Article

Estimating the CSLE Biological Conservation Measures' B-Factor Using Google Earth's Engine

Youfu Wu [1,2], **Haijing Shi** [1,3,4,*] and **Xihua Yang** [5,6]

1   Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Xianyang 712100, China; wuyoufu@nwafu.edu.cn
2   Geothermal and Geological Party, Tibet Bureau of Geology and Mineral Exploration and Development, Lhasa 850000, China
3   Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Xianyang 712100, China
4   University of Chinese Academy of Sciences, Beijing 100049, China
5   New South Wales Department of Climate Change, Energy, the Environment and Water, Parramatta, NSW 2150, Australia; xihua.yang@environment.nsw.gov.au
6   School of Life Sciences, University of Technology Sydney, Broadway, NSW 2007, Australia
*   Correspondence: shihaijingcn@nwafu.edu.cn

Abstract: The biological conservation measures factor (B) in the Chinese Soil loss Equation (CSLE) model is one of the main components in evaluating soil erosion, and the accurate calculation of the B-factor at the regional scale is fundamental in predicting regional soil erosion and dynamic changes. In this study, we developed an optimal computational procedure for estimating and mapping the B-factor in the Google Earth Engine (GEE) cloud computing environment using multiple data sources through data suitability assessment and image fusion. Taking the Yanhe River Basin in the Loess Plateau of China as an example, we evaluated the availability of daily precipitation data (CHIRPS, ERA5, and PERSIANN-CDR data) against the data at national meteorological stations. We estimated the B-factor from Sentinel-2 data and proposed a new method, namely the trend migration method, to patch the missing values in Sentinel-2 data using three other remote sensing data (MOD09GA, Landsat 7, and Landsat 8). We then calculated and mapped the B-factor in the Yanhe River Basin based on rainfall erosivity, vegetation coverage, and land use types. The results show that the ERA5 precipitation dataset outperforms the CHIRPS and PERSIANN-CDR data in estimating rainfall and rainfall erosivity, and it can be utilized as an alternative data source for meteorological stations in soil erosion modeling. Compared to the harmonic analysis of time series (HANTS), the trend migration method proposed in this study is more suitable for patching the missing parts of Sentinel-2 data. The restored high-resolution Sentinel-2 data fit nicely with the 10 m resolution land use data, enhancing the B-factor calculation accuracy at local and region scales. The B-factor computation procedure developed in this study is applicable to various river basin and regional scales for soil erosion monitoring.

Keywords: B-factor; CSLE; Google Earth Engine; soil erosion; data patching

1. Introduction

Soil erosion is one of the most serious environmental issues in many parts of the world; it not only causes soil degradation but also poses threats to agricultural production and food security [1–3]. Dynamic monitoring of soil erosion based on remote sensing data and GIS technology can provide important data support for assessing soil erosion conditions and developing soil and water conservation measures [4]. So far, the most widely applied and accepted empirical models are the Universal Soil Loss Equation (USLE) and its revised version (RUSLE), which are used to predict soil erosion rates at regional and global scales, including applications in China [5,6]. Based on the USLE, the Chinese Soil Loss Equation

(CSLE) was developed with the consideration of soil and water conservation measures, which is simple in structure and has stronger applicability in China [7,8]. The USLE/RUSLE uses two dimensionless factors (C and P) to represent the effects of vegetation cover (C-factor) and erosion control management (P-factor) [9–11]. The CSLE model better describes these anthropogenic components by using three factors (B, E, and T, or BET) to represent the effects of biological control (B), engineering control (E) and tillage practices (T) on soil erosion. Among them, the B-factor is the key component representing the effect of vegetation cover (or land use) and biological conservation measures on soil erosion.

The B-factor refers to the ratio of the amount of soil loss on slopes covered with vegetation to that on fallow land under the same conditions (including rainfall, soil, slope, slope length, engineering measures, and tillage measures) [7,12]. The B-factor is influenced by erosive rainfall during the growing season and the vegetation cover. Seasonal changes in vegetation, vertical vegetation structure, and land cover all influence the value of the B-factor to varying degrees, and obtaining the structured and time-series vegetation data is difficult, making the accurate estimation of the B-factor a challenging task.

Same as the cover and management factor (C) in RUSLE, rainfall erosivity is used as a weighting factor to subdivide the B-factor into a certain period (normally half-monthly). Rainfall data are fundamental for B-factor computing, as rainfall significantly affects the dynamic of the B-factor because rainfall varies in time and space, causing temporal instability and uneven spatial distribution [13,14]. Rainfall data are obtainable from ground meteorological observation stations, but the spatial distribution of meteorological stations is often uneven and sparse in remote areas, resulting in a substantial number of missing values [15,16]. Grid-based high temporal and spatial resolution large-scale rainfall products have been developed and employed for soil erosion model computation in recent years. However, the usefulness of multiple grid rainfall products in soil erosion modeling remains unknown; consequently, determining the quality and applicability of different rainfall datasets in soil erosion modeling is crucial.

As an important indicator of surface vegetation conditions and regional ecosystems, fractional vegetation cover (FVC) is important in estimating the B-factor [17]. Many studies have been undertaken over the last few decades, monitoring and assessing FVC using remote sensing satellites and low-altitude unmanned aerial vehicle (UAV) technologies [18,19]. The majority of remote sensing satellite data are derived from low-resolution satellite sensors such as MODIS, which are typically used for large-scale vegetation cover estimation and have poor accuracy. Although low-altitude UAVs have improved data accuracy to some extent, their coverage is limited and cannot meet the requirements for large-scale vegetation cover estimation. For B-factor estimation and monitoring, it is often necessary to integrate remote sensing imaging at various scales to improve the resolution, but it remains a research challenge due to issues such as data availability, compatibility, registration, calibration, and resolution, as well as the trade-offs between spatial, spectral, and temporal information.

The Google Earth Engine (GEE) is a big data platform that aggregates nearly 40 years of huge global remote sensing imagery data, and it has received significant attention from academics recently [20]. The fundamental advantage of the GEE is its remarkable computational power, which enables it to process and analyze data without downloading it, considerably enhancing the efficiency of data processing and analysis. It offers a practical method for conducting extensive regional research, therefore providing a reliable way for the dynamic monitoring of regional B-factors [21].

In this study, we chose the Yanhe River Basin as the study area as it represents the typical vegetation types and landforms of China's Loess Plateau, one of the most severely eroded areas in the world. We assessed the availability and suitability of three daily rainfall datasets in estimating rainfall erosivity in the GEE cloud platform. We estimated the NDVI and FVC from Sentinel-2 and the missing values were filled using other remote sensing data. With these high-precision estimation of regional rainfall and vegetation cover, we proposed and implemented a method for evaluating and mapping the B-factor in the GEE cloud computing environment, in conjunction with land use types.

