

Article

Estimating Sub-Pixel Soybean Fraction from Time-Series MODIS Data Using an Optimized Geographically Weighted Regression Model

Qiong Hu ^{1,2}, Yaxiong Ma ², Baodong Xu ^{2,3} , Qian Song ¹, Huajun Tang ¹ and Wenbin Wu ^{1,*}

¹ Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; huqiong02@caas.cn (Q.H.); songqian01@caas.cn (Q.S.); tanghuajun@caas.cn (H.T.)

² Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, Boston, MA 02215, USA; yxma@bu.edu (Y.M.); xubd@radi.ac.cn (B.X.)

³ State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China

* Correspondence: wuwenbin@caas.cn; Tel.: +86-10-82105070

Received: 26 December 2017; Accepted: 19 March 2018; Published: 21 March 2018



Abstract: Soybean cultivation in China has significantly decreased due to the rising import of genetically modified soybeans from other countries. Understanding soybean's extent and change information is of great value for national agricultural policy implications and global food security. Some previous studies have explored the quantitative relationships between crop area and spectral variables derived from remote sensing data. However, both those linear or non-linear relationships were expressed by global regression models, which ignored the spatial non-stationarity of crop spectral signature and may limit the prediction accuracy. This study presented a geographically weighted regression model (GWR) to estimate fractional soybean at 250 m spatial resolution in Heilongjiang Province, one of the most important food production regions in China, using time-series MODIS data and high-quality calibration information derived from Landsat data. A forward stepwise optimization strategy was embedded with the GWR model to select the optimal subset of independent variables for soybeans. Normalized Difference Vegetation Index (NDVI) of Julian day 233 to 257 when soybeans are filling seed was found to be the most important temporal period for sub-pixel soybean area estimation. Our MODIS-based soybean area compared well with Landsat-based results at pixel-level. Also, there was a good agreement between the MODIS-based result and census data at county level, with the coefficient of determination (R^2) of 0.80 and the root mean square error (RMSE) was 340.21 km². Additionally, F -test results showed GWR model had better model goodness-of-fit and higher prediction accuracy than the traditional ordinary least squares (OLS) model. These promising results suggest crop spectral variations both at temporal and spatial scales should be considered when exploring its relationship with pixel-level crop acreage. The optimized GWR model by combining an automated feature selection strategy has great potential for estimating sub-pixel crop area at regional scale based on remote sensing time-series data.

Keywords: geographically weighted regression; sub-pixel crop area; MODIS time series; forward stepwise selection; soybeans; Heilongjiang

1. Introduction

Soybean, one of the most important crop types and commodities for vegetable oil and protein in the global agricultural market, comprises 5% of global cropland and ranks 5th among major crop categories [1,2]. During the last decade, China imported more than 80% of its soybean consumption from

Brazil and the United States and has become the largest soybean importer [3,4]. The spatial-temporal change of soybean in China is greatly associated with the land cover and land use in those soybean-exporting countries such as Brazil. Therefore, reliable and timely information on soybean planting area is not only essential for global food security and agricultural policy implications, but also of great value in understanding global land use changes under the changing climate. In addition, crop planting area is a key input for a wide range of studies and models including natural environment, social and economic factors.

Previous studies have shown that Moderate Resolution Imaging Spectroradiometer (MODIS) is particularly suited for monitoring agricultural system [5,6]. The short temporal revisit cycle and rich spectral information of such data allows for the identification of the subtle differences in phenological characteristics between various crop types and thus improvement for their spectral separability [7,8]. Also, since these data are available on a global scale and with nearly twenty years of data, there is great potential for large-region crop mapping including changes to crop cultivation through time. However, the one substantial challenge for using MODIS time series data for these purposes is their inherently coarse spatial resolution (>230 m), which increases the occurrence of mixtures of different land cover classes within a pixel and therefore may result in significant errors in the estimated crop areas [9,10].

To address the “mixed-pixel” problem on crop classification, many previous studies were made to explore the quantitative relationships between sub-pixel crop fraction and a set of input variables involving spectral signature from different temporal periods. For example, Pan et al. [10] estimated winter wheat planting area in two regions in China from MODIS-EVI time series data by using the Crop Proportion Phenology index, which expresses the linear relationship between EVI time series and actual crop acreage. Lobel et al. [11] developed a probabilistic linear unmixing approach with MODIS that estimates sub-pixel fractions of crop area based on the temporal reflectance signatures throughout the growing season. In addition to the linear unmixing models, Atzberger et al. [12] examined the nonlinear relationships between sub-pixel crops area in Central Italy and AVHRR NDVI using a Neural Network model. Both those linear or non-linear sub-pixel unmixing methods are global regression models which assume the relationships between dependent and independent variables are homogenous across space. In other words, those traditional global models use constant regression parameters to estimate sub-pixel crop fraction throughout the whole study area without taking account of the local variation in the data. However, since the spectral reflectance signature of the crop highly depends on the agricultural intensive management, landscape fragmentation, local weather conditions and soil quality [13,14], the spectral variability of crop including intra-class and inter-class variability may vary with locations [15]. As a result, the relationships between sub-pixel crop fraction and spectral signature can vary across the geographical space. Furthermore, numerous studies have reported that ignoring the spatial non-stationarity of environmental variables would not be scientifically justifiable and may cause significant biases in the predictions [16–18]. Hence, the importance of spatial heterogeneity in crop spectral signature should not be overlooked for the large-region estimation of sub-pixel crop fractions.

One of the most widely used tools in exploring spatial heterogeneity of relationships is geographically weighted regression (GWR), which is a local spatial statistical technique to examine the spatial variation and non-stationarity of the dataset [19,20]. This technique estimates one set of coefficient values at every chosen fit point by moving a weighted window over the data [19]. The ability to not only express spatial variation but also to map the regression parameters and goodness-of-fit statistics makes GWR a very desired tool for remotely sensed variable modelling at large scale. Moreover, a few recent related studies have reported the implications of GWR to explore the spatial instability of the relationship between ground-level PM_{2.5} (Particulate Matter 2.5) concentrations and the aerosol optical depth (AOD) [21], aboveground biomass and multi-spectral vegetation indices (VI) [22], sub-pixel cropland proportion and agreements of three land cover products [23], and so on. Their results have shown considering the spatial variation in the studied relationship based on GWR model can significantly improve the accuracy of estimated dependent variable as well as the explanatory power of the model. However, despite the successful application of GWR models in remote

sensing fields, its potential in modelling the spatial distribution of crop type is unclear. In addition, how to optimize the GWR model to deal with many potential independent variables has not yet been explored.

To address the knowledge gap, the objective of this study is to use an optimized GWR model by combining a forward stepwise feature selection strategy (1) to evaluate its potential in estimating sub-pixel soybean fraction based on MODIS time-series data; (2) to characterize the non-stationary spatial relationships between sub-pixel crop fraction and MODIS-NDVI time series; and (3) to determine the optimal temporal features for soybean area estimation. The Heilongjiang province, one of the most important soybean production regions in China, was chosen as the study area. This paper is organized as follows: Section 2 introduces the basic geographical information of the study area and describes the study dataset that was used to calibrate and evaluate the GWR model. Section 3 shows the details of the specific implication of the optimized GWR in this study and accuracy assessment for GWR results. Sections 4 and 5 show the results and discussion, respectively. Summary and conclusion are given in Section 6.

2. Study Area and Datasets

2.1. Study Area

Heilongjiang province, the most northeastern part of China, is located approximately between 43°N and 53°N latitude and 121°W and 135°W longitudes (Figure 1). It has a continental monsoon climate between temperate and boreal zones, with an average annual temperature of -4°C to 4°C decreasing from southeast to northwest. Its annual average precipitation ranges from 300 mm to 700 mm, which mostly concentrated from June through August [24]. Nicknamed as the “Great northern granary”, this subarctic province produced more than 60.71 million tons of grain in 2016 on 11.87 million ha of cultivated land [25]. Its main crops include soybean, rice, and corn, all of which are harvested once per year due to the limited hours of sunshine and accumulated heat. During the past decade, the soybean production of Heilongjiang accounted for more than one-quarter of the nation’s annual soybean production. However, due to the rising import of genetically modified (GM) soybean from Brazil and United States, the non-GM soybean planting area in Heilongjiang decreased from 40,322 km² in 2005 to 24,761 km² in 2015, with corn replacing soybean as the dominant dry-land crop in this region [4,25]. A field survey in 2013 showed that the soybeans were mainly distributed in northwestern part of Heilongjiang Province, and soybean fields in regions near Songnen plain are almost continuous and large, while those in the northwest part close to Lesser Khingan Mountains are relatively small and fragment. Soybean in Heilongjiang is generally sowed in late-April-followed by emergence, three leaves, seven leaves, blooming, bearing pod, filling seed, senescence and harvested in early-October (Figure 2). The distinct difference in growing period length and biochemical properties between soybean and other crops in this region will be expressed as the variance of spectral signature (e.g., NDVI) across the growing seasons, which is the crucial characteristic to identify soybeans.

