Plant litter estimation and its correlation with sediment concentration in the Loess Plateau

Qian Li [1], **Ligang Ma** [Corresp., 1, 2], **Suhong Liu** [3, 4], **Adilai Wufu** [1], **Yinbo Li** [1], **Shengtian Yang** [1], **Xiaodong Yang** [1]

[1] College of Resource and Environmental Science, Xinjiang University, Urumqi, Xinjiang, China

[2] General College Key Laboratory of Smart City and Environmental Modeling, Xinjiang University, Urumqi, Xinjiang, China

[3] Faculty of Geographical Science, Beijing Normal University, Beijing, Beijing, China

[4] Beijing Key Laboratory of Environmental Remote Sensing and Digital Cities, Beijing Normal University, Beijing, Beijing, China

Corresponding Author: Ligang Ma
Email address: mosanzju@126.com

Background. Sediment concentration in the water of the loess Plateau region has dramatically decreased during the past two decades. Plant litter is considered to be one of the most important factors for this change. Existing remote sensing studies that focus on plant litter mainly use extraction methods based on vegetation indices or changes in the plant litter. Few studies have conducted time series analyses of plant litter or considered the correlation between plant litter and soil erosion. In addition, social factors are not given enough consideration in the remote sensing and soil community. **Methods.** This study performs time series estimation of plant litter by integrating three-scale remotely sensed data and a random forest (RF) modeling algorithm. Predictive models are used to estimate the spatially explicit plant litter cover for the entire Loess Plateau over the last two decades (2000–2018). Then, the sediment concentration in the water was classified into 9 grades based on environmental and social-economic factors. **Results.** Our results demonstrate the effectiveness of the proposed predictive models at the regional scale. The areas with increased plant litter cover accounted for 67% of the total area, while the areas with decreased plant litter cover accounted for 33% of the total area. In addition, plant litter is demonstrated to be one of the top three factors contributing to the decrease in the river sediment concentration. Social-economic factors were also important for the decrease of the sediment concentration in the water, for example, the population of the rural area.

Plant litter estimation and its correlation with sediment concentration in the Loess Plateau

Qian Li[1], Ligang Ma[1,2,*], Suhong Liu[3,4], Adilai Wufu[1] , Yinbo Li[1], Shengtian Yang[1], Xiaodong Yang[1]

[1] College of Resource and Environmental Science, Xinjiang University, Urumqi 830046, Xinjiang, China
[2] General College Key Laboratory of Smart City and Environmental Modeling, Xinjiang University, Urumqi 830046, Xinjiang, China
[3] Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[4] Beijing Key Laboratory of Environmental Remote Sensing and Digital Cities, Beijing Normal University, Beijing 100875, China

Corresponding Author:
Ligang Ma [1]
No. 666 Shengli Street, Tianshan District, Urumqi, Xinjiang, 830046, China
Email address: mosanzju@126.com

keywords

multiscale; spatiotemporal; prediction; classification; upscaling

Abstract:

Background. Sediment concentration in the water of the loess Plateau region has dramatically decreased during the past two decades. Plant litter is considered to be one of the most important factors for this change. Existing remote sensing studies that focus on plant litter mainly use extraction methods based on vegetation indices or changes in the plant litter. Few studies have conducted time series analyses of plant litter or considered the correlation between plant litter and soil erosion. In addition, social factors are not given enough consideration in the remote sensing and soil community.

Methods. This study performs time series estimation of plant litter by integrating three-scale remotely sensed data and a random forest (RF) modeling algorithm. Predictive models are used to estimate the spatially explicit plant litter cover for the entire Loess Plateau over the last two decades (2000–2018). Then, the sediment concentration in the water was classified into 9 grades based on environmental and social-economic factors.

Results. Our results demonstrate the effectiveness of the proposed predictive models at the regional scale. The areas with increased plant litter cover accounted for 67% of the total area, while the areas with decreased plant litter cover accounted for 33% of the total area. In addition, plant litter is demonstrated to be one of the top three factors contributing to the decrease in the river sediment concentration. Social-economic factors were also important for the decrease of the sediment concentration in the water, for example, the population of the rural area.

1 Introduction

The practice of farming is believed to significantly accelerate the rate of soil erosion relative to soil production, and soil is lost at rates that are several orders of magnitude greater than those of the soil replenishment mechanisms (Ronald Amundson 2015). The Loess Plateau in China has long suffered from soil erosion for the same reason (Hao Wang 2018; Zhi Li 2012). The sediment load of the Yellow River is approximately 6% of the global river sediment load (Baoxian Tao 2018). Soils were lost at annual rates between 5000 to 10,000 Mg km−2yr−1 in most areas of the Loess Plateau (Bing Wang 2015). Further, the middle reaches of the Yellow River catchment are estimated to account for 90% of the sediment (Yang Zhao 2019). However, the mean annual sediment yield decreased by approximately 94% in comparison with the yield during the period from 1919-1959. Particularly, various catchments of the loess plateau have seen great decreases (approximately 53.0%–88.2%) in sediment loads of the same grade in extreme rainfall events after 2000 compared with those in the 1960s–1990s. In addition, the maximum sediment concentration in floodwaters when extreme rainfall events occur is consistently decreasing. For example, the maximum sediment concentration of the Wuding River in flood events has decreased by over 47.2% in this century (Yang Zhao 2019). Therefore, the river water in most catchments of the plateau has become much clearer than it used to be due to the decrease of the sediment concentration (Yang Zhao 2019). Many factors have contributed to the drastic reduction in river sediment content, such as the restoration of farmland to forests and grasslands; the declining rural labor force (due to urbanization), which has reduced the land use intensity; and climate change. However, the effects of changes in the land use policy and labor force on soil erosion are usually neglected.

Restoring farmland to forests and grassland shall increase plant litter cover, thus enhance soil and water conservation. The mechanism of soil conservation is normally considered to be the mechanical binding effects of plant litter that mitigated soil detachment by overland flows(Hao Wang 2019; Long Sun 2016). In addition, plant litter (mainly in the form of herbaceous plants) can prevent soil erosion in the form of maintaining soil moisture and partitioning rainfall as interception, infiltration, and lateral flows(Du et al. 2019; Xiangyang SUN 2013). However, the spatial pattern and change of plant litter in the entire Loess Plateau over the last 20 years, during which significant changes have occurred in the sediment concentration, remain unclear.

Remote sensing technique has long been used to monitor grasslands(Jan de Leeuw 2019), investigate spatial variations(Boyan Li 2018), and conduct time series analyses of photosynthetic vegetation (Christopher J. Watson 2019). To achieve the quantitative estimation of the non-photosynthetic vegetation coverage, most studies use vegetation indices, such as the NDI, NDTI, NDSVI, SACRI, MSACRI, DFI, STI, and SWIR32, that are derived from multispectral data (Wang Guangzhen 2018), and features that are derived from the gray-level cooccurrence matrix (GLCM)(Fernando Roberti de Siqueira 2013; Xun Zhou 2018). Indices that are derived from hyperspectral data are considered to be the most effective indices for the quantitative estimation of the non-photosynthetic vegetation coverage (XIE Xiao-yan 2016). In particular, the normalized difference lignin index (NDLI) and the cellulose absorption index (CAI) are considered to be the best indices for estimating the senescent grass biomass. However, the costs of hyperspectral data significantly increase with the amount of data. Moreover, the data are often unobtainable for grassland regions, unavailable for long-term research, and difficult to apply at large scales.

From the algorithm aspect, researchers usually use multivariate stepwise regressions (F.S. Peterson 2013; Sibel Taskinsu-Meydan 2010), convolutional neural networks (CNNs) (Xun Zhou 2018), and spectral mixture analysis (SMA)(Wang Guangzhen 2018). However, CNN-based methods rarely provide consistently high accuracy. In addition, the potential accuracy of large-scale modeling based on traditional methods is usually pessimistic. Overall, previous studies (i) are small in scale, (ii) do not elucidate how the plant litter cover changes in the entire plateau, (iii) lack time series analysis, or (iv) do not provide sufficient accuracy. The implementation of multiscale remote sensing is expected to solve these problems. In particular, the potential of using unmanned aerial vehicles (UAVs) data for regional scale modeling need to be examined.

In summary, the main purposes of this study are (1) to integrate multiscale remote sensing data for time series analysis of the plant litter in the entire plateau in order to determine the temporal and spatial dynamic changes and (2) to explore the relationship between plant litter and

sediment concentration change in order to determine the contribution of plant litter to sediment reduction in the Loess Plateau.

2 Materials

2.1 Study area

The Yellow River catchment originates from the Bayankala Mountains in Qinghai Province and flows into the Bohai Sea. It extends from 96°E to 119°E and from 32°N to 42°N, spanning 1900 km from east to west and 1100 km from north to south, covering a total area of 795,000 square kilometers. It flows through Qinghai, Sichuan, Gansu, Ningxia, Inner Mongolia, Shaanxi, Shanxi, Henan, Shandong, and nine other provinces. The entire river is divided into eight secondary watershed partitions: above Longyangxia, Longyangxia to Lanzhou, Lanzhou to Hekouzhen, Hekouzhen to Longmen, Longmen to Sanmenxia, Internal Flow Zone, Sanmenxia to Huayuankou, and below Huayuankou.

The terrain of the Yellow River catchment is high in the west and low in the east. The western headwater region has an average elevation of >4000 m. It consists of a series of high mountains. The central region has an elevation of 1000–2000 m, and it is characterized by yellow land and severe soil erosion. The eastern region has an elevation of no more than 100 m, and it mainly comprises an alluvial plain. The Yellow River catchment has a continental climate with a semi-humid climate in the southeast, a semiarid climate in the middle, and a dry climate in the northwest. The diverse landforms and complex habitats in the catchment create favorable conditions for the development of various vegetation types, and the main land use types are grassland, woodland, and agricultural land. The location of the study area is shown in Figure 1.

2.2 Dataset

2.2.1 Field data

Field data were collected by a Sony EXMOR 1/2.3 inch CMOS digital camera on board a Phantom 3 Professional Quad-Rotor intelligent UAV, which was manufactured by DJ-Innovations (DJI) incorporation. The total number of effective pixels in one photograph is 12 million with 4000 columns and 3000 rows. The field survey data of plant litter cover were collected from October 12–18, 2016 when the plants were in their mature period. We collected thirteen plots in this region, which were considered to be representative because of their diverse terrain and growth status.

In this study, the UAV recorded aerial images (Figure 2) every 3 s in the nadir view, the flying height was approximately 40 m at which the ground resolution of each photograph was approximately 7 cm. Thus, a single aerial photograph can cover a ground area of approximately 58800 m2 (280 m × 210 m). Approximately 10 photographs were taken of each sample plot (700 m × 700 m).

