UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INVESTIGATING THE EFFECTIVENESS OF SUBSEASONAL TO SEASONAL (S2S) PRECIPITATION FORECASTS IN STREAMFLOW PREDICTION AT A LARGE NUMBER OF WATERSHEDS ACROSS CONTIGUOUS UNITED STATES (CONUS)

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

AAYUSH ADHIKARI
Norman, Oklahoma
2024

INVESTIGATING THE EFFECTIVENESS OF SUBSEASONAL TO SEASONAL (S2S) PRECIPITATION FORECASTS IN STREAMFLOW PREDICTION AT A LARGE NUMBER OF WATERSHEDS ACROSS CONTIGUOUS UNITED STATES (CONUS)

A THESIS APPROVED FOR THE
SCHOOL OF CIVIL ENGINEERING AND ENVIRONMENTAL SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Tiantian Yang, Chair

Dr. Yang Hong

Dr. Kendra Dresback

Dedication

Dedicated to my family for their unwavering love and support, my friends back in Nepal and Oklahoma for their encouragement, and my mentors who have guided and inspired me throughout this journey.

Acknowledgments

I would like to express my heartfelt gratitude to my advisor, Dr. Tiantian Yang, for his unwavering support, encouragement, and expertise throughout this journey. His guidance has been the cornerstone of my growth and accomplishments. I am deeply thankful to my committee members, Dr. Yang Hong and Dr. Kendra Dresback, for their time and invaluable insights, which have enriched my work significantly.

A special note of appreciation goes to Dr. Lujun Zhang, whose close mentorship has been a blessing; his patience, expertise, and commitment to my progress have been pivotal. I feel truly fortunate for the many opportunities he provided and for the hours he spent offering detailed feedback and answering my questions.

I also extend my sincere thanks to Dr. Yang's research group, especially Dilip Neupane, for their support. Further, I would also like to thank the University of Oklahoma for providing an excellent learning and growth environment. Additionally, to my wonderful friends here in Norman—thank you for being an incredible and cherished part of my life, filling my journey with unforgettable memories and warmth.

Finally, I want to express my gratitude to all the funding sources that supported the research for this thesis. The research works for this thesis is supported by the National Science Foundation (NSF) CAREER Award (No. 2236926). This work is also supported by the U.S. Bureau of Reclamation (USBR) Project No. R24AC00032, Department of Defense, Army Corps of Engineers (DOD-COR) Engineering With Nature (EWN) Program (Award No. W912HZ-21-2-0038), and the NSF Grant No. OIA-1946093 and its subaward No. EPSCoR-2020-3.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BCSD = Bias Correction and Statistical Downscaling

BoM = Bureau of Meteorology

CAMELS = Catchment Attributes and Meteorology for Large-Sample Studies

CanCM3 = Canadian Coupled Climate Model versions 3

CanCM4 = Canadian Coupled Climate Model versions 4

CCSM4 = Community Climate System Model version 4

CFSv2 = Climate Forecast System version 2

CLT = Critical Lead Time

CONUS = Contiguous United States

ENSO = El Niño-Southern Oscillation

ESP = Ensemble Streamflow Prediction

EFAS = European Flood Awareness System

GCMs = General Circulation Models

GEOS5 = Goddard Earth Observing System model version 5

HEFS = Hydrologic Ensemble Forecast System

IHCs = Initial Hydrological Conditions

IUH = Instantaneous Unit Hygrograph

IQR = Inter-Quartile Range

KGE = Kling Gupta Efficiency

MJO = Madden-Julian Oscillation

NARR = North American Regional Reanalysis

NMME-2 = North American Multi-Model Ensemble Phase-2

NMME-ESP = North American Multi-Model Ensemble Phase -2 driven ESP

NWP = Numerical Weather Prediction

PET = Potential Evapotranspiration

PRISM = Parameter-elevation Regressions on the Independent Slopes Model

QBO = Quasi-Biennial Oscillation

RevESP = Reverse Ensemble Streamflow Prediction

S2S = Subseasonal to Seasonal

SAC-SMA = Sacramento Soil Moisture Accounting

SC-SAHEL = Shuffled Complex-Self Adaptive Hybrid Evolution

SCE-UA = Shuffled Complex Evolution

SST = Sea Surface Temperature

Abstract

Accurate streamflow prediction at the Subseasonal-to-Seasonal timescale (S2S, i.e., up to 30 days into the future) could greatly benefit water resources planning, flood control, and the management of water infrastructures. Correct estimation of initial hydrological conditions (IHCs) and accurate precipitation forecasts are the two key factors leading to skillful S2S streamflow prediction. Ensemble Streamflow Prediction (ESP) is an operational approach for S2S streamflow predictions. However, ESP has been critiqued by many studies for its use of randomly resampled precipitation as conceptual future forecasts. The S2S precipitation forecasts from the North American Multi-Model Ensemble Phase II (NMME-2) can be an alternative to the randomly resampled precipitation for hydrological forecasting in the ESP framework. But, the effectiveness of the S2S precipitation forecasts from NMME-2 has not been fully understood in terms of its capability of streamflow prediction. In this thesis, the predictive skill of ESP driven by the S2S precipitation forecasts from NMME-2 is evaluated against conventional ESP driven by randomly resampled precipitation at 121 watersheds across the Contiguous United States (CONUS). Additionally, the relative importance of the NMME-2 precipitation forecasts and IHCs in S2S streamflow prediction at different forecast lead times is investigated using the ESP/Reverse Ensemble Streamflow Prediction (RevESP) framework. The results indicate that the integration of the S2S precipitation forecasts from NMME-2 in the ESP framework will lead to more skillful streamflow predictions across CONUS. Furthermore, the results indicate that the S2S precipitation forecasts are relatively more important and informative than the IHCs in streamflow predictions at most of the watersheds, especially at longer forecast lead times. However, certain hydrological properties, such as snow and large drainage areas, may weaken the effectiveness of available S2S

precipitation forecasts. In general, this study highlights the broader application of available S2S precipitation forecasts in hydrology, which is likely to benefit a wide range of forecast end-users.

1. Introduction

Accurate streamflow prediction at Subseasonal to Seasonal (S2S) timescale, which is defined as 10 to 30 days into the future (Vitart et al., 2017), could provide essential information for various human and socio-economic activities. These activities include water resource management, flood forecasting, reservoir management, irrigation planning, and several others (Anghileri et al., 2019; Roy et al., 2022; White et al., 2015).

The future precipitation scenarios, along with the initial hydrological conditions (IHCs) at the "current time," are the two major factors affecting future streamflow volumes at the S2S timescale (Arnal et al., 2018; Shukla & Lettenmaier, 2011). The precise estimation of IHCs would allow the lagged relationship between the watershed's past conditions and future streamflow to be carefully considered and hence is important to derive compelling streamflow prediction at S2S timescale (Mahanama et al., 2012; Wood & Lettenmaier, 2008). Additionally, since future precipitation events would enable runoff generation through various hydrological processes, accurate precipitation forecasts would also greatly benefit S2S streamflow prediction (Duan et al., 2019; Shrestha et al., 2013).

The Ensemble Streamflow Prediction (ESP) framework is commonly used for streamflow prediction at the S2S timescale (Day, 1985; Schaake & Larson, 1998). Under the ESP framework, the IHCs and future precipitation scenarios are jointly considered for probabilistic streamflow predictions (Troin et al., 2021). Specifically, when applying conventional ESP, historical precipitation measurements are randomly selected (i.e., resampled precipitation forecasts) and jointly employed with the estimated IHCs to produce an ensemble of streamflow predictions.

The interactions between IHCs and future precipitation scenarios in ESP comprise complex dynamics. When performing streamflow predictions under the ESP framework, the effect coming from the estimated IHCs diminishes over forecast iterations (i.e., the increase of forecast lead time), and the “randomly resampled precipitation” eventually dominates the streamflow predictions at an extended timescale (Li et al., 2009; Paiva et al., 2012). Most commonly, the predictability of streamflow from IHCs only sustains at shorter forecast lead times (Cao et al., 2021; Li et al., 2009; Shukla & Lettenmaier, 2011). Nevertheless, it has also been reported that in some locations, the IHCs’ effect could last up to 6 months into the future if complex hydrometeorological conditions are present (Mahanama et al., 2012; Wood & Lettenmaier, 2008).

Despite the potentially varying effect, IHCs are often considered the major predictability source of streamflow under the conventional ESP framework as the “randomly resampled precipitation” is not skillful in general. This is because “randomly resampled precipitation” does not consider subsequent future weather conditions, which are important for the reliable representation of future precipitation scenarios (Farmer & Vogel, 2016; Zhang et al., 2024). As a result, the overall predictive skill of ESP is compromised, especially at longer lead times. To address this issue, more advanced dynamical forecast products can provide a reliable representation of future precipitation scenarios (or ensemble) at the S2S timescale, potentially improving the streamflow predictions' predictive skills under the ESP framework. Dynamical S2S precipitation forecast products are promising alternatives to the “resampled precipitation forecasts” in ESP to generate more accurate and reliable streamflow predictions. Available S2S precipitation forecast products are generated by coupled General Circulation Models (GCMs).

In contrast to randomly resampling historical precipitation, GCMs incorporate various earth system physical processes when generating precipitation forecasts (Duan et al., 2019). In addition, GCM-generated precipitation forecasts are produced with a consideration of the “real-time” state of the Earth System (Robertson & Vitart, 2018). Therefore, by theory, available S2S precipitation forecast products should perform better than the “resampled precipitation forecasts” used in the current ESP framework.

However, accurately forecasting precipitation at the S2S timescale is notoriously challenging. Due to the unique meteorological and climatological physics at the transitional period from the weather to climate scales, S2S forecasts are often considered a “desert of predictability.” As illustrated in Figure 1.1, the S2S timescale has notably lower forecast skill than weather and climate outlooks, generally falling within the 'Poor' to 'Zero' categories. As a result, several studies have reported that available S2S precipitation forecasts have unsatisfactory performance over the Contiguous United States (CONUS) (Levey & Sankarasubramanian, 2024; Tian et al., 2017; Zhang et al., 2021).

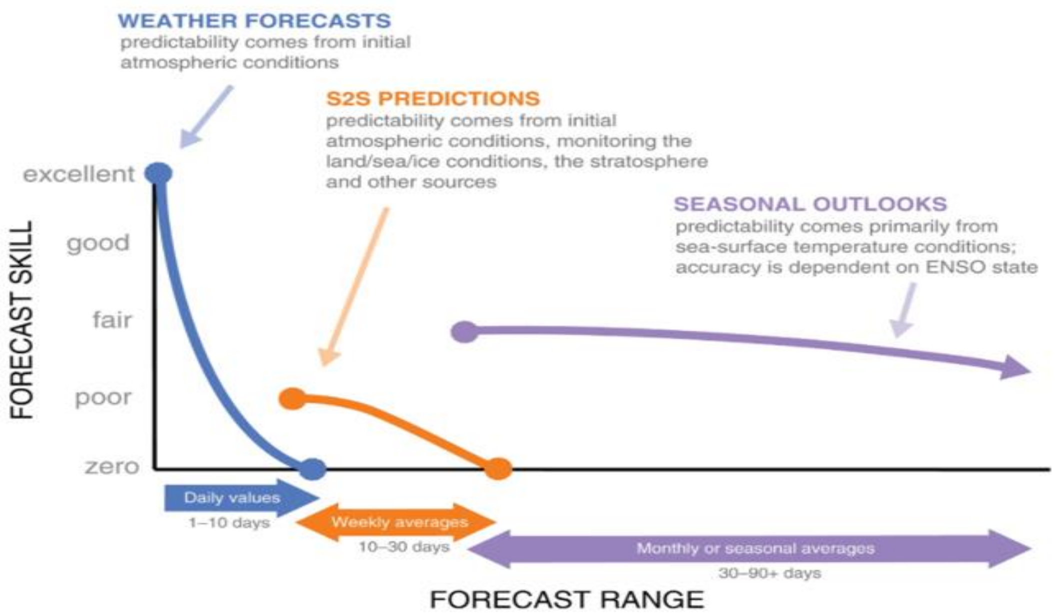


Figure 1.1: Schematic diagram showing the forecast skill and potential sources of predictability at different timescales. In particular, the predictive skill is extremely limited during the transforming S2S timescale. (White et al. 2017)

The unsatisfactory performance of available S2S precipitation forecasts raises doubts about their effectiveness in assisting streamflow predictions. A few pioneering studies have shown that S2S precipitation forecast-driven ESP provides more skillful streamflow predictions than conventional ESP. However, the added advantage offered by S2S precipitation forecasts is often marginal (Baker et al., 2019; Cao et al., 2021; Zhang et al., 2024). One major limitation of these studies is that they focus on a few watersheds in specific regions, limiting their scope. Consequently, a comprehensive examination of S2S precipitation forecasts for streamflow prediction at a broader scale remains largely unexplored.

In addition, the effectiveness of available S2S precipitation forecasts should be investigated at different forecast lead times across watersheds in different geospatial locations. Given the generally short-living impact of IHCs (Cao et al., 2021; Zhang et al., 2021), the S2S precipitation forecast is expected to be the dominating predictability source of streamflow at the extended forecast lead times. However, the exact forecast lead time at which S2S precipitation forecasts overtake IHCs as the dominant source of streamflow predictability remains unclear. This question is particularly important, as certain hydrological characteristics can extend the impact of IHCs on future streamflow for up to six months, thereby limiting the effectiveness of S2S precipitation forecasts (Mendoza et al., 2017; Shukla et al., 2013; Wood & Lettenmaier, 2008). As a result, it is reasonable to expect that the transition point from IHCs to S2S precipitation in dominating streamflow differs across watersheds within the CONUS due to varying hydrological properties. Unfortunately, such detailed information is currently unavailable from existing literature, underscoring the need for further studies to enhance streamflow predictions at the S2S timescale.

Therefore, the goal of this thesis is to comprehensively examine and validate the effectiveness of available S2S precipitation forecasts in streamflow prediction under the ESP framework at a larger scale across the CONUS. To evaluate the effectiveness, the predictive skill of S2S precipitation forecast-driven ESP is evaluated against the predictive skill of randomly resampled precipitation-driven ESP at a total of 121 watersheds across the CONUS. Furthermore, this thesis seeks to investigate the relative importance of IHCs and S2S precipitation forecasts in contributing to the predictive skill of streamflow at different forecast lead times and across different regions. To investigate the relative importance of S2S precipitation forecasts with IHCs, the S2S precipitation forecast-driven ESP is jointly considered with Reverse Ensemble Streamflow prediction (RevESP) introduced by Wood & Lettenmaier (2008) across the study watersheds.

2. Literature review

2.1 Streamflow Predictions at Subseasonal to Seasonal (S2S) timescale

The streamflow prediction at the S2S timescale is crucial for proactive disaster mitigation, such as optimizing reservoir operations for flood control. Accurate S2S streamflow predictions can provide early insights into incoming water volumes, enabling optimization of the allocation of water uses such as environmental flows, irrigation, flood management, and hydroelectric power generation (Chiew et al., 2003; Liu et al., 2022; Troin et al., 2021).

One of the major predictability sources of S2S streamflow is IHCs. IHCs are defined as the hydrological conditions of watersheds, mainly soil moisture and snow conditions, at the time of issuing a streamflow forecast (i.e., streamflow forecast initialization time). Soil moisture significantly influences streamflow generation by partitioning water storage into infiltration and surface runoff. Specifically, when the soil is saturated or near saturation, its ability to absorb water diminishes, leading to increased surface runoff volumes. Conversely, when the soil is dry, it absorbs more water, which reduces immediate runoff (Li et al., 2009). Similarily, snow conditions are also critical for accurate streamflow predictions. In snow-dominated watersheds, accumulated snow during the past winter melts in the coming warm seasons of spring and summer for streamflow generation (Berghuijs et al., 2014; Koster et al., 2010). For some geospatial locations, such as the mountain ranges in the Western Rockies, Sierra Nevada, and Cascades in the United States, up to 78% of the streamflow originates from snowmelt (Li et al., 2017).

Most commonly, the predictability of streamflow from initial hydrologic conditions (IHCs) often declines rapidly as forecast lead time increases (Cao et al., 2021; Zhang et al., 2024). Consequently, watershed runoff becomes increasingly influenced by variations in excess rainfall, which simultaneously alter the IHC over time. However, several studies have reported that the predictability of streamflow coming from IHCs can be sustained at longer timescales to months. For example, Mahanama et al. (2012) evaluated the impact of IHCs on streamflow predictability across 23 basins in the CONUS. Mahanama et al. (2012) found that accurate estimation of IHCs can extend streamflow predictability for up to six months in certain basins within the western United States. Similarly, Maurer et al. (2004) evaluated the influence of IHCs in streamflow prediction across various regions of North America. Maurer et al. (2004) found that accurate characterization of IHCs can enhance streamflow predictability for up to 4.5 months across various regions, specifically in western CONUS and the Great Lake regions. Similarly, Greuell et al. (2019) evaluated the influence of IHCs on streamflow prediction in snow-dominated and temperate regions in Europe. Greuell et al. (2019) found that the predictability of IHCs in streamflow prediction can be sustained for up to four months in those regions. In other words, the existing studies are still controversial and are actively debating how long the IHC would determine the predictability of streamflows for different watersheds.

Another major predictability source of S2S streamflow is future precipitation scenarios. Future precipitation events can influence runoff generation through various hydrological processes. Specifically, when precipitation occurs, it may generate surface runoff rapidly if it falls over areas with limited infiltration capacity. This immediate runoff contributes directly to streamflow. In contrast, when precipitation falls on areas with high infiltration capacity, the water is more likely to infiltrate into the soil rather than immediately contribute to runoff. This infiltrated water then moves through subsurface pathways and eventually contributes to baseflow, sustaining streamflow over longer periods. Therefore, accurate knowledge of future precipitation scenarios could greatly enhance streamflow prediction (Krishnamurti et al., 2007; Shrestha et al., 2013).

The operational streamflow prediction technique ESP heavily leans on exploiting the streamflow predictability from IHCs, leaving future precipitation conditions less considered (Day, 1985; Schaake & Larson, 1998). However, the predictability of streamflow from IHCs generally cannot be sustained at an extended timescale. As a result, ESP-generated streamflow predictions demonstrate little deterministic skill at longer lead times (Cao et al., 2021; Shukla et al., 2013; Wood & Lettenmaier, 2008; Zhang et al., 2024).

One of the major reasons behind this is that available S2S precipitation forecasts have limited predictive skill after a few days, and performance is inconsistent across different geospatial locations (Cao et al., 2021; Levey & Sankarasubramanian, 2024). Additionally, S2S precipitation forecasts are also criticized for their coarse resolution to be used for hydrological applications (Baker et al., 2019; De Andrade et al., 2019; Zhang et al., 2021). For this reason, the operational forecast centers have made little progress in incorporating the potential predictability of future precipitation scenarios in S2S streamflow prediction (Baker et al., 2019; Li et al., 2009; Wood & Lettenmaier, 2008).

In summary, S2S streamflow prediction is important for proactive water management decisions. Accurate S2S streamflow prediction relies on two main sources of predictability: IHCs and the quality of future precipitation forecasts. However, operational forecast centers have made limited progress in incorporating the potential predictability from future precipitation scenarios due to the limitations of available S2S precipitation forecasts.

