Feasibility Study of Soil Quality Survey using Visible and Near Infrared Spectroscopy in Rice Paddy Fields in China

Hongyi Li

1 School of Tourism and Urban Management, Jiangxi University of Finance and Economics, Nanchang, 330013, China
2 College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China
Tel.: +86 571 86038602, fax: +86 571 86038602
E-mail: lihongyi1981@zju.edu.cn

Abstract: Survey and monitoring of soil quality are needed to prevent soil degradation and are important for sustainable farming and food production. Conventional soil survey involves intensive soil sampling and laboratory analysis, which are time consuming and expensive. Visible and near infrared spectroscopy of soil has proved to be accurate, cheap and robust and has huge potential for survey of soil quality. To test its potential, 327 soil samples were taken from long-term paddy rice fields in four provinces in south of China and covered a wide range of soil types and texture. The samples were air-dried, ground and passed through a 2 mm sieve. They were then scanned by an ASD vis–NIR spectrometer with wavelength range from 350 to 2500 nm. Organic matter (OM), pH, total nitrogen (TN) and available nitrogen (N_av) were also measured on soil samples to build calibration models and also to validate the models’ accuracy. On the basis of the ratio of prediction deviation (RPD), which is standard deviation (SD) of prediction divided by the root mean square error of prediction (RMSEP), the accuracy of leave-one-out cross-validation of soil N_av model was classified very good (RPD=1.96) and soil OM and TN was good (RPD=1.78 and RPD=1.81, respectively). However, the model accuracy of pH was poor due to non-direct soil spectral response for soil pH in vis–NIR spectroscopy. The independent validation results showed excellent accuracy for soil N_av (RPD=3.26), good accuracy for OM and TN (RPD=1.76 and RPD=1.78) and relative poor accuracy for soil pH (RPD=1.27). This feasibility study is encouraging for the application of vis–NIR surveys of soil quality accuracy at regional and national scales; it found good to excellent accuracy for some important soil properties in quality survey.

Keywords: Soil quality survey, Proximal soil sensor, Vis–NIR, PLSR, Rice paddy field.

1. Introduction

Rice (Oryza sativa L.) is the primary staple food for more than half of the world’s population [1]. China, the world's most populated country with more than 1.3 billion people in 2013, and the world's second largest area of rice paddy fields, accounts for 19% of cultivation area of rice in the world [2]. However, the loss of cultivated land to manufacturing industry and urbanization, especially of the rice paddy fields, has increasingly become of national concern in China. The Chinese government
acknowledges the importance of the soil quality of paddy rice fields, as well as the quantity. It also recognizes the production of fiber and the soil’s role in maintaining freshwater, its contribution to energy and climate sustainability, and that is maintains the biodiversity and the overall protection of the ecosystem [3]. So, monitoring the soil quality of paddy rice fields is very important for decision making by Chinese government.

Since the early 1990s, the concept of soil quality has received much attention. Agricultural scientists have developed quantifiable indicators of soil quality. Larson and Pierce defined a ‘Minimum Data Set (MDS)’ [4], such as nutrient availability, pH, total organic C, maximum rooting depth, texture, and five other physical and chemical indicators, which could be used to monitor soil quality. Then, Doran and Parkin expanded the MDS to include biological properties, in addition to physical and chemical properties of the soil [5]. Based on the physical, chemical, and biological indicators, soil quality was widely recognized by thousands of papers devoted to it [6-8]. In addition, much research is exclusively focused on the soil quality of rice paddy fields [9-11]. Despite the above, we find little information on assessment of soil quality at a regional or national scale. Why have pedometricians been reluctant to make such studies? One of the most important reasons is the cost and time of obtaining soil samples and laboratory analysis over such large region [12].

Sampling the soil, however, has been revolutionized by the development of proximal soil sensing (PSS), with techniques such as electromagnetic induction [13] and visible and near infrared spectroscopy (vis–NIR) [14-16]. In the recent years, the vis–NIR technique has been applied to predict soil fertility. Viscarra Rossel et al. assembled a minimum data set to develop a soil fertility index for sugarcane for bio-fuel production in tropical soils [17]. Soriano-Disla et al. applied visible (VIS), near-infrared (NIR), and mid-infrared (MIR) reflectance spectroscopy for the prediction of soil properties [18]. They successfully predicted soil water content, texture, soil carbon (C), total nitrogen (TN), and pH, and several other physical, chemical, and biological properties. But no one has applied this technique into soil of rice paddy fields exclusively at a regional scale in China.

