Estimation of crop biomass using GF−3 polarization SAR data based on genetic algorithm feature selection

XU Kunpeng[1], ZHAO Lei[1], LI Kun[2], CHEN Erxue[1], ZHANG Wangfei[3], YANG Hao[4]

1. *Institute of Forest Resources Information Technique, Chinese Academy of Forestry, Beijing* 100091*, China*;
2. *Beijing Institute of Spacecraft System Engineering, Beijing* 100094*, China*;
3. *College of Forestry, Southwest Forestry University, Kunming* 650224*, China*;
4. *Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing* 100097*, China*

Abstract: In recent years, Polarization SAR (PolSAR) has been widely used in the filed of crop biomass estimation. However, high dimensional features extracted from PolSAR data will lead to information redundancy which will result in low accuracy and poor transfer ability of the estimation model. Aiming at this problem, we proposed a estimation method of crop biomass based on automatic feature selection method using Genetic Algorithm (GA). Firstly, the backscattering coefficient, the polarization parameters and texture features were extracted from PolSAR data. Then, these features were automatically pre-selected by GA to obtain the optimal feature subset. Finally, based on this subset, a Support Vector Regression machine (SVR) model was applied to estimate crop biomass. The proposed method was validated using the GaoFen-3 (GF-3) QPSI (C-band, quad-polarization) SAR data. Based on wheat and rape biomass samples acquired from a synchronous field measurement campaign, the proposed method achieve relative high validation accuracy (over 80%) in both crop types. For further analyzing the improvement of proposed method, validation accuracies of biomass estimation models based on several different feature selection methods were compared. Compared with feature selection based on linear correlation, GA method has increased by 5.77% in wheat biomass estimation and 11.84% in rape biomass estimation. Compared with the method of Recursive Feature Elimination (RFE) selection, the proposed method has improved crops biomass estimation accuracy by 3.90% and 5.21%, respectively.

Key words: polarization SAR, estimation of crop biomass, genetic algorithm, feature selection, GaoFen-3

Citation format: Xu K P,Zhao L,Li K,Chen E,Zhang W F and Yang H. 2020. Estimation of crop biomass using GF−3 polarization SAR data based on genetic algorithm feature selection. Journal of Remote Sensing (Chinese). 24(S1): 1−9

1 INTRODUCTION

Crop biomass is an important indicator reflecting the growth status and carbon stocks of crops. The use of remote sensing technique to monitor crops biomass in large area plays an important role in global carbon cycle research and agricultural management. Compared to optical remote sensing, Synthetic Aperture Radar (SAR) is more sensitive to vegetation biomass. Therefore, SAR has been widely used for crop biomass estimation (Mcnairn H, et al., 2004; Mattia F, et al., 2003). The fully polarimetric SAR (PolSAR) system provides complete polarization information of the target, which could be beneficial to the identification of main scattering mechanisms of crops. Therefore, the use of PolSAR data for crop monitoring has become the current research focus (Tao L, et al., 2016; Hosseini M, et al., 2017).

Based on the PolSAR data, many features such as backscattering coefficients, polarization parameters and texture parameters can be extracted. Some studies analyzed the relationship between the characteristics and crop parameters, and then established several estimation models (Hosseini M, et al., 2015; Yang H, et al., 2016). However, most of these estimation models are parametric model based on linear regression, which cannot characterize the complex multivariate nonlinear relationship between the features and vegetation parameters, resulting in low estimation accuracy. Therefore, non-parametric models with multi-factor nonlinear fitting ability have been widely used (Zhang W, et al, 2017; Vafaei S, et al., 2018) in recent years. Typical non-parametric model includes K Nearest Neighbor (k-NN), Random Forest (RF) and Support Vector Regression machine (SVR). As SVR can achieve a relative high accuracy using small sample set, it has been widely used in the field of quantitative remote sensing applications. However, the information redundancy caused by the excessive feature dimension severely limits the prediction accuracy and transfer abil-

Received: 2019−10−21; **Accepted:** 2020-04-16

Foundation: the National Key R&D Program of China (No.2017YFB0502700); the Technique of Accurate Surface Parameters Inversion using GF-3 Images (No.03-Y20A11-9001-15/16); the National Natural Science Foundation of China (No. 41801289).

First author biography: XU Kunpeng (1992—), male, PhD student, His research interests are polarimetric SAR applications to forestry and agriculture. E-mail: xkp1231@163.com

Corresponding author biography: ZHAO lei (1988—), male, assistant professor. His research interests are related to radar remote sensing with applications to forestry. E-mail: zhaolei@ifrit.ac.cn

ity of the method. Therefore, how to reduce the feature dimensions and select the appropriate feature combination is extremely important.

The typical feature selection method uses the indicators that are independent of the model training to evaluate the candidate features, such as linear correlation coefficient, rank correlation coefficient, and information gain (Chandrashekar G, et al., 2014). The feature set selected by this kind method usually cannot guarantee high estimation accuracy since the operations are independent of model training. Recursive Feature Elimination (RFE) is a well-known feature selection method commonly used in machine learning model such as SVM and RF for image classification and quantitative parameter estimation (Guyon I, et al, 2002; Xiaoxiao L, et al, 2017). This method evaluated feature combinations based on their performance in the model. So that the selected feature combination usually can achieve better accuracy. Nevertheless, the RFE method adopt a greedy searching strategy with searching direction decided by the model feedback of the training process. Therefore, the method has a fixed searching direction and limited searching range, which make the results easily fall into the local optimal solution.

Genetic Algorithm (GA) is a heuristic model that simulates the mechanism of natural selection, and has excellent global searching ability. Unlike the RFE algorithm, the searching direction of the GA doesn't depend on model feedback. Thus, the searching direction is not fixed and the potential searching range is all possible combinations of the features. In practice, GA feature selection method has been widely used in remote sensing classification (Maryam S, et al, 2014; Ataollah H G, et al, 2011), but it is rarely used in quantitative vegetation parameter estimation.

