scientific **data**

DATA DESCRIPTOR

Check for updates

An improved global vegetation health index dataset in detecting vegetation drought

Jingyu Zeng [1,2], Tao Zhou[1], Yanping Qu[3], Virgílio A. Bento[4], Junyu Qi[5], Yixin Xu[1], Ying Li[1,6] & Qianfeng Wang [2] ✉

Due to global warming, drought events have become more frequent, which resulted in aggravated crop failures, food shortage, larger and more energetic wildfires, and have seriously affected socio-economic development and agricultural production. In this study, a global long-term (1981–2021), high-resolution (4 km) improved vegetation health index (VHI) dataset integrating climate, vegetation and soil moisture was developed. Based on drought records from the Emergency Event Database, we compared the detection efficiency of the VHI before and after its improvement in the occurrence and scope of observed drought events. The global drought detection efficiency of the improved high-resolution VHI dataset reached values as high as 85%, which is 14% higher than the original VHI dataset. The improved VHI dataset was also more sensitive to mild droughts and more accurate regarding the extent of droughts. This improved dataset can play an important role in long-term drought monitoring but also has the potential to assess the impact of drought on the agricultural, forestry, ecological and environmental sectors.

Background & Summary

Drought is a complex phenomenon[1], and can trigger crop failures, food shortages, famines, epidemics, and even mass migration[2,3]. Severe drought events were responsible for widespread negative impacts on natural and socioeconomic systems[4,5]. Therefore, it is urgent to improve our understanding of the spatiotemporal characteristics and the evolutionary trends of droughts[6,7]. This would provide a basis for quantifying drought impacts and the social, economic and natural ecological responses to droughts across regions and times[8].

According to different subjects, drought usually presents four processes or types: meteorological (insufficient precipitation), agricultural, also called vegetation drought (insufficient soil moisture), hydrological (insufficient runoff and/or groundwater), and socioeconomic (social response to water supply and demand)[9]. Among these, vegetation drought is highly correlated with food production, and the extent of vegetation drought have been increasing[10,11]. The global average harvested area has decreased by 4.1% during droughts from 1964–2007[12]. The number and duration of severe drought events in the world are expected to maintain its growth in the future[13,14]. An efficient technique to detect vegetation drought events and monitor the potential risk of droughts in the long term is thus urgently needed[15,16]. This may be seen as a key asset to help policymakers implement timely mitigation and adaptation strategies[17].

Traditional meteorological drought index has been widely used for vegetation drought assessment[18–20]. For example, the self-calibrating Palmer drought severity index (sc-PDSI)[21] significantly improves spatial homogeneity, taking into account soil moisture conditions. Four values related to the soil moisture are computed along with their complementary potential values. These eight values are evapotranspiration (ET), recharge (R), runoff (RO), loss (L), potential evapotranspiration (PE), potential recharge (PR), potential runoff (PRO), and potential loss (PL). The PDSI depends on a two-stage "bucket" model of the soil[22]. The top layer of soil is

[1]Key Laboratory of Environmental Change and Natural Disasters of Ministry of Education, Beijing Normal University, Beijing, 100875, China. [2]College of Environment & Safety Engineering, Fuzhou University, Fuzhou, 350116, China. [3]Research Center on Flood and Drought Disaster Reduction, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China. [4]Universidade de Lisboa, Faculdade de Ciências, Instituto Dom Luiz, 1749-016, Lisboa, Portugal. [5]Earth System Science Interdisciplinary Center, University of Maryland, 5825 University Research Ct, College Park, MD, 20740, USA. [6]Zhejiang Institute of Meteorological Sciences, Hangzhou, 310008, China. ✉e-mail: wangqianfeng@fzu.edu.cn

assumed to hold one inch of moisture. The amount of moisture that can be held by the rest of the underlying soil is a location-dependent value, which must be provided as an input parameter to the program. The sc-PDSI automatically calibrates the behavior of the index at any location by replacing empirical constants in the index computation with dynamically calculated values[21]. Compared with other meteorological drought indices [such as the standard precipitation index (SPI) and the standardized precipitation-evapotranspiration index (SPEI)], the sc-PDSI has a higher correlation with vegetation drought[23]. However, these indices still face problems such as inaccurate station observation and delayed data collection[7,24], and do not consider either the impact of water and heat stress on vegetation growth or do not fully consider land cover and vegetation information of the underlying surface[25]. Therefore, the traditional meteorological drought index is not suitable for direct use in vegetation drought research. But using it as an auxiliary information supplement tool and combining it with other relevant technologies can help us better understand and study vegetation drought. Currently, satellite data are widely used for drought assessment, because of its ability to identify drought conditions on different underlying surfaces[26]. Remotely sensed drought monitoring has advantages such as being suitable for large-scale drought monitoring[27], having a powerful real-time update function[28,29], and high accuracy, as well as unmatched cost-effectiveness when compared to other methods[30,31]. Therefore, the combination of satellite remote sensing technology and traditional drought index is a very promising drought assessment method.

The vegetation health index (VHI) is one of the most popular remote sensing drought monitoring indices[32–35]. VHI is composed by two terms: the vegetation condition index (VCI) and the thermal condition index (TCI). VHI considers local biophysical and climatic conditions, and can be used for actual plant drought monitoring in various agrometeorological regions[36]. The basic principles of VHI are as follows: (1) a low normalized differences vegetation index (NDVI) and high land surface temperature (LST) suggest poor vegetation health[37,38]; and (2) the contributions of VCI and TCI to VHI are assumed to be equal, since there are no data on the relative contributions of other conditions to vegetation health[35]. However, the contributions of TCI and VCI to VHI depend on climatic and other environmental factor[39]. Indeed, environmental conditions in different regions are usually different, which implies that giving equal weights to VCI and TCI may reduce the application prospects of the VHI, increasing the uncertainty in drought detection[33]. Therefore, evaluating the contributions of VCI and TCI to VHI while considering the underlying surface and hydrothermal conditions is a key issue that needs to be addressed. Bento, et al.[39] used SPEI index combined with Pearson correlation analysis to evaluate the contribution of TCI and VCI to VHI in arid regions, while Zeng, et al.[23] compared SPEI and sc-PDSI as a control drought index to evaluate the contribution of TCI and VCI to VHI.

The purpose of this study is to produce a global, long-term (1981–2021), high-spatial-resolution, improved vegetation drought detection dataset and a reference parameter dataset for calculating the improved VHI for different regions. Here, we supplemented meteorological and soil moisture information on the basis of the VHI of the original algorithm and used the sc-PDSI as the control drought index, combined with the Pearson correlation analysis method, to evaluate the best contributions of VCI and TCI to VHI in different regions. Then, we generated a global long-term dataset for detecting vegetation drought. We also used drought event records from the Emergency Events Database (EM-DAT) to assess the drought detection ability of our improved VHI dataset. The developed global vegetation drought dataset has the potential to monitor and assess drought and its impact on the agricultural, forestry, ecological, and environmental sectors.

Data and Methods

Dataset coverage. Dataset coverage ranges from approximately from −50°S to 70°N and from −180°W to 180°E. Antarctica, the high latitude regions of the Northern Hemisphere (due to the lack of VCI and TCI data in this area), and the Sahara Desert region in Africa (due to the lack of sc-PDSI in this area) were excluded from the study area. These areas have little vegetation or extremely low vegetation coverage, so this part is usually excluded when studying vegetation drought[40–43], and these areas are not considered in this dataset. The colors on Figure 1 represent the number of drought events since 1900. Figure 1 also shows information on the change in global annual mean temperature (Fig. 1b), change in global annual precipitation (Fig. 1c), and the change in global annual number of droughts (Fig. 1d). Since the 20th century, global temperatures have substantially increased, and precipitation also showed a positive trend, but with high fluctuations and large uncertainties. Significantly higher temperatures and greater uncertainty in precipitation could dramatically increase drought risk. According to statistics, drought events have been on the rise since 1900 ($p < 0.001$), and especially since the 1980s.