Soil organic matter (SOM) is the most often reported attribute from long-term studies and is regarded as the most important indicator of soil quality and agronomic sustainability because of its impact on other physical, chemical and biological indicators of soil quality [19]. To obtain large crop yields, farmers apply much inorganic-N fertilizers. The total amount of N in synthetic fertilizers used in China increased from 9.43 Tg in 1980 to 24.00 Tg in 2012 [20]. However, over-application N fertilizer causes the leaching of N to lakes, rivers and ground water which are the main sources of drinking water in China. Evaluation of available N (N_av) in soil is more meaningful, because in that form N can be directly taken up by crops. However, unlike SOM and TN, N_av cannot be detected in NIR spectra of soil. Soil pH depends on the nature of the material from which it has developed. It ranges from about pH 4 (very acid), when most crops will fail, to about pH 8 for soils naturally rich in calcium carbonate (lime) or magnesium carbonate. A pH much in excess of 8 indicates the influence of exchangeable sodium. For soils with a pH less than 7, natural processes (e.g. rain, crop growth and especially leaching of calcium in drainage water) and some farming practices (e.g. use of large amounts of some nitrogen fertilizers) tend to acidify soil. Acidifying processes can cause the soil’s pH to fall quickly, and regular pH checks are advisable. Like N_av, pH also has no direct response in soil NIR spectra and there are a few papers showing acceptable soil pH predictions by NIR spectroscopy [21-22]. Such successes arise because the pH co-varies with SOM, TN and clay content in the soil.

Soil organic matter, TN, N_av and pH are important in the evaluation of soil quality. Their successful prediction from vis–NIR spectroscopy over large areas is still difficult, but if we can overcome the difficulties, the economic advantages over traditional methods are likely to be substantial. So, in this study, we try to use visible and near infrared spectroscopy to determine SOM, TN, N_av and pH in rice paddy fields in China, by analyzing the feasibility of vis–NIR as a supplementary method in soil quality assessment at a regional scale.

2. Material and Methods

2.1. Soil Samples

Soil samples were collected from four provinces, namely Zhejiang, Fujian, Sichuan and Hunan, in south of China. The samples, 327 in all, were chosen randomly from long-term cultivated rice paddy fields and taken by core sampling from 0-20 cm depth. They cover a wide range of soil types. The sampling was done in autumn after the rice had been harvested and fields were left dry for 10 days. At each sampling site, about 200 g soil was collected for laboratory analysis and optical measurement. The samples were dried at 40°C for 24 hour, ground and passed through a 2 mm sieve. Then, the samples were divided into two halves for chemical analysis and optical measurement. The pH and SOM was measured on all samples, but TN and N_av were measured on 155 and 145 samples since they are available history data, respectively.

2.2. Soil Treatment and Chemical Analysis

Soil organic matter content was determined calorimetrically after H2SO4-dichromate oxidation at 150 °C. Soil TN was measured by a TrusSpecCNS
spectrometer (LECO Corporation, St. Joseph, MI, USA) the Dumas combustion method. Soil N$_{av}$ was measured by the alkaline hydrolysis diffusion method. Soil pH was measured by a glass electrode in a 1:5 (volume basis) suspension of soil in a solution of 1 M KCl after the suspension was shaken on a side-to-side shaker (300 revolutions per minute) for 60 minutes. A summary of the statistics is in Table 1.

Table 1. Summary statistics of SOM, TN, N$_{av}$ and pH calibration and validation sets.

