2022

A statistical evaluation of Earth-observation-based composite drought indices for a localized assessment of agricultural drought in Pakistan

C. Schwartz

W. Lee Ellenburg

V. Mishra

T. Mayer

R. Griffin

See next page for additional authors

Authors
C. Schwartz, W. Lee Ellenburg, V. Mishra, T. Mayer, R. Griffin, F. Qamer, M. Matin, and Tsegaye Tadesse

International Journal of Applied Earth Observations and Geoinformation 106 (2022) 102646

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

A statistical evaluation of Earth-observation-based composite drought indices for a localized assessment of agricultural drought in Pakistan

Caily Schwartz [a,b,*], W. Lee Ellenburg [b,c,**], Vikalp Mishra [b,c], Timothy Mayer [b,c], Robert Griffin [a,b,c], Faisal Qamer [d], Mir Matin [d], Tsegaye Tadesse [e]

[a] Department of Atmospheric and Earth Science, The University of Alabama in Huntsville, 320 Sparkman Drive, Huntsville, AL 35805, USA
[b] SERVIR Science Coordination Office, NASA Marshall Space Flight Center, 320 Sparkman Drive, Huntsville, AL 35805, USA
[c] Earth System Science Center, The University of Alabama in Huntsville, 320 Sparkman Drive, Huntsville, AL 5805, USA
[d] International Centre for Integrated Mountain Development, GPO Box 3226, Kathmandu, Nepal
[e] National Drought Mitigation Center, School of Natural Resources, University of Nebraska-Lincoln, 811 Hardin Hall, 3310 Holdredge Street, Lincoln, NE 68583, USA

A R T I C L E I N F O

Keywords:
Composite drought index
Agricultural drought
Principal component analysis
Variable selection
Pakistan
Earth observations

A B S T R A C T

Drought is a complex phenomenon that impacts a multitude of sectors globally and is difficult to characterize due to the variation in defining conditions. With over 2.5 billion people dependent on agriculture for their livelihoods, agricultural drought impacts are particularly acute. Specific indicators based on different aspects of the hydrologic cycle can be used to better characterize drought and mitigate its impacts. The overall objective of this study is to develop a monthly Composite Drought Index (CDI) using Earth observation datasets to provide an assessment of droughts in Pakistan. Further, district level wheat production data was used to optimize variables to create a customized composite drought index (CCDI) specifically for agriculture and evaluate the two indices. Pakistans economy and communities rely heavily on the agricultural sector and many areas are at high risk of crop failure due to the dependence on precipitation and other environmental conditions. A total of 10 environmental variables are considered that account for supply and demand of water, soil moisture, and vegetative conditions. Statistically important variables are chosen for each district with respect to wheat production, creating a subset of the original inputs. Weights for each dataset are calculated using a Principal Component Analysis (PCA), identifying what variables contribute the most to the index. The CDI and CCDI are evaluated with nationally reported, district level, wheat production data, for the harvest years 2005–2017, carried out specifically for the Rabi season, October - March. The CCDI, on average used 5 variables, as compared to the full 10 in the CDI. Overall, the two composite indices were highly correlated and both captured well known climatological events. When compared to production trends, the CDI has the ability to identify agricultural droughts with a true negative rate of 0.742 in rainfed districts, while irrigated districts have a true negative rate of 0.568. For the CCDI the true negative rates in rainfed and irrigated districts were 0.667 and 0.602, respectively. The results distinguished the importance of each variable in the contribution to the CDI and CCDI. The findings from this study demonstrate the potential of using the methodology for the CDI to enhance pre-existing drought monitoring and forecasting systems in the region.

1. Introduction

In the past 20 years, droughts have impacted approximately 1.43 billion people, while only accounting for 5% of the recorded disasters documented by EM-DAT, the international disasters database from the Centre for Research on the Epidemiology of Disasters (Nations Office of Disaster Risk Reduction). Economic costs of drought vary greatly by sector and region. Of all the disasters recorded by the Food and Agriculture Organization (FAO) of the United Nations between 2009 and 2019, drought caused the majority of losses in the agricultural sector, resulting in $29 billion in damage in developing countries alone (FAO, 2018). Due to the significant impact of drought on developing economies, there is a need for increased drought characterization for monitoring and early warning systems.

* Corresponding author.
* Corresponding author.
E-mail addresses: cschwartz8@huskers.unl.edu (C. Schwartz), wle0001@uah.edu (W.L. Ellenburg).

https://doi.org/10.1016/j.jag.2021.102646
Received 20 October 2021; Received in revised form 6 December 2021; Accepted 7 December 2021
Available online 16 December 2021

The definition of drought varies depending on the application, but all droughts originate from a precipitation deficit (Wilhite and Glantz, 1985). Drought is broadly classified into 5 categories, meteorological (lack of precipitation over a period of time), agricultural (loss of soil moisture leading to crop failure), hydrological (insufficient surface water resources), socioeconomic (water demands do not meet domestic supply), and ecological (ecosystems become vulnerable and services are impacted) (Wilhite and Glantz, 1985; Crausbay et al., 2017). Moreover, drought is particularly challenging to characterize, as there is often no clear start or end date. The onset of drought can be slow, extend over large areas, and persist over multiple months or years (Wilhite, 2000; NOAA). Comparatively, flash droughts are short-lived and intensify rapidly, often at smaller spatial scales (Otkin et al., 2018; Pendergrass et al., 2020). The duration, intensity, and persistence depends on many biophysical characteristics that work together to create a drought event.

With over 60% of the rural population in Pakistan dependent on agriculture for their livelihoods (Azam and Shafique, 2017), agricultural drought impacts are particularly acute. Drought has the ability to extensively impact crop production at various stages of development (Saini and Westgate, 1999; Yu et al., 2018) and, as a result, contribute to food insecurity of a region (Mbow et al., 2019). To meet the increasing demand, most farmers depend on irrigation in Pakistan. Irrigated agriculture accounts for the vast majority of crop production in the Indus River Basin. Even with the prominence of irrigation, rainfed agricultural comprises ~25% of the arable land (Baig et al., 2013). Rainfed agriculture is a vital resource in the country accounting for a significant share of the national economy. Additionally, rainfed Pakistani farmlands are usually smaller and more fragmented compared to irrigated areas. Rainfed farms also are more often subsistence based, thus directly tied to the local food security (Devendra, 2012; Sibhatu and Qaim, 2017). Crops in rainfed areas are more susceptible to drought conditions than the same crop in irrigated areas (Ray et al., 2018). This susceptibility is largely due to irregular precipitation patterns which decrease the ability of farmers to mitigate and cope with drought events. Furthermore, as climatic variability and temperatures increase, drought frequency and severity in many parts of the country are expected to intensify (Ahmed et al., 2018).

Satellite and model based drought indicators have been used and studied to better characterize drought, its impacts, and to provide early warning, as reported by previous research studies (Hao and Singh, 2015; Zargar et al., 2011; Mishra and Singh, 2010). Drought monitoring and identification efforts typically involve the creation of an index that can be comprised of a single relevant variable or a combination of variables. For example, the Standardized Precipitation Index (SPI) uses only precipitation. SPI is commonly used for meteorological drought (Hayes et al., 2011) due to its simplicity and ability to be scaled over a wide range of time periods, from 1 month to up to two years (World Meteorological Organization, 2012). However, when considering longer term conditions, such as that of an agricultural drought, precipitation alone may not be enough to characterize other significant ground conditions such as soil moisture and the atmospheric demand from the land surface (Heim, 2002).

The need to incorporate more than one variable into a drought index has led researchers to create combinations of indices and variables to produce multivariate indices to characterize the persistence, duration, and intensity of drought (Bayissa et al., 2019; Rajsekhar et al., 2015; Hao and Aghakouchak, 2014; Brown et al., 2008; Svoboda et al., 2002; Liu et al., 2020a,b; Park et al., 2016; Sepulcre-Canto et al., 2012; Wu et al., 2013; Feng et al., 2019; Keyantash and Dracup, 2004; Zhang et al., 2021). By using the strengths of different drought indicators and combining them into a single index, multivariate indices capture different aspects of the hydrologic cycle that aim to create an inclusive drought index. Knowing what set of variables to consider is a challenge in developing a comprehensive drought index. While the specific set of variables may vary from index to index, generally, variables such as precipitation, soil moisture, evapotranspiration, and runoff are used to capture the major elements of the surface water balance model (Rajsekhar et al., 2015; Keyantash and Dracup, 2004; Sheffield et al., 2004; Mendicino et al., 2008; Tian et al., 2020). In practice, a compromise between data sets that are readily available and easily accessible at appropriate temporal and spatial scales is needed. Variable selection methods can be beneficial to reduce dimensions by removing redundant or irrelevant variables (Chandrashekar and Sahin, 2014; Zhang et al., 2021). This approach is used in drought indices that apply machine learning methods (Feng et al., 2019). A reduction in variables can also allow for the customization of an index at a specific location (i.e. at unique agro-climate zones or to distinguish from rainfed and irrigated systems).

When considering agricultural drought, soil moisture (available water to plants) and vegetative conditions (Brown et al., 2008; Liu and Kogan, 1995; Gao, 1996) are common metrics. Brown et al. (2008) introduced VegDRI which uses Palmer Drought Severity Index (PDSI), Standard Precipitation Index (SPI) (36-week), and satellite-derived vegetation conditions including Percent Average Seasonal Greenness (PASG) and Start of Season Anomaly (SOSA), calculated from satellite based Normalized Difference Vegetation Index (NDVI) from NOAAs Advanced Very High Resolution Radiometer (AVHRR). Weekly precipitation total, average temperature, previous history of the indices and division constants are inputs for the Crop Moisture Index (CMI) (Palmer, 1968), determining any deficit between potential evapotranspiration and moisture. Vicente-Serrano et al. (2010) incorporated precipitation, temperature and potential evapotranspiration to calculate the Standardized Precipitation Evapotranspiration Index (SPEI), another common index to measure agricultural drought (Vicente-Serrano et al., 2010).

