Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 5, No. 3, 2023

Study on Nitrite Nitrogen Based on Ultraviolet Visible Absorption Spectrometry

Li Zhang, Yitong Yin and Jinrui Zeng

School of CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS, Chongqing 400065, China

Abstract: In order to improve the accuracy of nitrite nitrogen detection, this template proposed a method for rapid determination of nitrite nitrogen in water based on ultraviolet visible absorption spectroscopy. The experimental object is the absorption spectrum of sodium nitrite standard solution with a concentration range of 0.1 to 17 $mg \cdot L^{-1}$, the continuous projection algorithm SPA (Continuous Projections Algorithm) is used to screen out the characteristic wavelengths related to nitrite nitrogen. the absorbance at the characteristic wavelengths and the sample concentration are fitted using Support Vector Regression (SVR) to establish a regression model for nitrite nitrogen, The decision coefficient R^2 and root mean square error RMSE are used as the evaluation indicators of the model. The experiment found that the R^2 and RMSE of the mixed prediction model established using the continuous projection algorithm support vector regression (SPA-SVR) modeling method were 0.999654 and 0.000479 $mg \cdot L^{-1}$ respectively, and their modeling effects were better than those of the three mixed prediction models, KPCA-SVR, PCA-SVR, and Lasso-SVR, achieving rapid and accurate measurement of nitrite nitrogen.

Keywords: Nitrite nitrogen, Absorption spectrum, Water quality detecting.

1. Introduction

With the continuous improvement of contemporary human living standards and the rapid development of society, the problem of water pollution is becoming increasingly serious, and its detection and remediation has become a social hotspot. When the nitrogen content in the water body is too high, eutrophication will occur in the water body, resulting in excessive algae, thereby reducing the dissolved oxygen content in the water[1]. When excessive nitrite enters the human body, it can harm human organs and ultimately cause serious damage[2]. Therefore, it is necessary to quickly and accurately detect nitrite nitrogen in water.

For the detection of nitrite nitrogen, traditional methods mainly include chromatography, chemiluminescence, spectrophotometry, electrochemical methods, etc.[3]. Traditional methods have shortcomings such as long experimental cycles, harsh experimental conditions, and environmental damage.

Compared to traditional detection methods, spectral methods have the advantages of simple operation, continuous detection, and no pollution. Therefore, this method has been a hot spot in the field of water detection since its inception [4]. WANG[5] et al. determined nitrite nitrogen in water through second derivative analysis of UV absorption spectrum, but its standard color scale is unstable and its sensitivity is low. Li Qingbo[6] and others used partial least square regression to detect the content of nitrite in surface water, but the sample size is too small, and the relative error of low concentration is large. UUSHEIMO[7] et al. used ultraviolet visible spectroscopic sensors with two different optical paths of 5 nm and 35 nm to detect cold water dissolved in nitrate nitrogen, nitrite nitrogen, and organic carbon. Due to insufficient use of useful spectral information, measurement accuracy is limited.

In order to improve the detection accuracy of nitrite nitrogen in water, a hybrid prediction model based on UV visible absorption spectroscopy, Continuous Projections Algorithm Support Vector Regression (SPA-SVR), is proposed in this paper. First, perform SG filtering on the original spectrum, then select the characteristic wavelength of the spectral data using the Continuous Projections Algorithm (SPA), and then use the Support Vector Regression (SVR) algorithm to establish a regression model to analyze a series of concentrations of nitrite nitrogen solutions, achieving rapid and real-time detection of nitrite nitrogen.

2. Experiment

2.1. Experimental Device

An experimental platform for ultraviolet visible absorption spectroscopy was established, mainly including light sources, sample cells, fiber optic collimators, spectrometers, and computers, as shown in Figure 1. The deuterium halide lamp (Wenyi Optoelectronics, DH-mini compact type) is used as the light source for the UV visible absorption spectrum. The light emitted by the light source is coupled through optical fibers, and then emitted to the sample pool through an optical fiber collimator (Wenyi Optoelectronics, 74UV). The generated output light is collected through the collimator, and received by a spectrometer (Shared Optics, PG2000 pro) through optical fibers. Finally, the collected data is stored, processed, and modeled on the computer side.