2. The Study Area

The study area is located in the Yanhe River Basin, in the northern part of Shaanxi Province, with geographical coordinates ranging from 36°23′ to 37°17′N and 108°45′ to 110°28′E. This basin is a primary tributary of the middle reaches of the Yellow River, spanning from Huokou Town to Longmen District (Figure 1). The Yanhe River has a total length of 286.9 km, covering a watershed area of 7687 km^2. This basin is within the continental semi-arid monsoon climate zone and influenced by seasonal monsoons throughout the year. The annual average temperature is 8.8 °C, with an average annual rainfall of 520 mm, primarily concentrated from June to September, accounting for over 70% of the total annual amount. The climate and temperature in the basin vary dramatically from southeast to northwest. The topography is complex, characterized by numerous gullies and ravines, representing a typical Loess hilly and gully region on the Loess Plateau. Vegetation distribution in the area exhibits distinct zonality [22]. The vegetation types change from forested areas to forest–grassland areas and then to typical grassland areas from the southeast to the northwest.
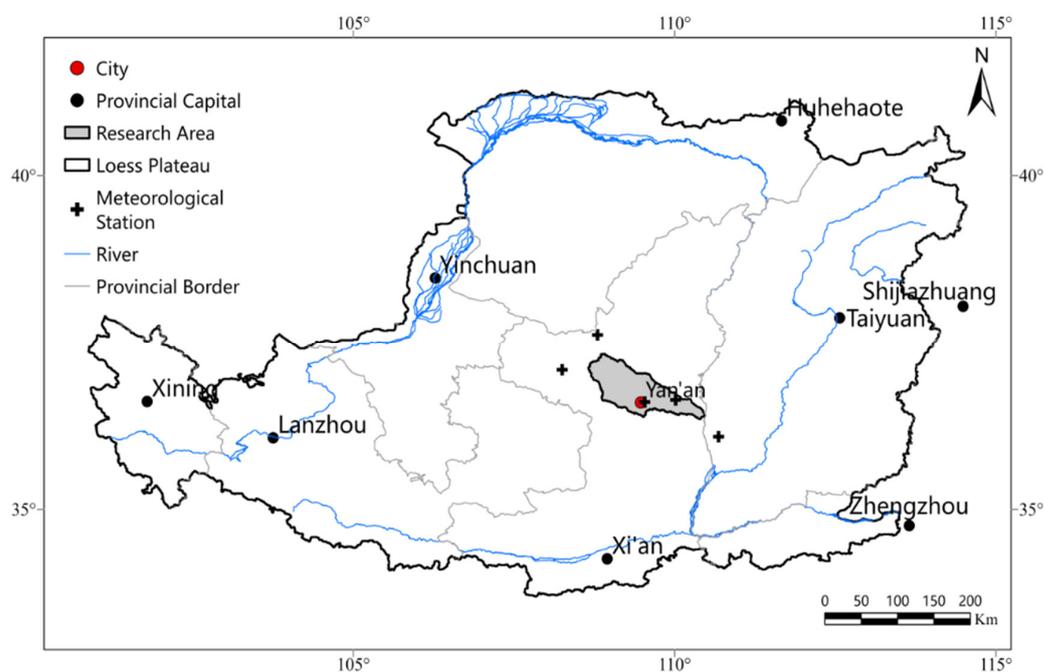
Figure 1. Geographical location of the Yanhe River Basin.

3. Materials and Methods

3.1. Data Sources

The rainfall remote sensing datasets (CHIRPS, ERA5, and PERSIANN-CDR), land use data (WorldCover 10 m, 2020), and surface reflectance remote sensing data (MOD09GA, Landsat 7, Landsat 8, and Sentinel-2) were sourced from the GEE cloud platform. The ground station data were obtained from China's national-level ground meteorological station basic meteorological element daily value dataset (V3.0). The vector boundary data for the Yanhe River Basin were acquired from the Earth System Science Data Center of the Loess Plateau Division (https://www.loess.geodata.cn (accessed on 24 September 2023)).

3.1.1. Rainfall Data

The rainfall datasets used in this study included CHIRPS, ERA5, and PERSIANN-CDR, all of which are available on the GEE platform. CHIRPS is a global grid rainfall dataset with a high resolution that combines 0.05° resolution satellite imagery with in situ site data to create a gridded rainfall time series that records daily rainfall data from 1981 to the present. PERSIANN-CDR is the precipitation estimation from remotely sensed information using

artificial neural networks, which is developed using the PERSIANN algorithm on GridSat-B1 infrared satellite data, and provides daily rainfall estimates at a spatial resolution of 0.25° from 1983 to the near present [23]. The ERA5 data are the fifth generation of climate reanalysis datasets produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), and they provide multi-parameter hourly data on the status of the atmosphere, land surface, and oceans. The data accessibility of the three rainfall datasets listed above was evaluated using daily values of basic meteorological elements (V3.0) from national surface weather stations of China.

3.1.2. Land Use Data

The land use data utilized in this study were the ESA WorldCover v100, a product of the European Space Agency (ESA) based on Sentinel-1 and Sentinel-2 imaging data. This global land use cover product is part of ESA's Fifth Earth Observation Envelope Programme (EOEP-5), with a resolution of 10 m. WorldCover products provide 11 land cover categories, and the data have been independently validated to obtain an overall accuracy of 75% worldwide. The land use types in the Yanhe River Basin include forestland, shrubland, grassland, cultivated land, building land, water body, wetland, and unused land (Figure 2), and the proportion of each land use type is shown in Table 1.
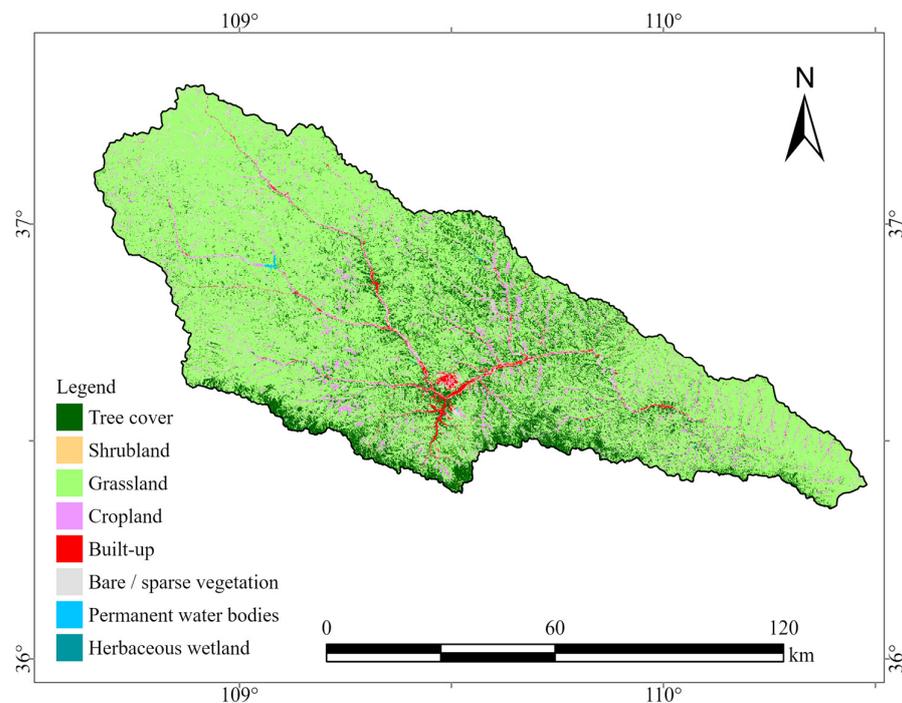
Figure 2. Land use types in Yanhe River Basin.

Table 1. The area and proportions of different land use types in Yanhe River Basin.

Land Use Types	Area (km^2)	Proportions
Forestland	1151.02	14.98%
Shrubland	0.0166	0.00%
Grassland	5358.8	69.76%
Cultivated land	528.086	6.87%
Building land	101.481	1.32%
Unused land	535.108	6.97%
Water body	6.9805	0.09%
Wetland	0.0569	0.00%

3.1.3. Remote Sensing Data

The surface reflectance remote sensing data products used in this study included MOD09GA, Landsat 7, Landsat 8, and Sentinel-2 (Table 2). MOD09GA is a surface spectral reflectance product obtained by the Terra Moderate Resolution Imaging Spectroradiometer (MODIS), which provides Bands 1–7, including 500 m surface reflectance values and 1 km observation and geolocation statistics. Landsat 7 ETM+ (the enhanced thematic Mapper) data consist of eight bands, bands 1–5 and 7 have a spatial resolution of 30 m, band 6 has a spatial resolution of 60 m, and band 8 has a spatial resolution of 15 m. The Landsat 8 Operational Land Imager (OLI) data product consists of 9 bands with a spatial resolution of 30 m, including a panchromatic band of 15 m. Sentinel-2 was launched on 7 March 2017 by the European Space Agency and it provides 13 operating bands ranging from visible to short-wave infrared wavelengths.

Table 2. Summary of the remote sensing products used in this study.

Datasets	Name	Period (Day)	Resolution (m)	Period
MODIS	MOD09GA	1	500	2000–2023
Landsat 7	LE07/C02/T1_L2	8	30	1999–2023
Landsat 8	LC08/C02/T1_L2	8	30	2013–2023
Sentinel 2	S2_SR HARMONIZED	5	10	2017–2023

3.2. Technological Procedure

The B-factor of the Yan River Basin was calculated in the GEE cloud computing environment using Python coding. The specific processing steps were as follows:

(1) The applicability of three gridded daily precipitation datasets CHIRPS, ERA5, and PERSIANN-CDR in the GEE for the Yan River Basin was assessed against the data from national meteorological stations. The dataset that demonstrated the highest suitability was chosen to calculate rainfall erosivity and its proportion.

