Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

A Wavelength Selection Model Based on Successive Projections Algorithm for pH Detection of Water by VIS-NIR Spectroscopy

To cite this article: Lina Li and Shangxing Guo 2021 *J. Phys.: Conf. Ser.* **1813** 012002

View the article online for updates and enhancements.

You may also like

- Metrological challenges for measurements of key climatological observables. Part 3: seawater pH
A G Dickson, M F Camões, P Spitzer et al.
- Uncertainty estimation of primary pH-measurement system by Monte-Carlo simulation method
R O Cristancho and A M Castillo
- Establishment of Traceability Chain For pH Measurement in Indonesia
Ayu Hindayani, Nuryatini Hamim, Christine Elishian et al.

This content was downloaded from IP address 27.71.108.77 on 27/09/2023 at 08:40

A Wavelength Selection Model Based on Successive Projections Algorithm for pH Detection of Water by VIS-NIR Spectroscopy

Lina Li* and Shangxing Guo

College of Mechanical Engineering and Automation, Huaqiao University, Xiamen, China

*Corresponding author email: lilina@hqu.edu.cn

Abstract. A novel pH detection method using VIS-NIR (visible near infrared) spectroscopy for water is introduced in this paper. In order to improve prediction accuracy and calculation speed of pH detection using VIS-NIR spectroscopy, a wavelength selection model of SPA-MLR (successive projections algorithm and multiple linear regression) is proposed. Two experiments (pH measurement experiment by VIS-NIR spectrophotometer and pH measurement experiment by VIS-NIR grating spectrograph) are employed to evaluate the performance of SPA-MLR. At meanwhile, PLS (partial least squares), GA-PLS (genetic algorithm and partial least squares) are introduced to compare with SPA-MLR. The results indicate that, the prediction accuracy and calculation speed of SPA-MLR are all better than that of PLS and GA-PLS. Under the pH measurement experiment using VIS-NIR spectrophotometer, only one wavelength is selected from 351 original wavelengths for calibration, the RMSEP (root mean square errors of prediction) is 0.35, and the model training speed is 2.36s. Under the pH measurement experiment using VIS-NIR grating spectrograph, 3 wavelengths are selected from 2860 original wavelengths for calibration, the RMSEP is 0.50, and the model training speed is 56.45s. For the two experiments, the selected wavelength variables by SPA are distributed mainly in the band of 870 nm to 990 nm.
Keywords: Visible near infrared spectroscopy; Wavelength selection; Multivariate calibration; pH; Water.

1. Introduction

pH value as the most basic property of water, can affect the dissociation degree of weak acid and alkali in water body, and reduce the toxicity of chloride, ammonia, hydrogen sulfide, etc. It has influence on the change of water quality, growth and decline of biological reproduction, corrosiveness and water treatment effect. It is an important parameter to evaluate water quality, so it is of great significance to measure the pH value of water body. [1-5]

Based on the traditional glass electrode method for pH value measurement, this kind of instrument is small and with high-precision ordinarily[6-7], but it also has some shortcomings, such as, it needs to be calibrated before use, and can not guarantee real-time and continuous monitoring and so on.

In recent years, VIS-NIR (visible near infrared) spectroscopy has been used in environmental monitoring, especially in the study of water quality detection[8-11]. It can be seen that VIS-NIR technology is simple, rapid, nondestructive, reagent free and can be used for online measurement.

In this research, the feature wavelength selection method is discussed to improving measurement accuracy for pH value measurement in water by VIS-NIR spectroscopy. Wavelength variables

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

selection can be benefit not only to modeling efficiency since several variables involved in calibration, it is more especially when the amount of original experimental data is very large, but also to improve the prediction accuracy since the feature information is employ in calibration, which is also one of the reasons for the stability of the prediction model optimized by wavelength selection.

Wavelength variables selection is a key point for constructing high-precision calibration model in determination of the content or nature of the substance using spectroscopy, which has attracted the attention of many researchers[12-17]. Nowadays, there are four main kinds of wavelength variable selection method: (1) the calculation method based on statistics, such as correlation coefficient method[18-19]; (2) the algorithms based on PLS (partial least squares), such as UVE (uninformative variables elimination), iPLS (interval partial least squares), and so on[20-21]; (3) the strategies based on intelligent computing, such as GA (genetic algorithm), PSO (particle swarm optimization), SAA (simulated annealing algorithm), and so on[22-25]; (4) the methods based on calculating certain distance parameter of spectra, e.g. SPA (successive projections algorithm) [26-29].

