



# Enhancing a machine learning model for predicting agricultural drought through feature selection techniques

Pardis Nikdad<sup>1</sup> · Mehdi Mohammadi Ghaleni<sup>1,2</sup> · Mahnoosh Moghaddasi<sup>1,2</sup> · Biswajeet Pradhan<sup>3,4</sup>

Received: 2 January 2024 / Accepted: 16 April 2024  
© The Author(s) 2024

## Abstract

This study aims to determine the crucial variables for predicting agricultural drought in various climates of Iran by employing feature selection methods. To achieve this, two databases were used, one consisting of ground-based measurements and the other containing six reanalysis products for temperature ( $T$ ), root zone soil moisture (SM), potential evapotranspiration (PET), and precipitation ( $P$ ) variables during the 1987–2019 period. The accuracy of the global database data was assessed using statistical criteria in both single- and multi-product approaches for the aforementioned four variables. In addition, five different feature selection methods were employed to select the best single condition indices (SCIs) as input for the support vector regression (SVR) model. The superior multi-products based on time series (SMT) showed increased accuracy for  $P$ ,  $T$ , PET, and SM variables, with an average 47%, 41%, 42%, and 52% reduction in mean absolute error compared to SSP. In hyperarid climate regions, PET condition index was found to have high relative importance with 40% and 36% contributions to SPEI-3 and SPEI-6, respectively. This suggests that PET plays a key role in agricultural drought in hyperarid regions because of very low precipitation. Additionally, the accuracy results of different feature selection methods show that ReliefF outperformed other feature selection methods in agricultural drought modeling. The characteristics of agricultural drought indicate the occurrence of drought in 2017 and 2018 in various climates in Iran, particularly arid and semi-arid climates, with five instances and an average duration of 12 months of drought in humid climates.

**Keywords** Agricultural drought · Climates of Iran · Combined condition indices · Feature selection methods · Support vector regression

✉ Mehdi Mohammadi Ghaleni  
m-mohammadighaleni@araku.ac.ir

Pardis Nikdad  
pardiiis78nk@gmail.com

Mahnoosh Moghaddasi  
m-moghaddasi@araku.ac.ir

Biswajeet Pradhan  
Biswajeet.Pradhan@uts.edu.au

<sup>1</sup> Department of Water Science and Engineering, Faculty of Agriculture and Environment, Arak University, Arak 38156-88138, Iran

<sup>2</sup> Research Institute for Water Science and Engineering, Arak University, Arak, Iran

<sup>3</sup> Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia

<sup>4</sup> Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

## Introduction

Agricultural drought monitoring is a pre-requirement for agricultural management and is of great importance for sustainable development, especially in arid regions. Therefore, it is crucial to develop an integrated agricultural drought index to understand the evolution of agricultural drought monitoring, forecasting, risk assessment, and crisis management (Feng et al. 2019; Tian et al. 2022). Challenges related to monitoring agricultural drought include, but are not limited to, obtaining the required datasets. Global gridded and satellite products with high spatial resolution and temporally continuous datasets can be very useful in monitoring any type of drought, particularly agricultural drought (Al-Yaari et al. 2019).

One of the main challenges in using reanalysis or other types of global products for drought analysis is dealing with errors in these products compared to ground-based measurements (Ma et al. 2019; Fan et al. 2022). Despite extensive

research on error analysis of global products for precipitation (Azizi et al. 2020; Saemian et al. 2021; Ghomlaghi et al. 2022; Fooladi et al. 2023), temperature (Gleixner et al. 2020; Rodrigues 2021; Zhang et al. 2021; Xu et al. 2023), evapotranspiration (Shirmohammadi-Aliakbarkhani and Saberali 2020; Elnashar et al. 2021; Ochege et al. 2021; Panahi et al. 2021; Zhu et al. 2022; Liu et al. 2023; Yao et al. 2023), and soil moisture (Li et al. 2020, 2022; Wu et al. 2021; Fan et al. 2022; Huang et al. 2022; Sion et al. 2022), there are still research gaps when it comes to the combination of global products (Allies et al. 2022; Min et al. 2022). Many studies have evaluated combination models to increase the accuracy of multi-product approaches compared to single-product approaches (Fooladi et al. 2021, 2023; Jiao et al. 2021; Chen et al. 2022). In recent years, multi-product datasets have been used for developing integrated drought indices (Jiao et al. 2021; Alkaraki and Hazaymeh 2023a).

The past few decades have been accompanied by a transformation in methods for developing drought indices, shifting from single input variables (e.g., standardized precipitation index (SPI) (McKee et al. 1993) and soil moisture index (SSI) (Hao and AghaKouchak 2013)) to multi-variate indices (e.g., standardized precipitation evapotranspiration index (SPEI) (Vicente-Serrano et al. 2010), standardized precipitation temperature index (SPTI) (Wable et al. 2019), and new multivariate standardized drought index (MSDI) (Hao and AghaKouchak 2013)). Additionally, indices based on parametric probability distribution functions have been replaced by nonparametric frameworks that describe drought more effectively (Hao and Aghakouchak 2014; Farahmand and AghaKouchak 2015; Alizadeh and Nikoo 2018). Furthermore, the development of combined condition indices (CCIs) such as vegetation health index (VHI) (Kogan 1995), scaled drought condition index (SDCI) (Rhee et al. 2010b), microwave integrated drought index (MIDI) (Zhang and Jia 2013), and integrated drought monitoring index (IDMI) (Arun Kumar et al. 2021) has been widely advanced using single condition indices (SCIs) such as precipitation condition index (PCI) (Rhee et al. 2010a), temperature condition index (TCI) (Kogan 1995), potential evapotranspiration condition index (PETCI) (Allen et al. 2007), soil moisture condition index (SMCI) (Zhang and Jia 2013), and vegetation condition index (VCI) (Kogan 1995).

Identifying suitable variables that contribute to the prediction accuracy of drought indices poses significant challenges in drought management (Ghazipour and Mahjouri 2022). To develop accurate drought prediction models, prior to selection, feature selection (FS) methods have been employed to identify the most effective predictors (Feng et al. 2019). Although feature selection methods have been widely utilized in various disciplines, their application in drought research has been limited to a few studies, such as variable selection using random forests (VSURF) (Feng et al.

2019), principal component analysis (PCA) (Arun Kumar et al. 2021), and Boruta random forest (Jamei et al. 2023). This study attempts to bridge this gap by proposing a comprehensive framework for predicting agricultural drought using feature selection methods based on different climates. Various machine learning methods such as artificial neural networks (ANNs) (Deo and Şahin 2015; Feng et al. 2019) and support vector regression (SVR) (Azamathulla and Ghani 2011; Malik et al. 2021; Ghazipour and Mahjouri 2022) have been employed thus far.

Based on the comprehensive review of the literature, there is a potential to enhance the performance of integrated agricultural drought indices. This study makes significant contributions to the literature by (1) presenting a comprehensive accuracy evaluation by combining six products (i.e., CRU TS4.05, TerraClimate, ERA5, MERRA2, GLDAS2.1, and GLEAM3.6a) for four key agrometeorological variables including mean temperature (T), precipitation (P), root-zone soil moisture (SM), and potential evapotranspiration (PET) using in situ observations under various climates in Iran; (2) determining the significance of key drought variables for predicting agricultural drought based on various FS methods; (3) developing SVR models with the best predictors for accurate agricultural drought prediction.

This study developed a machine learning feature selection-based model for predicting agricultural drought prediction multi-product datasets. The main research aims included (i) determining the superior combination of products for P, T, PET, and SM variables with the highest accuracy using ground-based measurements; (ii) finding appropriate drought-affecting variables for predicting drought indices between SCIs, and (iii) comparing new integrated agricultural drought indices based on SVR FS-based with CCIs (i.e., VHI, SDCI, MIDI, and IDMI) using statistical criteria and drought characteristics.

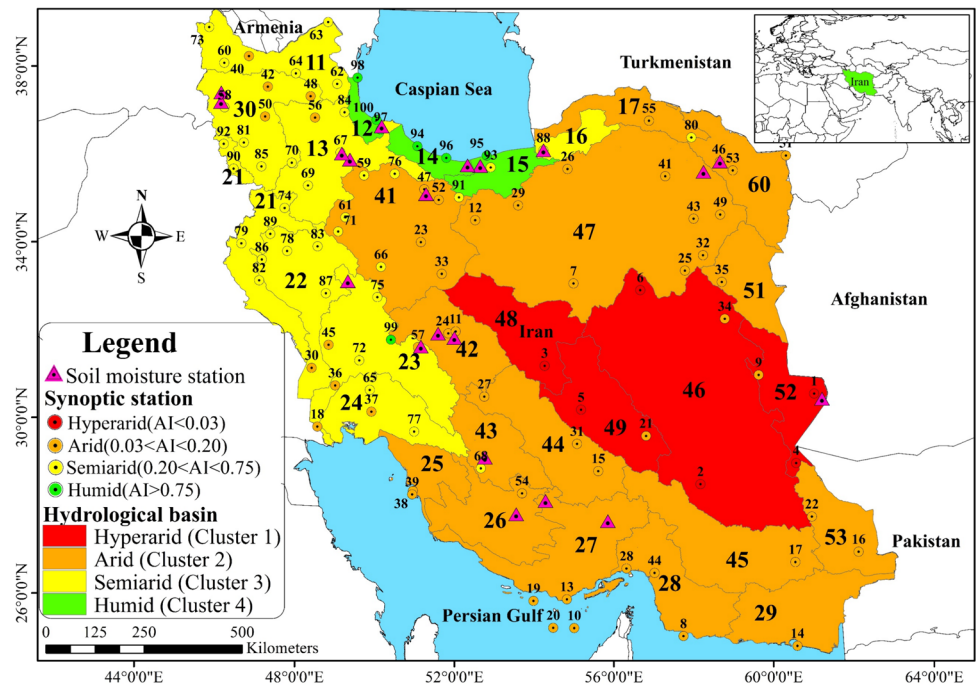
## Material and methods

### Case study

The study area, Iran, is located between the latitude of 25°–40° N and longitude 44°–63° E (Fig. 1). As the second biggest country in the Middle East, Iran has a population of ~82 million and covers an area of 1,648,195 km<sup>2</sup> (Bazrafshan and Cheraghalizadeh 2021; Rezaei 2021). The lowest and the highest points in Iran are 29 m below sea level and 5597 m above mean sea level located on the southern coast of the Caspian Sea and Mount Damavand, respectively (Bazrafshan and Cheraghalizadeh 2021).

The mean annual  $P$  in the 1987–2019 period was 357 mm, and the difference between the lowest (52 mm  $y^{-1}$ ) and highest (1692 mm  $y^{-1}$ )  $P$  was 1642 mm (Madani

**Fig. 1** Climate classification of the stations and basins based on the aridity index across the study area



2014; Moshir Panahi et al. 2020). PET for the study site was computed based on the FAO 56 Penman–Monteith (PMF56) method (Allen et al. 1998). The mean difference of PET in Iran is approximately 2868 mm y<sup>-1</sup> due to climate diversity from the Caspian Sea’s southern coast to the southeastern regions in Iran (Sharafi and Mohammadi Ghalehi 2021).

The aridity index (AI) value for the study site was calculated based on the ratio of mean annual P to the mean annual PET ( $AI = P/PET$ ). This index was used to categorize the aridity of the basins, including hyperarid ( $AI \leq 0.03$ ), Arid ( $0.03 < AI \leq 0.2$ ), semiarid ( $0.2 < AI \leq 0.75$ ), and humid ( $AI > 0.75$ ) regions (UNESCO 1979; Tsiros et al. 2020). In this study, AI was used to classify the hydrological sub-basin climates of the case study (Fig. 1).

Iran has 30 main hydrological sub-basins, 9 of which (37%) were classified as hyperarid (HA), 11 (42%) as arid (AR), 7 (18%) as semiarid (SA), and 3 (2%) as humid (HU) climates (Fig. 1). The specifications of these basins are presented in Table 1.

The workflow of the research is summarized in six steps (Fig. 2).

### Sources of datasets

The research datasets were obtained from three sources including ground-based measurements, reanalysis products, and remote sensing datasets. These data sources are described in the following subsections.

### Ground-based measurements

Two types of ground-based datasets, including synoptic and agrometeorological station measurements, were sourced from Iran's Meteorological Organization (i.e., IRIMO <https://data.irimo.ir/login/login.aspx>). The daily meteorological variables including  $P$ , minimum and maximum air temperature ( $T_{min}$  and  $T_{max}$ ), sunshine duration ( $S_d$ ), relative humidity (RH), and wind speed ( $U_2$ ) for the period 1987–2019 were obtained from IRIMO for 100 synoptic stations (Fig. 1). Daily PET was obtained using the FAO 56 Penman–Monteith (PMF56) method (Allen et al. 2006) and measured meteorological variables for 100 synoptic stations (Table 2).

