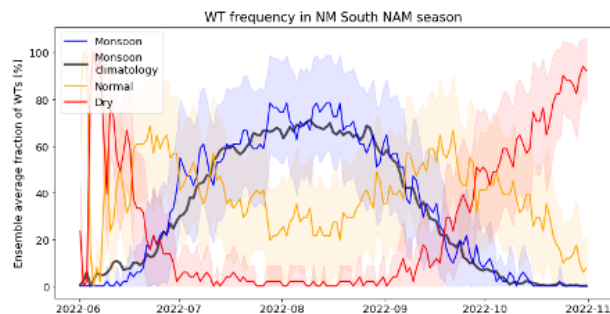
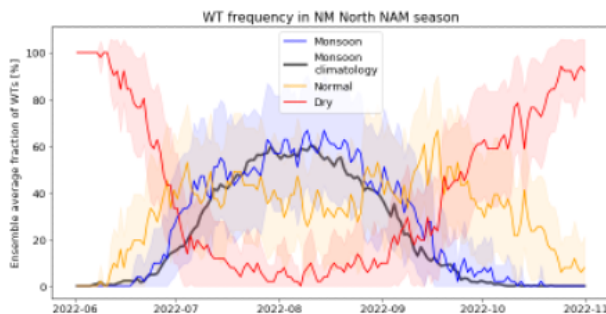




— BUREAU OF —
RECLAMATION

An Experimental Monsoon Forecast for Water Management

**Research and Development Office
Science and Technology Program
(Final Report) ST-2020-20032-01**



Mission Statements

The U.S. Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; honors its trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated Island Communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Disclaimer:

Information in this report may not be used for advertising or promotional purposes.

The enclosed data and findings should not be construed as an endorsement of any product or firm by the Bureau of Reclamation (Reclamation), U.S. Department of the Interior, or the Federal Government. The products evaluated in this report were evaluated in environmental conditions and for purposes specific to Reclamation's mission. Reclamation gives no warranties or guarantees, expressed or implied, for the products evaluated in this report, including merchantability or fitness for a particular purpose.

REPORT DOCUMENTATION PAGE		<i>Form Approved</i> <i>OMB No. 0704-0188</i>
T1. REPORT DATE: September 2023	T2. REPORT TYPE: Research	T3. DATES COVERED
T4. TITLE AND SUBTITLE An Experimental Monsoon Forecast for Water Management		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER 1541 (S&T)
6. AUTHOR(S) Dagmar Llewellyn, Bureau of Reclamation, Albuquerque Area Office; 505-462-3594, dllewellyn@usbr.gov Erin Towler, National Center for Atmospheric Research, Boulder, CO, 303-497-2724, towler@ucar.edu Andreas Prein, National Center for Atmospheric Research, Boulder, CO, 303-497-8200, prein@ucar.edu PARTICIPANT(S) Lucas Barrett, Bureau of Reclamation, Albuquerque Area Office Sarah Baker, Bureau of Reclamation, Boulder, CO Shana Tighi, Bureau of Reclamation, Boulder City, NV Taylor Adams, Hydros Consulting Inc. Boulder CO Nick Mander, Hydros Consulting, Inc. Boulder, CO		5d. PROJECT NUMBER ST-2020-20032-01
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER ALB-402
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Albuquerque Area Office 555 Broadway Blvd. NE, Suite 100, Albuquerque, NM 87102 Mesoscale and Microscale Meteorology Laboratory National Center for Atmospheric Research P.O. Box 3000, Boulder CO 80307		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007		10. SPONSOR/MONITOR'S ACRONYM(S) R&D: Research and Development Office BOR/USBR: Bureau of Reclamation DOI: Department of the Interior
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) ST-2020-20032-01
12. DISTRIBUTION / AVAILABILITY STATEMENT Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT (Maximum 200 words) This research has two main components: first, we characterize the predictability of monsoon season precipitation by evaluating the skill of monsoon weather types in forecast ensemble products. Second, we investigate the potential usability of these forecast ensemble products by identifying monsoon forecasts the provide opportunity to improve water management and transferring them to a web-based experimental real-time forecast platform. Leveraging prior research and partnerships, the monsoon forecast is applied to regions in New Mexico and Arizona, but the process of defining weather types and developing an experimental forecast system can be extended to other Reclamation regions.		

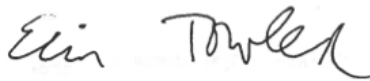
15. SUBJECT TERMS Precipitation, Monsoon, Forecasts, Predictability, Weather Types, New Mexico, Arizona			
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT U
18. NUMBER OF PAGES 65			19a. NAME OF RESPONSIBLE PERSON Dagmar Llewellyn
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	19b. TELEPHONE NUMBER 505-462-3594

S Standard Form 298 (Rev. 8/98)
P Prescribed by ANSI Std. Z39-18

BUREAU OF RECLAMATION
Research and Development Office
Science and Technology Program

(Final Report) ST-2020-20032-01

**An Experimental Monsoon Forecast for Water
Management**



Prepared by: Erin Towler
Project Scientist, NCAR



Checked by: Dagmar Llewellyn
Hydrologist, Albuquerque Area Office, UC Region



Checked by: Lucas Barrett
Hydrologist, Albuquerque Area Office, UC Region



Peer Review: James Done
Project Scientist, NCAR



Peer Review: Taylor Adams, Hydros Consulting

For Reclamation disseminated reports, a disclaimer is required for final reports and other research products, this language can be found in the peer review policy:

This document has been reviewed under the Research and Development Office Discretionary peer review process https://www.usbr.gov/research/peer_review.pdf consistent with Reclamation's Peer Review Policy CMP P14. It does not represent and should not be construed to represent Reclamation's determination, concurrence, or policy.

Cover Image: Experimental weather type (WT) frequency forecasts for New Mexico (NM) North (left) and NM South (right) for the monsoon season (June through October); the blue line is the forecast average of number of monsoon WT days, the black line is the climatological number of monsoon WT days, yellow is the forecast average of number of normal WT days, and red is the forecast average of number of dry WT days. The shaded areas show the forecast ensemble spread.

Executive Summary

Problem Statement

River Basins in New Mexico and Arizona, including the Lower Colorado River Basin, Rio Grande, and Pecos, are heavily impacted by monsoon season precipitation. However, monsoon season rain events are difficult to predict and there is a lack of robust forecasting products that provide actionable information in the time horizon needed. Reclamation would benefit from new tools to forecast monsoon season precipitation that can inform its water management.

Research Activities and Results

This project explores the predictability and useability of monsoon season precipitation forecasts for Reclamation. Although seasonal forecasts of monsoon precipitation for the US Southwest are not typically skillful, forecasts of recurring large-scale weather patterns, or “weather types” have shown promise. In this study, an experimental monsoon precipitation forecast is prototyped and tested using weather types developed for Arizona and New Mexico.

This research has two main components: first, we characterize the predictability of monsoon season precipitation by evaluating the skill of monsoon weather types in forecast ensemble products. Second, we investigate the potential usability of these forecast ensemble products by identifying monsoon forecasts that provide opportunity to improve water management and transferring them to a web-based experimental real-time forecast platform. Leveraging prior research and partnerships, the monsoon forecast is applied to regions in New Mexico and Arizona, but the process of defining weather types and developing an experimental forecast system can be extended to other Reclamation regions.

Future Plans

From these encouraging results, the next steps include work to further operationalize monsoon forecasts in New Mexico and Arizona. This includes additional testing of these tools, improving the usability of workflows, and refining NCAR forecast tools to improve predictability in the Rio Grande, for example by incorporating information from the Gulf of Mexico. Further, additional efforts will investigate how monsoon prediction skill translates to streamflow forecasts and associated operational conditions and reservoir elevations.

Contents

Executive Summary iii

 Problem Statement..... iii

 Research Activities and Results..... iii

 Future Plans iii

Main Report 1

 Problem Statement..... 1

 Research Activities and Results..... 1

 Summary Findings and Future Plans 2

References 3

 Appendix A – Component 1 Paper A-1

 Appendix B – Component 2 Paper B-1

 Appendix C – Component 3 Summary..... C-1

 Appendix D – Component 4 Summary..... D-1

 Appendix E – Research Products..... E-1

Main Report

Problem Statement

River Basins in New Mexico and Arizona, including the Lower Colorado River Basin, Rio Grande, and Pecos, are heavily impacted by monsoon season precipitation. However, monsoon season rain events are difficult to predict and there is a lack of robust forecasting products that provide actionable information in the time horizon needed. Reclamation would benefit from new tools to forecast monsoon season precipitation that can inform its water management.

Research Activities and Results

This research has two main components: first, we characterize the predictability of monsoon season precipitation by evaluating the skill of monsoon weather types in forecast ensemble products. Second, we investigate the potential usability of these forecast ensemble products by identifying monsoon forecasts that provide opportunity to improve water management and transferring them to a web-based experimental real-time forecast platform. Leveraging prior research and partnerships, the monsoon forecast is applied to regions in New Mexico and Arizona, but the process of defining weather types and developing an experimental forecast system can be extended to other Reclamation regions. The main approach and findings from each component are detailed below.

For the first component, we develop monsoon season weather types and evaluate the weather type skill in existing forecast products. The weather types are developed using a clustering method that includes water vapor mixing ratio at 850 hPa (Q850) as a predictor. Here we show that monsoon season (June–October) precipitation can be forecasted by the European Centre for Medium-Range Weather Forecast's model months ahead at catchment scales. This is possible by identifying the frequency of days with synoptic-scale moisture advection into the monsoon region, which greatly improves predictability over directly utilizing modeled precipitation. Other forecasting systems fail to provide useful guidance due to deficiencies in their data assimilation systems and biases in representing key synoptic features of the monsoon including its teleconnections. The outcomes of this component were published in *Geophysical Research Letters*. This publication is provided in Appendix A: “Sub-Seasonal Predictability of North American Monsoon Precipitation,” (Prein et al. 2022).

In the second component, we develop an experimental monsoon precipitation forecast using the weather types developed in the first component. 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. Forecasts were more skillful in Arizona than in New Mexico, likely because monsoonal moisture for Arizona comes mostly from the Pacific, while New Mexico gets monsoonal moisture from additional sources, for example, the Gulf of Mexico. 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. The outcomes of this component were published as part of the *Proceedings of the 2023 Federal Interagency Sedimentation and Hydrologic Modeling Conference (SEDHYD)* conference that took place in St. Louis, Missouri. This is provided in Appendix B: “Seasonal forecasting of monsoon precipitation characteristics using weather types and generalized linear modeling,” (Towler et al. 2023).

To complement the second component, we also provide a description of how the experimental monsoon forecasts were implemented in one of Reclamation’s operational models: the Upper Rio Grande Water Operations Model (URGWOM). Results showed that the experimental monsoon forecasts showed improvement in Elephant Butte deliveries in 9 out of 10 hindcast years. Results from this work are provided in Appendix C: “Implementing the Experimental Monsoon Forecasts in the Upper Rio Grande Water Operations Model Annual Operating Plan Runs.”

Finally, we include preliminary findings from applying machine learning techniques to predict monsoon season precipitation. This component corroborated findings from the first component, namely that water vapor mixing ratio at 850 mb (Q850) is the most important variable for monsoon prediction. While there is still more research to be done on this, preliminary results from Random Forest methods are encouraging and their potential for real-time forecast applications could be further explored. Key findings are provided in Appendix D: “Predicting the Summer Rains of the Southwestern United States Using Machine Learning.”

Appendix E to this report provides a complete list of the research products from this effort, including the papers, presentations, and outreach efforts.

Summary Findings and Future Plans

This research showed that monsoon precipitation can be skillfully predicted by existing seasonal forecast products, and that these forecasts are potentially useable for water management. A key result was a workflow that was transferred into an online Google Colab Notebook, which allowed operators to download real-time forecasts, assign the weather types, and to make probabilistic precipitation predictions in the 2022 monsoon season. These results and products are ripe for a new facilitated adoption phase, whereby these experimental forecasts and workflows can be further tested and refined for operational use. To this end, there are several recommendations for future work:

- Extend experimental forecast testing.
- Improve usability of the online user interface and workflows.
- Test an additional forecast product that could enhance forecast skill.
- Refine NCAR forecast tools to improve predictability in the Rio Grande.

These activities also lend themselves to collaboration with ongoing research being conducted in another Reclamation-funded project, “*Translating Monsoon Forecasts to Streamflow Predictability in the Lower Colorado River Basin*”, being conducted by Reclamation and NCAR colleagues.

Data and Code Availability

ERA-Interim data can be accessed from <https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>, PRISM precipitation data from <https://prism.oregonstate.edu/>, and NMME seasonal hindcasts can be downloaded from <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-multi-model-ensemble>. Finally, the ECMWF-IFS hindcasts and forecasts can be accessed from <https://climate.copernicus.eu/seasonal-forecasts>.

The code for the statistical analysis and visualization of data in this report is openly available through GitHub (https://github.com/AndreasPrein/NAM_S2S_predictability).

References

Prein AF, 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, Prein AF, Barrett L, (2023) Seasonal forecasting of monsoon precipitation characteristics using Weather Types and Generalized Linear Modeling. Proceedings of the Federal Interagency Sedimentation and Hydrologic Modeling Conference (SEDHYD), St. Louis, MO.

Appendix A – Component 1 Paper

Sub-Seasonal Predictability of North American Monsoon Precipitation

Andreas Prein¹, Erin Towler¹, Ming Ge¹, Dagmar Llewellyn², Sarah Baker³, Shana Tighi³, and Lucas Barrett²

¹ National Center for Atmospheric Research (NCAR), Boulder, CO

² Bureau of Reclamation, Albuquerque, NM

³ Bureau of Reclamation, Boulder, CO

⁴ Bureau of Reclamation, Boulder City, NV

Paper published in the Geophysical Research Letters 2022

<https://doi.org/10.1029/2021GL095602>

Abstract

North American Monsoon (NAM) rainfall is a vital water resource in the United States Southwest, providing 60–80% of the region's annual precipitation. However, NAM rainfall is highly variable and water managers lack skillful guidance on summer rainfall that could help inform their management decisions and operations. Here we show that NAM season (June–October) precipitation can be forecasted by the European Centre for Medium-Range Weather Forecast's model months ahead at catchment scales. This is possible by identifying the frequency of days with synoptic-scale moisture advection into the NAM region, which greatly improves predictability over directly utilizing modeled precipitation. Other forecasting systems fail to provide useful guidance due to deficiencies in their data assimilation systems and biases in representing key synoptic features of the NAM including its teleconnections.