The correlation coefficient method, UVE method and iPLS method are all needs some application experience in wavelength selection. GA, PSO and SAA, they are global optimization algorithms based on intelligent calculation, which are also commonly used in wavelength selection. Especially GA, which is used in many spectra analysis applications, and the application results are very well. However, the limitation of these intelligent optimization algorithms is that, the selected wavelength variables are random, which will affect the robustness of the calibration model. And the iterative calculation time is long, especially when there are many wavelength variables involved, the calculation efficiency is low.

SPA is a method based on projection calculation. Its advantage is that it can eliminate the problem of collinearity among variables[30-31]. Many application studies show that the calibration model established with selected wavelength variables by this method has high precision. And the analysis speed is fast because only a few variables employed in prediction model.

Therefore, in this paper, SPA-MLR (successive projections algorithm and multiple linear regression) is introduced for measuring pH value of water using VIS-NIR spectroscopy. Under the SPA, based on the projection algorithm, the feature wavelength variables selected is determinacy. The number of selected wavelength variables is relatively small, and the selected variables are not as random as some intelligent algorithms do. Using MLR, the math relationship of feature wavelength variables and pH value is constructed. Under the SPA-MLR, the least wavelength variables are selected, and the most relevant information with no collinearity is included in calibration, which is benefit to improve calculation speed and prediction performance for pH value measurement using VIS-NIR spectroscopy. Two experiments (pH measurement experiment by VIS-NIR spectrophotometer and pH measurement experiment by VIS-NIR grating spectrograph) were employed to evaluate the performance of SPA-MLR. Two different spectrometers were used in this research, and the number of wavelength variables of original spectrum was different in the two experiments. At meanwhile, PLS, GA-PLS (genetic algorithm and partial least squares) were introduced to compare with SPA-MLR. This research mainly includes three parts:

(1) The common used intelligent algorithm of GA-PLS for pH value measurement in distilled water using VIS-NIR spectroscopy was constructed;

(2) SPA-MLR proposed here for pH value measurement to the same experiment samples by VIS-NIR spectroscopy was constructed;

(3) The performance of PLS, GA-PLS and SPA-MLR in the determination of pH value in distilled water by VIS-NIR spectroscopy was compared and discussed, which including prediction accuracy, distribution characteristics of selected variables, stability of the algorithm, calculation speed, and simplicity.

2. Model and Algorithm

2.1. Wavelength Selection Method Based on GA

GA was first proposed by J. Holland of the University of Michigan in 1975[22]. The natural selection and genetic mechanism of biology is referenced in this algorithm, such as selection, exchange, mutation and so on. With the continuous genetic iteration, the variables with better value are retained,

and the variables with worse value are eliminated, and the optimal result is achieved finally. GA has been widely used in the field of spectral analysis, and it shows that better results have been obtained in the selection of wavelength variables.

GA is the commonly used wavelength selection method nowadays in spectral analysis because of its advantages of global optimization and easy operation[23]. But we should pay attention to the following problems in practical use. Firstly, many running steps of GA have strong randomness, such as initial group, selection, crossover, variation, and so on. So it is impossible to guarantee the repeatability of each wavelength selection result. Secondly, many research experiences show that, if there are too few samples or too many variables the risk of overfitting is high, and the result is unreliable. Under the GA, the suggested ratio of the number of wavelength variables and samples is generally less than 4 in calibration. And it is also suggested to have a number of starting variables not greater than 200 to reduce the risk of overfiting. Thirdly, it is very important for GA to choose the appropriate fitness function. The results of different fitness functions will be very different.

The ordinary algorithm of GA mainly includes five parts: code for parameter, initialization of population, design of fitness function, design of genetic operation, selection of stopping criterion and variables. For more detailed algorithm description of GA for wavelength selection, see Reference 22.

2.2. Wavelength Selection Method Based on SPA

SPA is a deterministic algorithm and has good repeatability, which does not employ stochastic operations such as GA does. SPA employs simple projection operations in a vector space to obtain subsets of variables with small collinearity[30-31]. The main principle of variable selection by SPA is that the new variable selected is the one among all the remaining variables, which has the maximum projection value on the orthogonal subspace of the previous selected variable.

As shown in Figure 1, it is a simple example to explain how to select the first informative wavelength variable by SPA. As x_1 is the start variable, the two dimensional subspace is given which is orthogonal to vector x_1. Then the projection lengths (as p_2, p_3, p_4 shown in Figure 1) of the other variables can be obtained. Thus the first informative wavelength variable (x_3) can be chosen, which has the max projection length in subspace orthogonal to vector x_1. And the vector x_3 will be the next start variable for selecting next variable. This means that the next selected variable is the variable with the largest projection length in the subspace orthogonal to the previous one[31].