Additionally, SM data in seven depths (including, 5, 10, 20, 30, 50, 70, and 100 cm below the soil surface) with a temporal resolution of 3-h from 43 agrometeorological stations during 2014–2021 were used (Table 2). Among the 43 stations, 20 stations were selected after preprocessing, based on having the least missing data for all depths within the 2014–2019 period. To determine the average root zone soil moisture (0–100 cm), the soil moisture measurements for each station from various depths are weighted after calculating the monthly soil moisture at all depths (Zhang et al. 2020; Xu et al. 2021; Ji et al. 2022). Weights were assigned to measurement depths based on the ratio of the distance between adjacent depths and the total root-zone depth. The assigned weights were 0.05, 0.05, 0.10, 0.10, 0.20, 0.20, and 0.30 for depth intervals of 0–5, 5–10, 10–20, 20–30, 30–50, 50–70, and 70–100 cm, respectively. In this analysis, various clusters or climates were analyzed (Table 1). The variation

**Table 1** Specifications of basins in the present study

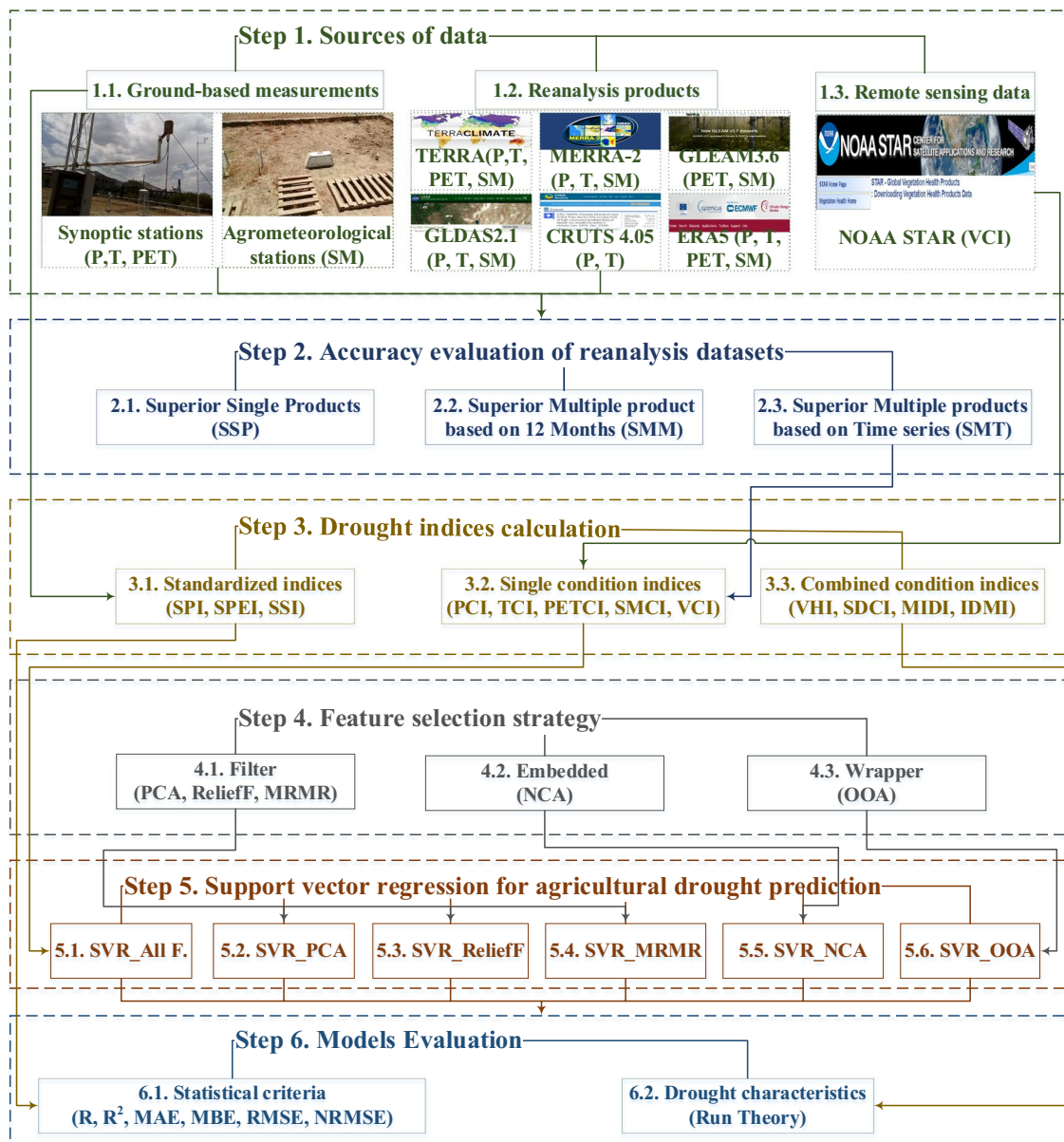
Cluster	Climate	AI value	Basin code	Basin name	Area (10 <sup>3</sup> km <sup>2</sup> )	<i>P</i> (mm year <sup>-1</sup> )	<i>T</i> (OC)	PET (mm year <sup>-1</sup> )
Cluster 1	Hyper-arid (HA)	0.03	46	Lut Desert	206	117	22.2	3819
		0.03	52	Hamun Hirmand	34	98	18.0	3130
		0.03	49	Daranjir Desert	51	108	23.4	3455
		0.03	48	Siahkuh Desert	49	123	26.1	3837
Cluster 2	Arid (AR)	0.04	53	Hamun Meshkil	37	117	24.5	3299
		0.04	29	Baluchestan	49	118	20.7	3177
		0.04	45	Hamun Jazmurian	69	123	20.1	3042
		0.06	28	Bandarabbas	45	156	19.0	2733
		0.06	27	Kal Mehran	63	166	21.5	2882
		0.06	44	Abarghu	57	149	19.9	2337
		0.08	51	Khaf	33	188	15.5	2472
		0.08	47	Central Desert	227	221	20.7	2855
		0.09	26	Mand Karian	48	245	22.9	2861
		0.11	43	Tashk Bakhtegan	31	243	15.5	2160
		0.12	60	Ghareghum	44	270	17.5	2182
		0.13	25	Heleh	21	308	16.6	2393
		0.18	41	Salt Lake	93	324	14.9	1791
		0.19	17	Atrak	26	327	15.0	1735
Cluster 3	Semiarid (SA)	0.19	42	Gavkhuni	42	271	12.6	1209
		0.20	11	Aras	40	331	23.2	1669
		0.20	30	Urmia	52	359	18.2	1808
		0.20	24	Jarrahi Zohreh	41	330	15.3	1850
		0.21	23	Karoon	67	451	24.9	2182
		0.27	22	Karkkeh	52	418	17.1	1570
		0.31	21	West Border	39	488	16.4	1570
		0.31	16	Gharesu Gorgan	13	361	12.1	1156
Cluster 4	Humid (HU)	0.38	13	Sefidrood	59	434	12.3	1146
		0.82	14	Sefidrood Haraz	11	938	16.4	1150
		0.92	15	Haraz Gharesu	19	588	16.3	1301
		1.02	12	Talesh	7	986	13.9	969

and anomalies of *P*, *T*, PET, and SM in four climates (i.e., HA, AR, SA, and HU) are presented in Fig. 3.

The yearly changes in variables (solid black line), trend of variation (dashed cyan line), and long-term anomalies (red and green bar charts) are presented in Fig. 3. A declining trend for *P* was detected across all climates (Fig. 3a1–a4), which influenced the overall precipitation trend in Iran. The negative precipitation anomalies during 2016–2018 were noticeable in hyper-arid (Fig. 3a1) and arid climates (Fig. 3a2). In recent years (2014–2019), the wettest year was 2019, with positive anomalies of 21, 67, 89, and 164 mm in HA, AR, SA, and HU climates, respectively. Meanwhile, 2017 was the driest year with anomalies of -35, -27, -150, and -328 mm from the long-term mean in HA, AR, SA, and HU climates, respectively. A clear upward trend in temperature was evident across all studied

climates (Fig. 3b1–b4) with positive temperature anomalies starting from 1998. The average annual temperature increased by 1.24, 0.57, 0.88, and 1.40 °C from 1987 to 2019 in HA, AR, SA, and HU climates, respectively.

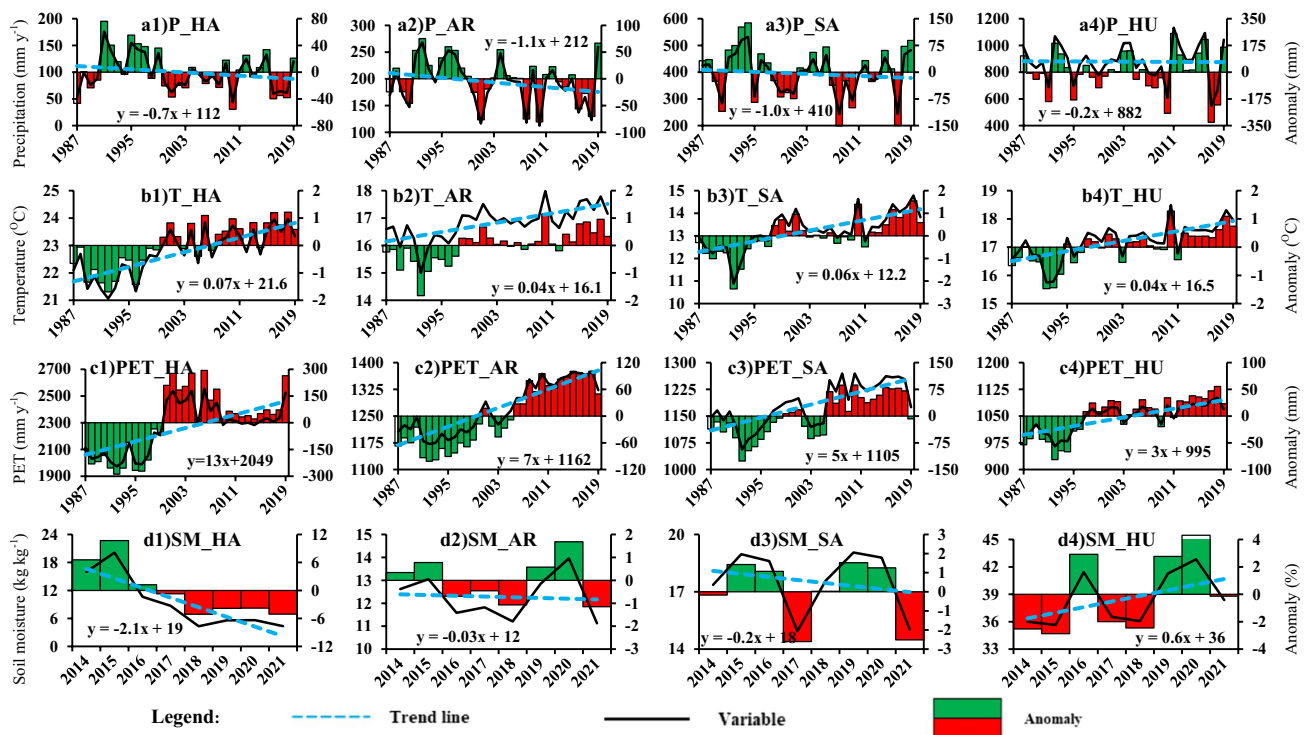
As observed, the increasing trend of PET (Fig. 3c1–c4) corresponds to an increasing trend of temperature in all climates. The PET exhibited a significant rising trend, with an approximately 13, 7, 5, and 3 mm per year increase from 1987 to 2019, in HA, AR, SA, and HU climates, respectively. Positive anomalies of PET were seen in 2000 in different climates. The variation in soil moisture showed a downtrend in all climates except for the humid climate (Fig. 3d4), which had an increasing trend. The greatest decrease in SM is associated with the highest increase in



**Fig. 2** Multistep workflow for agricultural drought prediction based on ensemble models with various feature selection methods

**Table 2** Ground-based measurement datasets used in the present study

Station	Duration	Temporal	Spatial	Variables	Sources
Synoptic	1987–2019	daily	100 st	$P, T_{\min}, T_{\max}, U_2, RH, S_d, PET_{PMF56}$	IRIMO
Agrometeorological	2014–2021	3-h	43 st	SM (5, 10, 20, 30, 50, 70 and 100 cm)	



**Fig. 3** Variation and anomaly of **a** P, **b** T, **c** PET, and **d** SM in four climates

T and PET, occurring in the hyperarid climate (Fig. 3d1). The negative anomaly of SM can be observed in 2017, indicating an agricultural drought.

**Reanalysis products**

In this study, six global gridded products including CRU TS4.05, ERA5, TerraClimate, MERRA-2, GLDAS2.1, and GLEAM 3.6a were used for two aims: (i) to evaluate the variable accuracy of these products compared to ground measurements and (ii) determine a superior combination of these products for calculating the drought condition indices.

A summary of this study’s product characteristics is demonstrated in Table 3.

In this study, a range of 0–100 cm was defined for the root-zone depth, and different datasets with varying soil layer depths were used. To obtain the averaged soil moisture in this range, the datasets were weighted based on their corresponding depths. For example, soil moisture measurements from GLDAS at depths of 0–10 cm, 10–40 cm, and 40–100 cm were weighted with values of 0.1, 0.3, and 0.6, respectively, to calculate the root zone soil moisture (Xu et al. 2021; Ji et al. 2022). Here, the TerraClimate dataset represents soil moisture as a root zone depth due to the absence of 100 cm depth segmentation in its variables. As

**Table 3** Global gridded products reanalysis-based specifications

Group	Data source	Product's name	Temporal	Spatial	Resolution	Variable	References
Interpolated	Gauge-based	CRU TS4.05	1901–2021	Global	0.5°	P, T, PET	Harris et al. (2020)
Modeled	Reanalysis	ERA5	1979–presents	Global	~0.28°	P, T, PET, SM (0–7, 7–28, 28–100 cm)	Hersbach et al. (2020)
	Ensemble, reanalysis, gauge	TerraClimate	1958–present	Global	~0.04°	P, T, PET, SM	Abatzoglou et al. (2018)
Multisource	Reanalysis, gauge, satellite	MERRA2	1958–present	Global	0.5°×0.625°	P, T, PET, SM (10–100 cm)	Gelaro et al. (2017)
		GLDASv.2.1	1980–present	Global	0.25°	P, T, PET, SM (0–10, 10–40, 40–100 cm)	Rodell et al. (2004)
		GLEAM3.6a	1980–2021	Global	0.25°	PET, SM	Martens et al. (2017)

for the GLEAM products, the downloaded variable for root-zone soil moisture was used.

### Remote sensing data

The VCI datasets used here are based on data from vegetation health products developed by the National Oceanic and Atmospheric Administration and processed by the Center for Satellite Applications and Research (i.e., NOAA STAR) ([https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh\\_ftp.php](https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_ftp.php)).

The vegetation health products include the VCI obtained from the radiance data captured by the advanced very high-resolution radiometer (AVHRR). The vegetation health products derived from AVHRR utilize the NOAA global area coverage dataset, spanning from 1981 to the present. The data and images in the vegetation health products (4-km spatial resolution) are composited weekly. The VCI time series from 1981 to 2021 was downloaded in the NETCDF format ([https://www.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/Blended\\_VH\\_4km/VH/](https://www.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/Blended_VH_4km/VH/)) with a weekly temporal resolution (Fernández-Tizón et al. 2020).

### Accuracy evaluation of single- and multi-products

The accuracy of different products varies with climate, elevation, and location (Fallah et al. 2020; Fooladi et al. 2023). To address this gap, it is recommended to create a combined dataset by integrating multiple products. This approach aims to enhance the accuracy of the resulting dataset. However, before combining products, it is necessary to evaluate single products against in situ observations to select the best options.