Introduction

The U.S. Southwest is a global hotspot for water scarcity (Liu et al., 2017; Mekonnen & Hoekstra, 2016) and is currently battling one of its most severe droughts in decades. With diminishing water resources from snowmelt (Ikeda et al., 2021; Milly & Dunne, 2020; Mote et al., 2005; Rasmussen et al., 2011) and steadily increasing population and freshwater demand (MacDonald, 2010), U.S. Southwest water resource managers are interested in potential opportunities from the secondary source of rainfall from the North American Monsoon (NAM). The NAM contributes approximately 35%–45% to the annual precipitation in the desert Southwest (Higgins et al., 1999) and up to 60% in New Mexico. However, NAM rainfall is extremely variable on inter-seasonal (Carleton, 1986), inter-annual (Carleton et al., 1990; Higgins et al., 1998, 1999; Higgins & Shi, 2000), and decadal time-scales (Castro et al., 2001; Yu & Wallace, 2000).

Meteorologically, the NAM is related to the poleward propagation of the subtropical high, that is, the monsoon high, over North America, which starts in May and June in Mexico and is centered over New Mexico in July and August (Adams & Comrie, 1997; Higgins et al., 1998).

This shift results in anomalous flow patterns from the southeast over Arizona and New Mexico, transporting warm and moist air from the Gulf of Mexico and the Pacific Ocean onto the continent. This leads to convective precipitation and a distinct summer maximum in the precipitation annual cycle. The onset of the monsoon in Arizona and New Mexico is variable but typically happens in early July. The NAM season is characterized by intermittent monsoonal moisture surges and dry spells related to shifts in the monsoonal high-pressure system (Higgins et al., 2004; Jiang & Lau, 2008; Pascale & Bordoni, 2016; Schiffer & Nesbitt, 2012; Seastrand et al., 2015).

Predicting NAM precipitation on seasonal to sub-seasonal (S2S) time scales remains challenging. Historic observations show that NAM characteristics are modulated by Pacific sea surface temperature patterns (Castro et al., 2001). Observations also show that wet winters tend to be followed by dry monsoon seasons and vice versa in New Mexico and Arizona (Gutzler & Preston, 1997; Higgins et al., 1998). However, tree-ring reconstructions covering the last five centuries indicate that this link is weak, at best, and not stable over time (Griffin et al., 2013). Zhu et al. (2005) found no significant relationships between the antecedent soil moisture conditions during NAM onset and total precipitation. Additionally, S2S forecasting systems have no significant skill in predicting precipitation in the U.S. Southwest after two to 3 weeks of leadtime (Krishnamurti et al., 2002; Li & Robertson, 2015; Slater et al., 2019). The dominant processes that could contribute to S2S predictability of early season NAM precipitation are warm-season atmospheric teleconnection modes such as the West Pacific North America pattern or quasi-stationary Rossby wave trains (Castro et al., 2012; Ciancarelli et al., 2014). These teleconnections influence the positioning and seasonality of the monsoon high and thereby the amount of precipitation in the region (Carleton et al., 1990).

In many snowmelt-driven basins in the Western U.S., operational seasonal water supply forecasts decisions are mainly based upon snowpack and don't explicitly consider any information about monsoon precipitation due to its low predictability. Skillful forecasting products that intersect with decision points on seasonal and annual planning horizons are extremely valuable to water managers and are a research priority (Bureau of Reclamation, 2018, 2021; Raff et al., 2013).

Here we present a forecasting framework that leverages the ability of current seasonal forecasting systems in simulating large-scale circulation patterns that are associated with monsoonal rainfall rather than utilizing erroneously modeled rainfall directly (Crochemore et al., 2016). The goal of this study is to establish a set of weather types (WTs) that can be used in a seasonal forecasting framework to investigate the potential to predict NAM precipitation variability. Therefore, we use a novel data-driven weather typing algorithm based on clustering (Prein & Mearns, 2021) that uses objective criteria to identify archetypal WTs. Specifically, we derive WTs that are optimized for NAM precipitation in catchments in Arizona and New Mexico, which include the Lower Colorado River Basin and Rio Grande Basin, respectively. The algorithm tests many combinations of atmospheric predictor variables, establishing the optimal set of WTs from two skill scores based on the precipitation average and standard deviation of the derived WTs. Similar approaches have been shown to provide novel insights into the origin of

precipitation changes on climate time scales (Lehner et al., 2017; Prein et al., 2016). Further, we investigate the potential for skillful forecasts of NAM season rainfall by examining the ability of seasonal forecasts to predict the identified WTs, versus predicting precipitation directly.

Data and Methods

Season and Region of Interest

We investigate June to October conditions, which incorporates the core monsoon season from July to September and helps to assess the onset and decay of the monsoonal high. Our focus region includes eight catchments in Arizona and six in New Mexico (Figure 1a). Those catchments were selected due to their importance for water management in these states. We derive WTs for each of these catchments based on the skill scores discussed below.

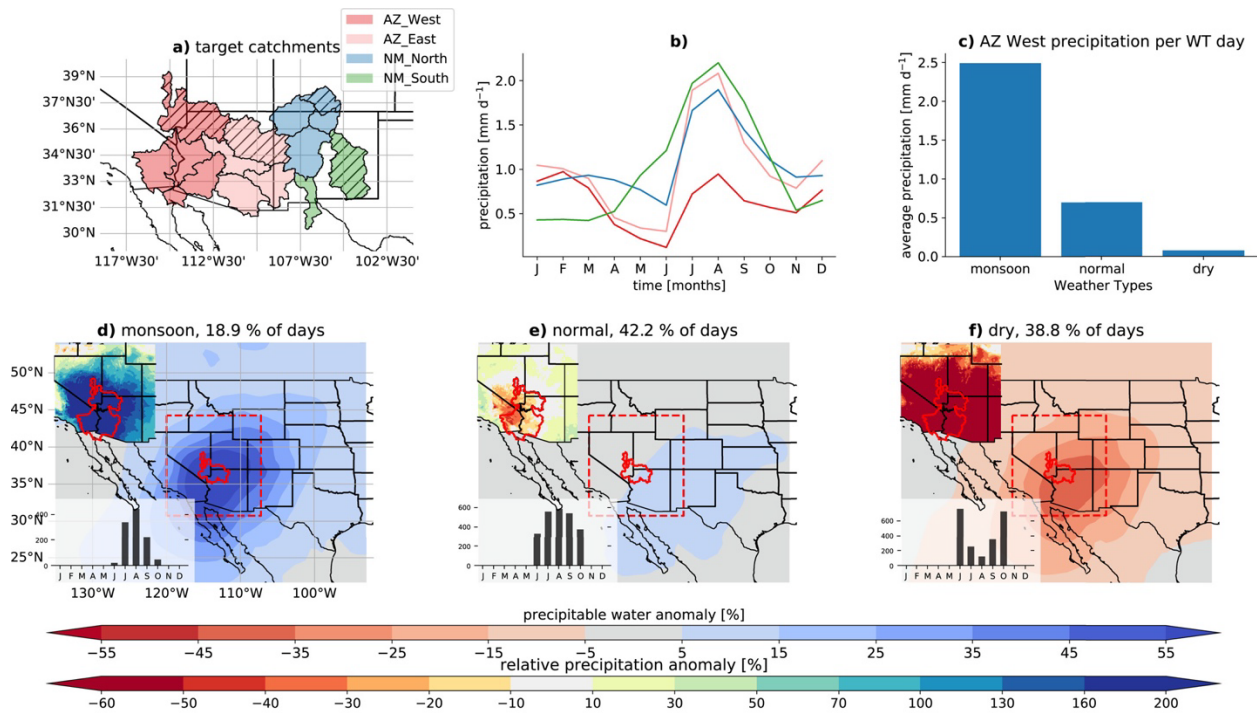


Figure 1. Three weather types—a wet (monsoon), normal, and dry weather type (WT)—characterize the major modes of synoptic-scale variability during the North American monsoon season. (a) Basin clusters that feature similar WTs include Arizona West (AZ West, dark red), Arizona East (AZ East; light red), New Mexico North (NM North, blue), and New Mexico South (NM South, green). Hatching shows the basin that was used to derive the WTs. (b) Monthly average precipitation in each region. (c) Average June to October precipitation during monsoon (d), normal (e), and dry (f) days in the AZ West region. Average WT precipitable water anomalies (colored contour) for each WT (d–f) in the AZ West region (red contour). The histogram in the lower left shows the monthly frequency of WT days. The inset in the top left shows the precipitation anomaly for each WT and the WT frequency is shown in the title of each panel.

Observations and Reanalysis

We use the Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al., 1997) gridded daily precipitation data for calculating precipitation statistics. PRISM provides data for the 1982–2018 period on a 4 km horizontal grid over the conterminous United States.

For the WT classification, we use daily average atmospheric variables from ECMWF's Interim Reanalysis within the period 1982–2018 (Dee et al., 2011). The selection of predictor variables is limited by variables that are commonly stored by seasonal forecasting centers; we include the following 12 variables in the analysis: sea level pressure, 850 hPa zonal, meridional, and total wind speed, 850 hPa and 500 hPa moisture flux, water vapor mixing ratio, and air temperature, 500 hPa geopotential height, and 200 hPa wind speed. The same variables are also important on weather forecast timescales, as they related to dynamic forcing and thermodynamic instability.

Weather Typing Methodology

The WT algorithm was modified from the one presented in (Prein & Mearns, 2021). In short, it is a data-driven clustering algorithm that tests atmospheric predictor combinations to find optimal WTs based on objective skill scores associated with NAM precipitation. The details of the WTing algorithm are outlined below.

All days between June and October within the period 1982–2018 are considered in the clustering. We consider all 12 above-mentioned variables and up to combinations of three variables (220 possible combinations) as predictors for catchment precipitation. First, we calculate daily anomalies from the mean climate state of each predictor on a grid cell basis and apply Gaussian spatial smoothing ($\sigma = 0.5$) to remove small-scale variability from the predictor variables. Next, we normalize the anomalies by their regional average mean and daily standard deviation. The normalized anomaly fields are then clustered using the results of a hierarchical (using Ward's linkage on a condensed distance matrix) cluster algorithm as initial seed for a k-means clustering (Romesburg, 2004). The combination of these two cluster algorithms showed high skill in a WT intercomparison study (Schiemann & Frei, 2010) and was successfully applied to capture precipitation characteristics over the US (Prein et al., 2016) and weather conditions in other regions (Comrie, 1996; Ekstrom et al., 2002).

For each catchment, we assess how these skill scores change if WTs from the remaining 13 basins are used as predictors of the selected catchment's NAM precipitation (Figure S1 in Supporting Information S1). In Arizona, using the WTs from the two northernmost catchments (Lower Colorado-Lake Mead subregion—HUC1501; Little Colorado subregion—HUC1502) results in higher skill in characterizing precipitation in southern (upwind during NAM conditions) catchments. Based on these results we cluster the catchments in Arizona in two regions—Arizona West (AZ West containing HUC1501, HUC1503, HUC1507, HUC1810) and Arizona East (AZ East containing HUC1502, HUC1504, HUC1506, HUC1508). Similarly, we cluster the catchments in New Mexico into Northern catchments (NM North containing HUC1301, HUC140801, HUC130201, HUC130202) and southern catchments (NM South containing HUC130301 and HUC1306). The differences between Arizona and New Mexico

WTs is not surprising since previous literature showed that spatial variability of monsoon precipitation differs east and west of the U.S. continental divide (Castro et al., 2012; Ciancarelli et al., 2014).

We derive clusters for each of the 14 target catchments (Figure 1a) and test the sensitivity of the cluster domain size by adding 2°, 5°, and 10° around a rectangle that includes the catchment of interest to test sensitivities to the clustering region (Beck et al., 2016).

We use two skill scores to assess the quality of derived WT dependent on the predictor variables and WT domain size. The first is the absolute average of WT centroid (cluster mean state) precipitation anomalies (PRanom), which should be maximized—the precipitation in the derived WT should be as different as possible from the climatological precipitation. The second skill score is the ratio of the intra- versus inter-cluster standard deviation of precipitation (IvI), which should also be maximized. Although these two scores are correlated (Figure S2 in Supporting Information S1) combining them helps to improve the robustness of the derived WT. This score favors WT that have days with similar precipitation within WT and whose precipitation is different between WT. The combination of these two scores has been successfully used in previous WTing applications (Prein et al., 2016).

Based on these skill scores we select the top 10% of the tested predictor combinations and count how often each variable is included in a well-performing setting (Figure S3 in Supporting Information S1). Water vapor mixing ratio at 850 hPa (Q850) is by far the most frequently used predictor in well-performing settings in all catchments and results in close to optimal performance as a single predictor. Performance differences between different WT domain sizes are small but adding 5° around each catchment performed best in Arizona catchments and adding 2° was best in New Mexico. WT are derived for each basin based on ERA-Interim data. The WT centroids are then use to associate each day in the seasonal forecasts to the most similar WT centroid according to their Euclidean distance.

Seasonal Prediction Systems

We use seasonal forecasts from the North American Multi-Model Ensemble (NMME; Kirtman et al., 2014) and the Integrative Forecasting System (IFS, Version 5) seasonal forecasts from ECMWF from the Copernicus Climate Change Service. We included all NMME models that provide daily Q850 for their hindcast (retrospective forecasts) runs, which are: NCAR CESM1, UM-RSMAS CCSM4, and ECMWF IFS. We also include NASA-GMAO GEOS-5 and CCCMA CanCM4 but use Q650 and Q675 respectively instead due to missing data at the 850 hPa level. This should have little impact on the predictive skill of those models since moisture patterns are well correlated between the 850 hPa and 650 hPa level (absolute differences do not matter due to the use of daily anomalies in the WTing). All models have a horizontal grid spacing of one degree. NMME models run a ten-member ensemble forecast while ECMWF runs 25 members. All models are initialized on the first of each month and forecast 365 days except for NASA-GMAO GEOS-5 and ECMWF IFS, which forecast 273 and 215 days respectively. The common hindcast period for the NMME models is 1982–2010 while ECMWF IFS provides hindcasts for 1993–2016.

To associate seasonal forecasts days with one of the derived WTs we use the same pre-processing as described above (i.e., for the predictor variables we derive daily anomalies, apply spatial smoothing, and normalize the anomalies). Next, we regrid the ERA-Interim based WT centroids to the one-degree model grid and calculate Euclidean distances to all WTs for each forecast day. A forecast day is assigned to a WT according to the lowest Euclidean Distance. We use linearly detrended time series for calculating anomaly correlation coefficients.

Results

Monsoon season (June–October) precipitation contributes between 60% (AZ West) to 80% (NM South) to the annual rainfall in the four study areas (Figure 1b). The precipitation in this period typically comes from moisture surges from the Gulf of Mexico and tropical Pacific (Favors & Abatzoglou, 2013; Higgins et al., 1997) intercepted by dry periods. We aim to capture this variability with our WTing analysis.