Data sources. Global VCI and TCI data were downloaded from the National Oceanic and Atmospheric Administration (NOAA) Center for Satellite Application and Research (STAR) (https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_ftp.php), with a spatial resolution of 4 km, a weekly temporal resolution, and a time span ranging from 1981 to 2021. We organized the data in annual values through the arithmetic mean, and processed the background value (−9999) as a null value through the spatial masking technique to facilitate subsequent spatial and related statistical analyses.

Global sc-PDSI data were downloaded from the Climate Research Unit (CRU), with a spatial resolution of 0.5°, a monthly temporal resolution, and a time span ranging from 1901 to 2020 (https://crudata.uea.ac.uk/cru/data//drought/#global)[44,45]. We organized the data in annual data by the arithmetic mean, then resampled to 4 km by the nearest neighbor method to match the VCI and TCI data. Resampling the 0.5 ° sc-PDSI to 4-km will not lose the information it contains, making the results acceptable. However, the use of sc-PDSI with higher spatial resolution may provide more information on hydrothermal and soil characteristics and may reduce the uncertainty of results.

Data on global drought events from 1900 to 2021 were obtained from the EM-DAT (https://public.emdat.be/)[46]. After filtering and eliminating invalid data, a total of 592 valid records from 1981 to 2021 were selected. We have
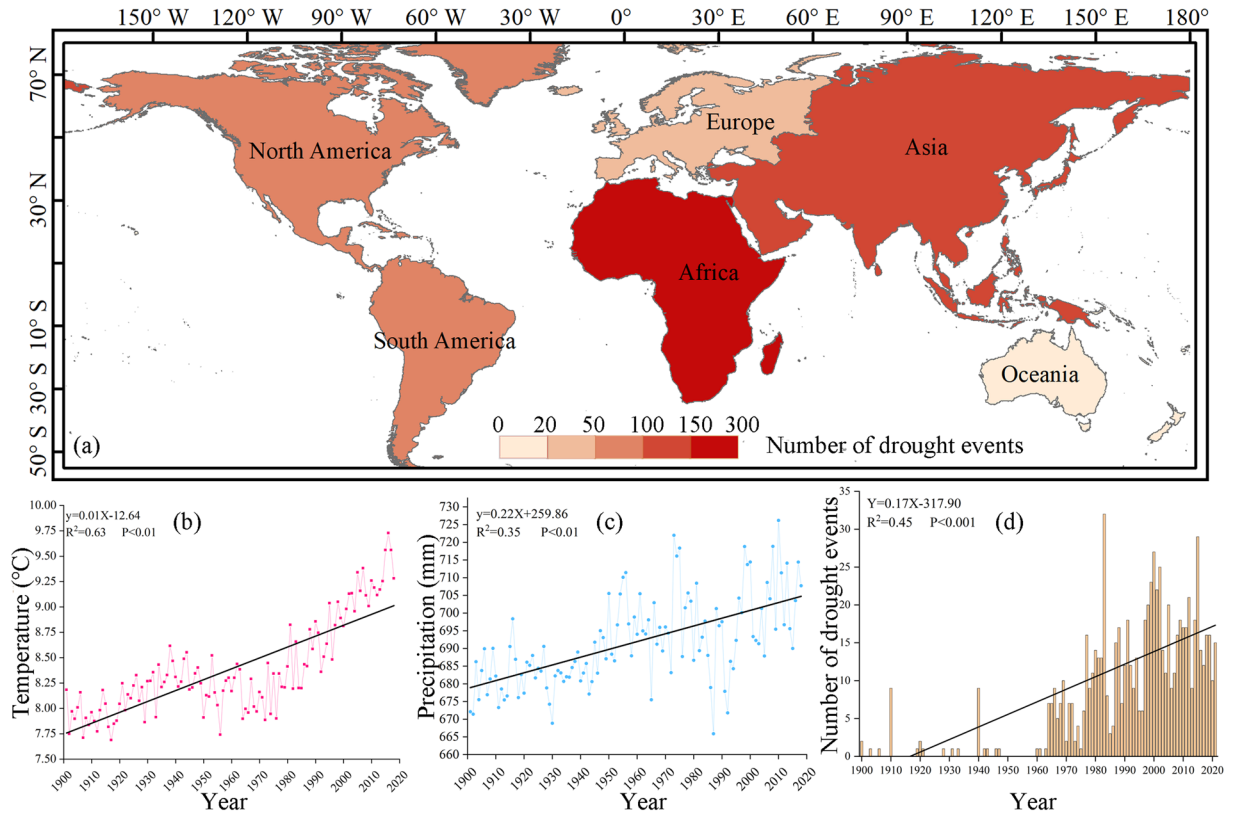
Fig. 1 Dataset Coverage (**a**) Number of global drought events, (**b**) the change in global annual mean temperature, (**c**) the change in global annual precipitation, and (**d**) the change in the number of global annual droughts. The black lines in (**b**–**d**) are the best-fit lines from linear regression. The relations, coefficients of determination, and p-values are given in each panel. The data in a and d are sourced from EM-DAT, and the data in b and c are sourced from CRU.

uploaded the drought event data records for verification to figshare, which can be obtained from https://doi.org/10.6084/m9.figshare.19811854.v5[47].

Annual VHI calculation. According to Kogan[35], the VCI for each pixel and period in a given year was calculated as follows:

$$VCI = \frac{NDVI - NDVI_{MIN}}{NDVI_{MAX} - NDVI_{MIN}} \times 100 \tag{1}$$

where NDVI is the value of a given pixel and period, and $NDVI_{MIN}$ and $NDVI_{MAX}$ are the minimum and maximum values of NDVI for all pixels and periods, respectively. Equation (2) was used to calculate the TCI:

$$TCI = \frac{LST_{MAX} - LST}{LST_{MAX} - LST_{MIN}} \times 100 \tag{2}$$

where LST is the value of a given pixel and period. LST_{MIN} and LST_{MAX} are the minimum and maximum values of LST for all pixels and periods, respectively. The VHI represents the overall health of the vegetation and is used to identify drought[48]. It is calculated by combining the VCI and the TCI as follows:

$$VHI = a \times VCI + (1 - a) \times TCI \tag{3}$$

where a determines the contributions of VCI and TCI to VHI, which varies depending on the environment of the study area[33]. The original VHI (VHI_ori) assumes equal contributions from water demand (here, a proxy of NDVI) and temperature during plant growth, and the coefficient a is assigned the value of 0.5. Following other studies[35,49,50], we classified drought levels on the basis of the VHI (Table 1):

Improvement of the VHI algorithm. We chose the meteorological drought index sc-PDSI, which considers hydrothermal conditions and soil moisture, as the control drought index, and combined it with Pearson correlation analysis to evaluate the contribution of VCI and TCI on a grid by grid basis, thereby obtaining an improved VHI index. To improve VHI, the following steps are taken: (1) The parameter a was set to vary in steps of 0.02,

Category	VHI
Extremely dry	[0, 10]
Severely dry	(10, 20]
Moderately dry	(20, 30]
Mild dry	(30, 40]
Normal	(40, 50]
Good	(50, 60]
Excellent	(60, 100]

Table 1. Vegetation drought classification based on the VHI.

starting from 0.02 and gradually increasing to 0.98 (49 values in total). The VHI was then calculated in each step for each year from 1981 to 2021 based on different a values. The specific formula is as follows:

$$VHI_{i,t,a} = a \times VCI_{i,t} + (1 - a) \times TCI_{i,t}(a = 0.02, 0.04, \ldots 0.98)$$ (4)

Where $VHI_{i,t,a}$ represents the VHI when the VCI contribution of the ith pixel is a at time t. $VCI_{i,t}$ represents the VCI of pixel i at time t, and $TCI_{i,t}$ represents the TCI of pixel i at time t.

