water

MDPI

Article

Assessment of Subseasonal-to-Seasonal (S2S) Precipitation Forecast Skill for Reservoir Operation in the Yaque Del Norte River, Dominican Republic

Norman Pelak *, Eylon Shamir, Theresa Modrick Hansen and Zhengyang Cheng

Hydrologic Research Center, San Diego, CA 92127, USA; eshamir@hrcwater.org (E.S.); tmodrick@hrcwater.org (T.M.H.); zcheng@hrcwater.org (Z.C.)
* Correspondence: npelak@hrcwater.org

Abstract: Operational forecasters desire information about how their reservoir and riverine systems will evolve over monthly to seasonal timescales. Seasonal traces of hydrometeorological variables at a daily or sub-daily resolution are needed to drive hydrological models at this timescale. Operationally available models such as the Climate Forecast System (CFS) provide seasonal precipitation forecasts, but their coarse spatial scale requires further processing for use in local or regional hydrologic models. We focus on three methods to generate such forecasts: (1) a bias-adjustment method, in which the CFS forecasts are bias-corrected by ground-based observations; (2) a weather generator (WG) method, in which historical precipitation data, conditioned on an index of the El Niño–Southern Oscillation, are used to generate synthetic daily precipitation time series; and (3) a historical analog method, in which the CFS forecasts are used to condition the selection of historical satellite-based mean areal precipitation (MAP) traces. The Yaque del Norte River basin in the Dominican Republic is presented herein as a case study, using an independent dataset of rainfall and reservoir inflows to assess the relative performance of the methods. The methods showed seasonal variations in skill, with the MAP historical analog method having the strongest overall performance, but the CFS and WG methods also exhibited strong performance during certain seasons. These results indicate that the strengths of each method may be combined to produce an ensemble forecast product.

Keywords: subseasonal-to-seasonal forecasting; weather generator; CFS bias adjustment; conditional historical analog

Citation: Pelak, N.; Shamir, E.; Hansen, T.M.; Cheng, Z. Assessment of Subseasonal-to-Seasonal (S2S) Precipitation Forecast Skill for Reservoir Operation in the Yaque Del Norte River, Dominican Republic. *Water* **2024**, *16*, 2032. https://doi.org/10.3390/w16142032

Academic Editor: Christos S. Akratos

Received: 14 June 2024
Revised: 16 July 2024
Accepted: 16 July 2024
Published: 18 July 2024

1. Introduction

Skillful predictions of hydrometeorological variables at the subseasonal-to-seasonal (S2S) timescale are needed for forecasters to provide timely and accurate information to stakeholders, which can aid in drought management, flood mitigation, and water resources decision-making. The S2S timescale, generally defined as spanning from beyond 2 weeks to about 3 months [1], is challenging to forecast accurately, as it bridges the short-term, where initial conditions dominate weather patterns, and the seasonal, where longer-term climate effects are more pronounced.

Methods of generating S2S hydrometeorological forecasts can be broadly categorized as statistically based, physically based, or some combination thereof [2,3]. Statistical methods make primary use of observed relationships between a predictor (such as a climate index) and a predictand (such as seasonal rainfall), while physical methods rely on global circulation models (GCMs) and may be further downscaled using numerical weather prediction (NWP) models. Mixed methods combine elements of both approaches, for example, by downscaling a physically based GCM using statistical relationships with historical observations (e.g., [4]).

The high uncertainty of S2S forecasts is typically characterized using forecast ensembles [2,3]. Model uncertainty may be accounted for using multi-model ensembles (for

example, the North American Multi-Model Ensemble [5]) or using different parameterizations of a single model. More recent efforts have attempted to use machine learning to enhance the skill of existing forecast methods (e.g., [6]).

In this study, which was conducted in the Yaque del Norte River basin in the northwestern Dominican Republic (Figure 1), we evaluated the skill of three methods of generating S2S forecasts as part of the development of an operational forecast system for the basin. It has an area of roughly 7000 km^2 and ranges in elevation from approximately 3097 m to sea level at its outlet. We focus, in particular, on the portion of the basin that is upstream of the Moncion reservoir (hereafter the Moncion basin), which has an area of 612 km^2.

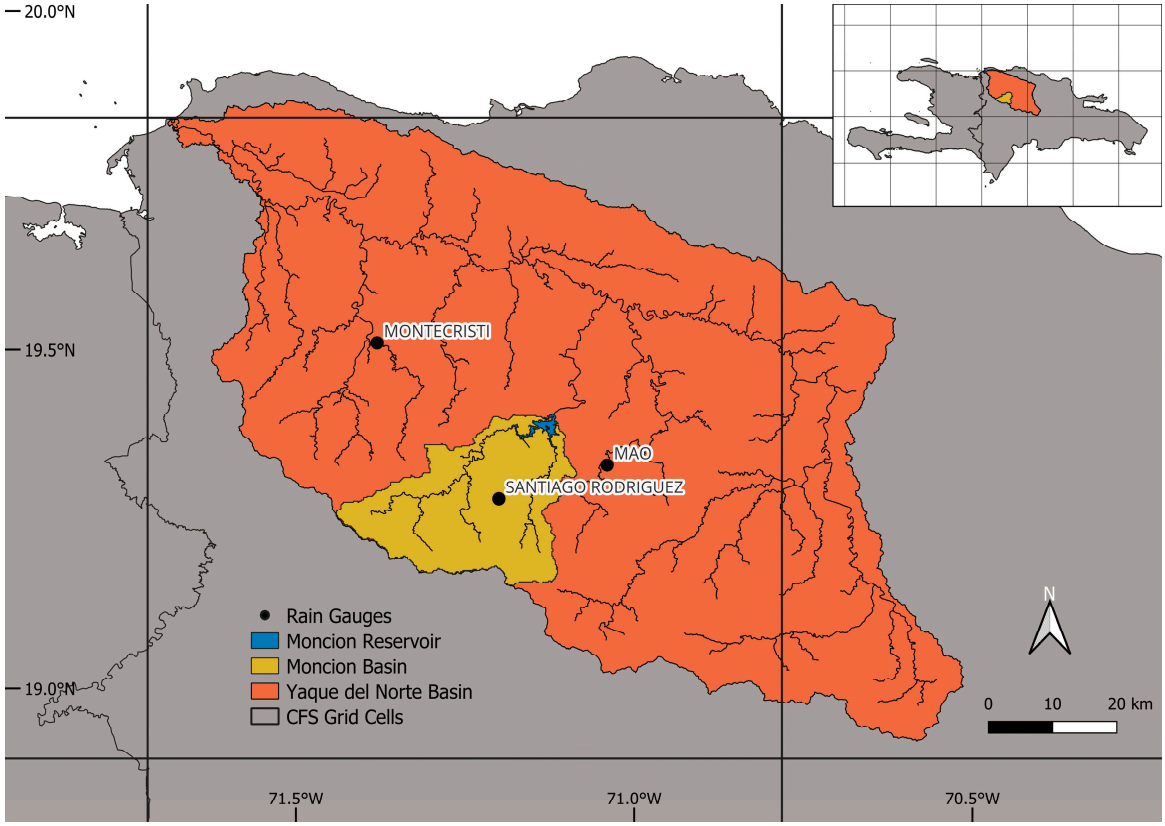

Figure 1. Map of the study area, showing the Moncion reservoir (blue) and its upstream basin (yellow), the Yaque del Norte basin (orange), and the CFS grid cells (black lines). The rain gauges are shown as black circles, with the gauge within the Moncion watershed being the Santiago Rodriguez rain gauge that was used in this study. The inset in the top right shows the location of these features in the northwestern Dominican Republic.

The annual precipitation pattern in the Dominican Republic is typically bimodal, with peaks occurring in late spring and in the fall and lower precipitation during the summer and winter [7]. In this study, we use the following season definitions: Spring (March, April, May—MAM), Summer (June, July, August—JJA), Fall (September, October, November—SON) and Winter (December, January, and February—DJF). This is a commonly used seasonal breakdown in which each season roughly overlaps with either one of the peaks or one of the dips in annual precipitation.