Dominant Weather Types

Three weather patterns are sufficient to characterize the dominant effects of synoptic-scale variability on precipitation in the four target regions. The centroids of the three WTs in the AZ West region are shown in Figures 1d–1f. The first WT (monsoon) is associated with monsoonal moisture surges and characterized by tropical moisture advection from the Gulf of California and the eastern tropical Pacific further south into the target catchments resulting in a high water vapor mixing ratio at 850 hPa (Q850) and precipitable water anomalies. Monsoon WT days have a rapid onset in July, peak in August, and decay in September. This WT resembles the published definition of the NAM large-scale circulation and seasonal occurrence (Vera et al., 2006) and are often referred to as “moisture surges” (Pascale & Bordoni, 2016). Although only 19% of days within June to October are associated with this WT, they produce the majority of monsoon season precipitation (Figure 1c). Our definition of wet NAM days is more general than most published definition of synoptic patterns that cause wet surges. This is because we aim to use these patterns in a probabilistic predictive framework rather than understanding the drivers of wet surges in this region, which is already well studied (Higgins et al., 2004; Jiang & Lau, 2008; Schiffer & Nesbitt, 2012; Seastrand et al., 2015). Our wet WT definition includes the classical definition of wet surges but also allows capturing wet days with other synoptic-scale patterns (see Figure S13 in Supporting Information S1 and Movie S1 for an example). The second weather pattern (normal, Figure 1e) is associated with more zonal flow, a subtropical height that is weaker and centered on the coast of Texas, and climatologically average precipitable water anomalies. Days within this WT are 20%–40% dryer than average days and only contribute about one fifth of the precipitation of monsoonal flow days. The third WT (dry, Figures 1c and 1f) is associated with anomalously dry air advection, zonal flow, and extremely dry precipitation anomalies and resembles previously found dry patterns (Schiffer & Nesbitt, 2012). Dry WT days occur most frequently in June and October. Deriving WTs based on July–September conditions results in similar patterns (not shown), which indicates that these patterns are robust and inherent to the core monsoon season. An example of the occurrence of wet, normal, and dry WT days compared to daily average precipitation in the AZ West region for the dry 2009 and wet 2018 NAM season is shown in Figure S12 in Supporting Information S1.

The WTs for the other focus areas are shown in the supplement (Figures S4–S11 in Supporting Information S1). Each area features a monsoon, normal, and dry WT. The monsoonal WTs mainly differ in the size and location of the subtropical high-pressure system (Figure S15 in Supporting Information S1 shows 500 hPa geopotential height anomalies). The high-pressure system is shifted further to the east and south particularly in the NM North and South region, which facilitates the advection of Gulf of Mexico moisture rather than moisture from the Pacific as in the Arizona catchments.

The pronounced relationship between WTs and precipitation anomalies in the target regions results in a significant (Pearson's $r = 0.65$) correlation between the seasonal average frequency of monsoon WT days and observed monsoon season mean precipitation in the AZ west region (Figures 2a and 2b). Similarly high correlations are found in the other three regions (Figure 2c). Also, a high frequency of dry WT days is significantly negatively correlated with monsoon seasonal average precipitation.

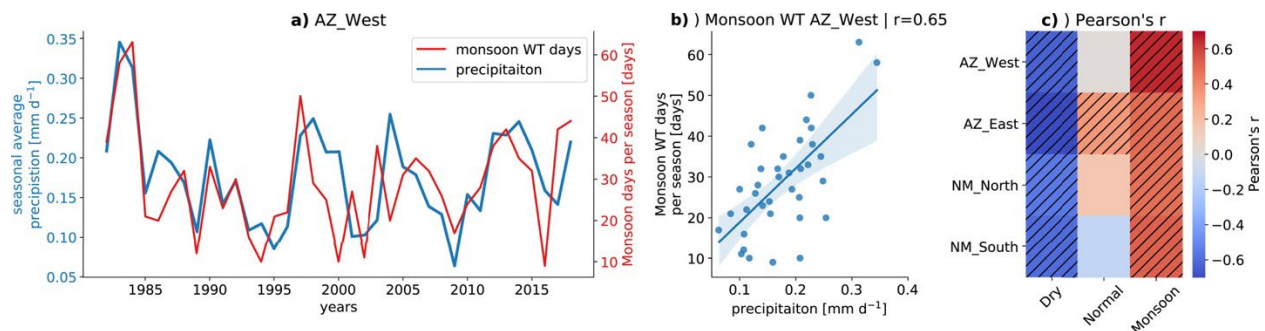


Figure 2. The frequency of North American monsoon season total monsoon weather type (WT) days is significantly correlated with seasonal catchment precipitation. (a) June to October daily average precipitation over the Arizona (AZ) West region (blue) and seasonal frequency of monsoon WT days (red) between 1982 and 2018. (b) Scatter plot and linear relationship estimate for the data shown in (a). The Pearson correlation coefficient is shown in the title. (c) Pearson correlation coefficient for all regions and WTs. Hatching indicated significant correlation coefficient (two-tailed p -value < 0.05).

Seasonal Forecasts of Monsoon Precipitation in Hindcasts

There are large differences between the performance of the analyzed models in simulating the position and seasonal evolution of the monsoonal high-pressure system that partly depend on the lead time of the forecast (Figure S15 in Supporting Information S1). The ECMWF-IFS system has the best climatological representation of the monsoonal high, while the other models feature a too far northward extension and partly erroneous seasonal progression.

Also, the ability to simulate the seasonal evolution of Q850 over the AZ West WT domain strongly depends on the forecasting system and the forecast lead time (Figures 3a and 3b). National Center for Atmospheric Research's Community Earth System Model v1 (NCAR-CESM1) forecasts feature a dry bias in Q850 during the peak monsoon season. This is not necessarily a problem as long as this bias is systematic since we perform the WTing with Q850 anomalies. A more severe issue is the considerable initiation shock—visible in the June,

July, and August forecasts—which typically has strong negative impacts on the forecast quality (Mulholland et al., 2015). Similar to the positioning of the monsoonal high, the ECMWF-IFS v5 model forecasts show a much better climatological evolution of the Q850 seasonal cycle and do not feature an initialization shock. The other forecasting systems (not shown) feature initialization shocks that are typically not as severe as the ones in the NCAR-CESM1 system.

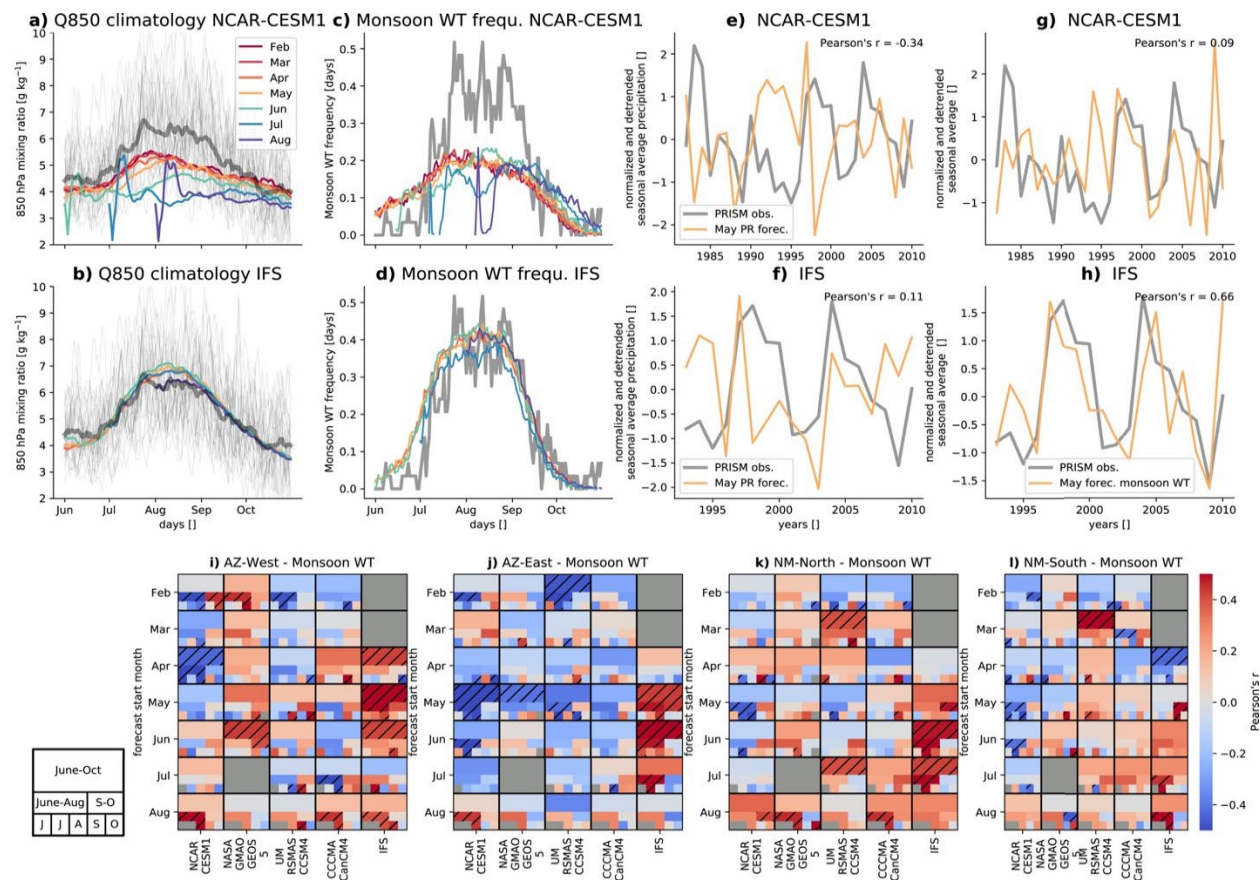


Figure 3. ECMWF's IFS model can skillfully forecast North American monsoon season precipitation starting in April when using monsoon weather type (WT) frequencies as a proxy for catchment average precipitation. ERA-Interim (gray), (a) NCAR-CESM1, and (b) ECMWF-IFS climatological average Q850 over the AZ West region. Thin gray lines show individual years from ERA-Interim and colors show different forecast initialization months. (c and d) Similar to (a and b) but for climatological average monsoon day frequencies. (e and f) Annual average detrended and normalized regional averaged observed (gray) and May forecast precipitation (orange). (g and h) Similar to (e and f) but for May forecast average detrended and normalized monsoon day frequency (orange) versus observed precipitation. (i–l) Heatmaps showing the Pearson correlation coefficient between the detrended and normalized observed basin average precipitation and forecasted monsoon WT frequency. Panels (i–l) show results from the AZ West, AZ East, NM North, and NM South watershed (from left to right). Each panel shows results from each modeling system (columns) and forecast start months (rows). Each month and model segment (see explanation on the bottom left) shows correlation coefficients for June–October (top two rows), June–August and September–October (second row from below), and each month (bottom row). Hatching indicates significant correlation coefficients ($p < 0.1$).

The quality of simulating climatological Q850 conditions is reflected in the model's ability to capture the frequency and seasonality of monsoonal flow WT days (Figure 3c and 3d and Figure S16 in Supporting Information S1). NCAR-CESM1 has too many monsoon WT days in June and too few during the peak monsoon season. The impacts of the initialization shock are also visible in the forecasted monsoon WT frequencies. IFS has a much better simulation of the amount and seasonality of monsoon WT days (Figure 3d) and generally outperforms any of the other forecasting systems (Figure S16 in Supporting Information S1).

Using the modeled precipitation for forecasting observed NAM season precipitation results in no significant skill—measured by the anomaly correlation coefficient (Figures 3e and 3f)—confirming results from previous studies (Slater et al., 2019). This is also true for using NAM season monsoon WT frequencies from the NCAR-CESM1 forecast (Figure 3g; and the other NMME models). However, using the May forecast from the IFS model allows us to skillfully predict monsoon season precipitation in the AZ West region (Figure 3h; Pearson's $r = 0.66$). Even the April forecast is skillful (Figure 3i). Predicting the precipitation in individual months is more challenging except for the July and August forecast, which are skillful in predicting precipitation during the first forecasting month. Generally, June to August precipitation is more predictable than September and October rainfall.

IFS has similar high skill in forecasting precipitation in the AZ East region (Figure 3j; except for the April forecast) compared to AZ West. Predictability is generally lower in New Mexico and skillfully forecasting NAM season precipitation in the NM North region becomes feasible in June (Figure 3k) while no significant predictability is found in the NM South region (Figure 3l).

Sources for Predictability

Here we investigate the potential sources of predictability that allow skillful sub-seasonal precipitation forecasts by the IFS model.

Compositing the top 25% of years (top 9 years out of 37) with the highest frequency of monsoonal flow WT days in June and July in the AZ West region show significant 500 hPa geopotential height anomalies that resemble a Rossby wave train with alternating high and low anomalies (Figure 4a). Using the top 10% of years results in similar patterns (not shown). We focus on June on July conditions since they have the highest predictability (see Figure 3). A similar pattern has been identified previously as driver of continental-scale precipitation variability in the U.S. (Castro et al., 2012; Ciancarelli et al., 2014). This pattern is associated with La Nina like tropical Pacific sea surface temperatures (Figure 4b), which is in line with some previous studies (Ciancarelli et al., 2014; Higgins & Shi, 2001) but in opposition to others (Grantz et al., 2007). The anomaly patterns change gradually when moving from the AZ-West to the NM-South region with the high anomaly over the CONUS moving eastward and weakening and the low anomalies in the Gulf of Alaska and near the Gulf of California becoming stronger (Figure S17 in Supporting Information S1).

