Lin J *et al.* (2024)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 52, Issue 4, Article number 14099
DOI:10.15835/nbha52414099
Research Article

Modeling of spectral monitoring of cotton aphid population based on feature bands

Jiao LIN[1,2], Zi-Wei BAI[1,2], Jia-Qi ZHANG[1,2], Qiang HU[1,2], Nan CAO[1,2], Su-Mei WAN[1,2]*

[1]*Tarim University, College of Agronomy, Alar 843300, Xinjiang, China; linjiao@taru.edu.cn (J.L.); 10757232006@stumail.taru.edu.cn (Z-W.B); 2587074907@qq.com (J-Q.Z.); 610359291@qq.com (Q.H.); caonan0209@163.com (N.C.); wansumei510@163.com (S-M.W.) (*corresponding author)*

[2]*Key Laboratory of Genetic Improvement and Efficient Production for Specialty Crops in Arid Southern Xinjiang of Xinjiang Corps, Alar 843300, Xinjiang, China*

Abstract

Cotton aphid is the most common and harmful insect in the process of cotton growth and development, and it is the most serious pest that restricts the high yield and high quality of cotton in China and even in the world. The field-scale experiments were conducted for two years to estimate the population of cotton aphids model on cotton leaves using remote sensing. Hyperspectral data from single leaves and aphid populations were obtained at different growth stages of various cotton varieties. Preprocessing techniques including savitzky-golay (SG), first derivative (FD), second derivative (SD), logarithm transformation (LG), and reciprocal transformation (RG) were applied to the original spectral data. The successive projections algorithm (SPA) combined with the Pearson correlation coefficient (Pearson) method were used to construct three types of hyperspectral monitoring models: linear regression (LR), Extreme gradient boosting (XGBoost), and partial least squares regression (PLSR). Results indicated that LG significantly improved model accuracy, while SPA effectively reduced the number of bands required for analysis. Among the three models constructed using selected feature bands, the XGBoost model outperforms LR and PLSR models in terms of prediction accuracy. LG-Person-XGBoost identified ten feature bands with a coefficient of determination (R^2), root mean square error (RMSE), and relative percent deviation (RPD) values reaching 0.76, 65.74 heads/leaf, and 0.91 respectively for the modeling set; whereas predicted R^2, RMSE, and RPD values were found to be 0.36, 57.62 heads/leaf, and 0.99 respectively. LG-SPA-XGBoost achieved superior prediction accuracy by selecting eight feature bands with R^2, RMSE, and RPD values reaching 0.87, 46.7 heads/leaf, and 1.43 respectively for the modeling set; whereas predicted R^2, RMSE, and RPD values were found to be 0.77, 42.66 heads/leaf, and 1.27 respectively. This indicates that the hyperspectral remote sensing model can be used to estimate the population of cotton aphids on cotton leaves based on the selection of feature bands providing are reference for the nondestructive monitoring of cotton aphids in cotton fields and offering a significant supplement to the traditional methods of crop pest quantity monitoring.

Keywords: cotton aphid; feature band; machine learning; modeling; population

Received: 03 Sep 2024. Received in revised form: 11 Oct 2024. Accepted: 01 Nov 2024. Published online: 18 Dec 2024.
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

Introduction

Rapid, non-destructive and accurate monitoring methods have important application value for the acquisition and prediction of cotton aphid information. China is the world's largest consumer, producer, and cultivator of cotton (Luo *et al*., 2011). Xinjiang is the primary cotton-growing region in China, boasts the highest quality cotton production, and plays a crucial role in the sustainable development of the local cotton industry (*https://www.stats.gov.cn/sj/zxfb/202312/t20231225_1945745.html*). Pests in cotton fields are significant factors that hinder its healthy growth (Hu *et al*., 2018), thereby restricting the overall progress of China's cotton industry. Among these pests, the cotton aphids stand out as one of the main culprits (Zhang *et al*., 2020), primarily infesting the undersides of leaves and young stems to extract sap. The results in symptoms such as leaf curling and stunted plant growth while severely impeding photosynthesis and nutrient-organ development (Zeng *et al*., 2015). Besides, it can transmit various plant viral diseases and induce mildew formation, ultimately diminishing product quality grades (Lacitignol *et al*., 2005; Gomez, 2006; Zhou *et al*., 2011). In the process of cotton cultivation and management, the acquisition of insect information mainly depends on manual survey, and its lag makes the excessive spraying of pesticides affect cotton yield, quality and environmental pollution.