Data clustering algorithms are crucial for unsupervised classification, and high-density data analysis is of importance for resource management(Md Saifuzzaman 2019). The true-color UAV images were first classified into 30 classes using the iterative self-organizing (ISO) data analysis and clustered into binary litter/non-litter images. Then, the plant litter binary images were aggregated into a pixel size of 30 m. The percentage of plant litter for the 30-m pixels is deemed to be the fraction of the 7 cm pixels that are labeled as "plant litter" within each 30-m cell. It was calculated as the total number of 7-cm UAV plant litter pixels divided by the 30-m cell. This plant litter cover has been demonstrated to be more accurate than the average cover based on the traditional method(Jing Ge 2018). The corresponding parameters of those images can be seen in Table 1.

2.2.2 Landsat data

Landsat data were used in the first level of modeling. We download the Landsat data for the year 2016 when the UAV data were collected. The Landsat-8 Operational Land Imager (OLI) data were derived from the Geospatial Data Cloud (www.gscloud.cn) with a spatial resolution of 30 m. Scenes overlapping with the 13 UAV sample plots, path/row: 128/32, 127/32, 127/34, 128/34, 129/34, 129/35, 130/35, and 131/35, were selected. The reflectivity was calculated based on the Landsat DN value and then the tasseled cap transformation was conducted using six bands (red, green, blue, near infrared, shortwave infrared 1, and shortwave infrared. Finally, the green component and wetness component were selected as predictor variables.

2.2.3 Temperature data

Meteorological data were downloaded from the China Meteorological Science data sharing service network (Http://data.cma.cn/site/index.html).

There are 64 standard meteorological sites within the Yellow River catchment and approximately 40 standard meteorological sites near and outside the Yellow River catchment. The time range of the data (years) is from 2000 to 2018. To obtain a more widely coverage of temperature distribution, we used the Kriging spatial interpolation method based on the abovementioned 64 and 40 standard meteorological sites. Then, the temperature grid data were resampled to a 500-m resolution and projected to a projection of UTM 48N.

2.2.4 SRTM data

The SRTM data were derived from the NASA Shuttle Radar Topography Mission. We used the SRTMGL1 Global 1 arc second V003 DEM product which was downloaded from the NASA data system (http://earthdata.nasa.gov/). The data are in hgt format with a 30-m spatial resolution. A total of 181 images, with orbit numbers from N32E101 to N43E111, were acquired to characterize the elevation across the Yellow River catchment.

The processing of the SRTM data includes the following steps: (1) mosaic the 181 images to one raster dataset; (2) extract the regions of interest from the 30-m mosaicked images based on

the 13 UAV sample plots, and they serve as topographical feature variables in the first-level upscaling; and (3) resample the 30-m images to a 500-m resolution. This topographical feature variable was used in the second-level upscaling. The 181 SRTM images with a 30-m spatial resolution are illustrated in Figure 3.

2.2.5 MODIS data

The MODIS data were downloaded from NASA EOSDIS's data search portal (http://search.earthdata.nasa.gov), including the evapotranspiration (ET), the enhanced vegetation index (EVI), and the reflectance in the blue and green band. These data sources were used as variables in the second-level upscaling. Specifically, the following data were adopted: the MODIS/Terra Vegetation Indices 16-Day L3 Global 500 m SIN Grid V006 (MOD13A1), the MODIS/Terra Evapotranspiration 8-Day Level-4 Global 500 m SIN Grid (MOD16A2), and the MODIS/Terra Surface Reflectance 8-Day L3 Global 500 m SIN Grid (MOD09A1) in the blue and green bands.

We used the data for the 289th day of the year. Four scenes are required to cover the Yellow River catchment: h26v04, h25v05, h26v05, and h27v05. The time range (years) is from 2000 to 2018. The four tiles were mosaicked and projected to the UTM 48N projection and WGS 84 datum.

2.2.6 Sediment concentration and its explanatory variables

The factors affecting sediment concentration are typically classified into runoff, vegetation, soil, and topographical factors. In addition, human activities can also affect sediment concentrations. They include plant litter, vegetation type, population, and GDP. In this study, the population, soil type, vegetation type, and GDP data were downloaded from the Resource and Environment Data Cloud platform (http://www.resdc.cn/), while the runoff, sediment transport, and sediment concentration data were downloaded from the China River Sediment Bulletin (http://www.mwr.gov.cn/sj/tjgb/zghlnsgb/). As data on the GDP and population were available for only four years between 2000 and 2018 (2000, 2005, 2010, and 2015), they were analyzed only in this period.

3 Methods

We used the random forest (RF) regression models that are implemented in R to predict the fractional plant litter cover for the Yellow River catchment over the last two decades (2000–2018). The two-level regression modeling are: (1) upscaling of the plant litter cover from the UAV 0.07-m pixel scale to the 30-m Landsat pixel scale, and (2) upscaling of the plant litter cover from the 30-m Landsat pixel scale to the 500-m MODIS pixel scale.

We used the decision tree (DT) classification algorithm that was implemented in C5.0 for examining the relationship between sediment concentration in the water and those explanatory

variables. Decision tree is a non-parametric algorithm and has no special requirements for the sample distribution (Tianjun Wu 2018).

3.1 Random forest regression

Originally proposed by Breiman (2001), the random forest is an ensemble learning algorithm, which belongs to the bagging (bootstrap aggregating) type (Andrew Mellor 2017). By combining multiple weak classifiers, the final result is voted or averaged (L. Collins 2018). In each round of random sampling in Bagging, the data that are not collected are called OOB, which can be used to determine the generalization ability of the model so that the results of the overall model have higher accuracy and generalization performance. This approach can achieve good results, mainly due to being "random" and the "forest", which makes this algorithm resistant to overfitting and, thus, more precise. There are two important parameters in the random forest model: “ntree”, which stands for the number of decision trees, and “mtry”, for the number of predictor variables at each splitting node. We conducted the random forest modeling in R using the randomForest 4.6-14 package. The random forest can automatically find the optimal ntree and mtry based on the training dataset. In this study, ntree is set as 500, and mtry is set as 1.

The upscaling of plant litter cover from in situ measurements to the catchment scale was based on Random forest regression. Specifically, for the first-level upscaling, the plant litter cover from UAV images was the dependent variable, and DEM, wetness and the green component were the independent variables. For the second-level upscaling, the plant litter cover estimated from the first-level was the dependent variable, and temperature, DEM, EVI, ET, and the reflectance in the blue and green bands were the independent variables. The sample plots of second level upscaling are illustrated in Figure 2.

Variable importance was acquired based on the average contribution rate of each feature on each tree in the random forest. There are two common calculation methods. One is the mean decreased accuracy, which is measured by the change of the out of bag (OOB) error when adding random noise to the feature (Wayne S. Walker 2007). The other is the mean decreased impurity, which uses the gini/entropy/information gain indices. The “IncMSE” and “IncNodePurity” correspond to the above two methods, respectively. We used the IncMSE to measure the variable importance in this study. The variable importance (varImpPlot) analyses were conducted using the R package to assess the relative contributions of the predictor variables to the modeling process (Ranjeet John 2018).

The percentage of training and testing samples vary in the two- scale modeling. For the first scale, the percentage of training and testing were 80% and 20% respectively, while for the second scale, 50% and 50%.

3.2 Decision tree classification

The decision process of the decision tree is as follows: first, it starts from the root node and then tests the corresponding feature attributes in the item to be classified. Then, the output branch is selected according to its value until the leaf node is reached, and the category that is stored by the leaf node is used as the decision result.

Among the various DT platforms, C5.0 has been widely used in land cover classification and soil mapping. The See5.0/C5.0 machine learning rule software is a data mining tool that was developed by USGS for the National Land Cover Mapping (NLCD) project to automatically extract classification rules. The C5.0 model supports boosting, and the decreasing rate of the information entropy is used as the basis for determining the optimal branching variable and segmentation threshold(Tianjun Wu 2018). The C5.0 nonlinear model can handle multisource big data sets. For detailed principles of the C5.0 algorithm, the reader can refer to Quinlan (2001) and Jansson (2016). The website of the source code is http://www.rulequest.com/see5-info.html.

We established a classification model based on the grade of sediment concentration in the water and the social-environmental factors that affect the sediment. So, the following indices were calculated: plant litter coverage, total population, major vegetation type, major soil type, average elevation, and total GDP. The measured runoff and sediment concentration data from the six main hydrological control stations in the main streams are selected as the hydrological variables. Previous studies have demonstrated that sediment concentration can be classified into 16 grades(LI Xungui 2010). In this study nine grades were considered.

3.3 Accuracy assessment

To determine the model performance of RF, we used four indices: the root-mean-square error (RMSE), mean absolute error (MAE), coefficient of the variation of the root mean square error (CVRMSE), and the coe□cient of determination (R2). The definition of the former three were as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}[E(y_i) - y_i]^2}{n}} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(E(y_i) - y_i)| \quad (2)$$

$$CVRMSE(\%) = \sqrt{\frac{\sum_{i=1}^{n}[E(y_i) - y_i]^2}{n}} \times \frac{100}{\bar{y}} \quad (3)$$

where E(yi) and yi are the predicted and field measured values for sample point i, respectively, and n represents the sample size (n = 1353 for the first-level upscaling and n =110,916 for the second-level upscaling), $\bar{y}$ is the measured mean value;. Lower RMSE values represent higher accuracy and are generally desirable in verification (Jing Ge 2018; Julian D. Olden 2000).

When CVRMSE > 30%, the predictive ability of the model is poor; when CVRMSE is 20%–30%, the predictive ability is fair; when CVRMSE is 10%–20%, the predictive power is

good; and when CVRMSE < 10%, the predictive ability is excellent (Jie Pan 2006; Jing Ge 2018).

To determine the model performance of DT, a confusion matrix was constructed and Kappa value was derived. The confusion matrix is a comparison array that indicates the number of samples that are classified into a certain category and its field-measured categories. Typically, the columns in the array represent the field-measured category and the rows represent the category resulting from the classification. The calculation of Kappa value is as below:

$$\text{Kappa} = \frac{N\sum_{i=1}^{n} x_{ii} - \sum_{i=1}^{n}(x_{i+} x_{+i})}{N^2 - \sum_{i=1}^{n}(x_{i+} x_{+i})} \quad (4)$$

where n is the total number of columns in the confusion matrix, i.e., the number of categories; xii is the number of samples in the i-th row and i-th column in the confusion matrix, i.e., the number of correct classifications; xi+ and x+i are the numbers of samples in the i-th row and i-th column, respectively; and N is the total number of samples.

3.4 Change detection

The plant litter coverage in the Yellow River catchment was estimated for the last 19 years (2000 to 2018). The average of the 19-year plant litter coverage was used to imply the spatial distribution.

The time-varying characteristics of the plant litter cover during 2000-2018 are represented by using a linear slope model (Jing Ge 2018; Rasmus Fensholt 2009). This model is simple and robust, and the dynamic of each pixel is calculated by using the slope function(Haidi Zhao 2015; Jing Ge 2018), as shown in function (5). When the slope is greater than 0, it indicates an upward trend, and when it is less than 0, it indicates a downward trend.

$$\text{slope} = \frac{n \times \sum_{i=1}^{n} i \times cover_i - \sum_{i=1}^{n} i \sum_{i-1}^{n} cover_i}{n \times \sum_{i=1}^{n} i^2 - (\sum_{i=1}^{n} i)^2} \quad (5)$$

where n represents the total number of years; i ranges from 1-19, corresponding to the years from 2000 to 2018; and coveri represents the plant litter coverage of the i-th year.

