Water Supply

Water Supply Vol 25 No 3, 647 doi: 10.2166/ws.2025.029

Evaluation of CanCM3 and CanCM4 models from the North American Multi-Model Ensemble (NMME) for drought prediction in arid and semi-arid basins of Iran

Mehdi Moghasemi[a], Narges Zohrabi [b,*], Hossein Fathian[a], Alireza Nikbakht Shahbazi[a] and Mohammad Reza Yeganegi[c]

[a] Department of Water Resources Engineering, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
[b] Department of Water Sciences and Engineering, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
[c] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
*Corresponding author. E-mail: narges.zohrabi@iau.ac.ir

NZ, 0000-0003-2187-3110

ABSTRACT

This study evaluates the potential of two models within the North American Multi-Model Ensemble (NMME) system, i.e., CanCM3 and CanCM4, for improving drought risk management through reliable prediction. By employing the Standardized Precipitation Evapotranspiration Index (SPEI) and gridded datasets (GPCC and CRU), this study assesses their drought forecast capabilities across four semi-arid to arid basins in Iran. The results reveal that both models effectively capture drought events at short lead times (0.5 months), achieving correlation coefficients exceeding 0.93. The performance decline at longer lead times (3.5 months) is less severe in spring and autumn, maintaining correlations of >0.6 compared to summer. A Critical Success Index (CSI) analysis further highlights the models' skill in detecting summer drought events at a 1.5-month lead time (CSI >0.94), underscoring their utility for critical agricultural and water resource planning. Seasonal analysis shows CanCM4 outperforming CanCM3, particularly regarding CSI and correlation stability. These findings offer a novel contribution to understanding the applicability of CanCM3 and CanCM4 for drought forecast purposes in arid and semi-arid basins and underline their value for enhancing drought early warning systems and supporting efficient resource allocation to mitigate drought impacts.

Key words: CanCM3, CanCM4, drought, forecast, NMME, SPEI

HIGHLIGHTS

- The study evaluated North American Multi-Model Ensemble models, i.e., CanCM3 and CanCM4, for drought prediction using the standardized precipitation evapotranspiration index.
- CanCM4 outperformed CanCM3, showing a higher critical success index and more stable correlation.
- The study provides insights into variations in model performance, which are key for enhancing drought warnings and water management.

1. INTRODUCTION

In recent decades, climate change and global warming have significantly increased the frequency of droughts worldwide (Jasim & Awchi 2020; Kesarwani *et al.* 2023; Lotfirad *et al.* 2023). Drought is one of the costliest and least understood natural disasters, devastatingly impacting agriculture, water supply, ecosystems, public health, and basin management. Addressing this complex issue requires effective tools and strategies to mitigate its impact. To this end, researchers and scientists have focused on improving predictive models and developing proactive measures.

Seasonal forecasting models have been utilized to develop drought early warning systems and predict hydroclimatic variables. Advancing these systems is crucial for water resource managers and decision-makers, as reliable drought prediction could lead to more effective basin management and provide a significant socio-economic added value.

Effective risk management requires accurate and timely predictions, enabling proactive measures to mitigate potential impacts. By enhancing early warning systems, stakeholders can better address the risks of droughts and reduce their far-reaching consequences.

The advancement of dynamic climate models, such as the North American Multi-Model Ensemble (NMME) (Kirtman *et al.* 2014), has introduced new opportunities for improving drought forecasting and enhancing decision-making processes in water resource management. These models integrate state-of-the-art climate data and provide multi-lead time forecasts, allowing stakeholders to better prepare for and respond to drought events.

The NMME is a system for monthly-to-seasonal climate prediction that combines multiple global climate models. The NMME comprises real-time initialized predictions and a research database of retrospective and real-time forecasts, which are available for free. It consists of 16 fully coupled global climate models from various institutions in North America (Becker *et al.* 2022). CMC1-CanCM3 and CMC2-CanCM4 (hereafter referred to as CanCM3 and CanCM4) are two important global climate models that contribute to the multi-model ensemble predictions of the NMME system, particularly for temperature and precipitation forecasts. CanCM3 and CanCM4 are coupled climate models developed by the Canadian Centre for Climate Modelling and Analysis (CCCma) (Merryfield *et al.* 2013). The CanCM4 model is an upgraded version of CanCM3 and has been incorporated into the Canadian Seasonal to Interannual Prediction System Version 2 (CanSIPSv2) (Lin *et al.* 2020).

In recent years, numerous studies have evaluated the NMME models' ability to predict temperature and precipitation at the basin scale, including the Northern Hemisphere (Becker & van den Dool 2018), the United States (Tian *et al.* 2014), East Asia (Xu *et al.* 2018), Europe (Slater *et al.* 2017), East Africa (Shukla *et al.* 2014), South Asia (Cash *et al.* 2019), and Iran (Shirvani & Landman 2016; Moghasemi *et al.* 2022; Yazdandoost *et al.* 2023).

Sheffield *et al.* (2014) investigated an online system designed for drought monitoring and seasonal forecasting of water resources. This system utilized microscale (downscaled) temperature and precipitation data from the CFSV2 model to enhance the predictions of drought conditions. The study's findings demonstrated the system's superior performance to traditional statistical methods, notably highlighting its robustness in predicting the severe drought that affected Africa in 2010–2011. Shukla *et al.* (2019) demonstrated the limited skill of the NMME in forecasting precipitation in East Africa. They found that accurate precipitation predictions with a 2-month lead time could only be achieved in northern Ethiopia. Barbero *et al.* (2017) investigated the capability of the NMME in forecasting precipitation across 17 hydroclimatological regions in the United States. Their findings confirmed that each basin's seasonal variations and geographical location influence the forecast performance. Najafi *et al.* (2018) focused on predicting seasonal temperature in Iran. They utilized the Climatic Research Unit (CRU) dataset to evaluate the temperature output of the NMME models. The results highlighted the superior performance of the CSFv2 and GFDL-FLOR-BO5 models compared to others. Slater *et al.* (2019) downscaled the NMME models for precipitation and temperature datasets in the western region of the United States. Their findings revealed that seasonal and geographical changes, not evident in the raw output, could be discerned in the downscaled data, with a spatial resolution of 1°. Dehban *et al.* (2019) utilized NMME models to forecast seasonal precipitation in the Sefidroud River basin in northwestern Iran. Their results indicated that the individual NMME models exhibited inadequate accuracy; however, combining them enhanced the accuracy by 20%. Additionally, they assessed the uncertainty of the precipitation prediction and concluded that the combined model approach significantly reduced uncertainty.

The CanCM3 model has been extensively used in various studies to assess its performance in predicting precipitation and temperature. One study evaluated the performance of CanCM3 in predicting global surface temperature for boreal summer and winter seasons from 1982 to 2011 (Yang *et al.* 2022). The results showed high prediction skills in the tropics, particularly in the equatorial Pacific, while the performance was relatively poor over land areas. Compared to models like JMA/MRI-CPS2 and NCEP CFSv2, CanCM3 exhibited slightly lower prediction skills in the tropics and extratropical regions.

Despite significant advances in drought prediction, there remains a critical research gap in evaluating NMME models' performance in regions with high climate variability, particularly in semi-arid and arid basins. This study addresses this gap through two key contributions: (1) comprehensive assessment of CanCM3 and CanCM4 models from the NMME system across four semi-arid to arid basins in Iran and (2) implementation of the standardized precipitation evapotranspiration index (SPEI) for assessing drought prediction accuracy, which provides a more nuanced drought assessment than the conventional SPI by incorporating both precipitation and evapotranspiration. The findings aim to provide actionable insights for integrating these forecasting tools into drought management frameworks, enhancing resource allocation, and supporting the development of early warning systems. This study contributes to a growing body of research that emphasizes the value of predictive modeling in mitigating the socio-economic impacts of drought.

2. MATERIALS AND METHODS

2.1. The study area

This research focuses on four second-degree basins in western and southwestern Iran (Karkhe, Great Karoon, Hale, and Jarahi and Zohre; Figure 1), which are located between 46° to 52° 30′ east longitude and 28° to 34° north latitude. These basins, encompassing roughly 180,718 km^2, constitute nearly half (42%) of the Persian Gulf and Oman Sea's first-degree basin area. The geographical and climatic characteristics of the studied basins are shown in Table 1.

The Great Karoon and Karkheh basins are among the most important watersheds in Iran. The Great Karoon basin covers an area of 67,257 km^2, while the Karkheh basin spans 51,407 km^2; both are characterized by a semi-arid climate. Three of Iran's major rivers regarding discharge volume (the Karoon, Dez, and Karkheh rivers) are located within these basins. Agriculture represents the largest consumer of water in both regions. In recent years, due to recurring droughts, the Great Karoon and Karkheh basins have faced numerous challenges in meeting water demands for various uses, underscoring the necessity of accurate drought forecasting in these areas.