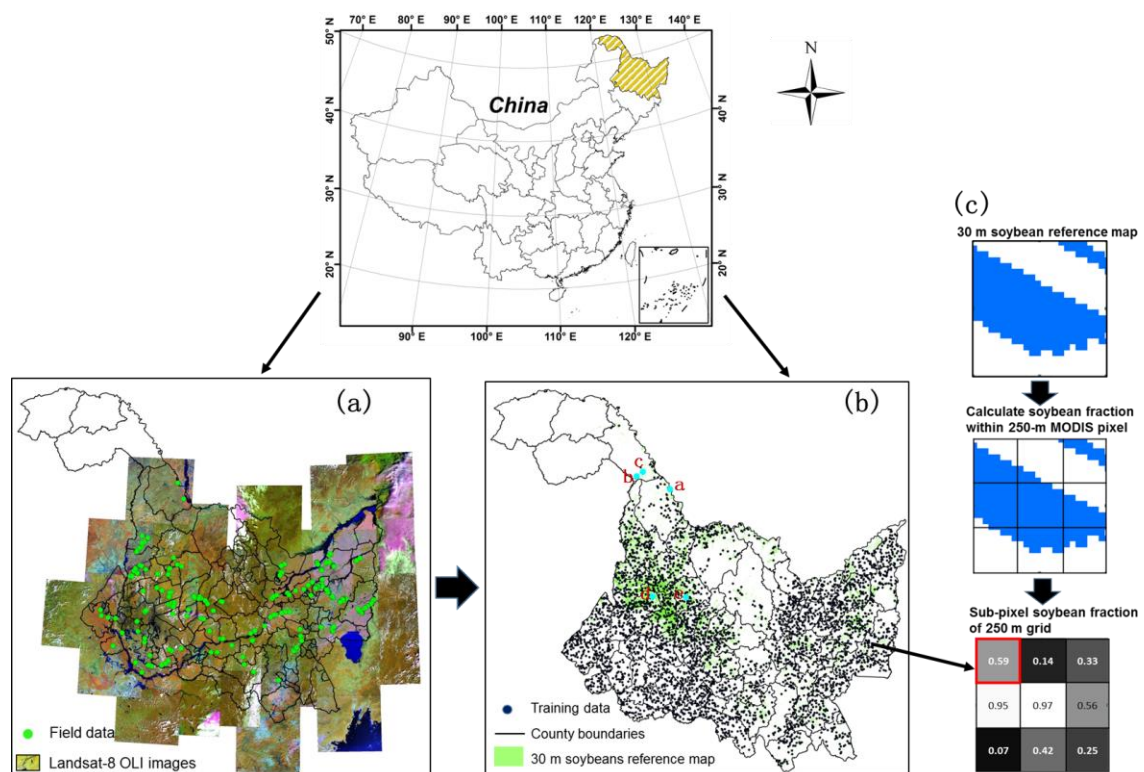


Figure 1. (a) Landsat-8 OLI images and field data (green dots) used to derive the 30-m soybean reference map; (b) 30 m soybean reference map, from which 4000 training points (blue dots) were selected as training data for GWR model; (c) An example to illustrate how soybean fraction was calculated at 250-m grid based on the 30-m soybean reference map.



Figure 2. Key phenological events of soybean cultivation in Heilongjiang Province.

2.2. Sample Data

The reference data used to calibrate the GWR model and assess the prediction results was from a classified soybean map of partial regions in Heilongjiang, which was generated by supervised classifications based on 45 cloud-free Landsat-8 OLI images in 2013 and partial expert visual interpretation. The accuracy of this mapped soybean map was 93.84% that was assessed by 341 field data, which was shown in Figure 1a [26]. The reference soybean map with a spatial resolution of 30 m was re-projected to the MODIS standard Sinusoidal projection. Then, the sub-pixel soybean fraction for each 250-m grid was calculated by dividing the number of 30-m soybean pixels by the total number of 30-m pixels within the 250-m grid cell (An example was shown in Figure 1c). Finally,

a total of 4000 samples points (Figure 1b) were randomly selected from the resampled reference map to guarantee the representativeness of training data in geographical space and values distribution. The estimated sub-pixel soybean fractions of those 4000 points range from 0 to 1 and were used as the values of dependent variables for GWR model. The histogram for the dependent variable was illustrated in Section 4.3. In addition to training points, a total of 2000 points with estimated soybean fractions were kept aside for validation.

2.3. Feature Set

In our study, time series of the 8-day composite MODIS/Terra Surface Reflectance Collection 6 products (MOD09Q1 V006) were used as the main data input. The MOD09 Q1 product, with a spatial resolution of ~250 m (Band 1, red, 620–670 nm; Band 2, near infrared, 841–876 nm) and a temporal repeat of 8-days, has already been cloud screened, atmospherically corrected, and transformed to a standard Sinusoidal projection. Due to the trade-offs involving spatial details, temporal information and areal coverage, MOD09Q1 has been deemed particularly suitable to characterize the seasonal dynamic of land cover classes and monitor large-region land cover change [9,27]. Four MODIS tiles (i.e., h25v03, h26v03, h26v04, and h27v04) covering the entire Heilongjiang region during the key growing season, from Julian days 65 to 305 (31 time points) in 2013 were selected and downloaded from the USGS EROS Data Center (<http://edc.usgs.gov/>). These tiled MODIS data were mosaicked and clipped to the extent of the study area. Also, NDVI was calculated based on Band₁ and Band₂ for each time point and used as the candidate input variables of GWR (shown as Equation (1)).

$$NDVI = \frac{Band_2 - Band_1}{Band_2 + Band_1} \quad (1)$$

where Band₁ is the MODIS surface reflectance of red (620–670 nm) and Band₂ is the MODIS surface reflectance of near infrared (841–876 nm).

In addition to MODIS-NDVI time series, the cropland map of Heilongjiang Province extracted from GlobeLand30 [28], the world's first global land cover dataset at a 30 m resolution, was resampled to MODIS pixel grid and then was used to mask non-agriculture areas within the satellite data sets.

2.4. Census Data

In this study, soybean cultivation statistics of 2013 were acquired for county level from the China Agricultural Statistics Yearbook published by Chinese Ministry of Agriculture [25]. The soybean statistical data of 80 counties in Heilongjiang were used to compare the soybean planting area estimated from GWR-derived soybean fraction maps.

3. Methodology

This paper is focused on using an optimized GWR model to estimate the sub-pixel soybean fractions in Heilongjiang, province based on time-series MODIS data. 4000 training points were used to calibrate the GWR model, where soybean fractions of 250-m MODIS pixels estimated from 30-m soybean reference map was used as the dependent variable for the GWR model. A total of 31 time-series NDVI, from Julian days 65 to 305, were used as the candidate input features. A forward stepwise strategy was used to select the optimal features by minimizing the corrected Akaike Information Criterion (AICc). The optimal NDVI temporal features selected from the forward stepwise strategy were used as the final independent variables for the GWR model. There were three ways to evaluate the performances of GWR model in estimating the sub-pixel crop fractions: (1) comparisons with OLS-based result; (2) assessment by using 2000 validation points with soybean fractions estimated from Landsat-based results; and (3) comparisons with county-level census data. The flow diagram in Figure 3 outlines the main analysis steps, which are described in detail in the following sections.

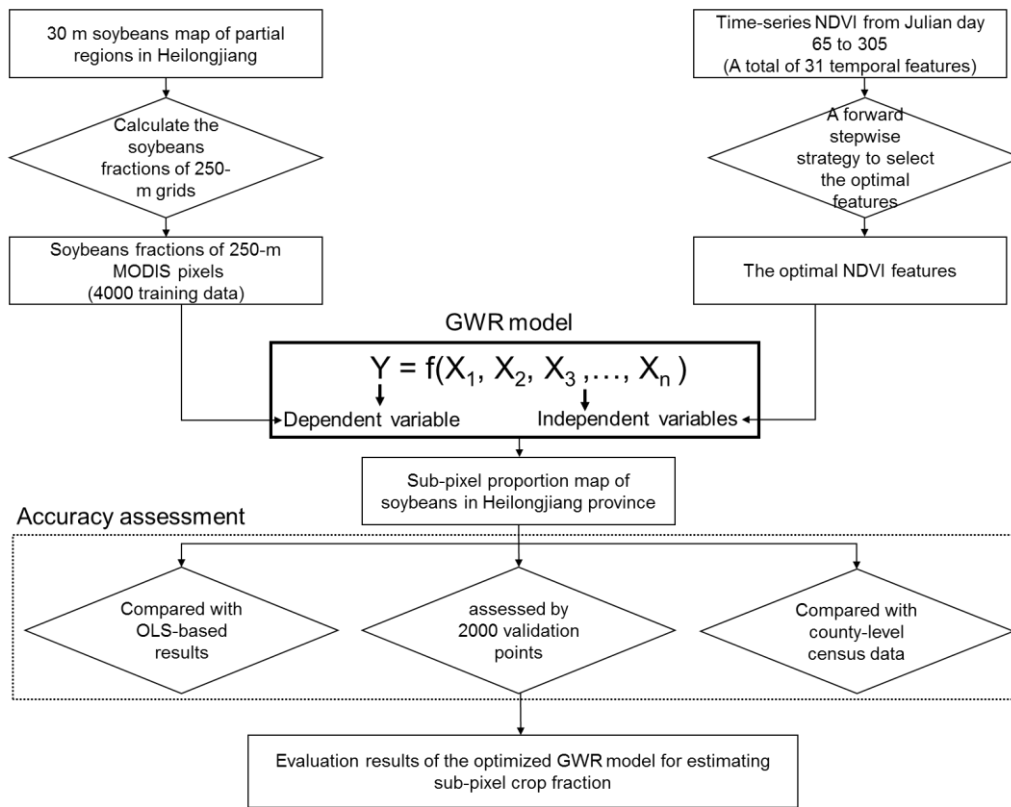


Figure 3. Overview of the optimized GWR model for estimating the sub-pixel soybean fractions and the evaluation methods. The equation of GWR here is simplified and the detailed equation of GWR and associated parameters are described in Section 3.1.1.