Based on the F test, this study further analyzed the significance level of the change of plant litter coverage. When slope > 0%/yr and F > 3.03, it follows a significant upward trend. When slope > 0%/yr and F < 3.03, it follows a general upward trend. When slope < 0%/yr and F < 3.03, it follows a general downward trend. When slope < 0%/yr and F > 3.03, it follows a significant downward trend. The F test is defined as:

$$F = \frac{r^2(n-2)}{1-r^2} \quad (6)$$

$$r = \frac{\sum_{i=1}^{n}(i - \bar{i})(cover_i - \overline{cover})}{\sqrt{\sum_{i=1}^{n}(i - \bar{i})^2 \sum_{i=1}^{n}(cover_i - \overline{cover})^2}} \quad (7)$$

where n represents the total number of years; r is the correlation coefficient between the plant litter cover and the time series; $\bar{i}$ represents the average of 1-19, which is 10 in this study; coveri represents plant litter coverage of the i-th year; and $\overline{cover}$ represents the average plant litter coverage during 2000-2018. The 90% confidence value $F0.1(1, n-2) = F0.1(1,17) = 3.03$ (Jing Ge 2018; Yi Song 2010).

4 Results

4.1 Model accuracy

For the first-level upscaling, the R2 was 0.37 and P-value was 4.35e-138. Though the R square is not high, the P-value indicate the significance of this model at 0.001 confidence level. What's more, the RMSE for plant litter was 18.455% which is acceptable. For the second-level upscaling, the R2 for plant litter increased markedly to 0.891 while RMSE decreased markedly to 2.936% (Table 2). The second-level upscaling had higher model precision probably due to the much larger sample size used for training the model. The model was further evaluated using the scatterplots of the predicted and observed plant litter cover. As shown in Figure 4, the slope of the two model is 0.97 and 1.03, respectively, very close to 1:1 straight line. Compare the two model, it can be seen that the scatter agglomeration and quantity of the second-level model is better than that of the first-level model.

The VIPs for the two levels of upscaling are shown in Figure 5. In the first-level model, DEM is of the greatest importance, greenness and wetness were of equal importance, slightly lower than that of dem. This shows that among the three independent variables, the topography plays a decisive role in the distribution of plant litter. Similarly, in the second level of modeling, DEM also has the greatest importance, EVI is of second importance, and the importance of other independent variables is roughly equivalent, with lower importance.

4.2 Spatial pattern of plant litter cover

From the spatial aspect, it can be seen that the plant litter coverin the upper reaches of the Yellow River catchment is higher than that in the downstream area which is coincide with the allocation of water (precipitation and runoff). Specifically, the high covering areas of plant litter are located in: above Longyangxia, Longyangxia to Lanzhou, south area of Lanzhou to Hekouzhen, west area of Longmen to Sanmenxia, and Internal Flow Zone. In contrast, the low covering areas of plant litter are located in: Hekouzhen to Longmen, east area of Longmen to Sanmenxia, and Sanmenxia to Huayuankou. The lowest plant litter cover is found in the north area of Lanzhou to Hekouzhen and below Huayuankou.

4.3 Time series analysis of plant litter cover

The slope distribution map of the plant litter cover is illustrated in Figure 7. The plant litter coverin most areas is increasing, especially in the Qingshui River and Kushui River basins, Lanzhou to Xiahe Yan basin, Jinghe River basin. However, the plant litter cover is decreasing along the north bank of Shizuishan to Hekou Town.

We calculated the spatial distribution of the F results, as shown in Figure 8, and then combined the slope and F results to make a significant division (see Figure 9) of the change of plat litter. It can be seen that the area where the plant litter covershowed a significant increasing trend accounts for 1.683% of the total area and is located in the middle of Lanzhou to Hekouzhen, Hekouzhen to Longmen, and Longmen to Sanmenxia regions. The plant litter cover in most regions shows an increasing trend, accounting for 65.466% of the total area. However, in the regions north of Longyangxia, north and south of Longyangxia to Lanzhou, north of Lanzhou to Hekouzhen, and east and south of Longmen to Sanmenxia, Sanmenxia to Huayuankou, and the Internal Flow Zone, the plant litter cover showed a decreasing trend, accounting for 32.765% of the Yellow River catchment. Further, the cover of dead branches and deciduous leaves showed a significant decreasing trend in only 0.086% of the area. Overall, the area with an increase in the plant litter cover accounted for 67% of the Yellow River catchment, while in the remaining 33% of the catchment, the plant litter coverdecreased.

From 2000 to 2018, the plant litter cover in the Yellow River catchment increased as a whole, which is consistent with the actual situation. In recent decades, considerable effort has been devoted toward restoring the vegetation to reduce the soil and water loss in the Loess Plateau. Plant litter increases as the green vegetation increases. With the implementation of the "Three North Protection Forest", "Returning Farmland to Forests and Grasses" and other ecological civilization projects(Jing Ge 2018; Kenneth Michael Bauer 2011), more than 2 Mha of sloping arable land on the Loess Plateau had been converted into forestland or grassland by 2012(Lei Deng 2014; Long Sun 2016). From 2000 to 2008, the main land use/cover change in the Loess Plateau was transformation from cultivated land to forest land and grassland(ZHOU Shugui 2016).

4.4 Response of plant litter to sediment concentration

The changes in runoff, sediment transport, and sediment concentration from 2000 to 2018 are shown in Figure 10. From 2000 to 2018, the runoff increased while the sediment transport and sediment concentration decreased. Previous studies have shown that from 2000 to 2015, the rainfall in the Yellow River catchment has increased (TIAN Zhihui 2019). This implies that the increase in the runoff is consistent with the growth trend of the precipitation.

The structure of DT model is illustrated in Figure 11. Seven decision rules were extracted in this study (as shown in the Supplementary Data). The classification results indicate that the

overall classification accuracy was 81.5%, and the kappa coefficient was 0.775 (see Table 3). The importance of the classification feature variable is illustrated in Figure 12. The soil type, vegetation type, and plant litter cover are the top three factors contributing to the reduction of sediment concentration. The importance of DEM was 0%, probably because the DEM is the topographical mean value of the secondary subcatchment and its contribution to the classification is not as significant as that of the other six variables.

5 Discussion

5.1 Regression and classification consideration

It should be noted that the core issue of this study is to examine the response of change of plant litter on sediment concentration in the water. Considering the shortage of large scale information of plant litter, it was estimated using RF regression model prior to response examining. In addition, we included UAV data and two-step upscaling method to improve prediction accuracy. Cubist model can be used to estimate soil organic carbon content assessment(Shangshi Liu 2019), but over-fitting occurs when modeling plant litter cover.

For another aspect, the reduction of sediment concentration in the water can be attributed by many environmental and anthropogenic factors. The mechanism of their relationship is rather complex and in most cases non-linear. The classification of the degree of sediment concentration in the water by means of both environmental and anthropogenic factors is a preliminary trial for this task. The result is very optimistic indicated by accuracy indices of classification, though those sample regions still need to be expanded.

As for those factors, anthropogenic factors (GDP and population) were demonstrated to be very important, but not as important as those of the top three. It can be hypothesized that environmental factors might be the control of the reduction of sediment concentration in the water in the entire plateau scale. While the urbanization process in this region could be the control of the reduction of sediment concentration in a relatively small scale and a specific sub-catchment. The above hypothesis has been partly demonstrated by our further classification test in which the original nine classes were reclassified into six grades according to the histogram of the data.

5.2 Accuracy comparison with other models

Through two-level regression modeling established by random forest, this paper derived plant litter cover at 30-m Landsat pixel scale and 500-m MODIS scale. For the first scale, the percentage of training and testing were 80% and 20% respectively according to the statistical significance and previous article (Jing Ge 2018), while for the second scale, 50% and 50% due to large size of sample.

Previous studies have shown that the maximum annual grassland coverage of the Yellow River headwater region increased from 2001 to 2016(Jing Ge 2018), and the Net Primary Productivity (NPP) of the Yellow River catchment showed a slightly fluctuating increasing trend overall from 2000 to 2015 (TIAN Zhihui 2019), which is consistent with the increasing trend of the plant litter coverage in this study. However, the accuracy is slightly lower than that of the two abovementioned cases because the study area is extremely large, and the UAV data are not abundant, which implies that the model accuracy was limited by the number of UAV sample and its size with respect to the entire Yellow River catchment. The UAV sample coverage is approximately 1-5 pixels at the MODIS scale (500 m). However, the prediction results are consistent with the relevant research, which further demonstrate the feasibility of the proposed method. With the improvement of the spatiotemporal and spectral resolutions of remote sensing images, using remote sensing images to infer the plant litter cover is a more efficient and accurate method for large-scale dynamic monitoring. The CAI that is derived from hyperspectral images is applicable to arid and semiarid areas with relatively uniform vegetation types and soil textures and low vegetation coverage (Wang Guangzhen 2018). In addition, synthetic aperture radar (SAR) and light detection and ranging (LiDAR) techniques can obtain vertical structural information with high precision, regardless of the weather conditions, and they can estimate the fractional non-photosynthetic vegetation with high accuracy (Baojuan Zheng 2014).

5.3 Potential drivers for plant litter cover trend

Increased greenness may partly be due to the effect of climate warming and humidification (Jing Ge 2018; Yurong Hu 2012). And partly due to human activities. The major factor is different over different regions.

From the previous article on the temporal and spatial variation of precipitation over the Yellow River Basin (Yanli Liu & Wang 2019), it can be seen that during the period from 1961 to 2015, the total annual precipitation in the south of Longyangxia-Lanzhou, the southwestern part of Longmen-Sanmenxia, and north of Sanmenxia-Huayuankou showed a decreasing trend, the same as the plant litter cover's decreasing trend from 2000 to 2018, indicating that precipitation reduction plays a major role in the reduction of plant litter. However, the total annual precipitation in the north of Longyangxia-Lanzhou and above Longyangxia showed a increasing trend from 1961 to 2015, while in these areas, the plant litter cover has decreased from 2000 to 2018, indicating that the plant litter cover is mainly affected by human activities, returning farmland to forests and grasses is the major factor for plant litter reduction, the litter of forest and grassland is generally less than that of crops, the natural plants might be sparse, and the artificial planting can be more dense, so after returning farmland to forests and grass, the plant litter will be reduced. During the period from 1961 to 2015, in the central area of Hekouzhen-Longmen, the total annual precipitation showed an increasing trend, which was

consistent with the increasing trend of plant litter, indicating that climate change is the main factor for the increase of plant litter.

6 Conclusions

Based on multiscale remote sensing and random forest, this paper analyzed the temporal-spatial distribution characteristics of the plant litter in the Yellow River catchment and analyzed the effects of the plant litter cover on the sediment concentration based on C5.0 classification. This paper confirms that multiscale remote sensing images used with the random forest algorithm present a feasible method to upscale the ground observations of plant litter cover up to the regional scale. The plant litter cover in the upper and middle reaches of the Yellow River catchment were higher than that in the downstream area from 2000 to 2018. The plant litter cover in most regions showed an increasing trend. Thus, the policy of restoring farmland to forests and grassland yielded positive results.