The Jarahi and Zohreh basin and the Haleh basin are situated south of the Great Karoon and Karkheh basins and are characterized by arid climates. The main rivers within the Jarahi and Zohreh basin, which spans 40,744 km^2, and the

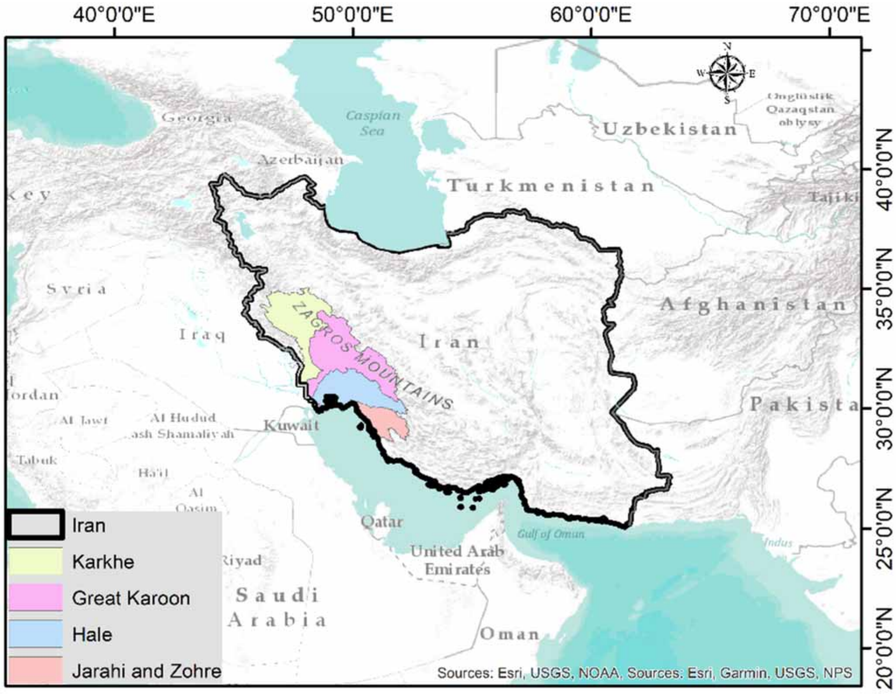

Figure 1 | The study area (including the Karkhe, Great Karoon, Hale, and Jarahi and Zohre basins).

Table 1 | Geographical and climatic characteristics of the studied basins

Basin	Northern latitude (°N)	Eastern longitude (°E)	Annual mean precipitation (mm)	Annual mean maximum temperature (°C)	Annual mean minimum temperature (°C)	Mean elevation (m)	Climate classification
Karkhe	30–35	46–49	383.27	25.20	8.60	1,396.51	Semi-arid
Great Karoon	29–32	48–52	256.76	27.10	12.10	1,752.53	Semi-arid
Hale	28–30	50–52	205.70	30.00	20.30	258.81	Arid
Jarahi and Zohre	28–30	49–51	266.47	32.70	18.03	706.29	Arid

Haleh basin, covering 21,309 km^2, eventually flow into the Persian Gulf. In these basins as well, agriculture is the predominant consumer of water.

The diverse topography of the study area, including mountains, plains, and river valleys, significantly influences the region's hydrology. The semi-arid to arid climate features highly variable precipitation patterns, making all four basins particularly vulnerable to drought. Seasonal precipitation and snowmelt from the Zagros Mountains are crucial water sources feeding the rivers. Given their importance to agriculture, industry, and domestic water use, these basins are vital to the regional socio-economy. Drought-induced variability in water availability can have profound consequences, affecting agricultural productivity, water resource management, and the livelihoods of communities dependent on these basins.

Figure 2 presents the digital elevation model (DEM) map of the study area and the characteristics of the synoptic stations located within the basins. Only six stations – Ahvaz, Abadan, Aligudarz, Borujen, Bostan, and Bushehr – possessed complete data (less than 8% missing) for the 1982–2018 period.

2.2. Precipitation and temperature datasets

This study evaluated the SPEI using gridded precipitation data from the Global Precipitation Climatology Center (GPCC) and minimum and maximum temperature data from the Climatic Research Unit gridded Time Series (CRU TS) for the 1982–2018 period, thereby addressing the limitations of the sparse synoptic station network within the basins.

2.2.1. Global Precipitation Climatology Center

The GPCC provides high-quality, global precipitation data derived from rain gauge measurements. The GPCC Full Data Monthly Product Version 2020 covers the monthly period from 1901 to the end of 2020 with a spatial resolution of 0.5° × 0.5° (Schneider *et al.* 2020). Previous studies have demonstrated that GPCC data exhibit high accuracy in capturing precipitation patterns, as evidenced by strong correlations and low errors compared to observed data (Nassaj *et al.* 2022).

2.2.2. Climatic Research Unit

The CRU, affiliated with the University of East Anglia and renowned for its work on climate change, provides various climate datasets at different resolutions. This study utilizes CRU TS version 4 maximum and minimum temperatures, offering data from 1901 to 2021 at a spatial resolution of 0.5°. Notably, previous studies by Navidi Nassaj *et al.* (2021) and Jafarpour *et al.* (2022) have confirmed the CRU dataset's exceptional effectiveness in accurately estimating air temperature.

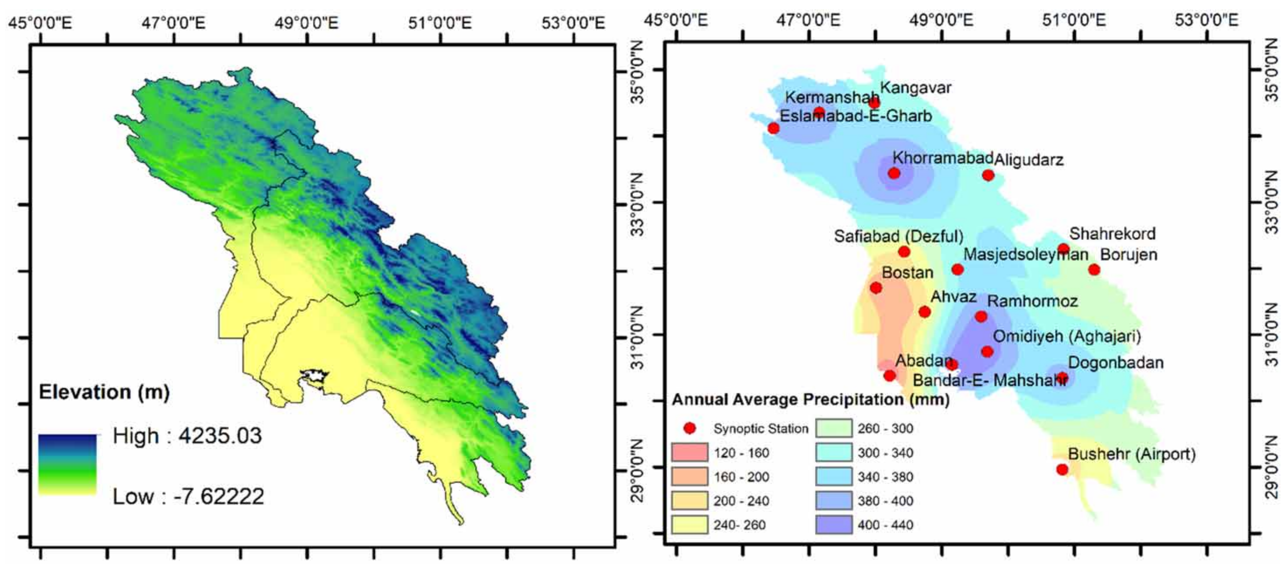

Figure 2 | DEM map of the study area and the characteristics of the synoptic stations located within the basins.

2.3. North American Multi-Model Ensemble

The NMME, a collaborative effort involving various North American institutions, is dedicated to enhancing seasonal climate predictions. This study utilizes two models from the NMME (CanCM3 and CanCM4), developed by the CCCma. With their lead times ranging from 1 to 12 months, these models are well-suited for capturing drought dynamics. CanCM3 incorporates state-of-the-art physical parameterizations for atmospheric processes, ocean dynamics, land-surface processes, and sea ice interactions during its development. CanCM4 builds upon these by incorporating advancements like higher resolution and improved representations of physical processes. For instance, it includes better cloud physics, improved land-surface interactions, and more sophisticated ocean–atmosphere coupling. While the NMME framework allows ensemble modeling by combining outputs from multiple models, this study evaluates the standalone predictive capabilities of CanCM3 and CanCM4. This approach aims to understand these models' individual strengths and limitations before exploring their potential contributions to ensemble predictions. To assess drought conditions, the SPEI was calculated using precipitation and air temperature (minimum and maximum) data from CanCM3 and CanCM4 for 1982–2018. For the monthly analysis, lead times of 0.5 and 3.5 months were selected. Lead times of 1.5 and 3.5 months were chosen for the seasonal evaluation. The selection of these lead times is of utmost importance for meteorological drought forecasting. The 0.5- and 1.5-month lead times provide near-term forecasts crucial for early warning systems, underlining the urgency of the information. This actionable information allows water resource managers, farmers, and policymakers to respond swiftly to emerging drought conditions. The 3.5-month lead time offers crucial information for seasonal planning, informing medium-term strategies for agriculture, water management, and resource allocation. Focusing on these lead times enhances the practical value of the forecasts, ensuring they directly address the needs of stakeholders managing drought impacts.

