Comparison of Multi-Year Reanalysis, Models, and Satellite Remote Sensing Products for Agricultural Drought Monitoring over South Asian Countries

Muhammad Shahzaman 1,2,3,*, Weijun Zhu 1,2,3,*, Irfan Ullah 1,*, Farhan Mustafa 2,3,*, Muhammad Bilal 4,*, Shazia Ishfaq 5, Shazia Nisar 5, Muhammad Arshad 1, Rashid Iqbal 6,* and Rana Waqar Aslam 7,*

Abstract: The substantial reliance of South Asia (SA) to rain-based agriculture makes the region susceptible to food scarcity due to droughts. Previously, most research on SA has emphasized the meteorological aspects with little consideration of agrarian drought impressions. The insufficient amount of in situ precipitation data across SA has also hindered thorough investigation in the agriculture sector. In recent times, models, satellite remote sensing, and reanalysis products have increased the amount of data. Hence, soil moisture, precipitation, terrestrial water storage (TWS), and vegetation condition index (VCI) products have been employed to illustrate SA droughts from 1982 to 2019 using a standardized index/anomaly approach. Besides, the relationships of these products towards crop production are evaluated using the annual national production of barley, maize, rice, and wheat by computing the yield anomaly index (YAI). Our findings indicate that MERRA-2, CPC, FLDAS (soil moisture), GPCC, and CHIRPS (precipitation) are alike and constant over the entire four regions of South Asia (northwest, southwest, northeast, and southeast). On the other hand, GLDAS and ERA5 remain poor when compared to other soil moisture products and identified drought conditions in regions one (northwest) and three (northeast). Likewise, TWS products such as MERRA-2 TWS and GRACE TWS (2002–2014) followed the patterns of ERA5 and GLDAS and presented divergent and inconsistent drought patterns. Furthermore, the vegetation condition index (VCI) remained less responsive in regions three (northeast) and four (southeast) only. Based on annual crop production data, MERRA-2, CPC, FLDAS, GPCC, and CHIRPS performed fairly well and indicated stronger and more significant associations (0.80 to 0.96) when compared to others. Thus, the current outcomes are imperative for gauging the deficient amount of data in the SA region, as they provide substitutes for agricultural drought monitoring.

Keywords: South Asia; agricultural drought; rainfall; soil moisture; TWS; VCI; standardized index; standardized anomalies; YAI
1. Introduction

South Asia (defined as Pakistan, Bangladesh, India, Afghanistan, Nepal, Bhutan, and Sri Lanka) relies deeply on rain-fed subsistence agriculture, which is gradually becoming more susceptible to droughts [1–3]. South Asia (SA) is one of the most disaster-prone and densely populated regions in the world, comprising over one-fifth of the world’s population [4]. Population increases and frequent drought events have been leading issues that have furthered food shortages and water crises [5]. According to a recent study conducted by Zhai et al. [6], a significant increasing trend of droughts has been recorded over SA. Moreover, strong increases in average drought frequencies and durations have been projected. A drought is an abnormally dry climate period and is one of the most destructive natural hazard types, and droughts can be distinguished into hydrological, meteorological, and agricultural droughts [7]. An agrarian drought includes a discrepancy in accessible water that could compromise crop production [8], and this can have overwhelming effects on crop production and water supply [9]. About 31% of the agricultural lands of SA have been influenced by droughts during the past two decades, representing a severe hazard to the economic and social development of the region [10]. According to the World Economic Forum (WEF), a global annual economic loss of USD 6–8 billion is caused by agricultural droughts and related activities [11]. To increase security against droughts caused by global changes, it is essential to find comprehensive, reliable, and timely drought monitoring methods [10,12]. Inclusive descriptions of droughts in SA, like in several other regions worldwide, faces challenges when using ground-based precipitation datasets. For instance, rainfall variability (spatial) cannot be measured satisfactorily because of an uneven and sparse distribution of rain gauges [12]. Aadhar et al. [5] developed a high resolution (0.05°) precipitation (bias-corrected) data collection method which can be used for near real-time drought monitoring in SA. Additionally, breaks/gaps in records and poor proceedings in data handling confound the use of weather station data [9,13]. In many studies, ground/station-based precipitation datasets replaced with satellite, model output, and reanalysis products have provided consistent and reliable data with regional and global coverage at several spatial/temporal levels, which is considered satisfactory for drought detection and characterization [14]. Sharma et al. [15] evaluated satellite-based precipitation products in terms of gauge stations precipitation data and found that their performance was better for drought monitoring in Nepal. Similarly, a study carried by Bai et al. [16] showed consistent performance for satellite-based precipitation datasets for drought detection in China.

In addition to model outputs and satellite products, normalized difference vegetation index (NDVI) [17–20] and Gravity Recovery and Climate Experiment (GRACE) water storage datasets [12,21] have been used to assess drought. The NDVI has been used directly, or occasionally in a derivative form, such as a vegetation condition index (VCI), [22–24], in many studies to monitor the effects of a drought on agriculture. Many drought studies have been carried out in different regions of world and have been purely based on precipitation data [13,25,26]. The aforementioned studies and some others [27,28] have used standardized precipitation index (SPI), NDVI, and soil moisture (SM) products for agricultural drought assessment; however, for a region like SA, the majority of people depend upon rain-fed agriculture [1]. Further studies concentrating on crop production and agrarian drought influences would be more helpful to people in SA. Hence, the current study focuses on drought behaviors in agricultural regions in terms of soil moisture, precipitation, water storage, and NDVI data derived from satellite, model, and reanalysis products. Further, the existing study considers the effectiveness of products in the context of annual crop yields, which has not been of significant focus in prior studies.