(2) The Pearson correlation was then used to evaluate the correlation between sc-PDSI and the VHI calculated in each iterative step of contribution values a pixel by pixel. The sc-PDSI based on the water balance theory fully considers meteorological factors and soil moisture conditions, suitable for improving the VHI[23]. The Pearson correlation coefficient is calculated as follows:

$$R_i = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_i)(y_i - \overline{y}_i)}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x}_i)^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}}$$ (5)

Where R_i represents the correlation coefficient between VHI and sc-PDSI of pixel i, x_i represents the VHI of pixel i, $\overline{x}_i$ represents the average value of VHI of pixel i from 1981 to 2021, y_i represents the sc-PDSI of pixel i, and $\overline{y}_i$ denotes the mean value of sc-PDSI for pixel i from 1981 to 2021. The range of $|R|$ is 0 to 1.

(3) After obtaining the spatial correlation coefficient map of sc-PDSI and VHI calculated with different a values, the optimal contribution value a was estimated pixel by pixel according to the following formula:

$$a_{i,opt} = MAX\ R(VHI_{i,a}, scPDSI_i)$$ (6)

Where $a_{i,opt}$ represents the best contribution value a of VCI to VHI at pixel i, and (1-$a_{i,opt}$) represents the best contribution value of TCI to VHI. R is the correlation coefficient between VHI and sc-PDSI, $VHI_{i,a}$ is the VHI of pixel i when the contribution value of VCI is a, and $scPDSI_i$ is the sc-PDSI of the ith pixel. By comparing the correlation coefficients between sc-PDSI and VHI calculated for the 49 values of a pixel by pixel, the a value of the largest correlation coefficient was selected as the best contribution value a of the VCI to VHI of this pixel.

(4) Finally, according to the best contribution value a and the calculation formula of VHI, the improved VHI, namely VHI_{opt}, is obtained. The specific formula is as follows:

$$VHI_{i,t,opt} = a_{i,opt} \times VCI_{i,t} + \left(1 - a_{i,opt}\right) \times TCI_{i,t}$$ (7)

Where $VHI_{i,t,opt}$ represents the VHI obtained by the best contribution value a of the ith pixel at time t, and $a_{i,opt}$ represents the best contribution value of the VCI to the VHI of the ith pixel. The algorithms used to improve the VHI index have been uploaded to Github and can be obtained from https://github.com/BNUJingyuZeng/A-new-global-VHI-dataset-code. A general overview of the working scheme is given in Fig. 2.

We compared the scatterplots and linear regression fittings of sc-PDSI and VHI before and after the improvement using detrending. The method of detrending is as follows:

$$VHIde_{i,t} = VHI_{i,t} - VHI_{i,t-1}$$ (8)

Where $\boldsymbol{VHIde_{i,t}}$ represents the VHI for the ith pixel at time t after detrending, $VHI_{i,t}$ represents the VHI for ith pixel at time t before detrending, and $VHI_{i,t-1}$ represents the VHI of pixel ith at time t-1 before detrending.

Evaluation of the drought detection efficiency. Drought detection efficiency of the VHI dataset was evaluated before and after the improvement, based on drought event records in the EM-DAT. The specific formula is as follows:

$$DTE = \frac{s}{TDE} \times 100$$ (9)

DTE is the drought detection efficiency, S is the score, and TDE is the total number of drought events. The score is evaluated according to the following principles. The VHI-based vegetation drought rating scale threshold is selected as 40. When the VHI is lower than 40, this means dryness (Table 1) - let's assume drought, and if it is higher than 40, this means normal or wet conditions (Table 1). Given the occurrence time and location
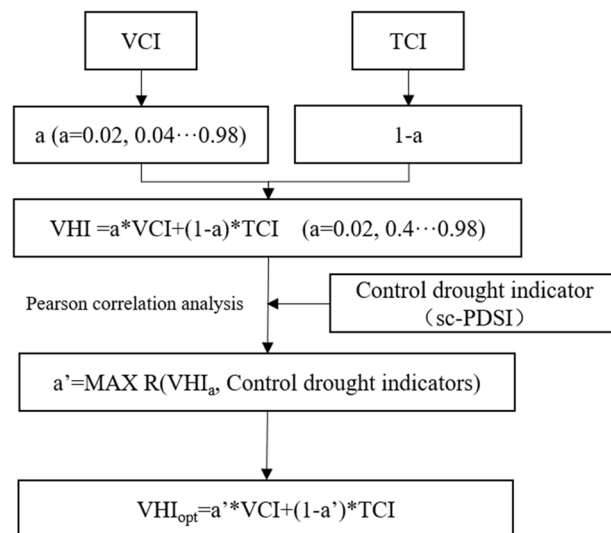
Fig. 2 Flowchart of improving VHI index based on sc-PDSI and Pearson correlation analysis.

of the drought event, VHIs before and after the improvement are compared one by one. When the number of pixels detected by drought accounts for more than 80% of the number of pixels in the area where the drought event occurred, the score is 1, when the number of pixels detected by drought accounts for more than 40% of the number of pixels in the area where the drought event occurred, the score is 0.5, when the number of pixels detected by drought accounts for less than 40% of the number of pixels in the area where the drought event occurred, the score is 0.

Statistical methods.   The Mann–Kendell (MK) method is widely used in meteorology, ecology, environmental research[51–54]. It is a nonparametric test method[55,56]. We used the Theil-sen trend analysis and the MK trend detection methods to study the temporal and spatial trends of global vegetation drought.

Data Records
Extent, projection, resolution and data format.   The VHI$_{opt}$ dataset covers most of the world's land areas except Antarctica, with an approximate range of $-50°$S to $70°$N and $-180°$W to $180°$E. The data projection is GCS_WGS_1984. This dataset is available as a 4-km GeoTIFF accessible from a data repository on figshare (https://doi.org/10.6084/m9.figshare.19811854.v5)[47]. For each year's VHI$_{opt}$ image, the globe is divided into 10000×3616 1-km $\times$ 1-km grid cells. In addition to annual data, we also provide monthly data, which can be used to analyze vegetation drought and related research from the seasonal scale analysis, and also provide a list to explain which months' data are missing due to the lack of data, and how to combine weekly data into monthly data. We also provide a global best contribution parameter dataset based on the 1981–2021 VHI dataset, i.e., the a_{opt} parameter file. The file's extent, projection, spatial resolution and data format are consistent with the VHI$_{opt}$ dataset. The annual VHI$_{opt}$ dataset does not exceed 12 GB (about 3 GB for the compressed download). To preserve as much information as possible, we kept the value of each cell to six decimal places. Missing data are represented by "*Nodata*".

Data naming and availability.   The global 1981–2021 VHI$_{opt}$ dataset is named "VHIopt_year.tif", where year corresponds to the year of the data, with a total of 41 files. The VHI$_{opt}$ data of the corresponding year is calculated from annual VCI and TCI data of the year with the best contribution parameter graph from the a_{opt} file. The parameter file of the global best contribution value a is directly named a_{opt}, which can be used to calculate VHI data from small to macro scales and from daily to seasonal scales.

Technical Validation
Drought detection efficiency of the VHI$_{opt}$ dataset.   The best contribution value a_{opt} is less than 0.5 in most regions of the world. The proportion of regions dominated by TCI reaching 70% and the proportion of regions dominated by VCI of ~28% (Fig. 3). VHI$_{opt}$ is affected by TCI and VCI differently in different regions. Most of Africa is dominated by VCI, while South America, Australia, and regions north of 30° latitude are dominated by TCI. This is especially visible in Europe, where TCI-dominated regions are concentrated and contiguous, with a_{opt} values generally lower than 0.3. Overall, an abnormally high surface temperature is the main driving factor affecting vegetation drought.

