

# Attribution of climate and human activities to vegetation change in China using machine learning techniques

Yu Shi<sup>a,b</sup>, Ning Jin<sup>c,\*</sup>, Xuanlong Ma<sup>d</sup>, Bingyan Wu<sup>a</sup>, Qinsi He<sup>e</sup>, Chao Yue<sup>b</sup>, Qiang Yu<sup>b,e,f,\*\*</sup>

<sup>a</sup> College of Natural Resources and Environment, Northwest A&F University, Yangling 712100, China

<sup>b</sup> State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling 712100, China

<sup>c</sup> Department of Resources and Environment, Shanxi Institute of Energy, Jinzhong 030600, China

<sup>d</sup> College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China

<sup>e</sup> School of Life Sciences, University of Technology Sydney, Broadway, NSW 2007, Australia

<sup>f</sup> College of Resources and Environment, University of Chinese Academy of Science, Beijing 100049, China

## ARTICLE INFO

### Keywords:

Climate change  
Human activity  
Climatic zone  
Vegetation greening  
Attribution analysis

## ABSTRACT

A series of policies and laws have been implemented to address climate change impacts in China since the 1980s. One of the most notable policies is ecological restoration engineering. However, there are many environmental factors that affect vegetation in the ecological restoration engineering zones. The relationships among different factors cannot be explained well by traditional statistical methods due to the existence of hidden non-linear features. Moreover, it is difficult to adopt threshold methods to accurately define vegetation areas fully, or to quantitatively analyze and assess the effects of climate factors and human activities on vegetation changes. The objective of this study was to determine vegetation area and distribution using Landsat TM/ETM/OLI images combined with a support vector machine (SVM) classification model. We analyzed the dynamic characteristics of vegetation area and greenness (NDVI, Normalized Difference Vegetation Index) in China's ecological restoration engineering zones from 1990 to 2015. Based on random forest regression (RFR) with a residual analysis method, the contributions of meteorological factors and human activities to vegetation greenness changes were quantitatively evaluated. Vegetation area and NDVI changed significantly in the study areas, increasing by more than 50% and 40%, respectively, from 1990 to 2015. Temperature, sunshine hours, and precipitation impacted vegetation greenness, which caused NDVI fluctuations in specific years. However, the NDVI increase was difficult to explain fully with meteorological factors. Using cross-validation, we predicted about 80% of the observed NDVI variation occurring from 1984 to 1994. Nine meteorological factors were related to vegetation growth, of which the average temperature, minimum temperature, maximum temperature, and average relative humidity were most critical. The combined effect of the nine climatic factors contributed less to NDVI increase than human activities. Human activity was the most important factor associated with NDVI increase, with contributions of more than 100% in most study areas. Human activities derived from national or local policies had large impacts on vegetation changes. The methods and results of this study can help to understand vegetation changes observed in ecological zones and provide guidance for evaluating ecological restoration policies.

## 1. Introduction

Climate change, especially global warming caused by the greenhouse effect, is today's common challenge faced by all humanity. Climate change directly affects the growth of vegetation, and may play a positive or negative role. Meanwhile, vegetation can also reflect and regulate climate, and is therefore an important part of terrestrial

ecosystems (Baldocchi et al., 2001; Bakwin et al., 2002). In a global context, areas of green vegetation have been increasing in recent decades, especially in China and India which account for one-third of the global net growth in green leaf area (Jong et al., 2012; Zhang et al., 2017; Chen et al., 2019a). Some scholars have attributed this greening to human factors which have changed land management methods (Song et al., 2018; Tong et al., 2018). In general, vegetation greening

\* Corresponding author.

\*\* Corresponding author at: State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling 712100, China.

E-mail addresses: [jinn.13b@igsrr.ac.cn](mailto:jinn.13b@igsrr.ac.cn) (N. Jin), [yuq@nwfau.edu.cn](mailto:yuq@nwfau.edu.cn) (Q. Yu).

<https://doi.org/10.1016/j.agrformet.2020.108146>

Received 17 January 2020; Received in revised form 19 July 2020; Accepted 10 August 2020

0168-1923/ © 2020 Elsevier B.V. All rights reserved.

and browning are the result of a combination of natural and human factors (Kuenzer et al., 2015; Zheng et al., 2019). Therefore, dynamic vegetation monitoring and attribution of drivers are of great value for understanding and mitigating the effects of climate change and adjusting strategic policies.

Vegetation greening or browning occur under the background of climate change. Previous studies have demonstrated that climate affects vegetation changes in different regions. Niemand et al. (2005) and Piao et al. (2008) have shown that climate warming could stimulate vegetation growth by extending the growing season and increasing photosynthesis, especially in areas without water limitations. Adole et al. (2018) reported that precipitation had a great influence on large-scale greening in Africa. Rainfall amount is related to vegetation greening and browning in arid and semi-arid regions (Adole et al., 2018; Lamchin et al., 2018).

Human activities also have a profound impact on vegetation changes. Li et al. (2017) indicated that human activities had a significant influence on the growth and greening of vegetation. Previous studies have evaluated and validated the contribution of human activities to vegetation change (Chen et al., 2019b; Liu et al., 2019b; Tong et al., 2019). China has implemented a series of related policies and laws affecting agriculture, forestry, environmental protection, and urban planning to improve the environment in ecologically fragile regions. These policies aimed to reduce soil erosion (Chen, 2007; Liu et al., 2008) and land desertification (Zhang et al., 2020), and to protecting biodiversity (Narain et al., 2020) and alleviate water shortages (Xu et al., 2020). The most notable policies deal with vegetation restoration engineering in different climate zones of China. These projects have been implemented since the end of the 1980s, with forest land, grassland, wetland, and water bodies as the targeted land cover types to be converted from cropland and unused land. Trillions of dollars have been invested in these projects to address ecological problems (Chang et al., 2019; Du et al., 2019; Sun et al., 2019). These ecological restoration projects have improved the ecological environment in the project areas (Li et al., 2020; Xu et al., 2020). However, the factors dominating vegetation changes are still unknown with regards to ecological restoration projects. Quantifying the relative contribution of various factors to vegetation change is still a challenge (i.e., determining to what extent the observed greening is caused by human activities versus climate change). Moreover, we are not yet able to determine whether China's policies that required large financial and labor investments have had their intended consequences. In the long run, the variable climate conditions across different ecological regions make it difficult to clearly understand whether or not vegetation has really been improved by these projects.

The Normalized Difference Vegetation Index (NDVI) is commonly used to monitor vegetation growth (Xiao and Moody, 2004), determine vegetation distribution, and assess productivity (Xiao and Moody, 2004; Evans et al., 2006). The relationship between NDVI and environmental factors is increasingly important in ecological research (Liu et al., 2019a; Peng et al., 2019). NDVI quantifies the total amount of light absorbed and reflected by vegetation, and can indicate the degree of vegetation recovery after disturbance (Franks et al., 2017). Vegetation growth is closely connected with meteorological factors such as temperature, precipitation, and solar radiation (Fensholt et al., 2012; Cong et al., 2013; Emmett et al., 2018). In fact, there is a problem of collinearity between meteorological factors. Some studies have considered the contribution of climate factors and human factors to NDVI. Li et al. (2017) analyzed the impact of climate and human factors on NDVI in the Loess Plateau using linear regression and residual analysis. Zheng et al. (2019) distinguished the impact of climate and human activities on NDVI using a linear model and NPP. These studies did not always explain more of the variation in NDVI when using linear models to construct the relationship between several environmental variables and NDVI. There are many environmental factors that can affect NDVI non-linearly, and the correlation among multiple factors

cannot be explained well by traditional statistical methods. Because interactions among environmental factors can be complex, outliers could interfere with linear models (e.g., vegetation greening caused by sudden short-precipitation events in arid areas). In view of the above problems, we must find a more efficient and accurate method to assess vegetation changes.

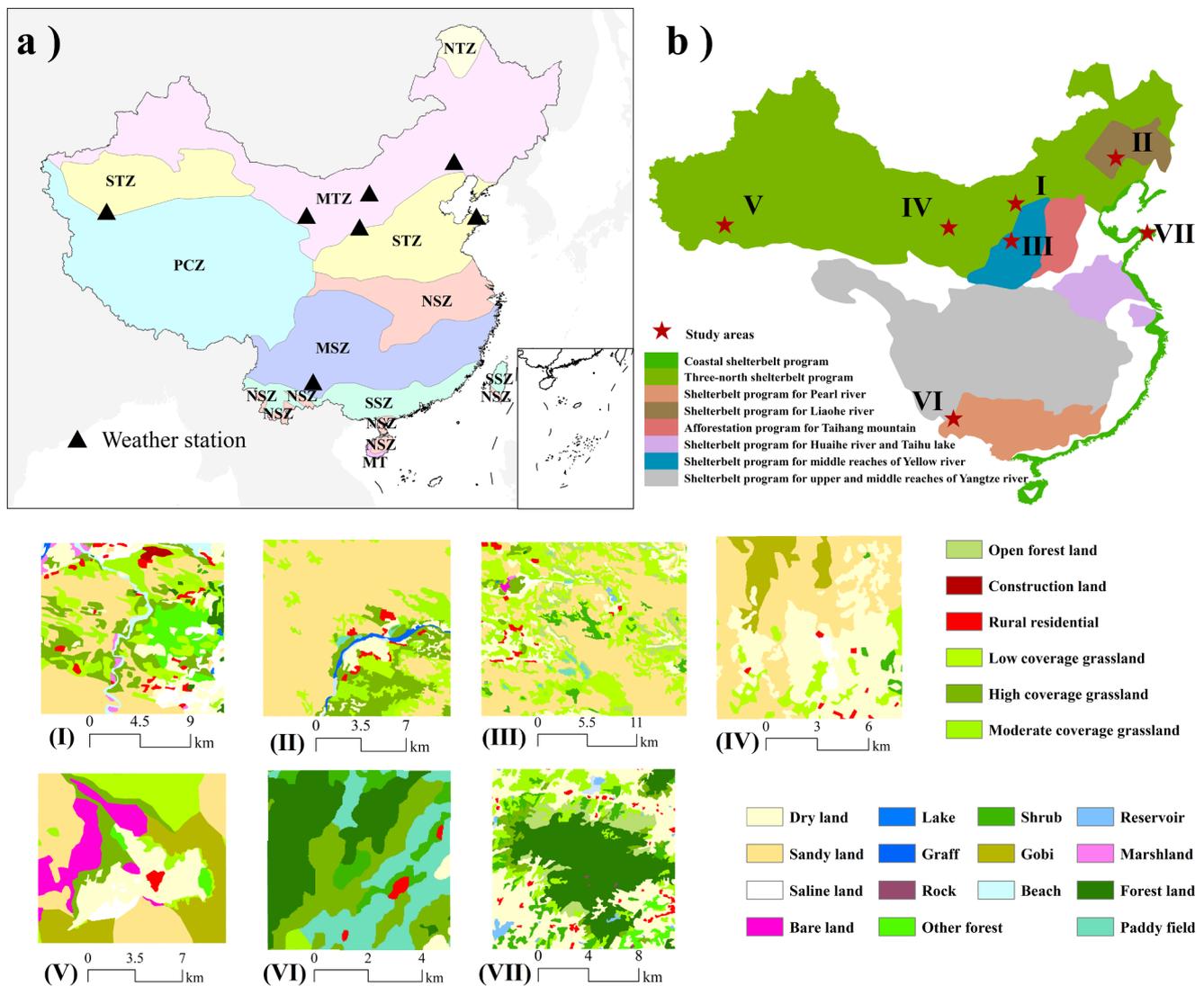
The assessment of human activities is usually based on survey data from yearbooks, but farmers in the study areas do not directly submit production reports to government agencies, which means that the data used are highly uncertain with potential bias (Hofmann et al., 2005). For example, Clauss et al. (2018) found that the data released by provincial statistics offices in Vietnam disagreed with data provided by the country. Machine learning methods came into being in order to solve nonlinear problems and big data processing issues. Machine learning is a cross-discipline study area focusing on how computers can be used to simulate or implement human learning behaviors so as to acquire new knowledge or skills. Machine learning also studies how to reorganize existing knowledge structures for continuous performance improvement. Machine learning methods have been applied to the field of vegetation recovery and performed well. Vidal-Macua et al. (2020) used boosted regression trees combined with Landsat images to assess vegetation restoration in coal reclamation areas, thereby demonstrating the negative influence of drought on vegetation. Zaines et al. (2019) used a vegetation index and random forest to assess the impact of a dam on the vegetation of a riparian delta, and the results showed that low vegetation cover was more affected human activities than higher density vegetation cover.

In summary, this study used NDVI to quantify vegetation greenness. We attempted to use machine learning methods to analyze the relationships between multiple factors and NDVI in different ecological engineering zones, accurately identify vegetation changes, and quantitatively evaluate the driving factors. The objectives of this study were to: (1) interpret and analyze the distribution characteristics of vegetation using SVM (support vector machine) classifier in seven study areas; (2) analyze vegetation greenness changes and the influence of major climatic factors (temperature, precipitation, and sunshine hours) on those changes over a 25-year period; and (3) quantify the contribution of climatic and human activities/policies to vegetation changes based on a residual analysis.