For a more accurate combination of products, observations of soil moisture as a key variable in agricultural drought prediction were classified in different climates based on AI. The accuracy of P, T, PET, and SM variables was assessed separately in each climate.

Then, the accuracy ratio of the six products (i.e., CRU, ERA, TerraClimate, MERRA, GLDAS, and GLEAM) for the four variables was assessed using three different methods, namely (i) superior single product (SSP), (ii) superior multiple products in each month from January to December (SMM), and (iii) superior multiple products in each month from first to end of time series (SMT).

In the SSP method, the product with the highest accuracy was selected for each of the four variables. In SMM, first, the time series of the variables was broken into 12 (months). Then, the product with the highest accuracy in each month was selected. In SMT, for each month from the start to the end of the time series, the products with the nearest value to the observations were selected. Figure 4 shows the violin plots of variables related to six products and three

combination-based approaches (i.e., SSP, SMM, and SMT) versus in situ observations for each variable (P, T, PET, and SM) in different climates.

Figure 4 indicates the violin plots of T, P, PET, and SM in four climates including HA, AR, SA, and HU related to six products and three combined-based approaches compared to observations.

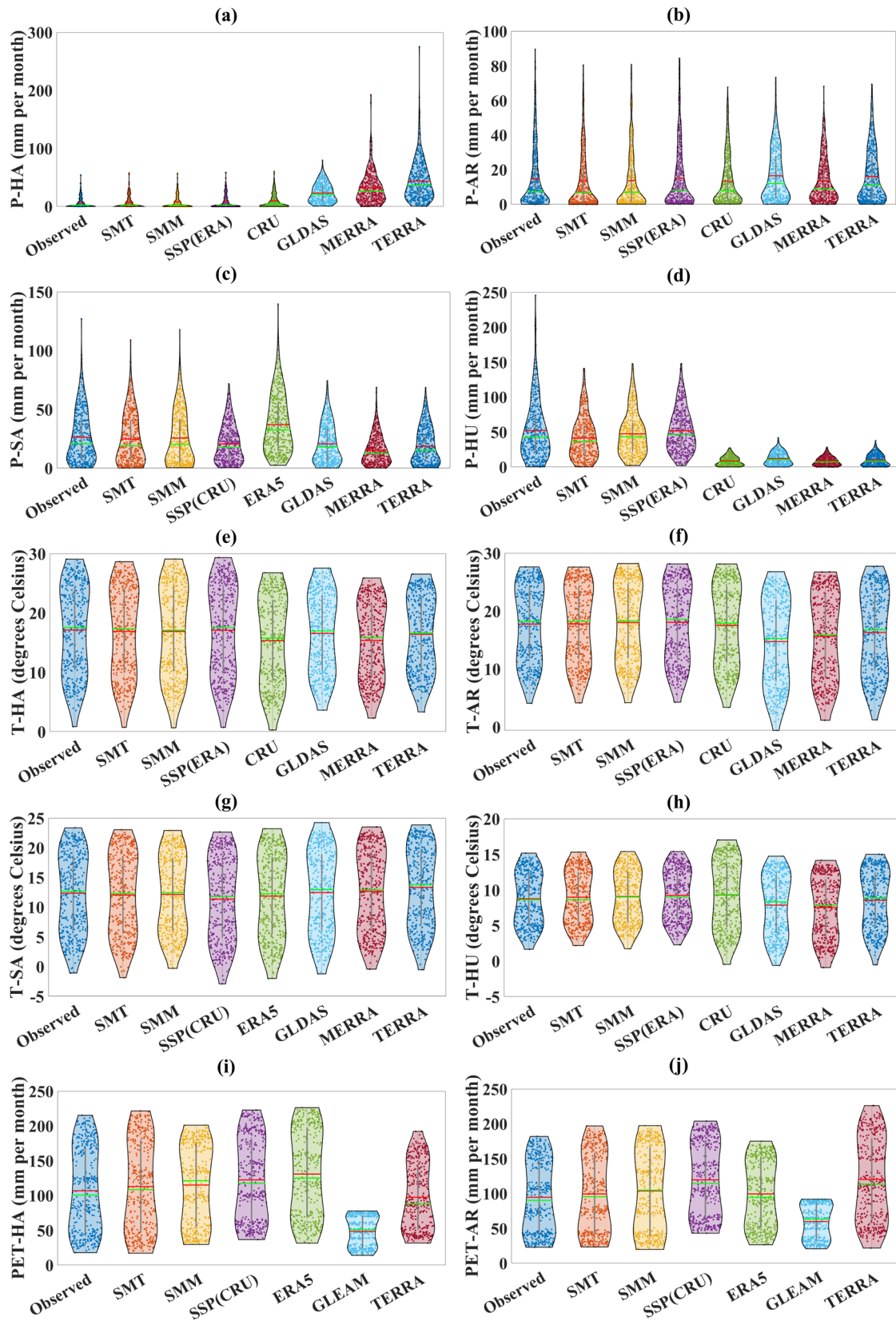
The mean and median values (represented by the red and green lines) indicate that the ranges of variable data across all climates close estimations to the observations. This demonstrates the accuracy of the combined products in capturing the variability in the variable datasets. In contrast, the results of single products for each variable deviate from observations in all climates. The combined approach, which involves integrating multiple reanalysis products, demonstrates a higher level of accuracy and lower uncertainty when compared to observations. As shown in Fig. 4 a-g, the superior single product for P and T variables was ERA5 in HA, AR, and HU climates, and CRU in the SA climate. SSP for PET (Fig. 4i-l) was CRU in the HA, AR, and SA climates and GLEAM in the HU climate. SSP for SM (Fig. 4p) was GLEAM in HA, AR, and SA (Fig. 4m-o) and ERA5 in HU. According to the violin plot in Fig. 4, the soil moisture data series in the reanalysis product exhibited a higher number of anomalies compared to other variables. This finding highlights the complexity of modeling soil moisture, particularly in the root zone depth, which is a crucial variable for predicting agricultural drought.

### Drought indices

The drought indices used in this research were divided into three categories including standardized indices (SIs), SCIs, as well as CCIs. SIs include SPEI, SPI, and SSI, which were calculated using ground-based data, as target indices for agricultural drought prediction. SCIs were assumed as features (inputs or predictors), which are calculated based on the most accurate combined products (SMT procedure) to model agricultural drought. These included the PCI, TCI, PETCI, SMCI, and VCI. To assess the accuracy of predicting agricultural indices, several widely used combination indices such as VHI, SDCI, IDMI, and MIDI were taken into account. These three types of indices are explained in the following sections.

#### Single condition indices (SCIs)

Many studies have used single products to calculate SCIs, such as calculating PCI using global precipitation climatology center (GPCC) (Zhang et al. 2017), climate hazards group infrared precipitation with stations (CHIRPS) (Kumar et al. 2021), global precipitation measurement (GPM) (Alkaraki and Hazaymeh 2023b), and tropical



**Fig. 4** Probability distribution function of single and combined products compared to observed variables as violin plots in climates



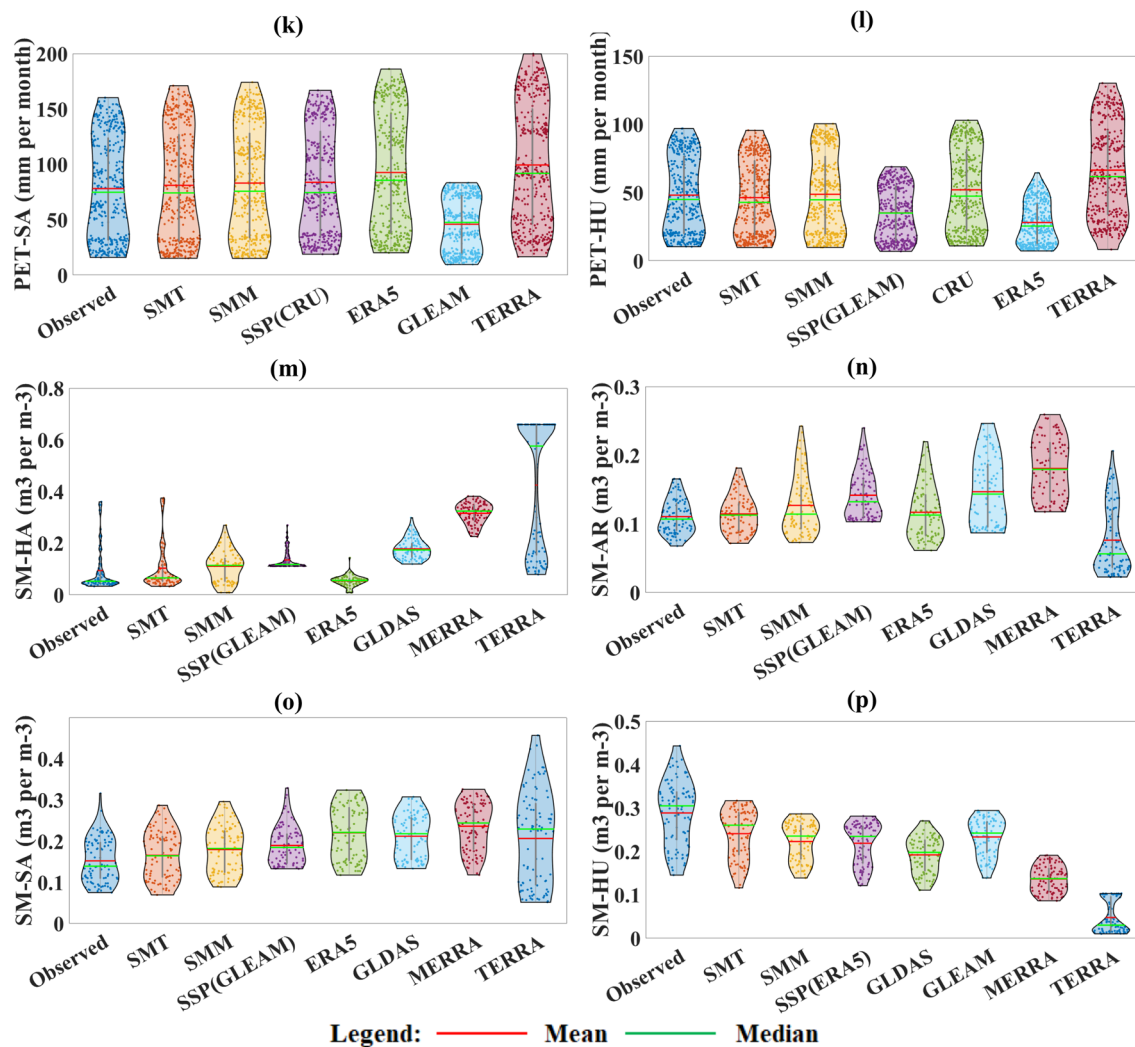


Fig. 4 (continued)

rainfall measuring mission (TRMM) (Cai et al. 2023; Zhang et al. 2022a, b), TCI using MODIS (Arun Kumar et al. 2021; Zhang et al. 2022b; Alkaraki and Hazaymeh 2023b; Cai et al. 2023), PETCI using MODIS (Alkaraki and Hazaymeh 2023b), SMCI using GLDAS (Zhang et al. 2017), ESA-CCI (Arun Kumar et al. 2021), and SMAP(Alkaraki and Hazaymeh 2023b), and VCI using AVHRR (Zhang et al. 2017) and MODIS (Arun Kumar et al. 2021; Cai et al. 2023; Zhang et al. 2022a, b). In this study, the selected five SCIs including PCI, TCI, PETCI, SMCI, and VCI were calculated using the SMT datasets. The formulas for obtaining SCIs are listed in Table 4.

All these SCIs range from zero to one, with smaller value indicating a more severe drought (Zhang et al. 2022a). For all five SCIs, multiple timescales including 1, 3, 6, 9, and 12 months, were calculated in all climates. In this way, by calculating five SCIs in five-time scales, a matrix with 25 columns (features) by 72 rows (monthly

datasets from Jun-14 to Dec-19 based on SM ground-based data) was prepared as an input matrix for the prediction model in each climate.

### Standardized indices (SIs)

So far, several studies have used standardized indices at multiple timescales, especially SPI (Zhang et al. 2017; Alkaraki and Hazaymeh 2023b; Yin and Zhang 2023) and SPEI (Tian et al. 2020; Ali et al. 2022; Yang et al. 2023), to evaluate the newly developed indices. To avoid probability distribution function fitting, which may not always be the best selection of a distribution function, a nonparametric approach was proposed for deriving SIs (Farahmand and AghaKouchak 2015; Fooladi et al. 2021). The SIs including SPI, SPEI, and SSI were assumed for target selection. To achieve this aim, first,  $SPI_n$ ,  $SPEI_n$ , and  $SSI_n$  were calculated at multiple timescales (i.e.,  $n = 1, 3, 6, 9,$  and  $12$  months) based on ground

**Table 4** Descriptions of single condition indices used in the study

Index	Formula	References	Equation number
PCI	$PCI_{i,j,k} = \frac{PSMT_{i,j,k} - PSMT_{i,j,\min}}{PSMT_{i,j,\max} - PSMT_{i,j,\min}}$	Rhee et al. (2010b)	(1)
TCI	$TCI_{i,j,k} = \frac{LSTSMT_{i,j,k} - LSTSMT_{i,j,\min}}{LSTSMT_{i,j,\max} - LSTSMT_{i,j,\min}}$	Kogan (1995)	(2)
PETCI	$PETCI_{i,j,k} = \frac{PETSMT_{i,j,k} - PETSMT_{i,j,\min}}{PETSMT_{i,j,\max} - PETSMT_{i,j,\min}}$	Allen et al. (2007)	(3)
SMCI	$SMCI_{i,j,k} = \frac{SMSMT_{i,j,k} - SMSMT_{i,j,\min}}{SMSMT_{i,j,\max} - SMSMT_{i,j,\min}}$	Zhang and Jia (2013)	(4)
VCI	$VCI_{i,j,k} = \frac{NDVI_{i,j,k} - NDVI_{i,j,\min}}{NDVI_{i,j,\max} - NDVI_{i,j,\min}}$	Kogan (1995)	(5)

The *i*, *j*, and *k* in the above formulas represent the *i*-th pixel, the *j*-th month, and the *k*-th year, respectively. PSMT<sub>*i,j,min*</sub>, LSTSMT<sub>*i,j,min*</sub>, PETSMT<sub>*i,j,min*</sub>, SMSMT<sub>*i,j,min*</sub>, and PSMT<sub>*i,j,max*</sub>, LSTSMT<sub>*i,j,max*</sub>, PETSMT<sub>*i,j,max*</sub>, SMSMT<sub>*i,j,max*</sub> denote the multi-year minimum and maximum of pixel *i* in month *j* of the most accurate precipitation, LST, PET, and soil moisture from global gridded products. In addition, NDVI<sub>*i,j,min*</sub> and NDVI<sub>*i,j,max*</sub> represent the multi-year minimum and maximum of the pixel *i* in month *j* of NDVI

measurements. Then, the correlation between SCIs and SIs was examined at various timescales to determine which SI timescale is more correlated with SCIs and select optimal SIs as agricultural target indices.