3.1. An Optimized Geographically Weighted Regression Model

3.1.1. The Basic Framework of GWR Model

The traditional ordinary least squares (OLS) regression, often called a global logistic regression model, assumes that relationships between independent variables and the dependent variable are spatially stationary and the regression coefficients are also constant across space. The OLS model structure can be expressed as follows:

$$\hat{y}_i = \hat{\beta}_{0i} + \hat{\beta}_{1i}X_{1i} + \dots + \hat{\beta}_{ni}X_{ni} + \varepsilon_i \quad (2)$$

where \hat{y}_i is the dependent variable, X_{1i} to X_{ni} are independent variables, $\hat{\beta}_{0i}$ is the constant parameter specific to location i , $\hat{\beta}_{1i}$ to $\hat{\beta}_{ni}$ are estimated regression coefficients and ε_i is the independent error term at location i with distribution $N(0, \sigma^2)$;

GWR, a local regression approach, recognizes that the parameter estimates in a regression relationship can vary across the study area and thereby allows the parameter estimates to be a function of location. The above equation can be modified to the following form [19]:

$$\hat{y}_i = \hat{\beta}_{0i}(u, v) + \hat{\beta}_{1i}(u, v)X_{1i}(u, v) + \dots + \hat{\beta}_{ni}(u, v)X_{ni}(u, v) + \varepsilon_i \quad (3)$$

This regression equation supposes the regression parameters to be estimated at a location for which the spatial coordinates are provided by the variables u and v . GWR accounts for the spatial drift in relationships by localizing the regression through a moving window (kernel). In this model, the regression

- Step 1: Start by calibrating all possible bivariate GW regressions by in turn regressing a single explanatory variable against the dependent variable. Calculate AICc in each case (31 runs in this study). Select the variable that produces the lowest AICc.
- Step 2: Sequentially introduce a variable from the remaining $n - 1$ features to construct new models with the permanently included independent variable in step 1. Calculate the change in AICc between step1 and step 2. Select variable yielding greatest reduction in AICc. Add this variable to the model.
- Step 3: Repeat step 2 until no independent variables among the candidate variables can be added into the model, and the model at this point is the final model.

The function in GWmodel package to perform this procedure is `model.selection.gwr`, where AICc values were sorted using function `model.sort.gwr` and then were visually presented using `model.view.gwr`. Figure 4 shows the iteration process of the stepwise selection strategy in selecting the optimal features from the candidate thirty-one features, where the dependent variable (i.e., the soybean proportion within MODIS pixel) is located at the center of the vortex and different shape in different color represents different independent variable (i.e., 31 NDVI acquired at different dates).

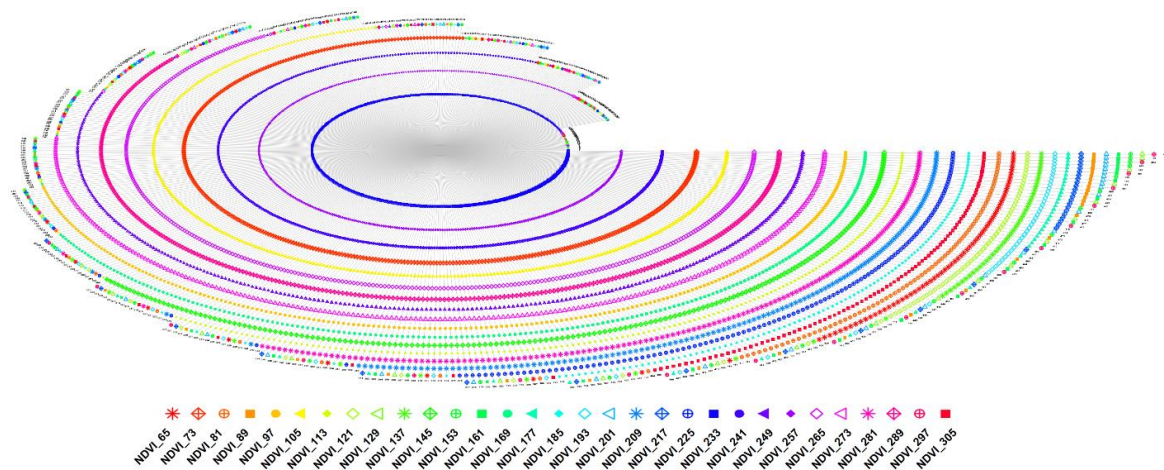


Figure 4. The iteration process of the forward stepwise strategy for selecting the optimal NDVI features (independent variables) from the thirty-one candidate NDVI variables. The dependent variable is located at the center of the vortex. Different symbol in different shape represents different NDVI temporal variable. The iteration process of the forward stepwise selection strategy starts from the first vortex circle (inside) and ends at the last vortex (outside), where the symbol with the largest amount in each vortex circle was the optimal NDVI variable of this iteration.

3.1.3. *F*-Test Statistics

With the selected optimal subset of independent variables, we run the GWR and OLS models based on the 4000 training pixels and then used the calibrated models to predict soybean fractions for all MODIS pixels of the whole study area. Descriptive statistics, including the overall AICc, the residual sum of squares (RSS), R^2 and adjusted R^2 , were used to reflect the model fitting performance and the spatially inhomogeneous of the GWR. Then, an *F*-test statistical testing method proposed by Leung et al. [33], was utilized to test the improvements of GWR over OLS in fitting the model and to understand whether or not each variable in the GWR model vary significantly.

To test whether GWR is superior than OLS, we used the F_1 -test statistics in Leung et al. [31]. Specifically, if the null hypothesis— H_0 : there is no significant difference between OLS and GWR models for the given

data—is true, the quantity RSS_g/RSS_o is close to one. Otherwise, it tends to be small. The F_1 -value can be calculated as [33]:

$$F_1 = \frac{RSS_g/\sigma_1}{RSS_o/(n-p-1)} \quad (8)$$

where RSS_g and RSS_o are the residual sum of squares of GWR and OLS model, respectively. σ_1 is the standard deviation of error and also equal to the mean of RSS_g/σ^2 where σ^2 is constant variance. $n-p-1$ represents the degrees of freedom in the denominator. The distribution of F_1 -value may be approximated by an F -distribution with σ_1^2/σ_2 degrees of freedom in the numerator and $n-p-1$ degrees of freedom in the denominator. If F_1 -value $< F_{1-\alpha}(\frac{\sigma_1^2}{\sigma_2}, n-p-1)$, where α is the given significance level, the null hypothesis will be rejected and it is thus believed that the GWR model is significantly better than the OLS in describing the data. Otherwise, we will conclude that the GWR model cannot improve the model fitting significantly compared to the OLS model.

To test whether each variable varies significantly across the study region, we used the F_3 -test statistics in Leung et al. [31]. The null hypothesis is: $H_0: \beta_{1k} = \beta_{2k} = \dots = \beta_{nk}$ (for a given k), the F_3 -value can be calculated as:

$$F_3(k) = \frac{V_k^2/\gamma_1}{\hat{\sigma}^2} \quad (9)$$

where V_k^2 is a statistic that can reflect the spatial variation of the given set of parameters $\{\beta_{ik}; i = 1, 2, \dots, n\}$. $\hat{\sigma}^2$ is an unbiased estimator of σ^2 . Its distribution can be approximated by an F -distribution with γ_1^2/γ_2 degrees of freedom in the numerator and δ_1^2/δ_2 degrees of freedom in the denominator. If $F_3 \geq F(\gamma_1^2/\gamma_2, \delta_1^2/\delta_2)$, reject H_0 ; otherwise, accept H_0 . More details about the F -test statistical testing method are provided by Leung et al. [33].

3.2. Accuracy Assessment

In this study, the performances of the GWR model in estimating the sub-pixel crop area were assessed in three ways: first was to compare the model fitting result of GWR and OLS (described in Section 3.1.3), second was the most traditional accuracy assessment using a separate validation dataset and third was comparing area estimates to available national statistics data. For the second way, the sub-pixel soybean fraction map derived from GWR was assessed using 2000 points derived from Landsat-based reference map that is described in Section 2.2. These 2000 validation points were separate from the 4000 training samples. For the third evaluation on area estimates, the sub-pixel soybean map was aggregated to the total acreage of soybean cultivation for each pixel by multiplying the prediction values (between 0 to 1) by the area of each pixel. Then, the derived soybean acreage was compared with census data at county level by overlaying administrative boundaries with the soybean cultivation area map. The results of these two assessments were quantified by computing several statistical measures, including the RMSE, NRMSE (normalized root mean square error) and R^2 . They are defined as:

$$RMSE = \sqrt{\sum_{i=1}^k (a_i' - a_i)^2 / k} \quad (10)$$

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \quad (11)$$

$$R^2 = \frac{cov(a, a')^2}{var(a)var(a')} \quad (12)$$

For the assessment of using validation points, a_i' was the estimated soybean fraction from GWR, a_i was the soybean fraction of 250 m pixel that was aggregated from Landsat-based reference map, and k , the number of validation points, is 2000. y_{\max} is the maximum sub-pixel soybean fractions and y_{\min} is the minimum sub-pixel soybean fractions of those 2000 validation points. For the assessment of

using census data, a_i/t is the GWR-derived soybean acreage, and a_i is the census data, and k , the number of counties in Heilongjiang, is 80. y_{max} is the maximum area and y_{min} is the minimum areas of those 80 counties.