## 2. Study area

Major vegetation restoration projects in China include Natural Forest Protection Project (NFPP), "Grain for Green" Project (GGP), Three-North Shelterbelt Program (TSP), and Beijing-Tianjin Sandstorm Source Control Project (BSSCP). NFPP began in 1998 in southwestern China to reduce commercial logging and accelerate construction of plantations to protect natural forests and reduce natural disasters (Liu et al., 2008). BSSCP was launched in 2001 with the aim of establishing an ecological protection system to deal with the hazards of sandstorms in the north, especially Beijing (Wu et al., 2013). TSP encompassed a wide geographic range (including Northwest China, North China, and Northeast China) over a long time period (1978–2050) to form a barrier against desertification (Duan et al., 2011). GGP has been called the world's largest ecological restoration project. The project started in 1999 and promoted the conversion of farmland to forest in hilly areas over most of the ecologically fragile regions in China (Liu et al., 2008).

We randomly selected three to five areas in different ecological engineering projects, and left seven representative areas through comparative analysis for further investigation (Fig. 1b). The representative study areas were distributed in different climatic zones across China (Fig. 1a) because we wanted to evaluate the relative contributions of meteorological factors to vegetation changes. The land use status of each study area was different due to different climatic factors and human activities (Fig. 1). The sizes of the study areas were selected to



**Fig. 1.** Study areas and major forest and grassland projects. (a) Climate zones in China and weather station locations. NTZ, MTZ, STZ, PCZ, NSZ, MSZ, SSZ, and MT are north temperate zone, mid temperate zone, south temperate zone, plateau climate zone, north subtropical zone, mid subtropical zone, south subtropics zone, and mid tropical zone, respectively. (b) Vegetation restoration projects and representative research regions selected for the study. The bottom of the figure shows more details of the study areas, where (I) KBQ, (II) KEQ, (III) MWS, (IV) MQX, (V) TKLMG, (VI) QJS, and (VII) YTS. See Table 1 for abbreviation definitions for the study areas.

**Table 1**  
Basic information of the study areas.

Area	Location	Coordinate	Temperature	Precipitation	Climatic zone
KBQ	The north of the Ordos Plateau	109.5E, 40.4N	7.37 °C	316 mm	MTZ
KEQ	The southeast Inner Mongolia	119.8E, 42.9N	6.00 °C	420 mm	MTZ
MWS	The between Yulin and Ordos City	108.9E, 37.7N	8.41 °C	340 mm	STZ
MQX	Wuwei City in Gansu	102.9E, 38.7N	7.40 °C	113 mm	MTZ
TKLMG	The Yutian area of Xinjiang	81.9E, 36.8N	11.40 °C	> 100 mm	STZ
QJS	The southwestern of China	103.7E, 24.8N	15.10 °C	< 1000 mm	MSZ
YTS	The Shandong Peninsula of China	121.7E, 37.3N	12.20 °C	650 mm	STZ

Note: KBQ = Kubuqi Desert; KEQ = Keerqin sandy land; MWS = Maowusu sandy; MQX = Minqinxian; TKLMG = Taklimakan Desert; QJS = Qujingshi; YTS = Yantaishi; MTZ = mid temperate zone; STZ = south temperate zone; MSZ = mid subtropical zone.

match the coverage of nearby weather stations because the weather conditions can be considered as relatively uniform in a certain area (Fig. 1a). The more details about study regions are as follows (Table 1):

Table 1 shows basic information about the study areas. The Kubuqi Desert (KBQ) area is located in the north of the Ordos Plateau which belongs to the mid-temperate arid and semi-arid regions. The Keerqin Sandy Land (KEQ) area is located in the semi-arid area of southeast

Inner Mongolia with an annual evaporation of 1500–2500 mm. Desertification of KEQ has been caused by increased human activities in recent decades. The Maowusu Sandy (MWS) area is located between Yulin City and Ordos City. Desert formed here after the loss of vegetation due to reclamation, climate change, and war. The Minqinxian (MQX) region, under the jurisdiction of Wuwei city in Gansu province, has a temperate continental climate, with an average annual

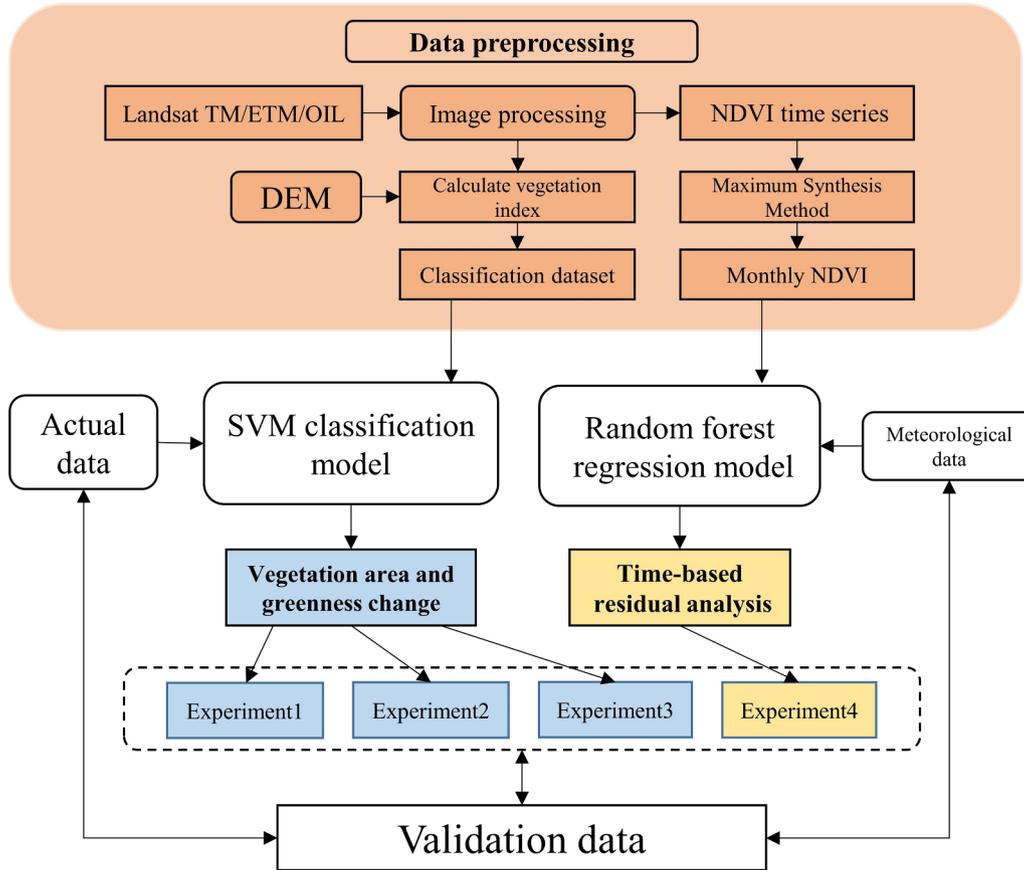


Fig. 2. Study workflow.

**Table 2.**  
Features of the SVM classifier.

Type	Input data
Reflectance	Band 1-Band 7 from Landsat images
Vegetation index	NDVI, EVI, RVI from Landsat images
Other	DEM (See Supplementary materials)

evaporation of 2600 mm. The Taklimakan Desert (TKLMG) area in the Yutian area of Xinjiang is the largest desert in China. The region is hot and dry year around, with sparse vegetation. The Qujingshi (QJS) area is located in southwestern China. The Yantaishi (YTS) area is located in the Shandong Peninsula of China, surrounded by the sea on three sides.

### 3. Materials and methods

In this study, Landsat imagery was used as the data source, and the pre-processed imagery was analyzed to study changes in vegetation area (SVM, support vector machine classifier) and greenness with meteorological data (MLR, multiple linear regression). A random forest regression (RFR) model was constructed using actual NDVI data and data for nine monthly meteorological parameters from 1984 to 1994, and the contribution of human activities to vegetation changes was quantitatively assessed through time-based residual analysis (Fig. 2).

#### 3.1. Landsat TM/ETM/OLI images pre-processing

This study used 793 Landsat TM / ETM / OLI images from the United States Geological Survey (USGS, <http://glovis.usgs.gov/>). The images have 30 m-resolution with the UTM projected coordinate system and the WGS84 geographic coordinate system.

The raw images were subjected to pre-processing by radiometric calibration and FLAASH atmospheric correction using ENVI software. We used the radiometric calibration module to convert the DN value of the original image into radiance based on the radiation calibration parameters that came with the Landsat satellite. The formula is:

$$L_{\lambda} = G \cdot DN + O \tag{1}$$

where  $L_{\lambda}$  is radiance, and  $G$  and  $O$  are gain and offset from the image metafile, respectively. The DN value is from the pixel value of the raw images. Further, we used the FLAASH atmospheric correction module based on the radiation transmission model to convert the radiance data to surface reflectance data. For specific details about FLAASH atmospheric correction, see Adler-Golden et al. (1999).

We used pre-processed remote sensing images to calculate normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), and ratio vegetation index (RVI) as follows:

$$EVI = G \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + C_1 \times \rho_{Red} - C_2 \times \rho_{Blue} + L} \tag{2}$$

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \tag{3}$$

$$RVI = \frac{\rho_{NIR}}{\rho_{Red}} \tag{4}$$

where  $\rho_{Red}$ ,  $\rho_{NIR}$ ,  $\rho_{SWIR}$ , and  $\rho_{Blue}$  are the surface reflectance values of the red band, near infrared band, shortwave infrared band, and blue band, respectively, in the Landsat TM/ETM/OLI sensor.  $L$  is the canopy background adjustment that addresses non-linear, differential near infrared and red radiant transfer through a canopy.  $C_1$  and  $C_2$  are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band (Huete et al., 2002).

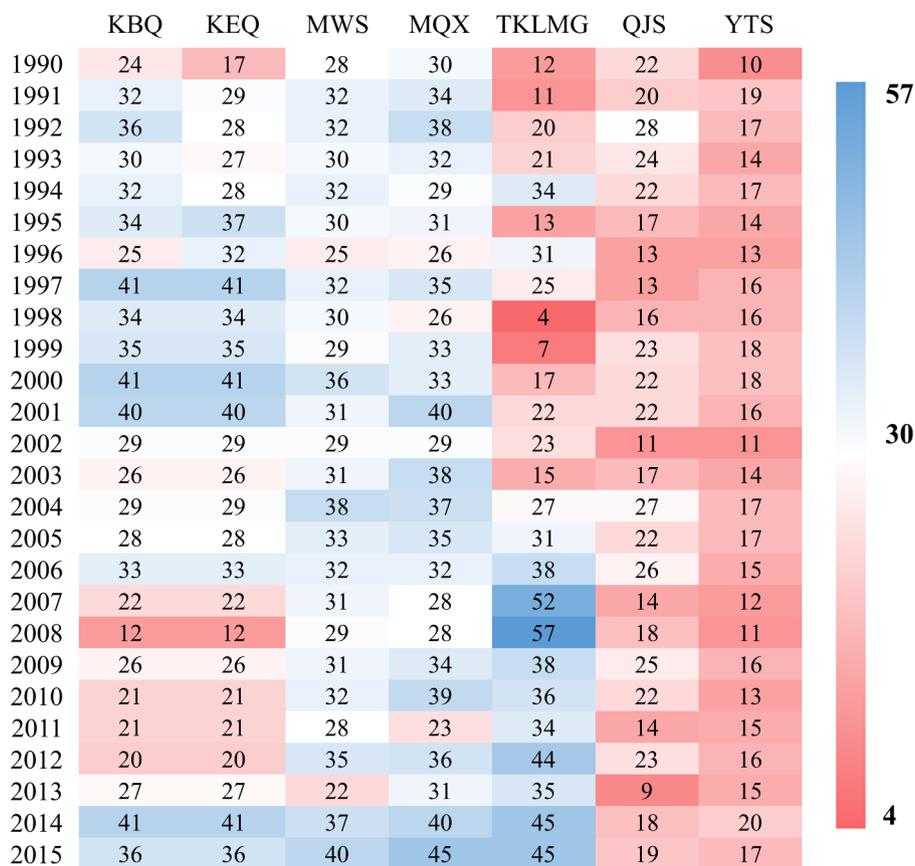


Fig. 3. The number of annual Landsat images used for calculating annual NDVI by the maximum value composites method at seven study sites in China. (KBQ, Kubuqi Desert; KEQ, Keerqin sandy land; MWS, Maowusu sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).

