DOI 10.1007/s10812-024-01801-0
Journal of Applied Spectroscopy, Vol. 91, No. 4, September, 2024 (Russian Original Vol. 91, No. 4, July–August, 2024)

VARIABLE SELECTION IN NEAR-INFRARED SPECTRA FOR MODELING OF HEMOGLOBIN CONTENT IN BIO-WATER SOLUTIONS

Renjie Fang,[a,b] Xin Han,[a,*] Xiangxian Li,[a] Jingjing Tong,[a] Minguang Gao,[a] and Yang Wang[a]

UDC 535.34:547.963.4

The background differences in water content of different samples have a very strong influence on the robustness of near-infrared spectroscopy (NIRS). For this reason, this study simulated typical biological water matrix samples with formulated hemoglobin (Hb), glucose (Glc), and distilled water, and attempted to use four different intelligent spectral variable selection algorithms [Competitive Adaptive Reweighted Sampling (CARS), Randomized Frog Hopping Algorithm (RF), Genetic Algorithm (GA), and Variable Projection Importance Algorithm (VIP)] to perform the Hb water interference-resistant feature band preferences, while combining partial least squares (PLS) in parallel to build a robust quantitative model of Hb. In addition, the applicability and validity of the model were validated using three prediction sets P_1, P_2, P_3 with different water backgrounds (the formulation method and composition were kept the same, and only the water content increased sequentially). The results showed that RF, GA, and VIP could effectively screen out the characteristic wavelengths of Hb with low sensitivity to water changes and successfully correct the water effect, but due to the large number of characteristic variables they screened out and the existence of a large number of redundant and water interference variables, this ultimately made the model's robustness less than ideal. The CARS algorithm performed the best, and the RMSEP of the three prediction sets were 0.016, 0.017, and 0.038, which is closer to the RMSECV of the calibration set. Therefore, NIRS combined with the variable selection can reduce the effect of water on model robustness and improve the prediction accuracy of the model by the method of selecting effective wave number intervals, and CARS may be one of the ideal algorithms to solve such problems.

Keywords: *near-infrared spectroscopy, hemoglobin, variable selection algorithm, water robust model.*

Introduction. Hemoglobin (Hb) is a protein found in red blood cells that is responsible for transporting oxygen. It is commonly used for oxygen transport in biological blood and is involved in the development of various diseases. As a result, research on Hb is of great importance in fields such as biochemistry, medicine, and diagnostics [1, 2]. In recent years, NIRS analysis technology has become a popular research topic in the field of Hb detection due to its advantages of being efficient, fast, nondestructive, and capable of simultaneously analyzing multiple components [3–5]. Water is often used as a solvent for infrared spectroscopic analysis of Hb; however, due to the strong polarity of liquid water, its O–H group exhibits broad and strong absorption bands in the NIRS region, often overlapping or partially overlapping with the absorption bands of Hb. At the same time, the NIRS absorption signal of Hb is very weak, making the NIRS study of Hb complex and highly difficult [6–8]. Therefore, reducing the impact of water on Hb's NIRS predictive analysis and establishing a robust analysis model is crucial for accurately analyzing Hb.

A robust NIRS analysis model refers to a model that is not affected by external factors such as water and temperature or whose impact is within an acceptable range, i.e., it has robustness and anti-interference capabilities. Currently, commonly used robust modeling methods include global modeling [9, 10]: collecting spectra of different samples under different external conditions to form a mixed training set, including different interference factors, so as to make the model more widely applicable; various preprocessing methods have also been developed to improve the changes in the spectral data set

*To whom correspondence should be addressed.

