



Spatial pattern of cultivated land fragmentation in mainland China: Characteristics, dominant factors, and countermeasures

Sijing Ye^{a,b,1}, Shuyi Ren^b, Changqing Song^{a,b,*}, Zhenbo Du^c, Kuangxu Wang^b, Bin Du^b, Feng Cheng^d, Dehai Zhu^e

^a State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China

^b Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

^c Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

^d China Land Surveying and Planning Institute, Beijing 100035, China

^e College of Land Science and Technology, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Keywords:

Cultivated land fragmentation
Land use change
Arable land use intensity
Arable land health
Land computing
Grid

ABSTRACT

Systematically recognizing spatial patterns and driving factors of cultivated land fragmentation is of great significance for the exploration of locally appropriate path to relieve cultivated land fragmentation. This study aims to estimate cultivated land density, mean patch size and area-weighted mean shape index to respectively indicate the characteristics of cultivated land fragmentation from three dimensions, namely, natural resource endowment, spatial partition, and convenience of utilization. The regional leading factors of cultivated land fragmentation are also analyzed. The results demonstrate that the distribution of cultivated land density is higher in Northern regions compared with those of southern regions. The significant positive correlation between cultivated land density and mean patch size is found to be universal across nearly all cities, exceeding differences in terrain, elevation, climate, soil, and social economic condition. For cities in the southern part of China, cultivated land of regular shape is partitioned to smaller blocks compared with irregular ones; alternatively, intensive and meticulous farming under small-scale agricultural operation leads to clusters with low mean patch size - low area-weighted mean shape index. Random forest model explains the impact of driving factors on cultivated land fragmentation, with an explanatory power ranging from 66% to 95%. The terrain factor emerges as the primary driver negatively affecting cultivated land density. Gross domestic product emerges as the dominant factor with a significant ($p < 0.01$) negative correlation to mean patch size for nearly all agricultural climatic zones. Terrain, gross domestic product and population is the most important factor affecting area-weighted mean shape index. Rural development degree influences correlation between dominant factors and cultivated land fragmentation. This study is greatly instructive for recognizing the spatial patterns of cultivated land fragmentation at the national scale and for exploring the barriers that impede regionally scaled cultivated land use.

1. Introduction

Cultivated land fragmentation (CLF) refers to that driven by natural and human factors, cultivated land is divided into fragmented plots of different sizes with a discrete or disorderly land use pattern, which renders the large-scaled operation of agricultural production difficult (Ntihinyurwa, de Vries, 2020; Abubakari et al., 2016). CLF is a universal challenge related to sustainable cultivated land use around the world (Sklenicka, 2016; Lu et al., 2018; Liu and Li, 2017b). Despite the

advantage of CLF in terms of diversifying crop planting and resisting risks under a few specific conditions (e.g., labor surplus), an increasing number of empirical studies emphasize its negative effects, such as the restriction of application of agricultural machines (Chaudhary et al., 2020; Cao et al., 2020), reduction of crop yields and household income (Wan, Cheng, 2001; Tran, Vu, 2019), exacerbated habitat loss and cultivated land shrinking (Teillard et al., 2014; Qiu et al., 2015; Qian et al., 2020), aggravated cultivated land marginalization, and even hindrance of rural development (Long, 2014, 2019; Liu, 2018a; Looga

* Corresponding authors at: State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China.
E-mail address: songcq@bnu.edu.cn (C. Song).

¹ Ye Sijing (1988–), Associate Professor, specialized in arable land evaluation and protection.

et al., 2018). Thus, a systematic recognition of spatial patterns and the driving factors of CLF is of great significance for the exploration of locally appropriate path to relieve it and, thereby, promote intensive land use (Jin et al., 2024; Liu and Zhou, 2021).

Challenges related to the recognition of characteristics of CLF are complex. Initial studies focus on the influence of scattered landowners and scattered plot distribution related to CLF at the family scale from the economic perspective (Looga et al., 2018; Ciaian et al., 2018). CLF is mainly indicated by total household cultivated land area, land plot number, and average plot area (Falco et al., 2010; Demetris et al., 2013; Cao et al., 2020). Research data are mainly derived through questionnaire surveys and interviews, which are costly and lead to difficulty in achieving a large-scale estimation. The development of medium- to high-resolution remote sensing observation technology introduces support for data in estimating the characteristics of insufficient operation scale, spatial dispersion, and inconvenient farming conditions of cultivated land at the national or regional scale (Ye et al., 2018, 2022a; Liu et al., 2019b). Using land use/land cover (LULC) raster datasets, several landscape fragmentation indexes (e.g., mean patch size (MPS), patch density, largest patch, area-weighted mean shape index (AWMSI), the Simpson index, the Januszewski index, edge density, division index, and cohesion index) can be calculated to identify the multi-dimensional features of CLF, from the perspective of landscape ecology (Yan et al., 2016; Deininger et al., 2017; Looga et al., 2018). However, presenting CLF due to dispersive cultivated land ownership (or tenure) and diversified crop planting by using LULC datasets is difficult (Yu et al., 2018; Ye et al., 2020; Xu et al., 2021). Moreover, in certain cases, comprehensive estimation is conducted using the weighted average method to present the overall characteristic of CLF (Gonzalez et al., 2007; Hartvigsen, 2014; Liu et al., 2019a). The application of weighted average method exerts a smooth effect that may cover up serious shortcomings of CLF.

Another challenge lies in understanding the occurrence and development of CLF. On the one hand, the influence of land reform, inheritance custom, agricultural development and terrain features on CLF exhibits similarities across various study cases. Land reform events in history are generally considered to be the leading factors causing abrupt changes of CLF degrees (Kopeva et al., 1994; Hartvigsen, 2014; Viet Nguyen, Ngoc Tran, 2014). At the peasant household scale, family size and the proportion of labor force members within the family emerge as important factors influencing CLF (Tan et al., 2006). The presence of complex terrain features physically limits the contiguous distribution of cultivated land, and increases farming costs and risks, driving farmers to choose diversified planting patterns, thus exacerbating CLF (Ye et al., 2022a, 2023). On the other hand, the driving effect of natural—social—economic factors on CLF exhibits regional differences and multi-scale characteristics. At the scale of peasant households, factors such as household income, number of non-farmers, and the stress of parenting life may drive some households to participate in agricultural land circulation, leading to a reduction in CLF (Chen et al., 2010; Xie, Lu, 2017). But in mountainous areas, the driving effect of these factors is limited by the complex terrain features (Yucer et al., 2016; Qian et al., 2020). At the regional scale, an increase in GDP, urbanization rate, and the proportion of industry and services has a positive impact on improving the opportunity cost of farming and promoting agricultural land circulation. Nevertheless, these factors also contribute to the occupation of cultivated land by the construction of residential land, roads, industrial and mining facilities, and agricultural facilities, resulting in an increase in CLF (Liu et al., 2019a). For certain study cases in developed provinces in plain areas (e.g., Jiangsu province of China), the influence of the distance from the nearest water plot and density of water network have been verified to be non-significant (Liu et al., 2019a). Conversely, these factors have been found to significantly exacerbate CLF in multiple mountainous case studies (Guo et al., 2017; Wang, Xu, 2022a). This phenomenon can be explained by Liu and Jin's theory (2022), which suggests that for areas in the "extreme rural stage",

CLF is more influenced by spatial location and less affected by external social and economic factors.

The overall goal of the current study is to estimate spatial patterns in CLF and identify regional driving factors at the national scale by taking mainland China as case study. Cultivated land density, mean patch size (MPS) and area weighted mean shape index (AWMSI) are used to indicate the characteristics of CLF from three dimensions, namely, natural resource endowment, spatial partition, and convenience of utilization, respectively. This study contributes in two main ways. First, it presents the national spatial pattern of CLF indicators using a 1-km grid as the basic statistical unit. The initial spatial data on cultivated land are organized in a vector format and derived from the second national land use survey of China. This approach enables a better description of the spatial heterogeneity of CLF within administrative units. And it provides a clearer depiction of the distribution characteristics of cultivated land plots compared to the land use type data based on remote sensing classification. Second, the study employs a clustering method to express spatial clustering patterns of different CLF dimensions, providing a more effective representation of the main shortcomings of regional CLF compared to the traditional weighted average method. Additionally, by using the Random Forest model and partial correlation analysis method, the study identifies the regional dominant factors of CLF and offers a broader understanding of the spatial differences in their influence on CLF. This study is mainly instructive for recognizing the spatial patterns of CLF at the national scale and for exploring barriers that impede regionally scaled cultivated land use. It can also provide a narrative of China's experiences in CLF management that can serve as a model for other countries.

2. Material and methods

2.1. Data

The indicators of cultivated land fragmentation (CLF) are calculated on the basis of vector data of cultivated land plots. The initial spatial data on cultivated land plots are organized in vector format from the second national land use survey of China, which was conducted on 2007–2009. The total quantity of plots is approximately 67 million. Using the vector data of cultivated land plots, the study intends to elucidate CLF due to decentralized land use rights and diversified planting structure compared to using raster data on cultivated land cover. The division of the study area refers to the agricultural climatic zoning plan in China. Toward this end, the study calculates six factors (i. e., elevation; slope; distance to urban boundary; farming distance; GDP; population) within 1-km grids to quantitatively estimate the driving forces of regional CLF. Raster images on elevation and slope are derived from the ASTER GDEM V3 dataset. The distance of each grid to urban boundary is indicated by the Euclidean distance to the nearest urban—rural boundary grid for urban land use data for 2010 within a 1-km resolution (He and Liu, 2018b, 2019). For each cultivated land plot, farming distance is indicated by the Euclidean distance to the nearest rural residential area (excluding roads). Farming distance within the 1-km resolution grid is calculated as the average farming distance of all cultivated land plots inside the grid (Liu et al., 2023). Data on rural residential areas are extracted from the GlobeLand30 dataset for 2010 (Chen et al., 2017). Data on GDP and population based on the 1-km grid for 2010 are obtained from the Resources and Environmental Science Data Center (Xu, 2017a, 2017b). Table 1 provides detailed information.

2.2. Generation of 1-km grid-based cultivated land fragmentation (CLF) maps

The study categorizes the initial dataset on cultivated land plots by taking the county as data organization unit (i.e., cultivated land plots in each county correspond to one independent spatial data file [format: *.shp]). The study utilizes various spatial coordinate references of these

Table 1

Detailed dataset information related to estimate cultivated land fragmentation indicators and analyze their driving factors.