4. Results

4.1. Descriptive Statistics

The histograms of independent variables (NDVI values) as well as the dependent variable (sub-pixel soybean fraction) are illustrated in terms of frequency distribution in Figure 5. Five NDVI images that are acquired at different Julian days and refer to different phenological phases of soybeans are shown as examples for the total thirty-one candidate variables. Among the five variables, Julian day 121 when soybeans are still at emergence stage has the lowest frequency for NDVI values of larger than 0.4, with a mean value of 0.2485 and a standard deviation (Std.Dev) of 0.0446. As time goes on, the pixel number of NDVI larger than 0.4 increases and that smaller than 0.4 decreases dramatically. Both the frequency for values between 0.8–1.0 and the mean NDVI value approach the peaks at Julian day 217 when soybeans are bearing pods and contain higher vegetation density than other periods. From bearing pod stage to filling seed stage, partial green leaves begin to turn yellow, which makes NDVI gradually decrease with the mean value dropping from 0.7769 to 0.6740. Overall, the changes of NDVI frequency distribution across seasons well reflect the phenological dynamics of soybean cultivation.

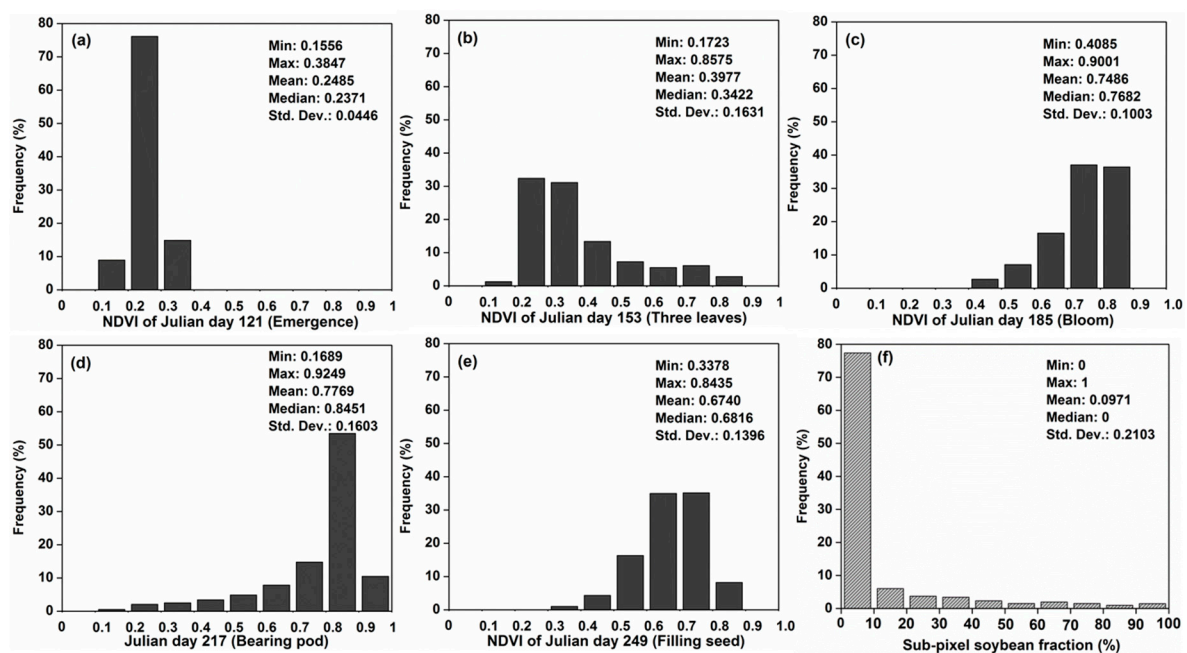


Figure 5. Histograms and descriptive statistics for the independent variables (a–e) and dependent variable (f). These five NDVI data acquired from different Julian day respond to different phenological phases and are shown as examples to represent the candidate independent variables for the 31 NDVI variables.

For the dependent variable, the sub-pixel soybean fraction value ranges from 0 to 1, with a mean value of 0.0971 and a standard deviation of 0.2103. However, the frequency distribution of these 4000 training pixels is uneven because a clear majority of pixel values are less than 0.1. This high frequency for low values indicates the soybean field is generally fragment in Heilongjiang Province, which also implies the necessity of sub-pixel calculation of crop area when using MODIS images.

In addition to analyzing the value distribution characteristics of independent and dependent variables, we also observed whether these values changes are related to spatial distances. The MODIS-NDVI time

series spectral curves of three pure soybean pixels in different spatial positions (a–c in Figure 1a) and two mixed pixels (d, e in Figure 1a) were illustrated to characterize soybean temporal-spectral curves of different soybean fractions and to examine whether the distance of spatial locations impact the variations of crop time-series spectral curves. Their NDVI time-series curves ranging from day 65 to day 305 were shown in Figure 6. We observed that pixel “c”, which is spatially close to pixel “b”, presents much similar spectral curve characteristic with pixel “b”. By contrast, the pixel “a” further away from pixel “b”, shows a large discrepancy in spectral curve with the pixel “b”, especially for the peak value of curve and associated dates. This implies the spatial instability of crop signature should be considered when the relationship between sub-pixel soybean area and time-series NDVI is explored.

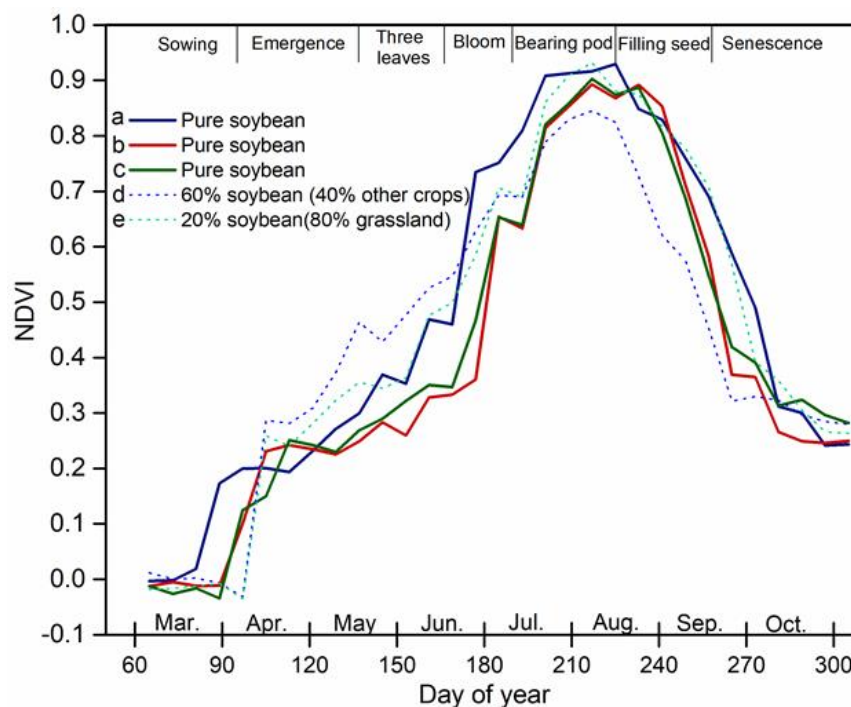


Figure 6. Time-series NDVI curves of pure soybean pixels (a–c) and mixed pixels (d,e) during the entire crop growing period in Heilongjiang. The specific spatial locations of those five pixels were described in Figure 1. These five pixels were presented here to show the time-series curve characteristics of different soybean fractions and to examine whether the distance of spatial locations is related to the variations of crop time-series spectral curves.

4.2. The Optimal Independent Variables for Soybean Cultivation

In this study, the best adaptive bandwidth (number of nearest neighbors) is 736, with the minimum AICc value of -3087.201 . This best value means 736 nearest neighbors (points) are included as local samples to perform local regression. Figure 7 displays the corresponding AICc values of models composed by different features derived from the forward stepwise selection strategy, which corresponds to the vortices in Figure 4. Obviously, AICc value continues to decrease as the number of independent variable increases to 31. The model (point B in Figure 7) containing all 31 NDVI variables is the best for calculating sub-pixel soybean fraction in terms of the minimum AICc value. In addition, it can be found the model at point A has the biggest change rate of AICc, indicating those associated variables play the most critical roles in reducing AICc values and thus are considered as the optimal features among the 31 NDVI variables for sub-pixel soybean area calculation. Specifically, the number of independent variables at point A is four, and there are NDVI of Julian day 233, 257, 241 and 249, respectively. During this period (day 233 to 257), soybean cultivation is at the stage of filling seeds when the yellowing leaves and changing canopy structure due to the growing seeds provides good separation with other classes.

