



A New Ensemble Index for Extracting Predictable Drought Features from Multiple Historical Simulations of Climate

SUN YUANBIN

SADIA QAMAR

ZULFIQAR ALI

TAO YANG

AMNA NAZEER

RABIA FAYYAZ

**Author affiliations can be found in the back matter of this article*

ORIGINAL RESEARCH
PAPER



STOCKHOLM
UNIVERSITY PRESS

ABSTRACT

Drought assessment and forecasting under an ensemble of multiple climate simulation models play important role in early warning drought mitigation policies. This research provides a new ensemble drought measure – the Multivariate Multi-Scaler Forecastable Standardized Drought Index (MMFSDI). At a particular georeferenced point, the MMFSDI uses time-series data of precipitation from multiple climate simulation models for the characterization of drought. The methodology of MMFSDI is mainly based on Forecastable Component Analysis (FCA) and K-Component Gaussian Mixture Distribution (K-CGMD). In application, historical simulated data of precipitation from 23 climate models of Coupled Model Intercomparison Project Phase 6 (CMIP6) at fifty grid points scattered over the Tibet Plateau region are considered to evaluate the applicability of MMFSDI. For comparative analysis, the forecasting performance of MMFSDI is compared with Standardized Precipitation Index (SPI) using Residual Mean Square Error (RMSE) and Mean Average Error (MAE) under Auto-Regressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN) models. Outcomes associated with this research shows that 1) the first component of FCA as an ensemble of multiple climate models is more describable than simple model averaging, 2) the strong consistency between MMFSDI and SPI makes MMFSDI as an alternative multi-model drought measure, and 3) the implications of ARIMA and ANN revealed that MMFSDI has inherited feature for forecasting drought. In summary, the finding of the research argues to substitute SPI with MMFSDI as MMFSDI has inherited forecasting ability.

CORRESPONDING AUTHOR:

Zulfiqar Ali

State Key Laboratory of Hydro-Science and Engineering and Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, PR China

Department of Statistics, University of Sialkot, Pakistan

zulfiqarali@stat.qau.edu.pk

KEYWORDS:

Drought; Ensemble drought measure; Forecastable Component Analysis; K-Component Gaussian Mixture Distribution

TO CITE THIS ARTICLE:

Yuanbin, S, Qamar, S, Ali, Z, Yang, T, Nazeer, A and Fayyaz, R. 2022. A New Ensemble Index for Extracting Predictable Drought Features from Multiple Historical Simulations of Climate. *Tellus A: Dynamic Meteorology and Oceanography*, 74(2022), 236–249. DOI: <https://doi.org/10.16993/tellusa.46>

1. INTRODUCTION

Due to climate warming, many studies have shown the increasing trend of recurrent occurrences of drought hazards in several parts of the world (Sharma and Goyal, 2020; Shao and Kam, 2020; Liang et al. 2018; Nam et al. 2015; Kousari et al. 2014; Damberg and AghaKouchak, 2014; Gocic and Trajkovic, 2013). Many consequential studies, statistical facts, and figures have revealed the significant damages of drought in various discipline of life (Andriano and Behrman, 2020; Carroll et al. 2009). On the same lines, many studies have reported the continuation of drought in many parts of the globe. Drought directly affects human lives (Yao et al. 2018), agriculture crop yield, livestock (Nembilwi et al. 2021), and ecosystem (Haile et al. 2019). In addition, drought creates several issues related to public health such as poor sanitation and low quality of drinking water (Mullin et al. 2020). However, the adverse impact of drought can be reduced by proper planning and appropriate drought mitigation policies.

In contrast to other natural hazard, drought is complex and difficult to assess accurately as well as efficiently. Therefore, accurate drought monitoring and forecasting are necessary for early warning drought mitigation policies. For drought monitoring and forecasting, many authors have proposed and recommended several drought monitoring tools and forecasting methods. Some popular and new drought monitoring tools are, Palmer Drought Severity Index (PDSI) (Palmer, 1965), Standardized Precipitation Index (SPI) (McKee et al. 1993), Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente et al. 2010), Joint Deficit Index (JDI) (Kao et al. 2010), Multivariate Standardized Drought Index (MSDI) (Hao et al. 2013), Multivariate Standardized Precipitation Index (MSPI) (Bazrafshan et al. 2014), Composite Drought Index (CDI) (Liu et al. 2020), Seasonally Combinative Regional Drought Indicator (SCRDI) (Ali et al. 2020), Regional Multi-Component Gaussian Hydrological Drought Assessment (RMcGHDA) (Ali et al. 2021), etc. Similarly, many researchers have developed several forecasting and prediction methods using various time series and machine learning approaches. For instance, Aghelpour et al. (2020) forecasted Joint Deficit Index (JDI) and MSPI under machine learning methods and entropy theory. Khan et al. (2020) utilized the strength of wavelet transformation and has proposed a hybrid model of Autoregressive Integrated Moving Average Model (ARIMA) and Artificial neural network (ANN) for drought predication. Pham et al. (2021) have forecasted drought based on SPI using Singular Spectrum Analysis (SSA) and Single Least Square Support Vector Machine (LSSVM). Dikshit et al. (2021) have used deep learning approach called the Long Short-Term Memory (LSTM) for the prediction of SPEI at different time scales.

In the last decade, several authors have assessed and forecasted drought using historical simulations

of precipitation data of various climate models of Coupled Model Intercomparison Project (CMIP). For instances, Huang et al. (2018) have used SPI and SPEI to investigate and explore various drought characteristics such as a trend, intensity, and duration under historical simulations of precipitation and temperature of CMIP5 regional climate model. Wu et al. (2021) have used SPI to assess global uncertainties in drought features based on the historical simulations of 28 general circulation models of CMIP5. Moon et al. (2018) have used different global climate models for assessing the persistence of drought under the Markov chain method. Tam et al. (2019) have used SPEI to project drought from seasonal to annual time scales for Canada region using 29 global climate models of CMIP5. Zhai et al. (2020) have used simulated data from five climate models of CMIP6 for projecting drought characteristics in South Asia. Yang et al. (2020) have used sixteen climate models of CMIP5 for assessing spatio-temporal trends in drought under the PDSI drought index. Campozano et al. (2020) have used SPI to evaluate the spatio-temporal features of drought CMIP5 at Ecuador region of South America.

However, spatio-temporal variation is a common in all the climate models (Yazdandoost., et al. 2021, Norris et al. 2021, Srivastava et al. 2020, Séférian et al. 2019). Furthermore, each model of CMIP5 and CMIP6 have a certain amount of uncertainties in the quantification of climate variables. Consequently, use of single climate model for the assessment of meteorological events like drought etc., can reduce the reliability of the results and predication. On the other hand, many researchers have used multiple climate models simultaneously (Ying et al. 2012; Su et al. 2021).

On the other hand, many researchers have applied ensemble techniques for the assessment of drought features under multiple climate simulations settings. For example, Crawford et al. (2019) have developed an ensemble approach for 34 climate models of CMIP5 using statistical and machine learning tools. Ruan et al. (2019) have used multi-model ensemble and delta method for the projection of future temperature changes. Similarly, Chhin et al. (2020) have explored the future changes of precipitation, temperature and drought characteristics in Indochina Region using the optimal ensemble of global climate models of the CMIP5. However, a certain amount of estimation errors and biases are the part each ensemble models (Ficklin et al. 2016). These errors and biases reduce the reliability of climate ensemble models. Therefore, a composite method enabled with inherited forecasting characteristics of drought is essential for accurate drought monitoring and forecasting. In this research, we aimed to improve drought monitoring and forecasting module by providing a compact drought indicator that have capabilities to define drought under multiple climate models and forecasting ability.

Statistics and machine learning data mining provide a large number of dimension reduction methods (Aziz et al. 2017; Alasadi and Bhaya, 2017; Härdle and Rönz, 2012). Depending upon the research questions and data, each method has certain assumptions and limitations. This research incorporates Forecastable Component Analysis (FCA) (Goerg, 2013) as a dimension reduction tool. Unlike other compatible techniques such as Principle Component Analysis (PCA), the main feature of FCA is that the reduced data have the inherited ability to forecast future values. A brief overview, literature, and methodology of FCA have been provided in section 2.1.