Different approaches have been used to combine variables, either assigning weights (equal or unequal) or through the use of statistical methods such as Principal Component Analysis (PCA) to reduce the dimension and aggregate variables. Methods include machine learning (Brown et al., 2008; Feng et al., 2019; Liu et al., 2020b; Park et al., 2016; Wu et al., 2013), because of the flexibility in handling many different data types, copula-based (Hao and Aghakouchak, 2014), effective for models with fewer variables but are not as flexible in modelling the dependence structure, entropy (Rajsekhar et al., 2015), allowing for the assumption of non-linearity between variables or linear combination (Keyantash and Dracup, 2004; Bayissa et al., 2019). A common approach from combining variables is using a objective statistical approach, PCA, to quantify the contribution of each input variable (Kourgialas et al., 2015; Bayissa et al., 2019). PCA has been applied in many studies to reduce the dimensionality of datasets to identify the most relevant predictive variables (Hao and Singh, 2015; Bazrafshan et al., 2014; Rajsekhar et al., 2015). PCA has also see wide use in creating composite drought indices from multiple variables (Keyantash and Dracup, 2004; Bayissa et al., 2019; Barua et al., 2009). The PCA-based approach is essentially a linear combination of the original variables that are weighted based on their contribution to the overall variance to the fist (linear) component. The main limitations to the PCA approach for variable compositing is the assumption of linearity of the variables (Hao and Singh, 2015; Rajsekhar et al., 2015) and the assumption the most important drought characterization is represented in 1st component (maximum variance). However, the first principal component can retain the information from the original data (Bayissa et al., 2019; Keyantash and Dracup, 2004) making its interpretation physically based. Additionally, PCA is widely used in practice because it is relatively straightforward to implement, provides an aggregate perspective from many variables, and can be easily customized based on a regions characteristics.

Specifically in Pakistan, studies have determined the ability of different data sources in monitoring droughts. These studies assessed the ability of NDVI anomalies (Haroon et al., 2016) while out of 15 variables, the SPI, SPEI and the Reconnaissance drought index (RDI) were recommended for monitoring drought conditions in Pakistan (Adnan et al., 2018). Other studies assessed the characteristics of drought (Adnan et al., 2015; Jamro et al., 2 2020) and the performance of a composite drought index (Qaiser et al., 2021) for specific provinces in Pakistan. Three month SPEI was used to assess the variability of drought throughout the country (Jamro et al., 2019). Because of Pakistan's dependence on agriculture and the irregularity in drought conditions, there is a need for a country-wide agricultural composite drought index that can provide localized information and incorporates multiple geophysical variables.

The objective of this paper is to create a multivariate composite drought index (CDI) and evaluate the ability of the developed index to provide a localized assessment of drought events in Pakistan for the years 2004–2017. These years include wide spread drought events and historical flooding, capturing both extremes and a variety of years in-between. Using the CDI, a customized CDI (CCDI) is created to further understand the key variables for assessing agricultural drought at a district level. Utilizing knowledge gained from existing drought studies in Pakistan, this research introduces a new multivariate agricultural drought index. The methodology calculates weights for each variable, allowing for an in depth analysis of localized drought, both in space and time. The weights account for different climatic variables that are most prevalent in producing droughts. Further, a customized selection of variables also provides information regarding which inputs are most important in determining agricultural drought. The CDI and CCDI are compared and evaluated alongside district level wheat production data as a proxy for agricultural drought. The methodology developed can enhance existing drought monitoring and forecasting systems to better understand and characterize drought conditions in the country.

2. Data and methods

2.1. Study area

Pakistan is defined by the Food and Agriculture Organization (FAO) as an arid country, characterized by low rainfall and high evapotranspiration. Pakistan has five administrative divisions and two federally administered territories and are all further divided into districts. The country's range of topographic conditions creates a variety of climatic and ecological zones where precipitation and temperatures can vary greatly both geographically as well as seasonally. Pakistan on average receives 307.40 mm of rainfall (1991–2020). Precipitation varies widely depending on the region. The southern and western parts receive less than 150 mm a year while the north-eastern regions can receive between 400 and 1,000 mm. The mean temperature throughout the country is 20.74 degrees Celsius (1991–2020) (World Bank, 2020).

This study was conducted over agricultural areas in 97 districts, 43 of the total are primarily rainfed while the remaining 54 are predominantly irrigated districts. Fig. 1 highlights the agricultural areas, irrigated and rainfed, in solid blue and green respectively. Irrigated and rainfed districts are highlighted with a blue or green outline. Districts that were not included in the analysis are shown with a crosshatch pattern.

2.2. Production and cropland data

Pakistan has two cropping seasons, Kharif, starting between April and June, during the monsoon season, and Rabi, starting between October and December (FAO, 2004), during the relatively drier winter

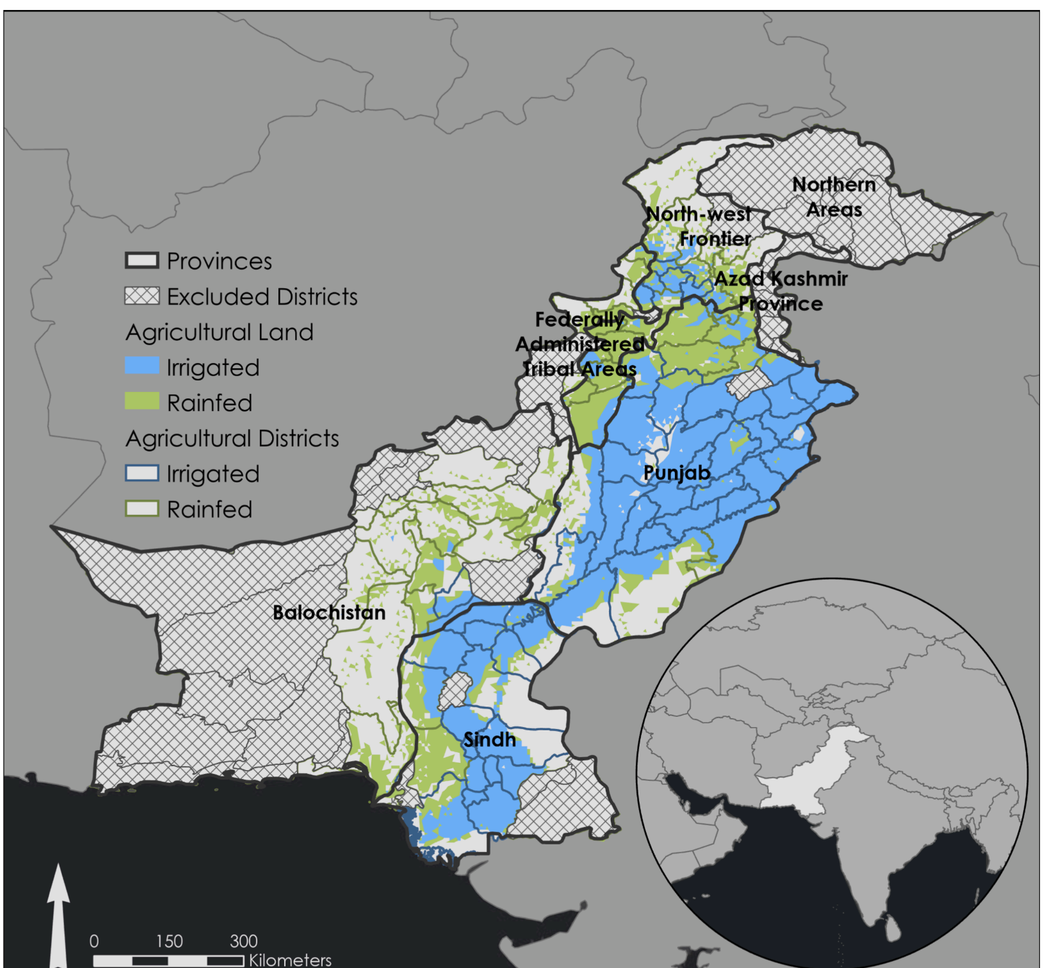

Fig. 1. Study area map of Pakistan and the agricultural land and districts, classified as irrigated or rainfed based on cropland datasets outlined below.

season. Crops that are planted during the Kharif season include rice, sugarcane, cotton, and maize. Crops that are planted during the Rabi season include wheat, lentil, and barley, among others. For this study, yearly wheat area ('000' hectares) and production ('000' tonnes) data for each district from 2004 to 2017 were obtained from the Pakistan Bureau of Statistics. The detrended and standardized wheat production data were used to evaluate the CDI and CCDI as a proxy for agricultural drought. Wheat accounts for 8.9% (approx.) value added to the agricultural sector and contributes 1.6% to the countrys GDP (Government of Pakistan Finance Division, 2020). Throughout Pakistan, wheat is consistently in the top 3 crops by area by district. In Pakistan, around 33% of the wheat crop is grown in rainfed areas (Baig et al., 2013).

To identify areas that are typically under agriculture, the 30 m South Asia, Afghanistan and Iran Cropland Extent map from the Land Processes Distributed Active Archive Center (LP DAAC) (Gumma et al., 2017) was used. Since many districts in Pakistan have rocky terrain or barren land, only areas identified as cropland was used to assess the agricultural drought. Furthermore, only districts that have 5 or more years of wheat production data and at least 10% of agricultural pixels have been included in the analysis. To determine possible differences, this study will identify and analyze irrigated and rainfed districts separately. The Global Food Security Support Analysis Data (GFSAD) Crop Dominance 2010 Global 1 km (Teluguntla et al., 2016) obtained from the LP DAAC was used to determine the rainfed and irrigated pixels within the agricultural areas from the 30 m South Asia, Afghanistan and Iran Cropland Extent dataset. The districts are determined to be irrigated or rainfed based on majority of cropland pixels falling in either category.

2.3. Remote sensing and model derived data

2.3.1. Precipitation data

2.3.1.1. Climate Hazards group Infrared Precipitation with Stations (CHIRPS). CHIRPS (Funk et al., 2015) data is used as a source of precipitation in this study. This product uses a combination of station data and satellite imagery to create a global gridded rainfall dataset. Daily data was downloaded from ClimateSERV with a spatial resolution of 0.05 degrees.