In summary, PolSAR can provide a large number of features for the crop biomass estimation, but how to automatically select the feature subset for modeling is the key to achieve high-precision estimates. In this paper, we introduced GA into the estimation process of crop biomass. Firstly, the backscattering coefficients, the polarization and the texture features were extracted from GF-3 Quad-polarimetric data. Then, selected the optimal feature subset based on GA. Finally, SVR model is applied to estimate the crop biomass using the optimal feature subset.

2 STUDY AREA AND DATA

2.1 Study Area

The study area within the Shangkuli Farm (120° 46′E—120°53′E; 50°17′N—50°23′N) (2800 hm^2) in Inner Mongolia, China, is located at the northwest of Greater Khingan Mountain and the north of Hulunbuir steppe (Fig.1). The terrain of the study area is relatively flat and contains some gentle hills. Climate in the study area is the cold temperate continental monsoon climate. It has a cold and dry long winter and warm and wet short summer, which supports one harvest per year. The main crops in the region are wheat and rape, which are cultivated annually from May to September.

2.2 PolSAR Data

GF-3 PolSAR (C-band) data was acquired over the study area on 3 August 2017, with level 1.1 quad polarization strip 1 (QPSI) mode. The azimuth and range resolution of this model was about 8 m, with the pixel spacing of azimuth and range were 5.01 m and 4.50 m, respectively. The coverage of the data (Fig.1) was 25 km wide (east west) and 30 km long (north south). In addition, the center incidence angle was 48.8°. The Pauli RGB display of the GF-3 PolSAR data was shown in Fig.1.

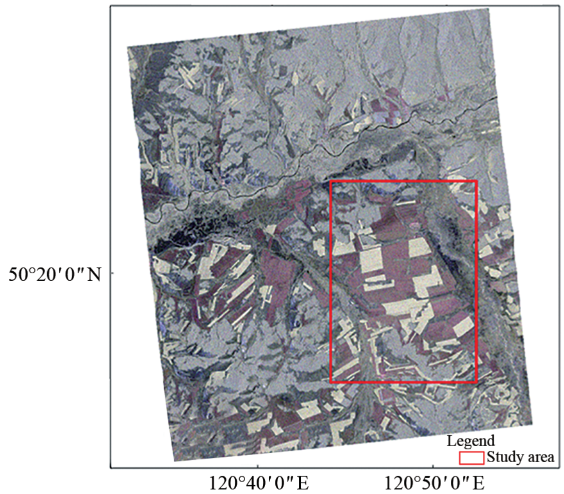


Fig. 1 Location map of the study area in Inner Mongolia, China

2.3 Ground truth data

Synchronous field measurement campaign was carried out for the satellite overpass from August 1st to 9th, 2017. As shown in Fig.2, 52 plots were investigated, including 30 wheat plots and 22 rape plots. For each plot, three repeated measurements were performed and their geographical positions were recorded with a Trimble Pathfinder® Differential GPS (50cm precision). For the wheat plot, 20 plants were taken as samples in a square of 0.5 m × 2 rows for each measurement. For the rape plot, 5 plants were taken as samples in a square of 2.0 m × 2 rows for each measurement. Then, the collected crop samples were dried by a 95°C oven and weighed to obtain the dry biomass of the samples. Then, the biomass per square meter was calculated. The biomass of each plot is the mean of three measurements. The location of Each plot is represented by the center of three measurement points.

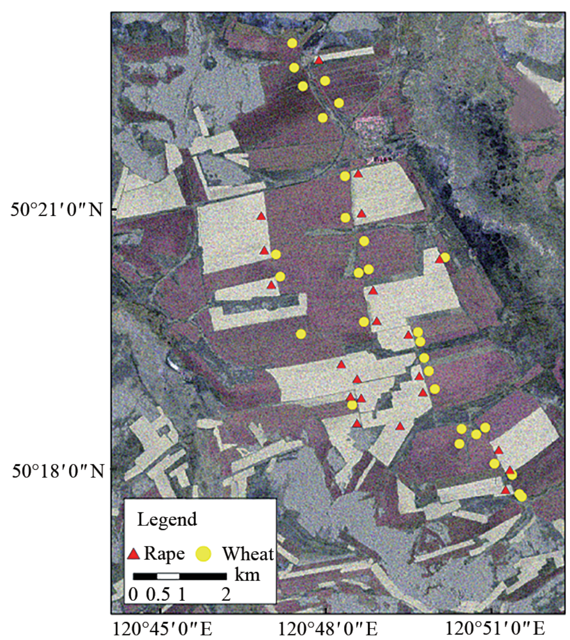


Fig. 2 Location of the ground measurements plots

3 METHODS

3.1 Framework of the Proposed Method

The processing flow chart of this study was presented in Fig.3, which includes four main data processing steps. Firstly, it is PolSAR data pre-processing. Normally, the level 1.1 Single-Look Complex (SLC) data should be calibrated, multi-looked and geocoded. Moreover, the polarization orientation angle and effective scattering area correction are performed on the PolSAR data in consideration of the influence of the terrain (Zhao L. et al., 2017). Secondly, the original features set is extracted from the PolSAR data. It mainly includes backscattering coefficient, polarization features and texture features. Then, it is feature selection processed by GA. This is the crucial step for the method. Finally, the estimation of crop biomass is performed based on the SVR model.

3.2 Feature extraction

Polarimetric SAR data can provide a large numbers of polarization and texture features. In this study, we extracted 21 features from GF-3 PolSAR data. Among them, polarization characteristics can be divided into three categories. The first category is the characteristics of the backscattering coefficients corresponding to different polarizations and their combinations, such as |HH|, |VV|, |HH+VV|, |HH-VV|, |HV|, SPAN (total backscattering power). In addition, the second one is the characteristics obtained by polarization decomposition methods, such as Freeman-Durden decomposition (Freeman A, et al., 1998), H-Alpha decomposition (Cloude, S. R. et al., 1997), Adaptive Model-Based Decomposition (Arii M, et al., 2011). The third type is the polarization parameters proposed by some studies according to various rules, such as Radar Vegetation Index (RVI) (Yamada Y, et al., 2015), Lueneburg entropy (Luneburg E., 2001). Texture features was calculated by using the Gray Level Co-occurrence Matrix (GLCM) method (De Siqueira F R, et al., 2013) based on span image. A total of five texture features were calculated, namely, the mean, homogeneity, uniformity, dissimilarity and contrast parameters. The texture features extracted above adopt a uniform window (7×7), direction orientation (0°) and gray level (32) in the calculation. Features used in the study are shown in Table 1.