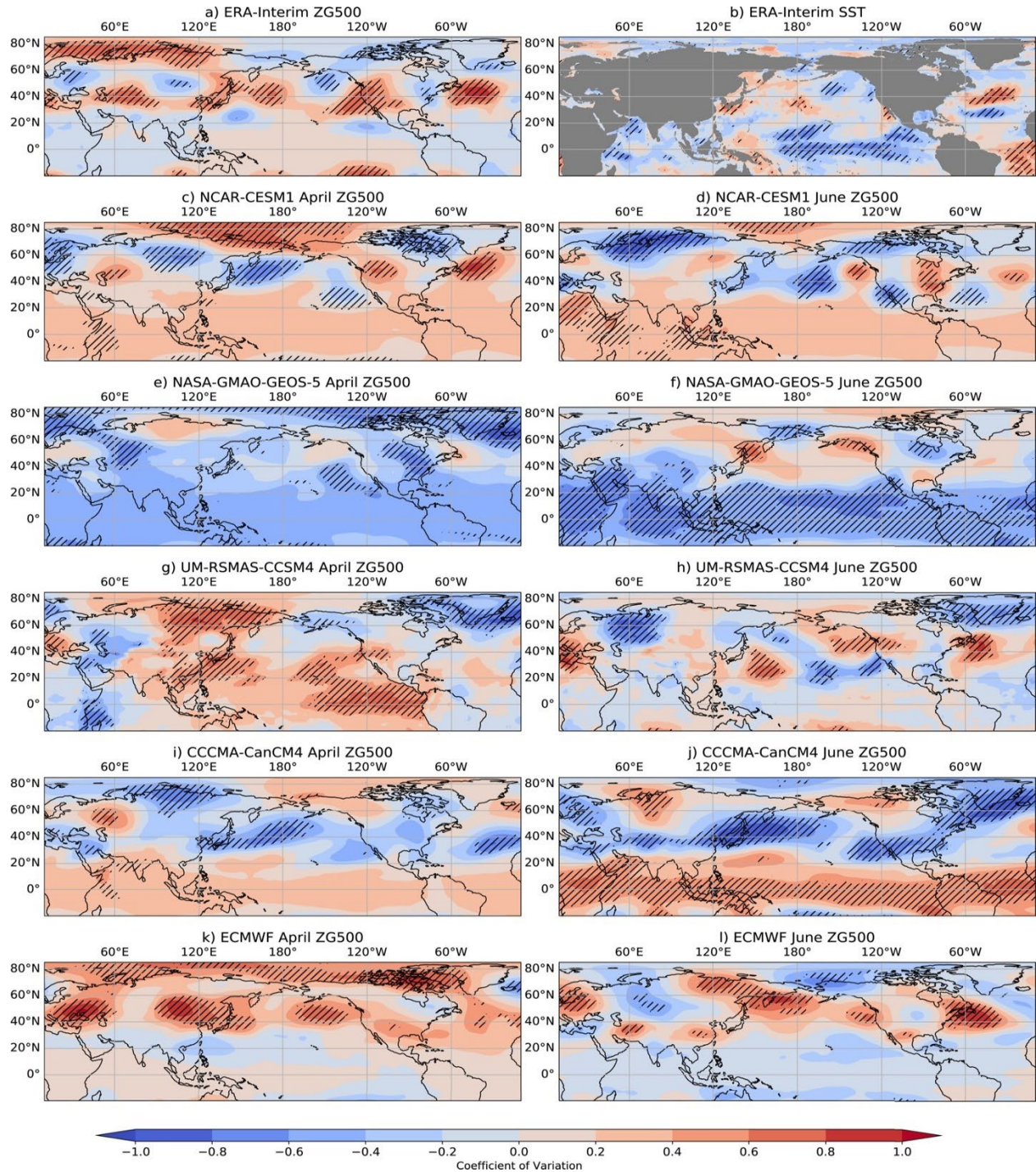


Figure 4. Coefficient of variation of the 500 hPa geopotential height (ZG500) anomaly of the 25% years with the highest June and July monsoon weather type frequencies compared to the climatological average. Shown is data from ERA-Interim (a; 1982–2018), the four included North American multi-model ensemble models (c–j; 1982–2010); and the ECMWF forecast (k–l; 1993–2016). Coefficients of variation are shown for April (left) and June (right) forecasts for each modeling system. Additionally, coefficient of variation of sea surface temperatures are shown based on ERA-Interim data (b). Hatched areas show significant anomalies based on the Mann-Whitney U test ($p \leq 0.1$).

None of the forecasting systems is able to capture these anomaly patterns perfectly although we should expect some differences due to the different simulation periods. The June forecasts from the IFS model resemble the observed pattern best (Figure 4l) while the model's April forecasts miss the low anomaly in the Gulf of Alaska but still feature a correctly but too weak high anomaly over the central U.S. (Figure 4k). Nevertheless, IFS's April forecast better captures the observed pattern than most of the models forecast in June. For example, the GEOS-5 model has too strong and negative connections to the tropical Pacific (Figure 4f) while CanCM4 has too strong positive connections (Figure 4j). Most models struggle with correctly simulating the positive height anomaly in the western U.S. and many models show fundamentally different patterns in their April forecasts compared to their June forecasts.

Discussion and Conclusion

We assess the ability of seasonal forecasting systems to predict the North American Monsoon (NAM, June to October) precipitation in Arizona and New Mexico catchments that are important for water management. Due to significant biases in simulating precipitation in state-of-the-art seasonal forecasting systems (Crochemore et al., 2016), we developed a forecasting framework that uses hydrologically important synoptic-scale weather types (WTs) for catchment-scale precipitation predictions. We found that three WTs—a dry, normal, and wet (monsoonal flow) WT—are sufficient to characterize the interseasonal and interannual variability of precipitation in the U.S. NAM region.

We show that the observed seasonal average frequency of monsoonal flow WTs significantly correlates with catchment-scale seasonal average precipitation in all regions. Most importantly, we show that ECMWF's IFS forecasting system can skillfully predict NAM season rainfall at catchment scales with several months lead-time (i.e., the April forecast is skillful over the AZ West catchments) except for the NM South catchments while using the model's precipitation as a predictor does not provide predictive skill. The other evaluated forecasting systems do not have predictive skills partly due to large initialization shocks (Mulholland et al., 2015), and an erroneous simulation of the low-level moisture transport into the study regions. Additionally, ECMWF-IFS has a superior simulation of the location and seasonal evolution of the monsoonal high-pressure ridge and outperforms the other models in simulating teleconnection patterns (Castro et al., 2012; Ciancarelli et al., 2014) that control the position of the monsoon high and the amount of regional precipitation. The superior performance of the ECMWF-IFS system is not dependent on the use of ERA5 reanalysis data, since using MERRA2 data (Gelaro et al., 2017) results in similar seasonal and inter-annual WT properties (not shown).

These results push the boundaries of seasonal predictability of regional precipitation (Slater et al., 2019) and offer novel opportunities for improved water resource management on S2S time scales in the U.S. Southwest. The presented WTing framework is flexible and could offer enhanced predictive skill also for sub-seasonal forecasts and in other drought-prone regions around the world. The presented results also show that many S2S systems have to advance their data assimilation systems and reduce biases in their climatological representation of regional phenomena such as monsoon circulations. Having multiple prediction systems with the quality of ECMWF-IFS would allow us to construct more skillful multi-model ensemble forecasts resulting in further improvements of predictive skill. Importantly, even skillful models such as ECMWF-IFS are not able to simulate the mesoscale processes that drive local precipitation. Future

research should focus on the potential added value of using convection-permitting models (Prein et al., 2015) for S2S prediction, which could result in a step-improvement in our skill to simulate mesoscale weather phenomena and their extremes (Clark et al., 2016; Prein et al., 2021), snowpack dynamics (Rasmussen et al., 2011), land-atmosphere interactions (Barlage et al., 2021), and could result in improved S2S predictive skills by better simulating teleconnections due to the dynamic interaction of convective scales with larger-scale processes (Weber & Mass, 2019).

Acknowledgements

This work was sponsored by the U.S. Bureau of Reclamation Interagency Agreements R20PG00065 and R20PG00083. We would like to acknowledge high-performance computing support from Cheyenne (<https://doi.org/10.5065/D6RX99HX>) provided by NCAR's Computational and Information Systems Laboratory. NCAR is sponsored by the National Science Foundation under Cooperative Agreement 1852977.

Data and Code Availability

ERA-Interim data can be accessed from <https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>, PRISM precipitation data from <https://prism.oregonstate.edu/>, and NMME seasonal hindcasts can be downloaded from <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-multi-model-ensemble>. Finally, the ECMWF-IFS hindcasts and forecasts can be accessed from <https://climate.copernicus.eu/seasonal-forecasts>. The code for the statistical analysis and visualization of data in this report is openly available through GitHub (https://github.com/AndreasPrein/NAM_S2S_predictability).

References

- Adams, D. K., & Comrie, A. C. (1997). The North American monsoon. *Bulletin of the American Meteorological Society*, 78(10), 2197– 2213. [https://doi.org/10.1175/1520-0477\(1997\)078<2197:tnam>2.0.co;2](https://doi.org/10.1175/1520-0477(1997)078<2197:tnam>2.0.co;2)
- Barlage, M., Chen, F., Rasmussen, R., Zhang, Z., & Miguez-Macho, G. (2021). The importance of scale-dependent groundwater processes in land-atmosphere interactions over the Central United States. *Geophysical Research Letters*, 48(5), e2020GL092171. <https://doi.org/10.1029/2020gl092171>
- Beck, C., Philipp, A., & Streicher, F. (2016). The effect of domain size on the relationship between circulation type classifications and surface climate. <https://doi.org/10.1002/joc.3688>
- Bureau of Reclamation, U. D. o. t. I. (2018). Annual operating plan for Colorado river reservoirs 2019. Tech. Rep. Retrieved from <https://www.usbr.gov/lc/region/g4000/aop/AOP19.pdf>
- Bureau of Reclamation, U. D. o. t. I. (2021). Glen canyon dam adaptive management program. Accessed July 8, 2021.

- Carleton, A. M. (1986). Synoptic-dynamic character of “bursts” and “breaks” in the South-West US summer precipitation singularity. *Journal of Climatology*, 6(6), 605– 623. <https://doi.org/10.1002/joc.3370060604>
- Carleton, A. M., Carpenter, D. A., & Weser, P. J. (1990). Mechanisms of interannual variability of the southwest United States summer rainfall maximum. *Journal of Climate*, 3(9), 999– 1015. [https://doi.org/10.1175/1520-0442\(1990\)003<0999:moivot>2.0.co;2](https://doi.org/10.1175/1520-0442(1990)003<0999:moivot>2.0.co;2)
- Castro, C. L., Chang, H.-I., Dominguez, F., Carrillo, C., Schemm, J.-K., & Henry Juang, H.-M. (2012). Can a regional climate model improve the ability to forecast the North American monsoon? *Journal of Climate*, 25(23), 8212– 8237. <https://doi.org/10.1175/jcli-d-11-00441.1>
- Castro, C. L., McKee, T. B., & Pielke Sr, R. A. (2001). The relationship of the North American monsoon to tropical and North Pacific Sea surface temperatures as revealed by observational analyses. *Journal of Climate*, 14(24), 4449– 4473. [https://doi.org/10.1175/1520-0442\(2001\)014<4449:trotna>2.0.co;2](https://doi.org/10.1175/1520-0442(2001)014<4449:trotna>2.0.co;2)
- Ciancarelli, B., Castro, C. L., Woodhouse, C., Dominguez, F., Chang, H.-I., Carrillo, C., & Griffin, D. (2014). Dominant patterns of US warm season precipitation variability in a fine resolution observational record, with focus on the southwest. *International Journal of Climatology*, 34(3), 687– 707. <https://doi.org/10.1002/joc.3716>
- Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016). Convection-permitting models: A step-change in rainfall forecasting. *Meteorological Applications*, 23(2), 165– 181. <https://doi.org/10.1002/met.1538>
- Comrie, A. C. (1996). An all-season synoptic climatology of air pollution in the US-Mexico border region. *The Professional Geographer*, 48(3), 237– 251. <https://doi.org/10.1111/j.0033-0124.1996.00237.x>
- Crochemore, L., Ramos, M.-H., & Pappenberger, F. (2016). Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, 20(9), 3601– 3618. <https://doi.org/10.5194/hess-20-3601-2016>
- Daly, C., Taylor, G., & Gibson, W. (1997). The PRISM approach to mapping precipitation and temperature. In *Proc., 10th AMS conf. on applied climatology* (pp. 20– 23).
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553– 597. <https://doi.org/10.1002/qj.828>
- Ekstrom, M., Jonsson, P., & Barring, L. (2002). Synoptic pressure patterns associated with major wind erosion events in southern Sweden. *Climate Research*, 23, 51– 66. <https://doi.org/10.3354/cr023051>

- Favors, J. E., & Abatzoglou, J. T. (2013). Regional surges of monsoonal moisture into the southwestern United States. *Monthly Weather Review*, **141**(1), 182– 191. <https://doi.org/10.1175/mwr-d-12-00037.1>
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, **30**(14), 5419– 5454. <https://doi.org/10.1175/jcli-d-16-0758.1>
- Grantz, K., Rajagopalan, B., Clark, M., & Zagona, E. (2007). Seasonal shifts in the North American monsoon. *Journal of Climate*, **20**(9), 1923– 1935. <https://doi.org/10.1175/jcli4091.1>
- Griffin, D., Woodhouse, C. A., Meko, D. M., Stahle, D. W., Faulstich, H. L., Carrillo, C., et al. (2013). North American monsoon precipitation reconstructed from tree-ring latewood. *Geophysical Research Letters*, **40**(5), 954– 958. <https://doi.org/10.1002/grl.50184>
- Gutzler, D. S., & Preston, J. W. (1997). Evidence for a relationship between spring snow cover in North America and summer rainfall in New Mexico. *Geophysical Research Letters*, **24**(17), 2207– 2210. <https://doi.org/10.1029/97gl02099>
- Higgins, R., Mo, K., & Yao, Y. (1998). Interannual variability of the US summer precipitation regime with emphasis on the southwestern monsoon. *Journal of Climate*, **11**(10), 2582– 2606. [https://doi.org/10.1175/1520-0442\(1998\)011<2582:ivotus>2.0.co;2](https://doi.org/10.1175/1520-0442(1998)011<2582:ivotus>2.0.co;2)
- Higgins, R., & Shi, W. (2000). Dominant factors responsible for interannual variability of the summer monsoon in the Southwestern United States. *Journal of Climate*, **13**(4), 759– 776. [https://doi.org/10.1175/1520-0442\(2000\)013<0759:dfrfiv>2.0.co;2](https://doi.org/10.1175/1520-0442(2000)013<0759:dfrfiv>2.0.co;2)
- Higgins, R., & Shi, W. (2001). Intercomparison of the principal modes of interannual and intraseasonal variability of the North American monsoon system. *Journal of Climate*, **14**(3), 403– 417. [https://doi.org/10.1175/1520-0442\(2001\)014<0403:iotpmo>2.0.co;2](https://doi.org/10.1175/1520-0442(2001)014<0403:iotpmo>2.0.co;2)
- Higgins, R., Shi, W., & Hain, C. (2004). Relationships between Gulf of California moisture surges and precipitation in the southwestern United States. *Journal of Climate*, **17**(15), 2983– 2997. [https://doi.org/10.1175/1520-0442\(2004\)017<2983:rbgocm>2.0.co;2](https://doi.org/10.1175/1520-0442(2004)017<2983:rbgocm>2.0.co;2)
- Higgins, R., Yao, Y., & Wang, X. (1997). Influence of the North American monsoon system on the US summer precipitation regime. *Journal of Climate*, **10**(10), 2600– 2622. [https://doi.org/10.1175/1520-0442\(1997\)010<2600:iotnam>2.0.co;2](https://doi.org/10.1175/1520-0442(1997)010<2600:iotnam>2.0.co;2)
- Higgins, R. W., Chen, Y., & Douglas, A. V. (1999). Interannual variability of the North American warm season precipitation regime. *Journal of Climate*, **12**(3), 653– 680. [https://doi.org/10.1175/1520-0442\(1999\)012<0653:ivotna>2.0.co;2](https://doi.org/10.1175/1520-0442(1999)012<0653:ivotna>2.0.co;2)
- Ikeda, K., Rasmussen, R., Liu, C., Newman, A., Chen, F., Barlage, M., et al. (2021). Snowfall and snowpack in the Western US as captured by convection permitting climate simulations: Current climate and pseudo global warming future climate. *Climate Dynamics*, **57**(7), 2191– 2215. <https://doi.org/10.1007/s00382-021-05805-w>