### 3.2. Meteorological data pre-processing

The meteorological data were acquired from the National Meteorological Centre (Fig. 1a). The impact of the meteorological parameters on NDVI were assessed by processing the meteorological data into two forms: (1) annual meteorological data – we used daily temperature, precipitation, and sunshine hours from 1990 to 2015 to generate mean annual temperature (MAT), total annual precipitation (TAP), and total annual sunshine hours (TAS). (2) monthly meteorological data – we used daily temperature, precipitation, and sunshine hours from 1984 to 1994 to generate mean monthly temperature (MMT), mean monthly minimum temperature (MMMT1), mean monthly maximum temperature (MMMT2), mean monthly relative humidity (MMRH), mean monthly precipitation (MMP), mean monthly wind speed (MMWS), mean monthly sunshine hours (MMSH), total monthly precipitation (TMP), and total monthly sunshine hours (TMSH).

Monthly meteorological data were used to build the regression model. Annual meteorological data were used to analyze the impact of meteorological factors on NDVI.

### 3.3. Extracting distribution of vegetation

SVM is a machine learning algorithm based on the structural risk minimization principle and statistical theory (Cortes and Vapnik, 1995). SVM uses a kernel function to map the nonlinear problem to a high-dimensional space so as to be transformed into a linear problem (Vapnik, 2013). In addition, the values for cost factor and gamma, will affect the punishment imposed for misclassification of samples and the complexity of the algorithm for the SVM model. The detailed description and formulas for SVM are provided in the

supplementary materials.

This paper used the SVM model to establish a nonlinear-implicit relationship between the object category and features (Table 2) of the study areas and to identify vegetation and non-vegetation areas from Landsat images of vigorous vegetation growth in 1990, 1995, 2000, 2005, 2010, and 2015. We used the radial basis kernel function and the optimal parameters (cost and gamma by the lattice search) in SVM. This work was accomplished with the R program by using the e1071 package. (<https://cran.r-project.org/web/packages/e1071/>).

Considering that the interpretation results of SVM will directly affect the analysis of vegetation area, we generated training data and verification data using Google Earth. First, we divided the research areas into several layers, including cultivated land, forest, grassland, water bodies, construction land, and unused land. Second, random points were generated on each layer. We then used very high resolution images from Google Earth to check each random point to determine if the point fit the correct distribution, and random points without clear land cover information were removed. Third, we reclassified each layer as vegetation and non-vegetation. Finally, about 200 validation points in the study areas were generated to construct the SVM classification model and test vegetation maps, of which 50% were used as training points and the remaining were used as validation points. The overall accuracy and kappa coefficient (calculated by confusion matrix) were used as the indicators to assess the interpretation accuracy.

### 3.4. Impact of climate on NDVI

Usually the limiting factors for vegetation growth are solar radiation, temperature, and precipitation. Therefore, we used multiple linear regression (MLR) to identify relationships between annual NDVI (aNDVI) and three meteorological factors: mean annual temperature

**Table 3.**  
The accuracy of image interpretation using the SVM model in seven study areas in China from 1990 to 2015.

Region	Year	Class	Veg.	Non-veg.	Overall acc.	Kappa coef.
KBQ	1991	Veg./Non-veg.	50/0	0/48	1	1
	1995	Veg./Non-veg.	49/1	0/50	0.99	0.98
	2000	Veg./Non-veg.	47/0	0/53	1	1
	2005	Veg./Non-veg.	50/0	0/48	1	1
	2010	Veg./Non-veg.	52/2	0/44	0.9796	0.9589
2015	Veg./Non-veg.	48/1	0/47	0.9896	0.9792	
KEQ	1992	Veg./Non-veg.	49/0	0/53	1	1
	1995	Veg./Non-veg.	50/0	0/51	1	1
	2000	Veg./Non-veg.	47/0	0/48	1	1
	2005	Veg./Non-veg.	46/0	0/54	1	1
	2009	Veg./Non-veg.	52/0	0/46	1	1
2015	Veg./Non-veg.	49/0	0/51	1	1	
MWS	1991	Veg./Non-veg.	47/0	0/51	1	1
	1995	Veg./Non-veg.	47/0	1/56	0.9904	0.9806
	2000	Veg./Non-veg.	49/1	0/47	0.9897	0.9794
	2005	Veg./Non-veg.	45/0	0/52	1	1
	2010	Veg./Non-veg.	51/0	0/49	1	1
2015	Veg./Non-veg.	59/0	0/46	1	1	
MQX	1991	Veg./Non-veg.	43/0	0/56	1	1
	1995	Veg./Non-veg.	50/0	0/52	1	1
	2000	Veg./Non-veg.	45/0	0/52	1	1
	2006	Veg./Non-veg.	49/0	0/51	1	1
	2010	Veg./Non-veg.	45/0	0/54	1	1
2015	Veg./Non-veg.	44/0	0/55	1	1	
TKLMG	1993	Veg./Non-veg.	55/0	0/44	1	1
	1996	Veg./Non-veg.	50/0	1/45	0.9896	0.9791
	2001	Veg./Non-veg.	45/0	3/50	0.9694	0.9387
	2006	Veg./Non-veg.	45/0	0/52	1	1
	2010	Veg./Non-veg.	45/0	1/54	0.99	0.9798
2015	Veg./Non-veg.	45/0	0/50	1	1	
QJS	1989	Veg./Non-veg.	50/0	0/48	1	1
	1994	Veg./Non-veg.	43/0	0/56	1	1
	2000	Veg./Non-veg.	46/0	0/51	1	1
	2005	Veg./Non-veg.	43/0	0/56	1	1
	2010	Veg./Non-veg.	52/0	0/46	1	1
2015	Veg./Non-veg.	46/0	0/52	1	1	
YTS	1990	Veg./Non-veg.	49/0	0/50	1	1
	1996	Veg./Non-veg.	50/0	0/50	1	1
	2000	Veg./Non-veg.	48/0	0/49	1	1
	2005	Veg./Non-veg.	48/0	0/52	1	1
	2010	Veg./Non-veg.	48/0	0/51	1	1
2015	Veg./Non-veg.	46/1	0/50	0.9897	0.9793	

Note: KBQ, Kubuqi Desert; KEQ, Keerqin Sandy land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi.

(MAT), total annual precipitation (TAP), and total annual sunshine hours (TAS). The aNDVI was calculated in two steps: (1) All NDVI images from the processed Landsat images in every year were used to generate an annual maximum NDVI image by the maximum value composites (MVC) method. (2) Then the mean NDVI values were calculated from the NDVI images obtained in step (1). Fig. 3 shows the number of available Landsat images for every year from 1990 to 2015 in the seven different study areas.

MLR analysis was completed in the R program. The analysis of variance (ANOVA) was used to test the significance of different variables in the final linear regression model. The F-statistic was used to test the significance of overall significance of aNDVI and three independent variables. To quantify the relative contribution of different climate factors to the regression model, we wrote an R program function to derive the relative importance of the three meteorological factors by referring to research from Johnson (2004).

### 3.5. Attribution of vegetation change

The RFR model generates different regression trees by using random multiple training sets and features (Tian et al., 2016). Each regression tree is sampled independently and distributed identically in RFR (Fig. S1). Each regression tree produces different results through branching,

and the prediction from the RFR model is the average of all trees. For more details about the RF model, please see Breiman (2001) and the supplementary materials.

Our purpose in using RFR was to reestablish NDVI affected by climate (NDVI<sub>c</sub>) from 2000 to 2015. Specifically, we used actual monthly mean NDVI (mNDVI) from Landsat in 1984–1994 as NDVI<sub>c</sub> due to the nonlinear change of vegetation greenness and less human disturbance. Then we established the relationship between mNDVI and nine monthly meteorological factors (MMT, MMT1, MMT2, MMRH, MMP, MMWS, MMSH, TMP, TMSH). Finally, the established RFR model was used to predict NDVI<sub>c</sub> from 2000 to 2015.

Values of mNDVI were calculated in two steps: (1) NDVI images of the processed Landsat images obtained in every month were used to generate a monthly maximum NDVI image by the MVC method. (2) The mean value of the NDVI images of step (1) were calculated. Table S1 (Supplementary material) shows the number of available Landsat images for every month from 1984 to 1994 in the seven different study areas. The RFR model was obtained through the “randomForest” package in the R program, in which ntree was set as 500 and mtry was set as 3 (square root of the number of predictor variables) after debugging.

To evaluate RFR model performance, we used a stratified 10-fold cross-validation. Two statistical parameters were used to evaluate the results of cross-validation: (1) coefficient of determination ( $R^2_{CV}$ ); and (2) normalized root mean squared error (nRMSE<sub>CV</sub>), where the CV subscript represents the data obtained from the validation datasets. We used the Cal subscript to represent the data obtained from the calibration datasets. These parameters were calculated as:

$$R^2 = \frac{[\sum_{i=1}^n (x(i) - x_m)(y(i) - y_m)]^2}{\sum_{i=1}^n (x(i) - x_m)^2 \sum_{i=1}^n (y(i) - y_m)^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (pe_i - p_i)^2}{n}} \quad (6)$$

$$nRMSE = \frac{RMSE}{\bar{pe}_i} \quad (7)$$

where  $y(i)$  and  $x(i)$  are the simulated and observed values of the NDVI, respectively;  $n$  is the number of samples,  $pe_i$  and  $p_i$  are the observed and simulated values, respectively.  $\bar{pe}_i$  is the mean value of observed NDVI.

A brief explanation of the process of stratified cross-validation follows. The data were first randomly divided into 10 folds by year (using the createFolds function in the R caret package). Then validation data from the first folder and training data from the remaining data were used to build the model. Finally, the previous steps were repeated 10 times in turn, and the average value was calculated at the end. We also used function importance to compute the variable importance in the “randomForest” package. The variable importance was quantified as the MeanDecreaseAccuracy (defined as the mean decrease value after disturbance of the variable) to quantify the contribution of each meteorological factor.

NDVI changes in the study area were divided into two groups. One group was NDVI affected by climate factors (NDVI<sub>c</sub>), and the other group was NDVI affected by human factors (NDVI<sub>h</sub>) (Evans and Geerken, 2004; Wessels et al., 2007; Jing et al., 2015). Therefore we considered using residual analysis to separate NDVI<sub>h</sub>. NDVI<sub>c</sub> was calculated in the RFR model. NDVI<sub>h</sub> was calculated as follows (Zheng et al., 2019):

$$NDVI_h = NDVI_a - NDVI_c + \varphi \quad (8)$$

where NDVI<sub>a</sub> is the mNDVI of the Landsat images of luxuriant vegetation, and is affected by both human and climatic factors.  $\varphi$  is uncertainty from hidden or unknown environmental factors, such as soil/ground water change, atmospheric concentration, rainfall seasonal distribution, etc. In this paper, we did not consider  $\varphi$ , but showed the influence of  $\varphi$  on the model in the results to ensure the stability of the

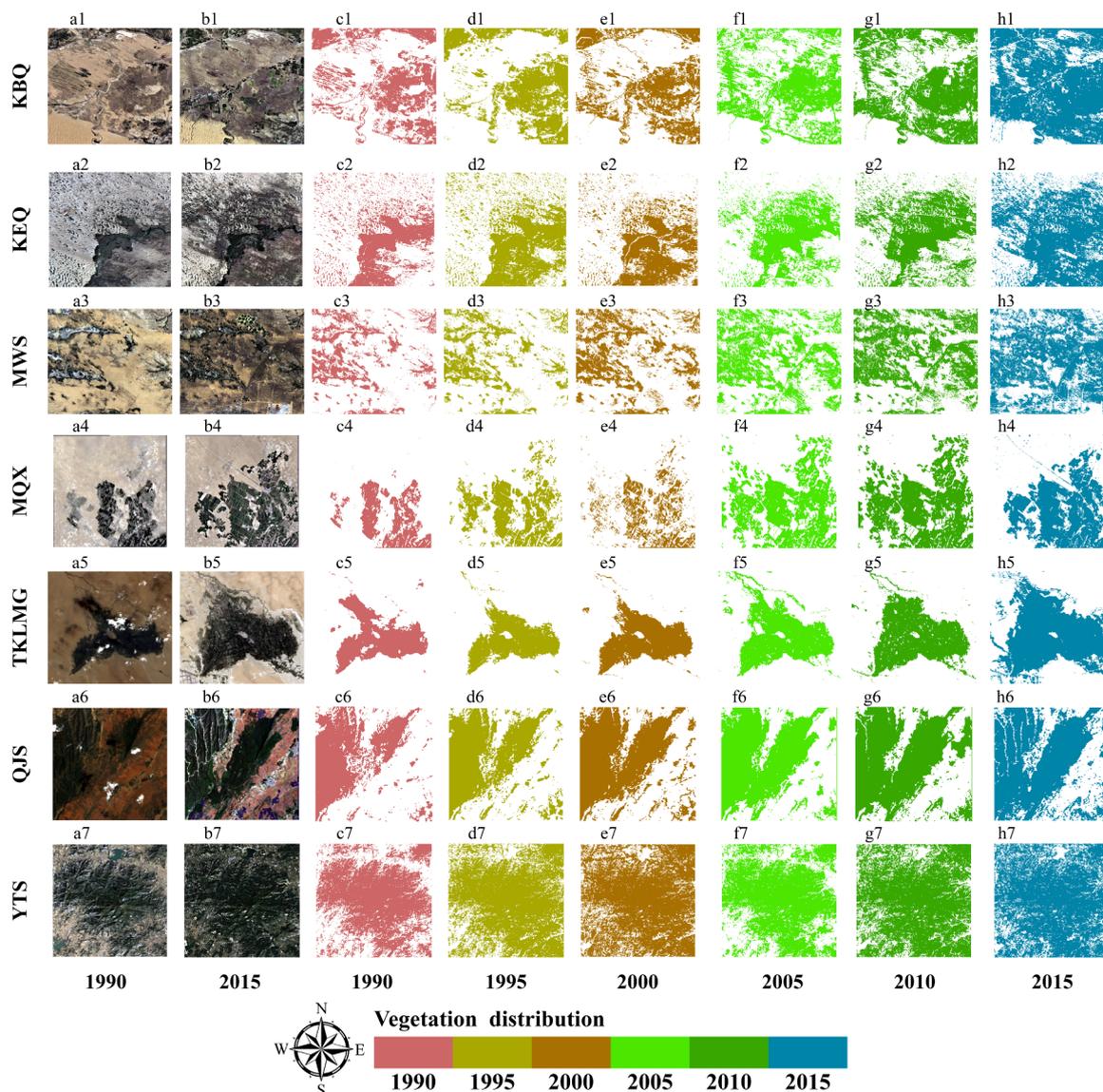


Fig. 4. Image interpretation using SVM classifier from 1990 to 2015 in seven study areas in China. (a) True color synthesis image in 1990; (b) true color synthesis image in 2015; (c) vegetation distribution in 1990; (d) vegetation distribution in 1995; (e) vegetation distribution in 2000; (f) vegetation distribution in 2005; (g) vegetation distribution in 2010; (h) vegetation distribution in 2015. (KBQ, Kubeqi Desert; KEQ, Keerqin Sandy land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).