The climate of the Dominican Republic is strongly influenced by the El Niño–Southern Oscillation (ENSO) phenomenon [8]. Typically, El Niño events in the Caribbean are associated with drier conditions, while La Niña events tend to be wetter [9,10]. However, this pattern is not strongly apparent in the Santiago Rodriguez gauge record, where most of the seasons that had the heaviest precipitation occurred during either Neutral or El Niño periods (Figure 2b). Total rainfall in the region is also heavily influenced by hurricanes and

tropical storms, with the fall season peak in precipitation roughly coinciding with the peak of the Atlantic hurricane season. Such events are difficult to predict at seasonal timescales, although historically higher (lower) hurricane activity occurs during La Niña (El Niño) events [8], indicating one possible source of predictability.

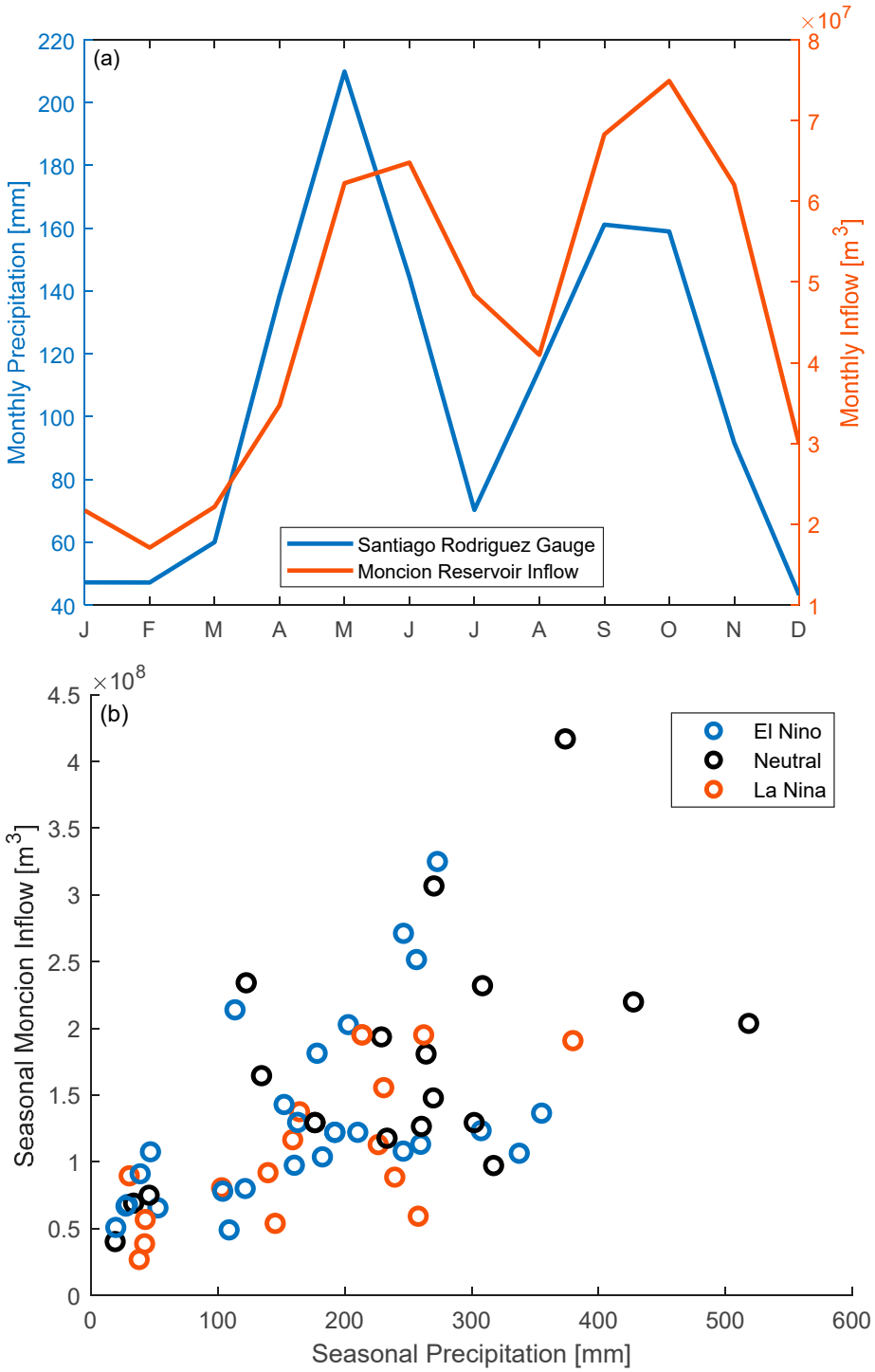

Figure 2. (**a**) Mean monthly precipitation at the Santiago Rodriguez precipitation gauge from 1938 to 2023 (blue line) and mean monthly inflow to the Moncion reservoir from 2009 to 2023 (red line). (**b**) Seasonal inflow to the Moncion reservoir vs. seasonal precipitation at the Santiago Rodriguez gauge. Colors indicate the ENSO status as determined by the mean Niño-3.4 index during the previous season: El Niño conditions (red), Neutral conditions (black), and La Niña conditions (blue). For further details, see Section 2.

To drive the hydrologic model for the Yaque del Norte basin, S2S precipitation forecasts must be generated as input. A variety of methods exist for generating S2S precipitation forecasts for regional hydrologic studies, each with its own advantages and disadvantages, and all may not be equally suitable for the region. Other studies have examined the skill of seasonal forecasts in the Caribbean (e.g., [11–13]), but few have focused on evaluating daily or sub-daily precipitation forecasts at the S2S timescale, as we do in this study. This study focuses on three contrasting methods (described in detail in Section 2), which represent a range of ways of obtaining the necessary rainfall forecasts, all of which are feasible to apply in an operational setting. The goal of this study is to determine which of these methods provides skillful S2S forecasts and how that skill varies throughout the year. The generation of temperature forecasts was not included in this study due to the low variability in seasonal temperature in the tropical climate of the Dominican Republic.

2. Methods

Three methods of generating forecasts of precipitation on the S2S timescale are examined in this study: (1) seasonal Climate Forecast System (CFS) forecasts [14] are bias-corrected using monthly change factors that are calculated from past forecasts and observations, (2) historical records of precipitation and ENSO indices are used to parameterize a stochastic weather generator (WG), and (3) an ensemble of historical satellite-based mean areal precipitation (MAP) time series are selected by conditioning on the state of the CFS forecast at the beginning of the season. These methods do not represent all possible means of generating S2S forecasts but were chosen to represent a range of possible methods, and while not completely independent of one another (two of the methods make use of the same precipitation gauge data, and two make use of the CFS model output), each is driven by a different primary data source (CFS precipitation, historical precipitation gauge records, and past archives of satellite-derived MAP, respectively). The methods and data used for each are described below.

2.1. CFS Bias Adjustment

The operational Climate Forecast System (CFS) available from the Climate Prediction Center of the U.S. National Atmospheric and Oceanic Administration (NOAA) provides forecasts of precipitation initialized 4 times per day with a forecast lead time of up to 9 months. In this study, we refer to these unaltered model outputs as the 'raw' CFS forecasts, or CFS-RAW, which we make use of in the analysis of the results in Section 3. However, the large spatial scale (~56 km) of the CFS grid cells generally requires that the outputs be bias-corrected to be useful for hydrological modeling, which typically requires inputs on much smaller scales (for instance, the total area of the Moncion basin is 612 km^2). Numerous methods to downscale and/or bias-correct climate and weather model outputs exist (e.g., [3,15–18]). In this study, given that we are aggregating all outputs to a seasonal timescale, we deemed it sufficient to only correct for the mean bias in precipitation depths. This was accomplished using the change factor method (also known as the perturbation factor method, e.g., [17]), which involves calculating both the forecasted and observed monthly precipitation over a reference period to find a monthly change factor. This monthly change factor represents the average ratio between the observed and modeled monthly precipitation, which can then be applied to the model outputs as a bias correction.