Dataset	Definition	Data source	Applications
Cultivated land plots	Cultivated land plots of each county generated by remote-sensing and field investigation	Second national land survey of China, performed during 2007–2009 (Cheng et al., 2014)	(Ye et al., 2014, 2019, 2020, 2022b; Yao et al., 2016)
China agricultural climatic zoning data	Nine agricultural zones have been generated according to climatic and topographical conditions	Resource and environment science and data center [http://www.resdc.cn/]	(Ren et al., 2022, 2023; Ye et al., 2020)
ASTER GDEM V3 dataset	Digital elevation model dataset in 30-meter spatial resolution, published by NASA and METI	Geospatial Data Cloud site (http://www.gscloud.cn)	(Ye et al., 2017; Wan et al., 2021)
1-km resolution urban land use data	Global urban land in 1992, 1996, 2000, 2006, 2010, 2016. In each data, the digital number of 1 denotes urban land	(He and Liu, 2018, 2019) (https://doi.org/10.1088/1748-9326/aaf936)	(Huang et al., 2021; Kuang et al., 2022; Ye et al., 2022a)
2010 GlobeLand30 dataset	A peer-reviewed 30-metre resolution global land cover dataset in 30-metre spatial resolution	(Chen et al., 2014, 2017)	(Yang et al., 2017; Lu et al., 2016; Arsanjani et al., 2016)
1-km resolution GDP dataset	County-level GDP data is scaled down to kilometer-grid units, in 1995, 2000, 2005, 2010, 2015 and 2019	(Xu, 2017a) (http://www.resdc.cn/, doi:10.12078/2017121101)	(Adnan et al., 2020; Zhang et al., 2022)
1-km resolution population density dataset	County-level demographic data is scaled down to kilometer-grid units, in 1995, 2000, 2005, 2010, 2015 and 2019	(Xu, 2017b) (http://www.resdc.cn/, doi:10.12078/2017121101)	(Liu, Wang, 2016; Carrao et al., 2016)

files following the Gauss 3-degree zonal projection and ellipsoid parameters of GCS 2000. Hence, merging the vector data files of numerous cultivated land plots into one grid map is difficult (Ye et al., 2016; Yao et al., 2017; Wang et al., 2022b).

Ye et al. (2022a) generated a 1-km grid-based cultivated land potential yield map using the Raster Dataset Clean and Reconstitution Multi-Grid (RDCRMG) architecture to guide data reorganization (Ye et al., 2018). A similar method was used to generate 1-km grid-based CLF maps. According to RDCRMG, a 1-km square grid map that cover mainland China was generated based on the Albers equal-area conic projection: “grids in the same level have been generated line by line with uniform size, shape and orientation” (Ye et al., 2018). The study integrated a strict grid coding rule and a storage strategy called *no metadata* model-based files for rapid data retrieval and merging. For each grid, cultivated land density, MPS, and AWMSI has been calculated. Cultivated land density and MPS increase as with the decrease in the degree of CLF, whereas an increase in AWMSI indicates an increased degree of CLF. The subsequent text outlines the generation process of the 1-km grid-based CLF maps. Steps 1 to 4 can be executed in parallel among multiple grids.

Step 1: convert the spatial coordinate system of the vector data file of cultivated land plots for each county from the Gauss 3-degree zonal projection to the Albers equal-area conic projection.

Step 2: generate 1-km square grid maps according to the bounding box of mainland China based on the Albers equal-area conic projection.

Step 3: for each 1-km square grid, reorganize cultivated land plots

that intersect with the grid into one unique vector data file through data extraction and merging. The original shape of each cultivated land plot should not be clipped.

Step 4: extract the corresponding vector data file (if any) of each 1-km square grid; calculate *MPS* using Eq. (1) and assign the result to the grid; and output a 1-km grid-based *MPS* map. *A* presents the total cultivated land area of a specific vector data file (i.e., grid); *N* presents the quantity of cultivated land plots. Similarly, 1-km grid-based *AWMSI* map can be generated using Eq. (2). For specific cultivated land plot *i*, its perimeter and area are presented as *P_i* and *a_i*, respectively. Lastly, the corresponding vector data file of each grid has been clipped using the bounding box of the grid. The cultivated land density of each grid is then calculated as the proportion of cultivated land area to the total area of the grid (i.e., 1 km²).

$$MPS = A/N \quad (1)$$

$$AWMSI = \sum_{i=1}^N \frac{0.25 \times P_i}{\sqrt{a_i}} \times \frac{a_i}{A} \quad (2)$$

2.3. Pearson correlation coefficient

For this study, for each prefecture-level city, Pearson correlation coefficient *r* between any two variables (be expressed as *x* and *y*) of the three cultivated land fragmentation indicators (i.e. mean patch size; area weighted mean shape index; cultivated land density) has been calculated, as Eq. (3) shows (Rodgers, Nicewander, 1988). σ_x and σ_y are standard deviations of *x* and *y*, and \bar{x} and \bar{y} are mean value of *x* and *y*, and *m* is the number of cultivated land grids in the prefecture-level city. Then we ran tests of *t* distribution for each prefecture-level city to evaluate the significance of *r*, according to Eq. (4).

$$r = \frac{1}{m-1} \sum_{i=1}^m \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right), \quad (3)$$

$$t = \frac{|r - 0|}{\sqrt{\frac{1-r^2}{v}}} \sim T(m-2), \quad v = m-2, \quad (4)$$

2.4. k-means clustering algorithm

The k-means clustering algorithm was originally derived from a vector quantization method in signal processing and is one of the most widely used algorithms in cluster analysis (MacQueen, 1967). The algorithm is easy to describe, and its processing efficiency is high in large datasets. Moreover, it can produce a good clustering result when the sample distribution is close within the classes and far away between the classes. It has been widely used in soil analysis and natural language processing at home and abroad (Wilpon and Rabiner, 1985; Kanungo et al., 2002; Brus et al., 2006). The core idea is to divide *n* samples into *k* cluster categories, minimizing the sum of the square distance of each sample with its cluster center, i.e., within-cluster sum of squares is the smallest sum that satisfies the clustering result (Eq. (5)), where *x* is the sample value; *Y* is the category set; *y_j* is a category in the category set, and *z_j* is the cluster center in the category *y_j*.

$$\arg \min \sum_{j=1}^k \sum_{x \in y_j} \|x - z_j\|^2 \quad (5)$$

The algorithm steps are as follows:

- (1) Determine the appropriate number of categories *k* based on prior knowledge. The category set *Y* is {*y*₁, *y*₂, ..., *y*_{*k*}};
- (2) The observation sets {*x*₁, *x*₂, *x*₃, ..., *x*_{*n*}} are the known samples while the optional *k* samples are given as the initial cluster center {*z*₁, *z*₂, ..., *z*_{*k*}}, corresponding to *k* number of categories;

- (3) Calculate the distance between the sample x_i and each cluster center z_j . Determine the closest cluster center z_j , and assign x_i to the category y_j to which z_j belongs;
- (4) After all samples are allocated, the samples in each category y_j ($j = 1, 2, \dots, k$) are recalculated to obtain a new cluster center z_j ;

Finally, if all the category center positions $\{z_1, z_2, \dots, z_k\}$ remain unchanged and the results tend to converge, then the output category set is $\{y_1, y_2, \dots, y_k\}$. Otherwise, go to step (3) for further iterative calculations until the classification result converges.

2.5. Random Forest model

Random Forest (RF) is an integrated machine learning model that aggregates multiple decision trees to make repeated predictions (Breiman, 2001). Large-scale data sets can be handled efficiently and randomness is introduced to make overfitting less likely. For each decision tree, the random forest method performs self-sampling so that error estimates can be based on out-of-bag sample data. During tree generation, each split node of each tree is generated randomly and its partition consists of few variables. When the random forest is used to solve a regression problem, the average of the results of each decision tree becomes the predicted value.

The steps of applying the random forest model are as follows:

- (1) The training set and the test set are divided according to the ratio of 75% and 25%.
- (2) The Bootstrap algorithm is applied, and n samples are drawn from the training set in a put-back manner from among N samples.
- (3) When each node of the decision tree needs to be split, m variables are randomly selected from M variables, satisfying the condition $m \ll M$. The Gini index is chosen as the information gain strategy for node splitting.
- (4) Each decision tree is recursively split from top to bottom until the termination condition is satisfied.
- (5) The predicted model is applied to the test set, and the accuracy is evaluated using R^2 and Root Mean Square Error (RMSE) as evaluation metrics.
- (6) The importance score of the variables is obtained by reordering the eigenvalues of a column in the trained model and observing how much the accuracy is reduced. For unimportant variables, being eliminated has little effect on the accuracy of the model, while the opposite is true for important variables.

2.6. Partial correlation coefficient

Cultivated land systems are subject to multiple influences and the correlation between related variables is complex. Thus, the direct study of simple correlation coefficients often does not correctly describe the relationship between variables. In this case, partial correlation analysis can be applied to measure the degree of linear correlation between two variables, controlling for the effects of others. The partial correlation coefficient can be calculated using Eq. (6).

$$\rho_{XY|Z} = \text{cor}(\hat{\epsilon}, \hat{\delta}) = \frac{\text{cov}(\hat{\epsilon}, \hat{\delta})}{\sqrt{\text{var}(\hat{\epsilon})} \sqrt{\text{var}(\hat{\delta})}} \quad (6)$$

where $\rho_{XY|Z}$ is the partial correlation coefficient of the control variables Z for variables X and Y , $\hat{\epsilon}, \hat{\delta}$ are the residuals of the multiple linear regression established between X, Y and Z , respectively. $\text{cov}(\hat{\epsilon}, \hat{\delta})$ is the correlation coefficient of $\hat{\epsilon}, \hat{\delta}$, $\text{cov}(\hat{\epsilon}, \hat{\delta})$ is the covariance of $\hat{\epsilon}, \hat{\delta}$, and $\text{var}(\hat{\epsilon}), \text{var}(\hat{\delta})$ are the variances of $\hat{\epsilon}, \hat{\delta}$. The residuals $\hat{\epsilon}, \hat{\delta}$ eliminate the

linear correlation between X, Y and Z respectively. Thus, calculating the correlation coefficient between $\hat{\epsilon}, \hat{\delta}$ gives its partial correlation coefficient.

3. Results

3.1. Spatial pattern of Cultivated Land Fragmentation (CLF) in mainland China

Fig. 1(a) indicates that level of cultivated land density is higher in plain regions compared with those of mountainous regions and higher in the Northern regions than those of the southern regions. Cultivated land distribution in the “four plains of China” (i.e., the Northeast China Plain; the Huang–Huai–Hai Plain; the Middle-Lower Yangtze Plain; the Central Shaanxi Plain) exhibits various characteristics, such as high aggregation, which indicate the sizable endowment of natural resources to support an increase in the scale of agricultural operation. A similar, highly concentrated distribution of cultivated land is also observed in the Chengdu Plain. The spatial distribution of cultivated land density denotes significant global autocorrelation features (Moran's $I = 0.574586$, $P < 0.001$, $Z = 1225.333528$). It indicates that that the contiguous grids of those with high levels of cultivated land density also tend to display high levels of cultivated land density and form high–high clusters and vice versa (low–low cluster). The local spatial autocorrelation index (Fig. 1(b)) illustrates that the high–high cluster is mainly distributed in flat areas of the Northeast China Plain, Huang–Huai–Hai Plain, Sichuan Basin and surrounding regions, and the Central Shaanxi Plain in the Loess Plateau. For the Middle-Lower Yangtze Plain, the high–high cluster only covers parts of the north and midlands. The other high–high cluster regions are distributed in the southern part of the Northern arid and semiarid region, the central and southeast part of the Yunnan–Guizhou Plateau. The distribution of low–low cluster areas is highly consistent with the those of the hilly and mountainous regions in China, especially Xing'an Mountains, Yanshan–Taihang–Qinling Mountains, Wushan–Xuefeng Mountains, Wuyi Mountains, and Hengduan Mountains.

