Visible/near infrared spectroscopy method applied research in wetland soil nutrients rapid test

G X Ren\textsuperscript{1,2}, Z Q Wei\textsuperscript{1,4}, P P Fan\textsuperscript{2} and X Y Wang\textsuperscript{3}

\textsuperscript{1}College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China
\textsuperscript{2}Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao 266100, China
\textsuperscript{3}Qingdao Ke Rui Tong Science and Technology Consulting Co., Ltd., Qingdao 266100, China

E-mail: weizhiqangouc@126.com

Abstract. Visible/near infrared spectrum method has the advantages of fast and green, there is much better technical superiority in soil nutrients on-site detection. Studying on wetland soil samples from 4 districts of Qingdao with higher heterogeneity, according to its total nitrogen (TN), available nitrogen (AN), total phosphorus (TP), available phosphorus (AP), total potassium (TK), slowly available potassium (SK) and available potassium (FK) content and visible/near infrared reflectance spectroscopy, adopted SPA and PLSR to choose characteristic wavelength and build mode. Modeling results and experimental data indicated that this method to the content of TN, AN, TP, AP and FK in wetland, can be predicted accurately, it also can be predicted the content of TK and SK roughly.

1. Introduction

Nitrogen, phosphorus and potassium are the main nutrients in wetland soil, play an important role in promoting plant growth, increasing throughput, and they are the important index for evaluating soil fertility. But these nutrient elements in different regions field show the wide disparity due to environmental change, human influence or variety of plants coverage. Even in the same area soil, it also presents different nutrients content as time going [1]. So accurate analysis to know the content of nitrogen, phosphorus, potassium in time, in order to provide basis for nutrient analysis of wetland soil and scientific management, this is a matter of great significance.

Currently, detection methods of N, P, K in wetland soil are mainly traditional laboratory method and based colorimetric detection methods. Laboratory detection methods require complex pre-treatment of the soil, moreover, for different nutrients, it is necessary to use different methods for separate operation detection. The analysis process is lengthy and complicated [2]; compared with the laboratory method, the colorimetric method is more convenience, but it also need a large amount of pre-processing, much longer detection time (More than 2 hours), beside, it needs to undertake the risk of secondary pollution from chemical agents. Visible/near infrared reflectance spectrum functional groups that reflect the different characteristics and quantities of NPK and its active constituents in the soil. Spectral analysis technology is a fast, simple, non-destructive green analysis technology with fast analysis, simultaneous determination of multiple parameters, samples do not require pre treatment or simply require pre-treatment and easy to operation etc. It can meet on-site, real-time testing
requirements, then become a research hotpot [3-5]. This article stated to use visible/near infrared spectroscopy method, four different wetland soils in different areas of Qingdao were collected as research objects. Based on their N, P, K content and reflectance spectra, application of Successive Projection Algorithm (SPA) to choose TN, AN, TP, AP, TK, SK and FK reflected spectral characteristic wavelength, establishing an analytical model by partial least squares regression, to achieve the real-time detection of nitrogen, phosphorus and potassium in freshly collected wetland soils.

2. Sample collection and data acquisition

Based on existing research, the much greater difference of nutrients in soil samples, the better effect of the established model [6]. Therefore, in this project, four regional wet spots were sampled from Qingdao Jiaozhou city, Chengyang district, Laoshan district and Licang district. Select one region with uniform colour and vegetation coverage at each sampling point, keeping soil nutrients are uniform at each sampling point, the basic overview of each sampling point is shown in table 1.

| No. | Area   | Place         | Longitude   | Latitude   | Vegetation                  |
|-----|--------|---------------|-------------|------------|-----------------------------|
| S1  | Jiaozhou| Shaohai wetland| E120°4'35"  | N36°14'53" | Poorly growing land         |
| S2  | Chengyang| Nugukou wetland| E120°19'16" | N36°14'85" | Weed                       |
| S3  | Laoshan | Reservoir     | E120°30'19" | N36°15'29" | Reed and Polygonaceae      |
| S4  | Licang | Licun River   | E120°26'30" | N36°9'26"  | Weed                       |

When taking sample, scooping 0-20 cm surface soil with a shovel, around 1 kg per soil sample, put them in the sealed bag, sticking the label, then brought back to the lab, grinding after air drying, passed the filter of 2 mm, 0.5 mm and 0.1 mm respectively, put it in sealed bag and mark to be used for later testing.