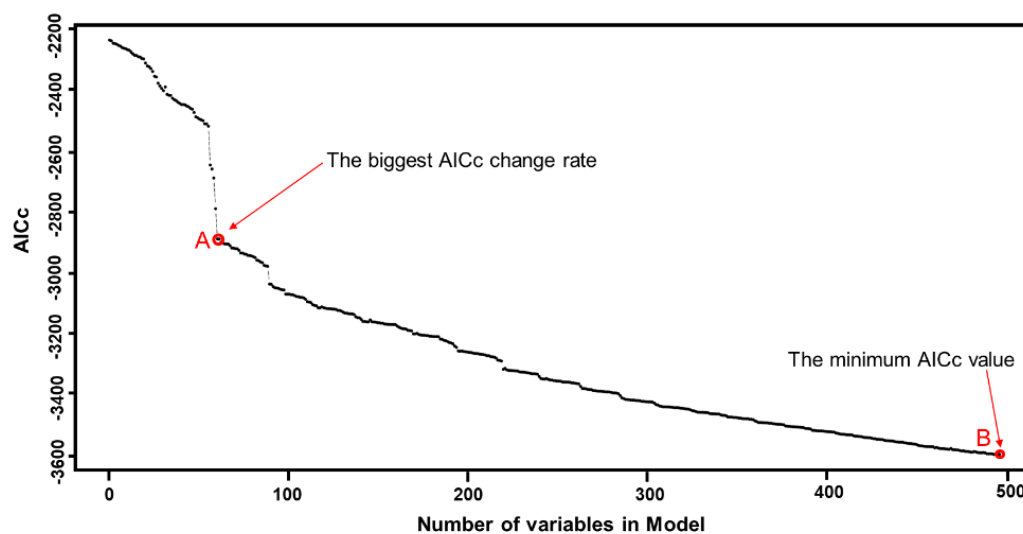


Figure 7. AICc values of different feature combinations derived from the forward stepwise strategy in Section 3.1.2. Each point corresponds to one vortex circle in Figure 4.

4.3. GWR Results of Sub-Pixel Soybean Area Estimation

Model fitting results (AICc, RSS, R^2 , and adjusted R^2) of GWR and OLS are summarized as Table 1. The adjusted overall mean local R^2 for GWR model is 0.4377, meaning 43.77% of the variability in the sub-pixel soybean proportion can be explained in the modeling dataset. Compared to OLS model, significant improvements for all descriptive statistics indicates GWR is superior in exploring the relationships between the time-series NDVI values and sub-pixel crop area. In addition, the non-stationarity F -test results including F_1 -test and F_3 -test between GWR and OLS is shown in Table 2. The F_1 -value indicates the null hypotheses is rejected and there is a significant difference between GWR and OLS model. Additionally, it is observed 21 variables of the total 31 NDVI variables as well as intercept show highly significant non-stationarity ($\alpha = 0.001$ level), suggesting the response of the sub-pixel soybean area to those variables is not constant across the study area. By contrast, only six NDVI variables (i.e., NDVI of Julian day 81, 89, 137, 153, 201 and 209) are not statistically significant, indicating they keep relatively stable relationships with the dependent variable. Not surprisingly, the four optimal variables, i.e., NDVI of Julian day 233, 241, 249, and 257 (Section 4.2) are found statistically significant, which confirms the importance of images acquired at filling seed stage for mapping sub-pixel soybean area. This F -test results also highlight the advantages of GWR over OLS model in expressing the spatially variant relationships between crop area and remote sensing images.

Table 1. Statistics of model fitting for GWR and OLS models.

Model	AICc	RSS	R^2	Adjusted R^2
GWR	−3087.201	85.8289	0.5144	0.4377
OLS	−2220.059	132.1226	0.2524	0.2466

It should be noted here the coordinate system based on which the distance between different points are calculated may impact the bandwidth and consequently the regression accuracy of GWR. For our work, the geographic coordinate system (GCS/WGS84) was used to calculate great circle distance between different points as we have demonstrated it outperforms the Euclidean distance based on projected coordinate system (PCS/Sinusoidal). The performances of model fitting based on these two distance calculation ways are displayed in Table 3. It shows their optimal bandwidth is different, and WGS84 has lower AICc and RSS, and higher R^2 and adjusted R^2 . The higher accuracy derived by calculating great

circle distance may be explained by the fact the spherical distance can characterize the real geographical correlations between different points more accurately than the plane distance.

Table 2. Statistics of non-stationarity F -test between GWR and OLS model.

	Variables	F -Value	NDF	DDF	p -Value (Significance)
F_1 -test	/	0.75	3606.49	3966	$<2.20 \times 10^{-16}$ ***
F_3 -test	Intercept	6.85	348.37	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_65	1.81	523.21	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_73	2.59	457.76	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_81	1.78	445.29	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_89	0.72	353.16	3606.5	0.999973
	NDVI_97	4.75	332.68	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_105	2.08	258.34	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_113	3.61	269.81	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_121	1.15	484.77	3606.5	0.019104 *
	NDVI_129	1.02	485.38	3606.5	0.354737
	NDVI_137	1.15	736.19	3606.5	0.006462 **
	NDVI_145	1.32	589.97	3606.5	2.18×10^{-6} ***
	NDVI_153	0.53	666.26	3606.5	1
	NDVI_161	0.86	907.46	3606.5	0.998103
	NDVI_169	2.46	904.98	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_177	1.25	610.68	3606.5	8.12×10^{-5} ***
	NDVI_185	1.04	407.12	3606.5	0.296108
	NDVI_193	1.40	629.78	3606.5	6.44×10^{-9} ***
	NDVI_201	0.85	292.13	3606.5	0.969006
	NDVI_209	0.77	226.61	3606.5	0.995028
	NDVI_217	1.64	252.09	3606.5	4.06×10^{-9} ***
	NDVI_225	4.43	160.93	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_233	7.19	227.22	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_241	13.18	336.43	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_249	2.70	632.45	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_257	11.21	711.35	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_265	2.93	772.44	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_273	3.30	955.55	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_281	2.40	737.06	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_289	2.13	622.15	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_297	1.80	418.44	3606.5	$<2.20 \times 10^{-16}$ ***
	NDVI_305	1.55	634.59	3606.5	1.28×10^{-14} ***

NDF and DDF represent the degrees of freedom of the numerator and denominator, respectively. More details see Leung et al. [33]. *** Significant at the $\alpha = 0.001$ level; ** Significant at the $\alpha = 0.01$ level; * Significant at the $\alpha = 0.05$ level.

Table 3. A comparison of model-fitting performances for GWR based on different distance calculation ways.

Distance Calculation Ways	Bandwidth	AICc	RSS	R^2	Adjusted R^2
GCS/WGS84 (Great circle)	736	−3087.201	85.8289	0.5144	0.4377
Sinusoidal (Euclidean distance)	687	−2928.25	88.4218	0.4997	0.4177

4.4. Sub-Pixel Soybean Map and Accuracy Assessment

Figure 8 shows the continuous map of soybean cultivation derived from the optimized GWR model using MODIS-NDVI time series, where the value presents the proportion of each MODIS pixel occupied by soybean cultivation and can also reflect the field heterogeneity as its patch sizes are smaller than the spatial resolution of MODIS (~250 m). As expected, the mapped soybean cultivation mainly concentrated in the northwest part close to Lesser Khingan Mountains, which is one of the most important soybean production regions in China. In addition, Sanjiang Plain to the northeast and the south part of Heilongjiang are also observed have some small fractions of soybean cultivation. Those regions were reported their soybean planting areas had been dramatically decreased during the

last five years due to a great deal of importation of GM soybean from other countries, such as Brazil and the United States [4].

A total of 2000 random validation points selected from Landsat-based fractional soybean map were used to assess our MODIS-based soybean extent in a spatially explicit fashion. A direct differencing comparison based on 10% bins of increase in percentage of soybean cover are illustrated in Figure 9. Overall, the regression line composed of median values was close to 1:1 line as hoped, indicating the spatial distribution of soybean cultivation derived from time-series MODIS data generally agree with Landsat-based soybean maps. Nevertheless, relative to Landsat data, MODIS results underestimate the values of larger than 0.2 while overestimates those of less than 0.2. Additionally, the vertical bars that respond to standard deviations of soybean fractions derived from MODIS show soybean with the proportion between 0.4–0.6 has relatively large bias compared to others. These big biases tend to mostly happen for those pixels which are mixtures of soybean and corn. This is because field survey (Figure 1a) shows the soybean fields always adjoin to corn fields and some of them are even intercropped with each other in Heilongjiang. Furthermore, previous studies have reported there are high levels of spectral confusion between soybean and corn [6,34]. Hence, those pixels occupied by a large proportion of corn will significantly reduce the identifiability of their soybean areas and consequently increase the regression bias.