The CRU dataset and the CanCM3 and CanCM4 model outputs were regridded using bilinear interpolation to match the spatial resolution of the GPCC data (0.5°). Notably, the seasonal scales considered were December to January as winter, March to May as spring, June to August as summer, and September to November as autumn.

2.4. Standardized precipitation evapotranspiration index

The SPEI is a meteorological drought index that utilizes the difference between precipitation and potential evapotranspiration (PET) across various timescales, as expressed by the simple water balance equation (Vicente-Serrano *et al.* 2010). The SPEI has been widely applied to assess drought conditions across different regions (Navidi Nassaj *et al.* 2021; Lotfirad *et al.* 2022; Adib *et al.* 2024).

This study specifically employs the 3-month SPEI to capture medium-term droughts. The calculation of the SPEI involves estimating PET and standardizing the resulting water balance series using a log-logistic probability distribution to generate the SPEI. The negative SPEI indicates drought conditions, while the positive SPEI indicates wet conditions.

The flexibility of the SPEI stems from its combination of precipitation and PET, making it particularly useful in regions where PET plays a significant role in the water balance, such as arid and semi-arid areas. As Vicente-Serrano *et al.* (2010) highlighted, the SPEI can be applied across various geographical and climatic conditions, offering a more comprehensive assessment of drought than indices based solely on precipitation, such as the standardized precipitation index (SPI).

Moreover, studies like that of Lotfirad *et al.* (2022) have shown that the SPEI outperforms the SPI in arid and semi-arid regions of Iran because it considers evapotranspiration. Despite the widespread use of the SPEI for drought assessment, it has not been applied in studies evaluating the performance of NMME models in drought forecast. Therefore, this study aims to fill this gap by exploring the potential of the SPEI in assessing the prediction capability of CanCM3 and CanCM4 models for drought.

2.5. Statistical criteria

This study employs three evaluation metrics to assess the performance of the CanCM3 and CanCM4 models from the NMME system in predicting drought using the SPEI (Table 2). The correlation coefficient (CC) measures the linear relationship between the predicted and observed SPEI values. The root mean squared error (RMSE) indicates the magnitude of the error between predicted and observed values. Finally, the critical success index (CSI) evaluates the models' ability to correctly classify drought events. A contingency table is the foundation for calculating the CSI, highlighting the importance of accurate drought detection in this evaluation process.

Table 2 | Statistical criteria

Index	Equation	Optimal value
CC	$\dfrac{\sum_{i=1}^{n}(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X-\bar{X})^2\sum_{i=1}^{n}(Y-\bar{Y})^2}}$	1
RMSE	$\sqrt{\dfrac{\sum_{i=1}^{n}(X-Y)^2}{n}}$	0
CSI	$\dfrac{\text{Hits}}{\text{Hits}+\text{False Alarm}+\text{Misses}}$	1

Note: X represents the observed SPEI, *Y* represents the SPEI predicted by NMME models, Hits represents the number of correctly predicted drought events, False Alarms represents the number of events predicted as droughts that did not actually occur, and Misses represents the number of drought events that occurred but were not predicted.

3. RESULTS AND DISCUSSION

3.1. Validation of GPCC and CRU datasets against synoptic station data

The initial analysis focused on the agreement between GPCC precipitation and CRU temperature data with corresponding observational records (Figure 3). The results revealed a strong correlation (CC $>$ 0.6) between GPCC precipitation estimates and observed data across most basins during the 1982–2018 period. This positive correlation was particularly evident in the Karkhe, Great Karoon, and Jarahi and Zohre basins. However, some of the Hale basin exhibited a moderate correlation (0.2 $<$ CC $<$ 0.6).

The analysis employed the RMSE to capture average annual precipitation accurately. Extensive areas within the Karkhe, Great Karoon, and Jarahi and Zohre basins displayed RMSE values below 30 mm/year, indicating good agreement between GPCC estimates and observed precipitation. The central and western regions of the Hale basin demonstrated the highest RMSE, suggesting a more considerable discrepancy between estimated and observed precipitation in these areas.

As anticipated, due to its inherently lower spatial variability than precipitation (Figure 3), the CRU temperature data exhibited a strong correlation (0.8–0.96) with observed temperature across the entire study area throughout the 1982–2018 period. The analysis of the RMSE for temperature revealed generally low values (1–2 °C) across most of the basins, particularly in low-elevation regions. Higher elevation areas experienced a maximum RMSE of 4 °C.

This positive evaluation demonstrates the suitability of GPCC precipitation and CRU temperature data for representing observed climate conditions within the study area. The strong correlations and low RMSE values between the gridded datasets and observational records suggest good agreement. Consequently, SPEI values derived from CanCM3 and CanCM4 model outputs (precipitation and temperature) can be reliably compared with the SPEI calculated from GPCC and CRU data for both hindcast (past) and forecast simulations. This comparison allows for assessing the models' ability to capture drought conditions and their potential value for drought prediction.

3.2. Evaluation of NMME-based SPEI and observed-based SPEI

3.2.1. Monthly evaluation

To evaluate the performance of the CanCM3 and CanCM4 models in capturing drought conditions, this study compared the models' SPEI estimation with the observed SPEI, considering lead times and calculating performance metrics for the 1982–2018 period (Figure 4). At a 0.5-month lead time, both models generally aligned with the observed SPEI, although they displayed limitations in capturing the severity of extreme droughts (SPEI < -1.5). For instance, the severe drought with the SPEI of -3.33 observed in 1983 was significantly underestimated by both models. Conversely, the drought with the SPEI of -1.7 in 2001 was overestimated. Both models showed similar performance during the forecast period, but CanCM3 accurately estimated the 2013 drought, while CanCM4 underestimated it.

At the 3.5-month lead time, the models tended to both overestimate drought onset and underestimate the severity of extreme droughts. This earlier prediction of drought conditions, combined with an inability to capture extreme events, could compromise their utility for long-term drought prediction.

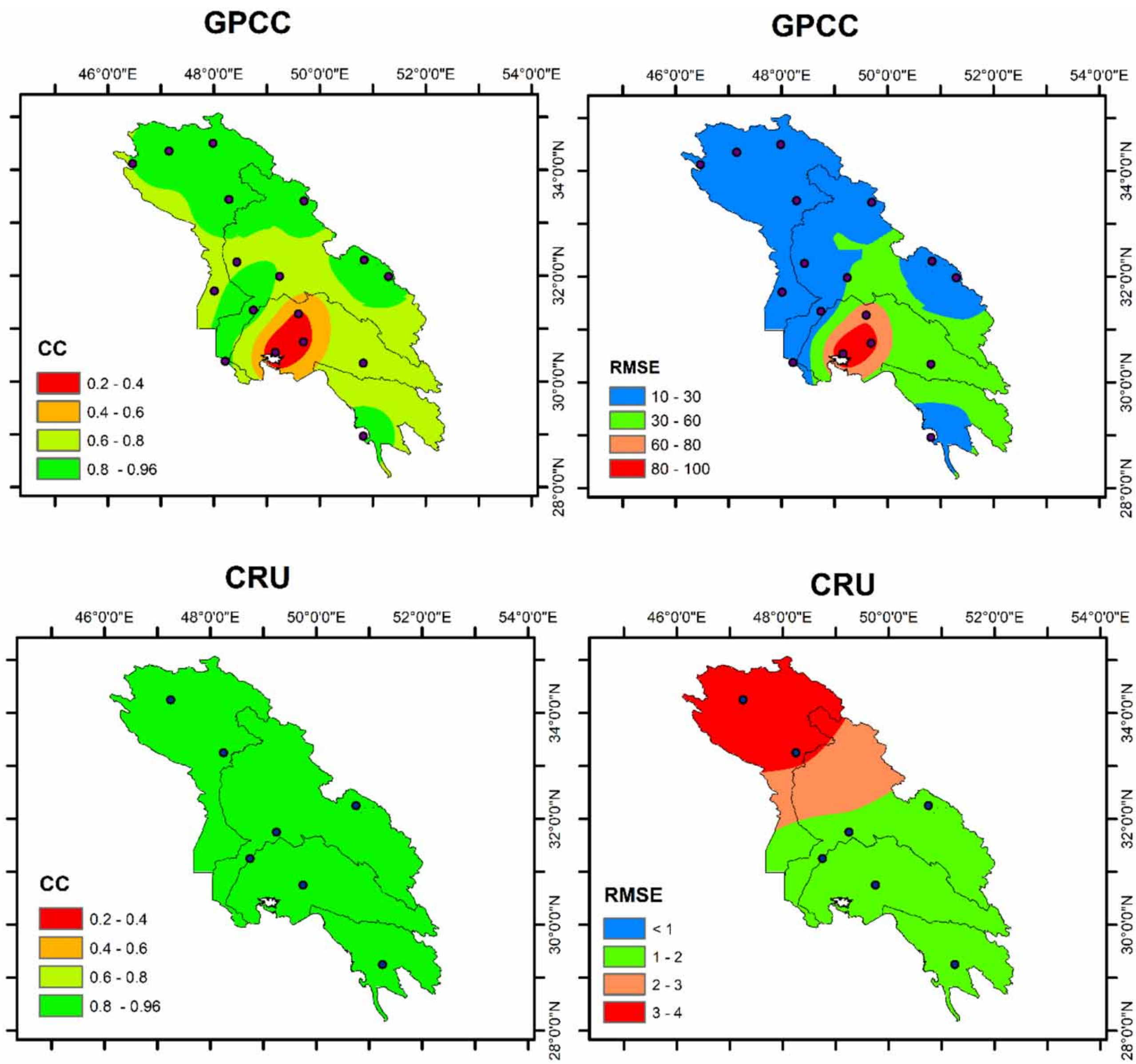

Figure 3 | The performance of GPCC precipitation data (top) and CRU average air temperature data (bottom) compared to gauge observations in the study area. Blue dots represent stations for the precipitation/temperature variables.

