

Research on apple orchard classification and yield estimation model based on GF-1 and Sentinel-2

Yu Liu¹, Xiaoping Wang^{1,2*}, Jiaxin Qian¹

¹ Xi'an University of Science and Technology, Xi'an, 710054 China

² Xi'an Dadi Surveying And Mapping Co., Ltd, Xi'an, 710054 China

Abstract: China is a big apple planting country and attaches great importance to the development of apple industry in agricultural economy. There are many mountainous areas in Shaanxi Province, which has obvious geographical advantages and is one of the important areas for apple production in my country. A quick and effective forecast of the apple output in Shaanxi Province can not only strengthen the management of apple planting and production, improve the varieties of apple production, and improve the quality of apple production, but also provide technical support for regional agricultural departments to expand the apple market and improve the base construction. It is of great significance to promote the rapid development of my country's apple planting industry. In this study, Luochuan County, Yana's City, Shaanxi Province was used as the research area, using GF-1 and Sentinel-2 multispectral remote sensing images and their vegetation indices from 2013 to 2019, and using RF to extract orchards in the research area. Secondly, combining the classification results with rainfall, temperature, sunshine hours, air pressure, humidity, wind speed, drought indicators and remote sensing vegetation index, using RFR and SVR methods, establish a comprehensive production estimation model suitable for Luochuan County apples, and compare different types Model accuracy. The main conclusions are drawn through the research: Using RF classification method can effectively extract the luochuan orchard distribution and high precision, based on RFR and SVR method combined with meteorological factor, the drought index and remote sensing vegetation index to establish basic quite, crop yield estimation model precision machine learning regression algorithm for subsequent apple luochuan orchard management, and provide strong decision basis for the development of apple industry.

1 Introduction

Agriculture is our country's primary industry, an important industrial sector in the national economy, and an important foundation for social stability and national economic development. Cash crops are an important part of agriculture. Our country attaches great importance to the development of the apple industry in the agricultural economy and is a big country in apple cultivation. There are many mountainous areas in Shaanxi Province, with obvious geographical advantages. It is one of the important areas for apple production in my country. Shaanxi Province has a strong industrial foundation. The area of apple planting and the output of apples rank first in the country, and it has become the main way for farmers to increase their income^[1]. Therefore, remote sensing monitoring of apple production areas in Shaanxi Province is carried out to accurately grasp the changing trend and distribution of apple orchards. Long-term continuous remote sensing monitoring can timely reflect the growth of apples, which can further improve the quality of apples and strengthen Apple nutrition diagnosis and increase of apple production are of great significance to the rapid

development of my country's apple planting industry^[2].

In recent years, remote sensing technology has been widely used due to its fast data acquisition speed, short observation period and wide observation range^[3-4], especially in the fields of meteorology, agriculture and forestry^[5-6]. The application of remote sensing technology to orchard classification research has a certain foundation, including the combination of orchard texture characteristics, spectral characteristics, vegetation index, multi-source remote sensing images, DEM, multiple classification methods, measured data, to improve orchard information extraction Accuracy, with more significant results^[7-13]. Remote sensing technology has the advantages of wide range, fast speed, short return visit period, and large amount of stored information, so that it can realize the needs of macro, fast, accurate and dynamic monitoring in many fields, thus also promoting the rapid development of crop yield estimation research^[14]. The model of crop yield estimation is established by using the remote sensing vegetation index of crops and the growing period of crops, which improves the accuracy of crop yield estimation and achieves very high results^[15-19]. Therefore, this study uses Luochuan County in Shaanxi Province as the research area to achieve timely and accurate crop yield

*Corresponding author : Xiaoping Wang Email: 542781966@qq.com

prediction. This study will explore the indication effect of vegetation index obtained from remote sensing data on orchard yield, use Random Forest algorithm to accurately extract apple orchards in Luochuan County from 2013 to 2019, and use two regression methods were used to establish an orchard yield estimation model involving remote sensing vegetation index and meteorological factors, drought index and greenhouse gas, and to predict the apple yield in the study area.

1.1 Methods

The purpose of this research is to explore the relationship between various spectral characteristics and their related vegetation index and orchard yield, reveal the influence of meteorological factors, greenhouse gases and drought indicators on orchard yield, and establish consideration of relevant spectral characteristics, meteorological factors, and greenhouse gases. And a comprehensive yield estimation model of drought indicators.