Among all drought monitoring tools, the standardized procedure of drought has gained more popularity (Erhardt and Czado, 2018). This is due to the fact that the standardized values are comparable and independent from regional scaling. However, the main limitation of SDIs is their standardization that involved in the time-series data. This makes the series whiten and unpredictable. Therefore, the main contribution of this research is the integration of FCA in SDIs procedures. Here, the Implication of FCA preserves the forecastability in the time series data of the drought index.

By following Ali et al. (2021a, 2021b), this research standardized the Cumulative Distribution Function (CDF) of K-Component Gaussian Mixture Distribution (K-CGMD) fitted on the first component of FCA. Consequently, this article proposes a new drought monitoring and forecasting indicator – the Multivariate Multi-Scaler Forecastable Standardized Drought Index (MMFSDI). In application, the research includes 50 randomly selected grid points scattered across the Tibet Plateau. To assess the efficiency of MMFSDI, the research compares the prediction and forecasting matrices of MMFSDI with SPI under ARIMA and ANN.

The organization of this article is as follows: Section 2 presents a brief introduction on the methodology of FCA and K-CGMD. Section 3 presents the proposal of MMFSDI. Section 4 provides the way and methods for assessing the efficiency of MMFSDI over existing ones. For the application of the MMFSDI, section 5 describes the data and study area. Section 6 consists of the results and discussion. Finally, the summary and conclusion are presented in section 7.

2. METHODS

2.1. FORECASTABLE COMPONENT ANALYSIS (FCA) AS A DIMENSION REDUCTION TOOL

The primary interests in multivariate time series analysis are dimension reduction and handling redundancy problems within multiple variables. To do so, many authors have proposed several machine learning methods for various nature of multivariate datasets. Examples include Principle Component Analysis (PCA),

non-negative matrix factorization (Lee and Seung, 1999), t-distributed stochastic neighbor embedding (t-SNE) (Van, 2008, 2014), autoencoders (Rumelhart et al. 1985).

In this research, we propose to employ Forecastable Component Analysis (FCA) for generating time series vectors of standardized drought index using multiple climate simulated model based time series data of precipitation. In multivariate time series analysis, FPCA is a new dimension reduction method that optimally transforms the set of multivariate time series data into forecastable series. Unlike PCA, FCA explicitly addresses forecastability from the data set. That is, FCA accounts temporal dependency within data and discovers forecastable subspace.

The idea of FCA is based on the minimization of differential entropy as a measure of uncertainty.

Mathematically, let $L \times K$ be the data matrix U , f be the vector defined by FCA that maximize the forecastability from such that $U = U^T f$, subject to the constraint that $f^T \tilde{R} f = 1$. Here, $\tilde{R}$ is the covariance matrix. In first step, the methods whiten the data matrix U . After whitening the model proceeds to obtain a multivariate spectrum matrix. In general, the main aim of FCA is to examine an extraction vector f in such a way that the predictable signals $U = U^T f$ are as much forecastable as possible. For detailed mathematical description, readers are referred to see Goerg, (2013, 2016a).

2.2. UNIVARIATE PROBABILITY DISTRIBUTIONS AND K – COMPONENT GAUSSIAN MIXTURE DISTRIBUTION (K-CGMD)

From last three decades, univariate probability models have several application in hydrological data. Specifically, the methodologies of SDIs are mainly based on the CDF of appropriate univariate probability distribution. In contrast, new research is based on the mixture models of multiple distributions. In our recent research, we have shown that the accuracy of SDIs can be increase by employing the CDF of multiple distributions. Hence, the selection of mixture probability model is more suitable alternative choice for modeling hydrological data. On the same lines of Ali et al. (2021a and 2021b), this research is based on K-CGMD. A brief description on K-CGMD is as follows:

Let x be the time series data having multi-components. That is the distribution of x have several center. Then the two type of parameters formulate the K-CGMD. These two types of parameters are called locations and scale parameters. Mathematically, K-CGMD can be define as

$$f(x) = \sum_{i=1}^K \delta_i N(\theta_i, \sigma_i) \quad (1)$$

where,

$$N(\theta_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{(x - \theta_i)^2}{2\sigma_i^2}\right\} \quad (2)$$

In above model, θ is the mean, σ_i^2 is the variance. While, δ is the weight for i^{th} component k such that:

$$\sum_{i=1}^K \delta_i = 1 \quad (3)$$

In previous research, many researchers have used K-CGMD in different applications. Some of them are Lu et al. (2014), Ozonder and Miller, (2021), Janelidze et al. (2020), Zoonomia and Consortium, (2020) and Xu et al., (2020).

3. THE PROPOSED ENSEMBLE INDICATOR – THE MULTIVARIATE MULTI-SCALER FORECASTABLE STANDARDIZED DROUGHT INDEX (MMFSDI)

This section provides the two-stage procedure of MMFSDI. The first stage is based on the implication of FCA on the multivariate time series data of precipitation. While the second stage provides the standardization process for obtaining MMFSDI values. The details on each stage are provided in the following subsection.

3.1. PHASE 1 – DIMENSION REDUCTION – THE INTEGRATION AND ASSESSMENT OF THE APPROPRIATENESS OF FCA FOR MULTIPLE CLIMATE MODELS

For the sake of generality, consider the set of climate models say Σ . Symbolically, $\Sigma \in (m_1, m_2, m_3, \dots, m_k)$, where, m_1, m_2 , and m_k are the time series data of historical simulated data of precipitation under various climate models. This phase proposes the integration climate models using FPCA in dimension reduction context. Here, a multidimensional spatio-temporal matrix of rainfall vectors defined in Σ are used to maximize the variance and directional information in the data set of the first component (FC_1) of FCA by keeping maximum forecastability. After observing substantial amount forecastability (Ω) in FC_1 , the research suggests to standardized FC_1 appropriate standardization technique.

3.2. STANDARDIZATION UNDER 12-COMPONENT GAUSSIAN MIXTURE DISTRIBUTION

After the integration of climate models, this phase standardized the time series data of FC_1 by fitting CDF of K-CGMD. The adoption of K-CGMD is due to (Ali et al. 2021a, 2021b). For easy to understand, the standardization process has been divided in the following two steps:

- i) In the first step, we compute the CDF of K-CGMD for the time series data of FC_1 .

Mathematically speaking, let $F(x)$ denotes the CDF of K-CGMD. The CDF of i^{th} component can be represented as follows.

$$F(\mathbf{x}) = F_1(\mathbf{FC}_1) + F_2(\mathbf{FC}_1) + F_3(\mathbf{FC}_1) + \dots, F(\mathbf{FC}_1) \quad (4)$$

- ii) In the second step, the time series vector of $F(x)$ is standardized by following approximation.

$$MMFSDI = - \left(\Psi + \frac{v_o + v_1 \Psi + v_c \Psi^2}{1 + w_1 \Psi + w_2 \Psi^2 + w_3 \Psi^3} \right) \quad (5)$$

for

$$\Psi = \sqrt{\ln \left[\frac{1}{\{F(x)\}^2} \right]}$$

when

$$0 < F(x) \leq 0.5 \quad (10)$$

$$MMFSDI = + \left(\Psi + \frac{v_o + v_1 \Psi + v_c \Psi^2}{1 + w_1 \Psi + w_2 \Psi^2 + w_3 \Psi^3} \right) \quad (6)$$

and for,

$$\Psi = \sqrt{\ln \left[\frac{1}{\{1 - F(x)\}^2} \right]}$$

when,

$$0.5 < F(x) \leq 1 \quad (7)$$

Where,

$v_o = 2.515517$, $v_1 = 0.802853$, $v_2 = 0.010328$, $w_1 = 1.432788$, $w_2 = 0.985269$, $w_3 = 0.0018$ are constant values.

The resultant standardized time series can be categorizing into various drought classes. **Table 1** provides the classification of drought under MMFSDI and SPI.

For various time scales, the estimation of RMSDI can be done by just taking moving the total of FC_1 . Accordingly, the repetition of step 1 and step 2 is required.

In computation, *mixtools* (Benaglia et al. 2009) of R package are employed for estimation of CDF the K-CGMD, while the standardized data of MMFSDI and SPI is obtained by writing R codes.