CHIRPS data was used to calculate the Standard Precipitation Index (SPI) for 1 month, 3 months, 6 months, and 12 months (SPI1, SPI3, SPI6, SPI12). SPI was used to identify the intensity of meteorological drought by determining a lack of precipitation based on historical records (Keyantash and Atmospheric Research Staff, Eds, 2018).

2.3.2. Evaporative demands

Two datasets that identify the evaporative demand are used in this research due to their ability to capture drought conditions. The demand for water will increase during drought conditions. When demand is greater than supply, an increase in evapotranspiration will occur.

2.3.2.1. Evaporative Stress Index (ESI). ESI defined as the ratio of actual to potential evapotranspiration, can be used as an drought indicator (Anderson et al., 2011, 2013). Under drought conditions, plants gets stressed due to shortage of moisture and rate of evapotranspiration is reduced to compensate for moisture shortage. The higher the difference between the potential and actual ET, the higher the stress. The actual and potential ET data is obtained from a well matured remotely sensed based model - the Atmosphere-Land Exchange Inverse (ALEXI) (Anderson et al., 1997, 2007). Daily ALEXI ESI data was obtained from ClimateSERV with a spatial resolution of 5 km.

2.3.2.2. Evaporative Demand Drought Index (EDDI). EDDI is the anomaly in evaporative demand which increases in all types of drought (Hobbins et al., 2016; McEvoy et al., 2016). The evaporative demand is calculated from the reference evapotranspiration FAO Penman-Monteith equation. EDDI incorporates physically based variables including air temperature, wind speed, relative humidity and shortwave radiation. This dataset was obtained from NOAA, has a 0.125 degree resolution and is produced at dekadal and monthly time steps. This research used the 1 month product. The physical forcing variables used to compute EDDI are obtained from the University of Idaho (Abatzoglou, 2013; Abatzoglou, 2021).

2.3.3. Soil moisture and vapor pressure deficit

2.3.3.1. Gravity Recovery and Climate Experiment (GRACE). GRACE provides a monthly report of total water storage across the globe. This research used root zone soil moisture percentiles which are based on observations of terrestrial water storage. This dataset was obtained from NASA GRACE global data and has a resolution of 0.25 degrees and is produced weekly. The root zone is defined as the top 1 meter of soil. GRACE allows for the detection of variability in storage conditions of the availability of water to plants.

2.3.3.2. South Asia Land Data Assimilation System (South Asia LDAS). The SLDAS (Zhou et al., 2021; Zaitchik et al., 2017) model is used to acquire soil moisture (SM) and calculate vapor pressure deficit (VPD) input datasets. SLDAS is a regionally calibrated implementation of the Noah land surface model that provides information on surface states and fluxes and is implemented using remotely sensed observations over South Asia. SALDAS is run at the daily scale and aggregated at the dekadal and monthly time scales. For this study the monthly output at 5 km resolution was used and is accessed through the tethys platform at the International Centre for Integrated Mountain Development (http://tethys.icimod.org/apps/).

SLDAS has multiple soil moisture layers [0–10; 10–40; 40–100 and 100–200 cms]. This study creates a composite soil moisture dataset using a weighted average profile depth for 0–100 cm. Since moisture content within the rootzone is the only water that is directly utilized by plants from the entire hydrological cycle, information on rootzone soil moisture is critical for drought assessment.

Vapor pressure deficit (VPD) is derived using Tetens Formula (Stull, 2000) from the specific humidity and surface pressure output by the SLDAS model. VPD is the difference between the saturation vapor pressure (e_s) and the actual vapor pressure (e_a) at a given temperature and is a driving force for water movement between the leaf and the atmosphere (Zhang et al., 2017).

2.3.4. Vegetative conditions

2.3.4.1. Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI). The monthly 1 km NDVI data product obtained from the eMODIS collection V6 dataset via USGS EarthExplorer is used. The eMODIS is a 10-day composite data collection based of MODIS Aqua sensor. NDVI captures vegetation stress based on the reflectivity of near infrared and red wavelengths. Healthy vegetation absorbs red light while reflecting green and near-infrared.

2.4. Methodology

This study assesses the effectiveness of customizing variable inputs to the CDI at the district level in Pakistan. First a seasonal, CDI by district was created using a range of climatic and hydrologic variables derived from satellite based remote sensing and land modeling systems. These input datasets included SPI (1, 3, 6, and 12 months), ESI, EDDI, GRACE, SM, NDVI, and VPD (Table 1). All variables were aggregated to the monthly scale and re-sampled to a 5 km resolution for this analysis. Standard anomalies for each month in each year for each input were calculated at the pixel level. A PCA-based approach was used to weigh each variable by their contribution to the total variance. Next, a

Table 1
Source of remote sensing and model derived data.

Variable (Abbreviation)	Source/ Model	Native Temporal Resolution	Native Spatial Resolution
Standard precipitation index for 1 month (SPI1)	CHIRPS	Daily	0.05°
Standard precipitation index for 3 months (SPI3)	CHIRPS	Daily	0.05°
Standard precipitation index for 6 months (SPI6)	CHIRPS	Daily	0.05°
Standard precipitation index for 12 months (SPI12)	CHIRPS	Daily	0.05°
Soil Moisture (SM)	SLDAS	Daily	0.05°
Vapor Pressure Deficit (VPD)	SLDAS	Daily	0.05°
Evaporative Stress Index (ESI)	ALEXI	Daily	0.05°
Evaporative Demand Drought Index (EDDI)	NOAA	Daily	12 km
Root zone soil moisture percentiles (GRACE)	GRACE	Monthly	0.25°
Normalized difference vegetation index (NDVI)	MODIS	Monthly	1 km

customized CDI (CCDI) was created for each month in the wheat growing season for each district. The CCDI was created to best capture the most optimum variables at a localized scale, specifically for agricultural drought. The CCDI uses the 10 processed variables (Table 1) utilized in the CDI. The importance of the variables was ranked using a Random Forest model by district, using the recorded wheat production data as the target variable. Evaluation took place at the district level where monthly CDI and CCDI values were aggregated to a single value to represent the entire growing season. The CDI and CCDI were then evaluated with wheat production data. The following sections will first describe the methodology to calculate the CDI in more detail, then how the variables were chosen for the CCDI will be described, followed by the methodology for evaluation of both the CDI and CCDI.

2.4.1. *Calculating the Composite Drought Index (CDI)*

To capture historical agricultural drought, 10 environmental variables were selected to create a CDI. The datasets are highlighted in Table 1. These variables represent multiple components of the hydrologic cycle and the broader climatology of the region.

The first step in combining the input variables is to process the data to account for different spatial resolutions. The input variables were resampled to have a consistent spatial resolution of 5 km and averaged to the monthly timestep. The sum of precipitation for each month was used instead of mean for precipitation data (SPI1, 3, 6, 12). Then, a temporal standard anomaly for each month in each year was calculated at the pixel level. The resulting data were clipped to agricultural land (Fig. 1) . Standarized anomalies are a proxy of z-score with a typical value range is from −3 to 3. The values close to zero represent near normal conditions while higher values (either side of zero) signify extreme or abnormal conditions.

A monthly CDI was created by weighing the standardized anomalies based on the percent contribution to their combined variance using a principal component approach. At each pixel, the variable weights were calculated from the eigenvectors of an $n \ x \ n$ covariance matrix of the input variables, where n is the number of variables, n is equal to 10 for the CDI. Eigenvectors describe the relationship between the original data and the principal components. Specifically, they express the variability in the data as orthogonal vectors. The first component, being the direction of the maximum variance, was used in this study, similar to the methodology of Bayissa et al. (2019). The square of each dimension of the eigenvector represents the proportion of the variance explained by each variable for that eigenvector. The proportions, or weights, were applied to each respective variables' standard anomaly creating a weighted average of the input variables. This was applied for each month of the year and at each pixel resulting in a CDI for every month at a 5 km scale across the study area.

2.4.2. *Calculating the Customized Composite Drought Index (CCDI)*

An agriculture specific CCDI was created to determine the use of a customized drought index, where each district uses a subset of the original 10 input datasets based off their importance in explaining wheat production trends. The CCDI follows the same methodology as the CDI to calculate weights for each variable but a variable selection step has been included before weights are calculated, seen in the methodology workflow (Fig. 2) in the box labelled 'Variable selection'.

Variable selection takes place after preprocessing. To reduce redundancy in the 10 input datasets, Variable Selection Using Random Forests (VSURF), version 1.1.0, (Genuer et al., 2015) method was applied to rank the variables by importance with respect to historical district level wheat production data. VSURF is a three step procedure: thresholding, interpretation, and prediction. The predicting variable is the standardized, detrended wheat production data. Only the thresholding step is used for variable selection because it removes statistically insignificant variables using a model specified calculated threshold of variable importance and variables that exceed the threshold are kept. VSURF was executed using the VSURF package in R. An in depth description of VSURF can be found in Genuer et al. (2015).

VSURF identified a set of statistically significant variables for each grid cell. Hence all grids may have distinct set of variables, however for district level analysis, all the grid level variable subsets were combined into a unique subset that represented the agriculture areas. This allows for all pixels within a district to use the same set of variables in the CCDI. A variable was selected only if it was represented in 25% or more of the agricultural pixels in a district. This resulted in a unique combination of downselected, important variables for each district for use in the CCDI. Once a subset of the datasets were chosen for each district, weights were calculated for each input. The calculation of the weights follow the same methodology as the CDI, using the eigenvectors of the first principal component. The difference lies in the number of variables inputted into the covariance matrix, resulting in different $n \ x \ n$ covariance matrices sizes based on district for the CCDI.

