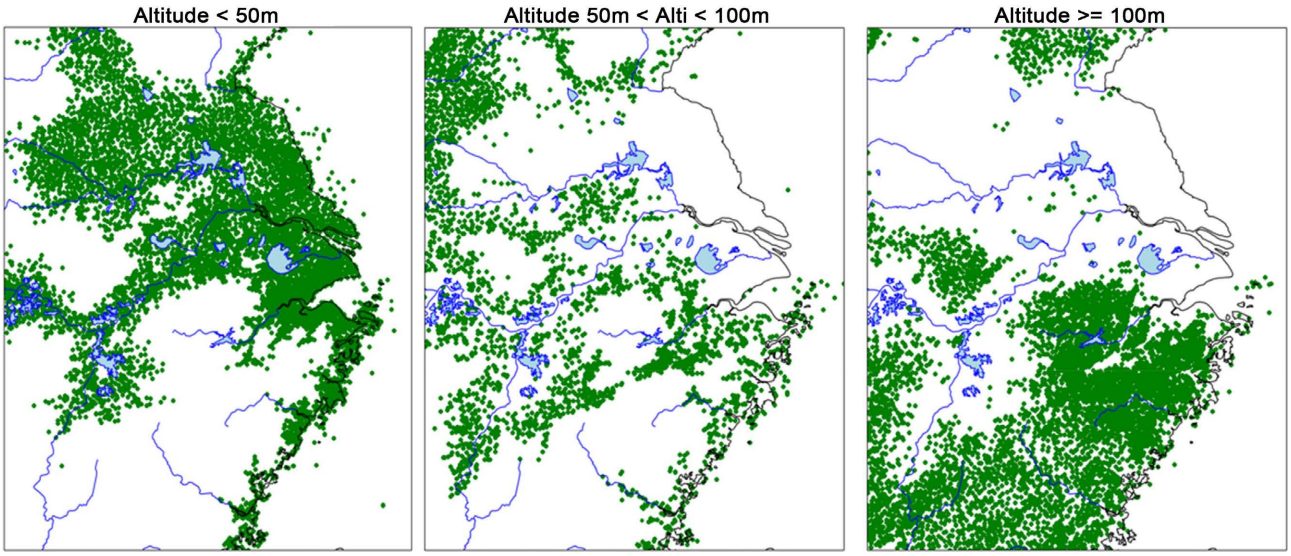

Figure 13. Observation points available for the three altitude ranges.