MMFSDI AND SPI VALUES	DROUGHT CLASSIFICATION
2.00 and above	Extremely Wet
1.50 to 1.99	Very Wet
1.00 to 1.49	Moderate Wet
-0.99 to 0.99	Near Normal
-1.00 to -1.49	Moderate Drought
-1.50 to -1.99	Severe Drought
-2.00 and less	Extremely Drought

Table 1 Drought classification criterion.

4. PERFORMANCE ASSESSMENT UNDER FORECASTING MODELS AND COMPARATIVE MATRICES

There are several popular forecasting models for modeling univariate time series. In comparative analysis, the following forecasting models and evaluation matrices have been considered to assess and compare the forecasting efficiency of MMFSDI over SPI.

4.1. ARTIFICIAL NEURAL NETWORK (ANN) TECHNIQUE AND AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

ANN technique and ARIMA models are the most popular forecasting models for univariate time series data (Zhang, 2003). In several disciplines, many studies have assessed the future states of the process under ANN and ARIMA models. The examples include, biology (Wang et al. 2021), economic (Mallikarjuna and Rao, 2019), chemistry (Bornéo et al. 2021), hydrology (Khan et al. 2020), meteorology (Zhang et al. 2020), environmental studies (Khairuddin et al. 2019), agriculture (Paul and Garai, 2021) and civil engineering (Maleki et al. 2018) etc. Accordingly, many authors have forecasted drought based on the ANN technique and ARIMA models. Such works have been done by Belayneh et al. (2014), Xu et al. (2020), Pham et al. (2021), and many more. Therefore, this research compares the forecasting efficacy of MMFSDI with SPI under both the ARIMA and ANN models.

In this research, the numerical computation of the ARIMA model is based on *auto.arima()* function of *forecast* (Hyndman et al. 2020) package of R statistical software. While, *nnetar()* function is used to predict and forecast MMFSDI and SPI values under ANN techniques. Further, the error matrices in MMSFDI and SPI under ANN and ARIMA are compared to assess the forecastability power of MMFSDI over SPI.

4.2. QUALITY MEASURES

To assess the forecasting performances of the proposed ensemble index, time series data of MMFSDI and SPI have been divided into two sets. The first set contains 80% of the time series data, while the second set consists on the remaining 20% of the data. The first set of data is used in training the models. While the second set is used to test the forecasting performance. In both phases, errors matrices are investigated, independently. In this research, we have included two errors matrices- the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE). The mathematical formula of these two matrices are given in the following two equations.

$$RMSE = \sqrt{\frac{1}{h+1} \sum_{t=T}^{T+h} (y_{t+1} - \hat{y}_{t+1|T})^2} \quad (8)$$

$$MAE = \frac{1}{h+1} \sum_{t=T}^{T+h} (|y_{t+1} - \hat{y}_{t+1|T}|) \quad (9)$$

In above equations, y is the actual value of the time series data, $\hat{y}$ is the predicted or forecasted values, h is the step, and T is the time index. In model comparison, the low values of RMSE and MAE suggest the appropriateness of one model over another.

5. APPLICATION

In application, this research considers 50 random grid points scattered over the Tibet Plateau region. The Tibet Plateau is the highest and largest plateau in the world, and has a total area of more than 2.5 million km². As the Third Pole, it has profound impacts on regional and even global weather pattern. Our study mainly focuses on the area 25–40° N and 73–105° E with mean elevation of about 4252m (see [Figure 1](#)). The Tibet Plateau boundary is defined as above the 2500m contour line using the National Aeronautics and Space Administration Shuttle

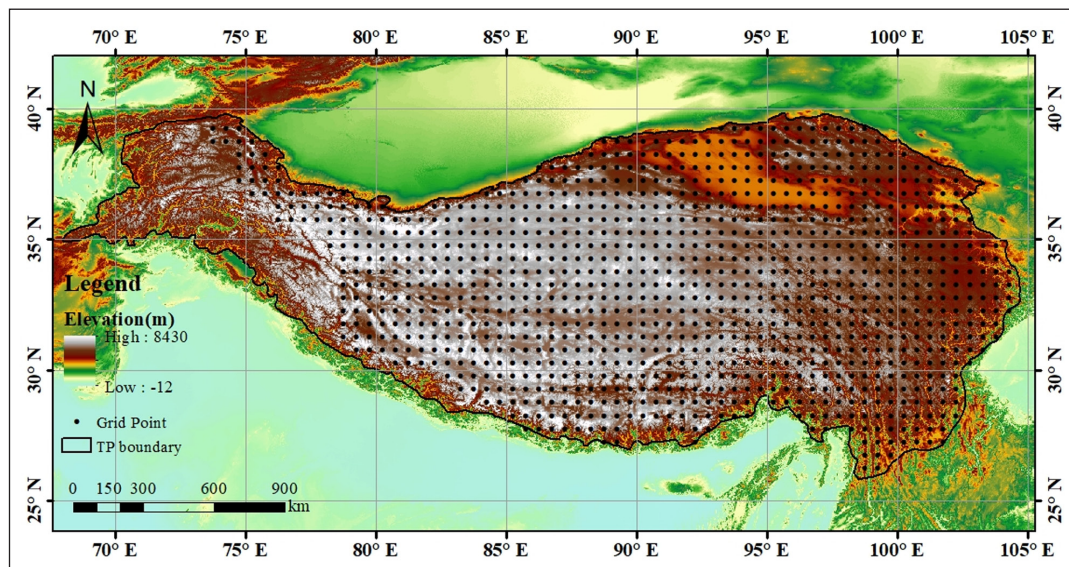


Figure 1 Geographical coverage of the study area and CMIP6 0.5°×0.5° grid points (black dots).

Radar Topographic Mission 90m DEM (Li et al. 2021). In this research, time series data (ranged from 1850–2014) of precipitation, minimum and maximum temperature is used to estimate SPEI value under ten climate models of CMIP6. The data set has been obtained from the CMIP6 data archive <https://esgf-node.llnl.gov/search/cmip6/> (see **Table 2**). **Figure 2** shows the geographical reference of fifty random locations. Here, the selection of these locations are based on simple random sampling.

6. RESULTS AND DISCUSSION

6.1. INFERENCE OF FORECASTABLE COMPONENT ANALYSIS ON MULTI MODEL CLIMATE MODEL OF CMIP-6

The crucial aspect in multivariate time series analysis is to reduce the dimension of data by keeping the forecasting characteristics. This section is related to the implication of FCA in dimension reduction context. Here, we present and discuss the results associated with

NUMBER	MODEL NAME	MODELING CENTER	RESOLUTION
			(LONGITUDE × LATITUDE)
1	ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation, Australia	1.875° × 1.25°
2	ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation, Australia	1.875° × 1.2143°
3	AWI-CM-1-1-MR	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany	0.9375° × 0.9375°
4	BCC-CSM2-MR	Beijing Climate Center and China Meteorological Administration, China	1.125° × 1.125°
5	CanESM5	Canadian Centre for Climate Modeling and Analysis, Canada	2.8125° × 2.8125°
6	CanESM5-CanOE5	Canadian Centre for Climate Modeling and Analysis, Canada	2.8125° × 2.8125°
7	CNRM-CM6-1	National Centre for Meteorological Research and European Centre for Research and Advanced Training in Scientific Computation, France	1.40625° × 1.40625°
8	CNRM-CM6-1-HR	National Centre for Meteorological Research and European Centre for Research and Advanced Training in Scientific Computation, France	0.5° × 0.5°
9	CNRM-ESM2-1	National Centre for Meteorological Research and European Centre for Research and Advanced Training in Scientific Computation, France	1.40625° × 1.40625°
10	EC-Earth3-Veg	EC-Earth consortium, Europe	0.703125° × 0.703125°

Table 2 The information of the selected CMIP6 models in this study.

