

Precipitation Forecasting for Hydrologic Modeling in West-Central Florida using Seasonal Climate Outlooks

Manoj Shrestha¹, Hui Wang², Jeffrey S. Geurink², Kshitij Parajuli², Tirusew Asefa², Fanzhang Zeng¹, and Dingbao Wang¹

¹Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, Florida, USA

²Tampa Bay Water, 2575 Enterprise Rd., Clearwater, Florida, USA

*Correspondence to: Hui Wang(hwang@tampabaywater.org)

Abstract: Seasonal precipitation forecasts play a vital role in short-term decision-making for water resources management, agriculture, and wildfire preparedness. NOAA’s seasonal precipitation forecasts can be used at the local scale to further develop precipitation forecasts. Rather than evaluating the forecasting skill at the scale at which forecasts are provided, this study applies NOAA forecasts at the local basin scale and evaluates the skill of such localized forecasts. This study evaluates the skill of NOAA’s 3-month precipitation outlooks at a 0.5-month lead for the Alafia and Hillsborough River Basins in west-central Florida, using hindcasts from 1995 to 2019. Forecast performance is assessed seasonally using categorical and probabilistic metrics [against basin-scale observed precipitation from Bayesian gauge-radar dataset](#). To translate categorical outlooks into basin-scale rainfall estimates, two non-parametric ensemble generation methods are introduced: Proportional Tercile Sampling (PTS) and Dominant Tercile Sampling (DTS). These methods sample from pre-generated rainfall realizations conditioned on seasonal forecasts to capture uncertainty and support operational planning. Results indicate that forecast skill peaks during the dry season (October to February), particularly for wet-tercile forecasts issued during El Niño years. DTS performs best during high-skill seasons by leveraging dominant climate signals, while PTS proves more reliable during low-skill periods. Based on these findings, a hybrid strategy is recommended: apply DTS during late fall and winter to capitalize on strong climate signals and use PTS during other seasons to maintain reliability and operational value. This study contributes a strategic approach to applying NOAA’s forecasts in the study area and demonstrates that this method of applying NOAA’s forecasts at the local scale is general and can be applied to other regions.

1. Introduction

Seasonal precipitation forecasts play a critical role in decision making across water resources management, agriculture, and disaster preparedness. By offering lead times of several months, these forecasts help mitigate risks and support proactive planning. Numerous studies have demonstrated their value across diverse sectors. For example, Portele et al. (2021) and MacLeod et al. (2023) applied seasonal climate forecasts to drought management, while Bett et al. (2018) explored their use in flood control. Alexander and Block (2022) and Garcia et al. (2025) highlighted their relevance to agricultural planning.

30 Hurkmans et al. (2023) and Graham et al. (2021) investigated their utility in reservoir operations, and Turco et al. (2019) examined their application in assessing fire risk.

Over the past century, seasonal climate forecasting has evolved through successive waves of research and operational advancement. Namias (1968) laid the groundwork by synthesizing long-range forecasting methods, including analog techniques and emerging statistical approaches. Building on this foundation, Nicholls (1980) classified a broad spectrum of
35 methodologies such as analog, teleconnection, cosmic cycle, time series, and early numerical models, highlighting the diversity of forecasting strategies. Barnett and Preisendorfer (1987) advanced the field by applying canonical correlation analysis to identify key drivers of seasonal temperature predictability in North America. A major milestone came with the establishment of the National Oceanic and Atmospheric Administration’s (NOAA) Climate Prediction Center (CPC) in 1995, which marked the beginning of modern operational seasonal climate forecasting. Digitized monthly and three-month “tercile” forecasts were
40 officially archived and made publicly accessible (Hartmann et al., 2002), enhancing transparency and usability. Nowadays, NOAA’s CPC issues three-month seasonal precipitation outlooks with a 0.5-month lead time, extending up to 12.5 months into the future.

NOAA’s CPC generates seasonal precipitation forecasts by integrating statistical and dynamic modeling approaches. On the statistical side, techniques such as Canonical Correlation Analysis (CCA) and Screening Multiple Linear Regression (SMLR)
45 establish relationships between global sea surface temperature (SST) indices to precipitation anomalies, effectively capturing teleconnection patterns like the El Niño–Southern Oscillation (ENSO) and the North Atlantic Oscillation (O’Lenic et al., 2008; NOAA Climate Prediction Center, 2021). Dynamic models, such as the Climate Forecast System (CFS) and the North American Multi-Model Ensemble (NMME), simulate coupled ocean-atmosphere processes. CFS initializes oceanic conditions using the Global Ocean Data Assimilation System (GODAS), while NMME improves forecast reliability through multi-model
50 consensus (Kirtman et al., 2014). The NOAA CPC operationally maintains the NMME, a dynamic multimodel ensemble system comprising NCEP-CFSv2, ECCO-CanESM5, ECCO-GEM5.2-NEMO, NCAR-CESM1, NCAR-CCSM4, and NASA-GEOS-S2S-2 (NOAA CPC, n.d.), which provides more skillful probabilistic seasonal forecasts than single-model systems (Kirtman et al., 2014). Evaluation studies have shown that NMME precipitation skill varies substantially by season, region, and lead time, with the multimodel ensemble generally outperforming individual members, and models such as CCSM4 and CFSv2 showing stronger skill in the southeastern USA (Wang, 2014; Slater et al., 2019; Becker & van den Dool, 2016). When raw GCM precipitation fields are used directly rather than probabilistic outlooks, postprocessing such as quantile mapping is often applied to correct systematic biases, though its effectiveness depends on the forecast objective (Zhao et al., 2017).

Forecast skill tends to be highest when statistical and dynamical models agree, particularly during ENSO events. In contrast, when large-scale climate signals are weak, forecast skill often reverts toward climatology (Hartmann et al., 2002). Livezey &
55 Timofeyeva (2008) observed that forecast skill generally declines with increasing lead time, except during ENSO winters when skill remains elevated. These short-term seasonal forecasts provide critical lead time for water resource managers, supporting proactive planning, optimized reservoir operations, and effective drought mitigation strategies (Peng et al., 2012).

60

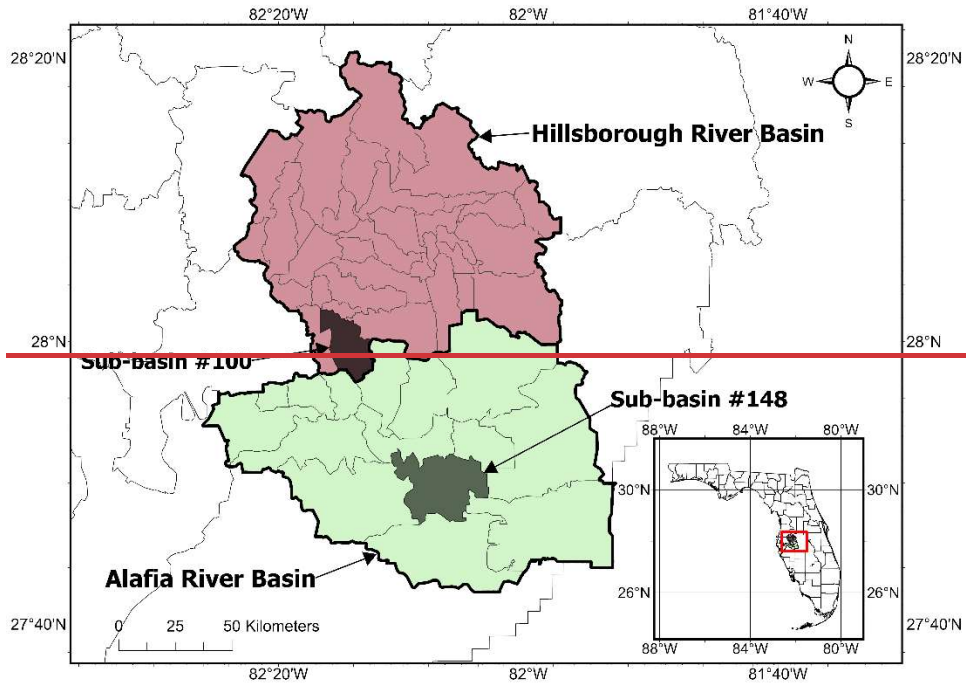
65 Despite advances in seasonal climate forecasting, its application in basin-scale water management remains limited. Water managers continue to evaluate whether these forecasts can meaningfully inform decision-making at the basin level. Tampa Bay Water (TBW), the largest wholesale water provider in the southeastern United States, employs a rainfall–streamflow statistical modeling approach to generate stochastic flow estimates for two key basins that serve as primary surface water sources (Wang et al., 2024). However, the limited skill of seasonal precipitation forecasts has constrained their operational utility in water resources planning and management (Wang et al., 2019). If forecast skill improves, probabilistic information could be ~~integrated into existing hydrological models to anticipate anomalously wet or dry periods and adjust operations~~ proactively used to support proactive water resources planning and management. Wang and Asefa (2018) demonstrated that ENSO phases exert a strong influence on winter rainfall in Florida, highlighting the potential value of seasonal climate outlooks in enhancing regional water management strategies.

70 This study focuses on two subtropical river basins in Southwest-central Florida and pursues two primary objectives. First, it evaluates the categorical and probabilistic skill of CPC’s seasonal precipitation outlooks, with a 0.5-month lead time, over a 25-year period (1995–2019), assessing performance by season (wet vs. dry) and forecast category (above-normal vs. below-normal). Second, it translates CPC’s tercile forecasts into basin-scale precipitation estimates by sampling stochastically generated rainfall realizations within the basins. Two sampling approaches are proposed: the proportional probabilistic sampling method and the dominant category-based sampling method. The study examines whether conditioning ensembles of simulated seasonal rainfall on forecast probabilities enhances prediction accuracy relative to climatology. By evaluating both forecast skill and operational utility at the basin scale, this research aims to assess the reliability of CPC’s seasonal precipitation forecasts for informing water supply planning in the Tampa Bay region and identifies the most effective method for translating probabilistic seasonal outlooks into quantitative precipitation estimates ~~that TBW can use as inputs to the hydrologic models applied for operational decision-making~~ operational water supply planning in the Tampa Bay region.

85 2. Data and Methodology

85 2.1. Study Area

This study focuses on two river basins in west-central Florida: one within the Hillsborough River system and the other in the Alafia River system (Figure 1). Both basins lie within the TBW supply region, serving more than 2.6 million people through its member governments. The agency employs a statistical rainfall-runoff model to generate monthly streamflow forecasts that inform surface water management decisions (Wang et al., 2024). As primary sources of surface water supply, the Alafia River basin and Hillsborough River basin are particularly sensitive to improvements in seasonal rainfall predictability, which can enhance short-term water supply planning. Rainfall in these basins is influenced by large-scale climate variability, especially the ENSO. Notably, El Niño events tend to increase cool-season (winter to early spring) rainfall, while La Niña events typically suppress it. This sensitivity to ENSO highlights the potential for improved predictive skill during certain seasons (Wang and Asefa, 2018).