<table>
<thead>
<tr>
<th></th>
<th>Calibration set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Samples</td>
<td>Min.</td>
</tr>
<tr>
<td>SOM (g kg$^{-1}$)</td>
<td>313</td>
<td>3.74</td>
</tr>
<tr>
<td>TN (g kg$^{-1}$)</td>
<td>141</td>
<td>0.34</td>
</tr>
<tr>
<td>N$_{av}$ (mg kg$^{-1}$)</td>
<td>131</td>
<td>15.80</td>
</tr>
<tr>
<td>pH</td>
<td>313</td>
<td>4.56</td>
</tr>
</tbody>
</table>

2.3. Optical Measurement

A Fieldspec ProFR vis–NIR spectrometer (Analytical Spectral Device, Boulder, CO, USA) was used for the optical measurement. This spectrometer measures wavelengths from 350 nm to 2500 nm. Each soil sample was placed into a container of 10 cm radius and 1 cm depth, and the soil surface was leveled to achieve a high signal to noise ratio. Each sample was scanned 10 times to get 10 spectra which were finally averaged to one spectrum. A 100% white reference was used before scanning, and repeated every 10 samples.

2.4. Model Calibration and Validation

Scanned spectra were first cut at both ends to leave them from 400 nm to 2450 nm for removing noise. We then reduced each spectrum to 411 data points by averaging over five consecutive wavelengths, followed by maximum normalization, Savitzky–Golay first derivation, and Savitzky–Golay smoothing. This data pre-treatment reduced the noise further and smoothed the spectra.

The pre-treated spectra were divided into a calibration set and a validation set. For all the four soil properties, fourteen soil samples were randomly selected for model validation, and the others samples were used to build robust calibration models. Partial least squares regression (PLSR) was employed to build the calibration models and leave-one-out cross-validation and independent validation were used to evaluate the models’ performances. We determined the number of latent variables for a model by examining a plot of leave-one-out cross validation residual variance against the number of latent variables obtained from PLSR. The latent variable of the first minimum value of residual variance was selected [23]. All calculations were done with Unscrambler 7.8 software (Camo Inc.; Oslo, Norway).

Models’ accuracies were assessed by the standard introduced by Viscarra Rossel et al. who classified residual prediction deviation (RPD), standard deviation / root mean square error of prediction, as follows: RPD<1.0 indicates very poor performance, any model with such a RPD is not recommended; RPD between 1.0 and 1.4 indicates poor performance, and only large and small values are distinguishable; RPD between 1.4 and 1.8 indicates fair performance, which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good performance, where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates very good, quantitative performance, and RPD>2.5 indicates excellent performance [24].

3. Results and Discussion

Both cross-validation and independent validation showed acceptable accuracy for SOM (RPD=1.78 and RPD=1.76, respectively). As SOM or soil organic C is essential for soil management and carbon sequestration, they were the main properties considered for soil vis–NIR spectroscopy [25-27]. They were some significant absorbance peaks for C in both the VIS and NIR spectral ranges. Some of the literature showed good calibration for [25-26]. However, there was no independent validation in the studies described. Models calibrated with samples of soil from the field at the farm scale are usually accurate described, because those samples are from the same materials; they are face homogeneous in colour and texture. Soil colour and texture are two main factors affecting vis–NIR spectra [28]. Once the area expands from field and farm to larger regions, the accuracy deteriorates significantly [29], and we might expect it to be even worse at a continental or global scale [30]. In this study, soil samples were collected from four provinces in Southern China covering around 1 million km$^2$ with varied climate
and geographical conditions, which make prediction from vis–NIR spectroscopy less accurate.

Table 2. Summary of model cross-validation independent validation results for SOM, TN, N_av and pH

<table>
<thead>
<tr>
<th>Properties</th>
<th>Calibration</th>
<th></th>
<th></th>
<th>Validation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSEP</td>
<td>RPD</td>
<td>R²</td>
<td>RMSEP</td>
<td>RPD</td>
<td>R²</td>
</tr>
<tr>
<td>SOM</td>
<td>5.84</td>
<td>1.78</td>
<td>0.69</td>
<td>5.82</td>
<td>1.76</td>
<td>0.71</td>
</tr>
<tr>
<td>TN</td>
<td>0.37</td>
<td>1.81</td>
<td>0.69</td>
<td>0.37</td>
<td>1.85</td>
<td>0.70</td>
</tr>
<tr>
<td>N_av</td>
<td>26.11</td>
<td>1.96</td>
<td>0.74</td>
<td>17.02</td>
<td>3.26</td>
<td>0.90</td>
</tr>
<tr>
<td>PH</td>
<td>0.50</td>
<td>1.67</td>
<td>0.58</td>
<td>0.77</td>
<td>1.31</td>
<td>0.49</td>
</tr>
</tbody>
</table>

RMSEP: Root mean square error of prediction

The cross-validation for both soil TN and soil N_av are classified as good with RPD=1.81 and 1.96, respectively. In particular, the independent validation for N_av was excellent; it was also good for TN. However, excess N fertilizer is a waste and it contamination of ground water. Thus, soil TN or N_av monitoring have attracted much research on vis–NIR spectroscopy [31-33]. Other workers have found similar satisfactory results of calibration for SOM, but only if soil samples are from fairly small area. Exceptionally, in this study, the calibration accuracy for N_av was excellent, and that contrasts with most other studies reported in the literature.