Between SPI, PSEI, and SSI at multiple timescales, SPEI-3 and SPEI-6 were selected as target indices for agricultural drought prediction. Some studies have selected SPEI-3 (Guo et al. 2019) or SPEI-6 (Tian et al. 2020) as agricultural target indices. The SPI index does not account for the influence of global warming and climate change (Kazemzadeh et al. 2022). SPEI is considered more useful for assessing drought events as it takes into account both precipitation and evapotranspiration, providing a comprehensive perspective on drought conditions. This study employed SPEI-3 with a seasonal accumulation period and SPI-6 with a semi-yearly accumulation period to characterize seasonal drought and semi-annual agricultural drought events.

**Combined condition indices (CCIs)**

Table 5 presents four CCIs derived from weighted linear combinations of SCIs. The CCIs include the MIDI, SDCI, VHI, and IDMI. The formulas for calculating these CCIs, along with their respective references, are provided in Table 5.

**Table 5** Characteristics of combined condition indices used in the study

Index	Equation	References	Eq. number
VHI	$VHI = 0.5 \times VCI + 0.5 \times TCI$	Kogan (1995)	(6)
SDCI	$SDCI = 0.25 \times VCI + 0.25 \times TCI + 0.5 \times PCI$	Rhee et al. (2010a)	(7)
MIDI	$MIDI = 0.2 \times TCI + 0.3 \times SMCI + 0.5 \times PCI$	Zhang and Jia (2013)	(8)
IDMI	$IDMI = PCA(PCI, TCI, SMCI, VCI)$	Arun Kumar et al. (2021)	(9)

The linear combination method, similar to the SCIs illustrated in Table 5, is the simplest method for generating new multivariate drought indices (Guo et al. 2019). However, calculating the weights of different variables due to various accumulation periods can be challenging (Hao and Singh 2015). This study's aims were accomplished by considering the research background on linear weighted CCIs by (a) adding PETCI as one of the most important variables for developing an integrated agricultural condition drought index to IDMI, (b) considering accumulation period of SCIs for accurate agricultural drought prediction using multi-timescale SCIs as predictors, and (c) clustering the datasets based on different climates to consider the effects of climates on the relations between key variables.

**Feature selection strategy**

The FS strategy plays a crucial role in drought prediction by identifying the most relevant and informative features or variables that contribute to accurate prediction of drought events (Feng et al. 2019; Rostami et al. 2021). Feature selection methods have several advantages such as reducing variance, increasing the model's accuracy, and shortening training time (Lap et al. 2023). The feature selection methods in this research were applied for two reasons; (i) selecting

the more efficient and relevant SCIs (i.e., PCI, TCI, PETCI, SMCI, VCI) or predictors for agricultural drought prediction in various climates and (ii) increasing the accuracy of the machine learning prediction model by eliminating redundant features.

Feature selection methods can be classified into three primary categories based on the level of supervision available: unsupervised, semi-supervised, and supervised. Additionally, these methods are categorized into five groups based on the selection strategy employed: Filter, Embedded, Wrapper, Hybrid, and Ensemble (Alhenawi et al. 2022). In this study, five feature selection methods including PCA, ReliefF, minimal redundancy maximal relevance (MRMR), neighborhood components analysis (NCA), and Osprey algorithm optimization (OOA) were used as appropriate predictors in agricultural drought prediction. Among the five selected FS methods, PCA, ReliefF, and MRMR belong to the Filter group, NCA to the Embedded group, and OOA to the Wrapper group. These FS methods are explained briefly in the following sections.

### Filter methods

Within filter methods, features are chosen based on the statistical characteristics of the data, without employing any learning models. In this particular model, features are assessed and prioritized using information-theoretic measures, and subsequently, the features with the highest ranks are chosen (Labani et al. 2018). In this study, three methods including PCA, ReliefF, and MRMR from filter approaches were used for selecting the best inputs in agricultural drought modeling.

PCA is a mathematical method utilized to decrease the dimensionality of a dataset (Jackson 1983). Due to its simplicity and nonparametric nature, PCA has been extensively employed in various types of analyses as a means of extracting pertinent information from intricate datasets (Wold et al. 1987; Arun Kumar et al. 2021). In PCA, as a feature selection method, large absolute factor loadings indicate that the corresponding variables have a greater impact on the factor compared to other variables (Guo et al. 2002).

The ReliefF algorithm (Kira and Rendell 1992), initially developed for binary classification problems, has been improved to address multiclass problems through the ReliefF algorithm (Kononenko 1994). The ReliefF algorithm was developed by adapting ReliefF to address continuous class or regression problems (Robnik-Šikonja and Kononenko 1997). In this research, the SVR model was constructed using the first  $k$ -PC factors, and for each factor, the top five SCIs with the highest loadings were chosen as inputs for the PCA.

This study employed ReliefF and MRMR, widely recognized univariate filters (Alhenawi et al. 2022). The ReliefF method chooses random instances and looks for a specified

number of nearest neighbors with the same class, as well as  $K$ -nearest neighbors with different classes. After that, the process is repeated for each feature by computing the average of all classes over a specific number of iterations (Zhang et al. 2003; Alhenawi et al. 2022). The MRMR method chooses features that exhibit higher relevance (demonstrating a strong correlation with the target class) and lower redundancy (displaying a weak correlation with other features) (Mandal and Mukhopadhyay 2013).

### Embedded strategy

In the embedded model, a learning algorithm is employed to explore and find the optimal set of features (Zhang et al. 2015). As an embedded strategy, the NCA algorithm (Goldberger et al. 2005) was incorporated in this study as one of the learning methods. NCA does not rely on any parametric assumptions regarding the data distribution and can naturally handle multiclass problems (Yang et al. 2012). Here, NCA was used for selecting SCIs with more relevance to agricultural drought index targets (i.e., SPEI-3 and SPEI-6).

### Wrapper strategy

In the wrapper approach, a search method is employed to discover the best feature subset. At each step of the search strategy, a subset of features is generated and evaluated using a classifier or another learning model. Ultimately, the optimal generated feature subset is chosen as the final feature set (Chandrashekar and Sahin 2014; Rostami et al. 2021). The OOA, developed by Dehghani and Trojovský (2023), is inspired by the hunting behavior of Osprey. The algorithm mimics the hunting strategies of ospreys, showcasing an exceptional ability to optimize and swiftly converge. Drawing upon the inborn capability of ospreys, the algorithm presented in this study demonstrates a remarkable ability to efficiently identify optimal solutions (Hu et al. 2023). In this study, OOA was applied as a search strategy for selecting more relevant SCIs in each climate or clustering region (Fig. 1).

### Developed indices

SVR, functioning as a supervised learning mechanism, employs the concept of structural risk minimization to improve the handling of multi-dimensional problems. The SVR approach utilizes linear equations within the simulation algorithm, resulting in improved performance by effectively employing a kernel function. In the SVR method, a non-linear mapping of  $\phi$  is computed in the trait space, where  $X_t$  represents the input data in  $R_m$  and  $Y(X_t)$  represents the output data in  $R$  (Drucker et al. 1996; Vapnik 1999):

$$Y(X_t) = w^T \times \phi(X_t) + b \tag{10}$$

Here, the weights ( $w$ ) and biases ( $b$ ) represent the values of the regression function, which are determined by minimizing the following function.

$$\min_{w,b,e_i} j(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=1}^n e_t^2 \tag{11}$$

$$S.T : w^T \times \phi(X_t) + b \quad t = 1, 2, \dots, n$$

In Eq. (11),  $e$  represents the errors, while gamma ( $\gamma$ ) corresponds to the regularization parameters utilized in the model to regulate the smoothness of the approximation function. The optimal figures for these parameters are defined by the users. A review of the literature shows that SVR is commonly used for agricultural drought prediction (Tian et al. 2018; Prodhan et al. 2022).

Here, SVR was performed with six input datasets including all 25 features, i.e., five SCIs (PCI, TCI, PETCI, SMCI, and VCI), by five multi-timescales (1, 3, 6, 9, and 12 months) as five SCIs selected using PCA, RReliefF, MRMR, NCA, and OOA FS methods. The predicted agricultural drought indices with each input dataset were compared with SPEI-3 and SPEI-6 as target agricultural indices in different climates using statistical criteria.

To prioritize the selection of pertinent inputs in the study, the cross-correlation coefficient approach is employed to assess the time delay between SPEI3 and SPEI6, and each input variable. The analysis reveals that the lag times for each index (PCI, TCI, PETCI, SMCI, and VCI) in relation to SPEI3 and SPEI6 mostly fall within the same month (zero lag time) across the majority of the four climates. Furthermore, the correlation coefficient between the chosen input indices for the preceding 0–12 months and the current SPEI3 and SPEI6 generally diminishes as the lag time increases, particularly beyond 3 months, in most regions examined. Consequently, the input indices are considered to be contemporaneous with SPEI3 and SPEI6 in the modeling process, without any lag time.

### Drought characteristics

The run theory (Herbst et al. 1966) is commonly used for drought event identification. Typically, a threshold of -1 is

employed to identify drought events. Droughts generally occur when there is a persistence of below-average precipitation for a duration exceeding three consecutive months (Liu et al. 2016; Haile et al. 2020). A drought event was defined here as a period lasting for at least three consecutive months, during which the drought index values remained below -1. Three classes of drought including moderate, severe, and extreme are distinguished for different indices based on Table 6.

Drought characteristics such as the number of events ( $N$ ), duration ( $D$ ), frequency ( $F$ ), and intensity ( $I$ ) were calculated for each drought event using the categories outlined in Table 6. The drought duration of a drought is computed using Eq. (12) (Xu et al. 2019):

$$D = \frac{\sum_{i=1}^n d_i}{n} \tag{12}$$

The  $D$  is determined by summing the durations of individual drought events ( $d_i$ ) for each station, where  $n$  represents the total drought events. The drought frequency is calculated through Eq. (13) (Spinoni et al. 2014; Wang et al. 2018):

$$F = \frac{n_m}{N_m} \times 100\% \tag{13}$$

$F$  is calculated using drought months ( $n_m$ ) divided by total months ( $N_m$ ). The drought intensity ( $I$ ) is defined by Eq. (14) (Wang et al. 2018; Haile et al. 2020).

$$I = \left| \frac{1}{n} \sum_{i=1}^n DI_i \right| \tag{14}$$

$I$  is determined by summing the accumulated drought index values below the threshold (-1) for each drought event ( $i$ ), where  $n$  represents the total number of drought events. In this study, drought characteristics including  $N$ ,  $D$ ,  $F$ , and  $I$  were calculated for target indices (i.e., SPEI-3 and SPEI-6), CCIs (i.e., VHI, SDCI, MIDI, and IDMI) and six modeled indices using SVR (i.e., all features as well as PCA, ReliefF, MRMR, NCA, and OOA methods) in various climates from Jan-14 to Dec-19.

**Table 6** Drought classification based on SIs and CCIs used in this study

Indices (references)	Category	Threshold	Indices (references)	Category	Threshold
SPEI, SPEI, SSI, IDMI (McKee et al. 1993; Vicente-Serrano et al. 2010; Arun Kumar et al. 2021; Salimi et al. 2021)	No drought (ND)	> -1	VHI, SDCI, MIDI (Kogan 1995; Rhee et al. 2010a; Zhang and Jia 2013)	No drought (ND)	> 0.4
	Moderate (MD)	(-1.5, -1]		Moderate (MD)	(0.3, 0.4]
	Severe (SD)	(-2, -1.5]		Severen (SD)	(0.2, 0.3]
	Extreme (ED)	≤ -2		Extreme (ED)	≤ 0.2

**Evaluation criteria**

Various performance criteria have been employed thus far to assess the quality of reanalysis datasets and evaluate the accuracy of prediction results. In this study, the accuracy of the reanalysis data and each model's predictions was evaluated using six statistical measures: the Pearson correlation coefficient (R), coefficient of determination (R<sup>2</sup>), mean bias error (MBE), root mean square error (RMSE), mean absolute error (MAE), and normalized root mean square error (NRMSE). The explanations for each of the statistical measures are provided in the following section.