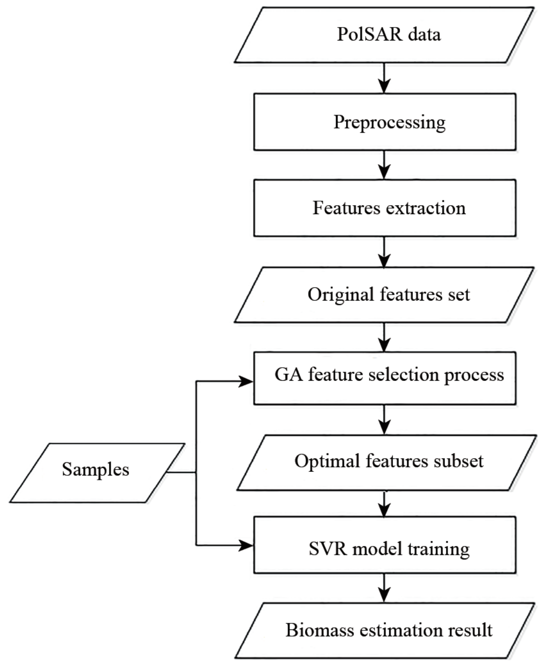


Fig. 3 Flow chart of the proposed method

Table 1 Features description

Feature type	Feature name	Description	Number
Polarization features	|HH|, |HV|, |VV|, |HH+VV|, |HH−VV|, SPAN	Backscattering coefficients parameters	16
	$Freeman_{Odd}$, $Freeman_{Vol}$, $Freeman_{Dbl}$	Freeman−Durden decomposition parameters	
	Entropy, alpha	H−Alpha decomposition parameters	
	$ANNED_{Odd}$, $ANNED_{Vol}$, $ANNED_{Dbl}$	Adaptive model−based decomposition	
	RVI, Lueneburg entropy	Radar vegetation index and Lueneburg entropy	
Textural features	Mean, Homogeneity, Uniformity, Dissimilarity, Contrast	GLCM textural features	5

3.3 Feature selection method based on GA

Genetic algorithm is implemented in a simulation in which a population of abstract representation of candidate solutions (called individuals) to an optimization problem evolves toward better solution. Corresponding to the feature selection process, each individual represents a feature combination. A binary number represents each individual. The number of bits of the binary number is equal to the number of features in the original feature set, and 0 or 1 on each bit represents whether the corresponding feature is selected or not. For example, let us assume that the original feature set has five features. Then each individual in GA is a feature subset, which is a 5-digit binary number, such as 10001, 01111, and so on. Therefore, the individual "10001" represents that this feature subset contains two features: the first and last features in the original feature set. The feature selection process based on GA is shown in Fig.4.

(1) It is population initialization and update. A population is initialized based on the original feature set. The binary code for each individual is randomly generated and the number of individuals within the population can be determined empirically. It should be noted that individuals in the population are constantly updated as the algorithm is iterated. However, the number of individuals in the population remains the same throughout the iteration.

(2) It is individual performance evaluation. This step evaluates the performance of each individual in each generation of the population. In this paper, we use the estimation accuracy (K-fold cross-validation) of the SVR model as an evaluation index. After this step, we can select the best individual in this generation of population and rank the performance of all individuals.

(3) It is genetic operations. This process will generate the next generation of population through genetic operations. It consists of three main steps: selection, crossover and mutation. In the selec-

tion operation, according to the ranking of the individual performance obtained in the previous step, individuals are randomly selected from the current population to form a new population (the higher the ranking, the greater the probability of being selected). In the crossover operation, two individuals are randomly selected, and the partial values of their binary code are exchanged to generate two new individuals. In the mutation operation, the partial coding of some individuals in the population is locally inverted according to a certain probability (0 to 1, 1 to 0). After the above operation, a new generation of population will be formed.