(2) MOD09GA, Sentinel-2, Landsat 8, and Landsat imagery data were obtained and preprocessed, and the missing areas in the Sentinel-2 data were then identified.

(3) The normalized difference vegetation index (NDVI) was computed separately for each of these four datasets. The missing NDVI data in Sentinel-2 were patched using NDVI calculated from MOD09GA, Landsat 8, and Landsat 7.

(4) The patched NDVI data were analyzed using the pixel binary method to determine the average vegetation coverage (FVC) for 24 half-months over three years.

(5) Based on the average FVC of the 24 half-months, the B-factor was assessed by considering land use types and the proportion of rainfall erosivity factors during this period (Figure 3).

3.3. Image Processing

MOD09GA, Sentinel-2, Landsat 8, and Landsat 7 data from 2019–2021 were used to calculate the NDVI in this study. The total number of effective images for the four products in the Yanhe River Basin was 1096, 749, 133, and 155, respectively, and the missing data rates were 0%, 5%, 20%, and 27%, with a screening condition of less than 20% cloud cover (Table 3). Since Sentinel-2 had the highest coverage and resolution in the entire research region, it was chosen to compute the B-factor in the Yanhe River basin. Due to the missing data in the 7th and 19th half- months, Sentinel-2 images were patched before NDVI computing. The NDVI was calculated as follows:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \tag{1}$$

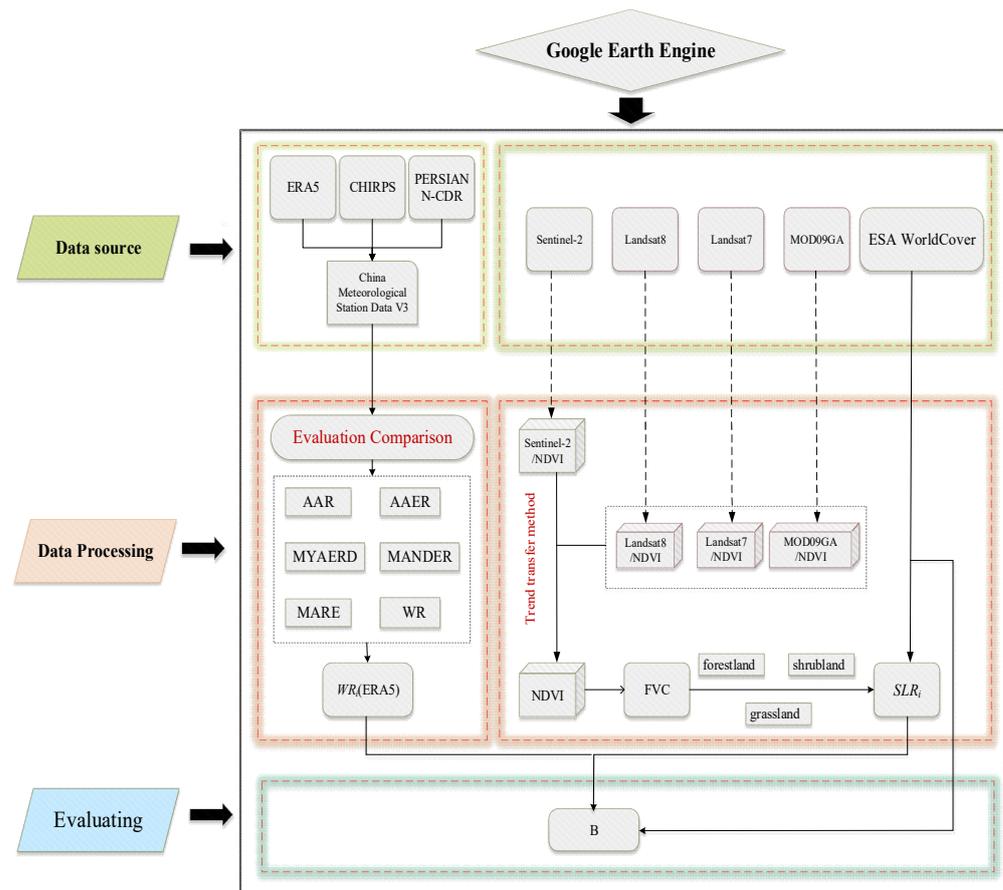where NIR represents the near-infrared band and RED represents the red band of the respective remote sensing images.

**Figure 3.** Flowchart of the data processing in GEE environment.

Table 3. The missing rate of the four remote sensing datasets in 24 half-months.

Half-Month	MOD09GA		Sentinel-2		Landsat 8		Landsat 7	
	Number of Available Images	Missing Rate	Number of Available Images	Missing Rate	Number of Available Images	Missing Rate	Number of Available Images	Missing Rate
1	45	0%	27	0%	6	10%	6	55%
2	48	0%	35	0%	5	0%	8	6%
3	45	0%	31	0%	5	41%	9	3%
4	40	0%	18	0%	6	0%	3	46%
5	45	0%	21	0%	7	9%	9	6%
6	48	0%	32	0%	8	0%	10	17%
7	45	0%	14	56%	3	1%	2	92%
8	45	0%	33	0%	8	0%	9	11%
9	45	0%	35	0%	8	0%	9	10%
10	48	0%	43	0%	6	0%	8	2%
11	45	0%	34	0%	8	45%	3	40%
12	45	0%	21	0%	2	81%	1	55%
13	45	0%	33	0%	6	16%	9	0%
14	48	0%	30	0%	2	30%	2	91%
15	45	0%	35	0%	5	6%	5	35%
16	48	0%	33	0%	6	0%	1	90%
17	45	0%	38	0%	6	63%	9	2%
18	45	0%	31	0%	6	3%	7	9%
19	45	0%	9	56%	0	100%	2	48%
20	48	0%	35	0%	2	1%	8	25%
21	45	0%	51	0%	7	62%	9	2%
22	45	0%	27	0%	5	1%	9	1%
23	45	0%	45	0%	6	0%	9	8%
24	48	0%	38	0%	10	0%	8	7%
Total	1096	0%	749	5%	133	20%	155	27%

In terms of spatial distribution and time series, Sentinel-2 data were highly consistent with MOD09GA, Landsat 8, and Landsat 7 data. Therefore, MOD09GA, Landsat 8, and Landsat 7 remote sensing images were used to patch the missing data in the Sentinel-2 imagery. Figure 4 demonstrates that the average NDVI variance among the four groups of images had a highly consistent trend throughout the 24 half-monthly time series. It was found that the correlation of each two groups of data in the time series was above 0.96. Furthermore, the NDVI values for the four datasets from 2019 to 2021 were synthesized by the maximum synthesis method, and the correlation of the four datasets was obtained by using the Pearsons correlation function. The heatmap of Pearson correlation coefficient matrix shows that the spatial relationships between Sentinel-2 imagery, Landsat 7, Landsat 8, and MOD09GA data were fitted, with all correlation coefficients exceeding 0.5 (Figure 5).
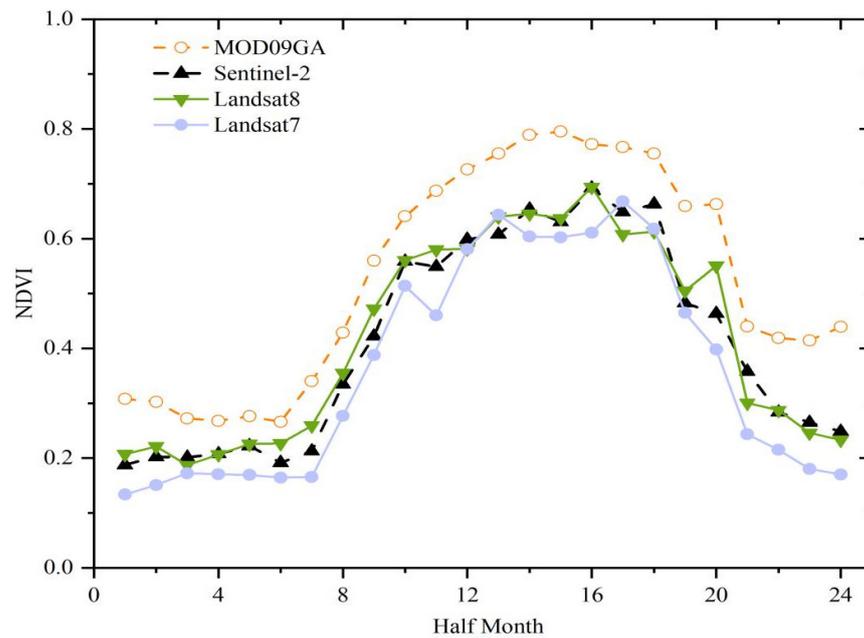
Figure 4. The seasonal changes of NDVI values derived from MOD09GA, Sentinel-2, Landsat 8, and Landsat 7 image data.
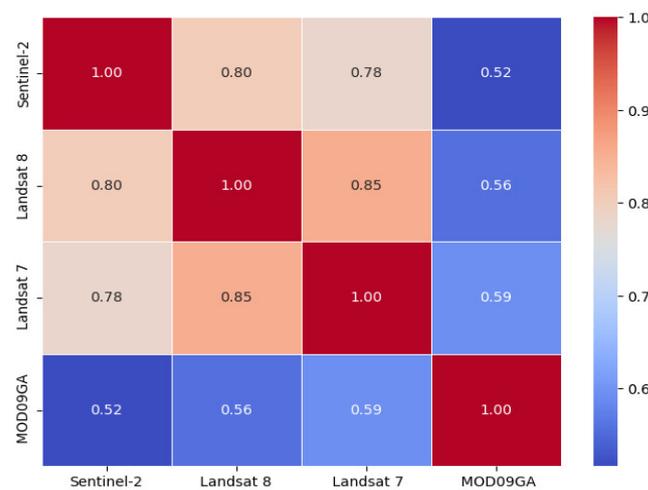
Figure 5. Heatmap of Pearson correlation coefficient matrix.

