Seasonal forecasting of monsoon precipitation characteristics using weather types and generalized linear modeling

Erin Towler, Project Scientist, National Center for Atmospheric Research (NCAR), Boulder, CO, towler@ucar.edu
Dagmar Llewellyn*, Civil Engineer (Hydrologic), US Dept. of Interior Bureau of Reclamation (Reclamation), Albuquerque, NM, dllewellyn@usbr.gov
Andreas Prein, Project Scientist, NCAR, Boulder, CO, prein@ucar.edu
Lucas Barrett*, Civil Engineer (Hydrologic), Reclamation, Albuquerque, NM, lbarrett@usbr.gov
*Opinions expressed represent those of the authors and not necessarily those of the Bureau of Reclamation.

Abstract

River Basins in New Mexico and Arizona are heavily impacted by monsoon season precipitation. Seasonal forecasts of monsoon precipitation for the US Southwest are not typically skillful, but forecasts of recurring large-scale weather patterns, or "weather types" have shown promise. In this study, we develop an experimental monsoon precipitation forecast using weather types developed for Arizona and New Mexico. We use a generalized linear modeling statistical framework with historical reanalysis data to develop functional relationships between monsoon-season precipitation and the number of days associated with specific weather types. Specifically, we predict the categorical precipitation likelihood (i.e., above- or below-median, or above-average, average, or below-average tercile). Further, using hindcasts (i.e., retrospective forecasts) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), we demonstrate when these forecasts are skillful as compared to climatology. Finally, we describe an online Google Colab Notebook that has been developed to allow managers to download real-time ECMWF forecasts, assign the weather types associated with each forecast day, and make probabilistic precipitation predictions.

Introduction

Previous reviews of forecasting products for the US Southwest indicate that seasonal forecasts tend to underpredict monsoonal precipitation (Hartmann et al. 1999) and recent work shows that available monsoon precipitation forecasts are not skillful (Prein et al. 2022). This is not surprising given the small-scale processes that contribute to monsoonal convection, which are not resolved at the coarse spatial scales at which most forecast models are run. However, monsoonal moisture can be a critical component of summertime water supply in the Southwestern US (Towler et al. 2019), and key water management decisions are made in late spring and early summer based on monsoon forecasts. This study seeks to improve these seasonal monsoon forecasts.

To understand monsoonal changes, Seneviratne et al. (2012) recommend the consideration of large-scale circulation and dynamics, rather than just precipitation. One appealing approach is to identify large-scale atmospheric patterns that can be related statistically to precipitation (Maraun et al. 2010; Wilby et al. 2004). Prein et. al. (2016) identified so-called "weather types"

(WTs), or large-scale atmospheric patterns that are associated with precipitation, and developed WTs for the continental US to examine recent precipitation trends. Prein and Mearns (2021) identify extreme-precipitation-producing WTs for major watersheds in the continental US. Towler et al. (2020) use WTs developed for New Mexico with extreme value theory to characterize extreme monsoonal precipitation. In Prein et al. (2022), WTs for Arizona and New Mexico monsoon seasons were developed and shown to skillfully capture monsoonal moisture in retrospective forecasts produced by the European Centre for Medium-Range Weather Forecasts (ECMWF).

This study uses the WTs developed in Prein et. al. (2022) to develop probabilistic forecasts of monsoon precipitation. We use a generalized linear modeling (GLM) statistical framework with historical reanalysis data to develop functional relationships between the number of days associated with specific WTs and monsoon-season precipitation characteristics. Specifically, we predict the categorical precipitation likelihood (i.e., above- or below-median, or above-average, average, or below-average tercile). Further, we utilize ECMWF hindcasts (i.e., retrospective forecasts) to quantify the skill of this approach as compared to climatology at different forecast lead times. Finally, we describe an online Google Colab Notebook that has been developed to allow operators to download real-time ECMWF forecasts, assign the WTs associated with each forecast day, and make experimental precipitation predictions.

Data

Region and Season

Precipitation associated with the North American Monsoon (NAM) exhibits spatial variability (Castro et al., 2012; Ciancarelli et al., 2014). Our analysis examines four regions affected by the NAM (Figure 1): western and eastern Arizona (AZ-West and AZ-East) and northern and southern New Mexico (NM-North and NM-South). These regions include catchments that are important for water management in the region: the AZ regions are relevant to the Lower Colorado River Basin and the NM regions are relevant to the Upper Rio Grande and Upper Pecos watersheds. For each region, catchments are combined based on a clustering assessment conducted in Prein et al. (2022). AZ-West contains HUC1501, HUC1503, HUC1507, HUC1810, and AZ-East contains HUC1502, HUC1504, HUC1506, HUC1508. NM-North includes HUC1301, HUC140801, HUC130201, HUC130202, and NM-South contains HUC130301 and HUC1306.

This study focuses on monsoon precipitation associated with the NAM for the months of June through October, examining individual and multi-month prediction periods. In total, there were 14 prediction periods considered: June through October (JJASO), July through October (JASO), June through August (JJA), July through September (JAS), August through October (ASO), June through July (JJ), July through August (JA), August through September (AS), September through October (SO), June (Jun), July (Jul), August (Aug), September (Sep), and October (Oct).

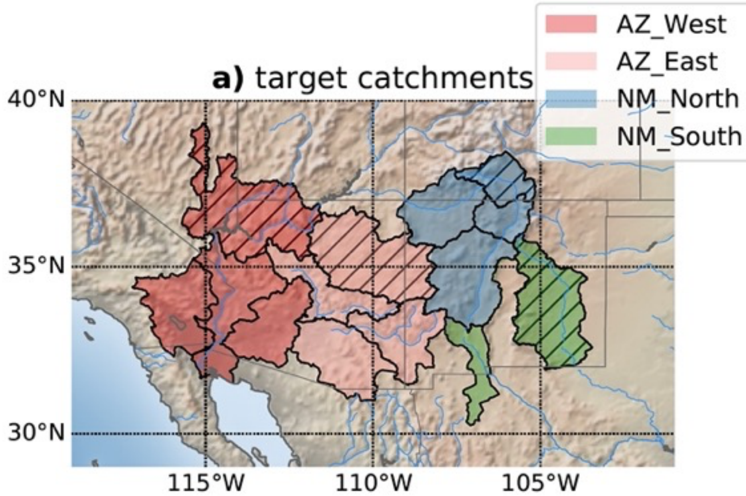

Figure 1. Hydrologic unit codes (HUCs) that exhibit similar weather types (WTs) are combined to create 4 regions: Arizona West (AZ-West, red), Arizona East (AZ-East, pink), New Mexico North (NM-North, blue) and New Mexico South (NM-South, green). Hatching shows the basin used to derive the WTs. Reproduced from Prein et al. (2022).