Sample total nitrogen content using elemental analyzer (Germany Perkin 2400 determination), the alkali nitrogen content of the sample was determined by sodium hydroxide indirect diffusion method [7]. Determination of total phosphorus content in samples by alkali solution-molybdenum antimony Resistance Spectrophotometry [8], determination of available phosphorus content by sodium carbonate extraction-Molybdenum Antimony Anti-colorimetric method [9]. Determination of total potassium in samples by atomic absorption spectrophotometry [10], determination of available potassium in samples by neutral ammonium acetate-atomic absorption spectrophotometry, determination of slowly available potassium in samples by atomic absorption spectrophotometry [11]. The results are shown in table 2.

| soil nutrients | Average Value | Variance |
|---------------|--------------|----------|
| TN(g/kg)      | 0.78         | 0.00     |
| AN(mg/kg)     | 66.7         | 3.0      |
| TP(g/kg)      | 0.51         | 0.03     |
| AP(mg/kg)     | 19.9         | 1.4      |
| TK(g/kg)      | 16.3         | 0.1      |
| SK(mg/kg)     | 246.2        | 12.3     |
| FK(mg/kg)     | 70.6         | 2.8      |

The reflection spectroscopy of the soil sample is obtained by using the prototype designed in the study, the prototype applied to Ocean Optics' DH-2000 as the light source, acquisition of spectra by QE-65000 spectrometer (slit 10 um) connected to Ocean Optics with Y-mode fiber. Spectral sampling interval is set to 1 nm, integration time is set to 600ms, visible/near infrared spectroscopy with a spectral range of 200-1100 nm, eliminating the front and back of the spectrum with loud noise, the
obtained soil sample reflection spectrum is shown in figure 1.

![Sample visible/near infrared reflected spectrum](image)

**Figure 1.** Sample visible/near infrared reflected spectrum.

3. Extracting characteristic wavelength and establishing an analytical model

In order to minimize the interference factors, the full-band spectral data of all 120 soil samples will be modeled to verify the effect of several pre-processing methods, as shown in table 3. According to the results in table 3, utilizing the multiple scattering correction (MSC) method of spectral data processing to TN, AP, TK, SK and FK; utilizing the standard normal variable (SNV) method of spectral data processing to AN; utilizing the Gaussian filter smoothing method of spectral data processing to TP; utilizing S-G derivative method of spectral data processing to FK. Making use of robust principal component analysis to remove the abnormal samples; adopting Kennard-Stone algorithm as 2:1 to classify these 120 samples [9], to choose representative integrated modelling and prediction set.

|                | S-G derivative | Gaussian filter | MSC     | SNV     |
|----------------|----------------|-----------------|---------|---------|
|                | $R^2$          | RMSE            | $R^2$   | RMSE    | $R^2$ | RMSE |
| TN             | 0.96           | 0.09            | 0.98    | 0.07    | 0.99 | 0.05 | 0.98 | 0.07 |
| AN             | 0.94           | 5.83            | 0.93    | 6.01    | 0.96 | 5.49 | 0.97 | 5.27 |
| TP             | 0.91           | 0.22            | 0.98    | 0.05    | 0.97 | 0.13 | 0.98 | 0.07 |
| AP             | 0.93           | 28.4            | 0.95    | 24.7    | 0.98 | 24.0 | 0.95 | 25.6 |
| TK             | 0.97           | 0.71            | 0.97    | 0.78    | 0.98 | 0.64 | 0.94 | 0.81 |
| SK             | 0.95           | 25.4            | 0.97    | 27.3    | 0.98 | 21.5 | 0.97 | 21.7 |
| FK             | 0.98           | 4.57            | 0.91    | 6.28    | 0.96 | 4.88 | 0.95 | 4.72 |

3.1. Model evaluation parameter

Determination coefficient $R^2$ is one of the criteria for judging the correlation between quantitative models, the more $R^2$ approximates 1, the better the model effect will be.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

(1)

Inside, $n$ is calibration set (or prediction set) samples number, $y_i$ is calibration set (or prediction
set) the No. \( i \) sample’s actual value, \( y_i \) is calibration set (or prediction set) the No. \( i \) sample’s predictive value, \( \hat{y}_i \) is calibration set (or prediction set) all actual values average.

**RMSE** means actual error, it is that measured value subtract predictive value squared differences with sample number subtract 1’s quotient square root.

\[
RMSE = \sqrt{\frac{1}{n_c-1} \sum_{i=1}^{c} (y_i - \hat{y}_i)^2}
\]  

(2)

Inside, \( n \) is calibration set (or prediction set) samples number, \( y_i \) is calibration set (or prediction set) the No. \( i \) sample’s actual value, \( \hat{y}_i \) is calibration set (or prediction set) the No.\( i \) sample’s predictive value, \( \overline{y}_i \) is calibration set (or prediction set) all actual values average.