Not surprisingly, the model for pH gave only fair accuracy in the cross-validation and poor accuracy in independent validation. Vis–NIR spectroscopy for pH is not as robust as it is for those properties with direct spectral response in soil NIR spectrum, such as soil SOM and TN. Stenberg et al. concluded that occasionally successful reports for the measurement of soil pH without direct spectral response in the NIR range are because of co-variation through other properties that have direct spectral responses in the NIR, e.g. carbon, nitrogen and clay [34]. Therefore, we need further research to understand and probably improve the calibration accuracy of soil properties without direct spectral responses in the NIR range.

Nevertheless, even with an accuracy that distinguishes only high and low values, the technique might be of practical value.

![Fig. 1. Scatter plots for model cross-validation (hollow dots) and independent validation (solid dots) for SOM, TN, N_av and pH (the oblique line is 1:1).](image)

To understand the variation of measurement accuracy among the several soil properties further, we have plotted the regression coefficient curves for SOM, TN, N_av and pH (Fig. 2). The distribution of regression coefficients over the entire wavelength range (403–2448 nm) shows few significant wavelength bands for almost all soil properties considered.

These bands are in visible and NIR regions and might be attributed to energy absorption because of colour, water, organic constituents and clay minerals in the soil. Two major peaks in visible range at about
490 nm and 640 nm are related with the blue region around 450 nm and the red region around 680 nm [35]. In the NIR region between 750 nm and 1800 nm, two relative distinguished peaks can be seen for all properties. These two peaks are associated with the water absorption bands in the third overtone region (960 nm) and in the second overtone region (1450 nm). However, these peaks are more evident with organic C, TN and N_av, than with pH. It is clear that the most active spectral region is between 1800 nm and 2450 nm, where more distinguished bands can be seen. These bands are associated with the water absorption band in the first overtone region (1950 nm), clay minerals around 2300 nm and C–H + C–H, C–H + C–C and N–H combinations. Those bonds are clear and significant for SOM, TN and organic C, and they show why SOM, TN and N_av can be predicted more accurately than pH, the regression coefficients of which have fewer peaks and smoother curves [14].

![Fig. 2. Regression coefficient for SOM, TN, N_av and pH.](image)

4. Conclusion

Surveys of soil quality are important for the increasingly intensive rice farming in south China. In this feasibility study, we have built calibration models for large region for several important soil properties to enable accurate prediction. The cross-validation and independent validation lead to the following conclusions.

Regional calibration models for TN and N_av seem to be good both in cross validation and independent validation. The models can be used for prediction sample data collected from those areas to evaluate the total and available nitrogen in the soil.

The regional calibration model for organic matter was not so good, but it is acceptable for monitoring in those areas. The model for soil pH obtained was the least accurate because of the weak spectral response in the NIR spectrum, or perhaps because we had relatively few data.

The most significant advantage of vis–NIR calibration models is the cost. The development of model might be expensive as it requires many data. However, once a robust and accurate model has been built, the initial cost is one-off, and the model can be used for soil survey in the region for many years with little additional cost.

This is a huge advantage for the vis–NIR soil spectroscopy over traditional soil survey methods. This feasibility showed encouraging result for the vis–NIR soil sensing to be used in the survey of soil quality in rice paddy field in the south of China. The research is efficiently encouraging for us to explore expanding of the technique for other soil constituents (such as phosphorous).

5. Acknowledgements

This work was supported by the National Natural Science Foundation of China (41101197, 41271234), Ministry of Education, Humanities and social science project (No. 10YJC910002) and the Natural Science Foundation of Jiangxi Province (20114BAB213017).

**References**