[a]Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China; email: xhan@aiofm.ac.cn; [b]University of Science and Technology of China, Hefei, China. Abstract of article is published in Zhurnal Prikladnoi Spektroskopii, Vol. 91, No. 4, p. 613, July–August, 2024.

caused by external influences [11, 12]. In addition, some literature attempts to locate the spectral part that is not sensitive to external influences by discarding noninformational or regions with large noncorrelated changes, making the model more robust and easier to understand [13, 14]. In fact, among all the variables in the spectrum, only a few or very few are related to the analysis target. The existence of other variables is not only unhelpful for modeling but may even lead to overfitting and reduce the predictive power of the model. These variables belong to interference variables — for example, in this study, the water characteristic band is an interference variable. Therefore, before modeling, we can use variable selection methods to eliminate such variables. Many accounts in the literature also point out that variable selection is very important for establishing a good robust model. It can effectively eliminate the influence of external changes on the spectrum, simplify the model, find out the most effective spectral region, and more importantly, due to the elimination of irrelevant and redundant variables, a correction model with strong predictive ability and good robustness can be obtained [15]. However, there are few studies on the application of multiple variable selection algorithms combined with NIRS in the process of establishing a robust Hb water model [16–18].

To this end, we have configured Hb, Glc, and distilled water to simulate a typical biological water-based matrix sample dataset. Hb is used as the target component, and water is used as an external interference factor to study the robust calibration model of Hb. The spectra after the best preprocessing are used as the dataset, and four currently highly representative intelligent spectral variable selection algorithms — namely, competitive adaptive reweighted sampling (CARS), randomizer frog algorithm (RF), genetic algorithm (GA), and variable projection importance algorithm (VIP) — are used to select the anti-water interference characteristic spectral interval of Hb, combined with partial least squares (PLS) analysis method to establish a robust Hb water model. The purpose of this paper is to compare the performance and robustness of different variable selection algorithms in the process of establishing a robust NIRS model for the Hb aqueous solution, and to try to find the most suitable variable selection algorithm to provide a theoretical reference for the establishment of a high-precision and strong robust NIRS prediction model.

Materials and Methods. Hb lyophilized powder (model H7379, purity ≥98%, purchased from Merck, Germany), glucose powder (analytical grade), and distilled water were used to prepare Hb–Glc aqueous solution to simulate typical biological water-based matrix samples, with the temperature maintained at room temperature (25.6°C). During the preparation of the aqueous solution samples, 20 mg of Hb lyophilized powder and 100 mg of Glc powder weighed by an electronic balance (model PWN124ZH/E, purchased from Ohaus Corporation, USA) were added to 9.5 mL of distilled water, respectively, gently shaken until completely dissolved, then water was added to a volume of 10 mL to obtain a 2 mg/mL Hb–Glc aqueous solution. During the preparation of the samples, a magnetic stirrer (model WH260, purchased from Wiggens GmbH, Germany) was used to continuously stir the samples to ensure complete dissolution. To simulate the Hb content in real biological samples, 203 groups of samples were prepared using this method, with Hb concentration ranging from ~0.19–2.89 wt.%, Glc concentration ranging from ~0.49–28.24 wt.%, and water content ranging from ~68.87–99.01 wt.%. The experimental dataset is denoted as V_0 and is used to analyze the influence of water content on the NIRS spectral analysis of Hb content and to locate the anti-water interference characteristic interval. According to the same method as already mentioned, three datasets with different water contents were prepared for the verification of the water robust analysis method in this paper, denoted as V_1, V_2, and V_3, respectively. They were uniformly divided into calibration sets C_1, C_2, and C_3 and prediction sets P_1, P_2, and P_3 using the KS (Kennard–Stone) division method at a ratio of 8:2. The water content is shown in Table 1.

The sample spectra were collected using a near-infrared spectrometer (MB3600, ABB Ltd., Canada) and a quartz cuvette liquid pool with a light path length of 1 mm was used for the experiment. The spectrum collected by the empty cuvette was used as the background spectrum, and the collection method was transmission. The resolution was 4 cm^{-1}, the number of scans was 128, and the spectral range was controlled at 4000–11000 cm^{-1}.