- Jiang, X., & Lau, N.-C. (2008). Intraseasonal teleconnection between North American and Western North Pacific monsoons with 20-day time scale. *Journal of Climate*, 21(11), 2664– 2679. <https://doi.org/10.1175/2007jcli2024.1>
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585– 601. <https://doi.org/10.1175/bams-d-12-00050.1>
- Krishnamurti, T., Stefanova, L., Chakraborty, A., Tsv, V., Cocke, S., Bachiochi, D., & Mackey, B. (2002). Seasonal forecasts of precipitation anomalies for North American and Asian monsoons. *Journal of the Meteorological Society of Japan. Ser. II*, 80(6), 1415– 1426. <https://doi.org/10.2151/jmsj.80.1415>
- Lehner, F., Deser, C., & Terray, L. (2017). Toward a new estimate of “Time of emergence” of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, 30(19), 7739– 7756. <https://doi.org/10.1175/jcli-d-16-0792.1>
- Li, S., & Robertson, A. W. (2015). Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Monthly Weather Review*, 143(7), 2871– 2889. <https://doi.org/10.1175/mwr-d-14-00277.1>
- Liu, J., Yang, H., Gosling, S. N., Kummerow, M., Flörke, M., Pfister, S., et al. (2017). Water scarcity assessments in the past, present, and future. *Earth's Future*, 5(6), 545– 559. <https://doi.org/10.1002/2016ef000518>
- MacDonald, G. M. (2010). Water, climate change, and sustainability in the southwest. *Proceedings of the National Academy of Sciences*, 107(50), 21256– 21262. <https://doi.org/10.1073/pnas.0909651107>
- Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science Advances*, 2(2), e1500323. <https://doi.org/10.1126/sciadv.1500323>
- Milly, P. C., & Dunne, K. A. (2020). Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. *Science*, 367(6483), 1252– 1255. <https://doi.org/10.1126/science.aay9187>
- Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in Western North America. *Bulletin of the American Meteorological Society*, 86(1), 39– 50. <https://doi.org/10.1175/bams-86-1-39>
- Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Monthly Weather Review*, 143(11), 4631– 4644. <https://doi.org/10.1175/mwr-d-15-0076.1>
- Pascale, S., & Bordoni, S. (2016). Tropical and extratropical controls of Gulf of California surges and summertime precipitation over the southwestern United States. *Monthly Weather Review*, 144(7), 2695– 2718. <https://doi.org/10.1175/mwr-d-15-0429.1>

- Prein, A., Rasmussen, R., Wang, D., & Giangrande, S. (2021). Sensitivity of organized convective storms to model grid spacing in current and future climates. *Philosophical Transactions of the Royal Society A*, 379(2195), 20190546. <https://doi.org/10.1098/rsta.2019.0546>
- Prein, A. F. (2021). NAM_S2S_predictability. Retrieved from https://github.com/AndreasPrein/NAM_S2S_predictability
- Prein, A. F., Holland, G. J., Rasmussen, R. M., Clark, M. P., & Tye, M. R. (2016). Running dry: The US Southwest's drift into a drier climate state. *Geophysical Research Letters*, 43(3), 1272– 1279. <https://doi.org/10.1002/2015gl066727>
- Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53(2), 323– 361. <https://doi.org/10.1002/2014rg000475>
- Prein, A. F., & Mearns, L. O. (2021). US Extreme precipitation weather types increased in frequency during the 20th century. *Journal of Geophysical Research: Atmospheres*, 126(7), e2020JD034287. <https://doi.org/10.1029/2020jd034287>
- Raff, D., Brekke, L., Werner, K., Wood, A., & White, K. (2013). Short-term water management decisions: User needs for improved climate, weather, and hydrologic information. Tech. Rep. U.S. Army Corps of Engineers. Retrieved from <https://www.usbr.gov/research/st/roadmaps/WaterSupply.pdf>
- Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., et al. (2011). High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *Journal of Climate*, 24(12), 3015– 3048. <https://doi.org/10.1175/2010jcli3985.1>
- Romesburg, C. (2004). Cluster analysis for researchers. Lulu. com.
- Schiemann, R., & Frei, C. (2010). How to quantify the resolution of surface climate by circulation types: An example for alpine precipitation. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(9), 403– 410. <https://doi.org/10.1016/j.pce.2009.09.005>
- Schiffer, N. J., & Nesbitt, S. W. (2012). Flow, moisture, and thermodynamic variability associated with Gulf of California surges within the North American monsoon. *Journal of Climate*, 25(12), 4220– 4241. <https://doi.org/10.1175/jcli-d-11-00266.1>
- Seastrand, S., Serra, Y., Castro, C., & Ritchie, E. (2015). The dominant synoptic-scale modes of North American monsoon precipitation. *International Journal of Climatology*, 35(8), 2019– 2032. <https://doi.org/10.1002/joc.4104>
- Slater, L. J., Villarini, G., & Bradley, A. A. (2019). Evaluation of the skill of North-American Multi-Model Ensemble (NMME) global climate models in predicting average and extreme precipitation and temperature over the continental USA. *Climate Dynamics*, 53(12), 7381– 7396. <https://doi.org/10.1007/s00382-016-3286-1>

- Vera, C., Higgins, W., Amador, J., Ambrizzi, T., Garreaud, R., Gochis, D., et al. (2006). Toward a unified view of the American monsoon systems. *Journal of Climate*, 19(20), 4977– 5000. <https://doi.org/10.1175/jcli3896.1>
- Weber, N. J., & Mass, C. F. (2019). Subseasonal weather prediction in a global convection-permitting model. *Bulletin of the American Meteorological Society*, 100(6), 1079– 1089. <https://doi.org/10.1175/bams-d-18-0210.1>
- Yu, B., & Wallace, J. M. (2000). The principal mode of interannual variability of the North American monsoon system. *Journal of Climate*, 13(15), 2794– 2800. [https://doi.org/10.1175/1520-0442\(2000\)013<2794:tpmoiv>2.0.co;2](https://doi.org/10.1175/1520-0442(2000)013<2794:tpmoiv>2.0.co;2)
- Zhu, C., Lettenmaier, D. P., & Cavazos, T. (2005). Role of antecedent land surface conditions on North American monsoon rainfall variability. *Journal of Climate*, 18(16), 3104– 3121. <https://doi.org/10.1175/jcli3387.1>

Appendix B – Component 2 Paper

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

Erin Towler¹, Dagmar Llewellyn², Andreas Prein¹, and Lucas Barrett²

¹ National Center for Atmospheric Research (NCAR), Boulder, CO

² Bureau of Reclamation, Albuquerque, NM

Paper was published in the Proceedings of the 2023 Federal Interagency Sedimentation and Hydrologic Modeling Conference (SEDHYD) in St. Louis, MO

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, a collaboration between scientists at the National Center for Atmospheric Research (NCAR) and water managers at the Bureau of Reclamation (Reclamation), seeks to improve these seasonal monsoon forecasts. It is funded by Reclamation’s Science & Technology Program (S&T project number 20032). 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 United States. 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 skillfully captured monsoonal moisture in retrospective forecasts produced by the European Centre for Medium-Range Weather Forecasts (ECMWF).

The purpose of this study is to explore the use of the WTs developed in Prein et al., 2022 (Appendix A) to develop seasonal forecasts of monsoon precipitation characteristics. Specifically, 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. 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 Rio Grande and 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).

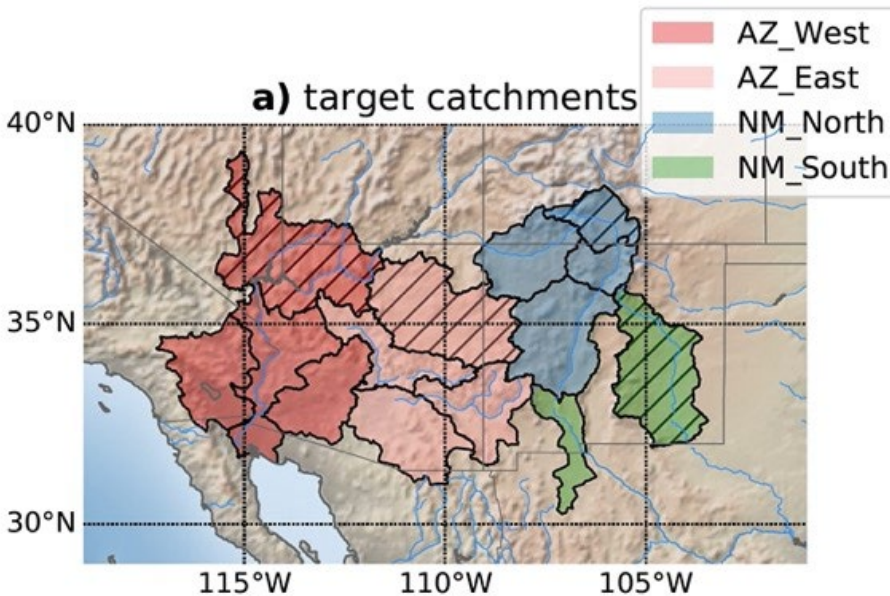


Figure 1. Hydrologic unit codes (HUCs) that exhibit similar weather types (WTs) are combined to create 4 regions: Arizona West (AZ-West, red), Arizona East (AZ-East, pink), New Mexico North (NM-North, blue) and New Mexico South (NM-South, green). Hatching shows the basin that was used to derive the WTs. 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 North American Monsoon 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 within the period 1982 to 2018 (Dee et al. 2011) 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 used the ECMWF forecasts were used here. We downloaded seasonal forecasts of Q850 from ECMWF from 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 are pooled together, from 1993–2018, and are

referred to 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 $G(\cdot)$ 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 33rd percentile (Q33), “2” if it is between Q33 and the 66th 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:

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

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

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

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:

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

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

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

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:

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

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:

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

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

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

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):

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

and

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

where for a given year, i , $p = (p_{i,1}, p_{i,2}, \dots, p_{i,k})$ 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).

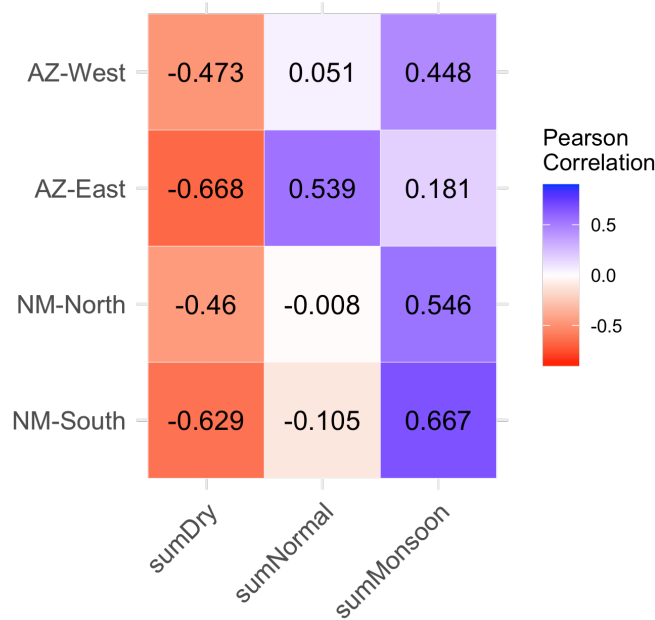


Figure 2. Pearson’s correlation between prediction period of June-October (JJASO) average historical precipitation from PRISM and the sum of weather types (WTs) from the historical reanalysis.