Temperature fluctuations significantly influence spatial distribution patterns of cotton aphids; with rising temperatures, they tend to migrate towards lower parts of plants while their population numbers decrease considerably. However, over time, there is an overall upward trend observed for these aphids' presence on plants (Feng *et al*., 2015). Currently, acquiring information about cotton aphid infestations relies mainly on manual investigations which are time-consuming and labor-intensive. Moreover, due to localized occurrences of these pests within a field or plot area (point pieces), accurate real-time data cannot be obtained through manual surveys alone. Consequently, this greatly hampers effective prevention and control measures against this pest species (Wang, 2020). Therefore, timely access to comprehensive information regarding outbreaks caused by this pest serves as an essential prerequisite for implementing integrated management strategies (Xu, 2020). Hyperspectral remote sensing monitoring technology offers real-time, rapid, and non-destructive advantages for studying various aspects of crops. This technology has been widely applied in crop identification, acreage extraction, yield estimation, and nutrient detection (Liu *et al*., 2015; Guo *et al*., 2017; Guo *et al*., 2017; Wang *et al*., 2017), thereby greatly promoting the development of agricultural informatization. Previous studies have demonstrated that analyzing the spectral information changes in crop canopy and single leaf can determine whether a crop is infested with insects and assess the degree of infestation (Chen, 2006; Zhang, 2011; Yu, 2015). The 713-831 nm band plays a crucial role in detecting cotton aphid damage, while vegetation indices such as SAVI and NDVI are effective in recognizing cotton aphids (Dilixiati *et al*., 2022). Random forest models constructed based on vegetation indices and machine learning techniques can accurately evaluate tomato late blight under field conditions (Felipe *et al*., 2023). Additionally, applying the standard normal-variable data transformation method to hyperspectral imaging technology enables early diagnosis of crown rot in wheat (Xie *et al*., 2024). Through remote sensing technology, the classification of aphid infestation can be realized. At present, there are few studies using remote sensing technology to directly obtain the number of cotton aphids.

In summary, this study focuses on the number and spectral data of individual cotton aphid leaves. By employing five pretreatment methods to process original spectral data and utilizing correlation coefficient analysis along with successive projections algorithm (SPA) for feature band selection, we developed spectral monitoring models for three types of cotton aphids using linear regression analysis on selected bands as well as extreme gradient boosting tree (XGBoost) and partial least squares regression (PLSR). Three parameters - relative coefficient (R^2), root mean square error (RMSE), and residual predictive deviation (RPD) were employed to evaluate model performance and select the best monitoring model. This approach allows for non-destructive, rapid, accurate monitoring of cotton aphids while providing a scientific theoretical basis and methodology for obtaining crop pest information.

Materials and Methods

Test area overview

The research area is located in Alar City, Xinjiang, at 40°32'31.9"N, 81°16'37.5"E. The area is rich in solar and thermal resources, the average annual solar radiation is 559.4 ~ 612.1 KJ·cm, the sunshine duration is about 2,996 h/year, the sunshine is 100%, the annual accumulated temperature of ≥10 °C is above 4,000 °C, the frost-free period is 180-224 days, and the average annual temperature is 10.8 °C. Rainfall is rare, with an average annual precipitation of 50 mm and an average annual evaporation of 1,976.6-2,558.9 mm, which is a typical warm temperate continental arid desert climate and a typical irrigated agricultural area. The meteorological factors in the study area during 2022-2023 are shown in Figure 1. The rainfall in 2022 is mainly concentrated 98-125 days after seeding, the highest temperature in history is 95 days after seeding, and the lowest temperature is 11 days after seeding. In 2023, the rainfall was mainly concentrated 78-112 days after sowing, the highest temperature was 95 days after sowing, and the lowest temperature was 22 days after sowing.

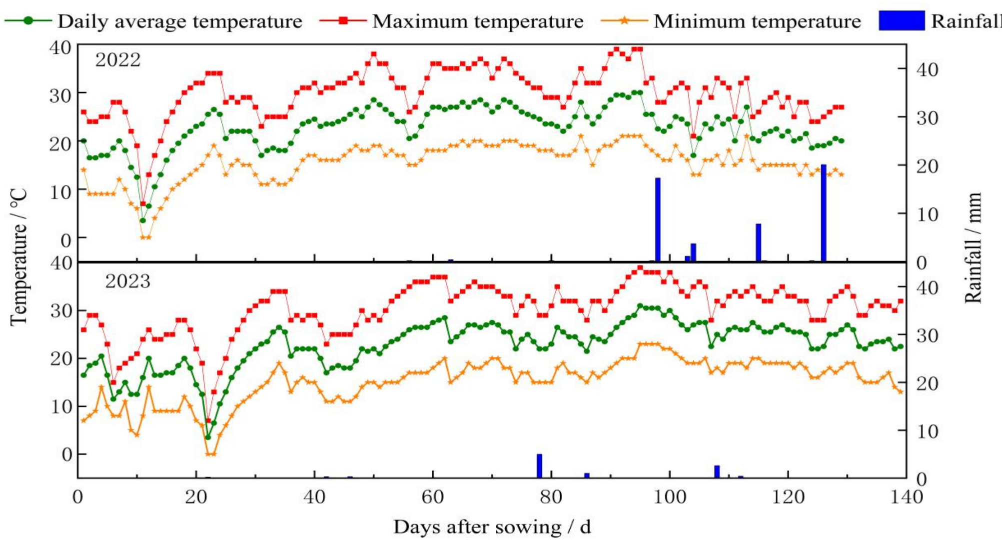

Figure 1. Maximum, minimum, daily average temperature and rainfall from 2022 to 2023