NIRS is an indirect analysis method that requires the establishment of a corresponding calibration model. Due to the high dimensionality of the data and the strong correlation between adjacent variables, model training is complex and time-consuming. Therefore, the selection of feature wavelength variables has become a key step in spectral analysis, mainly used to extract effective information, simplify the model, and eliminate data redundancy [15]. In this study, the Hb content in the sample set was selected as the research indicator, and the following several spectral regions were mainly included in the full wavelength range: wavelengths unrelated to Hb; wavelengths related to Hb but sensitive to water changes; wavelengths related to Hb but insensitive to water changes. The purpose of selecting spectral feature wavelengths is to accurately locate the third type of wavelength, so as to establish a robust quantitative model with little influence from water. In this study, four currently highly representative and widely used variable selection algorithms were used for performance com-

TABLE 1. Water Content Statistics for Data Sets

Dataset	Sample size	Water content range, wt.%	Mean value, wt.%
V_1	60	69.13 ~ 78.55	73.88
V_2	60	80.10 ~ 92.76	85.76
V_3	60	96.34 ~ 98.91	97.68

parison, namely competitive adaptive reweighted sampling (CARS), randomizer frog algorithm (RF), genetic algorithm (GA), and variable projection importance algorithm (VIP).

CARS is a popular algorithm for spectral data analysis [19]. CARS uses an iterative and competitive approach to calibrate sample sets based on Monte Carlo random sampling to establish a calibration model. This method treats each variable as an individual, and the variable selection process is iterative. At the same time, an exponential decay function (EDF) and adaptive reweighted sampling (ARS) are introduced to control the ratio of remaining variables, with high computational efficiency, and then screen out the optimal variable subset. RF is a newly proposed feature wavelength selection algorithm based on the reverse jump Markov chain Monte Carlo (RJMCMC) model dimension transformation technology and model cluster analysis (MPA) ideas [20]. It calculates the selection frequency of each variable and evaluates the importance of variables by using a large number of sequentially sampled submodels, thereby selecting feature wavelengths. GA is an adaptive heuristic global search algorithm that mimics the biological evolution process in nature [21]. It uses evolutionary operators such as selection, exchange, and mutation to make the target function value variable ("survival of the fittest") and ultimately achieve the optimal result. Applied to NIRS analysis, it can handle large-scale variable data well. VIP analysis is an auxiliary analysis technique of PLS [22]. VIP comprehensively considers the contribution of spectra to the construction of PLS scores and the explanatory power of PLS scores for concentration variables, representing the importance of wavelength variables to model fitting.

After collecting the spectra of the Hb aqueous solution, the raw spectral data was preprocessed and the spectra after the best preprocessing method were divided into a calibration set (163 samples) and a prediction set (40 samples) using the KS division method at a ratio of 8:2. Then, variable selection algorithms were used to select the anti-water interference characteristic interval of Hb, and at the same time, combined with the partial least squares (PLS), the robust quantitative model of Hb was established. During the modeling process, the root mean square error of cross-validation (RMSECV) and the root mean square error of prediction (RMSEP) of the prediction set were used to evaluate the performance of the calibration model. The specific formulas are as follows:

$$\mathrm{RMSECV/RMSEP} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}\,, \tag{1}$$

where n is the number of samples in the cross-validation set or prediction set, y is the true value of the sample, $\bar{y}$ is the mean value of the true value of the sample, and y_i is the predicted value of the model. A good model usually has a smaller root mean square error. When selecting feature bands, each variable selection algorithm was allowed to run 50 times repeatedly to obtain more stable and consistent variable selection results, and the subset of variables with the highest frequency of repetition was used as the final selection result. All calculations were implemented on MATLAB R2021b (MathWorks, USA), and Python 3.7 platforms.