$$R = \frac{\sum_{t=1}^N (M_t - \bar{M})(E_t - \bar{E})}{\sqrt{\left[\sum_{t=1}^N (M_t - \bar{M})^2\right] \left[\sum_{t=1}^N (E_t - \bar{E})^2\right]}} \tag{15}$$

$$R^2 = \left[ \frac{\sum_{t=1}^N (M_t - \bar{M})(E_t - \bar{E})}{\sqrt{\left[\sum_{t=1}^N (M_t - \bar{M})^2\right] \left[\sum_{t=1}^N (E_t - \bar{E})^2\right]}} \right]^2 \tag{16}$$

$$MBE = \frac{1}{N} \sum_{t=1}^N (E_t - M_t) \tag{17}$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |E_t - M_t| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (E_t - M_t)^2} \tag{19}$$

$$NRMSE = \frac{RMSE}{\bar{M}} \tag{20}$$

**Table 7** Accuracy of single and combined time series using reanalysis products

Climate		Hyperarid		Arid		Semiarid		Humid	
Variable	Product	MAE	R2	MAE	R2	MAE	R2	MAE	R2
P	CRU	5.93	0.64	8.62	0.64	15.24	0.63	83.54	0.13
	GLDAS	23.03	0.14	18.52	0.23	24.47	0.36	81.24	0.02
	MERRA	32.76	0.05	17.40	0.16	26.01	0.27	85.51	0.03
	TERRA	45.21	0.08	17.72	0.24	25.40	0.35	84.30	0.01
	SSP(ERA5)	4.81	0.66	8.03	0.72	17.68	0.65	43.20	0.59
	SMM	4.59	0.64	7.11	0.73	11.86	0.66	41.74	0.59
	SMT	2.48	0.84	3.90	0.88	6.74	0.82	32.15	0.73
T	CRU	2.07	0.99	1.30	0.98	1.56	0.88	2.85	0.97
	GLDAS	4.73	0.90	7.12	0.93	5.20	0.92	5.12	0.90
	MERRA	5.47	0.93	7.68	0.94	5.66	0.92	5.56	0.91
	TERRA	4.76	0.93	7.20	0.94	5.12	0.94	4.51	0.91
	SSP(ERA5)	1.15	0.99	1.22	0.98	1.51	0.88	1.23	0.99
	SMM	1.17	0.99	1.12	0.97	1.27	0.87	1.07	0.98
	SMT	0.70	0.99	0.69	0.98	0.75	0.98	0.84	0.99
PET	ERA5	29.94	0.94	48.69	0.78	28.05	0.85	43.72	0.66
	GLDAS	69.22	0.92	39.42	0.83	52.20	0.89	26.99	0.95
	MERRA	50.78	0.90	39.52	0.91	31.27	0.88	52.42	0.92
	SSP(CRU)	25.86	0.95	29.86	0.94	17.91	0.86	16.99	0.98
	SMM	30.69	0.80	28.01	0.84	21.36	0.79	18.83	0.95
	SMT	16.04	0.95	15.19	0.94	9.98	0.89	10.60	0.98
SM	ERA	0.05	0.00	0.06	0.34	0.09	0.56	0.12	0.59
	GLDAS	0.10	0.16	0.07	0.33	0.08	0.55	0.14	0.58
	MERRA	0.22	0.13	0.09	0.21	0.12	0.33	0.20	0.29
	TERRA	0.33	0.03	0.10	0.16	0.19	0.30	0.32	0.13
	SSP(GLEAM)	0.07	0.21	0.05	0.38	0.06	0.48	0.10	0.48
	SMM	0.06	0.20	0.05	0.35	0.06	0.48	0.11	0.37
	SMT	0.02	0.94	0.02	0.62	0.04	0.71	0.07	0.63

In Eqs. (15) to (20),  $M_t$  and  $\bar{E}_t$  are the variables based on measured and modeled data,  $\bar{M}$  and  $\bar{E}$  are the mean values of measured and modeled variables, respectively, and N represents the number of datasets.

## Results and discussions

### Accuracy evaluation of products

A single reanalysis product has uncertainties in temporal and spatial spaces, which can be reduced by combining multiple products. The results in Table 7 indicated that SMT can significantly improve the accuracy of variables in all climates.

The accuracy of SMT based on MAE for the P variable showed an increase of 49%, 51%, 62%, and 26% in comparison with superior single products (SSP\_ERA5) in HA, AR, SA, and HU climates, respectively. Overall, the accuracy ranking with regard to MAE and  $R^2$  of the five P and T products was  $SMT > SMM > SSP(ERA5) > CRU > GLDAS > TERRA > MERRA$ .

The CRU product was selected as the SSP for PET in all climates. The mean absolute error was reduced by 10%, 15%, 18%, and 6%, regarding SMT in comparison with SSP for PET in HA, AR, SA, and HU climates, respectively. The single products, even for SSP, showed a low correlation with ground-based soil moisture measurements, especially in the HA climate. In contrast, SMT indicated high accuracy with a maximum MAE of 7% for soil moisture volume in HA climates.

Taylor diagrams, originally introduced by Taylor (2001), were utilized here to provide a succinct summary and overview of the performance of the reanalysis products. The Taylor diagram (Fig. 5) presents three statistical indices, namely root mean square deviation (RMSD), standard deviation (SD), and R. Figure 5 illustrates the Taylor diagram for each individual and combined dataset, categorized by different climates in Iran.

As shown in Fig. 5a1–a4, the Taylor diagrams provide a visual representation of the accuracy of precipitation data for each individual reanalysis product dataset as well as two combined datasets (SMM and SMT) compared to ground-based measurements across all climates. In this regard, it can be observed that both combined datasets (SMM and SMT)

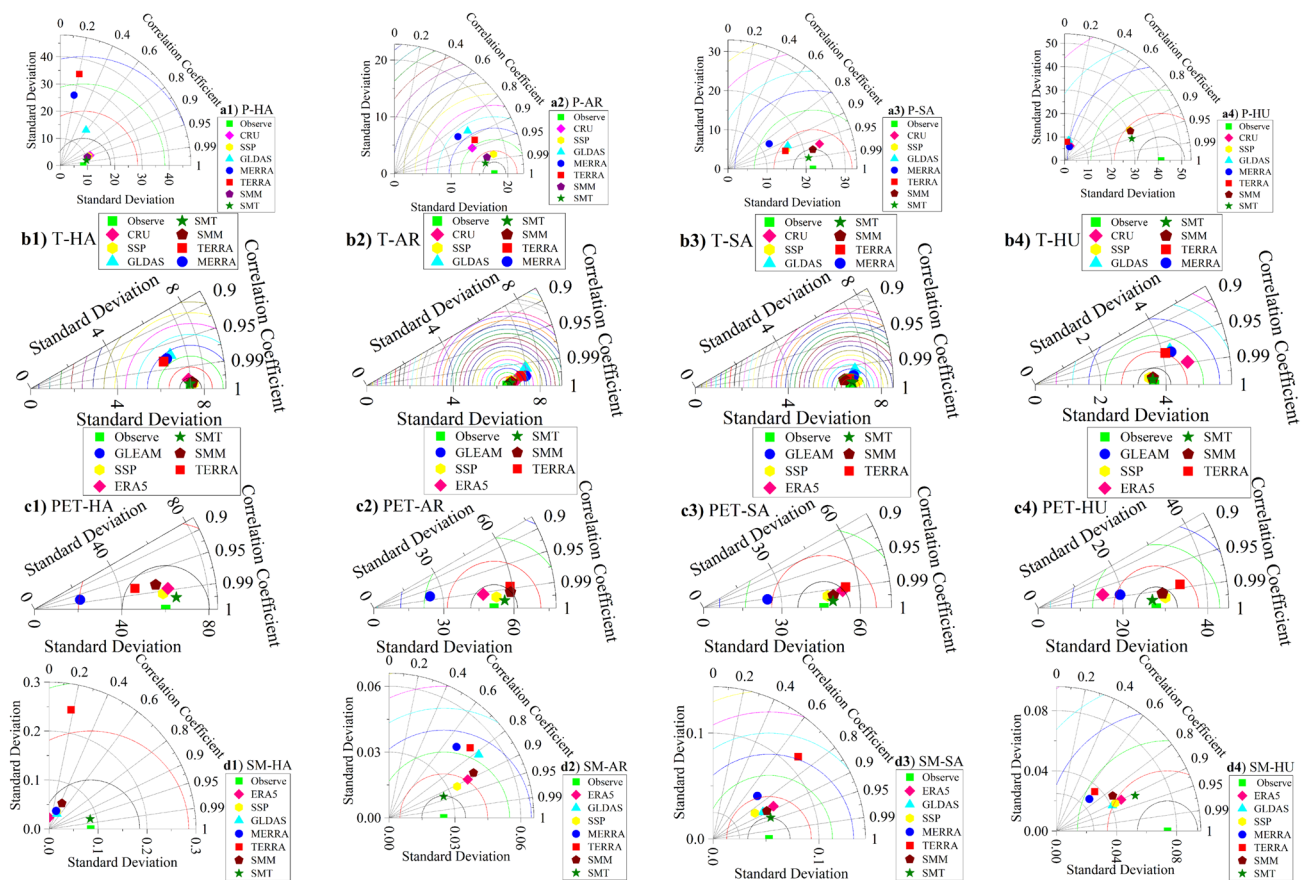


Fig. 5 Taylor diagram of single and combined reanalysis products for each climate

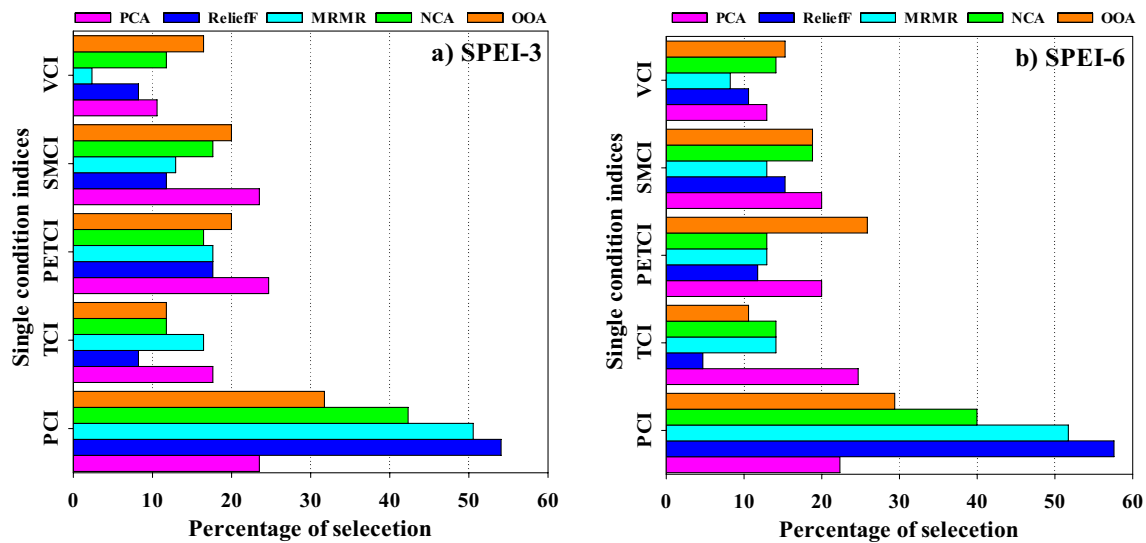


Fig. 6 Average percentage of selection of SCIs in each FS method for SPEI-3 and SPI-6

exhibit higher accuracy compared to single precipitation products. Furthermore, the SMT procedure demonstrated superior performance compared to SMM in terms of precipitation data across all climates.

As observed from the Taylor diagrams, ERA5 and CRU products exhibited better agreement with ground-based measurements of P, T, and PET. This is evident from their higher R and lower RMSD when compared to the ground-based data series. The SD and RMSD are crucial indicators for evaluating dataset accuracy. In this context, both combined datasets exhibit reduced RMSD, while maintaining a similar SD for all variables. This highlights the effectiveness of the combined datasets in improving the accuracy of the data. In general, Fig. 5 demonstrates that the SMT generated using single reanalysis products achieves a notable enhancement in the accuracy of variable time series when compared to ground-based measurements.

### Feature selection results

Figure 6a, b demonstrates the average percentage of SCIs in each FS method, in which a specific feature is selected for SPEI-3 and SPI-6, respectively. The results for SPEI-3 (Fig. 6a) and SPEI-6 (Fig. 6b) show similarity in the average percentage of selection by SCIs in each FS method. As can be seen among all SCIs, PCI, PETCI, and SMCI were the most important and selected 40, 18, and 17 percent of the features through the five feature selection methods. PCI has been reported by other researchers as the most important index in agricultural drought prediction (Arun Kumar et al. 2021; Zhang et al. 2022b; Cai et al. 2023). Out of the five SCIs, TCI and VCI were selected less than other SCIs for agricultural drought prediction in SPEI-3 and SPEI-6.

In Table 8, the average percentage of relative importance for SCIs is provided separately for each climate. These results show that PCI, in all climates except HA, has the highest relative importance among SCIs, which is marked with a double underline.

Table 8 Average percentage of the relative importance of SCIs for each target and climate

Climate SCIs	Hyperarid		Arid		Semiarid		Humid	
	SPEI-3	SPEI-6	SPEI-3	SPEI-6	SPEI-3	SPEI-6	SPEI-3	SPEI-6
PCI	16.0	<u><b>20.0</b></u>	<u><b>40.5</b></u>	<u><b>42.0</b></u>	<u><b>47.2</b></u>	<u><b>40.8</b></u>	<u><b>37.3</b></u>	<u><b>41.3</b></u>
TCI	<u><b>21.0</b></u>	16.0	15.5	8.0	9.6	<u><b>19.2</b></u>	10.7	18.7
PETCI	<u><b>40.0</b></u>	<u><b>36.0</b></u>	<u><b>20.5</b></u>	16.5	14.4	14.4	17.3	14.7
SMCI	4.0	12.0	15.5	<u><b>21.5</b></u>	<u><b>16.0</b></u>	9.6	<u><b>28.0</b></u>	<u><b>20.0</b></u>
VCI	19.0	16.0	8.0	12.0	12.8	16	6.7	5.3

The highest relative importance is indicated by text that is both bold and double underlined for both the SPEI-3 and SPEI-6 indices. Additionally, the bold and underlined number signifies the highest relative importance in either the SPEI-3 or SPEI-6 index

**Fig. 7** Scatter plots of modeled and target indices SPEI-3 (a1–d1) and SPEI-6 (a2–d2) for all CCI in different climates