Fig. 1(c) depicts that the distribution of MPS tends to be consistent with those of cultivated land density in the Northeast China Plain and the Huang–Huai–Hai Plain. It demonstrates that in these regions, a suitable farming scale has been implemented with the support of the high levels of natural resource endowment. For the hilly and mountainous regions in the Loess Plateau, Southern China, and the Yunnan–Guizhou Plateau, MPS is relatively low due to the low levels of natural resource endowment. For the Middle-Lower Yangtze Plain, the Central Shaanxi Plain, and the Chengdu Plain, initially concentrated cultivated land has been spatially partitioned into numerous small patches and form a contrast between high levels of cultivated land density and low level of MPS. The reason is the equal allocation of cultivated land given their soil quality and farming conditions, which are driven by the policy on the household contract responsibility system and diverse patterns of crop planting. The level of MPS is relatively high in the Northern arid and semiarid region, whereas the level of cultivated land density is low, because the population is relatively smaller in this area than those in the southeastern coastal area. Moreover, in the Yunnan–Guizhou Plateau and Southern China (Fig. 1(b)), several high–high cluster regions of cultivated land density exist; however, the level of MPS in these regions is low. The reason underlying this result is that cultivated land in these regions is “aggregated” only from the plane perspective. In fact, the wavy terrain has partitioned cultivated land into many patches (e.g., terraced fields). The spatial distribution of MPS point to the significant features of global autocorrelation (Moran's $I = 0.473866$, $p < 0.001$, $z = 464.548603$). According to Fig. 1(d), the high–high and low–low clusters of MPS indicate a north–south divide.

As shown in Fig. 1(e), the Northeast China Plain and Huang–Huai–Hai Plain illustrate the highest surface regularities in the

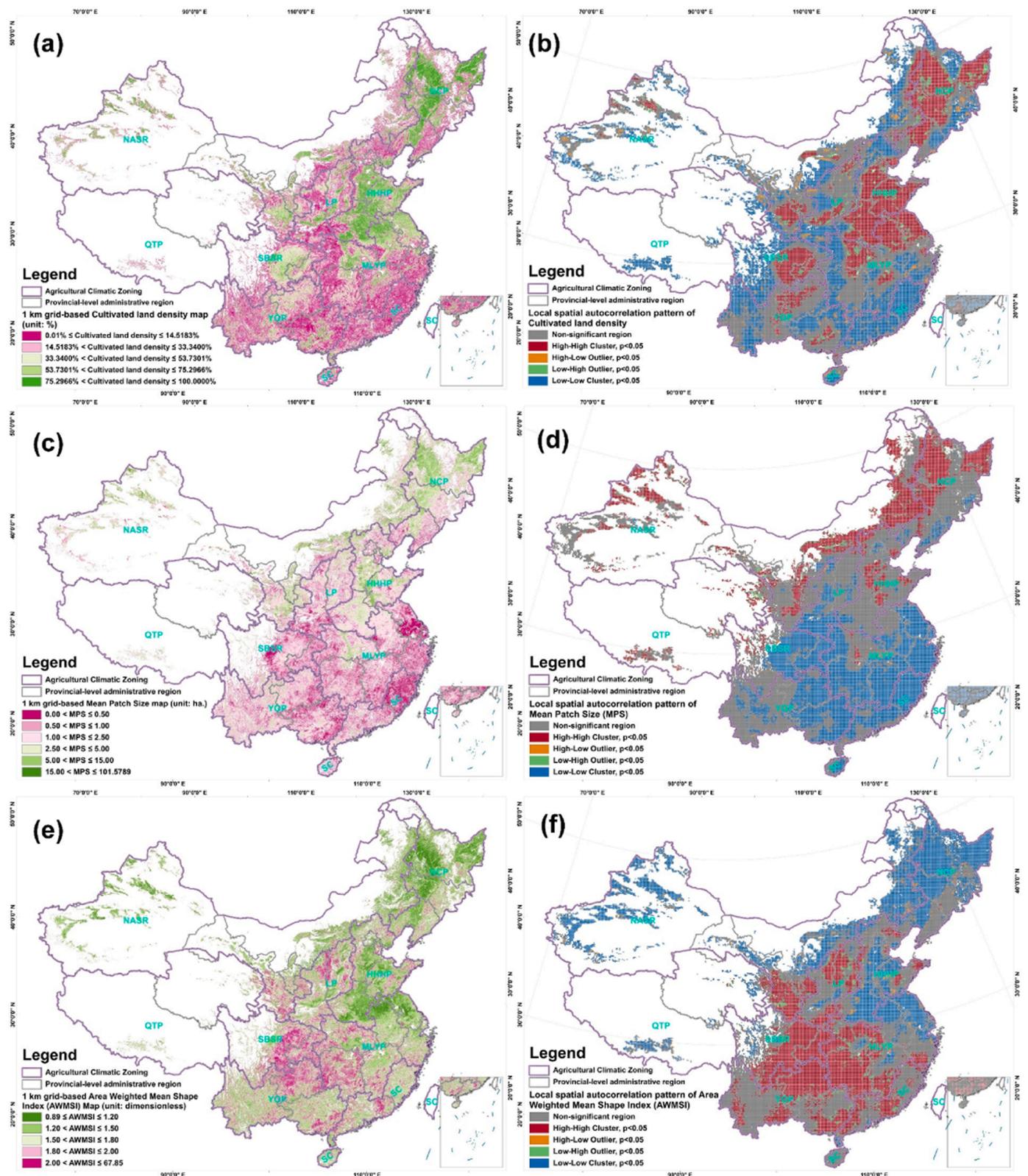


Fig. 1. (a) 1 km grid-based cultivated land density map (unit: %). Jenk's natural breaks method had been used for the classification of cultivated land density. (b) 10 km grid-based local spatial autocorrelation pattern of cultivated land density. (c) 1 km grid-based cultivated land mean patch size (MPS) map (unit: ha.). The numerical section of MPS has considered the fractile quantiles. (d) 10 km grid-based local spatial autocorrelation pattern of cultivated land MPS. (e) 1 km grid-based cultivated land area weighted mean shape index (AWMSI) map (unit: dimensionless). The numerical section of AWMSI has considered the fractile quantiles. (f) 10 km grid-based local spatial autocorrelation pattern of cultivated land AWMSI. Abbreviation of Agricultural Climatic Zoning: NCP (Northeast China Plain); NASR (Northern arid and semiarid region); HHHP (Huang-Huai-Hai Plain); LP (Loess Plateau); QTP (Qinghai Tibet Plateau); MLYP (Middle-lower Yangtze Plain); SBSR (Sichuan Basin and surrounding regions); YGP (Yunnan-Guizhou Plateau); SC (Southern China). Blank area expresses few cultivated land (e.g. Qinghai-Tibet Plateau, urban region) or no data (e.g. Taiwan).

formation of cultivated land plots, which indicates added convenient conditions for the use of agricultural machinery. The AWMSI of the Northern arid and semiarid region and the Middle-Lower Yangtze Plain is slightly higher than those the Northeast China Plain and Huang–Huai–Hai Plain. Cultivated land with the highest AWMSI is concentrated in the eastern part of the Sichuan Basin, western part of the Middle-Lower Yangtze Plain, eastern part of the Yunnan–Guizhou Plateau, and the Northern part of the Loess Plateau. The local spatial autocorrelation index results in Fig. 1(e) highlights the northeast–southwest divide of AWMSI. The spatial distribution of AWMSI demonstrates significant global autocorrelation features (Moran's $I = 0.630093$, $p < 0.001$, $z = 1111.597444$).

3.2. Pearson's correlation coefficients among CLF indicators

Pearson's correlation coefficients are calculated to present the order of degree of the spatial interaction among multi-dimensional CLF indicators from a holistic perspective at the scale of prefecture-level cities. As shown in Fig. 2(a), Pearson's correlation between cultivated land density and AWMSI points to evident differences between east and west. Prefecture-level cities in the *four plains of China* (i.e., the Northeast China Plain; Huang–Huai–Hai Plain; Middle-Lower Yangtze Plain; Central Shaanxi Plain) tend to exhibit significant weak or medium negative correlation ($p < 0.05$). This result demonstrates that grids in these cities with high cultivated land density tend to display regular shapes and, thereby are more convenient for mechanized farming compared with those with lower ones. From this perspective, cities with lower values in Pearson's correlation index exhibit more rational use of cultivated land than those of higher correlation index. In cities located in the Sichuan, Chongqing, Yunnan, and Guizhou provinces, the relationship between cultivated land density and AWMSI tend to exhibit a statistically significant ($p < 0.05$) low-to-medium positive correlation or an extremely low correlation. Compared to the negatively correlated cities, the terrain of positively correlated cities is more complex and varied, and rural development is also relatively backward, which increases the difficulty of cultivated land consolidation.

As shown in Fig. 2(b), Pearson's correlation between cultivated land density and MPS displays a generally significant weak or medium positive correlation ($p < 0.05$). This finding demonstrates that grids with high cultivated land density for the majority of cities tend to be partitioned to relatively large cultivated land plots, which indicates larger agricultural operation scales. From this perspective, cities with higher values in Pearson's correlation index between cultivated land density and MPS exhibit more rational use of cultivated land than those with lower values. The spatial pattern revealing a significant positive correlation between cultivated land density and MPS appears to be universal across nearly all cities, transcending variations in terrain, elevation, climate, soil, and socioeconomic conditions. Cities with a significant medium positive correlation are distributed widely in a "Y" shape.

Fig. 2(c) denotes that Pearson's correlation between MPS and AWMSI points to evident differences between north and south. Prefecture-level cities in the southern part of China tend to obtain significantly weak or medium positive correlation ($p < 0.05$). In these cities, cultivated land of regular shape has been partitioned to smaller blocks than those of irregular ones. Alternatively, intensive and meticulous farming under small-scale agricultural operation leads to clusters of low MPS–low AWMSI; thus, cultivated land with higher levels of MPS is more likely driven by inappropriate land expansion and in a state of extensive utilization with high levels of AWMSI. From this perspective, cities with lower values in Pearson's correlation index between MPS and AWMSI display more rational use of cultivated land than those cities with higher values. Prefecture-level cities with a weak negative correlation are mainly distributed along a line from Heilongjiang province to Hubei province, where local grids with higher levels of MPS tend to be matched with more convenient conditions for mechanized farming.