The basic methods in the research are as follows: (1) Calculate multiple vegetation indices and use RF methods to classify apple orchards in Luochuan County from 2013 to 2019. (2) Through Person correlation analysis, analyze the period during the apple growth period that is most closely related to the yield and the vegetation index, use different machine learning regression algorithms to explore the quantitative relationship between the vegetation index and the apple yield, and combine meteorological factors and drought indicators Wait for the establishment of an apple production estimation model. The following describes the remote sensing image classification algorithm and production estimation modeling method.

Random forest algorithm: Random forest algorithm can predict the effect of up to thousands of explanatory variables very well, and is known as one of the best machine learning algorithms [20-21]. Experiments show that the classification OOB error gradually converges and stabilizes as the number of decision trees increases ($N \geq 100$). Therefore, N is set to 100, and the number of features is set to arithmetic square root (Sqrt).

Support Vector algorithm: Support Vector Machine is an algorithm used for classification (SVC), and it can also be used for regression, namely support vector regression (SVR) [22]. SVM does not use the traditional derivation process, which simplifies the usual classification and regression problems. A small number of support vectors determine the final decision function of the SVM, and the complexity of the calculation depends on the support vector, rather than the entire sample space, which can avoid the "dimensionality disaster". This method is not only simple in algorithm, but also has better robustness. SVM itself is proposed for two classification problems, and SVR is an important application branch in SVM. The difference between SVR and SVC is that the sample points of SVR ultimately have only one category. The optimal hyperplane it seeks does not separate two or more sample points like SVM, but minimizes the total deviation of all sample points from the hyperplane.

Regression analysis is a statistical analysis method to

determine the quantitative relationship between two or more variables. This study will use this method to establish a remote sensing yield estimation model for apple production, a remote sensing combined with meteorological factor yield estimation model, and a comprehensive yield estimation model combining remote sensing, meteorological factors, and drought indicators.

1.2 Technical process

According to the above research content, the technical route of this research is shown in Figure 1:

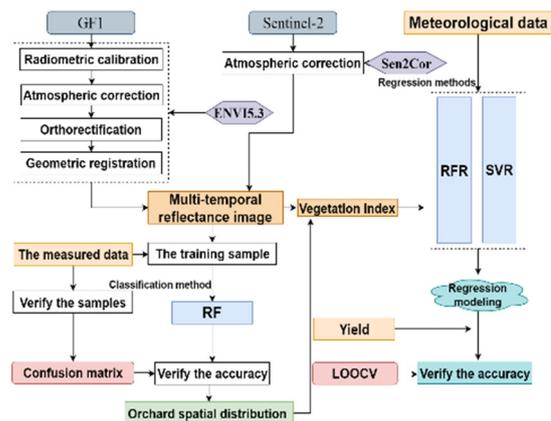


Figure 1 Technical process

2 Study area and data

2.1 Overview of study area

Luochuan County is located in the central part of Shaanxi Province and the south of Yan'an City. As shown in Figure 2, the terrain of the study area is higher in the northeast and lower in the southwest, with an altitude of about 625~1524m and an average altitude of about 1100m. The soil layer in the study area is about 80-200m thick, the annual average temperature is 9.2°C, the frost-free period is 167 days, the annual precipitation is 622mm, the sunshine is sufficient, the rain and heat are in the same season, and the temperature difference between day and night is large. The study area is densely covered with orchards, stretching for hundreds of miles. Luochuan County belongs to the north temperate continental humid and dry monsoon climate. The climate in the territory is relatively mild, the solar radiation energy is rich, the annual average temperature is 9.2 °C , the annual precipitation is 622mm, the frost-free period is 167 days, the sunshine is sufficient, the temperature difference between day and night is large, and the rain and heat are in the same season.