To provide reliable information for agricultural drought detection when using various products, it is indispensable to recognize the most active agricultural drought indicators for the SA region. Consequently, the primary goals of study are: (1) to highlight/illustrate droughts in agricultural regions of countries in SA in terms of the spatial magnitude and severity by using satellite, model, and reanalysis products. Further, (2) to assess how
well the data are related to agricultural droughts in the context of comparison with crop production data. This is a unique and comprehensive study regarding the performance of satellite remote sensing, reanalysis, and model products. Besides, GRACE satellite data are used for agricultural drought detection over all countries in SA for the first time here, thus proving the link between total water storage and crop yields.

2. Materials and Methods

2.1. Study Area

This study focuses on SA, which is situated between 5° to 40°N and 60° to 100°E (Figure 1A). This area has a diverse range of climatic zones, including tropical, sub-tropical, mountainous, humid, alpine, dry land, and desert areas with bimodal rainfall (summer and winter monsoon) regimes. SA features four seasons (spring, summer, autumn, and winter). This region is famous for its reversal of winds during the monsoon season. Monsoon rainfall characteristics vary between areas in terms of the spatial and temporal scales [2]. The summer monsoon season, along with southwesterly winds, represents about 75% of annual rainfall [29]. Besides, the winter monsoon season typically covers a vast area, including the northern and northwest regions. The western disturbances caused by the Mediterranean Sea and Atlantic Ocean (MSAO) remain influential in these areas during the winter season. South Asian agriculture mainly depends upon the summer monsoon season. About 70% of people in SA related with direct or indirect agricultural practices. The major source of income in SA is agriculture, which depends on rainfall patterns [1]. India is the leading country in South Asia and second largest country worldwide, having 160 million hectares of agricultural land [30]. Pakistan is the second largest country in SA and contributes about 23.3 million hectares of land in the agriculture sector [31]. Bangladesh has 9.3 million hectares of agricultural land and Afghanistan has 8 million hectares [32]. The other countries in SA contribute far lower proportions. South Asia has a total land area of 5.2 million km² [10]. Figure 1A demonstrates the typical elevation above sea level and the overall geographical location, while Figure 1B indicates MODIS-based land cover for SA.

2.2. Data

2.2.1. Precipitation

(I). The Climate Hazard Group Infrared Precipitation with Stations (CHIRPS) dataset provides global data (50°N–50°S) at monthly, five-day, and daily intervals with a 0.05° spatial resolution. CHIRPS data comprise the combination of satellite estimation and in-situ/station observation data based upon cold cloud duration (CCD) data. Primarily, CHIRPS data have been produced for agricultural drought monitoring [33]. Monthly product version 2.0 was downloaded from ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/ (accessed on 9 February 2021) for the period 1982 to 2019. CHIRPS precipitation data are frequently used for hydrology and drought-related studies [12,28,34].

(II). The Global Precipitation Climatology Center (GPCC) reanalysis dataset, version 7, at a monthly basis with a 0.5° spatial resolution, has been accessed via ftp://ftp.dwd.de/pub/data/gpcc/html/fulldata_v7_doi_download.html (accessed on 12 February 2021) from 1982 to 2019. Additionally, this product is gauge-gridded based on 75,000 global rain gauge stations and has been used for drought monitoring [12].
Figure 1. (A) Elevation map of South Asia. (B) Land cover types in the study area as generated with the MODIS MCD12 dataset.

2.2. Soil Moisture

(I). The second Modern-Era Retrospective Analysis for Research and Application (MERRA-2) soil moisture (reanalysis) product is a replacement of the original MERRA reanalysis product of the US National Aeronautics and Space Administration [35]. The monthly soil moisture product, having a 0.625° by 0.5° root zone, has been downloaded from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/ (accessed on 12 February 2021) for the duration of 1982 to 2019. Due to better data forcing (observation-corrected precipitation) and an upgraded assimilation system (upgraded canopy interception), MERRA-2 provides enhanced soil moisture (SM) evaluation when compared to MERRA [36].

(II). The ERA5 reanalysis dataset is a replacement of ERA-Interim (reanalysis) dataset and is based on a four-dimensional integration which enhances essential dynamics and model physics [37]. The European Center for Medium Range Weather Forecasts offers a temporal series of various variables from 1981 to the present. The monthly ERA5 average reanalysis soil moisture (SM) product is available with a regular grid size of 0.1° × 0.1° [38]. These data have been used for the period of 1982–2019 as per the https://www.ecmwf.int/en/forecasts/datasets website (accessed on 6 March 2021).

(III). The Climate Prediction Center (CPC) provides several SM products. Soil moisture version 2 data, at a monthly basis with a 0.5° spatial resolution, are used in the existing study for the period from 1982 to 2019. This dataset was acquired from the laboratory database of the National Oceanic and Atmospheric Administration.
website http://www.esrl.noaa.gov/psd/data/gridded/data.cpcsoil.html (accessed on 16 February 2021). This dataset is considered in present study because it includes in situ precipitation as an input and thus is considered to provide data that would be very close to the real soil moisture data [39].

(IV). The Global Land Assimilation System (GLDAS), with the soil moisture version 2 dataset, was used at a monthly basis with a 1° spatial resolution as downloaded from http://disc.sci.gsfc.nasa.gov/services/grads-gds/glades (accessed on 20 February 2021) for the period of 1982 to 2014. All three layers of soil moisture were aggregated as one for further processing [40].