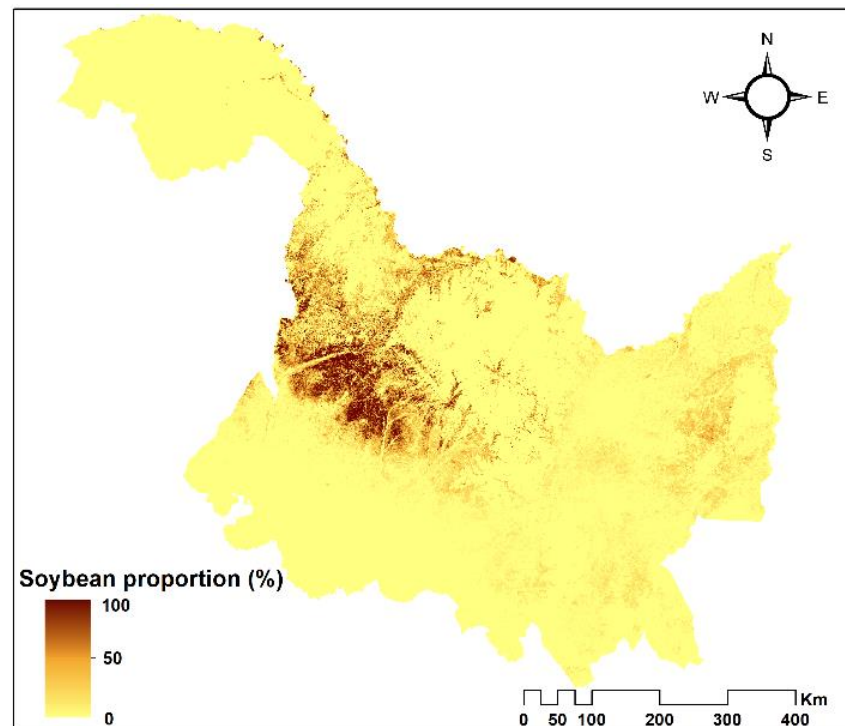


Figure 8. Sub-pixel proportion map of soybean cultivation in Heilongjiang province derived from the optimized GWR model.

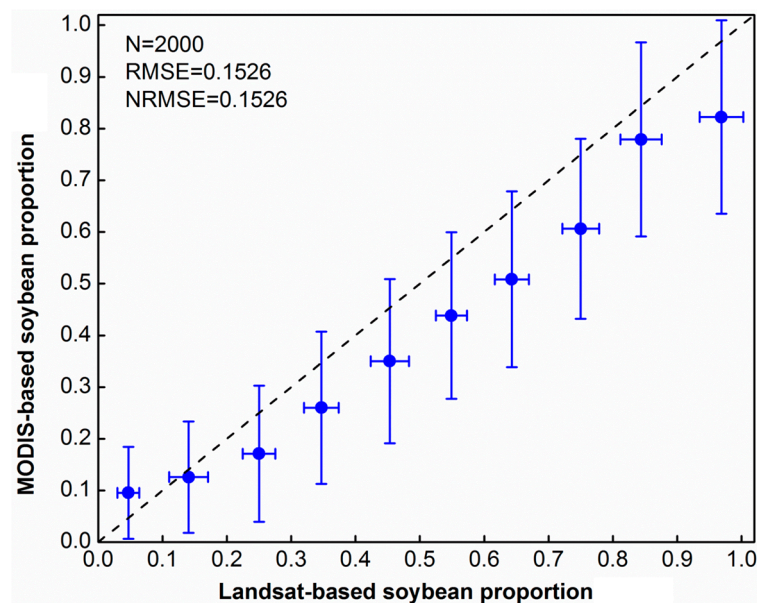


Figure 9. MODIS versus Landsat sub-pixel soybean fractions in Heilongjiang based on 2000 validation samples. Results are summarized for each 0.1 bin (0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, 0.4–0.5, 0.5–0.6, 0.6–0.7, 0.7–0.8, 0.8–0.9, 0.9–1 increase). The dots indicate median soybean proportion at each bin. The bars correspond to the standard deviation of sub-pixel soybean fractions derived from Landsat (horizontal direction) and MODIS (vertical direction), which can reflect the regression bias.

A comparison on the soybean cultivation areas estimated from the optimized GWR and the agricultural census data at the county level is shown in Figure 10. The two estimates of soybean acreage were significantly correlated ($\alpha < 0.01$) at county levels, with R^2 of 0.80 and RMSE of 340.21 km² and NRMSE of 0.1054 across 80 counties. A closer observation for this figure shows our GWR model based on MODIS data overestimated soybean areas of more than 750 km², while almost underestimated those areas of less than 500 km².

The map in Figure 11 displays the spatial distribution of differences (both in soybean area and percentage) between the MODIS-based and census-based soybean cultivation area estimates for each county in Heilongjiang Province. It is observed that the largest percentage differences between these two datasets are in those counties, which are in the center part of Heilongjiang and has a soybean area of less than 500 km², such as Yichun, Hegan, Suiling, Hailun, Tangyuan and Baiquan. The high landscape heterogeneity and small field size are responsible for the relatively large confusion between soybean and non-soybean spectral signature in those regions. Overall, in the north part of Heilongjiang, our GWR model based on MODIS data overestimated soybean cultivation. In contrast, our GWR model slightly underestimated soybean acreage in northern districts, such as Shanzhi, Dongning, Mulin, Linkou and Hailin. The obvious regional similarity in the regression bias may be because the adjoining areas present similar spectral characteristics for soybean cultivation, and thereby they may have the similar or the same regression relationships between sub-pixel soybean fractions and time-series NDVI. It should be noted here some errors could be due to the uncertainty of the census data, since crop area information is acquired from sampling survey or communication with farmers which are not always accurate [13].

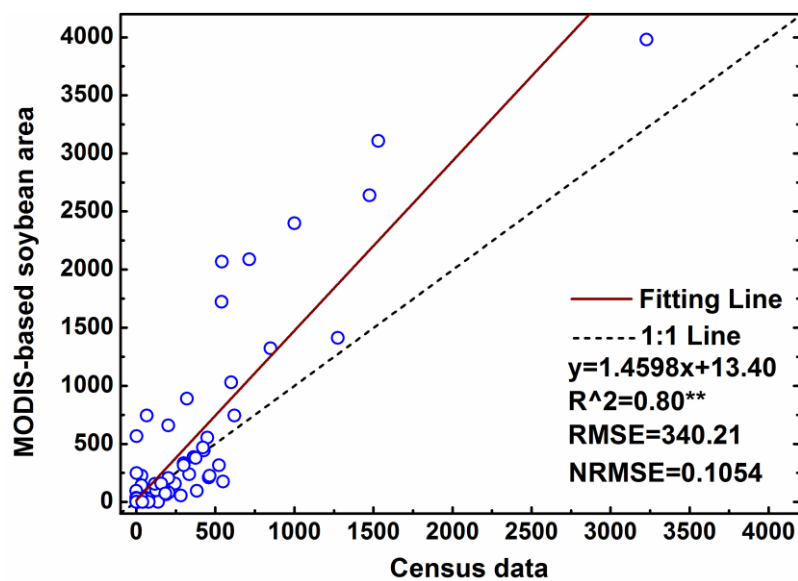


Figure 10. Regression-based comparisons between the MODIS-based soybean area estimates and census data respectively at county level. ** Significant at the $\alpha = 0.001$ level.

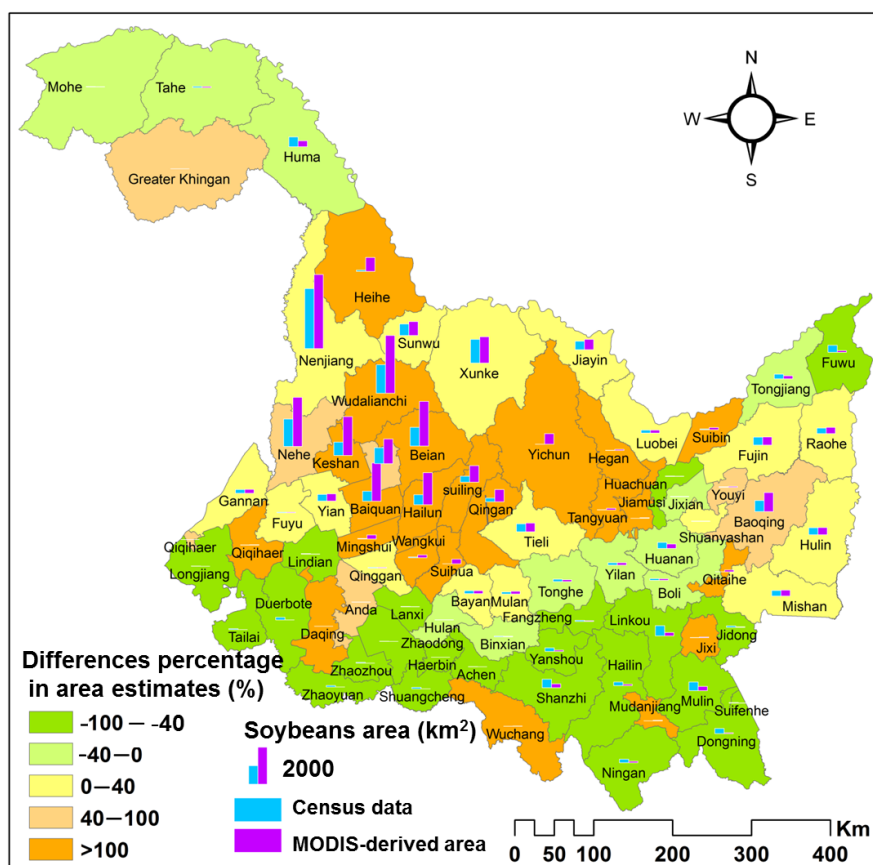


Figure 11. The spatial distribution of agreement between MODIS-based result and census data on soybean cultivation area for each county in Heilongjiang Province. The colors show the percentage (%) differences between the two estimates and the accompanying bar plots show the actual area comparison of the two estimates (km^2).