Predictor, Forecasts, and Predictand Data

The predictors used in this study were based on WTs that were defined in Prein et al. (2022). Prein et al. (2022) uses historical reanalysis data to define synoptic-scale WTs to characterize NAM rainfall variability for the same regions in AZ and NM that are analyzed in this paper. To characterize the WTs, daily average atmospheric variables from ECMWF's Interim Reanalysis (Dee et al. 2011) within the period 1982 to 2018 were examined. Results from Prein et al. (2022) show that the best available variable to characterize the WTs is synoptic-scale moisture advection, as represented by the water vapor mixing ratio at 850 hPa (Q850). For each region, Q850 is used in a clustering technique to identify three distinct WTs, i.e., days with dry, normal, or wet (monsoonal) warm season flow patterns. For this study, the predictors considered are the sum of the number of days associated with each defined WT (i.e., dry, normal, or wet) for the prediction period.

The next step was to obtain seasonal forecasts of Q850. Initially, we examined seasonal forecasts from the North American Multi-Model Ensemble (NMME; Kirtman et al. (2014)) and from ECMWF's Integrative Forecasting System (IFS, Version 5). However, Prein et al. (2022) found that the NMME did not produce skillful Q850 forecasts, so only the ECMWF forecasts were used here. We downloaded seasonal forecasts of Q850 from ECMWF through the Copernicus Climate Change Service (C3S; https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-pressure-levels?tab=form), including retrospective forecasts for 1993–2016 and operational forecasts from 2017-2018. These were pooled together, from 1993-2018, and are referred to in this paper as the ECMWF hindcasts. Each ECMWF forecast is initialized on the first of the month and runs an ensemble forecast that includes 25 members. Each ensemble member is run 215 days out (~7 months). Using the same clustering technique that is used for the historical reanalysis, each forecast day is assigned to a WT (dry, normal, wet).

In this study, precipitation characteristics (i.e., above- or below-median, or above-average, average, or below-average tercile precipitation) were the targets for prediction. To calculate historical precipitation statistics, we use PRISM (Daly et al. 1997), which is a gridded 4-

kilometer (km) observational dataset. The dataset is available from 1982-2018, but we used 1993-2018 as the climatological period to overlap with the available ECMWF hindcasts.

Methodology

Generalized Linear Modeling

The predictive statistical framework used in this application is the Generalized Linear Model (GLM). In GLM, the response variable, Y, can be assumed to be from a distribution in the exponential family, with the specific distribution depending on the response being predicted (continuous, discrete, categorical, etc.). A link function is used to specify the distribution and relate the expected value of Y to a set of predictors (McCullagh and Nelder 1989):

$$G(E(Y)) = X\beta^T + e \quad \text{(Equation 1)}$$

Where G(.) is the link function, E(Y) is the expected value of the predictand, β^T is the transposed vector of fitted model coefficients, X is the predictor matrix, and e is the error. An appropriate link function is identified based on the attributes of the predictand. In this case, we use the logit link function because we are predicting categorical responses, and the logit link function converts the distribution of values into a scale of probability. Specifically, we are interested in the likelihood that (i) precipitation is above (or below) the climatological median and (ii) precipitation is in the above-average, average, or below-average climatological tercile. For the former, the binomial distribution is appropriate, with the logit link function (i.e., logistic regression). In that case, the predictand (i.e., precipitation) is set to a categorical value of "1" if the value is greater than the climatological median (Q50) and "0" if the value is lower. For the latter, the multinomial logit, an extension of logistic regression, is used (i.e., multinomial regression). We use the proportional odds model, since the tercile categories are ordered (if the categories were not ordered, we would use the ordinal multinomial). The predictand is assigned based on the climatological terciles: "1" if it is less than or equal to the 33^{rd} percentile (Q33), "2" if it is between Q33 and the 66^{th} percentile (Q66), and "3" if it is above or equal to Q66.

McCullagh and Nelder (1989) provide details on distributions and link functions, as well as on coefficient estimation. Here, the GLMs were fitted in the R package VGAM (Yee and Moler 2022) using the vector generalized linear models (vglm) function. For the logistic, *family = binomialff*, and for the multinomial, *family = propodds*. The coefficients can be estimated and applied internally in the VGAM package. To use the estimated coefficients directly, the following equations are employed. From Helsel and Hirsh (1995), the probability of exceeding the median, Q50, is estimated as:

$$P(Y > Q50) = \frac{\exp\,(B_0 + B_1 x)}{(1 + \exp\,(B_{0+} B_1 x)} \quad (\textit{Equation 2})$$

From McNulty (2022), to predict the probability of being in the lowest ordered tercile, i.e., less than or equal Q33:

$$P(Y \leq Q33) = \frac{\exp\,(-1 * (B_{0,1} + B_1 x))}{(1 + \exp\left(-1 * (B_{0,1} + B_1 x)\right))} \quad (\textit{Equation 3})$$

where $B_{0,1}$ is the first intercept, and B_1 is the slope. To predict the probability of being in the upper tercile, i.e., greater than or equal to Q66:

$$P(Y \geq Q66) = \frac{1}{(1+\exp\left(-1*\left(B_{0,2}+B_1 x\right)\right))} \quad (\textit{Equation 4})$$

where $B_{0,2}$ is the second intercept and B_1 is the slope. The estimated slope is the same for both equations. Finally:

$$P(Q66 > Y > Q33) = 1 - (P(Y \leq Q33) + P(Y \geq Q66)) \quad (\textit{Equation 5})$$

As mentioned in the Data section, the predictor, x, was based on the WTs developed in Prein et al. (2022). We only allow univariate regression (i.e., a single predictor), which could be either the sum of the number of dry WT days (sumDry) or the sum of the number of monsoon WT days (sumMonsoon) over the prediction period. Further, the predictand and predictors were standardized; this is shown here for the predictor:

$$x_{Stand} = \frac{x - avg(x)}{sd(x)} \quad (\textit{Equation 6})$$

where x_{Stand} is the standardized variable, *avg(x)* is the variable average and *sd(x)* is the variable standard deviation.

