Vectorized dataset of check dams on the Chinese Loess Plateau using object-based classification method from Google Earth images

Yi Zeng[1, 2¶], Tongge Jing[1, 2¶], Baodong Xu[3], Xiankun Yang[4], Jinshi Jian[1, 2], Renjie Zong[1, 2], Bing Wang[1, 2], Wei Dai[1, 2], Lei Deng[1, 2], Nufang Fang[1, 2*], Zhihua Shi[3]

5   [1]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, 26 Xinong Road, Yangling, Shaanxi Province 712100, PR China
[2]Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, 26 Xinong Road, Yangling, Shaanxi Province 712100, PR China
[3]College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, PR China
10   [4]School of Geographical Sciences, Guangzhou University, Guangzhou, 510006, PR China

¶These authors contribute equally

Correspondence to: Nufang Fang (fnf@ms.iswc.ac.cn)

Abstract. The Chinese government has invested tens of billions of dollars and about 60 years to implement a large-scale
15   check dam project on the Chinese Loess Plateau (CLP) to control severe soil erosion. These check dams have trapped billions of tons of eroded sediment over the past few decades, significantly reducing the sediment load of the Yellow River, which was once the river with the largest sediment load in the world. However, there is still great uncertainty about how much sediment is trapped by check dams and what roles they play in the flow and sediment variability in the Yellow River, because the number and spatial distribution of check dams are still unclear. In this study, we produced the first vectorized
20   dataset of check dam on the CLP, combining high-resolution and easily accessible Google Earth images with object-based classification methods. We first investigated and analysed the key characteristics of check dams, and obtained the 0.3-1.0 m resolution Google Earth image of the best extraction period. Then we preliminarily obtained the rough check dam layer through multi-scale segmentation, threshold classification, and river network superposition. Finally, a self-developed human-computer interaction program combined with auxiliary data, visual interpretation, and expert knowledge is used to improve
25   the classification accuracy of check dams. The accuracy of the dataset is verified by 1947 collected test samples, and the producer's accuracy and user's accuracy of the check dam are 88.9% and 99.5%, respectively. Furthermore, at the provincial level, the area and number of check dams in our dataset are highly consistent with the latest official statistics of check dams, with $R^2 > 0.99$. Our study provides fundamental dataset for accurately assessing the ecosystem service functions of check dams, including sediment retention, carbon sequestration, grain supply, and will help to interpret current changes in sediment
30   delivery of the Yellow River and plan future soil and water conservation projects. The check dam dataset introduced in this article is freely available at https://doi.org/10.5281/zenodo.7857443 (Zeng et al., 2023).

1 Introduction

Accelerated soil erosion has caused worldwide land degradation, water pollution, and grain yield reduction, which has become one of the most pressing environmental issues threatening human sustainable development (Alewell et al., 2020;
35 Wuepper et al., 2020). To alleviate serious soil erosion, many countries have implemented a series of soil and water conservation projects according to local conditions in recent years, including vegetation restoration, conservation tillage (Klik and Rosner, 2020), terrace and check dam construction (Klik and Rosner, 2020; Liu et al., 2021b; Osman and Sauerborn, 2001). Among them, check dams have played an amazing role in soil and water conservation and are widely applied in global arid and semi-arid regions with serious soil erosion, such as China, Spain, Australia, America, India, Iran,
40 and Ethiopia (Abbasi et al., 2019; Lucas-Borja et al., 2021). The check dams built in the gullies can directly intercept the eroded sediment from the watershed (Zeng et al., 2023). Additionally, the silted land behind the check dam formed by the eroded sediment reduces the gully slope and runoff velocity, stabilizes the gully bed, and weakens erosion kinetic energy (Wang et al., 2021). The "sediment retaining" and "erosion reduction" of the check dam synergistically reduce the watershed sediment yield. According to data from different study areas such as China, Spain, and America, check dams can reduce
45 sediment yield by about 50-77% at the watershed scale (Boix-Fayos et al., 2008; Polyakov et al., 2014; Zhao et al., 2017). In addition to preventing soil erosion, check dams also provide more unexpected ecosystem services, including carbon sequestration and grain supply. Large amounts of organic-rich eroded sediments are buried and effectively preserved by check dams, becoming an important terrestrial carbon sink (Yao et al., 2022). Moreover, the fertile silted land behind the check dam, with its high water content and strong drought resistance, is often used as cropland with high and stable grain
50 production (Zeng et al., 2022b).

The Chinese Loess Plateau (CLP) is the region with the largest number and the densest distribution of check dams in the world due to serious soil erosion in the past few decades (Zeng et al., 2022a). Since the 1970s, the Chinese government and local farmers have widely built check dams in the gullies to prevent sediment loss to the Yellow River, which used to have the largest riverine sediment flux in the world. Tens of thousands of check dams cooperate with other soil and water
55 conservation measures such as vegetation restoration and terraces, have greatly reduced the sediment load into the Yellow River from the CLP by ~85% in recent decades (from ~1.6 Pg yr^{-1} in the period 1919-1960 to ~0.25 Pg yr^{-1} in the period 2010-2016) (Wang et al., 2016). However, there is still great uncertainty about how much sediment is intercepted by check dams and how many roles they play in the flow and sediment variation in the Yellow River because the number and spatial distribution of check dams are still unclear (Liu et al., 2021a). The Ministry of Water Resource of P.R. China reported that
60 more than 110,000 check dams had been constructed and captured about 21×10^9 m^3 sediment on the CLP (CMWR, 2003). Yet, the Bulletin of First National Census for Water reported only 58,446 check dams had been constructed and remained intact till 2013 and the silt sediment is about 7×10^9 m^3 (CMWR, 2013). Almost all regional studies regarding the sediment budget of the Yellow River referred to one of these two reports, and the huge discrepancy between the check dam data in these reports results in that all sediment retention and carbon storage estimated using these data may be overestimated or

65   underestimated (Ran et al., 2018; Wang et al., 2016; Wang et al., 2011). Additionally, although large check dams designed with capacities between 5×10^5 to 5×10^6 m^3 and funded by the central government had detail records, tens of thousands check dams constructed by local governments or farmers on the CLP remains unclear (Fang et al., 2019). Therefore, it is necessary to generate the spatial distribution data of check dams on the CLP, which is an important prerequisite to accurately quantify the contribution of check dams on the CLP to the variation of sediment load in the Yellow River and to evaluate the carbon

70   sequestration and grain supply benefits of check dams.

The early check dam data mainly come from the field survey organized by government departments, which may lead to incalculable statistical errors at the regional scale (Jin et al., 2012). In recent decades, remote sensing has been widely used to obtain regional, national, and even global land cover or land use, and has made great progress in cropland, terraces, water identification, and so on (Cao et al., 2021; Pazúr et al., 2022; Tortini et al., 2020). However, it is incredible that, although the

75   dataset of check dams is very important, only two studies have explored the possibility of obtaining check dam data based on remote sensing technology (Li et al., 2021; Tian et al., 2013). Zhao et al. (2013) obtained the spatial distribution of check dams in the Huangfuchuan watershed based on the supervised classification and Landsat images. Li et al. (2021) proposed a method integrating deep learning and object-based classification to extract check dams within a range of tens to hundreds of km^2. Nevertheless, these studies have only explored different methods to extract check dams on a very small scale, and could

80   not be extended to the whole CLP. That is, the current dataset of check dams on the CLP is still blank.

Several key difficulties limit the current extraction of check dams at the regional scale. Firstly, check dams on the CLP are mainly small and medium-sized check dams, with a corresponding silted area of 0.2-2 hm^2 (Bai et al., 2020). The decametric-resolution images (e.g., Landsat-7/8) may produce large silted land extraction errors due to the limited number of pixels and jagged edges (Ling et al., 2019). Secondly, the silted land formed by check dams is usually used for planting

85   crops (Li et al., 2019a), which is difficult to distinguish from the surrounding slope cropland and terrace in terms of spectral characteristics. Finally, silted land has the characteristics of large spatial heterogeneity and patch fragmentation, which is more difficult to identify than other land use types (Li et al., 2021). The traditional pixel-based extraction methods usually focus on medium resolution images, and seldom consider the structure and texture within the category and the correlation information between adjacent pixels (Duro et al., 2012; Hussain et al., 2013). Meanwhile, this method also produces salt-

90   and-pepper noise, which reduces the integrity of classification (Zhang et al., 2022). In contrast, the object-based classification method comprehensively considers a series of factors, such as spectral statistical features, shape, size, texture, and adjacency, and can obtain high-precision silted land extraction results combined with high-resolution images (Belgiu and Csillik, 2018). Therefore, combined with high-resolution and easily accessible Google Earth images and object-based classification, we provide the check dam dataset on the CLP for the first time. The self-developed computer program

95   combined with auxiliary data, visual interpretation and expert knowledge is used to improve the extraction accuracy of check dams. The accuracy of this dataset is verified by a visually interpreted test set and the latest official statistics. This dataset provides basic data for researchers to quantify the ecosystem service function of check dams, and to provide useful information for policy makers to plan soil and water conservation projects.