The key problem is how many variables are appropriately needed for calibration. Usually, the optimal number of wavelength variables is determined by RMSECV (root mean square error of cross validation) of multivariable calibration with different number of variables. Ordinarily, the number of selected variables is not beyond the number of calibration samples.

Based on projection calculation, SPA extracts informative variable with minimum redundant information, which is benefit to elimination of colinearity between calibration variables. More details on the theoretical description of SPA can be obtained from Reference 31.

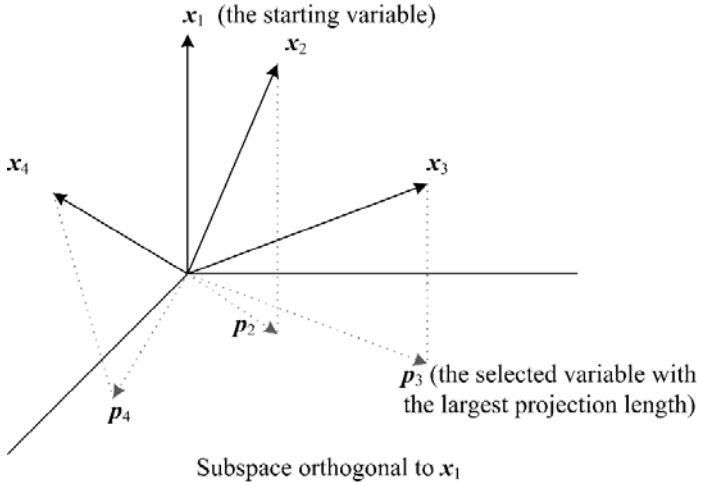


Figure 1. Select the first informative wavelength variable by SPA.

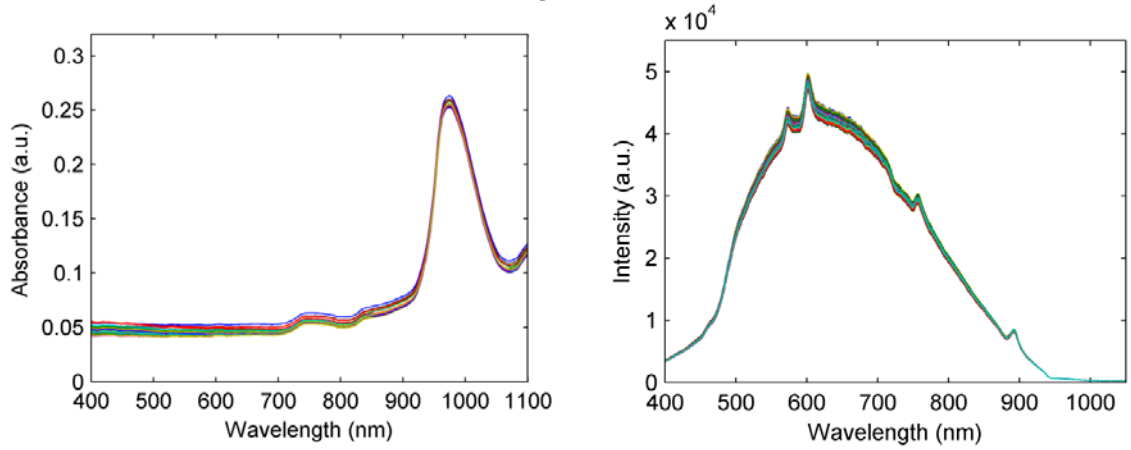
3. Experimental Part

3.1. Distilled Water pH Measurement Experiment by VIS-NIR Spectrophotometer

Instrumentation: The VIS-NIR spectra data sets of distilled water with different pH value were got by a VIS-NIR spectrophotometer V-1800PC (Shanghai Mapada Instrument Co., Ltd., China) with 2 nm interval. Reference value of pH was got by acidometer (Shanghai Lichen-BX Instrument Technology Co., Ltd., China) with ±0.1 units' precision.

Experiment part: The distilled water samples with different pH value were got by adding reagent (0.1 mmol/L HCL or 0.1 mmol/L NaOH) randomly. It was transmission measurement with 1cm cuvette, and empty light path for background. Every spectrum was the average of 10 spectra of the sample. At mean while, the pH value of the water sample was measured by acidometer, which as the reference value for calibration.

Sample data: The spectra band was from 400 to 1100 nm with 351 variables. There were 34 distilled water samples got by this experiment. The range of pH value is from 3.2 to 10.6. The average value is 6.15, and the standard deviation is 2.08. The original spectrum is shown in Figure 2(a).



(a) (b)

Figure 2. Original spectrum of distilled water pH measurement experiment: (a) using VIS-NIR spectrophotometer; (b) using VIS-NIR grating spectrograph.