Experimental design

The test varieties were 'Tahe 2', this variety has a growth period of 136 days, with stable growth and good uniformity throughout the entire period. The plant is cylindrical in shape, with type II fruit branches. The stem is thick and sturdy, and the plant has a clear shape. The leaves are medium-sized and have a dark green color. 'Xinluzhong 81', belonging to non-transgenic insect resistant early to mid-maturity conventional cotton varieties, with a growth period of 137 days. Quick emergence, stable growth, strong bell setting ability, and smooth opening of catkins. The plant has a height of 70.3 cm and a relatively loose shape. It has type II fruit branches, with more fuzz on the stem. The leaves are medium-sized and dark green in color, and the bell is oval shaped and moderately large. The third variety 'J8031' characterises through disease resistance, high clothing density, and good quality. The growth period is 134 days, with good emergence, stable growth, no premature aging in the later stage, good uniformity, good bell setting, concentrated shedding, loose plant type, type II fruit branches, and more hairy stems. The sowing dates were April 25, 2022, and April 14, 2023, respectively. At each sampling, five cotton plants with uniform growth were randomly selected from each variety, cotton aphids on each main stem leaf were counted, and spectral data of each main stem leaf were determined. Conduct a pest investigation every 10 days and measure the spectral data of individual leaves until early August. All production management in this experiment was carried out according to the field production. If the weather was bad during the collection of data, the experiment was postponed and records were kept.

Measurement indicators

Cotton aphid population survey

The total number of cotton aphids on each main stem leaf (L0) was counted by removing the main stem leaf (L0) of each plant, with the first main stem leaf at the bottom as the inverted one, and marking the remaining main stem leaves from the bottom to the top and so on as shown in Figure 2.

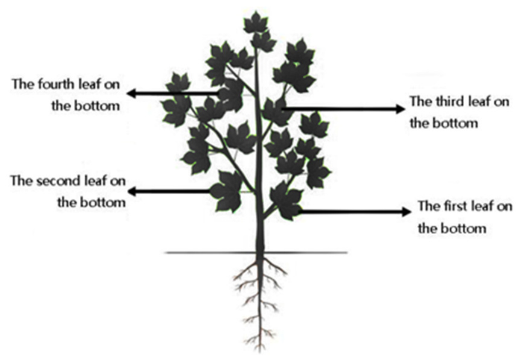

Figure 2. Schematic diagram of cotton plant division

Measurement of hyperspectral data in cotton after aphid infestation

Spectral acquisition was performed by Spectral Evolution's PSR-1100f portable ground object spectrometer. The spectral measurement range was 320-1,100 nm (UV-Vision-near-infrared), the probe was 25° field of view Angle, the resolution was 3.2 nm, and the canopy spectrum averaged more than 10 scans. During spectral data collection, the whiteboard should be corrected every 10-15 min. The whiteboard should be placed flat, and the surface of the whiteboard should not be blocked by shadows during correction.

All main stem leaves were cut from the main stem of the cotton plant respectively, and the spectral reflectance of each main stem leaf was collected. The leaf veins were avoided during data collection, as shown in Figure 3. Each leaf was divided into three parts: upper, left, and right, and one spectral curve was collected for each part. Canopy spectra of cotton plants at each fixed point were collected in the non-destructive sampling area.

Figure 3. Schematic diagram of cotton leaf division

Spectral data preprocessing

In this study, hyperspectral data in the range of 400-1,000 nm were selected for analysis and modelling (Zhao *et al*., 2005; Wang *et al*., 2019), and spectral data were pre-processed by Savitzky-Golay convolutional smoothing (SG), first-order differential (FD), second-order differential (SD), logarithm (LG) and reciprocal (RG) (Gao *et al*., 2011; Lu, 2019; Wang, 2021). The formula is as follows:

$$FD: \quad R'_{\lambda(i)} = \frac{R_{\lambda(i+1)} - R_{\lambda(i-1)}}{2\Delta\lambda} \tag{1}$$

$$SD: R''_{\lambda(i)} = \frac{R'_{\lambda(i+1)} - R'_{\lambda(i-1)}}{2\Delta\lambda} \tag{2}$$

Among them, $R'_{\lambda(i)}$ is the first-order differential value at wavelength i, $R_{\lambda(i+1)}$ and $R_{\lambda(i-1)}$ are the spectral reflectance values at wavelengths i+1 and i-1 respectively, $R''_{\lambda(i)}$ are the second-order differential values at wavelength i, $R'_{\lambda(i+1)}$ and $R'_{\lambda(i-1)}$ are the first-order differential values at wavelengths i+1 and i-1 respectively, representing the difference interval.

$$RG: R(i) = \frac{1}{r(i)} \tag{3}$$

In the formula, i represents the number of bands, r(i) represents the i spectral reflectance, and R(i) represents the i spectral reflectance after reciprocal transformation. Among them, i=400, 401, ... 1,000.

$$LG: R(i) = \ln(r(i)) \tag{4}$$

In the formula, i represents the number of bands, r(i) represents the i spectral reflectance, and R(i) represents the i spectral reflectance after logarithmic transformation. Among them, i=400, 401, ... 1,000.