There are several methods available for data patching, which can be broadly categorized into two types. The first type is spatial interpolation, which is useful for filling small-scale missing data, but is ineffective for large-scale data gaps. The second type is

linear/non-linear interpolation in the time series. Linear interpolation methods, such as the best index slope extraction (BISE) method [19], use a sliding window approach to identify and replace missing values in a series. However, determining the optimal window size is challenging, and it has limited applicability in capturing the non-linear patterns of real-world imagery. Non-linear interpolation methods, such as the Savitzky–Golay (S-G) filtering method, asymmetric Gaussian (AG) model filtering method [24], and harmonic analysis of time series (HANTS) method [25], have more complex calculations and often suffer from overfitting, resulting in data value distortion. Based on these concerns, this study proposed a trend migration method, a new approach for repairing missing data based on the rate of change in remote sensing imagery over a time and the surrounding raster values in the time series of the missing values. The HANTS approach, which is commonly used for time series analysis, was also utilized for comparative analysis. The formula of the trend migration is as follows:

$$2S_i = S_{i-1}\frac{D_i}{D_{i-1}} + S_{i+1}\frac{D_i}{D_{i+1}} \qquad (2)$$

where S_{i-1} is the value of the Sentinel-2 data in the ith half-month, Di represents the corresponding values of an alternative dataset (i.e., MOD09GA, Landsat 8, or Landsat 7) that is in high temporal consistency with Sentinel-2 data in the time series for the same half-month.

The Sentinel-2 images after data patching were used to calculate the NDVI for each half-month period from 2019 to 2021, resulting in a total of 24 half-months per year. Based on NDVI data, the pixel-based binary model was applied to compute the vegetation cover, and the average vegetation cover for the 24 half-month periods over a period of three years was then derived. The formula for calculating FVC is as follows:

$$FVC = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \qquad (3)$$

where FVC is vegetation coverage (for grassland, shrubland, and forestland). NDVI is the pixel value of the normalized difference vegetation index; $NDVI_{min}$ and $NDVI_{max}$ are NDVI values of the pure vegetation cover pixel and the bare soil cover pixel, respectively.

3.4. Calculation of the B-Factor

The B-factor was calculated by combining vegetation cover (FVC) data with the 10 m resolution land use types and the rainfall erosion factor proportions over one year in 24 half-monthly intervals [26]. Specific formulas were used to calculate the B-factor values for shrublands, forestland, and grasslands, while the remaining land use classifications were allocated B-factor values directly based on land use types (Table 4). The annual average value of the B-factor was estimated based on the sum of soil loss ratio (SLR_i) for each half-month (1–24) in a year and multiplied by the rainfall erosivity ratio of the same half-month as a weighting factor [27]. The formula is as follows:

$$B = \sum_{i=0}^{24} SLR_i \times WR_i \qquad (4)$$

where WR_i is the ratio of rainfall erosivity in the *i*th half-month to annual erosivity, with a value range of 0–1; SLR_i is the soil loss ratio of forestland, shrubland, and grassland in the *i*th half-month, dimensionless, with a value range of 0–1.

For forestland, the SLR_i is calculated as:

$$SLR_i = 0.44468 \times e^{(-3.20096 \times GD)} - 0.04099 \times e^{FVC - FVC \times GD} + 0.025 \qquad (5)$$

Table 4. The B-factor value based on land use classes.

Land Use (Class I)	Land Use (Class II)	B-Factor Value	Description
Cultivated land	Paddy field	1	Soil and water conservation benefits reflected by tillage measure factor
	Irrigated land	1	Soil and water conservation benefits reflected by tillage measure factor
	Dry land	1	Soil and water conservation benefits reflected by tillage measure factor
Residential, industrial, and mining lands	Urban residential lands	0.01	Equivalent to 80% of vegetation cover
	Rural residential lands	0.025	Equivalent to 60% of vegetation cover
	Independent industrial and mining lands	1	Equivalent to no vegetation cover
	Commercial service and public lands	0.01	Equivalent to 80% of vegetation cover
	Special land use	0.1	
Land for transportation		0.01	Equivalent to 80% of vegetation cover
Land for water area and water conservancy facilities		0	Forced to 0, so that the amount of erosion is equal to 0
Other land	Saline alkali soil	0	-
	Sandy land	0	-
	Swamp land	0	-
	Bare rock	0	-
	Bare soil	1	-
	Glaciers and permanent snow cover	0	-

For shrubland, the SLR_i is calculated as:

$$SLR_i = \frac{1}{1.17647 + 0.86242 \times 1.05905^{100 \times FVC}} \tag{6}$$

For grassland, the SLR_i is calculated as:

$$SLR_i = \frac{1}{1.25 + 0.78845 \times 1.05968^{100 \times FVC}} \tag{7}$$

where GD is the understory coverage of the tree forest, with a value range of 0–1, including the understory coverage of all vegetation (shrubs, herbs, and litter) except for the tree canopy layer, based on field investigation or expert experience. FVC is vegetation coverage calculated as in Formula (3).

4. Results

### 4.1. Applicability of Different Rainfall Datasets in the Yanhe River Basin

Multiple rainfall indices for the Yanhe River Basin from 1986 to 2015 were generated using three different daily rainfall datasets in the GEE platform (Table 5). The results reveal that the multi-year mean erosive rainfall, daily erosive rainfall, and rainfall erosivity estimated from ERA5 dataset are the closest to those from the meteorological stations. The average annual erosive rainfall measured by meteorological stations was 283 mm, while the ERA5, GHIRPS, and PERSIANN-CDR datasets measured 288 mm, 332 mm, and 128 mm, respectively. The absolute errors between the three datasets and the meteorological stations were 5.17 mm, 49.24 mm, and −154.82 mm, respectively. The average annual rainfall erosivity calculated by meteorological station data was 1327 MJ mm hm^{-2} h^{-1} yr^{-1}, whereas the mean annual rainfall erosivity values of the ERA5, GHIRPS, and PERSIANN-CDR datasets were 1327, 1876, and 502 MJ mm hm^{-2} h^{-1} yr^{-1}, respectively. The differences

between them and meteorological stations were -105.23, 549.1, and -825.18 MJ mm hm^{-2} h^{-1} yr^{-1}, respectively.

Table 5. Comparisons of rainfall parameters from the national meteorological stations and the three daily rainfall datasets.