(V). The Famine Early Warning System Network (FEWS NET) Land Data Assimilation System (FLDAS) is a conventional system of NASA that has been upgraded to work with domains and data streams and provide monitoring and forecasting within the context of nutrition safety estimation in data-sparse and underdeveloped countries [12]. The VIC and NOAH land surface models and supplement FLDAS. The FLDAS NOAH soil moisture product, at a monthly basis with a 0.1° spatial resolution, was downloaded for the period of 1982 to 2019 from ftp://hydro1.sci.gsfc.nasa.gov/data/s4pa/FLDAS/FLDAS_NOAH01_C_EA_M.001/ (accessed on 26 February 2021). This dataset resulted from an arrangement of the CHIRPS and MERA-2 datasets. FLDAS and GLDAS (NOAH model) were considered because of their extensive usage by atmospheric and land modelling societies. Consequently, the model limitations are satisfactory reliable and have been tested well. Additionally, several studies have used these limitations around the world [25,41,42].

2.2.3. Terrestrial Water Storage (TWS)

(I). Gravity Recovery and Climate Experiment (GRACE) satellite operations have been providing monthly temporal gravity variations (global) in terms of spherical harmonic coefficients (SHCs) since 2002 [21,43]. The Center for Space Research’s (CSR) release five (RL05) monthly spherical harmonic coefficients, downloaded from http://igem.gfz-potsdam.de/ICGEM/shms/monthly/csr-rl05/ (accessed on 20 February 2021) for the period of 2002 to 2017. The GRACE terrestrial water storage measurements agree with the geophysical model and Earth rotation-derived variations [44], and hence are used globally for drought-related studies [45,46].

(II). The MERRA-2 land water storage product, used in addition to GRACE, was acquired from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/ (accessed on 26 February 2021) on a monthly basis with a 0.5° latitude to 0.625° longitude resolution. This product is commonly used for agricultural drought monitoring in the period of 2002–2019 [12].

2.2.4. Normalized Difference Vegetation Index

NOAA Advanced Very High Resolution Radiometer (AVHRR) normalized difference vegetation index (NDVI) data, with 15-day temporal and 0.083° × 0.083° spatial resolutions, have been considered to calculate VCI data [47]. The data were downloaded from http://ecocast.arc.nasa.gov/data/pub/gimms/3g.v0/ (accessed on 28 February 2021) for the period of 1982 to 2019. NDVI data are advantageous and are extensively used in local and global scales to characterize weather-related vegetation stresses [9,17,18,48].

2.2.5. Crop Yield

Annual barley, maize, rice, and wheat yield data for Afghanistan, Pakistan, India, Bangladesh, Nepal, and Bhutan for 1982 to 2019 were used, obtained from the Food and Agricultural Organization (FAO) data portal http://www.fao.org/faostat/en/#data/QC (accessed on 20 February 2021). For Sri Lanka, the same crop product data could not be acquired because of the unavailability of datasets. These data are regularly used as the most credible crop production data [12,49] to assess the efficiency of model and satellite
products in terms of monitoring agrarian droughts. The summary of the datasets is given in Table 1.

**Table 1. Summary of the datasets.**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Products</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
<th>Duration</th>
<th>Primary References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>CHIRPS</td>
<td>Monthly</td>
<td>0.05° × 0.05°</td>
<td>1982–2019</td>
<td>Funk et al., 2015; McNally et al., 2017, 2016; Shukla et al., 2014 [26,33,34,42].</td>
</tr>
<tr>
<td></td>
<td>GPCC</td>
<td>Monthly</td>
<td>0.5° × 0.5°</td>
<td>1982–2019</td>
<td>Agutu et al., 2017; Dutra et al., 2014; C. Funk et al., 2014; Kurnik et al., 2011 [12,50–52].</td>
</tr>
<tr>
<td>Soil moisture</td>
<td>FLDAS</td>
<td>Monthly</td>
<td>0.1° × 0.1°</td>
<td>1982–2019</td>
<td>Agutu et al., 2017; Jung et al., 2020; McNally et al., 2017 [12,34,53].</td>
</tr>
<tr>
<td></td>
<td>GLDAS</td>
<td>Monthly</td>
<td>1° × 1°</td>
<td>1982–2014</td>
<td>Agutu et al., 2017; Chen et al., 2020; Jung et al., 2020 [12,53,54].</td>
</tr>
<tr>
<td></td>
<td>CPC</td>
<td>Monthly</td>
<td>0.5° × 0.5°</td>
<td>1982–2019</td>
<td>Fan and van den Dool, 2004; Wu, 2014 [55,56].</td>
</tr>
<tr>
<td></td>
<td>MERRA-2</td>
<td>Monthly</td>
<td>0.625° × 0.5°</td>
<td>1982–2019</td>
<td>Chen et al., 2020, 2019; Le et al., 2020 [54,57,58].</td>
</tr>
<tr>
<td></td>
<td>ERA5</td>
<td>Monthly</td>
<td>1° × 1°</td>
<td>1982–2019</td>
<td>Almendra-Martín et al., 2021; Cheng et al., 2019 [38,59].</td>
</tr>
<tr>
<td>TWS</td>
<td>MERRA-2</td>
<td>Monthly</td>
<td>0.625° × 0.5°</td>
<td>1982–2019</td>
<td>Agutu et al., 2017; Bosilovich et al., 2015 [12,36].</td>
</tr>
<tr>
<td>VCI</td>
<td>NDVI</td>
<td>15 days</td>
<td>0.083° × 0.083°</td>
<td>1982–2019</td>
<td>Ali et al., 2019a; Bhuiyan et al., 2006; Karnieli et al., 2010; Wang et al., 2001 [20,48,61,62].</td>
</tr>
</tbody>
</table>

**2.3. Methods**

Because of the connection between precipitation anomalies and agricultural drought [38,51], standardized indices (SI) were calculated, e.g., standardized precipitation indices (SPI) for CHIRPS and GPCC precipitation data products and standardized soil moisture indices (SSMI) for the MERRA-2, ERA5, GLDAS, FLDAS, and CPC soil moisture data products. Likewise, a standardized terrestrial water storage index (TWSI) for MERRA-2 terrestrial water storage data was computed to characterize agricultural droughts [63,64]; however, standardized anomaly (SA)/Z-scores were calculated for the GRACE water storage product because of the short time period from 2002–2017 [12]. The subsequent standardized indices (SIs) and standardized anomalies (SAs) were subjected to k-means clustering analysis to identify the main temporal and spatial occurrences over homogenous sub-regions. Finally, proportional variance and correlation in each region was
assessed, along with the relationships between field/crop data and significant testing at 0.05 significance level. The SPI, SSMI, TWSI, SA, and VCI conditions are only presented and discussed here for the June to August period. The scale and duration were taken under consideration of South Asian agriculture depending upon monsoon rainfall [65].