Evaluation Metrics

To evaluate the relative performance of our forecasts compared to a reference forecast, two skill scores were applied: the Brier Skill Score (BSS) (Wilks, 1995) and the ranked probability skill score (RPSS) (Wilks, 1995). Climatology was used as the reference forecast (see details in subsequent paragraphs). The BSS is used to evaluate the performance of the categorical forecast from the logistic regression:

$$BSS = 1 - \frac{BS_{Forecast}}{BS_{Climatology}} \quad (\textit{Equation 7})$$

where the $BS_{Forecast}$ is the Brier Score (BS) for the forecast, defined as:

$$BS_{Forecast} = \frac{\sum_{i=1}^{N}(p_i - o_i)^2}{N} \quad (\textit{Equation 8})$$

where p_i refers to the forecast probabilities, o_i refers to the observed probabilities (o_i = 1 if the observed precipitation exceeds the median, 0 otherwise), and N is the sample size (i.e., number of years). $BS_{Climatology}$ is the BS for climatology, which is also calculated from the above equation, but for every year uses climatological probabilities, i.e., p_i 0.50 (since there are two categories: above or below the median). BSS values range from negative infinity to 1. BSS<0 indicates that the forecast has less skill than climatology (equal chances). BSS 0 indicates equal skill, and a BSS>0 indicates more skill, with 1 being a “perfect” forecast.

The ranked probability skill score (RPSS) is used to evaluate the multinomial logit forecast performance (Wilks, 1995) for multiple categories (below-average, average, and above-average precipitation terciles):

$$RPSS = 1 - \frac{RPS_{Forecast}}{RPS_{Climatology}} \quad (Equation\ 9)$$

and

$$RPS = \sum_{i=1}^{N} \left[\sum_{j=1}^{k} \left(p_{i,j} - o_{i,j} \right)^2 \right] \quad (Equation\ 10)$$

where for a given year, i, p=($p_{i,1}$, $p_{i,2}$, ...$p_{i,k}$) and k is the number of categories (=3 in our case); RPS is calculated for the forecast using the multinomial logit, and RPS is calculated for climatology using the climatological probabilities (=.33).

For both the BSS and RPSS, the data are standardized and evaluated using leave-one-out cross-validation; where in this case one year is left out of the total of 26 years that are available. For example, if 1993 is being predicted, only 1994-2018 are used in the prediction, and so on. We point out that cross-validated scores are more representative of actual model performance since they are predicting blindly, like a real forecast would.

Results

Precipitation Relationship with Weather Types (June – October)

For each of the regions for the June through October (JJASO) prediction period, we examine the linear relationship between historical precipitation from PRISM and the WT frequencies derived from the historical reanalysis. As expected, Figure 2 shows that there is a negative correlation between precipitation and the number of dry WT days (sumDry), and a positive relationship with the number of monsoon WT days (sumMonsoon). For AZ-West, the magnitude of the relationship with sumDry and sumMonsoon is similar (-0.47 and 0.45). For AZ-East, there is a strong relationship with the sum of the number of normal WT days (sumNormal) – a predictor that is not considered here; but even so, the sumDry has a higher absolute value (-0.67). For NM-North and NM-South, the magnitude of the relationship with sumMonsoon is the strongest, where correlation is 0.55 and 0.67, respectively.

Applying the WTs to a historical reanalysis gives a sense of the upper limit of predictability, i.e., since it is based on the historical observations. But in this study, we are interested in predictability based on existing forecasts (not historical observations), so we examine the predictability based on the ECMWF hindcasts for different lead times. Figure 3a shows the correlation between historical precipitation and the hindcasted number of monsoon days for different leads. Leads are referred to by their month number (i.e., 4 refers to an April issued forecast, etc.). We note that the predictand period decreases as lead months get closer to the prediction period, i.e., leads 4, 5, and 6 predict 5 months (June-October; JJAOS); lead 7 predicts 4 months (July-October; JAOS); and lead 8 predicts 3 months (August-October; ASO). The correlations with the number of monsoon days are positive, though the magnitude is variable, as shown in Figure 3a. The highest correlation seen is in AZ-East, which has a correlation of r=0.7 for the lead month 6 prediction of JJASO, whereas the highest correlation for both NM regions is 0.4, but is seen consistently for NM-South across lead months 6 and later and in NM-North for lead month 7. Overall, NM-North tends to have lower correlations than the other regions, and NM does not have any skill in for lead month 4 (April) in the North or South. Some of the possible reasons for this are noted in the Discussion and Conclusions. Figure 3b shows the correlation between historical precipitation and the number of dry WT days from the ECMWF hindcasts. The correlations with the number of dry WT days are negative (Figure 3b), with lower magnitudes than the correlation of the number of monsoon WT days (Figure 3a).

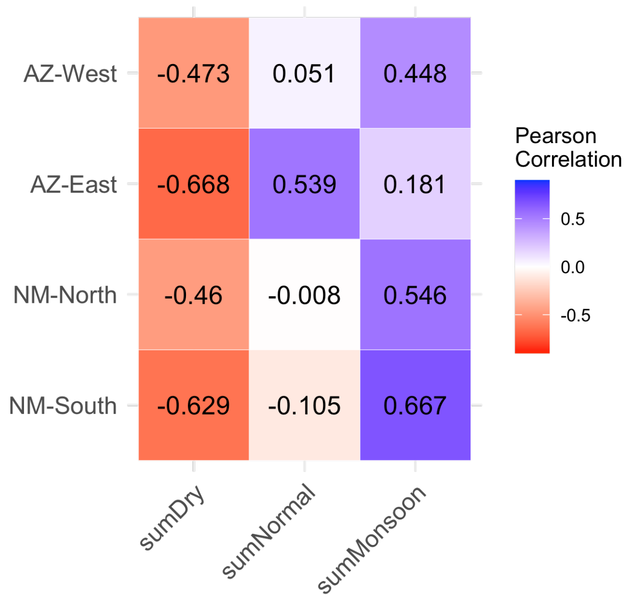

Figure 2. Pearson's correlation between prediction period of June-October (JJASO) average historical precipitation from PRISM and the sum of weather types (WTs) from the historical reanalysis.