model. Therefore, the modified formula is given as:

$$NDVI_h = NDVI_a - NDVI_c \tag{9}$$

Because ecological projects began before 2000, this study analyzed NDVI changes from 2000 to 2015 (i.e.,  $t_1 = 2000, t_2 = 2015$ ). The actual NDVI changes based on formula (8) were calculated as follows:

$$\Delta NDVI_a = NDVI_{a,t_2} - NDVI_{a,t_1} \tag{10}$$

$\Delta NDVI_a$  was composed of two parts, which were  $\Delta NDVI_c$  affected by climate factors and  $\Delta NDVI_h$  affected by human factors. Their formulas were:

$$\Delta NDVI_h = NDVI_{h,t_2} - NDVI_{h,t_1} \tag{11}$$

$$\Delta NDVI_c = NDVI_{c,t_2} - NDVI_{c,t_1} \tag{12}$$

The contributions of human activities and climate factors to NDVI changes were quantified by formulas (9)–(12).

### 3.6. Verification of attribution results

For the verification of attribution analysis results, we used the land use database of 1990 and 2015 to establish a transition matrix. The objective was to demonstrate the changes in six land use classes from 1990 to 2015 to indirectly assess the accuracy of attribution results. The transition matrix showed the value of the number of land use class pixels. Specifically, the diagonal values of the matrix represented the number of unchanged land use classes from 1990 to 2015, the values in the lower left portion of the matrix are the sum of all classes in 1990 transferred to a certain class in 2015, and the values in the upper right portion of the matrix are the sum of a certain class in 1990 transferred to all classes in 2015.

The land use database was a multi-temporal land use status database that covered the national land area that included 1990, 1995, 2000, 2005, 2010, and 2015 with a main data source of the Landsat TM/ETM images, which was generated by visual interpretation (Liu et al., 2014). The land use dataset includes six landform types: cultivated land (CL1), forest land (FL), grassland (GL), waters (W),

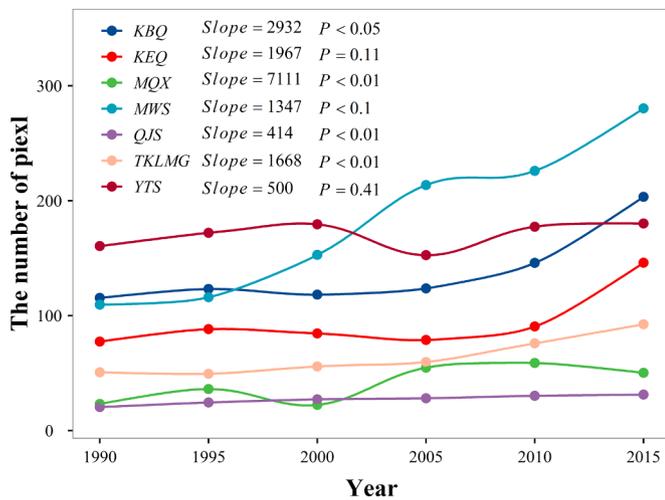


Fig. 5. Changes in vegetation area every five years from 1990 to 2015 in seven study areas in China. The value of the y-axis tick labels is pixels divided by 1000. (KBQ, Kubuqi Desert; KEQ, Keerqin Sandy Land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).

construction land (CL2), and unused land (UL). See Supplementary materials Section 1.4 for the results and analysis of the transition matrix.

#### 4. Results

##### 4.1. SVM interpretation and changes in vegetation area

The SVM model was used to extract the vegetation distribution and area, and the validation points were used to verify the interpretation accuracy (Table 3). Results showed that the interpretation of each study area had a high precision, with values of over 95% (i.e., 0.95) for overall accuracy and kappa coefficient (Table 3). Fig. 4 shows the image interpretation of vegetation distribution, and the result is basically the same as the actual vegetation distribution. The high classification accuracy of the SVM model is the basis of extracting vegetation area.

Further, we extracted the vegetation area every five years from 1990 to 2015 to analyze the area change. Fig. 5 shows that the slope of the vegetation area change in different study areas was greater than 0 ( $p$ -value  $< 0.05$  at KBQ, MWS, TKLMG, and, QJS,  $p$ -value  $< 0.1$  in MQX). The area of vegetation changed greatly from 1990 to 2015, and the vegetation area increased more than 50% at six of the seven areas (YTS was 12.26%) (Table 4). Some study areas showed rapid increases in vegetation (e.g., KBQ, MWS), while other areas showed much slower increases (QJS, YTS) (Fig. 5).

Vegetation expanded in varying degrees and directions from 1990 to 2015 (Fig. 4c–h). Vegetation expanded toward the desert in some study areas containing desert (KEQ, KBQ, MWS, MQX, and TKLMG). In other areas, vegetation expanded from the center of the area to the

Table 4. Growth of vegetation area and NDVI in 2015 compared with 1990.

Change in	Study areas						
	TKLMG <sup>a</sup>	QJS	YTS	KEQ	KBQ	MWS	MQX
NDVI (%)	-0.62	63.77	43.70	130.97	10.79	10.78	37.44
Vegetation area (%)	82.40	53.04	12.26	88.59	76.15	141.59	116.66
Vegetation area (ha)	3760	974	1771	6174	7913	15369	2436

<sup>a</sup> TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi; KEQ, Keerqin Sandy Land; KBQ, Kubuqi Desert; MWS, Maowusu Sandy; MQX, Minqinxian.

periphery. Some areas showed obvious vegetation expansion (KEQ and TKLMG), while other areas had less dramatic vegetation expansion (YTS).

##### 4.2. Temperature, precipitation, and sunshine affect NDVI change

We analyzed changes in aNDVI from 1990 to 2015. aNDVI changed somewhat differently in each study area (Fig. 6), although all seven study areas exhibited generally increasing aNDVI over the study period, (all slopes  $> 0$  and  $p$ -values  $< 0.05$ ) (Fig. 6). The vegetation generally showed a high greenness in each region, even though different regions had different climate types, underlying surfaces, and human activities. Fig. 7 shows the spatial distribution of aNDVI every five years (the non-vegetation areas have been masked out).

In this study, the environmental factors which were closely related to vegetation growth (temperature, precipitation, and solar radiation) were selected to analyze the influence of meteorological factors on aNDVI (Figs. S2–S8). Vegetation greening was observed to be closely related to meteorological factors. Different study areas had different limiting factors for aNDVI (Figs. S2–S8). Specifically, temperature influenced aNDVI at KBQ ( $P < 0.01$ ,  $R^2 = 0.41$ ); precipitation influenced aNDVI at KEQ ( $P = 0.09$ ,  $R^2 = 0.11$ ); temperature influenced aNDVI at MQX ( $P = 0.02$ ,  $R^2 = 0.21$ ); temperature influenced aNDVI at MWS ( $P = 0.01$ ,  $R^2 = 0.24$ ); sunshine hours influenced aNDVI at TKLMG ( $P = 0.07$ ,  $R^2 = 0.16$ ); sunshine hours influenced aNDVI at QJS ( $P = 0.07$ ,  $R^2 = 0.13$ ). Figs. S2–S8 indicated that a single meteorological element cannot explain the change of aNDVI well.

We further used MLR to explore the impact of the three meteorological parameters on aNDVI. The MLR of aNDVI included temperature, precipitation, and sunshine hours (Table 5). ANOVA indicated that MAT exerted a significant control over aNDVI at KBQ, MWS, and, MQX (all with  $p$ -value  $< 0.05$ ). TAP exerted a significant control over aNDVI at MQX (with  $p$ -value  $< 0.05$ ). TAS exerted a significant control over aNDVI at KEQ and MQX (with all  $p$ -value  $< 0.05$ ). MAT contributed the most to the explanatory power of the regression model at KBQ, MWS, MQX, and YTS ( $R^2 = 0.51, 0.34, 0.45,$  and  $0.15$ , respectively). TAS contributed the most to the explanatory power of the regression model at KEQ, TKLMG, and QJS ( $R^2 = 0.29, 0.19,$  and  $0.17$ , respectively).

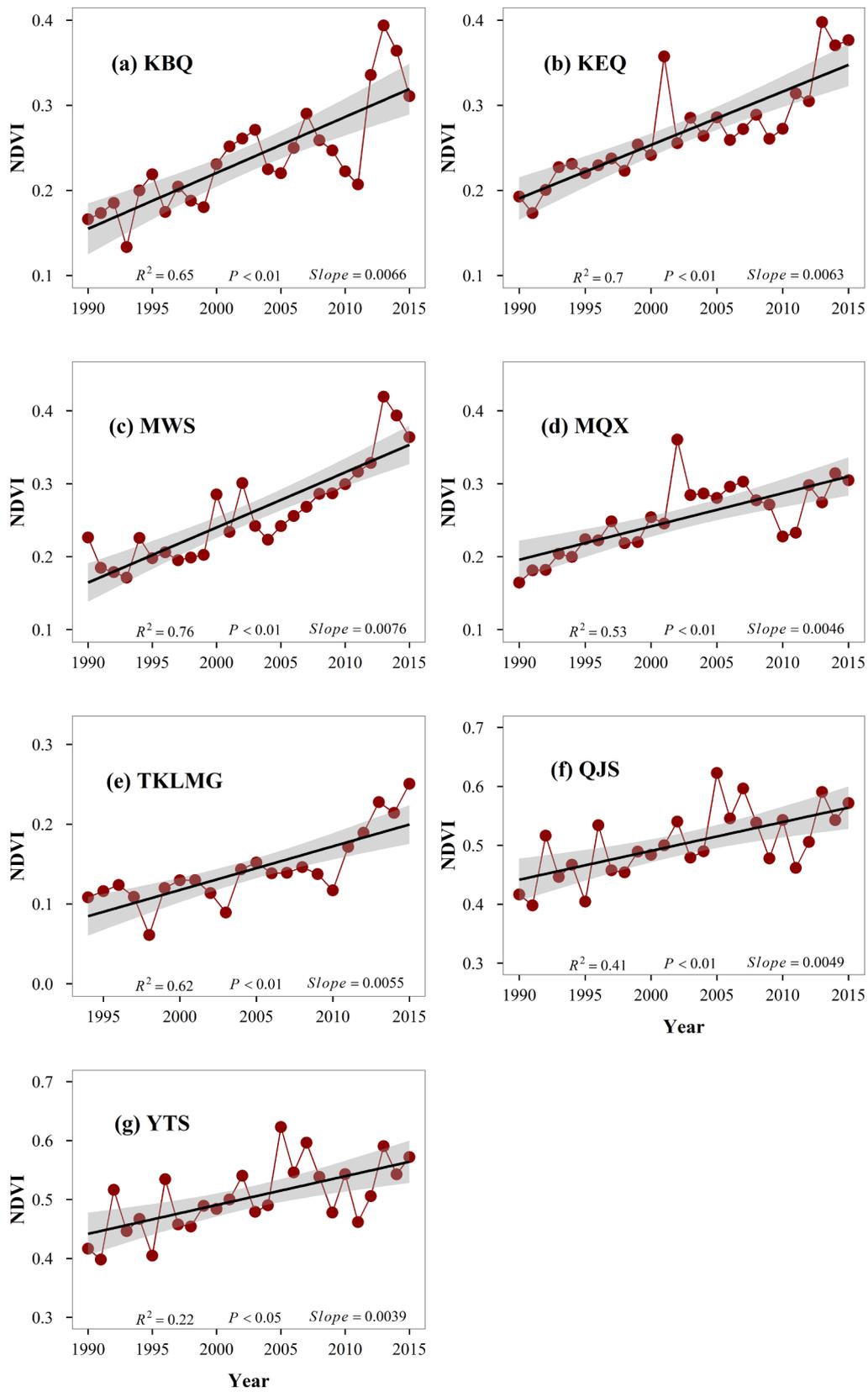
Fig. 8 shows the changes in aNDVI and the three meteorological factors every five years from 1990 to 2015. The temperature at QJS varied the most (coefficient of variation = 10.49%); the range of precipitation variation was large for all regions (all coefficients of variation  $> 30\%$ ); the sunshine hours at MQX and QJS varied the most (coefficients of variation = 26.50% and 30.66%).