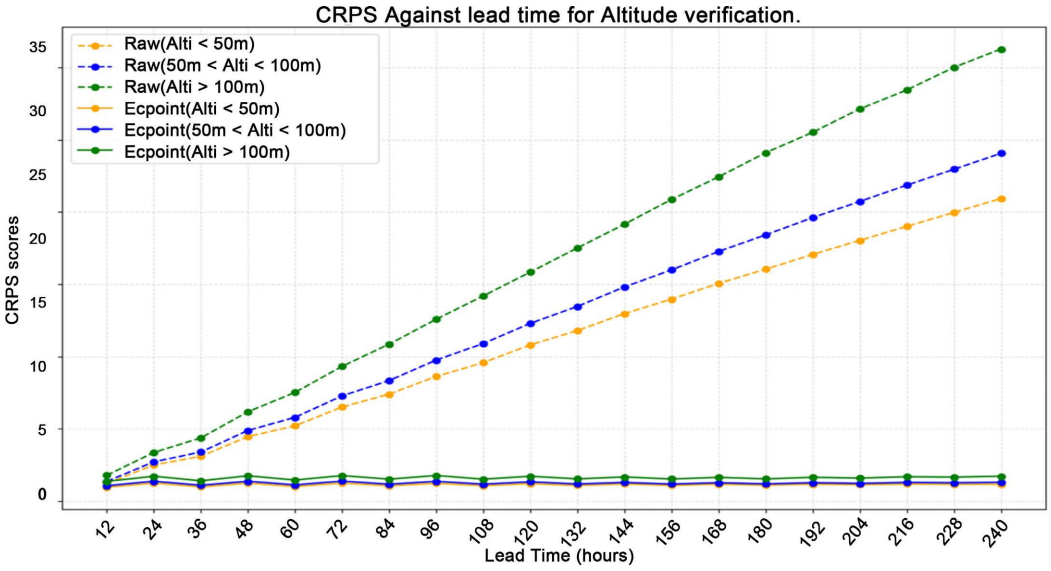

Figure 14. CRPS versus lead time (up to 10 days) for the altitude-based performance comparison of the two forecasting systems illustrating relative accuracy over time across different altitude ranges.

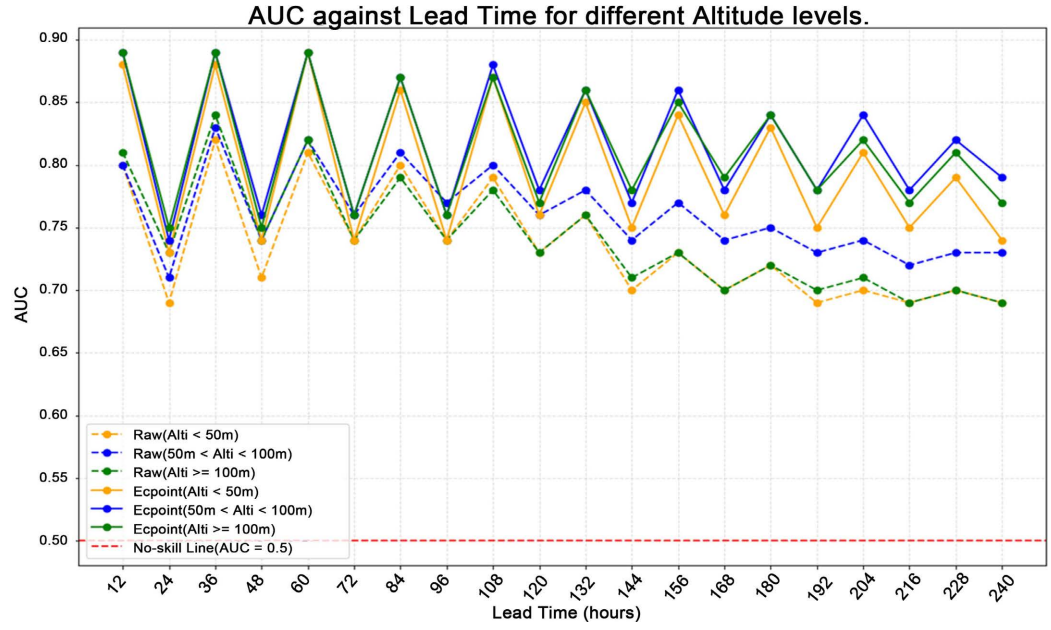

Figure 15. Area Under the Curve (AUC) versus lead time (up to 10 days) for the altitude-based performance comparison of the two forecasting systems illustrating how predictive skill varies over time across different altitude ranges.

For the ability to discriminate between two events (AUC), EcPoint seems, encouragingly, to add the most value to Raw Ensembles in mountainous regions. See the blue curves in **Figure 15** compared to the orange ones at a lower altitude. It may be because there is more discrimination ability added in mountainous areas, and because EcPoint post-processing is targeting this effectively via its decision tree subdivisions as was noted by [9]. However, despite its effectiveness in event discrimination, EcPoint is less effective at improving reliability in mountainous

areas, as observed in CRPS results. At shorter lead times, the performance differences between the two systems are relatively minor. However, as lead time increases, EcPoint's performance remains consistent, whereas Raw Ensembles show greater variability, particularly at lower altitudes (<50 m), where AUC values decline significantly. The performance gap between Raw and EcPoint widens at higher altitudes and higher lead times, with EcPoint maintaining consistently higher AUC values. At longer lead times (beyond 120 hours), AUC values for both systems tend to stabilize, reflecting the inherent difficulty of long-term weather prediction. However, performance degradation occurs across all terrain types and forecast systems, with Raw Ensembles experiencing a sharper decline in AUC values. This trend highlights the challenge of maintaining forecast skill over extended horizons and underscores the need to address the pronounced performance drop especially for Raw Ensembles at longer lead times.