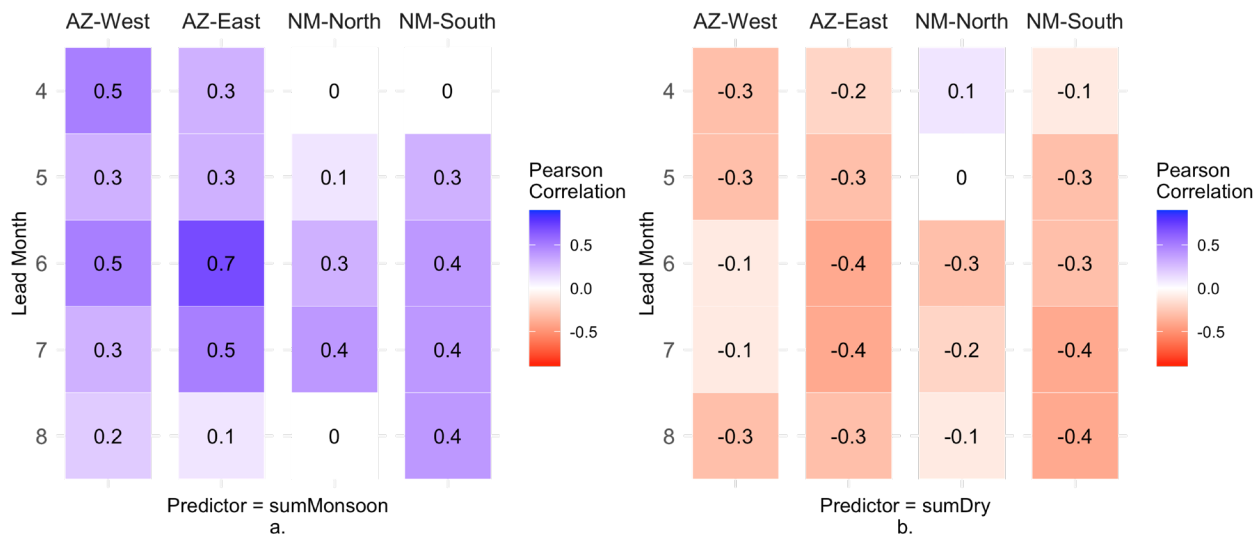


Figure 3. Pearson's correlation between seasonal average historical precipitation and the sum of the number of a) monsoon WT days (sumMonsoon) and b) dry WTs (sumDry) from the ECMWF hindcasts by lead time. Leads 4, 5, and 6 predict June-October (JJAOS); lead 7 predicts July-October (JAOS); lead 8 predicts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

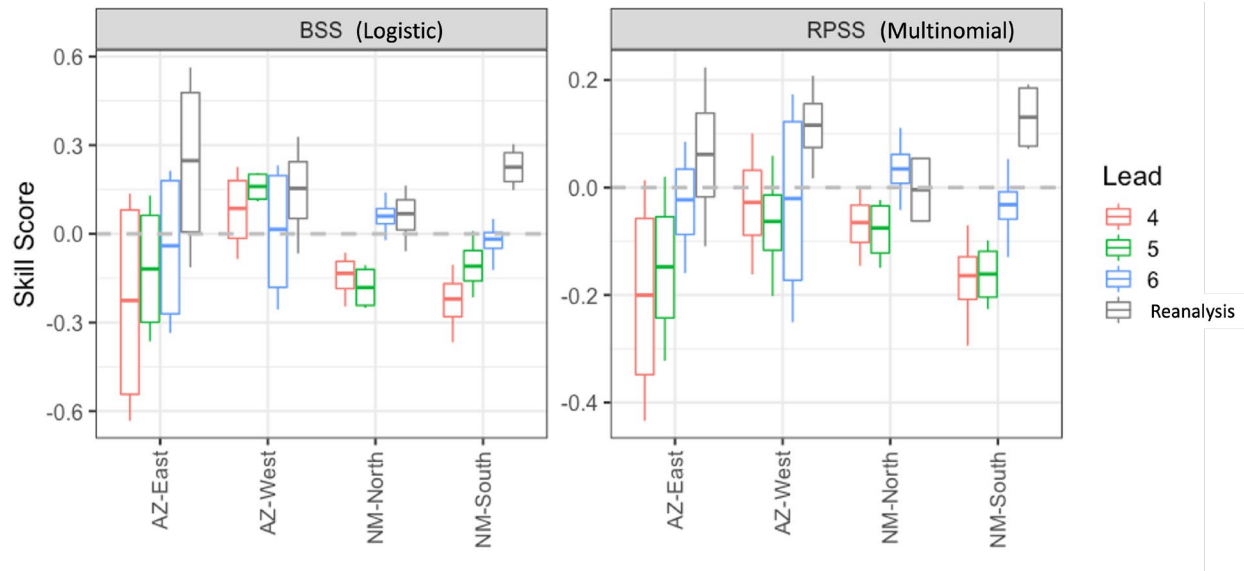


Figure 4. For the June-October prediction season, Brier Skill Score (BSS; left) for the logistic regression and Rank Probability Skill Score (RPSS, right) for the multinomial regression using ECMWF hindcasts for leads 4, 5, and 6, as well as the historical reanalysis. Each boxplot contains 4 skill scores (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.

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

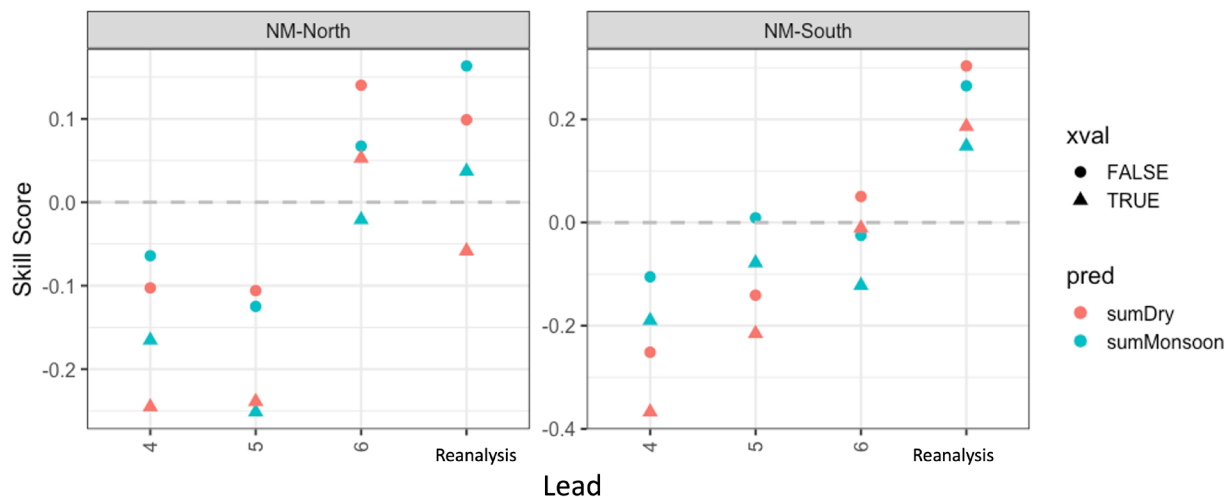


Figure 5. For the June-October prediction season, Brier Skill Score (BSS) for logistic regression using ECMWF weather type (WT) hindcasts for leads 4, 5, and 6, as well as the WTs from the historical reanalysis. Colors indicate if the predictor is the number of dry days (sumDry) or monsoon days (sumMonsoon),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

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.

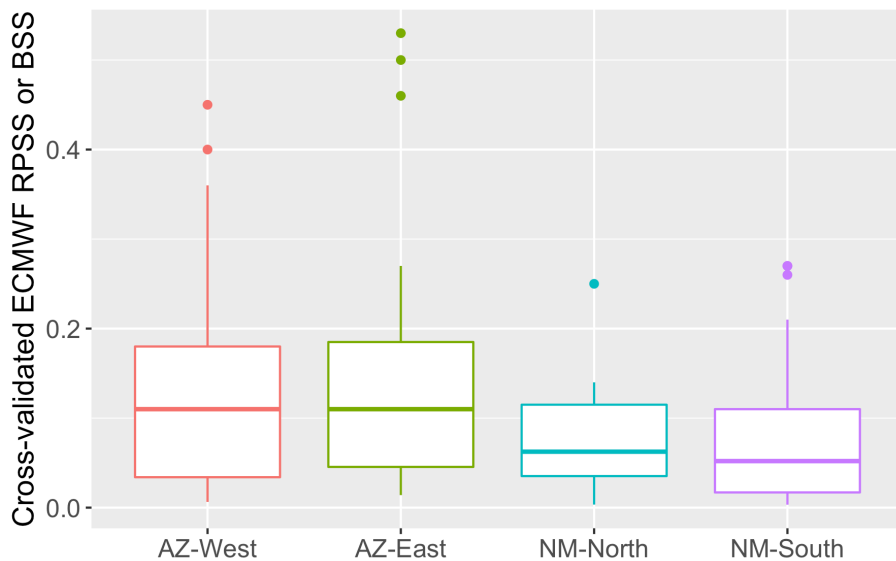


Figure 6. Cross-validated skill scores for each region; BSS = Brier Skill Score; RPSS=Rank Probability Skill Score.

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). Although not discussed here, NM-North regression results are included at the end of this Appendix (Tables 4 and 5).

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.

Table 1. Logistic (binomial) regression models for NM-South that were skillful in terms of the cross-validated

Lead Month	Predicted Season	BSS (Ecmwf_xval)	BSS (Reanalysis)	BSS (Rean_xval)	x	avg(x)	sd(x)	intercept	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

ECMWF (Ecmwf_xval) and reanalysis (Rean_xval); BSS = Brier skill score; M = sumMonsoon, D = sumDry

Table 2. Multinomial (propodds) models for NM-South that were skillful in terms of the cross-validated

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	4.1	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	3.2	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

ECMWF and Reanalysis; RPSS = Rank Probability Skill Score; M = sumMonsoon, D = sumDry

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.

Plot the ensemble mean WT frequency during the NAM season



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

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.

Table 3. Transposed Colab Notebook output showing logistic (binomial) regression models for NM-South for lead 6 of 2020

	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

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

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.

Table 4. Logistic (binomial) regression models for NM-North that were skillful in terms of the cross-validated

Lead Month	Predicted Season	BSS (Ecmwf_xval)	BSS (Reanalysis)	BSS (Rean_xval)	x	avg(x)	sd(x)	intercept1	slope	Q50 (mm/Day)
7	Jul	0.25	0.23	0.11	M	9.6	3.0	0.026	1.1	1.6
7	Jul	0.14	0.33	0.23	D	8.1	4.3	-0.091	-1.5	1.6
7	JA	0.066	0.13	0.0019	M	24	3.7	0.011	0.89	1.7
7	JASO	0.059	0.21	0.088	M	28	4.3	0.013	1.16	1.5
7	JAS	0.042	0.19	0.075	M	28	4.3	0.019	1.16	1.6
7	JASO	0.027	0.22	0.090	D	49	6.8	0.020	-1.17	1.5
7	JAS	0.011	0.23	0.11	D	23	5.7	-0.011	-1.15	1.6
8	Aug	0.13	0.39	0.29	M	14	4.6	0.16	2.0	1.7
8	Aug	0.10	0.23	0.11	D	3.9	2.5	-0.25	-1.8	1.7

ECMWF (Ecmwf_xval) and reanalysis (Rean_xval); BSS = Brier skill score; M = sumMonsoon, D = sumDry

Table 5. Multinomial (propodds) models for NM-North that were skillful in terms of the cross-validated

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)
6	JJA	0.039	0.15	0.047	M	30	3.3	0.80	-0.81	1.2	1.1	1.4
6	JA	0.0034	0.16	0.052	M	29	3.1	0.81	-0.79	1.2	1.4	1.9
7	JA	0.13	0.16	0.052	M	24	3.7	0.81	-0.79	1.2	1.4	1.9
7	JASO	0.11	0.15	0.042	M	28	4.3	0.84	-0.76	1.2	1.3	1.5
7	JASO	0.11	0.16	0.043	D	49	6.8	0.82	-0.85	-1.3	1.3	1.5
7	JA	0.038	0.27	0.16	D	11	4.7	0.78	-1.2	-1.7	1.4	1.9
7	JAS	0.0058	0.15	0.045	M	28	4.3	0.95	-0.64	1.1	1.4	1.7

ECMWF and Reanalysis; RPSS = Rank Probability Skill Score; M = sumMonsoon, D = sumDry

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).

Appendix C – Component 3 Summary

Implementing the Experimental Monsoon Forecasts in the Upper Rio Grande Water Operations Model Annual Operating Plan Runs

Taylor Adams¹, Nick Mander¹, Erin Towler²

¹ Hydros Consulting, Boulder CO

² National Center for Atmospheric Research, Boulder, CO

Derived from Hydros Consulting, Inc. memo to Cindy Stokes, New Mexico Interstate Stream Commission.

1. Background

To assess the sensitivity of the system hydrology and management to the forecasts, the experimental monsoon forecasts developed by the National Center for Atmospheric Research (NCAR) were implemented in one of Reclamation’s operational models: the Upper Rio Grande Water Operations Model (URGWOM). The runs were conducted by Hydros Consulting, Inc, as part of a collaboration under a WaterSMART Applied Science grant from Reclamation to the New Mexico Interstate Stream Commission entitled: “Developing a Projection Tool for Otowi Index Supply and Elephant Butte Effective Index Supply for Rio Grande Compact Compliance”. The purpose of this Appendix is to document how the monsoon tercile forecasts generated by NCAR were implemented, as well as the results.

2. Methods

URGWOM is used for short-term planning by various stakeholders in the Rio Grande Basin. By combining operating rules with physical process modeling, accounting, and hydrologic forecast inputs, URGWOM is used for model runs called “Annual Operating Plan” (AOP) runs. AOP runs are used by many agencies in the spring to predict end-of-year conditions. The URGWOM AOP runs are used to project the annual Elephant Butte Delivery, which is relevant to the New Mexico Credit component of Rio Grande Compact (Compact) accounting. At the end of each year, the New Mexico Credit at Elephant Butte is adjusted based on the annual Elephant Butte Delivery (calculated as the change in Elephant Butte storage plus outflow) minus the annual delivery obligation (Elephant Butte Effective Index Supply, which is defined in the Rio Grande Compact based on the annual Otowi Index Supply).

Historically, the AOP model runs have used USDA/NRCS (United States Department of Agriculture/ Natural Resources Conservation Service) snowmelt forecasts to inform the hydrology for the entire year. However, this method does not consider summer and fall monsoon events, which can have a large effect on Elephant Butte deliveries, particularly when they occur in the Middle Rio Grande (between Cochiti Lake and Elephant Butte Reservoir). To try to improve the accuracy of Elephant Butte delivery predictions in URGWOM AOP model runs,

Hydros tested AOP runs using the monsoon tercile forecast (dry, average, or wet) generated by NCAR combined with the NRCS/USDA snowmelt hindcast, and compared them to the AOP runs forced with only the NRCS/USDA snowmelt hindcast.

Hydros used AOP runs from the URGWOM Tech Team from the last 10 years. Hydros attempted to use runs from May, but in some cases, only runs from February or April could be found (Table 1). NCAR provided probabilistic monsoon hindcasts for the corresponding years (2012-2021), which were used to identify the July-December hydrologic year type at Otowi. The NCAR hindcasts are shown in Table 2.