3.3. Clustering characteristics of CLF indicators

The clustering characteristics of the three CLF indicators have been grouped into seven types using the k-means algorithm (see *SI Appendix A.1* for the *Davies–Bouldin index and ANOVA result*). As depicted in Fig. 3, clustering type F indicates grids with the lowest degrees of CLF (i.e., high density; low AWMSI; high MPS), which should be valued and protected. Type F grids account for 15.38% of the total cultivated land quantity. A total of 92.96% of cultivated land in type F is distributed in the Northeast China Plain, Northern arid and semiarid region, and the Huang–Huai–Hai Plain. The core shortcoming of CLF for cluster type B is low MPS driven by excessive spatial partition. Natural resource endowment and the degree of the shape regularity of the spatial form of cultivated land under type B are denoted as high level. Thus, promoting cultivated land circulation in grids that fall under type B presents several advantages such as high potential and low cost. The core shortcoming of cluster type A is similar to that of type B, where the cultivated land density and AWMSI of type A are in moderate levels. Grids that fall under types A and B account for 34.95% and 22.97% of the total quantity of cultivated land, respectively. Cultivated land in type B is mainly distributed in the Huang–Huai–Hai Plain, Middle-Lower Yangtze Plain, Northeast China Plain, and the Northern arid and semiarid region accounts for 27.09%, 23.04%, 20.80%, and 13.06%, respectively, of the total quantity in type B. Cultivated land in type A is mainly distributed in the Middle-Lower Yangtze Plain (23.31%) and the Yunnan–Guizhou Plateau (20.72%). For clustering type G, cultivated land density is in moderate level, and high levels of AWMSI and low levels of MPS are the main shortcomings of CLF. To improve the degree of grids in type G in relation to CLF, land consolidation and circulation should be implemented simultaneously. Grids under type G account for 8.03% of the total quantity of cultivated land. Cultivated land in type G is mainly distributed in the Sichuan Basin and surrounding regions and the Yunnan–Guizhou Plateau. Cluster type D indicates a status with severe CLF, that is low cultivated land density, high AWMSI, and low MPS. The cost of controlling CLF in grids under type D is high with a low potential for development. The core shortcoming of cluster type E is the low density of cultivated land and low MPS. Grids under types D and E account for 5.93% and 9.02% of the total quantity of cultivated land, respectively, and are mainly distributed in the Middle-Lower Yangtze Plain and the Yunnan–Guizhou Plateau. For clustering type C, the degree of CLF is mainly restricted by low levels of natural resource endowment of cultivated land. Grids under type C are mainly distributed in the Northern arid and semiarid region. Thus, expanding cultivated land in grids under type C is more appropriate.

3.4. Regional main driving factors of CLF

For each agricultural climatic zone, the driving force of six factors (i.e., elevation; slope; distance to urban; farming distance; GDP; population) on CLF has been quantitatively estimated using the random forest model. Sample data of the six factors are organized based on 1-km resolution grids that are consistent with the CLF maps. The Cross-verified R^2 values of the model fit of cultivated land density, MPS, and AWMSI are [0.87, 0.95], [0.66, 0.79], and [0.69, 0.89], respectively (see *SI Appendix A.2* for details on R^2 and RMSE). This result indicates that the established random forest regression model is applicable to explain the driving force of each factor on CLF. The driving force of the six factors on CLF is indicated by the increase in mean squared error (Inc. MSE; Fig. 4). Partial correlation coefficients have been generated to demonstrate the positive or negative relationship of the six factors to CLF (see *SI Appendix A.3* for details on Inc. MSE and partial correlation coefficients).

According to Fig. 4, the driving force of multiple factors indicates regional differentiation. Slope is the dominant factor that limits cultivated land density in the Northern plains (i.e., the Northeast China Plain and the Huang–Huai–Hai Plain) and Southern China with a significant

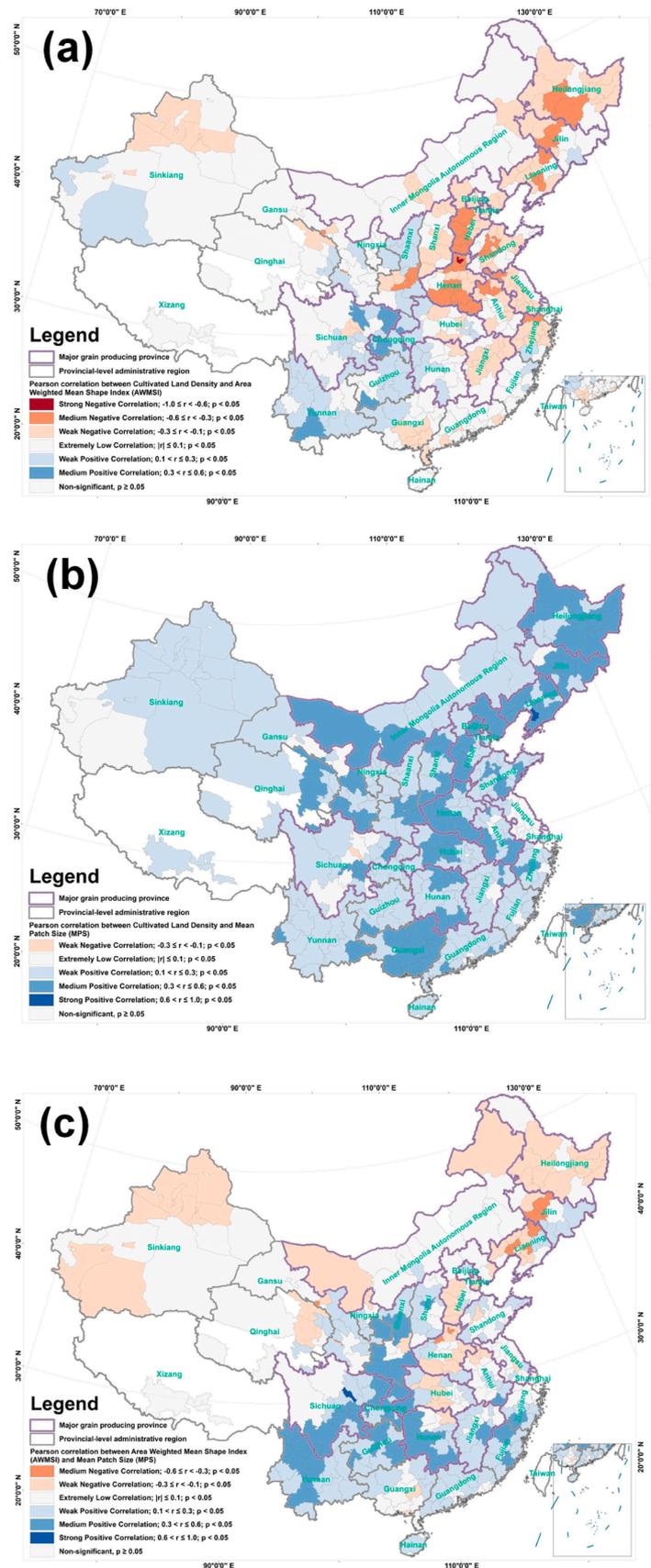


Fig. 2. (a) Pearson correlation between cultivated land density and area weighted mean shape index (AWMSI). (b) Pearson correlation between cultivated land density and mean patch size (MPS). (c) Pearson correlation between mean patch size (MPS) and area weighted mean shape index (AWMSI). Blank area expresses few cultivated land (e.g. Qinghai-Tibet Plateau, urban region) or no data (e.g. Taiwan).