Applying this method requires a suitable observed dataset. Historical precipitation records for several gauges in the vicinity of the Yaque del Norte basin were provided by the Oficina Nacional de Meteorología (ONAMET) of the Dominican Republic. While three of these gauges were located within the Yaque del Norte basin, only one (Santiago Rodriguez) was located within the sub-basin, which flows into the Moncion reservoir (see Figure 1). We therefore chose this gauge, which has a long and relatively complete daily precipitation record from 1938 to 2023, to bias-correct the CFS output.

The current version of the CFS model became operational in March 2011. However, it was also run retrospectively by NOAA's National Centers for Environmental Information

every 5 days from 1982 to 2010 (this dataset is known as the CFS Reforecast). To avoid using information that would not be available at the time that the forecasts begin, we calculate the change factors over the 29-year reforecast period and apply them to the CFS forecasts from 2012 to 2023. Because the reforecasts are available every 5 days, the forecast for the first day of each season is not always available. In such cases, the next available set of CFS forecasts was used and treated as though they were made at the beginning of the season. The total monthly precipitation from the CFS reforecast was calculated and compared to the precipitation over the same time period from the Santiago Rodriguez rain gauge. In cases when more than 50% of days were missing, the monthly precipitation total was also set to missing, and that month was excluded from the change factor calculation. Change factors for each month were calculated as the ratio between the rain gauge values and the CFS values. Averages for each calendar month were found and linearly interpolated to the daily timescale, with monthly values being centered on the 15th of each month.

To apply the change factors over the forecast period of 2012–2023, we downscaled each of the 4 CFS forecasts initialized on the first day of each season during this period. This variation in initial conditions provides a 4-member ensemble. In a small number of cases, one or more of the forecasts from the first day of the season were unavailable, in which case we substituted the forecast from the second day of the month. The change factors calculated using the data and model output from the reforecast period were applied to each member of the ensemble over the forecast period. We refer to the results of this process as the bias-adjusted CFS or CFS-BA. Sample results for the March 2014 forecast, alongside the corresponding observations, are shown in Figure 3a.

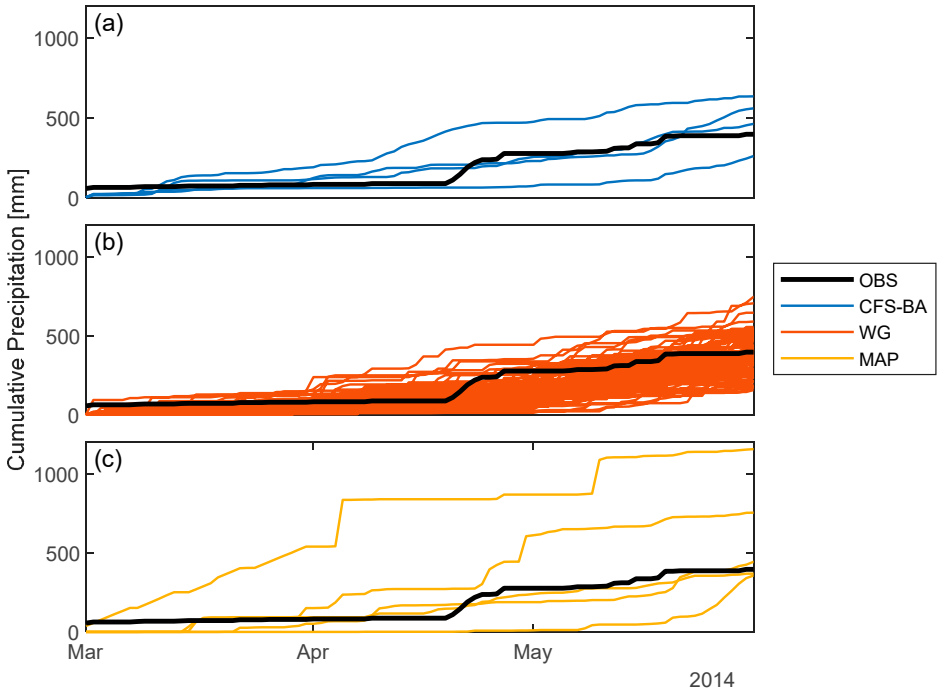

Figure 3. Cumulative seasonal precipitation for a sample forecast (initialized 1 March 2014) showing the precipitation traces from each of the 3 methods vs. the observed precipitation time series: (**a**) CFS-BA: an ensemble of bias-adjusted outputs from Climate Forecast System; (**b**) WG: a 100-member ensemble of outputs from the weather generator method; (**c**) MAP: an ensemble of outputs from the MAP historical analog method.

2.2. Weather Generator

Stochastic weather generators (WGs) use a statistical characterization of historical weather observations to produce synthetic precipitation traces that are statistically similar to the original time series. One of the main advantages of WGs when compared to the

other two forecast methods in this study, for which limited numbers of ensemble members are available due to computational cost (in the case of CFS forecasts) or to the length of the observational record (in the case of the MAP historical analog method, see Section 2.3), is that an arbitrarily large number of precipitation traces can be generated at minimal computational cost. If properly structured and calibrated, an ensemble with a sufficient number of traces represents well the probabilistic characteristics of the historical rainfall record [19].

As in Section 2.1, we make use of the Santiago Rodriguez rain gauge due to its location within our basin of interest and for its long and relatively complete record and use it to parameterize the WG. To avoid using data that would not have been available at the time of forecast, only gauge data prior to 2012 (1938–2011) was used.

The method used in this study is a modification of previous WGs developed elsewhere [20–22]. The first step was to categorize the total seasonal precipitation of the historical record into Dry, Medium, and Wet categories based on the terciles (separate tercile values were calculated for each season). If more than 50% of dates in a season were missing, the precipitation total for that season was set to missing and was not included in the wetness category calculations. As there are about 61 years of data available, splitting into 3 wetness categories resulted in ~20 seasons per category.

For each wetness category of each season, the 2-parameter gamma distribution (Equation (1)) was fit to the positive daily precipitation values (Table 1). To better represent the seasonal pattern of precipitation in this region, which is bimodal with the two peaks occurring in late spring and early autumn, separate parameter values for each calendar month within a season were calculated. The two-parameter gamma distribution was chosen for this purpose, as it is commonly used for modeling precipitation depth [23,24] and had the best fit among several alternatives that were tested. The form of the gamma distribution is

$$p(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}, \quad (1)$$

where x is the precipitation depth in millimeters, a is the shape parameter (unitless), b is the scale parameter (millimeters), and $\Gamma(\ )$ is the gamma function. Each day in the gauge time series was categorized as Wet (if daily precipitation > 0 mm) or Dry (if daily precipitation = 0 mm) (we note that these are distinct from the Dry and Wet wetness categories discussed above). For each season and wetness category, daily conditional transition probabilities were calculated for each of 4 possibilities: Wet–Wet, Wet–Dry, Dry–Dry, and Dry–Wet (for example, the Wet–Wet transition probability gives the probability that a day will experience rain given that it also rained on the previous day). As before, separate values were obtained for each calendar month within a season (Table 2). Note that Table 2 reports only two of the four sets of transition probabilities, but the Wet–Dry and Dry–Dry values can be obtained by subtracting from 1 the Wet–Wet and Dry–Wet values, respectively. To determine if rainfall occurs on a particular day, the WG first determines if rain occurred on the previous day, then selects the appropriate conditional transition probability for the wetness category and time of year, and finally uses that probability to determine if rainfall will occur on that day or not.