##### 4.3. Contributions of climate and human factors to NDVI changes

QJS is located in the monsoon climate zone where conditions were often rainy and cloudy, and the imaging times were later at TKLMG and KBQ. Therefore, the training points were less able to meet the RFR model requirements in those three study areas. We selected the other four study areas to quantitatively assess the contribution of meteorological factors and human factors to NDVI changes.

In order to further evaluate NDVIc changes, NDVIa from remote sensing images and nine meteorological parameters from 1984 to 1994 were used to construct the RFR model in this study. We tested the stability of the model to accurately predict NDVIc in four of the ecological engineering zones (Fig. 9). In those zones, the RFR model performed well (cross-validation  $R_{CV}^2 = 0.81, 0.78, 0.82,$  and  $0.76$ ;  $nRMSE_{CV} = 0.5\%, 0.79\%, 1.82\%,$  and  $1.76\%$ , respectively,  $n_{tree} = 500$ ,  $m_{try} = 3$ ) (Fig. 9). The RFR model reestablished NDVIa well from 1984 to 1994, and we therefore used NDVIa driven by nine meteorological factors as NDVIc to predict NDVIc from 2000 to 2015.

As mentioned earlier, the RFR model considered only nine major meteorological parameters instead of all of the environmental factors. The cross-validation results showed that  $R^2$  was not close to 1.00



**Fig. 6.** Changes in annual NDVI from 1990 to 2015 in seven study areas in China. NDVI data were missing from 1990 to 1993 at TKLMG. The gray shaded band represents the 95% confidence band for the linear regression. (KBQ, Kubuqi Desert; KEQ, Keerqin Sandy Land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).

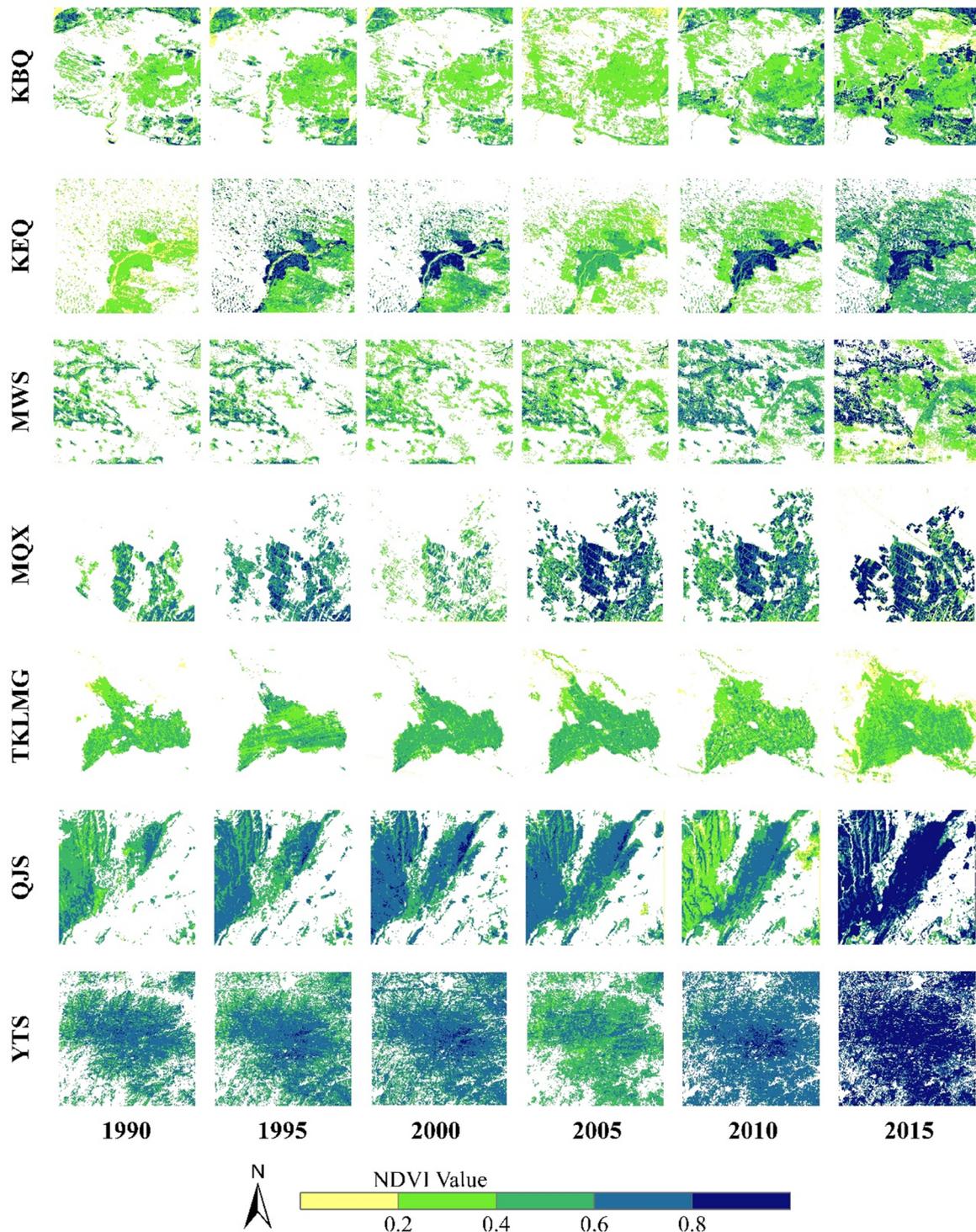


Fig. 7. Changes in NDVI from Landsat images of luxuriant vegetation from 1990 to 2015 at seven study areas in China. (KBQ, Kubeqi Desert; KEQ, Keerqin Sandy Land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).

because of unaccounted factors (Fig. 9). We also tested the variable importance of the nine climate factors (MMT, MMMT1, MMMT2, MMRH, MMP, MMWS, MMSH, TMP, TMSH). MMT, MMMT1, MMMT2, MMRH contributed more to model stability than the other climate factors (Fig. 10).

We then used the RFR model to predict NDVI<sub>c</sub> from 2000 to 2015, and obtained NDVI<sub>h</sub> by the difference of NDVI<sub>a</sub> and NDVI<sub>c</sub>. During the 2000–2015 period, human activities had a positive effect on NDVI growth at MWS, KEQ, YTS, and MQX ( $\Delta\text{NDVI}_{h,\%} = 100.02\%, 101.39\%$ ,

$124.86\%$ , and  $51.38\%$ , respectively) (Table 6). The contribution of climatic factors to NDVI growth was different in different regions (Table 6).  $\Delta\text{NDVI}_{c,\%}$  values at MWS, KEQ, and YTS were less than 0. This phenomenon indicates that human activities have a major controlling influence on NDVI growth, while climate factors play a negligible role in these areas. The  $\Delta\text{NDVI}_{c,\%}$  value at MQX was  $-24.86\%$  and indicated NDVI growth was a result of contributions of both human activities and meteorological factors (Table 6).

We analyzed the changes in NDVI<sub>c</sub> (affected by only climate) and

**Table 5.** Multiple linear regression results for annual NDVI changes in response to three meteorological factors at seven study areas in China.

		Estimate (and standard error)	ANOVA F-value	Pr (> F)	Relative importance	R <sup>2</sup>
KBQ	MAT	0.70 (0.16)	18.673	***	83.87%	0.51
	TAP	0.31 (0.15)	4.204	0.05	12.39%	
	TAS	0.01 (0.16)	0.003	0.96	3.74%	
KEQ	MAT	0.07 (0.20)	0.129	0.72	3.04%	0.29
	TAP	0.48 (0.20)	3.661	0.07	46.53%	
	TAS	0.43 (0.19)	5.275	*	50.43%	
MWS	MAT	0.54 (0.18)	8.081	**	75.48%	0.34
	TAP	0.33 (0.19)	3.099	0.09	20.87%	
	TAS	0.05 (0.19)	0.071	0.79	3.65%	
MQX	MAT	0.41 (0.16)	7.962	**	41.03%	0.45
	TAP	0.40 (0.16)	4.976	*	27.99%	
	TAS	0.36 (0.17)	4.687	*	30.98%	
TKLMG	MAT	0.11 (0.22)	1.021	0.33	14.74%	0.19
	TAP	0.20 (0.24)	0.001	0.98	7.52%	
	TAS	0.46 (0.25)	3.372	0.08	77.74%	
QJS	MAT	-0.23 (0.30)	1.015	0.32	15.01%	0.17
	TAP	0.22 (0.22)	0.035	0.85	9.15%	
	TAS	0.57 (0.31)	3.467	0.07	75.84%	
YTS	MAT	0.29 (0.23)	1.517	0.23	54.47%	0.15
	TAP	-0.06 (0.24)	0.003	0.96	3.65%	
	TAS	-0.28 (0.23)	1.416	0.25	41.88%	

Note: The *P*-values generated by the *F* test for the overall significance of the regression models in the seven different regions were 0.001, 0.05, 0.02, 0.004, 0.26, 0.24, and 0.43, respectively. Pr (> *F*) is the probability of an *F*-value greater than that given in the “*F*-value” column. Symbols for the significance test are: \* for *p* < 0.05, \*\* for *p* < 0.01, and \*\*\* for *p* < 0.001. MAT, TAP, and TAS are mean annual temperature, total annual precipitation, and total annual sunshine hours, respectively. (KBQ, Kubuqi Desert; KEQ, Keerqin Sandy Land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).

NDVIa (affected by both climate and human activity) every five years from 2000 to 2015 (Fig. 11). On the whole, the effect of human activities on NDVI growth increased over time, and the effect of human activities on NDVI growth was similar in all four regions (Fig. 11). Meteorological elements in different study areas had different effects on vegetation greenness, which may depend on local land use conditions, hydrothermal conditions, and vegetation types. However, vegetation development in different environments can be directed towards increased greening by implementation of production policies (human factors). Such policies may be beneficial for mitigating negative climate change impacts.

## 5. Discussion

### 5.1. Application of machine learning and residual analysis methods

This study used an RFR model to construct a relationship between NDVIc and monthly meteorological data from 1984 to 1994. NDVIc from 2000 to 2015 was then predicted by climatic factors. The residual between the observed and predicted values was considered to be the effect of human activities and non-considered factors on NDVI. We predicted about 80% of the variation in observed NDVI through 10-fold cross-validation. The key step of this method was to reconstruct the vegetation index affected by climatic factors. Therefore, it was very important to select an appropriate time period under natural conditions. Although the predicted period is usually determined by statistical methods to detect the year of an abrupt change, the period we chose was reasonable for two primary reasons: (1) The results of cross-validation showed that NDVI driven by climate factors was well reconstructed (Fig. 9). (2) Although land use types changed in some areas due to human activities before 1994, these changes were insignificant on a regional scale.

We advise caution when using traditional analysis methods to attribute causes of vegetation changes. First, users need to consider whether there is a linear relationship between the variables. If so, users will eliminate non-linear variables, thus ignoring the causal relationships of non-linear changes. There is less robustness to noise (e.g., the noise resulting from the impact of changes in the natural environment on vegetation) with traditional statistical models. Moreover, it is difficult to use several variables to explain more vegetation changes, especially in complex environments. Therefore, we chose the machine learning model. Machine learning models can overcome the effects of outliers and simulate non-linear vegetation changes but they lack a mechanistic explanation.

Previous research has also used residual analysis to separate human contributions (Li et al., 2017, 2020). We have made some changes on this basis. Specifically, we analyzed vegetation changes in small areas, with representative meteorological data from a single station in each study area. However, previous studies have used large-scale areas such as watersheds, plateaus, and countries (Li et al., 2017), or made spatial interpolations of meteorological factors and regression analysis in grid studies (Liu et al., 2018). Supplementary Fig. S9 clearly shows the process of the method.

The interpolation method itself has large errors and the calculation results also have large deviations. Therefore, we did not adopt the interpolation method, but used single station data. This method can better deal with the problem of local areas. However, using this method (i.e., without interpolation calculations) does not allow us to show the spatial distribution of relevant information and spatial dynamic variation characteristics. Wang et al. (2018) used linear regression of two meteorological factors and NDVI to analyze NDVI trends. In practice, it is difficult to assure a linear relationship between meteorological factors and NDVI because NDVI is affected by many different factors. This study used a machine learning method to build the non-linear relationship, which may further improve the prediction of NDVI.