95

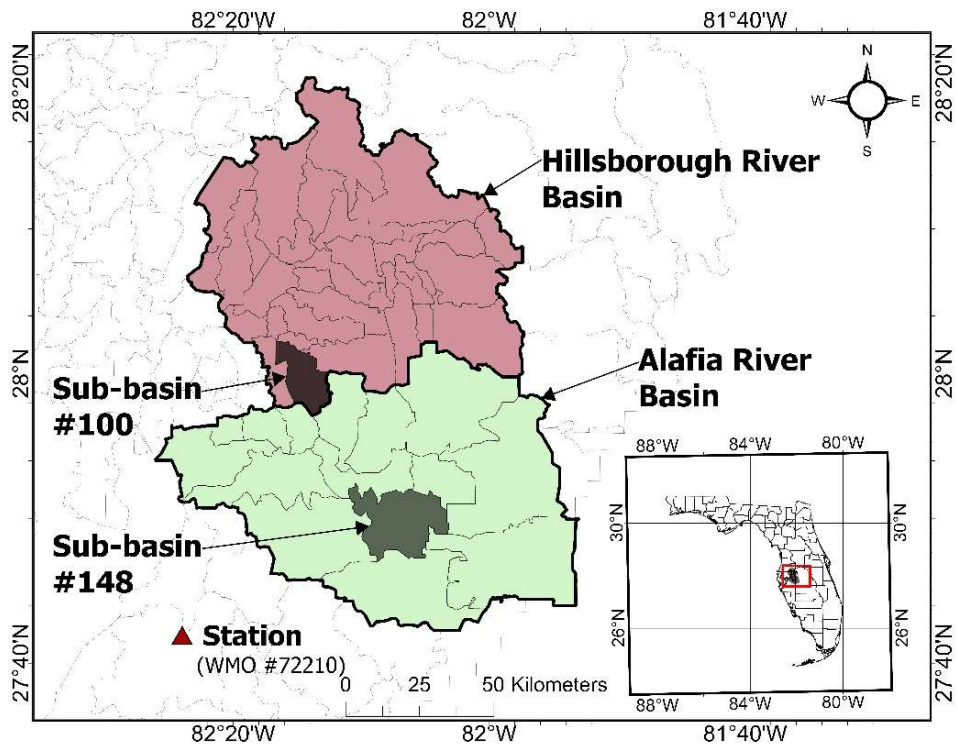


Figure 1: Hillsborough River Basin and Alafia River Basin in west-central Florida, with sub-basins #100, #148 and forecast station (WMO #72210) highlighted.

2.2. Seasonal Forecasts and Observation Data

100 Seasonal precipitation forecasts from CPC with a 0.5-month lead time were obtained for the period 1995–2019 from the publicly accessible database (https://ftp.cpc.ncep.noaa.gov/cpcfcsts/archives/long_lead/data/). Gridded forecasts corresponding to the $2^{\circ} \times 2^{\circ}$ grid cell encompassing 28°N , 82°W , where the study area is located, were used. Beginning in 2000, CPC transitioned to station-specific forecasts; therefore, data from World Meteorological Organization (WMO) Station #72210 (Tampa Bay, FL, at 27.7°N , 82.4°W) were utilized. A 25-year compilation of seasonal forecasts was assembled, each
105 represented as a tercile probability distribution (below-normal, near-normal, and above-normal). Forecast categories were assigned based on the highest probability among the three. If all categories had equal probabilities (33.33%), the forecast was classified as “Equal Chances (EC).”

Tercile boundaries were defined using a uniform 30-year baseline climatology for each season, updated every decade in accordance with guidelines from the WMO and the National Weather Service (NWS). For forecasts spanning 2010–2020, tercile limits were based on the 1981–2010 climatology (NOAA Climate Prediction Center, 2019). CPC forecasts were
110 available with lead times ranging from 0.5 to 12.5 months, offering long-range seasonal outlooks. The earliest forecasts were issued mid-month, with a 0.5-month lead indicating release approximately two weeks prior to the start of the target season. For example, the January–February–March (JFM) forecast issued in mid-December (0.5-month lead) was used to predict precipitation for the JFM period. For five seasons during the study period, 0.5-month lead forecasts were unavailable; in these cases, the next-shortest available lead time was used.

115 Observed basin precipitation for the study basins was obtained from Tampa Bay Water’s Bayesian rainfall dataset for the Integrated Northern Tampa Bay (INTB) domin, which integrates rain-gauge observations and Doppler radar measurements within a Bayesian statistical framework and interpolates the resulting daily rainfall estimates to the basins (GSI Environmental Inc., 2021a,b,c). The resulting basin-scale daily precipitation record was then aggregated to seasonal totals for use in forecast verification.

To analyze the influence of ENSO phases on forecast skill, Oceanic Niño Index (ONI) values were collected for each season from 1995 to 2019 (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php). The ONI is defined as the three-month running mean of SST anomalies in the Niño 3.4 region (5°S – 5°N , 120° – 170°W). El Niño (La Niña) events are identified by Niño 3.4 index values $\geq +0.5^{\circ}\text{C}$ ($\leq -0.5^{\circ}\text{C}$) sustained for at least five consecutive overlapping
125 seasons, while neutral conditions are defined by $-0.5^{\circ}\text{C} < \text{ONI} < +0.5^{\circ}\text{C}$.

2.3. Stochastic Rainfall Simulation

To generate ensemble rainfall scenarios for the study basins, this research employed a stochastic weather simulation approach. Stochastic rainfall simulators have been widely used to produce synthetic precipitation time series that preserve the statistical

characteristics of observed climates (Wilks & Wilby, 1999; Srikanthan & McMahon, 2001). In this study, Monte Carlo-simulated realizations of monthly rainfall, developed by GSI Environmental Inc. and TBW for 172 basins within the **Integrated Northern Tampa Bay** (INTB) hydrologic model domain, were utilized. These realizations were produced using a stochastic weather simulation framework specifically designed to maintain the historical spatial-temporal variability and cross-basin correlations of rainfall across the Tampa Bay region.

Over the 25-year simulation period, a total of 12,300 monthly gauge rainfall observations (41 gauges \times 12 months \times 25 years) were collected and subsequently transformed into standard normal variables as follows:

$$Z_{i,m,t} = \Phi^{-1}[F_{i,m}(X_{i,m,t})] \quad (1)$$

where $X_{i,m,t}$ represents the observed rainfall at station i , in month m , and year t ; $F_{i,m}$ denotes the empirical cumulative distribution function (CDF); and Φ^{-1} is the inverse standard normal CDF. The resulting standard normal variable is denoted as $Z_{i,m,t}$. To model the spatial and temporal dependence of rainfall, lagged cross-correlations among the normalized variables were computed using:

$$\rho_{i,j,m,t} = \text{Corr}(Z_{i,m,t}, Z_{j,m,t-\tau}) \quad (2)$$

where i and j denote the gauge stations, and τ represents the lag in months, extending up to 96 months (8 years). This process yields a comprehensive correlation matrix that captures the spatial correlation among gauges across years. The correlation matrix over 8-years period is structured as:

$$\Sigma_{8years} = \begin{bmatrix} \Sigma_0 & \cdots & \Sigma_{1month-8yea} \\ \vdots & \ddots & \vdots \\ \Sigma_{1month-8yea}^T & \cdots & \Sigma_0 \end{bmatrix} \quad (3)$$

where $\Sigma_0 = \rho_{i,j,m,t}$ denotes the correlation matrix between gauges i and j for the same month, and $\Sigma_{1month-8yea} = \rho_{i,j,m,1month-8yea}$ represents the correlation matrix between gauges i and j across different months and years. Using this correlation matrix, realizations of the joint multivariate normal distribution were generated to represent spatially and temporally correlated rainfall fields in the standard-normal space. These simulated Z -scores were then transformed back into original rainfall values ($X_{i,m,t}$) using the inverse of the empirical CDF:

$$x_{i,m,t} = (F_{i,m,t})^{-1}[\Phi(Z_{i,m,t})] \quad (4)$$

Simulated values that exceeded historical maximum or minimum observations were truncated to remain within those bounds. Finally, monthly gauge rainfall was spatially interpolated to generate basin-specific rainfall time series.

2.4. Ensemble Rainfall Forecast Generation

To incorporate forecast probabilities into the selection of members from the stochastic rainfall simulations described in Section 2.3, two ensemble generation schemes were evaluated for Sub-basin #100 and Sub-basin #148 (Figure 1). A nonparametric, tercile-stratified sampling method was employed to condition synthetic rainfall on CPC tercile probabilities, ensuring the

generated ensemble is probabilistically consistent with the forecast without assuming a parametric rainfall distribution. For each season and basin, the simulated total rainfall was defined as:

$$160 \quad X = \{ R_{r,t} \mid r = 1,2, \dots, 1000; t = 1,2, \dots, 25 \} \quad (5)$$

where $R_{r,t}$ denotes the total rainfall in year t in realization r . The empirical CDF of X , denoted as $F(X)$, is used to stratify the rainfall totals into tercile categories:

$$\text{Below - Normal (BN): } \{X: X \leq F^{-1}(0.33)\} \quad (6)$$

$$\text{Near - Normal (NN): } \{X: F^{-1}(0.33) < X \leq F^{-1}(0.66)\} \quad (7)$$

$$165 \quad \text{Above - Normal (AN): } \{X: X > F^{-1}(0.66)\} \quad (8)$$

Given the CPC seasonal forecast probabilities, P_{BN} , P_{NN} , P_{AN} representing probabilities of below-normal, near-normal, and above-normal rainfall conditions, respectively, two sampling schemes were evaluated to generate ensemble forecasts conditioned on these probabilities. The full ensemble generation process, including stratification of synthetic rainfall and probabilistic sampling, is illustrated in Figure 2.