Table 1. AOP runs used for testing

Year	AOP Model	Ruleset	AOP start date
2012	TestCombinedAcct-AOP_5-9-12.mdl.gz	URGWOM_5.0.0_5-3-12_TestCombined.rls.gz	5/1/12
2013	URGWOM_DRAFT_2013AOPModel_Feb50percent_ManualLobatos_WODeviation_TestRGRconciliation.mdl.gz	URGWOM_5.0.2_2-12-13.rls.gz	2/6/13
2014	URGWOM_6.0_2014AOP_April50Percent.mdl.gz	URGWOM_6.0_4-21-14.rls.gz	4/18/14
2015	RiverWare 6.6.4_May_50percent_AOP_NM.mdl.gz	URGWOM_6.6_4-7-15.rls.gz	5/6/15
2016	URGWOM_6.8_2016AOP_May50Percent_DefaultUpdated_StoreAndRelease.mdl.gz	URGWOM_6.8_04-11-16_Merge.rls.gz	5/4/16
2017	URGWOM_7.0_2017AOP_May50Percent_Updated_06-13-17.mdl.gz	URGWOM_7.0_06-12-17_Merge.rls.gz	5/3/17
2018	Account_7.2_2018_May22_50pct_AOP.mdl.gz	URGWOM_7.2_05-02-18.rls.gz	5/22/18
2019	Account_7.4_2019_0505_mayAOP_50%.mdl.gz	URGWOM_7.4_03-05-19.rls.gz	5/6/19
2020	URGWOM_27May2020_AOP_50_AbiquiuEASetup.mdl.gz	URGWOM_8.0_05-27-20.rls.gz	5/27/20
2021	URGWOM_05Apr2021AOP70_UPDATED.mdl.gz	URGWOM_8.2_04-20-21.rls	4/5/21

Table 2. NCAR probabilistic hindcasts used to identify the Otowi hydrologic year type

Year	Probability of <33 rd percentile flow	Probability of 33 rd -66 th percentile flow	Probability of >66 th percentile flow	Most-likely hydrologic Category
2012	0.1	0.29	0.6	Wet
2013	0.16	0.33	0.51	Wet
2014	0.06	0.2	0.74	Wet
2015	0.03	0.1	0.87	Wet
2016	0.29	0.39	0.32	Normal
2017	0.28	0.35	0.38	Wet
2018	0.16	0.29	0.55	Wet
2019	0.26	0.38	0.36	Normal
2020	0.28	0.38	0.34	Normal
2021	0.4	0.37	0.23	Dry

URGWOM cannot use a probabilistic forecast at this time. Therefore, the most-likely hydrologic “Category” provided by NCAR was input to URGWOM using the “InputForecastData.SwitchToUseHydrologicCategoryYearForPostForecastPeriod” tableslot, in each of the 10 AOP runs. An example is shown below, for the year 2015, which had a very strong Wet hindcast in the NCAR tool (Figure 1).

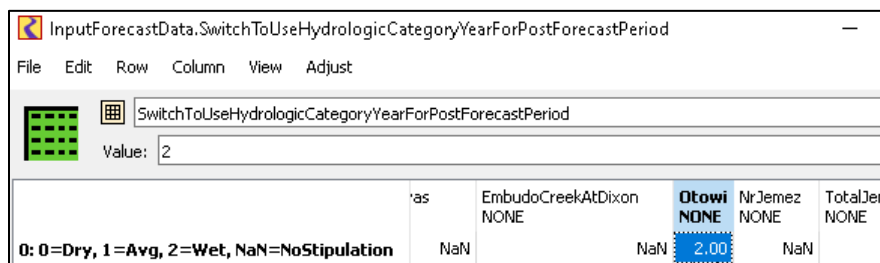


Figure 1. Implementing NCAR’s hydrologic category hindcast in URGWOM.

3. Results

NCAR’s hindcasts of the Otowi July-December hydrologic categories were input into the ten AOP runs. The effect on modeled Elephant Butte deliveries are shown in Table 3 and Figure 2. NCAR’s forecast showed improvement in 9 out of 10 hindcast years. The one year that was off was 2013: in 2013, URGWOM over-predicted the snowmelt delivery to Elephant Butte, so by adding in the NCAR 2013 wet monsoon prediction (which was correct, as there was historical flooding in September of 2013), the annual delivery was too high in URGWOM.

Table 3. Actual versus modeled Elephant Butte Delivery in acre-feet, with NCAR hindcasts

Year	Actual	Original AOP projected	AOP projected, with NCAR hindcast	Improvement*	Relative Improvement (%)**
2012	239,800	359,300	351,200	8,100	7%
2013	310,100	321,600	371,600	-50,000	-435%
2014	284,600	162,700	206,300	43,600	36%
2015	482,500	259,600	281,500	21,900	10%
2016	412,400	332,900	376,300	43,400	55%
2017	871,700	700,100	748,500	48,400	28%
2018	184,200	141,500	159,600	18,100	42%
2019	914,200	817,500	867,300	49,800	51%
2020	182,900	233,400	207,000	26,400	52%
2021	256,600	232,900	234,800	1,900	8%

* Improvement calculated as (AOP_NCARE - AOP_original) - (AOP_original - Actual)

** Relative improvement calculated as: (Improvement)/(AOP_original - Actual)

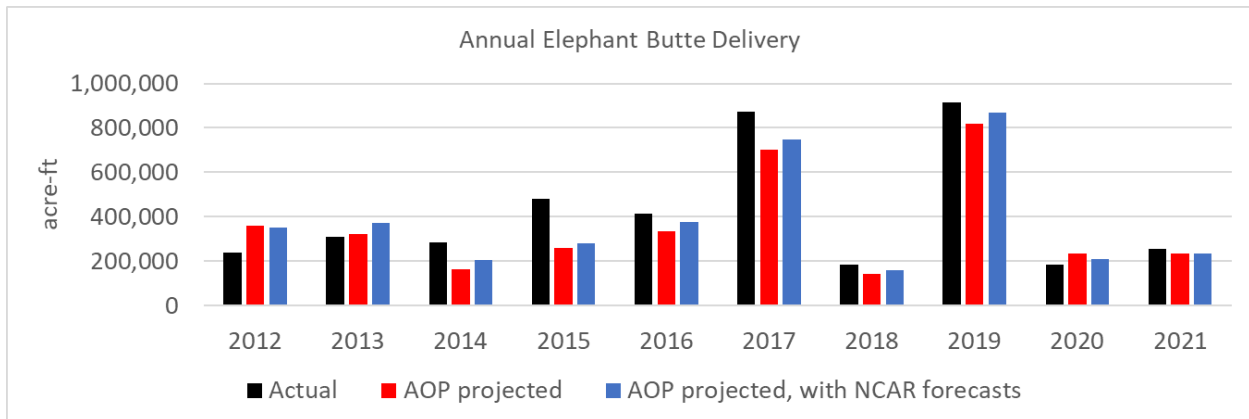


Figure 2. Results from implementing NCAR's hydrologic category hindcast in URGWOM.

Appendix D – Component 4 Summary

Predicting the Summer Rains of the Southwestern United States Using Machine Learning

Janelle McManaman^{1,2}, Erin Towler³, James Done³

¹ University Corporation for Atmospheric Research (UCAR), Boulder, CO

² Central Michigan University, Mount Pleasant, MI

³ National Center for Atmospheric Research (NCAR), Boulder, CO

Derived from the 2023 Final Report for the Significant Opportunities in Atmospheric Research and Science (SOARS) summer undergraduate research program at UCAR

Abstract

The summer rains in the Southwest United States (US) are a vital water source, but these seasonal rains have long been challenging to forecast. This paper explores whether tree-based machine learning techniques, such as Random Forest (RF), can improve rainfall prediction. In this paper, RF and linear regression are used to predict the summer rains for the four regions in the US Southwest defined in Prein et.al., 2022 (Appendix A): Arizona East, Arizona West, New Mexico North, and New Mexico South.

Regardless of technique, summer rainfall in Arizona East is the most predictable of all the regions. For all regions, the RF predictions show better performance compared to the single-variable linear regressions. Despite slight differences, both approaches agree that the water vapor mixing ratio at 850 mb (Q850) is, of 14 tested variables, the most important variable to consider when predicting rainfall. Moreover, both agree that the atmospheric variables are more relevant than the oceanic variables. While there is still more research to be done on this, preliminary results from the tree-based methods are encouraging and their potential for real-time forecast applications should be explored.

1. Review of the Literature

In the 1997 paper titled “The North American Monsoon” by Adams and Comrie, they define what the North American Monsoon is and the features of it. The North American Monsoon (NAM) is a seasonal reversal of winds that occurs from June to October in the southwest United States and northwest Mexico. The monsoon wind transports moisture onto land. Upper level moisture comes from the Gulf of Mexico and the lower level moisture comes from the Gulf of California. Both are needed for the transportation of water vapor in the atmosphere, but the Gulf of California has a stronger influence over the monsoon rains than the Gulf of Mexico. For these summer monsoons, there needs to be sufficient atmospheric moisture and warm land where the land is warmer than the water. This causes the cool moist air from the ocean to move over the dry warm land. This flow of moisture is driven by intense surface heating of the land and an associated thermal low to the southwest. When the air goes over the elevated surface of

the Southwest United States it cools orographically and it rains. A mid-level ridge of high pressure causes a moisture flow from the southeast, and the ridge also brings warmer temperatures and high pressure. Thermal and pressure contrasts between the cooler air mass and intensely heated surface in the desert amplify the moisture surge that happens. There is a surge of moist tropical Pacific air that is brought up the Gulf of California as a result of the low-level atmospheric pressure gradient. A northward influx of low level maritime tropical air into the desert and uplands promotes convective activity. Persistent low level jet flow means significant northerly transport of water vapor from the Gulf of California (Adams and Comrie 1997).

In the paper, “Sub-Seasonal Predictability of North American Monsoon Precipitation” by Prein et al. (2022), they show that weather types can be used to predict NAM precipitation (see Appendix A). To further their research, this paper explores the value added from tree-based machine learning approaches, for predicting monsoon precipitation.

2. Methods

Tree-based machine learning approaches from James (2021) are used to see if it can improve forecasting the summer rains of the Southwest United States. We perform our analysis in the coding language R and use the data and regions from Prein et al. (2022). The variables studied were: average precipitation, sum of precipitation, sea level pressure, u component of the wind at 850 mb, v component of the wind at 850 mb, moisture flux at 850 mb and 500 mb, water vapor mixing ratio at 850 mb and 500 mb, geopotential height at 500 mb, wind speed at 200 mb, dry air temperature at 850 mb and 500 mb, ENSO anomaly, ENSO value, and PDO. To get this information, Prein et al. (2022) used the fifth generation for atmospheric analysis of climate (ERA5) for the atmospheric data, PRISM data for precipitation observations, and data from NOAA’s Physical Science Laboratory (PSL) for ocean variables. These data are monthly for June through October from 1981 to 2021 for four locations: Arizona (AZ) East, AZ West, New Mexico (NM) North, and NM South.

Linear regression and tree-based machine learning methods are used. For linear regression, only one variable was used at a time, and the relationship between each variable and average precipitation is explored and used for prediction. Tree-based methods, including Random Forest (RF), were used to identify variables of greatest importance and to make predictions. RF was conducted using the “randomForest” package in R. For prediction, both the linear regression and RF models are trained on the same half of the data (training data) and predictions are evaluated using the remaining data (test data).

3. Results and Discussion

First, the linear association between each variable and average precipitation was assessed using the full dataset. For all four locations, the variable with the strongest association with average precipitation, as measured by the adjusted R^2 value, is water vapor mixing ratio at 850 mb (Q850). The scatterplots for Q850 versus average precipitation is shown in Figure 1. AZ East has the largest adjusted R^2 , with the value of 0.73 (Table 1). Of the four locations, NM South had the lowest adjusted R^2 value (0.47; Table 1), yet that is the highest adjusted R^2 value out of all the variables for that location. In general, atmospheric variables tended to have stronger associations with the average rainfall than the oceanic variables (results not shown).

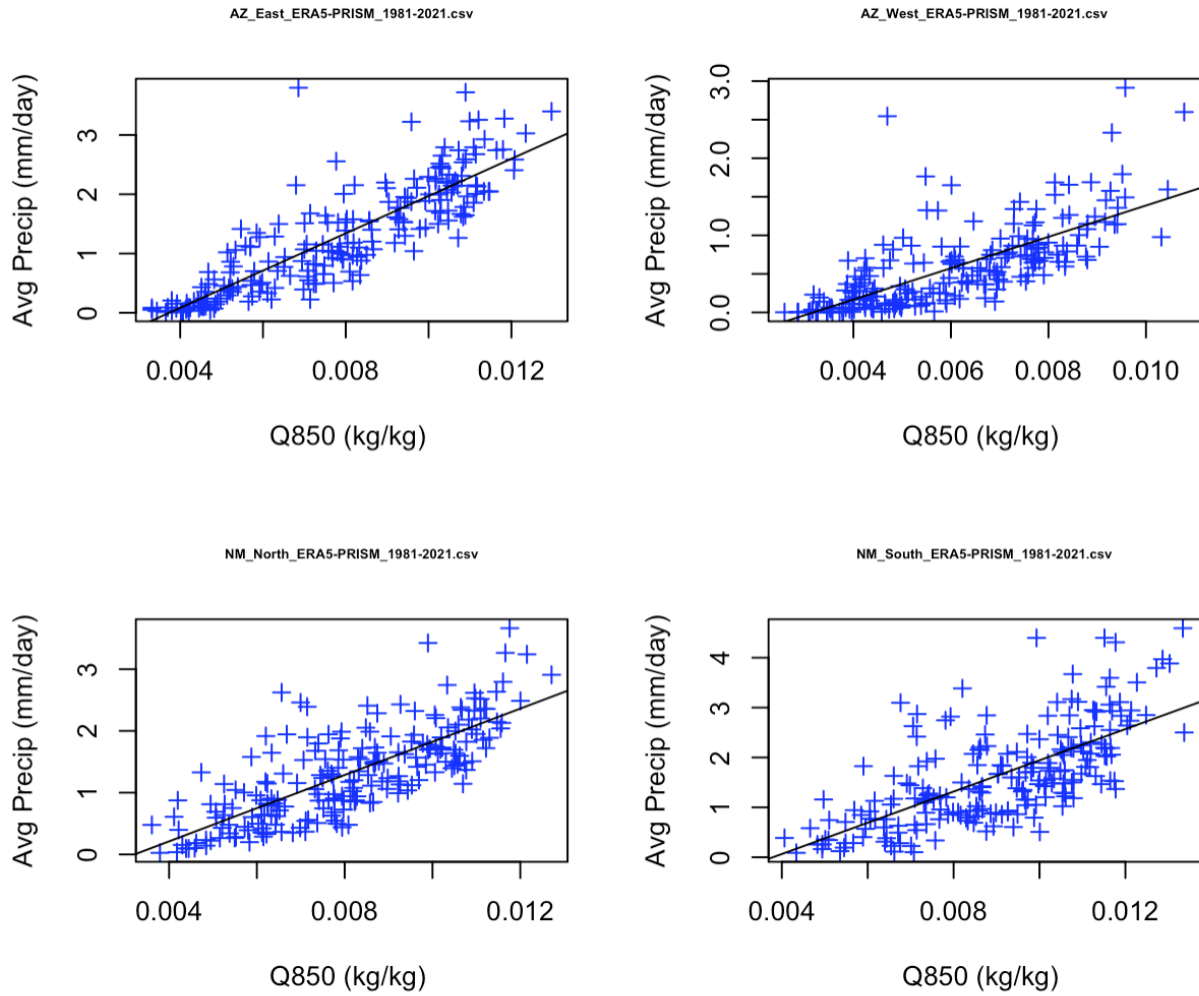


Figure 1. Water vapor mixing ratio (kg/kg) at 850 mb for AZ East (top left), AZ West (top right), NM North (bottom left), and NM South (bottom right). All have strong linear relationships between Q850 and average precipitation (mm).