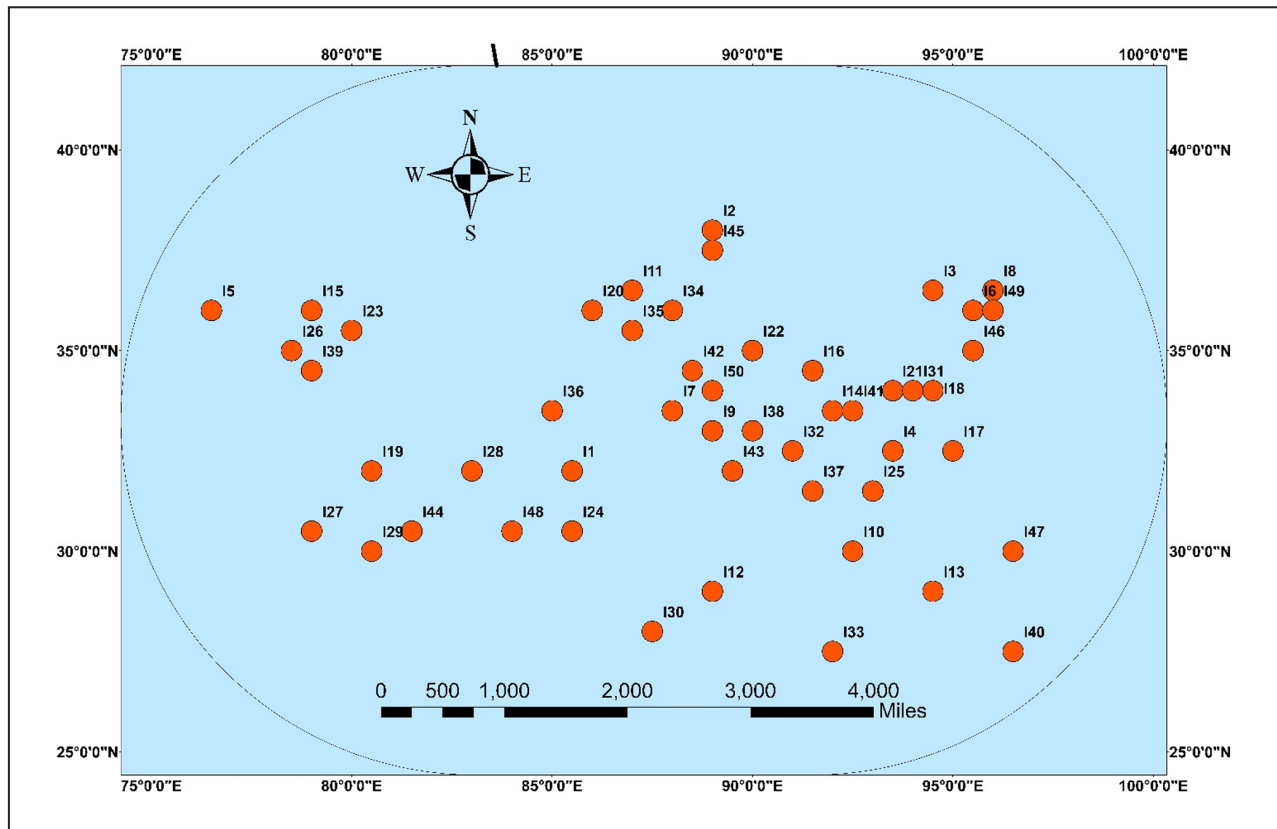


Figure 2 Spatial distribution fifty random location scattered over Tibet Plateau.

the application of FCA on the multivariate time series data set of 23 climate simulated models of CMIP-6. The main objective of using FCA is to achieve aggregate time-series data that have inherited forecastability features for drought forecasting. In computation and inference, the research employed *ForeCA* (Goerge, 2016a).

Figure 3 shows the graphical summary of FCA on the time series data of 23 climate models at one random grid point of Tibet Plateau. At this grid point, we have assessed that the first component has 77.08% forecasting ability. Forecasting features are inherently time-limited. Therefore, reducing prediction errors is always a difficult task. In general, drought is

complex and difficult to predict. However, the high forecastability make sure the more prediction accuracy under the proposed index. **Table 3** provides the values of Ω for the reaming components. Here, the highest values of Ω for the first components reveal its excellent forecasting ability. In other words, the highest value of Ω of the first component shows that it has is more forecastable.

Similarly, the FCA algorithm is performed for all the grid points of Tibet Plateau. **Figure 4** shows the spatial distribution of Ω . It is observed that only one or two grid points have a significantly small value of Ω (<50). In addition, the range Ω lies between 50–60 covers a comparatively small area. However, a large

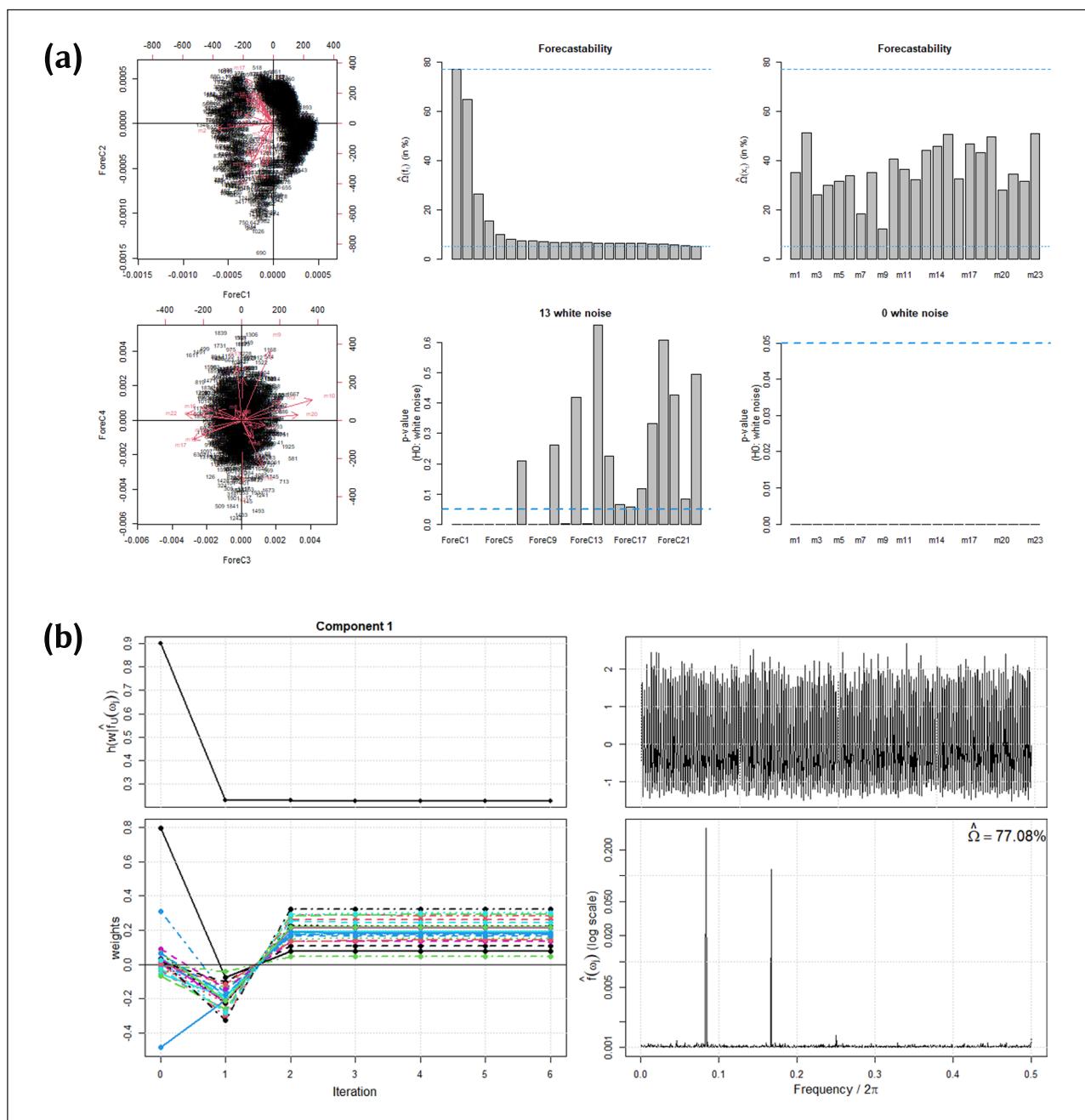


Figure 3 a) Biplots, forecastability and b) summary of single FPC.