170 **2.4.1. Proportional Tercile Sampling (PTS)**

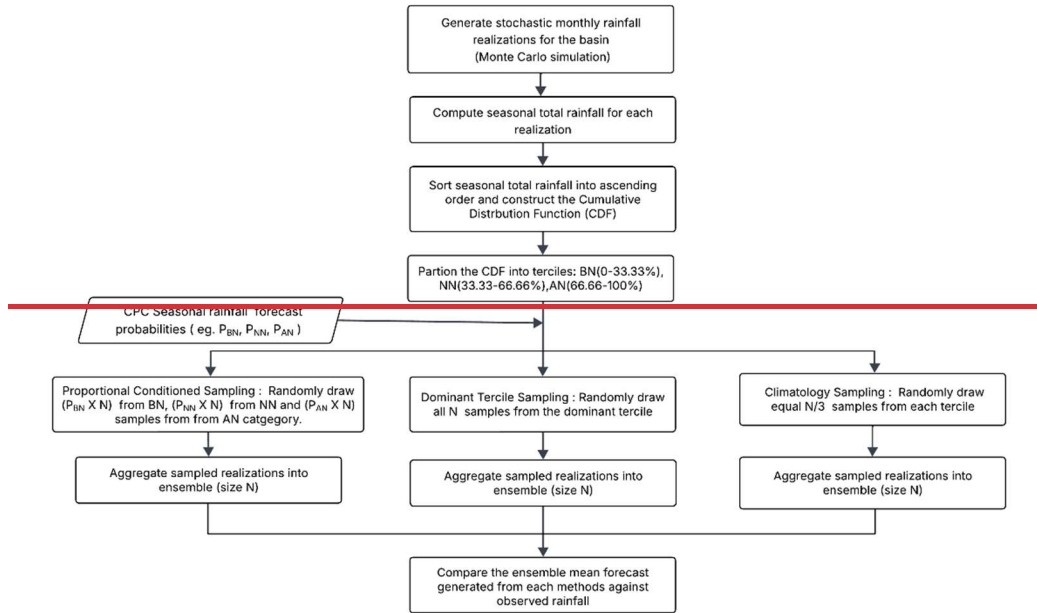
In this approach, ensemble members are randomly sampled with replacement from each pool, i.e., below-normal, near-normal, and above-normal as defined by Equations (6), (7), and (8), respectively. Sampling is performed in proportion to the CPC seasonal tercile forecast probabilities, ensuring that the ensemble reflects the relative likelihood of each rainfall category indicated by the forecast. For example, if the CPC forecast for a target season assigns a 50% probability to above-normal, 30% to near-normal, and 20% to below-normal rainfall, 50% of the ensemble members are randomly drawn from the pool defined by Equation (8), 30% from Equation (7), and 20% from Equation (6). By weighing the sampling according to the forecast probabilities, the likelihood of generating wet or dry scenarios is guided by the forecast signal.

175
|
In a prior study by MacLeod et al. (2023), tercile-based seasonal rainfall forecasts were used to resample the historical hydrological data. In contrast, the proposed PTS method utilizes a large ensemble of stochastically generated rainfall realizations, offering a more robust representation of the inherent variability and uncertainty in seasonal rainfall.

180 **2.4.2. Dominant Tercile Sampling (DTS)**

Under this scheme, ensemble members are randomly sampled with replacement exclusively from the tercile with the highest forecast probability. For instance, if the above-normal category has the highest probability among the three, all ensemble members are drawn from the above-normal pool defined by Equation (8). This method assumes high forecast confidence and may produce sharper predictions; however, it risks underrepresenting uncertainty when forecast probabilities are relatively close.

For comparison, unconditioned (climatology-based) ensembles were also generated, independent of seasonal forecasts. In this case, ensemble members were sampled uniformly from each of the terciles defined by Equations (6), (7), and (8), without regard to forecast probabilities. For EC forecasts, realizations were sampled equally from the below-normal, near-normal, and above-normal tercile pools for each method. This baseline represents a “no-skill” forecast and is equivalent to relying solely on historical climatology for each season. The climatology ensemble serves as a benchmark, and any climate-informed method should ideally demonstrate improved performance relative to this baseline (Liu et al., 2022). In this study, 1,000 ensemble members were generated for each method: PTS, DTS, and climatology, covering the period from 1995 to 2019.



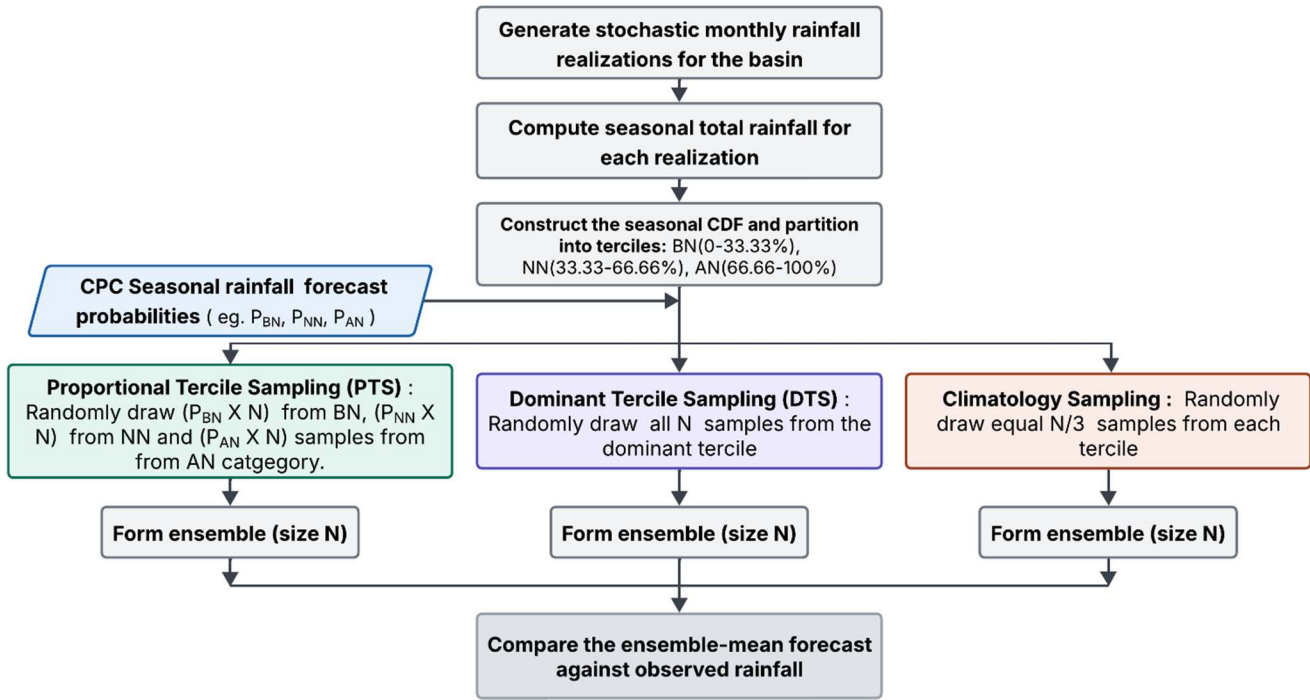


Figure 2: Schematic diagram illustrating the process of generating ensemble forecasts informed by seasonal climate outlooks.

2.5. Seasonal Precipitation Forecast Verification

Verification of seasonal forecasts is essential for evaluating the reliability and skill of climate forecasts (Huang & Zhao, 2022). Seasonal probabilistic precipitation forecasts are typically assessed using two primary approaches. The first is categorical verification, which compares the forecasted precipitation category (below-normal or above-normal) with the corresponding observed category. This comparison is summarized in a contingency table from which various verification metrics are derived. The second is probabilistic verification, which evaluates the forecasted probabilities assigned to each tercile category against the actual observed outcome, providing a more nuanced assessment of forecast performance (Zhang & Casey, 2000; Murphy & Winkler, 1987).

2.5.1. Categorical Forecast Verification

For categorical evaluation, the discrete distribution of forecasts and observations is assessed using a 2×2 contingency table, as illustrated in Table 1 (Doswell et al., 1990). The table includes four: ‘Hits’, when an event is both forecasted and observed; ‘Misses’, when the event occurs but is not forecasted; ‘False Alarms’, when the event is forecasted but does not occur; and ‘Correct Rejections’, when the event is neither forecasted nor observed. Separate contingency tables are constructed for wet events (above-normal) and dry events (below-normal) for each basin to enable more detailed evaluation [where equal chances](#)

cases forecast grouped into the non-event category. From these tables, several standard skill metrics are computed, consistent with methodologies used in prior studies (Wilks, 2006; Doswell et al., 1990):

$$POD = \frac{Hits}{Hits+Misses} \quad (9)$$

$$FAR = \frac{False\ alarms}{Hits+False\ alarms} \quad (10)$$

$$SR = \frac{Hits}{Hits+False\ Alarms} \quad (11)$$

The probability of detection (*POD*), or hit rate, defined in Equation (9), represents the proportion of observed events that were correctly forecasted. The false alarm rate ratio (*FAR*), shown in Equation (10), indicates the proportion of forecasted events that did not actually occur. The success rate ratio (*SR*), given in Equation (11), measures the proportion of correct forecasts relative to all forecasts issued. The forecast's ability to discriminate between events and non-events can be assessed by comparing *POD*, and *FAR*: a useful forecast should detect a substantial portion of events while minimizing false alarms (i.e., $POD > FAR$). If *POD* equals *FAR*, the forecast lacks skill in distinguishing events from non-events and effectively performs no better than random guessing. It is important to evaluate all three metrics collectively to obtain a comprehensive understanding of forecast performance. The ideal scenario is characterized by being $POD = 1$, $FAR = 0$, and $SR = 1$.

Table 1: A 2 x 2 contingency table summarizing outcomes of hits (H), misses (M), false alarms (FA), and correct rejections (CR) derived from observed versus forecasted events.

	Observed (Yes)	Observed (No)	Total
Forecast (Yes)	Hits (H)	False alarms (FA)	Total forecasted (Yes)
Forecast (No)	Misses (M)	Correct rejections (CR)	Total forecasted (No)
Total	Total Observed (Yes)	Total Observed (No)	Total Events (T)

To assess overall categorical skill relative to a no-skill baseline, the CPC-modified Heidke Skill Score (HSS), defined in Equation (12) (Heidke, 1926; NOAA Climate Prediction Center, 2005), is used. HSS quantifies the improvement of the number of correct forecasts over what would be expected by chance and is defined as:

$$HSS(\%) = 100 \frac{H-E}{T-E} \quad (12)$$

where T is the total number of forecasts issued; H is the number of correct forecasts; and E is the expected number of correct forecasts by chance. For tercile categorical forecasts, E is equal to one-third of T . A HSS score of 100 indicates a perfect forecast, -50 corresponds to the worst possible forecast with no correct predictions, and a score of 0 reflects a forecast with no skill, equivalent to random guessing.

235 **2.5.2. Probabilistic Forecast Verification**

In addition to categorical evaluations, probabilistic forecasts are assessed using metrics that quantify the quality of the predicted probability distributions. A key metric for this purpose is the Ranked Probability Skill Score (*RPSS*), which is derived from the Ranked Probability Score (*RPS*) (Wilks, 1995). The *RPS* evaluates the accuracy of forecast probability distribution relative to the observed outcome, while the *RPSS* compares this accuracy against a reference forecast, typically climatology (Wilks, 240 2006). For a single forecast–observation pair, the *RPS* is calculated as:

$$RPS = \sum_{k=1}^N (F_k - O_k)^2 \quad (13)$$

where $N = 3$ for tercile forecasts. F_k and O_k represent the cumulative forecast probabilities and cumulative observed probabilities for category k , respectively. An *RPS* value of 0 indicates a perfect forecast. To evaluate the skill of the forecast relative to a reference (typically climatology), the reference *RPS*, denoted as RPS_{Cl} , is computed as:

245
$$RPS_{Cl} = \sum_{k=1}^N (C_k - O_k)^2 \quad (14)$$

where C_k represents the cumulative climatological probability for category k . The *RPSS* metric is defined as:

$$RPSS = 1 - \frac{RPS}{RPS_{Cl}} \quad (15)$$

The *RPSS* value ranges from $-\infty$ to 1. An *RPSS* of 0 indicates that the forecast performs no better than the reference baseline, typically climatology. The metric penalizes more severely when high-confidence predictions are incorrect, while forecasts 250 made with low confidence receive smaller penalties if they turn out to be wrong.