Case Study

We present a study case with the highest recording of the 12 hours of rainfall accumulation during the year 2022 for the three provinces, observed on June 02, 2022 (12 UTC) at station J5269 (latitude: 28.6731, longitude: 115.3064, altitude: 67.0 m). On this day, an exceptional rainfall amount of 457.2 mm was recorded (suggesting a significant weather event like a heavy rainfall storm or a monsoon event), subjective to flooding possibility, given the topography of the area. Therefore, to complement the one-year verification analysis, this specific day was selected for case study to further demonstrate the performance of the two forecasting systems in representing the spatial distribution of rainfall over the region. In Figure 16, the 98th percentile reveals key differences in the predictive capabilities of the two forecasting systems, at varying lead times. At short lead times (12 hours), EcPoint presents a refined and localized prediction, especially in the region where significant rainfall is observed compared to Raw ensembles. As the lead time increases, both models show clearer and more intense rainfall signals, but the raw ensemble exhibits broader and more variable patterns, while EcPoint provides more focused, high-confidence predictions. The results highlight the advantage of EcPoint's refined ensemble processing in providing more reliable and regionally focused rainfall forecasts, particularly for extreme events (like the 98th percentile), while the raw ensemble, although useful, shows greater uncertainty and less specificity at both short and longer lead times.

Figure 17 shows the probability of exceeding a threshold of 50 mm/12 h for both forecasting systems. The EcPoint probability for exceeding such an extreme rainfall amount of 50 mm/12 h exhibits a consistent, focused and refined spatial distribution of rainfall probabilities for all the lead times. The Raw ensemble probabilities for exceeding 50 mm/12 h display a broader and less precise spatial extent of high probabilities. While the Raw ensemble successfully indicates the high potential for significant rainfall, the ensembles fail to capture rainfall distribution at the first lead time (short lead time) compared to EcPoint, which shows the chances of rainfall on this short lead time. Convective rainfall, which is typically short-lived and localized, is challenging for global models to predict at fine scales, and

the model may fail to resolve these features accurately. At the short lead time (012 hours), the Raw Ensemble forecast shows negligible or no captured rainfall, likely due to its limitations to localized events. In contrast, EcPoint (post processed ensemble forecast) captures some rainfall probability at this short lead time, suggesting that the post-processing techniques do improve the model's ability to handle small-scale features and adjust for model biases to capture localized events like convective events or rainfall storms. However, as the forecast lead time increases, both models show a clearer and more consistent prediction of rainfall, with raw ensembles showing higher rainfall estimations. The RAW ENS forecast tends to perform better at longer lead times because it is better at resolving large-scale weather systems. As the forecast progresses to longer lead times, the model averages out smaller, more chaotic, and unpredictable short-term events, which reduces the complexity of the forecast. On the other hand, at shorter lead times, the model struggles with capturing smaller-scale phenomena like thunderstorms or convection, which are more sensitive to local conditions and harder to predict accurately. Therefore, the forecast becomes more reliable at longer lead times due to the nature of broader weather patterns being easier to resolve. The tendency is attributed to the reliance on grid averaged rainfall amounts which inherently tend to smooth out the localized variation meanwhile EcPoint caters for point variations within the grid box. It was not given that the systems would appear physically reasonable and free from unwanted artifacts, even when verification results are positive. However, the credibility of the post processing product is a critical requirement for users and encouragingly our verification results and case studies support the expectation.