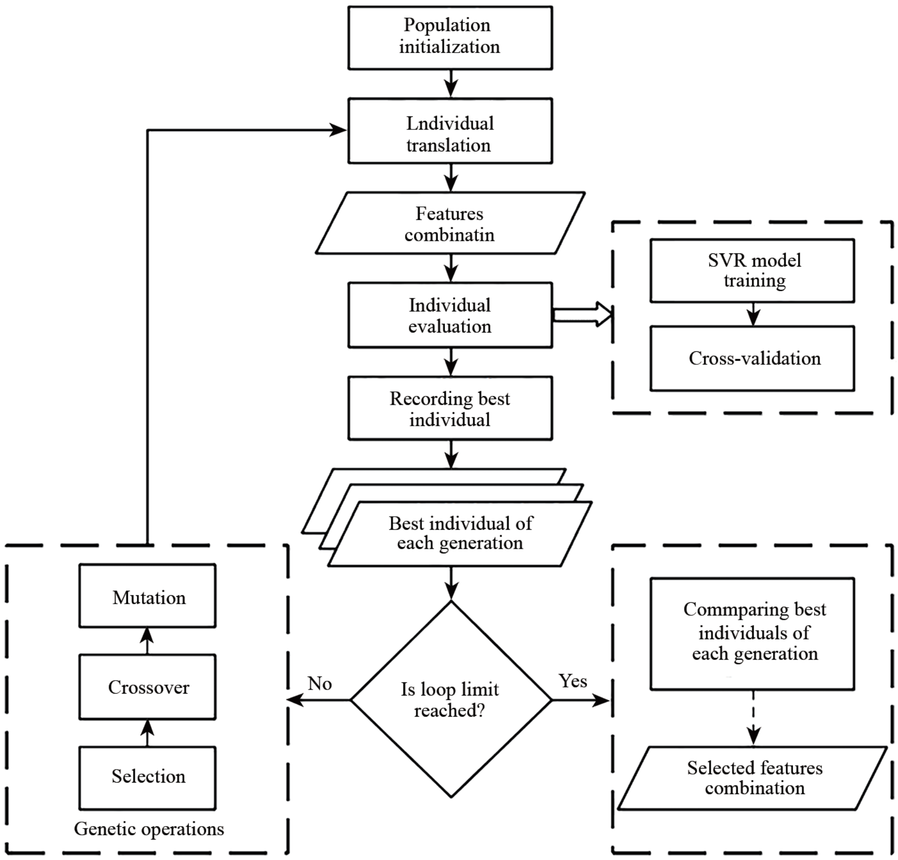


Fig. 4 Flowchart of the proposed feature selection method

(4) Select the best individual. After several generations of population updates, we recorded the best individual in each generation. Then, we can easily select the best individual of all generations. The feature combination corresponding to this individual is the result of feature selection.

Obviously, the genetic operation is the core of the genetic algorithm. The selection operation ensures the convergence of the algorithm results. The crossover and mutation operations ensure the potential of the global search, making the search results less likely to fall into the local optimal solution.

3.4 Estimation of crop biomass based on SVR

Based on the features combination selected from genetic algorithm, SVR is trained to estimate the biomass parameter. SVR algorithm is an excellent non-parameter model, which is a modified version of well-known SVM model for regression task. The model inherits the signature of SVM which it only uses partial samples in training process. Therefore, the model has the ability to focus on the main relation between input vector and output parameter. This training strategy also reduces the possibility of over fitting. Another signature of SVR model is that it uses kernel function to perform a non-linear transformation on the input vector. In result, SVR model has a strong non-linear fitting ability. In this paper, we use Radial Basis Function (RBF) as kernel function. Due to its strong performance, RBF is one of the most popular kernel functions. Since the user-defined parameters have influence on the performance of SVR, we use grid search method with cross validation to determine the user-defined parameters, which are RBF parameter (γ), penalty parameter (C) and relaxation parameter (ε).

3.5 Accuracy verification methods and indicators

In this paper, we use K-fold cross-validation method to evaluate estimation accuracy. Samples are equally divided into K parts. If there are undistributed samples, the remaining samples will be randomly added to one of the parts. Each time, we select one part of samples without repetition as the verification samples set and the remaining parts of the samples are called training samples set. Based on the model trained by the training samples set, estimation values of verification samples set are tested. Repeat k times until all samples have been tested and only tested once as validation samples. According to the field measurements of biomass and estimate value of biomass, we calculate three types of accuracy indicators to evaluate estimation model. The details of indicators are shown in the Table 2.

Table 2 Indicators of estimation model accuracy

Accuracy indicators	Symbol	Function
Root mean squared error	RMSE	$\mathrm{RMSE}=\sqrt{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2/n}$
Accuracy	Acc.	$\mathrm{Acc.}=1-\frac{\mathrm{RMSE}}{\bar{y}}=1-\frac{\sqrt{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2/n}}{\bar{y}}$
Coefficient of determination	R^2	$R^2=1-\sum_{i=1}^{n}(\hat{y}_i-y_i)^2/\sum_{i=1}^{n}(y_i-\bar{y})^2$

注：n：number of samples，$\hat{y}_i$：estimation values of sample i，y_i：field measurements of sample i，$\bar{y}$：average of the field measurement.

4 RESULTS AND ANALYSIS

4.1 Preprocessing and features extraction

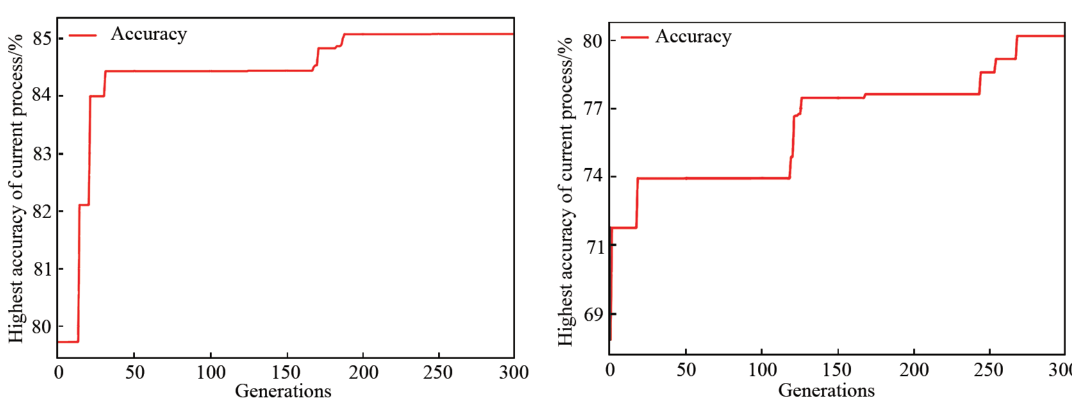
In this paper, the proposed method is verified by using a scene of GF-3 QPSI PolSAR data . First, the PolSAR image were calibrated and multi-looked with window size of 5×5. Then, the refine Lee filter with window size of 3×3 were adapted to reduce the effect of speckle noise. In the next step, the image were geocoded based on RD model. After geocoding, the resolution of the SAR image is 10 m×10 m. Finally, features extracted based on geocoded image formed the original feature set.

It is important to note that in order to ensure the operation efficiency of feature selection and modeling the extracted features were normalized. 12 features need to be converted to decibel form before normalization. They are |*HH*|, |*HV*|, |*VV*|, |*HH+VV*|, |*HH-VV*|, *SPAN, FreemanOdd, FreemanVol, FreemanDbl, ANNEDOdd, ANNEDVol, ANNEDDbl*.