Figure 5 demonstrates the spatial monthly performance (CC, RMSE, and CSI) of the CanCM3 and CanCM4 models in the study area for the lead times of 0.5 and 3.5, respectively.

Performance metrics revealed good agreement between the observed and modeled SPEI at the 0.5-month lead time. Both CanCM3 and CanCM4 exhibited strong correlations (average $CC = 0.92$) and low RMSE values ($RMSE < 0.6$ mm), indicating the accurate capture of drought conditions. The high CSI values (above 0.8) across most of the study areas further supported this finding, suggesting that the models effectively detected drought events. These findings align closely with the results of Kondal *et al.* (2024), who highlighted the overall skill of NMME models in providing seasonal forecasts that inform agricultural decision-making. Their study noted that NMME models showed reliable predictions for the short-term water balance.

However, performance significantly declined in the 3.5-month lead time. The CCs dropped to an average value of 0.13 for CanCM3 and 0.18 for CanCM4, indicating a substantial decrease in agreement with observations. This decline was particularly pronounced in elevated regions. Furthermore, the CSI values decreased considerably to 0.16 for CanCM3 and 0.2 for

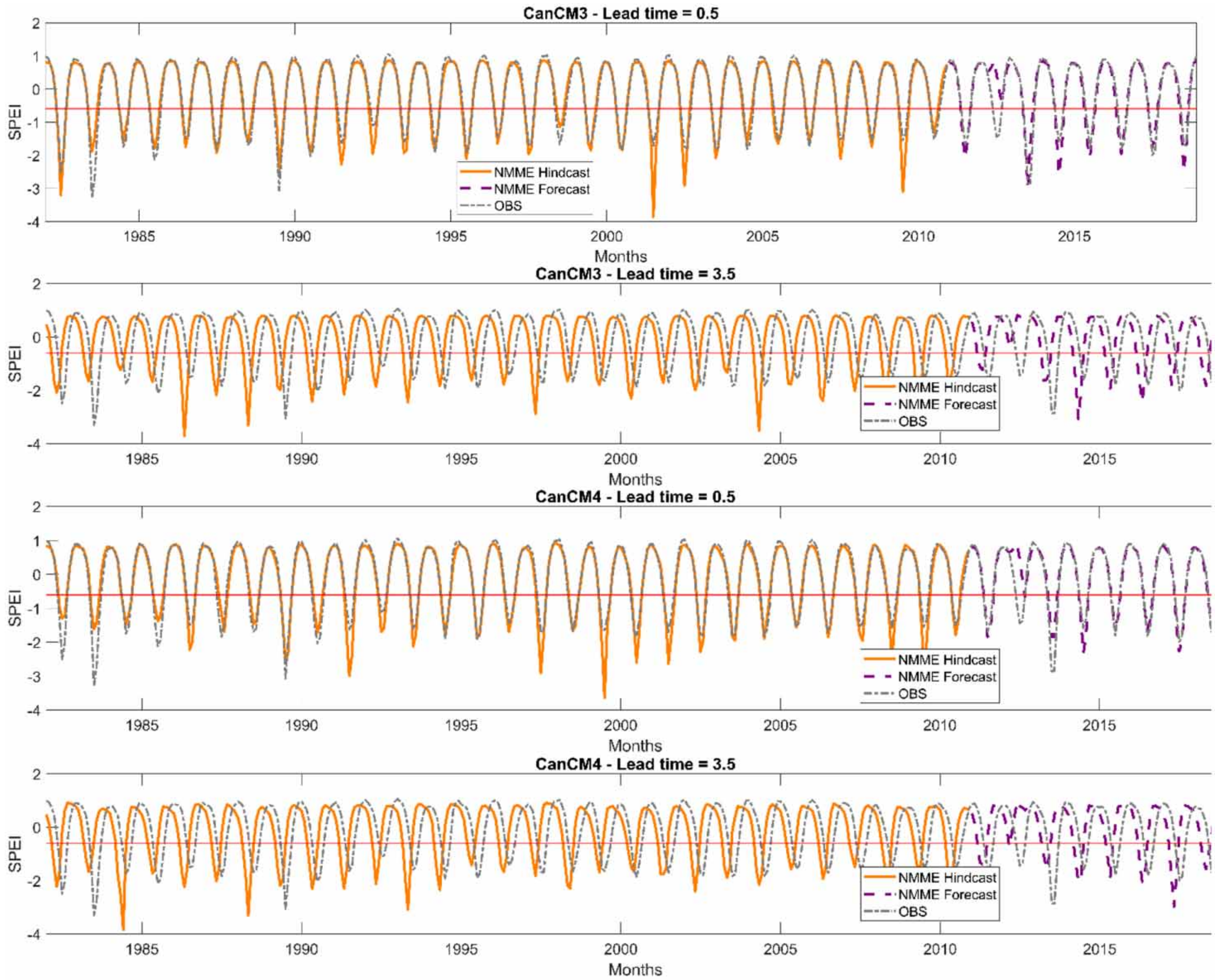

Figure 4 | CanCM3-based SPEI and CanCM4-based SPEI against the observed SPEI.

CanCM4, highlighting a decline in the model's ability to predict drought events at this longer lead time accurately. The observed decline in skill at extended lead times is consistent with the findings of Yazdandoost *et al.* (2023), who evaluated NMME precipitation forecasts over Iran's Karoon basin. Their study similarly highlighted a substantial drop in correlation beyond 2 months, attributing this to the models' inherent limitations in simulating the progression of drought events under complex climatic and hydrological conditions. Our study's performance disparity between short- and long-term forecasts reinforces the need for integrating higher-resolution datasets and ensemble approaches to improve extended forecasts.

3.2.2. Seasonal evaluation

Figures 6 and 7 reveal pronounced seasonal variations in the performance of the CanCM3 and CanCM4 models for SPEI prediction across different lead times.

At the 1.5-month lead time, both models exhibited the strongest CC (>0.7) with the observed SPEI during spring and autumn across the entire basin. This strong CC suggests good agreement between these seasons' modeled and observed drought conditions. Winter performance was moderate ($0.3 < CC < 0.7$), with higher agreement in the southern and south-western regions. Summer displayed the weakest performance, with CanCM3 showing a slight advantage in the southeast ($CC = 0.7$) but declining toward the north and northwest ($0.3 < CC < 0.4$). This difference in seasonal performance highlights the influence of spatial variability on model performance.