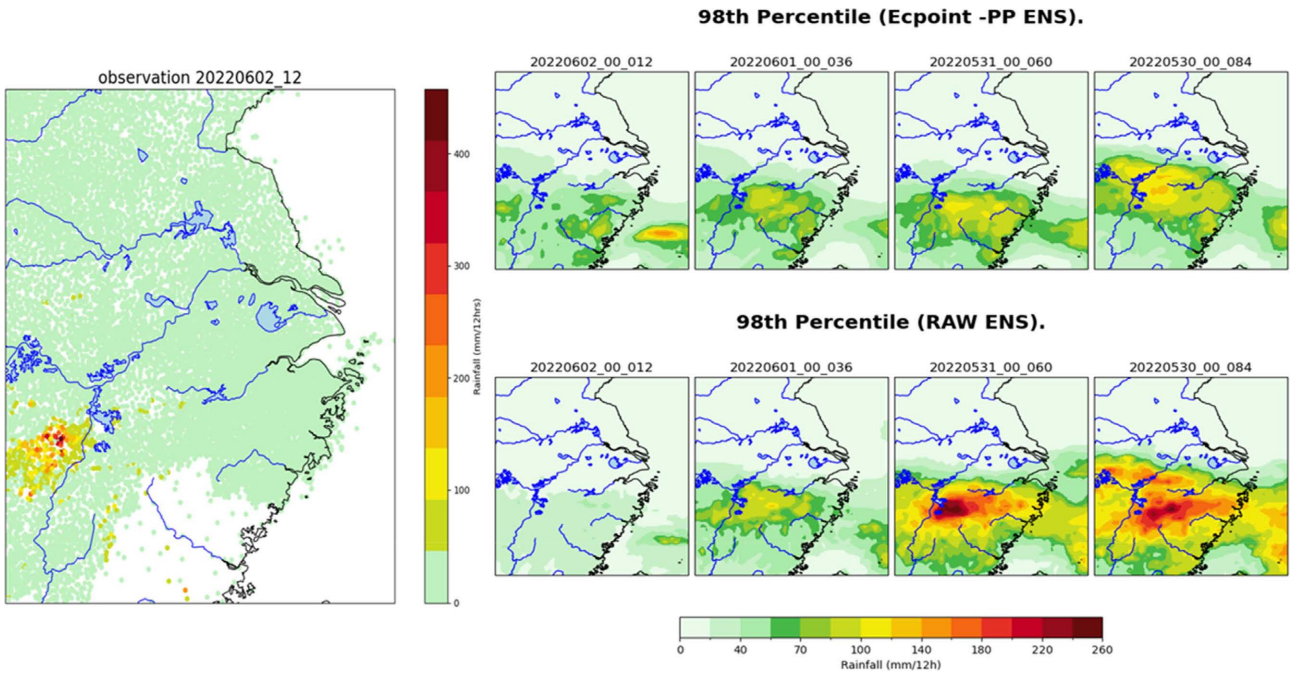

Figure 16. 98th percentiles of 12 h rainfall forecasts for Raw and EcPoint at nominal lead times of 12 h, 36 h, 60 h and 84 h, also the gauge observation for verifying the forecasts on the left.