The coefficient of determination of the VHI before and after the improvement and the sc-PDSI were compared, as well as the performance before and after detrending (Fig. 4). The coefficient of determination of VHI$_{opt}$ and sc-PDSI (Fig. 4a,c) agree better compared with VHI$_{ori}$ (Fig. 4b,d). The correlation between VHI$_{opt}$ and sc-PDSI was 0.13 higher than that between VHI$_{ori}$ and sc-PDSI (i.e., 0.51 and 0.38, respectively) after excluding time-dependent trends (Fig. 4a,b), which implies an improvement of 34% in correlation. The correlation between VHI$_{opt}$ and sc-PDSI without detrending was 0.04 higher than that between VHI$_{ori}$ and sc-PDSI (i.e.,
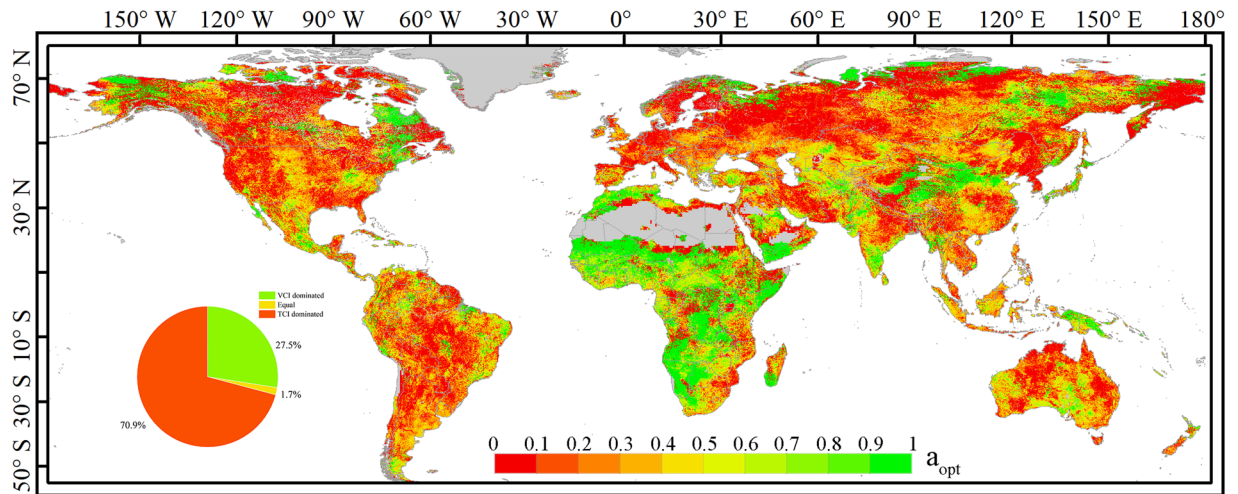
Fig. 3 Global distribution of the best contribution rate a_{opt} of global VCI to VHI$_{opt}$. The pie chart shows the proportion of regions dominated by TCI (red) or VCI (green).
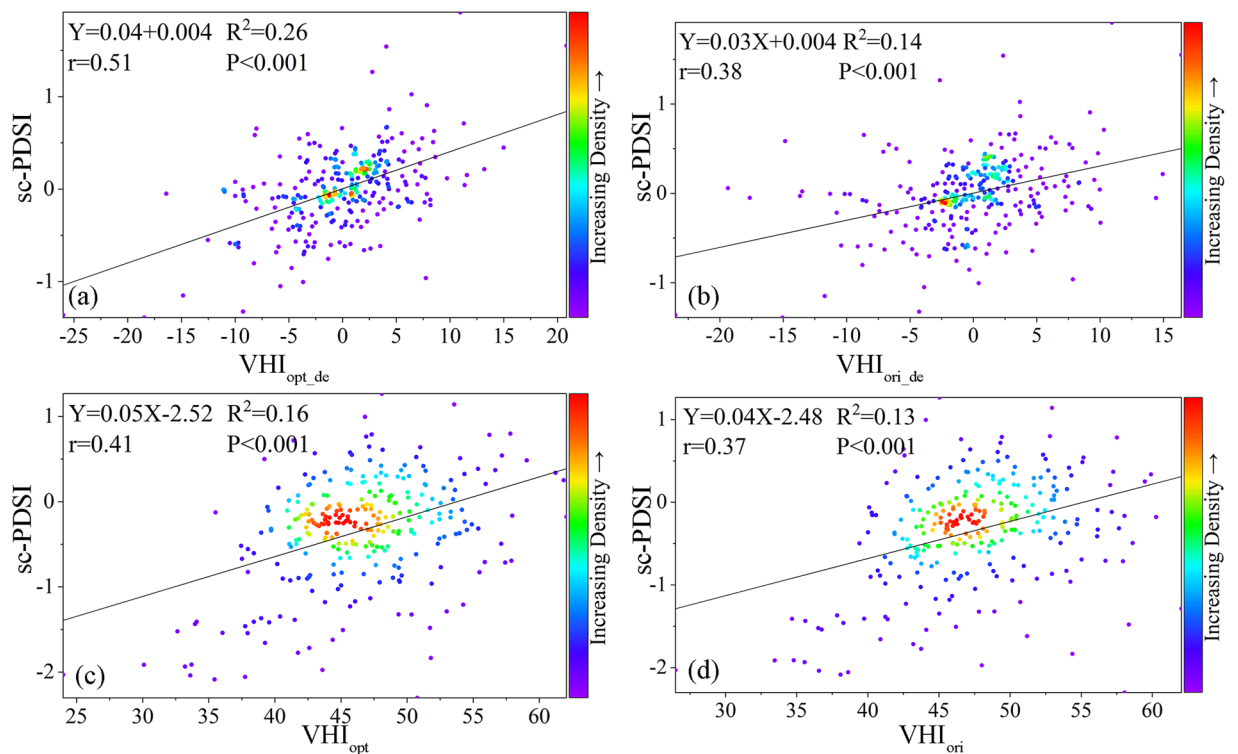

Fig. 4 Scatter plots of sc-PDSI as a function of (**a**) detrended VHI$_{opt}$, (**b**) detrended VHI$_{ori}$, (**c**) VHI$_{opt}$, and (**d**) VHI$_{ori}$. The black lines are the best-fit lines from linear regression. The relations, correlation coefficients, and p-values are given in each panel.

0.41 and 0.37, respectively; Fig. 4c,d), i.e., an improvement of 10%. The improved VHI dataset can better capture soil moisture anomalies and heat stress levels than the original VHI dataset, thus improving the detection of vegetation drought.

Based on global drought events recorded in the EM-DAT, we compared the ability of VHI$_{opt}$ and VHI$_{ori}$ to detect drought in a year-by-year globally (Fig. 5). The ability of VHI$_{opt}$ to detect drought was higher than that of VHI$_{ori}$, with drought detection efficiencies about 84.97% and 70.69%, respectively. The period from 1997 to 2002 was a period of frequent drought events in the world. The year 2001 experienced the largest number of drought events, i.e., 27. Drought events have declined in recent years.