2.2. Sample Preparation

This article uses the standard substance for nitrite nitrogen solution (1000 $mg \cdot L^{-1}$, Northern Weiye) as an experimental sample, and prepares a series of concentration gradient solutions through gradual dilution. The concentration range is 0.1~17 $mg \cdot L^{-1}$, and in the range of 0.1~2 $mg \cdot L^{-1}$, the concentration gradient is 0.1 $mg \cdot L^{-1}$; In the range of 2-5 $mg \cdot L^{-1}$, the concentration gradient is 0.25 $mg \cdot L^{-1}$; In the range of 5 to 17 $mg \cdot L^{-1}$, the concentration gradient is 1 $mg \cdot L^{-1}$. Using the experimental device built in the laboratory, set the spectrometer integration time to 45 ms, the average number of times is 11, and the smoothness is 4. Complete the collection of the absorption spectrum of nitrite nitrogen solution, and the obtained absorption spectrum of nitrite nitrogen solution is shown in Figure 2.

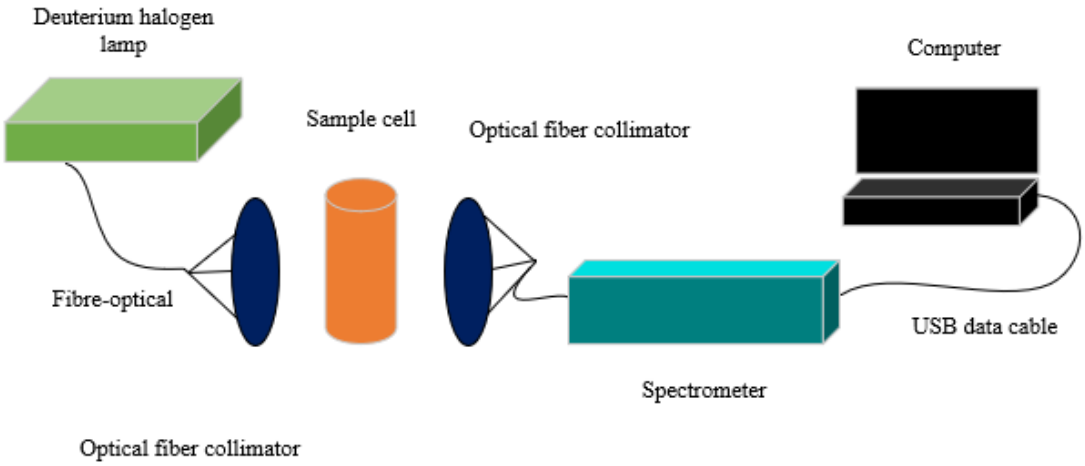

Figure 1. Experimental device diagram

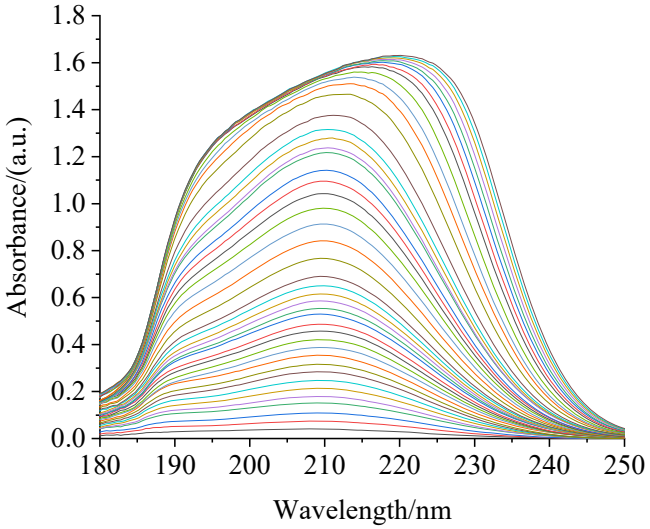

Figure 2. Ultraviolet absorption spectrogram of nitrite nitrogen standard solution

As can be seen from Figure 2, the absorption signal of nitrite nitrogen solution is mainly concentrated in the range of 190 to 250 nm. As the concentration increases, its absorbance increases, and the absorption peak position exhibits a red shift phenomenon. After 8 $mg \cdot L^{-1}$, as the concentration increases, the absorbance increases slowly and a saturation state begins to appear.

3. Data Processing

The spectral data processing process is shown in Figure 3. Firstly, the collected spectral data is subjected to SG filtering processing to remove various noises associated with the original spectrum. Then, SPA is used to extract the characteristic wavelengths to obtain three characteristic wavelengths, while effectively eliminating redundant data information. Then, based on the cross validation of the retention method, a regression model for nitrite nitrogen was established using SVR. Finally, the regression model was evaluated through the determination coefficient R^2 and root mean square error RMSE.