Datasets with different resolutions cannot be directly processed together [66–69]. All datasets other than GLDAS and GRACE were aggregated spatially at 1° by 1° resolution before standardization in order to maintain uniformity. Comparisons of agricultural drought information were primarily related to several datasets with the same variables, such as the soil moisture, precipitation product, and water storage product datasets as the existing study is completely focused on agricultural droughts over the region. The cropland cells were extracted from the MODIS land cover dataset, then keeping those cells as a standard for all other datasets (CHIRPS, GPCC, MERRA-2, CPC, FLDAS, GLDAS, ERA5, MERRA-2 TWS, GRACE TWS and AVHRR-VCI) during filtering/extraction.

### 2.3.1. Standardized Precipitation Index (SPI)

Precipitation is considered an essential constituent for long-term drought monitoring and estimation in any region of interest. The use of such indices is regarded as an ingenious standardized approach based on precipitation scarcity [68]. The SPI is the most important and frequently used index for drought monitoring [15,70–73]. This index computes a gamma probability distribution function for a precipitation time series. Furthermore, it involves the transformation of accumulated gamma probability to a cumulative distribution function for a standard normal distribution [74]. Because of the sensitivity in the calculated SPI values to the fitted parametric distribution, a non-parametric SPI fitting approach has been applied in the current study via use of the Standardized Drought Analysis Toolbox (SDAT) [74]. The various SPI drought intensities [12,75] used to characterize droughts are given in Table 2. The SDAT offers a widespread outline for the origins of non-parametric univariate and multivariate standardized indices (SI). A non-parametric framework is typically used for deriving non-parametric standardized drought indices against various land surface and climate variables, such as precipitation, soil moisture, humidity, and terrestrial water storage. The mathematical concept regarding the calculation of indices behind SDAT is given by Equations (1) and (2) [12,74]:

\[
SI = \phi - 1 \left( p \right)
\]  

where \( \phi \) represents a standardized normal distribution function and \( p \) is the probability, which can be calculated using Equation (2):

\[
p(xi) = \frac{i - 0.44}{n + 0.12}
\]

where \( p(xi) \) is the empirical probability, \( i \) indicates the rank of non-zero data (precipitation, soil moisture, and terrestrial water storage) from the smallest to largest, while \( n \) denotes the sample size. Standardized indices can be computed by moving the outputs of Equation (2) into Equation (1). The subsequent indices, i.e., the SSMI, SPI, and standardized terrestrial water storage index (TWSI), were computed using similar approaches. Moreover, the same drought categories (Table 2) were used for the SSMI and TWSI. The spatial patterns were scaled in the form of \( \pm 1 \). Thus, time-based assessment indicated the genuine magnitude (Table 2) of SI/SA, where spatial patterns have values near \( \pm 1 \). Interpretation of the spatial patterns has been taken under consideration with the context of temporal evaluation and indicates spatial drought development pattern at any time when the temporal values of evaluation drop below \(-0.84 \) [12].
Table 2. SPI drought categories [75].

<table>
<thead>
<tr>
<th>SPI</th>
<th>Drought Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;1.65</td>
<td>Extreme wetness</td>
</tr>
<tr>
<td>&gt;1.28</td>
<td>Severe wetness</td>
</tr>
<tr>
<td>&gt;0.84</td>
<td>Moderate wetness</td>
</tr>
<tr>
<td>&gt;−0.84 and &lt;0.84</td>
<td>Average (normal)</td>
</tr>
<tr>
<td>&lt;−0.84</td>
<td>Moderate dryness</td>
</tr>
<tr>
<td>&lt;−1.28</td>
<td>Severe dryness</td>
</tr>
<tr>
<td>&lt;−1.65</td>
<td>Extreme dryness</td>
</tr>
</tbody>
</table>

2.3.2. Standardized Anomalies (SA)

For GRACE data products, SAs were computed instead of SIs in order to categorize agricultural drought. This was because the GRACE products feature short timeframes. Here, TWSI was obtained like SPI [12]. The anomalies were computed for 3 months, removing the monthly mean, and then dividing by the standard deviation [76] as follows:

\[ XY_{abc} = \frac{X_{abc} - \frac{1}{n} \sum_{c=1}^{n} X_{abc}}{\sigma_{ab}} \]  

where \( XY_{abc} \) specifies the standardized TWS (monthly) for location \( a \), month \( b \), and year \( c \); \( X_{abc} \) is the average TWS (monthly) for location \( a \), month \( b \), and year \( c \); and \( n \) demonstrates the length of years; however, \( \sigma_{ab} \) is the standard deviation (long-term) for location \( a \), month \( b \), and year \( c \). The resulting SAs (z-scores) indicate the deviations of TWS near mean values and are typically used to monitor droughts globally [64,77,78]. Index values <0 designate drought conditions and values >0 indicate wet conditions, while a value of 0 indicates average conditions [12,64].