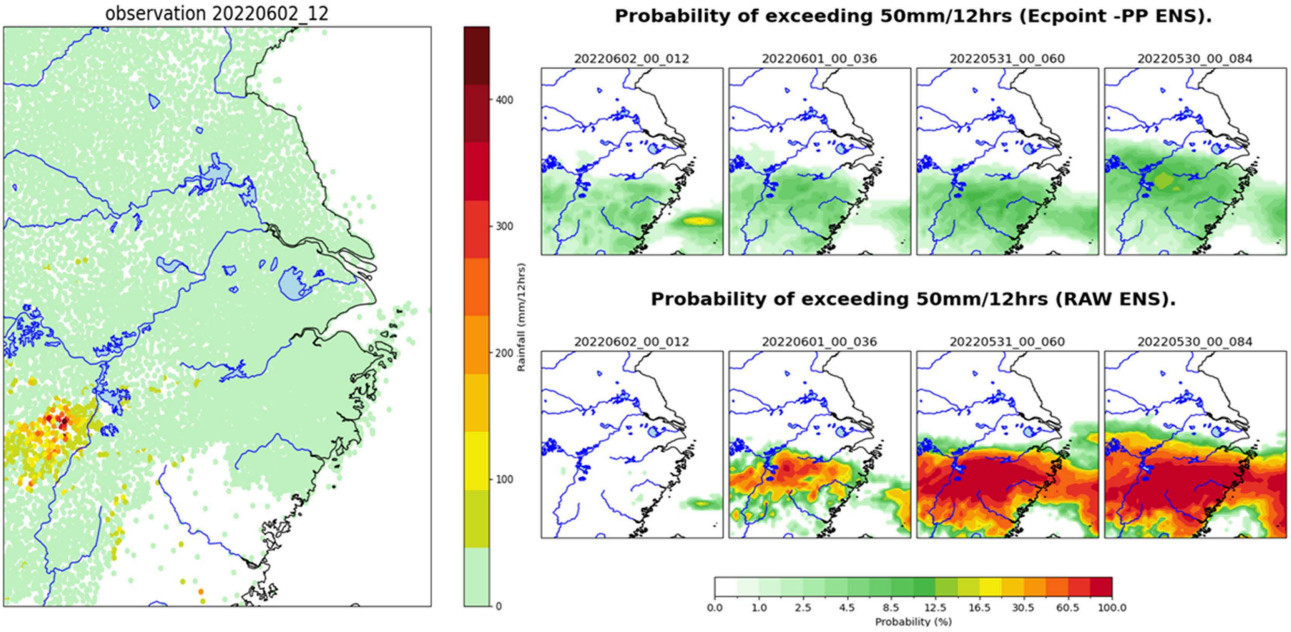

Figure 17. Forecast probabilities for exceeding 50 mm of 12 hours rainfall accumulation at nominal lead times of 12 h, 36 h, 60, and 84 h for post processed ensemble, *i.e.*, EcPoint and Raw Ensembles. Also, observations are on the left for verification.

4. Conclusion

The study highlighted the advancements brought by EcPoint, a state-of-the-art post-processing approach, in improving rainfall ensemble forecasts. The overall aim was to evaluate the EcPoint post-processing method over South China with the study area of three provinces, to determine if it can deliver much greater skill than can be achieved by using the Raw ensembles. By comparing the post-processed ensemble (EcPoint) with the Raw ensemble forecast, we demonstrated that EcPoint consistently outperforms the Raw ensemble in terms of discrimination and reliability for most rainfall thresholds and lead times. These improvements were more noticeable for low thresholds, such as 0.2 mm, and at shorter lead times. However, the differences were also pronounced for heavier rainfall thresholds like for a threshold of 30 mm particularly at long lead times. A key strength of EcPoint lies in its ability to balance true and false positive rates, especially at shorter lead times (e.g., at Day 1), where its post-processed outputs show clear added value in discriminating events. At longer lead times (e.g., Day 10), the gap between EcPoint and Raw Ensemble narrows due to increasing uncertainties in both systems. Nonetheless, EcPoint remains superior in reducing false alarms and maintaining better forecast skills, demonstrating its reliability in the face of inherent forecast uncertainties. The Raw ensemble's higher false positive rates during nighttime are likely exacerbated by the absence of diurnal forcing, while EcPoint's statistical adjustments help compensate for such biases, resulting in more dependable predictions even during nighttime hours. Combining CRPS results with threshold-based evaluation confirms EcPoint's overall improvement in forecast quality, particularly pronounced for specified rainfall intensities, reinforcing its value as a tool for high-impact weather forecasting.