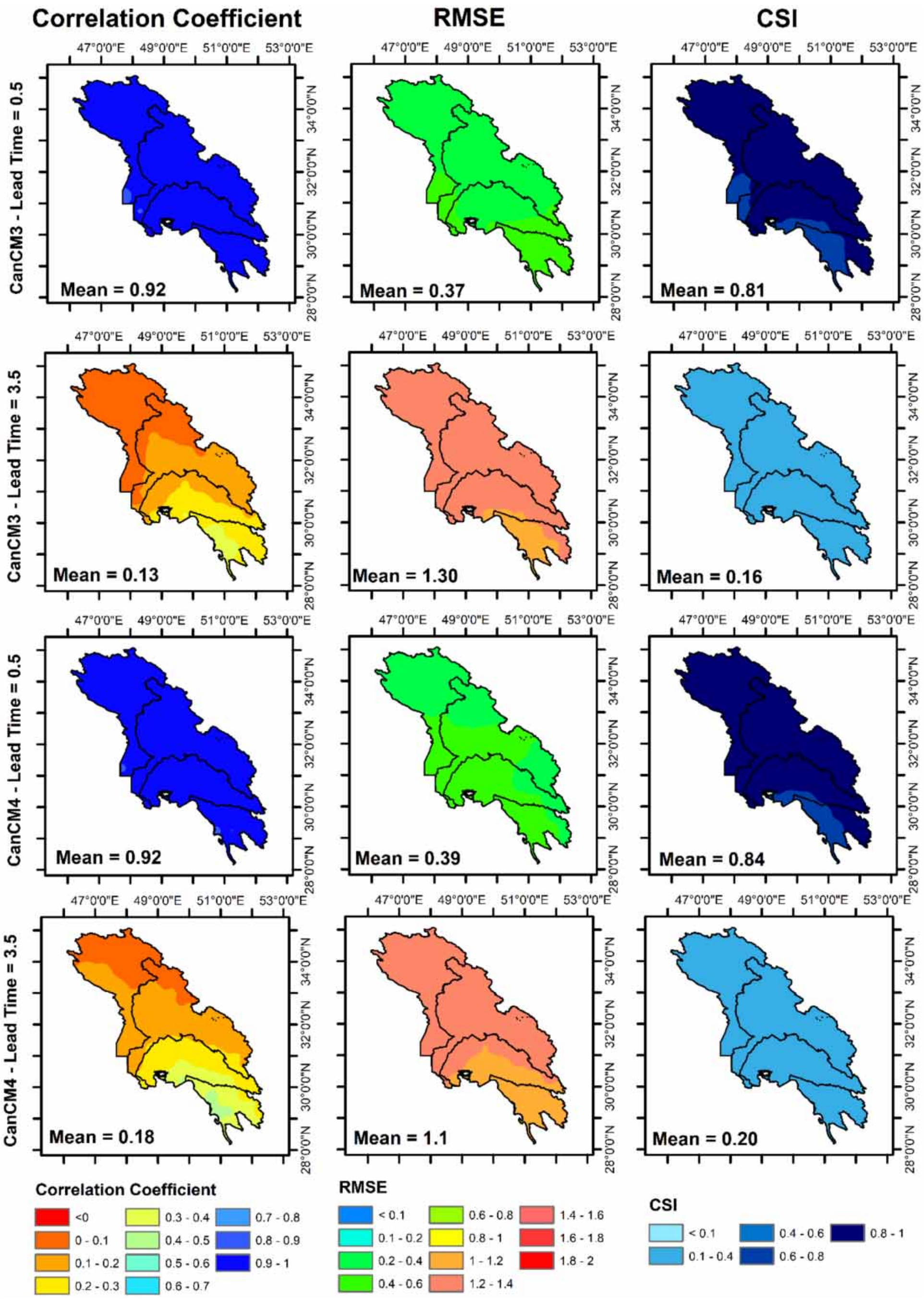

Figure 5 | Spatial performance of CanCM3 and CanCM4 models to detect the SPEI for 0.5- and 3.5-month lead times.

Performance significantly declined at the 3.5-month lead time, particularly for summer and winter, highlighting the variability in model performance across seasons. This variability echoes the findings of Turner *et al.* (2024), who analyzed the NMME's subseasonal precipitation prediction over sub-Saharan Africa. Turner *et al.* observed that regional characteristics and inter-model variability significantly influenced prediction accuracy, particularly in drought-prone areas. This is comparable to the lower performance of CanCM3 and CanCM4 in summer, as seen in the present study, highlighting the models' sensitivity to spatial and climatic variations.

CC for both models dropped below 0.1 in summer, indicating a near inability to capture drought trends. In winter and autumn, CanCM3 displayed marginally better performance than CanCM4 at this longer lead time. This seasonal disparity

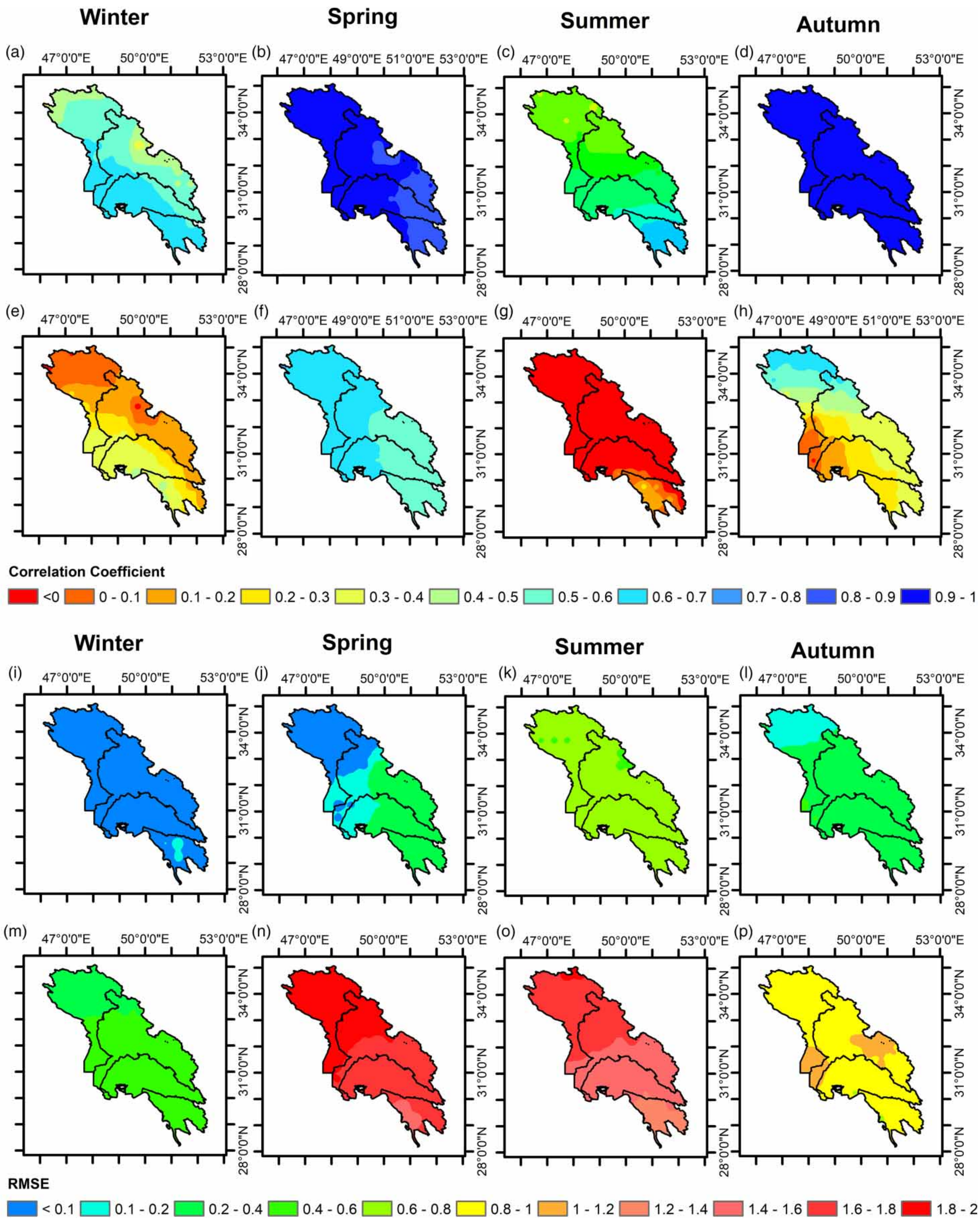

Figure 6 | The performance of CanCM3 in seasonal drought prediction based on the SPEI (a–d: correlation at a 1.5-month lead time; e–h: correlation at a 3.5-month lead time; i–l: RMSE at a 1.5-month lead time; and m–p: RMSE at a 3.5-month lead time).