Results and Discussion. *Effect of water content on the spectra of aqueous Hb solutions.* The water content of V_0 was counted and the median water content in V_0 was 85.76 wt.%, the mean was 86.04 wt.%, and the maximum and minimum values were 99.01 and 68.87 wt.%. Figure 1 shows the spectral and first derivative spectral graphs of V_0 at different water contents after eliminating baseline drift. From Fig. 1a, it can be seen that the change in water content has a significant impact on the entire NIRS spectral region. As the water content increases, the absorbance of the NIRS spectrum also increases, and the position of some absorption peaks also shifts, which has an impact on the prediction model of Hb content. From Fig. 1b, it can be seen that the influence of water content is not just a simple up-and-down drift of the spectrum. There is strong absorption at 5155 and 6920 cm^{-1}, where the large peak near 5155 cm^{-1} is produced by the combination frequency absorption of O–H antisymmetric and bending, and near 6920 cm^{-1} is the first harmonic absorption peak of O–H bond.

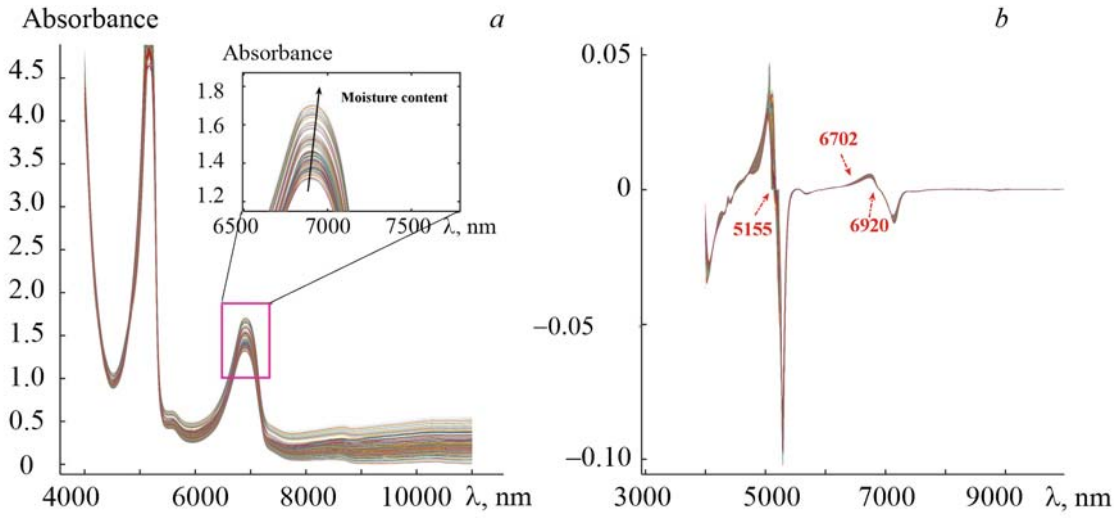

Fig. 1. Original spectrum (a) and first-order derivative spectrum (b).

It interferes with Hb's absorption peak near 6702 cm^{-1}, so the difference in water content will cause changes in Hb's characteristic absorption peak. In addition, due to the strong absorption of water in the spectrum, it is difficult to find Hb's spectral information in Fig. 1a. The introduction of variable selection algorithms can be very helpful in extracting Hb information bands, thereby establishing a concise and high-precision prediction regression model.

Spectral pre-processing. Figure 1a shows that the strong absorption of water near 5155 cm^{-1} exceeds the detection limit of the spectrum. Therefore, we focus our analysis on the wavenumber intervals of 4200–4920 and 5400–8500 cm^{-1}. The former arises from the combination of symmetric and antisymmetric O–H stretching and the first overtone of N–H stretching, while the latter is the overlap of the combination of N–H, C–N, and C–H stretching vibrations. These spectral regions provide valuable information for analyzing the composition of the sample. To eliminate the influence of high-frequency random noise, baseline drift, and sample inhomogeneity, we have adopted a variety of data preprocessing methods. These methods include baseline correction methods [first derivative (FD), second derivative (SD)]; scatter correction methods [multiple scatter correction (MSC) and standard normal variate transformation (SNV)]; smoothing correction methods (moving average smoothing (SMOOTH), Savitzky–Golay (S–G) convolution smoothing filtering); mean centering transformation (MD). After processing the raw spectral data using the aforementioned spectral preprocessing methods, a PLS regression model for Hb content was established. The model prediction results under different preprocessing methods obtained in the experiment are shown in Table 2. By analyzing the experimental data in Table 2, we found that the modeling effect of data after MD preprocessing is the best; therefore, it is used as the dataset required for subsequent feature wavelength selection.