In the Yellow River catchment, the runoff increased while the sediment concentration and sediment transport decreased from 2000 to 2018. The increase in the runoff is consistent with the growth trend of the precipitation over the last two decades. In summary, the soil type, vegetation type, and plant litter cover are the top three factors contributing to the sediment concentration reduction. Our findings can serve as a reference for the department of water resource management in terms of reducing the sediment concentration and obtaining clearer water.

Directions for future work include the evaluation plant litter cover by using hyperspectral, SAR, or LIDAR data. Moreover, besides the population and GDP, other human factors affecting the sediment concentration should be considered. Climate change is also one important environmental factor to be considered in the quantitative analysis of the sediment concentration. In addition, the study of basin-scale vegetation-hydrological models is also necessary for water resource utilization and ecological balance on the Loess Plateau.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2016YFC0402409) and the National Natural Science Foundation of China (41661079, U1603241, 41771470). The authors are also very grateful for the detailed and constructive suggestions that improved the manuscript from the editor and anonymous reviewers.

Conflict of Interest Statement

No potential conflict of interest was reported by the authors.

7 Reference

Andrew Mellor SB. 2017. Exploring diversity in ensemble classification: Applications in large

area land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* 129:151-161. 10.1016/j.isprsjprs.2017.04.017

Baojuan Zheng JBC, Guy Serbin, John M. Galbraith. 2014. Remote sensing of crop residue and tillage practices: Present capabilities and future prospects. *Soil & Tillage Research* 138:26-34. 10.1016/j.still.2013.12.009

Baoxian Tao YW, Yan Yu, Qinze Li, Chunyu Luo, Baohua Zhang. 2018. Interactive effects of nitrogen forms and temperature on soil organic carbon decomposition in the coastal wetland of the Yellow River Delta, China. *Catena* 165:408-413. 10.1016/j.catena.2018.02.025

Bing Wang G-HZ, Yang-Yang Shi, Zhen-Wei Li, Zhe-Jie Shan. 2015. Effects of Near Soil Surface Characteristics on the soil Detachment Process in a Chronological series of Vegetation Restoration. *Soil & Water Management & Conservation*:1213-1222. 10.2136/sssaj2015.03.0120

Boyan Li WW, Liang Bai, Nengcheng Chen & Wei Wang. 2018. Estimation of aboveground vegetation biomass based on Landsat-8 OLI satellite images in the Guanzhong Basin, China. *International Journal of Remote Sensing* 40:3927-3947. 10.1080/01431161.2018.1553323

Christopher J. Watson NR-CaARH. 2019. Multi-Scale Phenology of Temperate Grasslands: Improving Monitoring and Management With Near-Surface Phenocams. *Frontiers in Environmental Science* 7:1-18. 10.3389/fenvs.2019.00014

Du J, Niu J, Gao Z, Chen X, Zhang L, Li X, van Doorn NS, Luo Z, and Zhu Z. 2019. Effects of rainfall intensity and slope on interception and precipitation partitioning by forest litter layer. *Catena* 172:711-718. 10.1016/j.catena.2018.09.036

F.S. Peterson JS, K. Lajtha. 2013. Scaling litter fall in complex terrain: A study from the western Cascades Range, Oregon. *Forest Ecology and Management* 306:118-127. 10.1016/j.foreco.2013.05.052

Fernando Roberti de Siqueira WRS, Helio Pedrini 2013. Multi-scale gray level co-occurrence matrices for texture description. *Neurocomputing* 120:336-345. 10.1016/j.neucom.2012.09.042

Haidi Zhao SL, Shikui Dong, Xukun Su, Xuexia Wang, Xiaoyu Wu,Lei Wu and Xiang Zhang. 2015. Analysis of vegetation change associated with human disturbance using MODIS data on the rangelands of the Qinghai-Tibet Plateau. *Rangeland Journal* 37:77-87. 10.1071/RJ14061

Hao Wang G-hZ, Ning-ning Li,Bao-jun Zhang, Han-yue Yang. 2018. Soil erodibility influenced by natural restoration time of abandoned farmland on the Loess Plateau of China. *Geoderma* 325:18-27. 10.1016/j.geoderma.2018.03.037

Hao Wang GZ, Ning-ning Li, Bao-jun Zhang, Han-yue Yang. 2019. Variation in soil erodibility under five typical land uses in a small watershed on the Loess Plateau, China. *Catena* 174:24-35. 10.1016/j.catena.2018.11.003

Jan de Leeuw AR, Elmaddin Namazov, Emil Bayramov, Michael T.Marshalld,Jonathan Etzold, Regina Neudert. 2019. Application of the MODIS MOD 17 Net Primary Production

product in grassland carrying capacity assessment. *International Journal of Applied Earth Observation and Geoinformation* 78:66-76. 10.1016/j.jag.2018.09.014

Jie Pan YZ, Weixing Cao, Tingbo Dai, Dong Jiang. 2006. Predicting the protein content of grain in winter wheat with meteorological and genotypic factors. *Plant Production Science* 9:323-333. 10.1626/pps.9.323

Jing Ge BM, Tiangang Liang,Qisheng Feng,Jinlong Gao,Shuxia Yang,Xiaodong Huanga, Hongjie Xie. 2018. Modeling alpine grassland cover based on MODIS data and support vector machine regression in the headwater region of the Huanghe River, China. *Remote Sensing of Environment* 218:162-173. 10.1016/j.rse.2018.09.019

Julian D. Olden DAJ. 2000. Torturing data for the sake of generality: How valid are our regression models? *Ecoscience* 7:501-510. 10.1080/11956860.2000.11682622

Kenneth Michael Bauer DC. 2011. Laws and Regulations Impacting the Enclosure Movement on the Tibetan Plateau of China. *Himalaya* 30:23-37.

L. Collins PG, G. Newell, A. Mellor. 2018. The utility of Random Forests for wildfire severity mapping. *Remote Sensing of Environment* 216:374-384. 10.1016/j.rse.2018.07.005

Lei Deng ZZ, Zhouping Shangguan. 2014. Long-term fencing effects on plant diversity and soil properties in China. *Soil & Tillage Research* 137:7-15. 10.1016/j.still.2013.11.002

LI Xungui WN, WEI Xia. 2010. *A Maximum Rating Method for Determination of Abandoned Floodwater in Hyper-concentration Rivers*: Resources Science.

Long Sun G-hZ, Fa Liu, Li-li Luan. 2016. Effects of incorporated plant litter on soil resistance to flowing water erosion in the Loess Plateau of China. *Biosystems Engineering* 147:238-247. 10.1016/j.biosystemseng.2016.04.017

Md Saifuzzaman VA, Roberto Buelvas, Asim Biswas, Shiv Prasher, Shiv, Nicole Rabe, Doug Aspinall, Wenjun Ji 2019. Clustering Tools for Integration of Satellite Remote Sensing Imagery and Proximal Soil Sensing Data. *Remote Sensing* 11. 10.3390/rs11091036

Ranjeet John JC, Vincenzo Giannico,Hogeun Park, Jingfeng Xiao,Gabriela Shirkey, Zutao Ouyang,Changliang Shao, Raffaele Lafortezza, Jiaguo Qi. 2018. Grassland canopy cover and aboveground biomass in Mongolia and Inner Mongolia: Spatiotemporal estimates and controlling factors. *Remote Sensing of Environment* 213:34-48. 10.1016/j.rse.2018.05.002

Rasmus Fensholt KR, Thomas Theis Nielsen, Cheikh Mbow. 2009. Evaluation of earth observation based long term vegetation trends - Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. *Remote Sensing of Environment* 113:1886-1898. 10.1016/j.rse.2009.04.004

Ronald Amundson AAB, Jan W. Hopmans, Carolyn Olson,A. Ester Sztein, Donald L. Sparks. 2015. Soil and human security in the 21st century. *Science* 348:1261071. 10.1126/science.1261071

Shangshi Liu HS, Songchao Chen, Xia Zhao, Asim Biswas, Xiaolin Jia, Zhou Shi, Jingyun Fang. 2019. Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. *Geoderma* 348:37-44. 10.1016/j.geoderma.2019.04.003

559 Sibel Taskinsu-Meydan FE, Suha Berberoglu & Cenk Donmez. 2010. Modeling above-ground
560 litterfall in eastern Mediterranean conifer forests using fractional tree cover, and remotely
561 sensed and ground data. *Applied Vegetation Science* 13:485-497. 10.1111/j.1654-
562 109X.2010.01088.x
563 TIAN Zhihui ZD, HE Xiaohui, GUO Hengliang, WEI Haitao. 2019. Spatiotemporal Variations
564 in Vegetation Net Primary Productivity and their Driving Factors in Yellow River Basin
565 from 2000 to 2015. *Research of Soil and Water Conservation* 26:255-262.
566 Tianjun Wu WD, Jiancheng Luo,Yingwei Sun,Qiting Huang, Weizhi Wu,Xiaodong Hu. 2018.
567 Geo-parcel-based geographical thematic mapping using C5.0 decision tree: a case study
568 of evaluating sugarcane planting suitability. *Earth Science Informatics* 12:57-70.
569 10.1007/s12145-018-0360-8
570 Wang Guangzhen WJ, Zou Xueyong, Han Liu, Zong Min. 2018. A Review on Estimating
571 Fractional Cover of Non-photosynthetic Vegetation by Using Remote Sensing. *REMOTE*
572 *SENSING TECHNOLOGY AND APPLICATION* 33:1-9. 10.11873/j.issn.1004-
573 0323.2018.0001
574 Wayne S. Walker JMK, Elizabeth LaPoint,Michael Hoppus, James Westfall. 2007. An empirical
575 InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of*
576 *Environment* 109:482-499. 10.1016/j.rse.2007.02.001
577 Xiangyang SUN GW, Yun LIN, Linan LIU, Yang GAO. 2013. Intercepted rainfall in Abies fabri
578 forest with different-aged stands in southwestern China. *Turkish Journal of Agriculture*
579 *and Forestry*:495-504. 10.3906/tar-1207-36
580 XIE Xiao-yan LY-m, LI Jing-zhong,CHANG Wei,WANG Ling. 2016. Remote Sensing
581 Estimation of Plant Litter Cover Based on the Spectra of Plant Litter-Soil Mixed Scenes.
582 *Spectroscopy and Spectral Analysis* 36:2217-2223. 10.3964/j.issn.1000-0593(2016)07-
583 2217-07
584 Xun Zhou ZS, Suhong Liu, Peixin Yu, Xikai Wang & Yuebin Wang. 2018. A method for
585 extracting the leaf litter distribution area in forest using chip feature. *International*
586 *Journal of Remote Sensing* 39:5310-5329. 10.1080/01431161.2018.1484965
587 Yang Zhao WC, Chunhong Hu, Yousheng Wang, Zhaoyan Wang,Xiaoming Zhang, Bisheng Zhu,
588 Chen Cheng, Xiaolin Yin, Bing Liu, Gang Xie. 2019. Analysis of changes in
589 characteristics of flood and sediment yield in typical basins of the Yellow River under
590 extreme rainfall events. *Catena* 177:31-40. 10.1016/j.catena.2019.02.001
591 Yanli Liu XT, ouliang Sun, Jianyun Zhang, Guoqing Wang, Junliang Jin,, and Wang G.
592 2019. spatiotemporal precipitation variability and potential drivers during 1961-2015
593 over the yellow river basin. *Weather* 99:1-8. 10.1002/wea.3517
594 Yi Song MM, Frank Veroustraete. 2010. Comparison and conversion of AVHRR GIMMS and
595 SPOT VEGETATION NDVI data in China. *International Journal of Remote Sensing*
596 31:2377-2392. 10.1080/01431160903002409
597 Yurong Hu SM, Stefan Uhlenbrook. 2012. Trends in temperature and rainfall extremes in the
598 Yellow River source region, China. *Climatic Change* 110:403-429. 10.1007/s10584-011-
599 0056-2