COMPONENTS	OMEGA (Ω)
ForeC1	77.08283
ForeC2	64.98184
ForeC3	26.62885
ForeC4	15.64948
ForeC5	10.16605
ForeC6	8.119885
ForeC7	7.546159
ForeC8	7.281088
ForeC9	7.117616
ForeC10	7.077022
ForeC11	6.888564
ForeC12	6.882761
ForeC13	6.791538
ForeC14	6.623308
ForeC15	6.611346
ForeC16	6.607055
ForeC17	6.302802
ForeC18	6.256745
ForeC19	6.184776
ForeC20	6.074139
ForeC21	5.896357
ForeC22	5.723812
ForeC23	5.685728

Table 3 Values of Omega against each components.

TIME SCALES	SPI	MMFSDI
Scale 1	-4003.55	-4357.048
Scale 6	-8614.13	-9084.659
Scale 9	-7977.962	-8446.784
Scale 12	-4295.112	-4548.289
Scale 24	-5609.977	-5905.589
Scale 48	-6740.838	-6921.417

Table 4 BIC of KCGMD for SPI and MMFSDI.

number of portion of grid points have Ω greater than 60. In dimension reduction context, these inferences validate the use of FCA for increasing the forecastability of drought features under multiple climate models.

6.2. ESTIMATION OF MMFSDI AND SPI

After obtaining the time series data of the first component of FCA for each grid point, we have obtained the MMFSDI data by standardizing the CDF of L-CGMD. We have followed this new procedure due to Ali et al., (2021a, 2021b). For comparison, SPI time series data is obtained by averaging all the historical simulation models, accordingly. This section explores quantitative and graphical results related to the estimation of MMFSDI and SPI. For one random selected grid point, **Figure 5** shows the density plots corresponding with q-q plots observed on the precipitation (averaged of all the selected climate models) and the first component of FCA at a one-month time scale. Here, we have observed the BIC value for MMFSDI (BIC = -4357.048) is comparatively low than for the estimation of SPI (BIC = -4003.55). This shows that the data of the first component of FCA is more accurately describable than simple model averaging. Resultantly, we can say that MMFSDI is more accurate than SPI. Similar results have been found for the rest of the time scale (see **Table 4**). To check the consistency between SPI and MMFSDI, **Figure 6** is prepared to show the temporal behavior of MMFSDI and SPI. While the correlation analysis reveals that MMFSDI is strongly correlated with SPI (see **Figure 7**). This consistency leads towards the endorsement of MMFSDI as an alternative drought indicator.

To assess that MMFSDI is more forecastable than SPI, the next section presents inferences related to ARIMA and ANN modeling.

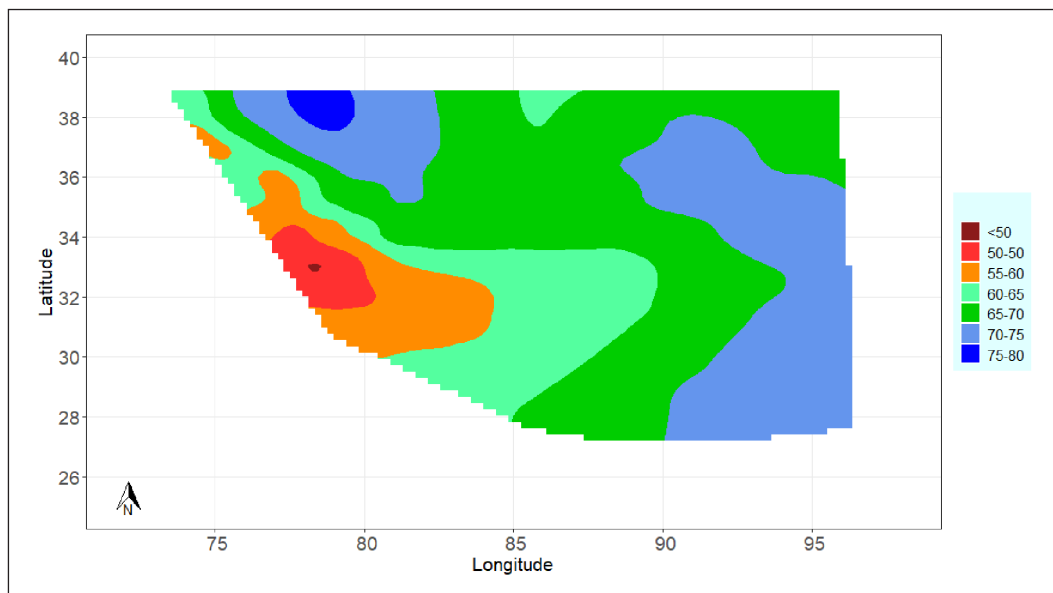


Figure 4 Spatial distribution of Omega (Ω).

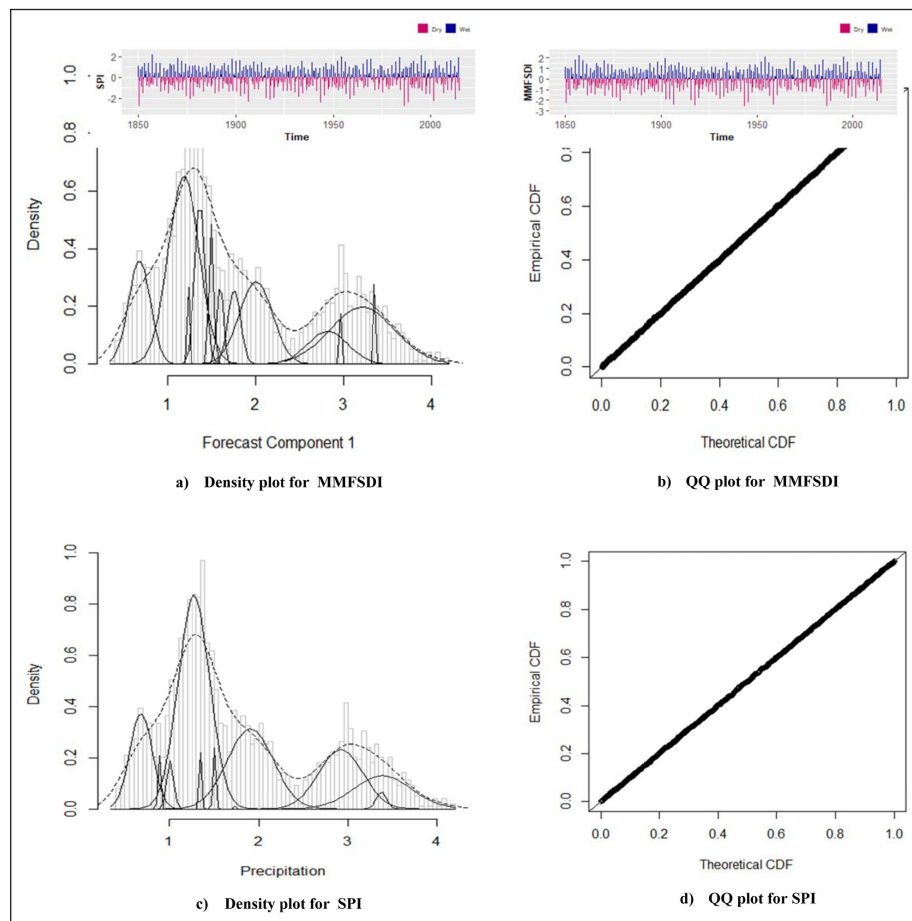


Figure 5 Density and qq plot of KCGMD for SPI and MMFSDI at one month time scale.