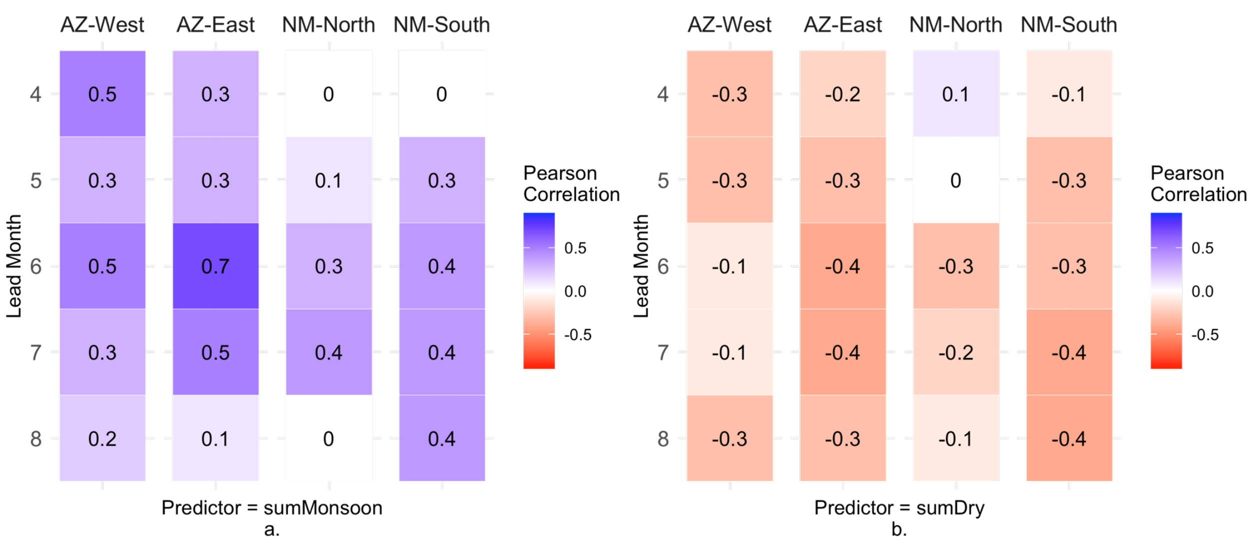

Figure 3. Pearson's correlation between seasonal average historical precipitation and the sum of the number of a) monsoon WT days (sumMonsoon) and b) dry WTs (sumDry) from the ECMWF hindcasts by lead time. Leads 4, 5, and 6 predict June-October (JJAOS); lead 7 predicts July-October (JAOS); lead 8 predicts August-October (ASO).

Prediction Skill for Generalized Linear Modeling (June – October)

In this section, we use the GLM framework to translate the WT information to precipitation characteristics (e.g., above- and below-median, and the terciles), for both the historical reanalysis and the hindcasts. We examine the skill scores (BSS and RPSS) for the JJASO prediction period (Figure 4) for several lead times and predictor combinations, i.e., the predictor can be the sum of the number of dry or monsoon WT days, and it includes both cross-validated and not cross-validated scores. We point out that cross-validated scores are more representative of actual model performance since they are predicting blindly, like a real forecast would. Pooling of results from several combinations of lead times allowed us to see general patterns for this prediction period.

AZ-East and NM-South show the expected pattern that as lead time decreases, the skill scores increase; this relationship is less clear for AZ-West and NM-North. The figure also shows the skill scores based on the historical reanalysis, which represent the upper limit of predictability, or of a "perfect forecast". However, since forecasts are never perfect, the skill scores from the historical reanalysis tend to be higher than the skill scores using the ECMWF hindcasts at the given leads. Only positive skill scores indicate that the skill is better than climatology. For this prediction period, we see that for lead months 4 and 5, all the GLMs from NM-North and NM-South are below zero. Both AZ-East and AZ-West have GLMs that are positive for all leads, as indicated by parts of the boxplot being above the zero line. NM-North and NM-South start showing positive skill for the GLMs for lead month 6.

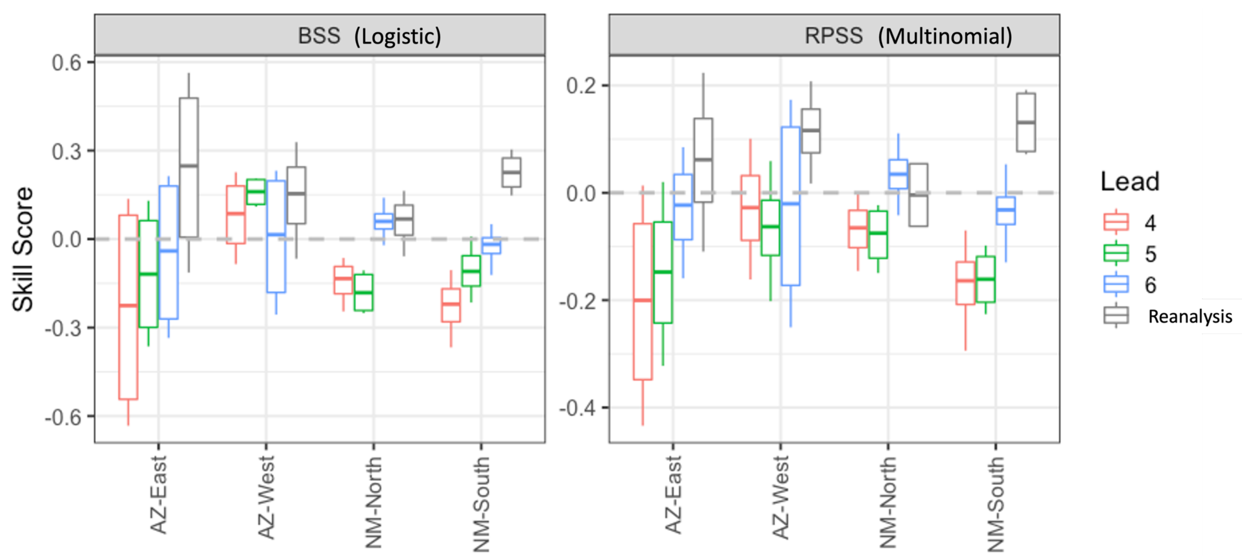

Figure 4. For the June-October prediction season, Brier Skill Score (BSS; left) for the logistic regression and Rank Probability Skill Score (RPSS, right) for the multinomial regression using ECMWF hindcasts for leads 4, 5, and 6, as well as the historical reanalysis. Each boxplot contains 4 skill scores (sumDry, sumDry/cross-validated, sumMonsoon, sumMonsoon/cross-validated). Skill scores greater than zero have skill over climatology.