**RPD** is the ration of sample standard deviation and prediction standard error.

\[
RPD = \frac{SD}{RMSEP} = \frac{\sqrt{\frac{1}{n_c-1} \sum_{i=1}^{c} (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n_d-1} \sum_{i=1}^{d} (y_i - \overline{y}_i)^2}}
\]

(3)

Inside, \( n \) is calibration set (or prediction set) samples number, \( y_i \) is calibration set (or prediction set) the No. \( i \) sample’s actual value, \( \hat{y}_i \) is calibration set (or prediction set) the No. \( i \) sample’s predictive value, \( \overline{y}_i \) is calibration set (or prediction set) all actual values average.

**RPD** and model quality quantitative relationship is shown in table 4 [10].

| **RPD value** | **Model quality evaluation** |
|---------------|-----------------------------|
| **RPD**<1.0    | Very Inferior model or prediction, Not recommended |
| 1.0<**RPD**<1.4 | Inferior model or prediction, make a distinction only in high and low value |
| 1.4<**RPD**<1.8 | Model or prediction is clearly, it can be evaluated and associated |
| 1.8<**RPD**<2.0 | Good model has the possibility of quantitative prediction |
| 2.0<**RPD**<2.5 | Very good model, it can be quantitative prediction |
| **RPD**>2.5    | Better model, it can be good quantitative prediction |

3.2. **Successive projection algorithm (SPA) optimization characteristic wavelength**

SPA algorithm uses vector projection analysis to find the minimally redundant information variable, it can overcome spectral collinear data and redundancy effectively, reduce modelling variable and modelling complexity [11,12].

To reduce the spectral range and avoid the interference of redundant information, first, adopting correlation coefficient spectrum to get a region with strong correlation with the component to be tested, as shown in figure 2. The high correlation coefficient spectrum of TN, AN, TP, AP and TK located at 250-600 nm, but in this wave band, only the correlation coefficient of total phosphorus less than 0.5; FK has more higher correlation coefficient (0.5-0.8) in the whole wave band, the most high correlation coefficient located in 250-350 nm and 500-650 nm; SK has more lower correlation
coefficient (less than 0.3) in whole wave band, among them, the most high wave band is in 226-250 nm and 400-500 nm.

![Correlation coefficient spectrum (wavelength).](image)

Figure 2. Correlation coefficient spectrum (wavelength).

After get the region which strong correlation with the component to be tested, via correlation coefficient spectrum, using SPA for optimizing characteristic wavelength to various soil nutrients, the results are shown in table 5.

| Nutrient | $R^2$ | RMSE | Characteristic wavelength |
|----------|-------|------|---------------------------|
| TN       | 0.83  | 0.19 | 11(545/575/488/472/325/270/289/342/290/266) |
| AN       | 0.77  | 15.74| 11(274/375/347/257/269/270/281/245/488/275/487) |
| TP       | 0.79  | 0.16 | 7(282/488/560/325/311/487/249) |
| AP       | 0.75  | 7.55 | 7(301/282/325/254/263/488/550) |
| TK       | 0.33  | 4.09 | 7(328/249/560/371/299/240/488) |
| SK       | 0.12  | 129.23| 4(487/522/325/302) |
| FK       | 0.75  | 15.83| 7(605/491/273/659/655/240/342) |

Table 5. Optimization wavelength by SPA.

3.3. Partial least squares regression (PLSR) build analysis model
PLSR will increase spectral variable step by step and extract the main component, build the regression model then examine its prominence till accepted, it is one main way to build model [13,14].

The result of building model is shown in table 6, although correlation coefficient is not high ($R^2$<0.9), but prediction effect of TN, TP, AP, FK is well. $RPD$ is higher than 2; the prediction effect of AN is very well too, $RPD$ is up to 1.91. Only the prediction effect of TK and SK is ordinary, $RPD$ is lower than 1.5.

| Soil nutrients | Calibration set | Prediction set |
|----------------|-----------------|----------------|
|                | $R^2$ | RMSEC | $R^2$ | RMSEP | RPD  |
| TN             | 0.83  | 0.18  | 0.85  | 0.21  | 2.40 |
| AN             | 0.78  | 15.55 | 0.76  | 17.24 | 1.90 |
| TP             | 0.80  | 0.16  | 0.82  | 0.15  | 2.33 |
| AP             | 0.72  | 8.06  | 0.83  | 6.19  | 2.37 |
| TK             | 0.30  | 4.23  | 0.41  | 3.80  | 1.49 |
| SK             | 0.14  | 127.72| 0.22  | 124.18| 1.40 |
| FK             | 0.74  | 16.27 | 0.80  | 13.94 | 2.20 |
From the table above, we know that the related parameters modelling effect of TK and SK is ordinary. Analyzing the reasons, one is larger noise, it influenced characteristic spectrum; third is systematic error. Therefore further research is needed on this.