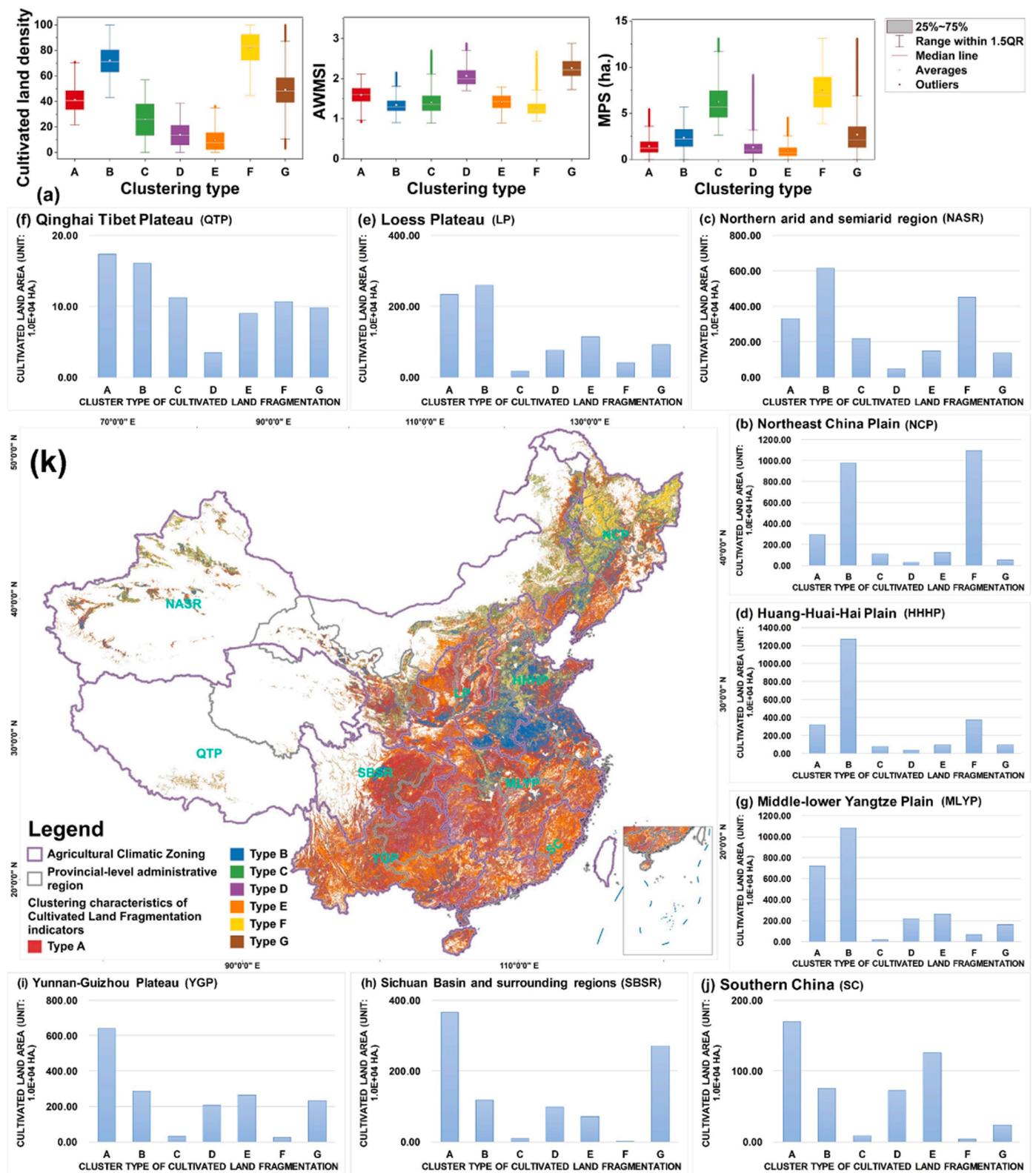


Fig. 3. (a) Box-plot of Cultivated Land Fragmentation (CLF) indicators (i.e., cultivated land density; area weighted mean shape index (AWMSI); mean patch size (MPS)) for multiple clustering types. (b)–(j) Cultivated land area (unit: 1.0E+04 ha.) of multiple clustering types in: (b) Northeast China Plain (abbr. NCP); (c) Northern arid and semiarid region (abbr. NASR); (d) Huang-Huai-Hai Plain (abbr. HHHP); (e) Loess Plateau (abbr. LP); (f) Qinghai Tibet Plateau (abbr. QTP); (g) Middle-lower Yangtze Plain (abbr. MLYP); (h) Sichuan Basin and surrounding regions (abbr. SBSR); (i) Yunnan-Guizhou Plateau (abbr. YGP); (j) Southern China (abbr. SC). (k) Spatial pattern of clustering characteristics of CLF indicators. Blank area expresses few cultivated land (e.g. Qinghai-Tibet Plateau, urban region) or no data (e.g. Taiwan).

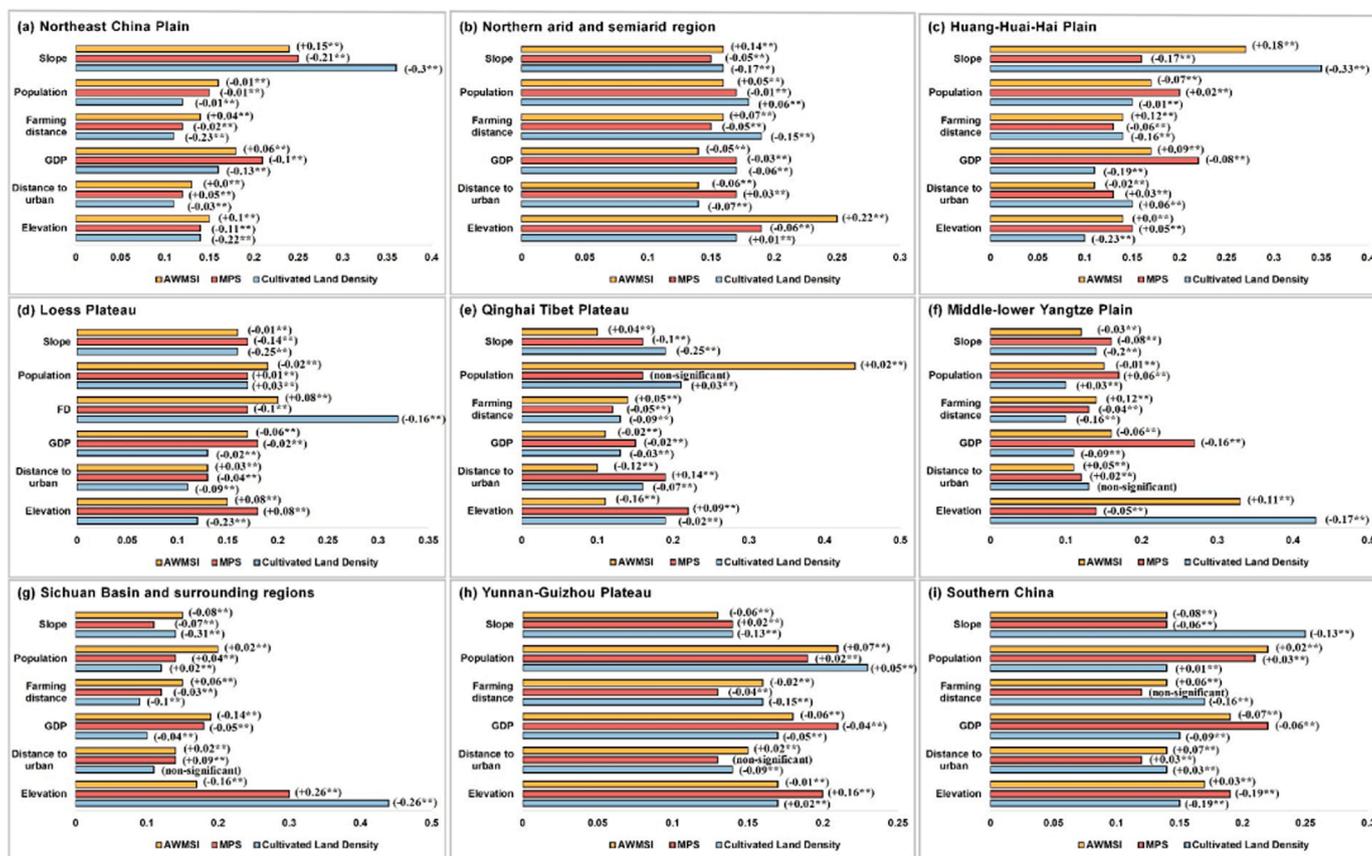


Fig. 4. The driving force of the six factors (i.e. elevation; slope; distance to urban; farming distance; GDP; population) to three cultivated land fragmentation indicators (i.e., cultivated land density, mean patch size (abbr. MPS) and area weighted mean shape index (abbr. AWMSI)), based on increased in mean squared error (Inc. MSE), for multiple regions: (a) Northeast China Plain; (b) Northern arid and semiarid region; (c) Huang-Huai-Hai Plain; (d) Loess Plateau; (e) Qinghai Tibet Plateau; (f) Middle-lower Yangtze Plain; (g) Sichuan Basin and surrounding regions; (h) Yunnan-Guizhou Plateau; (i) Southern China. Partial correlation coefficients between each factor and AWMSI, MPS, cultivated land density is listed. ** indicates $p < 0.01$.

negative impact ($p < 0.01$). Cultivated land is more concentrated on flat lands. In the southern plains (i.e., Middle-Lower Yangtze Plain and Sichuan Basin and surrounding regions), elevation plays the most important role in cultivated land density with a significant negative impact ($p < 0.01$). In other words, the distribution of cultivated land is more concentrated in regions with low elevation and within limits. In the Loess Plateau, the driving force of farming distance is two to three times as much as other factors with a significant negative correlation ($p < 0.01$). In the Yunnan-Guizhou Plateau and the Qinghai-Tibet Plateau, population exerts the highest positive influence on cultivated land density. The driving force of multiple factors is nearly equal in the Northern arid and semiarid region.

In terms of MPS, first, the difference in the quantitative driving force of multiple factors on MPS is less than that on cultivated land density. The reason underlying this notion is that the influencing process of these factors is more complicated under human-environment interactions. Second, GDP is the dominant factor of MPS for nearly all agricultural climatic zones, except for the Qinghai-Tibet Plateau. In all regions, the impact of GDP on MPS is significantly negative ($p < 0.01$). As such, cultivated land in grids with high levels of GDP have suffered from severe segmentation. Third, population is another important factor. In the Northeast China Plain and Northern arid and semiarid region, which are reliant on high levels of the application of agricultural mechanization, cultivated land in grids with low numbers of population exhibit high levels of MPS (i.e., a significant negative correlation). Alternatively, in other regions, especially in hilly and mountainous areas with high levels of labor dependence, the impact of population on MPS is positive. Fourth, in the Qinghai-Tibet Plateau, elevation and distance to urban

are the most important driving factors with a significant positive correlation ($p < 0.01$). In the Loess Plateau, the driving force of farming distance on MPS has decreased compared with that on cultivated land density.

Slope exhibits the largest impact on AWMSI in the Northeast China Plain and Huang-Huai-Hai Plain. In the Northern arid and semiarid region and Middle-Lower Yangtze Plain, the dominant factor of AWMSI is another terrain indicator, that is, elevation. The impacts of these terrain indicators are significantly positive ($p < 0.01$). In the south of China (i.e., the Qinghai-Tibet Plateau; Sichuan Basin and surrounding regions; the Yunnan-Guizhou Plateau; Southern China), the dominant factors of AWMSI are population and GDP. The impact of population is significantly positive ($p < 0.01$), whereas that of GDP is significantly negative ($p < 0.01$). In the Loess Plateau, farming distance, population, and GDP display an important impact on AWMSI. The impact of farming distance is significantly positive ($p < 0.01$), and those of population and GDP are significantly negative ($p < 0.01$).

4. Discussion

4.1. Comparative analysis with other studies on recognizing spatial pattern of Cultivated Land Fragmentation (CLF)

Many studies provide important inspiration for this research. Against this background, this study forms its contribution to the understanding of the spatial pattern of CLF in mainland China. [Chen et al. \(2018\)](#) estimate the CLF of China using the GlobeLand 30 raster dataset for 2010 and the patch density index. The result indicates evident differences

between north and south and between mountains and plains in terms of CLF. It also presents changes in CLF at the county level. Compared with Chen et al., the current study obtains a comprehensive cognition of CLF. The reason is that the result of Chen et al. (2018) cannot present CLF that due to the division of cultivated land use right. Therefore, it underestimates CLF in the Middle-Lower Yangtze Plain and Sichuan Basin. The result of the estimation of cultivated land density in the current study is consistent with that of Chen et al. Conversely, Liu et al. (2019a) estimate the CLF of Jiangsu province (China) at the township level from three perspectives, namely, resource scale, spatial agglomeration, and convenience of utilization. The current study refers to this theory. Although the selected indicators of CLF in this study are much less to avoid collinearity between indicators as much as possible, the results of the CLF estimation of Jiangsu province in the current study are consistent with those of Liu et al. (2019a) This finding demonstrates that

cultivated land density, MPS, and AWMSI are indicators representative of natural resource endowment, spatial partition, and convenience of utilization, respectively. The advantage of this study lies in its broader scope, encompassing a wider geographical range (i.e., mainland China), and its finer scale analysis using 1-km grids. By employing the 1-km grid as the unit of estimation, the present study unveils the intricate spatial heterogeneity of CLF. For instance, regions located near urban areas exhibit lower natural resource endowment, while experiencing higher convenience of utilization. This detailed information enhances the understanding of the driving factors of CLF. Moreover, the estimation result of CLF is comparable using 1-km grids. Another difference between the current study and Liu et al. (2019a) is that the current authors use the clustering method to explore regional weaknesses characteristic of CLF, which are referenced from Ye et al., (2020, 2022a; b, c). Compared with the weighted average method, the clustering method