3.2. Distilled Water pH Measurement Experiment by VIS-NIR Grating Spectrograph

Instrumentation: The VIS-NIR spectra data sets of distilled water with different pH value were got by a VIS-NIR grating spectrograph FLAME-T-XR1-RS (Ocean Optics, U.S.A) with 3648 pixsels. And the accessory includes light source HL-2000-LL, transmission fiber. Reference value of pH was got by acidometer (Shanghai Lichen-BX Instrument Technology Co., Ltd., China) with ±0.1 units precision.

Experiment part: The distilled water samples with different pH value were got by adding reagent (0.1 mmol/L HCL or 0.1 mmol/L NaOH) randomly. It was transmission measurement with 1cm cuvette, and empty light path for background. Scan settings included average "200", integral time "25 ms", smooth "3". At mean while, the pH value of the water sample was measured by acidometer, which as the reference value for calibration.

Sample data: The spectra band was from 400 to 1049 nm with 2860 variables. There were 60 distilled water samples got by this experiment. The range of pH value is from 2.7 to 12.2. The average value is 6.90, and the standard deviation is 2.55. The original spectrum is shown in Figure 2(b).

3.3. Evaluation Method and Calculation Software

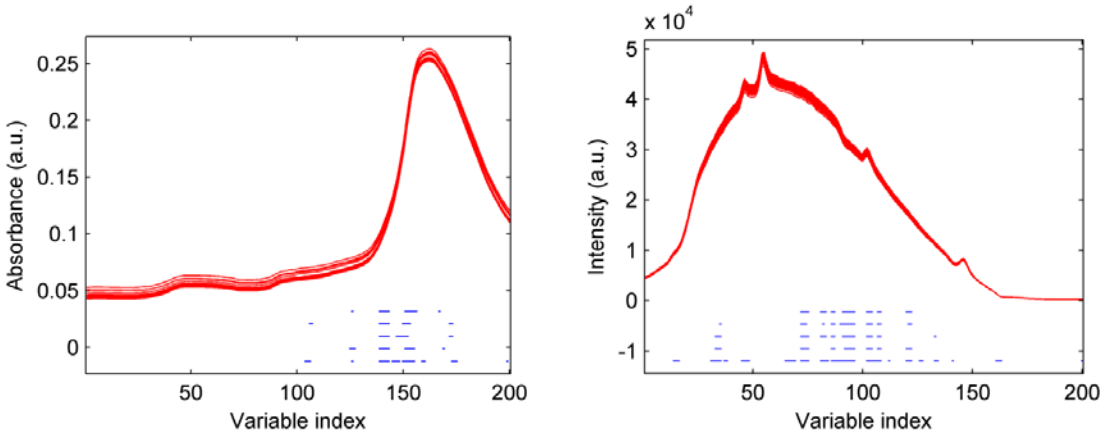
In this research, the wavelength selection methods, GA, and SPA were introduced firstly. Then PLS and MLR was employed respectively for multivariate calibration. The two experimental data sets were partitioned into two parts, training set and prediction set. For the pH measurement experiment with VIS-NIR spectrophotometer, 26 samples were employed for training and the rest of 8 samples were

employed for prediction. For the pH measurement experiment with VIS-NIR grating spectrograph, 45 samples were employed for training and the rest of 15 samples were employed for prediction.
3 points were discussed mainly in the two experiments: (1) RMSEP (root mean square errors of prediction) of prediction set is introduced for evaluating the prediction performance of multivariate calibration model with different wavelength selection method; (2) the distribution characteristics of wavelength variables selected by different wavelength selection methods is discussed; (3) the efficiency of different wavelength selection methods is discussed.
All the calculation programs were compiled by MATLAB. In this research, all the MATLAB codes were running on MATLAB R2013b (The Mathworks, Inc., Natick, MA, USA) software in Windows 7 computer operation system with 32-bit. The CPU of the computer utilized in this research is Intel Core i5-2450, the main frequency is 2.5GHz, and the memory is 4G.