Rainfall Parameters	Meteorological Station	ERA5	CHIRPS	PERSIANN-CDR
Annual average rainfall (mm)	467	570	481	455
Average annual erosive rainfall (mm)	283	288	332	128
Multi-year average erosive rainfall days (mm day^{-1})	22	22	28	20
Multi-year average number of days of erosive rainfall (day)	12	13	12	6
Multi-year average rainfall erosivity (MJ·mm hm^{-2} h^{-1} yr^{-1})	1327	1222	1876	502

Figure 6 depicts the fluctuation trend of annual average rainfall erosivity estimated from four independent rainfall datasets. The annual average rainfall erosivity of all four datasets peak in 2013, with the ERA5 dataset having the closest peak value and variation trend to the meteorological station. The WR change trend of the three datasets and the meteorological station data is substantially the same, as shown in Figure 7. The variation curves of WR by meteorological station data and the ERA5 dataset create a bimodal distribution, with WR reaching peaks of 16.2% and 16.0%, respectively, in the 13th and 16th half-months. The WR curves from the CHIRPS and PERSIANN-CDR datasets have a unipeak distribution and peak at 18.9% and 18.1%, both in the 15th half-months. The WR trend in the ERA5 dataset is the most similar to meteorological station data.
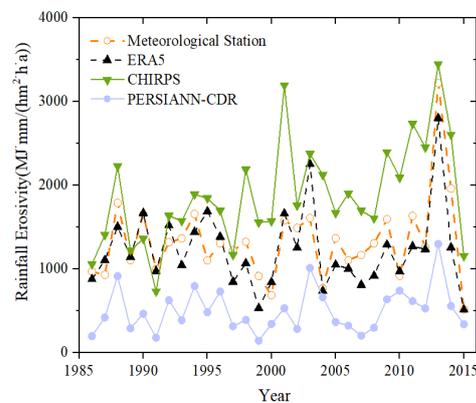

Figure 6. Annual average rainfall erosivity calculated by different datasets from 1986 to 2015.
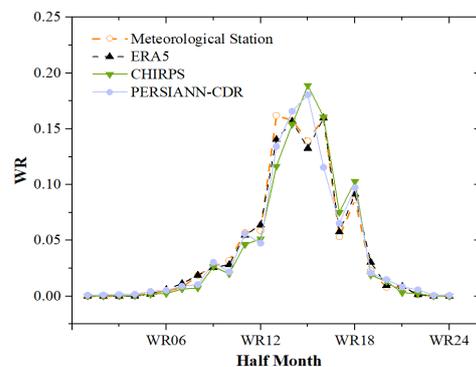
Figure 7. WR (the ratio of half-month rainfall erosivity to average annual rainfall erosivity) calculated by different datasets from 1986 to 2015.

*4.2. The NDVI Calculated by Patched Sentinel-2 Images*

The NDVI of the Yanhe River Basin was calculated for 24 half-months using the raw Sentinel-2 images, the Sentinel-2 images with 20% filtered cloud cover, the Sentinel-2 images patched by the HANTS method, and the Sentinel-2 images patched by the trend migration method (Figure 8). Except for the missing data in the 7th and 19th half-month periods, the NDVI variation curves generated from the patched images essentially coincide with the variation curves of the raw images and the images with 20% filtered cloud cover, maximizing the retention of the original data. The NDVI values calculated from the Sentinel-2 images patched by the HANTS method and raw images might overfit and distort images, particularly in the 4th to 8th half-month periods, resulting in errors in time-series-based remote sensing studies, such as vegetation phenology research. Compared to the HANTS method, the trend migration method proposed in this study is more consistent with the original data, and more reliable in patching the missing data in the Sentinel-2 images.
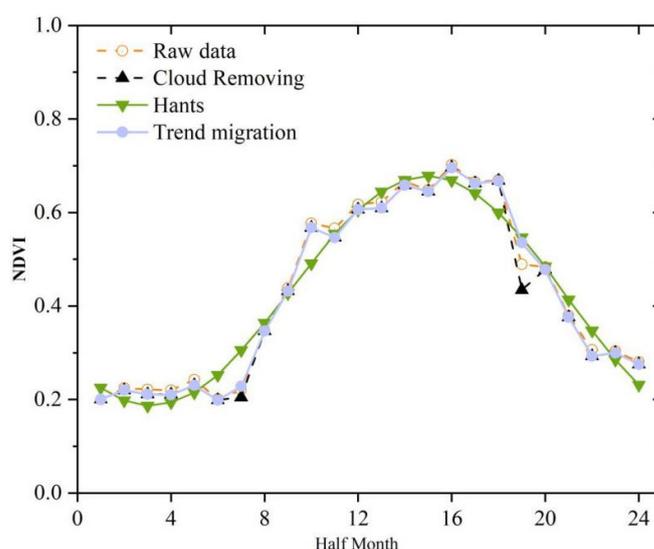
Figure 8. NDVI calculated for 24 half-months using the raw Sentinel-2 images, the Sentinel-2 images with 20% filtered cloud cover, the Sentinel-2 images patched by the HANTS method, and the Sentinel-2 images patched by the trend migration method.

The trend migration method was employed to patch the missing NDVI data from Sentinel-2 images in the Yanhe River Basin in the 7th and 19th half-months. Taking the 7th half-month as an example, when using raw Sentinel-2 images, there were visible cloud cover and localized anomalies in the NDVI images (Figure 9a). When using the Sentinel-2 images with 20% filtered cloud cover, the missing rate of NDVI images reached 56% (Figure 9b). By using the trend migration method, the missing parts were patched by MOD09GA, Landsat 7, and Landsat 8 images in the 7th half-month, the calculated NDVI image was more coordinated overall, and the average NDVI values in the Yanhe River Basin changed from 0.22 before image patching to 0.23 after image patching (Figure 9c,d). Similarly, the average NDVI values changed from 0.51 to 0.53 in the 19th half-month after image patching (Figure 9e,f).

Figure 10 displays the results of the NDVI calculated using Sentinel-2 images, MOD09GA images, Landsat 7 images, and Landsat 8 images. It reveals that the NDVI calculated from the patched Sentinel-2 data not only exhibited higher data integrity, but also possessed the finest resolution (10 m) showing more details in vegetation patterns and distribution.
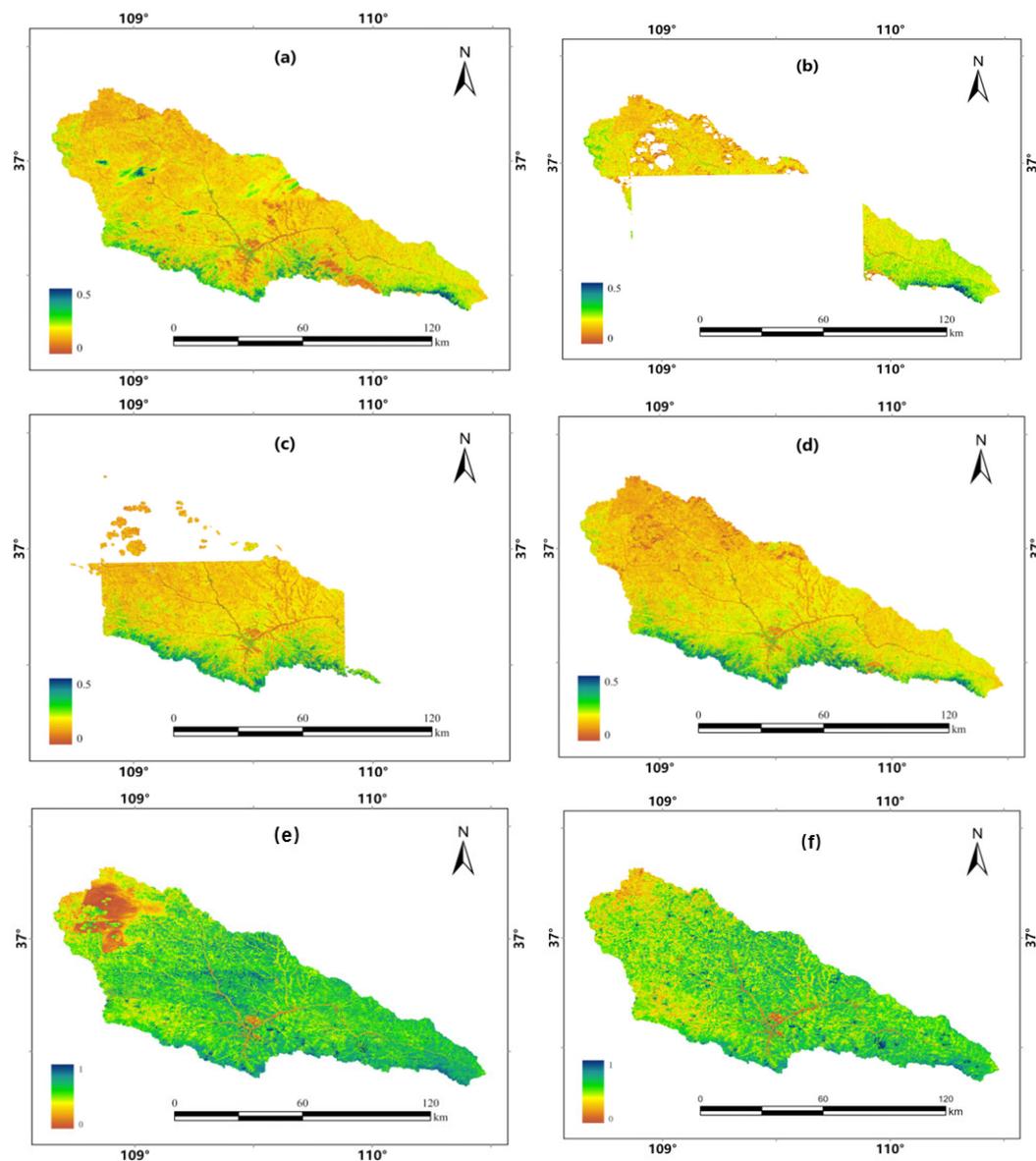
Figure 9. NDVI generated using (**a**) the raw Sentinel-2 images in the 7th half-month, (**b**) the Sentinel-2 images with clouds filtered in the 7th half-month, (**c**) the missing parts patched by MOD09GA, Landsat 7, and Landsat 8 images in the 7th half-month, (**d**) the Sentinel-2 images after patching in the 7th half-month, (**e**) the raw Sentinel-2 images in the 19th half-month, and (**f**) the Sentinel-2 images after patching in the 19th half-month.