2.2 Ensemble Streamflow Prediction (ESP) and Reverse ESP (RevESP)

ESP is an operational streamflow prediction method used at 13 National River Forecast Centers across the United States at the S2S timescale (Prudhomme et al., 2017; Schaake & Larson, 1998). ESP has been widely used in research applications across the globe and has proved to be effective (Harrigan et al., 2018; Kim et al., 2001; Paiva et al., 2012; Singh, 2016; Souza Filho & Lall, 2003).

The framework of ESP allows the joint consideration of both IHCs and future precipitation conditions when predicting streamflow. Specifically, in the conventional ESP, ensembles of precipitation measurements are randomly selected from historical precipitation records (i.e., randomly resampled precipitation) to represent future precipitation scenarios. Theoretically, randomly resampling sufficient historical precipitation records shall lead to reasonable and probabilistically accurate predictions of extreme streamflow values (Troin et al., 2021). This is under the classical "stationary" assumption in hydrology, where it is believed that historical precipitation events will repeat themselves with the same frequency and magnitude in the future (Matalas, 1998; Zhang et al., 2024).

However, the classical "stationary" assumption in hydrology may no longer apply across different geospatial locations, given the evidence of changing climate (Matalas, 1998; Milly et al., 2008; Yang et al., 2021). Additionally, randomly resampled precipitation does not account for the "real-time" state of the Earth System, which influences weather at the S2S timescale (Zhang et al., 2024). Consequently, the deterministic predictive skill of randomly resampled precipitation tends to be low (Farmer & Vogel, 2016; Zhang et al., 2024). As a result, the predictive skill of ESP mainly originates from the estimated IHCs (Cao et al., 2021; Shukla et al., 2013).

Recognizing the limited effectiveness of “randomly resampled precipitation” in ESP, Wood & Lettenmaier (2008) introduced the RevESP to identify the relative importance of IHCs and future precipitation conditions in streamflow predictions. In practice use, ESP method normally accompanies the implementation of RevESP approach to identify the relative importance of IHCs and precipitation forecasts in streamflow prediction at different geospatial locations. For instance, Wood & Lettenmaier (2008) evaluated the relative importance of IHCs and precipitation forecasts in two western United States basins. The evaluation showed that the relative importance of IHCs and precipitation forecasts varies in two western United States basins, with precipitation forecasts being relatively important in streamflow prediction as the influence of IHCs diminishes. Li et al. (2009) used a similar approach to quantify the relative contributions of IHCs and precipitation forecasts in the Ohio River basin in the Southeastern US. Li et al. (2009) found that IHCs are relatively more important at a short lead time (~1 month); however, at longer lead times, precipitation forecasts are the dominant factor in the predictive skill of streamflow. Shukla et al. (2013) applied this approach to evaluate the relative importance of IHCs and precipitation forecasts in the United States and globally. Shukla et al. (2013) reported spatial variability in the relative importance of IHCs and precipitation forecasts with lead time. Shukla et al. (2013) also reported some regions where precipitation forecasts are relatively more important than IHCs at all lead times. Yossef et al. (2013) analyzed multiple large river basins worldwide using ESP/RevESP and concluded that the relative importance of IHCs and precipitation forecasts varies geographically, with precipitation forecasts being relatively more important than IHCs in monsoonal basins.

Previous research using the ESP/RevESP framework highlights precipitation forecasts' crucial role in determining streamflow predictions' accuracy, especially at longer forecast lead times.

Hence, an accurate streamflow prediction is also contingent upon the quality of the precipitation forecasts. However, due to the reliance on inaccurate resampled precipitation forecasts, the overall predictive skill of ESP is compromised (Li et al., 2009; Paiva et al., 2012). To address this issue, there is a need for improved precipitation forecasts under the ESP framework to fully harness the predictability from both IHCs and precipitation forecasts.

To summarize, ESP is an operational method for streamflow prediction at the S2S timescale. However, using randomly resampled precipitation limits the performance of ESP in streamflow prediction. As a result, the predictive skill of ESP-generated streamflow predictions mainly originates from estimated IHCs. Previous research has shown the critical role of precipitation forecasts and IHCs in determining the overall accuracy of streamflow prediction using the ESP/RevESP framework. Hence, it is crucial to incorporate more reliable and accurate precipitation forecasts at the S2S timescale under the ESP framework.

2.3 Dynamical Precipitation Forecasts

Dynamical precipitation forecasts are obtained from forecasting models that leverage mathematical formulations to capture the processes constituting the Earth's geophysical system. By assimilating the “real-time” state of the Earth system and simulating interactions between different components of the Earth, dynamical forecasting models predict and project future precipitation volumes and patterns at different timescales (Christensen & Berner, 2019; Zhang et al., 2024).

The precipitation forecast products from the dynamical models can be categorized as short-, medium-, and long-range forecasts based on available lead times. In the short/medium range (hours to 3 days ahead), Numerical Weather Prediction (NWP) models are widely used to forecast precipitation (Bauer et al., 2015a). NWP models utilize the predictability of precipitation from initial atmospheric conditions to forecast precipitation (Bauer et al., 2015b; Shrestha et al., 2013) with considerable accuracy for up to 10 days, as shown in Figure 1. The precipitation forecasts from NWP have been used for many years, supporting activities ranging from water resources management to emergency planning (Gustafsson et al., 2018; Kalnay et al., 1998). In the long ranges (30+ days ahead), the predictability of precipitation is influenced by slower-changing ocean dynamics like Sea Surface Temperature (SST), as shown in Figure 1. To consider the effect of oceanic dynamics, GCMs coupled with dynamic oceanic and land surface components are widely used to forecast precipitation over longer periods (Robertson et al., 2015). The skill of precipitation forecasts obtained from coupled GCMs has been previously applied in agriculture, irrigation planning, and water resources management (Konzmann et al., 2013; Kundzewicz & Stakhiv, 2010; Remesan et al., 2014).

Consequently, a forecast gap lies between medium-range weather forecasts and the longer-range climate forecasts that cover 10-30 days into the future (White et al., 2017). This gap underscores a transitional period during which neither NWP nor the coupling of GCMs can achieve skillful precipitation forecasts. This is mainly because, at the S2S timescale, the lead time is too long, so most of the predictability of atmospheric initial conditions is already lost. Furthermore, the lead time is too short for the slower-changing dynamics of the ocean to influence S2S weather (Robertson & Vitart, 2018; Vitart et al., 2017). This unique characteristic makes forecasting at the S2S timescale very challenging. As a result, S2S precipitation forecasts are considered a "desert of predictability" as shown in Figure 1, since the skill of S2S precipitation forecasts is very poor compared to forecasts at other timescales (White et al., 2017).

To fill the gap between medium-range and long-range precipitation forecasts, studies have attempted to identify the additional predictability of forecasts at the S2S timescale. Specifically, planetary-scale oceanic patterns like the Madden-Julian Oscillation (MJO), El Nino-Southern Oscillation (ENSO), and the dynamics between the stratosphere and troposphere linked to the Quasi-Biennial Oscillation (QBO) have been found to influence precipitation patterns at the S2S timescale (Butler et al., 2019; Gerber et al., 2012; Lau & Waliser, 2011). Additionally, slowly varying processes like SST, sea ice, soil moisture, and snowpack conditions have also been shown to influence precipitation patterns at the S2S timescale (Chevallier et al., 2019; Holland et al., 2013; Koster et al., 2010).

To better utilize the present knowledge of the predictability of forecast at the S2S timescale, there have been several global collaborations among governmental agencies and researchers worldwide. Through these collaborative efforts, several precipitation hindcast and real-time forecast products have been made available utilizing coupled GCMs. Some notable S2S forecast products include SubX, The S2S Prediction Project, and North American Multi-Model Ensembles (NMME) (Kirtman et al., 2014; Pegion et al., 2019; Robertson & Vitart, 2018).

NMME is one of the collaborative programs initiated by mission agencies to enhance the accuracy and utility of precipitation forecasts at the S2S timescale. The NMME utilizes multiple coupled GCMs operated by governmental agencies and research institutions. NMME-2 is the second phase of this project, involving updated models and expanded datasets. NMME-2 provides precipitation forecasts and hindcasts with daily temporal resolution, which can be utilized in the hydrological applications of S2S precipitation forecasts (Kirtman et al., 2014; Wanders & Wood, 2016).

However, some recent evaluation studies have reported unsatisfactory performance from precipitation forecasts of NMME-2. For instance, Baker et al. (2019) evaluated biweekly precipitation forecasts from the Climate Forecast System Version2 (CFSv2; a contributing model to NMME-2) across the CONUS and found regional biases and a significant decline in forecast skill after the first two weeks. Guo & Nie (2020) evaluated the precipitation forecasts from CFSv2 over East China and highlighted the limitation of CFSv2 in capturing the variability of extreme weather events. Zhang et al. (2021) evaluated precipitation forecasts from five NMME-2 model members across the CONUS and reported spatial inconsistencies, with a rapid decline in forecast skill rapidly after week 1. Additionally, a global evaluation of precipitation forecasts from four model members of NMME-2 by Wanders & Wood (2016) showed a rapid decrease in forecast skill for all four models, with notable inconsistencies in model performance.

The compromised performance of the NMME-2 precipitation forecasts raises questions about their applicability in hydrological applications, particularly streamflow prediction. Despite this, the effectiveness of NMME-2 in streamflow prediction is largely unexplored. According to my understanding, only one study has investigated the effectiveness of NMME-2 precipitation forecasts in streamflow prediction. Zhang et al. (2024) evaluated the application of S2S precipitation forecasts from NMME-2 in four watersheds in the Southern CONUS under the ESP framework. Zhang et al. (2024) found that NMME-2-driven ESP provides more skillful streamflow predictions than ESP in all four study regions. However, the advantage of NMME-2-driven ESP was marginal, especially at longer forecast lead times. Considering the hydrological properties vary widely across the landscape of the CONUS (Rakovec et al., 2019), it is essential to validate the applicability of NMME-2 precipitation forecasts for streamflow prediction across a large number of watersheds. Furthermore, considering the importance of both IHCs and precipitation forecasts in S2S streamflow prediction, it is necessary to evaluate the relative contributions of NMME-2 precipitation forecasts and IHCs to streamflow predictive skill at different forecast lead times. However, the relative importance of IHCs and NMME-2 precipitation forecasts in S2S streamflow prediction has yet to be explored in the current literature.

To summarize, precipitation forecasts are obtained by using dynamical models that leverage mathematical formulations to capture the processes constituting the Earth's geophysical system. However, precipitation forecasts at the S2S timescale, which fall in the transitional period between medium and long range, are often considered a “desert of predictability”. To advance precipitation forecasting at the S2S timescale, several precipitation forecast products, such as NMME-2, have been made available using coupled GCMs. However, previous evaluations of the NMME-2 precipitation forecasts have shown poor performance, raising questions about their use in streamflow prediction. Additionally, there is a lack of comprehensive analysis of the NMME-2’s applicability and relative importance compared to IHCs at different forecast lead times in streamflow predictions across the CONUS.

3. Goals, Motivation and Hypothesis

It is evident from the literature review that the newly developed S2S precipitation forecasts can be an alternative to the randomly resampled precipitation in ESP for S2S streamflow prediction. However, previous evaluation studies reveal the inconsistent and limited performance of S2S precipitation forecasts, which raises questions about their effectiveness in assisting streamflow prediction. Despite being validated at a few watersheds, the effectiveness of available S2S precipitation forecasts needs to be examined at a large scale across the CONUS. Moreover, the specific forecast lead time at which S2S precipitation forecasts surpass IHCs as the dominant source of streamflow predictability is still unclear. Additionally, the diverse hydrometeorological characteristics of watersheds across CONUS are likely to result in different transition timings from IHC dominance to S2S precipitation forecasts in streamflow predictions; to the best of my knowledge, no available studies systematically investigated the relative importance of S2S precipitation forecasts and IHCs in streamflow prediction at different S2S forecast lead times across a large number of watersheds in the CONUS. Such a knowledge gap prevents the development of forecasting systems for more accurate and reliable S2S streamflow predictions.

This thesis attempts to close the above-mentioned knowledge gap. To do that, I choose to comprehensively evaluate the S2S precipitation forecasts from the NMME-2 in assisting streamflow prediction at a large number of watersheds across the CONUS. The selected watersheds in this thesis span different geospatial locations across the CONUS and represent diverse catchment characteristics, including variations in climate, topography, land cover, hydrological responses, and others. The selection of a large number of study watersheds with diverse conditions ensures a comprehensive assessment of the effectiveness of S2S precipitation forecasts in assisting S2S streamflow prediction. Furthermore, this thesis also investigates the relative importance of S2S precipitation forecasts from NMME-2 and IHCs in S2S streamflow prediction. Specifically, the relative importance of S2S precipitation forecasts and IHCs in contributing to streamflow predictive skills is derived at different S2S forecast lead times and watersheds.

The overarching goal of this thesis is to comprehensively evaluate the effectiveness of NMME-2 S2S precipitation forecasts in assisting streamflow prediction at a large number of watersheds across CONUS and over the S2S forecast horizon. By investigating the above-mentioned research items in this thesis, the following research hypotheses are tested:

1. The integration of S2S precipitation forecasts in ESP leads to overall more skillful streamflow predictions at S2S timescale.
2. The relative effectiveness of S2S precipitation forecasts in improving streamflow predictions varies at different forecast lead times and in different geospatial locations.

4. Data and Study Area

4.1 Data

The hydrometeorological datasets utilized in this thesis include (1) multiple dynamical S2S precipitation forecast products from the NMME-2; (2) gridded precipitation measurements from Parameter-elevation Regressions on the Independent Slopes Model (PRISM); (3) Reanalysis of Potential Evapotranspiration (PET) from North American Regional Reanalysis (NARR); (4) Snow Water Equivalent (SWE) from National Snow and Ice Data Center (NSIDC) and (5) streamflow measurements from the United States Geological Survey (USGS). For NMME-2, this study adopts the S2S precipitation hindcasts from four of its contributing models of Canadian Coupled Climate Model versions 3 and 4 (CanCM3, CanCM4), the Community Climate System Model version 4 (CCSM4), and the Goddard Earth Observing System Model version 5 (GEOS5). Each selected model has ten realizations (i.e., ensemble members). The above hydrometeorological datasets cover a 30-year period from 01/01/1982 to 12/31/2011. More details and references of the hydrometeorological datasets used in this research are listed in Table 1.

Table 1: A list of hydrometeorological data utilized

Hydrometeorological Datasets	Temporal Resolution	Spatial Resolution	References
CanCM3	Daily	1°×1°	(Merryfield et al., 2013)
CanCM4	Daily	1°×1°	(Merryfield et al., 2013)
CCSM4	3-hourly/Daily	1°×1°	(Gent et al., 2011)
GEOS5	Daily	1°×1°	(Vernieres et al., 2012)
PRISM	Daily	0.07°	(Daly et al., 1994)

NARR	Daily	0.25°	(Mesinger et al., 2006)
SWE	Daily	0.04º	(Broxton et al., 2019)
USGS streamflow measurements	Daily	N/A	(United States Geological Survey, 1994)

4.2. Study watersheds

In this thesis, a total of 121 study watersheds are selected for a series of streamflow prediction experiments. The 121 study watersheds are selected from the Catchment Attributes and Meteorology for Large-Sample Studies dataset (CAMELS; Newman, Andrew, 2014) across the CONUS. Based on long-term mean areal Snow Water Equivalent (SWE) values, the study watersheds were classified into two groups: snow-dominated (SWE > 25 mm) and non-snow-dominated (SWE < 25 mm). Out of the 121 watersheds, 10 are classified as snow-dominated, while the remaining 111 are non-snow-dominated.

The selected study watersheds cover a diverse range of hydrological and climate conditions. Specifically, the selected watersheds represent a wide range of characteristics, with areas spanning from 5.17×10^8 m² to 1.06×10^{10} m² and mean elevations ranging from 27.37 m to 2268.65 m. Additionally, the selected watersheds exhibit a wide variation in climatic factors, with mean annual precipitation values ranging from 1.07 mm/day to 8.23 mm/day and long-term mean areal snow water equivalent (SWE) values between 0 mm and 387.9 mm.

This thesis also employs region-based analysis by leveraging the climate region definition of the National Centers for Environmental Information (NCEI) (Karl & Koss, 1984). The nine climatic regions separate the CONUS into Northwest, Upper Midwest, Northern Rockies, Northeast, Ohio Valley, West, Southwest, South, and Southeast regions. Figure 4.1 shows the selected geospatial locations of the 121 study watersheds in conjunction with NCEI climatic regions over CONUS.

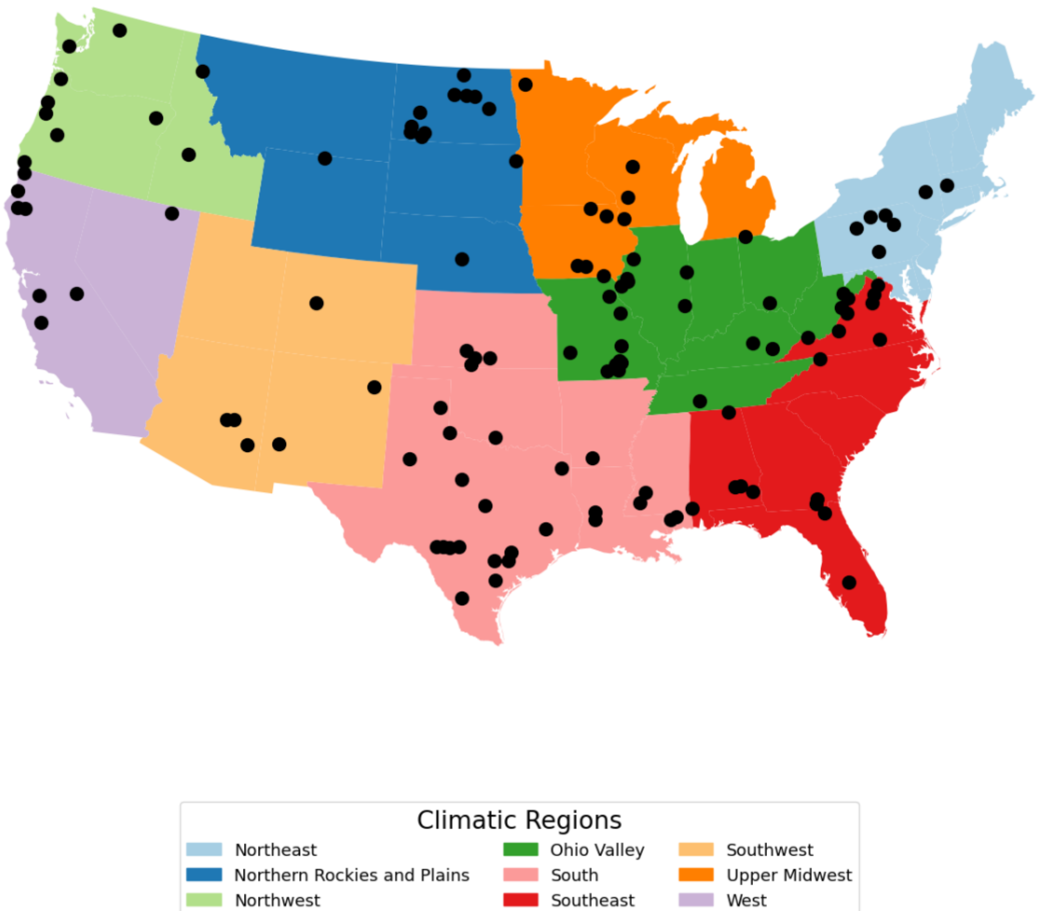


Figure 4.1: Study watersheds across CONUS climate regions

5. Methodology

This thesis adopts a systematic approach to evaluate the effectiveness of S2S precipitation forecasts from NMME-2 in assisting streamflow prediction across study watersheds. First, the hydrological model is calibrated to mimic the hydrological behavior of watersheds as accurately as possible. The calibration ensures the hydrological model delivers reasonable performance for later streamflow prediction experiments. Second, the collected raw precipitation forecast data from NMME-2 are subjected to downscaling and bias correction, given its substantial bias and coarse spatial resolution (Zhang et al., 2021). Specifically, the collected S2S precipitation forecasts from NMME-2 are re-gridded to a finer spatial resolution of 0.25° x 0.25° given its coarse native resolution of 1° x 1° with respect to our study domain. This is followed by bias correction, where the re-gridded NMME-2 precipitation forecasts are adjusted in reference to the observed precipitation dataset from PRISM using the Multi-Segment Quantile Mapping (MSQM) approach. The re-gridding and bias correction prepare the collected NMME-2 precipitation forecasts for later ESP and RevESP experiments.