4. Results and Discussion

4.1. Wavelength Variable Selection Results by GA

For the training set samples, GA is employed to wavelength selection. The test set samples are used for evaluating the GA-PLS algorithm. PLS-GA TOOLBOX is introduced here for analysis[22-23]. The default values are: population size is 30 chromosomes; there are 5 variables per chromosome in the original population; probability of mutation is 1%; probability of cross-over is 50%; the max number of components is 15. In general, the number of starting variables is not greater than 200 for reducing the risk of overfitting. And in order to improve the robustness of the GA method, GA program is performed 5 times for each data set. According to the cross validation results of PLS, the number of variables selected can be determined.
For the pH measurement experiment by VIS-NIR spectrophotometer, the 200 variables from 652 nm to 1050 nm are employed in this study. The ratio of variable number and sample number is 7.7. The number of evaluations is 50. GA running results of 5 times are shown in Figure 3(a). The sign '-' marks the position of selected variables' index. It is clear that, the results of the 5 times running are not the same, there are some differences. But it can be seen that, the distribution origin of these variables are all within index of 105 to 174 (wavelength from 860 nm to 998 nm), and most of them in the band of 125 to 174 (wavelength form 900 nm to 998 nm). The RMSECV results of GA-PLS for the 5 times running are given in Table 1. The results indicate that, the best result is got by the first running. Thus, the best number of variables selected by GA is 13. The variable indexes are 141, 140, 142, 152, 153, 154, 151, 139, 155, 126, 156, 143, and 167. The corresponding wavelengths are 932 nm, 930 nm, 934 nm, 954 nm, 956 nm, 958 nm, 952 nm, 928 nm, 960 nm, 902 nm, 962 nm, 936 nm, and 984 nm.
Similarly, for the pH measurement experiment by VIS-NIR grating spectrograph, the band of 415 nm to 1049 nm is employed in this study. Thinking of the number of starting variables not greater than 200 as the GA suggested, the average value of 14 contiguous wavelengths is got, thus there are 200 variables involved in GA program. The ratio of variable number and sample number is 4.4. The number of evaluations is 163. GA running results of 5 times are shown in Figure 3(b). The sign '-' marks the position of selected variable index. It is clear that, the results of the 5 times running are not the same too. But it can be seen that, most of the selected variables are distributed around variable index from 70 to 122 (wavelength from 653 nm to 820 nm), which is very different to the pH measurement experiment by VIS-NIR spectrophotometer. The RMSECV results of GA-PLS for the 5 times running are given in Table 1. The results indicate that, the best result is got by the second running. Although the accuracy of the fifth run is similar to that of the second run, relatively few variables are selected for the second run. Thus, in this experiment, the best number of variables selected by GA is 21. The selected wavelength are 731.46nm, 760.15 nm, 763.31 nm, 734.66 nm, 663.27 nm, 666.56 nm, 756.98 nm, 659.98 nm, 728.25 nm, 737.86 nm, 535.63 nm, 775.95 nm, 721.82 nm, 772.80 nm, 819.65 nm, 708.92 nm, 725.04 nm, 692.70 nm, 718.61 nm, 816.55 nm, 705.68 nm.



(a) (b)

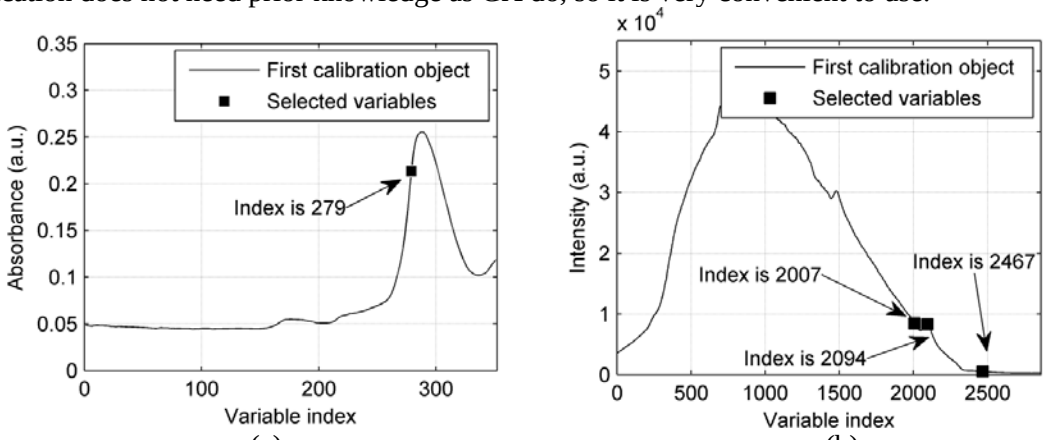
Figure 3. Variable selection results of 5 times running by GA for the pH measurement experiment: (a) using VIS-NIR spectrophotometer; (b) using VIS-NIR grating spectrograph.

Table 1. The cross validation results by GA-PLS for the two pH measurement experiments.