Spectral signature band screening

In this study, the Pearson correlation coefficient method (Pearson) (Wan, 2023) and the Successive Projections Algorithm (SPA) (Wang, 2011; Niu *et al*., 2021; Cai, 2022; Kong *et al*., 2022) were used to screen the characteristic bands with a strong correlation with cotton aphid populations, in which the Pearson correlation coefficient method selected the 10 bands with the highest correlation for the subsequent modelling analysis.

Model construction and verification

Based on the selected feature bands, the model was constructed by linear regression, XGBoost (Extreme Gradient Boosting), and Partial least squares regression (PLSR), respectively.

Coefficient of determination (R^2), Root Mean Square Error (RMSE), and Relative Percent Deviation were selected. Three indexes (RPD) were used to analyze and verify the accuracy of the model. The closer the R^2 value is to 1, the smaller the RMSE value is, and the closer the RPD is to 2, the more stable and accurate the model is, and the better the model is. Its calculation formula is as follows:

$$\rho = \left| \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{6}$$

$$RPD = \frac{SD}{RMSE} \tag{7}$$

Where i represents the data of the i th sample, xi is the measured value of the i sample data, and $\bar{x}$ is the average value of the measured value; yi is the predicted value of the i sample data, and $\bar{y}$ is the average value of the predicted value.

Results

Data set

The total number of samples, training set, and validation set for determining the number of cotton aphids in a single leaf in this study are shown in Table 1. The total sample set was 732, and the training and validation sets were divided by the assignment method, with 488 samples in the training set, the maximum value was 647 aphids/leaf, the mean value was 71 aphids/leaf, the SD was 108.76, and the coefficient of variation was 1.54; and 244 samples in the validation set, the maximum value was 1,372 aphids/leaf, the mean value was 71 aphids/leaf, the SD was 159.65, and the coefficient of variation was 2.25.

Table 1. Descriptive statistical analysis of the number of cotton aphids on cotton leaves

Sample	Number	Maximum	Minimum	Average	SD	CV/%
Total sample	732	1 372	0	71	127.88	1.81
Modeling set	488	647	0	71	108.76	1.54
Validation set	244	1 372	0	71	159.65	2.25

Correlation analysis

The correlation between the raw spectral reflectance of pre-processed spectral data and the number of cotton aphids is shown in Figure 4. In the range of 400-724 nm, the number of cotton aphids was negatively correlated with RS, SG, and LG, but positively correlated with RG. In the range of 724-1,000 nm, the number of cotton aphids was positively correlated with RS, SG, and LG, and negatively correlated with RG.

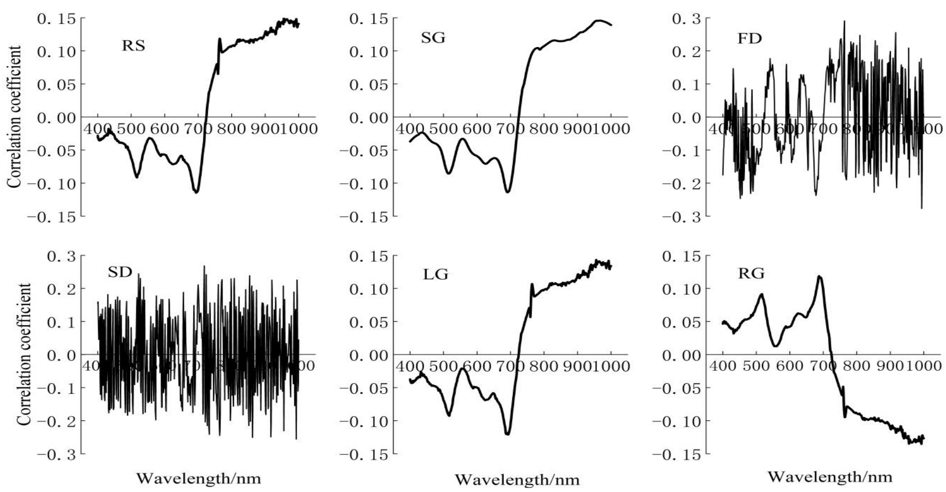

Figure 4. The correlation between spectral transformation and the number of cotton aphids

Spectral feature band screening

Screening feature bands significantly reduces the redundancy of spectral data and the input of model parameters, which plays a crucial role in spectral data model building. In this study, two methods, Pearson and SPA, were selected to screen the feature bands, as shown in Table 2. The feature bands selected by Pearson after RS, SG, and LG treatment are concentrated in the range of 956-967 nm, and most of them are adjacent bands. The feature bands selected by Pearson after FD, SD, and RG treatment were concentrated in the range of 522-808 nm.