To test hypothesis 1, the bias-corrected NMME-2 precipitation forecasts are applied to force the calibrated hydrological model for a series of streamflow prediction experiments under the ESP framework. The NMME-2 precipitation-driven ESP is benchmarked against the streamflow prediction obtained from the conventional ESP driven by the randomly resampled precipitation across the study watersheds. To test hypothesis 2, the relative importance of NMME-2 precipitation forecasts and IHCs is evaluated by jointly considering RevESP and NMME-2 precipitation-driven ESP across all study watersheds. More details about the hydrological model and calibration, bias correction, ESP, RevESP, and evaluation strategy and metrics are introduced in Sections 5.1 through 5.5, respectively.

5.1 Hydrological Model and Calibration

In this thesis, two suites of hydrological models are utilized based on the hydrological conditions of the watersheds. For non-snow-dominated watersheds, the Sacramento Soil Moisture Accounting (SAC-SMA) model is utilized for hydrological calibration and validation. In contrast, for snow-dominated watersheds, the Snow-17 coupled with Sac-SMA is utilized for hydrological calibration and validation.

The SAC-SMA model is one of the most extensively researched conceptual lumped rainfall-runoff models over the past decades (Chouaib et al., 2021; Roodsari et al., 2019; Zhang et al., 2012). The SAC-SMA model requires precipitation and potential evapotranspiration inputs at daily time steps and simulates surface runoff, interflow, and baseflow. To achieve this, the SAC-SMA model divides soil horizons into two-story layers: an upper zone and a lower zone. The upper zone and lower zone are further partitioned into free water, which drains by gravity, and tension water, driven by evapotranspiration and diffusion. Specifically, the upper zone primarily handles rapid hydrological responses to rainfall events. It initially stores water in its tension water storage, which represents the soil moisture available for evapotranspiration. Once this storage reaches its capacity, excess water is transferred to the free water storage, while any water that cannot infiltrate flows as surface runoff. From the free water storage, water can either flow laterally, as interflow or percolate downward into the lower zone. The lower zone, on the other hand, is responsible for slower hydrological processes. It also consists of both tension and free water storages, where the free water is further divided into primary and supplemental components. These storages release water gradually and contribute to the simulation of the baseflow of the stream.

On the other hand, the Snow-17 has a single-layer structure designed to simulate snowpack dynamics within a watershed using precipitation and temperature as inputs (Anderson, 1968). Snow-17 determines whether precipitation falls as snow, rain, or a mix, based on temperature and governs the accumulation and melting processes in the snowpack. When Snow-17 is coupled with the SAC-SMA model, Snow-17 runs first to determine the partition of precipitation into rain and snow and the evolution of the snowpack. Any snowmelt or rain passing unfrozen through the snowpack is used as input instead of precipitation for the SAC-SMA model to simulate streamflow (Raleigh & Lundquist, 2012).

In this thesis, the study period from 1981 to 2011 is divided into two separate periods: 1981-2001 (20 years) and 2001-2011 (10 years) for calibration and validation, respectively. For the first suite of the hydrological model (i.e., SAC-SMA), 13 tunable parameters of SAC-SMA are calibrated. For the second suite of the hydrological model (i.e., SAC-SMA + Snow17), 6 parameters of Snow-17, along with 13 from SAC-SMA, are calibrated. In both suites of the hydrological model, streamflow routing is accomplished through the use of the 2-parameters Nash-type Unit-Hydrograph (UH) model (Nash, 1957). All the parameters are fine-tuned iteratively with the Shuffled Complex-Self Adaptive Hybrid Evolution (SC-SAHEL) algorithm (Naeini et al., 2018). Note that the SC-SAHEL algorithm belongs to the Shuffled Complex Evolution (SCE-UA) algorithm family, which has been well-recognized and proven to be effective by numerous hydrological studies (Duan et al., 1994; Huang et al., 2018; Jeon et al., 2014; Kang & Lee, 2014). Specifically, the SC-SAHEL algorithm is employed to maximize the Kling Gupta Efficiency

(KGE; Gupta et al., 2009) of the daily simulated streamflow versus the observed streamflow. More information about the KGE is provided in section 5.5. With the calibrated parameters derived from SC-SAHEL during the calibration period, additional hydrological simulations are conducted during the validation period to ensure that the calibration of the Sac-SMA is free of overfitting.

5.2 Bias Correction of NMME-2 Precipitation Forecasts

The overall objective of bias correction is to adjust the NMME-2 precipitation forecast distribution to better align with the observed precipitation distribution from PRISM, thereby reducing systematic biases. In this thesis, the Multi-Segment Quantile Mapping (MSQM) method (Grillakis et al., 2013) is employed for the bias correction of the S2S precipitation forecast data from NMME-2.

In this thesis, both small non-zero precipitation and extreme precipitation are corrected individually within the MSQM process. This correction ensures that the frequency of rainy days (non-zero precipitation) and extreme precipitation in the NMME-2 precipitation forecasts match that of PRISM. To do that, the PRISM and NMME-2 precipitation forecast data are sorted at each grid. For small non-zero precipitation, the rank of the lowest non-zero value from the sorted PRISM dataset is identified as the “dry threshold.” Any forecast values below the “dry threshold” in the sorted NMME-2 precipitation dataset are set to zero to match the frequency of rainy days with PRISM.

To correct the extreme precipitation, the “extreme threshold” is identified by determining the value corresponding to the top 3% of the sorted PRISM precipitation data during the training period. Then, the extreme threshold is used to determine the extreme values in the sorted NMME-2 precipitation data. The extreme values from both the PRISM and the NMME-2 datasets are then used to fit a first-order linear regression model during the training period. This model is applied during the validation period to correct the extreme precipitation values in the NMME-2 precipitation forecasts. The correction follows the equation *(1)*:

$$X_{corrected,extreme} = a.X_{raw,extreme} + b \qquad (1)$$

Where, $X_{corrected,extreme}$ and $X_{raw,extreme}$ are the corrected and raw extreme precipitation values from the NMME-2 precipitation forecast data, respectively, during the validation period. a and b are the coefficients of the linear regression model derived from the training period.

To correct the remaining non-zero precipitation values that fall between the dry and extreme thresholds, both the NMME-2 and PRISM data are divided into three equal segments based on their rankings during the training period. Gamma Cumulative Distribution Functions (CDFs) are fitted separately for both PRISM and NMME-2 precipitation forecast datasets for each segment. Then, the Gamma CDF from the training period is used to correct the raw NMME-2 precipitation values from the validation period. Specifically, the correction involves mapping the CDF of the raw NMME-2 values to the corresponding quantile in the PRISM data using the inverse CDF of the PRISM distribution. This process replaces the NMME-2 precipitation values with the PRISM precipitation values at the same quantile. The correction follows the equation *(2):*

$$X_{corrected} = F_{obs}^{-1}(F_{raw}(X_{raw})) \qquad (2)$$

where X_{raw} is the raw NMME-2 precipitation from the validation period, $F_{raw}(X_{raw})$ is the CDF of the raw NMME-2 precipitation from the validation period, F_{obs}^{-1} is the inverse CDF of PRISM precipitation from the training period and $X_{corrected}$ is the final corrected NMME-2 precipitation.

In this thesis, a three-fold training and validation approach is employed, where the PRISM and NMME-2 datasets are split into three folds, with different combinations of years used for training and validation. Additionally, seasonality is considered for bias correction by dividing the dataset into four seasons: Winter (DJF), Spring (MAM), Summer (JJA), and Autumn (SON), ensuring the model accounts for varying precipitation patterns across the year. Further, the bias correction is also applied on a weekly basis: Week 1 (1–7 days), Week 2 (8–14 days), Week 3 (15–21 days), and Week 4 (22–28 days).

5.3. Ensemble Streamflow Prediction (ESP)

Two sets of streamflow prediction experiments are conducted at each of the 121 study watersheds under the ESP framework: conventional ESP and NMME-2-driven ESP (NMME-ESP). Both approaches start with the same initial step, in which the calibrated hydrological model (i.e., either SAC-SMA or SNOW17+SAC-SMA) is driven with the observed hydrometeorological data to get the best estimation of the IHCs of the watershed at forecast initialization time. This initial step is known as the "spin-up."

The conventional ESP and the NMME-ESP diverge in using precipitation data to generate streamflow predictions during the forecast period. For conventional ESP, ensembles of precipitation measurements are randomly selected and resampled using all available historical records from PRISM. Then, the resampled precipitation is used to drive the calibrated hydrological model with the estimated IHCs for ensemble streamflow predictions during the forecast period. Specifically, when conventional ESP is generating month X's streamflow predictions in a certain year Y, all month X's precipitation measurements from the collected 30 years of PRISM precipitation data, excluding year Y's, are used to create an ensemble of precipitation to drive the hydrological model with estimated IHCs for streamflow predictions. As a result, the randomly resampled precipitation, as well as the corresponding ESP-generated streamflow predictions, contain 29 ensemble members.

In contrast, NMME-ESP uses the bias-corrected NMME-2 precipitation forecasts to drive the calibrated hydrological model with the estimated IHCs to obtain ensembles of streamflow prediction during the forecast period. As a result, NMME-2-driven ESP consists of a total of 40 ensemble streamflow members, corresponding to the ensemble size of the members from NMME-2. A schematic diagram showing both NMME-ESP and conventional ESP is presented in Figure 5.1.

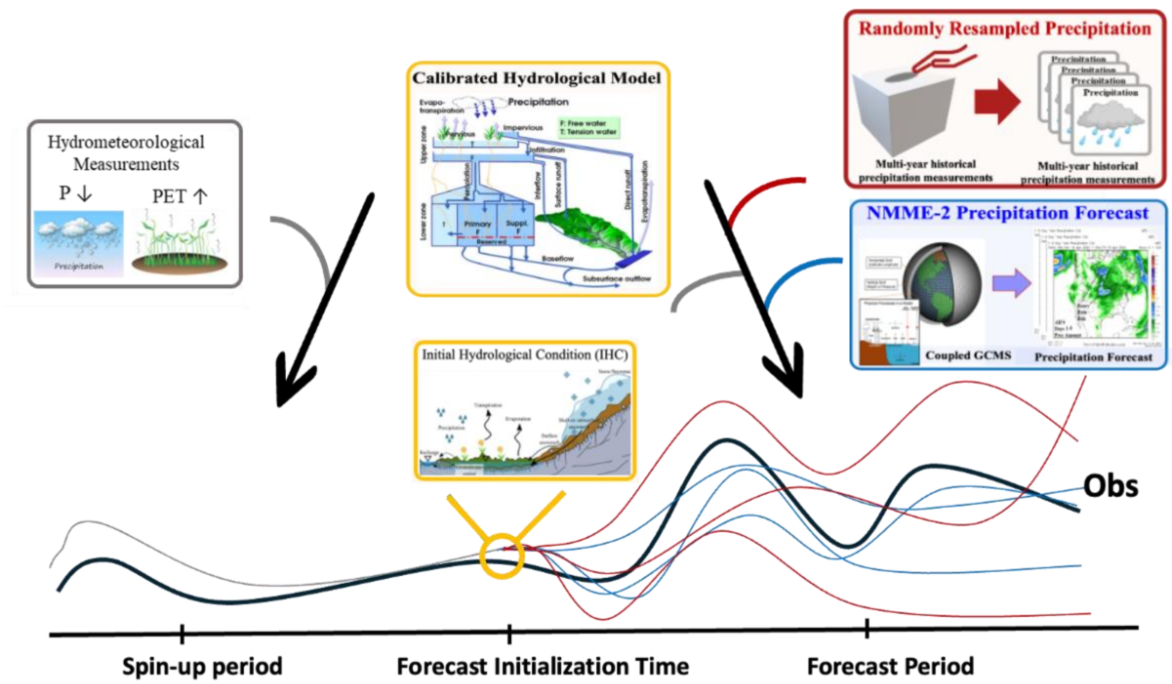


Figure 5.1: Schematic diagram showing the technical steps of ESP and NMME-ESP. The black line presents the observed streamflow from USGS. The grey line represents the simulation obtained from past hydrometeorological data from PRISM and NARR. The red line represents

the ensemble of streamflow prediction obtained by driving randomly resampled precipitation from PRISM with IHCs (conventional ESP), and the blue line represents the ensemble of streamflow prediction obtained by driving NMME-2 precipitation forecast with IHCs (NMME-ESP).

5.4. Reverse Ensemble Streamflow Prediction (RevESP)

RevESP operates during a historical period and is typically accompanied by conventional ESP. When conducting RevESP, ensembles of precipitation measurements are randomly selected and resampled from historical precipitation records. The ensembles of randomly resampled precipitation are then used to drive the calibrated hydrological model during the "spin-up" period to obtain the ensembles of IHCs at forecast initialization time. Finally, the obtained ensembles of IHCs are subsequently combined with the measured precipitation during the "forecast" period for a series of RevESP streamflow predictions. A schematic diagram showing the technical steps of the RevESP framework is presented in Figure 5.2.

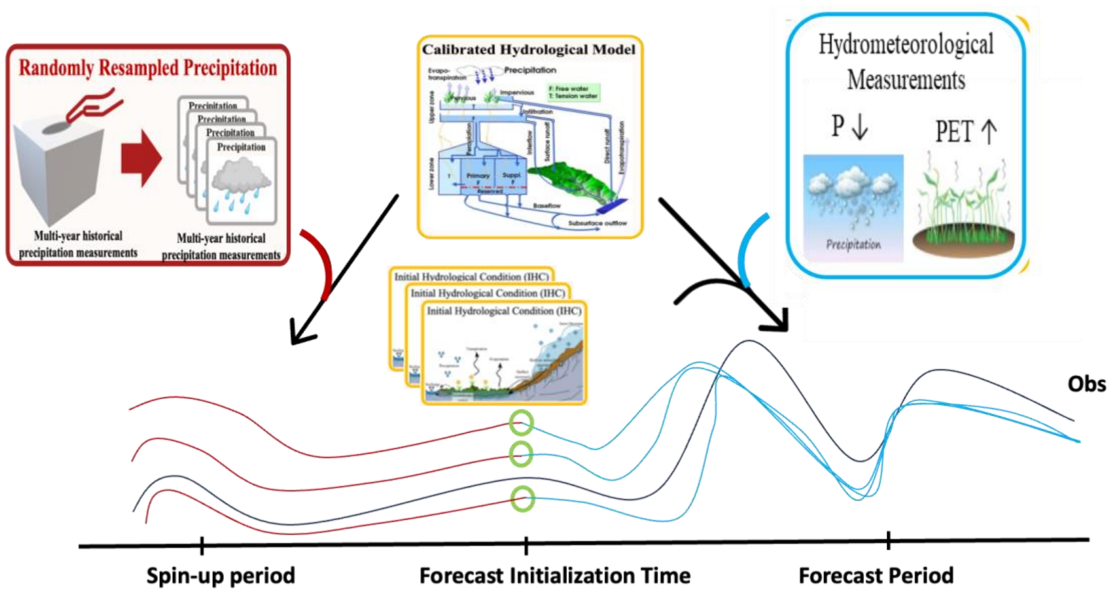


Figure 5.2: Schematic diagram showing the technical steps of the Reverse Ensemble Streamflow Prediction (RevESP) framework. The black line presents observed streamflow from USGS, and the red line represents the sequence of simulations using randomly resampled precipitation from PRISM. The blue line represents the ensemble of streamflow simulations during the "forecast period" using measured precipitation.

In this thesis, all available historical precipitation records from PRISM are resampled to generate ensembles of IHCs, which are then combined with observed precipitation to form 29-ensemble RevESP-based streamflow predictions. More specifically, when RevESP is generating month X's streamflow forecasts in a certain year Y, 12 months' precipitation measurements that are antecedent to month X from all 30 years' PRISM data records are used for the "spin-up" process to estimate IHCs, excluding those from year Y's. Then, the obtained ensemble IHCs are combined with measured precipitation of the month X to obtain 29 ensemble members of streamflow predictions.

5.5. Evaluation strategy and metrics

The evaluation strategy in this thesis is structured around obtaining and comparing the predictive skill curve of streamflow prediction across forecast lead times from 1 to 28 days in each study watershed. To obtain the skill curve for each method (conventional ESP, NMME-ESP, and RevESP), the ensemble members are averaged separately to produce a single deterministic time series of streamflow prediction. The ensemble mean of each method is then evaluated with observed streamflow data from the United States Geological Survey (USGS) to determine the predictive skill for each day within the 1 to 28-day forecast lead time.

To evaluate the effectiveness of NMME-2 precipitation forecasts in streamflow prediction, the predictive skill curve of NMME-ESP is compared with that of conventional ESP at different forecast lead times across the study watersheds. In general, both conventional ESP and NMME-2 ESP show decreasing skill curves as the forecast lead time increases.

To investigate the relative importance of IHCs and NMME-2 precipitation forecasts, the predictive skill curves of NMME-ESP and RevESP are jointly considered. Following the unique design of RevESP, the resulting streamflow hindcasts generally show an increasing skill curve over the S2S forecast horizon. By jointly considering the predictive skill curves of the streamflow predictions from both NMME-ESP and RevESP, a crossing point may be identified on the S2S forecast horizon. This crossing point is called Critical Lead Time (CLT). Theoretically, all forecast lead times before/after the CLT would be the timescales where IHCs/precipitation forecasts are the more dominant factor for streamflow predictions. A schematic diagram showing the theoretical predictive skill curves of RevESP, NMME-ESP, and CLT is presented in Figure 5.3.