4. Experimental results and discussion
According to the above method, we built the principle prototype on the basis of pro phase studying. Using the principle prototype to test 5 standard soil samples (National Institute of Metrology, China) TN, AN and AP, the test results are compared with the results of the laboratory national standard test, the results are shown in table 7.

| Sample | TN (g/kg) | AN(mg/kg) | AP(mg/kg) |
|--------|-----------|------------|-----------|
|        | Lab       | Prototype  | Error %   | Lab       | Prototype  | Error %   | Lab       | Prototype  | Error %   |
| W1     | 0.63      | 0.66       | 4.7       | 57.0      | 5938       | 4.8       | 100.0     | 95.9       | -4.1      |
| W2     | 0.77      | 0.73       | -4.3      | 76.0      | 79.2       | 4.2       | 23.3      | 22.3       | -4.5      |
| W3     | 1.07      | 1.13       | 5.1       | 97.0      | 101.4      | 4.2       | 29.0      | 30.3       | 4.7       |
| W4     | 1.97      | 1.87       | -4.9      | 165.0     | 173.0      | 4.8       | 1.5       | 1.5        | 4.8       |
| W5     | 0.54      | 0.52       | -4.6      | 44.0      | 41.8       | -5.0      | 0.5       | 0.6        | 5.2       |

The experiment examined three nutrient nutrients: total nitrogen, alkali nitrogen and available phosphorus in the modelling. The prototype test results are in good agreement with the laboratory test results. The detection error is mostly below ±5.0%. Only one is more than 5.0%. It is indicated that the rapid measurement method based on visible/near infrared reflectance spectra established in the study is feasible.

The method established in the study, the prototype further optimized in terms of hardware, real-time detection of nutrients such as nitrogen, phosphorus and potassium in the collected wetland soil samples can be achieved on site. Improve the detection efficiency of soil nutrients, at the same time, it can provide technical support for the establishment of a large database of wetland soils in China.

5. Conclusions
Studying on wetland soils with high heterogeneity in Qingdao, via visible/near infrared spectroscopy, Adopted SPA to choose characteristic wavelength and Partial least Squares Regression (PLSR) build model, established a method for rapid detection of nutrients in wetland soil. Compared to traditional laboratory methods and colorimetric soil velocity testing, it has the advantages of on-site, fast and green. The experimental results show that the method can be used to quickly and accurately detect soil nutrients in wetland field. It plays an active role in monitoring the information of wetland soil nutrients timely, planning and protecting wetland resources rationally in China.

Acknowledgments
Authors acknowledge funding from Natural Science Foundation of Shandong Province (ZR2018LD007) and Shandong Key R & D Project (2015GNC110016).

References
[1] Li X Y, Fan P P, Hou G L, Lv M R and Wang Q 2017 Spectrosc. Spect. Anal. 37 3562-6
[2] Liu B, Zhou F and Wang G X 2011 Acta. Ecol. Sin. 31 6947-58
[3] Li Z M, Zheng L H and An X F 2013 Trans. Chin. Soc. Agric. Mach. 44 73-87
[4] Rossel R A V and Webster R 2012 Eur. J. Soil Sci. 63 848-60
[5] Iznaga A C, Orozco M R and Alcantara E A 2014 Bio. Syst. Eng. 125 105-21
[6] Liu X M and Liu J S 2013 Trans. Chin. Soc. Agric. Mach. 44 88-91
[7] Fan P P, Li X Y, Lv M R, Wu N and Liu Y 2018 Spectrosc. Spect. Anal. 38 3210-14
[8] Wang R J, Chen T J, Wang Y B and Wang L S 2017 *Chin. J. Lumi.* **38** 109-16
[9] Galvao R K, Araujo M C and Jose G E 2005 *Talanta.* **67** 736-40
[10] Rossel R A V, Mcglynn R N and Mcbratney A B 2006 *Geoderma.* **137** 70-82
[11] Zhang H L, Luo W and Liu X M 2017 *Spectrosc. Spect. Anal.* **37** 584-7
[12] Gao H Z and Lu Q P 2011 *Spectrosc. Spect. Anal.* **31** 1246-9
[13] Bao C F, Liu T and Wang B 2009 *J. Jilin Uni. (Sci.)* **47** 362-6
[14] Zhang X M and Tang N 2018 *Mod. Elec. Tech.* **41** 126-9