2.5.3. Ensemble Forecast Evaluation

Following the generation of ensemble mean forecasts as described in Section 2.4, their performance in predicting observed seasonal precipitation from 1995 to 2019 was evaluated using two metrics. The first metric, Root Mean Square Error (*RMSE*), quantifies the average magnitude of the differences between the ensemble mean forecasts and the corresponding observed 255 precipitation. *RMSE* was computed using Equation (16):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2} \quad (16)$$

where f_i and o_i denote the i^{th} forecast and corresponding observation, respectively; and N is the total number of forecast–observation pairs. *RMSE* provides a measure of overall forecast accuracy and is particularly sensitive to outliers, penalizing larger errors more heavily than smaller ones.

260 To evaluate the agreement between forecasted and observed precipitation time series, the Pearson correlation coefficient was computed as:

$$r = \frac{\sum_{i=1}^N (f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^N (f_i - \bar{f})^2} \sqrt{\sum_{i=1}^N (o_i - \bar{o})^2}} \quad (17)$$

Where \bar{f} and \bar{o} denote the mean forecasted and observed precipitation, respectively. This metric evaluates how effectively the ensembled forecast captures interannual variability.

265 Forecasts were generated with a 0.5-month lead time to support short-term seasonal water resources management. By comparing the ensemble forecasts derived from sampling historical stochastic data, this study aims to identify the most effective method for each season and assess whether either approach yields skillful predictions beyond those provided by climatology-based forecasts.

3. Results and Discussion

270 3.1. Evaluation of localized NOAA forecasts

This section examines the frequency of CPC tercile-category forecast issuance over a 25-year study period, as illustrated in Figures 3a for the Alafia River Basin and Figure 3c for the Hillsborough River Basin. Of the 298 forecasts issued, 59.06% (176/298) favored a specific tercile category: 27.85% (83/298) predicted below-normal precipitation, while 31.21% (93/298) predicted above-normal precipitation. The remaining 40.94% (122/298) were classified as EC, indicating no clear preference for any category. Notably, none of the forecasts during the study period favored the near-normal category, likely reflecting the inherent difficulty in predicting the middle tercile, which is often associated with weak anomaly signals (Van den Dool & Toth, 1991; Mason et al., 2021).

275 A seasonal pattern emerges in the issuance of EC and non-EC forecasts, as shown in Figures 3a and 3c. From late spring through early fall (AMJ to OND), only about half of the forecasts fell into the non-EC category (i.e., either above-normal or below-normal precipitation). In contrast, the proportion of non-EC forecasts increased markedly during winter months: 79% for NDJ, 84% for FMA, 92% for DJF, and 96% for JFM. This trend suggests that CPC tends to issue forecasts with greater confidence during winter, consistent with the findings of Peng et al. (2012), who observed that predictive skill enhance the frequency of non-EC forecasts. The success ratio, as defined in Equation (11), was calculated based on Figures 3a and 3c over the 25-year evaluation period. In the Alafia River Basin, below-normal forecasts yielded a success **rate ratio** of 45.78% (38 hits out of 83), while above-normal forecasts achieved 43.01% (40 hits out of 93). The Hillsborough River Basin exhibited slightly lower but comparable success ratios: 42.17% (35/83) for below-normal and 40.86% (38/93) for above-normal forecasts.

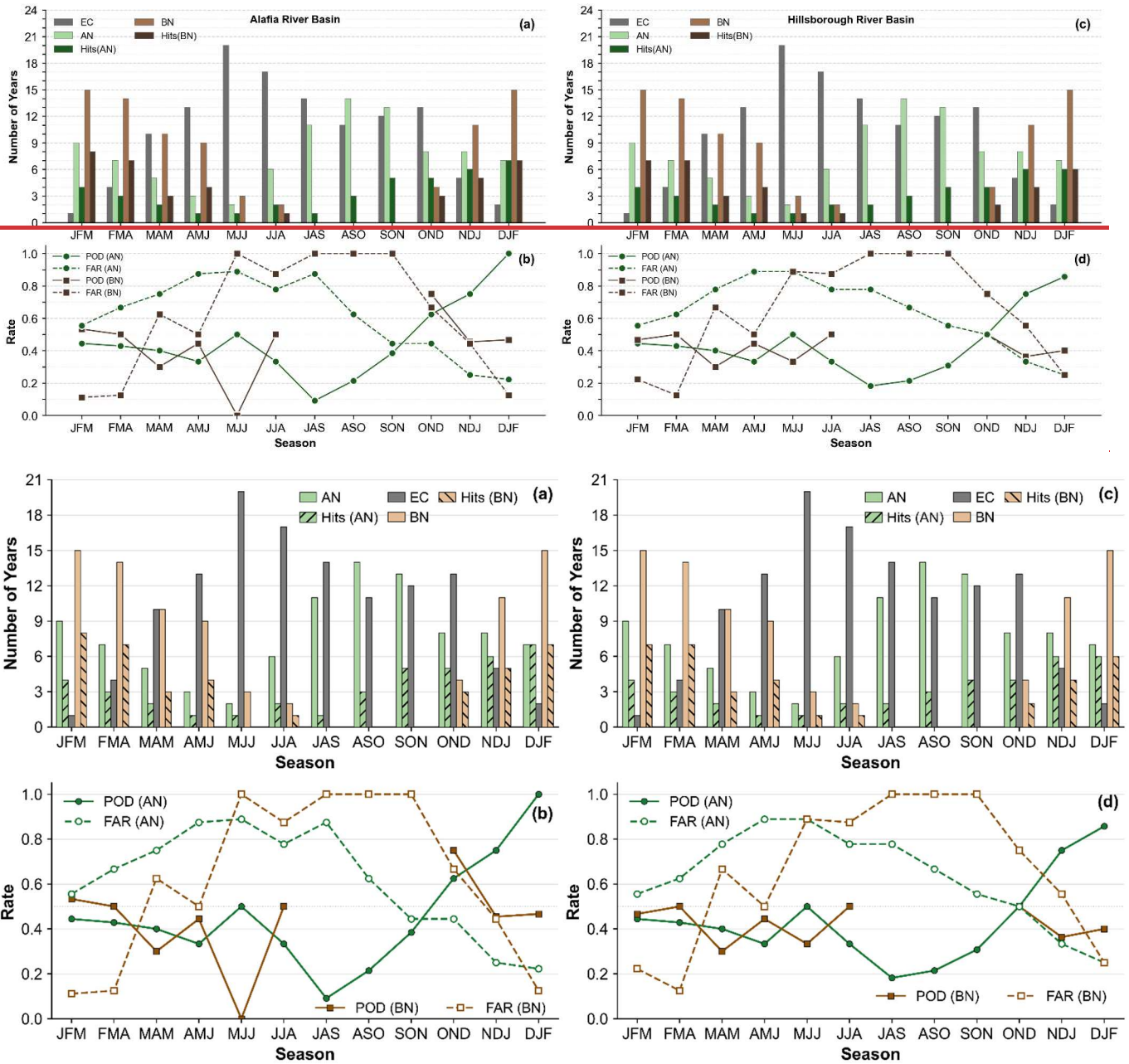


Figure 3: Seasonal distribution of categorical precipitation forecasts and associated performance metrics for the Alafia River Basin (a, b) and Hillsborough River Basin (c, d). Panels (a) and (c) display the number of years forecasted as above-normal (AN), below-normal (BN), and equal-chance (EC), along with the number of years correctly predicted as above-normal [Hits (AN)] and below-normal [Hits (BN)]. Panels (b) and (d) present the probability of detection (POD) and false alarm rate (FAR) for AN and BN categories across seasons from 1995-2019

290

To assess the seasonal discrimination ability of the forecasts, contingency statistics (i.e., hits, misses, false alarms, and correct rejections) were computed for each season. The POD and FAR were then derived and are presented in Figure 3b for the Alafia

295

River Basin and Figure 3d for the Hillsborough River Basin. Forecast discrimination skill is most pronounced during the dry season for the above-normal category, particularly in OND, NDJ, and DJF periods, where POD consistently exceeds FAR. In the Alafia River Basin, above-normal forecasts during OND (POD = 0.63, FAR = 0.44), NDJ (POD = 0.75, FAR = 0.25), and DJF (POD = 1.00, FAR = 0.22) demonstrate strong predictive capability. The Hillsborough River Basin shows comparable skill, with above-normal forecasts during OND (POD = 0.50, FAR = 0.50), NDJ (POD = 0.75, FAR = 0.33), and DJF (POD = 0.85, FAR = 0.25). In contrast, during the remaining dry-season months, POD values for above-normal generally fall below FAR, indicating weaker discrimination between observed and unobserved events. Forecast skill for the below-normal category is more limited, with POD exceeding FAR only during DJF, JFM, and FMA. This pattern suggests that the ability to correctly identify dry anomalies is largely confined to the early part of the year, when large-scale climate drivers such as ENSO enhance predictability. During the rest of the year, particularly throughout the wet season, FAR consistently exceeds POD, indicating a higher tendency for false alarms and reduced discrimination skills for both forecast category forecasts.

Overall, these findings indicate that while below-normal forecasts exhibit episodic skill during winter and early spring (DJF, JFM, FMA), above-normal forecasts show stronger and more consistent discrimination performance during late fall and early winter (OND–DJF).

Figure 4 presents the HSS, which adjusts the proportion of correct forecasts expected by chance, thereby isolating the true categorical forecast skill. Figure 5 displays the RPSS, which evaluates the probabilistic forecast skill relative to climatology. Both categorical and probabilistic skill metrics consistently show that forecast skill is concentrated during the dry season (OND–FMA), while the wet season exhibits limited or even negative skill. For both basins, overall HSS values exceed 20 throughout late fall and winter but become negative during the wet season (Figures 4a and 4b). When disaggregated by forecast category, above-normal forecasts contribute most significantly to overall skill. In both basins, HSS for above-normal forecasts increase from OND, peak in DJF with maximum scores of 100 for the Alafia River Basin (Figure 4a) and 79 for the Hillsborough River Basin (Figure 4b) and then gradually decline through JFM and FMA. These trends are mirrored in the RPSS results (Figure 5a and 5b), where above-normal forecasts consistently outperform climatology during OND–DJF, indicating higher forecast skill and reliability. In contrast, below-normal forecasts exhibit weaker and less consistent skill. HSS values for below-normal forecasts remain below 30 for most seasons, except for OND in the Alafia River Basin (Figure 4a). RPSS values for below-normal forecasts hover near or below zero for much of the year, indicating limited probabilistic skill.