2.4.3. *Evaluating the monthly CDI and CCDI*

Considering Rabi season crops, CDI and CCDI were evaluated from the period October-March using only pixels that are in agricultural land to identify drought that has impacted cropland. For this study, wheat crop was used as it contained the most complete data from the Pakistan Bureau of Statistics. The districts that were evaluated have 10% or more pixels identified as agricultural and have more than 7 years of recorded wheat production data, to ensure the districts have enough data. A total of 96 districts consisting of 42 rainfed and 54 irrigated districts were evaluated. The mean, maximum, and minimum value for every month in every year were calculated for each district for both indices (CDI and CCDI). The mean represented the CDI and CCDI value for each district for each month in the study period.

To determine the effectiveness of CDI and CCDI in observing agricultural drought, wheat production anomalies were used as a proxy for drought. Specifically for the CCDI, the wheat data was initially used to down-select variables, however, the CCDI was created independently from the production data and only a function of the co-variation of the input variables. The wheat production data was detrended and standardized using the standard anomaly for each district. Detrending accounts for improvements in agricultural technologies that would increase production. Wheat is harvested once a year, leading to a single value for the entire growing season, whereas the CDI and CCDI are monthly and have 6 values for the wheat growing season, October-March. To evaluate the CDI and CCDI with wheat production, different methods were evaluated to obtain a single value that best represents the wheat production trends for each harvest year in every district. These included minimums and means for the growing season,

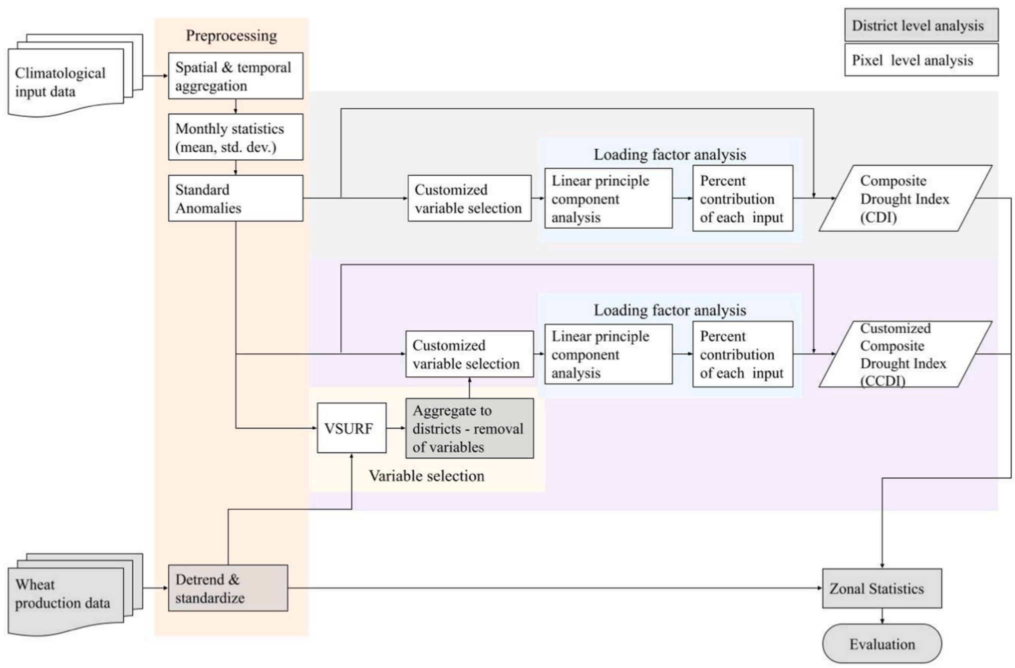

Fig. 2. Workflow of methodology employed in this study.

the number of negative index values in a growing season, and the number of consecutive negatives in a growing season. By comparing the results of the statistics using different combinations of CDI and CCDI values, a thresholding approach was chosen that maximized both the accuracy and specificity (true negative rate). In this approach, if two or more negative months are in a growing season and the sum of all negatives in that growing season is less than -0.60, then the minimum index value of the growing season is taken for the year value. These conditions account for persistence and intensity of drought events during a growing season. If these conditions are not met, the mean value of the growing season is used to represent the CDI or CCDI value for the harvest year.

The relationship between each index and wheat production was assessed by using confusion matrices and statistics of correlation, accuracy, precision, true negative rate, the area under a receiver operating characteristic curve (ROC/AUC), and sensitivity (Table 2). Where TP is the number of true positives, TN is true negatives, FN is false negatives, and FP is false positives. Wheat production data were used as the observed positives and negatives, while CCDI and CDI values were used as the predicted positive and negative values. True negatives occurred when the observed wheat value was negative and the CDI or CCDI value was also negative, indicating a drought. The statistics calculated allow for the assessment of the CDI and CCDI to identify an agricultural drought that impacted wheat production.

Table 2
Equations for statistical analysis.

Statistic	Equation	Reference
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Fawcett (2006)
True Negative Rate	$\frac{TN}{TN+FN}$	Fawcett (2006)
True Positive Rate	$\frac{TP}{TP+FN}$	Fawcett (2006)
AUC/ROC	Area under ROC curve	Fawcett (2006)
Correlation	$\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$	Stull (2000)

3. Results and discussion

3.1. Variable analysis of the CDI and CCDI

Because this study focuses on agricultural drought and uses wheat production as a proxy, only the main wheat growing season, October-March, was evaluated. When creating the CDI, the component loading analysis (PCA) weighs each variable according to amount of overall variance it explains when combined all together. Fig. 3 showcases the weights of each variable across all districts as a result of the CDI-PCA methodology. Fig. 3a shows the average weight for all 6 months for each variable. Districts are split up by rainfed and irrigated, in green and blue respectively. On average for rainfed districts, SPI1, SPI3 and SM are seen to have the highest median percent contribution while SPI12 and NDVI were the lowest. The average contribution for all variables in rainfed districts ranged between 6.0% (NDVI) to 14.1% (SPI3). Similarly, in irrigated districts, SPI1, SPI3 and SM show the highest median percent contribution and SPI12 and NDVI were the lowest. The average contribution for all variables in irrigated districts ranged between 6.7% (SPI12) to 13.5% (SPI3).The distributions of each variable's weights were similar between both irrigated and rainfed districts.

Fig. 3b shows the average percent contribution of each variable in the CDI, by district. Blue is representative of irrigated districts while green represents rainfed districts. Districts with a darker shade of either color indicate that the average weight over the study period during the growing season was high. Districts that are white indicate that the district is not included in the analysis. With exception of SPI3, the spatial pattern of weights are relatively evenly distributed across the country, highlighting the spread seen in Fig. 3a. SPI3, which contributes the most to the overall variance, shows a cluster of higher values in the eastern part of Balochistan and Western Punjab (majority rainfed districts) and in northern Punjab (majority irrigated districts). In these clusters, the SPI3 contribution is greater than 15%. From Fig. 3b, it is also seen that NDVI has a consistently low contribution, with 75% of the districts between 5 and 7.5%, in both rainfed and irrigated districts.

Fig. 3c shows the contribution of each variable by month. A pattern can be seen across the growing season, for both rainfed and irrigated districts, where SPI12 and NDVI decrease over time and SPI1 increases with each month. Interestingly, EDDI and VPD increase as the season

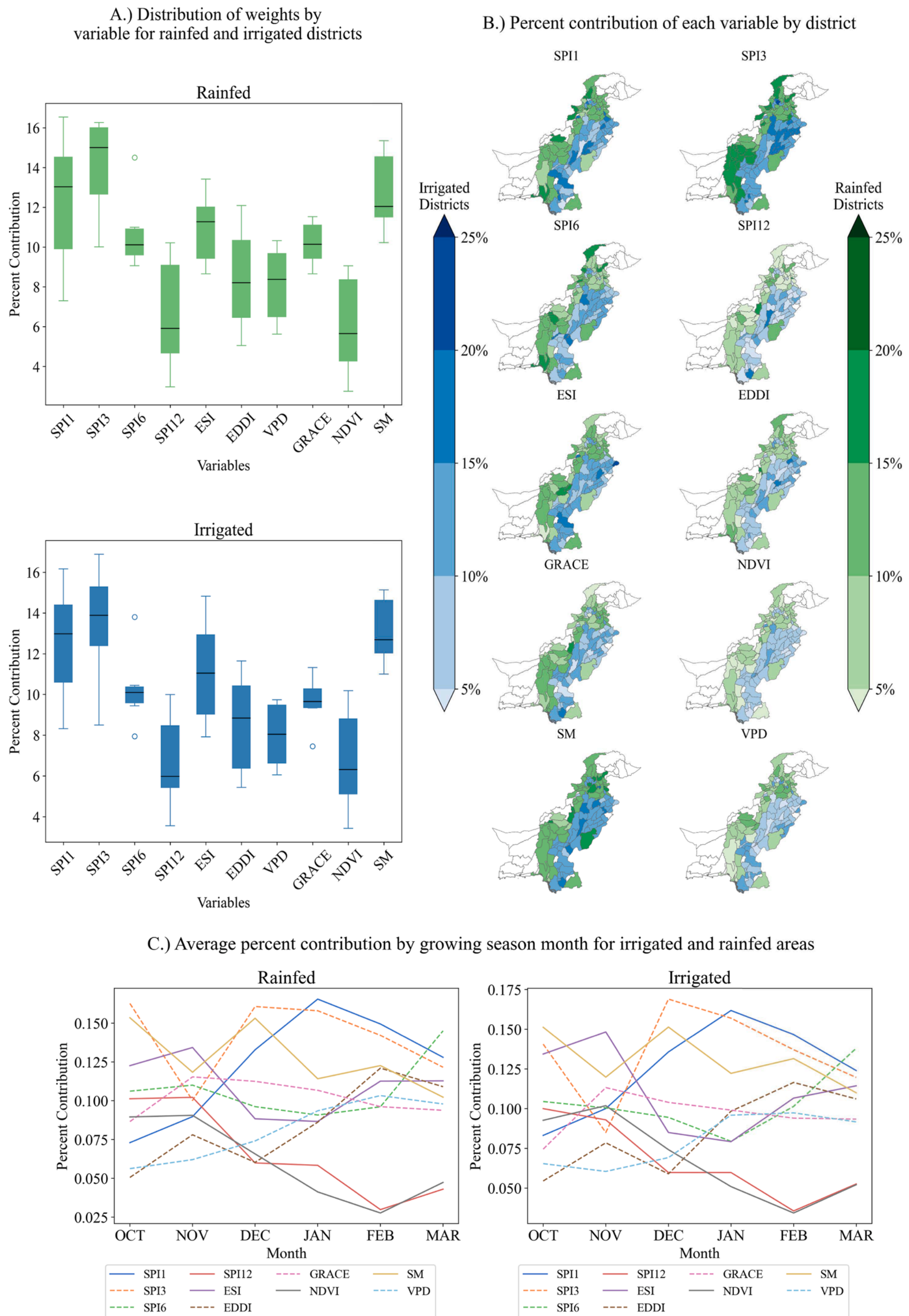

Fig. 3. (A) Visualization of the distribution of the weight of variables used in the CDI by irrigated and rainfed districts. (B) Spatial representation of average percent contribution of each variable in the CDI by district. Blue is irrigated districts and green is rainfed districts. White districts are either excluded from analysis or do not include that variable in the CDI. (C) Average contribution of each variable by month based on irrigated (blue) and rainfed (green).