Table 2. Feature band screening results

Pretreatment	Band selection method	Feature band numbers	Feature wavelength/nm
RS	Pearson	10	956, 960, 961, 962, 963, 964, 965, 966, 967, 993
	SPA	4	520, 759, 916, 936
SG	Pearson	10	960, 961, 962, 964, 965, 966, 967, 968, 960, 970
	SPA	7	516, 550, 746, 767, 787, 911, 938
FD	Pearson	10	751, 752, 753, 754, 762, 763, 782, 809, 855, 916
	SPA	3	860, 876, 955
SD	Pearson	10	522, 527, 718, 724, 771, 798, 808, 854, 920, 995
	SPA	2	724, 954
LG	Pearson	10	956, 961, 962, 963, 964, 965, 966, 967, 976, 993
	SPA	8	759, 764, 929, 938, 944, 948, 955, 956
RG	Pearson	10	685, 686, 687, 688, 689, 690, 691, 692, 693, 694
	SPA	5	400, 424, 517, 759, 993

Model construction and verification

LR

The spectral data of different pre-screened feature bands were used as input values to construct linear regression (LR) models, and the results are shown in Table 3. After pretreatment by SG, FD, SD, and RG, the model constructed by Pearson was better than that by RS. After pretreatment by SG and LG, the model constructed by using SPA to screen the feature bands has a better effect than RS. Among the 12 linear regression models constructed, the LG+SPA model had the best effect, with R^2 of 0.28, RMSE of 94.04 head/leaf, and RPD of 0.63.

Table 3. Evaluation of cotton aphid population estimation model based on LR

Pretreatment	Band selection method	Training set			Validation set		
		R^2	RMSE heads/leaf	RPD	R^2	RMSE heads/leaf	RPD
RS	Pearson	0.03	109.08	0.19	0.02	155	0.14
	SPA	0.17	101.08	0.45	0.11	147.55	0.31
SG	Pearson	0.11	104.41	0.36	0.05	152.45	0.23
	SPA	0.23	97.47	0.54	0.13	145.33	0.36
FD	Pearson	0.20	98.93	0.51	0.49	54.83	0.93
	SPA	0.10	104.97	0.34	0.08	150.06	0.23
SD	Pearson	0.16	101.63	0.44	0.08	149.59	0.3
	SPA	0.09	105.96	0.31	0.11	148.04	0.24
LG	Pearson	0.03	109.17	0.18	0.02	154.43	0.14
	SPA	0.28	94.04	0.63	0.16	142.99	0.4
RG	Pearson	0.12	104.33	0.36	0.06	151.29	0.27
	SPA	0.21	98.43	0.52	0.07	150.57	0.34

XGBoost

XGBoost models were established by taking the spectral data of different pre-screened feature bands as input values, as shown in Table 4. After FD and SD pretreatment, the XGBoost model constructed by the Pearson method was better. Compared with RS, the R^2 values were 0.79 and 0.81, RMSE values were 53.50

and 53.35 head/leaf, and RPD values were 1.43 and 1.42, respectively. After SG and LG pretreatment, the XGBoost model constructed by the feature bands screened by the SPA method has a better effect. Compared with RS, R^2 are 0.80 and 0.87, RMSE are 58.64 and 46.70 head/leaf, and RPD are 1.11 and 1.43, respectively. It can be seen from the table that among the models for estimating the number of cotton aphids based on XGBoost, the LG+SPA model has the best effect, with R^2 of 0.87, RMSE of 46.70 heads/leaves, and RPD of 1.43.

Table 4. Evaluation of cotton aphid population estimation model based on XGBoost

Pretreatment	Band selection method	Training set			Validation set		
		R^2	RMSE heads/leaf	RPD	R^2	RMSE heads/leaf	RPD
RS	Pearson	0.76	66.15	0.90	0.40	47.98	0.72
	SPA	0.79	55.93	1.20	0.62	55.90	0.65
SG	Pearson	0.70	71.65	0.77	0.01	164.04	0.39
	SPA	0.80	58.64	1.11	0.65	45.27	0.95
FD	Pearson	0.79	53.50	1.43	0.71	41.71	1.48
	SPA	0.59	74.91	0.81	0.09	149.39	0.33
SD	Pearson	0.81	53.35	1.42	0.66	43.96	1.41
	SPA	0.52	80.33	0.70	0.05	153.67	0.35
LG	Pearson	0.76	65.74	0.91	0.36	57.62	0.99
	SPA	0.87	46.70	1.43	0.77	42.66	1.27
RG	Pearson	0.74	70.62	0.75	0.14	73.37	0.39
	SPA	0.79	59.30	1.15	0.53	62.32	0.76

PLSR

PLSR models were established using the spectral data of different pre-screened feature bands as input values, as shown in Table 5. After pretreatment by SG, FD, SD, and RG, the model constructed by Pearson was better than that by RS. After pretreatment with SG, LG, and RG, the model constructed by using SPA to screen the feature bands has a better effect than RS. Among the 12 PLSR models constructed, the LG+SPA model had the best effect, with R^2 of 0.28, RMSE of 94.27 heads/leaves, and RPD of 0.62.