To validate the uniformity between the SA and SI data in terms of characterizing agricultural droughts over South Asia, CHIRPS-derived SPI and CHIRPS-derived SA decompositions (spatiotemporal) were compared over the study region, showing a parallel drought pattern. Additionally, a significant positive correlation (>0.90) was seen between the SI and SA temporal patterns. Because of the close bond/association between the SI and SA data, the SPI drought categories (Table 2) were considered to sufficiently distinguish droughts [12,64].

2.3.3. Vegetation Condition Index (VCI)

The VCI approach has been used to achieve greater precision and strength in determining drought verdicts as it is advantageous to indicate weather-linked vegetation stress [9,20,24,47]. Droughts based on vegetation are closely associated with climatic impacts. Higher VCI values denote wet or stable moisture contents, representing unstressed vegetation conditions, while lower values demonstrate vegetation stress or drought [20]. The VCI data were calculated by using Equation (4):

\[ \text{VCI} = \frac{\text{NDVI}_i - \text{NDVI}_{\text{min}}}{\text{NDVI}_{\text{max}} - \text{NDVI}_{\text{min}}} \times 100 \]  

The NDVI\(_i\) represents a monthly/yearly value of the NDVI, while NDVI\(_{\text{min}}\) and NDVI\(_{\text{max}}\) are the long-term minimum and maximum values of NDVI, respectively. The AVHRR NDVI data considered in this study are used globally for drought-related studies.

2.3.4. K-Means Clustering Algorithm

South Asia has multifaceted climatic regimes, where precipitation fluctuates considerably over the entire region [10]. Moreover, the region can also be affected other factors, such as the given topography, landscape, and seaside locality [79]. A K-means clustering technique was employed with three-month sets of SPI, SSMI, TWSI, and VCI data to assess drought occurrence over different homogenous sub-regions. The algorithm generates clusters with the aim of: (I) decreasing inconsistency with number of clusters \( k \); and (II)
increasing the erraticism of each centroid amongst all data. The algorithm provides better partition outputs, particularly when applied against a larger number of grids. In this algorithm, the data points were assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroid would be at a minimum. Less variation within clusters leads to more similar data points being located within the same cluster. The resulting clusters were used to divide large regions into sub-regions. Thus, this study followed the same pattern to divide the SA region. The names of the regions (northwest, southwest, northeast, and southeast) were given based on their locations in the South Asia region. This method is used worldwide, such as by Santos et al., who used k-means clustering in Portugal for the spatial and temporal monitoring of drought patterns [80]. Furthermore, Li et al. [81] have used k-means clustering in China.

2.3.5. Correlation between Crop Yield Anomaly (YAI) and Drought Indices

Scientific investigations based on remote sensing datasets pertaining to agricultural drought assessments require validation in order to understand the result certainty. Typically, ground/field-based measurements are considered for validation. As crop yield data are available at the country level, SIs/SAs were re-calculated against each country. In the current study, major crop statistics for wheat, maize, rice, and barley were used from 1982–2019. Consequently, a long-term yield anomaly index (YAI) over each of the countries in the study region was performed [65]. YAI values were computed to find the deviations in yield for specific years. Firstly, this approach was applied for each year separately. Secondly, the results were used to further identify an overall connotation amongst the YAI and drought indices. Generally, crop production experiences increasing trends because of advances in adaptation and technology in agriculture. Thus, a linear regression technique was applied to remove trends from the data [82]. This index can be calculated using the equation [65,83] given below:

\[ \text{YAI} = \frac{(\gamma - \mu)}{\sigma} \]

where \( \gamma \) denotes the yield of a precise year, whereas \( \mu \) indicates the average (long-term) yield; however, \( \sigma \) signifies the standard deviation. Besides, the correlation matrix among drought indices, percentage variance, significant testing at a 0.05 significance level, and general order of products based on their performance in each region were considered.

3. Results and Discussion

3.1. Spatial Variability

The four most substantial components (four regions) of the study area in terms of explaining variability indicated separate patterns (spatial) for all datasets (Figures 2 and 4). The MERRA-2, CPC, FLDAS (soil moisture), CHIRPS, and GPCC (precipitation) data remained almost alike and constant over the entire four regions of South Asia; however, in the case of the ERA5 and GLDAS data, the spatial patterns of the standardized soil moisture indexes (SSMI) were dissimilar when compared to the other SM products (Figure 2). Equally, the standardized terrestrial water storage index (TWSI) of MERRA-2 and GRACE products indicated different patterns for regions one and three (Figure 4). This could be endorsed to the circumstance that region two covers southern and southwestern Pakistan and almost all of western India. This region represents the core belt of the typical monsoon area and hence has consistent rainfall, causing smaller discrepancies to be produced [83,84]. On the other hand, region four comprises an entire region of southern India and Sri Lanka; however, this region features highly variability because of dry and wet extremes [2], thus representing significant variations in the SI/SA spatial patterns.
3. Results and Discussion

3.1. Spatial Variability

The four most substantial components (four regions) of the study area in terms of explaining variability indicated separate patterns (spatial) for all datasets (Figures 2 and 3). The MERRA-2, CPC, FLDAS (soil moisture), CHIRPS, and GPCC (precipitation) data remained almost alike and constant over the entire four regions of South Asia; however, in the case of the ERA5 and GLDAS data, the spatial patterns of the standardized soil moisture indexes (SSMI) were dissimilar when compared to the other SM products (Figure 2). Equally, the standardized terrestrial water storage index (TWSI) of MERRA-2 and GRACE products indicated different patterns for regions one and three (Figure 3). This could be endorsed to the circumstance that region two covers southern and southwestern Pakistan and almost all of western India. This region represents the core belt of the typical monsoon area and hence has consistent rainfall, causing smaller discrepancies to be produced [83,84]. On the other hand, region four comprises an entire region of southern India and Sri Lanka; however, this region features highly variability because of dry and wet extremes [2], thus representing significant variations in the SI/SA spatial patterns.