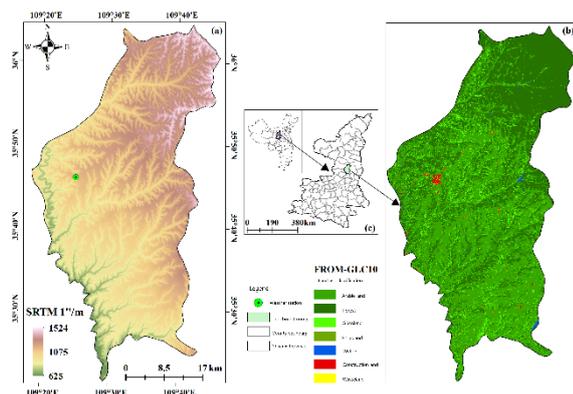


Figure 2 Geographical location of the study area

The data sources of this study mainly include Sentinel-2 remote sensing images, GF-1 remote sensing images, high-resolution Google remote sensing images, the second national land survey Luochuan County land use type data, Luochuan County apple planting area data and ground Investigate sampling data, etc. Because the medium and high resolution optical satellites are affected by factors such as cloud and rain, it is difficult to obtain effective images in a certain month in the long-term sequence. Therefore, based on the actual situation of the study area, the vegetation growth season (April-October) from 2013 to 2019 is selected. All cloud-free images in this section have a total of 26 issues.

2.2 Remote sensing image acquisition

The Sentinel-2 L1C level data comes from ESA. It is a product of atmospheric apparent reflectance that has undergone orthorectification and geometric precision correction. Atmospheric correction is required to obtain the true reflectance of the earth's surface. The Sentinel-2A satellite was successfully launched in 2015, but some domestic images were only available in 2016, so the Sentinel-2 data involved in this study is from 2016 to 2019. The GF-1 satellite was successfully launched on April 26, 2013. It is the first satellite of a major national science and technology project for China's high-resolution earth observation system. The remote sensing image data involved in this research is from 2013 to 2019, of which the data from 2013 to 2015 is GF-1, and the data from 2016 to 2019 is Sentinel-2. Therefore, classification can not only be guaranteed on the basis of the two types of remote sensing images. The results and the accuracy of production estimation can also compare the advantages and disadvantages of the two kinds of experimental data, and the accuracy differences of different data in classification and production estimation.

2.3 Remote sensing image preprocessing

Sentinel-2 preprocessing: first use Sen2cor 2.8.0 processing model for Sentinel-2 L1C level data to generate L2A level data, and call the Sentinel-2 toolbox module in SNAP 7.0 (Sentinel Application Platform) software to proceed with the nearest neighbor method Re-sampling, cropping and other operations are performed,

and finally a multi-temporal surface reflectance image with a spatial resolution of 10 m in the study area is obtained. Second, calculate the vegetation index and texture characteristics, and further classify and estimate the yield of the orchard.

GF-1 data preprocessing: In this study, GF-1 L1A-level data is selected as the data source for orchard classification in Luochuan County. The data comes from the China Resources Satellite Application Center. The data at this level has been corrected for relative radiation. First use the RPC file that comes with the image to perform orthorectification, and then perform radiometric correction (including radiometric calibration and atmospheric correction) to obtain the true surface reflectance, and perform geometry based on the 15m spatial resolution Landsat8 panchromatic band image of the same period Fine correction, and finally get the true reflectance image of the ground surface with 16m spatial resolution, and further classification and yield estimation of orchard.

2.4 Meteorological data of the study area

Select the daily value data of rainfall (mm), air pressure (hpa), temperature (°C), relative humidity (%) and wind speed (m/s) from the meteorological station of Luochuan County (site code: 53942), the data comes from National Meteorology Scientific data center. Among them, the time span of rainfall, air pressure, air temperature, relative humidity and wind speed data is 2013/01/01-2019/12/31.

2.5 Apple orchard planting area and production data

The data on the planting area and output of orchards obtained in this research are all from the agricultural statistics database of Shaanxi Agriculture Network (Table 1).

Table 1 Apple Planting Area and Yield in Luochuan County

Years	Planting area(hm ²)	Yield (t)
2013	33858	763500
2014	33889	793000
2015	34275	801000
2016	34274.533	864700
2017	35088.867	888494
2018	35593.467	553000
2019	390000	930000

2.6 Field survey data collection

The sampling data of the ground survey are mainly field sampling points. Field surveys of orchards in the study area were conducted from July 20 to 23, 2020. A total of 79 orchard plots, 50 arable land and 48 forest lands were measured. The spatial distribution is shown in the figure. 3 shown. During the survey of plot samples, Zhonghaida handheld GPS is used for precise positioning, with an accuracy of about 3m.