Table 1. Adjusted R² for Q850 using linear regression on the full data and then a random half of the data for all four locations

Q850	R ² Adj- Full Data	R ² Adj- Half Data
AZ East	0.73	0.70
AZ West	0.51	0.44
NM North	0.58	0.52
NM South	0.47	0.48

Best number is in bold

Next, we explore the tree-based methods. Figure 2 shows an example of a complex statistical tree produced for NM North; “complex” implies that the tree can use all of the available predictors (rather than just a subset). Figure 2 shows an example of a tree produced to predict if the summer rainfall in NM North would be above the 75th percentile (“Yes”) or not (“No”). Although it is only a single tree, and not an ensemble of trees, it is useful to examine individual trees for interpretation. First, Q850 is the top node, thereby indicating its high importance. For NM North, a Q850 value of 0.01077 (Figure 2) is a critical threshold for determining if the precipitation will be above the 75% percentile. We also explored pruned trees, whereby only a subset of predictors are used, but results from that exercise are not shown.

The next step is to produce multiple trees to reach a prediction consensus. In this case, we use RF, and we allow the RF model to include all of the available predictors; this is also referred to as bagging.

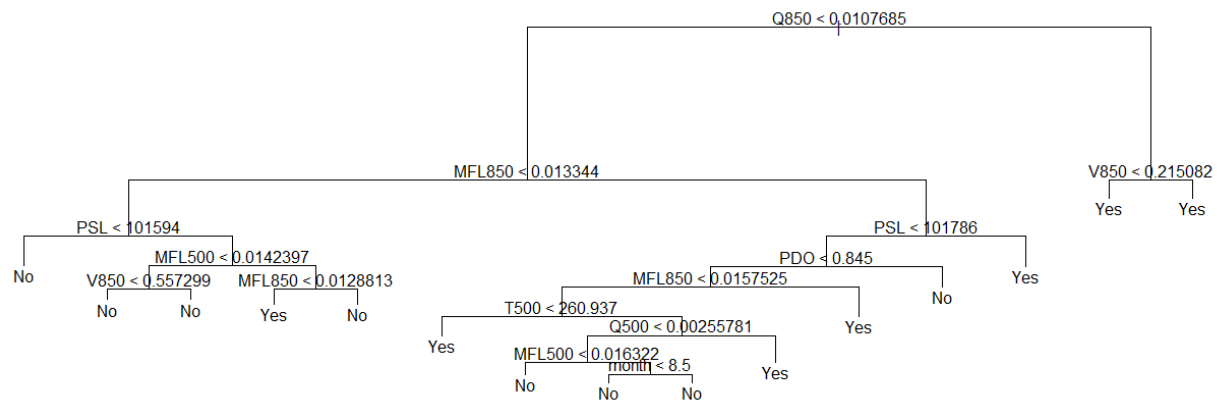


Figure 2. Complex statistical tree for NM North. “Yes” indicates precipitation value above the 75% quartile and “No” indicates below that. On each node is the variable name and the threshold value. A single tree, as shown here, is useful for interpretation, but an ensemble of multiple trees is better for prediction.

RF is used to predict the average summer rainfall. Figure 3 shows the increase of mean squared error if a certain variable were to be taken out. For NM North and NM South, Q850 is the most significant variable when predicting the summer rains, because without it, the error would increase dramatically. Similar results were found for AZ West and AZ East (results not shown). Further, ocean variables show less importance than atmospheric variables. Using the test data, scatterplots of the RF prediction versus what was observed can be seen in Figure 4 for NM North and South; a perfect prediction would lie on the one-to-one diagonal line.

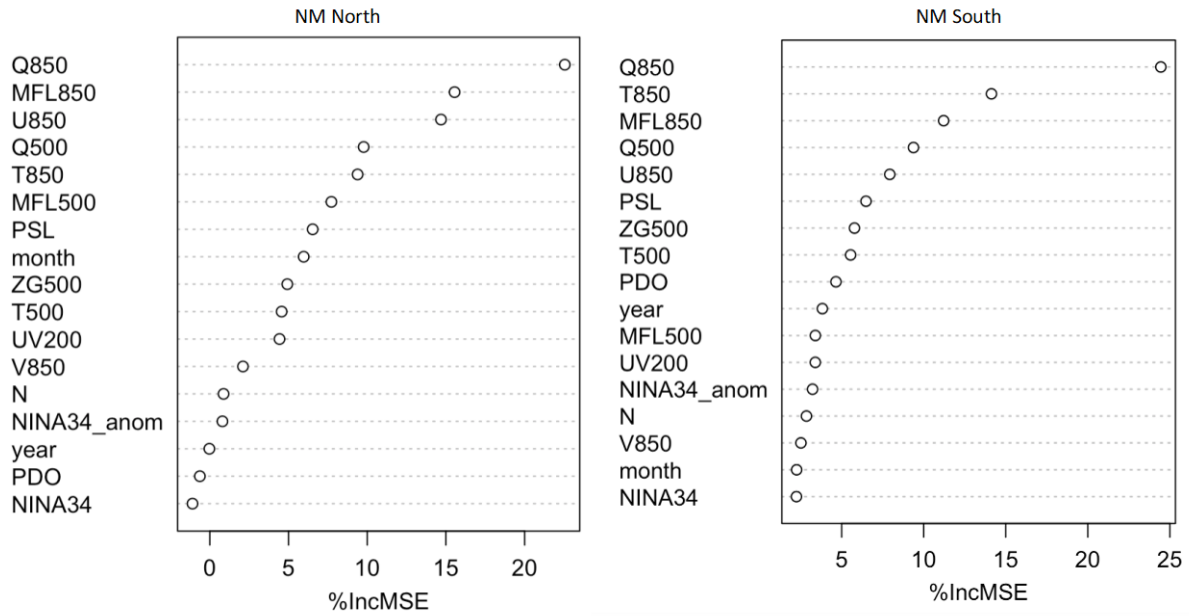


Figure 3. Variable importance for NM North (left) and NM South (right). Plots shows the increase of mean squared error if the variable is not included in the prediction of the rain.

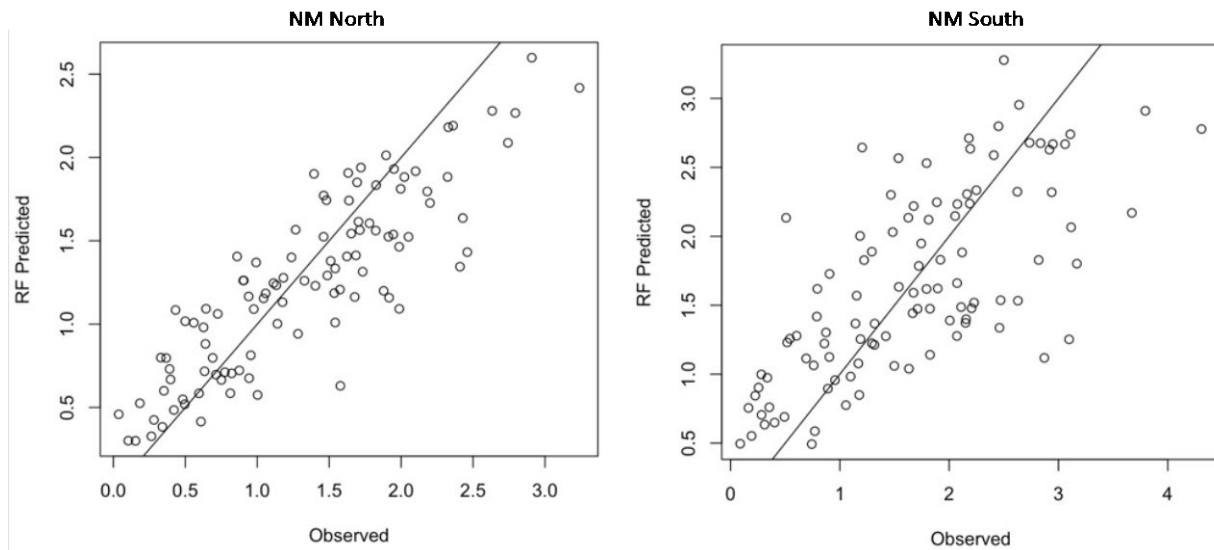


Figure 4. Using the test data, the observed versus what was predicted by the random forest for NM North (left) and NM South (right); a perfect prediction would lie on the one-to-one diagonal line.

Finally, we compare predictions from the linear regression and Random Forest using the test data. In this comparison, the prediction is from the linear regression model using Q850 as the predictor, whereas the RF model is able to use all of the predictors. The test data is used to compare the predictions using R^2 and RMSE. The highest R^2 value is 0.83 (Table 2) at AZ East from Random Forest. The lowest RMSE value is 0.23 (Table 2) at AZ West from Random Forest. Random forest is better than linear regression, however, not by much for some locations.

Table 2. R2 and root mean squared error from linear regression and random forest using the test data

	R2 Linear Regression	R2 Random Forest	RMSE	
			Linear Regression	Random Forest
AZ East	0.80	0.83	0.42	0.40
AZ West	0.58	0.77	0.31	0.23
NM North	0.62	0.67	0.48	0.46
NM South	0.55	0.64	0.71	0.68

Best numbers are in bold

Maps of summertime Q850 shown in Figure 5 provide context for the statistical results, suggesting why the outside locations of the two states are harder to predict than the more inner locations of the two states. There is a greater change and a steeper gradient in Q850 in the outer locations than in the inner locations making it harder to predict with more extreme values and less linearity.

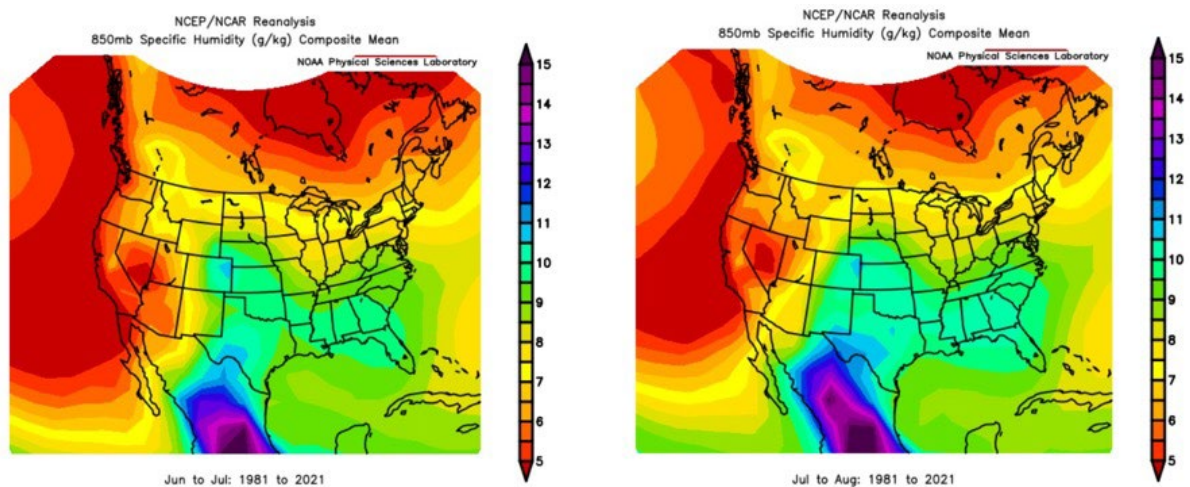


Figure 5. Water vapor mixing ratio at 850 mb (g/kg) from 1981 to 2021 in June and July (left panel) and July and August (right panel). The better predictability for the east and northern sides of the two states may be because the gradient shift is less intense than it is for the outside locations of the two states.

4. Conclusions

In summary, out of the four locations, summer rains for AZ East are the most predictable using RF, while NM South is the least predictable. RF shows better performance than linear regression in this analysis, but we only examined a linear regression model with a single predictor (Q850). Q850 is very important for predicting the summer rains, showing high correlation to the average amount of precipitation. Oceanic variables have less importance than atmospheric variables. Several recommendations for future work include:

- Compare predictions from multivariate linear regression
- Test other statistical learning approaches, e.g., develop random forest models that can only use a subset of the predictors.
- Explore possible physical reasons for the threshold values from the statistical trees.
- Identify other explanatory predictors (e.g., moisture from the Gulf of Mexico)

References

- Adams, D.K. and Comrie, A.C. 1997. “The North American monsoon,” *Bulletin of the American Meteorological Society*, 78(10): 2197-2213.
- James, G., Witten, D., Hastie, T., Tibshirani, R. 2021. “An Introduction to Statistical Learning with Applications in R”, 2nd ed, Springer, New York, available at: <https://www.statlearning.com/>
- Prein, A.F., Towler, E., Ge, M., Llewellyn, D, Baker, S., Tighi, S., and Barrett L. 2022. “Sub-Seasonal Predictability of North American Monsoon Precipitation”, *Geophysical Research Letters*, 49, e2020GL095602, <https://doi.org/10.1029/2021GL095602>.

Appendix E – Research Products

This Appendix provides a complete list of the research outcomes, including the papers, presentations, products, and outreach.

Papers & Presentations

Prein AF, 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>. [Published paper]

Towler E, Llewellyn D, Prein AF, Barrett L, (2023) Seasonal forecasting of monsoon precipitation characteristics using Weather Types and Generalized Linear Modeling. Proceedings of the Federal Interagency Sedimentation and Hydrologic Modeling Conference (SEDHYD), St. Louis, MO. [Published proceeding and presentation]

Towler E, Llewellyn, D, Prein AF, Mander N, Baker S, Barrett L, (2022), “An experimental monsoon forecast for water management”, Bureau of Reclamation Water Operations and Planning Seminar (Virtual), Nov 10, 2022. [Presentation]

Towler E, Baker S, Barrett L, Ge M, Prein A, Llewellyn D, Tighi S (2023) “An experimental monsoon forecast for water management”, AMS Annual Meeting, Denver, CO, Jan 10, 2023. [Presentation].

Products

Prein AF, Towler E (2023) “Experimental Monsoon Forecast”, Google Colab Notebook available at: <https://colab.research.google.com/drive/1O0LdKKvKf6yBO-AABMddkmNtprtm-6k8?usp=sharing>

Outreach

McManaman J, Towler E, Done J (2023) “Predicting the Summer Rains of the Southwest United States Using Machine Learning”, SOARS. [Final Report for the Significant Opportunities in Atmospheric Research and Science (SOARS) summer undergraduate research program at UCAR, with mentee McManaman (UCAR and Central Michigan University) and research mentors Erin Towler and James Done (NCAR)]