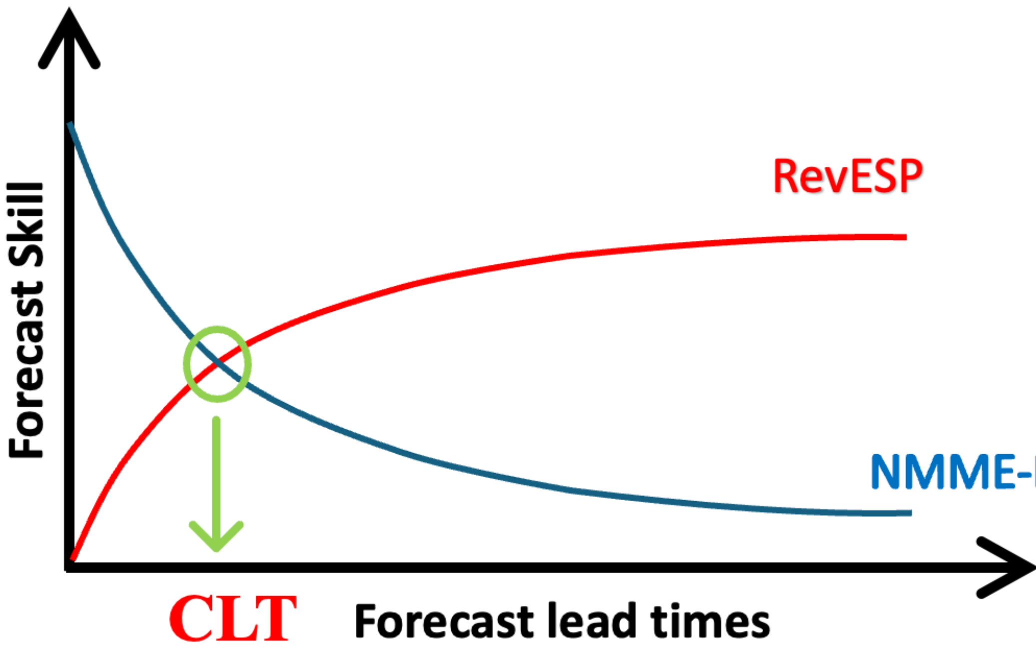


Figure 5.3: Schematic diagram showing the resulting predictive skill curves from Ensemble Streamflow Prediction (ESP) and Reverse ESP (RevESP). "CLT" demonstrates the conceptual transition from IHCs to precipitation as the more dominant factor in determining streamflow prediction.

In this thesis, KGE is used as an evaluation metric to calculate the predictive skill of the streamflow. KGE is a commonly used evaluation statistic in the field of hydrology and is derived from three separate components: the Correlation Coefficient (CC), the Bias Ratio (BR), and the Relative Variability (RV).

$$CC = \frac{\sum_{i=1}^{n}(Q_{obs,i}-\overline{Q_{obs}})(Q_{sim,i}-\overline{Q_{sim}})}{\sqrt{\sum_{i=1}^{n}(Q_{obs,i}-\overline{Q_{obs}})^2}\cdot\sqrt{\sum_{i=1}^{n}(Q_{sim,i}-\overline{Q_{sim}})^2}} \qquad (3)$$

$$BR = \frac{\frac{1}{n}\sum_{i=1}^{n}Q_{sim,i}}{\frac{1}{n}\sum_{i=1}^{n}Q_{obs,i}} \qquad (4)$$

$$RV = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{sim,i}-\overline{Q_{sim}}\right)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{obs,i}-\overline{Q_{obs}}\right)^2}} \quad (5)$$

$$KGE = 1 - \sqrt{(CC-1)^2 + (BR-1)^2 + (RV-1)^2} \quad (6)$$

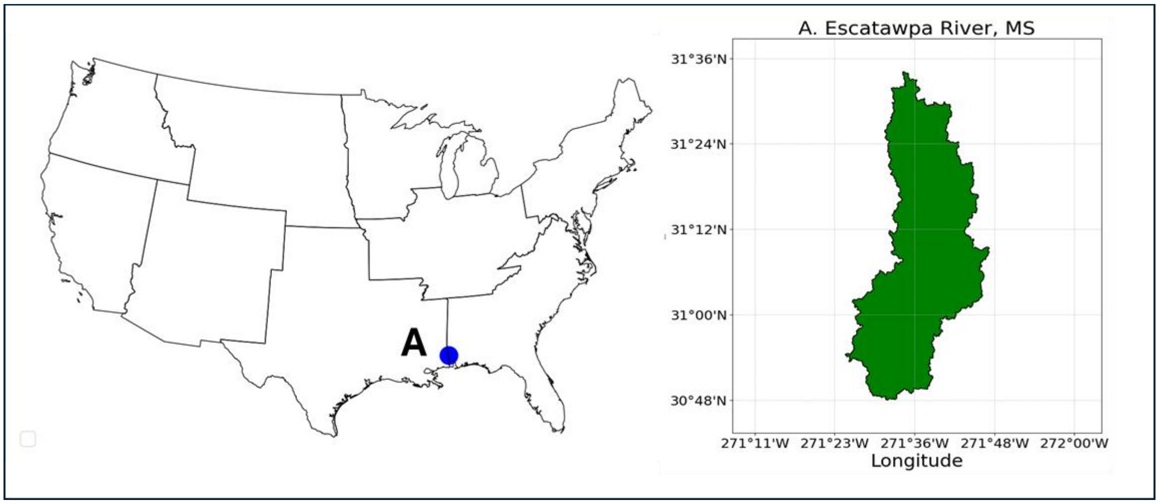
The term $Q_{sim,i}$ and $Q_{obs,i}$ in equations *(3)*, *(4)*, and *(5)* represent simulated and observed streamflow at time step *i*. The terms $\overline{Q_{sim}}$ and $\overline{Q_{obs}}$ denote the mean of the simulated and observed streamflow, respectively, calculated over the entire time period of analysis. The KGE value is computed as equation *(6)*:

KGE is a dimensionless metric that ranges from negative values to 1. A KGE of 1 indicates a perfect match between simulated and observed data. A KGE value below -0.41 indicates that the model's performance is less skillful than simply using climatology (Knoben et al., 2019; Zhang et al., 2024).

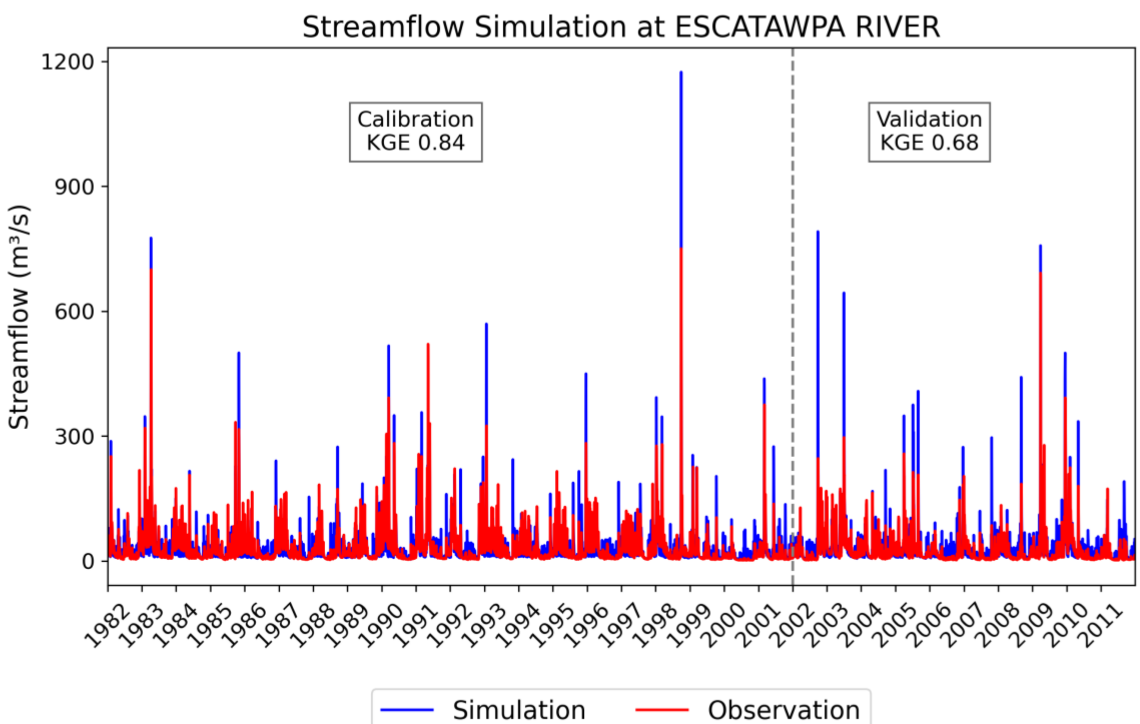
6. Results

6.1 Calibration and Validation of Hydrological Model

To ensure the hydrological model delivers reasonable performance for streamflow prediction experiments, hydrological calibration and validation are performed at each of the 121 study watersheds. Figure 6.1 presents the streamflow simulation hydrograph after the calibration and validation in one example watershed, the Escatawapa River Basin in Mississippi. According to the SWE dataset collected from the NSIDC, the long-term areal averaged SWE at the Escatawapa River Basin is less than 25 mm. Therefore, only the SAC-SMA is employed for calibration and validation. In Figure 6.1, the red line indicates streamflow measurements, and the blue line indicates the simulated streamflow. The vertical dashed line separates the entire study period into calibration and validation periods.



(A)



(B)

Figure 6.1. (A) Location of example watershed of Escatawpa River Basin in CONUS (B) Simulated and Observed hydrograph at Escatawpa River Basin

From Figure 6.1, it can be observed that the simulated streamflow reconstructs the observed streamflow with reasonable accuracy. Specifically, during the calibration period, the simulated streamflow delivers the KGE value of 0.84. During the validation period, the KGE value drops slightly but remains at 0.68.

The calibration and validation of the hydrological model (either Sac-SMA or SNOW17+Sac-SMA) are performed at each of the 121 study watersheds. Figure 6.2 shows the calibration and validation results across the study watersheds. In Figure 6.2, circles and squares mark the locations of the study watersheds. Specifically, the circle markers represent watersheds where only the Sac-SMA model is used, while square markers represent watersheds where the coupled SNOW-17 and Sac-SMA model is employed. The KGE values are shown using different shades of blue at the watershed locations, with warmer shades indicating higher KGE values and cooler shades indicating lower values. A KGE value of 1 indicates a perfect match between the simulated and observed streamflow (i.e., an optimal calibration result).

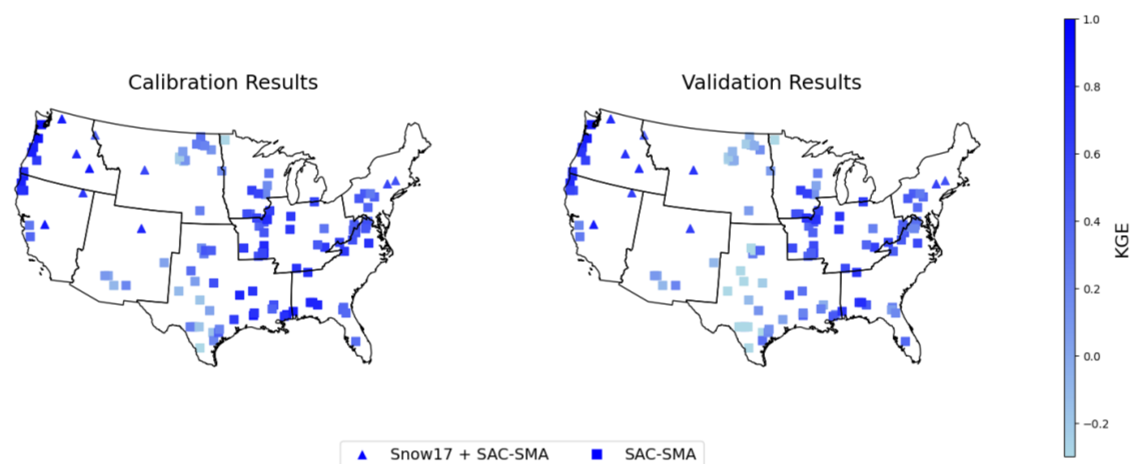


Figure 6.2. Calibration and Validation Results at all 121 Study Watersheds

According to Figure 6.2, spatial patterns can be observed in the calibration results across the study watersheds. Specifically, the calibration results are poor in watersheds in the central CONUS, mainly in the Northern Rockies, South, and Southwest regions. The poor results are indicated by lower KGE values in these regions, which range from -0.21 to 0.62. On the other hand, the calibration results are reasonably accurate in watersheds over the western and eastern CONUS (i.e., the Northwest, West, and Southeast regions.), as indicated by higher KGE values ranging from 0.4 to 0.91. In general, the KGE values for the calibration period range from -0.21 to 0.91 across the study watersheds.

Similar spatial patterns are also observed in both validation and calibration periods across most regions over CONUS, as shown in Figure 6.2. Specifically, in the Northwest, West, and Southeast regions, KGE values during validation are similar to those during calibration, ranging from 0.4 to 0.86. However, the South region is an exception, which experiences a noticeable decline in validation compared to calibration. During calibration, the KGE values in the South range from -0.15 to 0.85. In contrast, the KGE values in the South drop during the validation period, ranging from -0.31 to 0.68. Overall, the KGE values for validation across the study watersheds range from -0.31 to 0.86.

6.2 Evaluation of predictive skill of Subseasonal to Seasonal (S2S) precipitation forecast-driven ESP and conventional ESP

To evaluate the predictive skill of S2S precipitation forecasts in streamflow prediction, NMME-ESP experiments are conducted and compared against conventional ESP. Figure 6.3 presents the predictive skill of streamflow in terms of KGE from NMME-ESP and conventional

ESP across the 28-day forecast horizon in one example watershed, the Escatawapa River Basin in Mississippi. The higher the KGE value, the more skillful streamflow prediction is, with the optimal value being 1. In Figure 6.3, the blue and red lines indicate the KGE values of NMME-ESP and conventional ESP, respectively.

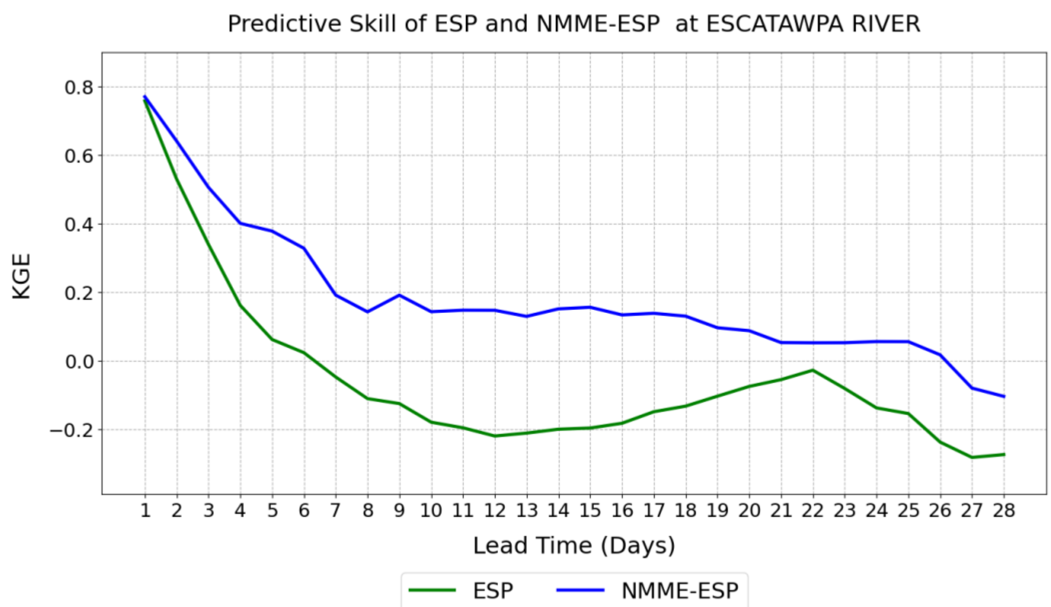


Figure 6.3. Predictive skill of conventional ESP and NMME-ESP across 1-28 days lead time at an example watershed

From Figure 6.3, it can be observed that there is a general decreasing trend in the skill curve of both NMME-ESP and conventional ESP over the 28-day forecast horizon. Specifically, the predictive skill of both NMME-ESP and conventional ESP decreases most rapidly between forecast lead times of 1 to 10 days. After day 10, the forecast skill remains at a marginal level, ranging from 0.1 to 0.2 for NMME-ESP and -0.1 to -0.2 for conventional ESP. Comparing the KGE values of NMME-ESP to those of conventional ESP, it can be observed that the NMME-ESP consistently presents higher KGE values across all the forecast lead times. This advantage of

NMME-ESP is more significant and apparent after a forecast lead time of 8 days. However, a less significant advantage of NMME-ESP is observed at both very short forecast lead times (within 3 days) and longer forecast lead times (beyond 18 days).

Building on the Escatawapa River Basin analysis, the NMME-ESP and conventional ESP experiments were conducted across all study watersheds to evaluate their predictive skill. Figure 6.4 presents the cumulative density functions (CDFs) of the KGE values for conventional ESP and NMME-ESP, grouped by weekly forecast lead times. The CDF describes the distribution of KGE values by showing the cumulative proportion of values that are less than or equal to a given threshold. In Figure 6.4, the x-axis represents the KGE values, where higher values indicate better predictive performance (1 being perfect). The y-axis shows the cumulative density, which reflects the proportion of data points with KGE values less than or equal to a specific threshold. In Figure 6.4, the blue line represents the CDF curve of NMME-ESP, while the green line represents the CDF curve of the conventional ESP.

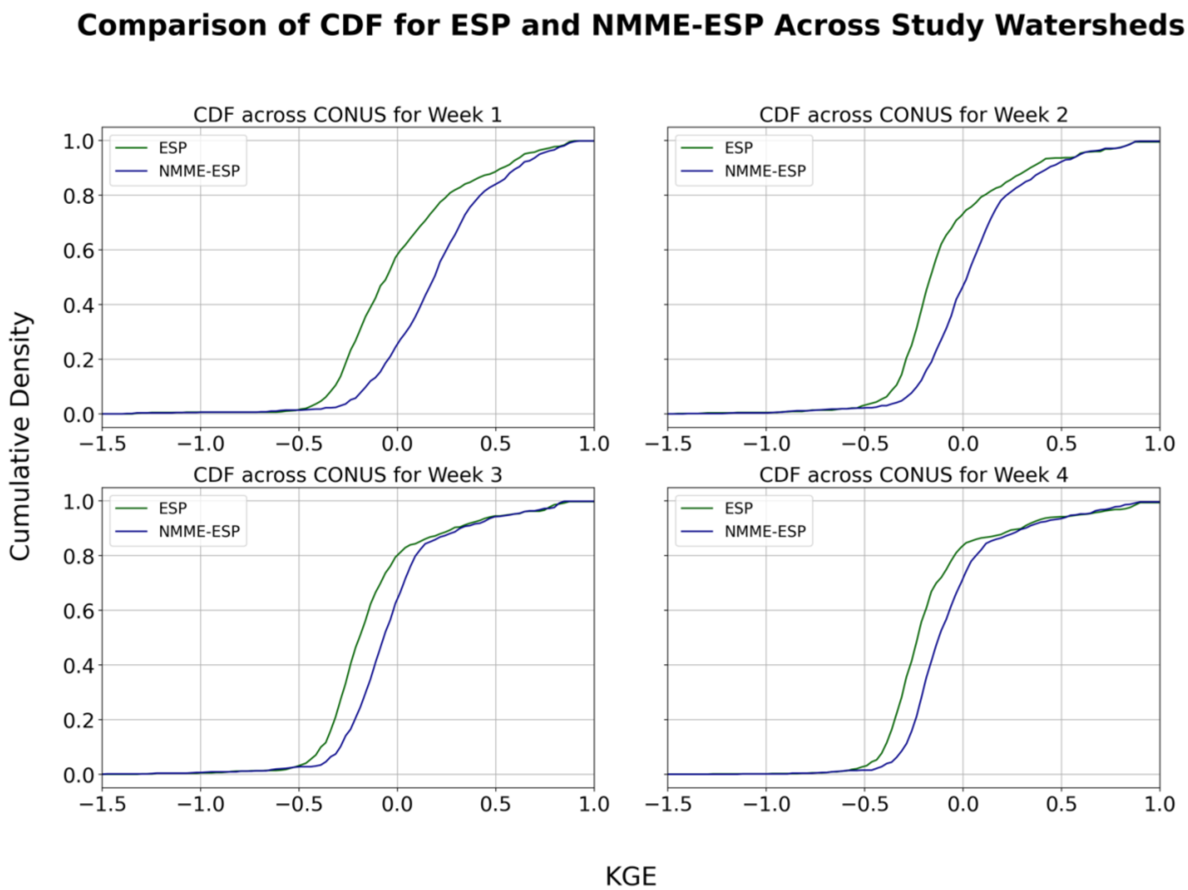


Figure 6.4. CDF of predictive skill of conventional ESP and NMME-ESP on a weekly basis at all study watersheds

Figure 6.4 shows that the KGE values for both NMME-ESP and conventional ESP generally decrease at all cumulative density levels as the forecast lead time increases. For example, at a cumulative density of 0.8 in Week 1, the KGE value for NMME-ESP is approximately 0.5. This means that 80% of the KGE values for NMME-ESP are less than or equal to 0.5 in Week 1. However, by Week 4, for the same cumulative density of 0.8, the KGE value for NMME-ESP decreases to about 0.1, indicating that 80% of the KGE values are less than or equal to 0.1. A similar trend is observed for conventional ESP. In Week 1, at a cumulative density of 0.8, the KGE value for ESP is approximately 0.3, meaning that 80% of its KGE values are less than or equal to 0.3. By Week 4, this KGE value drops to around -0.1 for the same cumulative density, indicating that 80% of the KGE values are less than or equal to -0.1.