progresses, where ESI (also estimating evaporative demand) decreases, though ESI is greater on average. Of the three products, ESI is least model dependant. SPI3 is consistently at or above 14% except for November (month 11). A statistical analysis shows that the mean contribution in the 1st three months is significantly different (p-value $<$ 0.05) than the last three months for all variables except SPI3, SPI6, GRACE, and SM.

These results depict a significant seasonal aspect that the climatology of the region exerts on the variables under analysis. In general, the shorter term (1–3 month) precipitation indices are more influential, with SPI1 contributing most to the overall variance later in the season. The shorter term indices are affected by the preceding monsoon rains (July-September) early in the season, whereas the departures from normal appear most important later in the season. For SPI3, the significant drop in November can be explained, in part, due to the fact that the 3 month aggregation includes one month of the prior monsoon period. This would diminish any affect the October/November values may have. In both SPI1 and SPI3, the maximum variance explained occurs in December-January, and are the highest monthly weights across all variable and all months, on average. The SLDAS modeled soil moisture is also an important variable, peaking in December and showing a similar artifact as SPI3. In addition to the short term precipitation, the results also highlight the diminishing contribution of NDVI. This trend seems to point to the fact that the vegetative signals are most important early in the season, near and after planting.

The CDI was further customized to only incorporate variables most import to wheat production, creating the CCDI. This step allowed for variables to be removed from the CDI at the district level. Fig. 4a visualizes the number of variables that were chosen for each district in the CCDI. A rainfed district on average used 5 variables while irrigated used 4. No district used more than 8 variables in the CCDI and two districts, one rainfed and one irrigated, used only one variable, while six rainfed districts used 8 variables compared to one irrigated district. From the bar graph in Fig. 4a, it can be seen the majority of districts used between 3 and 6 variables for the CCDI.

Fig. 4b visualizes the pattern of frequency for all input variables used in rainfed and irrigated districts. The green bar is representative of the percentage of rainfed districts using each variable and the blue is irrigated districts. All variables were used in at least 30% of the districts. Eight out of 10 variables are more common in rainfed districts than in irrigated districts. VPD and SPI3 are the two variables that are used more often in irrigated districts. NDVI and SPI1 had the greatest difference between the frequency in rainfed and irrigated districts. This would be expected because short term precipitation deficits would have a higher impact on rainfed crops than those with irrigated systems. NDVI was used most frequently in rainfed districts and has been found to be closely linked to precipitation (Jain et al., 2009; Ji and Peters, 2003; Rajpoot and Kumar, 2019). Though as seen in the CDI, this impact was most prominent in the early months of the growing season. NDVI should also be closely linked to irrigation, however, the results only show it was used in approximately 40% of districts. Whereas short term precipitation deficits are closely linked to production deficits in rainfed districts, the 3 month deficit (SPI3) was found most related to districts with majority irrigated systems, again aligning with the previous results discussed. SPI3 has a longer memory and can be indicative of broader hydrologic trends, especially early in the season. Overall, the distribution of variables in the irrigated districts were more uniformly distributed, used in 40% of districts on average. The rainfed districts had more variation

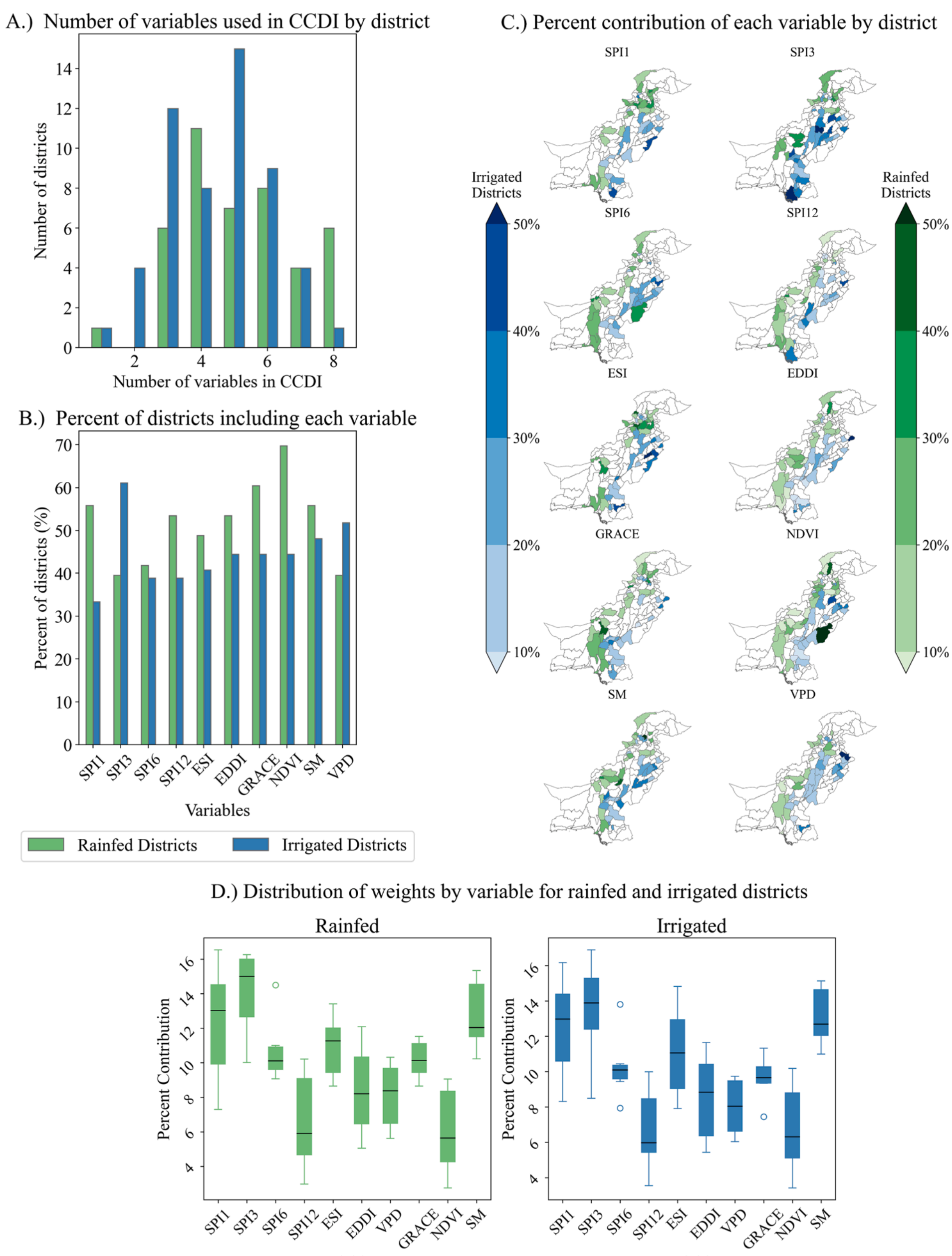

Fig. 4. (A) Visualization of the distribution of the number of variables used in the CCDI by district by irrigated and rainfed districts. (B) Percent of districts that include each variable in the CCDI by irrigated and rainfed districts. The green bars are rainfed and the blue bars are irrigated. (C) Spatial representation of average percent contribution of each variable in the CCDI by district. Blue is irrigated districts and green is rainfed districts. White districts are either excluded from analysis or do not include that variable in the CCDI. (D) Visualization of the distribution of the weight of variables used in the CCDI by irrigated and rainfed districts.

ranging from 40% (VPD, SPI3) to more than 70% (NDVI) of all rainfed districts.

Once a set of variables were identified for each district, the component loading analysis was applied as in the CDI. Fig. 4c visualizes the average weight of each variable by district and summarized by Fig. 4d by variable. Districts that are white indicate that there is zero percent contribution from that variable or it is not included in the CCDI. On average all variables have a higher weight since less variables were used. It can be seen that the general pattern is similar to the full CDI analysis with a few exceptions. Longer-term precipitation measures were weighted relatively higher, especially SPI12 in irrigated areas and ESI had a higher importance.

Overall, the CDI and CCDI showed similar spatial and temporal patterns. For those variables selected in the CCDI, the same monthly trends were observed. The original weight assigned by the CDI did not necessarily correlate to the variables being selected. For example, the NDVI has the least unique contribution to the variance, however, in the customization it was the most used variable in rainfed. In this case, we see that the NDVI does not include much weight (unique variance) in the original CDI, about 6% on average. However, its contribution to the overall variance is highly linked to wheat production (selected in 70% of rainfed districts). Thus, the CCDI is imparting information that is unique to wheat production, specifically. This makes sense in that wheat production itself can be a form of quasi-direct observation of production anomalies as it is a related to vegetation health. This distinction is a result of the added information in the CCDI.