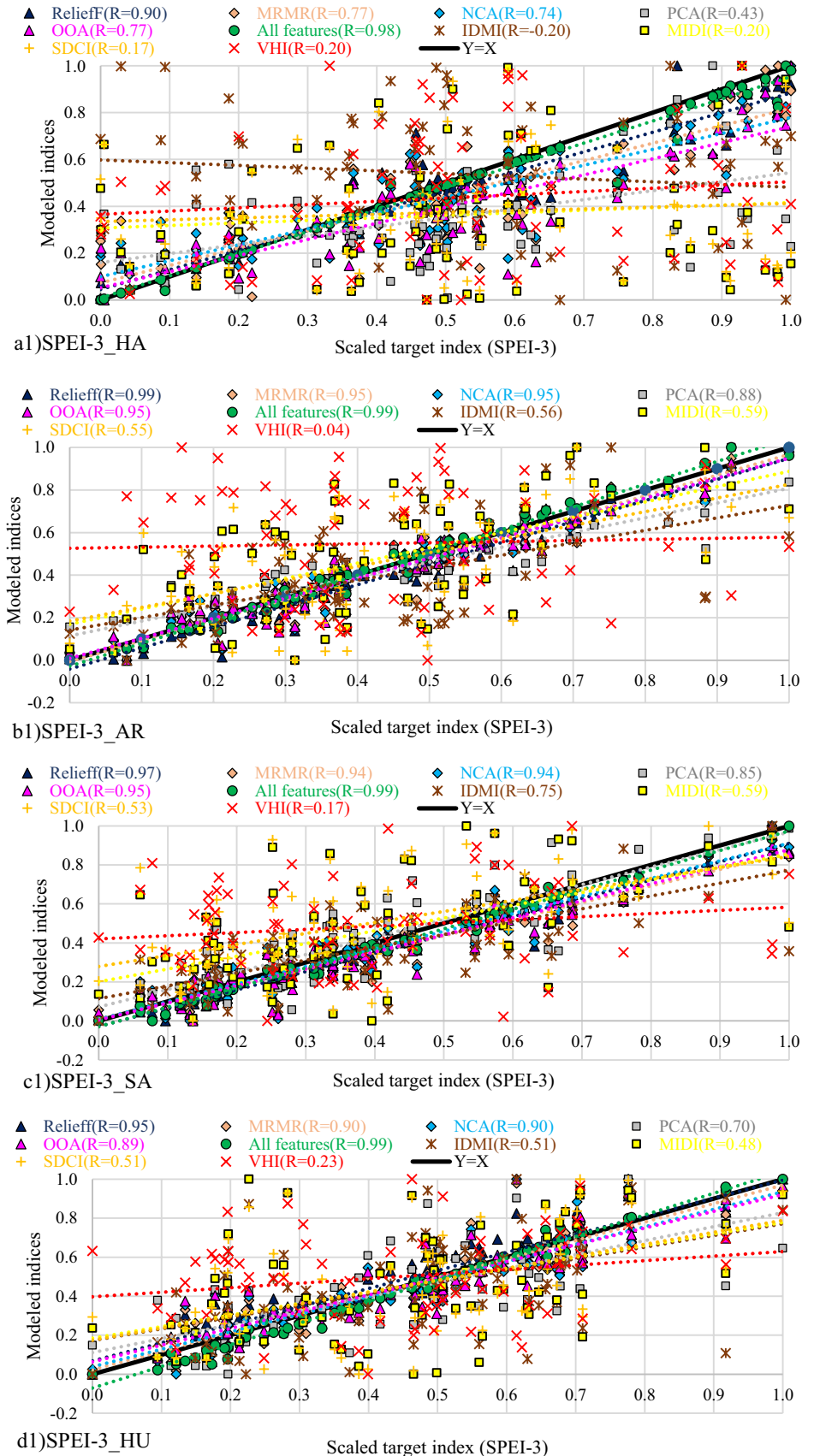
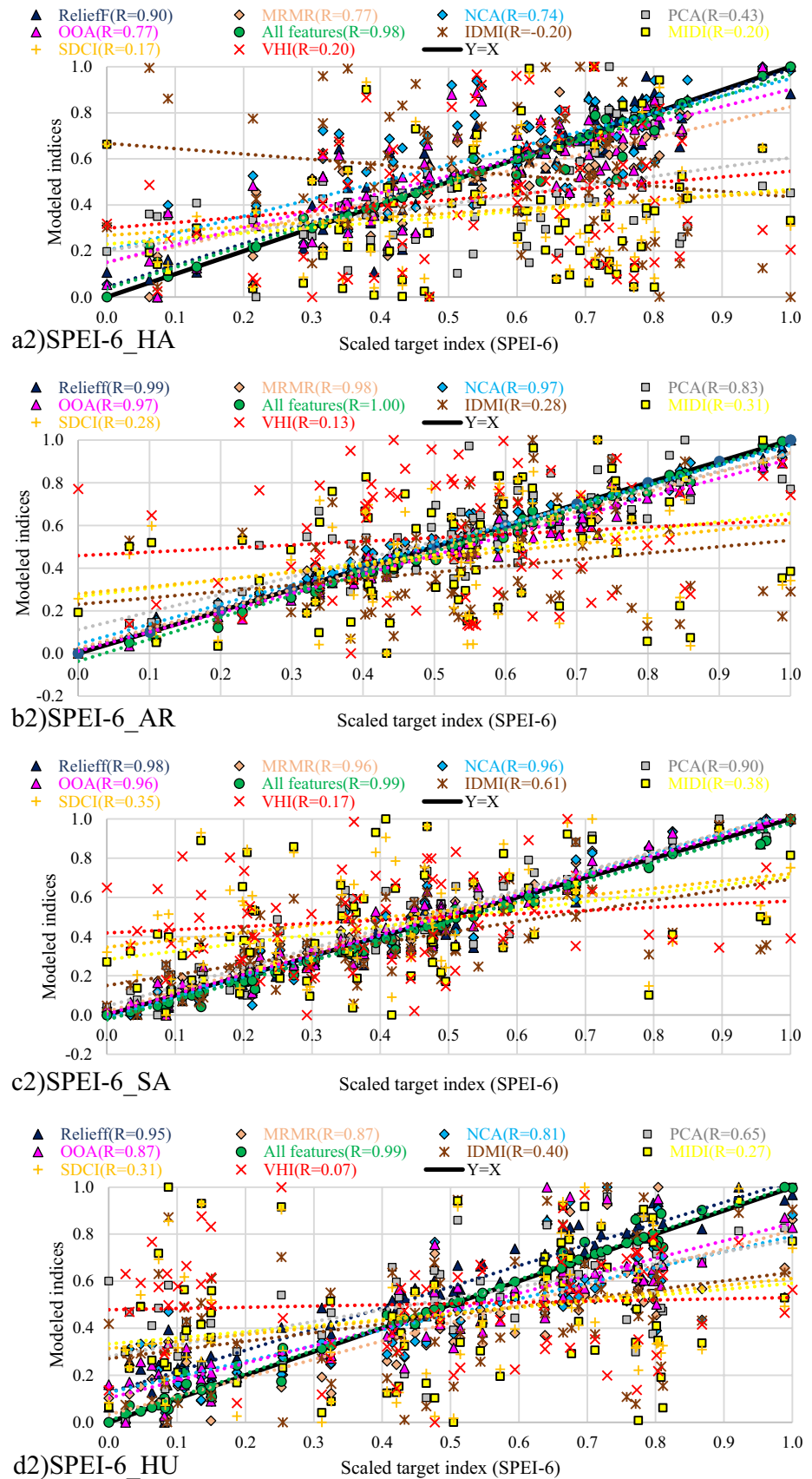




Fig. 7 (continued)



The highest relative importance of SCIs in the HA climate belongs to PETCI with 40 and 36 percent for SPEI-3 and SPEI-6, respectively. This shows that PET plays a crucial role in agricultural drought prediction in hyperarid regions because of very low precipitation. The relative importance of SMCi increases from HA to HU climate, which shows the greater effect of soil moisture in humid regions on agricultural drought. The correlation coefficients as scatter plots of modeled and target indices SPEI-3 and SPEI-6 for all CCIs are illustrated in Fig. 7a1–d1 and a2–d2.

The highest and lowest R in Fig. 7 belongs to SVR with all features as inputs (SVR-all f.). Furthermore, three CCIs (i.e., SDCI, MIDI, and IDMI) showed a similar range of R with an average correlation coefficient of 0.44. Besides, R had higher values in the SA climate (see Fig. 7c1 and c2), indicating the higher accuracy of agricultural drought indices in semiarid regions compared to other climates.

### Drought characteristics

The intensity for three drought classes, including extreme drought (dark red color), severe drought (red color), and moderate color (orange color), is analyzed based on the drought classes of Table 6. These drought events were shown for all combined condition indices (SVR models and CCIs) compared to SPEI-3 and SPEI-6 in all agrometeorological stations from Jan. 14 to Dec. 19. The most severe

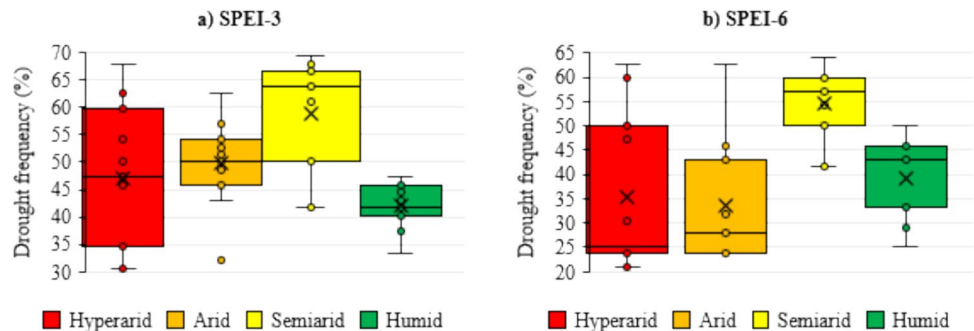
agricultural droughts occurred during Jun. 17 and May 18 (for SPEI-3) and Sept. 17 and May 18 (for SPEI-6).

The highest similarity between SPEI-3 and SPEI-6 as target agricultural drought indices and other indices belonged to SVR-all f. Moreover, SVR with inputs from the ReliefF method demonstrated the same drought events as target indices. Among all indices, VHI, IDMI, and PCA showed more disagreement with SPEI-3&6 in determining drought classes, especially in the period 2017–2018. The variations in the drought frequency based on studied indices are illustrated in Fig. 8 separately for each climate.

The multiplication signs (×) in box plots of Figure 8 show the average drought frequency in all studied drought indices including targets (i.e., SPEI-3&6), CCIs (i.e., VHI, SDCI, MIDI, and IDMI), and SVR results with different inputs (i.e., all features, PCA, ReliefF, MRMR, NCA, OOA). The average frequency of these agricultural drought indices for HA, AR, SA, and HU climates was 47, 50, 59, and 42 and 35, 33, 54, and 39 percent in 3- and 6-month timescales (Fig. 8a-b). The mean duration of drought events based on SVR and CCIs indices versus targets is shown in Table 9 for each climate.

By increasing the time scales of SPEI from 3 to 6 months, drought duration has increased significantly, especially in humid climates from 12.0 to 21.0 months. So far, many studies (Mesbahzadeh 2020; Kazemzadeh et al. 2022; Naderi and Moghaddasi 2022; Sharafi and Ghaleni

**Fig. 8** Box plots of drought frequency for studied indices in different climates



**Table 9** Duration of agricultural drought (month) based on CCIs and SVR indices in different climates

Target index	Clim type	Target	CCIs					SVR- based FS				
			VHI	SDCI	MIDI	IDMI	PCA	ReliefF	MRMR	NCA	OOA	All f
SPEI-3	HA	4.5	6.0	5.0	5.2	3.5	4.7	<b>4.6</b>	5.0	5.6	6.4	4.2
	AR	6.0	3.0	4.3	4.3	4.4	6.6	4.9	5.0	5.3	<b>6.0</b>	5.4
	SA	8.2	4.3	4.3	6.3	6.3	10.7	9.2	9.8	<b>8.0</b>	9.6	10.8
	HU	12.0	4.0	5.3	4.4	3.5	7.3	21.0	10.0	7.7	9.0	<b>13.0</b>
SPEI-6	HA	4.7	6.0	5.0	5.2	3.5	4.4	7.0	5.3	10.0	10.0	<b>4.7</b>
	AR	6.5	3.0	4.3	4.3	4.4	10.0	<b>7.5</b>	7.6	10.0	8.0	8.0
	SA	11.7	4.3	4.3	6.3	6.3	12.7	12.3	12.3	13.3	10.8	<b>11.7</b>
	HU	21.0	4.0	5.3	4.4	3.5	5.0	19.0	11.7	9.0	12.5	<b>21.0</b>

Bold numbers indicate the nearest value of drought duration among various indices to target values

2022) indicate a decrease in the number of events and drought frequency and an increase in severity and drought duration by increasing the timescale of drought indices.

Results of Table 9 indicated that SVR with all features as input had the nearest duration compared to the target indices. Meanwhile, SVR had the nearest duration to target indices in SPEI-3 for HA, AR, and SA climates and in SPEI-6 for AR climate with input features from ReliefF, OOA, NCA, and ReliefF.

### Discussion

In our study, we evaluated various feature selection techniques to determine their effectiveness in accurately monitoring drought across different climate regions in Iran. The results of our analysis indicate that ReliefF surpassed

other methods in agricultural drought modeling due to its capacity to handle multiclass issues and its robustness in managing incomplete and noisy data (Bolón-Canedo et al. 2013). The best SVR results using ReliefF methods beside target indices are presented in Fig. 9.

As shown in Fig. 9a1, a2, an increasing trend of drought intensity is observed in the hyperarid climate, especially for SPEI-6 (Fig. 9a2). The temporal drought intensity in semiarid and humid climates was severely low in comparison with other climates, especially in 2019.