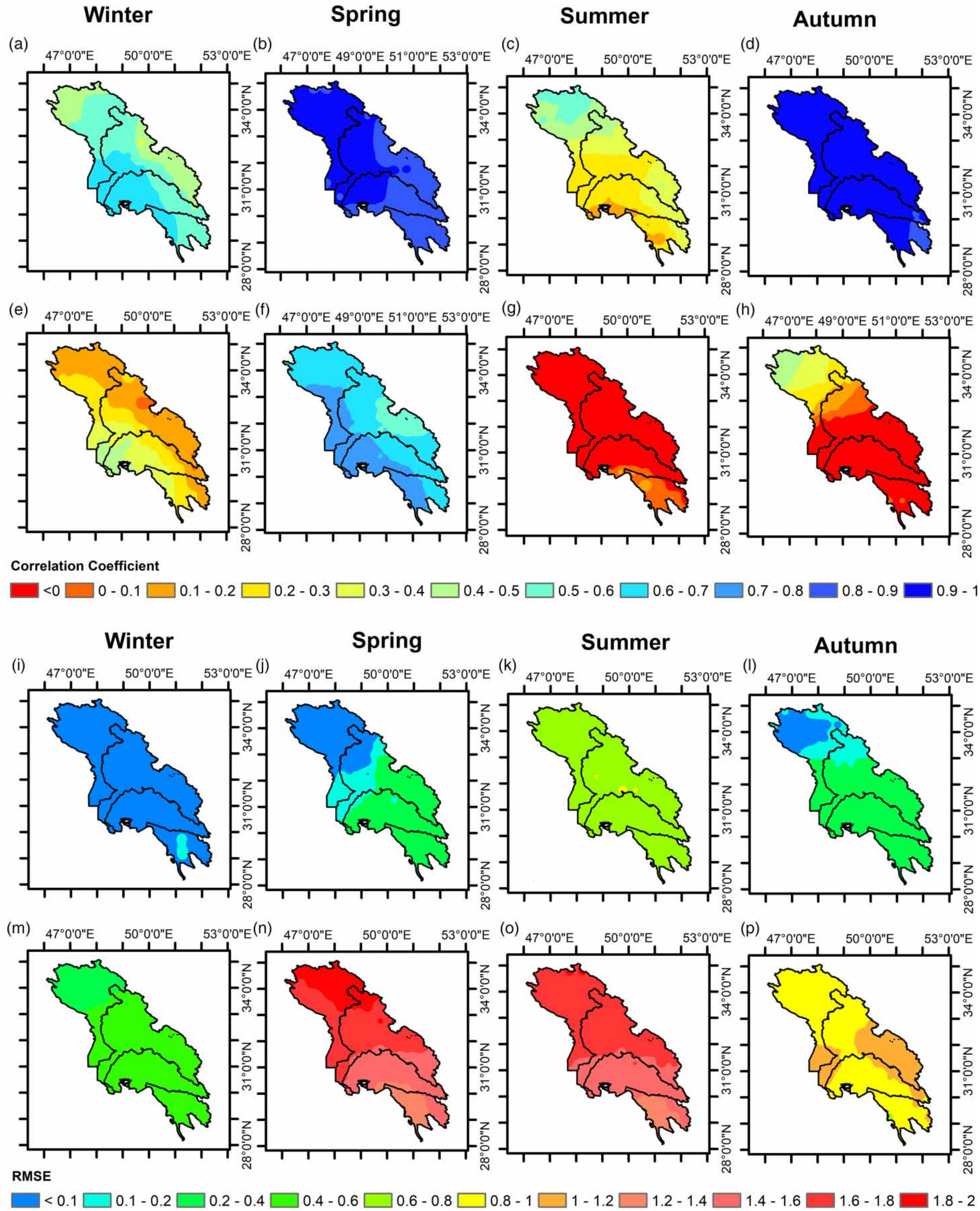

Figure 7 | The performance of CanCM4 in seasonal drought prediction based on the SPEI (a–d: correlation at a 1.5-month lead time; e–h: correlation at a 3.5-month lead time; i–l: RMSE at a 1.5-month lead time; and m–p: RMSE at a 3.5-month lead time).

aligns with findings by Seager *et al.* (2020), who investigated the onset and termination of seasonal droughts using NMME outputs at 2.5–4.5-month lead times. Their study emphasized significant seasonal and regional variability, noting reduced predictive accuracy during summer and comparatively better performance during spring and autumn. These patterns underscore seasonal atmospheric dynamics' influence on model performance, a factor observed consistently across multiple studies.

RMSE values generally reflect the seasonal trends in correlation. Both models exhibited the lowest RMSE (<0.1 mm) in winter at the 1.5-month lead time, indicating high accuracy. This value increased to 0.4 mm at the 3.5-month lead time, suggesting acceptable performance. Spring and autumn displayed a decline in accuracy with increasing lead time, with the RMSE exceeding 1.6 mm for the 3.5-month lead time in spring. Summer exhibited the highest RMSE, likely due to naturally lower precipitation, where minor discrepancies can significantly impact the error metric.

At the 1.5-month lead time, both CanCM3 and CanCM4 demonstrated high skill in detecting drought events during summer, as indicated by CSI values ranging from 0.8 to 1.0 (Figure 8). These results suggest a strong agreement between model predictions and observed drought occurrences. However, calculating the CSI for the 3.5-month lead time and for the winter and spring seasons was impossible due to limitations in the data used for evaluation. A CSI of Not a Number (NaN) or zero can indicate issues such as no predicted or actual drought events (leading to NaN due to division by zero in the CSI calculation) or a complete lack of accurate predictions (leading to a CSI of 0). These limitations often arise from data imbalance, model limitations in capturing drought events at longer lead times or specific seasons, or insufficient drought occurrences during the evaluation period. In this case, the limited number of drought events in most grids during winter and spring likely prevented a robust assessment of model performance using the CSI metric.

Table 3 summarizes the average performance metrics (CC, RMSE, and CSI) across the study area for different lead times and seasons. Both CanCM3 and CanCM4 displayed generally similar performances across most evaluation criteria. Notably, spring exhibited the strongest correlations (average CC > 0.7) between the observed and modeled SPEI for both models at the 1.5-month lead time (details in Table 3).

In evaluating the performance of models for drought forecasting, a critical aspect is comparing their performance during hindcast and forecast periods. Table 4 presents the results of performance indices for these periods, indicating no significant differences between them. Furthermore, both models exhibit remarkably similar performance levels, as evidenced by the CC of 0.89 for both models at the lead time of 0.5 months. This finding suggests that further investigations could focus on the

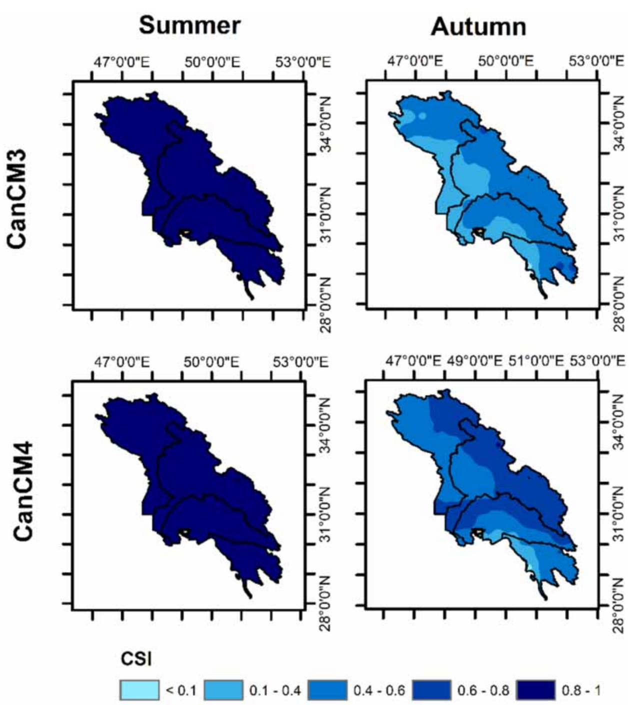

Figure 8 | The CSI of CanCM3 and CanCm4 for summer and autumn (lead time = 1.5).

Table 3 | Seasonal performance indices of NMME models per regional average

Model	Seasonal period	CC	RMSE (mm)	CSI
1.5-month lead time				
Cancm3	Overall	0.92	0.3	0.81
	Winter	0.56	0.11	–
	Spring	0.91	0.19	0.25
	Summer	0.41	0.67	0.95
	Autumn	0.93	0.23	0.43
Cancm4	Overall	0.91	0.39	0.84
	Winter	0.54	0.1	–
	Spring	0.9	0.19	0.14
	Summer	0.34	0.7	0.94
	Autumn	0.93	0.23	0.58
3.5-month lead time				
Cancm3	Overall	0.13	1.3	0.16
	Winter	0.19	0.46	–
	Spring	0.61	1.7	0.03
	Summer	−0.3	1.5	0.39
	Autumn	0.36	0.9	0
Cancm4	Overall	0.18	1.2	0.2
	Winter	0.23	0.47	–
	Spring	0.67	1.6	0.04
	Summer	−0.2	1.6	0.36
	Autumn	0.05	0.96	0

Note: CSI values are not available for some seasons due to a lack of drought events or limitations in model predictions, leading to a division-by-zero error in the calculation.

Table 4 | Performance of CanCM3 and CanCM4 models for the hindcast and forecast periods

Indices Model	CC Hindcast (mm)	RMSE	CSI	CC Forecast (mm)	RMSE	CSI
0.5-month lead time						
CanCM3	0.93	0.34	0.84	0.89	0.46	0.72
CanCM4	0.92	0.38	0.85	0.89	0.43	0.80
3.5-month lead time						
CanCM3	0.15	1.30	0.17	0.08	1.3	0.13
CanCM4	0.19	1.26	0.26	0.11	1.27	0.15

overall period without separating the evaluation into hindcast and forecast periods. Similar trends were observed in the study by Yazdandoost *et al.* (2023), where CanCM3 and CanCM4 exhibited comparable performance across these periods.

Figure 9 employs Taylor diagrams to visualize the models' performance across seasons and lead times. Figure 9 shows that at a lead time of 1.5, the highest correlation occurred in spring and autumn (CC > 0.9 in both models), and the standard deviation in both models was close to that of observations. In contrast, the correlation decreased in winter and summer, and the standard deviations diverged from the observations. At a lead time of 3.5, the best performance was in spring, with correlations above 0.6 for both models. However, this correlation was significantly lower than the lead time of 1.5. In other seasons, particularly summer, the models did not perform well at a lead time of 3.5; for instance, the correlation in winter was less than 0.3. Nevertheless, it is noteworthy that the performance of both the CanCM3 and CanCM4 models was very similar across all seasons and both lead times.