Figure 2. (a–e) The spatial patterns of different standardized soil moisture index (SSMI) over South Asia. These patterns have been scaled in the form of ±1, hence temporal assessment indicates the genuine magnitude (Figure 3) of SSMI, where spatial patterns have values near ±1.

Figure 3. (a–d) Temporal assessment of standardized soil moisture index data over the study region. Temporal valuation interpreted in conjunction with Table 2.
Figure 3. (a–e) Spatial patterns of (a) SPI from CHIRPS, (b) SPI from GPCC, (c) TWSI from MERRA-2, (d) SA from GRACE, and (e) VCI from AVHRR over South Asia. These patterns have been scaled in the form of ±1, and hence temporal assessment indicates the genuine magnitude (Figure 5) of SI/SA, where spatial patterns have values nearby ±1.

Figure 5. (a–d) Temporal assessment (SPI from CHIRPS, SPI from GPCC, TWSI from MERRA-2, SA from GRACE, and VCI from AVHRR) over the study region. Temporal valuation interpreted in conjunction with Table 2.
The vegetation condition index is only presented with the spatial drought pattern in regions three and four (Figure 4). This is because vegetation-based indices are only associated with green plant biomass. Hence, such indices are considered lethargic and less receptive to climatic variables [85]. Table 3 shows the topographical coverage of study region. The total variance of SI/SA for all corresponding products in all regions was explained to be between 22.71% (ERA5) to 84.43% (MERRA-2 TWS). Most of the products indicated the highest or lowest discrepancy over the study region, as given in Table 4.

Table 3. Geographical coverage of study region.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Countries/Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (northwest)</td>
<td>Afghanistan; northern, eastern, and central parts of Pakistan; Jammu and Kashmir, northern India</td>
</tr>
<tr>
<td>Region 2 (southwest)</td>
<td>Southern and southwestern Pakistan and western India</td>
</tr>
<tr>
<td>Region 3 (northeast)</td>
<td>Central and eastern India, Nepal, Bhutan, and Bangladesh</td>
</tr>
<tr>
<td>Region 4 (southeast)</td>
<td>Southern India and Sri Lanka</td>
</tr>
</tbody>
</table>

Table 4. Percentage of variance as exposed by different products for all four regions.

<table>
<thead>
<tr>
<th>Region</th>
<th>MERRA-2 (SSMI)</th>
<th>CPC (SSMI)</th>
<th>FLDAS (SSMI)</th>
<th>GLDAS (SSMI)</th>
<th>ERA5 (SSMI)</th>
<th>CHIRPS (SPI)</th>
<th>GPCC (SPI)</th>
<th>MERRA-2 (TWSI)</th>
<th>GRACE (SA)</th>
<th>AVHRR (VCI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.70</td>
<td>16.6</td>
<td>17.97</td>
<td>11.62</td>
<td>4.58</td>
<td>14.09</td>
<td>18.11</td>
<td>20.29</td>
<td>1.30</td>
<td>19.1</td>
</tr>
<tr>
<td>2</td>
<td>14.06</td>
<td>17.4</td>
<td>14.82</td>
<td>10.58</td>
<td>9.05</td>
<td>16.70</td>
<td>17.03</td>
<td>25.96</td>
<td>0.56</td>
<td>7.45</td>
</tr>
<tr>
<td>3</td>
<td>15.60</td>
<td>15.7</td>
<td>18.55</td>
<td>7.70</td>
<td>4.79</td>
<td>17.76</td>
<td>15.80</td>
<td>9.78</td>
<td>9.64</td>
<td>5.47</td>
</tr>
<tr>
<td>Mean</td>
<td>63.25</td>
<td>63.4</td>
<td>63.89</td>
<td>35.81</td>
<td>22.71</td>
<td>64.04</td>
<td>64.34</td>
<td>84.43</td>
<td>17.54</td>
<td>39.15</td>
</tr>
</tbody>
</table>

3.2. Temporal Variability

The temporal assessment of the aforementioned spatial patterns, pertaining to all regions, is presented in Figures 3 and 5. Generally, the temporal estimation is interpreted in combination with spatial patterns (Table 2) [12]. It has been revealed that maximums for the regions in terms of suffering from severe to extreme drought during 1988–89, 1991–92, 1999, 2000–04, and 2016–18 feature negative values ranging from –0.84 to –1.65. All products had different drought patterns in different regions, which may be attributed to the geographical locations of regions. Further, the tendencies and durations of droughts exceedingly depended upon the amount of precipitation received in a particular area [83,86]. The performances of the soil moisture products (MERRA-2, CPC, and FLDAS) were determined to be almost similar over the entire study region, having little differences between index values; however, inconsistency has been identified for the performances of the GLDAS and ERA5 soil moisture products (Figure 3), especially in regions one and three. Finally, the soil moisture datasets mainly appeared in two classes/groups based on their performances. The MERRA-2, CPC, and FLDAS data represent one category that was found to be interrelated in regard to the outputs over the study region, while ERA5 and GLDAS were not found to be interrelated. The performances of the precipitation products (GPCC and CHIRPS) were found to be better and were close to each other. This is because both datasets contain in situ measurements of rainfall. The CHIRPS data features precipitation estimation based on a combination of satellite data and in situ measurements, while the GPCC data purely features in situ information [33].