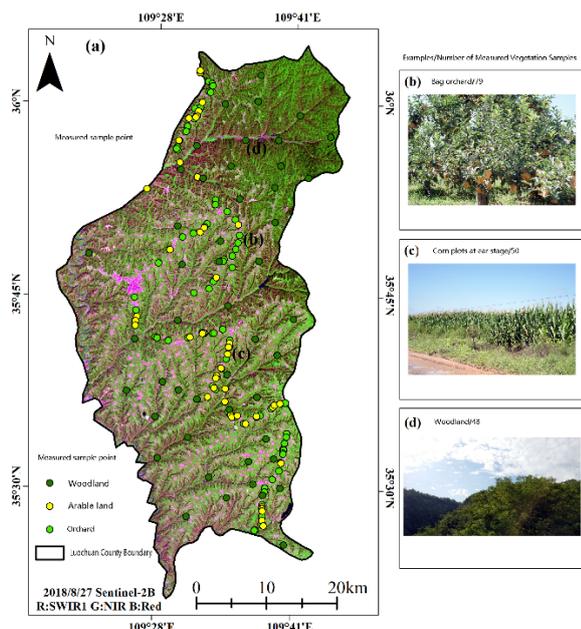


Figure 3 Spatial distribution of measured vegetation plots

Select the NDVI data of the 1000m spatial resolution MOD13A3 product (monthly maximum synthesis) and the land surface temperature data (LST) of the 1000m spatial resolution MOD11A2 product during the apple phenology period from 2013 to 2019 as the use of MODIS data to monitor Los Angeles Data source of monthly drought in Chuan County.

Select the GOSAT (Greenhouse Gases Observing Satellite) FTS L3 product (V2.90), which is based on the CO₂ and CH₄ concentration data and the global 2.5° spatial resolution monthly average concentration distribution data obtained through the Kriging difference, from JAXA (Japan Aerospace Exploration Agency).

3 Research on Orchard Classification Based on Satellite Remote Sensing

3.1 Orchard classification based on RF

This study uses the RF to classify the remote sensing images of Luochuan County from 2013 to 2019. First calculate the corresponding vegetation index, combine with the July field survey vector data, use the annual remote sensing image as the base map, use the Google Earth high-resolution image from September to November each year, and the second national land survey Luochuan County land use type data As a supplement, samples of orchards, arable land, woodland, waters, construction land, and bare land are selected. The JM (Jeffries-Matusita distance) distance is selected to qualitatively describe the separability between the selected samples, and to ensure that the JM distance between each two kinds of ground object samples is greater than 1.9 [23]. Calculate the overall

accuracy (OA) and Kappa coefficient of the classification results of different classification feature groups through the verification sample and evaluate the accuracy.

3.2 Orchard classification results

From the 2013-2019 orchard classification map (Figure 4), it can be seen that orchards are concentrated in the central and north-central regions of Luochuan County, with a small number in the northern region. The south is dominated by arable land, with relatively few orchards; some orchards and arable land are staggered, the arable land in the central and northern parts is mainly distributed in low-lying but relatively flat valley areas, and the arable land in the south is mainly distributed in flat open land; woodland is in the whole The counties are all distributed, concentrated in the mountainous areas with higher elevations in the north, while the woodlands in the central, western and southern parts are mainly distributed in relatively low-lying ridge areas. The construction land is mainly distributed in the central and western villages and towns, and the terrain is relatively high and flat; the waters are mainly three reservoirs and the river in the west. In 2013-2019, the classification accuracy exceeded 85%, and the Kappa coefficient exceeded 0.80, meeting the extraction needs.