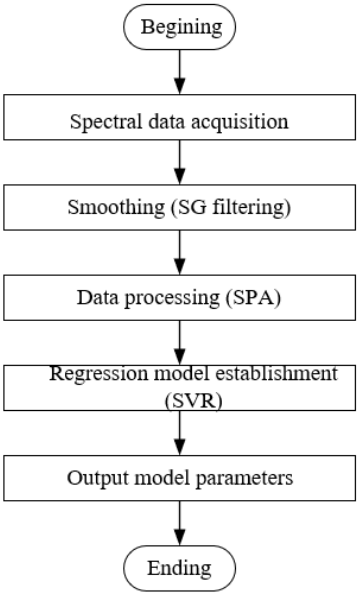

Figure 3. Data processing flow chart

3.1. SG Filtering

The collected spectral data is usually accompanied by high-frequency noise and random noise, so it is necessary to conduct noise reduction processing. There are many processing methods for noise reduction of spectral data, among which the most common and effective method is smooth filtering (Savitzky Golay, SG). The biggest advantage of this method is that it can maintain the trend and characteristics of the original spectrum while eliminating noise, and can achieve good results even in the case of large noise.

When using SG filtering, it is necessary to select the appropriate window size and polynomial order. Larger windows can better smooth the signal, but may lose details, especially for rapidly changing signals, while smaller windows can better preserve the details of the signal, but may generate noise or spurious peaks. Typically, the selection window size depends on the number and width of peaks in the data, as well as the level of smoothing required. From Figure 2, it can be seen that the original spectrum of nitrite nitrogen solution has fewer burrs. Therefore, the window size set for this experiment is 5, and the order of the polynomial is 3. The filtered absorption spectrum is obtained as shown in Figure 4. As can be seen from Figure 4, after SG smoothing filtering, the noise in the original spectrum has been effectively suppressed, indicating that using SG filtering as a preprocessing method is effective.

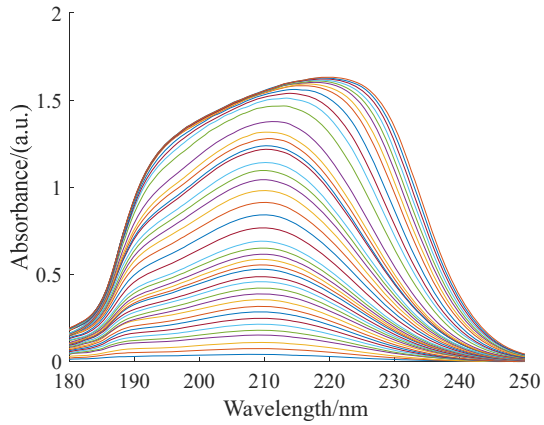

Figure 4. Absorption spectrum after SG filtering

3.2. Selection of Characteristic Wavelength

The collected spectral data are all high-dimensional data, including both useful spectral information and redundant and useless information. If not processed, it will affect the subsequent modeling effect, so dimensionality reduction processing is required. This paper uses the Continuous Projections Algorithm (SPA) to reduce the dimension of spectral data. SPA determines the most representative characteristic wavelength by projecting the vector at each wavelength point onto the vector at other wavelength points and comparing the size of the projection vector. These characteristic wavelengths have strong explanatory power and can represent all useful information in the spectrum. Extracting these characteristic wavelengths using SPA algorithm can significantly improve the performance and running speed of the model. At the same time, the SPA algorithm can also be optimized based on the correction model to obtain more accurate characteristic wavelengths.

The final selected characteristic wavelengths in this article are 190.46 nm, 203.39 nm, and 215.82 nm, as shown in Figure 5.

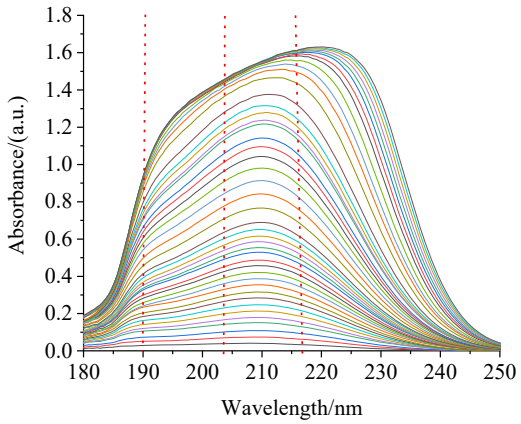

Figure 5. Characteristic wavelength points extracted by SPA

3.3. Establishment of Regression Model

Support Vector Regression (SVR) has advantages in solving small sample, nonlinear, and high-dimensional problems[8]. Unlike traditional linear or polynomial regression, SVR models nonlinear relationships by mapping data into a high-dimensional space and fitting in that space[9].

In SVR, we first need to determine a kernel function that maps low dimensional data into high dimensional space. Common kernel functions include linear, polynomial, and radial basis functions. Next, search for an optimal hyperplane in a high-dimensional space to minimize the sum of distances from all data points to the hyperplane. This hyperplane is called the "maximum boundary hyperplane.".