Table 5. Evaluation of cotton aphid population estimation model based on PLSR

Pretreatment	Band selection method	Training set			Validation set		
		R^2	RMSE heads/leaf	RPD	R^2	RMSE heads/leaf	RPD
RS	Pearson	0.03	109.16	0.18	0.03	154.28	0.13
	SPA	0.17	101.08	0.45	0.11	147.55	0.31
SG	Pearson	0.07	107.04	0.27	0.04	152.83	0.19
	SPA	0.21	98.74	0.51	0.11	146.83	0.34
FD	Pearson	0.20	99.01	0.51	0.28	80.99	0.65
	SPA	0.10	104.97	0.34	0.08	150.06	0.23
SD	Pearson	0.16	101.65	0.44	0.08	149.39	0.30
	SPA	0.09	105.96	0.31	0.11	148.04	0.24
LG	Pearson	0.03	109.39	0.17	0.02	154.58	0.13
	SPA	0.28	94.27	0.62	0.15	143.59	0.39
RG	Pearson	0.11	104.70	0.35	0.07	150.42	0.25
	SPA	0.21	98.43	0.52	0.07	150.57	0.34

Discussion

The analysis and processing technology of crop disease and pest monitoring data is the key to remote sensing monitoring of disease and pests (Liao *et al*., 2023). Hyperspectral data can reflect the growth status of plants, and it is easy to operate and is often used for monitoring plant diseases and insect pests. In this study, five spectral data preprocessing methods, SG, FD, SD, LG, and RG, are selected to reduce the interference of noise. In this study, two methods Person and SPA were used to screen the feature bands. The Person selected the bands with the top 10 correlation coefficients, and SPA selected 2-8 bands. As can be seen from Tables 3, 4, and 5, the number of bands selected by SPA is small, and the modeling effect is better, which is the same as the research results of Xu (2020). The feature band 746 nm screened by SPA is consistent with the research results of Shukla *et al*. (2023). In this study, LR, XGBoost, and PLSR population monitoring models were constructed based on the characteristic bands. As can be seen from Tables 3, 4, and 5, the XGBoost model has obtained better monitoring effects, and better monitoring results have been obtained under different pretreatment and feature band screening. The results show that LR and PLSR are unsuitable for the model based on the feature band. It is different from the findings of Döring and Kirchner (2023).

This experiment used three methods: LR, XGBoost, and PLSR to establish prediction models. Based on linear regression and partial least squares regression, the RPD of each model was less than 1.4, indicating low prediction accuracy. This may be due to the model being too simple to capture the true distribution of the data. The FD+Pearson model performs the best in models based on the XGBoost.The vegetation index is calculated by four operations between bands, which can effectively improve the accuracy of models. Pei (2019) selected multiple vegetation indices as variables in the remote sensing estimation model of aphid damage index and used a partial least squares regression model, and the II-PLSR model R^2=0.89, which expressed a high accuracy. Based on this, the monitoring model of the cotton aphid population should also consider a vegetation index for monitoring. In the future, we will further utilize characteristic bands combined with physiological parameter changes to construct a cotton aphid population model.

Conclusions

In this study, based on the pretreatment of the reflectance spectra of single leaves, it was determined that the characteristic bands could realize the non-destructive monitoring of cotton aphid population, and the spectral monitoring model of cotton aphid population was established with the spectral data of the characteristic bands as the input variables. Spectral data preprocessing can reduce noise pollution and spectral data redundancy. In this study, the model effect after SG and LG preprocessing is due to RS. Choosing different models to build, its model effect is different. In this study, the linear regression and partial least squares regression models based on the characteristic bands were ineffective and could not be used to monitor the number of cotton aphids. LG+SPA model is the most stable. The selected feature bands are 759 nm, 764 nm, 929 nm, 938 nm, 944 nm, 948 nm, 955 nm, and 956 nm. The training set R^2 is 0.87. RMSE=46.70 head/leaf, RPD=1.43, validation set R^2=0.77, RMSE=42.66 head/leaf, RPD=1.27, can predict the number of cotton aphids well.

Authors' Contributions

JL and ZWB were responsible for determining some indicators and writing this manuscript. Make suggestions and opinions on the manuscript JL and ZWB are the coauthors of this manuscript. JQZ conducted experiments and collected all data sets. NC is responsible for editing and typesetting.QH is responsible for revision. Corresponding author SMW is responsible for the revision and quality control of the paper.

All authors read and approved the final manuscript.