Test index	pH measurement experiment by VIS-NIR spectrophotometer			pH measurement experiment by VIS-NIR grating spectrograph		
	The number of variables	Components	RMSECV	The number of variables	Components	RMSECV
1	13	4	1.19	23	6	0.98
2	13	4	1.28	21	6	0.90
3	13	8	1.20	24	5	0.92
4	14	4	1.22	22	6	0.94
5	25	5	1.29	58	7	0.90

4.2. Wavelength Variable Selection Results by SPA

Based on the theory of SPA[31], the max number of selected variables by SPA is not more than the number of the training samples in calibration. Thus, for the pH measurement experiment with VIS-NIR spectrophotometer, the minimum number of variables selected by SPA is set to 1, and the max number of variables selected by SPA is set to 25. The wavelength selection result is shown in Figure 4(a). It is clear that the selected variable index is 279, the corresponding wavelength is 956 nm. SPA is a deterministic algorithm, the repeated calculation results are completely consistent. And its application does not need prior knowledge as GA do, so it is very convenient to use.



(a) (b)

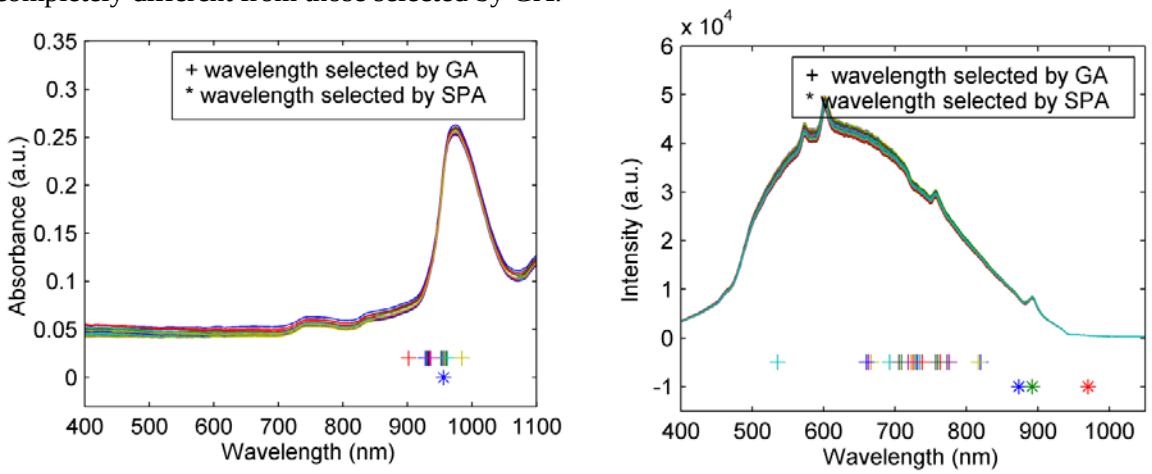
Figure 4. Variable selection results by SPA for the pH measurement experiment: (a) using VIS-NIR spectrophotometer; (b) using VIS-NIR grating spectrograph.

For the pH measurement experiment with VIS-NIR grating spectrograph, the minimum number of variables selected by SPA is set to 1, and the max number of variables selected by SPA is set to 44. The wavelength selection result is shown in Figure 4(b). It is clear that the selected 3 variables' index are 2007, 2094, and 2467, the corresponding wavelengths are 873.24 nm, 891.92 nm, and 970.13 nm.

4.3. Comparison Discussion of GA and SPA

For the pH measurement experiment by VIS-NIR spectrophotometer, the wavelength variables of the first run are selected as the optimal result of GA. There are 13 wavelength variables selected by GA and 1 wavelength variables selected by SPA. The distribution comparison of wavelength variables selected by GA and SPA is shown in Figure 5(a). It is clear that, the wavelength 956 nm selected by SPA is also included in the wavelength set selected by GA.

For the pH measurement experiment with VIS-NIR grating spectrograph, the wavelength variables of the second run are selected as the optimal result of GA. There are 21 wavelength variables selected by GA and 3 wavelength variables selected by SPA. The distribution comparison of wavelength variables selected by GA and SPA is shown in Figure 5(b). It shows that the selected 3 variables by SPA are completely different from those selected by GA.



(a) (b)

Figure 5. The distribution comparison of wavelength selected by GA and SPA for the two experiments: (a) using VIS-NIR spectrophotometer; (b) using VIS-NIR grating spectrograph.

And prediction set data is used to evaluate the performance of PLS, GA-PLS and SPA-MLR. The prediction parameters are given in Table 2. The results indicate that, under the pH experiment by VIS-NIR spectrophotometer, the prediction accuracy of GA-PLS and SPA-MLR are similar. And the models with selected wavelength are all better than the PLS model with full spectra. Moreover, for this experiment, the wavelength variables selected by the two methods are overlapped. Under the pH measurement experiment by VIS-NIR grating spectrograph, the prediction performance of SPA-MLR is the best. Based on the SPA-MLR, only 3 variables are employed in calibration from original 2860 variables, the RMSEP is 0.5. For this experiment, 21 variables are employed in GA-PLS, its prediction performance is similar to PLS, but GA-PLS has less calibration set variables than PLS. It is also noticed that the wavelength variables selected by GA are completely different from that of SPA.