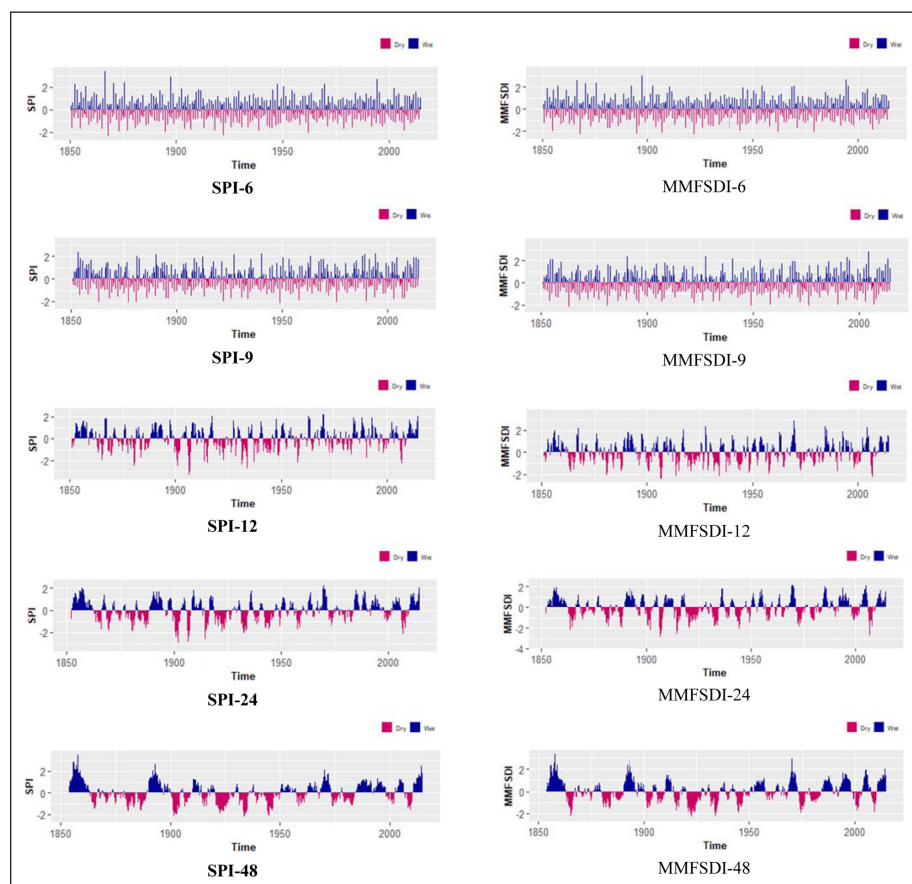


Figure 6 Temporal behaviour of SPI and MMFSDI in various time scale.

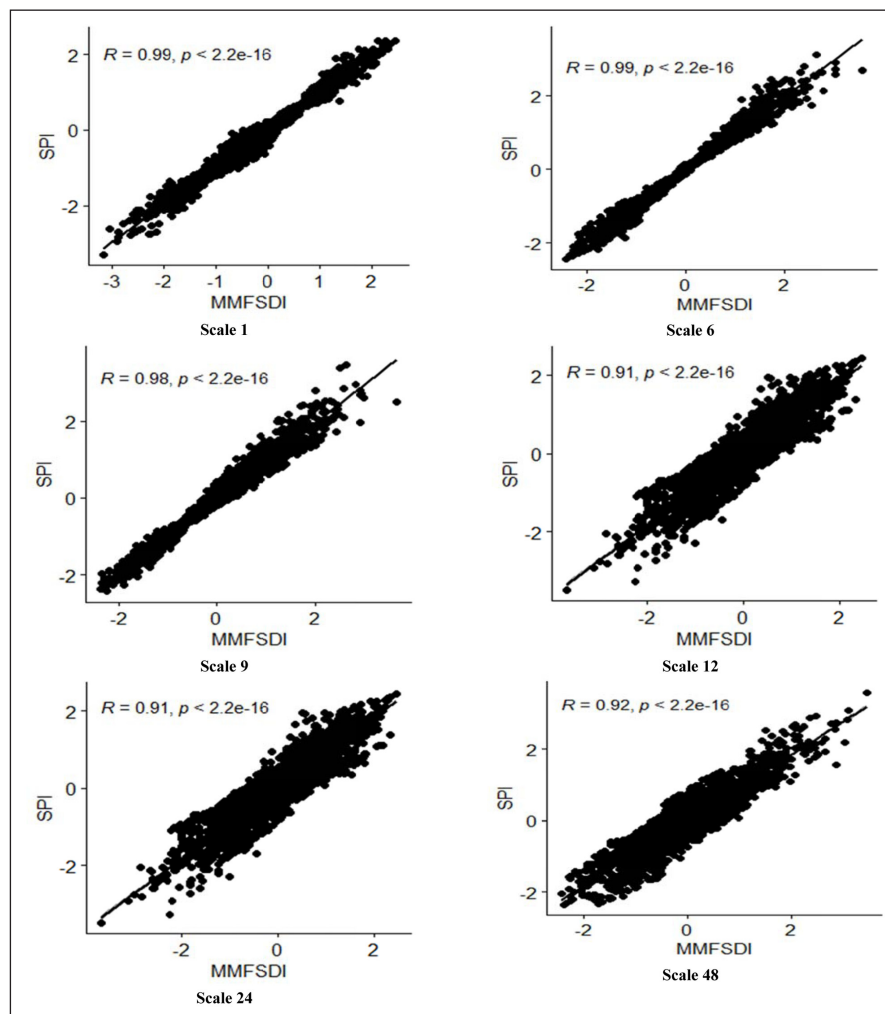


Figure 7 Correlation analysis between SPI and MMFSDI in various time scales.

6.3. COMPARATIVE EVALUATION

This section presents the results associated with comparative evaluation of the proposed ensemble approach with model averaging based drought index. Here, the forecasting efficiency of MMFSDI is compared with SPI using performing matrices of ARIMA and ANN techniques. In this article, the results related to the implications of ARIMA and ANN have been provided for one random grid point. The remaining results of all other grid points are in accordance with that random point. According to standard practice, the implications of ARIMA and ANN have consisted of training and testing phase. The training phase is employed for the estimation of model parameters and assessment of prediction accuracy, while the testing phase describes the forecasting accuracy. The two models ARIMA and NN are executed for each time scale of MMFSDI and SPI, separately. To do this, this research has employed *forecast* and *nnet* packages of R statistical software.

In both the training and testing phase, the performance assessment criterion of ARIMA and ANN have analogous evaluation (see [Table 5](#)). In training phase, ARIMA gives the comparative low value of RMSE (0.580) and MAE (0.445) for MMFSDI than SPI (RMSE = 0.760, MAE = 0.601) at one-

month time scales. Accordingly, the similar quantitative assessment of performing matrices in the testing phase proves that MMFSDI is more predictable than SPI at month time scale. Although, some inadmissible quantitation in performing matrices have been observed at a 24-month time scale. However, the forecasting ability of MMFSDI in most of the time scales is comparatively high. Contrary to ARIMA, the performing matrices in both the training and testing phase under ANN models show that the forecasting ability of MMFSDI is high from SPI in all time scales. Furthermore, the quantitation deviation in RMSE and MAE at the training and testing phase between ARIMA and ANN shows that ANN is more appropriate than ARIMA. Hence, these results substantiate that MMFSDI is a more appropriate indicator than SPI for forecasting future values of drought under multiple climate models.

7. SUMMARY AND CONCLUSION

Accurate and efficient ensemble procedures of historical simulations of various climate models are essential for drought monitoring. The appropriate choice for selecting an ensemble depends on the

TIME SCALES	TECHNIQUES	INDEX	TRAINING		TESTING	
			RMSE	MAE	RMSE	MAE
Scale 1	ARIMA	SPI	0.760	0.601	0.962	0.774
		MMFSDI	0.580	0.445	0.737	0.555
	MLP	SPI	0.302	0.233	1.010	0.831
		MMFSDI	0.191	0.145	0.752	0.588
Scale 6	ARIMA	SPI	0.474	0.341	1.089	0.885
		MMFSDI	0.229	0.172	0.655	0.491
	MLP	SPI	0.163	0.112	0.650	0.540
		MMFSDI	0.115	0.078	0.399	0.307
Scale 9	ARIMA	SPI	0.438	0.331	1.658	1.349
		MMFSDI	0.396	0.317	1.106	0.871
	MLP	SPI	0.150	0.108	0.826	0.644
		MMFSDI	0.116	0.085	0.440	0.335
Scale 12	ARIMA	SPI	0.346	0.266	0.683	0.548
		MMFSDI	0.343	0.258	1.398	1.138
	MLP	SPI	0.225	0.173	1.564	1.222
		MMFSDI	0.223	0.171	0.941	0.631
Scale 24	ARIMA	SPI	0.227	0.173	0.497	0.450
		MMFSDI	0.241	0.179	1.167	1.000
	MLP	SPI	0.163	0.125	1.339	0.903
		MMFSDI	0.162	0.121	0.406	0.324
Scale 48	ARIMA	SPI	0.153	0.108	1.250	1.102
		MMFSDI	0.141	0.106	1.135	1.121
	MLP	SPI	0.151	0.107	1.041	0.904
		MMFSDI	0.143	0.106	0.917	0.851

Table 5 Performance assessment of MMFSDI over SPI.

certain quantitative characteristics such as low biases and forecasting errors. This paper provides a novel way to ensemble climate models for drought forecasting. Here, a new drought indicator- the Multivariate Multi-Scaler Forecastable Standardized Drought Index (MMFSDI) has been proposed. Contrary to other drought indices, the distinctive ability of MMFSDI is the inherited forecasting feature. This is because the standardized values of MMFSDI are based on the first component of FCA. To assess the performance of MMFSDI, the paper included statistical inferences based on 50 random grid points of the Tibet Plateau. Here, historical simulated data of 23 climate models of CMIP-6 is used to obtain MMFSDI. While the comparative assessment is based on SPI under ANN and ARIMA models. Under both forecasting algorithms, outcomes associated with this research show that MMFSDI has an inherited ability to forecast drought under multiple climate models. In summation, the research provides an alternate approach for an ensemble of climate models for analyzing and forecasting drought features. Thus, this research strengthens the drought forecasting module by providing a new multiple climate simulated data based drought monitoring indicator.