To incorporate the effects of large-scale climate teleconnections, as a further step, we divided the seasonal precipitation time series based on the value of the Niño-3.4 index [25], a commonly used index of the El Niño–Southern Oscillation (ENSO) phenomenon. The mean monthly Niño-3.4 index anomaly is available from 1950 to the present from NOAA, and these values were converted to seasonal averages to correspond to the seasonal precipitation totals described above. The Niño-3.4 index is considered to be in the El Niño (La Niña) phase when the Niño-3.4 anomaly is above 0.4 °C (below −0.4 °C) for 6 consecutive months and in a neutral phase otherwise [26]. In this study, we divide the time series into 3 categories based on this index using a similar breakdown, but as we are focused on the seasonal timescale, we consider the Niño-3.4 values over a 3-month period rather than a 6-month period, and for simplicity, we use an average rather than considering each

month individually. Therefore, our delineation of El Niño/La Niña periods does not match exactly what would be obtained using the traditional definition. Rather than calculating separate sets of parameters for each ENSO category, we considered that different ENSO conditions may influence whether the seasonal precipitation total falls in the Dry, Medium, or Wet category. Therefore, for each season and wetness category, the conditional transition probability of the next season being in each of the wetness categories was calculated (Table 3).

Table 1. Gamma distribution shape (*a*) and scale (*b*) parameters by season and for the Dry, Medium, and Wet categories.

	a-Dry	a-Medium	a-Wet	b-Dry	b-Medium	b-Wet
March	0.98	0.66	0.68	8.16	16.07	21.21
April	0.66	0.77	0.90	21.95	20.54	26.31
May	0.72	0.70	0.96	17.12	21.35	21.47
June	0.66	0.78	0.84	16.33	15.83	27.94
July	0.73	0.55	0.88	14.81	19.30	22.08
August	0.76	0.69	0.81	13.60	16.71	23.79
September	0.67	0.70	0.74	15.75	20.65	26.73
October	0.65	0.62	0.77	16.62	21.04	24.87
November	0.62	0.60	0.66	18.73	20.11	23.35
December	0.96	0.73	0.62	4.93	12.90	20.80
January	0.81	0.80	0.63	7.16	11.49	22.15
February	0.77	0.69	0.65	6.23	11.46	22.02

Table 2. Transition probabilities p(W | D) (Wet day conditioned on the previous day being Dry) and p(W | W) (Wet day conditioned on the previous day being Wet) by season and for the Dry, Medium, and Wet categories.

	p(W | D) Dry	p(W | D) Medium	p(W | D) Wet	p(W | W) Dry	p(W | W) Medium	p(W | W) Wet
March	0.05	0.13	0.11	0.41	0.42	0.46
April	0.13	0.14	0.20	0.38	0.52	0.62
May	0.22	0.30	0.32	0.55	0.65	0.65
June	0.21	0.28	0.23	0.40	0.52	0.51
July	0.13	0.21	0.14	0.27	0.30	0.26
August	0.17	0.23	0.22	0.35	0.44	0.41
September	0.23	0.30	0.27	0.46	0.51	0.52
October	0.21	0.28	0.26	0.45	0.54	0.49
November	0.13	0.16	0.18	0.30	0.46	0.48
December	0.06	0.09	0.14	0.41	0.37	0.39
January	0.07	0.12	0.11	0.39	0.43	0.43
February	0.08	0.09	0.13	0.36	0.32	0.50

Table 3. Conditional transition probabilities of the seasonal wetness categories, conditioned on the wetness category, and for each possible ENSO status in the previous season (see main body of text).

	p(D | D)	p(D | M)	p(D | W)	p(M | D)	p(M | M)	p(M | W)	p(W | D)	p(W | M)	p(W | W)
La Niña	0.36	0.36	0.40	0.24	0.45	0.27	0.40	0.18	0.33
Neutral	0.33	0.42	0.31	0.46	0.29	0.17	0.21	0.29	0.52
El Niño	0.27	0.11	0.24	0.36	0.63	0.38	0.36	0.26	0.38

An example of the WG output is shown in Figure 3b, in the form of a cumulative time series of the observed and 100 synthetically generated daily precipitation traces. Forecasts of precipitation were generated for a 3-month period beginning on the first day of each season (i.e., the first day of December, March, June, and September).

This approach relies on two major assumptions: that the chosen seasonal breakdown is valid and that ENSO is a dominant driver of seasonal precipitation totals. We note that by using data from as far back as 1950, possible recent trends related to climate change may be less apparent.

2.3. MAP Historical Analog

The mean areal precipitation (MAP) historical analog method extracts a time series of observed bias-adjusted satellite-based precipitation from the historical record, conditional on the CFS forecasts of total seasonal precipitation at the time of the forecast being similar to those that were present at the beginning of those time series. CFS has skill in representing large-scale atmospheric processes but lacks the ability to represent local-scale precipitation due to its coarse spatial resolution. This method assumes that large-scale conditions will give rise to similar local precipitation patterns in the present as they did in the past.

The implementation of the MAP historical analog method builds off previous studies [27–29]. It calculates the terciles of the seasonal precipitation forecasts from CFS over a reference period, which are then used to divide each season into one of three wetness categories. Then, the observed satellite-based seasonal precipitation time series beginning on the same day that the forecasts were initiated are extracted, and each is associated with a CFS wetness category. During the forecast period, the seasonal precipitation total is computed for each forecast and the corresponding category is determined. The observed time series from that category provides the set of possible forecast members, from which several are randomly selected to make up a forecast ensemble.

A single CFS grid cell is roughly centered on the Moncion basin (see Figure 1), and therefore, the precipitation output from this cell alone was used to calculate the CFS seasonal precipitation totals. We use the unaltered CFS model outputs (the CFS-RAW version of the forecasts, as described in Section 2.1) for the purpose of conditioning the historical MAP timeseries to be selected as forecasts. For the observed precipitation time series, we made use of an archive of satellite-derived bias-adjusted MAP over the Moncion basin that is available from the operational Haiti-Dominican Republic Flash Flood Guidance System (HDRFFGS) that was implemented in 2010 at ONAMET and in collaboration with the national meteorological services of Haiti and the Dominican Republic, through a program with the World Meteorological Organization [30,31]. The HDRFFGS provides forecasters in these two countries with the data and tools to develop timely and location-specific warnings for flash floods [31,32]. We use the HDRFFGS merged MAP precipitation product, which combines the best available precipitation data from (in order of priority) the satellite-based Micro-Wave Global Hydro-Estimator (MWGHE) product [31,32], the GHE product [33], and the local networks of real-time rain gauge reports. The latest operational GHE product has been available since May 2012 [31], which leaves us with 11 years of data from which to draw the MAP forecasts.

To increase the number of possible forecast members, we additionally include the CFS forecasts of seasonal precipitation, and the MAP traces for the 15th day of the month for the month before each season begins and the month during which it begins (for example, the DJF forecast for a given year includes the CFS precipitation forecasts and corresponding season-long MAP traces beginning on 15 November, 1 December, and 15 December of that year). This effectively triples the number of potential ensemble members, with the downside that the bulk of the new members is composed of a time-shifted version of the original time series.

The short period of record of MAP, 2012–2023, which essentially coincides with the period over which we aim to evaluate our forecasts, presents a challenge in producing predictions that do not make use of information that would not have been available at the

forecast initiation date. To accomplish this, when randomly selecting forecast members, we ensure that any potential members whose time ranges overlap (completely or in part) with the time period over which a forecast is being made are excluded. For instance, when generating the MAM 2013 forecast, the season-long MAP time series beginning on 15 February, 1 March, and 15 March 2013 were not considered as possible forecast members, even if they fell in the same wetness category as the CFS seasonal forecast. We did, however, calculate the terciles of CFS precipitation using all CFS forecasts over this period rather than calculating separate terciles that exclude each forecast period.

Seasonal forecasts were generated using the MAP historical analog method for each 3-month season from June 2012 to March 2023. Five ensemble members were generated for each season, though in a limited number of seasons, only four traces were available. As an example, the forecast initialized on 1 March 2014 is shown in Figure 3c, alongside the observed precipitation for that season.