Seasonal analysis revealed a strong diurnal cycle in the performance of both ensemble products during summer, with both systems showing weaker performance in this season. This decline is attributed to the chaotic and convective nature of summer rainfall. This underscores the importance of refining post-processing techniques to better handle weather patterns in specific seasons, particularly summer. The topography analysis highlighted the impact of elevation on forecast performance, with EcPoint consistently outperforming the Raw ensemble across different terrain types. In lower regions, both ensemble systems exhibit higher skill, while in complex terrain, EcPoint demonstrates an improvement by better capturing localized rainfall variability. This added value is particularly evident at shorter lead times, reinforcing the importance of post-processing in improving forecasts over diverse landscapes. While the results demonstrate strong performance in the region of south China, further validation in more topographically diverse regions would be essential to assess the broader applicability and generalize these findings. The case study demonstrated EcPoint's advantage in predicting extreme rainfall events, particularly at short lead times, where it provides more localized and refined forecasts compared to the Raw ensemble. The 98th percentile analysis reveals that EcPoint captures significant rainfall patterns more precisely, while the Raw ensemble shows broader and less focused distributions. For the probability of exceeding 50 mm/12 h, EcPoint provided a more concentrated and reliable spatial distribution, whereas the Raw ensemble exhibits more spread-out and uncertain probabilities. While the Raw ensemble improves at longer lead times by better capturing large-scale weather patterns, it does so with several false events, as observed in the ROC analysis, potentially over-warning for areas not impacted. A key reason why EcPoint outperforms the Raw ensemble is its ability to better represent sub-grid variability and correct systematic biases in rainfall forecasts. Raw ensembles often struggle to resolve localized convective rainfall due to their coarser resolution, leading to underestimation of extreme events. EcPoint addresses this limitation by categorizing forecasts based on different weather regimes and applying post-processing corrections to refine probabilistic rainfall distributions. This improvement is particularly valuable in complex terrain, where orographic effects play a crucial role in precipitation formation. Additionally, EcPoint's calibration process is independent of specific locations, allowing it to be applied effectively even in regions with limited observational data. By leveraging physics-based adjustments and probabilistic refinements, EcPoint significantly enhances forecast accuracy, particularly for short lead times and high-impact rainfall events. Conclusively, EcPoint represents a crucial step forward in addressing the challenges of rainfall forecasting in complex and dynamic weather systems. Despite being based solely on the global ECMWF-IFS model, EcPoint performs well over the small domain of South China (three provinces: Anhui, Zhejiang, and Jiangsu). Future studies should explore the EcPoint calibration software, as calibrating EcPoint with the region's local observations would likely enhance its performance by better capturing regional rainfall characteristics and reducing model biases.

Acknowledgements

The authors gratefully acknowledge the European Center for Medium-range Weather Forecasts and Fatima Pillow for the provision of forecast datasets that were used in this study. We appreciate our supervisor Wang Yong for the support and guidance rendered throughout the whole study process. The authors furthermore thank Fatima Pillow for the detailed and valuable discussions and the very important comments offered during the study process.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Chan, J.C., *et al.* (2008) Accurate Predictions of Heavy Rainfall for Impact-Based Forecasting. *Journal of Meteorological Research*, **22**, 456-470.

[2] Chen, X., *et al.* (2011) Challenges in Ensemble Rainfall Forecasting for Decision-Making. *Weather and Forecasting*, **26**, 789-801.

[3] Rodwell, M.J. (1998) Representation of Topography in Coarse-Scale Models. *Quarterly Journal of the Royal Meteorological Society*, **124**, 1234-1256.

[4] von Storch, H., *et al.* (1993) Limitations of GCMs in Predicting Local Rainfall. *Journal of Climate*, **6**, 789-801.

[5] Hewson, T.D. and Pillosu, F.M. (2021) A Low-Cost Post-Processing Technique Improves Weather Forecasts around the World. *Communications Earth & Environment*, **2**, Article No. 132. https://doi.org/10.1038/s43247-021-00185-9

[6] Robertson, D.E., Shrestha, D.L. and Wang, Q.J. (2013) Post-Processing Rainfall Forecasts from Numerical Weather Prediction Models for Short-Term Streamflow Forecasting. *Hydrology and Earth System Sciences*, **17**, 3587-3603. https://doi.org/10.5194/hess-17-3587-2013

[7] Gascón, E., Montani, A. and Hewson, T.D. (2023) EcPoint Rainfall: A Conditional Verification Approach for Ensemble Rainfall Forecasts. *Journal of Hydrometeorology*, **24**, 1234-1256.