5. Discussion

In this study, we used an optimized GWR model to explore the non-stationary spatial relationships between soybean area and time-series NDVI values on a sub-pixel scale and then predict soybean fraction for the whole Heilongjiang Province, China. Our results have high consistency both with Landsat-based soybean estimates and census data, indicating that satellite remote sensing technology when given high quality reference data can provide a simple but effective way to estimate regional crop area compared to traditional sampling survey or communication with farmers. In addition, compared to the OLS model, the superior performance of the GWR model also implies the spatial variation of spectral signature across time and the spatial nonstationary within these NDVI variables should be considered when the relationships between crop area and time-series NDVI is explored. The regression results achieved for soybean cultivation also indicate the strengths of the MODIS time series data for crop mapping as its high temporal and spectral resolutions effectively preserve the dynamic phenological characteristics of crops. It is well known that one critical shortcoming in mapping land cover over large areas based on MODIS data is its coarse spatial resolution, which can be partially overcome by using this regression method that can calculate area fraction at the sub-pixel level. Since such data are available at global scale and with nearly twenty years of data, there is great potential for sub-pixel crop mapping at large region including changes to crop cultivation through time, especially when it combined with GWR model.

For the selection of optimal independent variables, the minimum AICc is achieved when all the 31 time-series NDVI are employed together, among which NDVI of Julian day 233, 241, 248 and 257 were found the most crucial variables in significantly reducing AICc. This period (233 to 257) responds to the filling seed stage of soybean cultivation in Heilongjiang, when the yellowing leaves and unique seed texture make the spectral signature of soybean significantly different from other classes. It should be noted here since phenological characteristics may vary from region to region, these selected spectro-temporal features for soybean cultivation in Heilongjiang may not necessarily be optimal for other study areas and other crop types. However, the forward stepwise regression method can effectively select the optimal combination of independent variables and provide valuable insights for understanding the variable importance, which is highly recommended for variables selection of GWR especially when there is a large number of candidate independent variables.

It is well known that spectral variations, including inter-class and intra-class spectral variation, are essential factors for the accuracy of spectral unmixing [15]. In this study, inter-class spectral variation is caused by the different land cover classes and their associated area proportions within MODIS pixels. Due to different agricultural intensive management (e.g., irrigation, fertilization, and sowing), local weather conditions and soil quality, the same crop type especially those spatially further away may present different spectral signature, thus resulting in intra-class spectral variations. The results of GWR model show the relationships between sub-pixel crop fraction and spectral signatures vary from location to location, with similar relationships for adjacent pixels and vice versa. This can be explained by the fact that the close soybean MODIS pixels may share more similar climate environment, agricultural planting conditions and phenological characteristics than those far away from each other, which makes them have smaller intra-class spectral variations and thus having similar relationships with sub-pixel crop fractions. Therefore, the specific advantage of GWR model in estimating sub-pixel crop area is that GWR can largely reduce intra-class spectral variation by building local regression relationships between sub-pixel crop fraction and spectral signatures.

Despite the wide application of GWR models in remote sensing fields, its potential in modelling the spatial distribution of crop type has never been explored before. Our study is the first effort to explore the spatial variation in the relationships between sub-pixel crop area and remote sensing data and to predict the crop area at regional scale based on the GWR model. Our conclusions extend and strengthen the findings of Pan et al. [10] and Lobel et al. [11], who found the quantitative relationships between sub-pixel crop fractions and MODIS time-series data but ignored the spatial non-stationarity of such relationship. In addition, our presented GWR model improved the regular GWR model by

using a forward stepwise strategy to select the optimal dependent variables from time-series remote sensing data, which shows great potential when having a large of candidate independent variables for GWR. Even though the GWR model was used in this study to estimate the sub-pixel soybean cultivation in a single province in China using MODIS time series data, it is easily generalizable to any other crop type (e.g., corn or rice) and any input data (e.g., VIIRS) since it is entirely automated. However, it is noteworthy that the landscape fragmentation has negative impacts on the accuracy of crop area estimation. The landscape heterogeneity impacts estimates of crop area by increasing variability in crop spectral reflectance at decreasing spatial scales [10]. Also, the fragmented cropland always increases the difficulty for sensors to detect the weak spectral signature within MODIS pixels and therefore will reduce the accuracy of GWR prediction. Since the landscape fragmentations may vary with locations, more regions with different agriculture landscapes and more remote sensing images with different spatial resolution needed to be tested in the future to improve the robustness of GWR model in estimating sub-pixel crop fraction.

However, some potential improvements could be considered before the presented GWR model is applied to larger areas and more crop types. First, a wide range of independent variables, such as different vegetation indices and phenological metrics (e.g., mean, standard deviation, minimum and maximum) in addition to NDVI could be tested to improve the regression accuracy. Previous studies have proved NDVI is not necessarily the optimal spectral feature for identifying specific crop type (e.g., rice) [6,35], and some studies also reported the inclusion of phenological metrics could significantly improve classification accuracy [1,36]. Second, an extension to GWR will be developed, which aims to realize the possibility of creating models where some variables are held constant across the study area while others can vary spatially. Such models, which are known as mixed or semi-parametric models [32,37], could be tested in the future for application to other study areas or other land cover classes. In addition, some models [38] that considered non-stationarity both in space and time will be also explored to improve the accuracy in regions where agricultural landscapes is fragment and cropping pattern is complex. Finally, it is well known that the representativeness in quality and quantity of training samples, as well as their spatial homogeneity are quite important for GWR [1,39]. However, one of the most challenging problems is how to collect and compile such a dataset for regions without existing high-quality reference data. Fortunately, more and more freely satellite data of finer spatial and temporal resolutions than MODIS becomes available, such as Sentinel1-A [40], Sentinel2-A [41], Landsat 8 [42], GaoFen-1/2 [43]. High-quality classification results from these datasets can be used as reference information to acquire reference samples for large-scale land cover mapping efforts on a sub-pixel level based on coarse spatial resolution images (e.g., MODIS and VIIRS).

6. Conclusions

In this study, we presented and evaluated a method for estimating the fractional soybean cultivation at 250 m spatial resolution using the MODIS-NDVI time series. We developed a geographically weighted regression model calibrated using sub-250 m percentage soybean information from Landsat, focusing on Heilongjiang region of China in 2013. A forward stepwise selection optimization strategy was specially used to select the optimal dependent variables from the time-series NDVI for the GWR model. All 31 time-series NDVI were identified as the optimal combination of independent variables with minimum AICc. NDVI of Julian day 233, 241, 249, and 257 were found the most important temporal variables for sub-pixel soybean area calculation when soybean is at the stage of filling seed. Our analyses showed that estimated percent soybean at 250 m compared well with Landsat-based results. There was also a good agreement between our MODIS-based results and census data, with R^2 of 0.80 and RMSE was 340.21 km² across 80 counties. The largest differences between these two datasets were observed in regions with high landscape heterogeneity and small field size. Additionally, the optimized GWR model had better performances with lower RSS and higher R^2 compared to OLS model, and their F -test results shown 23 variables of the total 31 time-series NDVI are statistically significant. These promising results indicate the spatial nonstationary within spectral-based variables should be considered when the

relationship between crop sub-pixel fraction and remote sensing images is explored. Our presented GWR model shows great potential in mapping sub-pixel crop distribution over large region based on coarse spatial resolution imagery such as MODIS. However, if the presented GWR model is to be extended to other regions and land cover types, further development of the GWR method is possible, such as testing a variety of the input variables (e.g., vegetation indices and phenological metrics) and developing a mixed GWR model that allow some variables held constant while others vary spatially. In addition, future work could address the issues on how to use multi-source finer spatial resolution data to improve the selection of training and validation samples.

Acknowledgments: This research was supported by the National Key Research and Development Program of China (2017YFD0300201 and 2017YFE0104600), by the China Academy of Engineering Consulting Project (2016-ZCQ-08), and by National Natural Science Foundation of China (Grant Nos. 41401116). We are grateful to the anonymous reviewers and academic editor for their valuable suggestions and comments. We also greatly appreciate the researchers at the Heilongjiang Academy of Agriculture Sciences for sharing essential crop reference maps and field soybean samples.