The same analysis was repeated but by continent (Fig. 6). The drought detection efficiency of VHI$_{opt}$ was higher than that of VHI$_{ori}$ in all continents except Oceania. The drought event frequency from high to low was
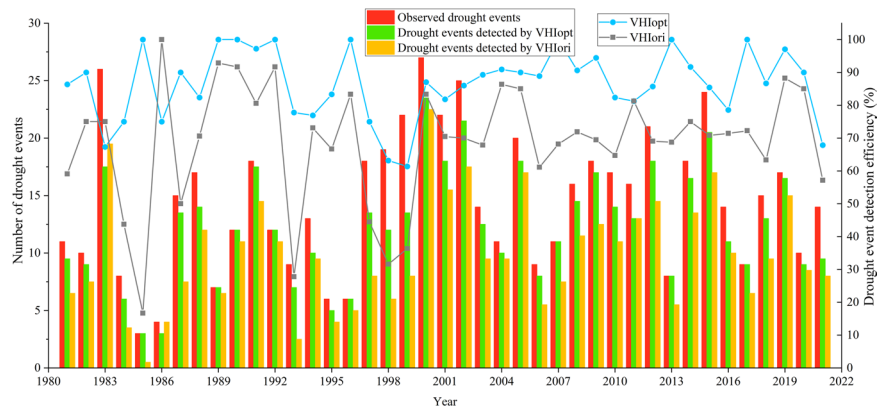
Fig. 5 Histogram of the number of observed (red), VHI$_{opt}$-detected (green), and VHI$_{ori}$-detected (orange) global drought events from 1981 to 2021. The blue and dark gray lines show the drought detection efficiencies (unit: %) of VHI$_{opt}$ and VHI$_{ori}$, respectively.
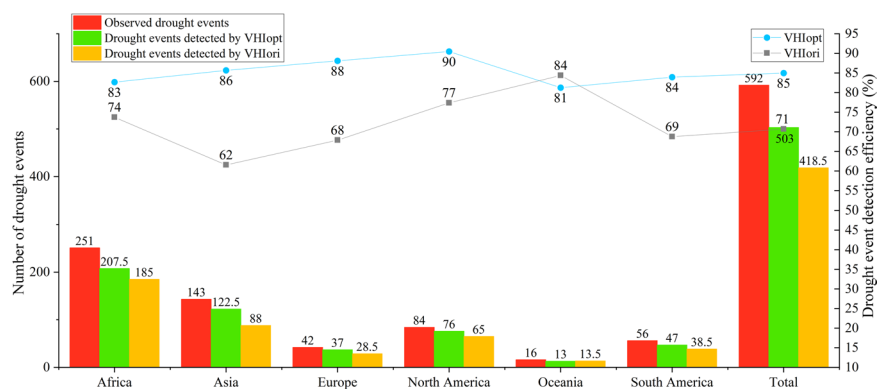

Fig. 6 Histogram of the number of observed (red), VHI$_{opt}$-detected (green), and VHI$_{ori}$-detected (orange) drought events from 1981 to 2021 in different continents. The blue and gray lines show the drought detection efficiencies (unit: %) of VHI$_{opt}$ and VHI$_{ori}$, respectively.

251, 143, 84, 56, 42, and 16 for Africa, Asia, North America, South America, Europe, and Oceania, respectively. The VHI$_{opt}$ drought detection efficiency was highest in North America at 90% and lowest in Oceania at 81%. The VHI$_{ori}$ drought detection efficiency is the highest in Oceania at 84% and lowest in Asia at 62%. The drought detection efficiencies of VHI$_{opt}$ and VHI$_{ori}$ in Oceania were not much different, related to the small number of drought events occurring there. In addition, in many areas of Oceania, drought occurs on small islands, which are affected by the resolution of the data. In Australia, the drought detection efficiency of VHI$_{opt}$ was still higher than that of VHI$_{ori}$.

In general, compared with the original VHI$_{ori}$, our improved VHI$_{opt}$ dataset has an improved drought detection efficiency in all continents. It has broad application prospects for the detection and long-term monitoring of vegetation drought.

Spatial differences between VHI$_{opt}$ and VHI$_{ori}$. Based on the estimated global best contribution value a_{opt}, we analyzed global annual VHI$_{opt}$ and VHI$_{ori}$ from 1981 to 2021 and compared their spatial patterns (Fig. 7). A relatively low degree of global vegetation drought is seen, with normal to good levels in most areas over the past 40 years, especially near the Equator, eastern North America, and southeastern South America. Clear differences between the two indices are seen (Fig. 7c). VHI$_{opt}$ effectively detected mild vegetation drought, while VHI$_{ori}$ underestimated the occurrence and impact range of mild vegetative drought. VHI$_{ori}$ was significantly higher than VHI$_{opt}$ along the western coasts of North America, Europe, South America, and southwest Asia.

We also compared VHI$_{opt}$ and VHI$_{ori}$ globally and by continents (Fig. 7b,d). VHI$_{ori}$ is higher than VHI$_{opt}$ in 66.5% of the cases. Oceania was the region with the lowest level of vegetation health in the world, as well as the region with the greatest variability, the highest uncertainty, and a higher risk of vegetation health. The vegetation health level in North America was the highest in the world, with average values of 49.02 and 47.98 for VHI$_{ori}$ and VHI$_{opt}$, respectively, which are significantly different from Oceania ($p < 0.05$).

We further zoomed in on the above regions to compare spatial differences between VHI$_{opt}$ and VHI$_{ori}$ (Fig. 8). In South America and Europe, VHI$_{opt}$ values below 40 indicates that this index effectively detected vegetation drought, especially in specific years. VHI$_{ori}$ values above 40 indicate normal conditions. VHI$_{opt}$ was more sensitive to mild vegetation drought than VHI$_{ori}$, illustrating its better ability to detect mild drought. VHI$_{opt}$ also
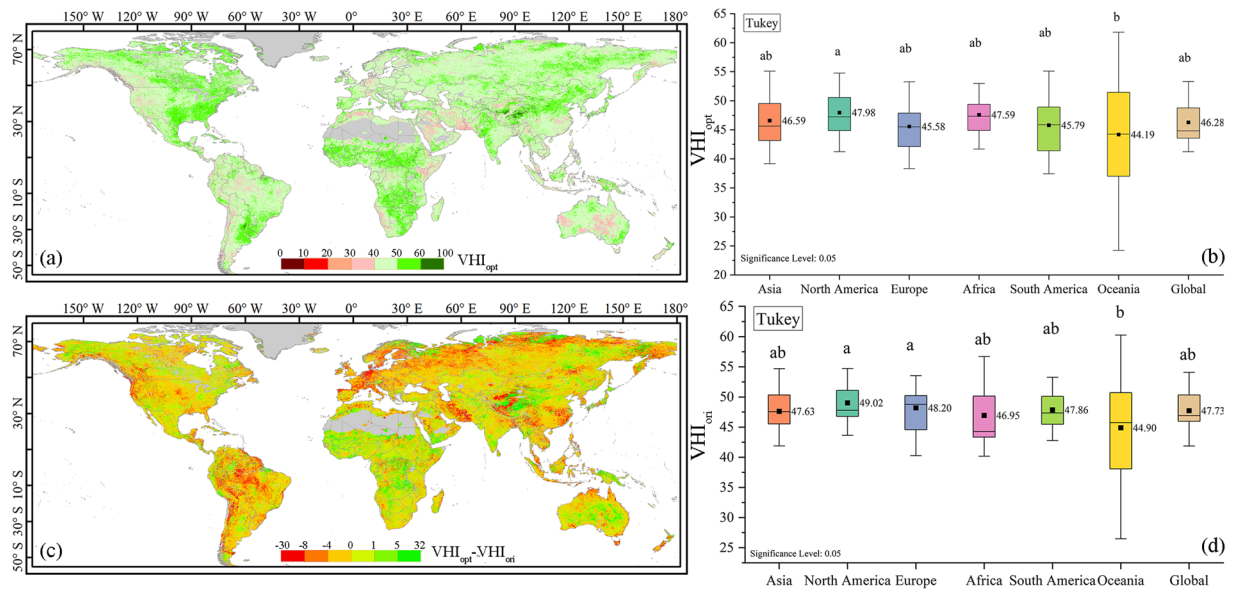
Fig. 7 Global distributions of (**a**) VHI$_{opt}$ and (**c**) differences between VHI$_{opt}$ and VHI$_{ori}$. Box plots of (**b**) VHI$_{opt}$ and (**d**) VHI$_{ori}$ globally and by continent. The letters above the boxes indicate significant differences at the p = 0.05 level.
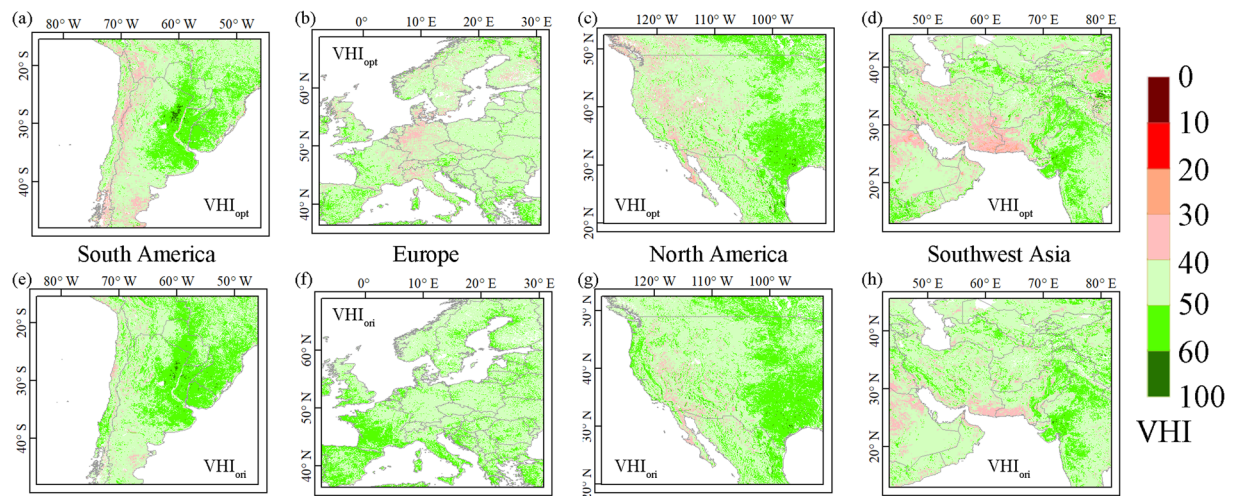