2 Method

100 It should be emphasized that the extraction of check dams on the CLP is not to extract the dam body, but to extract the silted land behind the check dam formed by the eroded sediment (hereinafter referred to as dam land). Because the sediment retention, carbon sequestration, and grain supply of check dams can only be determined if the dam land area is obtained. According to previous reports, the hilly and gully region occupies 33% (21.2 km^2) of the area of the CLP (64.0 km^2), but about 85% of the check dams are distributed here (Liu et al., 2021a). Therefore, to improve efficiency, we divide the CLP

105 into dam concentrated region (hilly and gully region) and dam sparse region (other region) according to the spatial distribution characteristics of check dams. For the dam concentrated region, we extracted dam land mainly through high-resolution Google Earth images and object-based classification method (Fig. 1). However, the area of dam sparse region is too large, and the acquisition and processing of corresponding Google Earth images are complex. In addition, only about 15% of the check dams are distributed in this region. Therefore, we extracted the dam land by artificial visual interpretation in

110 Google Earth. Noticeable, there are no check dams distributed in the Mu Us Desert in the northwest, the Guanzhong Plain in the middle, and the Rocky Mountains in the east of the dam sparse region (Fig. 2a) (Jin et al., 2012), so we masked these areas, which greatly reduced the workload of visual interpretation. Finally, we aggregate the dam land extraction results of these two regions and verify the accuracy.
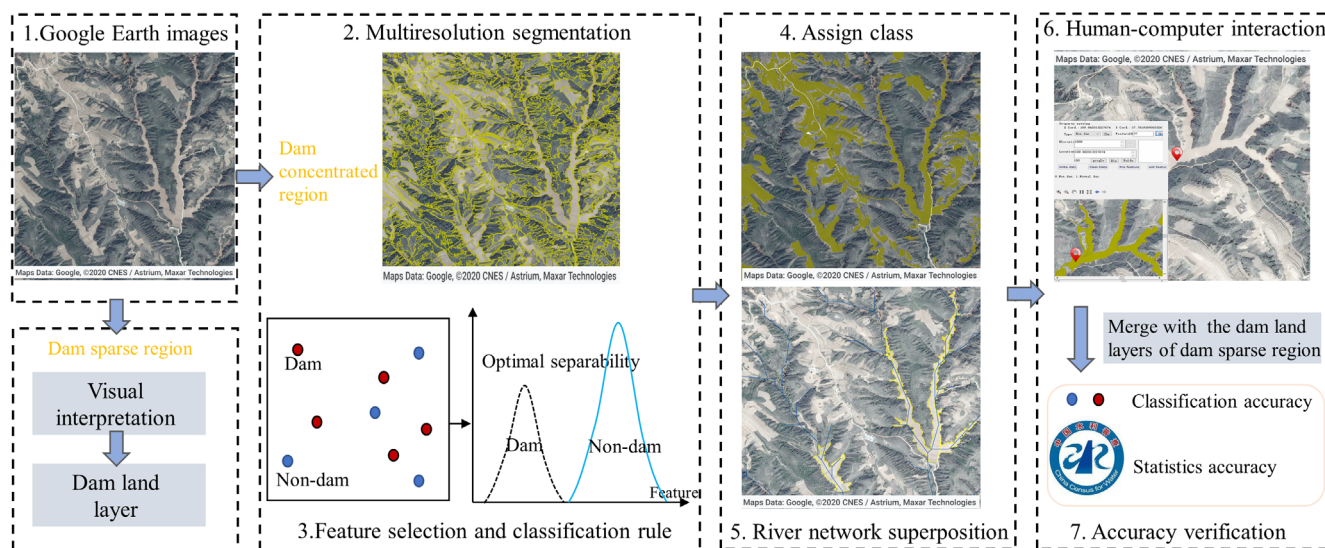
115 **Figure 1: The workflow of check dam land extraction on the Chinese Loess Plateau.**

2.1 Satellite images collection

Dam land is one of the most productive land types on the CLP, so the vast majority of dam land on the CLP is cultivated. Therefore, the extraction object of this study is the cultivated dam land on the CLP, excluding water-covered dams and abandoned dam land. Farmers on the CLP usually plough the dam land around May (Zhang et al., 2019), while the gullies

120   around the dam land are covered with vegetation, which makes it easy to distinguish the dam land from the surrounding landscape from satellite images. A large number of field investigations and Google Earth image observations have also confirmed this phenomenon (Fig. 2). Additionally, we randomly selected dam land (N_{dam}=127) and surrounding landscape samples (N_{dam}=130) in three counties with dense distribution of check dams (Baota county: N_{dam}=41 and N_{sur}=48; Zizhou county: N_{dam}=37 and N_{sur}=45; Lin county: N_{dam}=49 and N_{sur}=37). Then, combined with the Sentinel-2 image, the NDVI time

125   series of these sample points in 2019 are calculated in Google Earth Engine. We also found that there was the most significant difference in NDVI between the dam land and the surrounding landscape around May (Fig. 3). Therefore, we collected all available May images from 2016 to 2020 in the dam concentrated region in Google Earth, with a spatial resolution of 0.3-1.0 m. It is worth noting that the images in Google Earth are usually spliced from images taken in different periods or from different satellites, resulting in the possible chromatic aberration between the images of the two adjacent