Variable selection algorithm for selecting characteristic intervals of Hb resistance to water interference. After MD preprocessing, the V_0 spectrum was used as the modeling dataset, with Hb content as the prediction indicator. Four variable selection algorithms, CARS, RF, GA, and DE, were used to screen the full wavelength and extract the Hb characteristic information band and, finally, to establish a PLS quantitative regression model. During the modeling process, the important parameter settings of the four variable selection algorithms were as follows: the number of Monte Carlo samples for CARS was set to 100, and the smallest RMSECV subset was obtained when the sampling number was 61, screening out 18 effective wavelength points; when the RF window width was set to 20, the lowest RMSECV was obtained at the 55th window, screening out 119 effective wavelength points; in the GA algorithm, increasing the population size and iteration times helps the algorithm converge to the optimal solution, but it also means higher time cost. Therefore, the initial population size of GA is 100, the number of iterations is 20 times, and the crossover probability is 0.5, screening out 553 effective wavelength points; during the VIP evaluation variable importance process, wavelength points with VIP values greater than 1.2 were selected as feature variables, screening out 299 effective wavelength points. The screening results of four different variable selection algorithms are shown in Fig. 2.

From Fig. 2, it can be seen that CARS selects the fewest Hb variables (only 18) distributed at 4262, 4629–4916, 5400, 6609–6849, 7096, 7624–7672, 8112, and 8133 cm^{-1}. It can be seen that except for the O–H group combination

TABLE 2. The Prediction Results of PLS Regression Models for Hb Content with Different Preprocessing Methods

Pre-processing methods	RMSECV, wt.%	RMSEP, wt.%
Raw data	0.018	0.016
FD	0.019	0.017
MD	0.015	0.014
MSC	0.020	0.021
SNV	0.020	0.020
SMOOTH	0.019	0.016
S–G filtering	0.088	0.022

frequency absorption near 5400 cm^{-1} of water, most other characteristic variables show protein structure information bands, eliminating some irrelevant interference information variables and explaining the excellent performance of the CARS algorithm. RF selects 119 Hb characteristic variables, distributed at 4266–4272, 4584–4630, 4709–4771, 4862–4864, 5508–5560, 5919, 6549–6639, 6727, 6798–6920, 7047, 7095–7097, 7294, 7381–7398, 7566–7724, 8123–8141, and 8384 cm^{-1}. Among them are mainly protein structure information bands but there are still some water interference characteristic bands. GA selects the most Hb characteristic variables with a total of 553. They are mainly distributed at 4243–4293, 4551–4644, 4655–4800, 4823–4920, 5400–5431, 5460–5512, 6464–6607, 6844–6908, 7093–7132, and 7465–7708 cm^{-1}. It can be seen that in addition to containing protein structure information bands in the wavelengths selected by GA, there are also a large number of redundant and interfering variables. VIP selects a total of 299 Hb characteristic variables which are mainly distributed at 4200–4212, 4852–4920, 6952–7085, 8135–8399, and 4682–4773 cm^{-1}. In addition to containing protein information bands, it also contains some irrelevant interference variables. From the screening results in Fig. 2 it can be seen that CARS, RF, GA and VIP have all successfully selected Hb-related information bands. Compared with the full spectrum wavelength, the number of wavelengths has decreased significantly, indicating that for NIRS modeling analysis of Hb aqueous solution data set using variable selection algorithms can effectively screen target component information bands eliminate water interference information and improve model concision and robustness.