Region	Cluster characteristics of cultivated land fragmentation [*1]				dominant factors [*2]
	Main types	Fragmentized resource endowment (FRE)	Fragmentized spatial partition (FSP)	Fragmentized utilization convenience (FUC)	
Northeast China Plain	B (36.1%)	Low	High	Low	FRE & FSP & FUC: Slope +; GDP +;
	F (40.8%)	Low	Low	Low	
Northern arid and semiarid region	B (31.5%)	Low	High	Low	FRE: Farming distance +; Population -; FSP: Elevation +; FUC: Elevation +;
	F (23.3%)	Low	Low	Low	
	A (16.9%)	Moderate	High	Moderate	
Huang-Huai-Hai Plain	B (56.2%)	Low	High	Low	FRE: Slope +; FSP: GDP +; Population -; FUC: Slope +;
	F (16.3%)	Low	Low	Low	
	A (14.1%)	Moderate	High	Moderate	
Loess Plateau	B (31.1%)	Low	High	Low	FRE & FUC: Farming distance +; Population -; FSP: Elevation -; GDP +;
	A (28.1%)	Moderate	High	Moderate	
	E (13.8%)	High	High	Low	
Qinghai Tibet Plateau	A (22.3%)	Moderate	High	Moderate	FRE: Population -; Slope +; FSP: Elevation -; Distance to urban -; FUC: Population -;
	B (20.7%)	Low	High	Low	
	C (14.5%)	High	Low	Low	
	F (13.7%)	Low	Low	Low	
Middle-lower Yangtze Plain	B (42.8%)	Low	High	Low	FRE: Elevation +; FSP: GDP +; FUC: Elevation +;
	A (28.5%)	Moderate	High	Moderate	
Sichuan Basin and surrounding regions	A (39.2%)	Moderate	High	Moderate	FRE: Elevation +; FSP: Elevation -; GDP +; FUC: GDP -;
	G (29.1%)	Moderate	High	High	
Yunnan-Guizhou Plateau	A (38.1%)	Moderate	High	Moderate	FRE: Population -; Elevation -; FSP: GDP +; Elevation -; Population -; FUC: Population +; GDP -;
	B (17.9%)	Low	High	Low	
	E (15.6%)	High	High	Low	
	G (13.9%)	Moderate	High	High	
Southern China	A (35.3%)	Moderate	High	Moderate	FRE: Slope +; FSP: GDP +; Population -; FUC: Population +; GDP -;
	E (26.2%)	High	High	Low	
	B (15.7%)	Low	High	Low	
	D (15.1%)	High	High	High	

Fig. 5. Comprehensive analysis of spatial pattern of CLF and its regional driving factors. [* 1] Cluster characteristics of cultivated land fragmentation is from Fig. 3. Fragmentized resource endowment (FRE) presents the degree of CLF that indicated by cultivated land density. Lower cultivated land density indicates higher FRE. Fragmentized spatial partition (FSP) presents the degree of CLF that indicated by mean patch size (abbr. MPS). Lower MPS indicates higher FSP. Fragmentized utilization convenience (FUC) presents the degree of CLF that indicated by area weighted mean shape index (abbr. AWMSI). Higher AWMSI indicates higher FSP. [* 2] Regional dominant factors to CLF is from Fig. 4. The arrangement of factors is in descending order of importance. “+” means significant ($p < 0.01$) positive correlation. “-” means significant ($p < 0.01$) negative correlation.

can better preserve the extreme characteristics of CLF. Moreover, Liu et al. (2019a) utilized the multiple linear regression model to detect the primary factor influencing CLF. The simulated R^2 (0.622) of Liu et al. (2019a) study is noticeably lower than the Cross-verified R (0.66–0.95) obtained in this paper, which is based on the random forest model. The utilization of the random forest model allows for a more accurate and robust analysis, resulting in a wider range of R values, indicating a stronger predictive performance. Qian et al. (2020) analyze the spatial patterns of CLF in Liaoning Province using 10-km grids. The result demonstrates that CLF in the central plain area of Liaoning province is “generally external low, internal high because of the number and division of cultivated land patches.” The present study verifies this explanation, because the spatial pattern of MPS exhibits similar characteristics.

4.2. Dominant factors of CLF and appropriate strategies for controlling CLF

To explore strategies for controlling CLF that are locally appropriate, the study comprehensively analyzed the spatial patterns of CLF (Figs. 2 and 3) and its driving factors (Fig. 4), as shown in Fig. 5. The study identified and summarized two characteristics of CLF. First, the highly fragmented spatial partition of cultivated land, which is mainly due to the division of land use rights and diverse crop planting, is the most critical problem related to CLF in mainland China (fragmentized spatial partition presents the degree of CLF that indicated by mean patch size. Lower MPS indicates higher FSP, in Fig. 5). Liu et al. (2022) conducted an analysis of influence of rural development on the dominant factors affecting CLF in Jiangsu province. Consistent with Liu et al. (2022), this study reveals that the regional dominant factors and their impact characteristics on the fragmented spatial partition of cultivated land differ or even reverse in better-developed rural areas (mainly in the economically developed eastern plain areas) and backward rural areas (mainly in the economically backward mountainous or plateau areas).

In better-developed rural areas, contiguous plots of cultivated land tend to be equally allocated to many families given the soil properties (Tan et al., 2006; Chen et al., 2010). Economic development exacerbates the construction of residential land, roads and agricultural facilities, leading to an increase in the fragmented spatial partition of cultivated land. Studies by Liu et al. (2019a) and Xu et al. (2021) indicates that GDP per capita, population per unit area, the proportion of industry and services, and slope are the dominant factors influencing CLF in Jiangsu province. Wang et al. (2020) reports that population density, GDP per capita, forest area ratio, slope and elevation are the dominant factors affecting CLF in Guangdong province. In the current study, the authors propose that GDP and population are the most important factors of MPS, where terrain factors have the largest impacts on cultivated land density and AWMSI, in Middle-lower Yangtze Plain. In southern China, population, GDP and elevation show higher influence to MPS and AWMSI, and slope is the most important factor to cultivated land density. Thereinto, GDP is the dominant factor positively exacerbating the degree of fragmented spatial partition. Therefore, promoting cultivated land circulation and the consolidation of hollow villages are key strategies for controlling fragmented spatial partition in better-developed rural areas. This is because these areas generally have high potential for cultivated land circulation as their high cultivated land density, flat terrain, and dense populations. Moreover, high levels of GDP in these areas can provide better conditions for the circulation of cultivated land (Chen et al., 2010). In the implementation process, the erosion of cultivated land by urbanization requires strict management, and policies toward the circulation of cultivated land, and the consolidation of hollow villages should be improved by setting clear and sufficient responsibilities for both parties to protect the benefits of renters and to respect the wishes of lessors (Cao et al., 2020). Especially in the context of more developed rural areas in southern China, reinforcing cultivated land circulation and land consolidation becomes imperative. These

regions possess abundant agro-climatic resources, which present an opportunity to enhance the efficiency of large-scale agricultural management. The use of transferred cultivated land should be regulated to avoid the *non-food* phenomenon and overexploitation, and the livelihood of migrant farmers should be considered (Liu et al., 2014a,b,c, 2018b).

Several studies on mountainous counties have generally identified farming distance, slope, and elevation as significant factors influencing CLF, which aligns with findings in this paper (Zhang et al., 2016; Guo et al., 2017; Chang et al., 2021). This study provides a more detailed analysis of the regional variations in the impact of these dominant factors on CLF in mountainous areas. In backward rural areas characterized by undulating terrain, high farming and transportation costs, and low labor efficiency, the size of cultivated land plots tends to be small. In order to meet basic livelihood needs or enhance resilience against disaster risks, these small cultivated land plots are often further subdivided to grow various crops. Terrain factors (i.e., elevation and slope) are the positively dominant factors of the degree of fragmented spatial partition in backward rural areas. The implementation of land consolidation may not be appropriate due to its high cost and low efficiency (Wen et al., 2016; Ntuhinyurwa et al., 2019). The third national land survey's major data bulletin reveals that nearly 30.0 million hectares of cultivated land in mainland China have slopes exceeding 6° , accounting for 22.75% of the total cultivated land area. In regions with challenging natural conditions, it is essential to recognize the positive impact of CLF, such as increased economic income and reduced farming risks. The planting of cash crops should be properly allowed to ensure the livelihood of rural households or attract large-scale management subjects to participate in the cultivated land circulation, thus mitigating CLF issues. Additionally, promoting the "Grain for Green" program in cultivated land plots with high elevation or slope can be beneficial. The other important strategy is developing agricultural machinery suitable for mountainous regions are key strategies for mitigating the impact of wavy terrain and scarce labor on fragmented spatial partition in backward rural areas. Population density represents another crucial factor, exhibiting a significant negative correlation ($p < 0.01$) in these regions. On the one hand, as population density increases, so does the demand for the division and allocation of cultivated land, exacerbating the problem of fragmented spatial partitioning. On the other hand, sparsely populated areas often face the challenges of complex terrain and diverse cultivation practices. By contrast, negative impact of population density outweighs its positive impact to fragmented spatial partition of cultivated land. Therefore, when conducting research on CLF based on different spatial zoning modes (e.g., administrative districts, watershed divisions, developing functional zones), it is crucial to consider the variations in dominant factors between better-developed rural areas and backward rural areas.

Second, fragmented resource endowment and fragmented utilization convenience are mainly exacerbated in backward rural areas, which are positively driven by terrain factors and farming distance and negatively driven by population (Fragmented resource endowment presents the degree of CLF that indicated by cultivated land density. Lower cultivated land density indicates higher FRE. Fragmented utilization convenience presents the degree of CLF that indicated by AWMSI. Higher AWMSI indicates higher FSP, in Fig. 5). The negative correlation between population to fragmented resource endowment can be attributed to the tendency of village settlements to cluster in areas suitable for farming, resulting in higher cultivated land density. In the Yunnan-Guizhou Plateau and Southern China, population and GDP exhibit positive and negative correlations, respectively, with the fragmented utilization convenience of cultivated land. This suggests that increasing agricultural opportunity cost and promoting rural development can significantly reduce AWMSI in these regions. Conversely, in Northern arid and semiarid region and the Loess Plateau, farming distance shows a significant positive correlation with fragmented resource endowment. This finding underscores the importance of

concentrating cultivated land expansion and urgently addressing the increase in farming distance caused by "occupying near plots and compensating farther ones" in these regions.