Moreover, the calculation speed of the models in one time running is given in Table 2. It is showed that the computation speed of SPA-MLR is much faster than that of PLS and GA-PLS. SPA-MLR can get a result in a few seconds, while GA-PLS needs several minutes, ten minutes or even dozens of minutes according to the amount of original data. Moreover GA needs the auxiliary analysis of prior knowledge, and the analysis process needs to be adjusted constantly due to the different results of each operation. In fact, it usually takes engineers several hours or even many days to apply GA analysis. SPA does not need prior knowledge, and the best result can be obtained in one computation. The stability and accuracy of SPA are relatively high.

Table 2. The calibration model performance comparison of PLS, GA-PLS and SPA-MLR.

Model		pH experiment with spectrophotometer	pH experiment with grating spectrograph
PLS	Number of modeling variables	351	2860
	Number of latent variables	1	1
	RMSEP	0.69	1.15
	Modeling speed (s)	3.56	373.03
GA-PLS	Number of modeling variables	13	21
	Number of latent variables	1	1
	RMSEP	0.36	1.16
	Modeling speed (s)	369.19	965.51
SPA-MLR	Number of modeling variables	1	3
	RMSEP	0.35	0.50
	Modeling speed (s)	2.36	56.45

5. Conclusions

A novel pH detection method based on VIS-NIR spectroscopy for water is introduced in this paper. In order to improve prediction accuracy and calculation speed, a wavelength selection model of SPA-MLR is proposed for pH detection using VIS-NIR spectroscopy. Two experiments for water pH measurement are employed to evaluate the wavelength selection method of GA and SPA. The effects of GA and SPA for wavelength optimization are discussed from the aspects of algorithm characteristics, distribution property of selected wavelength variables, prediction performance of calibration and training speed of model. The conclusions are as follows:

(1) The calibration model of wavelength optimization is better than that of full spectrum, whether it is the prediction accuracy or the speed of model training, because only a few informative wavelength variables for target analyte are involved in calibration. In this experimental research, the prediction accuracy of SPA-MLR is obviously better than that of PLS and GA-PLS.

(2) For the two groups of experimental studies, there are obvious differences in the distribution characteristics of variables selected by SPA and GA. The SPA method based on projection algorithm avoids the collinearity, and the prediction performance of the established calibration model is relatively better. And the feature wavelengths selected by SPA are in the band of 870 nm to 990 nm for pH measurement in water samples.

(3) SPA is a deterministic algorithm based on projection operator, and the selected variables remain unchanged in each repeated run. GA is a stochastic optimization algorithm based on genetic strategy, and the results are different each time while it is run repeatedly. SPA is more stable than GA.

(4) The analysis speed of SPA is faster than that of GA, especially when there are many original variables. SPA is based on a certain projection algorithm, the number of variables selected is not more than the number of training samples, and the calculation speed is fast. The random iterative calculation process of GA is complex, and the optimal model needs to be trained many times, so the calculation speed is relatively slow.

(5) Prior knowledge is not needed for the application of SPA, which is very convenient to use of SPA. And a certain prior knowledge is needed for the application of GA, and the initial variables involved in GA should not exceed 200. The selection of initial variables will affect the performance of the calibration model based on GA.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 51805177), and the Scientific Research Fund Project of Huaqiao University (No. 11BS413).

References

[1] X. Zhang, W. Zhang, and S. Tang 2019 *J. Photochem. Photobiol. A* vol 372 (Amsterdam: Elsevier) p 71