Modeling of Hb after water removal. In order to effectively examine the performance of the four variable selection algorithms in screening Hb characteristic variables and evaluate the effect of eliminating water interference, the full spectrum band of medium water content background C_2 was used in combination with PLS to establish a Hb quantitative regression model. At the same time, the Hb characteristic variables selected by C_2 in combination with the four variable selection algorithms were used to establish an Hb water correction model. Low water-content background P_1, medium water-content background P_2, and high water-content background P_3 were used before and after to verify removal of the water interference, and the division methods and proportions of the three datasets were consistent. The results are shown in Table 3.

From Table 3, it can be seen that before removing water interference, using prediction sets with different water-content backgrounds to verify the model, the RMSEP of P_1 and P_3 are significantly higher than P_2, where the water-content background of P_2 is consistent with the calibration set C_2 used for modeling. This indicates that changes in external water conditions will have a greater impact on the accuracy of the model and need to be considered in the modeling process. From Table 3, it can be observed that no matter which variable selection algorithm is used to establish a robust Hb water model, the robustness of the model after removing water has been further improved, and the modeling impact brought by different water backgrounds has been effectively eliminated. The RMSEP of the three prediction sets is relatively consistent and closer to the RMSECV of the calibration set. The results show that by using variable selection algorithms to screen Hb information bands and remove characteristic bands related to water interference information, the influence of water background on quantitative models can be effectively calibrated, and model robustness and prediction accuracy can be improved.

In addition, we compared the performance of four variable selection algorithms in removing water modeling effects horizontally, as shown in Fig. 3. Among them, CARS showed a better ability to remove water interference. The RMSEP of

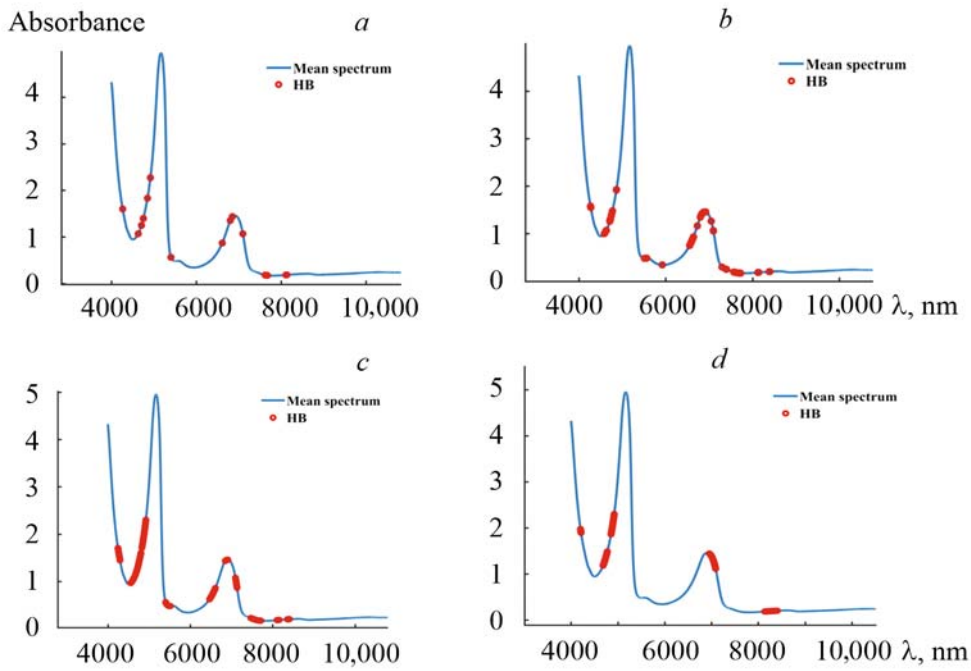

Fig. 2. Results of four different variable selection algorithms the Hb eigen bands (mean spectra of the V_0 dataset are shown): a) CARS; b) RF; c) GA; d) VIP.