Additionally, from Figure 6.4, a comparison can be made between the CDFs of NMME-ESP and conventional ESP across four weeks. Specifically, it can be observed that the CDFs of NMME-ESP consistently lie to the right of CDFs of conventional ESP in all four weeks. This rightward positioning indicates that, for any given cumulative density, the corresponding KGE value is higher for NMME-ESP than for ESP. For example, at a cumulative density of 0.6 in Week 1, the NMME-ESP curve is approximately 0.25 units to the right of the ESP curve, with KGE values of approximately 0.35 for NMME-ESP and 0.1 for ESP. In Week 4, for the same cumulative density of 0.6, the NMME-ESP curve is about 0.1 units to the right of the ESP curve, with KGE values of approximately 0.2 for NMME-ESP and 0.1 for ESP.

To understand more about the regional performance of NMME-ESP and ESP, a more detailed analysis was conducted by categorizing watersheds according to their climatic regions. Specifically, Figure 6.5 presents a box plot comparison between the KGE values of NMME-ESP and ESP in different climate regions over four weeks of forecast lead times. The light green box represents conventional ESP, and the light blue box represents NMME-ESP. The height of each box represents the interquartile range (IQR), and the whiskers extend to the 5th and 95th percentiles of KGE values across watersheds in a certain climate region. The horizontal line inside the box represents the median KGE values for each week. Any outliers are shown as dots outside the whiskers. The higher the median, IQR, whisker, and outliers, the more skillful streamflow prediction is.

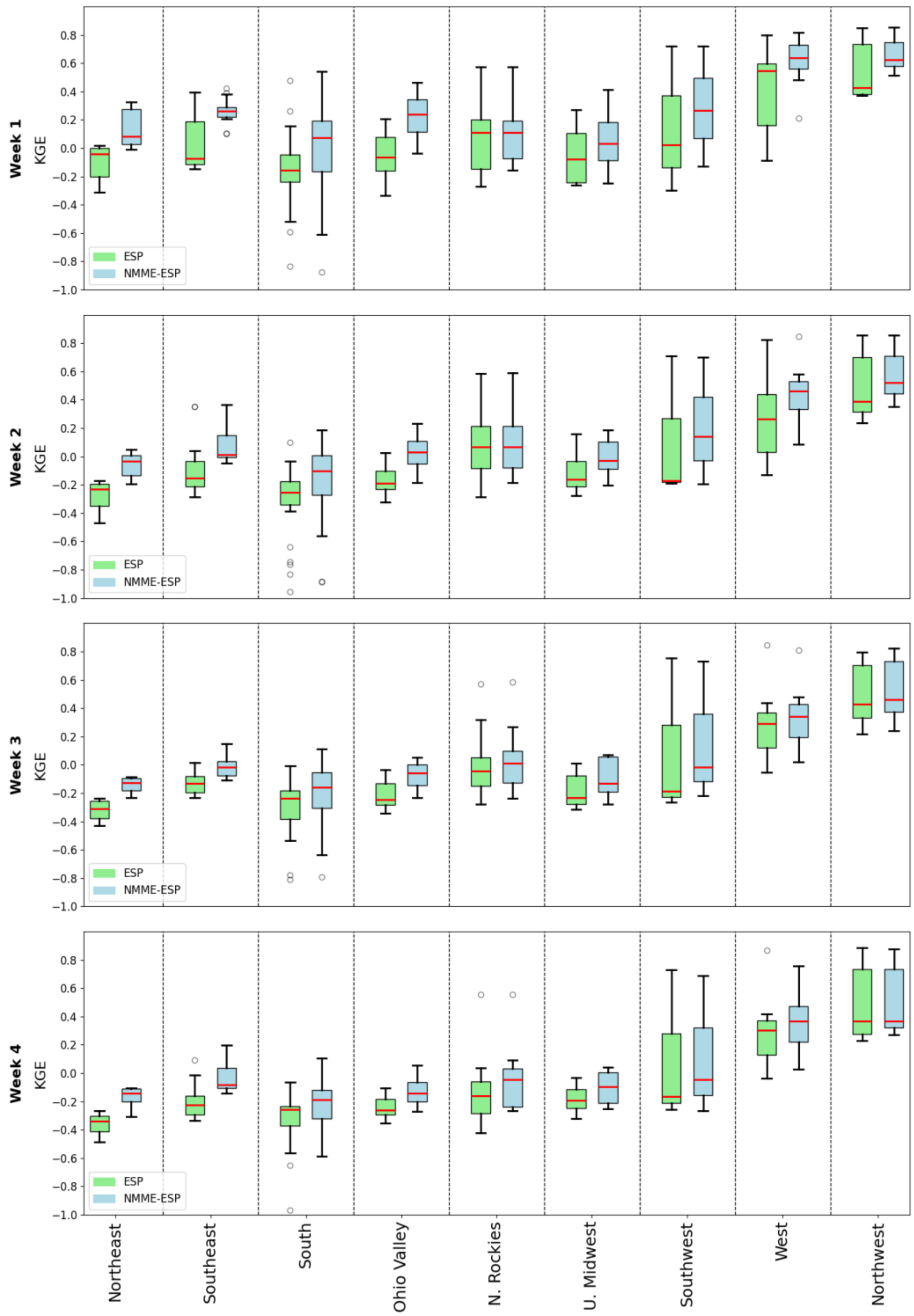


Figure 6.5. Box plot comparison of predictive skill of NMME-ESP and conventional ESP

across CONUS climate regions

The median, whiskers, and IQR of KGE for NMME-ESP and conventional ESP from Figure 6.5 show lead time-dependent performance. Specifically, it can be observed that the median, whiskers, and IQR of KGE are consistently higher for NMME-ESP than for conventional ESP in most regions across all four weeks of lead time. This advantage of NMME-ESP is more clearly visible in Week 1 and Week 2. As the lead time increases, the median, IQR, and whiskers of KGE for both NMME-ESP and conventional ESP become closer in most climatic regions. An exception can be seen in the Northern Rockies climate region, where the median, whiskers, and IQR of KGE for both NMME-ESP and conventional ESP remain close across all lead times.

Geospatial dependencies can also be observed in Figure 6.5. Specifically, in the South region, although NMME-ESP generally has higher medians and IQRs of KGE across all four weeks, some whiskers of KGE remain lower than those of conventional ESP. The lower whisker suggests that the KGE of NMME-ESP is occasionally lower compared to conventional ESP. Additionally, in the Northern Rockies, the median values of KGE of NMME-ESP stay relatively consistent with conventional ESP across the first two weeks. However, some advantages of NMME-ESP can be seen in Week 3 and Week 4. In contrast, in regions like the Southeast and Ohio Valley, the median, IQR, and whiskers of KGE of NMME-ESP are relatively higher than the KGE of conventional ESP. This advantage of NMME-ESP in Southeast and Ohio Valley is more pronounced in Week 1 and Week 2.

6.3 Investigation of the relative importance of S2S precipitation forecasts with IHCs in streamflow prediction using S2S precipitation forecast-driven ESP and RevESP

To further investigate the relative importance of NMME-2 precipitation forecasts and IHCs in determining the predictive skill of S2S streamflow prediction, NMME-ESP and RevESP experiments were jointly considered to identify the CLT at all study watersheds across the CONUS. To better visualize the CLT, the predictive skill of NMME-ESP and RevESP is plotted in Figure 6.6 for the example watershed of the Escatawapa River Basin in Mississippi. In Figure 6.6, the blue line represents the predictive skill curve associated with NMME-ESP, and the red line represents the skill curve associated with RevESP. Both the predictive skills of NMME-ESP and RevESP are quantified by KGE, with one being perfect.

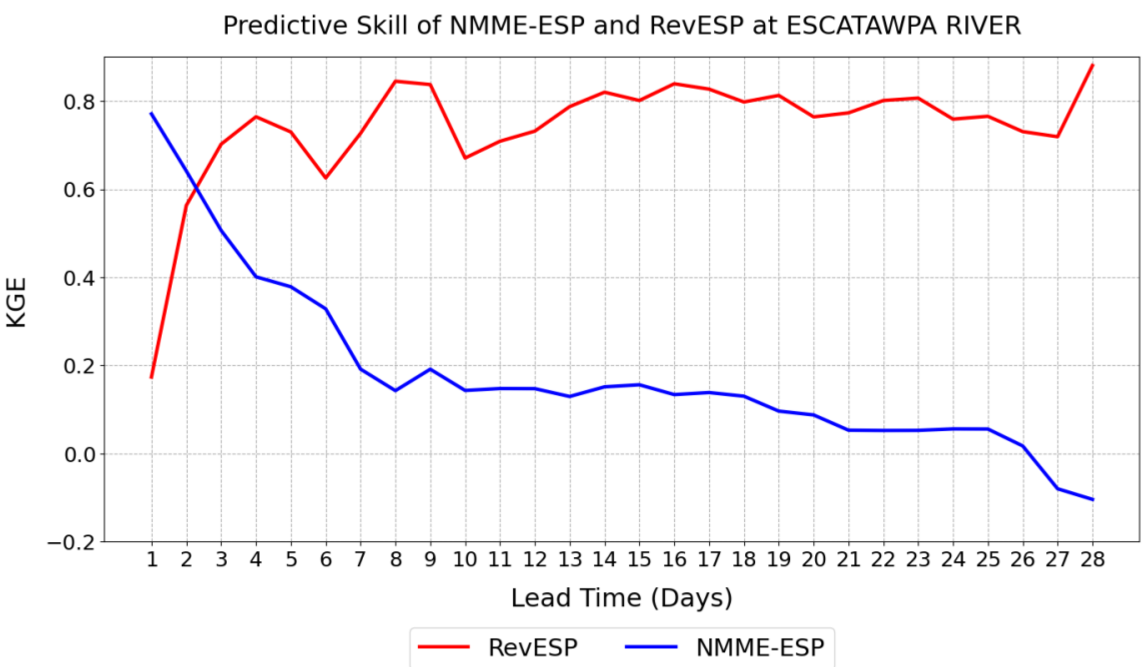


Figure 6.6. Predictive skill of NMME-ESP and RevESP across 1-28 days lead time at Escatawpa River Basin

According to Figure 6.6, NMME-ESP and RevESP exhibit distinctly different skill curves across the 28-day forecast horizon. From Figure 6.6, it can be observed that the predictive skill of NMME-ESP decreases rapidly from forecast lead time 1 to 10. After day 10, the forecast skill of NMME-ESP remains at a marginal level, ranging from 0.1 to -0.2. In contrast, the predictive skill of RevESP increases rapidly from forecast lead time 1 to 4. After day 4, the forecast skill of RevESP remains at a higher level, ranging from 0.6 to 0.85. As a result, the NMME-ESP skill curve and RevESP skill curve intersect at ~2.5 days over the 28-day forecast horizon. The observed intersection timing (i.e., 2.5 days) on the forecast horizon is the desired CLT for the Escatawpa River Basin.

Building on the analysis conducted in the Escatawapa River Basin, RevESP and NMME-ESP experiments are conducted in each of the study watersheds, and their predictive skills are jointly considered to identify the CLT. Figure 6.7 presents the identified CLT across all the study watersheds. In Figure 6.7, the circle markers represent the location of the study watersheds. The CLT of study watersheds is represented by green, red, black and different shades of blue. The cooler blue shades indicate smaller CLT values and warmer blue shades indicate larger CLT values over the 28-day S2S forecast horizon. Additionally, green markers represent watersheds with a CLT of less than 1 day, while red markers indicate watersheds where the CLT exceeds 28 days. Furthermore, black markers indicate watersheds with no distinct CLT due to the chaotic skill curves of NMME-ESP and RevESP.

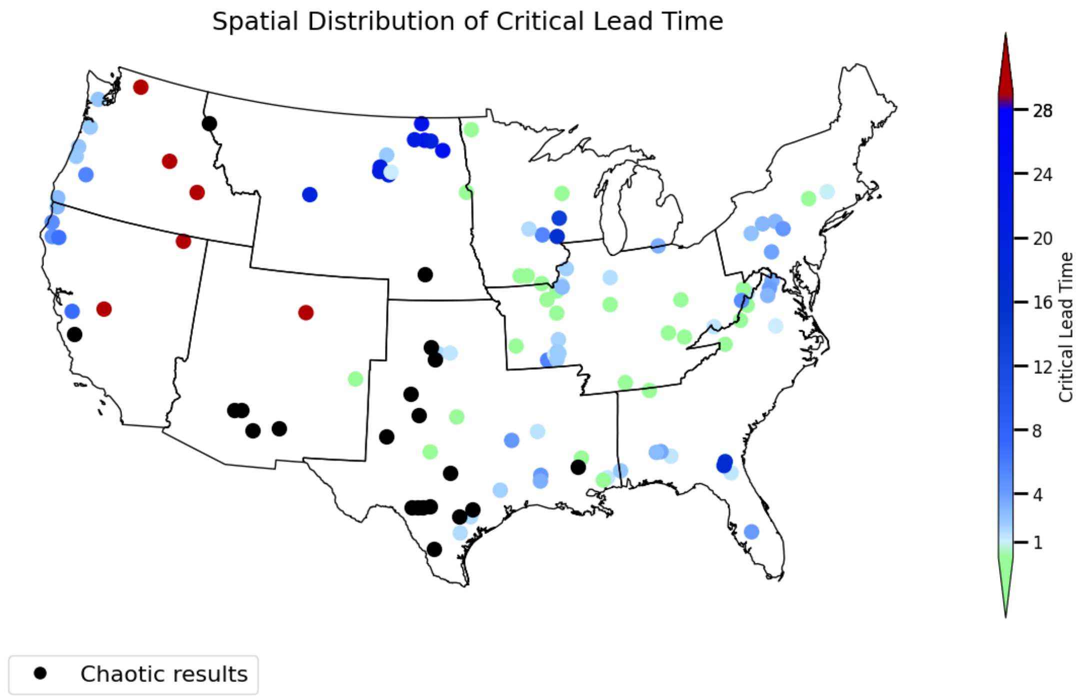


Figure 6.7. Spatial distribution of Critical Lead Time (CLT) across all 121 study watersheds

From Figure 6.7, geospatial variability can be observed in the CLTs across the CONUS. Specifically, the CLTs of the watersheds in the Ohio Valley and the South regions (i.e., Missouri, Illinois, Kentucky, and Mississippi) are predominantly shorter (less than 4 days), as indicated by green markers and lighter shades of blue. In contrast, watersheds located in the Northern Rockies (i.e., North and South Dakota) show notably longer CLTs (greater than 17 days), as indicated by warmer shades of blue. A similar pattern is also visible in snow-dominated watersheds found in the Northwest (i.e., Idaho, Washington) and West (i.e., Nevada and California), where CLTs often exceed 28 days, as indicated by red markers. Moreover, the watersheds in the Northwest (i.e., Oregon, Idaho, and Washington) region display notable variability in CLTs. Some watersheds in this region have shorter CLTs (1 to 4 days), while others exhibit much longer CLTs, exceeding 28 days. Additionally, most of the study watersheds in the Southwest (i.e., Arizona and New Mexico) and some watersheds in the South (i.e., Kansas, Oklahoma, and Texas) have chaotic results with no identifiable CLT, as represented by black markers. Specifically, chaotic results are observed in the watersheds where the calibration result is relatively poorer than in other regions. More detailed information regarding the different patterns of CLTs is presented in Appendix Section 9, which includes the NMME-ESP and RevESP skill curves for the example watersheds representing each pattern.

To understand the regional distribution of CLTs in different climatic regions, average CLT values are computed for each climate region and presented in Figure 6.8. Note that for this regional analysis, the watersheds with chaotic results are excluded from the regional CLT analysis in Figure 6.8. The Southwest climate region is also excluded from the analysis due to the limited number of study watersheds (i.e., six study watersheds in total, with four showing chaotic results).

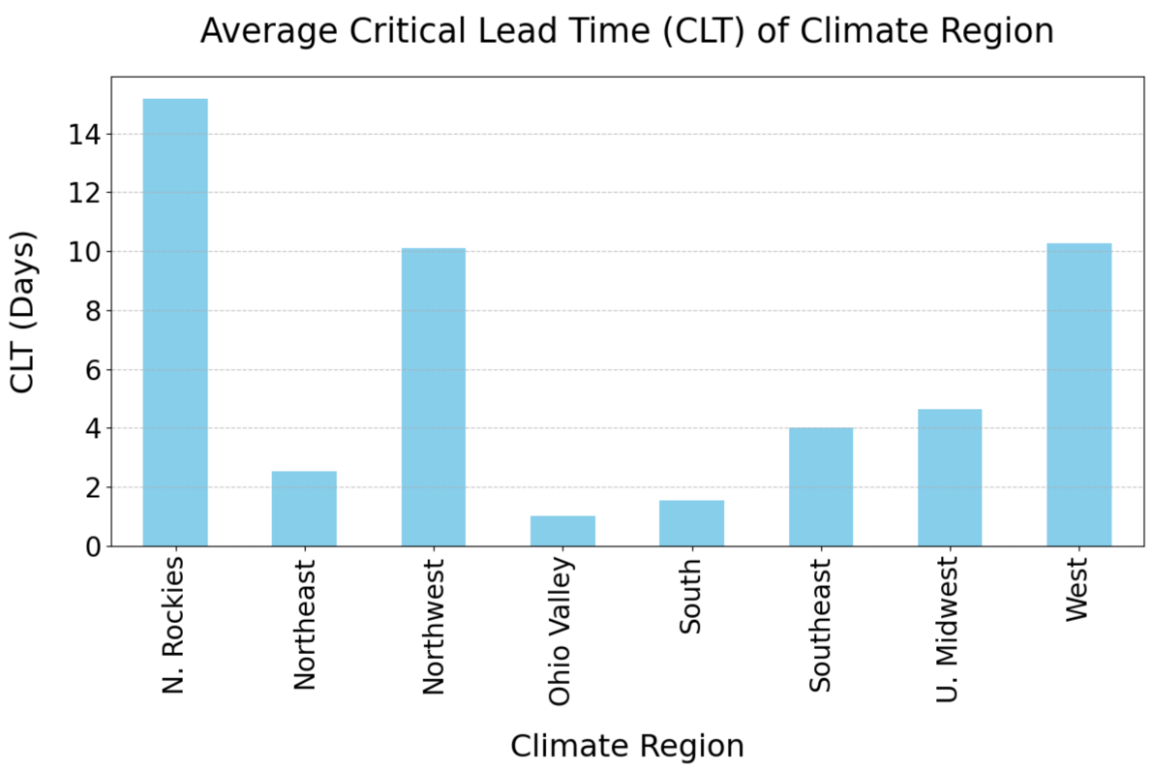


Figure 6.8. Regional Average of Critical Lead Time (CLT) across all 121 study watersheds

Figure 6.8 shows significant variation in CLT values across different regions, with some regions consistently showing higher CLTs than others. Specifically, the Northern Rockies have the highest average CLT, around 14.5 days. The Northwest and West regions also show relatively higher CLTs above 10 days. In contrast, the Ohio Valley and South regions have much shorter CLTs, averaging less than 2 days. For the rest of the regions, the CLTs are significantly shorter and range from 3 to 5 days.