130   scenes (Li et al., 2019b). To avoid subsequent segmentation errors caused by image chromatic aberration, we divided the images of the dam concentrated region according to the shooting date and obtained a total of 52 images.
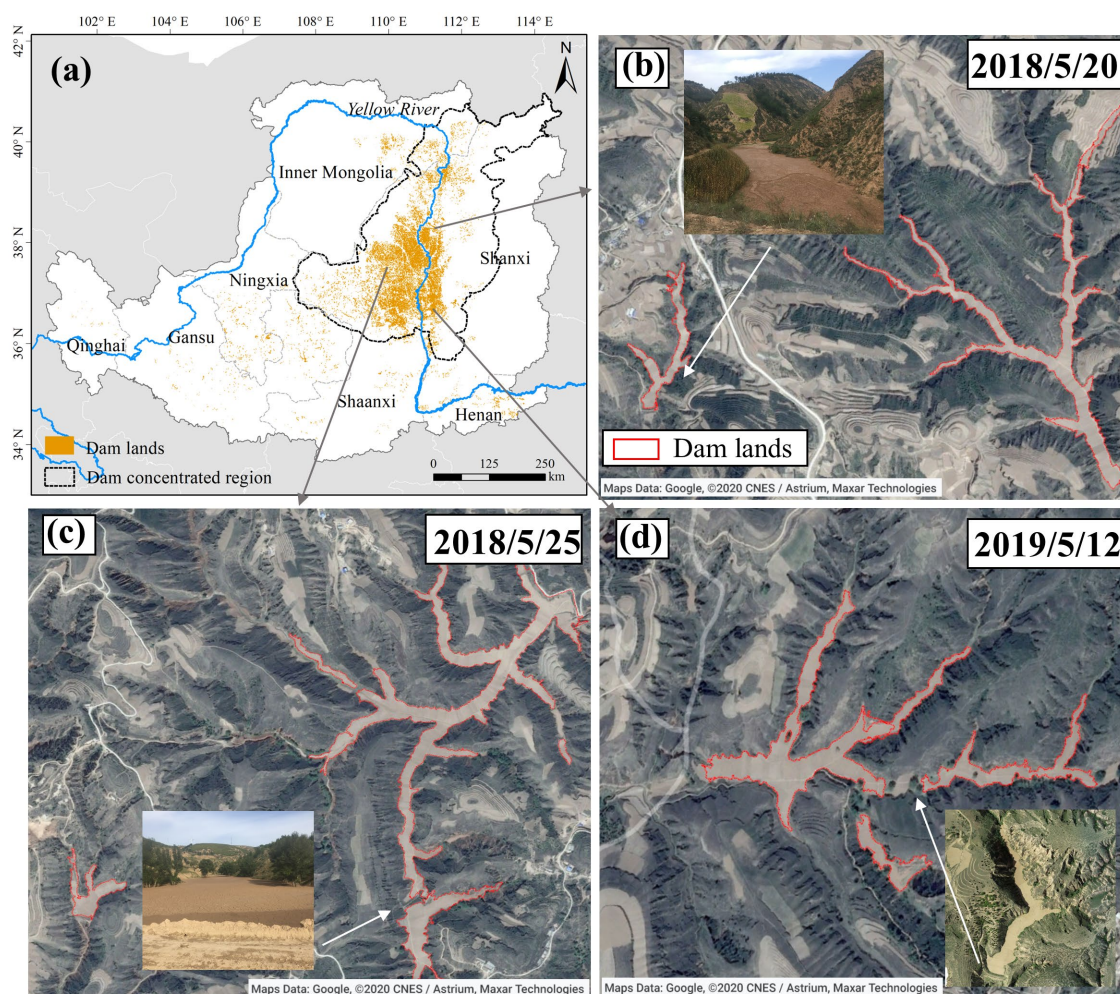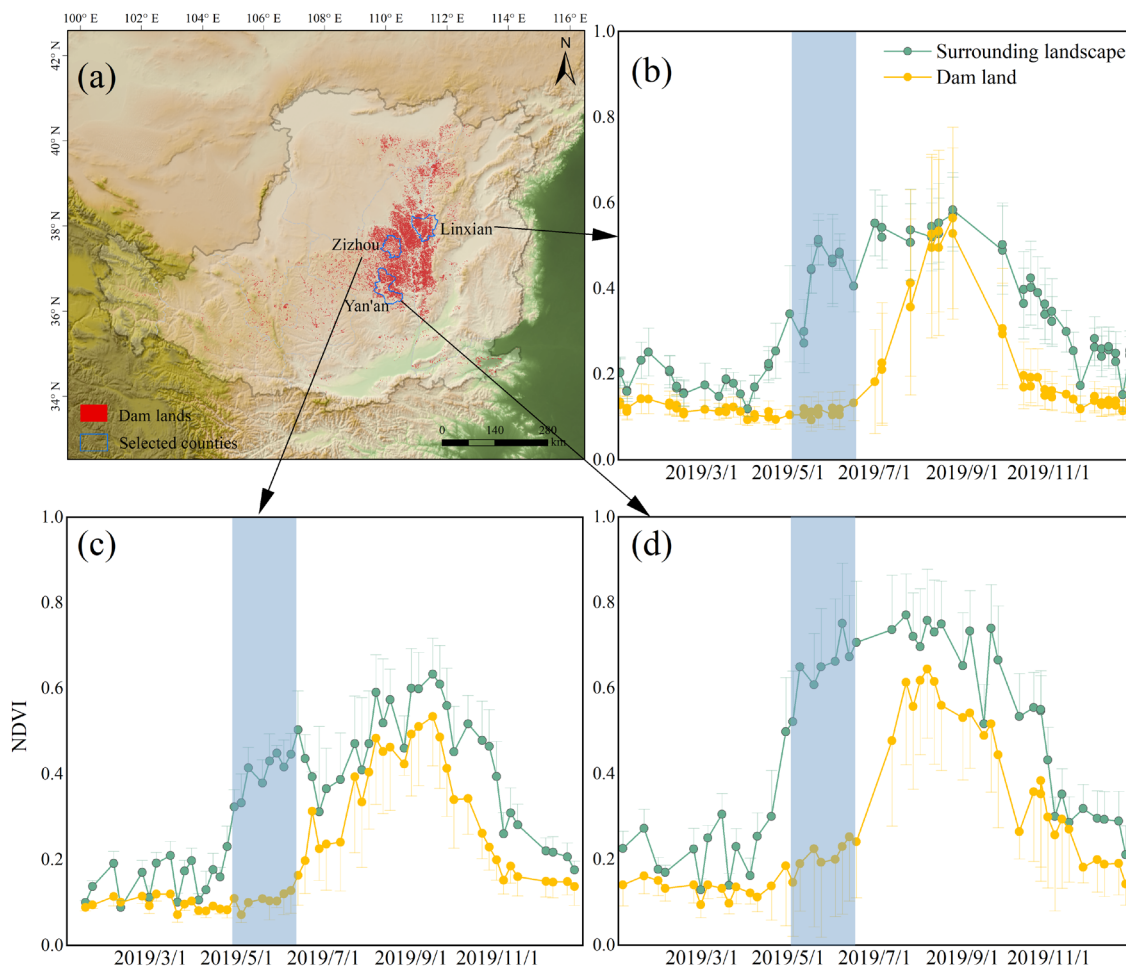
Figure 2: Check dams in the study area. (a) Dam concentration region on the Chinese Loess Plateau, (b-d) Google Earth images, photographic images, and unmanned aerial vehicle images of dam lands in May.

135

Figure 3: (a) Location of the three selected counties, (b-d) NDVI time series of the dam land and surrounding landscape of the three selected counties. The blue rectangle represents the period with the largest NDVI difference between the dam land and the surrounding landscape.

2.2 Object-Based Classification Method
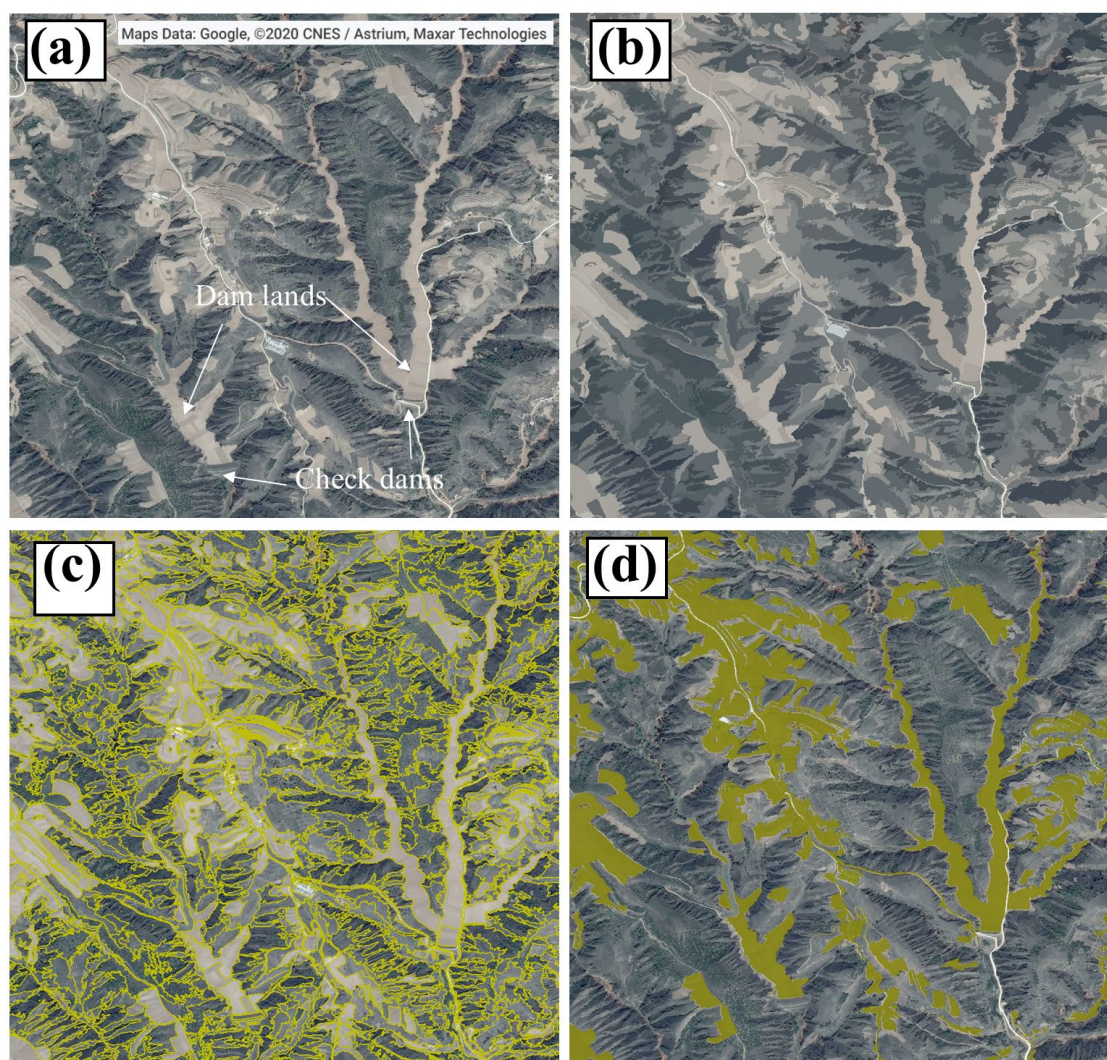
140 ### 2.2.1 Multiresolution segmentation

Here we used multiresolution segmentation algorithm in the eCognition Developer. This segmentation method achieves image segmentation by merging adjacent pixels with similar features under the premise of ensuring the maximum homogeneity between the pixels within the object (Gupta and Bhadauria, 2014). The scale, shape, and compactness parameters are the most important parameters in multiresolution segmentation algorithms, and their values will affect the

145 segmentation results (Munyati, 2018). The scale parameter is used to determine the maximum heterogeneity of the generated

Earth System
Science
Data

object and to control the size of the segmented object. The estimation of scale parameter 2 (ESP2) plugin in eCognition Developer can automatically evaluate the segmentation effect based on the local variance (LV) and its rate of change (ROC). Therefore, we used the ESP2 plugin to determine the optimal scale parameter for dam land extraction (Wei et al., 2021). The setting of the shape and compactness parameters is also crucial for the segmentation and subsequent classification of dam

150　land, as the dam land is usually narrow and irregular in shape. Therefore, we made different combinations of these parameters to test the best parameter settings in combination with visual inspection. Finally, we set the scale parameter, shape weight, and compactness weight to 100, 0.7, and 0.3 to obtain the best segmentation results (Fig. 4).