Ethical approval (for researches involving animals or humans)

Not applicable.

Acknowledgements

This work was supported by National Key R&D Program of China(Grant No.2023YFD2301200); the Xinjiang Uygur Autonomous Region Talent Development Fund 'Tianchi Talent' Introduction Programme;the President's Foundation of Tarim University, China (Grant No.TDZKBS202201).

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Cai HH (2022). Research on soil nutrient detection and fertilizer formulation in jujube orchards based on hyperspectral technology. MSc Dissertation, Tarim University. *https://doi.org/10.27708/d.cnki.gtlmd.2022.000435*

Chen PC (2006). Application of Groung-based Hyperspectral Remote Sensing to Detect Spider Mite. MSc Dissertation, Shihezi University. *https://doi.org/10.7666/d.y941979*

Dilixiati YMM, Zhou JP, Xu Y, Fan XP, Yalikun SWT (2022). Cotton pest monitoring based on logistic algorithm and remote sensing image. Journal of South China Agricultural University 43(02):87-95. *https://doi.org/10.7671/j.issn.1001-411X.202106004*

Döring TF, Kirchner SM (2022). A model for colour preference behaviour of spring migrant aphids[J]. Philosophical Transactions of the Royal Society B 377(1862):20210283. *https://doi.org/10.1098/rstb.2021.0283*

Felipe DOD, Sarvio DVM, Tavares CO, Franoise DD, Mariane GFC, Carlos N (2023). Remote sensing and machine learning techniques for high throughput phenotyping of late blight-resistant tomato plants in open field trials. International Journal of Remote Sensing 44(6):1900-1921. *https://doi.org/10.1080/01431161.2023.2192878*

Feng LK, Gao GZ, Lv ZZ, Jia HM, Wang PL (2015). Interspecific competition between *Aphis gossypii* glover and *Acyrthosiphon gossypii* Mordviiko at different temperatures. Journal of Applied Insects 52(03):557-565. *https://doi.org/CNKI:SUN:KCZS.0.2015-03-004*

Gao ZH, Bai LN, Wang BY, Li ZY, Li XS, Wang YK (2011). Estimation of soil organic matter content in desertified landsusing measured soil spectral data. Forestry Science 47(06):9-16. *https://doi.org/10.11707/j.1001-7488.20110602*

Gomez SK, Oosterhuis DM, Hendrix DL, Johnson DJ Steinkraus DC (2006). Diurnal pattern of aphid feeding and its effect on cotton leaf physiology. Environmental and Experimental Botany 55(1-2):77-86. *https://doi.org/10.1016/j.envexpbot.2004.10.001*

Guo P, Wu FD, Dai JG, Wang HJ, Xu LP, Zhang GS (2017). Comparison of farmland crop classification methods based on visible light images of unmanned aerial vehicles. Journal of Agricultural Engineering 33(13):112-119. *https://doi.org/10.11975/j.issn.1002-6819.2017.13.015*

Guo YS, Liu QS, Liu GH, Huang C (2017). Extraction of main crops in yellow river delta based on MODIS NDVI time series. Journal of Natural Resources 32(10):1808-1818. *https://doi.org/10.11849/zrzyxb.20160943*

Hu HY, Ren XL, Jiang WL, Ma XY, Ma YJ, Wang D,Ma T (2018). Pesticide spray distribution of plant protection UVA in cotton field. Journal of Huazhong Agricultural University 2018,37(05):59-64. *https://doi.org/10.13300/j.cnki.hnlkxb.20180622.007*

Kong YR, Wang LJ, Feng HK, Xu Y, Liang L, Xu L, Yang XD, Zhang QQ (2022). Leaf area index estimation based on UAV hyperspectral band selection. Spectroscopy and Spectral Analysis 42(03):933-939. *https://doi.org/10.3964/j.issn.1000-0593(2022)03-0933-07*

LACITIGNOL D, TEBALDI C (2005). Effects of ecological differentiation on Lotka–Volterra systems for species with behavioral adaptation and variable growth rates. Mathematical Biosciences 194(1):95-123. *https://doi.org/10.1016/j.mbs.2004.10.012*

Liao J, Tao WY, Zang Y, Zeng HY, Wang P, Luo XW (2023). Research progress and prospect of key technologies in crop disease and insect pest monitoring. Transactions of the Chinese Society for Agricultural Machinery 54 (11):1-19. *https://doi.org/10.6041/j.issn.1000-1298.2023.11.001*

Liu J, Wang LM, Yang FG, Yang LB, Wang XL (2015). Remote sensing estimation of crop planting area based on HJ time-series images. Journal of Agricultural Engineering 31(3):199-206. *https://doi.org/10.3969/j.issn.1002-6819.2015.03.026*

Lu LM (2019). Prediction of Soil Properties Based on Hyperspectral Characteristic. MSc Dissertation, Anhui University of Science and Technology.