2.4. Additional Data and Methods

We make use of two datasets to evaluate the seasonal precipitation forecasts. First, the Santiago Rodriguez gauge data from 2012 to 2023 was not used in any of the three methods and, therefore, provides an independent point of comparison. Second, we have also received daily inflow data to the Moncion reservoir from 2009 to 2023 from the Instituto Nacional de Recursos Hidráulicos (INDRHI) of the Dominican Republic. The reservoir's inflow data were calculated from the time series of the reservoir water level and outflows from the reservoir. The reservoir inflow exhibits a similar seasonal trend as the precipitation (the monthly averages for both precipitation and inflow are reported in Figure 2a).

3. Results

Seasonal totals were calculated from June 2012 to March 2023 (44 seasons) for each of the following: (a) the historical precipitation record from the Santiago Rodriguez station (in units of mm); (b) the Moncion reservoir inflows (in units of m^3); and (c) each ensemble member of the seasonal precipitation forecast for the three forecast methods (also in units of mm). Different ensemble sizes were available for each of the three methods: 4 members for the CFS bias-adjustment method (this is limited by the temporal resolution of the CFS model), 100 members for the WG method (determined to be sufficient to characterize the range of possible forecasts), and up to 5 members for the MAP historical analog method (it would be possible to generate more ensemble members, though maximizing the number of members would lead to many of the forecast ensembles being identical with one another).

As a first step to analyze the forecast skill, we calculate the probability of any given forecast member being in the same wetness category as the observations (each relative to their own climatology). These probabilities are computed with respect to the rain gauge observations (Table 4) and the reservoir inflow estimates (Table 5). The correct category can be selected randomly with 1/3 probability, so values greater than 0.333 indicate that the forecasts performed better than random chance (such values are shown in bold in Tables 4 and 5). Comparing these two tables reveals differences in the seasonal skill when compared to the gauge or the inflow data. The CFS-RAW and CFS-BA methods had good performance when compared to both the gauge and inflow datasets in 3 out of 4 seasons. The WG performed better than random chance in 3 and 2 seasons when compared to the gauge and inflow data, respectively. The MAP historical analog method had a noticeable difference between the two datasets, with values above 0.333 for all seasons when compared to the gauge but only for 2 seasons when compared to the inflow. Using this table, we can see that all methods had skill compared to the gauge in MAM and JJA, with 2 having skill in SON and 3 in DJF. Compared to the inflow, all methods had skill in SON, but only the CFS-based methods had skill in MAM, and only the MAP historical analog in DJF.

Table 4. Fraction of forecasts where the tercile of the forecasts matched that of the gauge observations, broken down by season. Bold entries indicate seasons when the forecasts performed better than random chance (i.e., had a value greater than 0.333).

	MAM	JJA	SON	DJF
CFS-RAW	**0.364**	**0.341**	**0.350**	0.295
CFS-BA	**0.386**	**0.341**	0.275	**0.364**
WG	**0.411**	**0.499**	0.287	**0.369**
MAP	**0.400**	**0.382**	**0.400**	**0.345**

Table 5. Fraction of forecasts where the tercile of the forecasts matched that of the reservoir inflow observations, broken down by season. Bold entries indicate seasons when the forecasts performed better than random chance (i.e., had a value greater than 0.333).

	MAM	JJA	SON	DJF
CFS-RAW	**0.396**	**0.432**	**0.409**	0.205
CFS-BA	**0.396**	**0.432**	**0.364**	0.227
WG	0.298	**0.400**	**0.350**	0.284
MAP	0.233	0.327	**0.418**	**0.473**

Given the uncertain nature of forecasts at the seasonal timescale, metrics that measure the skill of probabilistic forecasts often give a better indicator of forecast skill than deterministic metrics. While Tables 4 and 5 give a general indication of model performance relative to random chance, we can also use a metric that explicitly compares the model skill relative to climatology, the Ranked Probability Skill Score (*RPSS*) [34,35], which has been widely used in similar studies (e.g., [2,36,37]), to evaluate the relative skill of the three forecast methods described in this study. We use the debiased form of *RPSS* (*RPSS-D*), as described in [35], which corrects for the negative bias of *RPSS* for small ensemble sizes. This correction is particularly significant for the CFS and MAP historical analog methods, both of which have relatively small ensemble sizes (4 and 5 ensemble members per forecast, respectively).

The formula for *RPSS-D* is given by

$$RPSS\text{-}D = 1 - \frac{\langle RPS \rangle}{\langle RPS_{CLIM} \rangle + D}\,, \tag{2}$$

where <*RPS*> and <RPS_{CLIM}> are the average Ranked Probability Scores (*RPS*) of the forecast and the climatology, respectively, and *D* is a factor that adjusts RPSS for small ensemble sizes and is dependent on the number of ensemble members and probability categories [35]. Note that when *D* is equal to 0, Equation (2) reverts to the standard *RPSS* formula. The formula for *RPS* is

$$RPS = \sum_{k=1}^{K} (Y_k - O_k)^2, \tag{3}$$

where *K* is the number of probability categories in the forecast (*K* = 3 if terciles are used, as in this study) and Y_k and O_k are the forecast and observation probability vectors, respectively (a description of how these vectors are calculated is provided below–see [35] for a complete description).

To calculate *RPSS-D*, the terciles of the seasonal totals of forecasted precipitation, observed precipitation, and reservoir inflow were calculated. These terciles are used to categorize each ensemble member as being in the Dry, Medium, or Wet wetness category. The number of members in each category are then summed and divided by the size of the ensemble, providing the probability of the forecast being in each category, which is then summed to find the corresponding cumulative probability. For example, suppose a forecast has 5 ensemble members, with 2 falling in the Dry category, 1 in the Medium category, and 2 in the Wet category. Converting to a probabilistic forecast vector yields (2/5, 1/5, 2/5) as the forecast probabilities for each of the 3 categories and a cumulative forecast probability

of (2/5, 3/5, 5/5) = (0.4, 0.6, 1). It is this latter set of values that is used in the calculation of *RPSS-D*.

For the observation-based time series, the conversion to a probabilistic forecast vector involves determining the wetness category into which the observation falls, assigning a probability of 1 to that category, and then finding the cumulative probability. For example, if an observation is found to be in the Medium wetness category, its probability vector is (0, 1, 0), and its cumulative probability vector is (0, 1, 1).

An *RPSS-D* value of 1 indicates a perfect forecast (the forecast always falls in the same category as the observation), while a value of 0 indicates that the forecast has the same skill as the reference climatology (the forecast falls in each category 1/3 of time). Forecasts that have some degree of skill beyond that of the climatology have *RPSS-D* values between 0 and 1, and forecasts with negative values have less skill than would be achieved by the climatology. Between 0 and 1, there is no strict definition of a 'good' *RPSS-D* value, though they may be quite low—for example, [38] provides an approximate relationship between correlation and *RPSS-D*, which suggests that a correlation of 0.5 between forecasts and observations translates to an *RPSS-D* value of roughly 0.13.

We first consider the seasonal *RPSS-D* values of the three methods as compared to the precipitation gauge data, which provides the most direct comparison available to the precipitation predictions. These results are reported in Figure 4a. The MAP historical analog method had the best overall performance, with positive *RPSS-D* values (skill relative to the climatology) in all seasons and the highest skill in MAM and SON, the two wettest seasons of the year. The WG method exhibited the lowest overall skill of the three methods considered, with a positive *RPSS-D* value in JJA and negative values in other seasons, as well as the most variability in skill. The CFS-BA method had positive average *RPSS-D* values throughout the year (though slightly lower than those for the MAP historical analog), with the highest values occurring in DJF and MAM.