[2] L. Ma, W. Cao, J. Liu, D. Deng, Y. Wu, Y. Yan, and L. Yang 2012 *Sens. Actuators B Chem.* vol 169 (Amsterdam: Elsevier) p 243
[3] E. Hudson-Heck, and R. H. Byrne 2019 *Anal. Chim. Acta* vol 1090 (Amsterdam: Elsevier) p 91
[4] B. Yang, M. C. Patsavas, R. H. Byrne, and J. Ma 2014 *Mar. Chem.* vol 160 (Amsterdam: Elsevier) p 75
[5] V. M. C. Rérolle, C. F. A. Floquet, A. J. K. Harris, M. C. Mowlem, R. R. G. J. Bellerby, and E. P. Achterberg 2013 *Anal. Chim. Acta* vol 786 (Amsterdam: Elsevier) p 124
[6] D. K. Rajan, M. Patrikoski, J. Verho, J. Sivula, H. lhalainen, S. Miettinen, and J. Lekkala 2016 *Talanta* vol 161 (Amsterdam: Elsevier) p 755
[7] R. Gotor, P. Ashokkumar, M. Hecht, K. Keil, and K. Rurack 2017 *Anal. Chem.* vol 89 (Washington: ACS) p 8437
[8] J. Østergaard, H. Jensen, S. W. Larsen, C. Larsen, and J. Lenke 2014 *J Pharmaceut. Biomed.* Vol 100 (Amsterdam: Elsevier) p 290
[9] Y. Takeshita, K. S. Johnson, L. J. Coletti, H. W. Jannasch, P. M. Walz, and J. K. Warren 2020 *Mar. Chem.* vol 223 (Amsterdam: Elsevier) p 103801
[10] A. Inagawa, A. Sasaki, and N. Uehara 2020 *Talanta* vol 216 (Amsterdam: Elsevier) p 120952
[11] Z. Rasouli, H. Abdollahi, and M. Maeder 2020 *Anal. Chim. Acta.* vol 1109 (Amsterdam: Elsevier) p 90
[12] T. Lemos, and J. H. Kalivas 2017 *Chemom. Intell. Lab. Syst.* vol 168 (Amsterdam: Elsevier) p 121
[13] T. Lei, and D. W. Sun 2020 *Anal. Chim. Acta* vol 1110 (Amsterdam: Elsevier) p 169
[14] E. Bonah, X. Huang, J. H. Aheto, R.Yi, S. Yu, and H. Tu 2020 *Infrared Phys. Techn.* vol 107 (Amsterdam: Elsevier) p 103327
[15] W. Ng, B. Minasny, B. P. Malone, M. C. Sarathjith, and B. S. Das 2019 *Comput. Electron Agr.* vol 158 (Amsterdam: Elsevier) p 201
[16] M. Goodarzi, D. E. Bacelo, S. E. Fioressi, and P. R. Duchowicz 2019 *Microchem. J.* vol 145 (Amsterdam: Elsevier) p 872
[17] C. Q. Li, Z. Fang, and Q. S. Xu 2020 *Chemom. Intell. Lab. Syst.* vol 198 (Amsterdam: Elsevier) p 103935
[18] Z. Xu, S. Fan, W. Cheng, J. Liu, P. Zhang, Y. Yang, C. Xu, B. Liu, J. Liu, Q. Wang, and Y. Wu 2020 *Spectrochim. Acta A* vol 230 (Amsterdam: Elsevier) p 118053
[18] Y. H. Yun, H. D. Li, B. C. Deng, and D. S. Cao 2019 *Trend. Anal. Chem.* vol 113 (Amsterdam: Elsevier) p 102
[20] V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, and C. Sterna 1996 *Anal. Chem.* vol 68 (Washington: ACS) p 3851
[21] R. Shan, W. Cai, and X. Shao 2014 *Chemom. Intell. Lab. Syst.* vol 131 (Amsterdam: Elsevier) p 31
[22] R. Leardi, and A. L. González 1998 *Chemom. Intell. Lab. Syst.* vol 41 (Amsterdam: Elsevier) p 195
[23] R. Leardi 2000 *J. Chemometr.* vol 14 (Hoboken, New Jersey: John Wiley & Sons, Inc.) p 643
[24] Y. Liu, S. Zhou, W. Han, W. Liu, Z. Qiu, and C. Li 2019 *Anal. Chim. Acta* vol 1086 (Amsterdam: Elsevier) p 46
[25] X. Zou, J. Zhao, M. J.W. Povey, M. Holmes, and H. Mao 2010 *Anal. Chim. Acta* vol 667 (Amsterdam: Elsevier) p 14
[26] W. Windig, and J. Guilment 1991 *Anal. Chem.* vol 63 (Washington: ACS) p 1425
[27] S. Ye, D. Wang, and S. Min 2008 *Anal. Chem.* vol 91 (Washington: ACS) p 194
[28] D. Wu, Y. He, P. Nie, F. Cao, and Y. Bao 2010 *Anal. Chem.* vol 659 (Washington: ACS) p 229
[29] K. D. T. M. Milaneza, T. C. A. Nóbrega, D. S. Nascimento, R. K. H. Galvão, and M. J. C. Pontes 2017 *Anal. Chim. Acta* vol 984 (Amsterdam: Elsevier) p 76
[30] F. A. Honorato, R. K. H. Galvão, M. F. Pimentel, B. de B. Neto, M. C. U. Araújo, and F. R. de Carvalho 2005 *Chemom. Intell. Lab. Syst.* vol 76 (Amsterdam: Elsevier) p 65
[31] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão , T. Yoneyama, H. C. Chame, and V. Visani 2001 *Chemom. Intell. Lab. Syst.* vol 57 (Amsterdam: Elsevier) p 65