Also, the findings underscore the significant role of potential evapotranspiration (PET) as a key predictor for agricultural drought modeling compared to other climatic variables especially in hyper-arid climates. PET is impacted by atmospheric water content, meaning that in cases of adequate water supply, potential evapotranspiration primarily mirrors atmospheric conditions (Abrar Faiz et al. 2022), a

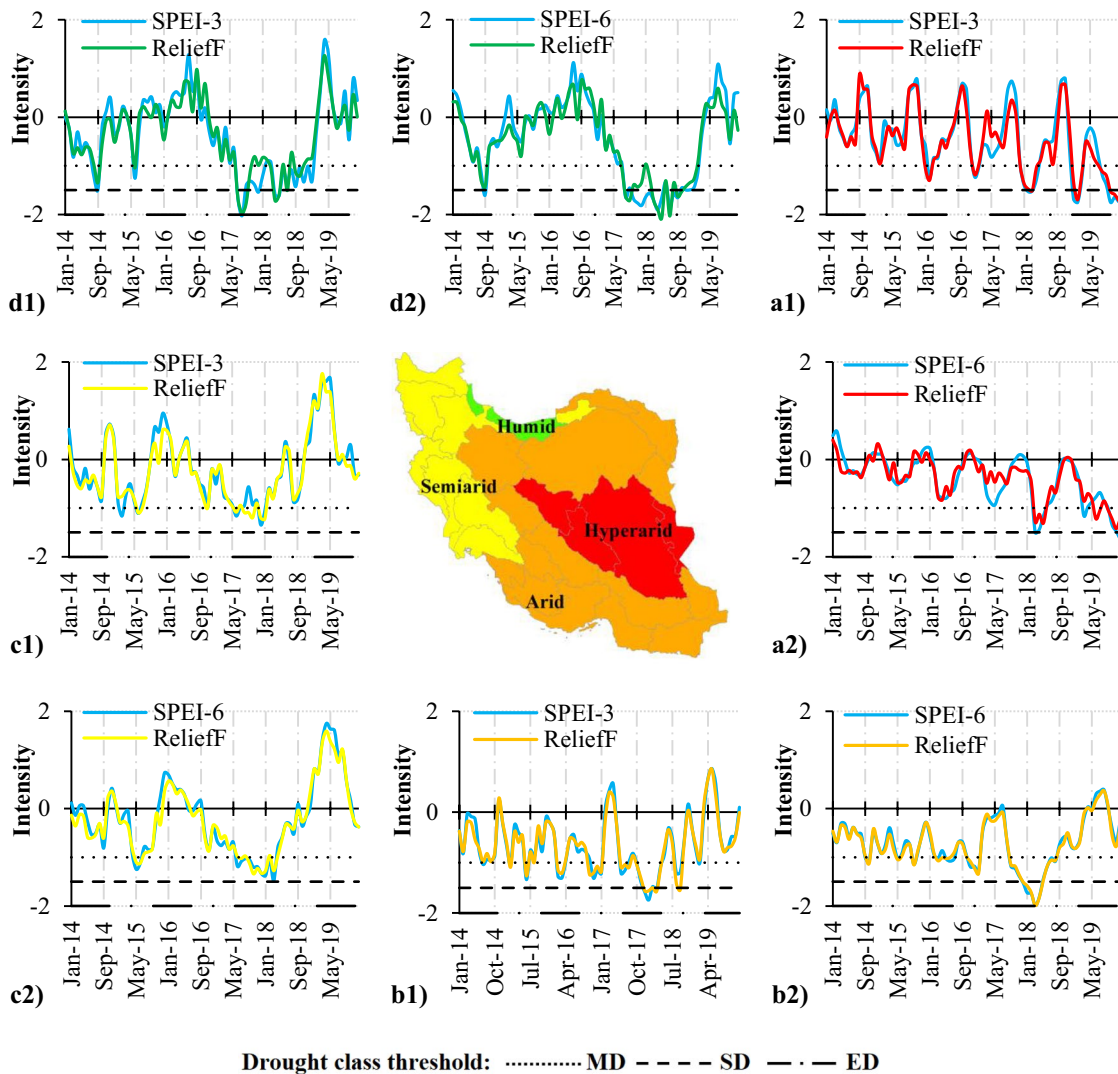


Fig. 9 Drought intensity in SVR ReliefF-based versus SPEI-3 and SPEI-6 indices in different climates

scenario commonly observed in humid climates. One reason for the heightened importance of PET in hyper-arid climates is the limited water supply in these regions and the crucial role of PET in the water balance within hyper-arid environments.

Before PET, P is considered the most critical feature for agricultural drought modeling across various climates. This is because precipitation is a primary determinant of water availability for agricultural activities and plant growth. In most climates, the amount, frequency, and distribution of precipitation play a crucial role in assessing drought conditions and their impact on agricultural productivity (Guo et al. 2019).

## Conclusion

Considering to spatial and temporal uncertainties of global gridded datasets based on single products, the combined datasets from six popular products significantly improved the accuracy of datasets compared to situ measurements. Moreover, the accuracy evaluation demonstrated the varying performance of single products under different temporal (month) and spatial (climate) conditions.

Drought prediction and monitoring plans often rely on information obtained based on different key variables used in drought indices. In the present study, agricultural drought indices were developed through SVR with different feature selection methods to determine the most relevant SCIs in different climates. The relative importance of variables based on reanalysis datasets for five single condition indices (i.e., PCI, TCI, PETCI, SMCI, and VCI) was analyzed and compared with ground-based measurements of SIs. The performance of the SVR feature-based methods was assessed by comparing their results with four combined condition indices, specifically SPEI-3 and SPEI-6. The findings revealed that SVR-ReliefF results are consistent with SPEI and the model outperforms other feature selection (FS) methods. The results showed the higher relative importance of PETCI in hyperarid climates compared to other SCIs. These results highlight the key role of the PET variable as a suitable predictor of agricultural drought in regions with very low precipitation. However, in AR, SA, and HU climates P was the most important variable in agricultural drought prediction.

Future works are recommended to evaluate two-stage feature selection methods based on spatial and temporal variations for selecting more relevant variables in agricultural drought prediction. Moreover, it is recommended to consider the lag time in different variables for assessing drought propagation from meteorological to agricultural.

**Author contributions** All authors contributed to the study conception and design. Pardis Nikdad took part in conceptualization, data curation, formal analysis. Mehdi Mohammadi Ghaleini involved in investigation, writing original draft, software, visualization. Mahnoosh Moghaddasi involved in supervision; validation, review & editing. Biswajeet Pradhan involved in supervision, review & editing. All authors read and approved the final manuscript.

**Funding** The authors declare that no funds, grants, or other support was received during the preparation of this manuscript. The authors have no relevant financial or non-financial interests to disclose.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethical approval** We confirm that the present research has been conducted in accordance with ethical standards, and the final version of this research has been approved by all authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC (2018) TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci Data* 5:1–12. <https://doi.org/10.1038/sdata.2017.191>
- Abrar Faiz M, Zhang Y, Tian X et al (2022) Drought index revisited to assess its response to vegetation in different agro-climatic zones. *J Hydrol* 614:128543. <https://doi.org/10.1016/j.jhydrol.2022.128543>
- Alhenawi E, Al-Sayyed R, Hudaib A, Mirjalili S (2022) Feature selection methods on gene expression microarray data for cancer classification: a systematic review. *Comput Biol Med* 140:105051. <https://doi.org/10.1016/j.combiomed.2021.105051>
- Ali S, Liu D, Fu Q et al (2022) Constructing high-resolution ground-water drought at spatio-temporal scale using GRACE satellite data based on machine learning in the Indus Basin. *J Hydrol* 612:128295. <https://doi.org/10.1016/j.jhydrol.2022.128295>
- Alizadeh MR, Nikoo MR (2018) A fusion-based methodology for meteorological drought estimation using remote sensing data. *Remote Sens Environ* 211:229–247. <https://doi.org/10.1016/j.rse.2018.04.001>
- Alkaraki KF, Hazaymeh K (2023) A comprehensive remote sensing-based Agriculture Drought Condition Indicator (CADCI) using machine learning. *Environ Challenges*. <https://doi.org/10.1016/j.envc.2023.100699>
- Alkaraki KF, Hazaymeh K (2023) A comprehensive remote sensing-based Agriculture Drought Condition Indicator (CADCI) using

- machine learning. *Environ Challenges*. <https://doi.org/10.1016/j.envc.2023.100699>
- Allen RG, Pruitt WO, Wright JL et al (2006) A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. *Agric Water Manag* 81:1–22
- Allen RG, Tasumi M, Morse A et al (2007) Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—applications. *J Irrig Drain Eng* 133:395–406. [https://doi.org/10.1061/\(asce\)0733-9437\(2007\)133:4\(395\)](https://doi.org/10.1061/(asce)0733-9437(2007)133:4(395))
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration guidelines for computing crop water requirements.
- Allies A, Olioso A, Cappelaere B et al (2022) A remote sensing data fusion method for continuous daily evapotranspiration mapping at kilometer scale in Sahelian areas. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2022.127504>
- Al-Yaari A, Wigneron JP, Dorigo W et al (2019) Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements. *Remote Sens Environ* 224:289–303. <https://doi.org/10.1016/j.rse.2019.02.008>
- Arun Kumar KC, Reddy GPO, Masilamani P et al (2021) Integrated drought monitoring index: a tool to monitor agricultural drought by using time-series datasets of space-based earth observation satellites. *Adv Sp Res* 67:298–315. <https://doi.org/10.1016/j.asr.2020.10.003>
- Azamathulla HM, Ghani AA (2011) Genetic programming for predicting longitudinal dispersion coefficients in streams. *Water Resour Manag* 25:1537–1544. <https://doi.org/10.1007/s11269-010-9759-9>
- Azizi J, Rasoulzadeh A, Rahmati A et al (2020) Evaluating the performance of Era-5 Re-analysis data in estimating daily and monthly precipitation, case study: ardabil province. *Iran J Soil Water Res*. 51:2937–2951. <https://doi.org/10.22059/ijswr.2020.302176.668600>
- Bazrafshan J, Cheraghalizadeh M (2021) Verification of abrupt and gradual shifts in Iranian precipitation and temperature data with statistical methods and stations metadata. *Environ Monit Assess*. <https://doi.org/10.1007/s10661-021-08925-2>
- Bolón-Canedo V, Sánchez-Maróño N, Alonso-Betanzos A (2013) A review of feature selection methods on synthetic data. *Knowl Inf Syst* 34:483–519. <https://doi.org/10.1007/s10115-012-0487-8>
- Cai S, Zuo D, Wang H et al (2023) Assessment of agricultural drought based on multi-source remote sensing data in a major grain producing area of Northwest China. *Agric Water Manag*. <https://doi.org/10.1016/j.agwat.2023.108142>
- Chandrashekar G, Sahin F (2014) A survey on feature selection methods. *Comput Electr Eng* 40:16–28. <https://doi.org/10.1016/j.compeleceng.2013.11.024>
- Chen C, He M, Chen Q et al (2022) Triple collocation-based error estimation and data fusion of global gridded precipitation products over the Yangtze River basin. *J Hydrol* 605:127307. <https://doi.org/10.1016/j.jhydrol.2021.127307>
- Dehghani M, Trojovský P (2023) Osprey optimization algorithm: a new bio-inspired metaheuristic algorithm for solving engineering optimization problems. *Front Mech Eng*. <https://doi.org/10.3389/fmech.2022.1126450>
- Deo RC, Şahin M (2015) Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmos Res* 153:512–525. <https://doi.org/10.1016/j.atmosres.2014.10.016>
- Drucker H, Burges CJC, Kaufman L et al (1996) Support vector regression machines. *Adv Neural Inf Process Syst* 9:155–161
- Elnashar A, Wang L, Wu B et al (2021) Synthesis of global actual evapotranspiration from 1982 to 2019. *Earth Syst Sci Data* 13:447–480. <https://doi.org/10.5194/essd-13-447-2021>
- Fallah A, Rakhshandehroo GR, Berg P et al (2020) Evaluation of precipitation datasets against local observations in southwestern Iran. *Int J Climatol* 40:4102–4116. <https://doi.org/10.1002/joc.6445>
- Fan L, Xing Z, De LG et al (2022) Evaluation of satellite and reanalysis estimates of surface and root-zone soil moisture in croplands of Jiangsu Province, China. *Remote Sens Environ*. <https://doi.org/10.1016/j.rse.2022.113283>
- Farahmand A, AghaKouchak A (2015) A generalized framework for deriving nonparametric standardized drought indicators. *Adv Water Resour* 76:140–145. <https://doi.org/10.1016/j.advwatres.2014.11.012>
- Feng P, Wang B, Liu DL, Yu Q (2019) Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agric Syst* 173:303–316. <https://doi.org/10.1016/j.agsy.2019.03.015>
- Fernández-Tizón M, Emmenegger T, Perner J, Hahn S (2020) Arthropod biomass increase in spring correlates with NDVI in grassland habitat. *Sci Nat*. <https://doi.org/10.1007/s00114-020-01698-7>
- Fooladi M, Golmohammadi MH, Safavi HR, Singh VP (2021) Fusion-based framework for meteorological drought modeling using remotely sensed datasets under climate change scenarios: Resilience, vulnerability, and frequency analysis. *J Environ Manag*. <https://doi.org/10.1016/j.jenvman.2021.113283>
- Fooladi M, Golmohammadi MH, Rahimi I et al (2023) Assessing the changeability of precipitation patterns using multiple remote sensing data and an efficient uncertainty method over different climate regions of Iran. *Expert Syst Appl* 221:119788. <https://doi.org/10.1016/j.eswa.2023.119788>
- Gelaro R, McCarty W, Suárez MJ et al (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J Clim* 30:5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- Ghazipour F, Mahjouri N (2022) A multi-model data fusion methodology for seasonal drought forecasting under uncertainty: application of Bayesian maximum entropy. *J Environ Manage* 304:114245. <https://doi.org/10.1016/j.jenvman.2021.114245>
- Ghomlaghi A, Nasserli M, Bayat B (2022) Comparing and contrasting the performance of high-resolution precipitation products via error decomposition and triple collocation: an application to different climate classes of the central Iran. *J Hydrol* 612:128298. <https://doi.org/10.1016/j.jhydrol.2022.128298>
- Gleixner S, Demissie T, Diro GT (2020) Did ERA5 improve temperature and precipitation reanalysis over East Africa? *Atmosphere* 11(9):996. <https://doi.org/10.3390/atmos11090996>
- Goldberger J, Roweis S, Hinton G, Salakhutdinov R (2005) Neighbourhood components analysis. *Adv Neural Inf Process Syst*
- Guo Q, Wu W, Massart DL et al (2002) Feature selection in principal component analysis of analytical data. *Chemom Intell Lab Syst* 61:123–132. [https://doi.org/10.1016/S0169-7439\(01\)00203-9](https://doi.org/10.1016/S0169-7439(01)00203-9)
- Guo H, Bao A, Liu T et al (2019) Determining variable weights for an optimal scaled drought condition index (OSDCI): evaluation in central Asia. *Remote Sens Environ* 231:111220. <https://doi.org/10.1016/j.rse.2019.111220>
- Haile GG, Tang Q, Hosseini-Moghari S-MM et al (2020) Projected impacts of climate change on drought patterns over East Africa. *Earth's Futur* 8:1–23. <https://doi.org/10.1029/2020EF001502>
- Hao Z, AghaKouchak A (2013) Multivariate standardized drought index: a parametric multi-index model. *Adv Water Resour* 57:12–18. <https://doi.org/10.1016/j.advwatres.2013.03.009>
- Hao Z, Aghakouchak A (2014) A nonparametric multivariate multi-index drought monitoring framework. *J Hydrometeorol* 15:89–101. <https://doi.org/10.1175/JHM-D-12-0160.1>
- Hao Z, Singh VP (2015) Drought characterization from a multivariate perspective: a review. *J Hydrol* 527:668–678. <https://doi.org/10.1016/j.jhydrol.2015.05.031>