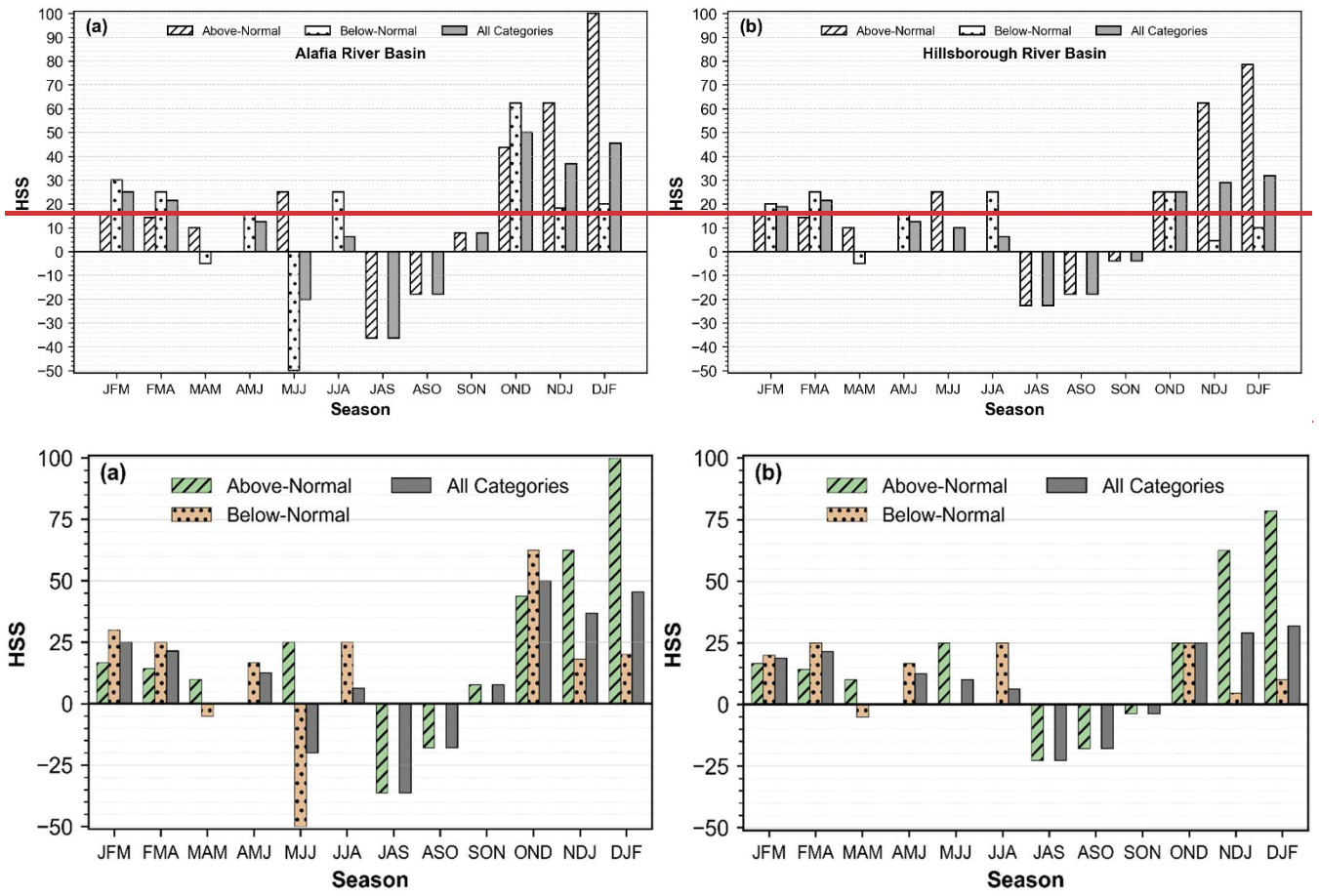
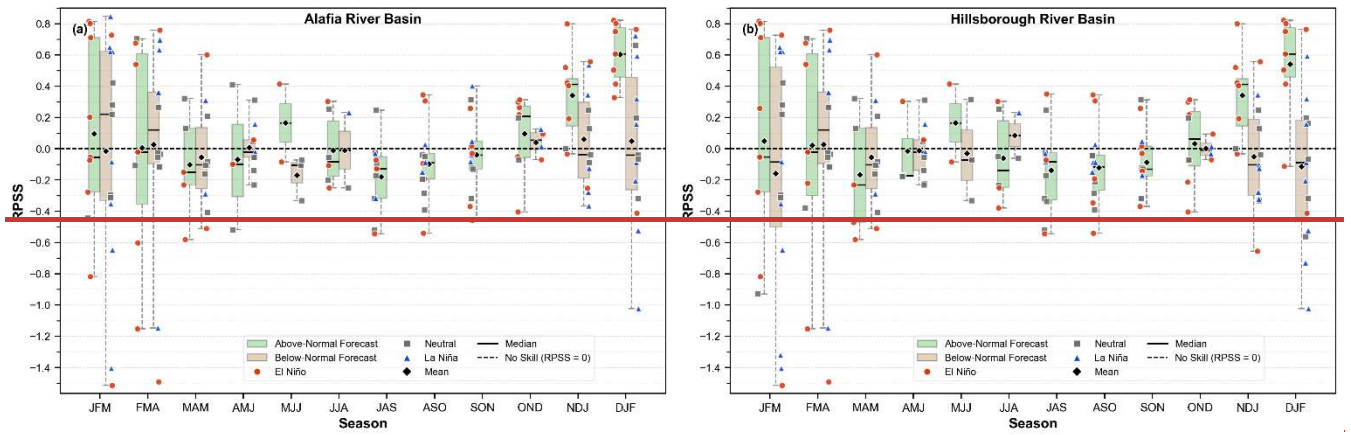
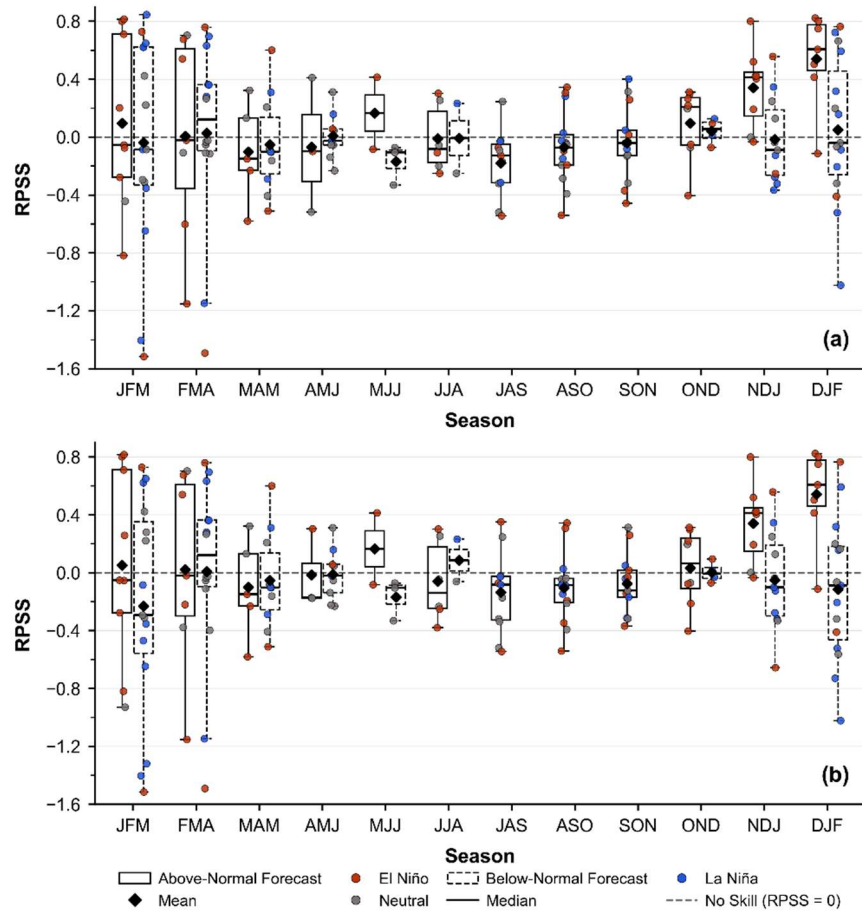


Figure 4: Heidke Skill Score (HSS) for CPC 0.5-month-lead seasonal precipitation forecasts in the Alafia River Basin (a) and Hillsborough River Basin (b), shown separately for above-normal, below-normal, and both forecast categories combined.

325





330 **Figure 5: Ranked probability skill score (RPSS) for CPC 0.5-month-lead seasonal precipitation forecasts in the Alafia River Basin (a) and Hillsborough River Basin (b). Boxplots depict the seasonal distribution of RPSS for above-normal (AN) and below-normal (BN) forecast categories from 1995 to 2019. Individual points represent seasonal cases, color-coded by ENSO phase (El Niño, La Niña, Neutral). The dashed horizontal line represents the no-skill reference (RPSS = 0), indicating performance equivalent to climatology.**

335 Overall, the categorical and probabilistic analyses highlight the strong seasonality of forecast skill, with enhanced performance during winter and early spring (OND–FMA). During this period, both categorical accuracy and probabilistic reliability substantially exceed climatology, underscoring the predictive value of CPC seasonal forecasts for hydrological applications in west-central Florida.

3.2. ENSO Influence on Forecast Skill

340 Figure 5 shows the seasonal distribution of RPSS for individual forecast categories, annotated by the corresponding ENSO phase during each observed season. Nearly all high-skill outcomes coincide with El Niño years. Across both the Alafia and Hillsborough basins, the highest RPSS values occur overwhelmingly during El Niño conditions, indicating that wintertime

345 predictability in this region is largely realized when the ENSO teleconnection provides a strong, coherent anomaly signal. Further analysis of dry season periods (NDJ and DJF) and classifying ENSO using the Oceanic Niño Index (ONI) further highlights the role of El Niño in both the frequency and accuracy of above-normal precipitation forecasts. In NDJ, eight above-normal forecasts were issued between 1995 and 2019 (1997/98, 2002/03, 2006/07, 2009/10, 2012/13, 2014/15, 2015/16, and 2018/19); seven align with El Niño conditions, with the exception of 2012/13, which occurred during a neutral phase. In DJF, all seven above-normal forecasts (1997/98, 2002/03, 2006/07, 2009/10, 2014/15, 2015/16, and 2018/19) were issued during El Niño conditions. By contrast, below-normal (dry) forecasts are distributed across El Niño, neutral, and La Niña phases and do not show a consistent alignment with a single large-scale driver, which is consistent with their weaker and more variable skill. Overall, these results underscore ENSO, particularly El Niño as a dominant source of enhanced wet-winter forecast skill for this region.