Figure 5 demonstrates the BSS for the New Mexico regions, where results are broken out by predictor and cross-validation method. For New Mexico for JJASO, as expected, Figure 5 shows that the cross-validated BSS is always lower than when it is not cross-validated; this is because cross-validation is a blind forecast, more like it would be in a real operational setting. Also, the skill generally improves with lead time, and is most skillful for the reanalysis. Results are similar for the Arizona regions (figures not shown). Interestingly, for NM-North sumMonsoon and sumDry tend to be similar in terms of their skill. For NM-South, both predictors are similar in terms of the reanalysis, but sumMonsoon is a better predictor for lead months 4 and 5, and sumDry is better for lead month 6.

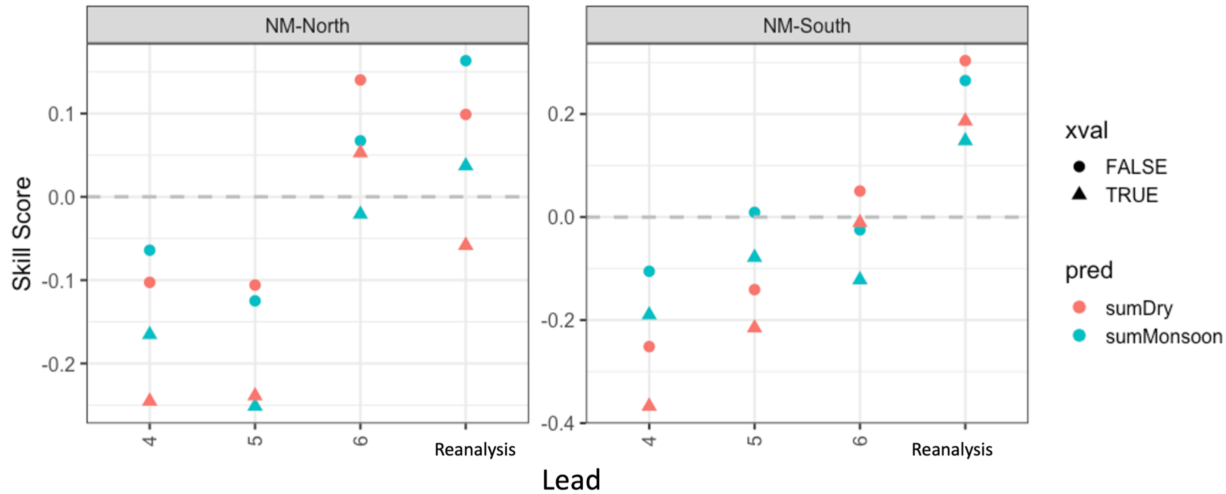

Figure 5. For the June-October prediction season, Brier Skill Score (BSS) for logistic regression using ECMWF weather type (WT) hindcasts for leads 4, 5, and 6, as well as the WTs from the historical reanalysis. Colors indicate if the predictor is the number of dry days (sumDry) or monsoon days (sumMonsoon), and the shape indicates if it is cross-validated (xval).

Skillful GLMs for all Prediction Periods

In the above section, we looked at skill diagnostics for the JJASO prediction period. However, in an experimental workflow, we are interested in all model combinations that are skillful for any of the 14 prediction periods (monthly or multi-monthly). As such, next we subset all GLMs that are skillful under cross-validation, that are standardized, and allow either sumMonsoon or sumDry as the univariate predictor. This is summarized below:

- BSS or RPSS must be > 0 when using the cross-validated Reanalysis
- BSS or RPSS must be > 0 when using the cross-validated ECMWF hindcasts
- Data is standardized (in a cross-validated manner)
- Can use either sum of monsoon or dry days as univariate predictor

Using the above criteria, Figure 6 pools the cross-validated ECMWF skill scores (RPSS and BSS) for all the models that were found to be skillful for each region. The median skill score for Arizona-West and Arizona-East is 0.11 (n=23 models for each region), whereas NM-North has a median = 0.063 (n=16 models) and NM-South has a median = 0.052 (n=37 models). In general, the medians for Arizona are higher than for New Mexico.

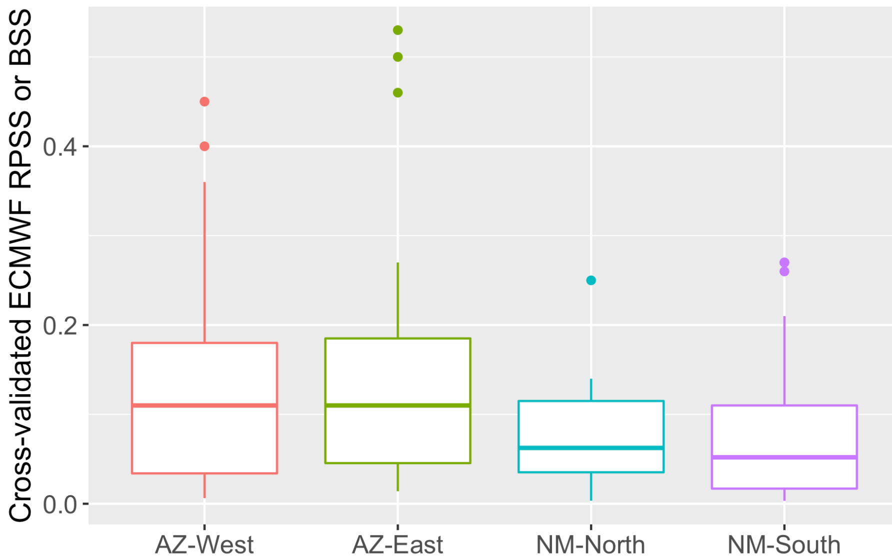

Figure 6. Cross-validated skill scores for each region; BSS = Brier Skill Score; RPSS=Rank Probability Skill Score.

Skillful Models for NM-South: Figure 6 pools all the skillful models for all the regions, but it is also illustrative to examine all the skillful models for a single region. Here, we use the example of NM-South, both for the logistic regression (Table 1) and for the multinomial proportional-odds regression (Table 2).

Of the 37 skillful models for NM-South, 19 of the models were from the logistic regression. Table 1 shows the logistic models by lead time, prediction period, and predictor. The table also includes the standardization values for the predictors from Equation 6 (*avg(x) and sd(x)*), as well as the intercept and slope terms from Equation 2; these are derived from data for the full period (i.e., they are not cross-validated, since in the cross-validated mode the values change with every value dropped). The table also includes the climatological median (Q50) for the season being predicted.