600 Zhi Li F-LZ, Wen-Zhao Liu, De-Juan Jiang. 2012. Spatially downscaling GCMs outputs to
601 project changes in extreme precipitation and temperature events on the Loess Plateau of
602 China during the 21st Century. *Global and Planetary Change* 82-83:65-73.
603 10.1016/j.gloplacha.2011.11.008
604 ZHOU Shugui SQ, CAO Wei. 2016. Characteristics of Land Use and Land Cover Change in the
605 Loess Plateau over the Past 20 Years. *Journal of Geo-information Science* 18.
606 10.3724/SP.J.1047.2016.00190

607

Figure 1

Figure 1 . Digital elevation model (DEM) and locations of UAV sample plots of plant litter cover in October 2016. The two denser sample areas are enlarged (red boxes) to clearly show the sample plots. The sample plot range of the second-level upscaling

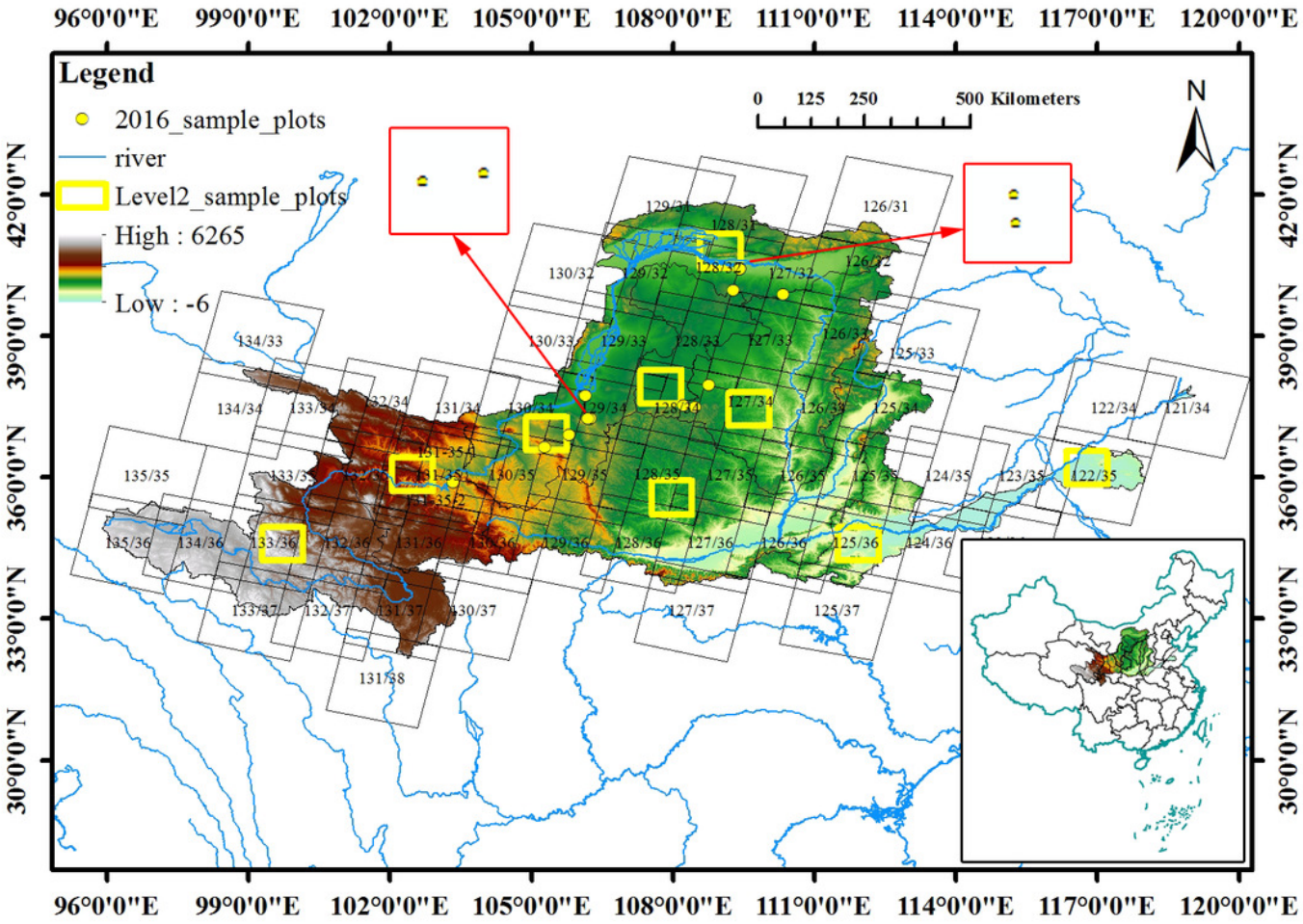

Figure 2

Figure 2 . (a) RGB true color UAV image; (b) Classification result for 30 classes (23 categories were omitted in the legend due to the size of the image); (c) Plant litter binarization result (1 stands for plant litter and 0 stands for not plant litter)

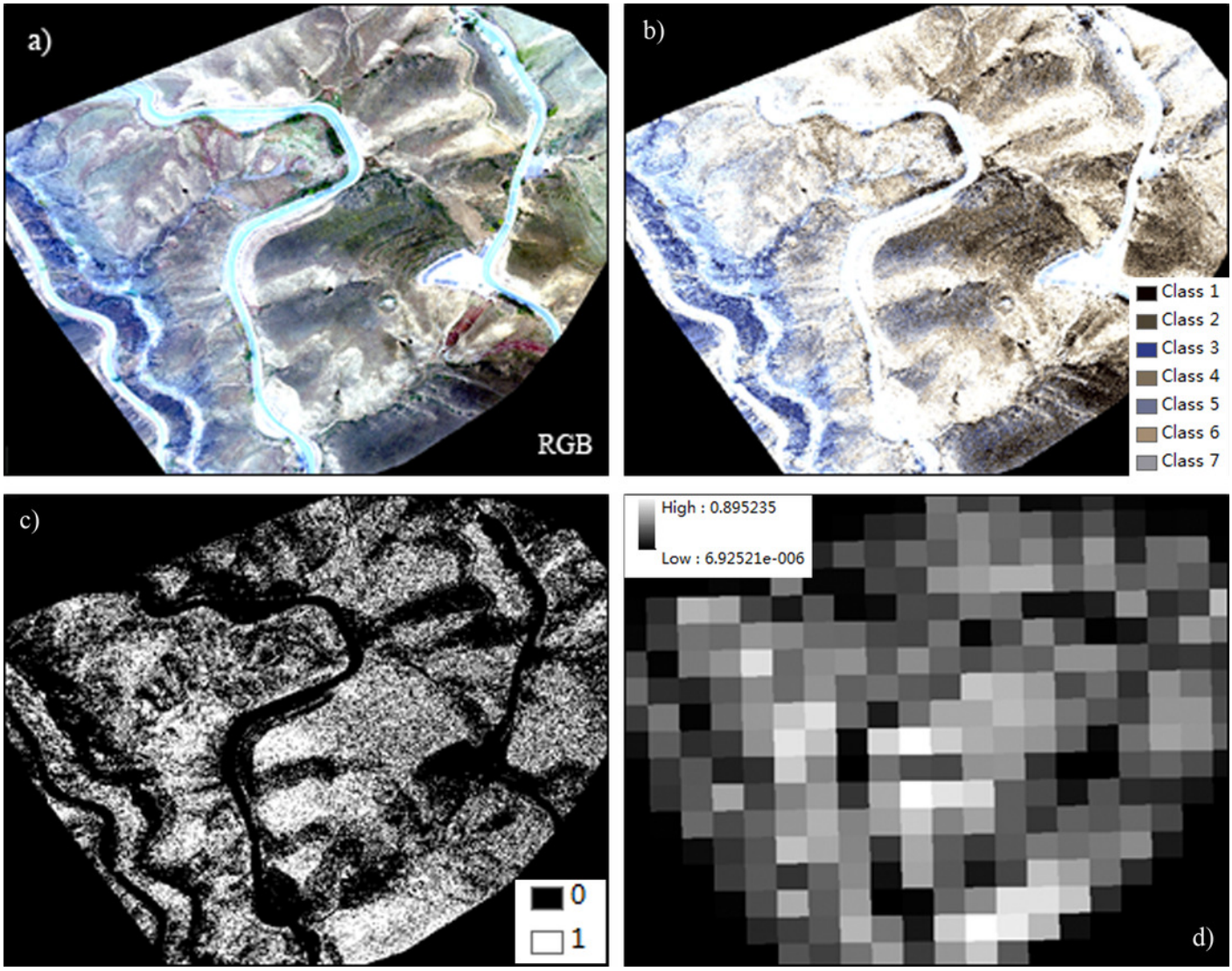

Figure 3

Figure 3 . Total of 181 SRTM images with a 30-m spatial resolution.