TABLE 3. Comparison of the Effectiveness of Different Variable Selection Algorithms for Building Robust Models

Model	Variable selection method	nLVs*	nVAR*	RMSECV, wt.%	RMSEP, wt.%		
					P_1	P_2	P_3
Before removing water interference	FULL	9	1931	0.011	0.049	0.020	0.291
After removing water interference	CARS	7	18	0.010	0.016	0.017	0.038
	RF	8	119	0.010	0.029	0.019	0.110
	GA	7	553	0.010	0.031	0.020	0.170
	VIP	8	299	0.010	0.045	0.020	0.142

*nLVs and nVAR denotes the number of latent variables and selected variables, respectively.

the three prediction sets remained consistent with minimal water impact. RF (randomized Frog algorithm) and VIP (variable projection importance algorithm) performed generally well while GA performed the worst. In an attempt to explain this phenomenon, CARS selected the fewest number of feature variables compared to full spectrum wavelengths (reducing by 99.07%), most of which are Hb-related information bands. Among them, 4262 cm^{-1} and 4629–4916 cm^{-1} belong to N–H, C–N and C–H stretching vibration combination frequency overlapping absorption; 6609–6849 cm^{-1} belongs to N–H bond harmonic absorption; and very few contain water-related bands. Therefore, for different water background datasets, the impact of water on modeling can be well eliminated. However, when using RF, VIP and GA for modeling, although some water interference was effectively eliminated, due to the retention of too many feature variables, there are still a large number of redundant and interfering variables such as O–H group first harmonic absorption near 6920 cm^{-1} that have a significant impact on the final model's robustness.

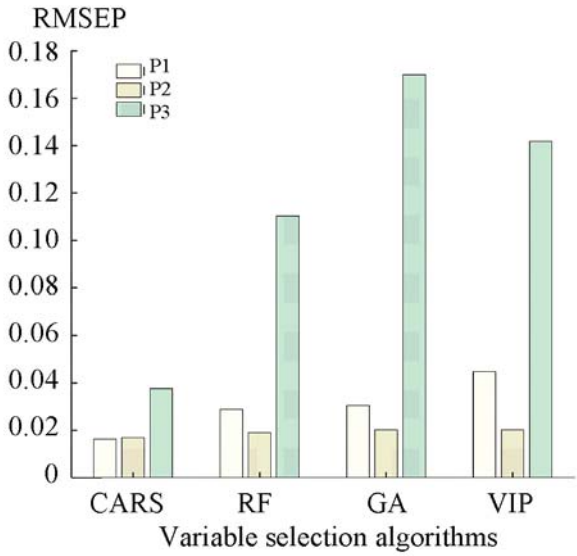

Fig. 3. Comparison of the effectiveness of different variable selection algorithms for prediction set modeling.

Conclusions. In analyzing the spectra of Hb aqueous solutions with different water contents, it was observed that as the water content increased, the absorbance spectrum of the spectra also changed, causing some absorption peaks to shift, which had a significant impact on the establishment of a quantitative model for Hb. To eliminate this effect, four variable selection algorithms were used to select the antiwater interference characteristic interval of Hb and then establish a robust Hb water model. At the same time, three prediction sets with different water backgrounds were used to verify the robustness of the model before and after modeling. The results showed that, regarding the four variable selection algorithms (CARS, RF, GA, and VIP), all selected Hb anti-water interference information bands. Compared with the full spectrum wavelength, the number of characteristic variables decreased significantly. After removing water interference, the RMSEP of the three prediction sets all decreased and were relatively consistent, indicating the effectiveness of the method. In addition, CARS performed best with fewer modeling variables used and more consistent RMSEP for P_1, P_2, and P_3. It is more suitable for Hb water robust model analysis. The fundamental reason is that, compared to the other three variable selection algorithms, CARS selected Hb characteristic variables with better robustness and less water interference information. Thus, it performs well in calibrating the impact of different water backgrounds on spectroscopy.

In subsequent research, therefore, for a large amount of water interference information in RF, GA, and VIP, it needs to be further screened out and characteristic variables need to be optimized to ensure that subsequent Hb characteristic bands used for modeling analysis have higher robustness and models are also more concise and achieve better prediction accuracy.