Next, we examine the seasonal *RPSS-D* values of the three methods by comparing the forecasts with inflow data to the Moncion reservoir (Figure 4b). The overall trend of the MAP historical analog and CFS methods performing better than the WG method is similar to the results in Figure 4a, with some differences, including that CFS-BA outperforms the MAP historical analog in some seasons and has significantly worse performance in DJF. WG method again showed a small amount of skill in JJA and negative values in other seasons. As we are now comparing two different variables (forecasted precipitation and observed inflow), we emphasize that this figure reports the ability of the forecasted precipitation to land in the same wetness category (relative to the forecast climatology) as the reservoir inflow (relative to its own climatology). This comparison of two linked but distinct variables likely accounts for at least some of the decline in performance when comparing the forecasts to inflow rather than precipitation data.

Next, we further break down the previous results by dividing them by seasons in which the raw CFS forecasts (CFS-RAW) did or did not have skill (i.e., had positive or negative *RPSS-D* values). This serves to shed more light on the differences in performance between the methods and possibly the predictive ability of the sources used by each. Figure 5a shows the skill of the three methods when CFS-RAW did have skill. As expected, CFS-BA had the best performance, while the MAP historical analog method still had positive skill in all seasons. WG showed a significantly different seasonal pattern of skill, with positive values in all seasons except SON and the highest skill now in MAM rather than JJA. When CFS-RAW did not have skill (Figure 5b), CFS-BA performed quite poorly, with less average skill than climatology in all seasons. This implies that the bias adjustments were not sufficient to overcome unskillful raw CFS forecasts. As in Figure 5a, Figure 5b shows positive (though lower) values for the MAP historical analog method in all seasons. WG had a similar trend as that found in Figure 4a, with positive *RPSS-D* values only in JJA.

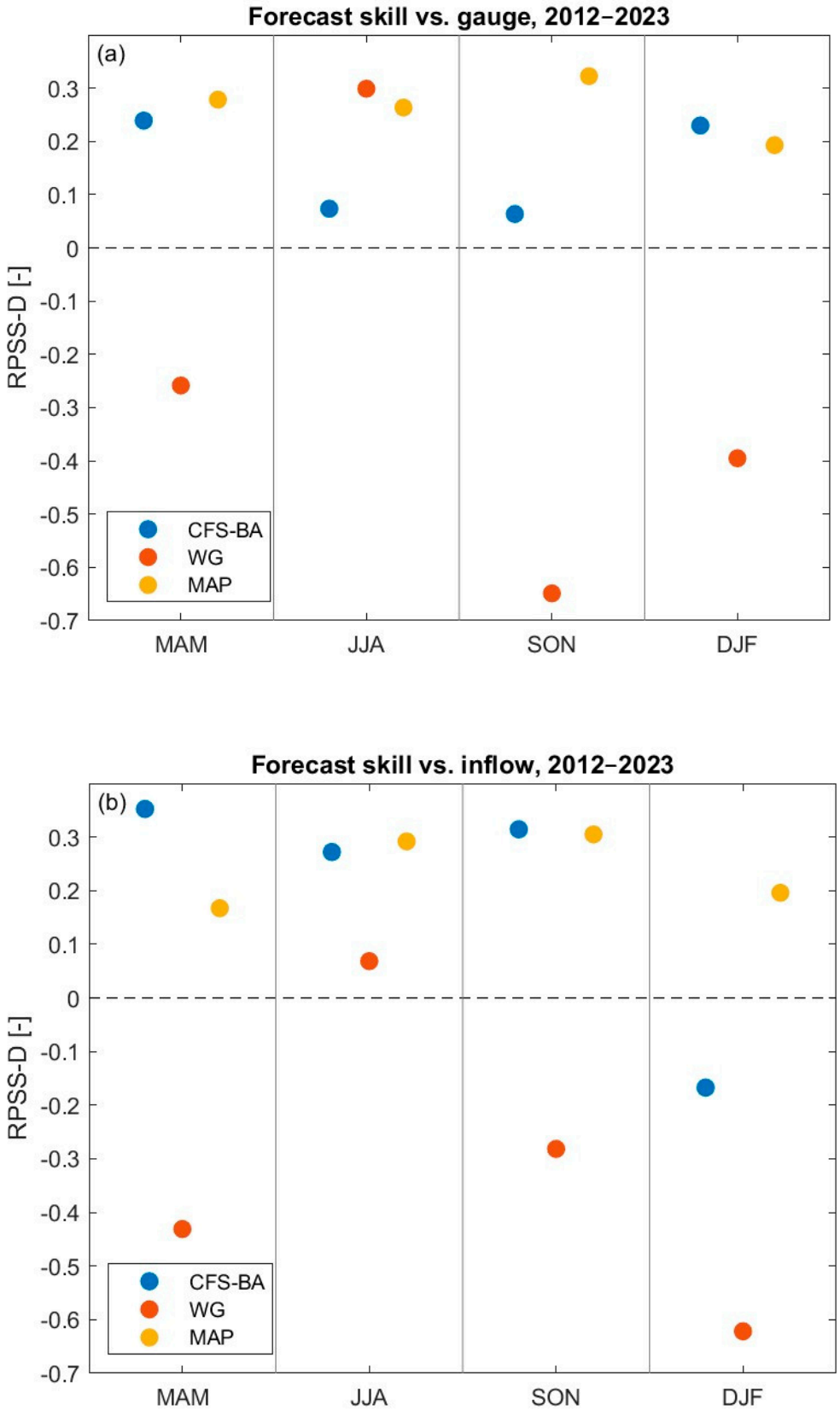

Figure 4. *RPSS-D* of each method by season, calculated by comparing the forecasts to (**a**) the precipitation gauge record and (**b**) the inflow to the Moncion basin.