4.3. Challenges in cognizing CLF and future work

To improve regional CLF, effective policies should be promulgated and implemented on the basis of cognizing the multi-dimensional characteristics and dominant factors of CLF. The cognition and supervision of the spatial pattern of CLF put forward high requirements for data, theory, and methodology in relation to land computing (Gong et al., 2023; Ye et al., 2018, 2020). First, the availability of high-resolution remote sensing data enhanced the research on changes in land cover and gradually lead to a clear understanding of spatial and temporal changes in cultivated land distribution worldwide (Liu et al., 2014a, 2018b). Despite this achievement, a gap remains in meeting research requirements of cognizing spatial pattern of CLF. One such challenge arises when using remote sensing images with ten meter-level spatial resolution, which may lead to the underestimation of the degree of CLF. This is mainly due to the difficulty in identifying the edges of cultivated land plots, such as ridges of fields, and the transformations occurring in extremely small plots. Additionally, the lack of spatial data on the types of crops hinders accurate estimation of CLF driven by diversified crop planting. Therefore, future efforts should focus on obtaining higher resolution spatial data, especially indicators of crop types, to enhance the identification of spatial partitions and spatio-temporal changes in cultivated land plots (Fang et al., 2022; Wan et al., 2021). Second, existing indicators of landscape fragmentation may have limitations when estimating CLF at the regional scale. For example, terrain factors, like terraced fields, are not adequately considered, and common methods used to assess landscape clustering may underestimate the degree of CLF due to the coexistence of discrete and aggregate distributions of cultivated land. To address these challenges, it is essential to promote the development of theory and innovative methods for determining the CLF estimation unit and estimating the degree of CLF while considering differences in terrain and distribution characteristics. Future research should explore the use of additional indicators to characterize CLF and investigate their interactions in different regions and scales.

The other major challenge lies in understanding complex driving mechanism of CLF. This study reveals that the dominant factors influencing CLF and their effects vary with the study area and the characteristic dimension of CLF. The correlation between dominant factors and CLF is affected by the degree of rural development. Whereas, in this study, agroclimatic zoning is used as a spatial unit to study the driving factors of CLF, without fully considering the variations in rural development degree within agricultural zoning. In future work, it is essential to explore the spatial heterogeneity of dominant factors' influence on CLF by employing different spatial partitioning schemes, such as administrative unit-based partitioning or sliding window-based partitioning. Additionally, future research should examine how changes in the numerical interval of dominant factors affect their correlation with CLF. It is also necessary to study the dominant factors of CLF at different grid scales, requiring the collection or production of higher spatial resolution driver datasets. Moreover, the impact of driving factors on CLF was explained by the random forest model with an accuracy range of 66–95% in this study. Hence in the future of work, more driving factors (e.g., urbanization rate, industrial structure, irrigation conditions, road density, etc.) should be included.

5. Conclusion

This study used cultivated land density, mean patch size (MPS) and area weighted mean shape index (AWMSI) to indicate the characteristics of Cultivated Land Fragmentation (CLF) in mainland China from three perspectives, namely, natural resource endowment, spatial partition,

and convenience of utilization, based on 1-km² grids. Pearson's correlation coefficients were calculated to assess the correlation among multi-dimensional CLF indicators from a holistic perspective at the prefecture-level city scale. Additionally, the study examined the clustering characteristics and regional driving factors of the three indicators of CLF. Regarding the spatial pattern of CLF, the results demonstrated that cultivated land density is higher in plain regions compared to mountainous regions, and it is also higher in Northern regions compared to Southern regions. The level of MPS tended to be consistent with that of cultivated land density for nearly all cities in mainland China, indicating that suitable farming scales have been implemented in these regions with support from high levels of natural resource endowment. Notably, the Northeast China Plain and Huang-Huai-Hai Plain exhibited the best coordination between CLF indicators, suggesting more favorable conditions for the use of agricultural machinery. Pearson's correlation between cultivated land density and AWMSI pointed to evident differences between the east and the west, whereas Pearson's correlation between MPS and AWMSI illustrated clear differences between the north and the south. High levels of fragmented spatial partition in cultivated land was considered the most critical problem related to CLF in mainland China. Regarding the driving factors of CLF, the random forest model explained 66–95% of the impact of these factors. The driving force of multiple factors indicates regional differentiation. The terrain factor emerged as the main driver negatively affecting cultivated land density, except for the Loess Plateau and the Yunnan–Guizhou Plateau. In the Loess Plateau, the driving force of farming distance is two to three times as much as other factors with a significant negative correlation to cultivated land density ($p < 0.01$). In the Yunnan–Guizhou Plateau and the Qinghai–Tibet Plateau, population has the most significant positive influence on cultivated land density. GDP emerges as the dominant factor, displaying a significant ($p < 0.01$) negative correlation to MPS across nearly all agricultural climatic zones. Population also remains a crucial factor influencing MPS. In the Northeast China Plain and the Northern arid and semiarid region, grids with low population density exhibit high levels of MPS. Conversely, in other regions, particularly in hilly and mountainous areas with high levels of labor dependence, the impact of population on MPS shows a positive association. Terrain, GDP and population are the most influential factors affecting AWMSI, demonstrating regional differences. The degree of rural development influences the correlation between dominant factors and CLF. Further exploration of spatial heterogeneity in the influence of dominant factors on CLF requires additional spatial partitioning schemes (e.g., administrative unit-based partitioning; sliding window-based partitioning). This study is greatly instructive for recognizing the spatial patterns of CLF at the national scale and for exploring the barriers that impede regionally scaled cultivated land use. Lastly, it can serve as a reference for other countries by reflecting on China's experiences in CLF management. The research method in this paper can provide technical reference for other studies related to CLF.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was funded by the second Tibetan Plateau Scientific Expedition and Research Program (STEP) [Grant No. 2019QZKK0608]; National Natural Science Foundation of China [Grant No. 42171250]; and the Strategic Priority Research Program of the Chinese Academy of

Sciences [Grant No. XDA23100303]; and Project Supported by State Key Laboratory of Earth Surface Processes and Resource Ecology [2022-ZD-04]. We would like to thank the high-performance computing support from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (<https://gda.bnu.edu.cn/>).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.landusepol.2024.107070](https://doi.org/10.1016/j.landusepol.2024.107070).