4.3. Assessment of Vegetation Cover and Biological Measures' B-Factor

The patched Sentinel-2 images were used to produce the FVC at a spatial resolution of 10 m for all 24 half-monthly periods of a year (Figure 11). The average value of FVC increases first and then declines with time, according to the shifting trend. The average FVC was less than 0.1 before the 7th half-month (early spring), while it increased dramatically after the 8th half-month (spring), and peaked at 0.8 in the 16th half-month (early August). Following that, the mean value of FVC progressively decreased, eventually going below 0.1 at the 24th half-month (December).
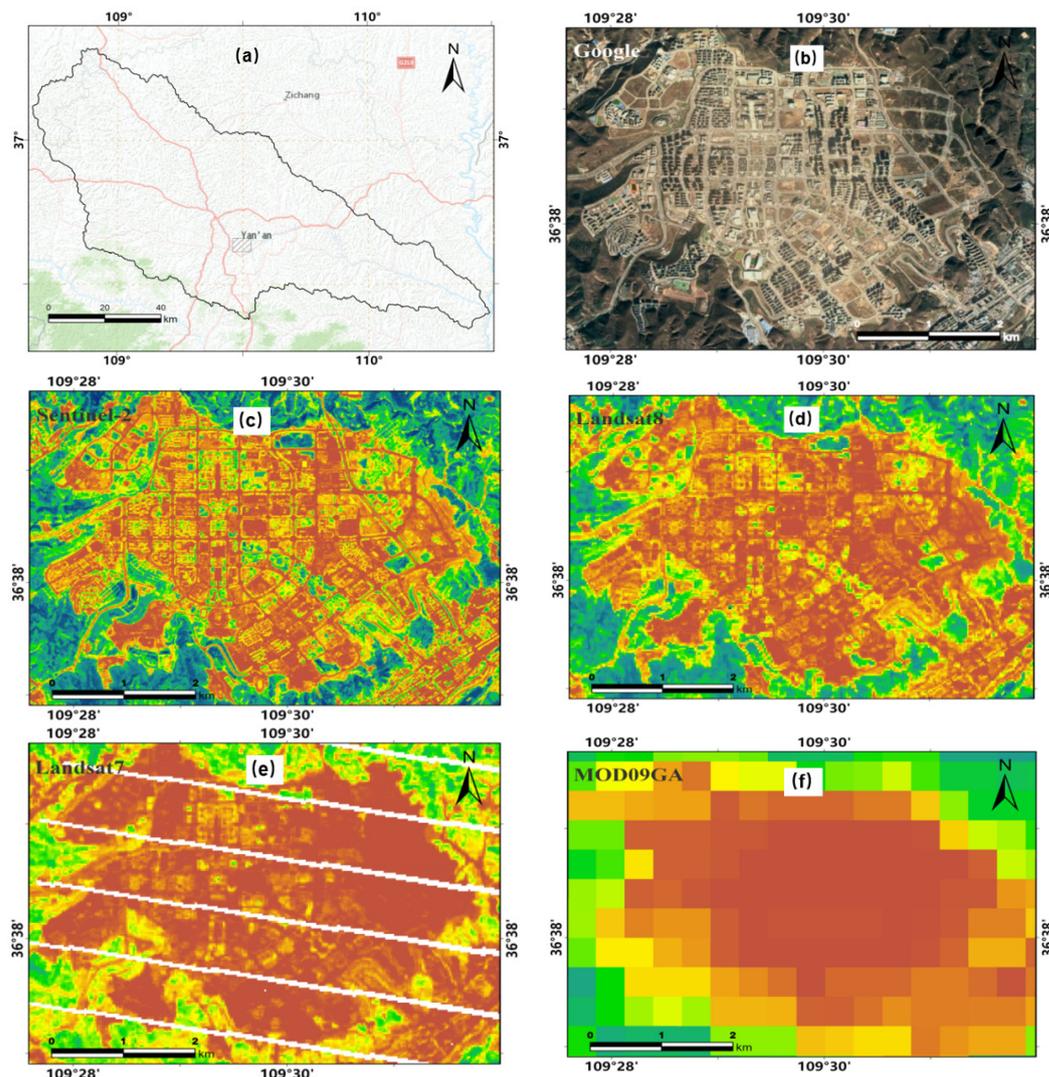
Figure 10. Details of NDVI calculated in the Yanhe River Basin. (**a**) The selected rectangular region, (**b**) the Google map of the selected region, (**c**) NDVI calculated from Sentinel-2, (**d**) NDVI calculated from MOD09GA, (**e**) NDVI calculated from Landsat 7, and (**f**) NDVI calculated from MODIS. The colors in the (**c**–**f**) figures represent the size of NDVI value, and the areas with green color have larger NDVI value.