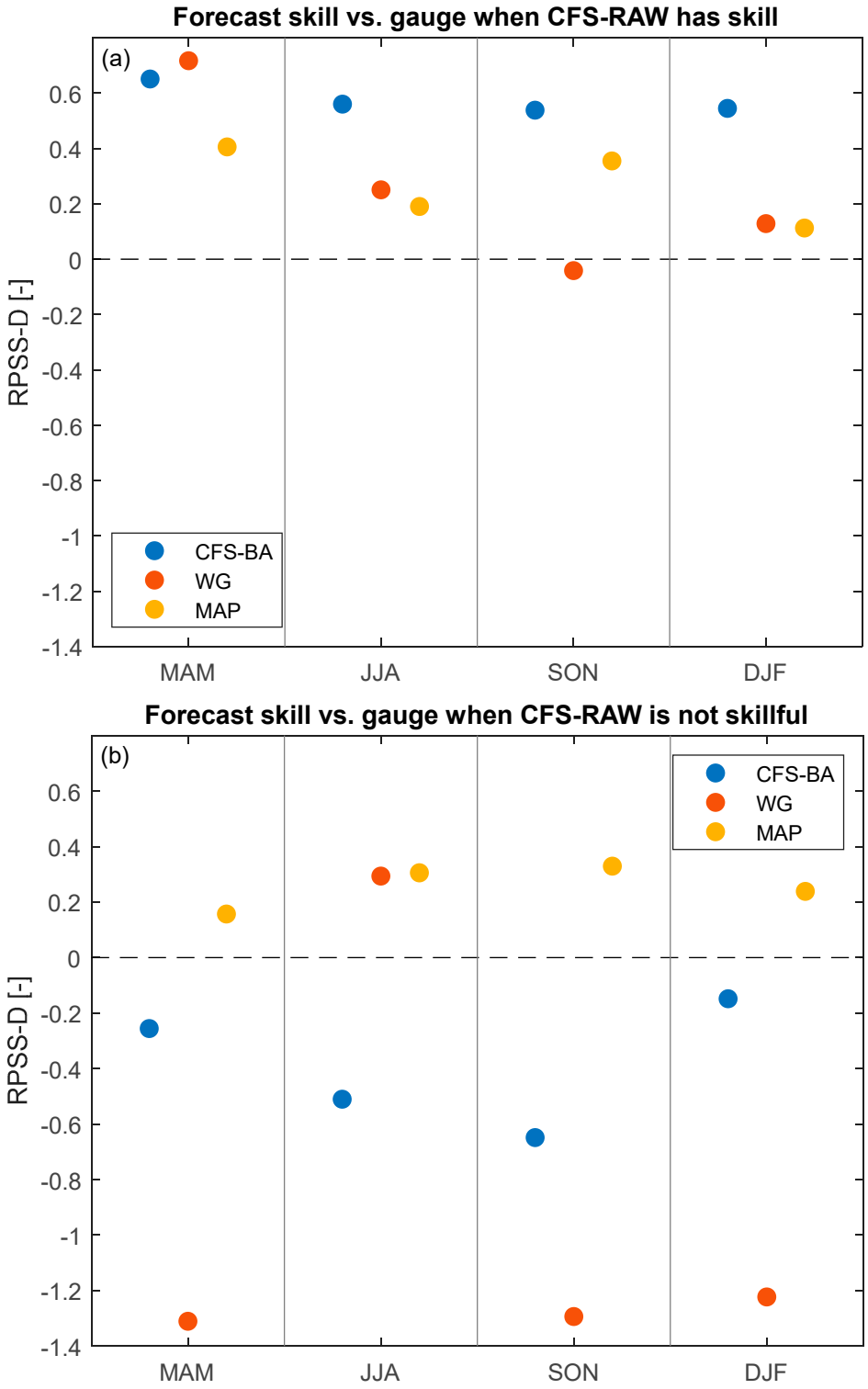

Figure 5. *RPSS-D* of each method by season, calculated by comparing the forecasts to the precipitation gauge record, for (**a**) seasons when the raw CFS forecast had skill (*RPSS-D* > 0) and (**b**) seasons when the raw CFS forecast had skill (*RPSS-D* > 0).

4. Discussion

The motivation for this study was to determine the relative performance of contrasting methods for producing seasonal precipitation forecasts, which will assist in the development of an operational S2S forecast system for the Yaque del Norte River basin. While we were not able to identify a single option that always gives the best performance throughout

the year, this study provides valuable information regarding the relative merits of the methods considered. Most importantly, the MAP historical analog method had relatively strong performance across seasons and for both seasonal precipitation and reservoir inflow when compared to the CFS and WG methods. However, the MAP historical analog method requires a sufficiently long historical archive that can be used for analog selection. The relatively good average performance of the CFS method and the ready availability of the CFS product also indicate it as a favorable alternative for generating seasonal-scale precipitation forecasts. Further refinements to the bias-adjustment method have the potential to improve the skill of this forecast. The reason that the WG method had the worst overall performance of the three methods, with positive *RPSS-D* values only in JJA, could be attributed to the ability of the chosen ENSO index to predict seasonal precipitation at other times of the year.

When conditioning the performance of the 3 methods by the performance of CFS-RAW, it is unsurprising that the CFS-BA method had the best (worst) performance when CFS-RAW had (did not have) skill, given their close relationship. Interestingly, despite being dependent on the CFS-RAW precipitation values for the conditional selection of MAP traces, the MAP historical analog method did not exhibit a notable relationship with the skill of the CFS-RAW forecasts, with positive *RPSS-D* values for all seasons regardless of the skill of CFS-RAW. This suggests that while the MAP historical analog method is producing more realistic seasonal totals than the CFS method, the CFS clearly has some ability to predict the conditions that give rise to particular MAP traces, perhaps by accurately capturing information about large-scale atmospheric conditions. The WG method also exhibited some differences, notably having much higher skill when CFS-RAW was skillful, which could be interpreted as the CFS forecasts having skill related to the external drivers of the WG method (i.e., ENSO and season-to-season wetness categories).

The short time period (11 years) of the forecast evaluation period was unavoidable due to the limited availability of much of the data needed to generate the forecasts (in particular, the CFS operational forecasts have only been available since 2011, and the MAP record since 2012). However, it is possible that this affected the results in ways that adjustments to the methods could alleviate. Using only a single rain gauge, as we have in this study, raises the possibility that biases or missing dates in this gauge could skew the results. We have attempted to address this by including the reservoir inflow data as an additional point of comparison. Also, we have not attempted to correct for long-term climate trends, which could have affected the performance of both the CFS and WG methods—a separate examination of the mean monthly rainfall before and after 2012 revealed differences that could have negatively impacted the performance of these methods (the WG and CFS methods use gauge records as far back as the 1930s and 1980s, respectively).

In the following, a number of extensions and methodological improvements are discussed. To begin, rather than focusing on seasonal aggregation, shorter time periods (e.g., monthly) could be considered, along with more start dates for each forecast (such as initializing a forecast at the beginning of each month rather than at the beginning of each season). The use of the CFS precipitation forecasts might be improved by alternate methods of bias-adjusting CFS (particularly by the use of a quantile mapping-type method, such as that described in [15], which corrects for variance in the precipitation distribution in addition to mean bias). For the weather generator method, one alternative is to include other operationally available information, such as an index of the North Atlantic Oscillation, alongside or in place of the ENSO index used in this study. Additionally, forecasts of ENSO indices are available, which could be used along with previous-season values to better parameterize the seasonal transition coefficients. The MAP historical analog method would be improved if this study were repeated in the future when more traces of satellite precipitation are available (more available traces would also make it feasible to condition their selection on additional variables, in particular, the CFS forecasts of surface temperature).

Future efforts will address these improvements and will also benefit from additional data that will be available in future years. They will also be informed by feedback from

users of the initial operational version of the S2S forecast method in the Dominican Republic. While the MAP method had the strongest overall performance, followed by the CFS and WG methods, the results indicate that the forecast skill varies by season and by validation dataset (either precipitation or reservoir inflow). For the deployment of an S2S precipitation forecast system in the Yaque del Norte River basin, we recommend developing a system that carefully combines the advantages of the three methods into one forecast ensemble. A better characterization of the skill of the available S2S forecasts, as well as the possibility of developing more skillful forecasts that combine different elements from multiple forecast methods, will allow for improved decision-making by water resources managers.

Author Contributions: Conceptualization, N.P., E.S. and T.M.H.; methodology, N.P., E.S. and T.M.H.; formal analysis, N.P. and E.S., writing—original draft preparation, N.P.; writing—review and editing, N.P., E.S., T.M.H. and Z.C.; visualization, N.P.; supervision, E.S. and T.M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Capacity Development for Weather, Water, and Climate Forecasting and Disaster Risk Reduction U.S. NWS Grant to the Hydrologic Research Center (NA22NWS4670002). The APC was funded by a waiver from MDPI.

Data Availability Statement: The CFS data used in this study are publicly available from the US National Centers for Climate Information (NCEI). This study also used satellite-derived MAP and precipitation gauge from the Oficina Nacional de Meteorología and reservoir inflow data from the Instituto Nacional de Recursos Hidráulicos (INDRHI) of the Dominican Republic. For data requests, interested individuals should contact these agencies directly.