DATA ACCESSIBILITY STATEMENTS

All the data were analyzed using *R* software. The data and code used to support the findings of this study are available from the corresponding author upon request.

ETHICS AND CONSENT

The manuscript is prepared in accordance with the ethical standards of the responsible committee on human experimentation and with the latest (2008) version of the Helsinki Declaration of 1975.

FUNDING INFORMATION

The authors are very grateful to the Natural Science Foundation of Jiangsu Province, China (BK20210369).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

All authors have equal contributions.

AUTHOR AFFILIATIONS

Sun Yuanbin

State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing 210098, China

Sadia Qamar

Department of Statistics, University of Sargodha, Pakistan

Zulfiqar Ali

State Key Laboratory of Hydro-Science and Engineering and Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, PR China
Department of Statistics, University of Sialkot, Pakistan

Tao Yang

State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing 210098, China

Amna Nazeer

Department of Statistics, COMSATS University, Islamabad, Pakistan

Rabia Fayyaz

Department of Statistics, COMSATS University, Islamabad, Pakistan

REFERENCES

- Aghelpour, P, Mohammadi, B, Biazar, SM, Kisi, O and Sourmirinezhad, Z. 2020. A theoretical approach for forecasting different types of drought simultaneously, using entropy theory and machine-learning methods. *ISPRS International Journal of Geo-Information*, 9(12): 701. DOI: <https://doi.org/10.3390/ijgi9120701>
- Alasadi, SA and Bhaya, WS. 2017. Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16): 4102–4107.
- Ali, F, Li, BZ and Ali, Z. 2021b. Strengthening drought monitoring module by ensembling auxiliary information based varying estimators. *Water Resources Management*, 1–18. DOI: <https://doi.org/10.1007/s11269-021-02888-2>
- Ali, Z, Ellahi, A, Hussain, I, Nazeer, A, Qamar, S, Ni, G and Faisal, M. 2021a. Reduction of Errors in Hydrological Drought Monitoring—A Novel Statistical Framework for Spatio-Temporal Assessment of Drought. *Water Resources Management*, 1–18. DOI: <https://doi.org/10.1007/s11269-021-02952-x>
- Ali, Z, Hussain, I, Grzegorzczak, MA, Ni, G, Faisal, M, Qamar, S, ... and Al-Deek, FF. 2020. Bayesian network based procedure for regional drought monitoring: The Seasonally Combinative Regional Drought Indicator. *Journal of Environmental Management*, 276: 111296. DOI: <https://doi.org/10.1016/j.jenvman.2020.111296>
- Andriano, L and Behrman, J. 2020. The effects of growing-season drought on young women's life course transitions in a sub-Saharan context. *Population Studies*, 74(3): 331–350. DOI: <https://doi.org/10.1080/00324728.2020.1819551>
- Aziz, R, Verma, CK and Srivastava, N. 2017. Dimension reduction methods for microarray data: a review. *AIMS Bioengineering*, 4(2): 179–197. DOI: <https://doi.org/10.3934/bioeng.2017.2.179>
- Bazrafshan, J, Hejabi, S and Rahimi, J. 2014. Drought monitoring using the multivariate standardized precipitation index (MSPI). *Water resources management*, 28(4): 1045–1060. DOI: <https://doi.org/10.1007/s11269-014-0533-2>
- Belayneh, A, Adamowski, J, Khalil, B and Ozga-Zielinski, B. 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508: 418–429. DOI: <https://doi.org/10.1016/j.jhydrol.2013.10.052>
- Benaglia, T, Chauveau, D, Hunter, D and Young, D. 2009. mixtools: An R package for analyzing finite mixture models. *Journal of statistical software*, 32(6): 1–29. DOI: <https://doi.org/10.18637/jss.v032.i06>
- Bornéo, R, Franco, MM, Orlandin, BC, Moraes, NT and Corso, LL. 2021. Application of Analytic Hierarchy Process Considering Artificial Neural Network and ARIMA for Selecting a Chemical Waste Plant. *Scientia cum Industria*, 9(1): 30–37. DOI: <https://doi.org/10.18226/23185279.v9iss1p30>
- Campozano, L, Ballari, D, Montenegro, M and Avilés, A. 2020. Future meteorological droughts in Ecuador: CMIP5 derived decreasing trends and spatio-temporal features associated. *Frontiers in Earth Science*, 8: 17. DOI: <https://doi.org/10.3389/feart.2020.00017>
- Carroll, N, Frijters, P and Shields, MA. 2009. Quantifying the costs of drought: new evidence from life satisfaction data. *Journal of Population Economics*, 22(2): 445–461. DOI: <https://doi.org/10.1007/s00148-007-0174-3>
- Chhin, R, Oeurng, C and Yoden, S. 2020. Drought projection in the Indochina Region based on the optimal ensemble subset of CMIP5 models. *Climatic Change*, 162(2): 687–705. DOI: <https://doi.org/10.1007/s10584-020-02850-y>
- Damberg, L and AghaKouchak, A. 2014. Global trends and patterns of drought from space. *Theoretical and applied climatology*, 117(3): 441–448. DOI: <https://doi.org/10.1007/s00704-013-1019-5>
- Dikshit, A, Pradhan, B and Huete, A. 2021. An improved SPEI drought forecasting approach using the long short-term memory neural network. *Journal of environmental management*, 283: 111979. DOI: <https://doi.org/10.1016/j.jenvman.2021.111979>
- Erhardt, TM and Czado, C. 2018. Standardized drought indices: a novel univariate and multivariate approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(3): 643–664. DOI: <https://doi.org/10.1111/rssc.12242>
- Gocic, M and Trajkovic, S. 2013. Analysis of precipitation and drought data in Serbia over the period 1980–2010. *Journal of Hydrology*, 494: 32–42. DOI: <https://doi.org/10.1016/j.jhydrol.2013.04.044>