With the computed CLT values, the entire 121 study watersheds are categorized into two groups using 14 days as a threshold value to investigate the effectiveness of NMME-2 precipitation forecast in streamflow prediction on watersheds based on CLT. Specifically, a threshold of 14 days was chosen because the Northern Rockies climate region, which exhibits the highest average CLT, shows improvements in predictive skill beyond 14 days of forecast lead time, as shown in Figure 6.5. Figure 6.9 shows the two groups of watersheds based on their CLT values. The blue-dotted watersheds on the right-hand side represent the watersheds with CLT values less than 14, whereas the watersheds plotted in green on the left-hand side represent the watersheds with CLT greater than 14.

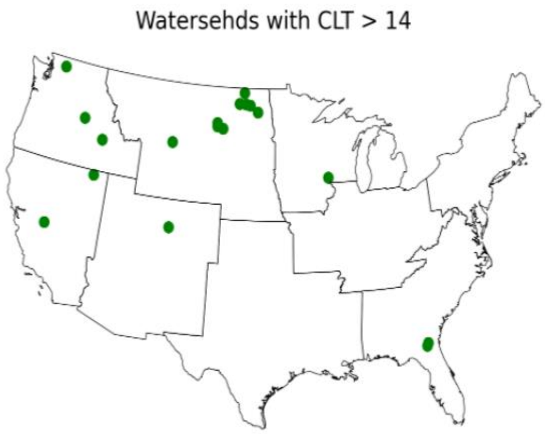


Figure 6.9. Study watersheds grouped according to Critical Lead Time (CLT)

According to Figure 6.9, there are relatively fewer watersheds that have CLT greater than 14 days. Specifically, these watersheds are located in the Northern Rockies and Northwest regions. In contrast, a majority of the study watersheds have CLT values of less than 14 days and are predominantly located in the eastern region of the CONUS (i.e., mostly in the Ohio Valley, South, Southeast, and Northeast).

Figure 6.10 shows the box plot of the added KGE value ($KGE_{NMME\text{-}ESP} – KGE_{ESP}$) brought by the incorporation of NMME-2 in ESP among the two groups of watersheds (CLT < 14 and CLT > 14) across four weeks of forecast lead times. The light red box represents the box plot of added KGE values in watersheds with CLT > 14, and the light blue box represents the box plot of added KGE values in watersheds with CLT < 14. The height of each box illustrates the IQR, and the whiskers extend to the 5th and 95th percentiles of added KGE value across watersheds. The horizontal line inside each box shows the median of the added KGE value for the respective week. Any outliers are displayed as dots beyond the whiskers. The higher the median, IQR, whiskers, and outliers indicate greater added KGE value brought by the incorporation of NMME-2 in ESP.

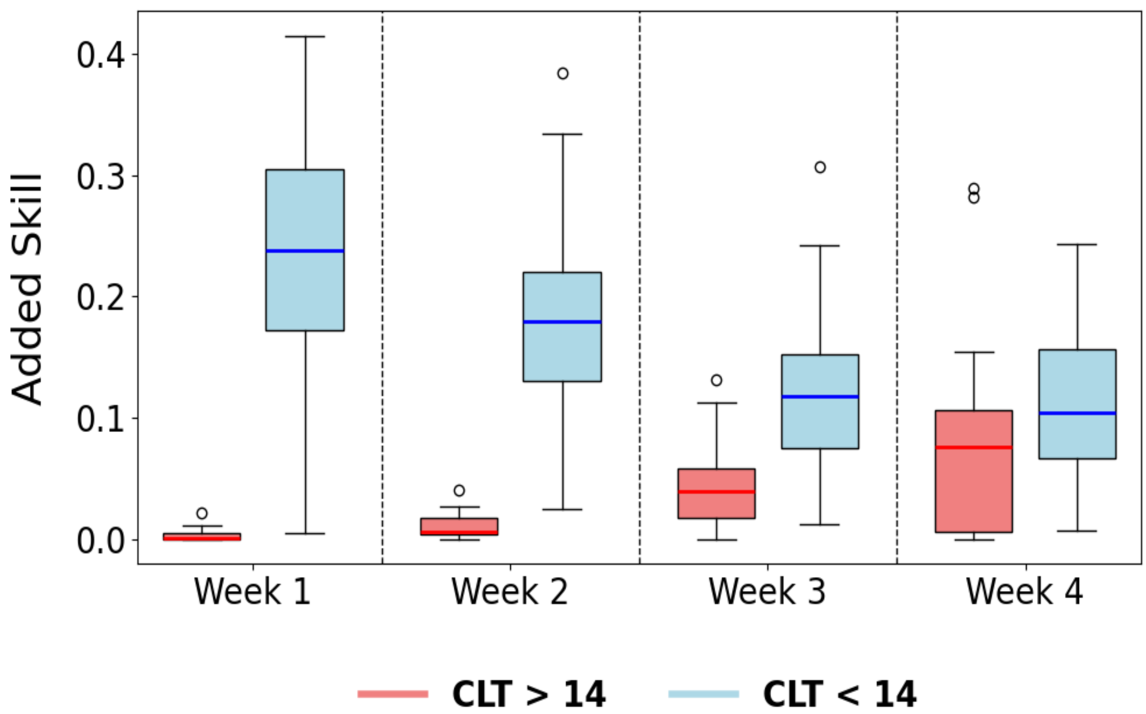


Figure 6.10. Comparison of added skill brought by NMME-2 precipitation forecast in ESP ($KGE_{NMME\text{-}ESP}$ – KGE_{ESP}) at watersheds with CLT >14 and CLT < 14

Figure 6.10 shows a distinctive pattern in the median values, IQR, whiskers, and outliers of the added KGE values for the two groups of watersheds over four weeks. In watersheds with CLT greater than 14, the added skill steadily increases with lead time, as indicated by the rising median value, IQR, whiskers, and outliers of the added KGE values as forecast lead time increases. Specifically, the added skill is marginal in weeks 1 and 2 but becomes more noticeable in weeks 3 and 4. In contrast, watersheds with CLT less than 14 show a decreasing trend in added skill over the forecast lead time, as indicated by the decreasing median, IQR, whiskers, and outliers of the added KGE values as forecast lead time increases. Specifically, the added skill is more substantial in weeks 1 and 2 but becomes less significant in weeks 3 and 4.

Additionally, a comparison of the added skill brought by NMME-2 in ESP is made between watersheds with CLT > 14 and CLT < 14, evaluated on a weekly basis. From Figure 6.8, it can be observed that the median, IQR, whiskers, and outliers of the added KGE values for CLT < 14 are higher than those for CLT > 14 across all four weeks. This implies that the added skill brought by NMME-2 into ESP is higher in watersheds with CLT < 14 than in those with CLT > 14. However, as lead time increases, the median, IQR, whiskers, and outliers of the added skill for CLT < 14 tend to move closer to those of CLT > 14. Some outliers of added skill for CLT > 14 even show higher added skill than those for CLT < 14 in Week 4.

7. Discussion

Accurate streamflow predictions at the S2S timescale are extremely challenging. Conventional ESP is a standard approach for performing S2S streamflow predictions. Despite being operational for decades, the skill of conventional ESP is generally limited at longer forecast lead times because of its reliance on the "randomly resampled precipitation". To overcome this issue, more advanced dynamical forecast products that provide reliable and accurate precipitation forecasts at the S2S timescale are critically needed under the ESP framework. Recognizing the limitations of ESP, this thesis comprehensively assessed the hydrological capability of newly developed S2S dynamical precipitation forecasts within the ESP framework. The analysis focused on their effectiveness in improving streamflow predictions across a large number of watersheds in the CONUS. The results suggest that the application of dynamical S2S precipitation forecasts brings overall more skillful streamflow predictions across a larger number of watersheds in CONUS.

Geospatially, the effectiveness of S2S precipitation forecasts in streamflow predictions varies across watersheds in the CONUS. According to the results in section 6.2, available S2S precipitation forecasts tend to be more effective in assisting streamflow predictions over certain regions than others. For instance, NMME-ESP (driven by S2S precipitation forecasts) demonstrates higher skill than the conventional ESP in regions including the Southeast and Ohio Valley (including states such as Indiana, Tennessee, Illinois, and North Carolina). However, in other regions, such as the Northern Rockies (including states like North Dakota, Idaho, and South Dakota), the advantage of NMME-ESP over the conventional ESP is marginal. These geospatial variabilities suggest that while S2S precipitation forecasts can offer better predictive skills of streamflow in general, their effectiveness is not uniformly distributed across all regions.

Further analysis by jointly considering NMME-ESP and RevESP (Section 6.3) reveals the complex lead time-dependent and geospatially varying dynamics between the employed S2S precipitation forecasts and the IHCs in streamflow predictions across the CONUS. In general, the IHCs are more important factors than future precipitation conditions for streamflow predictions at shorter forecast lead times, within ~5 days for most of the study watersheds. In contrast, future precipitation scenarios become more important and likely to determine the S2S streamflow conditions at much longer forecast lead times. However, in the Northwest, Northern Rockies, and West (including states like Idaho, North Dakota, California, and Montana) regions, the IHCs remain relatively more important than precipitation forecasts, even at longer lead times beyond 10 days. Such a result suggests that in the above-mentioned regions, better estimation of IHCs could contribute to more skillful streamflow predictions at an extended timescale.

Based on a preselected threshold value of 14 for the CLT, the study watersheds are further divided into two groups, excluding the watersheds with chaotic results. Examining the watersheds with CLT larger than 14, it can be determined that certain hydrological properties of watersheds might lead to longer-lasting effects of IHCs in streamflow predictions, thereby affecting the effectiveness of available S2S precipitation forecasts in streamflow predictions. Among the 18 watersheds that are identified with CLT larger than 14, 8 of them are heavily affected by snow (i.e., with long-term watershed average SWE values larger than 25 mm). As extensively discussed by previous studies, antecedently accumulated snow significantly contributes to runoff during the warm spring and summer seasons (Berghuijs et al., 2014; Brands et al., 2012; Li et al., 2017; Wood & Lettenmaier, 2008). Therefore, it is reasonable that for snow-dominated watersheds, IHCs, which comprise antecedent snow conditions, contribute more to the final runoff generation than incident precipitation.

Additionally, the watershed area and a specific hydrological feature termed "prairie pothole" are suspected to have an impact on the effectiveness of S2S precipitation forecasts in streamflow predictions. Specifically, 10 out of 18 watersheds that are identified with a CLT larger than 14 days demonstrate large drainage areas or present "prairie potholes". For watersheds with large drainage areas, the time of concentration is generally longer. As a result, it would take a longer time for incident precipitation to travel from the extremities of the watershed to its outlet compared to watersheds with smaller drainage areas (Li et al., 2009; Seck et al., 2015). Therefore, it is suspected that such an effect would prolong the consequence of IHCs on future streamflow conditions and weaken the effectiveness of available S2S precipitation forecasts. On the other hand, the "prairie potholes" are depressional wetlands found primarily in North Dakota and South Dakota (Kharel et al., 2016). These potholes have the capability to retain a considerable amount of water, often isolating portions of watersheds from direct drainage to streams (Mizukami et al., 2017; Wu & Lane, 2016). As a result, runoff generally takes longer to reach the watershed outlet. Hence, it is suspected that the effect of IHCs in the watersheds with prairie potholes will be prolonged, weakening the effectiveness of S2S precipitation forecasts in assisting streamflow predictions.

Given that the major objective of this thesis lies in validating and examining the effectiveness of available S2S precipitation forecasts in assisting streamflow predictions, the above-mentioned hydrological properties are not investigated in detail in terms of their effect on streamflow predictions. However, I believe future efforts should be made in this regard. Specifically, I recommend that future research focus on a detailed examination of the dynamics between S2S precipitation forecasts and various hydrological processes and properties involved in streamflow predictions.

Additionally, the relative importance of IHCs and S2S precipitation forecasts in streamflow prediction remains challenging to determine in the South (including states like Oklahoma, Texas, and Kansas) and Southwest (including states like Arizona and New Mexico) regions. Specifically, chaotic predictive skill curves of both NMME-ESP and RevESP are observed in the above-mentioned regions. As a result, multiple cross-section points between the skill curves of NMME-ESP and RevESP are observed at numerous watersheds. Given such phenomena, the exact forecast lead time at which S2S precipitation forecasts overtake IHCs as the dominant source of streamflow predictability remains unclear. It is suspected that the unclear relationships between S2S precipitation and IHCs in the South and Southwest regions stem from the poor simulation capability of available hydrological models (Section 6.1). In this thesis, an operational hydrological model, along with a well-recognized optimization algorithm for hydrologic model calibration, has been used. However, the accuracy of the hydrological model simulation is still unsatisfactory in some of the study watersheds. The observed poor hydrological simulation performance (especially in the South and Southwest) may result from the limitation of spatially lumped hydrological models to accurately simulate hydrological behavior in arid and semi-arid regions, which has been reported by several previous studies (Clark et al., 2008; Martinez & Gupta, 2010; Newman et al., 2015). While alternative semi-distributed or distributed hydrological models may better capture the hydrological behavior of the Southwest and South regions, significant challenges remain, including the availability of high-quality input data and the need for region-specific model calibration. These limitations are beyond the scope of this study. Therefore, I recommend that future efforts focus on identifying and implementing hydrologic models that are better suited for these regions, coupled with enhanced calibration techniques, to improve the accuracy and reliability of S2S streamflow predictions.

Another caveat of this thesis is that only the S2S precipitation forecasts are investigated in terms of their effectiveness in streamflow predictions. However, available S2S temperature forecasts could also play an important role in assisting streamflow predictions. As briefly mentioned, S2S precipitation forecasts tend to add less skill to streamflow predictions if a substantial amount of snow is present in a watershed. Therefore, I suspect that available S2S temperature forecasts will be more effective in assisting streamflow predictions in snow-dominated watersheds. Additionally, I suspect that the available S2S temperature forecast should be informative in predicting the snow melt runoff volumes during the warm seasons of spring and summer, thereby improving the overall streamflow predictions. Therefore, I recommend future research to investigate and incorporate dynamic temperature forecasts to assist streamflow predictions at the S2S timescale.

Moreover, In this thesis, only precipitation forecasts from NMME-2 are evaluated for their effectiveness in streamflow prediction. In recent decades, several S2S forecast products in recent decades are now available through governmental and research agency collaborations like SubX and The S2S Prediction Project (Pegion et al., 2019; Robertson & Vitart, 2018). However, in the current literature, these S2S forecast products remain underexplored for their effectiveness in streamflow prediction on a larger scale across CONUS. Expanding their use in streamflow prediction could improve our understanding of the hydrological applicability of S2S forecast products. I recommend future studies systematically evaluate these products to enhance regional solutions and improve streamflow prediction skills.

8. Conclusion

In this thesis, I comprehensively evaluated the effectiveness of S2S precipitation forecasts from the NMME-2 in streamflow prediction across 121 watersheds, which span diverse geographic and hydrological conditions in the CONUS. The evaluation is conducted by benchmarking the predictive skill of S2S precipitation forecast-driven ESP against that of conventional ESP, which uses randomly resampled precipitation. Additionally, this thesis investigates the relative importance of S2S precipitation forecasts and IHCs in contributing to the predictive skill of streamflow at different S2S forecast lead times across the 121 study watersheds. The relative importance of IHCs and S2S precipitation forecasts is examined by jointly considering the skill curves produced by NMME-ESP and the RevESP. The major conclusions from this thesis are listed as follows:

1. The integration of the S2S precipitation forecast from NMME-2 in ESP leads to overall more skillful streamflow predictions across a larger number of watersheds across CONUS.
2. The effectiveness of available S2S precipitation forecasts in assisting streamflow predictions demonstrates geospatial and forecast lead time-dependent variations.
3. The S2S precipitation forecasts are relatively more important and informative than the IHCs in streamflow predictions at most of the watersheds, especially at longer forecast lead times.
4. Certain hydrological properties, such as larger watershed areas or substantial snow conditions, may weaken the effectiveness of dynamical S2S precipitation forecasts in assisting streamflow predictions.

Recognizing areas where the current analysis can be further expanded, this work suggests several avenues for future research. For example, further investigations are recommended to explore the impact of watershed-specific characteristics on S2S streamflow prediction. Further, it is also recommended that future studies focus on improving hydrological simulations (particularly in the South and Southwest regions of the U.S.) and investigate the combined influence of S2S precipitation and temperature forecasts, especially in snow-dominated watersheds for overall more accurate and reliable S2S streamflow predictions.

9. *Appendix*

In this thesis, spatially varied patterns can be observed in CLT across different watersheds, as explained in section 6.3. Figure 9.1 illustrates the examples of these spatial patterns in five representative watersheds in different locations. In Figure 9.1, Cooler blue shades indicate watersheds with shorter CLTs, while warmer blue shades represent watersheds with longer CLTs. Dark red indicates watersheds with CLT values greater than 28 days, and light green indicates watersheds with CLT values shorter than 1 day. Black represents watersheds with no distinct CLT. Detailed analyses of the skill curves and CLT values for these examples are provided in sub-sections 9.1 to 9.5. Figures 9.2 to 9.6 present the skill curves for representative watersheds, illustrating the comparative performance of NMME-ESP (blue line) and RevESP (red line). The predictive skill is quantified using the KGE, where a value of 1 represents perfect skill.