In relation to precipitation products, the GRACE terrestrial water storage (GRACE TWS), MERRA-2 terrestrial water storage (MERRA-2 TWS), and vegetation condition index (VCI) data showed comparatively diverse responses in the temporal evaluation. This is visible in Figure 3. In the case of the MERRA-2 TWS and GRACE TWS data, no
substantial performance was observed, especially in regions one and two. The SI/SA values had remained either positive or slightly negative, and typically almost near to zero. This could be attributed to the delayed response of the variation in the TWS to rainfall and soil moisture [12]. Similarly, the VCI demonstrated minor variances in index values and no tendency towards drought detection, except for regions three and four. The VCI indicated dominantly wet data in regions 1 and 2, and some extent of wetness in region 3, which is opposite to the general trends of the soil moisture and precipitation products. These drought indices, which are based on vegetation alone, were considered to be less responsive when compared to the evaporative indices because the evaporative indices feature less influence by vegetation growth and diagnose instant deficits in wetness through differences in evapotranspiration [85]. The current temporal evaluations clearly support the spatial analyses.

In addition, the correlation analysis confirmed the precision of the spatial and temporal analyses, indicating close associations between various products (Figure 6). Among soil moisture products, MERRA-2, CPC, and FLDAS revealed strong relationships across all regions when compared to GLDAS and ERA5. Further, the connotations between the precipitation products (CHIRPS and GPCC) were similar to those between the previously mentioned soil moisture products (MERRA-2, CPC, and FLDAS).

![Figure 6. Relationships between drought indices for all four regions of South Asia.](image)

Commented [M65]: The hyphen of Figures needs to be changed to a minus sign. Please provide more high-resolution pictures and replace it.

Commented [n66R65]: Suggested changes have been done. Picture is replaced.
Significant \((p < 0.05)\) correlations between drought indices (MERRA-2, CPC, FLDAS, CHIRPS, and GPCC) have been observed in regions two and three, which supports the same performance, as observed in Figures 2 and 4. As a whole, the coefficient matrix exposed the strong relationships for soil moisture products (MERRA-2, CPC, and FLDAS) between each other, ranging from 0.74 to 0.87; however, their association towards ERA5 and GLDAS has been noted to be weak and non-significant \((p > 0.05)\), ranging from 0.14 to 0.50 in regions one, two, and four. Only region two indicated the highest value of 0.63 between MERRA2 and GLDAS. In the case of the precipitation products (CHIRPS and GPCC), they were found to be significantly correlated to each other over the entire region, while their relationships with MERRA-2, CPC, and FLDAS remained mixed. In regions one, two, and three, they showed fairly positive and strong correlations. On the other hand, in region four, GPCC only specified a good relationship (0.73) with MERRA-2. In addition to the soil moisture and precipitation, the water storage products (MERRA-2 TWS and GRACE TWS) demonstrated a weak correlation with all products, except MERRA-2 TWS and CPC (0.74) in region three. Likewise, the VCI data had fragile, negative, or non-significant relationships.

All the products (GLDAS, ERA5, MERRA-2 TWS and GRACE TWS, and VCI) demonstrated dissimilar outputs in the spatial drought severity maps (Figures 2 and 4) temporal evaluations (Figures 3 and 5), and correlation matrix analyses (Figure 6) when compared to other products. These variations could be ascribed to the differences in the preparation of products. Besides, various environments and impacting factors may also influence drought patterns. For instance, local soil characteristics may affect drought severity, as suggested by the soil moisture products. Similarly, rainfall properties (amount, duration, and intensity) may change the drought harshness, as shown by the precipitation products [12].

3.3. General Ranking of Products Based on Overall Performances

We used Taylor diagrams to assess and elaborate upon the capabilities of all products and their rankings in terms of their overall performance within the context of drought detection (Figure 7). The spatial, temporal, and crop yield analyses confirmed that the MERRA-2 soil moisture (SM) product performed equally well over the entire region when compared to the other products. Hence, the ranking of all other products was carried out relative to MERRA-2. From the Taylor diagram (Figure 7), the products were separately grouped based upon the standard deviation, correlation coefficient, and root mean square difference (RMSD) in each region. Out of 10 various products, the MERRA-2, CPC, and FLDAS soil moisture products seemed to be closely connected and performed fairly well and performed better, presenting a high correlation coefficient and low RMSD in each region. The orders of ranking were recorded as first, second, and third for MERRA-2, FLDAS, and CPC in regions one, two, and four, respectively; however, in region two, CPC remained second after MERRA-2. Further, it was perceived that the GPCC and CHIRPS rainfall products persisted in the fourth and fifth positions for regions one, two, and four, respectively. In region two, CHIRPS has remained fourth and was after the CPC soil moisture product. All other products showed the least performance for drought capturing and a low coefficient for correlation. Thus, they were placed at lower positions.
Figure 7. (a–d) Taylor diagrams indicating the overall rankings of products (SSMI from MERRA-2, FLDAS, CPC, GLDAS and ERA5, SPI from CHIRPS and GPCC, TWSI from MERRA-2, SA from GRACE, and VCI from AVHRR) relative to MERRA-2 soil moisture over all regions of South Asia. The angular axis denotes the coefficient of correlation and the radial axis indicates the standard deviation. The semi-circle demonstrates the root mean square difference (RMSD).