Similarly, Yazdandoost *et al.* (2020) found that CanCM4 performed slightly better than CanCM3 in capturing spatial variability and precipitation patterns. However, both models struggled with reduced accuracy in summer and winter, especially at

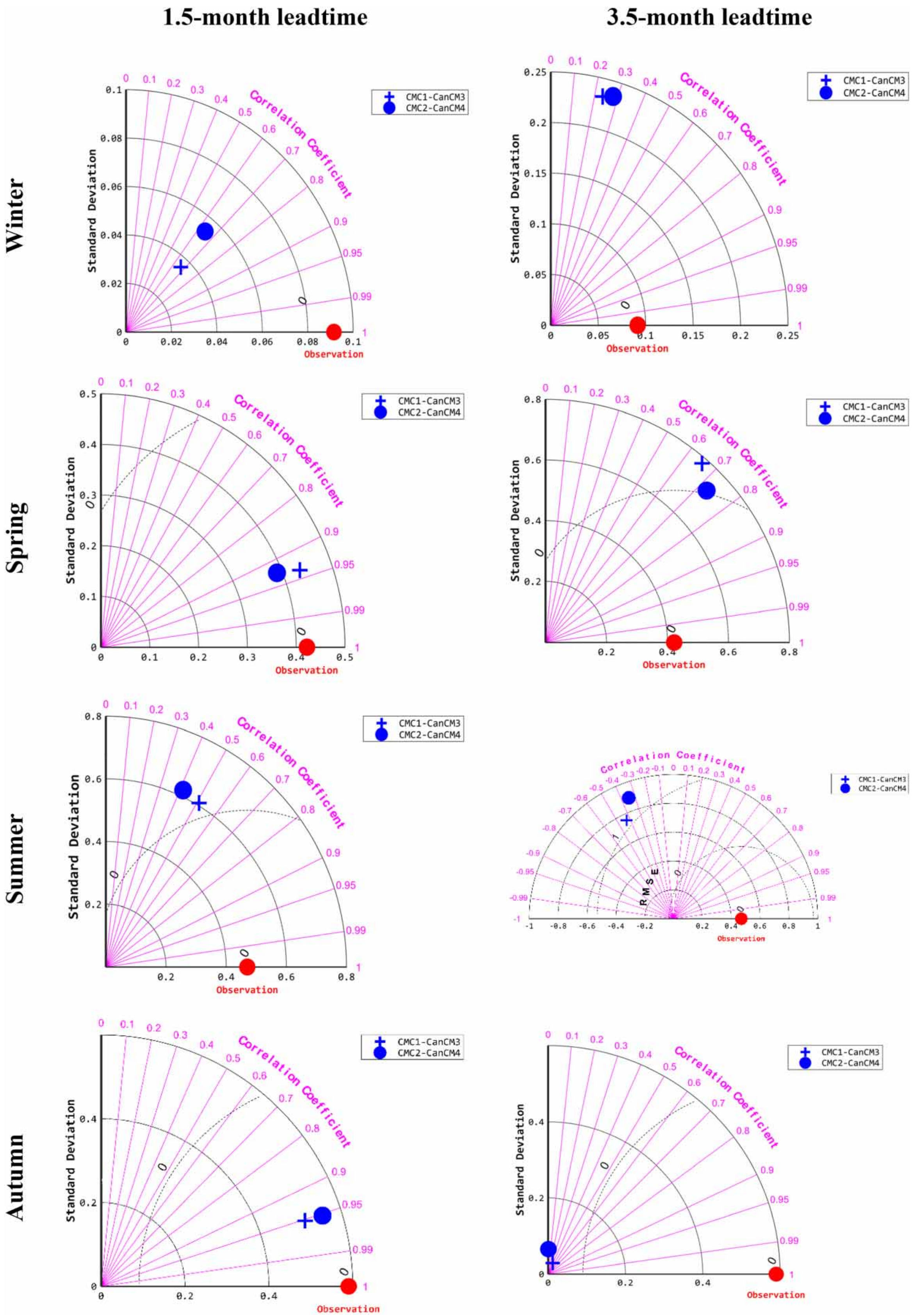

Figure 9 | Taylor diagram depicting the seasonal performance of CanCM3 and CanCM4 models in detecting the observational SPEI.

extended lead times. This consistency highlights the inherent strengths and limitations of the models, emphasizing their better performance in short-term forecasts during transitional seasons like spring and autumn.

The seasonal performance of the CanCM3 and CanCM4 models, as demonstrated in this study, aligns with the findings of several previous investigations. At a lead time of 1.5 months, both models achieved their highest correlation in spring and autumn (CC > 0.9), which indicates their ability to capture seasonal drought dynamics during transitional periods. This pattern is consistent with Barbero *et al.* (2017), who reported that NMME models generally perform better during transitional seasons (spring and autumn). However, the observed decline in correlation during summer and winter, particularly at a lead time of 3.5 months (e.g., CC < 0.3 in winter), mirrors findings by Shukla *et al.* (2019), who noted significant variability in NMME model skill during extreme seasons across East Africa. Shukla *et al.* emphasized that summer months, characterized by low precipitation and high evapotranspiration, challenge models because of the heightened sensitivity to minor errors in precipitation forecasts. Similarly, in our study, the reduced summer performance (e.g., CanCM3 correlation dropping below 0.1) underscores the limitations of the models in capturing the complex interplay of climatic variables during drier periods in semi-arid basins.

The ability of the models to achieve better accuracy in spring and autumn may also be attributed to the dominant role of climatic indices such as the El Niño-Southern Oscillation (ENSO), which exerts a more predictable influence during these seasons. Slater *et al.* (2019) demonstrated that the NMME model skill is highly dependent on the seasonality of large-scale climate drivers. Furthermore, as Moradian & Yazdandoost (2021) highlighted in their analysis of Iran's basins, the topographical heterogeneity of semi-arid regions exacerbates spatial prediction errors, particularly during winter. The reduced performance of the models during winter in the present study may therefore be partially attributed to the coarse spatial resolution of the models, which limits their capacity to resolve localized effects such as terrain-induced precipitation variability.

Although CanCM3 and CanCM4 provide valuable short-term predictions for drought monitoring in the study area, their performance diminishes with increasing lead time and during extreme climatic seasons. Addressing these limitations will require further refinement of model physics and the incorporation of additional predictors, such as soil moisture and surface temperature anomalies, as suggested by Shukla *et al.* (2019) and Barbero *et al.* (2017).

While studies such as that by Yao & Yuan (2018) have demonstrated significant advancements through bias correction and hydrological model integration, this study focuses on evaluating the raw outputs of the NMME models (CanCM3 and CanCM4) using the SPEI, which is tailored for semi-arid regions. Future research could build upon this foundational analysis by incorporating techniques like bias correction and ensemble hydrological modeling to enhance forecast accuracy.

4. CONCLUSION

This study comprehensively evaluates the CanCM3 and CanCM4 models from the NMME for forecasting meteorological droughts in semi-arid to arid basins. By employing the SPEI, the research offers valuable insights into the models' capabilities, limitations, and practical implications for drought risk management. Both models demonstrated high skill (correlation exceeding 0.93) in capturing short-term drought events (0.5-month lead time). The seasonal analysis further revealed strong performance in spring and autumn (CC > 0.7), making these models particularly valuable for planning during these critical periods. The slightly superior performance of CanCM4, as indicated by its higher CSI, highlights its applicability for more accurate and stable drought predictions.

Despite their strengths, both models exhibited declining accuracy with longer lead times (3.5 months). This limitation hinders their ability to predict complex drought dynamics and extreme events effectively. Moreover, spatial variations in performance, particularly during summer, underscore the need for region-specific model calibration and seasonally tailored approaches. Addressing these challenges will require incorporating additional climatic and hydrological variables, improving model resolution, and employing advanced statistical or machine-learning techniques to refine long-term forecasts.

4.1. Limitations and future research directions

The findings of this study emphasize the practical value of integrating the NMME models into drought risk management frameworks. By providing reliable short-term forecasts and highlighting seasonal patterns, these models can support the development of early warning systems, optimize resource allocation, and reduce the socio-economic impacts of droughts. Their integration into decision-making processes can enhance the resilience of communities and sectors dependent on water resources, particularly agriculture and energy.

However, this study has several limitations that warrant attention. First, while the evaluated lead times provided critical insights into the NMME models' performance for short- and medium-term drought prediction, investigating additional lead times (e.g., 2.5 and 4.5 months) could offer a more comprehensive understanding of drought dynamics. Second, the exclusive use of the 3-month SPEI drought index, though adequate for medium-term droughts, may not fully capture the models' performance in predicting short-term (1-month) or long-term (6- or 12-month) drought conditions. Third, broader applicability could be achieved through the adoption of multi-model ensemble approaches, which integrate outputs from models to enhance reliability and robustness. Finally, although the SPEI was utilized in this study to assess drought conditions, future investigations could involve employing other indices, such as agricultural or hydrological drought indices.

Future research should address these limitations. One key direction is the integration of statistical forecast models with the NMME models, as previous research by Xu *et al.* (2018) has shown that statistical drought prediction methods can outperform dynamic drought prediction in some regions. This hybrid approach could leverage the strengths of both methodologies to enhance prediction accuracy. Additionally, investigating the impact of elevation on prediction performance, particularly in regions with complex terrain, could provide valuable insights into topographic influences on NMME drought prediction. Furthermore, evaluating the effects of bias correction on NMME precipitation and temperature data, especially when calibrated with ground-based observations, could significantly reduce uncertainties and improve forecast accuracy.

ETHICAL APPROVAL

The manuscript is original work with its own merit, has not been previously published in whole or in part, and is not being considered for publication elsewhere.

CONSENT TO PARTICIPATE

M.M., N.Z., H.F., A.N.S. and M.Y. have read the final manuscript, have approved the submission to the journal, and have accepted full responsibilities pertaining to the manuscript's delivery and contents.

CONSENT TO PUBLISH

M.M., N.Z., H.F., A.N.S. and M.Y. agree to publish this manuscript upon acceptance.

AUTHORS' CONTRIBUTIONS

M.M., N.Z., H.F., A.N.S. and M.Y. declare that they have contributed to the preparation of this manuscript.