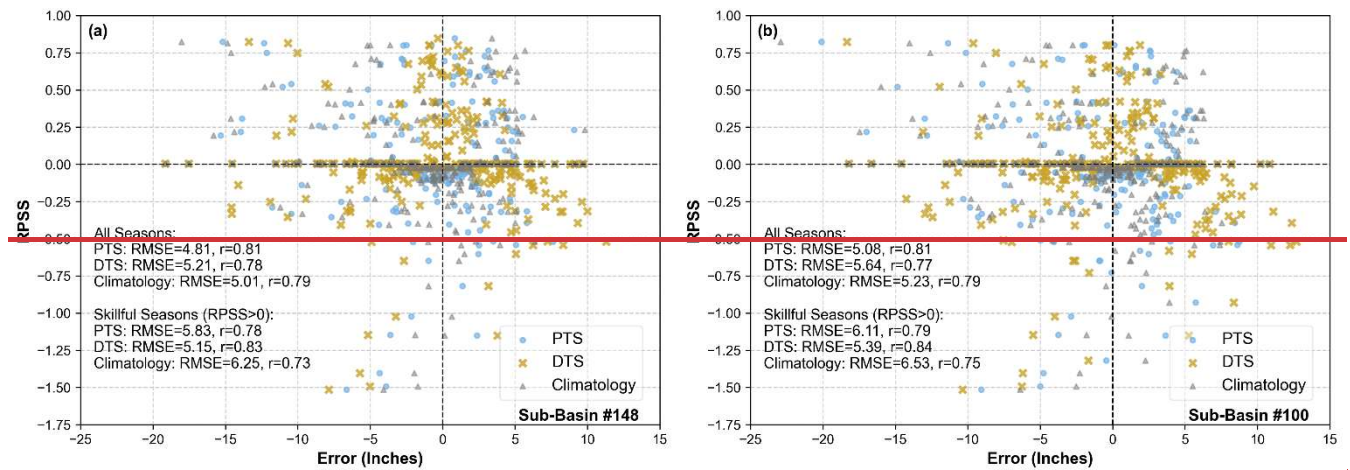
350

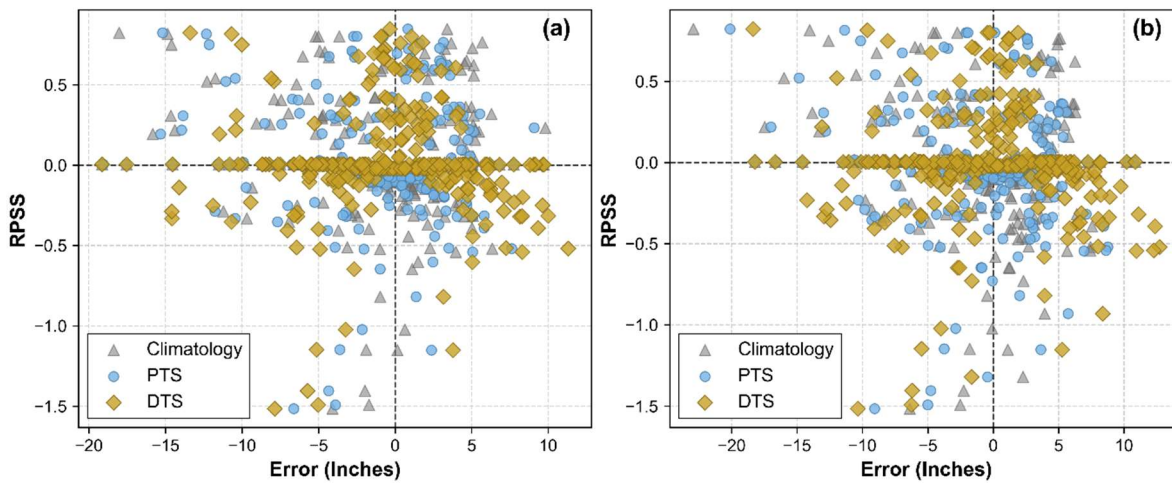
3.3. Performance of Sampling Schemes

Figure 6 summarizes how seasonal forecast error (ensemble-mean minus observed precipitation) varies with probabilistic skill (RPSS) for PTS, DTS, and climatology in sub-basin #148 (Figure 6a) and sub-basin #100 (Figure 6b). A consistent visual pattern emerges in both panels: near the no-skill threshold (RPSS \approx 0), the PTS and climatology points largely overlap, indicating comparable errors when the seasonal forecast provides little usable information, whereas DTS is typically farther from the zero-error line, reflecting larger errors under weak-signal conditions. In the skillful regime (RPSS > 0), DTS points cluster closest to the zero-error line, with PTS generally falling between DTS and climatology, suggesting that concentrating sampling on the dominant tercile is most beneficial when the seasonal signal is strong enough to yield positive RPSS. In contrast, for unskillful seasons (RPSS < 0), DTS shows the largest departures from the zero-error line, indicating that DTS can amplify errors when the forecast probabilities are unreliable.

355

360





365

Figure 6: Scatterplot of forecast error (mean ensemble forecast minus observed precipitation, in inches) for each season from 1995 to 2019, corresponding to proportional tercile sampling (PTS), dominant tercile sampling (DTS), and climatology sampling. Errors are plotted against the ranked probability skill score (RPSS) of that season for sub-basin #148 (a) and sub-basin #100 (b).

Table 2. Summary of PTS, DTS, and climatology performance in sub-basins #148 and #100

<u>Sub-basin / season set</u>	<u>PTS</u>	<u>PTS</u>	<u>DTS</u>	<u>DTS</u>	<u>Climatology</u>	<u>Climatology</u>
	<u>RMSE (in.)</u>	<u>r</u>	<u>RMSE (in.)</u>	<u>r</u>	<u>RMSE (in.)</u>	<u>r</u>
<u>#148 – All seasons</u>	<u>4.81</u>	<u>0.81</u>	<u>5.21</u>	<u>0.78</u>	<u>5.01</u>	<u>0.79</u>
<u>#148 – Skillful only (RPSS > 0)</u>	<u>5.83</u>	<u>0.78</u>	<u>5.15</u>	<u>0.83</u>	<u>6.25</u>	<u>0.73</u>
<u>#100 – All seasons</u>	<u>5.08</u>	<u>0.81</u>	<u>5.64</u>	<u>0.77</u>	<u>5.23</u>	<u>0.79</u>
<u>#100 – Skillful only (RPSS > 0)</u>	<u>6.11</u>	<u>0.79</u>	<u>5.39</u>	<u>0.84</u>	<u>6.53</u>	<u>0.75</u>

370

375

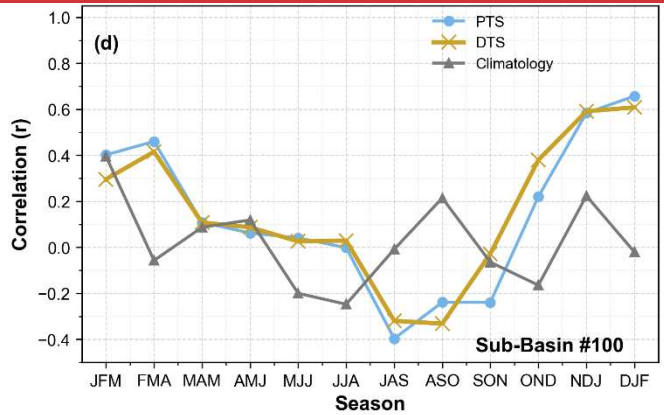
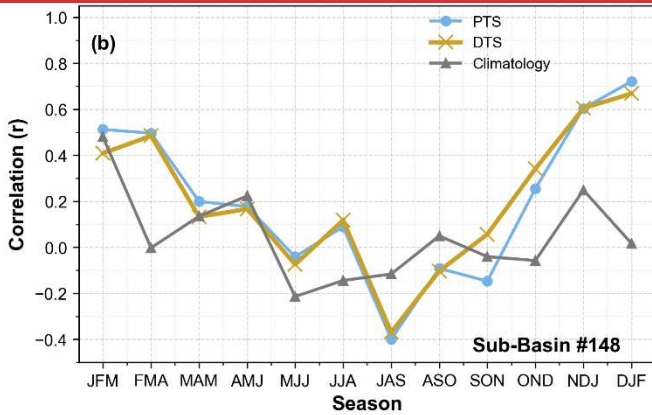
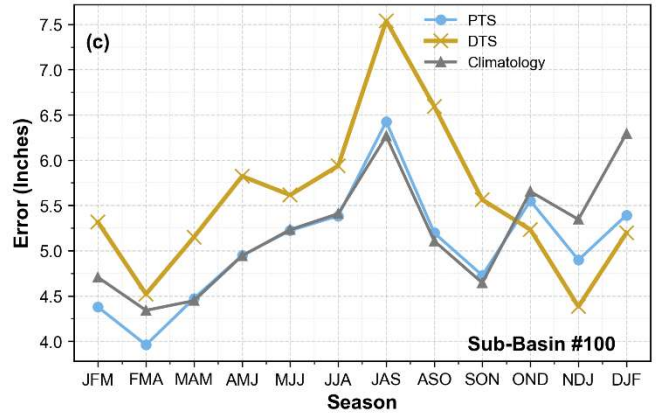
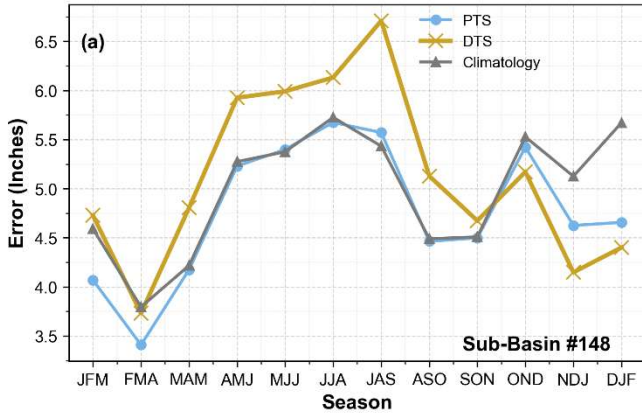
380