Luo JY, Cui JJ, Huang Q (2011). The relationship between the content of cotton leaf chlorophyll and waxiness and leaf thickness and the cotton resistance to *Apolygus lucorum*. Journal of Plant Protection 38(04):320-326. *https://doi.org/10.13802/j.cnki.zwbhxb.2011.04.016*

National Bureau of Statistics. *https://www.stats.gov.cn/sj/zxfb/202312/t20231225_1945745.html*

Niu FP, Li XG, Maimaitituerxun AZT, Zhao H (2021). Hyperspectral estimation of soil organic carbon content in the West Lakeside Oasis of Bosten Lakebased on successive projection algorithm. Journal of Zhejiang University (Agriculture and Life Sciences) 47(05):673-682. *https://doi.org/10.3785/j.issn.1008-9209.2021.01.181*

Pei PC (2019). Study on monitoring of cotton aphids damage level based on UVA hyperspectral images. MSc Dissertation, Henan Agricultural University. *https://doi.org/10.27117/d.cnki.ghenu.2019.000186*

Shukla KK, Nigam R, Birah A, Kanojia AK, Kumar A, Bhattacharya BK, Chander S (2023). Detection of aphid-infested mustard crop using ground spectroscopy. Remote Sensing 16(1):47. *https://doi.org/10.3390/rs16010047*

Wan Y (2023). Research and application of correlation analysis in wavelength selection in near-infrared spectroscopy. MSc Dissertation, Heilongjiang Bayi Agricultural University.

Wang CH (2020). Exploration of comprehensive prevention and control technology for cotton diseases and pests. China Cotton Processing (06):28-30. *https://doi.org/10.3969/j.issn.1003-0662.2020.06.011*

Wang PX, Xun L, Li L, Xie Y, Wang L (2017). Extraction of planting areas of main crops based on fourier transformed characteristics of time series leaf area index products. Journal of Agricultural Engineering 33(21):207-215. *https://doi.org/CNKI:SUN:NYGU.0.2017-21-025*

Wang QL (2015). Prediction Soil Total Nitrogen Based on the Soil Visible Near Infrared Spectral Library. MSc Dissertation, Zhejiang University.

Wang SD, Shi PJ, Zhang HB, Wang XC (2019). Retrieval of soil total nitrogen content in reclaimed farmland of mining area based on hyperspectral imaging. Chinese Journal of Ecology 38(01):294-301. *https://doi.org/10.13292/j.1000-4890.201901.034*

Wang SQ (2021). Study on Variation Characteristics and Estimation Modeling of Cotton Leaf Nitrogen Based on Hyperspectral. MSc Dissertation, Tarim University.

Xie YT, Plett D, Evans M, Garrard T, Butt M, Clarke K, Liu HJ (2024). Hyperspectral imaging detects biological stress of wheat for early diagnosis of crown rot disease. Computers and Electronics in Agriculture 217108571-.*https://doi.org/10.1016/J.COMPAG.2023.108571*

Xu JC (2020). The Research and Application of Monitoring Method of Cotton Aphid Hazard Based on Hyperspectral Imaging. MSc Dissertation, Shihezi University. *https://doi.org/10.27332/d.cnki.gshzu.2020.000679*

Yuan L (2015). Identification and Differentiation of Wheat Diseases and Insects with Multi-source and Multi-scale Remote Sensing Data[D]. PhD Thesis, Zhejiang University.

Zeng HL, Ye PS, He L, Jiang QP, Liu CH, Wei SG, Zhang QF, Dai SD, Li QY (2015). Occurrence and control strategies of seedling insect pests for transgenic insect-resistant cotton. Chinese Cotton Society. Compilation of Papers from the 2015 Annual Conference of the Chinese Cotton Society. Sichuan Academy of Agricultural Sciences Economic Crop Breeding and Cultivation Research Institute, 2015:218-219. *https://navi.cnki.net/knavi/conferences/proceedings/MHXB201508002/detail?uniplatform=NZKPT*

Zhang Q, Li XC, Yu L, Zhao BM, L H (2020). Analysis on the Occurrence and cause of main cotton diseases and insect pests in Xinjiang Production and construction corps in 2020. China Cotton 47(12):38-40. *https://doi.org/10.11963/1000-632X.zqlh.20201207*

Zhang ZG (2011). The Early Identification of Remote Sensing about Bursaphelenchus Xylophilus Based on Process Model. MSc Dissertation, Beijing Forestry University.

Zhao Q, Zhang GL, Chen XD (2005). Effects of multiplicative scatter correction on a calibration model of near infrared spectral analysis. Optics and Precision Engineering (01):53-58. *https://doi.org/10.3321/j.issn:1004-924X.2005.01.010*

Zhou YG, Wang JG, Ma TW, He ZM (2011). Effects of Mepiquat Chloride (DPC) on species reproduction of aphis gossypii glover. Acta Agriculturae Boreali-Occidentalis Sinica 20(05):199-202. *https://doi.org/10.3969/j.issn.1004-1389.2011.05.039*