Figure 4: Multiresolution segmentation of dam land. (a) Original Google Earth image; (b) Object mean view in eCognition
155　**Developer; (c) View after multiresolution segmentation; (d) Classify the segmented objects. The yellow polygon is the extracted target category.**

2.2.2 Feature selection and classification rule

Here we used the assign class algorithm in the eCognition Developer, which determines the category of segmented objects by setting threshold conditions. When there were significant differences between the target and background categories for some features, the assign class algorithm could be used to construct classification rules (Tamta et al., 2015). The assign class algorithm was more suitable for our study than other classification methods, as our goal was to include more dam land in the target category. For each segmented image, we first randomly selected 30-80 dam land and non-dam land at different locations according to the size of the image. Then, we used the separability and thresholds algorithm in eCognition Developer to automatically select classification features (e.g., red, green, and blue band, shape, texture, and brightness) and determine the threshold of the selected feature (Guirado et al., 2019) (Fig. 1). Finally, we used the feature threshold to classify the segmented objects and manually adjusted the threshold range combined with visual interpretation to ensure that all dam land was included in the classification range. Through the above steps, we obtained the bare land layer of the dam concentrated region in May, mainly including the dam land in gullies and the cropland on slope land (Fig. 4).

2.3 River network superposition and human-computer interaction

All dam lands on the CLP are distributed in gullies, which can be easily separated from non-dam land on the slope through the river network. We first used SRTM-DEM with a resolution of 30 m to extract river networks in ArcGIS. Then we superimpose the river network and the bare land layer to extract the dam land in the gullies (Fig. 5a, b). To further improve the accuracy of dam land extraction, we developed a convenient human-computer interaction program to mark and eliminate incorrect classifications. The left window of the program can call the high-resolution Google Earth image, and the right side can call the dam land layer we extracted in the previous step (Fig. 5c). We assign values to each extracted vector polygon (e.g., dam land is 1, non-dam land is 2) based on auxiliary data, visual interpretation, and expert knowledge. Finally, we merge the vector polygons with a value of 1 obtained from 52 images in ArcGIS, which is the final dam land layer in the dam concentration region.
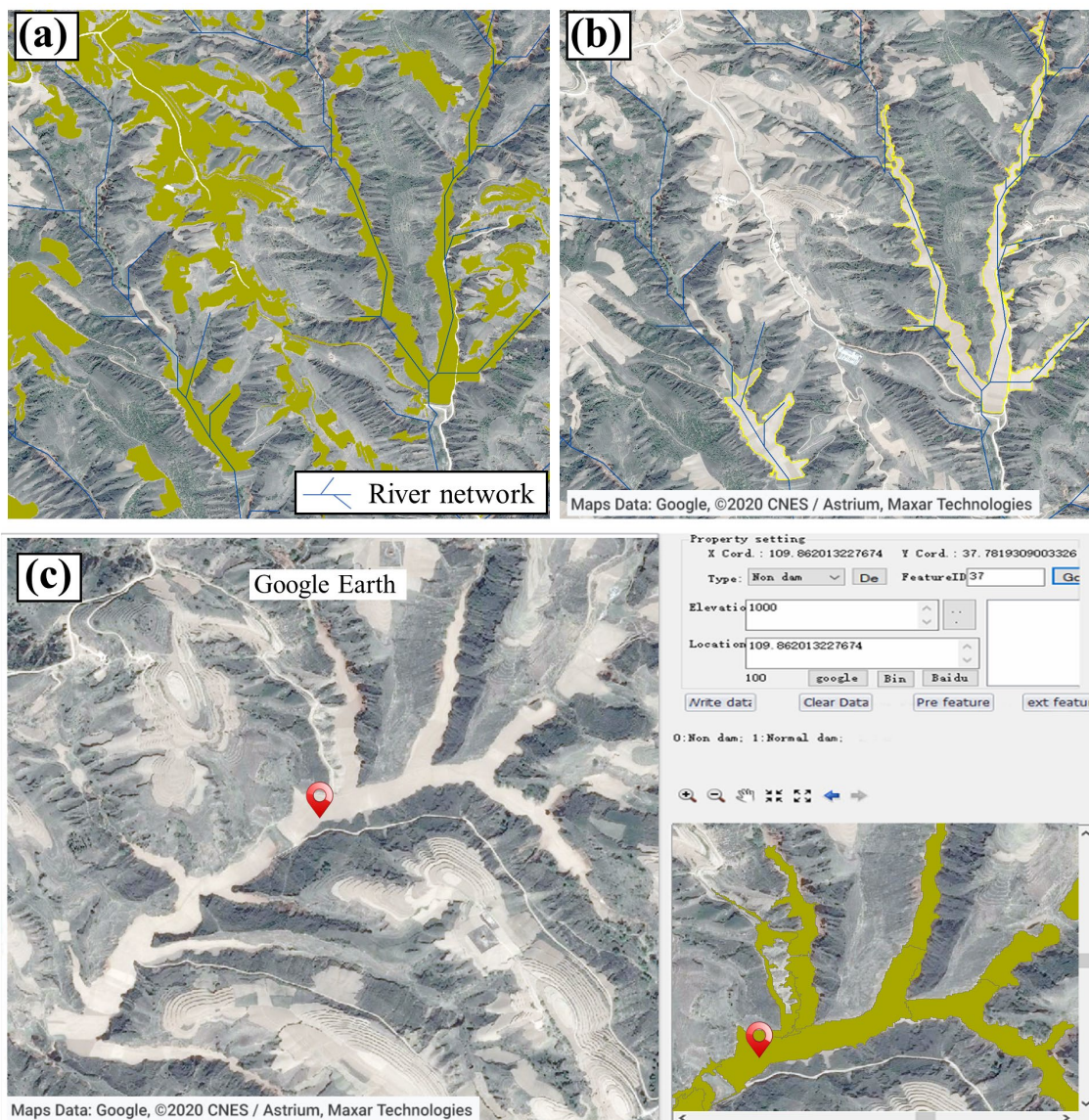
Figure 5: (a) Classified objects superimposed on river network, (b) removal of non-dam land by superimposed river network; (c) human-computer interaction.

2.4 Accuracy verification

At present, there is no available spatial distribution dataset of check dams on the CLP. Therefore, we take the test samples obtained from Google Earth by visual interpretation to verify our dataset. Due to the spatial distribution of check dams is not uniform, the traditional uniform sampling may lead to the deviation of accuracy evaluation. To improve the reliability of verification, we determine the number of test samples in each county according to the number of check dams in the dataset at the county level. That is, we allocated more test samples in counties where check dams are concentrated. A total of 1947 test

samples were acquired within the study area, of which 949 samples were interpreted as silted dam land and 998 samples as non-dam land. We evaluated the accuracy of the check dam dataset by calculating the confusion matrix, including producer's

190   (PA), user's accuracy (UA), and Kappa coefficient. Additionally, the Ministry of Water Resources of the People's Republic of China has provided the area and number of check dams in different provinces on the CLP according to statistical data of county-level water conservancy departments (CMWR, 2013). We compare our check dam dataset with this official statistical data for mutual verification.

2.5 Estimation of sediment volume of check dam

195   Our recent research has established a method for estimating the sediment volume of check dams based on unmanned aerial Vehicle (UAV) photogrammetry and proposed an area-volume empirical formula for estimating the volume of check dams at regional scale (Zeng et al., 2022a). Based on this area-volume empirical formula, we can accurately estimate the sediment volume of check dams according to the area of each dam land. A more detailed description and accuracy verification of this sediment volume estimation method for check dam can be found in Zeng et al. (2022).

$$V = \begin{cases} (6.31 \pm 0.03) A_{area}^{(1.33 \pm 0.01)} & (0 \leq A_{area} < 2) \\ (3.80 \pm 0.07) A_{area}^{(1.51 \pm 0.01)} & (2 \leq A_{area} < 5) \\ (8.48 \pm 0.89) A_{area}^{(0.96 \pm 0.02)} & (5 \leq A_{area}) \end{cases} \tag{1}$$

200

where V is the sediment volume of check dam ($\times 10^4$ m^3); Area is the silted land area (hm^2).