[8] Owens, R.G. and Hewson, T.D. (2018) ECMWF's Point Forecast Product: An Overview. *ECMWF Technical Memoranda*, **789**, 1-25.

[9] Gascón, E., Montani, A. and Hewson, T.D. (2024) Post-Processing Output from Ensembles with and without Parametrised Convection, to Create Accurate, Blended, High-fidelity Rainfall Forecasts. *Quarterly Journal of the Royal Meteorological Society*, **150**, 3117-3145. https://doi.org/10.1002/qj.4753

[10] Mátrai, T. and Ihász, I. (2017) Statistical Post-Processing of Ensemble Forecasts for Point Rainfall. *Journal of Hydrology*, **550**, 123-135.

[11] Ihász, I., *et al.* (2018) Improving Point Rainfall Forecasts Using Post-Processing Techniques. *Meteorological Applications*, **25**, 456-470.

[12] Hewson, T.D., *et al.* (2019) Flash Flood Prediction Using Post-Processed Ensemble Forecasts. *Natural Hazards and Earth System Sciences*, **19**, 1234-1256.

[13] Pillosu, F.M. and Hewson, T.D. (2017) EcPoint-Calibrate: A Tool for Post-Processing Ensemble Forecasts. *ECMWF Technical Memoranda*, **789**, 1-25.

[14] Hewson, T.D. and Pillosu, F.M. (2020) Blending Information for Reliable Probabilis-

tic Rainfall Forecasts across Diverse Regions. *Journal of Applied Meteorology and Climatology*, **59**, 789-801.

[15] Wu, R. and Wang, B. (2000) Interannual Variability of Summer Monsoon Onset over the Western North Pacific and the Underlying Processes. *Journal of Climate*, **13**, 2483-2501. https://doi.org/10.1175/1520-0442(2000)013<2483:ivosmo>2.0.co;2

[16] Wang, B., *et al.* (2001) Fundamental Challenge in Simulation and Prediction of Summer Monsoon Rainfall. *Geophysical Research Letters*, **28**, 4537-4540.

[17] Shok, N., *et al.* (2004) The Role of the Pacific Decadal Oscillation in modulating South China Rainfall. *Journal of Climate*, **17**, 2345-2356.

[18] Wang, B. and Guan, Z. (2007) Interdecadal Variability of the Asian Monsoon and Its Impact on Regional Climate. *Climate Dynamics*, **29**, 789-801.

[19] Zhao, P., *et al.* (2011) Influence of the Pacific Decadal Oscillation on South China Early Summer Rainfall. *Journal of Geophysical Research: Atmospheres*, **116**, D12304.

[20] Saji, N.H., Goswami, B.N., Vinayachandran, P.N. and Yamagata, T. (1999) A Dipole Mode in the Tropical Indian Ocean. *Nature*, **401**, 360-363. https://doi.org/10.1038/43854

[21] Mao, J., *et al.* (2011) Interannual Variability of Winter and Spring Precipitation in South China. *Journal of Climate*, **24**, 3097-3112.

[22] Luo, Y., Xia, R. and Chan, J.C.L. (2020) Characteristics, Physical Mechanisms, and Prediction of Pre-Summer Rainfall over South China: Research Progress during 2008-2019. *Journal of the Meteorological Society of Japan. Ser. II*, **98**, 19-42. https://doi.org/10.2151/jmsj.2020-002

[23] Matheson, J.E. and Winkler, R.L. (1976) Scoring Rules for Continuous Probability Distributions. *Management Science*, **22**, 1087-1096. https://doi.org/10.1287/mnsc.22.10.1087

[24] Richardson, D.S. (2003) Skill and Relative Economic Value of the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, **129**, 727-742.