Author Contributions: Qiong Hu, Wenbin Wu and Huajun Tang conceived and designed the experiments. Qiong Hu, Yaxiong Ma, Baodong Xu and Qian Song performed the experiments and analyzed the data. All authors contributed to the writing of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhong, L.; Gong, P.; Biging, G.S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 1–13. [[CrossRef](#)]
2. Monfreda, C.; Ramankutty, N.; Foley, J.A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Glob. Biogeochem. Cycles* **2008**, *22*, GB1022. [[CrossRef](#)]
3. Zheng, H.; Chen, L.; Han, X.; Zhao, X.; Ma, Y. Classification and regression tree (cart) for analysis of soybean yield variability among fields in northeast china: The importance of phosphorus application rates under drought conditions. *Agric. Ecosyst. Environ.* **2009**, *132*, 98–105. [[CrossRef](#)]
4. Sun, J.; Wu, W.; Tang, H.; Liu, J. Spatiotemporal patterns of non-genetically modified crops in the era of expansion of genetically modified food. *Sci. Rep.* **2015**, *5*, 14180. [[CrossRef](#)] [[PubMed](#)]
5. Wardlow, B.; Egbert, S.; Kastens, J. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the u.S. Central great plains. *Remote Sens. Environ.* **2007**, *108*, 290–310. [[CrossRef](#)]
6. Hu, Q.; Wu, W.; Song, Q.; Yu, Q.; Lu, M.; Yang, P.; Tang, H.; Long, Y. Extending the pairwise separability index for multicrop identification using time-series MODIS images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6349–6361. [[CrossRef](#)]
7. Hu, Q.; Wu, W.-B.; Song, Q.; Lu, M.; Chen, D.; Yu, Q.-Y.; Tang, H.-J. How do temporal and spectral features matter in crop classification in heilongjiang province, china? *J. Integr. Agric.* **2017**, *16*, 324–336. [[CrossRef](#)]
8. Chang, J.; Hansen, M.C.; Pittman, K.; Carroll, M.; DiMiceli, C. Corn and soybean mapping in the united states using MODIS time-series data sets. *Agron. J.* **2007**, *99*, 1654. [[CrossRef](#)]
9. Huang, X.; Schneider, A.; Friedl, M.A. Mapping sub-pixel urban expansion in china using MODIS and DMSP/OLS nighttime lights. *Remote Sens. Environ.* **2016**, *175*, 92–108. [[CrossRef](#)]
10. Pan, Y.; Li, L.; Zhang, J.; Liang, S.; Zhu, X.; Sulla-Menashe, D. Winter wheat area estimation from MODIS-evi time series data using the crop proportion phenology index. *Remote Sens. Environ.* **2012**, *119*, 232–242. [[CrossRef](#)]
11. Lobell, D.B.; Asner, G.P. Cropland distributions from temporal unmixing of MODIS data. *Remote Sens. Environ.* **2004**, *93*, 412–422. [[CrossRef](#)]
12. Atzberger, C.; Rembold, F. Mapping the spatial distribution of winter crops at sub-pixel level using avhrr ndvi time series and neural nets. *Remote Sens.* **2013**, *5*, 1335–1354. [[CrossRef](#)]
13. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sens. Environ.* **2011**, *115*, 1301–1316. [[CrossRef](#)]

14. Löw, F.; Michel, U.; Dech, S.; Conrad, C. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS J. Photogramm. Remote Sens.* **2013**, *85*, 102–119. [[CrossRef](#)]
15. Drumetz, L.; Chanussot, J.; Jutten, C. Variability of the endmembers in spectral unmixing: Recent advances. In Proceedings of the 8th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Los Angeles, CA, USA, 21–24 August 2016.
16. Zhang, L.; Wei, Y.; Meng, R. Spatiotemporal dynamics and spatial determinants of urban growth in Suzhou, China. *Sustainability* **2017**, *9*, 393. [[CrossRef](#)]
17. Propastin, P.A. Spatial non-stationarity and scale-dependency of prediction accuracy in the remote estimation of lai over a tropical rainforest in sulawesi, indonesia. *Remote Sens. Environ.* **2009**, *113*, 2234–2242. [[CrossRef](#)]
18. Wu, J.; Yao, F.; Li, W.; Si, M. Viirs-based remote sensing estimation of ground-level PM_{2.5} concentrations in Beijing–Tianjin–Hebei: A spatiotemporal statistical model. *Remote Sens. Environ.* **2016**, *184*, 316–328. [[CrossRef](#)]
19. Fotheringham, A.S.; Brunson, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2002.
20. You, W.; Zang, Z.; Zhang, L.; Li, Z.; Chen, D.; Zhang, G. Estimating ground-level PM₁₀ concentration in northwestern china using geographically weighted regression based on satellite aod combined with calipso and MODIS fire count. *Remote Sens. Environ.* **2015**, *168*, 276–285. [[CrossRef](#)]
21. Fang, X.; Zou, B.; Liu, X.; Sternberg, T.; Zhai, L. Satellite-based ground PM_{2.5} estimation using timely structure adaptive modeling. *Remote Sens. Environ.* **2016**, *186*, 152–163. [[CrossRef](#)]
22. Propastin, P. Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 82–90. [[CrossRef](#)]
23. See, L.; Schepaschenko, D.; Lesiv, M.; McCallum, I.; Fritz, S.; Comber, A.; Perger, C.; Schill, C.; Zhao, Y.; Maus, V.; et al. Building a hybrid land cover map with crowdsourcing and geographically weighted regression. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 48–56. [[CrossRef](#)]
24. Hu, Q.; Wu, W.; Xia, T.; Yu, Q.; Yang, P.; Li, Z.; Song, Q. Exploring the use of google earth imagery and object-based methods in land use/cover mapping. *Remote Sens.* **2013**, *5*, 6026–6042. [[CrossRef](#)]
25. Agriculture, C.M.O. *China Agricultural Statistics Yearbook*; China Statistics Press: Beijing, China, 2015.
26. Rui, X.; Zhongjun, L.; Yang, L.; Bin, F.; Kebao, L. Comparison on linear feature real width and interpretation width using landsat tm8 images and gf-1 images. *Trans. Chin. Soc. Agric. Eng.* **2015**, *16*. [[CrossRef](#)]
27. Wardlow, B.D.; Egbert, S.L. Large-area crop mapping using time-series MODIS 250 m ndvi data: An assessment for the u.s. Central great plains. *Remote Sens. Environ.* **2008**, *112*, 1096–1116. [[CrossRef](#)]
28. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [[CrossRef](#)]
29. Lesiv, M.; Moltchanova, E.; Schepaschenko, D.; See, L.; Shvidenko, A.; Comber, A.; Fritz, S. Comparison of data fusion methods using crowdsourced data in creating a hybrid forest cover map. *Remote Sens.* **2016**, *8*, 261. [[CrossRef](#)]
30. Salas, C.; Ene, L.; Gregoire, T.G.; Næsset, E.; Gobakken, T. Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sens. Environ.* **2010**, *114*, 1277–1285. [[CrossRef](#)]
31. Guo, L.; Ma, Z.; Zhang, L. Comparison of bandwidth selection in application of geographically weighted regression: A case study. *Can. J. Forest Res.* **2008**, *38*, 2526–2534. [[CrossRef](#)]
32. Gollini, I.; Lu, B.B.; Charlton, M.; Brunson, C.; Harris, P. Gwmodel: An r package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.* **2015**, *63*, 1–50. [[CrossRef](#)]
33. Leung, Y.; Mei, C.L.; Zhang, W.X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ. Plan. A* **2000**, *32*, 9–32. [[CrossRef](#)]
34. Maxwell, S.K.; Nuckols, J.R.; Ward, M.H.; Hoffer, R.M. An automated approach to mapping corn from landsat imagery. *Comput. Electron. Agric.* **2004**, *43*, 43–54. [[CrossRef](#)]
35. Xiao, X.; Boles, S.; Frolking, S.; Li, C.; Babu, J.Y.; Salas, W.; Moore, B. Mapping paddy rice agriculture in south and southeast asia using multi-temporal MODIS images. *Remote Sens. Environ.* **2006**, *100*, 95–113. [[CrossRef](#)]
36. Colditz, R.R.; López Saldaña, G.; Maeda, P.; Espinoza, J.A.; Tovar, C.M.; Hernández, A.V.; Benítez, C.Z.; Cruz López, I.; Ressler, R. Generation and analysis of the 2005 land cover map for mexico using 250 m MODIS data. *Remote Sens. Environ.* **2012**, *123*, 541–552. [[CrossRef](#)]

37. Mei, C.L.; He, S.Y.; Fang, K.T. A note on the mixed geographically weighted regression model. *J. Reg. Sci.* **2004**, *44*, 143–157. [[CrossRef](#)]
38. Huang, B.; Wu, B.; Barry, M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 383–401. [[CrossRef](#)]
39. Schepaschenko, D.; See, L.; Lesiv, M.; McCallum, I.; Fritz, S.; Salk, C.; Moltchanova, E.; Perger, C.; Shchepaschenko, M.; Shvidenko, A.; et al. Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and fao statistics. *Remote Sens. Environ.* **2015**, *162*, 208–220. [[CrossRef](#)]
40. Torbick, N.; Chowdhury, D.; Salas, W.; Qi, J. Monitoring rice agriculture across myanmar using time series sentinel-1 assisted by landsat-8 and palsar-2. *Remote Sens.* **2017**, *9*, 119. [[CrossRef](#)]
41. Radoux, J.; Chomé, G.; Jacques, D.; Waldner, F.; Bellemans, N.; Matton, N.; Lamarche, C.; d’Andrimont, R.; Defourny, P. Sentinel-2’s potential for sub-pixel landscape feature detection. *Remote Sens.* **2016**, *8*, 488. [[CrossRef](#)]
42. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [[CrossRef](#)]
43. Li, Z.; Shen, H.; Li, H.; Xia, G.; Gamba, P.; Zhang, L. Multi-feature combined cloud and cloud shadow detection in gaofen-1 wide field of view imagery. *Remote Sens. Environ.* **2017**, *191*, 342–358. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).