3 Results and discussion

3.1 Accuracy assessment

We first evaluated our check dam dataset with 1947 collected test samples. The results of the confusion matrix show that the

205   overall accuracy (OA) of the check dam map is 94.4%, in which the PA and UA of the check dam map are 88.9% and 99.5%, respectively. The non-dam land class had both UA and PA over 90%. Additionally, the kappa coefficient is 0.89, indicating that the dataset has high classification accuracy (Table 1).

Table 1. Accuracy evaluation of check dam maps with 2000 random test samples.

Class	Non-dam land	Dam land	PA (%)	UA (%)
Non-dam land	994	105	99.6	90.4
Dam land	4	844	88.9	99.5
	OA=94.4%		Kappa=0.89	

210   We further combined with official statistics of check dams to verify the overall accuracy of our dataset. At the provincial level, the dam land areas of our dataset are quite consistent with that in the Bulletin of First National Census for Water (CMWR, 2013), and the R^2 is 0.999 (Fig. 6a). Similarly, the number of check dams in our dataset shows very high correlation with the official statistics, with a slope of 1.1185 and R^2 of 0.995 (Fig. 6b). There results show that compared with the official statistics of check dams in 2003 (110,000 check dams), the official statistics of check dams in 2013 is more

215   reliable (58,446 check dams). Additionally, the high reliability of our check dam dataset is also confirmed through mutual verification.
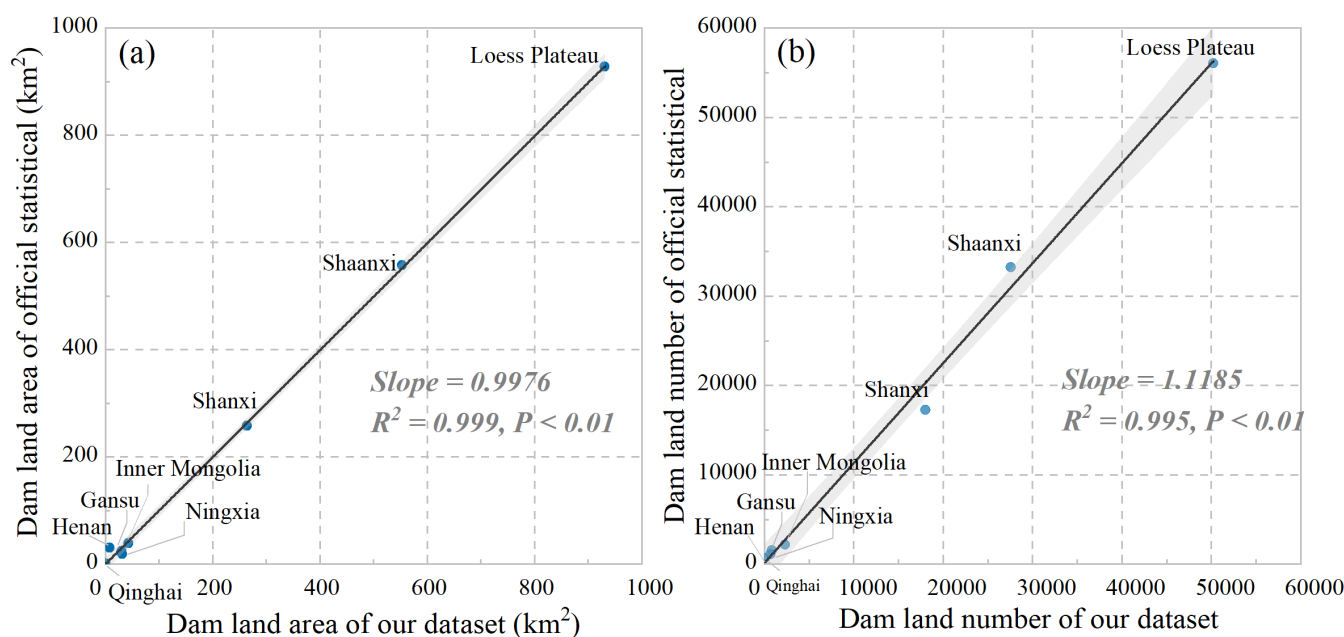

Figure 6: Comparison with the area (a) and number (b) of official statistics of check dams. The grey shade areas represent 95% confidence interval.

220

3.2 Spatial distribution, silted area, and sediment volume of check dams
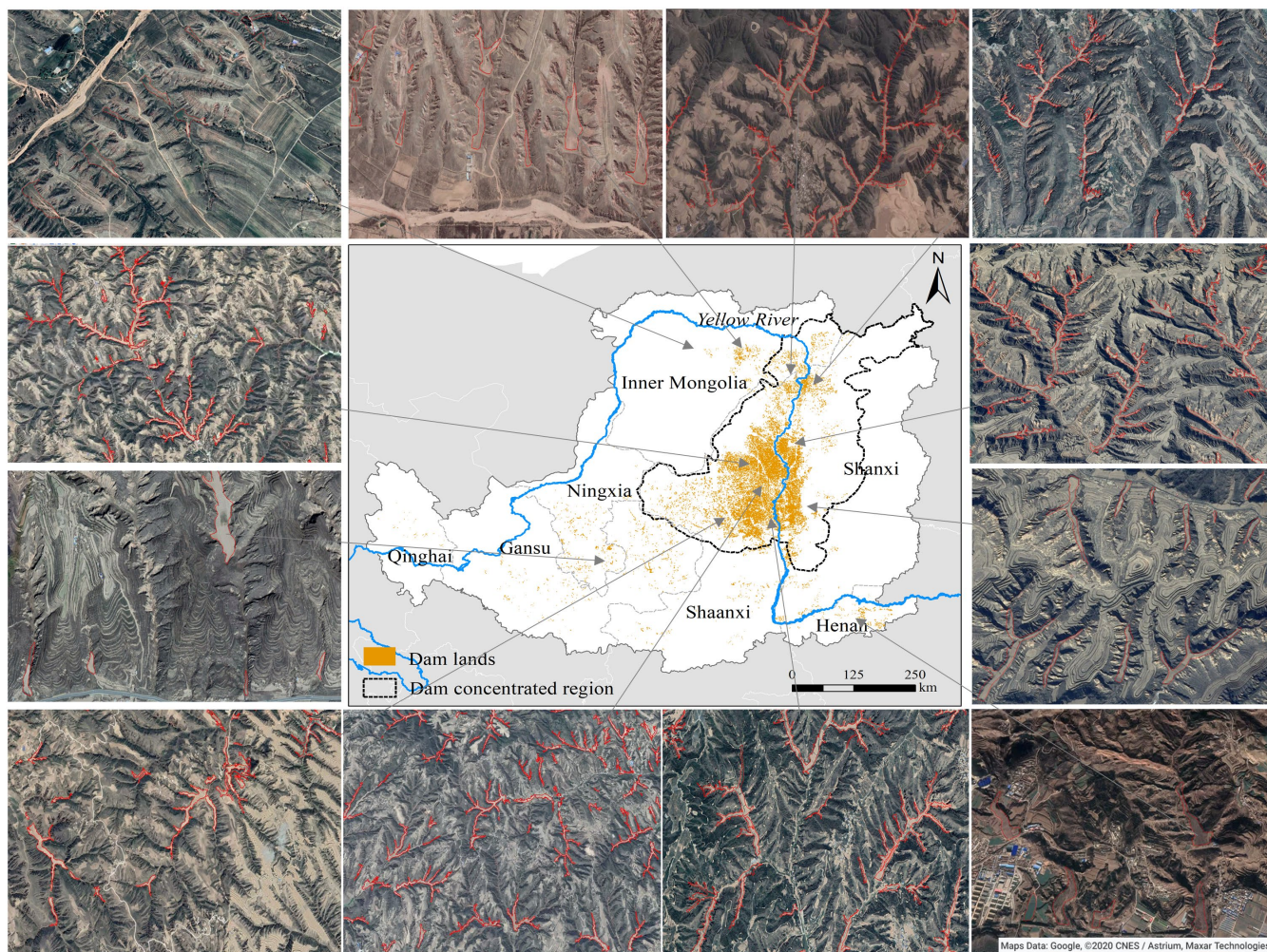

Figure 7: Spatial distribution of check dams on the Chinese Loess Plateau.