3.4. Relationship between Drought Indices and Crop Yield Anomaly

Figure 8 stipulates the association between the YAI and drought indices (all products) during the study period over the entire region. The relationships varied substantially among all the indices for all countries. As the time durations of the GRACE and GLDAS products were different and comparatively shorter than the other products, the data were divided and the YAI was computed two times. Figure 8 separately represents the performances of the drought indices against field data for the periods of 2002–2014 (a–f) and 1982–2019 (g–l) over the six countries of South Asia. Standardized indices (SI) were calculated for the longer period (1982–2019), whereas standardized anomalies (SA) were used for the shorter period (2002–2014). Sri Lanka was excluded because of the unavailability of crop yield data (specifically for wheat and barley). During the short time period from 2002–2014 (a–f), MERRA-2, CPC, and FLDAS (soil moisture products), along with CHIRPS and GPCC (precipitation products), performed well for all crop yield data over the four major countries, i.e., Pakistan, India, Bangladesh, and Afghanistan. Significant correlations ($p < 0.05$) were recorded, ranging from 0.80 to 0.96; however, it has been noted that the performance remained diverse for all crops in the other two countries, where they showed the least performance overall for barley in Nepal and wheat in Bhutan. The results specified that the products such as GLDAS, ERA5, MERRA-2 TWS, GRACE TWS, and VCI performed poorly as compared to MERRA-2, CPC, FLDAS, CHIRPS and GPCC over the study region.
specified that the products such as GLDAS, ERA5, MERRA-2 TWS, GRACE TWS, and VCI performed poorly as compared to MERRA-2, CPC, FLDAS, CHIRPS and GPCC over the study region.

In the case of the long-term YAI from 1982–2019 (g–l), the above-mentioned products indicate a miscellaneous relationship. MERRA-2, CPC, FLDAS, CHIRPS, and GPCC showed mixed but significant associations against different crops in all countries, except Bhutan. CHIRPS and GPCC demonstrated strong correlations with wheat. As a whole, both precipitation products performed equally during 2002–2014; however, CHIRPS performed relatively better than GPCC during 1982–2019. This could be attributed to the fact that CHIRPS has satellite-based estimation data in addition to as well as rain gauge data. According to investigations carried out by Chen et al. [87], spatial resolution plays substantial role in the performance of gridded products, unlike GPCC, which only features a rain gauge record. Hence, its performance is reliant on topographical changes and gauge density data [33]. On the other hand, ERA5, GLDAS, MERRA-2 TWS, GRACE TWS, and VCI have been identified as the least responsive datasets, with lower correlation compared to other products. The poor performances of these products could be connected to poor performances in drought detection, as seen in Sections 3.1 and 3.2. The products related to soil moisture indicate the rainwater which left behind after evaporation and run-off. Consequently, they are more likely to expose variability in crop yield rather than rainfall. Thus, their weak presentation could be related to how well they fit the area/region [12]. Moreover, both FLDAS and GLDAS are products of the same model, while FLDAS demonstrated better performance than GLDAS. This is because, for FLDAS, the NOAH model is forced by CHIRPS [12]. The inconsistent performance of ERA5, GLDAS, MERRA-2 TWS, GRACE TWS, and VCI products positively endorsed our spatial and temporal analyses. In the Figure 8, the dark red color indicates a significantly positive correlation, while the dark green color indicates a significantly negative correlation. The study led by Hamal et al. [82]
supports our results more effectively and is in line with outputs regarding the correlations among crop yields and SIs/SAs.

4. Conclusions

The current study has investigated the performances of soil moisture products (MERRA-2, CPC, FLDAS, GLDAS, and ERA5), precipitation products (CHIRPS and GPCC), and terrestrial water storage products (MERRA-2 TWS and GRACE TWS) in order to support agricultural drought characterization in South Asia. This was performed using standardized index/standardized anomaly and K-means algorithms. Additionally, the study computed YAI data to assess the relationships between field measurements (crop yield data) and drought indices.

The drought characterizations (spatial and temporal) elucidated MERRA-2, CPC, FLDAS (soil moisture), GPCC, and CHIRPS (precipitation) as being alike and constant over the four considered regions (northwest, southwest, northeast, and southeast) of South Asia. On the other hand, GLDAS and ERA5 performed poorly when compared to the other soil moisture products and identified droughts in regions one (northwest) and three (northeast). The TWS products (MERRA-2 TWS and GRACE TWS) followed the patterns of ERA5 and GLDAS and showed dissimilar and inconsistent drought patterns. GRACE TWS and MERRA-2 TWS highlighted droughts in the regions three (northeast) and four (southeast). In addition to the soil moisture and precipitation products, the VCI data identified droughts in regions three and four. From ground/field measurement analysis, stronger and significant \( p < 0.05 \) associations (0.80 to 0.96) were recorded for MERRA-2, CPC, and FLDAS for soil moisture and CHIRPS and GPCC for precipitation over the six considered countries (Pakistan, India, Bangladesh, Afghanistan, Nepal, and Bhutan). Besides, low consistency and weaker relationships were identified for all other products (ERA5, GLDAS, MERRA-2 TWS, GRACE TWS, and VCI). Further, based on the performances of all products, this study suggests the use of the MERRA-2, CPC and FLDAS datasets for soil moisture and CHIRPS and GPCC datasets for precipitation as optimal indicators for agrarian drought detection in South Asian countries. Further investigations should examine how well the model, reanalysis, and satellite-based datasets fit the area/region.

Author Contributions: M.S. and W.Z. established the unique idea and methodology for the current study; F.M., I.U. and M.B. contributed formal analysis and interpretation; M.A., S.I., S.N., R.W.A. and R.I. helped in the manuscript writing process. All authors have read and agreed to the published version of the manuscript.

Funding: The National Natural Science Foundation of China (grant number 41575070) supported the present study.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: M.S., the principal author, and all co-authors are extremely grateful to all remote sensing data suppliers and their staff for developing and upholding the websites used in existing study. We owe our deepest appreciation to Nanjing University of Information Science and Technology (NUIST) and the China Scholarship Council (CSC) for providing obligatory support and fellowship. We would extend our heartfelt thanks to the anonymous reviewers for their valuable time and courtesy.

Conflicts of Interest: The authors declare no conflict of interest.

References