- Harris I, Osborn TJ, Jones P, Lister D (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data* 7:1–18. <https://doi.org/10.1038/s41597-020-0453-3>
- Herbst PH, Db B, Barker HMG (1966) A technique for the evaluation of drought from rainfall data. *J Hydrol* 4:264–272
- Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- Hu Y, Wang F, Chen J et al (2023) Support vector regression model for determining optimal parameters of HfAlO<sub>3</sub>-based charge trapping memory devices. *Electron* 12:1–13. <https://doi.org/10.3390/electronics12143139>
- Huang S, Eisner S, Haddeland I, Tadege Mengistu Z (2022) Evaluation of two new-generation global soil databases for macro-scale hydrological modelling in Norway. *J Hydrol* 610:127895. <https://doi.org/10.1016/j.jhydrol.2022.127895>
- Jackson BB (1983) *Multivariate Data Analysis: An Introduction* Irwin. Homewood, Illinois, USA 244
- Jamei M, Ahmadianfar I, Karbasi M et al (2023) Development of wavelet-based kalman online sequential extreme learning machine optimized with boruta-random forest for drought index forecasting. *Eng Appl Artif Intell* 117:105545. <https://doi.org/10.1016/j.engappai.2022.105545>
- Ji Y, Li Y, Yao N et al (2022) Multivariate global agricultural drought frequency analysis using kernel density estimation. *Ecol Eng.* <https://doi.org/10.1016/j.ecoleng.2022.106550>
- Jiao W, Wang L, McCabe MF (2021) Multi-sensor remote sensing for drought characterization: current status, opportunities and a roadmap for the future. *Remote Sens Environ* 256:112313. <https://doi.org/10.1016/j.rse.2021.112313>
- Kazemzadeh M, Noori Z, Alipour H et al (2022) Detecting drought events over Iran during 1983–2017 using satellite and ground-based precipitation observations. *Atmos Res* 269:106052. <https://doi.org/10.1016/j.atmosres.2022.106052>
- Kira K, Rendell LA (1992) A practical approach to feature selection. In: *Machine learning proceedings*. Elsevier, pp 249–256
- Kogan FN (1995) Application of vegetation index and brightness temperature for drought detection. *Adv Sp Res* 15:91–100
- Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: *European conference on machine learning*. pp 171–182
- Labani M, Moradi P, Ahmadizar F, Jalili M (2018) A novel multivariate filter method for feature selection in text classification problems. *Eng Appl Artif Intell* 70:25–37
- Lap BQ, Phan TTH, Du NH et al (2023) Predicting water quality index (WQI) by feature selection and machine learning: a case study of an kim hai irrigation system. *Ecol Inform* 74:101991. <https://doi.org/10.1016/j.ecoinf.2023.101991>
- Li M, Wu P, Ma Z (2020) A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *Int J Climatol* 40:5744–5766. <https://doi.org/10.1002/joc.6549>
- Li L, Liu Y, Zhu Q et al (2022) Evaluation of nine major satellite soil moisture products in a typical subtropical monsoon region with complex land surface characteristics. *Int Soil Water Conserv Res* 10:518–529. <https://doi.org/10.1016/j.iswcr.2022.02.003>
- Liu Z, Wang Y, Shao M et al (2016) Spatiotemporal analysis of multi-scalar drought characteristics across the Loess Plateau of China. *J Hydrol* 534:281–299
- Liu H, Xin X, Su Z et al (2023) Intercomparison and evaluation of ten global ET products at site and basin scales. *J Hydrol.* <https://doi.org/10.1016/j.jhydrol.2022.128887>
- Ma H, Zeng J, Chen N et al (2019) Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sens Environ.* <https://doi.org/10.1016/j.rse.2019.111215>
- Madani K (2014) Water management in Iran: What is causing the looming crisis? *J Environ Stud Sci* 4:315–328. <https://doi.org/10.1007/s13412-014-0182-z>
- Malik A, Tikhamarine Y, Souag-Gamane D et al (2021) Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. *Meteorol Atmos Phys* 133:891–909. <https://doi.org/10.1007/s00703-021-00787-0>
- Mandal M, Mukhopadhyay A (2013) An Improved minimum redundancy maximum relevance approach for feature selection in gene expression data. *Proc Technol* 10:20–27. <https://doi.org/10.1016/j.protcy.2013.12.332>
- Martens B, Miralles DG, Lievens H et al (2017) GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci Model Dev* 10:1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>
- McKee TB, Doesken NJ, Kleist J et al (1993) The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th Conference on Applied Climatology*. pp 179–183
- Mesbahzadeh T (2020) Meteorological drought analysis using copula theory and drought indicators under climate change scenarios (RCP). 1–20. <https://doi.org/10.1002/met.1856>
- Min X, Shangguan Y, Li D, Shi Z (2022) Improving the fusion of global soil moisture datasets from SMAP, SMOS, ASCAT, and MERRA2 by considering the non-zero error covariance. *Int J Appl Earth Obs Geoinf* 113:103016. <https://doi.org/10.1016/j.jag.2022.103016>
- Moshir Panahi D, Kalantari Z, Ghajarnia N et al (2020) Variability and change in the hydro-climate and water resources of Iran over a recent 30-year period. *Sci Rep* 10:1–9. <https://doi.org/10.1038/s41598-020-64089-y>
- Naderi K, Moghaddasi M (2022) Standardized drought index and copula function under. 2865–2888
- Ochege FU, Shi H, Li C et al (2021) Assessing satellite, land surface model and reanalysis evapotranspiration products in the absence of in-situ in central asia. *Remote Sens* 13:1–25. <https://doi.org/10.3390/rs13245148>
- Panahi DM, Tabas SS, Kalantari Z et al (2021) Spatio-temporal assessment of global gridded evapotranspiration datasets across Iran. *Remote Sens* 13:1–20. <https://doi.org/10.3390/rs13091816>
- Proadhan FA, Zhang J, Hasan SS et al (2022) A review of machine learning methods for drought hazard monitoring and forecasting: current research trends, challenges, and future research directions. *Environ Model Softw* 149:105327. <https://doi.org/10.1016/j.envsoft.2022.105327>
- Rezaei A (2021) Ocean-atmosphere circulation controls on integrated meteorological and agricultural drought over Iran. *J Hydrol.* <https://doi.org/10.1016/j.jhydrol.2021.126928>
- Rhee J, Im J, Carbone GJ (2010a) Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. *Remote Sens Environ* 114:2875–2887. <https://doi.org/10.1016/j.rse.2010.07.005>
- Rhee J, Im J, Carbone GJ (2010b) Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. *Remote Sens Environ* 114:2875–2887
- Robnik-Šikonja M, Kononenko I (1997) An adaptation of {R}elief for attribute estimation in regression. *Mach Learn Proc Fourteenth Int Conf* 5:296–304
- Rodell M, Houser PR, Jambor U et al (2004) The global land data assimilation system. *Bull Am Meteorol Soc* 85:381–394. <https://doi.org/10.1175/BAMS-85-3-381>
- Rodrigues GC (2021) Evaluation of NASA POWER reanalysis products to estimate daily weather variables in a hot summer mediterranean Climate. pp 1–17
- Rostami M, Berahmand K, Nasiri E, Forouzande S (2021) Review of swarm intelligence-based feature selection methods. *Eng Appl*

- Artif Intell 100:104210. <https://doi.org/10.1016/j.engappai.2021.104210>
- Saemian P, Hosseini-Moghari SM, Fatehi I et al (2021) Comprehensive evaluation of precipitation datasets over Iran. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2021.127054>
- Salimi H, Asadi E, Darbandi S (2021) Meteorological and hydrological drought monitoring using several drought indices. *Appl Water Sci* 11:1–10. <https://doi.org/10.1007/s13201-020-01345-6>
- Sharafi S, Ghaleni MM (2022) Spatial assessment of drought features over different climates and seasons across Iran. *Theor Appl Climatol* 147:941–957. <https://doi.org/10.1007/s00704-021-03853-0>
- Sharafi S, Mohammadi Ghaleni M (2021) Calibration of empirical equations for estimating reference evapotranspiration in different climates of Iran. *Theor Appl Climatol* 145:925–939. <https://doi.org/10.1007/s00704-021-03654-5>
- Shirmohammadi-Aliakbarkhani Z, Saberali SF (2020) Evaluating of eight evapotranspiration estimation methods in arid regions of Iran. *Agric Water Manag*. <https://doi.org/10.1016/j.agwat.2020.106243>
- Sion BD, Shoop SA, McDonald EV (2022) Evaluation of in-situ relationships between variable soil moisture and soil strength using a plot-scale experimental design. *J Terramechanics* 103:33–51. <https://doi.org/10.1016/j.jterra.2022.07.002>
- Spinoni J, Naumann G, Carrao H et al (2014) World drought frequency, duration, and severity for 1951–2010. *Int J Climatol* 34:2792–2804. <https://doi.org/10.1002/joc.3875>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192
- Tian Y, Xu Y-P, Wang G (2018) Agricultural drought prediction using climate indices based on support vector regression in Xiangjiang River basin. *Sci Total Environ* 622:710–720
- Tian L, Leasor ZT, Quiring SM (2020) Developing a hybrid drought index: precipitation evapotranspiration difference condition index. *Clim Risk Manag* 29:1–17. <https://doi.org/10.1016/j.crm.2020.100238>
- Tian Q, Lu J, Chen X (2022) A novel comprehensive agricultural drought index reflecting time lag of soil moisture to meteorology: a case study in the Yangtze River basin, China. *Catena*. <https://doi.org/10.1016/j.catena.2021.105804>
- Tsiros IX, Nastos P, Proutos ND, Tsaousidis A (2020) Variability of the aridity index and related drought parameters in Greece using climatological data over the last century (1900–1997). *Atmos Res* 240:104914. <https://doi.org/10.1016/j.atmosres.2020.104914>
- UNESCO (1979) Map of the world distribution of arid regions: explanatory note
- Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10:988–999
- Vicente-Serrano SM, Beguer`via S, López-Moreno JI, et al (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J Clim* 23:1696–1718. <https://doi.org/10.1175/2009JCLI2909.1>
- Wable PS, Jha MK, Shekhar A (2019) Comparison of drought indices in a semi-arid river basin of India. *Water Resour Manag* 33:75–102. <https://doi.org/10.1007/s11269-018-2089-z>
- Wang F, Wang Z, Yang H et al (2018) Study of the temporal and spatial patterns of drought in the Yellow River basin based on SPEI. *Sci China Earth Sci* 61:1098–1111. <https://doi.org/10.1007/s11430-017-9198-2>
- Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* 2:37–52
- Wu Z, Feng H, He H et al (2021) Evaluation of soil moisture climatology and anomaly components derived from ERA5-land and GLDAS-2.1 in China. *Water Resour Manag* 35:629–643. <https://doi.org/10.1007/s11269-020-02743-w>
- Xu L, Chen N, Zhang X (2019) Global drought trends under 1.5 and 2 C warming. *Int J Climatol* 39:2375–2385
- Xu L, Chen N, Zhang X et al (2021) In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. *Remote Sens Environ* 254:112248. <https://doi.org/10.1016/j.rse.2020.112248>
- Xu Y, Han S, Shi C, et al (2023) Comparative analysis of three near-surface air temperature reanalysis datasets in inner mongolia region
- Yang W, Wang K, Zuo W (2012) Neighborhood component feature selection for high-dimensional data. *J Comput* 7:162–168. <https://doi.org/10.4304/jcp.7.1.161-168>
- Yang B, Cui Q, Meng Y et al (2023) Combined multivariate drought index for drought assessment in China from 2003 to 2020. *Agric Water Manag* 281:108241. <https://doi.org/10.1016/j.agwat.2023.108241>
- Yao T, Lu H, Yu Q et al (2023) Uncertainties of three high-resolution actual evapotranspiration products across China: comparisons and applications. *Atmos Res*. <https://doi.org/10.1016/j.atmosres.2023.106682>
- Yin G, Zhang H (2023) A new integrated index for drought stress monitoring based on decomposed vegetation response factors. *J Hydrol* 618:129252. <https://doi.org/10.1016/j.jhydrol.2023.129252>
- Zhang A, Jia G (2013) Monitoring meteorological drought in semi-arid regions using multi-sensor microwave remote sensing data. *Remote Sens Environ* 134:12–23
- Zhang LX, Wang JX, Zhao YN, Yang ZH (2003) A novel hybrid feature selection algorithm: using ReliefF estimation for GA-Wrapper search. *Int Conf Mach Learn Cybern* 1:380–384. <https://doi.org/10.1109/icmlc.2003.1264506>
- Zhang X, Wu G, Dong Z, Crawford C (2015) Embedded feature-selection support vector machine for driving pattern recognition. *J Franklin Inst* 352:669–685
- Zhang X, Chen N, Li J et al (2017) Multi-sensor integrated framework and index for agricultural drought monitoring. *Remote Sens Environ* 188:141–163. <https://doi.org/10.1016/j.rse.2016.10.045>
- Zhang M, Yuan X, Otkin JA (2020) Remote sensing of the impact of flash drought events on terrestrial carbon dynamics over China. *Carbon Balance Manag* 15:1–11. <https://doi.org/10.1186/s13021-020-00156-1>
- Zhang H, Immerzeel WW, Zhang F et al (2021) Creating 1-km long-term (1980–2014) daily average air temperatures over the Tibetan Plateau by integrating eight types of reanalysis and land data assimilation products downscaled. *Int J Appl Earth Obs Geoinf* 97:102295. <https://doi.org/10.1016/j.jag.2021.102295>
- Zhang H, Yin G, Zhang L (2022a) Evaluating the impact of different normalization strategies on the construction of drought condition indices. *Agric For Meteorol* 323:109045. <https://doi.org/10.1016/j.agrformet.2022.109045>
- Zhang Q, Shi R, Xu CY et al (2022b) Multisource data-based integrated drought monitoring index: Model development and application. *J Hydrol* 615:128644. <https://doi.org/10.1016/j.jhydrol.2022.128644>
- Zhu W, Tian S, Wei J et al (2022) Multi-scale evaluation of global evapotranspiration products derived from remote sensing images: Accuracy and uncertainty. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2022.127982>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.