4.2 Feature combination selected by different feature selection methods

In the feature selection processing, not only the GA feature selection method were used, the feature selection method based on linear correlation and the RFE algorithm were selected as the control method to experiment. For the two types of crops in the study area, the feature combination needs to be selected separately.

Based on the GA feature selection process introduced in this paper, feature selection operation was adapted to the samples set. In the population initialization operation, the number of individuals in population was set to 20. In genetic operation, the crossover probability was set to 70% and the mutation probability was 10%. The total number of iterations for the entire population was set to 300. It is important to note that there is no limit on the number of iterations. The higher the iteration number setting, the easier it is to find the global optimal feature combination. Nevertheless, there are corresponding computing costs to be paid. As shown in Fig. 5, are diagrams of iterative process of genetic algorithm based on wheat and rape samples.



（a）Feature selection using GA based on wheat samples （b）Feature selection using GA based on rape samples

Fig. 5 Feature Selection Process of GA

Fig. 5 shows the accuracy of biomass estimation using the optimal feature combination in the population of past generations in the process of population iteration. After 300 generations, wheat and rape biomass can achieve more than 80% estimation accuracy. The selected feature combinations are shown in Table 3.

The feature selection method based on linear correlation is a typical feature selection method. The method selects features based the correlation between each feature and the estimation parameter. Fig. 6 and Fig. 7 are linear coefficients between the extracted features and the biomass of wheat and rape, respectively.

According to the coefficients between the features and biomass, one-third of the total features, that is the first seven features, were selected to participate in the final biomass estimation model. As can be seen from Fig. 6, the linear coefficients between the seven features having the best correlation with wheat biomass are greater than or equal to 0.6. As can be seen from Fig. 7, the linear coefficients between the seven features having the best correlation with rape biomass are greater than 0.4. In general, the correlation between wheat biomass and features is better than that between rape biomass and features.

Table 3 Feature combinations based on different feature selection methods

Feature selection method	Crop type	Selected feature combinations	Quantity
Feature selection based on linear correlation	Wheat	|HH|, |VV|, |HH+VV|, |HH−VV|, SPAN, Mean, Contrast	7
	Rape	|HH|, | HH−VV |, SPAN, FreemanDbl, ANNEDDbl, Uniformity, Contrast	7
Feature selection of GA	Wheat	|HV|, alpha, RVI, FreemanVol, ANNEDOdd, ANNEDVol, Contrast	7
	Rape	|HV|, SPAN, Lueneburg Entropy, FreemanVol, ANNEDDbl, ANNEDVol, Mean, Contrast	9
RFE	Wheat	|VV|, |HH+VV|, |HH−VV|, FreemanOdd, ANNEDOdd, SPAN, Homogeneity, Dissimilarity, Contrast	9
	Rape	FreemanDbl, FreemanOdd, Homogeneity, Dissimilarity, Contrast	5

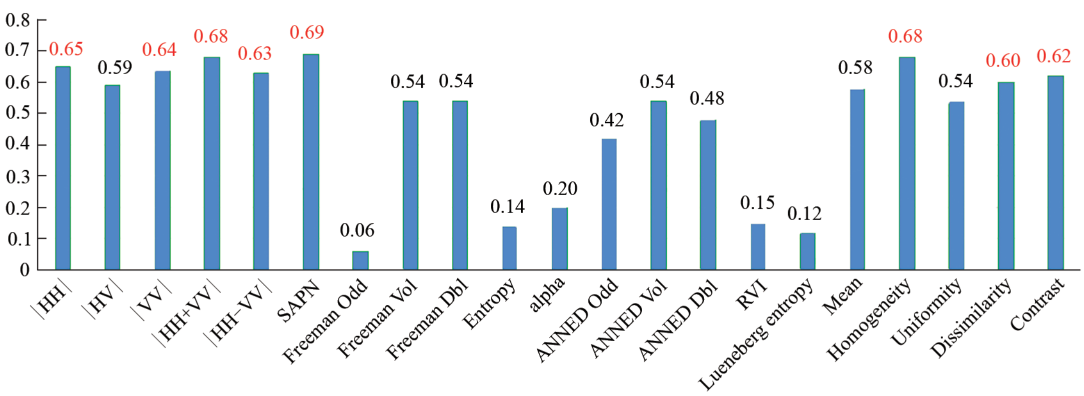


Fig. 6 Linear coefficient between features and wheat biomass

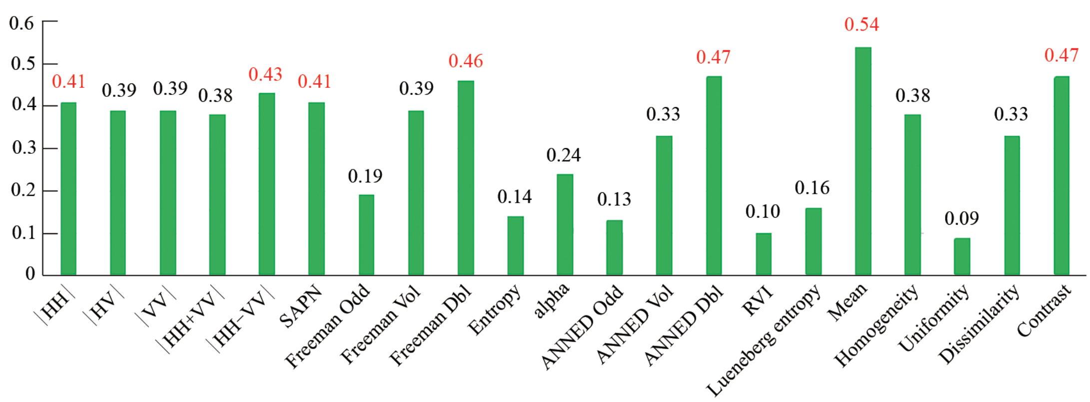


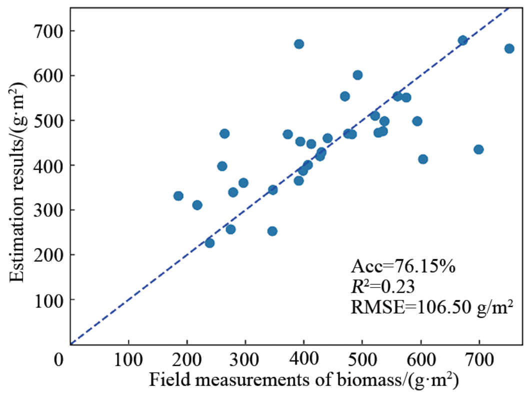
Fig. 7 Linear coefficient between features and rape biomass