A total of 50,226 existing check dams were identified. The number of check dams in Shaanxi, Shanxi, Inner Mongolia,
Henan, Gansu, Ningxia, and Qinghai are 27,557, 18,000, 2,288, 833, 827, 653, and 68, accounting for 54.9%, 35.8%, 4.6%,
1.7%, 1.6%, 1.3%, and 0.1% of the number of check dams on the CLP, respectively (Table 2). Check dams are mainly
distributed in Northern Shaanxi and Western Shanxi (Fig. 7), that is, the loess hilly and gully region, which is one of the
most seriously eroded areas on the CLP and even in the world. The spatial distribution of check dams is closely related to
regional soil erosion, which is in line with the soil conservation function of check dams on the CLP. According to the dam
land area in the classification standard of check dams in the Regulation of techniques for comprehensive control of soil
erosion-Technique for erosion control of gullies (GBT16453.3-1996), the numbers of the micro (dam land area ranges from
0-0.2 hm^2), small (0.2-2 hm^2), medium (2-7 hm^2), and large check dams (>7 hm^2) are 4,207, 34,577, 9,192 and 2,250

12

respectively (Fig. 8). The small and medium check dams with the highest proportion account for about 69% and 18% of the total number of check dams, which is basically consistent with the results previously reported (70% and 18%) (CMWR, 2013).

235

Table 2. Number, silted area, and sediment volume of check dams in different regions on the Chinese Loess Plateau.

Region	Number	Area (km^2)	Volume ($\times 10^6$ m^3)
Shaanxi	27557 (54.9%)	553.6 (59.5%)	3946.5 (60.0%)
Shanxi	18000 (35.8%)	264.0 (28.4%)	1814.3 (27.6%)
Inner Mongolia	2288 (4.6%)	42.7 (4.6%)	308.1 (4.7%)
Henan	833 (1.7%)	8.5 (0.9%)	55.4 (0.8%)
Gansu	827 (1.6%)	30.2 (3.2%)	212.8 (3.2%)
Ningxia	653 (1.3%)	31.7 (3.4%)	234.3 (3.6%)
Qinghai	68 (0.1%)	0.5 (0.1%)	3.3 (0.1%)
Chinese Loess Plateau	50226	931.0	6574.7

The dam land area of all check dams on the CLP is 931 km^2, of which Shaanxi province (553.6 km^2) and Shanxi province (264.0 km^2) account for about 88% (Table 2). The existing 50,226 check dams intercept a total of 6.6 billion m^3 of sediment

240 (Table 2). If the bulk density of dam land is estimated as 1.5 g cm^{-3} previously reported (Fang et al., 2019), check dams intercept about 10.2 billion tons of sediment, which equals about 46% of the Yellow River's sediment load to the Bohai Sea during the period 1970-2020.
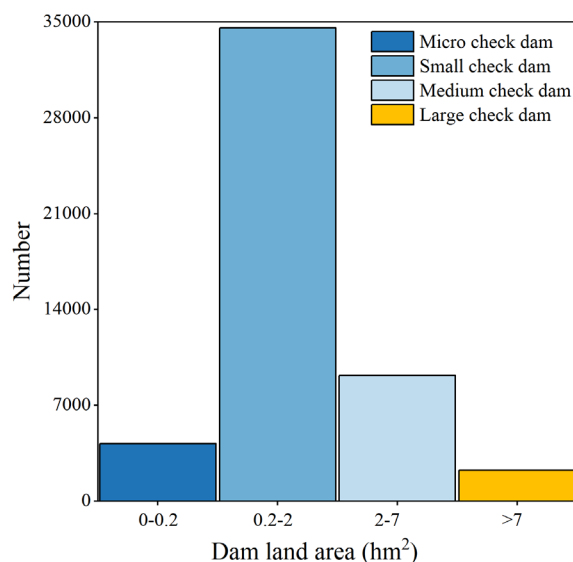
Figure 8: Frequency distribution of dam land area of check dams. The classification of check dams adopts the classification
245 **standard of check dams in the Regulation of techniques for comprehensive control of soil erosion-Technique for erosion control of gullies (GBT16453.3-1996).**

13

3.3 Broader implications

In recent years, check dams on the CLP have been paid more and more attention by researchers because of their important role in soil and water conservation. Using "check dam" and "Loess Plateau" as keywords, 8450 research articles were

250   searched on Google Scholar, of which 3540 were published after 2018. Most studies to date have relied on time-consuming and laborious field investigations to find check dams suitable for the research objectives (Liu et al., 2022; Zhang et al., 2020). Our dataset can provide the precise spatial location of check dams, which greatly saves the manpower and material resources of researchers. More importantly, our dataset provides key parameters such as silted land area and sediment volume of check dams, which are crucial for those studies using check dams to calculate soil erosion rate (Wang et al., 2020), estimate

255   watershed sediment delivery ratio (Zeng et al., 2023), and validate the soil erosion model (Zhao et al., 2017).

For policymakers, our database will fill the gap in the current check dam dataset and provide unprecedented detailed data on the spatial distribution, silted land area, and sediment volume of check dams on the CLP. When these key parameters of check dams are obtained, we can conveniently and accurately evaluate the ecosystem services functions of check dams, including sediment retention (Zhang and She, 2021), carbon sequestration (Ran et al., 2014), and grain supply (Shi et al.,

260   2020). We preliminarily estimated that the 50,226 active check dams on the CLP have intercepted about 10.2 billion tons eroded sediment during the period 1970-2020, which equals 46% of the sediment load of the Yellow River, once the largest sediment contributor to the global ocean. These data will be beneficial to understand the anthropogenic influences on the unprecedented changes of sediment reduction in the Yellow River (Wang et al., 2016). Additionally, recent review studies emphasize the important carbon sequestration potential of check dams, and point out that there are still great uncertainties

265   and challenges in the estimation of carbon storage of check dams on regional scale (Yao et al., 2022). When the sediment carbon content of check dams is obtained by combining large scale filed sampling or literature compilation, the carbon storage of check dams can be easily estimated by using our check dam dataset (Fang et al., 2023). Moreover, the check dams silt 93,150 hm^2 of farmland has better moisture conditions than terraced and slope cropland. Previous study shows the grain yield of silted dam land is 2~3 times that of terraced fields and 6~10 times that of slope cropland (Shi et al., 2020; Zeng et al.,

270   2022b). Our check dam dataset can be used to determine the grain supply of check dams, combined with the grain yield per unit area of the silted dam land. Finally, according to the Chinese government's plan, 56,161 new check dams will be built on the CLP by 2030 (NDRC, 2010). Our spatial distribution data of check dams provide important information for optimizing the location of the new check dams in the future.

3.4 Uncertainty, limitation, and future work

275   Combined with high-resolution and easily accessible Google Earth images and object-based classification strategy, we provide the check dam dataset on the CLP for the first time. However, there are still some limitations in the extraction of check dams on the regional scale. Firstly, although this method has high accuracy and good visual effect, the whole extraction process is still semi-automatic, and visual interpretation and expert experience are still needed in the steps of

280 classification feature selection and human-computer interaction. Secondly, there are a few water-covered check dams on the CLP, and the characteristics and functions of these water-covered check dams and small reservoirs are the same. It is therefore difficult to separate water-covered check dams from reservoirs, which may have led to a slight underestimation of the number of check dams in our dataset. Finally, due to the randomness and accessibility of historical Google Earth images, it is difficult to obtain April and May images for the entire CLP in the same year. Considering that the large-scale ecological restoration projects (e.g., Grain for Green project) on the CLP have significantly reduced soil erosion, the change of silted

285 land area of check dams in the short term is basically negligible. Therefore, we collected all available May images from 2016 to 2020 to cover more study areas. However, some newly constructed check dams may have been ignored (e.g., the acquired image is from 2016, but the check dam was constructed in 2018), which may also lead to a slight underestimation of the number of check dams. In future work, we can use our check dam dataset as the training set, combined with the high-resolution and long-time series satellites (e.g., Sentinel-2) and deep learning, to propose a more automated and convenient

290 check dam extraction process. Additionally, we will further evaluate the carbon storage and grain supply of check dams by combining a large amount of measured organic carbon content and grain yield data of dam land.

4 Data availability

The Digital elevation Model (DEM) of Chinese Loess Plateau was obtained from the Shuttle Radar Topography Mission (SRTM) with a resolution of 30 m (https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-

295 radar-topography-mission-srtm-1); The official statistics of the area and number of check dams in different provinces were obtained from the Yellow River Water Resources Commission of the Ministry of Water Resources (http://www.yrcc.gov.cn/sylm/2022stbcgb/). The generated check dam dataset in this article is freely available at https://doi.org/10.5281/zenodo.7619098 (Zeng et al., 2023). The vectorized dataset of check dams includes the following attributes: longitude, latitude, dam land area, dam land perimeter, sediment volume, and sediment mass.