Fig. 8 Spatial distributions of (**a–d**) VHI$_{opt}$ and (**e–h**) VHI$_{ori}$ in South America, Europe, the American Midwest, and southwest Asia, respectively.

improved the ability to assess the occurrence range of vegetation drought compared with VHI$_{ori}$. For example, in the western United States and southwestern Asia, the spatial range of VHI$_{ori}$ values below 40 was smaller than that of VHI$_{opt}$, and there was an overly optimistic estimate of vegetation health. Furthermore, VHI$_{opt}$ also more effectively assessed the extent of vegetation drought in specific years. In general, VHI$_{opt}$ has clear advantages over VHI$_{ori}$ in both drought detection efficiency and drought occurrence range identification.

Trends seen in the global VHI$_{opt}$ dataset. We analyzed global and continental trends in VHI$_{opt}$ and VHI$_{ori}$ from 1981 to 2021 (Fig. 9). Results show that over the past 40 years, the VHI$_{opt}$ in Europe, South America, and Oceania has decreased significantly by about 0.14, 0.31, and 0.35 per year, respectively. There was no significant change in Asia, North America, and Africa. Globally, VHI$_{opt}$ showed an overall downward trend, with an average annual decrease of 0.16. VHI$_{ori}$ showed a significant increase in Europe, with an annual increase of about 0.2, and a significant decline in Africa, with a decrease of about 0.24 per year. A further analysis showed that South America was the region with the most obvious downward trend in VHI$_{opt}$ ($R^2 = 0.54$). From 1981 to 2005, inter-annual differences in Asia were relatively large, and from 2005 to 2021, changes in VHI$_{opt}$ and VHI$_{ori}$ gradually stabilized. This suggests that the future uncertainty of vegetation health in Asia may be less than in other regions.
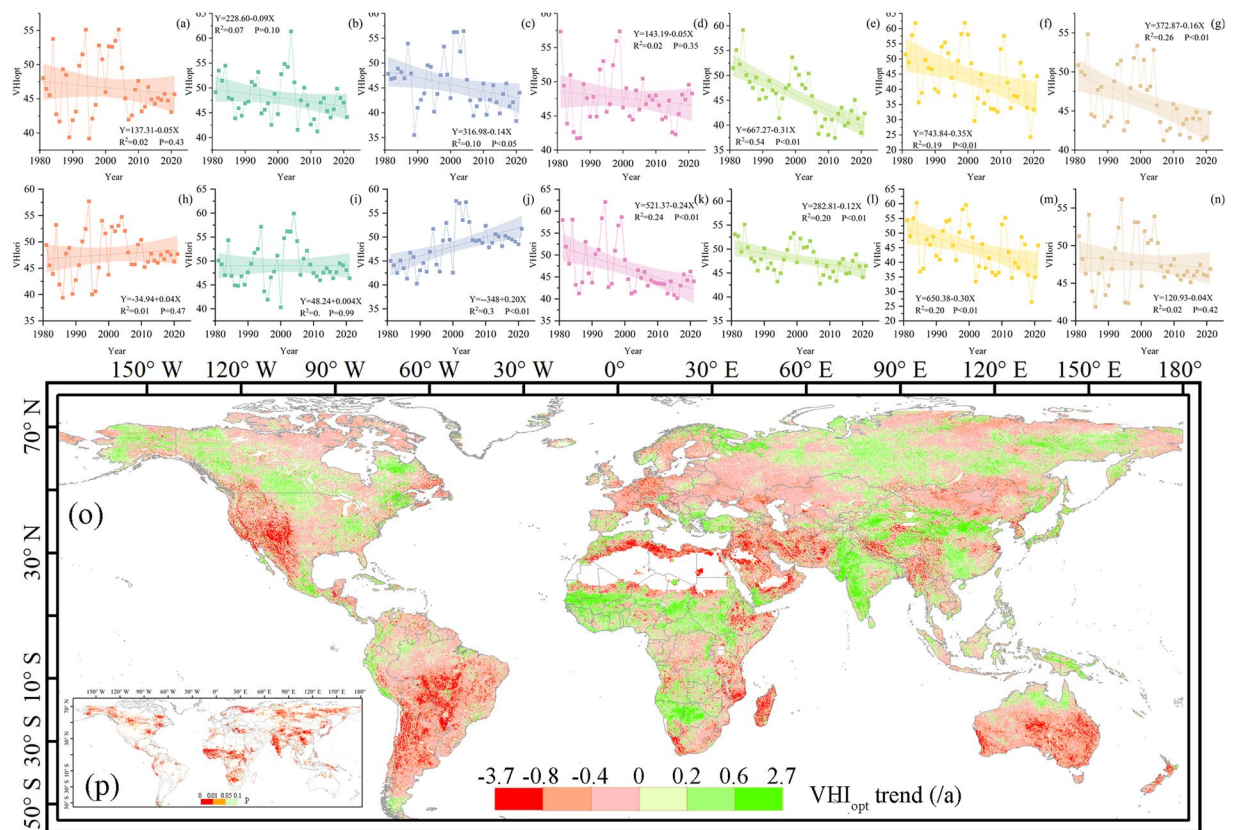
Fig. 9 Time series of (**a**–**g**) VHI_{opt} and (**h**–**n**) VHI_{ori} in Asia, North America, Europe, Africa, South America, Oceania, and globally, respectively. Global (**o**) trends in VHI_{opt} (unit: per year) and (**p**) p-values. The relations, coefficients of determination, and p-values are given in each (**a**–**n**) panel.

Since VHI_{opt} has noticeable advantages over VHI_{ori} in both drought detection efficiency and drought occurrence range identification, we used the VHI_{opt} dataset to evaluate the changing trend of global vegetation drought to identify hot spots where vegetation drought may be further aggravated in the future and areas with improved vegetation health (Fig. 9). Results show that the spatial heterogeneity of global vegetation drought changes may be high in the future. The western United States, South America, central and southern Australia, and southwestern Asia are regions where VHI_{opt} will decreases. On the other hand, central Africa, India, southern China, North America, and high-latitude parts of Asia are areas where vegetation droughts will be alleviated and vegetation health levels will improve. These changes have passed the significance assessment in the MK trend test. In the context of climate change, the changes in vegetation drought in areas where VHI_{opt} has declined are subject to large uncertainties and required more attention.