We used monthly data to establish regression relationships because annual average data may reduce the ability to identify the effects of abnormal changes. For example, a single large precipitation event can have a large impact on vegetation growth in arid or semi-arid areas (Miao et al., 2014). Annual data is sometimes not able to adequately demonstrate the relationships between NDVI and meteorological parameters.

### 5.2. Vegetation greening and vegetative area expansion

This paper made some changes for the method of vegetation extraction. We used an SVM that supported less validation data to interpret vegetation distribution. Previous studies extracted vegetation by setting a threshold NDVI value (Neinavaz et al., 2020). However, environmental changes (such as solar elevation angle) can cause changes in vegetation NDVI (Ma et al., 2019). A larger threshold will underestimate vegetation area, and a smaller threshold will increase the chances of errors associated with extracting other land types.

From 1990 to 2015, vegetation area and vegetation greenness increased by different degrees in each study area (Table 2), with vegetation area increasing by more than 50% in all but one area. This result is consistent with the results from Zhu et al. (2016), in which the importance of considering land cover change was demonstrated in a greening analysis. Although the geographical environment, climate, and human activity in each research area were quite different, increases in vegetation and greening occurred simultaneously. However, NDVI decreased in one study area even though vegetation area increased. Perhaps at that particular site vegetation growth may have been inhibited by past precipitation or temperature changes, resulting in a decline in NDVI (Li et al., 2019). It is also possible that the negative impact of some human activities led to poor vegetation growth (Qiu et al., 2017).

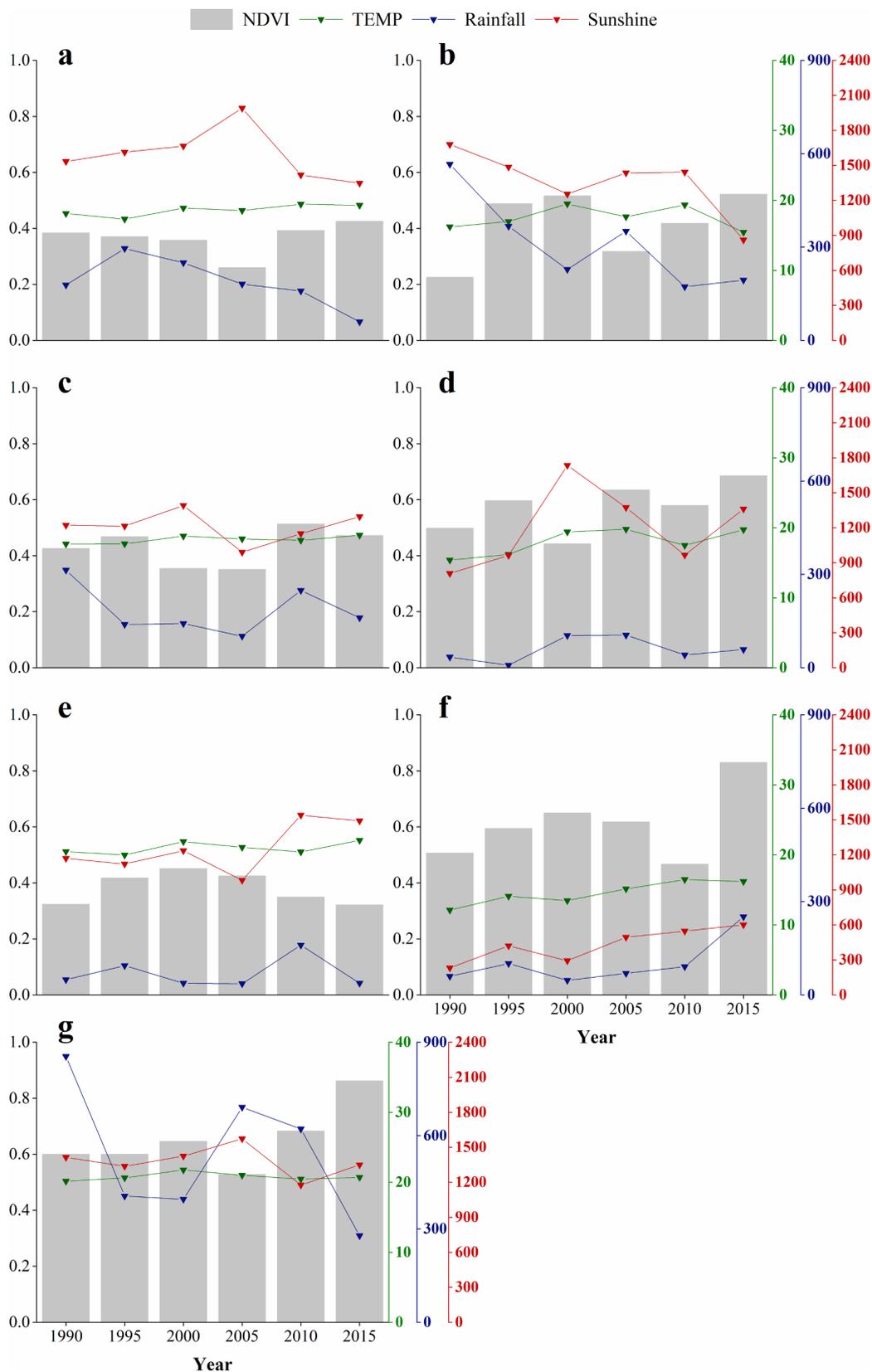
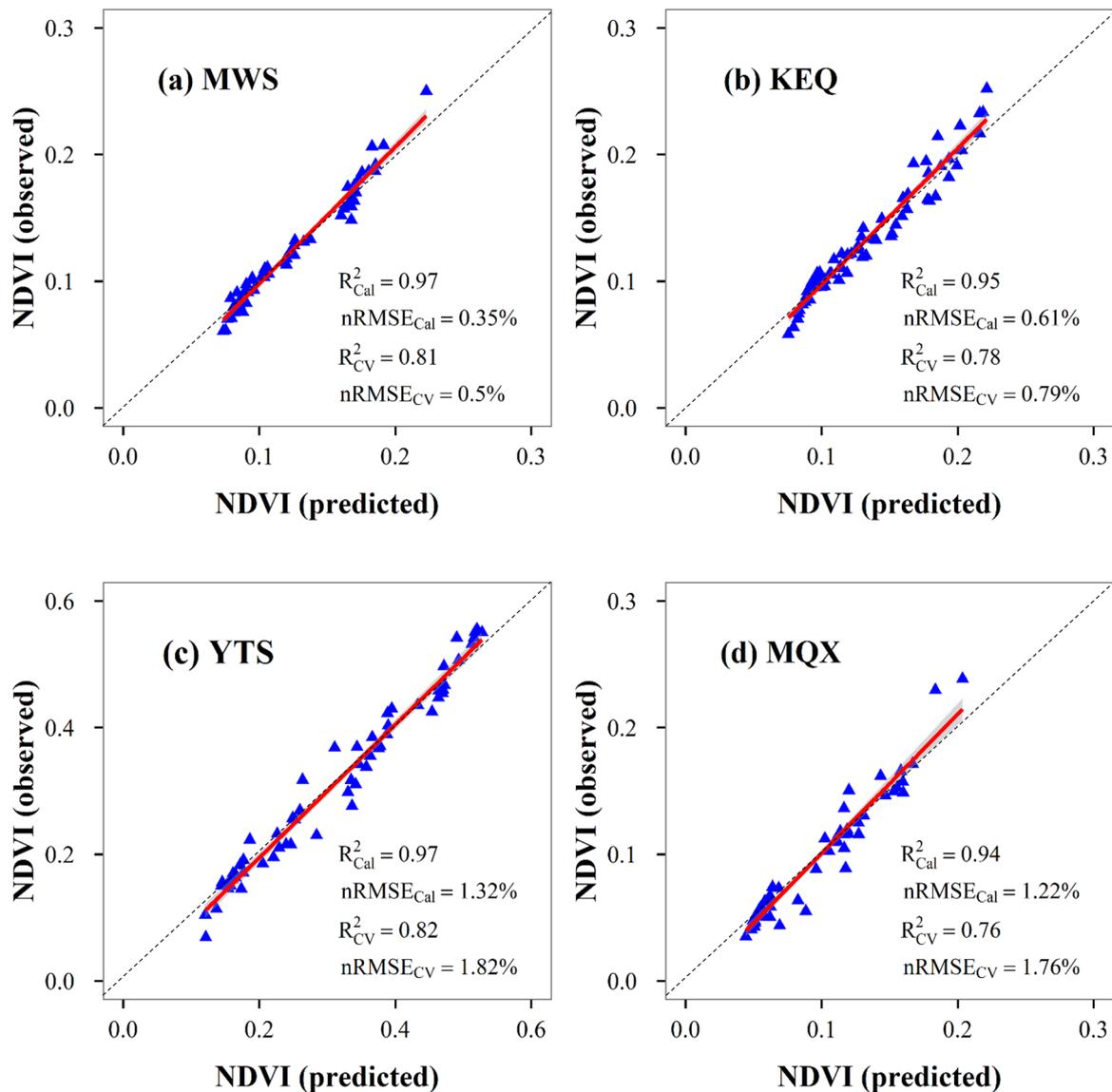


Fig. 8. Changes in annual NDVI and meteorological factors at seven study areas in China from 1990 to 2015. (a–g) Changes in NDVI and meteorological factors (temperature [°C], rainfall [mm], and sunshine hours) during the growing season (KBQ, Kubuqi Desert; KEQ, Keerqin Sandy Land; MWS, Maowusu Sandy; MQX, Minqinxian; TKLMG, Taklimakan Desert; QJS, Qujingshi; YTS, Yantaishi).



**Fig. 9.** Cross-validation results for the RFR model to reestablish NDVI in four study areas in China. NDVI during 1984–1994 was predicted using nine meteorological factors in the RFR model. The subscripts Cal and CV indicate results from calibration and cross-validation datasets, respectively. (MWS, Maowusu Sandy; KEQ, Keerqin Sandy Land; YTS, Yantaishi; MQX, Minqinxian).

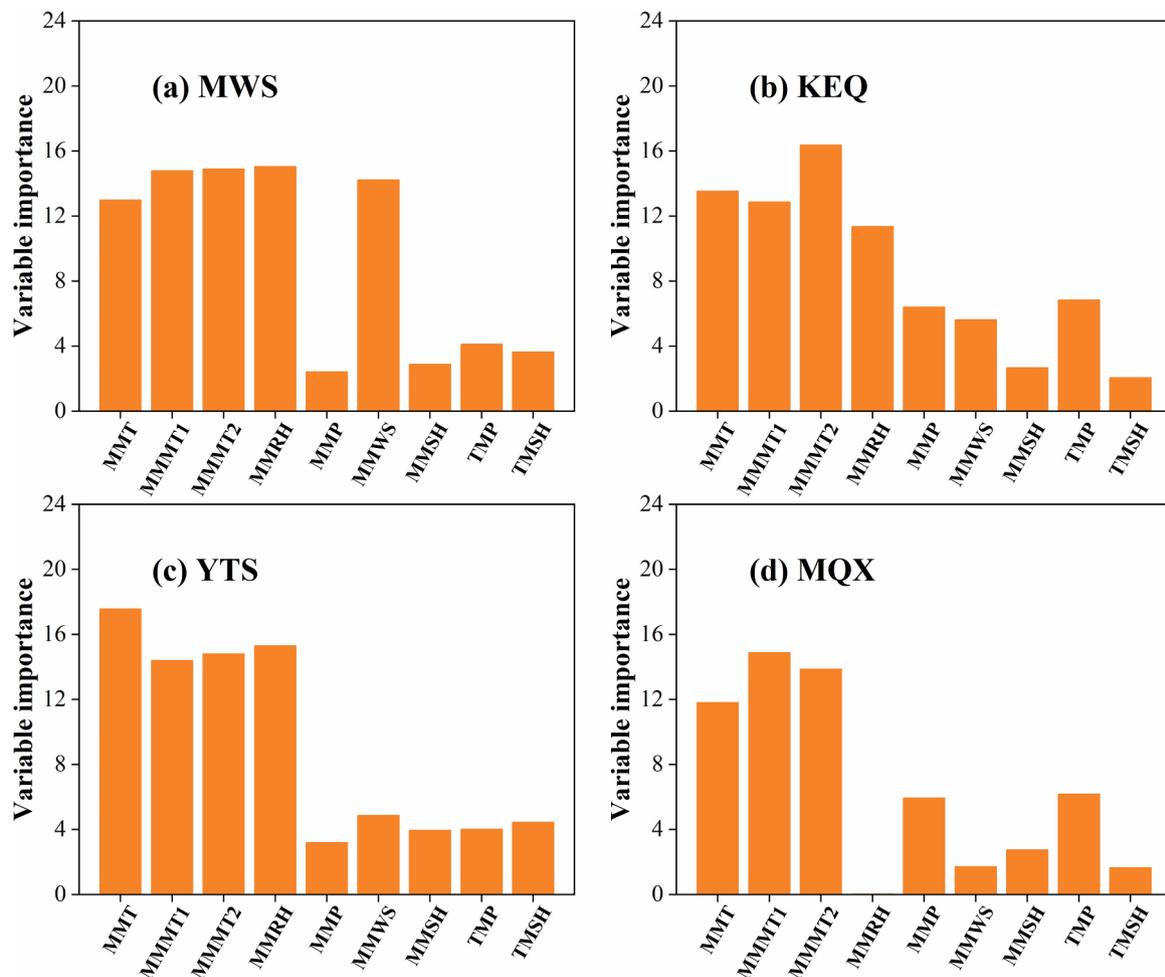
Vegetation greening is often associated with meteorological elements. This study selected three meteorological factors (solar radiation, temperature, and precipitation) that directly affect vegetation to further identify the important driving forces of vegetation greening. Meteorological factors will affect NDVI changes, and extreme drought and precipitation can greatly change the growth of plants (Ying et al., 2020). Wen et al. (2017) reported that solar radiation, temperature, and precipitation had significant effects on NDVI variations, and human activities (especially ecological engineering projects) had great influence on NDVI changes. However, some changes in NDVI are difficult to fully explain with meteorological factors, such as NDVI increases in specific years with low precipitation or high temperature (Zhang et al., 2016). Moreover, the combined effects of other meteorological and non-meteorological factors feed back into vegetation development (Wang et al., 2015). Collinearity among meteorological variables may also affect vegetation development. Therefore, it is difficult to qualitatively analyze the changes of vegetation NDVI with key meteorological factors.