3.2. Spatial drought assessment

When aggreated at the district level the CDI and CCDI are strongly correlated and both indices align well with known climatologically

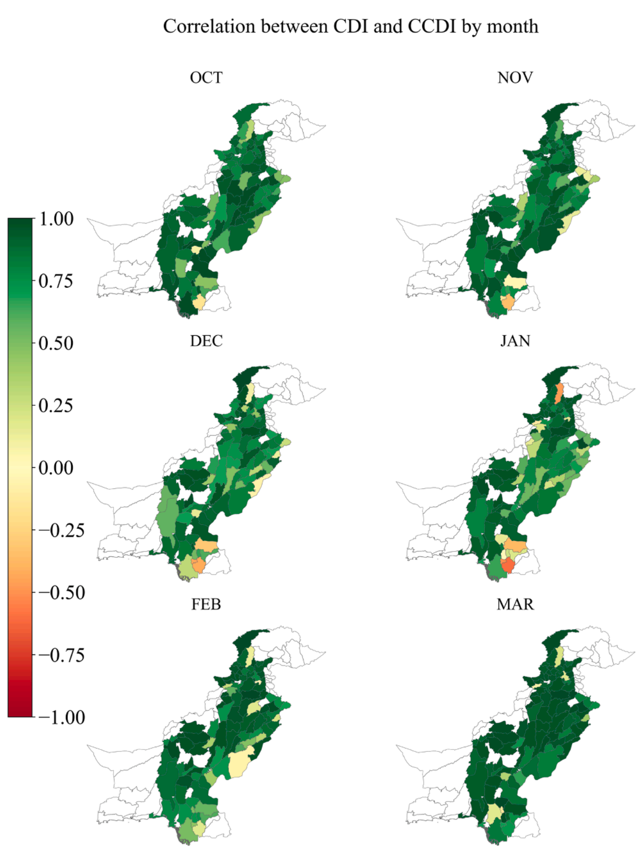

Fig. 5. Pearson's correlation coefficient between CDI and CCDI for the wheat growing season months, October-March. Dark green represents higher correlation while reds indicate negative correlation.

extreme events. Fig. 5 shows the correlation between the CDI and CCDI by district. As the number of variables used in the CCDI increases, the correlation between the CCDI and CDI also increases. Those districts where CCDI used four or more variables, correlation averaged 0.85 and those with less, 0.68. March had the highest average correlation overall of the districts, and January was the lowest, 0.86 and 0.70 respectively. December was the second lowest, at 0.72, indicating that the middle of the growing season saw the most differences between the CDI and CCDI. The second highest correlation was seen in October, the beginning of the growing season, revealing that the CDI and CCDI were the most similar at the beginning and end of the growing season.

Monthly CDI and CCDI values were calculated for each grid cell during the study period of 2004–2017. Because the CDI and CCDI had similar values, looking similar when visualized, only the CDI is shown in Fig. 6. Fig. 6 is a visualization of the CDI for the years 2009 and 2010, where red indicates below normal climatological conditions, indicating drier than usual while green indicates the opposite. The years 2009–2010 included a known dry spell that heavily impacted the provinces of Khyber Pakhtunkhwa and Punjab in the northern part of the country. In many parts of the country, wheat experienced failure due to the lack of rainfall. The CDI demonstrates this phenomenon. For this use case, the CDI shows normal and below normal climatological conditions beginning in June 2009 that grow in intensity through 2010. The catastrophic flooding event that occurred at the end of July and beginning of August 2010 can also be observed.

3.3. CDI and CCDI evaluation with wheat production

Wheat production was used as a proxy for agricultural drought to evaluate the trends of CDI and CCDI. When crop production is below normal, it is expected to see low CDI and CCDI, however, it is understood that not all stressors to production are climate induced. To evaluate the CDI and CCDI with yearly wheat production data, the six monthly values of the winter wheat growing season were composited into a single value that characterized drought conditions. During the assessment of the methods in determining a CDI or CCDI value that represents the growing season conditions, it was found that if the mean index value was used for all growing seasons, the true positive rate was high but the true negative rate was low. When using the minimum, the opposite was true. Using a combination of minimum and mean index values for different growing seasons provided a dynamic approach that was able to classify true positives and true negatives well. The results showed a trade off between sensitivity and specificity. If means were taken, sensitivity would increase, but minimums favored specificity. As stated in methodology, it was determined that the minimum CDI or CCDI value of the growing season would be used to represent the CDI or CCDI if two conditions were met; two or more negative CDI or CCDI months in a growing season and the sum of all negatives in that growing season is less than -0.60. These conditions account for persistence and intensity of a drought event. The mean is used if these conditions are not met. The dynamic thresholding approach allowed for sensitivity and specificity to balance out. The results from this approach for irrigated and rainfed districts are in Table 3.

On average, the two indices were similar with an accuracy ranging from 0.587 (rainfed CDI) to 0.515 (CCDI irrigated) using the dynamic compositing, despite the CCDI only using a fraction of the variables. Fig. 7 shows the average accuracy of the CDI and CCDI across each district. Overall, the statistical analysis showed consistently higher performance metrics for rainfed districts compared to irrigated districts in both the CDI and CCDI. This pattern is expected because wheat production was used as a proxy for drought. Short term droughts are unlikely to impact agricultural land that is supported by large water reservoirs in irrigated areas. Irrigation would be used to alleviate a water deficit before wheat production is affected, creating normal or above normal wheat production. The difference in performance between irrigated and rainfed regions support the findings of Ozelkan et al. (2016).

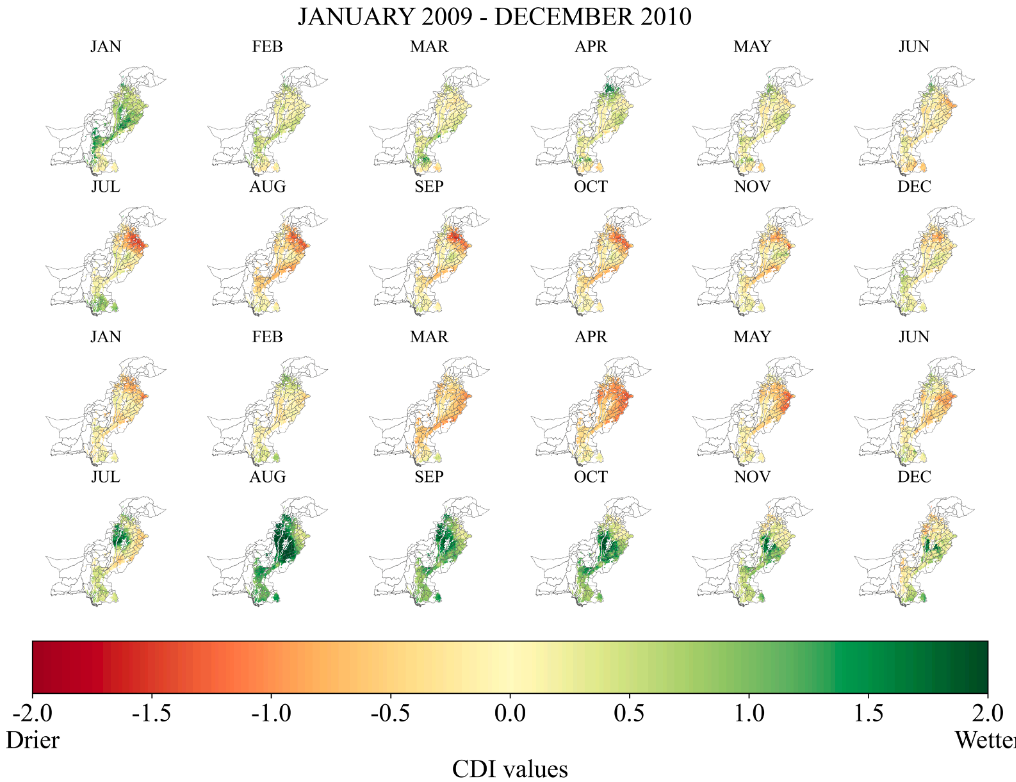

Fig. 6. Visualization of the monthly CDI for the years 2009–2010. Red indicates drier than normal conditions.

Table 3
Statistical analysis results for all irrigated and rainfed districts for the CDI and CCDI.

	Methodology	Accuracy	True Negative Rate	True Positive Rate	AUC	Correlation
CDI - Rainfed	Dynamic	0.587	0.742	0.519	0.631	0.245
CCDI - Rainfed	Dynamic	0.569	0.667	0.533	0.600	0.181
CDI - Irrigated	Dynamic	0.519	0.568	0.491	0.530	0.103
CCDI - Irrigated	Dynamic	0.515	0.602	0.479	0.541	0.135
CDI - Rainfed	Mean	0.646	0.548	0.729	0.638	0.275
CCDI - Rainfed	Mean	0.625	0.530	0.709	0.619	0.210
CDI - Irrigated	Mean	0.567	0.418	0.631	0.526	0.092
CCDI - Irrigated	Mean	0.548	0.428	0.608	0.521	0.112
CDI - Rainfed	Minimun	0.435	0.946	0.135	0.540	0.173
CCDI - Rainfed	Minimun	0.415	0.872	0.154	0.513	0.153
CDI - Irrigated	Minimun	0.355	0.847	0.112	0.480	0.128
CCDI - Irrigated	Minimun	0.380	0.853	0.164	0.509	0.135

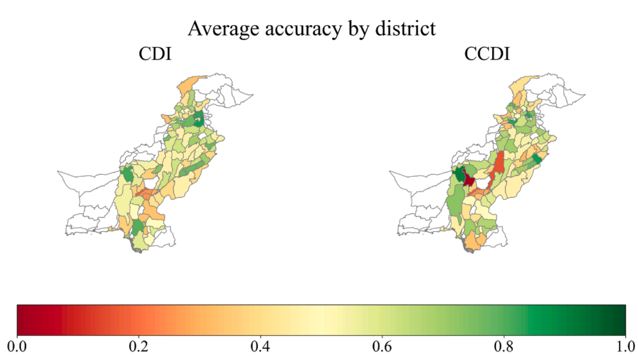

Fig. 7. Visualization of the average accuracy for the CDI and CCDI by district.