FUNDING

The authors did not receive support from any organization for the submitted work.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Adib, A., Soleimani, M., Ashrafi, S. M., Lotfirad, M., Sofla, M. S. D. & Liaghat, A. (2024) Monitoring annual meteorological drought in arid and semi-arid watersheds by SPI12 drought index and spatial autocorrelation pattern analysis: a case study of the Khuzestan province, Southwest Iran. *Sustainable Water Resources Management*, **10** (5), 160.

Barbero, R., Abatzoglou, J. T. & Hegewisch, K. C. (2017) Evaluation of statistical downscaling of North American multimodel ensemble forecasts over the Western United States, *Weather and Forecasting*, **32** (1), 327–341.

Becker, E. J. & Van den Dool, H. M. (2018) 'Assessing Probabilistic Forecasts of S2S Climate Extremes Using the NMME', *AGU Fall Meeting Abstracts*, Vol. 2018, pp. A33E–A05B.

Becker, E. J., Kirtman, B. P., L'Heureux, M., Muñoz, Á. G. & Pegion, K. (2022) A decade of the North American Multimodel Ensemble (NMME): research, application, and future directions, *Bulletin of the American Meteorological Society*, **103**, E973.

Cash, B. A., Manganello, J. V. & Kinter, J. L. (2019) Evaluation of NMME temperature and precipitation bias and forecast skill for South Asia, *Climate Dynamics*, **53**, 7363–7380.

Dehban, H., Ebrahimi, K., Araghinejad, S. & Bazrafshan, J. (2019) Evaluation of NMME models in forecasting of monthly rainfall (case study: Sefidrood Basin), *Journal of Agricultural Meteorology*, **7** (1), 3–12.

Jafarpour, M., Adib, A. & Lotfirad, M. (2022) Improving the accuracy of satellite and reanalysis precipitation data by their ensemble usage, *Applied Water Science*, **12** (9), 232.

Jasim, A. I. & Awchi, T. A. (2020) Regional meteorological drought assessment in Iraq, *Arabian Journal of Geosciences*, **13** (7), 284. doi:10.1007/s12517-020-5234-y.

Kesarwani, M., Neeti, N. & Chowdary, V. M. (2023) Evaluation of different gridded precipitation products for drought monitoring: a case study of Central India, *Theoretical and Applied Climatology*, **151** (1), 817–841.

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., Rienecker, M., Suarez, M., Li, Z. E., Marshak, J., Lim, Y-K., Tribbia, J., Pegion, K., Merryfield, W. J., Denis, B. & 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.

Kondal, A., Hegewisch, K., Liu, M., Abatzoglou, J. T., Adam, J. C., Nijssen, B. & Rajagopalan, K. (2024) Seasonal forecasts have sufficient skill to inform some agricultural decisions, *Environmental Research Letters*, **19** (12), 124049. https://doi.org/10.1088/1748-9326/ad8bde.

Lin, H., Merryfield, W. J., Muncaster, R., Smith, G. C., Markovic, M., Dupont, F., Roy, F., Lemieux, J. F., Dirkson, A., Kharin, V. V. & Lee, W. S. (2020) The Canadian seasonal to interannual prediction system version 2 (CanSIPSv2), *Weather and Forecasting*, **35** (4), 1317–1343.

Lotfirad, M., Esmaeili-Gisavandani, H. & Adib, A. (2022) Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran, *Journal of Water and Climate Change*, **13** (2), 383–406.

Lotfirad, M., Adib, A., Riyahi, M. M. & Jafarpour, M. (2023) Evaluating the effect of the uncertainty of CMIP6 models on extreme flows of the Caspian Hyrcanian forest watersheds using the BMA method, *Stochastic Environmental Research and Risk Assessment*, **37** (2), 491–505.

Merryfield, W. J., Lee, W. S., Wang, W., Chen, M. & Kumar, A. (2013) Multi-system seasonal predictions of Arctic sea ice, *Geophysical Research Letters*, **40** (8), 1551–1556.

Moghasemi, M., Zohrabi, N., Fathian, H., Nikbakht Shahbazi, A. & Yeganegi, M. (2022) Drought prediction using North American Multimodel Ensemble (NMME) over Western Regions of Iran, *Journal of Water and Soil Resources Conservation*, **11** (4), 25–40.

Moradian, S. & Yazdandoost, F. (2021) Seasonal meteorological drought projections over Iran using the NMME data, *Natural Hazards*, **108**, 1089–1107.

Najafi, H., Massah Bavani, A. R., Irannejad, P. & Robertson, A. W. (2018) Evaluation of NMME seasonal temperature forecasts over Iran's river basins, *Journal of Agricultural Meteorology*, **6** (1), 19–30.

Nassaj, B. N., Zohrabi, N., Shahbazi, A. N. & Fathian, H. (2022) Evaluating the performance of eight global gridded precipitation datasets across Iran, *Dynamics of Atmospheres and Oceans*, **98**, 101297.

Navidi Nassaj, B., Zohrabi, N., Shahbazi, A. N. & Fathian, H. (2021) Evaluation of three gauge-based global gridded precipitation datasets for drought monitoring over Iran, *Hydrological Sciences Journal*, **66** (14), 2033–2045. doi:10.1080/02626667.2021.1978444.

Schneider, U., Becker, A., Finger, P., Rustemeier, E. & Ziese, M. (2020) *GPCC Full Data Monthly Product Version 2020 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges Built on GTS-Based and Historical Data*. Offenbach, Germany: Deutscher Wetterdienst (DWD). doi:10.5676/DWD_GPCC/FD_M_V2020_050.

Seager, R., Nakamura, J. & Ting, M. (2020) Prediction of seasonal meteorological drought onset and termination over the southern Great Plains in the North American Multimodel ensemble, *Journal of Hydrometeorology*, **21** (10), 2237–2255. https://doi.org/10.1175/jhm-d-20-0023.1.

Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S. & Ogallo, L. (2014) A drought monitoring and forecasting system for sub-Sahara African water resources and food security, *Bulletin of the American Meteorological Society*, **95** (6), 861–882.

Shirvani, A. & Landman, W. A. (2016) Seasonal precipitation forecast skill over Iran, *International Journal of Climatology*, **36** (4), 1887–1900.

Shukla, S., McNally, A., Husak, G. & Funk, C. (2014) A seasonal agricultural drought forecast system for food-insecure regions of East Africa, *Hydrology and Earth System Sciences*, **18** (10), 3907–3921.

Shukla, S., Roberts, J., Hoell, A., Funk, C. C., Robertson, F. & Kirtman, B. (2019) Assessing North American multimodel ensemble (NMME) seasonal forecast skill to assist in the early warning of anomalous hydrometeorological events over East Africa, *Climate Dynamics*, **53**, 7411–7427.

Slater, L. J., Villarini, G. & Bradley, A. A. (2017) Weighting of NMME temperature and precipitation forecasts across Europe, *Journal of Hydrology*, **552**, 646–659.

Slater, L. J., Villarini, G. & Bradley, A. A. (2019) Evaluation of the skill of North-American Multimodel Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA, *Climate Dynamics*, **53**, 7381–7396.

Tian, D., Martinez, C. J., Graham, W. D. & Hwang, S. (2014) Statistical downscaling multimodel forecasts for seasonal precipitation and surface temperature over the southeastern United States, *Journal of Climate*, **27** (22), 8384–8411.

Turner, W., Shukla, S. & Husak, G. (2024) *NMME Experimental Subseasonal Precipitation Forecasts (SubX) Provide Enhanced Predictions of End-of-Season Water Balance Over sub-Saharan Africa*. EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-13863. https://doi.org/10.5194/egusphere-egu24-13863.

Vicente-Serrano, S., Beguería, S. & López-Moreno, J. I. (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *Journal of Climate*, **23**, 1696–1718. 10.1175/2009JCLI2909.1.

Xu, L., Chen, N., Zhang, X. & Chen, Z. (2018) An evaluation of statistical, NMME and hybrid models for drought prediction in China, *Journal of Hydrology*, **566**, 235–249.

Yang, Y., Sun, W., Zou, M., Qiao, S. & Li, Q. (2022) Multi-model seasonal prediction of global surface temperature based on partial regression correction method, *Frontiers in Environmental Science*, **10**, 1–15. https://doi.org/10.3389/fenvs.2022.1036006.

Yao, M. & Yuan, X. (2018) Evaluation of summer drought ensemble prediction over the Yellow River basin, *Atmospheric and Oceanic Science Letters*, **11** (4), 314–321. https://doi.org/10.1080/16742834.2018.1484253.

Yazdandoost, F., Moradian, S., Zakipour, M., Izadi, A. & Bavandpour, M. (2020) Improving the precipitation forecasts of the North-American Multi-Model Ensemble (NMME) over Sistan basin, *Journal of Hydrology*, **590**, 125263. https://doi.org/10.1016/j.jhydrol.2020.125263.

Yazdandoost, F., Zakipour, M. & Izadi, A. (2023) Statistical refinement of the North American Multi-Model Ensemble precipitation forecasts over Karoon basin, Iran, *Journal of Water and Climate Change*, **14** (8), 2517–2530.

First received 10 November 2024; accepted in revised form 7 February 2025. Available online 24 February 2025