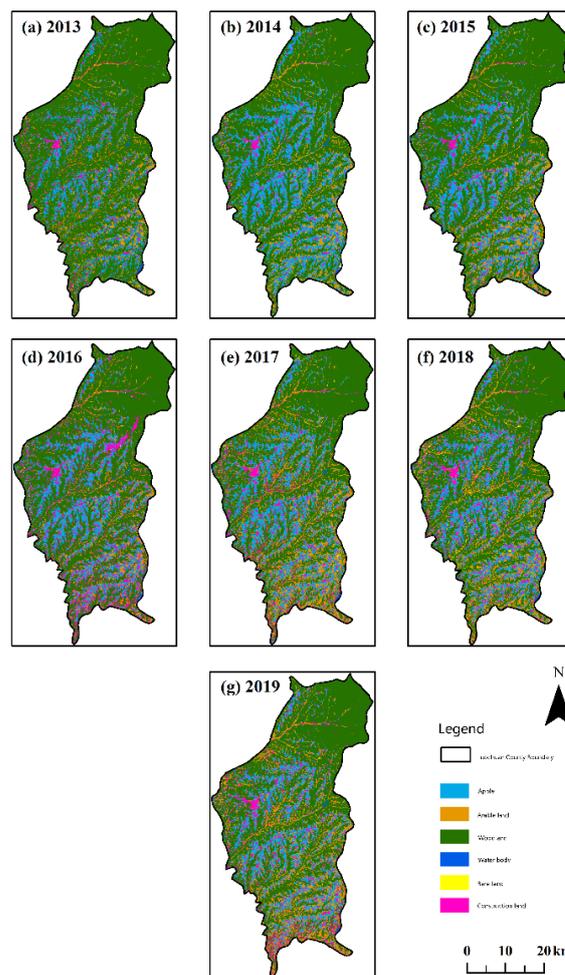


Figure 4 Classification of feature types in Luochuan County from 2013 to 2019

4 Establishment of orchard yield estimation model

Machine learning regression algorithms have powerful fitting capabilities and can better establish the relationship between different data sources and crop yields. Therefore, this chapter uses two machine learning algorithms (SVR and RFR) combined with meteorological factors, drought indicators and remote sensing vegetation index to establish crop yield estimation models, and compare the yield estimation results obtained by different regression methods.

The remote sensing images involved in this research include GF-1 and Sentinel-2, including NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), and NDWI (Normalized Water Index). It involves blue, green, red and near-infrared bands. TVDI coupled vegetation index and surface temperature are common indexes used to characterize the surface dry and wet conditions. Studies have shown that drought is an important factor affecting crop yields. Therefore, TVDI is also used as one of the input features of the production estimation model. The feature groups used in machine learning regression are shown in Table 2. According to correlation analysis, judge the important influencing factors that affect apple yield, including the influence of different influencing factors on apple growth in different phenological periods. Secondly, it is preferable to pass the characteristics of significance test and use RFR and SVR to estimate orchard yield.

Table2 Feature variables used in machine learning regression

Group	Feature description
(a)	Meteorological factors: Average annual precipitation(a1), sunshine hours(b1), annual average temperature(c1), air pressure(d1), wind speed(e1), humidity(f1)
(b)	Spectral characteristics: NDVI, EVI, NDWI
(c)	Drought indicator: TVDI
(d)	Greenhouse Gas Factor: CO ₂ , CH ₄

It can be seen from Table 3 and Table 2 that from 2013 to 2019, there was a significant positive correlation between the rainfall and yield during the mature period. The sunshine hours during the germination period were significantly positively correlated with the yield, and the sunshine hours during the maturity period were significantly negatively correlated with the yield. The average young fruit period Humidity has a significant negative correlation with yield, and the average humidity during maturity has a very significant positive correlation with yield. From 2013 to 2019, the mean value of TVDI

during the budding period was extremely negatively correlated with yield. From 2013 to 2019, the mean value, NDVI during the germination period and the young fruit period were extremely significantly positively correlated with yield. The NDVI during the flowering, swelling, and mature periods were significantly correlated with yield, and NDWI during the germination period was significantly negatively correlated with yield. Except for 2018, greenhouse gases and yields are extremely positively correlated, indicating that greenhouse gases are beneficial to the photosynthesis and usefulness of crops, and have a certain promotion effect on yield, but they are susceptible to extreme weather.

The characteristics that have passed the significance test for the correlation with the output are selected as the modeling characteristics of the production estimation, a total of 13 characteristics. Since the remote sensing images of Luochuan County's orchards meet the selection requirements only from 2013 to 2019, only the production data from 2013 to 2019 are used in the production estimation modeling. In order to make full use of the small sample data, a leave-one-out cross-validation method is adopted. Carry out the establishment of the orchard yield estimation model and verify the accuracy.

In this study, for RFR, the number of decision trees is set to 500, the minimum number of samples required at leaf nodes is set to 1, and the pattern of the maximum number of features is set to SQRT. For SVR, linear kernel function and Gaussian kernel function are used, and hyperparameters are automatically optimized through two-fold cross-validation and FitSVM (the root mean square error is the smallest) and the apple output is further inverted.