RFE algorithm is one of the most used feature selection methods at present. The algorithm is usually coupled with the RF or SVR algorithm. In this paper, feature selection process of RFE algorithm were adapted based on SVR-RFE module in scikit-learn (Pedregosa F, et al., 2013). Feature selection results based on linear correlation and RFE are shown in Table 3.

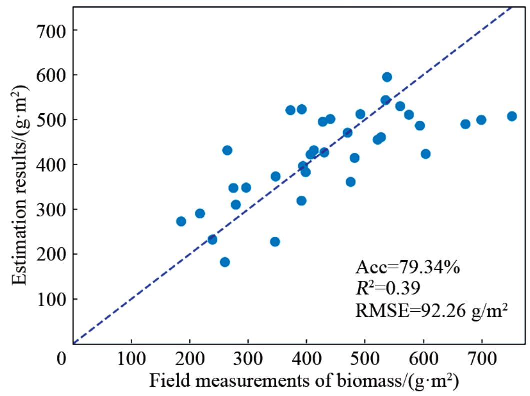
As can be seen from Table 3, features selected based on linear coefficient are mainly the backscatter coefficient feature and texture feature. Two polarization decomposition features are selected for rape samples only. Features selected based on RFE algorithm mainly include three kinds of features, such as backscatter coefficient feature, polarization decomposition feature and texture feature. Features selected based on GA contain all four kinds of characteristics extracted in this paper. Comparing with the first two methods, the search process of GA has more randomness, also pay more attention to the wholeness of feature combination. This makes the feature with poorly individual evaluations, while combining with other features with better modeling performance has the opportunity to be selected.

4.3 Estimation results based on different feature selection methods using SVR

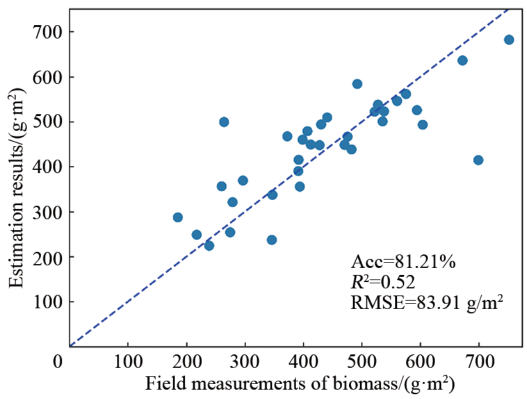
Based on the optimal feature combinations selected by the above methods and biomass sample data for both corps, SVR models for wheat and rape were trained separately. The SVR model uses RBF as the kernel function, and the model parameters were obtained by grid search method. Fig. 8 and Fig. 9 show the accuracy of biomass estimation of wheat and rape with different feature selection strategies using 10-fold cross-validation.



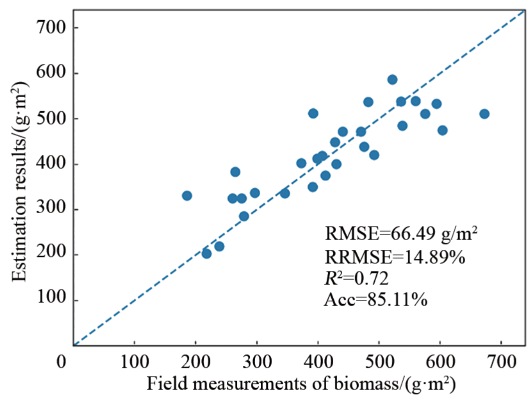
(a) Wheat biomass estimation result without feature selection



(b) Wheat biomass estimation result using linear correlation feature selection method



(c) Wheat biomass estimation result using RFE algorithm

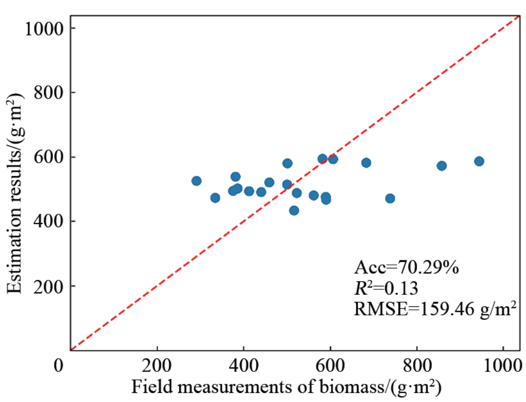


(d) Wheat biomass estimation result using proposed feature selection method

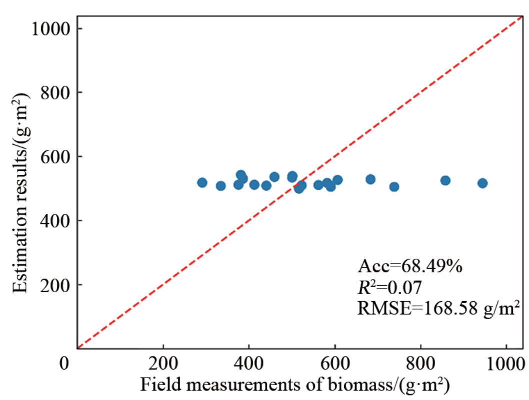
Fig. 8 Biomass estimation result of wheat with different feature selection strategies

As can be seen from Fig. 8, comparing with the estimation results without feature selection; three feature selection algorithms all have improved the estimation accuracy. Among which the processing using GA achieve the highest improvement, up to 8.96%. Comparing with linear correlation algorithm and RFE algorithm, the estimation accuracy using GA is increased by 5.77% and 3.90% respectively.