Usage Notes

We produced a long-term (1981–2021) high-spatial-resolution (4 km) improved vegetation drought detection dataset and the best contribution parameter dataset that can be used to calculate improved VHIs in different regions. We also evaluated the efficiency of the VHI_{opt} dataset to detect drought and the spatial pattern and changing trends of global vegetation drought to improve our techniques for long-term monitoring of vegetation drought. Based on the best global contribution parameter dataset, VHI_{opt} data in different regions can be calculated for performing drought assessment and studying food production and vegetation carbon sinks.

The data set can help people to carry out drought assessment more conveniently and efficiently, but there are areas where improvements are needed. First, the spatial resolution of the data may have an impacts on findings[57]. Using higher-spatial-resolution remote sensing products and sc-PDSI data will help further improve the detection efficiency of vegetation drought. Second, increased human activities may complicate changes in vegetation aridity[25]. Human interventions may have more complex impacts on vegetation health, requiring further research[58].

Our research was conducted on a global scale, providing a tool aimed at understanding global and regional vegetation drought characteristics. This dataset greatly improves the ability of VHI to detect vegetation drought, and can help people better understand the impact of temperature on vegetation drought in different regions. This research is also conducive to the effective implementation of vegetation drought resistance ecological engineering, and helps local governments and farmers reduce the losses caused by vegetation drought. The global vegetation drought dataset developed here also has the potential to be used to monitor and assess drought and its impact on the agricultural, forestry, ecological, and environmental sectors.

Code availability

All calculations of a global long term (1981–2021), high resolution (4 km) improved vegetation health index (VHI) dataset are based on MATLAB, and relevant step codes can be obtained from Github: https://github.com/BNUJingyuZeng/A-new-global-VHI-dataset-code.git.

References

1. Van Loon, A. F. Hydrological drought explained. *Wiley Interdisciplinary Reviews-Water* **2**, 359–392, https://doi.org/10.1002/wat2.1085 (2015).
2. Hameed, M., Ahmadalipour, A. & Moradkhani, H. Drought and food security in the middle east: An analytical framework. *Agricultural and Forest Meteorology* **281**, 107816, https://doi.org/10.1016/j.agrformet.2019.107816 (2020).
3. Kogan, F., Guo, W. & Yang, W. Z. Drought and food security prediction from NOAA new generation of operational satellites. *Geomatics Natural Hazards & Risk* **10**, 651–666, https://doi.org/10.1080/19475705.2018.1541257 (2019).
4. Dadson, S. J., Lopez, H. P., Peng, J. & Vora, S. in *Water Science, Policy, and Management* 11–28 (2019).
5. Marvel, K. *et al*. Twentieth-century hydroclimate changes consistent with human influence. *Nature* **569**, 59–65, https://doi.org/10.1038/s41586-019-1149-8 (2019).
6. Bento, V. A., Russo, A., Vieira, I. & Gouveia, C. M. Identification of forest vulnerability to droughts in the Iberian Peninsula. *Theoretical and Applied Climatology* **152**, 559–579, https://doi.org/10.1007/s00704-023-04427-y (2023).
7. Zhang, R. *et al*. The first high spatial resolution multi-scale daily SPI and SPEI raster dataset for drought monitoring and evaluating over China from 1979 to 2018. *Big Earth Data*, 1–26, https://doi.org/10.1080/20964471.2022.2148331 (2023).
8. AghaKouchak, A. *et al*. Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics* **53**, 452–480, https://doi.org/10.1002/2014rg000456 (2015).
9. AghaKouchak, A. A multivariate approach for persistence-based drought prediction: Application to the 2010–2011 East Africa drought. *Journal of Hydrology* **526**, 127–135, https://doi.org/10.1016/j.jhydrol.2014.09.063 (2015).
10. Wu, X. *et al*. The Effect of Drought on Vegetation Gross Primary Productivity under Different Vegetation Types across China from 2001 to 2020. *Remote Sensing* **14**, 4658 (2022).
11. Wu, Z. *et al*. Recent changes in the drought of China from 1960 to 2014. *International Journal of Climatology* **40**, 3281–3296, https://doi.org/10.1002/joc.6397 (2020).
12. Lesk, C., Rowhani, P. & Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **529**, 84–87, https://doi.org/10.1038/nature16467 (2016).
13. Wang, Q. *et al*. Temporal-spatial characteristics of severe drought events and their impact on agriculture on a global scale. *Quaternary International* **349**, 10–21, https://doi.org/10.1016/j.quaint.2014.06.021 (2014).
14. Zhang, Z. *et al*. Spatial pattern and decadal change of agro-meteorological disasters in the main wheat production area of China during 1991–2009. *Journal of Geographical Sciences* **24**, 387–396, https://doi.org/10.1007/s11442-014-1095-1 (2014).
15. He, B., Huang, L. & Wang, Q. Precipitation deficits increase high diurnal temperature range extremes. *Scientific Reports* **5**, 12004, https://doi.org/10.1038/srep12004 (2015).
16. He, B., Wang, H. L., Wang, Q. F. & Di, Z. H. A quantitative assessment of the relationship between precipitation deficits and air temperature variations. *Journal of Geophysical Research: Atmospheres* **120**, 5951–5961, https://doi.org/10.1002/2015JD023463 (2015).
17. Johansson, R., Luebehusen, E., Morris, B., Shannon, H. & Meyer, S. Monitoring the impacts of weather and climate extremes on global agricultural production. *Weather and Climate Extremes* **10**, 65–71, https://doi.org/10.1016/j.wace.2015.11.003 (2015).
18. Wang, Q. *et al*. A multi-scale daily SPEI dataset for drought characterization at observation stations over mainland China from 1961 to 2018. *Earth Syst. Sci. Data* **13**, 331–341, https://doi.org/10.5194/essd-13-331-2021 (2021).
19. Wang, Q. *et al*. An improved daily standardized precipitation index dataset for mainland China from 1961 to 2018. *Scientific Data* **9**, 124, https://doi.org/10.1038/s41597-022-01201-z (2022).
20. Zhang, R. *et al*. Investigating the effect of improved drought events extraction method on spatiotemporal characteristics of drought. *Theoretical and Applied Climatology* **147**, 395–408, https://doi.org/10.1007/s00704-021-03838-z (2022).
21. Wells, N., Goddard, S. & Hayes, M. J. A self-calibrating Palmer Drought Severity Index. *Journal of Climate* **17**, 2335–2351, 10.1175/1520-0442(2004)017<2335:aspdsi>2.0.co;2 (2004).
22. Palmer, W. C. in *U.S. Weather Bureau, Res. Pap. No. 45* 58–58 (1965).
23. Zeng, J. *et al*. Improving the drought monitoring capability of VHI at the global scale via ensemble indices for various vegetation types from 2001 to 2018. *Weather and Climate Extremes* **35**, 100412, https://doi.org/10.1016/j.wace.2022.100412 (2022).
24. Wang, Q. *et al*. The alleviating trend of drought in the Huang-Huai-Hai Plain of China based on the daily SPEI. *International Journal of Climatology* **35**, 3760–3769, https://doi.org/10.1002/joc.4244 (2015).
25. Zeng, J. *et al*. Drought frequency characteristics of China, 1981–2019, based on the vegetation health index. *Climate Research* **81**, 131–147, https://doi.org/10.3354/cr01616 (2020).
26. Wu, Z., Yu, L., Zhang, X., Du, Z. & Zhang, H. Satellite-based large-scale vegetation dynamics in ecological restoration programmes of Northern China. *International Journal of Remote Sensing* **40**, 2296–2312, https://doi.org/10.1080/01431161.2018.1519286 (2019).
27. Leng, S. *et al*. Assessing the Impact of Extreme Droughts on Dryland Vegetation by Multi-Satellite Solar-Induced Chlorophyll Fluorescence. *Remote Sensing* **14**, 1581 (2022).
28. Zhang, R., Qi, J., Leng, S. & Wang, Q. Long-Term Vegetation Phenology Changes and Responses to Preseason Temperature and Precipitation in Northern China. *Remote Sensing* **14**, 1396 (2022).
29. Leng, S. *et al*. Spatiotemporal Variations of Dryland Vegetation Phenology Revealed by Satellite-Observed Fluorescence and Greenness across the North Australian Tropical Transect. *Remote Sensing* **14**, 2985 (2022).
30. Tang, J. *et al*. A modified flexible spatiotemporal data fusion model. *Frontiers of Earth Science* **14**, 601–614, https://doi.org/10.1007/s11707-019-0800-x (2020).
31. Ezzine, H., Bouziane, A. & Ouazar, D. Seasonal comparisons of meteorological and agricultural drought indices in Morocco using open short time-series data. *International Journal of Applied Earth Observation and Geoinformation* **26**, 36–48, https://doi.org/10.1016/j.jag.2013.05.005 (2014).
32. Li, Y., Strapasson, A. & Rojas, O. Assessment of El Niño and La Niña impacts on China: Enhancing the Early Warning System on Food and Agriculture. *Weather and Climate Extremes* **27**, 100208, https://doi.org/10.1016/j.wace.2019.100208 (2020).
33. Bento, V. A., Gouveia, C. M., DaCamara, C. C. & Trigo, I. F. A climatological assessment of drought impact on vegetation health index. *Agricultural and Forest Meteorology* **259**, 286–295, https://doi.org/10.1016/j.agrformet.2018.05.014 (2018).
34. Kogan, F. N. in *Natural Hazards: Monitoring and Assessment Using Remote Sensing Technique* Vol. 15 *Advances in Space Research-Series* (eds Singh, R. P. & Furrer, R.) 91–100 (1995).