From Table 4 and Figures 5 and 6, it can be seen that for the support vector regression estimation, the accuracy of the estimated output in 2014 was the lowest, the absolute error and the relative error were the highest, which were 41038.62 and 5.18, respectively, and the accuracy of the estimated output in 2013 and 2015 was the highest. , The absolute error and relative error are the smallest, 168.50, 80 and 0.02, 0.01 respectively, followed by the higher estimated output in 2017, the absolute error and relative error are 541.53 and 0.1 respectively. For random forest regression estimation, the accuracy of the estimated output in 2018 is the lowest, with the highest absolute and relative errors, 98527.35t and 17.82%, respectively. The estimated output in 2015 has the highest accuracy, with the lowest absolute and relative errors, 7849.50. t and -0.98%, followed by the higher estimated output in 2013, with absolute and relative errors of 11204.91t and -1.47%, respectively.

Table 3 Correlation of different characteristics on apple yield

index	Characteristic variable	Mean	Germination stage	Flower stage	Young fruit stage	Expand stage	Aging stage
Meteorological factors	a1	0.020	-0.420	-0.032	-0.256	0.003	.856*
	b1	0.393	-0.010	-0.439	0.265	-0.003	0.060
	c1	-0.657	0.841*	-0.310	0.142	-0.257	-.873*
	d1	-0.009	0.152	-0.416	-.804*	-0.226	0.911**

	e1	-0.274	-0.445	-0.150	-0.274	-0.269	-0.333
	f1	0.015	-0.016	0.116	0.043	0.575	-0.664
Drought factor	TVDI	-0.573	-0.880**	-0.692	-0.477	-0.492	-0.503
Greenhouse Gas Factor	CO2	0.138	0.028	0.451	0.269	0.163	0.165
	CH4	0.243	0.035	0.246	0.529	-0.106	-0.009
Vegetation Index	NDVI	0.915**	0.895**	0.829*	0.908**	0.766*	0.851*
	EVI	0.561	0.433	0.514	0.476	0.464	0.483
	NDWI	0.156	-0.865*	0.038	0.298	0.388	0.101

Table 4 2013-2019 production estimation model predicts production

years	Statistical output	SVR/t	RFR/t
2013	763500	763331.49	774704.90
2014	793000	834038.62	811063.90
2015	801000	800920	793150.50
2016	864700	894977.16	816543.73
2017	888494	553541.53	863159.68
2018	553000	873400.05	651527.34
2019	930000	927263.09	876754.25

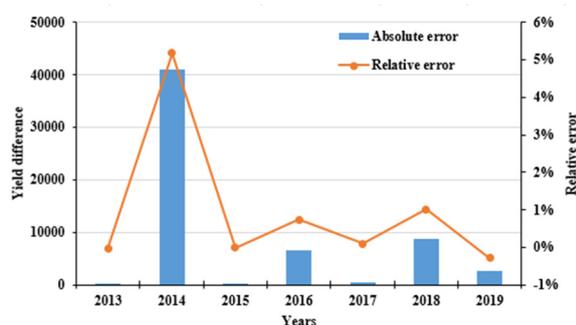


Figure 5 Yield estimation accuracy based on preferred features (SVR)

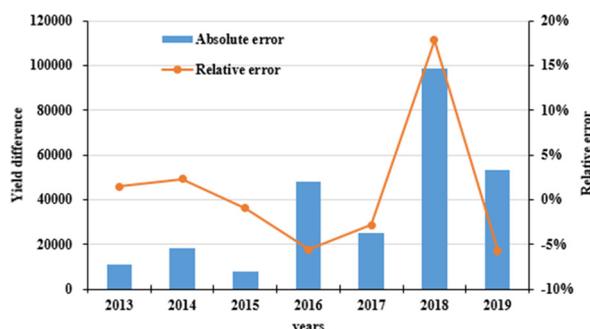


Figure 6 Yield estimation accuracy based on preferred features (RFR)

5 Conclusion

In this study, using multi-temporal Sentinel-2 and GF-1 images, combined with actual survey data, the Random Forest method was used to accurately extract orchards in Luochuan County. Combining the relevant spectral characteristics and its vegetation index, meteorological factors and drought indicators, the analysis shows that the characteristics with higher correlation are obtained. Random Forest regression and support vector regression methods are used to establish the orchard yield estimation model and evaluate the accuracy. The main conclusions

are as follows: Combining meteorological factors, remote sensing vegetation index factors and drought index factors to estimate yield can achieve higher yield estimation accuracy. The accuracy of yield estimation models established based on Random Forest Regression and Support Vector Regression methods is basically equivalent, and the machine learning regression algorithm can provide a strong decision-making basis for the follow-up apple management of Luochuan orchards and the development of the apple industry.