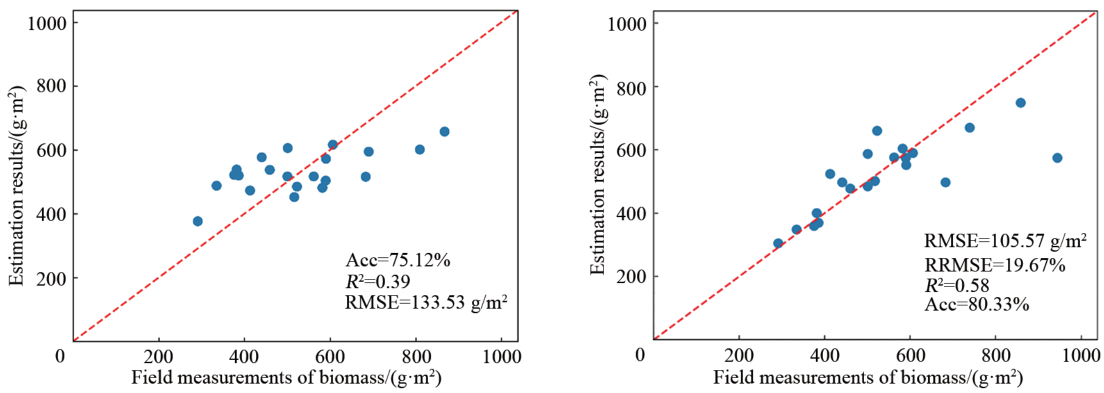
As shown in Fig. 8(d), the method proposed in this paper achieves 85.11% verification accuracy in wheat biomass estimation, and the decision coefficient is 0.72 in general, the measured value of biomass is strongly correlated with the predicted value, which shows that this method can achieve relatively good crop estimation result.



(a) Rape biomass estimation result without feature selection



(b) Rape biomass estimation result using linear correlation feature selection method



(c) Rape biomass estimation result using RFE algorithm

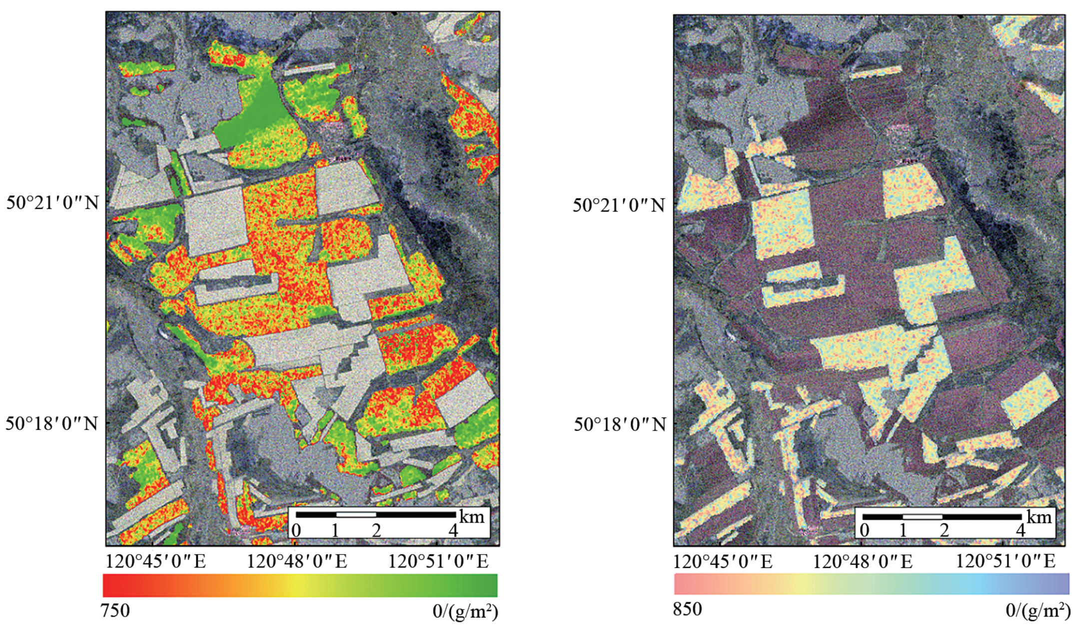
(d) Rape biomass estimation result using proposed feature selection method

Fig. 9 Biomass estimation result of rape with different feature selection strategies

As shown in Fig. 9, comparing with the estimation results without feature selection, the estimation accuracy based on linear correlation algorithm is reduced by 1.8%. Which reflects that this kind of algorithm cannot guarantee the performance of the selected feature combination in the model estimation. The estimation accuracy of RFE algorithm and GA is increased by 4.83% and 10.04% respectively. Comparing with linear correlation algorithm and RFE algorithm, the accuracy of GA is increased by 11.84% and 5.21% respectively, which demonstrate that the feature selection method of GA adopted in this paper can effectively improve the accuracy of crop biomass estimation both rape and wheat.

As can be seen from Fig. 9(d), the verification accuracy of 80.33% is achieved in rape biomass estimation, and the decision coefficient is 0.58, which shows a relatively strong correlation between the measured value of rape biomass and the predicted value based on proposed method. However, compared with wheat biomass, rape biomass estimation accuracy is lower. It may be that the biomass level of rape in the experimental area is generally higher and the difference is relatively small, so that the relation between the features and biomass is not obvious, resulting in relatively low accuracy of the estimation.

More detailed estimation results based on proposed method are shown in Fig. 10.