The remaining 18 skillful models for NM-South came from the multinomial proportional-odds regression. The results are shown in Table 2 for each lead, prediction period, and predictor. The table also includes the standardization values for the predictors from Equation 6, as well as the intercept and slope terms from Equation 3 and 4; again, these are derived from data for the full period. The table also includes the climatological terciles (Q33 and Q66) for the season being predicted. These tables are queried in the Google Colab Notebook developed as part of this study, which is described next.

Online Google Colab Experimental Forecast Notebook

To facilitate an experimental real-time forecast for water managers, the workflow and results from this study have been used to develop an online Google Colab Notebook. The Notebook is developed in Python and also ingests R code. The Notebook provides instructions for downloading real-time forecasts from ECMWF. Once the ECMWF forecasts are downloaded, the Notebook can be run by a user for the operational workflow described. The user first selects a region; here, we will continue with the example of NM-South. Next, the Notebook assigns WTs for each day of the ECMWF forecast ensemble; Figure 7 plots the ensemble mean WT frequency for June through

October of 2020. Then, the ensemble average WT predictor is used in the skillful statistical model(s) for that lead time (e.g., Tables 1 and 2). As mentioned, Tables 1-2 include the predictor averages and standard deviations needed to standardize in Equation 6, as well as the intercept and slope terms needed in Equations 2, 3, and 4. We note that the intercept and slope coefficients come from the fitting of all the available reanalysis data (1993-2018), and is not cross-validated (this is because there are different coefficients for every cross-validated fit, and we are now using all the available data for a future forecast, rather in an evaluative hindcast mode). For lead 6, there are 7 skillful binomial models for NM-South (Table 1). After downloading the lead 6 (June) ECMWF forecast from 2020, we run the Notebook, resulting in the output shown in Table 3.

Table 1. Logistic (binomial) regression models for NM-South that were skillful in terms of the cross-validated ECMWF (Ecmwf_xval) and reanalysis (Rean_xval); BSS = Brier skill score; M = sumMonsoon, D = sumDry.

Lead Month	Predicted Season	BSS (Ecmwf_xval)	BSS (Reanalysis)	BSS (Rean_xval)	x	avg(x)	sd(x)	intercept1	slope	Q50 (mm/Day)
4	SO	0.048	0.38	0.27	M	10	1.8	0.097	2.0	1.3
5	SO	0.27	0.38	0.27	M	9.3	1.3	0.097	2.0	1.3
5	SO	0.26	0.44	0.33	D	28	2.4	-0.013	-2.5	1.3
5	AS	0.13	0.16	0.020	D	6.8	1.2	-0.044	-1.0	1.7
5	ASO	0.12	0.54	0.46	D	29	2.4	0.079	-3.3	1.6
5	ASO	0.11	0.35	0.24	M	29	2.7	-0.11	1.8	1.6
5	Sep	0.0096	0.46	0.36	M	8.7	1.2	0.28	2.7	1.5
6	SO	0.21	0.38	0.27	M	9.6	1.4	0.097	2.0	1.3
6	SO	0.11	0.44	0.33	D	28	2.9	-0.013	-2.5	1.3
6	Oct	0.073	0.18	0.061	M	0.61	0.39	0.16	1.4	0.8
6	Jun	0.035	0.45	0.33	D	13	4.0	-0.19	-3.2	1.0
6	Oct	0.015	0.48	0.38	D	22	2.1	0.066	-2.4	0.8
6	ASO	0.011	0.35	0.24	M	30	2.6	-0.11	1.8	1.6
6	JJA	0.0063	0.17	0.046	M	41	3.9	0.0077	1.0	1.6
7	SO	0.11	0.38	0.27	M	8.1	1.7	0.097	2.0	1.3
7	Jul	0.034	0.19	0.057	D	2.7	2.5	-0.11	-1.1	1.9
7	JASO	0.0067	0.22	0.10	M	40	6.3	-0.055	1.3	1.6
8	Aug	0.27	0.37	0.26	M	19	3.6	0.0064	1.9	1.7
8	SO	0.13	0.44	0.33	D	27	3.7	-0.013	-2.5	1.3

Table 2. Multinomial (propodds) models for NM-South that were skillful in terms of the cross-validated ECMWF and Reanalysis; RPSS = Rank Probability Skill Score; M = sumMonsoon, D = sumDry.

Lead Month	Predicted Season	RPSS (Ecmwf_xval)	RPSS (Reanalysis)	RPSS (Rean_xval)	x	avg(x)	sd(x)	intercept1	intercept2	slope	Q33 (mm/Day)	Q66 (mm/Day)
5	Sep	0.081	0.30	0.19	M	8.7	1.2	1.2	-0.97	2.1	1.1	2.0
5	ASO	0.073	0.30	0.19	D	29	2.4	1.1	-1.0	-2.2	1.2	1.7
5	AS	0.048	0.12	0.010	D	6.8	1.2	0.69	-0.86	-1.1	1.4	2.0
5	SO	0.029	0.24	0.13	D	28	2.4	0.96	-0.98	-1.8	1.1	1.6
5	SO	0.029	0.15	0.031	M	9.3	1.3	0.79	-0.86	1.2	1.1	1.6
5	Sep	0.025	0.13	0.011	D	6.1	1.1	0.76	-0.81	-1.1	1.1	2.0
6	Jun	0.16	0.33	0.23	D	13	4.0	0.95	-1.2	-2.4	0.8	1.2
6	Jun	0.081	0.13	0.0025	M	3.4	1.5	0.87	-0.72	1.1	0.8	1.2
6	SO	0.060	0.24	0.13	D	28	2.9	0.96	-0.98	-1.8	1.1	1.6
6	SO	0.016	0.15	0.031	M	10	1.4	0.79	-0.86	1.2	1.1	1.6
6	JJA	0.0033	0.18	0.071	M	41	3.9	0.94	-0.76	1.3	1.4	1.8
7	Jul	0.076	0.34	0.23	M	15	4.1	1.1	-1.3	2.4	1.4	2.0
7	Jul	0.075	0.19	0.076	D	2.7	2.5	0.70	-1.1	-1.6	1.4	2.0
7	JA	0.0086	0.15	0.041	M	32	5.4	0.83	-0.81	1.2	1.7	2.1
8	Aug	0.052	0.25	0.15	M	19	3.6	0.95	-0.87	1.6	1.4	2.0
8	Aug	0.024	0.13	0.012	D	0.57	0.54	0.70	-0.86	-1.2	1.4	2.0
8	AS	0.017	0.34	0.23	M	29	4.7	1.0	-1.4	2.4	1.4	2.0
8	ASO	0.016	0.22	0.10	M	30	4.8	0.82	-1.1	1.7	1.2	1.7