When using crop production as a proxy for drought, it is important to know if agriculture is irrigated because drought is more difficult to characterize using Earth observations in these areas (Ozelkan et al., 2016). The largest difference in rainfed and irrigated statistics is in the true negative rate (specificity), 0.742 and 0.568 respectively for the CDI and 0.667 and 0.602 for the CCDI. The true negative rate is the fraction of times the CDI value or CCDI value was negative and the wheat production anomaly was negative.

Overall the CDI and CCDI performed similarly across all statistics and the results show they both are useful drought metrics. Removing variables provides a closer look into what variables are most important in defining agricultural drought event at the district level. Whereas SPI3 and SM were the most heavily weighted, on average, NDVI as found to the variable most used in the rainfed CCDI.

3.4. Sources of error

Crop production data as the proxy for drought may not always provide the most accurate depiction of ground truth data. In this case, there are multiple environmental factors that could contribute to low wheat production. Another source of error in the wheat data would be the consistent practices and protocols of reporting by districts and lack of reporting in some years. In this study, if a district did not report wheat production, then it was not included. This may cause inaccuracies because, in some cases, districts might not have reported data because the production was so low, due to drought or other environmental factors. Other sources of error include how the irrigated and rainfed districts are selected. In some districts, the amount of irrigated and rainfed agriculture is similar (45%/65%) and could cause errors in the evaluation of the CDI and CCDI. To overcome this, it is recommended that irrigated and rainfed land be split within the boundaries of each district. It would be necessary that wheat production reporting also be at a sub-district scale. This would more accurately and specifically identify the ability of the CCDI in irrigated and rainfed areas.

4. Conclusion

The overall objective of this study is to develop a monthly Composite Drought Index (CDI) to provide an assessment of droughts in Pakistan. Further, district level wheat production data was used to optimize variables to create a customized composite drought index (CCDI) specifically for agriculture and compare the two indices.

This study created a monthly composite drought index (CDI) to provide an assessment of drought in Pakistan. The input variables were further optimized to develop the CCDI for a localized drought index and the CDI and CCDI were compared. The variable contribution to the overall variance was evaluated seasonally and spatially. Both indices were evaluated with yearly wheat production data to assess the ability of the index in determining agricultural drought. This study found that, on average the shorter term precipitation and soil moisture metrics (SPI1, SPI3, & SM) contributed the most to the overall combined variance, whereas metrics such as NDVI and SPI12 were the least. There is a distinct seasonal pattern in the variables considered. The influence of SPI1, SPI3, and SM peak mid season, however, the importance of NDVI and SPI12 diminish throughout the season. When customized using wheat production, the CCDI on average used 5 variables compared to the original 10. The down selection of variables in the CCDI was not necessarily dependant on the variables weight in the CDI. The CDI and CCDI were highly correlated and both showed skill in capturing well known climatological events.

District wise wheat production was used as a proxy to agricultural drought to evaluate the performance of the CDI and CCDI. The CDI and CCDI performed better in rainfed districts compared to irrigated districts when using wheat production as a proxy for drought. This is expected since irrigated areas would be able to mitigate the consequences of short term droughts, preventing crop failure. The true negative rates for rainfed districts for the CDI and CCDI were 0.742 and 0.667, respectively. Irrigated districts saw a true negative rate for the CDI of 0.568 and for the CCDI of 0.602, indicating the importance of understanding and separating the type of agriculture within the study area for evaluation. Distinguishing rainfed from irrigated agricultural land when evaluating a drought index created using Earth observations allows for a better understanding of agricultural drought in rainfed areas. Overall the CDI and CCDI performed similarly. The fact that the CCDI uses, on average, 5 less variables than CDI, suggest that there is value in identifying variables specific to agricultural drought. However, though the CCDI does provide a localized context, removing variables by district adds a layer of complexity when integrating into preexisting systems. Thus this study shows that a generalized CDI utilizing all variables achieves an equivalent (and slightly more accurate) characterization of agriculture drought, in both space and time. The findings from this study can be used to improve methodologies and knowledge regarding the identification of drought which could enhance preexisting regional drought monitoring and forecasting systems.

CRediT authorship contribution statement

Caily Schwartz: Conceptualization, Formal analysis, Methodology, Software, Investigation, Visualization, Writing - original draft. **Walter Lee Ellenburg:** Conceptualization, Methodology, Software, Data curation, Investigation, Writing - review & editing. **Vikalp Mishra:** Conceptualization, Methodology, Software, Data curation, Writing - review & editing. **Timothy Mayer:** Conceptualization, Methodology, Writing - review & editing. **Robert Griffin:** Supervision. **Faisal Qamer:** Resources, Writing - review & editing. **Mir Matin:** Conceptualization, Resources. **Tsegaye Tadesse:** Conceptualization, Methodology.

Declaration of Competing 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.

Acknowledgements

We would like to thank SERVIR - Hindu Kush Himalaya for their collaboration through this process, Helen Parache for digitizing production data and providing valuable feedback and Ronan Lucey, also for digitizing production data. Funding for this research was provided through the joint U.S. Agency for International Development (USAID) and National Aeronautics and Space Administration (NASA) initiative SERVIR, particularly through the NASA Applied Sciences Capacity Building Program, NASA Cooperative Agreement NNM11AA01A.

References

Abatzoglou, J., 2021. gridmet. URL http://www.climatologylab.org/gridmet.html.

Abatzoglou, J.T., 2013. Development of gridded surface meteorological data for ecological applications and modelling. Int. J. Climatol. 33.

Adnan, S., Ullah, K., Shouting, G., 2015. Characterization of drought and its assessment over sindh, Pakistan during 1951-2010. URL http://www.esrl.noaa.gov/psd/data/gridded/data.gp.

Adnan, S., Ullah, K., Shuanglin, L., Gao, S., Khan, A.H., Mahmood, R., 2018. Comparison of various drought indices to monitor drought status in Pakistan. Clim. Dyn. 51, 1885–1899. https://doi.org/10.1007/s00382-017-3987-0.

Ahmed, K., Shahid, S., Nawaz, N., 2018. Impacts of climate variability and change on seasonal drought characteristics of Pakistan. Atmos. Res. 214, 364–374.

Anderson, M., Hain, C., Otkin, J., Zhan, X., Mo, K., Svoboda, M., Wardlow, B., Pimstein, A., 2013. An intercomparison of drought indicators based on thermal remote sensing and nldas-2 simulations with u.s. drought monitor classifications. J. Hydrometeorol. 14, 1035–1056.

Anderson, M., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J., Kustas, W., 2011. Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental united states. J. Clim. 24, 2025–2044.

Anderson, M.C., Norman, J.M., Diak, G.R., Kustas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. Remote Sens. Environ. 60.

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental united states based on thermal remote sensing: 2. surface moisture climatology. J. Geophys. Res. 112.

Azam, A., Shafique, M., 2017. Agriculture in Pakistan and its impact on economy-a review. Int. J. Adv. Sci. Technol. 103, 47–60.

Baig, M.B., Shahid, S.A., Straquadine, G.S., 2013. Making rainfed agriculture sustainable through environmental friendly technologies in Pakistan: A review. Int. Soil Water Conserv. Res. 1, 36–52.

Barua, S., Perera, B.J.C., Ng, A.W.M., 2009. A comparative drought assessment of yarra river catchment in victoria, australia. URL http://mssanz.org.au/modsim09.

Bayissa, Y.A., Tadesse, T., Svoboda, M., Wardlow, B., Poulsen, C., Swigart, J., Andel, S.J. V., 2019. Developing a satellite-based combined drought indicator to monitor agricultural drought: a case study for ethiopia. GISci. Remote Sens. 56, 718–748.

Bazrafshan, J., Hejabi, S., Rahimi, J., 2014. Drought monitoring using the multivariate standardized precipitation index (mspi). Water Resour. Manage. 28.

Brown, J.F., Wardlow, B.D., Tadesse, T., Hayes, M.J., Reed, B.C., 2008. The vegetation drought response index (vegdri): A new integrated approach for monitoring drought stress in vegetation. GISci. Remote Sens. 45, 16–46.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40, 16–28.

Crausbay, S.D., Ramirez, A.R., Carter, S.L., Cross, M.S., Hall, K.R., Bathke, D.J., Betancourt, J.L., Colt, S., Cravens, A.E., Dalton, M.S., Dunham, J.B., Hay, L.E., Hayes, M.J., McEvoy, J., McNutt, C.A., Moritz, M.A., Nislow, K.H., Raheem, N., Sanford, T., 2017. Defining ecological drought for the twenty-first century. Bull. Am. Meteorol. Soc. 98, 2543–2550.