Acknowledgments: The authors are grateful to Wagner Rivera and the Oficina Nacional de Meteorología (ONAMET) and to Israel Acosta and the Instituto Nacional de Recursos Hidráulicos (INDRHI) of the Dominican Republic for sharing the precipitation and reservoir inflow data that were used in this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Robertson, A.W.; Kumar, A.; Peña, M.; Vitart, F. Improving and Promoting Subseasonal to Seasonal Prediction. *Bull. Am. Meteorol. Soc.* **2015**, *96*, ES49–ES53. [CrossRef]
2. Yuan, X.; Wood, E.F.; Luo, L.; Pan, M. A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys. Res. Lett.* **2011**, *38*, L13402. [CrossRef]
3. Crochemore, L.; Ramos, M.-H.; Pappenberger, F. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3601–3618. [CrossRef]
4. Pierce, D.W.; Cayan, D.R.; Thrasher, B.L. Statistical downscaling using localized constructed analogs (LOCA). *J. Hydrometeorol.* **2014**, *15*, 2558–2585. [CrossRef]
5. Zhang, L.; Kim, T.; Yang, T.; Hong, Y.; Zhu, Q. Evaluation of Subseasonal-to-Seasonal (S2S) precipitation forecast from the North American Multi-Model ensemble phase II (NMME-2) over the contiguous US. *J. Hydrol.* **2021**, *603*, 127058. [CrossRef]
6. Cohen, J.; Coumou, D.; Hwang, J.; Mackey, L.; Orenstein, P.; Totz, S.; Tziperman, E. S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdiscip. Rev. Clim. Change* **2019**, *10*, e00567. [CrossRef]
7. Izzo, M.; Aucelli, P.P.C.; Maratea, A. Historical trends of rain and air temperature in the Dominican Republic. *Int. J. Climatol.* **2020**, *41*, E563–E581. [CrossRef]
8. Klotzbach, P.J. The Influence of El Niño–Southern Oscillation and the Atlantic Multidecadal Oscillation on Caribbean Tropical Cyclone Activity. *J. Clim.* **2011**, *24*, 721–731. [CrossRef]
9. Giannini, A.; Kushnir, Y.; Cane, M.A. Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. *J. Clim.* **2000**, *13*, 297–311. [CrossRef]
10. Taylor, M.A.; Enfield, D.B.; Chen, A.A. Influence of the tropical Atlantic versus the tropical Pacific on Caribbean rainfall. *J. Geophys. Res. Ocean.* **2002**, *107*, 10-1–10-14. [CrossRef]
11. Ashby, S.A.; Taylor, M.A.; Chen, A.A. Statistical models for predicting rainfall in the Caribbean. *Theor. Appl. Climatol.* **2005**, *82*, 65–80. [CrossRef]
12. Miller, P.W.; Ramseyer, C.A. Did the Climate Forecast System Anticipate the 2015 Caribbean Drought? *J. Hydrometeorol.* **2020**, *21*, 1245–1258. [CrossRef]

13. Martinez, C.; Muñoz, Á.G.; Goddard, L.; Kushnir, Y.; Ting, M. Seasonal prediction of the Caribbean rainfall cycle. *Clim. Serv.* **2022**, *27*, 100309. [CrossRef]
14. Saha, S.; Moorthi, S.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Behringer, D.; Hou, Y.-T.; Chuang, H.-y.; Iredell, M.; et al. The NCEP Climate Forecast System Version 2. *J. Clim.* **2014**, *27*, 2185–2208. [CrossRef]
15. Wood, A.W.; Leung, L.R.; Sridhar, V.; Lettenmaier, D. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim. Chang.* **2004**, *62*, 189–216. [CrossRef]
16. Abatzoglou, J.T.; Brown, T.J. A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.* **2012**, *32*, 772–780. [CrossRef]
17. Salathé Jr, E.P.; Mote, P.W.; Wiley, M.W. Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States Pacific Northwest. *Int. J. Climatol. A J. R. Meteorol. Soc.* **2007**, *27*, 1611–1621. [CrossRef]
18. Manzanas, R.; Lucero, A.; Weisheimer, A.; Gutiérrez, J.M. Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts? *Clim. Dyn.* **2018**, *50*, 1161–1176. [CrossRef]
19. Wilks, D.S.; Wilby, R.L. The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr.* **1999**, *23*, 329–357. [CrossRef]
20. Shamir, E.; Halper, E.; Modrick, T.; Georgakakos, K.P.; Chang, H.-I.; Lahmers, T.M.; Castro, C. Statistical and dynamical downscaling impact on projected hydrologic assessment in arid environment: A case study from Bill Williams River basin and Alamo Lake, Arizona. *J. Hydrol. X* **2019**, *2*, 100019. [CrossRef]
21. Shamir, E.; Meko, D.M.; Graham, N.E.; Georgakakos, K.P. Hydrologic Model Framework for Water Resources Planning in the Santa Cruz River, Southern Arizona1. *JAWRA J. Am. Water Resour. Assoc.* **2007**, *43*, 1155–1170. [CrossRef]
22. Shamir, E.; Wang, J.; Georgakakos, K.P. Probabilistic Streamflow Generation Model for Data Sparse Arid Watersheds1. *JAWRA J. Am. Water Resour. Assoc.* **2007**, *43*, 1142–1154. [CrossRef]
23. Ye, L.; Hanson, L.S.; Ding, P.; Wang, D.; Vogel, R.M. The probability distribution of daily precipitation at the point and catchment scales in the United States. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6519–6531. [CrossRef]
24. Martinez-Villalobos, C.; Neelin, J.D. Why Do Precipitation Intensities Tend to Follow Gamma Distributions? *J. Atmos. Sci.* **2019**, *76*, 3611–3631. [CrossRef]
25. Rayner, N.A.; Parker, D.E.; Horton, E.B.; Folland, C.K.; Alexander, L.V.; Rowell, D.P.; Kent, E.C.; Kaplan, A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* **2003**, *108*, 1. [CrossRef]
26. Trenberth, K.E. The definition of el nino. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 2771–2778. [CrossRef]
27. Carpenter, T.M.; Georgakakos, K.P. Assessment of Folsom lake response to historical and potential future climate scenarios: 1. Forecasting. *J. Hydrol.* **2001**, *249*, 148–175. [CrossRef]
28. Yao, H.; Georgakakos, A. Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management. *J. Hydrol.* **2001**, *249*, 176–196. [CrossRef]
29. Georgakakos, K.P.; Graham, N.E.; Modrick, T.M.; Murphy, M.J.; Shamir, E.; Spencer, C.R.; Sperfslage, J.A. Evaluation of real-time hydrometeorological ensemble prediction on hydrologic scales in Northern California. *J. Hydrol.* **2014**, *519*, 2978–3000. [CrossRef]
30. Shamir, E.; Georgakakos, K.P.; Spencer, C.; Modrick, T.; Murphy, M.; Jubach, R. Evaluation of real-time flash flood forecasts for Haiti during the passage of Hurricane Tomas, November 4–6, 2010. *Nat. Hazards* **2013**, *67*, 459–482. [CrossRef]
31. Georgakakos, K.P.; Modrick, T.M.; Shamir, E.; Campbell, R.; Cheng, Z.; Jubach, R.; Sperfslage, J.A.; Spencer, C.R.; Banks, R. The Flash Flood Guidance System Implementation Worldwide: A Successful Multidecadal Research-to-Operations Effort. *Bull. Am. Meteorol. Soc.* **2022**, *103*, E665–E679. [CrossRef]
32. Georgakakos, K.P.; Graham, R.; Jubach, R.; Modrick, T.; Shamir, E.; Spencer, C.; Sperfslage, J.A. Global Flash Flood Guidance System, Phase I. *HRC Technol. Rep.* **2013**, *9*, 134.
33. Scofield, R.A.; Kuligowski, R.J. Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Weather Forecast.* **2003**, *18*, 1037–1051. [CrossRef]
34. Epstein, E.S. A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol. (1962–1982)* **1969**, *8*, 985–987. [CrossRef]
35. Weigel, A.P.; Liniger, M.A.; Appenzeller, C. The Discrete Brier and Ranked Probability Skill Scores. *Mon. Weather Rev.* **2007**, *135*, 118–124. [CrossRef]
36. Wang, H. Evaluation of monthly precipitation forecasting skill of the National Multi-model Ensemble in the summer season. *Hydrol. Process.* **2014**, *28*, 4472–4486. [CrossRef]
37. Vigaud, N.; Tippett, M.K.; Robertson, A.W. Probabilistic Skill of Subseasonal Precipitation Forecasts for the East Africa–West Asia Sector during September–May. *Weather Forecast.* **2018**, *33*, 1513–1532. [CrossRef]
38. Tippett, M.K. Comments on "The Discrete Brier and Ranked Probability Skill Scores". *Mon. Weather Rev.* **2008**, *136*, 3629–3633. [CrossRef]