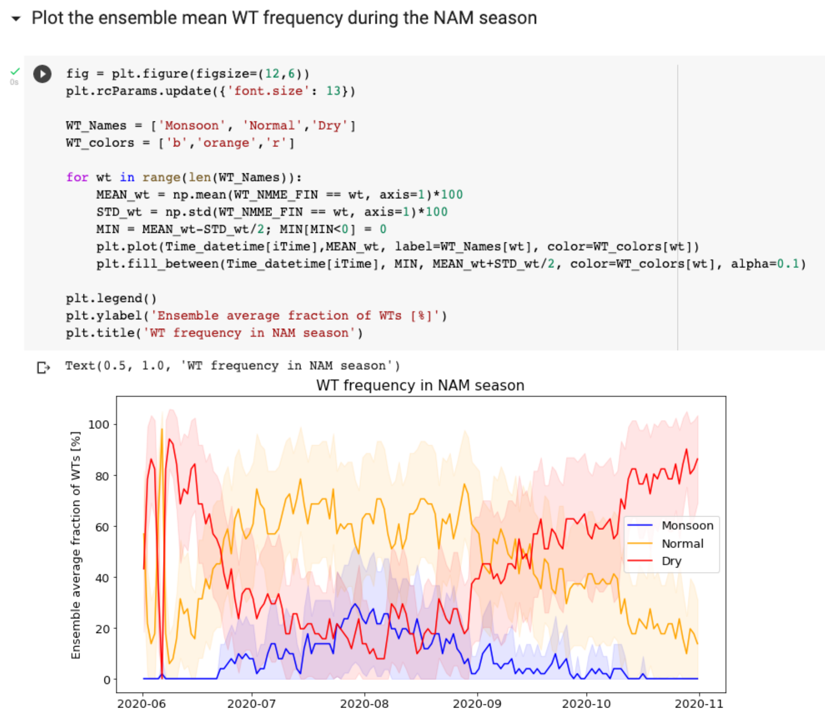

Figure 7. The Google Colab Notebook plots the ensemble mean WT frequency for June through October of 2020.

Table 3. Transposed Colab Notebook output showing logistic (binomial) regression models for NM-South for lead 6 of 2020. The forecasted probability of being greater than the median (Q50) is quantified by P_greaterQ50 (in bold). BSS = Brier skill score; M=sumMonsoon; D=sumDry.

	Skillful Model						
	1	2	3	4	5	6	7
Predicted season	SO	SO	Oct	Jun	Oct	ASO	JJA
Lead Month	6	6	6	6	6	6	6
BSS (Ecmwf_xval)	0.21	0.11	0.07	0.03	0.02	0.01	0.01
BSS (Reanalysis)	0.38	0.44	0.18	0.45	0.48	0.35	0.17
BSS (Rean_xval)	0.27	0.33	0.061	0.33	0.38	0.24	0.046
x	M	D	M	D	D	M	M
xnormMean	9.6	27.8	0.61	13.1	22	29.9	40.6
xnormSD	1.4	2.9	0.39	4.0	2.1	2.6	3.9
intercept	0.10	-0.013	0.16	-0.19	0.07	-0.11	0.0
slope	2.0	-2.5	1.4	-3.2	-2.4	1.8	1.0
Q50_mmDay	1.3	1.3	0.80	1.0	0.80	1.6	1.6
P_greaterQ50	**0%**	**0%**	**13%**	**2.7%**	**0%**	**0%**	**0%**
P_PRISM_mmDay	0.467	0.467	0.348	0.573	0.348	0.558	0.952
Correct?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

From Table 3, starting with Skillful Model 1 (the first column): the forecast for the prediction periods of September-October (SO) is based on the sumMonsoon WT predictor (summed for September and October); there are three BSS values reported: first, to understand how this logistic regression performed using the ECMWF hindcasts, we see the cross-validated BSS of 0.21. To put this into context, the BSS for the historical reanalysis is also output, which was quite high at 0.38, including in the cross-validated mode, where BSS = 0.27. The result shows a ~0% chance of being above the median precipitation for SO, which is 1.3 mm/day. Because this forecast has already happened, it can be checked using the observed precipitation that fell during the prediction period. For SO, the average precipitation from PRISM was 0.467 mm/day, which is NOT above the median (=1.3 mm/day), and the forecast was correct. We can see that all the ECMWF forecasted probabilities were low, and they all correctly verified.

Discussion and Conclusions

In this paper, we have demonstrated a technique that can provide skillful probabilistic forecasts of precipitation characteristics associated with the NAM in the Southwestern US. This technique was applied to two sub-basins of the Lower Colorado River in Arizona and two sub-basins of the Rio Grande in New Mexico. To facilitate the experimental execution of the workflow, we developed a Google Colab Notebook that allows a user to download a real-time ECMWF forecast, assign WTs, and run them through the predictive models that were found to be skillful for the retrospective ECMWF forecasts (i.e., hindcasts). This paper demonstrates results from the Colab Notebook on the NM-South region.

The technique in this study uses weather typing based on the ECMWF forecasts of synoptic-scale moisture advection, as represented by the water vapor mixing ratio at 850 hPa (Q850), in combination with a statistical technique called Generalized Linear Modeling (GLM). The skill was demonstrated using ECMWF hindcasts, with BSS and RPSS used as skill criteria. For the June through October prediction period, Arizona showed higher correlations, and had more predictability, including as early as April (i.e., lead 4). Further, pooling the results for all skillful models, the median skill scores for Arizona were higher than for New Mexico. We note that some of the reasons for the skillful correlations with the hindcasts are investigated in Prein et al. (2022); in short, they find that the ECMWF hindcasts faithfully represent key synoptic features of monsoon rainfall, including the ocean teleconnections. AZ precipitation has a strong ocean teleconnection, whereas NM's precipitation is more complicated, since its monsoon can come from two sources (Gulf of California or Gulf of Mexico), resulting in less predictability. Both regions in New Mexico were found to have skillful GLMs, with the skill varying depending on the region, lead, predictor, and prediction period. Overall, for NM-South, there were 37 skillful models, and for NM-North there were 16 skillful models. There were 23 skillful models for each region in Arizona. We also investigated using the number of normal WT days as a predictor. However, although this resulted in a few more skillful models, this in essence lumps the number of monsoon and number of dry days together, which is difficult to interpret.