Similar to classification issues, SVR also needs to consider balancing within the error range. Therefore, in addition to the maximum boundary hyperplane, two other parallel hyperplanes need to be considered. These three hyperplanes identify a region called a "support vector," which means that all sample points located within the region are considered vectors that play an important role in model construction. During model training, SVR will maximize the distance between support vectors as much as possible to obtain better generalization capabilities.

The characteristic of SVR is that it is not only suitable for linearly separable datasets, but also capable of processing non-linear datasets. In addition, SVR has a strong generalization ability and can have a certain tolerance for noise.

In this article, any one of the training data is selected as a test set, and the remaining data is used as a training set for cross validation. Repeat until all the training data have been selected as a test set. The parameter with the smallest mean square error is input into the SVR model.

3.4. Model Evaluation Indicators

The commonly used indicators for evaluating models include the determination coefficient R^2, the root mean square error RMSE, and the correlation coefficient r[10].

In this paper, the determination coefficient R^2 and the root mean square error RMSE are used to evaluate the performance of the model. The expressions for R2 and RMSE are shown in Equations (1) and (2).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(p_i-\bar{p})*(o_i-\bar{o})}{\sqrt{\sum_{i=1}^{n}(p_i-\bar{p})^2+\sum_{i=1}^{n}(o_i-\bar{o})^2}}\right)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i-o_i)^2}{n}}. \quad (2)$$

In the formula, $n, o_i, \bar{o}, \bar{p}, p_i$ are respectively the number of samples, sample concentration, mean value of sample concentration, sample predicted value, and mean value of sample predicted value.

The closer R^2 is to 1, the better the fitting effect of the nitrite nitrogen solution is. The closer RMSE is to 0, the smaller the error between the actual value of the nitrite nitrogen solution in the water and the predicted value of the model is, indicating that the prediction effect of the model is better.

4. Results and Discussion

After establishing a SPA-SVR mixed prediction model, the prediction results of spectral data under this model are shown in Table 1. Table 1 lists the true values, predicted values, and relative errors of the samples. From the table, it can be seen that when the sample concentration is 0.1 mg·L^{-1}, the relative error is 7.67%, not more than 10%, and the relative error is within 5% at other concentrations. This indicates that the modeling method used in this article is correct, and the applicability of the model is high, enabling accurate measurement of nitrite nitrogen

Table 1. Predicted values of nitrite nitrogen SPA-SVR mixed model

Sample number	Predicted value/(mg·L^{-1})	True value/(mg·L^{-1})	Relative error/%
1	0.10	0.1	7.67
2	0.20	0.2	0.73
3	0.29	0.3	1.29
4	0.41	0.4	3.99
5	0.49	0.5	0.96
6	0.59	0.6	0.80
……	……	……	……
43	16.05	16	0.37
44	16.86	17	0.80

In order to verify the effectiveness of the prediction model, three other models, namely KPCA-SVR, PCA-SVR, and LASSO-SVR, were established and compared with the above SPA-SVR model. The specific situation is shown in Figure 6.

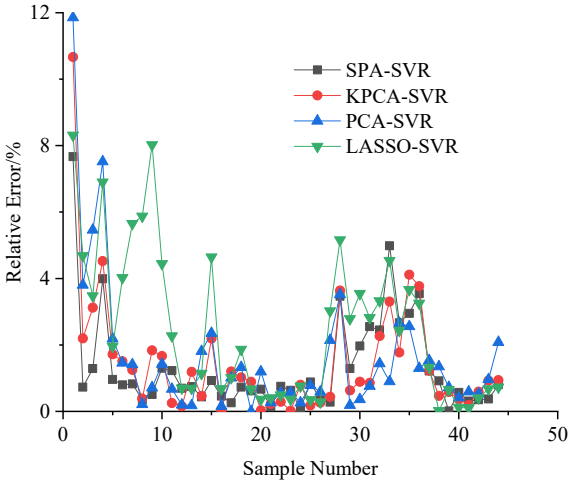

Figure 6. Relative error diagram of different models

From Figure 6, it can be seen that the SPA-SVR model has the smallest fluctuations, and the average relative errors of the four models are 1.33%, 1.93%, 3.00%, and 3.21%, respectively. Therefore, the SPA-SVR model is superior to the other three models.

Table 2 compares these four models in terms of R^2 and RMSE parameters. The R^2 of the SPA-SVR model is 0.999654, and the RMSE is 0.000479 mg·L^{-1}. Compared to the other three models, the R^2 has increased, while the RMSE has decreased, indicating that the model is superior to the other three models.