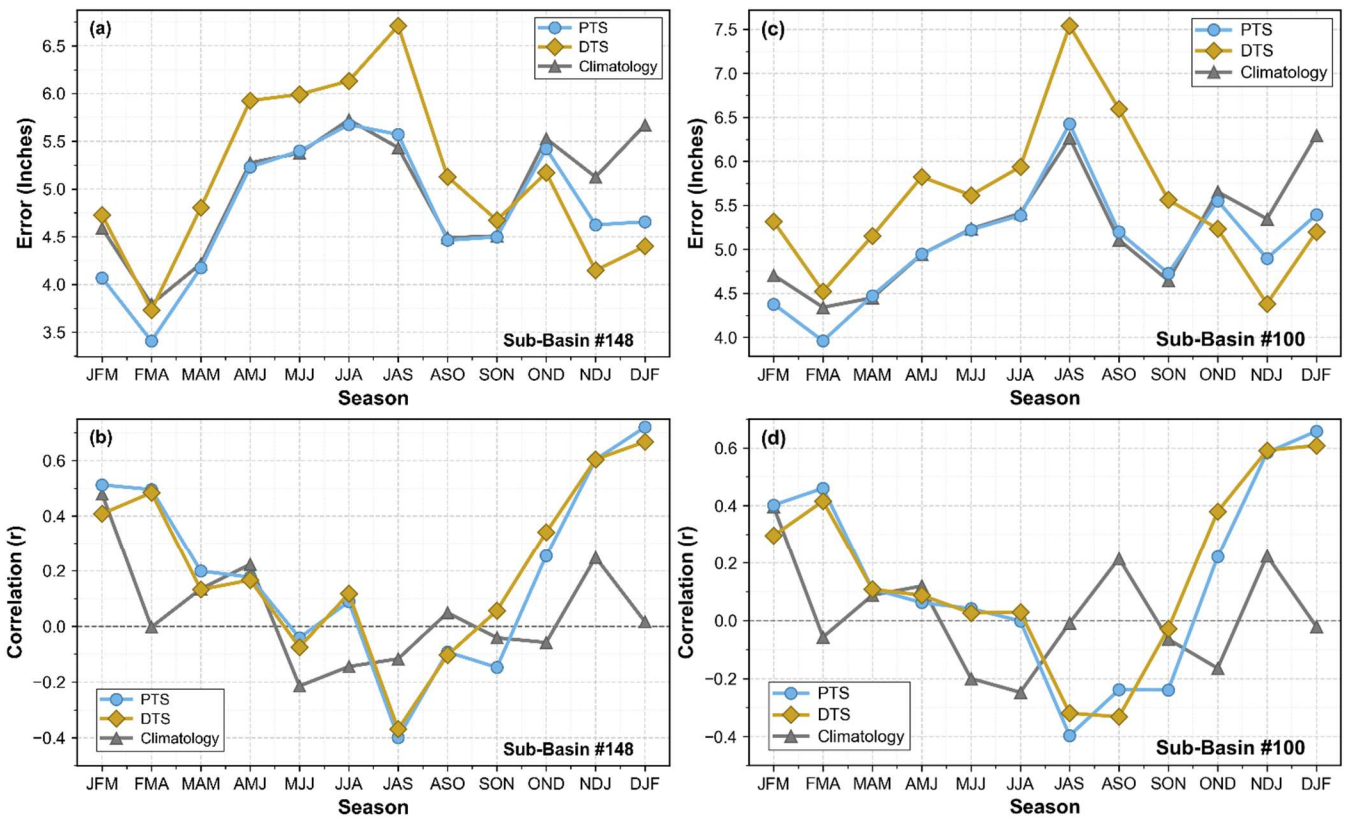
These qualitative patterns are reflected in the summary statistics shown in Table 2. Figure 6. Considering all seasons (1995–2019), PTS provides the lowest overall errors in both sub-basins. In sub-basin #148, PTS achieves RMSE = 4.81 in with $r = 0.81$, compared with climatology (RMSE = 5.01 in; $r = 0.79$) and DTS (RMSE = 5.21 in; $r = 0.78$). In sub-basin #100, PTS again yields the lowest RMSE (5.08 in) and highest correlation ($r = 0.81$), outperforming climatology (RMSE = 5.23 in; $r = 0.79$) and DTS (RMSE = 5.64 in; $r = 0.77$). However, when restricting to skillful seasons only (RPSS > 0), DTS becomes the best-performing approach in both basins. In sub-basin #148, DTS attains RMSE = 5.15 in and $r = 0.83$, improving over PTS (RMSE = 5.83 in; $r = 0.78$) and climatology (RMSE = 6.25 in; $r = 0.73$). In sub-basin #100, DTS similarly yields the lowest error (RMSE = 5.39 in) and highest correlation ($r = 0.84$), compared with PTS (RMSE = 6.11 in; $r = 0.79$) and climatology (RMSE = 6.53 in; $r = 0.75$). Across both sub-basins, a consistent performance pattern emerged for RMSE and correlation. When all seasons were included, the PTS approach produced only modest improvements, typically reducing RMSE by about 3–4% relative to climatology, while DTS actually performed worse than both methods. This indicates that PTS provides a relatively stable advantage under mixed-skill conditions. In contrast, when the analysis focused solely on skillful seasons (RPSS > 0), the DTS approach became clearly superior. DTS became clearly superior. DTS achieved RMSE reductions of

approximately 12% relative to PTS and 17–18% relative to climatology, while PTS still outperformed climatology by about 7%. Correlation patterns were consistent with these findings, with DTS yielding the highest r values during skillful seasons, compared to PTS and climatology.

385

Taken together, these results highlight that the relative value of the two sampling strategies is strongly dependent on forecast quality: PTS offers greater stability when all seasons are considered, whereas DTS is markedly more effective during seasons with demonstrable predictive skill, improving both error magnitude and correspondence with observed variability





390 **Figure 7: Seasonal variation in forecast performance for proportional tercile sampling (PTS), dominant tercile sampling (DTS), and climatology sampling for sub-basin #148 (a, b) and sub-basin #100 (c, d). Panels (a) and (c) display the root mean square error (RMSE, in inches) for each season, while panels (b) and (d) show the Pearson correlation coefficient (r) between ensemble forecasts and observed precipitation from 1995 to 2019.**

395 Figure 7 summarizes seasonal forecast skill using RMSE and Pearson correlation coefficient (r) for PTS, DTS, and climatology across sub-basins #148 and #100. Consistent with earlier RPSS diagnostics, the dry-season window, which spans OND, NDJ, DJF, and extends into JFM and MAM, exhibits the higher forecast skill. During this period, both CPC-informed methods (PTS and DTS) outperform climatology. At the onset of the dry season (OND), DTS achieves the lowest RMSE and highest correlation in both basins, maintaining its lead through NDJ and DJF. While PTS and DTS yield broadly comparable correlation values during these months, DTS consistently demonstrates a clear advantage in RMSE, reflecting the influence of strong large-scale climate signals such as El Niño.

400

In contrast, during late winter and spring (JFM–FMA–MAM–AMJ), when overall skill diminishes, PTS generally matches or exceeds climatology and outperforms DTS in both basins. As the wet-season begins (MJJ–JJA–JAS–ASO–SON), forecast performance deteriorates across all methods, with increased RMSE and near-zero or negative correlations. DTS frequently performs worse during this period, while PTS remains closest to the climatological baseline.

405 4. Conclusion

This study presents a comprehensive framework for applying NOAA's seasonal precipitation forecasts at the local scale and a thorough evaluation of the resulting localized forecasts. The evaluation of CPC 3-month precipitation outlooks at a 0.5-month lead for two basins in west-central Florida (Tampa Bay region) over the 1995–2019 period reveals that forecast skill is primarily concentrated in the dry half of the year. Non-equal chance (non-EC) outlooks were issued more frequently during late fall and winter, and verification confirms that OND and NDJ–DJF exhibit the strongest forecast performance, characterized by elevated Heidke Skill Scores, positive Ranked Probability Skill Scores, and strong categorical discrimination. In contrast, summer outlooks show minimal to no skill. Among forecast categories, above-normal (wet) forecasts are the primary contributors to wintertime skill, while below-normal (dry) forecasts demonstrate more modest but consistent performance throughout the dry season. Peak skill levels are closely aligned with El Niño winters, during which wet outcomes are more likely, a finding consistent with Infanti and Kirtman (2014), who documented elevated NMME skill during El Niño winters in the southeastern United States.

To translate categorical outlooks into quantitative basin-scale precipitation forecasts, two non-parametric ensemble methods were assessed. PTS offers a conservative approach, yielding the lowest overall errors and highest robustness across all seasons. DTS, by contrast, generally outperforms PTS during high-skill periods (fall–winter), particularly in terms of RMSE and correlation. Based on these findings, a hybrid operational strategy is recommended: employ DTS during high-skill fall and winter months (OND, NDJ, DJF) to capitalize on strong climate signals, and apply PTS during spring, summer, and other low-skill periods to consistent with the forecasted tercile probabilistic. This adaptive approach is expected to improve the accuracy of hydrologic simulation, e.g., streamflow forecasts, and subsequently benefit water resources management, which will be the focus of an extended study. The major contribution of this study is the development of a strategic approach to apply NOAA's forecasts at the local scale. The proposed methods, along with the approaches for forecast evaluation, are general and readily applicable to other regions.

Data availability

The precipitation forecast data used in this study are publicly available from NOAA CPC (https://ftp.cpc.ncep.noaa.gov/cpcfcsts/archives/long_lead/data/). Basin-scale observed precipitation data were obtained from Tampa Bay Water's rainfall dataset for the Integrated Northern Tampa Bay domain, available through Tampa Bay Water's data portal (<https://tampabaywater.sharefile.com/share/view/f3376a6f07534256/focac193-6df5-48bf-b3c0-6f8e7614a15f>). Oceanic Niño Index (ONI) values used to identify ENSO episodes for each season were obtained from the NOAA CPC (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php). The rainfall realizations ~~and the observed rainfall data for the~~ two study basins sisis will be made available on request.

435 **Author contributions**

MS conceptualized the study, curated the data, conducted the formal analysis, and prepared the original draft of the manuscript. HW contributed to the conceptualization of the study, provided supervision, and participated in manuscript review and editing. JSG, KP, TA, and FZ contributed to reviewing and editing the manuscript. DW contributed to the conceptualization and supervision of the study and participated in manuscript review and editing.

440 **Competing interests**

The contact author has declared that none of the authors has any competing interests.

Financial support

This research was funded by Tampa Bay Water.