- Goerg, G. 2013, May. Forecastable component analysis. In *International conference on machine learning* (pp. 64–72). PMLR.
- Haile, GG, Tang, Q, Sun, S, Huang, Z, Zhang, X and Liu, X. 2019. Droughts in East Africa: Causes, impacts and resilience. *Earth-science reviews*, 193: 146–161. DOI: <https://doi.org/10.1016/j.earscirev.2019.04.015>
- Härdle, W and Rönz, B. (Eds.). 2012. *Compstat: Proceedings in Computational Statistics*. Springer Science & Business Media.
- Huang, J, Zhai, J, Jiang, T, Wang, Y, Li, X, Wang, R, ... and Fischer, T. 2018. Analysis of future drought characteristics in China using the regional climate model CCLM. *Climate Dynamics*, 50(1): 507–525. DOI: <https://doi.org/10.1007/s00382-017-3623-z>
- Hyndman, RJ, Athanasopoulos, G, Bergmeir, C, Caceres, G, Chhay, L, O'Hara-Wild, M, ... and Wang, E. 2020. Package 'forecast'. [Online] <https://cran.r-project.org/web/packages/forecast/forecast.pdf>.
- Janelidze, S, Stomrud, E, Smith, R, Palmqvist, S, Mattsson, N, Airey, DC, ... and Hansson, O. 2020. Cerebrospinal fluid p-tau217 performs better than p-tau181 as a biomarker of Alzheimer's disease. *Nature communications*, 11(1): 1–12. DOI: <https://doi.org/10.1038/s41467-020-15436-0>
- Khairuddin, N, Aris, AZ, Elshafie, A, Sheikhy Narany, T, Ishak, MY and Isa, NM. 2019. Efficient forecasting model technique for river stream flow in tropical environment. *Urban Water Journal*, 16(3): 183–192. DOI: <https://doi.org/10.1080/1573062X.2019.1637906>
- Khan, MMH, Muhammad, NS and El-Shafie, A. 2020. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *Journal of Hydrology*, 590: 125380. DOI: <https://doi.org/10.1016/j.jhydrol.2020.125380>
- Kousari, MR, Dastorani, MT, Niazi, Y, Soheili, E, Hayatzadeh, M and Chezgi, J. 2014. Trend detection of drought in arid and semi-arid regions of Iran based on implementation of reconnaissance drought index (RDI) and application of non-parametrical statistical method. *Water resources management*, 28(7): 1857–1872. DOI: <https://doi.org/10.1007/s11269-014-0558-6>
- Lee, DD and Seung, HS. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788–791. DOI: <https://doi.org/10.1038/44565>
- Li, G, Yu, Z, Wang, W, Ju, Q and Chen, X. 2021. Analysis of the spatial Distribution of precipitation and topography with GPM data in the Tibetan Plateau. *Atmospheric Research*, 247: 105259. DOI: <https://doi.org/10.1016/j.atmosres.2020.105259>
- Liang, Y, Wang, Y, Yan, X, Liu, W, Jin, S and Han, M. 2018. Projection of drought hazards in China during twenty-first century. *Theoretical and Applied Climatology*, 133(1): 331–341. DOI: <https://doi.org/10.1007/s00704-017-2189-3>
- Liu, Q, Zhang, S, Zhang, H, Bai, Y and Zhang, J. 2020. Monitoring drought using composite drought indices based on remote sensing. *Science of The Total Environment*, 711: 134585. DOI: <https://doi.org/10.1016/j.scitotenv.2019.134585>
- Lu, Z, Guan, X, Schmidt, CA and Matera, AG. 2014. RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins. *Genome biology*, 15(1): 1–23. DOI: <https://doi.org/10.1186/gb-2014-15-1-r7>
- Maleki, A, Nasser, S, Aminabad, MS and Hadi, M. 2018. Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics. *KSCE Journal of Civil Engineering*, 22(9): 3233–3245. DOI: <https://doi.org/10.1007/s12205-018-1195-z>
- Mallikarjuna, M and Rao, RP. 2019. Application of ARIMA, ANN and hybrid models to forecast the SENSEX returns. *Wealth*, 8(1): 14–19.
- Mullin, M. 2020. The effects of drinking water service fragmentation on drought-related water security. *Science*, 368(6488): 274–277. DOI: <https://doi.org/10.1126/science.aba7353>
- Nam, WH, Hayes, MJ, Svoboda, MD, Tadesse, T and Wilhite, DA. 2015. Drought hazard assessment in the context of climate change for South Korea. *Agricultural Water Management*, 160: 106–117. DOI: <https://doi.org/10.1016/j.agwat.2015.06.029>
- Nembilwi, N, Chikoore, H, Kori, E, Munyai, RB and Manyanya, TC. 2021. The Occurrence of Drought in Mopani District Municipality, South Africa: Impacts, Vulnerability and Adaptation. *Climate*, 9(4): 61. DOI: <https://doi.org/10.3390/cli9040061>
- Ozonder, G and Miller, EJ. 2021. Longitudinal investigation of skeletal activity episode timing decisions—A copula approach. *Journal of choice modelling*, 40: 100306. DOI: <https://doi.org/10.1016/j.jocm.2021.100306>
- Paul, RK and Garai, S. 2021. Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. *Soft Computing*, 1–17. DOI: <https://doi.org/10.1007/s00500-021-06087-4>
- Pham, QB, Yang, TC, Kuo, CM, Tseng, HW and Yu, PS. 2021. Coupling Singular Spectrum Analysis with Least Square Support Vector Machine to Improve Accuracy of SPI Drought Forecasting. *Water Resources Management*, 35(3): 847–868. DOI: <https://doi.org/10.1007/s11269-020-02746-7>
- Ruan, Y, Liu, Z, Wang, R and Yao, Z. 2019. Assessing the performance of CMIP5 GCMs for projection of future temperature change over the lower Mekong Basin. *Atmosphere*, 10(2): 93. DOI: <https://doi.org/10.3390/atmos10020093>
- Rumelhart, DE, Hinton, GE and Williams, RJ. 1985. *Learning internal representations by error propagation*. California Univ San Diego La Jolla Inst for Cognitive Science. DOI: <https://doi.org/10.21236/ADA164453>
- Shao, W and Kam, J. 2020. Retrospective and prospective evaluations of drought and flood. *Science of The Total Environment*, 748: 141155. DOI: <https://doi.org/10.1016/j.scitotenv.2020.141155>
- Sharma, A and Goyal, MK. 2020. Assessment of drought trend and variability in India using wavelet transform. *Hydrological Sciences Journal*, 65(9): 1539–1554. DOI: <https://doi.org/10.1080/02626667.2020.1754422>

- Su, B, Huang, J, Mondal, SK, Zhai, J, Wang, Y, Wen, S, ... and Li, A.** 2021. Insight from CMIP6 SSP-RCP scenarios for future drought characteristics in China. *Atmospheric Research*, 250: 105375. DOI: <https://doi.org/10.1016/j.atmosres.2020.105375>
- Tam, BY, Szeto, K, Bonsal, B, Flato, G, Cannon, AJ and Rong, R.** 2019. CMIP5 drought projections in Canada based on the Standardized Precipitation Evapotranspiration Index. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 44(1): 90–107. DOI: <https://doi.org/10.1080/07011784.2018.1537812>
- Van Der Maaten, L.** 2014. Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1): 3221–3245.
- Wang, R, Wu, H, Wu, Y, Zheng, J and Li, Y.** 2021. Improving influenza surveillance based on multi-granularity deep spatiotemporal neural network. *Computers in Biology and Medicine*, 134: 104482. DOI: <https://doi.org/10.1016/j.combiomed.2021.104482>
- Wu, C, Yeh, PJF, Ju, J, Chen, YY, Xu, K, Dai, H, ... and Huang, G.** 2021. Assessing the spatiotemporal uncertainties in future meteorological droughts from CMIP5 models, emission scenarios, and bias corrections. *Journal of Climate*, 34(5): 1903–1922. DOI: <https://doi.org/10.1175/JCLI-D-20-0411.1>
- Yao, N, Li, Y, Lei, T and Peng, L.** 2018. Drought evolution, severity and trends in mainland China over 1961–2013. *Science of the Total Environment*, 616: 73–89. DOI: <https://doi.org/10.1016/j.scitotenv.2017.10.327>
- Ying, XU and Chong-Hai, XU.** 2012. Preliminary assessment of simulations of climate changes over China by CMIP5 multi-models. *Atmospheric and Oceanic Science Letters*, 5(6): 489–494. DOI: <https://doi.org/10.1080/16742834.2012.11447041>
- Zhang, GP.** 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50: 159–175. DOI: [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Zhang, Y, Yang, H, Cui, H and Chen, Q.** 2020. Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China. *Natural Resources Research*, 29(2): 1447–1464. DOI: <https://doi.org/10.1007/s11053-019-09512-6>
- Zoonomia Consortium.** 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature*, 587(7833): 240. DOI: <https://doi.org/10.1038/s41586-020-2876-6>

TO CITE THIS ARTICLE:

Yuanbin, S, Qamar, S, Ali, Z, Yang, T, Nazeer, A and Fayyaz, R. 2022. A New Ensemble Index for Extracting Predictable Drought Features from Multiple Historical Simulations of Climate. *Tellus A: Dynamic Meteorology and Oceanography*, 74(2022), 236–249. DOI: <https://doi.org/10.16993/tellusa.46>

Submitted: 10 March 2022 Accepted: 10 March 2022 Published: 21 April 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Tellus A: Dynamic Meteorology and Oceanography is a peer-reviewed open access journal published by Stockholm University Press.