35. Kogan, F. N. Global drought watch from space. *Bulletin of the American Meteorological Society* **78**, 621–636, 10.1175/1520-0477(1997)078<0621:gdwfs>2.0.co;2 (1997).

36. Bhuiyan, C., Singh, R. P. & Kogan, F. N. Monitoring drought dynamics in the Aravalli region (India) using different indices based on ground and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation* **8**, 289–302, https://doi.org/10.1016/j.jag.2006.03.002 (2006).

37. Tang, J. *et al*. Self-adapting extraction of cropland phenological transitions of rotation agroecosystems using dynamically fused NDVI images. *International Journal of Biometeorology* **64**, 1273–1283, https://doi.org/10.1007/s00484-020-01904-1 (2020).

38. Yousef, F., Gebremichael, M., Ghebremichael, L. & Perine, J. Remote-sensing Based Assessment of Long-term Riparian Vegetation Health in Proximity to Agricultural Lands with Herbicide Use History. *Integrated Environmental Assessment and Management* **15**, 528–543, https://doi.org/10.1002/ieam.4144 (2019).

39. Bento, V. A., Gouveia, C. M., DaCamara, C. C., Libonati, R. & Trigo, I. F. The roles of NDVI and Land Surface Temperature when using the Vegetation Health Index over dry regions. *Global and Planetary Change* **190**, 103198, https://doi.org/10.1016/j.gloplacha.2020.103198 (2020).

40. Bi, W. *et al*. A global 0.05 degrees dataset for gross primary production of sunlit and shaded vegetation canopies from 1992 to 2020. *Scientific Data* **9**, https://doi.org/10.1038/s41597-022-01309-2 (2022).

41. Liu, X. F. *et al*. Compound droughts slow down the greening of the Earth. *Global Change Biology* **00**, 1–13, https://doi.org/10.1111/gcb.16657 (2023).

42. Vicente-Serrano, S. M. *et al*. Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 52–57, https://doi.org/10.1073/pnas.1207068110 (2013).

43. Zhu, X. R. *et al*. Soil coarsening alleviates precipitation constraint on vegetation growth in global drylands. *Environmental Research Letters* **17**, https://doi.org/10.1088/1748-9326/ac953f (2022).

44. Dunn, R. J. H. *et al*. Global Climate. *Bulletin of the American Meteorological Society* **103**, S11–S142, https://doi.org/10.1175/BAMS-D-22-0092.1 (2022).

45. van der Schrier, G., Barichivich, J., Briffa, K. R. & Jones, P. D. A scPDSI-based global data set of dry and wet spells for 1901–2009. *Journal of Geophysical Research: Atmospheres* **118**, 4025–4048, https://doi.org/10.1002/jgrd.50355 (2013).

46. EM-DAT. *CRED/UCLouvain*, <Brussels, Belgium> (www.emdat.be).

47. Zeng, J. *et al*. Global_1981-2021_4km_Improved_VHI_Index_and_Best_Contribution_Parameter. *figshare* https://doi.org/10.6084/m9.figshare.19811854.v5 (2022).

48. Karnieli, A. *et al*. Use of NDVI and Land Surface Temperature for Drought Assessment: Merits and Limitations. *Journal of Climate* **23**, 618–633, https://doi.org/10.1175/2009jcli2900.1 (2010).

49. Monteleone, B., Bonaccorso, B. & Martina, M. A joint probabilistic index for objective drought identification: the case study of Haiti. *Nat. Hazards Earth Syst. Sci.* **20**, 471–487, https://doi.org/10.5194/nhess-20-471-2020 (2020).

50. Kogan, F., Salazar, L. & Roytman, L. Forecasting crop production using satellite-based vegetation health indices in Kansas, USA. *International Journal of Remote Sensing* **33**, 2798–2814, https://doi.org/10.1080/01431161.2011.621464 (2012).

51. Zeng, J. *et al*. Ecological sustainability assessment of the carbon footprint in Fujian Province, southeast China. *Frontiers of Earth Science* **15**, 12–22, https://doi.org/10.1007/s11707-020-0815-3 (2021).

52. Wang, Q. *et al*. Pronounced Increases in Future Soil Erosion and Sediment Deposition as Influenced by Freeze–Thaw Cycles in the Upper Mississippi River Basin. *Environmental Science & Technology* **55**, 9905–9915, https://doi.org/10.1021/acs.est.1c02692 (2021).

53. Wang, Q., Tang, J., Zeng, J., Leng, S. & Shui, W. Regional detection of multiple change points and workable application for precipitation by maximum likelihood approach. *Arabian Journal of Geosciences* **12**, 745, https://doi.org/10.1007/s12517-019-4790-5 (2019).

54. Wang, Q. *et al*. The effects of air temperature and precipitation on the net primary productivity in China during the early 21st century. *Frontiers of Earth Science* **12**, 818–833, https://doi.org/10.1007/s11707-018-0697-9 (2018).

55. Mann, H. B. Nonparametric Tests Against Trend. *Econometrica* **13**, 245–259, https://doi.org/10.2307/1907187 (1945).

56. Kendall, M. G. Rank correlation methods. *British Journal of Psychology* **25**, 86–91 (1990).

57. Jay, S. *et al*. Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. *Remote Sensing of Environment* **231**, 110898, https://doi.org/10.1016/j.rse.2018.09.011 (2019).

58. Kou, P. L. *et al*. Complex anthropogenic interaction on vegetation greening in the Chinese Loess Plateau. *Science of the Total Environment* **778**, 146065, https://doi.org/10.1016/j.scitotenv.2021.146065 (2021).

Acknowledgements

Author contributions

J.Z. developed the data set and methodology, collected and analyzed the data, provided statistics and material (figures and tables) for the paper, and wrote the manuscript. T.Z., Y.Q., V.B., J.Q., Y.X., Y.L. and Q.W. discussed the results of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Q.W.

Reprints and permissions information is available at www.nature.com/reprints.

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