5 Conclusion

In this study, the first vectorized dataset of check dam on the CLP containing spatial distribution, silted land area, and sediment volume of check dams was provided through object-based classification method combined with high-resolution and easily accessible Google Earth images. The self-developed computer program combined with auxiliary data, visual interpretation, and expert knowledge is used to improve the extraction accuracy of check dams, and finally 50,226 check

305 dams were extracted on the CLP. These check dams have formed 931 km^2 of dam land, intercepting about 10.2 billion tons of sediment, which equals about 46% of the Yellow River's sediment load to the Bohai Sea during the period 1970-2020. This dataset has good visual effect and high accuracy, of which the producer's accuracy and user's accuracy of the check dam are 88.9% and 99.5%, respectively. The area and number of check dams in our dataset are highly consistent with the

latest official statistics of check dams ($R^2 > 0.99$) to further verify the accuracy of this dataset. This dataset can be further
310 used to quantify the contribution of check dams on the CLP to the variation of sediment load in the Yellow River, and to assess other ecosystem services functions of check dams, such as carbon sequestration and grain supply. However, to ensure the accuracy of extraction, we involved more visual interpretation and expert experience in the extraction process, which affects the extraction efficiency of check dams. Future work should focus on proposing a more automated check dam extraction method, combining the evolving high-resolution and multispectral satellites and deep learning technologies.

Contributions

315

Y.Z. and N.F. conceived and designed the study. Y.Z. and T.J. wrote the paper. R.Z., B.W. and W.D. analyzed the data. B.X., J.J. and Z.S. contributed discussion. X.Y. provided assistance with the data analysis and software development. B.X., L.D., J.J. and Z.S. commented and edited the manuscript.

Competing interests

320 The authors declare that they have no conflict of interest.

Financial support

References

325 1. Abbasi, N. A., Xu, X., Lucas-Borja, M. E., Dang, W., and Liu, B.: The use of check dams in watershed management projects: Examples from around the world, Sci Total Environ, 676, 683-691, https://doi.org/10.1016/j.scitotenv.2019.04.249, 2019.
2. Alewell, C., Ringeval, B., Ballabio, C., Robinson, D. A., Panagos, P., and Borrelli, P.: Global phosphorus shortage will be aggravated by soil erosion, Nat Commun, 11, 4546, https://doi.org/10.1038/s41467-020-18326-7, 2020.
3. Bai, L. C., Wang, N., Jiao, J. Y., Chen, Y. X., Tang, B. Z., Wang, H. L., Chen, Y. L., Yan, X. Q., and Wang, Z. J.: Soil erosion and
330 sediment interception by check dams in a watershed for an extreme rainstorm on the Loess Plateau, China, International Journal Of Sediment Research, 35, 408-416, https://doi.org/10.1016/j.ijsrc.2020.03.005, 2020.
4. Belgiu, M. and Csillik, O.: Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, Remote Sensing Of Environment, 204, 509-523, https://doi.org/10.1016/j.rse.2017.10.005, 2018.
5. Boix-Fayos, C., de Vente, J., Martinez-Mena, M., Barbera, G. G., and Castillo, V.: The impact of land use change and check-dams
335 on catchment sediment yield, Hydrological Processes, 22, 4922-4935, https://doi.org/10.1002/hyp.7115, 2008.
6. Cao, B. W., Yu, L., Naipal, V., Ciais, P., Li, W., Zhao, Y. Y., Wei, W., Chen, D., Liu, Z., and Gong, P.: A 30 m terrace mapping in China using Landsat 8 imagery and digital elevation model based on the Google Earth Engine, Earth System Science Data, 13, 2437-2456, https://doi.org/10.5194/essd-13-2437-2021, 2021.
7. CMWR (Ministry of Natural Resources, People's Republic of China). Bulletin of First National Census for Water. Beijing, China,
340 2013.
8. CMWR (Ministry of Natural Resources, People's Republic of China). Programming for check dams in the Loess Plateau (Technical Report). Beijing, China, 2003

9. Duro, D. C., Franklin, S. E., and Dube, M. G.: A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery, Remote Sensing Of Environment, 118, 259-272, https://doi.org/10.1016/j.rse.2011.11.020, 2012.

10. Fang, N. F., Zeng, Y., Ni, L. S., and Shi, Z. H.: Estimation of sediment trapping behind check dams using high-density electrical resistivity tomography, Journal Of Hydrology, 568, 1007-1016, https://doi.org/10.1016/j.jhydrol.2018.11.062, 2019.

11. Fang, N. F., Zeng, Y., Ran, L. S., Wang, Z., Lu, X. X., Wang, Z. A., Yang, X. K., Jian, J. S., Yu, Q., Ni, L. S., Liu, C., Yue, C., and Shi, Z. H.: Substantial role of check dams in sediment trapping and carbon sequestration on the Chinese Loess Plateau, Communications Earth & Environment, 4, 65, ARTN 65, https://doi.org/10.1038/s43247-023-00728-2, 2023.

12. Guirado, E., Blanco-Sacristan, J., Rigol-Sanchez, J. P., Alcaraz-Segura, D., and Cabello, J.: A Multi-Temporal Object-Based Image Analysis to Detect Long-Lived Shrub Cover Changes in Drylands, Remote Sensing, 11, 2649, ARTN 2649, https://doi.org/10.3390/rs11222649, 2019.

13. Gupta, N. and Bhadauria, H.: Object based information extraction from high resolution satellite imagery using eCognition, International Journal of Computer Science Issues (IJCSI), 11, 139, 2014.

14. Hussain, M., Chen, D. M., Cheng, A., Wei, H., and Stanley, D.: Change detection from remotely sensed images: From pixel-based to object-based approaches, Isprs Journal Of Photogrammetry And Remote Sensing, 80, 91-106, https://doi.org/10.1016/j.isprsjprs.2013.03.006, 2013.

15. Jin, Z., Cui, B., Song, Y., Shi, W., Wang, K., Wang, Y., and Liang, J.: How many check dams do we need to build on the Loess Plateau?, Environ Sci Technol, 46, 8527-8528, https://doi.org/10.1021/es302835r, 2012.

16. Klik, A. and Rosner, J.: Long-term experience with conservation tillage practices in Austria: Impacts on soil erosion processes, Soil Till Res, 203, ARTN 104669, https://doi.org/10.1016/j.still.2020.104669, 2020.

17. Li, J. X., Liu, Q. L., Feng, X. M., Shi, W. Y., Fu, B. J., Lu, Y. H., and Liu, Y.: The synergistic effects of afforestation and the construction of check-dams on sediment trapping: Four decades of evolution on the Loess Plateau, China, Land Degradation & Development, 30, 622-635, https://doi.org/10.1002/ldr.3248, 2019a.

18. Li, S. J., Xiong, L. Y., Hu, G. H., Dang, W. Q., Tang, G. A., and Strobl, J.: Extracting check dam areas from high-resolution imagery based on the integration of object-based image analysis and deep learning, Land Degradation & Development, 32, 2303-2317, https://doi.org/10.1002/ldr.3908, 2021.

19. Li, X. H., Feng, R. T., Guan, X. B., Shen, H. F., and Zhang, L. P.: Remote Sensing Image Mosaicking Achievements and challenges, Ieee Geoscience And Remote Sensing Magazine, 7, 8-22, https://doi.org/10.1109/Mgrs.2019.2921780, 2019b.

20. Ling, F., Boyd, D., Ge, Y., Foody, G. M., Li, X. D., Wang, L. H., Zhang, Y. H., Shi, L. D., Shang, C., Li, X. Y., and Du, Y.: Measuring River Wetted Width From Remotely Sensed Imagery at the Subpixel Scale With a Deep Convolutional Neural Network, Water Resources Research, 55, 5631-5649, https://doi.org/10.1029/2018wr024136, 2019.

21. Liu, X., Gao, Y., Tian, Y., Li, X., and Ma, J.: SedimentIntercepted by Dams and the Sediment Production Situation Restoration of theLast 100 Years in the Yellow River Basin, Yellow River, 43, 19-23, 2021a.

22. Liu, X. Y., Xin, L. J., and Lu, Y. H.: National scale assessment of the soil erosion and conservation function of terraces in China, Ecological Indicators, 129, ARTN 107940, https://doi.org/10.1016/j.ecolind.2021.107940, 2021b.

23. Liu, Y., Walling, D. E., Yang, M., Zhang, F., Zhang, J., Xie, Y., and Zhao, T.: Using check dam deposits to reconstruct changes in the sediment response of a small catchment on the Loess Plateau in China, during the period 1960–1990, Hydrological Processes, 36, e14754, https://doi.org/10.1002/hyp.14754, 2022.