Table 2. Comparison of evaluation parameters of different analysis models

Modeling methods	Model evaluation	
	R^2	RMSE/(mg·L^{-1})
SPA-SVR	0.999654	0.000479
KPCA-SVR	0.999604	0.004186
PCA-SVR	0.999596	0.004469
LASSO-SVR	0.999504	0.004798

In order to verify the impact of this mixed prediction model on the stability of measured nitrite nitrogen, under this optimal model, eight measurements were conducted on a nitrite nitrogen standard solution with a concentration of 10 $mg \cdot L^{-1}$. The measured results are shown in Table 3:

Table 3. Measured values of nitrite nitrogen solution at 10 $mg \cdot L^{-1}$ for multiple tests

Sample number	Predicted value/($mg \cdot L^{-1}$)	True value/($mg \cdot L^{-1}$)	Relative error/%
1	10.75	10.00	7.5
2	10.75	10.00	7.6
3	10.74	10.00	7.4
4	10.75	10.00	7.5
5	10.77	10.00	7.7
6	10.79	10.00	8.0
7	10.78	10.00	7.9
8	10.80	10.00	8.1
Average value	10.77		7.7

As can be seen from Table 3, the relative error between the single measurement value and the actual value is within 10%, and the average value of the relative error for eight measurements is 7.7%, indicating that using the SPA-SVR mixed prediction model to measure nitrite nitrogen solution has good stability.

5. Summary

This paper proposes a SPA-SVR mixing research model based on nitrite nitrogen standard solution, and compares it with three other mixing models: KPCA-SVR, PCA-SVR, and LASSO-SVR. The research results show that the decision coefficient R^2 of the SPA-SVR hybrid prediction model is 0.999654, which is 0.0050%, 0.0058%, and 0.0150% higher than the other three hybrid models, respectively; The root mean square error (RMSE) is 0.0004798 $mg \cdot L^{-1}$, which is 88.54%, 89.26%, and 90% lower than the other three models, respectively. In addition, based on the mixed prediction model, the measurement stability of nitrite is 7.7%, which indicates that the SPA-SVR mixed model proposed in this article can provide a new solution for rapid and pollution-free monitoring of nitrite in water.

Acknowledgment

National Natural Science Foundation of China (61805030), Chongqing Basic and Frontier Technology Research Project (cstc2020jcyj-msxmX0147), Chongqing Education Commission Science and Technology Project (KJQN202000640, KJZD-M202200602)

References

[1] LU H, PENG M, ZHANG G, et al. Biokinetic and biotransformation of nitrogen during photosynthetic becteria wastewater treatment[J]. Environmental Technology, 2018, 41(15): 1-28.

[2] WANG C, WANG B, JI T, et al. Simulated estimation of nitrite content in water using transmission spectroscopy[J] Spectroscopy and Spectral Analysis, 2022, 42(7): 2181-2186.

[3] ALAHI M, MUKHOPADHYAY S. Detection methods of nitrate in water: A review[J]. Sensors and Actuators: A. Physical, 2018, 280: 210-221.

[4] SINGH P, SINGH M K, BEG Y R, et al. A review on spectroscopic methods for determination of nitrite and nitrate in environmental samples[J]. Talanta, 2019, 4(11): 364-381.

[5] WANG J, ZHANG J, ZHANG Z. Rapid determination of nitrate nitrogen and nitrite nitrogen by second derivative spectrophotometry[J]. Spectroscopy and Spectral Analysis, 2019, 39(1): 161-165.

[6] LI Q, HE L, CUI H, et al. Basic Research on Ultraviolet Visible Spectral Detection Method for Nitrite Nitrogen Concentration in Surface Water [J]. Spectroscopy and Spectral Analysis 2020, 40(04): 1127-1131.

[7] UUSHEIMO S, TULONEN T, ARVO-LA L, et al. Organic carbon causes interference with nitrate and nitrite measurements by UV/Vis spectrometers: the importance of local calibration[J]. Environmental Monitoring and Assessment, 2017, 189(7): 357.

[8] THARWAT A. Parameter investigation of support vector machine classifier with kernel functions[J]. Knowledge and Information Systems, 2019, 61(3): 1269-1302.

[9] KANG W. Research on the detection of pollutants in water based on SERS technology and GWO-SVR algorithm [D] Hebei: Yanshan University, 2021.

[10] ZHENG H, YAN Z, CHEN J, et al. Seasonal variations of dissolved organic matter in the east China sea using EEM-PARAFAC and implications for carbon and nutrient cycling[J]. Sustainability, 2018, 10(5): 1444.