The SLR was calculated based on FVC, and WR was calculated based on ERA5 data over a 24 half-monthly period on the GEE platform. We then calculated the B-factor of the Yanhe River Basin based on Equation (4), combined with land use types. Figure 12 depicts the spatial distribution of the B-factor in the Yanhe River Basin in 2020. Except for the agricultural land with a value of 1, high-value areas of the B-factor in the basin were concentrated along the river system due to human activities and the proximity of urban centers to water bodies, resulting in relatively low vegetation covering. Furthermore, in addition to places near water bodies, some high-value B-factor areas could be found in the northwest of the Yanhe River Basin, which was characterized by low-coverage grassland with weak development and sparse vegetation coverage, resulting in relatively high B-factor values. Low-value B-factor habitats, on the other hand, were mostly found in the forests and grasslands with high vegetation cover.
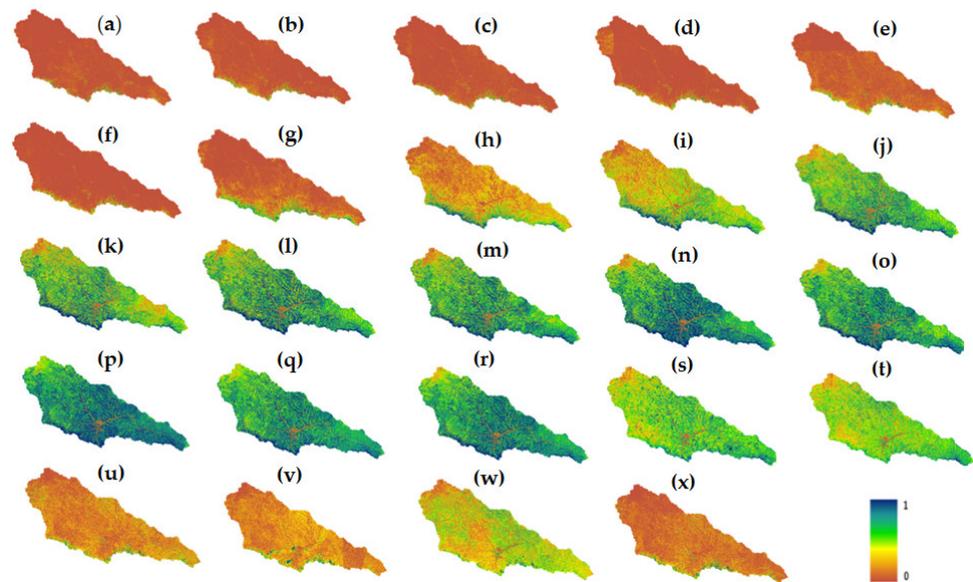
Figure 11. Mean 24 half-monthly FVC values from 2019 to 2021 in the Yanhe River Basin from January to December. Each subfigure represents the mean half month FVC.
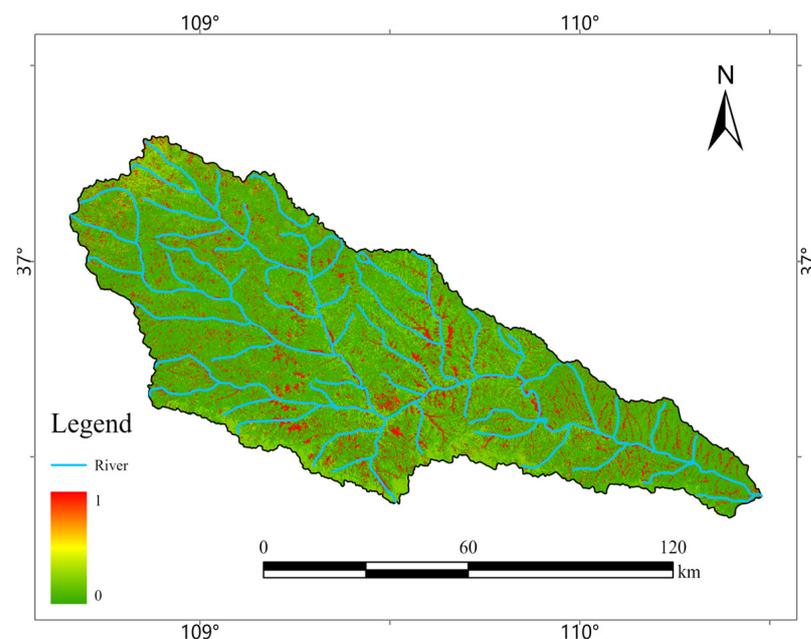
Figure 12. Spatial distribution of B-factor in the Yanhe River Basin.

5. Discussion

5.1. Applicability of Different Rainfall Datasets in the Yanhe River Basin

When comparing the performance of the three datasets, PERSIANN-CDR exceeded the other two in average annual rainfall, but fell shorts in terms of the other parameters. The CHIRPS dataset overestimated values for average annual erosive rainfall, average erosive rainfall days, and average rainfall erosion force calculation, and had poor consistency with data from meteorological stations in terms of WR. Although the resolution of the ERA5 dataset was not the highest, it demonstrated extremely high consistency with the meteorological station data in terms of average erosive rainfall, average erosive rainfall days, average erosive rainfall duration, average rainfall erosion force, and WR trends. Therefore, the ERA5 dataset was chosen and utilized as the alternative to meteorological station data

to calculate and map the B-factor in the CSLE model. In addition, ERA5 datasets are under continuous development with improvements in accuracy and spatiotemporal resolutions.

Many applications require rainfall erosivity to be estimated on an event basis or in near real time, such as post-fire erosion monitoring and sediment delivery at storm events. Weather radar rainfall data have high spatial (1 km) and temporal (10 min) resolutions with great potential for these near real time applications [12]. Our further study will explore the weather radar data and the application in event-based rainfall erosivity and sediment delivery estimation.

5.2. The NDVI and B-Factor Calculated Using Patched Sentinel-2 Images

Compared with the NDVI derived from MOD09GA images, Landsat 7 images, and Landsat 8 images, the NDVI derived from Sentinel-2 images not only had the highest spatial resolution but also the highest consistency. This demonstrates that Sentinel-2 images can be successfully used to estimate the NDVI and, subsequently, the B-factor for vegetation monitoring and erosion modeling.

The trend migration method was employed to patch the missing NDVI data from Sentinel-2 images in the Yanhe River Basin in the 7th and 19th half-months. The results imply that the method proposed in this study for patching Sentinel-2 images was simple, reliable, and highly useful. The patched Sentinel-2 images were compatible with the 10 m resolution land use data, which improved the accuracy of B-factor calculations from the regional to the spot level.

Though the satellite-derived NDVI is commonly used to estimate FVC and the cover and management factor (B-factor in CSLE or C-factor in RUSLE), it has several limitations. First, the NDVI is a green vegetation (or PV) index, it does not reflect non-green vegetation (or NPV); however, total vegetation cover (PV and NPV) is required in erosion modeling. Second, remotely sensed vegetation cover is canopy cover, not the surface cover or understory cover as required in erosion modeling. Though remote sensing has been successfully applied to derive overstory FVC [28], it remains a research challenge to extract understory FVC due to the complex environment within the forest ecosystem. Further studies on FVC and the B-factor may consider the applications of light detection and ranging (LiDAR) and unmanned aerial vehicles (UAVs) to quantify both overstory and understory vegetation cover.

5.3. Assessment of Vegetation Cover and Biological Measures' B-Factor

The average B-factor value in the entire Yanhe River Basin in 2020 was 0.19, which was slightly lower than the 0.22 average in the Loess hilly–gully region in 2017 and 0.38 in 2010, calculated by Huang [29]. The Yanhe River Basin had relatively high overall vegetation coverage, with increasing trends in woods and grasslands and a continuous decrease in the cultivated land area, which had decreased by about 25% between 2010 and 2020 [30]. The improvement of vegetation structure and changes in land use types have both contributed to the overall declining trend of the B-factor. Compared to previous studies, our projected average B-factor value in this study appears more consistent and appropriate.

The CSLE B-factor is compatible to the RUSLE C-factor, as both are defined and calculated in essentially the same way. It is possible that the RUSLE C-factor could be replaced by the CSLE B-factor or vice versa with some simple adjustments. With such compatibility, the CSLE is potentially applicable worldwide, along with the widely used USLE or RUSLE model.

6. Conclusions

The applicability of ERA5, CHIRPS, and PERSIANN-CDR daily rainfall data in the Yanhe River Basin was evaluated using the precipitation data from national meteorological stations for the period 1985 to 2015 in the GEE. The missing NDVI data in Sentinel-2 images were patched by incorporating MOD09GA, Landsat 7, and Landsat 8 images. An automated process was developed using Python codes for rainfall data assessment, image

patching and processing, as well as FVC, WR, and SLR calculation, thereby enabling the assessment and mapping of the B-factor in the GEE cloud computing environment. We can draw the following conclusions from this study:

(1) The ERA5 precipitation dataset exhibited strong agreement with meteorological station data across various metrics including average erosive rainfall, average erosive rainfall intensity, average rainfall erosivity, and WR over multiple years. It demonstrated better suitability for the Yanhe River Basin compared to the CHIRPS and PERSIANN-CDR datasets, making it a viable alternative for calculating rainfall erosivity in the absence of meteorological station data.

(2) This study proposed a new and simple method, the trend migration method, for patching missing Sentinel-2 data based on the rate of change of the image in time. Sentinel-2 modified by the trend migration approach outperformed the HANTS method in terms of data authenticity. The restored high-resolution Sentinel-2 data fit nicely with the 10 m resolution land use data, enhancing the B-factor calculation accuracy from the regional to the spot level.

(3) In the GEE environment, we developed a fully automated B-factor calculation procedure, and the B-factor was assessed and mapped efficiently based on rainfall erosivity, soil loss rate, vegetation covering, and land use types in the Yanhe River Basin. The computation procedure is broadly applicable and can be extended to various river basins as well as wider regional scales, which provides data and technical support for regional soil erosion monitoring.