Devendra, C., 2012. Review - rainfed areas and animal agriculture in asia: The wanting agenda for transforming productivity growth and rural poverty. Asian-Austral. J. Anim. Sci. 25, 122–142.
FAO, 2018. Disasters causing billions in agricultural losses, with drought leading the way. http://www.fao.org/news/story/en/item/1106977/icode/.
Fawcett, T., 2006. An introduction to roc analysis. Pattern Recogn. Lett. 27 (8), 861–874 rOC Analysis in Pattern Recognition. https://www.sciencedirect.com/science/article/pii/S016786550500303X.
Feng, P., Wang, B., Liu, D.L., 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.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. Sci. Data 2.
Gao, B.-C., 1996. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.
Genuer, R., michel Poggi, J., Tuleau-malot, C., Genuer, R., michel Poggi, J., malot Vsurf, C.T., Package, A.R., Genuer, R., michel Poggi, J., Tuleau-malot, C., 2015. Vsurf: An r package for variable selection using random forests to cite this version: Hal id: hal-01251924 vsurf: An r package for variable selection using random forests. R J. 7, 19–33.
Government of Pakistan Finance Division, 2020. Pakistan Economic Survey 2019–2020.
Gumma, M., Thenkabail, P., Teluguntla, P., Oliphant, A., Xiong, J., Congalton, R., Yadav, K., Phalke, A., Smith, C., 2017. Nasa making earth system data records for use in research environments (measures) global food security-support analysis data (gfsad) cropland extent 2015 south asia, afghanistan, iran 30 m v001 [dataset]. doi: https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30SAAFGIRCE.001.
Hao, Z., Aghakouchak, A., 2014. A nonparametric multivariate multi-index drought monitoring framework. J. Hydrometeorol. 15, 89–101.
Hao, Z., Singh, V.P., 2015. Drought characterization from a multivariate perspective: A review. J. Hydrol. 527, 668–678.
Haroon, M.A., Zhang, J., Yao, F., 2016. Drought monitoring and performance evaluation of modis-based drought severity index (dsi) over Pakistan. Nat. Hazards 84, 1349–1366.
Hayes, M., Svoboda, M., Wall, N., Widhalm, M., 2011. The Lincoln declaration on drought indices: universal meteorological drought index recommended. Bulletin of the American Meteorological Society 92 (4), 485–488.
Heim, R., 2002. A review of twentieth-century drought indices used in the united states. Bull. Am. Meteorol. Soc. 83.
Hobbins, M.T., Wood, A., McEvoy, D.J., Huntington, J.L., Morton, C., Anderson, M., Hain, C., 2016. The evaporative demand drought index. part i: Linking drought evolution to variations in evaporative demand. J. Hydrometeorol. 17, 1745–1761.
Jain, S.K., Keshri, R., Goswami, A., Sarkar, A., Chaudhry, A., 2009. Identification of drought-vulnerable areas using noaa avhrr data. Int. J. Remote Sens. 30, 2653–2668.
Jamro, S., Channa, F.N., Dars, G.H., Ansari, K., Krakauer, N.Y., 2020. Exploring the evolution of drought characteristics in balochistan, Pakistan. Appl. Sci. (Switzerland) 10.
Jamro, S., Dars, G.H., Ansari, K., Krakauer, N.Y., 2019. Spatio-temporal variability of drought in Pakistan using standardized precipitation evapotranspiration index. Appl. Sci. (Switzerland) 11, 9.
Ji, L., Peters, A.J., 2003. Assessing vegetation response to drought in the northern great plains using vegetation and drought indices. Remote Sens. Environ. 87, 85–98.
Keyantash, J., for Atmospheric Research Staff (Eds), N.C., 2018. The climate data guide: Standardized precipitation index (spi). https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi.
Keyantash, J.A., Dracup, J.A., 2004. An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage. Water Resour. Res. 40.
Kourgialas, N.N., Dokou, Z., Karatzas, G.P., 2015. Statistical analysis and ann modeling for predicting hydrological extremes under climate change scenarios: The example of a small mediterranean agro-watershed. J. Environ. Manage. 154.
Liu, Q., Zhang, S., Zhang, H., Bai, Y., Zhang, J., 2020a. Monitoring drought using composite drought indices based on remote sensing. Sci. Total Environ. 711, 134585. https://doi.org/10.1016/j.scitotenv.2019.134585.
Liu, W., Kogan, F., 1995. Monitoring regional drought using the vegetation condition index. Int. J. Remote Sens. 17 (14), 2761–2782.
Liu, X., Zhu, X., Zhang, Q., Yang, T., Pan, Y., Sun, P., 2020b. A remote sensing and artificial neural network-based integrated agricultural drought index: Index development and applications. Catena 186, 104394. https://doi.org/10.1016/j.catena.2019.104394.
Mbow, C., Rosenzweig, C., Barioni, L.G., Benton, T.G., Herrero, M., Krishnapillai, M., Liwenga, E., Pradham, P., Rivera-Ferre, M., Sapkota, T., Tubiello, F., Xu, Y., 2019. Food security. In: Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.
McEvoy, D.J., Huntington, J.L., Hobbins, M.T., Wood, A., Morton, C., Anderson, M., Hain, C., 2016. The evaporative demand drought index. part ii: Conus-wide assessment against common drought indicators. J. Hydrometeorol. 17.
Mendicino, G., Senatore, A., Versace, P., 2008. A groundwater resource index (gri) for drought monitoring and forecasting in a mediterranean climate. J. Hydrol. 357, 282–302.
Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216.
NOAA, Definition of drought. https://www.ncdc.noaa.gov/monitoring-references/dyk/drought-definition.
Otkin, J.A., Svoboda, M., Hunt, E.D., Ford, T.W., Anderson, M.C., Hain, C., Basara, J.B., 2018. Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the united states. Bull. Am. Meteorol. Soc. 99, 911–919.
Ozelkan, E., Chen, G., Ustundag, B.B., 2016. Multiscale object-based drought monitoring and comparison in rainfed and irrigated agriculture from landsat 8 oli imagery. Int. J. Appl. Earth Obs. Geoinf. 44, 159–170.
Palmer, W.C., 1968. Keeping track of crop moisture conditions, nationwide: The new crop moisture index. Weatherwise 21.
Park, S., Im, J., Jang, E., Rhee, J., 2016. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agric. For. Meteorol. 216, 157–169.
Pendergrass, A.G., Meehl, G.A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., Bonfils, C.J., Gallant, A.J., Hoerling, M., Hoffmann, D., Kaatz, L., Lehner, F., Llewellyn, D., Mote, P., Neale, R.B., Overpeck, J.T., Sheffield, A., Stahl, K., Svoboda, M., Wheeler, M.C., Wood, A.W., Woodhouse, C.A., 2020. Flash droughts present a new challenge for subseasonal-to-seasonal prediction. Nat. Clim. Change 10, 191–199. https://doi.org/10.1038/s41558-020-0709-0.
Qaiser, G., Tariq, S., Adnan, S., Latif, M., 2021. Evaluation of a composite drought index to identify seasonal drought and its associated atmospheric dynamics in northern punjab, Pakistan. J. Arid Environ. 185.
Rajpoot, P.S., Kumar, A., 2019. Impact assessment of meteorological drought on rainfed agriculture using drought index and ndvi modeling: a case study of Tikamgarh district, M. P., India. Appl. Geomat. 11, 15–23.
Rajsekhar, D., Singh, V.P., Mishra, A.K., 2015. Multivariate drought index: An information theory based approach for integrated drought assessment. J. Hydrol. 526, 164–182.
Ray, R.L., Fares, A., Risch, E., 2018. Effects of drought on crop production and cropping areas in texas. Agric. Environ. Lett. 3, 170037.
Saini, H.S., Westgate, M.E., 1999. Reproductive development in grain crops during drought. Adv. Agron. 68, 59–96.
Sepulcre-Canto, G., Horion, S., Singleton, A., Carrao, H., Vogt, J., 2012. Development of a combined drought indicator to detect agricultural drought in europe. Nat. Hazards Earth Syst. Sci. 12, 3519–3531.
Sheffield, J., Goteti, G., Wen, F., Wood, E.F., 2004. A simulated soil moisture based drought analysis for the united states. J. Geophys. Res. Atmos. 109 (D24). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD005182.
Sibhatu, K.T., Qaim, M., 2017. Rural food security, subsistence agriculture, and seasonality. PLoS ONE 12, 1–15.
Stull, R., 2000. Meteorology for Scientists and Engineers, second ed.
Svoboda, M., Lecomte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., Stephens, S., 2002. The drought monitor. Bull. Am. Meteorol. Soc. 83, 1181–1190.
Teluguntla, P., Thenkabail, P., Xiong, J., Gumma, M., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Tilton, J., Sankey, T., Massey, R., Phalke, A., Yadav, K., 2016. Global food security support analysis data (gfsad) crop mask 2010 global 1 km.
Tian, L., Leasor, Z., Quiring, S., 2020. Developing a hybrid drought index: Precipitation evapotranspiration difference condition index. Clim. Risk Manage. 29, 100238.
United Nations Office of Disaster Risk Reduction, Human Cost of Disasters.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718.
Wilhite, D., Glantz, M., 1985. Understanding the drought phenomenon: The role of definitions. Drought Mitigation Center Faculty Publications.
Wilhite, D.A., 2000. Chapter 1 drought as a natural hazard. Drought: A Global Assessment 1, 3–18.
World Bank, 2020. Climate change knowledge portal - Pakistan. URL https://climateknowledgeportal.worldbank.org/country/pakistan.
World Meteorological Organization, 2012. Standardized Precipitation Index User Guide.
Wu, J., Zhou, L., Liu, M., Zhang, J., Leng, S., Diao, C., 2013. Establishing and assessing the integrated surface drought index (isdi) for agricultural drought monitoring in mid-eastern china. Int. J. Appl. Earth Obs. Geoinf. 23, 397–410. https://doi.org/10.1016/j.jag.2012.11.003.
Yu, H., Zhang, Q., Sun, P., Song, C., 2018. Impact of droughts on winter wheat yield in different growth stages during 2001–2016 in eastern china. Int. J. Disaster Risk Sci. 9, 376–391. https://doi.org/10.1007/s13753-018-0187-4.
Zaitchik, B.F., Ghatak, D., Matin, M.A., Adhikary, B., Bajracharya, B., Nelson, J., Pulla, S. T., Ellenburg, W.L., 2017. Informing drought preparedness and response with the south asia land data assimilation system.
Zargar, A., Sadiq, R., Naser, B., Khan, F.I., 2011. A review of drought indices. Environ. Rev. 19, 333–349.
Zhang, D., Du, Q., Zhang, Z., Jiao, X., Song, X., Li, J., 2017. Vapour pressure deficit control in relation to water transport and water productivity in greenhouse tomato production during summer. Sci. Rep. 7.
Zhang, Z., Xu, W., Shi, Z., Qin, Q., 2021. Establishment of a comprehensive drought monitoring index based on multisource remote sensing data and agricultural drought monitoring. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 14.
Zhou, Y., Zaitchik, B.F., Kumar, S.V., Arsenault, K.R., Matin, M.A., Qamer, F.M., Zamora, R.A., Shakya, K., 2021. Developing a hydrological monitoring and sub-seasonal to seasonal forecasting system for south and southeast asian river basins. Hydrol. Earth Syst. Sci. 25.