(a) Mapping of wheat biomass result

(b) Mapping of rape biomass result

Fig. 10 Crop biomass mapping results based on GA and SVR shows that the wheat in the area are at different biomass levels. While the estimation result of rape indicates that most of the rape in the area are at a high biomass level

5 DISCUSSIONS

The method proposed in this paper is not only suitable for estimating the biomass of crops using PolSAR data, but also suitable for estimation modeling of other vegetation parameters using other type remote sensing data. It is worth noticing that, in addition tdiso feature combinations, model parameters are strongly affected the estimation performance. Considering the calculation cost, in this paper, the SVR model parameters in the feature selection process were not optimized. Only the SVR model parameters used in final estimation process are optimized. Therefore, in the future work, we will consider a method that can synthesize the feature selection and parameter optimization processing.

6 CONCLUSIONS

Aiming at the problem of high dimensional feature redundancy in the estimation of crop biomass using polarized SAR. In this paper, a method for estimating crop biomass based on the feature selection of GA is presented. The method is verified by the full polarization SAR data of GF-3. The main conclusions of this paper are as follows:

(1) Based on the method proposed in this paper, the estimation of crop biomass in the Shangkuli farm in Inner Mongolia is achieved. The accuracy of wheat biomass estimation reached 85.11%, R^2 is 0.71, rape biomass estimation accuracy is 80.33%, and R^2 is 0.58, both types of crops have achieved high estimation accuracy, which shows that the method proposed in this paper is suitable for crop biomass estimation based on polarized SAR data.

(2) Comparing with the estimation results of wheat and rape without feature selection, the improvement of GA method in accuracy is 8.96% and 10.04% respectively, which shows that the feature selection method based on GA can effectively improve the accuracy of non-parametric model in crop biomass estimation.

(3) Comparing the estimation results using the feature selection based on linear coefficient, the estimation accuracy of GA feature selection method is higher than 5.77% (wheat) and 11.84% (rapeseed) respectively. Comparing the estimation results of RFE algorithm, and the estimation accuracy proposed in this paper is higher than 3.90% (wheat) and 5.21% (rape) respectively. This shows that the GA feature selection method adopted in this paper can improve the accuracy of crop biomass estimation more effectively than other commonly used feature selection methods.

REFERENCES

Arii M, Zyl J J V, Kim Y. Adaptive Model-Based Decomposition of Polarimetric SAR Covariance Matrices [J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(3):1104-1113.

Ataollah H G, Sahebi M R, Mansourian A. Polarimetric SAR feature selection using a genetic algorithm [J]. Canadian Journal of Remote Sensing, 2011, 37(1):10.

Chandrashekar G, Sahin F. A survey on feature selection methods [J]. Computers & Electrical Engineering, 2014, 40(1):16-28.

Cloude, PottierS. R. and E., An entropy based classication scheme for land applications of polarimetric SAR. IEEE Transactions on Geoscience and Remote Sensing, 1997.

De Siqueira F R, Schwartz W R, Pedrini H. Multi-scale gray level cooccurrence matrices for texture description[J]. Neurocomputing, 2013, 120: 336-345.

Freeman A, Durden S L. A three-component scattering model for polarimetric SAR data [J]. IEEE Transactions on Geoscience & Remote Sensing, 1998, 36(3):963-973.

Guyon I , Weston J , Barnhill S , et al. Gene Selection for Cancer Classification using Support Vector Machines[J]. Machine Learning, 2002, 46(1-3):389-422.

Hosseini M, Mcnairn H, Merzouki A, et al. Estimation of Leaf Area Index (LAI) in corn and soybeans using multi-polarization C- and L-band radar data[J]. Remote Sensing of Environment, 2015, 170: 77-89.

Hosseini M, Mcnairn H. Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields [J]. International Journal of Applied Earth Observation & Geoinformation, 2017, 58:50-64.

Luneburg E. Foundations of the Mathematical Theory of Polarimetry [R]/Final Report PhaseI. N00014-00-M-0152, ConsultantsEML, 2001

Maryam S, Sahebi M R, Yasser M. Improving the Accuracy of Urban Land Cover Classification Using Radarsat-2 PolSAR Data [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(4):1394-1401.

Mcnairn H, Brisco B. The application of C-band polarimetric SAR for agriculture: a review [J]. Canadian Journal of Remote Sensing, 2004, 30(3):525-542.

Mattia F, Toan T L, Picard G, et al. Multitemporal C-band radar measurements on wheat fields [J]. IEEE Transactions on Geoscience & Remote Sensing, 2003, 41(7):1551-1560.

Pedregosa F, Gramfort A, Michel V, et al. Scikit-learn: Machine Learning in Python[J]. Journal of Machine Learning Research, 2013, 12 (10):2825-2830.

Sasan V , Javad S , Kamran A , et al. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran) [J]. Remote Sensing, 2018, 10(2):172-.

Tao L, Li J, Jiang J, et al. Leaf Area Index Inversion of Winter Wheat Using RADARSAT-2 Data and Modified Water-cloud Model [J]. Journal of Triticeae Crops, 2016, 36(2):236-242.

Xiaoxiao L, Liang W, Shenghua X U, et al. A remote sensing feature selection method of forest biomass estimation based on RF-RFE [J]. Science of Surveying and Mapping, 2017.

Yamada Y. Preliminary study on the radar vegetation index (RVI) application to actual paddy fields by ALOS/PALSAR full-polarimetry SAR data. [J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015: 129-131.

Yang H , Xie L , Chen E , et al. Biomass estimation of oilseed rape using simulated compact polarimtric SAR imagery[C]// IGARSS 2016 - 2016 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2016.

Zhang W, Li Z, Chen E, et al. Compact Polarimetric Response of Rape (Brassica napus L.) at C-Band: Analysis and Growth Parameters Inversion [J]. Remote Sensing, 2017, 9(6):591.

Zhao L, Chen E, Li Z, et al. Three-Step Semi-Empirical Radiometric Terrain Correction Approach for PolSAR Data Applied to Forested Areas[J], Remote Sensing, 2017, 9(3):269.