Acknowledgments. The authors gratefully acknowledge the support of the National Natural Science Foundation of China (No. 42075135) and the President's Fund Project of Hefei Institute of Materials Science, CAS (No. YYJJ2022QN09).

REFERENCES

1. D. Moorthy, R. Merrill, S. Namaste, and L. Iannotti, *Adv. Nutr.*, **11**, No. 6, 1631–1645 (2020).
2. D. Lelli, R. A. Incalzi, and C. Pedone, *J. Am. Geriat. Soc.*, **65**, No. 11, 2369–2373 (2017).
3. J. T. Kuenstner, K. H. Norris, and W. F. Mc Carthy, *Appl. Spectrosc.*, **48**, No. 4, 484–488 (1994).
4. K. Y. Wang, X. H. Bian, M. Zheng, P. Liu, L. G. Lin, and X. Y. Tan, *Spectrochim. Acta A*, **263**, Article ID 120138 (2021).
5. H. Tian, L. N. Zhang, M. Li, Y. Wang, D. G. Sheng, J. Liu, and C. M. Wang, *Infrared Phys. Technol.*, **102**, Article ID 103003 (2019).
6. H. Y. Yang, S. N. Yang, J. L. Kong, A. C. Dong, and S. N. Yu, *Nat. Protoc.*, **10**, No. 3, 382–396 (2015).
7. M. L. Fan, W. C. Cai, and X. S. Shao, *Appl. Spectrosc.*, **71**, No. 3, 472–479 (2017).

8. K. I. Izutsu, Y. Fujimaki， A. Kusabara, Y. Hiyama, C. Yomota, and N. Aogagi, *J. Pharm. Sci-Us.*, **95**, No. 4, 781–789 (2006).
9. J. A. Hageman, J. A. Westerhuis, and A. K. Smilde, *J. Near Infrared Spectrosc.*, **13**, 53–62 (2005).
10. N. K. Wijewardane, Y. Ge, and C. L. S. Morgan, *Eur. J. Soil Sci.*, **67**, No. 5, 605–615 (2016).
11. Y. P. Du, Y. Z. Liang, and Y. Ozaki, *Anal. Sci.,* **20**, No. 9, 1339–1345 (2004).
12. S. Chakraborty, B. Li, D. C. Weindorfc, and C. L. S. Morgan, *Geoderma*, **337**, 65–75 (2019).
13. H. Swierenga, P. J. de Groot, A. P. de Weijer, M. W. J. Derksen, and L. M. C. Buydens, *Chemometr. Intell. Lab.,* **41**, No. 2, 237–248 (1998).
14. B. Nadler and R. R. Coifman, *J. Chemometrics*, **19**, No. 2, 107–118 (2005).
15. Y. H. Yun, Li, B. C. Deng, and D. S. Cao, *Tract. Trend. Anal. Chem.,* **113**, 102–115 (2019).
16. K. Rahlow and W. Hubner, *Appl. Spectrosc.*, **51**, No. 2, 160–170 (1997).
17. S. S. Bai, H. Wang, Y. J. Chen, and H. S. Wang, *Spectrosc. Spectral. Anal.*, **35**, No. 4, 894–898 (2015).
18. B. Yuan, K. Murayama, Y. Q. Wu, R. Tsenkova, X. M. Dou, S. Era, and Y. Ozaki, *Appl. Spectrosc.*, **57**, No. 10, 1223–1229 (2003).
19. H. D. Li, Y. Z. Liang, Q. S. Xu, and D. S. Cao, *Anal. Chim. Acta*, **648**, 77–84 (2009).
20. H. D. Li, Q. S. Xu, and Y. Z. Liang, *Anal. Chim. Acta*, **740**, 20–26 (2012).
21. R. M. Jarvis and R. Goodacre, *Bioinformatics*, **21**, No. 7, 860–868 (2005).
22. I. G. Chong and C. H. Jun, *Chemometr. Intell. Lab.,* **41**, No. 2, 103–112 (2005).