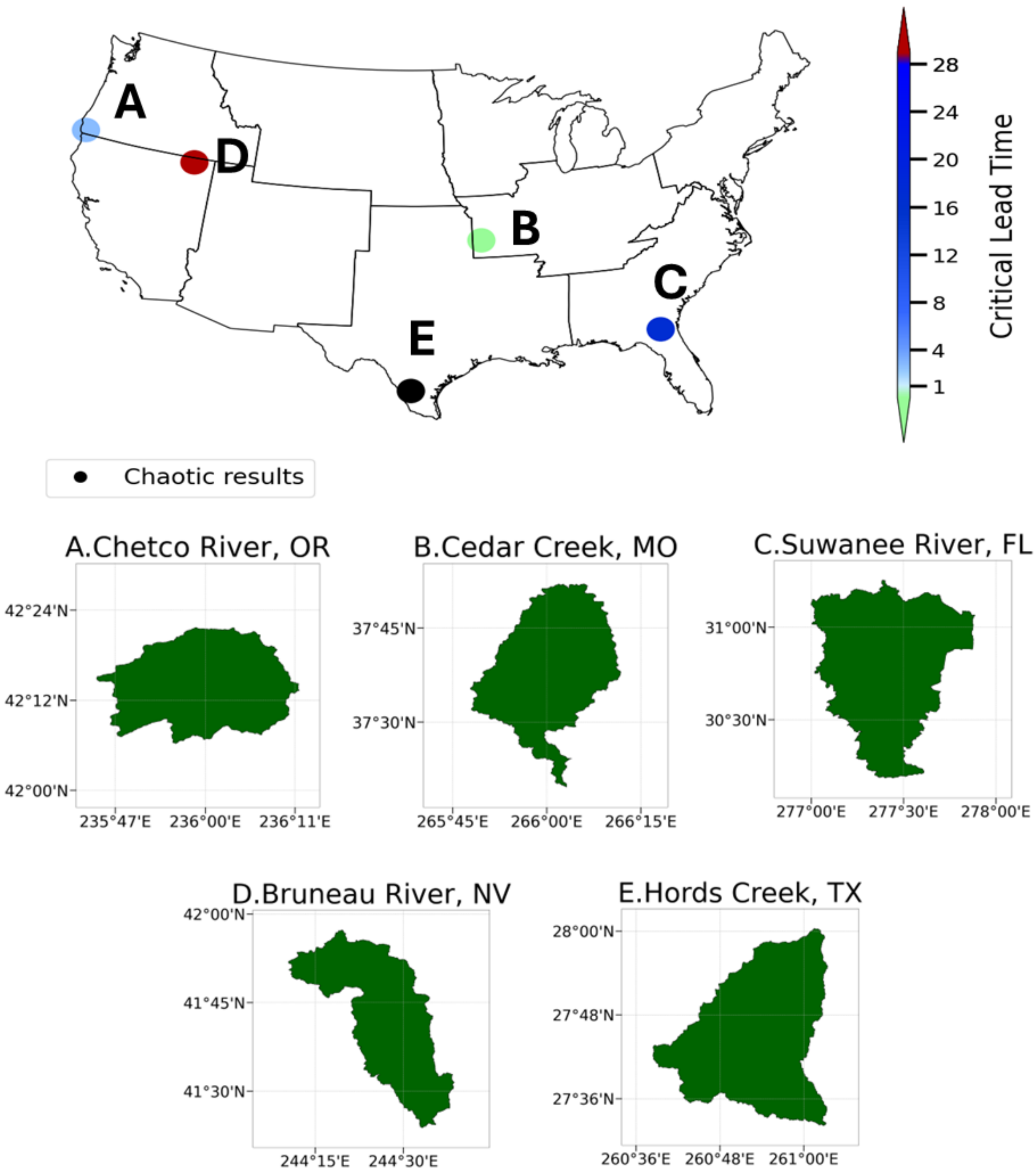


Figure 9.1. Location of representative watersheds across CONUS that show different spatial patterns in CLT

9.1 CLT at shorter forecast lead time

Figure 9.2 illustrates the skill curves for the Chetco River Basin in Oregon. The intersection of the NMME-ESP and RevESP curves occurs at approximately 3 days, defining the CLT for this watershed. This pattern (CLT between 1 and 5 days) is mainly found in the eastern and western CONUS watersheds with relatively high mean areal precipitation.

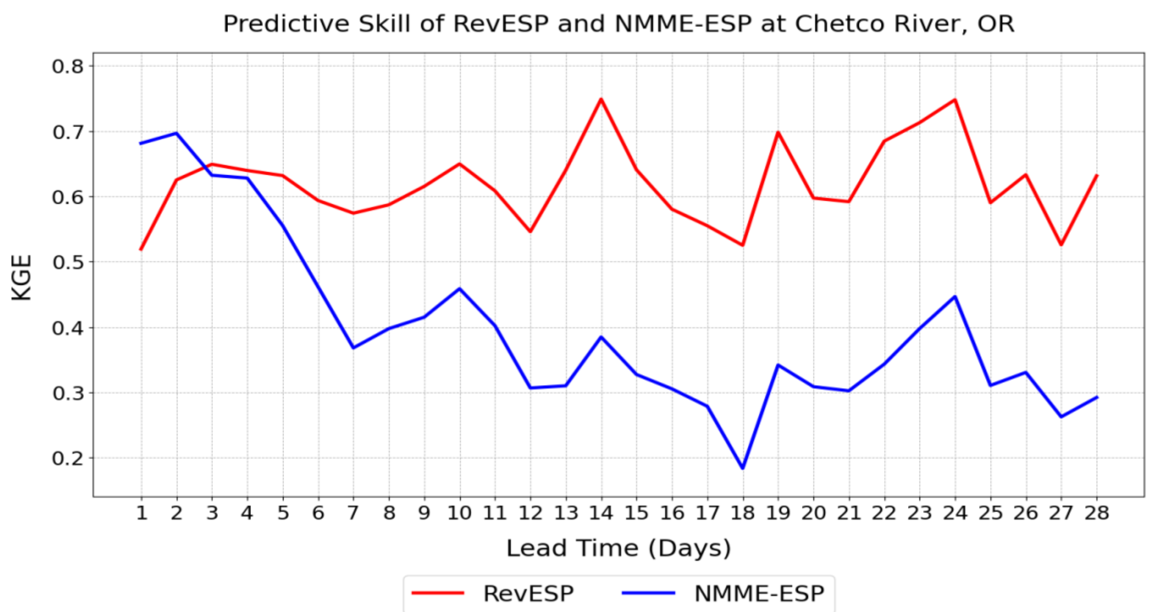


Figure 9.2. Predictive skill of NMME-ESP and RevESP across 1-28 days lead time at Chetco River Basin, OR

9.2 CLT at forecast lead time less than 1

Figure 9.3 illustrates the skill curves for Cedar Creek in Missouri. The RevESP curve remains consistently higher than the NMME-ESP curve, indicating a CLT of less than 1 day. This suggests that S2S precipitation forecasts are relatively more important than IHCs at all 28-day forecast lead times. Such patterns (CLT < 1 day) are commonly seen in humid regions with high mean areal precipitation, particularly in the Ohio Valley.

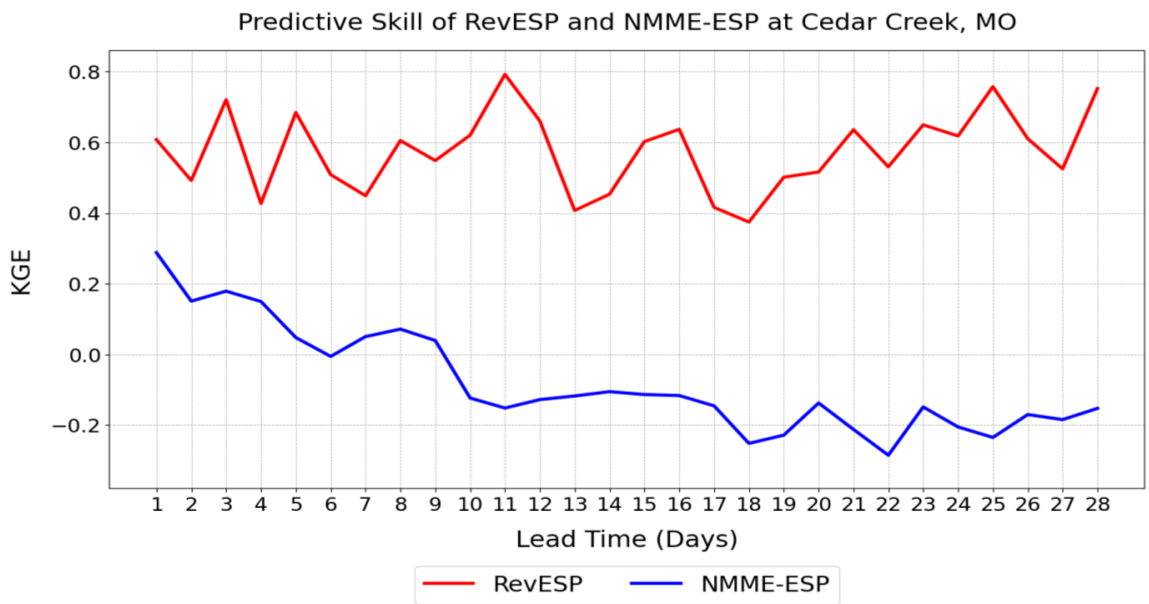


Figure 9.3. Predictive skill of NMME-ESP and RevESP across 1-28 days lead time at Cedar Creek Basin, MO

9.3 CLT at longer forecast lead time

Figure 9.4 illustrates the skill curve for the Suwanee River Basin in Florida. The intersection of NMME-ESP and RevESP occurs approximately 17, indicating that the relative importance of IHCs can sustain up to longer forecast lead time in the 28-day forecast horizon. This pattern (CLT > 14 days) is often observed in watersheds with large drainage areas and complex geological features like “Prarie Potholes”.

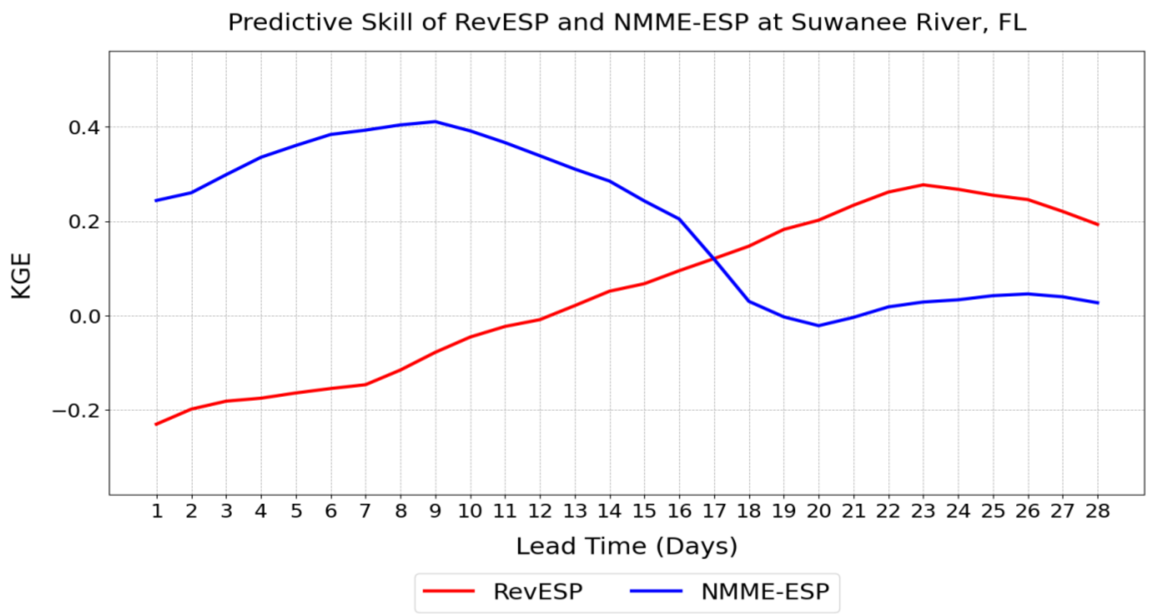


Figure 9.4. Predictive skill of NMME-ESP and RevESP across 1-28 days lead time at Suwanee River Basin, FL

9.4 CLT at forecast lead time greater than 28

Figure 9.5 shows the skill curves for the Bruneau River Basin in Nevada. The NMME-ESP curve is consistently higher than the RevESP curve across all 28-day forecast lead times, suggesting a CLT greater than 28 days. The skill curve indicates that IHCs are relatively more important across all 28-day forecast horizons. This skill curve pattern is mainly seen in watersheds with substantial snowpack.

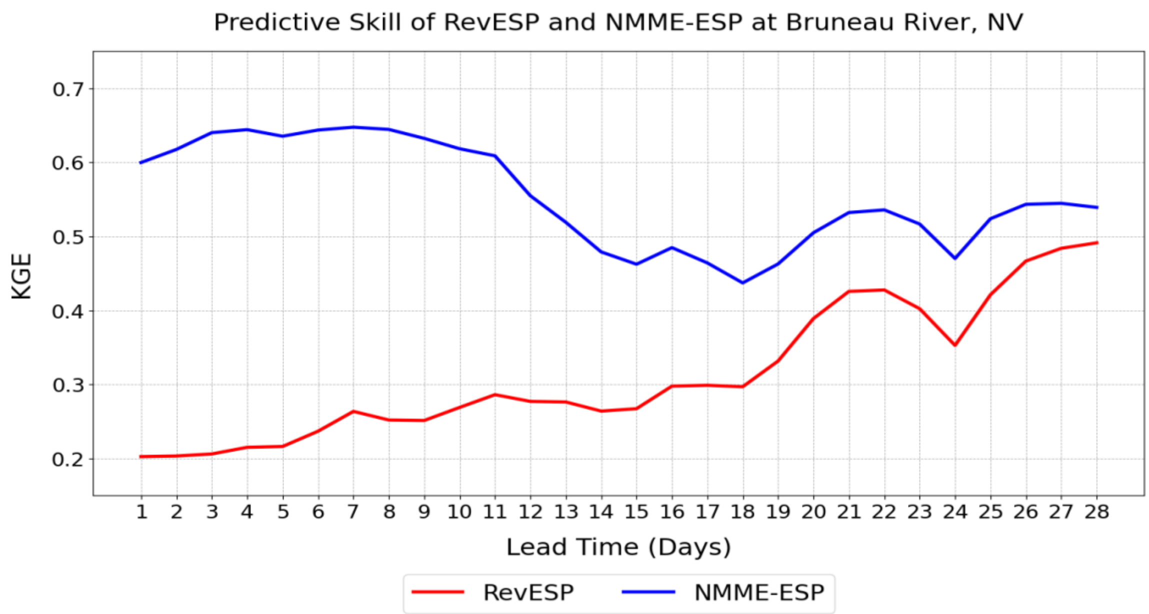


Figure 9.5. Predictive skill of NMME-ESP and RevESP across 1-28 days lead time at Bruneau River Basin, NV

9.5 No distinct CLT

Figure 9.6 displays the skill curves for a watershed with no distinct CLT. From Figure 9.6, it can be observed that multiple intersections occur between NMME-ESP and RevESP at 14, 16, 19, 24, 26, and 27 days, reflecting the chaotic behavior of NMME-ESP and RevESP. This pattern is typically observed in watersheds with low calibration results, particularly in the South and Southwest regions.

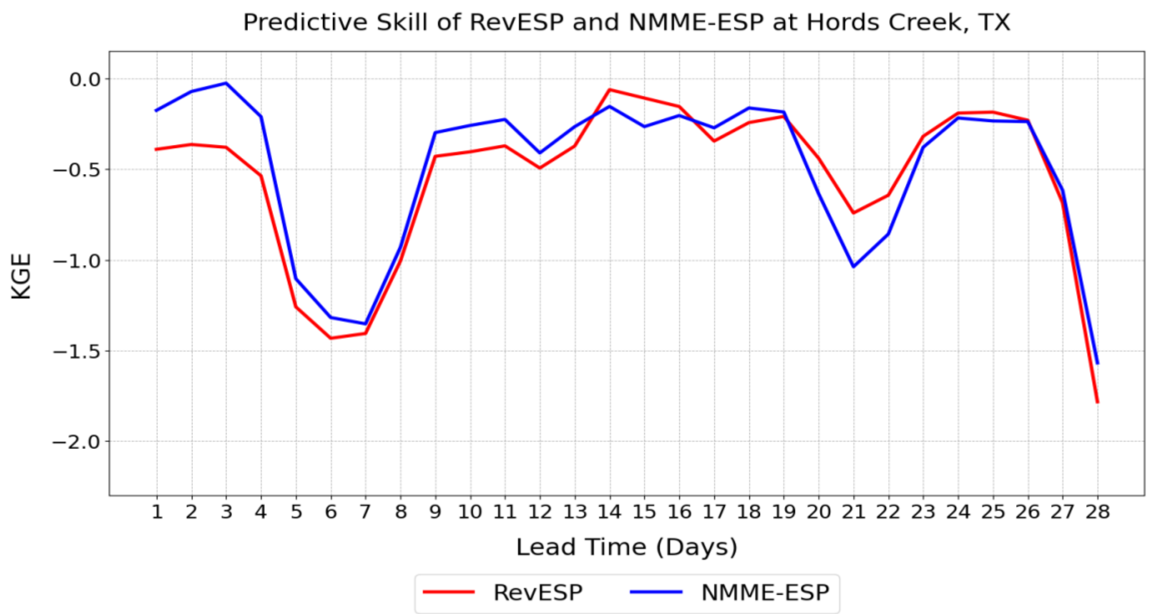


Figure 9.6. Predictive skill of NMME-ESP and RevESP across 1-28 days lead time at Hords Creek Basin, TX

References

Anderson, E. A. (1968). Development and testing of snow pack energy balance equations. *Water Resources Research*, *4*(1), 19–37. https://doi.org/10.1029/WR004i001p00019

Anghileri, D., Monhart, S., Zhou, C., Bogner, K., Castelletti, A., Burlando, P., & Zappa, M. (2019). The Value of Subseasonal Hydrometeorological Forecasts to Hydropower Operations: How Much Does Preprocessing Matter? *Water Resources Research*, *55*(12), 10159–10178. https://doi.org/10.1029/2019WR025280

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., & Pappenberger, F. (2018). Skilful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*, *22*(4), 2057–2072. https://doi.org/10.5194/hess-22-2057-2018

Baker, S. A., Wood, A. W., & Rajagopalan, B. (2019). Developing Subseasonal to Seasonal Climate Forecast Products for Hydrology and Water Management. *JAWRA Journal of the American Water Resources Association*, *55*(4), 1024–1037. https://doi.org/10.1111/1752-1688.12746

Bauer, P., Thorpe, A., & Brunet, G. (2015a). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Bauer, P., Thorpe, A., & Brunet, G. (2015b). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, *4*(7), 583–586. https://doi.org/10.1038/nclimate2246

Brands, S., Manzanas, R., Gutiérrez, J. M., & Cohen, J. (2012). Seasonal Predictability of Wintertime Precipitation in Europe Using the Snow Advance Index. *Journal of Climate*, *25*(12), 4023–4028. https://doi.org/10.1175/JCLI-D-12-00083.1

Broxton, P., Zeng, X., & Dawson, N. (2019). *Daily 4 km Gridded SWE and Snow Depth from Assimilated In-Situ and Modeled Data over the Conterminous US, Version 1* [Dataset].

NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/0GGPB220EX6A

Butler, A., Charlton-Perez, A., Domeisen, D. I. V., Garfinkel, C., Gerber, E. P., Hitchcock, P., Karpechko, A. Yu., Maycock, A. C., Sigmond, M., Simpson, I., & Son, S.-W. (2019). Sub-seasonal Predictability and the Stratosphere. In *Sub-Seasonal to Seasonal Prediction* (pp. 223–241). Elsevier. https://doi.org/10.1016/B978-0-12-811714-9.00011-5

Cao, Q., Shukla, S., DeFlorio, M. J., Ralph, F. M., & Lettenmaier, D. P. (2021). Evaluation of the subseasonal forecast skill of floods associated with atmospheric rivers in coastal Western U.S. watersheds. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-20-0219.1

Chevallier, M., Massonnet, F., Goessling, H., Guémas, V., & Jung, T. (2019). The Role of Sea Ice in Sub-seasonal Predictability. In *Sub-Seasonal to Seasonal Prediction* (pp. 201–221). Elsevier. https://doi.org/10.1016/B978-0-12-811714-9.00010-3

Chiew, F. H. S., Zhou, S. L., & McMahon, T. A. (2003). Use of seasonal streamflow forecasts in water resources management. *Journal of Hydrology*, *270*(1–2), 135–144. https://doi.org/10.1016/S0022-1694(02)00292-5

Chouaib, W., Alila, Y., & Caldwell, P. V. (2021). Implications of a Priori Parameters on Calibration in Conditions of Varying Terrain Characteristics: Case Study of the SAC-SMA Model in Eastern United States. *Hydrology*, *8*(2), 78. https://doi.org/10.3390/hydrology8020078

Christensen, H. M., & Berner, J. (2019). From reliable weather forecasts to skilful climate response: A dynamical systems approach. *Quarterly Journal of the Royal Meteorological Society*, *145*(720), 1052–1069. https://doi.org/10.1002/qj.3476

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12), 2007WR006735. https://doi.org/10.1029/2007WR006735

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal of Applied Meteorology*, *33*(2), 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Day, G. N. (1985). Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management*, *111*(2), 157–170. https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

De Andrade, F. M., Coelho, C. A. S., & Cavalcanti, I. F. A. (2019). Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dynamics*, *52*(9–10), 5451–5475. https://doi.org/10.1007/s00382-018-4457-z

Duan, Q., Pappenberger, F., Wood, A., Cloke, H. L., & Schaake, J. C. (Eds.). (2019). *Handbook of Hydrometeorological Ensemble Forecasting*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39925-1

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, *158*(3–4), 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Farmer, W. H., & Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research*, *52*(7), 5619–5633. https://doi.org/10.1002/2016WR019129

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., & Zhang, M. (2011). The Community Climate System Model Version 4. *Journal of Climate*, *24*(19), 4973–4991. https://doi.org/10.1175/2011JCLI4083.1

Gerber, E. P., Butler, A., Calvo, N., Charlton-Perez, A., Giorgetta, M., Manzini, E., Perlwitz, J., Polvani, L. M., Sassi, F., Scaife, A. A., Shaw, T. A., Son, S.-W., & Watanabe, S. (2012). Assessing and Understanding the Impact of Stratospheric Dynamics and Variability on the Earth System. *Bulletin of the American Meteorological Society*, *93*(6), 845–859. https://doi.org/10.1175/BAMS-D-11-00145.1

Gobena, A. K., & Gan, T. Y. (2010). Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology*, *385*(1–4), 336–352. https://doi.org/10.1016/j.jhydrol.2010.03.002

Greuell, W., Franssen, W. H. P., & Hutjes, R. W. A. (2019). Seasonal streamflow forecasts for Europe – Part 2: Sources of skill. *Hydrology and Earth System Sciences*, *23*(1), 371–391. https://doi.org/10.5194/hess-23-371-2019

Grillakis, M. G., Koutroulis, A. G., & Tsanis, I. K. (2013). Multisegment statistical bias correction of daily GCM precipitation output. *Journal of Geophysical Research: Atmospheres*, *118*(8), 3150–3162. https://doi.org/10.1002/jgrd.50323

Guo, Y., & Nie, H. (2020). Summertime daily precipitation statistics over East China in CFSv2. *Physics and Chemistry of the Earth, Parts A/B/C*, *115*, 102841. https://doi.org/10.1016/j.pce.2020.102841

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., … Fujita, T. (2018). Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1218–1256. https://doi.org/10.1002/qj.3179

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., & Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, *22*(3), 2023–2039. https://doi.org/10.5194/hess-22-2023-2018

Holland, M. M., Blanchard-Wrigglesworth, E., Kay, J., & Vavrus, S. (2013). Initial-value predictability of Antarctic sea ice in the Community Climate System Model 3. *Geophysical Research Letters*, *40*(10), 2121–2124. https://doi.org/10.1002/grl.50410

Huang, L., Wang, L., Zhang, Y., Xing, L., Hao, Q., Xiao, Y., Yang, L., & Zhu, H. (2018). Identification of Groundwater Pollution Sources by a SCE-UA Algorithm-Based Simulation/Optimization Model. *Water*, *10*(2), 193. https://doi.org/10.3390/w10020193

Jeon, J.-H., Park, C.-G., & Engel, B. (2014). Comparison of Performance between Genetic Algorithm and SCE-UA for Calibration of SCS-CN Surface Runoff Simulation. *Water*, *6*(11), 3433–3456. https://doi.org/10.3390/w6113433

Kalnay, E., Lord, S. J., & McPherson, R. D. (1998). Maturity of Operational Numerical Weather Prediction: Medium Range. *Bulletin of the American Meteorological Society*, *79*(12), 2753–2769. https://doi.org/10.1175/1520-0477(1998)079<2753:MOONWP>2.0.CO;2

Kang, T., & Lee, S. (2014). Modification of the SCE-UA to Include Constraints by Embedding an Adaptive Penalty Function and Application: Application Approach. *Water Resources Management*, *28*(8), 2145–2159. https://doi.org/10.1007/s11269-014-0602-6

Karl, T., & Koss, W. J. (1984). *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983*.

Kharel, G., Zheng, H., & Kirilenko, A. (2016). Can land-use change mitigate long-term flood risks in the Prairie Pothole Region? The case of Devils Lake, North Dakota, USA. *Regional Environmental Change*, *16*(8), 2443–2456. https://doi.org/10.1007/s10113-016-0970-y

Kim, Y.-O., Jeong, D.-I., & Kim, H.-S. (2001). Improving Water Supply Outlook in Korea with Ensemble Streamflow Prediction. *Water International*, *26*(4), 563–568. https://doi.org/10.1080/02508060108686957

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., … Wood, E. F. (2014). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331.

Konzmann, M., Gerten, D., & Heinke, J. (2013). Climate impacts on global irrigation requirements under 19 GCMs, simulated with a vegetation and hydrology model. *Hydrological Sciences Journal*, *58*(1), 88–105. https://doi.org/10.1080/02626667.2013.746495

Koster, R. D., Mahanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., Dirmeyer, P. A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J. -H., Lawrence, D. M., Lee, W. -S., Li, Z., Luo, L., Malyshev, S., Merryfield, W. J., Seneviratne, S. I., … Wood, E. F. (2010). Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophysical Research Letters*, *37*(2), 2009GL041677. https://doi.org/10.1029/2009GL041677

Krishnamurti, T. N., Gnanaseelan, C., & Chakraborty, A. (2007). Prediction of the Diurnal Change Using a Multimodel Superensemble. Part I: Precipitation. *Monthly Weather Review*, *135*(10), 3613–3632. https://doi.org/10.1175/MWR3446.1

Kundzewicz, Z. W., & Stakhiv, E. Z. (2010). Are climate models "ready for prime time" in water resources management applications, or is more research needed? *Hydrological Sciences Journal*, *55*(7), 1085–1089. https://doi.org/10.1080/02626667.2010.513211

Lau, W. K. M., & Waliser, D. E. (2011). *Intraseasonal Variability in the Atmosphere-Ocean Climate System*. Springer Berlin Heidelberg. https://books.google.com/books?id=qEOtLLjij-sC

Levey, J. R., & Sankarasubramanian, A. (2024). Spatial and Temporal Variation of Subseasonal-to-Seasonal (S2S) Precipitation Reforecast Skill across the CONUS. *Journal of Hydrometeorology*, *25*(5), 755–770. https://doi.org/10.1175/JHM-D-23-0159.1

Li, D., Wrzesien, M. L., Durand, M., Adam, J., & Lettenmaier, D. P. (2017). How much runoff originates as snow in the western United States, and how will that change in the future? *Geophysical Research Letters*, *44*(12), 6163–6172. https://doi.org/10.1002/2017GL073551

Li, H., Luo, L., Wood, E. F., & Schaake, J. (2009). The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research: Atmospheres*, *114*(D4), 2008JD010969. https://doi.org/10.1029/2008JD010969

Liu, J., Yuan, X., Zeng, J., Jiao, Y., Li, Y., Zhong, L., & Yao, L. (2022). Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrology and Earth System Sciences*, *26*(2), 265–278. https://doi.org/10.5194/hess-26-265-2022

Mahanama, S., Livneh, B., Koster, R., Lettenmaier, D., & Reichle, R. (2012). Soil Moisture, Snow, and Seasonal Streamflow Forecasts in the United States. *Journal of Hydrometeorology*, *13*(1), 189–203. https://doi.org/10.1175/JHM-D-11-046.1

Martinez, G. F., & Gupta, H. V. (2010). Toward improved identification of hydrological models: A diagnostic evaluation of the " *abcd* " monthly water balance model for the conterminous United States. *Water Resources Research*, *46*(8), 2009WR008294. https://doi.org/10.1029/2009WR008294

Matalas, N. C. (1998). Note on the assumption of hydrologic stationarity. *Journal of Contemporary Water Research and Education*, *112*(1), 11.

Maurer, E. P., Lettenmaier, D. P., & Mantua, N. J. (2004). Variability and potential sources of predictability of North American runoff. *Water Resources Research*, *40*(9), 2003WR002789. https://doi.org/10.1029/2003WR002789

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R. S., Fyfe, J. C., Tang, Y., & Polavarapu, S. (2013). The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization. *Monthly Weather Review*, *141*(8), 2910–2945. https://doi.org/10.1175/MWR-D-12-00216.1

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., & Shi, W. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360. https://doi.org/10.1175/BAMS-87-3-343

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity Is Dead: Whither Water Management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., & Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020–8040. https://doi.org/10.1002/2017WR020401

Najafi, M. R., Moradkhani, H., & Piechota, T. C. (2012). Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *Journal of Hydrology*, *442–443*, 105–116. https://doi.org/10.1016/j.jhydrol.2012.04.003

Nash, J. (1957). The form of the instantaneous unit hydrograph. *Comptes Rendus et Rapports Assemblee Generale de Toronto, 3*, 114–121.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., & Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223. https://doi.org/10.5194/hess-19-209-2015

Newman, Andrew. (2014). *A large-sample watershed-scale hydrometeorological dataset for the contiguous USA* (p. approximately 2.5 GB) [Text/plain, text/tab-separated-values, png, shp]. [object Object]. https://doi.org/10.5065/D6MW2F4D

Paiva, R. C. D., Collischonn, W., Bonnet, M. P., & De Gonçalves, L. G. G. (2012). On the sources of hydrological prediction uncertainty in the Amazon. *Hydrology and Earth System Sciences*, *16*(9), 3127–3137. https://doi.org/10.5194/hess-16-3127-2012

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., Bell, R., DelSole, T., Min, D., Zhu, Y., Li, W., Sinsky, E., Guan, H., Gottschalck, J., Metzger, E. J., Barton, N. P., Achuthavarier, D., Marshak, J., Koster, R. D., … Kim, H. (2019). The Subseasonal Experiment (SubX): A Multimodel Subseasonal Prediction Experiment. *Bulletin of the*

American Meteorological Society, *100*(10), 2043–2060. https://doi.org/10.1175/BAMS-D-18-0270.1

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R., & Jenkins, A. (2017). Hydrological Outlook UK: An operational streamflow and groundwater level forecasting system at monthly to seasonal time scales. *Hydrological Sciences Journal*, *62*(16), 2753–2768. https://doi.org/10.1080/02626667.2017.1395032

Rahnamay Naeini, M., Yang, T., Sadegh, M., AghaKouchak, A., Hsu, K., Sorooshian, S., Duan, Q., & Lei, X. (2018). Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL) optimization framework. *Environmental Modelling & Software*, *104*, 215–235. https://doi.org/10.1016/j.envsoft.2018.03.019

Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., Clark, M. P., & Samaniego, L. (2019). Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States. *Journal of Geophysical Research: Atmospheres*, *124*(24), 13991–14007. https://doi.org/10.1029/2019JD030767

Raleigh, M. S., & Lundquist, J. D. (2012). Comparing and combining SWE estimates from the SNOW-17 model using PRISM and SWE reconstruction. *Water Resources Research*, *48*(1), 2011WR010542. https://doi.org/10.1029/2011WR010542

Remesan, R., Bellerby, T., & Frostick, L. (2014). Hydrological modelling using data from monthly GCMs in a regional catchment: HYDROLOGICAL IMPACT ASSESSMENT USING MONTHLY GCM DATA. *Hydrological Processes*, *28*(8), 3241–3263. https://doi.org/10.1002/hyp.9872

Robertson, A., & Vitart, F. (2018). *Sub-seasonal to seasonal prediction: The gap between weather and climate forecasting*. Elsevier.

Robertson, A. W., Kumar, A., Peña, M., & Vitart, F. (2015). Improving and Promoting Subseasonal to Seasonal Prediction. *Bulletin of the American Meteorological Society*, *96*(3), ES49–ES53. https://doi.org/10.1175/BAMS-D-14-00139.1

Roodsari, B. K., Chandler, D. G., Kelleher, C., & Kroll, C. N. (2019). A comparison of SAC-SMA and Adaptive Neuro-fuzzy Inference System for real-time flood forecasting in small urban catchments. *Journal of Flood Risk Management*, *12*(S1), e12492. https://doi.org/10.1111/jfr3.12492

Roy, A., Murtugudde, R., Sahai, A. K., Narvekar, P., Shinde, V., & Ghosh, S. (2022). Water Savings with Irrigation Water Management at Multi-week Lead Time Using Extended Range Predictions. *Climate Services*, *27*, 100320. https://doi.org/10.1016/j.cliser.2022.100320

Schaake, J., & Larson, L. (1998). Ensemble streamflow prediction (ESP): Progress and research needs. *Special Symp. on Hydrology*, J19–J24.

Seck, A., Welty, C., & Maxwell, R. M. (2015). Spin-up behavior and effects of initial conditions for an integrated hydrologic model. *Water Resources Research*, *51*(4), 2188–2210. https://doi.org/10.1002/2014WR016371

Shrestha, D. L., Robertson, D. E., Wang, Q. J., Pagano, T. C., & Hapuarachchi, H. A. P. (2013). Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose. *Hydrology and Earth System Sciences*, *17*(5), 1913–1931. https://doi.org/10.5194/hess-17-1913-2013

Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrology and Earth System Sciences*, *15*(11), 3529–3538. https://doi.org/10.5194/hess-15-3529-2011

Shukla, S., Sheffield, J., Wood, E. F., & Lettenmaier, D. P. (2013). On the sources of global land surface hydrologic predictability. *Hydrology and Earth System Sciences*, *17*(7), 2781–2796. https://doi.org/10.5194/hess-17-2781-2013

Singh, S. K. (2016). Long-term Streamflow Forecasting Based on Ensemble Streamflow Prediction Technique: A Case Study in New Zealand. *Water Resources Management*, *30*(7), 2295–2309. https://doi.org/10.1007/s11269-016-1289-7

Souza Filho, F. A., & Lall, U. (2003). Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resources Research*, *39*(11), 2002WR001373. https://doi.org/10.1029/2002WR001373

Tian, D., Wood, E. F., & Yuan, X. (2017). CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States. *Hydrology and Earth System Sciences*, *21*(3), 1477–1490. https://doi.org/10.5194/hess-21-1477-2017

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. (2021). Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. *Water Resources Research*, *57*(7), e2020WR028392. https://doi.org/10.1029/2020WR028392

United States Geological Survey. (1994). *USGS Water Data for the Nation*. U.S. Geological Survey. https://doi.org/10.5066/F7P55KJN

Van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., & Beck, H. E. (2013). Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resources Research*, *49*(5), 2729–2746. https://doi.org/10.1002/wrcr.20251

Vernieres, G., Rienecker, M. M., Kovach, R., & Keppenne, C. L. (2012). *The GEOS-iODAS: Description and evaluation*.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., … Zhang, L. (2017). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Wanders, N., & Wood, E. F. (2016). Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, *11*(9), 094007. https://doi.org/10.1088/1748-9326/11/9/094007

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., Vitart, F., Coughlan De Perez, E., Ray, A. J., Murray, V., Bharwani, S., MacLeod, D., James, R., Fleming, L., Morse, A. P., Eggen, B., Graham, R., Kjellström, E., Becker, E., … Zebiak, S. E. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*(3), 315–325. https://doi.org/10.1002/met.1654

White, C. J., Franks, S. W., & McEvoy, D. (2015). Using subseasonal-to-seasonal (S2S) extreme rainfall forecasts for extended-range flood prediction in Australia. *Proceedings of the International Association of Hydrological Sciences*, *370,* 229–234. https://doi.org/10.5194/piahs-370-229-2015

Wood, A. W., Kumar, A., & Lettenmaier, D. P. (2005). A retrospective assessment of National Centers for Environmental Prediction climate model–based ensemble hydrologic forecasting in the western United States. *Journal of Geophysical Research: Atmospheres*, *110*(D4), 2004JD004508. https://doi.org/10.1029/2004JD004508

Wood, A. W., & Lettenmaier, D. P. (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters*, *35*(14), 2008GL034648. https://doi.org/10.1029/2008GL034648

Wu, Q., & Lane, C. R. (2016). Delineation and Quantification of Wetland Depressions in the Prairie Pothole Region of North Dakota. *Wetlands*, *36*(2), 215–227. https://doi.org/10.1007/s13157-015-0731-6

Yang, Y., Roderick, M. L., Yang, D., Wang, Z., Ruan, F., McVicar, T. R., Zhang, S., & Beck, H. E. (2021). Streamflow stationarity in a changing world. *Environmental Research Letters*, *16*(6), 064096. https://doi.org/10.1088/1748-9326/ac08c1

Yossef, N. C., Winsemius, H., Weerts, A., Van Beek, R., & Bierkens, M. F. P. (2013). Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resources Research*, *49*(8), 4687–4699. https://doi.org/10.1002/wrcr.20350

Zhang, L., Gao, S., & Yang, T. (2024). Adapting subseasonal-to-seasonal (S2S) precipitation forecast at watersheds for hydrologic ensemble streamflow forecasting with a machine learning-based post-processing approach. *Journal of Hydrology*, *631*, 130643. https://doi.org/10.1016/j.jhydrol.2024.130643

Zhang, L., Kim, T., Yang, T., Hong, Y., & Zhu, Q. (2021). Evaluation of Subseasonal-to-Seasonal (S2S) precipitation forecast from the North American Multi-Model ensemble phase II (NMME-2) over the contiguous U.S. *Journal of Hydrology*, *603*, 127058. https://doi.org/10.1016/j.jhydrol.2021.127058

Zhang, Z., Koren, V., Reed, S., Smith, M., Zhang, Y., Moreda, F., & Cosgrove, B. (2012). SAC-SMA a priori parameter differences and their impact on distributed hydrologic model simulations. *Journal of Hydrology*, *420–421*, 216–227. https://doi.org/10.1016/j.jhydrol.2011.12.004