A key aspect of this study was the close collaboration between NCAR scientists and Reclamation water management staff, who have been testing the forecast techniques to support current-year water operations. Continuing collaboration could be used to refine these forecasts. Our skill criteria was for the BSS or RPSS to be greater than zero for the cross-validated models, and this low threshold for skill needs to be considered when these forecasts are evaluated or used. In future studies, this skill criteria could be refined (e.g., made more conservative) by having a higher threshold for skill (e.g., >0.05).

The functional relationships between the precipitation and WTs are developed based on historical reanalysis, but we point out that climate is not stationary (Milly et al. 2001). As such, the GLMs should be regularly updated and re-evaluated as new data come available. Further, these models could be compared with other statistical approaches that are nonlinear, such as machine learning (random forests, neural networks etc).

Finally, the output from this work could be used to further inform water management. One successful approach has been to apply the probabilistic climate forecasts to streamflow trace weighting schemes (e.g., Werner et al. 2004, Baker et al. 2021, Towler et al. 2022). The importance of evaluating improvements in streamflow forecasts in terms of decision-relevant terms (e.g., operational reservoir projections) has also been underscored in recent research, (Towler et al. 2022; Woodson et al. 2021), and could be explored as an enhancement of these results.

References

Baker, S.A., Rajagopalan, B., and Wood, A.W. 2021. "Enhancing ensemble seasonal streamflow forecasts in the Upper Colorado River Basin using multi-model climate forecasts," JAWRA, 57 (6): 906–922, https://doi.org/10.1111/1752-1688.12960

Castro, C.L., Chang, H.-I., Dominguez, F., Carrillo, C., Schemm, J.-K., and Henry Juang, H.-M. 2012. "Can a regional climate model improve the ability to forecast the North American monsoon?" *J Climate*, 25 (23): 8212–8237.

Ciancarelli, B., Castro, C.L., Woodhouse, C., Dominguez, F., Chang, H.-I., Carrillo, C., and Griffin, D. 2014. "Dominant patterns of US warm season precipitation variability in a fine resolution observational record, with focus on the southwest," Int J Climatol, 34 (3):687.

Daly, C., Taylor, G., and Gibson, W. 1997. "The PRISM approach to mapping precipitation and temperature," Proc. 10th AMS Conf. on Applied Climatology, pp. 20–23.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., and others. 2011. "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system," *QJR Meteorol Soc*, 137 (656): 553–597.

Hartmann, H.C., Bales, R., Sorooshian, S. 1999. Weather, Climate, and Hydrologic Forecasting for the Southwest U.S. CLIMAS Report Series, CL2-99, University of Arizona, available at: https://climas.arizona.edu/sites/default/files/pdfcl2-99.pdf (Accessed Feb 17, 2022).

Helsel, D.R., and Hirsch, R.M. 1995. Studies in Environmental Science, in Statistical Methods in Water Resources, vol. 49, 529 pp., Elsevier, Amsterdam.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Paolino, D.A., and others. 2014. "The North American multimodel ensemble: phase-1 seasonal to-interannual prediction; phase-2 toward developing intraseasonal prediction," *BAMS*, 95 (4): 585–601.

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., and others. 2010. "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," Rev Geophys, 48(3): 1–34.

McCullagh, P., and Nelder, J.A. 1989. Generalized Linear Models, Chapman and Hall, London.

McNulty, K. 2022. Proportional Odds Logistic Regression for Ordered Category Outcomes, Chapter 7 in Handbook of Regression Modeling in People Analytics: With Examples in R and Python, Chapman and Hall/CRC, available online at: https://peopleanalytics-regression-book.org/ord-reg.html (Accessed Feb 17, 2022).

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., and others. 2008. "Stationarity Is Dead: Whither Water Management?" *Science*, 319(5863): 573-574.

Prein, A.F., Holland, G.J., Rasmussen, R.M., Clark, M.P., and Tye, M.R. 2016. "Running dry: The US Southwest's drift into a drier climate state," Geophys Res Lett, 43 (3): 1272–1279.

Prein, A.F., and Mearns, L.O. 2021. "US extreme precipitation weather types increased in frequency during the 20th century," *J Geophys Res (Atmos)*, 126 (7): e2020JD034287.
Prein, A.F., Towler, E., Ge, M., Llewellyn, D., Baker, S., Tighi, S., Barrett, L. 2022. "Sub-Seasonal Predictability of North American Monsoon Precipitation," Geophys Res Lett, 49(9), https://doi.org/10.1029/2021GL095602.
Towler, E., Llewellyn, D., Barrett, L., and Young, R. 2019. "Extremes of opportunity? A generalized approach to identify intersections between changing hydrology and water management," Proc. of SEDHYD, Reno, NV, https://www.sedhyd.org/2019/openconf/modules/request.php?module=oc_program&action=view.php&id=160&file=1/160.pdf
Towler, E., Llewellyn, D., Prein, A., and Gilleland, E. 2020. "Extreme-value analysis for the characterization of extremes in water resources: A generalized workflow and case study on New Mexico monsoon precipitation," Weather Clim Extrem, doi: 10.1016/j.wace.2020.100260.
Towler, E., Woodson, D., Baker, S., Ge, M., Prairie, J., Rajagopalan, B., Shanahan, S., Smith, R. 2022. "Incorporating mid-term temperature predictions into streamflow forecasts and operational reservoir projections in the Colorado River Basin," *J Water Resour Plan Manag (ASCE)*, doi: 10.1061/(ASCE)WR.1943-5452.0001534.
Werner, K., Brandon, D., Clark, M., Gangopadhyay, S. 2004. "Climate index weighting schemes for NWS ESP-based seasonal volume forecasts," *J Hydrometeorol* 5(6): 1076-.
Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods, Tech. report, Intergovt. Panel on Climate Change, Geneva, Switzerland.
Wilks, D.S. 1995. Statistical methods in the atmospheric sciences. Elsevier, New York.
Woodson, D., Rajagopalan, B., Baker, S., Smith, R., Prairie, J., Towler, E., Ge, M., Zagona, E. 2021. "Stochastic decadal projections of Colorado River streamflow and reservoir pool elevations conditioned on temperature projections," *Water Resour Res*, https://doi.org/10.1029/2021WR030936.
Yee, T. and Moler, C. 2022. VGAM: Vector Generalized Linear and Additive Models, available online at https://CRAN.R-project.org/package=VGAM (Accessed February 17, 2022).