This study quantitatively assessed the contribution of human activities to vegetation greening through residual analysis in four study

areas. Human activities may be caused by policy factors. China has implemented a series of forest and grassland policies (Fig. 1a) that have produced positive vegetation effects. NDVI has also increased greatly from 1998 to 2018 in China (Fig. 1b). Research from Qu et al. (2020) also evaluated the contribution of human activities to NDVI, and the results were highly consistent with this study. Wang et al. (2015) reported that vegetation coverage was improved or degraded by human activities, and the negative effects of extreme weather cannot be ignored. Changes in land use caused by ecological engineering can improve the vegetation situation (Qu et al., 2018).

### 5.3. Limitations and future perspectives

This study used three major meteorological factors related to vegetation growth to analyze NDVI changes. However, it is not only solar radiation, temperature and precipitation that affect vegetation growth and distribution. The NDVI residual value computed in this study contained not only human activity signals, but also residual climate effects because we used only nine meteorological parameters to build the RFR model. In future studies, we may add additional relevant



**Fig. 10.** Variable importance of meteorological factors in predicting NDVI at four study areas in China. The variable importance was determined by the mean decrease in accuracy. MMT, MMMT1, MMMT2, MMRH, MMP, MMWS, MMSH, TMP, and TMSH are mean monthly temperature, mean monthly minimum temperature, mean monthly maximum temperature, mean monthly relative humidity, mean monthly precipitation, mean monthly wind speed, mean monthly sunshine hours, total monthly precipitation, and total monthly sunshine hours, respectively. (MWS, Maowusu Sandy; KEQ, Keerqin Sandy Land; YTS, Yantaishi; MQX, Minqinxian).

**Table 6.**  
Mean NDVI changes in four study areas in China (2000–2015).

Area	$\Delta\text{NDVI}_c^b$	$\Delta\text{NDVI}_h$	$\Delta\text{NDVI}_a$	$\Delta\text{NDVI}_c\%$	$\Delta\text{NDVI}_h\%$
1 MWS <sup>a</sup>	-0.00002	0.10058	0.10056	-0.02	100.02
2 KEQ	-0.00188	0.13758	0.13570	-1.39	101.39
3 YTS	-0.04375	0.21970	0.17595	-24.86	124.86
4 MQX	0.05696	0.06020	0.11717	48.62	51.38

<sup>a</sup> (MWS, Maowusu Sandy; KEQ, Keerqin Sandy Land; YTS, Yantaishi; MQX, Minqinxian)

<sup>b</sup> Subscripts *c*, *h*, and *a* on NDVI refer to climate factors, human activities, and all factors.

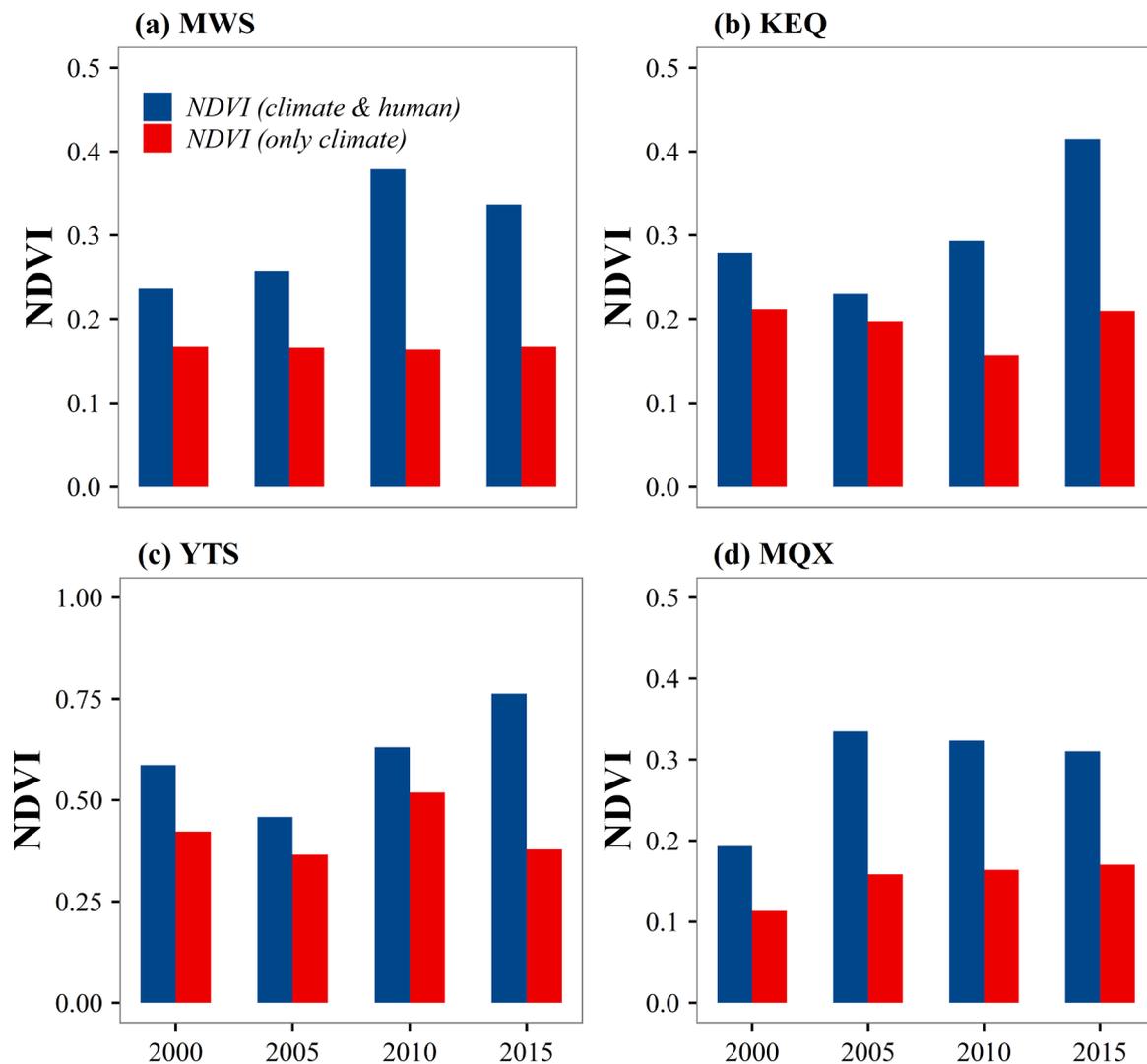
environmental variables to improve the adaptability of the model. An additional limitation of the current study is that ground verification was generally weak, as Google Image cannot be considered as ground truth. In future studies, we will use more actual ground-truth point data.

The vegetation change analysis and contribution assessment conducted in this study demonstrated that some uncertainties and limitations cannot be ignored. First of all, there was no trend analysis presented in this paper, and a relatively simple method was used to judge vegetation changes. In addition, the human factors affecting vegetation change were not explicitly determined, and it is difficult to judge which human activities have higher contributions to NDVI changes. For these

reasons, a greater degree of refinement for climate and human factors is necessary to establish a comprehensive model for quantifying the driving factors for changes in NDVI and vegetation amount and distribution. In the future, vegetation feedback effects will need to be considered to reflect vegetation changes more accurately.

## 6. Conclusions

This study used remote sensing images, meteorological data, DEM, and other data to quantify changes in vegetation area and greenness from 1990 to 2015 in selected ecological engineering zones in China. Our results indicated that the vegetation area and greenness increased in response to the ecological restoration projects across different climate regions in China from 1990 to 2015. There were different driving factors for vegetation changes in different projects. Specifically, temperature drove vegetation growth at KBQ, MWS, MQX, and YTS; sunshine hours promoted vegetation growth at KEQ, TKLMG, and YTS. However, the combined effects of climate factors had little effect on the growth of NDVI (and sometimes decreased growth), while human activities had a definitive impact, contributing more than 100% in most regions. The results of this paper can assist policy makers in knowing the ecological recovery status and the effect of ecological policies more clearly and accurately, allowing them to make timely adjustments to decision-making strategies. We used machine learning technology to



**Fig. 11.** NDVI as affected by only climate factors and by climate and human factors from 2000 to 2015 at four study areas in China. The red bars represent NDVI (affected by only climate factors) as predicted by the RFR model. The blue bars represent NDVI<sub>a</sub> (affected by climate and human factors) as assessed from actual data in remote sensing images. (MWS, Maowusu Sandy; KEQ, Keerqin Sandy Land; YTS, Yantaishi; MQX, Minqinxian).

improve the regression analysis and to obtain better results, reducing the additional uncertainties generated by traditional linear analysis, and accurately attributing the contributions of climate and human factors to changes in NDVI. This is an attractive method for understanding the impacts of human factors and environment on vegetative cover in local areas.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

This study was supported by Natural Science Foundation of China (No. 41961124006, 41730645), the International Partnership Program of the Chinese Academy of Sciences (161461KYSB20170013), Open Fund of State Key Laboratory of Remote Sensing Science (No. OFSLRSS201905), and “111 project” (No. B12007) of China. We thank Dr. Jie Zhao for data support in Land use database with 30-m spatial resolution, and Dr. Yan Zhang for technical support of image processing.

#### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.agrformet.2020.108146](https://doi.org/10.1016/j.agrformet.2020.108146).

#### References

- Adler-Golden, S., Matthew, M., Bernstein, L., Levine, R., Berk, A., Richtsmeier, S., Acharya, P., Anderson, G., Felde, J., Gardner, J., Hoke, M., Jeong, L., 1999. Atmospheric correction for shortwave spectral imagery based on MODTRAN4. *Proc. SPIE-Int. Soc. Opt. Eng.* 3753, 61–69.
- Adole, T., Dash, J., Atkinson, P.M., 2018. Large-scale prerain vegetation green-up across Africa. *Glob. Change Biol.* 24, 4054–4068.
- Bakwin, P.S., Davis, K.J., Jensen, N.O., Hollinger, D.Y., Falge, E.M., Law, B.E., Goldstein, A.H., Gu, L., Baldocchi, D.D., Berbigier, P., 2002. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. *Agric. For. Meteorol.* 113, 97–120.
- Baldocchi, D., Falge, E., Gu, L.H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X.H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K.T.P., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* 82, 2415–2434.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Chang, Y., Hou, K., Wu, Y.P., Li, X.X., Zhang, J.L., 2019. A conceptual framework for establishing the index system of ecological environment evaluation—A case study of the upper Hanjiang River, China. *Ecol. Indic.* 107, 105568.