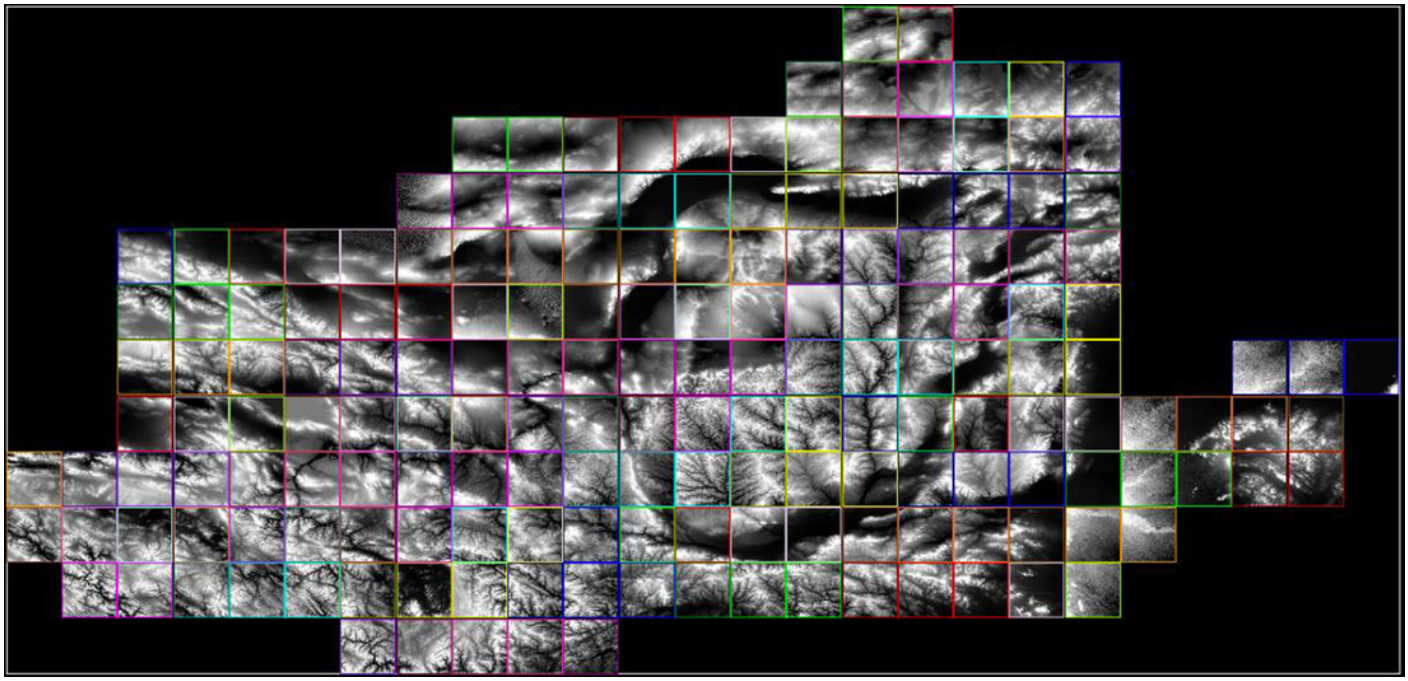

Figure 4

Figure 4 . Comparison of the plant litter covers that are estimated using the RF approach with reference plant litter cover for the (a) first-level upscaling and (b) second-level upscaling. Note: All models meet the condition of P < 0.001.

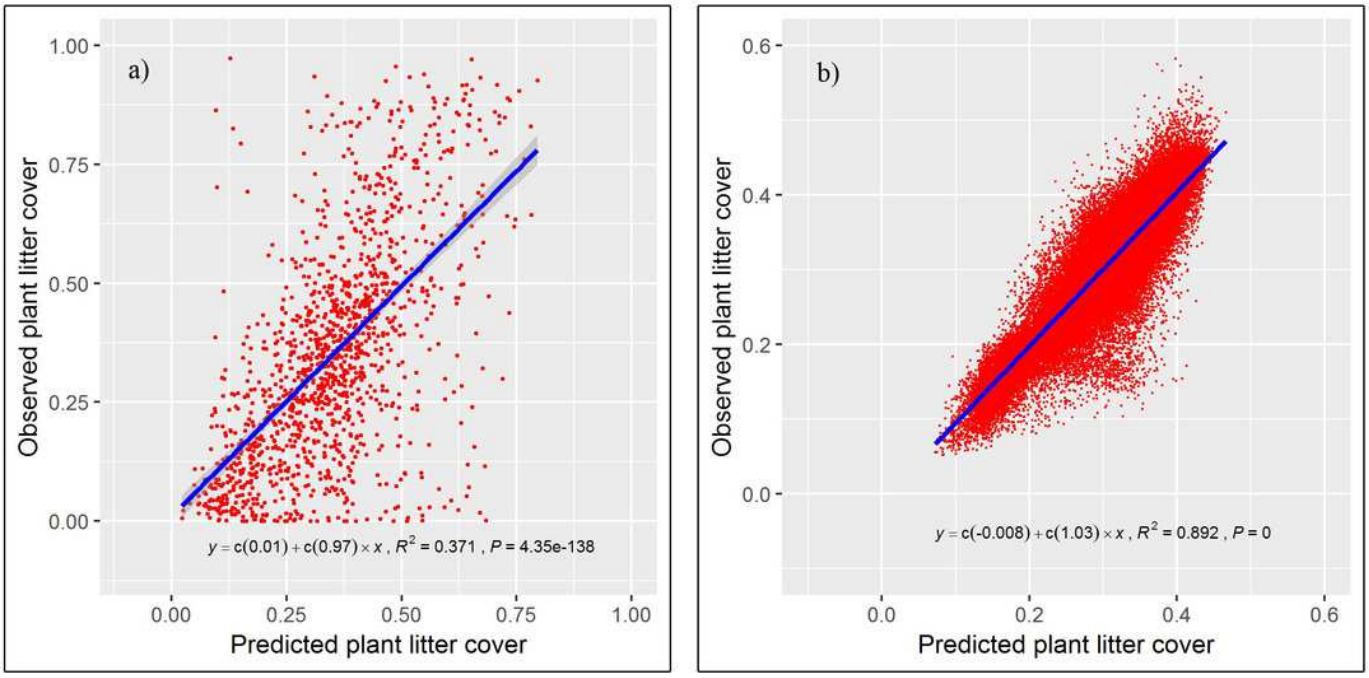

Figure 5

Figure 5 . VIPs that are generated in the random forest for the (a) first-level upscaling and (b) second-level upscaling (IncMSE stands for increasing MSE, B3ref stands for band3 reflectance, and B4ref stands for band4 reflectance. Band3 is blue and band

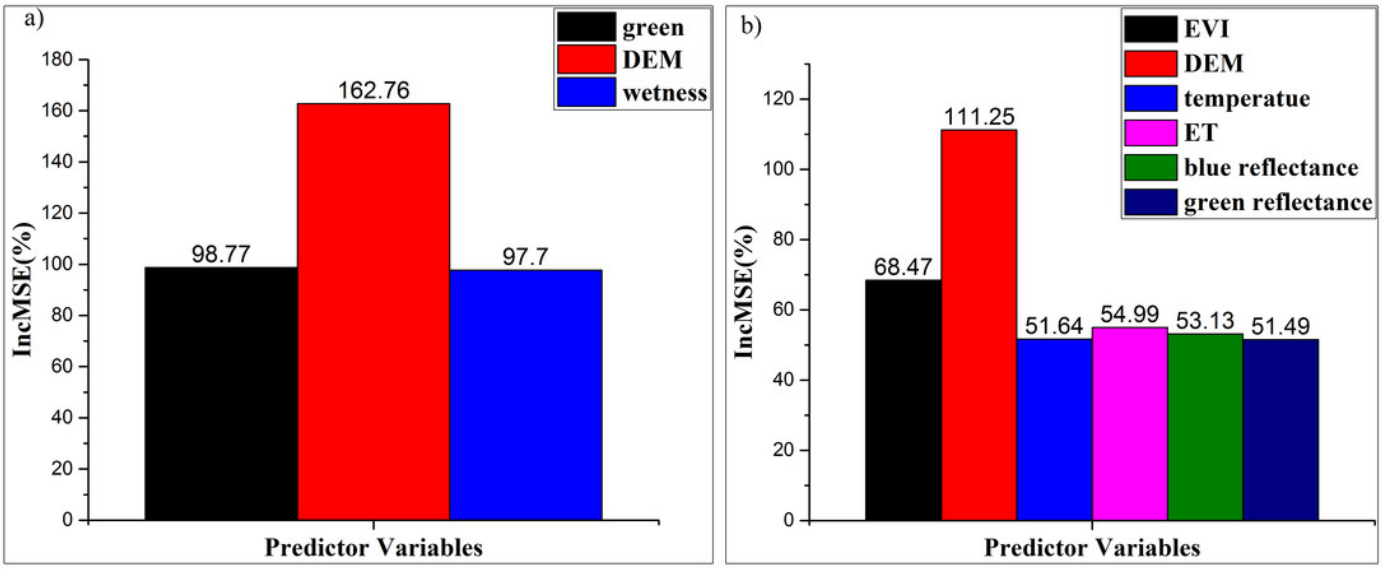

Figure 6

Figure 6 . Spatial distribution map of the average annual plant litter cover at the 500 m pixel scale in the Yellow River catchment from 2000 to 2018.

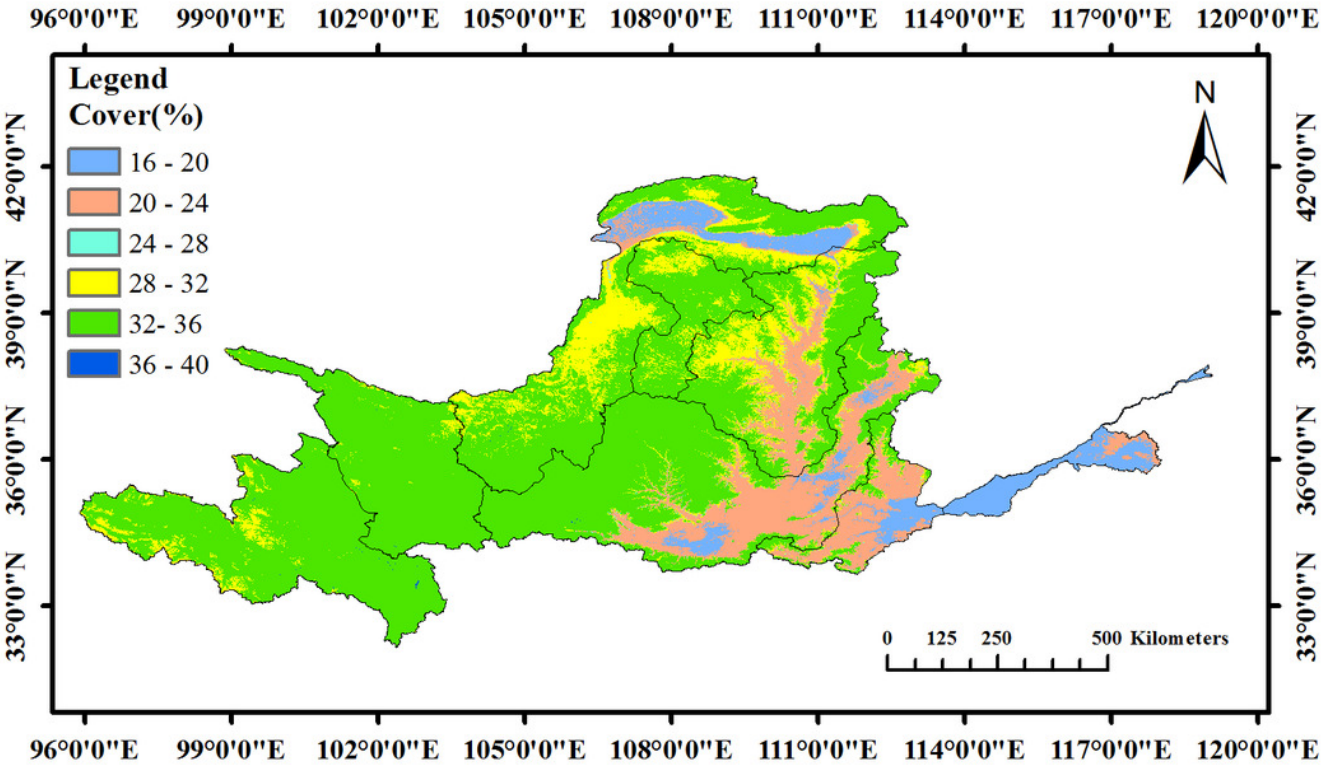

Figure 7

Figure 7 . Change trend of the annual plant litter cover at the 500 m pixel scale in the Yellow River catchment from 2000 to 2018.