References

1. Sun Z J. Shaanxi has become the first apple province in China [J]. China fruit industry information,2008(10):34.
2. Xin Q R. Research on the information extraction of Apple Garden in mountainous area based on multi-temporal high-resolution image [D]. Shandong University of Technology, 2017.
3. Meng L H. Cotton field yield estimation model based on time series remote sensing image [D]. Northeast Agricultural University, 2018.
4. Li S J ,Sun Y N ,Li M, Chen Z W. Advances in crop yield estimation by remote sensing at home and abroad [J]. The world's agricultural,2013(05):125-127+131+156.
5. Wang D P, Wang Z L, Li D Y. Information Recognition of SPOT-5 Image Orchard Based on NDVI Texture in Hilly Area of Shandong Province [J]. Science of Surveying and Mapping,2007(01):126-127+121+165.
6. Zhang A D, Peng D M ,Li D Y ,Wang Z L, Wang D P. Research on Extraction Technology of Orchard Spatial Information Based on TM Image [J]. Science of Surveying and Mapping,2007(05):121-123+205.
7. Luo W, Kuang R Y. Extraction of Orchard Information from Environmental Satellite Images in

- the Headwaters of Dongjiang River [J]. *Science of Surveying and Mapping*,2014,39(08):135-139.
8. Liu J Q. Research on information extraction of apple orchard in fufeng county based on Landsat8 remote sensing image [D]. Northwest Agriculture & Forestry University, 2015.
 9. Guo B, Wu Y Y, Zhou J Y, Wang Z Z ,Shi Q Y ,Tian Y. Simulation of the apple yield based on unmanned aerial vehicle image data ——Taking Luochuan county as an example [J]. *Journal of Xi'an University of Science and Technology*,2017,37(03):450-454.
 10. Dong F. Research on multi-source data based remote sensing extraction technology of Apple Garden information in hilly Area [D]. Shandong Agricultural University, 2012.
 11. Xu H Z Y, Liu C, Qi S H, Zhao G S. Remote sensing information extraction of citrus orchard in southern Jiangxi based on random forest algorithm [J]. *Journal of Jiangxi Normal University (Natural Science)*,2018,42(04):434-440.
 12. Zhang L, Gong Z N, Wang Q W, et al. Sentinel-2 Image Multi-feature Optimized Information extraction of Yellow River Delta Wetland [J]. *Journal of remote sensing*, 2019, 23(02): 313-326.
 13. Zhong Y . Information Extraction of Citrus Orchard Based on Improved Support Vector Machine and Spatial Distribution Characteristic Analysis [D]. Jiangxi University of Science and Technology,2019.
 14. Su T ,Wang P X ,Liu X K ,Yang B. Spring maize yield estimation based on entropy combined prediction and multi-temporal remote sensing [J]. *6/5000 Transactions of the Chinese Society for Agricultural Machinery*,2011,42(01):186-192.
 15. Franch B, Vermote E F, Becker-Reshef I, et al. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information[J]. *Remote Sensing of Environment*2015, 161: 131-148.
 16. Liu W T, Kogan F. Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices[J]. *International Journal of Remote Sensing*, 2002, 23(6):1161-1179.
 17. Johnson D M. An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States [J]. *Remote Sensing of Environment*, 2014, 141: 116-128.
 18. Moriondo M, Maselli F, Bindi M. A simple model of regional wheat yield based on NDVI data [J]. *European Journal of Agronomy*, 2007, 26(3): 266-274.
 19. Nuarsa I W, Nishio F, Hongo C, et al. Using variance analysis of multitemporal MODIS images for rice field mapping in Bali Province , Indonesia [J]. *International Journal of Remote Sensing*, 2012, 33(17): 5402-5417.
 20. Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24-31.
 21. Gislason P O, Benediktsson J A, Sveinsson J R. Random forest classification of multisource remote sensing and geographic data[C]. *2004 IEEE International Geoscience and Remote Sensing Symposium*, 2004, 2: 1049-1052.
 22. B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson, "Shrinking the tube: a new support vector regression algorithm," presented at the Proceedings of the 1998 conference on Advances in neural information processing systems II, 1999.
 23. Wang C,Liu Z,Yan C. An experimental study on imaging spectrometer data feature selection and wheat type identification[J]. *Journal of remote sensing*, 2006, 10(2): 24