References

- Abubakari, Z., van der Molen, P., Bennett, R.M., et al., 2016. Land consolidation, customary lands, and Ghana's Northern Savannah Ecological Zone: An evaluation of the possibilities and pitfalls. *Land Use Policy* 54, 386–398.
- Adnan, M.S.G., Abdullah, A.M., Dewan, A., et al., 2020. The effects of changing land use and flood hazard on poverty in coastal Bangladesh. *Land Use Policy* 99, 104868.
- Arsanjani, J.J., Tayyebi, A., Vaz, E., 2016. *GlobeLand30* as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat Int* 55, 25–31.
- Breiman, L., 2001. Random forests. *Mach. Learn* 45 (1), 5–32.
- Brus, D.J., De Groot, J.J., Van Groenigen, J.W., 2006. Designing spatial coverage samples using the k-means clustering algorithm. *Dev. Soil Sci.* 31 183–192.
- Cao, Y., Zou, J., Fang, X.Q., et al., 2020. Effect of land tenure fragmentation on the decision-making and scale of agricultural land transfer in China. *Land Use Policy* 99, 104996.
- Carrao, H., Naumann, G., Barbosa, P., 2016. Mapping global patterns of drought risk: An empirical framework based on sub-national estimates of hazard, exposure and vulnerability. *Glob. Environ. Chang* 39, 108–124.
- Chang, Y., Zhang, T., Zhang, F., et al., 2021. Analysis on the Spatiotemporal Variability and Driving Factors of Cultivated Land Fragmentation in the Hilly Region of Eastern Gansu - A Case of Wannian County of Jiangxi Province. *Res. Soil Water Conserv.* 28 (3), 264–271.
- Chaudhary, S., Wang, Y.K., Dixit, A.M., et al., 2020. A synopsis of farmland abandonment and its driving factors in Nepal. *Land-Base* 9 (3), 84.
- Chen, D., Wu, W., Zhou, Q., et al., 2018. Changes of cultivated land utilization Pattern in Asia from 2000 to 2010. *Sci. Agric. Sin.* 6, 1106–1120.
- Chen, J., Cao, X., Peng, S., et al., 2017. Analysis and applications of *GlobeLand30*: A review. *ISPRS Int J. Geo-Inf.* 6 (8), 230.
- Chen, M.Q., Zhong, T.Y., Zhou, B.J., et al., 2010. Empirical research on farm households' attitude and behaviour for cultivated land transferring and its influencing factors in China. *Apr. Environ.* -Czech 56 (9), 409–420.
- Ciaian, P., Guri, F., Rajcaniova, M., et al., 2018. Land fragmentation and production diversification: A case study from rural Albania. *Land Use Policy* 76, 589–599.
- Deininger, K., Monchuk, D., Nagarajan, H.K., et al., 2017. Does land fragmentation increase the cost of cultivation? Evidence from India. *J. Dev. Stud.* 53, 82–98.
- Demetris, D., Stillwell, J., See, L., 2013. A new methodology for measuring land fragmentation. *Comput. Environ. Urban* 39 (1), 71–80.
- Falco, D., Penov, I., Aleksiev, A., et al., 2010. Agrobiodiversity, farm profits and land fragmentation: evidence from Bulgaria. *Land Use Policy* 27 (3), 763–771.
- Fang, D.L., Cai, Q.N., Wu, F., et al., 2022. Modified linkage analysis for water-land nexus driven by interregional trade. *J. Clean. Prod.* 353, 131547.
- Gong, B., Liu, Z., Liu, Y., et al., 2023. Understanding advances and challenges of urban water security and sustainability in China based on water footprint dynamics. *Ecol. Indic.* 150, 110233.
- Gonzalez, X.P., Marey, M.F., Alvarez, C.J., 2007. Evaluation of productive rural land patterns with joint regard to the size, shape and dispersion of plots. *Agr. Syst.* 92, 52–62.
- Guo, S., Yang, W., Wei, M., et al., 2017. Cultivated Land Fragmentation and Impact Factors of Qinglong Manchu Autonomous County Based on Geographically Weighted Regression. *Res. Soil Water Conserv.* 24 (03), 264–269.
- Hartvigsen, M., 2014. Land reform and land fragmentation in Central and Eastern Europe. *Land Use Policy* 36, 330–341.
- Huang, Q.X., Zhang, H., van Vliet, J., et al., 2021. Patterns and distributions of urban expansion in global watersheds. *Earth's Future* 9 (8), e2021EF002062.
- Jin, X., Jiang, W., Fang, D., et al., 2024. Evaluation and driving force analysis of the water-energy-carbon nexus in agricultural trade for RCEP countries. *Appl. Energy* 353, 122143.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., et al., 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE T Pattern Anal.* 24, 881–892.
- Kopeva, D., Mishev, P., Howe, K., 1994. Land reform and liquidation of collective farm assets in Bulgarian agriculture: progress and prospects. *Communist Econ. Econ. Transform.* 6 (2), 203–217.
- Kuang, W.H., Liu, J.Y., Tian, H.Q., et al., 2022. Cropland redistribution to marginal lands undermines environmental sustainability. *Natl. Sci. Rev.* 9 (1), nwab091.
- Liu, C., Song, C., Ye, S., et al., 2023. Estimate provincial-level effectiveness of the arable land requisition-compensation balance policy in mainland China in the last 20 years. *Land Use Policy* 131, 106733.
- Liu, J., Jin, X.B., Xu, W.Y., et al., 2019a. Influential factors and classification of cultivated land fragmentation, and implications for future land consolidation: A case study of Jiangsu Province in Eastern China. *Land Use Policy* 88, 104185.
- Liu, J., Jin, X., Xu, W., et al., 2022. Evolution of cultivated land fragmentation and its driving mechanism in rural development: A case study of Jiangsu Province. *J. Rural Stud.* 91, 58–72.
- Liu, J.Y., Kuang, W.H., Zhang, Z.X., et al., 2014a. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *J. Geogr. Sci.* 24, 195–210.
- Liu, L., Liu, Z.J., Gong, J.Z., et al., 2019b. Quantifying the amount, heterogeneity, and pattern of farmland: Implications for China's requisition-compensation balance of farmland policy. *Land Use Policy* 81, 256–266.
- Liu, X.J., Wang, M.S., 2016. How polycentric is urban China and why? A case study of 318 cities. *Landsc. Urban Plan* 151, 10–20.
- Liu, Y., 2018a. Introduction to land use and rural sustainability in China. *Land Use Policy* 74, 1–4.
- Liu, Y., Li, Y., 2017b. Revitalize the world's countryside. *Nature* 548 (7667), 275–277.
- Liu, Y., Zhou, Y., 2021. Reflections on China's food security and land use policy under rapid urbanization. *Land Use Policy* 109, 109.
- Liu, Y., Fang, F., Li, Y., 2014c. Key issues of land use in China and implications for policy making. *Land Use Policy* 40, 6–12.
- Liu, Y., Yang, Ren, Long, H., et al., 2014b. Implications of land-use change in rural China: a case study of Yucheng, Shandong province. *Land Use Policy* 40, 111–118.
- Liu, Y., Li, J.T., Yang, Y.Y., 2018b. Strategic adjustment of land use policy under the economic transformation. *Land Use Policy* 74, 5–14.
- Long, H.L., 2014. Land consolidation: An indispensable way of spatial restructuring in rural China. *J. Geogr. Sci.* 24 (02), 211–225.
- Long, H.L., Zhang, Y.N., Tu, S.S., 2019. Rural vitalization in China: A perspective of land consolidation. *J. Geogr. Sci.* 29 (04), 517–530.
- Looga, J., Jürgenson, E., Sikk, K., et al., 2018. Land fragmentation and other determinants of agricultural farm productivity: The case of Estonia. *Land Use Policy* 79, 285–292.
- Lu, H., Xie, H.L., He, Y.F., et al., 2018. Assessing the impacts of land fragmentation and plot size on yields and costs: A translog production model and cost function approach. *Agr. Syst.* 161 (1), 81–88.
- Lu, M., Wu, W.B., Zhang, L., et al., 2016. A comparative analysis of five global cropland datasets in China. *Sci. China Earth Sci.* 59 (12), 2307–2317.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press., Berkeley, pp. 281–297.
- Nthinyurwa, P.D., de Vries, W.T., 2020. Farmland fragmentation and defragmentation nexus: Scoping the causes, impacts, and the conditions determining its management decisions. *Ecol. Indic.* 119, 1–16.
- Nthinyurwa, P.D., de Vries, W.T., Chigbu, U.E., et al., 2019. The positive impacts of farm land fragmentation in Rwanda. *Land Use Policy* 81, 565–581.
- Qian, F.K., Chi, Y.R., Lal, R., et al., 2020. Spatio-temporal characteristics of cultivated land fragmentation in different landform areas with a case study in Northeast China. *Ecosyst. Health Sust.* 6 (1), 1800415.
- Qiu, F., Laliberté, L., Swallow, B., et al., 2015. Impacts of fragmentation and neighbor influences on farmland conversion: A case study of the Edmonton-Calgary Corridor, Canada. *Land Use Policy* 48, 482–494.
- Ren, S., Song, C., Ye, S., Cheng, C., Gao, P., 2022. The spatiotemporal variation in heavy metals in china's farmland soil over the past 20years: a meta-analysis. *Sci. Total Environ.* 806, 150322.
- Ren, S., Song, C., Ye, S., et al., 2023. Land use evaluation considering soil properties and agricultural infrastructure in black soil region. *Land Degrad. Dev.* 34 (17), 5373–5388.
- Rodgers, J.L., Nicewander, W.A., 1988. 13 ways to look at the correlation-coefficient. *Am. Stat.* 42 (1), 59–66.
- Sklenicka, P., 2016. Classification of farmland ownership fragmentation as a cause of land degradation: A review on typology, consequences, and remedies. *Land Use Policy* 57 (7), 694–701.
- Tan, S., Heerink, N., Qu, F., 2006. Land fragmentation and its driving forces in China. *Land Use Policy* 23 (3), 272–285.
- Teillard, F., Antoniucci, D., Jiguet, F., et al., 2014. Contrasting distributions of grassland and arable birds in heterogeneous farmlands: Implications for conservation. *Biol. Conserv.* 176, 243–251.
- Tran, T.Q., Vu, H.V., 2019. Land fragmentation and household income: First evidence from rural Vietnam. *Land Use Policy* 89, 104247.
- Viet Nguyen, C., Ngoc Tran, A., 2014. The role of crop land during economic development: evidence from rural Vietnam. *Eur. Rev. Agric. Econ.* 41, 561–582.
- Wan, C.J., Kuzyakov, Y., Cheng, C.X., et al., 2021. A soil sampling design for arable land quality observation by using spcosa-clhs hybrid approach. *Land Degrad. Dev.* 32 (17), 4889–4906.
- Wan, G.H., Cheng, E.J., 2001. Effects of land fragmentation and returns to scale in the Chinese farming sector. *Appl. Econ.* 33 (2), 183–194.
- Wang, K.X., Ye, S.J., Gao, P.C., et al., 2022b. Optimization of numerical methods for transforming utm plane coordinates to lambert plane coordinates. *Remote Sens. Basel* 14 (9), 2056.
- Wang, X., Xu, X., 2022a. Spatiotemporal characteristics and influencing factors of landscape fragmentation of cultivated land in China. *Trans. CSAE* 38 (16), 11–20.
- Wen, G., Li, Y., Zhao, W., et al., 2016. The Governance Effect of Rural Land Consolidation on Cultivated Land Fragmentation and Its Causes: A Case Study on Jiangxia District, Xian'an District, Tongshan County in Hubei Province. *China Land. Sciences* 30 (9), 82–89.
- Wilpon, J.G., Rabiner, L.R., 1985. A modified K-means clustering algorithm for use in isolated work recognition. *IEEE T Acoust. Speech* 33 (3), 587–594.

- Xie, H.L., Lu, H., 2017. Impact of land fragmentation and non-agricultural labor supply on circulation of agricultural land management rights. *Land Use Policy* 68 (4), 355–364.
- Xu, W.Y., Jin, X.B., Liu, J., et al., 2021. Analysis of influencing factors of cultivated land fragmentation based on hierarchical linear model: A case study of Jiangsu Province, China. *Land Use Policy* 101, 105119.
- Xu, X.L., 2017b, 1 km Grid Population Dataset of China[J/DB/OL]. Resource and environment science and data center. (<http://www.resdc.cn/>, DOI:10.12078/2017121101).
- Xu, X.L., 2017a, 1 km Grid GDP Dataset of China[J/DB/OL]. Resource and environment science and data center. (<http://www.resdc.cn/>, DOI:10.12078/2017121101).
- Yan, F.Q., Zhang, S.W., Kuang, W.H., et al., 2016. Comparison of cultivated landscape changes under different management modes: A case study in Sanjiang Plain. *Sustain. -Basel* 8 (10), 1071.
- Yang, Y.K., Xiao, P.F., Feng, X.Z., et al., 2017. Accuracy assessment of seven global land cover datasets over China. *ISPRS-J. Photo Remote Sens* 125, 156–173.
- Yao, X.C., Mokbel, M.F., Alarabi, L., et al., 2017. Spatial coding-based approach for partitioning big spatial data in Hadoop. *Comput. Geosci. -UK* 106, 60–67.
- Ye, S., Liu, D., Yao, X., et al., 2018. RDRCMG: A raster dataset clean & reconstitution multi-grid architecture for remote sensing monitoring of vegetation dryness. *Remote Sens-Basel* 10 (9), 1376.
- Ye, S., Song, C., Cheng, F., et al., 2019. Cultivated land health-productivity comprehensive evaluation and its pilot evaluation in China. *Trans. Chin. Soc. Agric. Eng.* 35 (22), 66–78.
- Ye, S., Song, C., Shen, S., et al., 2020. Spatial pattern of arable land-use intensity in China. *Land Use Policy* 99, 104845.
- Ye, S., Ren, S., Song, C., et al., 2022a. Spatial patterns of county-level arable land productive-capacity and its coordination with land-use intensity in mainland China. *Agr. Ecosyst. Environ.* 326, 107757.
- Ye, S., Song, C., Gao, P., et al., 2022b. Visualizing clustering characteristics of multidimensional arable land quality indexes at the county level in mainland China. *Environ. Plan. A: Econ. Space* 54 (2), 222–225.
- Ye, S., Song, C., Kuzyakov, Y., et al., 2022c. Preface: arable land quality: observation, estimation, optimization, and application. *Land* 11, 947.
- Ye, S., Song, C., Gao, P., et al., 2023. Construction of the new cognitive system for arable land resources from geospatial perspective. *Trans. Chin. Soc. Agric. Eng.* 39 (9), 225–240.
- Ye, S.J., Zhu, D.H., Yao, X.C., et al., 2014. Development of a highly flexible mobile GIS-based system for collecting arable land quality data. *IEEE J. -STARS* 7 (11), 4432–4441.
- Ye, S.J., Yan, T.L., Yue, Y.L., et al., 2016. Developing a reversible rapid coordinate transformation model for the cylindrical projection. *Comput. Geosci. -UK* 89, 44–56.
- Yu, Q.Y., Hu, Q., van Vliet, J., et al., 2018. GlobeLand30 shows little cropland area loss but greater fragmentation in China. *Int. J. Appl. Earth Obs.* 66, 37–45.
- Yucer, A.A., Kan, M., Demirtas, M., et al., 2016. The importance of creating new inheritance policies and laws that reduce agricultural land fragmentation and its negative impacts in Turkey. *Land Use Policy* 56 (11), 1–7.
- Zhang, K.P., Li, Y.F., Wei, H.H., et al., 2022. Conservation tillage or plastic film mulching? A comprehensive global meta-analysis based on maize yield and nitrogen use efficiency. *Sci. Total Environ.* 831, 154869.