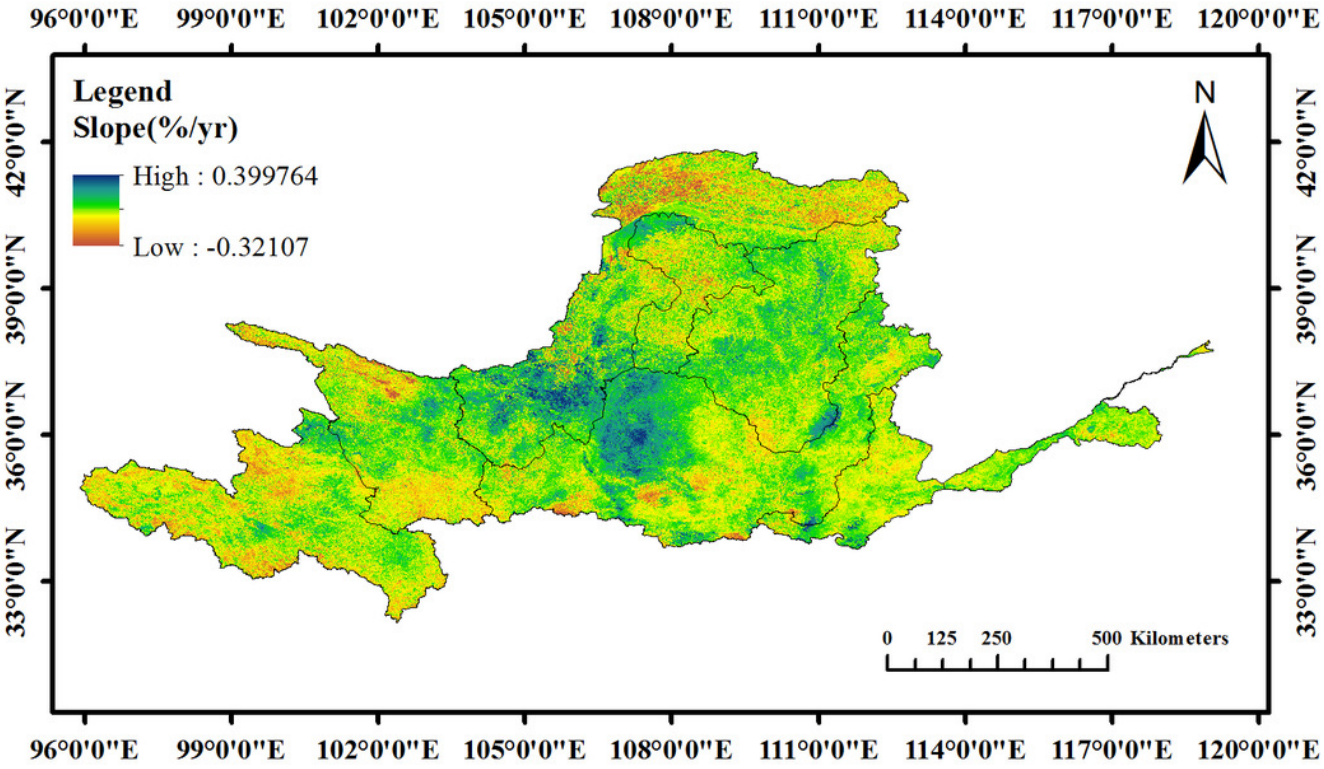

Figure 8

Figure 8 . F test result of the annual plant litter cover at the 500 m pixel scale in the Yellow River catchment from 2000 to 2018.

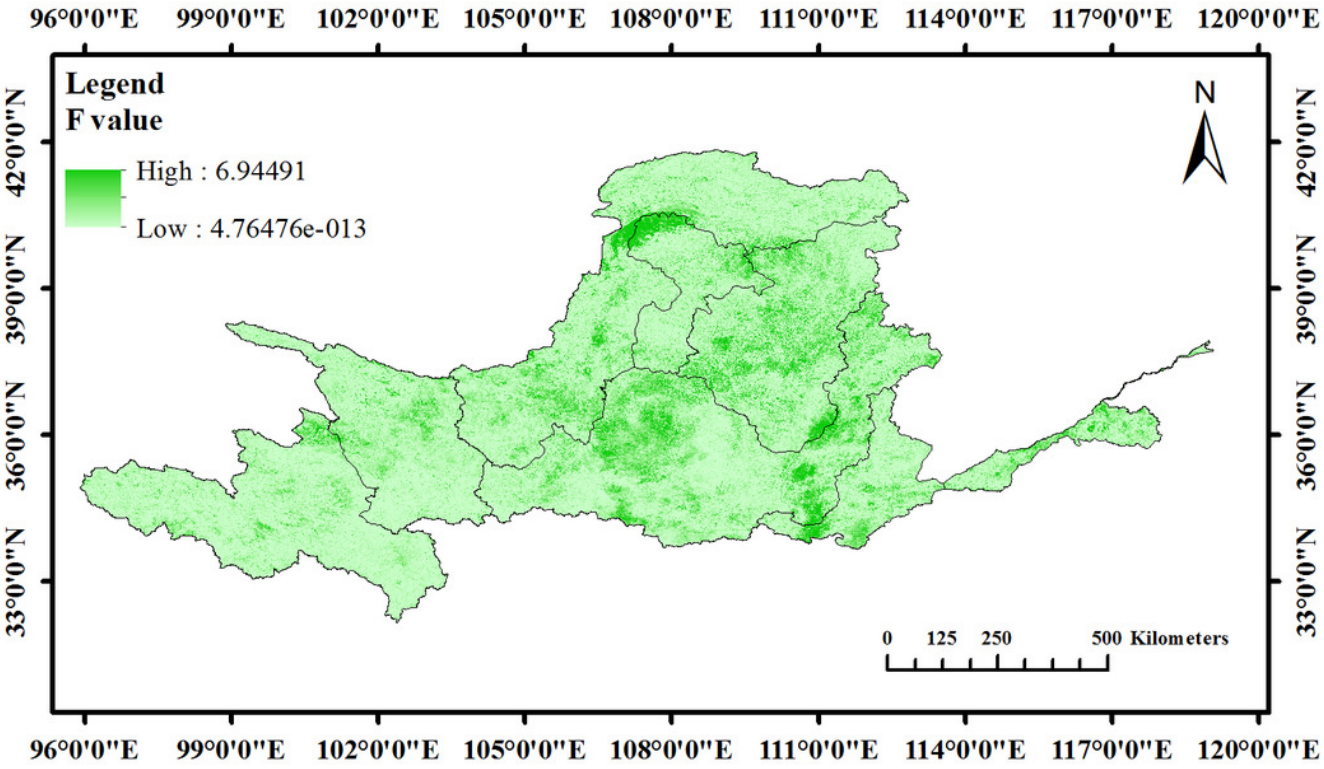

Figure 9

Figure 9 . Change grade of the annual plant litter cover at the 500 m pixel scale in the Yellow River catchment from 2000 to 2018.

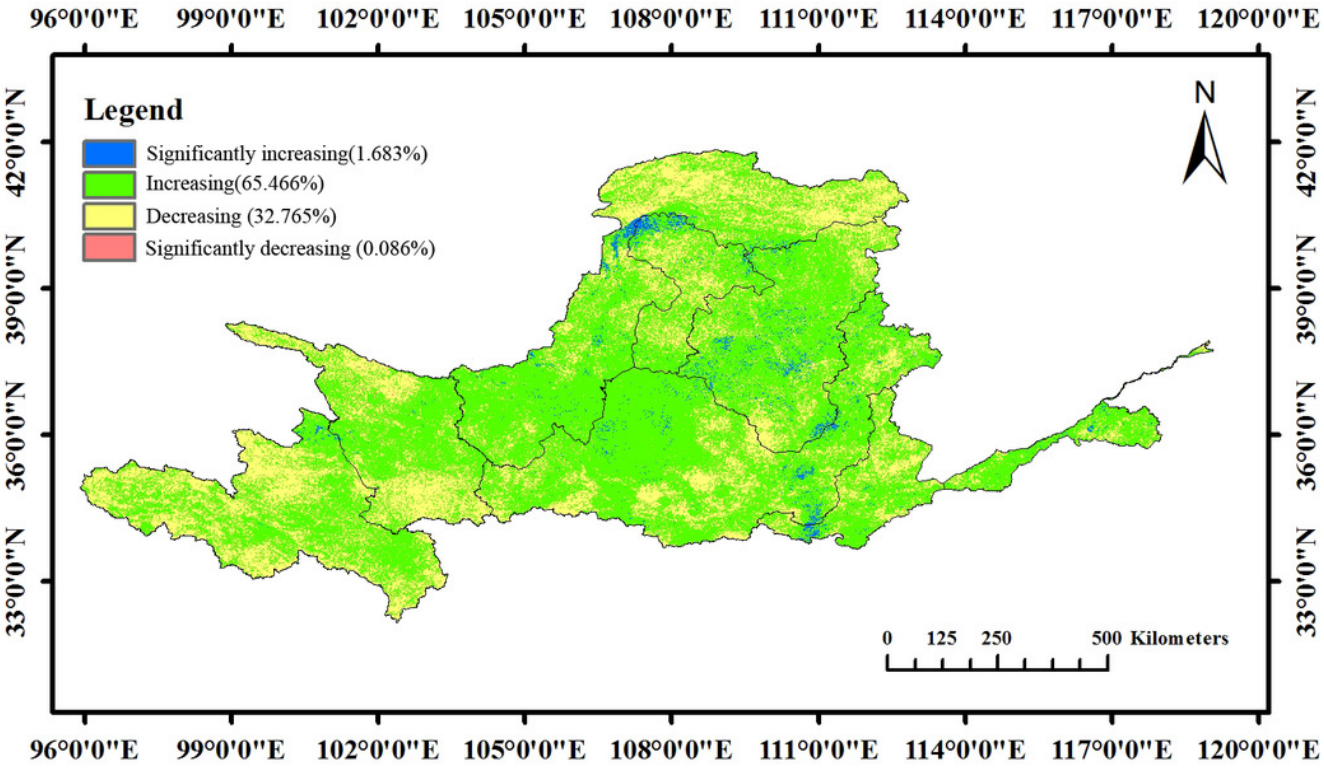

Figure 10

Figure 10 . Changes in the runoff, sediment transport, and sediment concentration from 2000 to 2018. The legends for the runoff and sediment transport are the same as those for the sediment concentration.