Author Contributions: Conceptualization, H.S.; formal analysis, Y.W. and H.S.; funding acquisition, H.S. and X.Y.; methodology, X.Y., Y.W. and H.S.; data curation, Y.W.; writing—original draft preparation, Y.W. and H.S.; writing—review and editing, X.Y., H.S. and Y.W. All authors have read and agreed to the published version of the manuscript.

References

1. Leng, S.Y.; Feng, R.G.; Li, R. Key Research Issues of Soil Erosion and Conservation in China. *J. Soil Water Conserv.* **2004**, *18*, 1–6.
2. Shi, P.J.; Liu, B.Y.; Zhang, K.L.; Jin, Z. Soil Erosion Process and Model Studies. *Resour. Sci.* **1999**, *21*, 9–18.
3. Li, Z.; Zhu, B.; Li, P. Advancement in Study on Soil Erosion and Soil and Water Conservation. *Acta Pedol.* **2008**, *45*, 802–809.
4. Cai, Q.; Liu, J. Evolution of Soil Erosion Models in China. *Prog. Geogr.* **2003**, *22*, 242–250.
5. Fu, B.; Liu, Y.; Lü, Y.; He, C.; Zeng, Y.; Wu, B. Assessing the Soil Erosion Control Service of Ecosystems Change in the Loess Plateau of China. *Ecol. Complex.* **2011**, *8*, 284–293. [CrossRef]
6. Lin, J.; Guan, Q.; Tian, J.; Wang, Q.; Tan, Z.; Li, Z.; Wang, N. Assessing Temporal Trends of Soil Erosion and Sediment Redistribution in the Hexi Corridor Region Using the Integrated RUSLE-TLSD Model. *CATENA* **2020**, *195*, 104756. [CrossRef]
7. Liu, B.; Zhang, K.; Xie, Y. An Empirical Soil Loss Equation. In Proceedings of the 12th International Soil Conservation Organization Conference, Beijing, China, 26 May 2002; pp. 26–31.
8. Shi, W.; Huang, M.; Barbour, S.L. Storm-Based CSLE That Incorporates the Estimated Runoff for Soil Loss Prediction on the Chinese Loess Plateau. *Soil Tillage Res.* **2018**, *180*, 137–147. [CrossRef]
9. Ayalew, D.A.; Deumlich, D.; Šarapatka, B.; Doktor, D. Quantifying the Sensitivity of NDVI-Based C Factor Estimation and Potential Soil Erosion Prediction using Spaceborne Earth Observation Data. *Remote Sens.* **2020**, *12*, 1136. [CrossRef]

10. Biddoccu, M.; Guzman, G.; Capello, G.; Thielke, T.; Strauss, P.; Winter, S.; Zaller, J.G.; Nicolai, A.; Cluzeau, D.; Popescu, D.; et al. Evaluation of soil erosion risk and identification of soil cover and management factor (C) for RUSLE in European vineyards with different soil management. *Int. Soil Water Conserv. Res.* **2020**, *8*, 337–353. [CrossRef]

11. Pinson, A.O.; AuBuchon, J.S. A new method for calculating C factor when projecting future soil loss using the Revised Universal soil loss equation (RUSLE) in semi-arid environments. *CATENA* **2023**, *226*, 107067. [CrossRef]

12. Yin, S.; Zhu, Z.; Wang, L.; Liu, B.; Xie, Y.; Wang, G.; Li, Y. Regional Soil Erosion Assessment Based on a Sample Survey and Geostatistics. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 1695–1712. [CrossRef]

13. Xu, Y.; Ding, Y.; Zhao, Z. Confidence Analysis of NCEP/NCAR 50-Year Global Reanalyzed Data in Climate Change Research in China. *J. Appl. Meteorol.* **2001**, *12*, 337–347. (In Chinese)

14. Zhao, T.; Ai, L.; Feng, J.M. An Intercomparison between NCEP Reanalysis and Observed Data over China. *Clim. Environ. Res.* **2004**, *9*, 278–294. (In Chinese)

15. Wang, Z.; Zhong, R.; Chen, J.; Huang, W. Evaluation of Drought Utility Assessment of TMPA Satellite-Remote-Sensing-Based Precipitation Product in Mainland China. *CSAE* **2017**, *33*, 163–170. (In Chinese)

16. Cai, Y.; Jin, C.; Wang, A.; Guan, D.; Wu, J.; Yuan, F.; Xu, L. Comprehensive Precipitation Evaluation of TRMM 3B42 with Dense Rain Gauge Networks in a Mid-Latitude Basin, Northeast, China. *Theor. Appl. Climatol.* **2016**, *126*, 659–671. [CrossRef]

17. Liu, H.; Zhang, M.; Lin, Z.; Xu, X. Spatial Heterogeneity of the Relationship between Vegetation Dynamics and Climate Change and Their Driving Forces at Multiple Time Scales in Southwest China. *Agric. For. Meteorol.* **2018**, *256–257*, 10–21. [CrossRef]

18. Li, M.; Zhang, Y. Temporal and spatial variation characteristics and influencing factors of vegetation cover in the middle Yellow River Basin. *J. Guizhou Norm. Univ. Nat. Sci.* **2023**, *41*, 10–20. (In Chinese)

19. Jiang, W.; Yuan, L.; Wang, W.; Cao, R.; Zhang, Y.; Shen, W. Spatio-Temporal Analysis of Vegetation Variation in the Yellow River Basin. *Ecol. Indic.* **2015**, *51*, 117–126. [CrossRef]

20. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

21. Zhang, B. Remotely Sensed Big Data Era and Intelligent Information Extraction. *Geomat. Inf. Sci. Wuhan Univ.* **2018**, *43*, 1861–1871. (In Chinese)

22. Wen, Z.; Jiao, F.; Jiao, J. Prediction and mapping of potential vegetation distribution in Yanhe River catchment in hilly area of Loess Plateau. *Chin. J. Appl. Ecol.* **2008**, *19*, 1897–1904. (In Chinese)

23. Ashouri, H.; Hsu, K.; Sorooshian, S.; Braithwaite, D.; Knapp, K.; Cecil, L.; Nelson, B.; Prat, O.P. PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bull. Amer. Meteor. Soc.* **2015**, *96*, 69–83. [CrossRef]

24. Roerink, G.J.; Menenti, M.; Verhoef, W. Reconstructing Cloudfree NDVI Composites Using Fourier Analysis of Time Series. *Int. J. Remote Sens.* **2000**, *21*, 1911–1917. [CrossRef]

25. Zhou, J.; Jia, L.; Menenti, M. Reconstruction of Global MODIS NDVI Time Series: Performance of Harmonic ANalysis of Time Series (HANTS). *Remote Sens. Environ.* **2015**, *163*, 217–228. [CrossRef]

26. Liu, B.Y.; Xie, Y.; Li, Z.Y.; Liang, Y.; Zhang, W.B.; Fu, S.H.; Yin, S.Q.; Wei, X.; Zhang, K.L.; Wang, Z.Q.; et al. The assessment of soil loss by water erosion in China. *Int. Soil Water Conserv. Res.* **2020**, *8*, 430–439. [CrossRef]

27. Yang, X.; Zhu, Q.; Tulan, M.; McInnes-Clarke, S.; Sun, L.; Zhang, X.P. Near real-time monitoring of post-fire erosion after storm events: A case study in Warrumbungle National Park, Australia. *Int. J. Wildland Fire* **2018**, *27*, 413–422. [CrossRef]

28. Guerschman, J.P.; Hill, M.J.; Renzullo, L.J.; Barrett, D.J.; Marks, A.S.; Botha, E.J. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sens. Environ.* **2009**, *113*, 928–945. [CrossRef]

29. Huang, C. Study on the Spatiotemporal Variation of Soil Erosion in the Loess Plateau over the Past 40 Years and Its Controlling Factors. Ph.D. Thesis, Northwest University, Kirkland, WA, USA, 2021. (In Chinese).

30. Yang, S.; Quan, Q.; Xu, J.; Zhang, F.; Liu, T. Water Conservation Function in the Process of Land Use Change in Yanhe River Basin. *Soil Water Conserv. Chin.* **2022**, 33–36. (In Chinese) [CrossRef]