- Chen, J., 2007. Rapid urbanization in China: a real challenge to soil protection and food security. *Catena* 69 (1), 1–15.
- Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R.K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tommervik, H., Bala, G., Zhu, Z., Nemani, R.R., Myneni, R.B., 2019a. China and India lead in greening of the world through land-use management. *Nat. Sustain.* 2, 122–129.
- Chen, T., Bao, A., Jiapaer, G., Guo, H., Zheng, G., Jiang, L., Chang, C., Tuerhanjiang, L., 2019b. Disentangling the relative impacts of climate change and human activities on arid and semiarid grasslands in Central Asia during 1982–2015. *Sci. Total Environ.* 653, 1311–1325.
- Clauss, K., Ottinger, M., Leinenkugel, P., Kuenzer, C., 2018. Estimating rice production in the Mekong Delta, Vietnam, utilizing time series of Sentinel-1 SAR data. *Int. J. Appl. Earth Obs. Geoinf.* 73, 574–585.
- Cong, N., Wang, T., Nan, H., Ma, Y., Wang, X., Myneni, R.B., Piao, S., 2013. Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: a multimethod analysis. *Glob. Change Biol.* 19, 881–891.
- Cortes, J., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297.
- Duan, H., Yan, C., Tsunekawa, A., Song, X., Li, S., Xie, J., 2011. Assessing vegetation dynamics in the Three-North Shelter Forest region of China using AVHRR NDVI data. *Environ. Earth Sci.* 64, 1011–1020.
- Du, J.Q., Fu, Q., Fang, S.F., Wu, J.H., He, P., Quan, Z.J., 2019. Effects of rapid urbanization on vegetation cover in the metropolises of China over the last four decades. *Ecol. Indic.* 107, 105458.
- Emmett, K.D., Renwick, K.M., Poulter, B., 2018. Disentangling climate and disturbance effects on regional vegetation greening trends. *Ecosystems* 22, 873–891.
- Evans, J., Geerken, R., 2004. Discrimination between climate and human-induced dryland degradation. *J. Arid. Environ.* 57, 535–554.
- Evans, K.L., James, N.A., Gaston, K.J., 2006. Abundance, species richness and energy availability in the North American avifauna. *Glob. Ecol. Biogeogr.* 15, 372–385.
- Fensholt, R., Langanke, T., Rasmussen, K., Reenberg, A., Prince, S.D., Tucker, C., Scholes, R.J., Le, Q.B., Bondeau, A., Eastman, R., Epstein, H., Gaughan, A.E., Hellden, U., Mbwo, C., Olsson, L., Paruelo, J., Schweitzer, C., Seaquist, J., Wessels, K., 2012. Greenness in semi-arid areas across the globe 1981–2007 – an Earth Observing Satellite based analysis of trends and drivers. *Remote Sens. Environ.* 121, 144–158.
- Franks, S., Masek, J.G., Turner, M.G., 2017. Monitoring forest regrowth following large scale fire using satellite data – a case study of Yellowstone National Park, USA. *Eur. J. Remote Sens.* 46, 551–569.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., Schmitt, M., 2005. A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personal. Soc. Psychol. Bull.* 31, 1369–1385.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Jing, W., Wang, K., Zhang, M., Zhang, C., 2015. Impacts of climate change and human activities on vegetation cover in hilly southern China. *Ecol. Eng.* 81, 451–461.
- Johnson, J.W., 2004. Factors affecting relative weights: the influence of sampling and measurement error. *Organ. Res. Methods* 7 (3), 283–299.
- Jong, R.D., Verbesselt, J., Schaepman, M.E., Bruin, S.d., 2012. Trend changes in global greening and browning: contribution of short-term trends to longer-term change. *Glob. Change Biol.* 18, 642–655.
- Kuenzer, C., Dech, S., Wagner, W., 2015. Remote sensing time series revealing land surfacedynamics: Status quo and the pathway ahead. *Remote Sens. Time Ser.* 1–24 Springer.
- Lamchin, M., Lee, W.K., Jeon, S.W., Wang, S.W., Lim, C.H., Song, C., Sung, M., 2018. Long-term trend and correlation between vegetation greenness and climate variables in Asia based on satellite data. *Sci. Total Environ.* 618, 1089–1095.
- Li, G., Sun, S., Han, J., Yan, J., Liu, W., Wei, Y., Lu, N., Sun, Y., 2019. Impacts of Chinese Grain for Green program and climate change on vegetation in the Loess Plateau during 1982–2015. *Sci. Total Environ.* 660, 177–187.
- Li, J., Peng, S., Li, Z., 2017. Detecting and attributing vegetation changes on China's Loess Plateau. *Agric. For. Meteorol.* 247, 260–270.
- Liu, J., Li, S., Ouyang, Z., Tam, C., Chen, X., 2008. Ecological and socioeconomic effects of China's policies for ecosystem services. *PNAS* 105, 9477–9482.
- Liu, J.Y., Kuang, W.H., Zhang, Z.X., Xu, X.L., Qin, Y.W., Ning, J., Zhou, W.C., Zhang, S.W., Li, R.D., Yan, C.Z., Wu, S.X., Shi, X.Z., Jiang, N., Yu, D.S., Pan, X.Z., Chi, W.F., 2014. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *J. Geog. Sci.* 24, 195–210.
- Li, K., Tong, Z., Liu, X., Zhang, J., Tong, S., 2020. Quantitative assessment and driving force analysis of vegetation drought risk to climate change: methodology and application in Northeast China. *Agric. For. Meteorol.* 282–283, 107865.
- Liu, Z., Liu, Y., Li, Y., 2018. Anthropogenic contributions dominate trends of vegetation cover change over the farming-pastoral ecotone of northern China. *Ecol. Indic.* 95, 370–378.
- Liu, L., Wang, Y., Wang, Z., Li, D., Zhang, Y., Qin, D., Li, S., 2019a. Elevation-dependent decline in vegetation greening rate driven by increasing dryness based on three satellite NDVI datasets on the Tibetan Plateau. *Ecol. Indic.* 107, 105569.
- Liu, Q.J., Zhang, H.Y., Gao, K.T., Xu, B., Wu, J.Z., Fang, N.F., 2019b. Time-frequency analysis and simulation of the watershed suspended sediment concentration based on the Hilbert-Huang transform (HHT) and artificial neural network (ANN) methods: A case study in the Loess Plateau of China. *Catena* 179, 107–118.
- Ma, X., Huete, A., Tran, N.N., 2019. Interaction of seasonal Sun-angle and savanna phenology observed and modelled using MODIS. *Remote Sens.* 11 (12), 19.
- Miao, L., Jiang, C., Xue, B., Liu, Q., He, B., Nath, R., Cui, X., 2014. Vegetation dynamics and factor analysis in arid and semi-arid Inner Mongolia. *Environ. Earth Sci.* 73, 2343–2352.
- Narain, D., Maron, M., Teo, H.C., Hussey, K., Lechner, A.M., 2020. Best-practice biodiversity safeguards for Belt and Road Initiative's financiers. *Nat. Sustain.* <https://doi.org/10.1038/s41893-020-0528-3>.
- Neinavaz, E., Skidmore, A.K., Darvishzadeh, R., 2020. Effects of prediction accuracy of the proportion of vegetation cover on land surface emissivity and temperature using the NDVI threshold method. *Int. J. Appl. Earth Obs. Geoinf.* 85, 101984.
- Niemand, C., Köstner, B., Prasse, H., Grünwald, T., Bernhofer, C., 2005. Relating tree phenology with annual carbon fluxes at Tharandt forest. *Meteorol. Z.* 14, 197–202.
- Peng, W., Kuang, T., Tao, S., 2019. Quantifying influences of natural factors on vegetation NDVI changes based on geographical detector in Sichuan, western China. *J. Clean. Prod.* 233, 353–367.
- Piao, S.L., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luysaert, S., Margolis, H., Fang, J., Barr, A., Chen, A., Grelle, A., Hollinger, D.Y., Laurila, T., Lindroth, A., Richardson, A.D., Vesala, T., 2008. Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* 451, 49–52.
- Qiu, B., Chen, G., Tang, Z., Lu, D., Wang, Z., Chen, C., 2017. Assessing the Three-North Shelter Forest Program in China by a novel framework for characterizing vegetation changes. *ISPRS J. Photogramm. Remote Sens.* 133, 75–88.
- Qu, S., Wang, L., Lin, A., Yu, D., Yuan, M., Li, C.a., 2020. Distinguishing the impacts of climate change and anthropogenic factors on vegetation dynamics in the Yangtze River Basin, China. *Ecol. Indic.* 108, 105724.
- Qu, S., Wang, L., Lin, A., Zhu, H., Yuan, M., 2018. What drives the vegetation restoration in Yangtze River basin, China: Climate change or anthropogenic factors? *Ecol. Indic.* 90, 438–450.
- Song, X.P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. *Nature* 560, 639–643.
- Sun, J., Yuan, X., Liu, G., Tian, K., 2019. Emergy and eco-exergy evaluation of wetland restoration based on the construction of a wetland landscape in the northwest Yunnan Plateau, China. *J. Environ. Manag.* 252, 109499.
- Tian, S., Zhang, X., Tian, J., Sun, Q., 2016. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sens.* 8 (11), 954.
- Tong, L., Liu, Y., Wang, Q., Zhang, Z., Li, J., Sun, Z., Khalifa, M., 2019. Relative effects of climate variation and human activities on grassland dynamics in Africa from 2000 to 2015. *Ecol. Inform.* 53, 100979.
- Tong, X., Brandt, M., Yue, Y., Horion, S., Wang, K., Keersmaecker, W.D., Tian, F., Schurgers, G., Xiao, X., Luo, Y., Chen, C., Myneni, R., Shi, Z., Chen, H., Fensholt, R., 2018. Increased vegetation growth and carbon stock in China karst via ecological engineering. *Nat. Sustain.* 1, 44–50.
- Vapnik, V., 2013. *The Nature of Statistical Learning Theory*. Springer Science & Business Media.
- Vidal-Macua, J.J., Nicolau, J.M., Vicente, E., Moreno-de Las Heras, M., 2020. Assessing vegetation recovery in reclaimed opencast mines of the Teruel coalfield (Spain) using Landsat time series and boosted regression trees. *Sci. Total Environ.* 717, 137250.
- Wang, F., An, P., Huang, C., Zhang, Z., Hao, J., 2018. Is afforestation-induced land use change the main contributor to vegetation dynamics in the semiarid region of North China? *Ecol. Indic.* 88, 282–291.
- Wang, J., Wang, K., Zhang, M., Zhang, C., 2015. Impacts of climate change and human activities on vegetation cover in hilly southern China. *Ecol. Eng.* 81, 451–461.
- Wen, Z., Wu, S., Chen, J., Lü, M., 2017. NDVI indicated long-term interannual changes in vegetation activities and their responses to climatic and anthropogenic factors in the Three Gorges Reservoir Region, China. *Sci. Total Environ.* 574, 947–959.
- Wessels, K.J., Prince, S.D., Malherbe, J., Small, J., Frost, P.E., VanZyl, D., 2007. Can human-induced land degradation be distinguished from the effects of rainfall variability? A case study in South Africa. *J. Arid. Environ.* 68, 271–297.
- Wu, Z., Wu, J., Liu, J., He, B., Lei, T., Wang, Q., 2013. Increasing terrestrial vegetation activity of ecological restoration program in the Beijing–Tianjin Sand Source Region of China. *Ecol. Eng.* 52, 37–50.
- Xiao, J.F., Moody, A., 2004. Trends in vegetation activity and their climatic correlates: China 1982 to 1998. *Int. J. Remote Sens.* 25, 5669–5689.
- Xu, X., Zhang, Y., Chen, Y., 2020. Projecting China's future water footprint under the shared socio-economic pathways. *J. Environ. Manag.* 260, 110102.
- Ying, H., Zhang, H., Zhao, J., Shan, Y., Zhang, Z., Guo, X., Rihan, W., Deng, G., 2020. Effects of spring and summer extreme climate events on the autumn phenology of different vegetation types of Inner Mongolia, China, from 1982 to 2015. *Ecol. Indic.* 111, 105974.
- Zaimes, G.N., Gounaridis, D., Symeonakis, E., 2019. Assessing the impact of dams on riparian and deltaic vegetation using remotely-sensed vegetation indices and Random Forests modelling. *Ecol. Indic.* 103, 630–641.
- Zhang, C., Wang, X., Li, J., Hua, T., 2020. Identifying the effect of climate change on desertification in northern China via trend analysis of potential evapotranspiration and precipitation. *Ecol. Indic.* 112, 106141.

- Zhang, Y., Peng, C., Li, W., Tian, L., Zhu, Q., Chen, H., Fang, X., Zhang, G., Liu, G., Mu, X., Li, Z., Li, S., Yang, Y., Wang, J., Xiao, X., 2016. Multiple afforestation programs accelerate the greenness in the 'Three North' region of China from 1982 to 2013. *Ecol. Indic.* 61, 404–412.
- Zhang, Y., Song, C., Band, L.E., Sun, G., Li, J., 2017. Reanalysis of global terrestrial vegetation trends from MODIS products: Browning or greening? *Remote Sens. Environ.* 191, 145–155.
- Zheng, K., Wei, J.Z., Pei, J.Y., Cheng, H., Zhang, X.L., Huang, F.Q., Li, F.M., Ye, J.S., 2019. Impacts of climate change and human activities on grassland vegetation variation in the Chinese Loess Plateau. *Sci. Total Environ.* 660, 236–244.
- Zhu, Z., Fu, Y., Woodcock, C.E., Olofsson, P., Vogelmann, J.E., Holden, C., Wang, M., Dai, S., Yu, Y., 2016. Including land cover change in analysis of greenness trends using all available Landsat 5, 7, and 8 images: A case study from Guangzhou, China (2000–2014). *Remote Sens. Environ.* 185, 243–257.