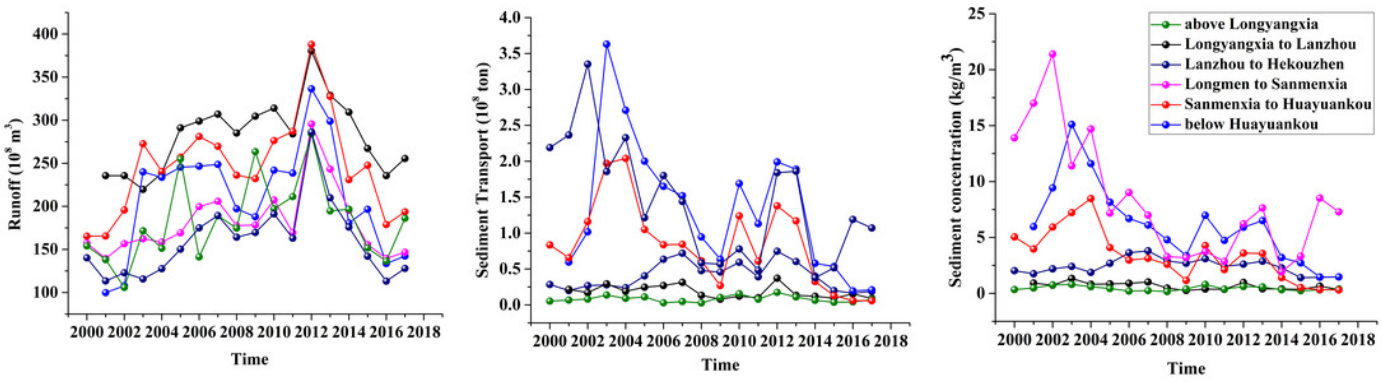

Figure 11

Figure 11 . C5.0 classification model diagram that is plotted with R studio.

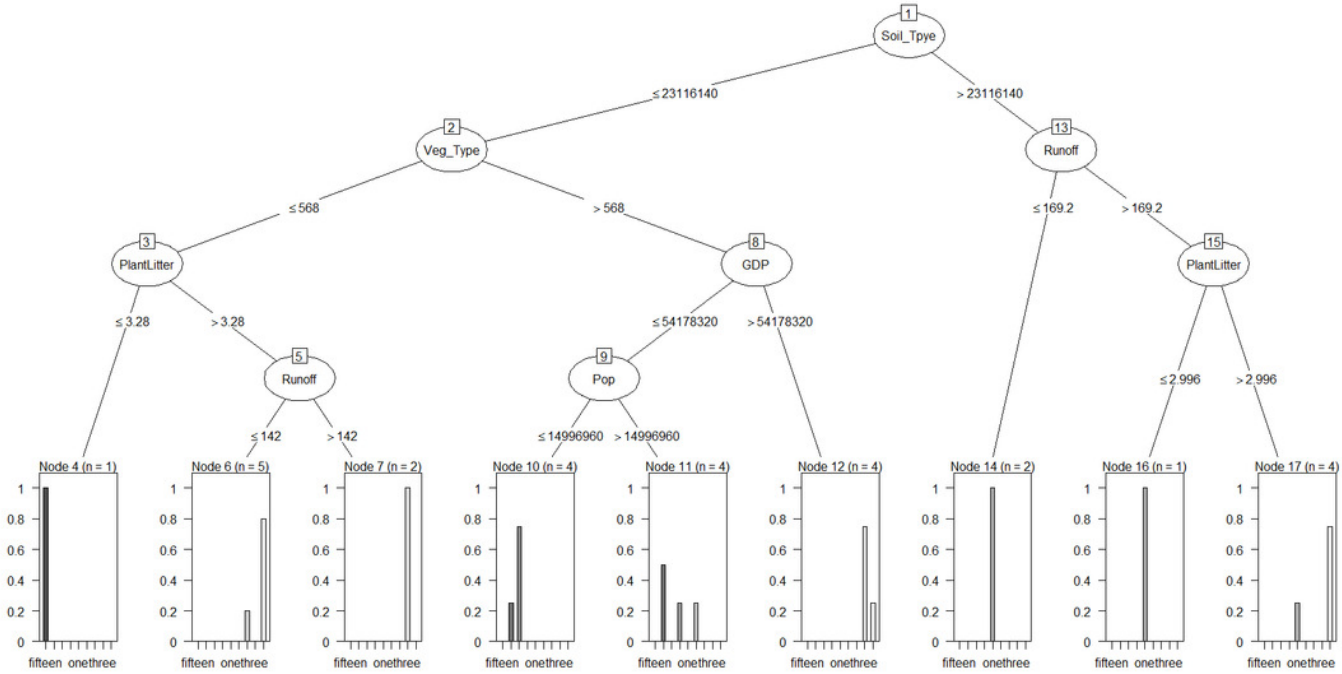

Figure 12

Figure 12 . Importance of variables affecting sediment concentration.

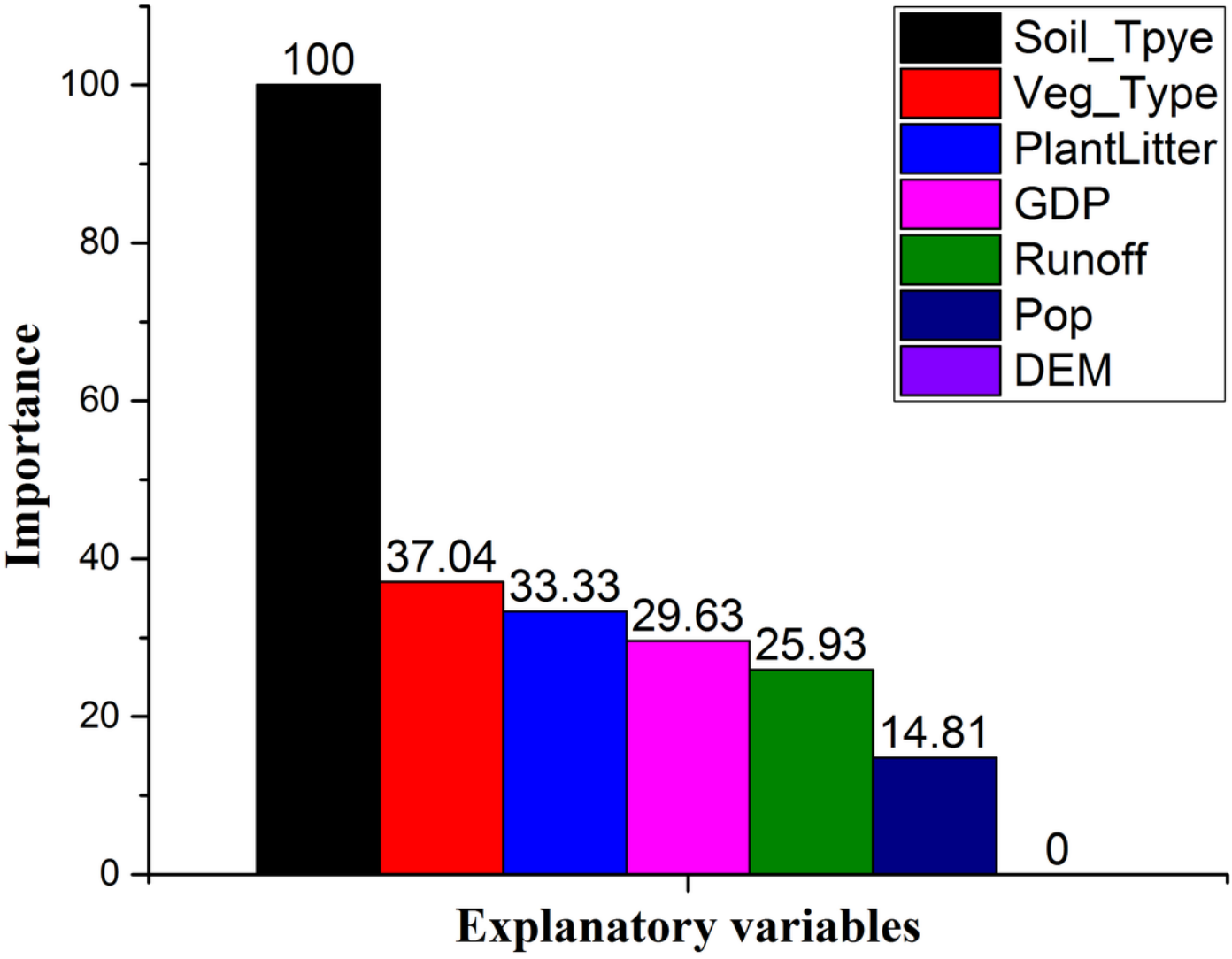

Table 1(on next page)

Table 1 . Image parameters.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27891v1 | CC BY 4.0 Open Access | rec: 10 Aug 2019, publ: 10 Aug 2019

FID	Columns	Rows	Cell size (m)	Column size (m)	Row size (m)	Number of valid cells after aggregation to 30 m
0	10991	10644	0.06778	744.96998	721.45032	426
1	13661	15118	0.04784	653.54224	723.24512	487
2	14811	17007	0.04393	650.64723	747.11751	475
3	10729	11090	0.06228	668.20212	690.6852	507
4	10896	11717	0.05383	586.53168	630.72611	322
5	11164	9111	0.06183	690.27012	563.33313	422
6	9482	8684	0.0539	511.0798	468.0676	249
7	22990	13065	0.05466	1256.6334	714.1329	704
8	21060	13047	0.1188	2501.928	1549.9836	1264
9	12406	11098	0.05872	728.48032	651.67456	499
10	12166	12944	0.06925	842.4955	896.372	741
11	6988	14250	0.05819	406.63172	829.2075	370
12	7493	6770	0.07893	591.42249	534.3561	296
Total	164837	154545	——	10832.8346	9720.35165	6762

1

Table 2(on next page)

Table 2 . Summary of the prediction accuracy of plant litter cover (PLC) (%) from the model development and validation using random forest regression trees for the first- and second-levels of upscaling.

Response variable	Plant litter cover (Test dataset n=1353) First level	Plant litter cover (Test dataset n=110,916) Second level
R^2	0.370	0.891
RMSE (%)	18.455	2.936
CVRMSE (%)	53.187	10.300
MAE (%)	13.583	2.000
Var explained (%)	38.54	88.87
Mean of squared residuals	0.035	0.0009

1

Table 3(on next page)

Table 3 . Confusion matrix for the sediment concentration classes. a-e stand for the grades of the sediment concentration: a(<0.4),b(0.4-2.4), c(2.4-4.0), d(4.0-4.8), e(4.8-8.0), f(8.0-9.6), g(13.6-16.0), h(16.0-19.2), i(32.8-40.0), unit:kg/m^3.

Classes	a	b	c	d	e	f	g	h	i
a	4	0	0	0	0	0	0	0	0
b	0	7	0	0	0	0	0	0	0
c	0	1	5	0	0	0	0	0	0
d	0	0	0	3	1	0	0	0	0
e	0	0	0	0	2	1	1	0	0
f	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	0	0	0	0
h	0	0	0	0	0	0	0	0	0
i	0	0	0	0	0	0	0	1	1

1