References

- 445 Alexander, S., & Block, P. (2022). Integration of seasonal precipitation forecast information into local-level agricultural decision-making using an agent-based model to support community adaptation. *Climate Risk Management*, 36, 100417. <https://doi.org/10.1016/j.crm.2022.100417>
- Barnett, T. P., & Preisendorfer, R. (1987). Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis. *Monthly Weather Review*, 115(9), 1825–1850. [https://doi.org/10.1175/1520-0493\(1987\)115%253C1825:OALOMA%253E2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115%253C1825:OALOMA%253E2.0.CO;2)
- 450 Bett, P. E., Scaife, A. A., Li, C., Hewitt, C., Golding, N., Zhang, P., Dunstone, N., Smith, D. M., Thornton, H. E., Lu, R., & Ren, H.-L. (2018). Seasonal Forecasts of the Summer 2016 Yangtze River Basin Rainfall. *Advances in Atmospheric Sciences*, 35(8), 918–926. <https://doi.org/10.1007/s00376-018-7210-y>
- 455 [Becker, E., & Van Den Dool, H. \(2016\). Probabilistic Seasonal Forecasts in the North American Multimodel Ensemble: A Baseline Skill Assessment. *Journal of Climate*, 29\(8\), 3015–3026. https://doi.org/10.1175/JCLI-D-14-00862.1](https://doi.org/10.1175/JCLI-D-14-00862.1)
- Doswell, C. A., Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting. *Weather and Forecasting*, 5(4), 576–585.
- Wilks, D. S. (2011). Statistical methods in the atmospheric sciences. Academic press. *Weather and Forecasting*, 5(4), 576–585. [https://doi.org/10.1175/1520-0434\(1990\)005<0576:OSMOSI>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2)
- 460 Garcia, D., Rolim, J., Cameira, M. D. R., Belaud, G., Dalezios, N. R., Karoutsos, G., Santos, J. A., & Paredes, P. (2025). Assessing seasonal forecast performance to predict crop irrigation requirements to support water management decision-making in the Mediterranean region. *Agricultural Water Management*, 313, 109467. <https://doi.org/10.1016/j.agwat.2025.109467>

- Graham, R. M., Browell, J., Bertram, D., & White, C. J. (2021). The application of sub-seasonal to seasonal (S2S) predictions for hydropower forecasting (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2108.06269>
- 465 [GSI Environmental Inc. \(2021a\). Evaluation of Rain Gauge and Doppler Radar Data in the Integrated Northern Tampa Bay Hydrologic Model Domain. Prepared for Tampa Bay Water, Clearwater, Florida.](#)
- [GSI Environmental Inc. \(2021b\). Integration of Rainfall Data From Rain Gauges and Doppler Radar in the Integrated Northern Tampa Bay Hydrologic Model Domain. Prepared for Tampa Bay Water, Clearwater, Florida.](#)
- 470 [GSI Environmental Inc. \(2021c\). Supplemental Report on Integration of Rainfall Data From Rain Gauges and Doppler Radar in the Integrated Northern Tampa Bay Hydrologic Model Domain for 2017 Through 2019. Prepared for Tampa Bay Water, Clearwater, Florida.](#)
- Hartmann, H. C., Pagano, T. C., Sorooshian, S., & Bales, R. (2002). Confidence Builders: Evaluating Seasonal Climate Forecasts from User Perspectives. *Bulletin of the American Meteorological Society*, 83(5), 683–698. [https://doi.org/10.1175/1520-0477\(2002\)083<0683:CBESCF>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2)
- 475 Heidke, P. (1926). Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst. *Geografiska Annaler*, 8(4), 301–349. <https://doi.org/10.1080/20014422.1926.11881138>
- Huang, Z., & Zhao, T. (2022). Predictive performance of ensemble hydroclimatic forecasts: Verification metrics, diagnostic plots and forecast attributes. *WIRES Water*, 9(2), e1580. <https://doi.org/10.1002/wat2.1580>
- Hurkmans, R. T. W. L., Van Den Hurk, B., Schmeits, M., Wetterhall, F., & Pechlivanidis, I. G. (2023). Seasonal Streamflow Forecasting for Fresh Water Reservoir Management in the Netherlands: An Assessment of Multiple Prediction Systems. *Journal of Hydrometeorology*, 24(7), 1275–1290. <https://doi.org/10.1175/JHM-D-22-0107.1>
- 480 [Infanti JM, Kirtman BP \(2014\) Southeastern U.S. rainfall prediction in the North American multi-model ensemble. J Hydrometeorol 15\(2\):529–550. doi:10.1175/JHM-D-13-072.1](#)
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., ... Wood, E. F. (2014). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>
- 485 Liu, J., Yuan, X., Zeng, J., Jiao, Y., Li, Y., Zhong, L., & Yao, L. (2022). Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrology and Earth System Sciences*, 26(2), 265–278. <https://doi.org/10.5194/hess-26-265-2022>
- 490 Livezey, R. E., & Timofeyeva, M. M. (2008). The First Decade of Long-Lead U.S. Seasonal Forecasts: Insights from a Skill Analysis. *Bulletin of the American Meteorological Society*, 89(6), 843–854. <https://doi.org/10.1175/2008BAMS2488.1>
- 495 MacLeod, D., Quichimbo, E. A., Michaelides, K., Asfaw, D. T., Rosolem, R., Cuthbert, M. O., Otenyo, E., Segele, Z., Rigby, J. M., Otieno, G., Hassaballah, K., Tadege, A., & Singer, M. B (2023). Translating seasonal climate forecasts into water balance forecasts for decision making. *PLOS Climate*, 2(3), e0000138. <https://doi.org/10.1371/journal.pclm.0000138>

- Mason, S. J., Ferro, C. A. T., & Landman, W. A. (2021). Forecasts of “normal.” *Quarterly Journal of the Royal Meteorological Society*, 147(735), 1225–1236. <https://doi.org/10.1002/qj.3968>
- 500 Murphy, A. H., & Winkler, R. L. (1987). A General Framework for Forecast Verification. *Monthly Weather Review*, 115(7), 1330–1338. [https://doi.org/10.1175/1520-0493\(1987\)115<1330:AGFFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFV>2.0.CO;2)
- Namias, J. (1968). long range weather forecasting—History, current status and outlook1. *Bulletin of the American Meteorological Society*, 49(5–1), 438–470. <https://doi.org/10.1175/1520-0477-49.5.438>
- Nicholls, N. (1980). Long-range weather forecasting: Value, status, and prospects. *Reviews of Geophysics*, 18(4), 771–788. <https://doi.org/10.1029/RG018i004p00771>
- 505 NOAA Climate Prediction Center. (2005, December 12). CPC verification summary: Tutorial (Heidke Skill Score). National Weather Service, National Centers for Environmental Prediction. Retrieved [your access date], from <https://www.cpc.ncep.noaa.gov/products/verification/summary/index.php?page=tutorial>
- NOAA Climate Prediction Center. (2021, July). Long-lead forecast tool discussion and analysis (Revised 07/2021). National Weather Service, National Oceanic and Atmospheric Administration. Retrieved October 12, 2025, from https://www.cpc.ncep.noaa.gov/products/predictions/long_range/tools.html
- 510 NOAA Climate Prediction Center. (2019, May 16). Prognostic discussion for long-lead seasonal outlooks. National Weather Service, National Centers for Environmental Prediction. Retrieved October 12, 2025, from https://www.cpc.ncep.noaa.gov/products/predictions/long_range/fxus05.php
- [NOAA Climate Prediction Center. \(n.d.\). North American Multi-Model Ensemble \(NMME\). National Oceanic and Atmospheric Administration. Retrieved March 16, 2026, from https://www.cpc.ncep.noaa.gov/products/NMME/NMME_description.html](https://www.cpc.ncep.noaa.gov/products/NMME/NMME_description.html)
- 515 O’Lenic, E. A., Unger, D. A., Halpert, M. S., & Pelman, K. S. (2008). Developments in Operational Long-Range Climate Prediction at CPC. *Weather and Forecasting*, 23(3), 496–515. <https://doi.org/10.1175/2007WAF2007042.1>
- Peng, P., Kumar, A., Halpert, M. S., & Barnston, A. G. (2012). An Analysis of CPC’s Operational 0.5-Month Lead Seasonal Outlooks. *Weather and Forecasting*, 27(4), 898–917. <https://doi.org/10.1175/WAF-D-11-00143.1>
- 520 Portele, T. C., Lorenz, C., Dibrani, B., Laux, P., Bliefernicht, J., & Kunstmann, H. (2021). Seasonal forecasts offer economic benefit for hydrological decision making in semi-arid regions. *Scientific Reports*, 11(1), 10581. <https://doi.org/10.1038/s41598-021-89564-y>
- Srikanthan, R., & McMahon, T. A. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, 5(4), 653–670. <https://doi.org/10.5194/hess-5-653-2001>
- 525 Turco, M., Marcos-Matamoros, R., Castro, X., Canyameras, E., & Llasat, M. C. (2019). Seasonal prediction of climate-driven fire risk for decision-making and operational applications in a Mediterranean region. *Science of The Total Environment*, 676, 577–583. <https://doi.org/10.1016/j.scitotenv.2019.04.296>

- 530 [Slater, L. J., Villarini, G., & Bradley, A. A. \(2019\). Evaluation of the skill of North-American Multi-Model Ensemble \(NMME\) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA. *Climate Dynamics*, 53\(12\), 7381–7396. <https://doi.org/10.1007/s00382-016-3286-1>](https://doi.org/10.1007/s00382-016-3286-1)
- Van Den Dool, H. M., & Toth, Z. (1991). Why Do Forecasts for “Near Normal” Often Fail? *Weather and Forecasting*, 6(1), 76–85. [https://doi.org/10.1175/1520-0434\(1991\)006<0076:WDFNO>2.0.CO;2](https://doi.org/10.1175/1520-0434(1991)006<0076:WDFNO>2.0.CO;2)
- 535 [Wang, H. \(2014\). Evaluation of monthly precipitation forecasting skill of the National Multi-model Ensemble in the summer season: FORECASTING SKILL OF THE NATIONAL MULTI-MODEL ENSEMBLE. *Hydrological Processes*, 28\(15\), 4472–4486. <https://doi.org/10.1002/hyp.9957>](https://doi.org/10.1002/hyp.9957)
- Wang, H., & Asefa, T. (2018). Impact of different types of ENSO conditions on seasonal precipitation and streamflow in the Southeastern United States. *International Journal of Climatology*, 38(3), 1438–1451. <https://doi.org/10.1002/joc.5257>
- Wang, H., Asefa, T., & Duncan, J. (2024). Event-based evaluation of operational ENSO forecasting models in 2002–2020: Implications for seasonal water resources management. *Journal of Hydrology*, 636, 131295. <https://doi.org/10.1016/j.jhydrol.2024.131295>
- 540 Wang, Hui, Tirusew Asefa, Dave Bracciano, Alison Adams, and Nisai Wanakule. “Proactive Water Shortage Mitigation Integrating System Optimization and Input Uncertainty.” *Journal of Hydrology* 571 (April 2019): 711–22. <https://doi.org/10.1016/j.jhydrol.2019.01.071>.
- 545 Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences: An introduction* (1st ed.). Academic Press. ISBN 978-0127519654.
- Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: A review of stochastic weather models. *Progress in Physical Geography: Earth and Environment*, 23(3), 329–357. <https://doi.org/10.1177/030913339902300302>
- Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed.). Academic Press. ISBN 978-0127519661.
- 550 Zhang, H., & Casey, T. (2000). Verification of Categorical Probability Forecasts. *Weather and Forecasting*, 15(1), 80–89. [https://doi.org/10.1175/1520-0434\(2000\)015<0080:VOCPF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0080:VOCPF>2.0.CO;2)