24. Lucas-Borja, M. E., Piton, G., Yu, Y., Castillo, C., and Zema, D. A.: Check dams worldwide: Objectives, functions, effectiveness and undesired effects, Catena, 204, ARTN 105390, https://doi.org/10.1016/j.catena.2021.105390, 2021.

25. Munyati, C.: Optimising multiresolution segmentation: delineating savannah vegetation boundaries in the Kruger National Park, South Africa, using Sentinel 2 MSI imagery, International Journal Of Remote Sensing, 39, 5997-6019, https://doi.org/10.1080/01431161.2018.1508922, 2018.

26. NDRC (National Development and Reform Commission People's Republic of China). Outline of the comprehensive management plan for the Loess Plateau (2010-2030). Beijing, China, 2010.

27. Osman, M. and Sauerborn, P.: Soil and water conservation in Ethiopia, Journal of soils and sediments, 1, 117-123, https://doi.org/10.1007/bf02987717, 2001.

28. Pazúr, R., Huber, N., Weber, D., Ginzler, C., and Price, B.: A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data, Earth System Science Data, 14, 295-305, https://doi.org/10.5194/essd-14-295-2022, 2022.

29. Polyakov, V. O., Nichols, M. H., McClaran, M. P., and Nearing, M. A.: Effect of check dams on runoff, sediment yield, and retention on small semiarid watersheds, Journal Of Soil And Water Conservation, 69, 414-421, https://doi.org/10.2489/jswc.69.5.414, 2014.

30. Ran, L., Lu, X. X., and Xin, Z.: Erosion-induced massive organic carbon burial and carbon emission in the Yellow River basin, China, Biogeosciences, 11, 945-959, https://doi.org/10.5194/bg-11-945-2014, 2014.

31. Ran, L., Lu, X., Fang, N., and Yang, X.: Effective soil erosion control represents a significant net carbon sequestration, Sci Rep, 8, 12018, https://doi.org/10.1038/s41598-018-30497-4, 2018.

32. Shi, P., Feng, Z., Gao, H., Li, P., Zhang, X., Zhu, T., Li, Z., Xu, G., Ren, Z., and Xiao, L.: Has "Grain for Green" threaten food security on the Loess Plateau of China?, Ecosystem Health and Sustainability, 6, 1709560, https://doi.org/10.1080/20964129.2019.1709560, 2020.

33. Tamta, K., Bhadauria, H., and Bhadauria, A.: Object-oriented approach of information extraction from high resolution satellite imagery, IOSR Journal of Computer Engineering (IOSR-JCE), 17, 47-52, https://doi.org/10.1109/CONFLUENCE.2014.6949274, 2015.

34. Tian, P., Zhao, G. J., Mu, X. M., Wang, F., Gao, P., and Mi, Z. J.: Check dam identification using multisource data and their effects on streamflow and sediment load in a Chinese Loess Plateau catchment, Journal Of Applied Remote Sensing, 7, Artn 073697, https://doi.org/10.1117/1.Jrs.7.073697, 2013.

35. Tortini, R., Noujdina, N., Yeo, S., Ricko, M., Birkett, C. M., Khandelwal, A., Kumar, V., Marlier, M. E., and Lettenmaier, D. P.: Satellite-based remote sensing data set of global surface water storage change from 1992 to 2018, Earth System Science Data, 12, 1141-1151, https://doi.org/10.5194/essd-12-1141-2020, 2020.

36. Wang, N., Jiao, J. Y., Bai, L. C., Zhang, Y. F., Chen, Y. X., Tang, B. Z., Liang, Y., Zhao, C. J., and Wang, H. L.: Magnitude of soil erosion in small catchments with different land use patterns under an extreme rainstorm event over the Northern Loess Plateau, China, Catena, 195, 104780, ARTN 104780, https://doi.org/10.1016/j.catena.2020.104780, 2020.

37. Wang, S. A., Fu, B. J., Piao, S. L., Lu, Y. H., Ciais, P., Feng, X. M., and Wang, Y. F.: Reduced sediment transport in the Yellow River due to anthropogenic changes, Nature Geoscience, 9, 38-41, https://doi.org/10.1038/Ngeo2602, 2016.

38. Wang, Y., Fu, B., Chen, L., Lu, Y., and Gao, Y.: Check dam in the Loess Plateau of China: engineering for environmental services and food security, Environ Sci Technol, 45, 10298-10299, https://doi.org/10.1021/es2038992, 2011.

39. Wang, Z. Y., Chen, Z. Y., Yu, S., Zhang, Q., Wang, Y., and Hao, J. W.: Erosion-control mechanism of sediment check dams on the Loess Plateau, International Journal Of Sediment Research, 36, 668-677, https://doi.org/10.1016/j.ijsrc.2021.02.002, 2021.

40. Wei, H. D., Hu, Q., Cai, Z. W., Yang, J. Y., Song, Q., Yin, G. F., and Xu, B. D.: An Object- and Topology-Based Analysis (OTBA) Method for Mapping Rice-Crayfish Fields in South China, Remote Sensing, 13, 4666, ARTN 4666, https://doi.org/10.3390/rs13224666, 2021.

41. Wuepper, D., Borrelli, P., and Finger, R.: Countries and the global rate of soil erosion, Nature Sustainability, 3, 51-55, https://doi.org/10.1038/s41893-019-0438-4, 2020.

42. Yao, Y. F., Song, J. X., and Wei, X. R.: The fate of carbon in check dam sediments, Earth-Science Reviews, 224, ARTN 103889, https://doi.org/10.1016/j.earscirev.2021.103889, 2022.

43. Zeng, Y., Meng, X. D., Zhang, Y., Dai, W., Fang, N. F., and Shi, Z. H.: Estimation of the volume of sediment deposited behind check dams based on UAV remote sensing, Journal Of Hydrology, 612, ARTN 128143, https://doi.org/10.1016/j.jhydrol.2022.128143, 2022a.

44. Zeng, Y., Meng, X. D., Wang, B., Li, M. J., Chen, D., Ran, L. S., Fang, N. F., Ni, L. S., and Shi, Z. H.: Effects of soil and water conservation measures on sediment delivery processes in a hilly and gully watershed, Journal Of Hydrology, 616, 128804, ARTN 128804, https://doi.org/10.1016/j.jhydrol.2022.128804, 2023.

45. Zeng, Y., Ran, L., Fang, N., Wang, Z., Xu, Z., Lu, X., Yu, Q., Wang, L., Yu, S., and Shi, Z.: How to balance green and grain in marginal mountainous areas?, Earth's Future, e2021EF002552, https://doi.org/10.1029/2021EF002552, 2022b.

46. Zeng Y., Jing, T., Xu, B., Yang, X., Jian, J., Zong, R., Wang, B., Dai, W., Deng, L., Fang, N., and Shi, Z.: Vectorized dataset of check dams on the Chinese Loess Plateau using object-based classification method from Google Earth images, Zenodo [data set], https://doi.org/10.5281/zenodo.7857443.

47. Zhang, J. Q., Yang, M. Y., Deng, X. X., Liu, Z., and Zhang, F. B.: The effects of tillage on sheet erosion on sloping fields in the wind-water erosion crisscross region of the Chinese Loess Plateau, Soil Till Res, 187, 235-245, https://doi.org/10.1016/j.still.2018.12.014, 2019.

48. Zhang, J. Q., Yang, M. Y., Zhang, F. B., Tang, Y., Wang, X. T., and Wang, Y. J.: Revealing soil erosion characteristics using deposited sediment sources in a complex small catchment in the wind-water erosion crisscross region of the Chinese Loess Plateau, Geoderma, 379, 114634, ARTN 114634, https://doi.org/10.1016/j.geoderma.2020.114634, 2020.

49. Zhang, X. and She, D. L.: Quantifying the sediment reduction efficiency of key dams in the Coarse Sandy Hilly Catchments region of the Yellow River basin, China, Journal Of Hydrology, 602, 126721, ARTN 126721, https://doi.org/10.1016/j.jhydrol.2021.126721, 2021.

50. Zhang, X. H., Xu, M., Wang, S. J., Huang, Y. K., and Xie, Z. Y.: Mapping photovoltaic power plants in China using Landsat, random forest, and Google Earth Engine, Earth System Science Data, 14, 3743-3755, https://doi.org/10.5194/essd-14-3743-2022, 2022.

51. Zhao, G. J., Kondolf, G. M., Mu, X. M., Han, M. W., He, Z., Rubin, Z., Wang, F., Gao, P., and Sun, W. Y.: Sediment yield reduction associated with land use changes and check dams in a catchment of the Loess Plateau, China, Catena, 148, 126-137, https://doi.org/10.1016/j.catena.2016.05.010, 2017.