

STATISTICAL UNDERSTANDING ON THE PRE-PROCESSING OF VNIR SPECTRA DATA FROM SOIL SAMPLES WITH DIFFERENT PREPARATIONS

Yiyun CHEN^a, Yaolin LIU^{a,*}, Teng FEI^a, Qinghu JIANG^a, Junjie WANG^a, Tiezhu SHI^a

^a School of Resource and Environmental Science, Wuhan University, No. 129, Luoyu Road, Wuhan 430079, P. R. China - (chenyy, yaolin610, feiteng, jiang8687, wjjlight, tiezhushi)@whu.edu.cn

ABSTRACT: Statistical analyses of the relationship between visible and near-infrared (VNIR) spectra and soil parameters usually require the normal distribution of wavelength variables. Most previous studies examined only the distribution of soil parameters, while the distribution of each VNIR spectra wavelength was ignored. Moreover, how the sample preparation process and spectral pre-processing procedure influence the distribution of each spectral wavelength has seldom been reported with data collected at a regional scale. With 71 soil samples collected from the Le'an River floodplain, China, our study analyzed the distribution of spectral wavelength variables in the VNIR region (350–2500 nm). More specifically, by means of the Lilliefors normality test plot and the kurtosis and skewness curve plot proposed by this study, we aimed to (1) test the normality of each wavelength using the Lilliefors normality test; (2) compare the kurtosis and skewness of wavelength variables between soil samples, with and without preparations. Results show that most wavelength variables from differently transformed spectra are not normally distributed. Soil sample preparation processes such as air-drying, grinding, and 0.2 mm sieving, can significantly change the distribution of wavelength variables. Thus, statistical methods with normality assumption in the wavelength variables would not be appropriate for data of this kind.

KEY WORDS: Soil, Visible and Near-Infrared Spectroscopy, Statistical analysis, Normality test

1. INTRODUCTION

Visible and near-infrared (VNIR) spectra are now being used by more researchers than ever before for characterizing a variety of soil parameters (e.g., organic matter, moisture, and heavy metal content) (Gomez et al., 2008; Viscarra Rossel and Behrens, 2010; Viscarra Rossel et al., 2009). In the study of VNIR-based soil constituents estimation, the VNIR spectral wavelengths (e.g., 350–2500 nm) act as predictor variables, and the soil parameters are predicted by a statistical model (e.g., a PLS model). Two aspects can largely influence the VNIR-based determination of soil parameters. One is the spectra transformation/pre-processing technique, while the other is soil sample preparation (Stenberg, 2010).

The spectra transformation/pre-processing technique has been considered to be an integral part of chemometrics modelling. Its application allowed the baseline shift, light scattering and non-linearities to be largely eliminated (Rinnan et al., 2009; Wold et al., 2001a; Wold et al., 2001b). Some of the most commonly used transformation/pre-processing techniques include absorbance ($\log [1/\text{Reflectance}]$), Savitzky-Golay smoothing, multiplicative scatter correction, standard normal variate (SNV), continuum removal (CR), and the first derivative. In practice, a variety of spectral transformation/pre-processing techniques and their combinations are used to give a better overall model (Mouazen et al., 2010; Viscarra Rossel and Behrens, 2010; Wu et al., 2009). To date, few studies, however, have investigated the data distribution of each wavelength before/after spectral transformation/pre-processing, despite the important role that this plays in the correlation analysis between spectra and soil parameters, as well as in statistical modelling (Bellon-Maurel et al., 2010). Moreover, the data distribution analysis of wavelength variables could be helpful for understanding how the spectral transformation/pre-processing technique influences the distribution of spectral wavelengths.

Soil sample preparation, such as grinding, sieving and air/oven-drying processes can largely change the reflectance of a soil sample. Some researchers have compared the spectral curves of soil samples with different surface roughness and moisture

conditions (Wu et al., 2009). Few of them, however, have investigated the distribution changes of each spectral wavelength variable before and after sample preparation. The investigation of such changes could be helpful for model developing and knowledge discovery, as the VNIR spectra are known to carry comprehensive information of both chemical and physical phenomena.

In addition, the recent trend of incorporating the VNIR spectra in the mapping of soil properties using methods from the kriging family also highlights the needs for normality examination of VNIR spectral wavelengths (Bilgili et al., 2011; Ge et al., 2007).

Despite the importance of data structure visualization in data mining (Daszykowski et al., 2003), traditional statistical analyses such as histogram of variables or QQ plots for normality analyses cannot simultaneously present the distribution of thousands of wavelength variables. To tackle such difficulties, two kinds of plots for the data distribution analysis of VNIR spectra are proposed: (1) a Lilliefors normality test plot of each wavelength variable, and (2) a kurtosis and skewness curve plot of each wavelength variable. As the distribution of wavelength variable is sample dependent, this study limited its scope to 71 soil samples collected from the Le'an River floodplain of China. More specifically, our study aimed to (1) test the normality of each wavelength variable using the Lilliefors normality test; (2) compare the kurtosis and skewness of wavelength variability between soil samples with and without preparations.

2. MATERIALS AND METHODS

2.1 Soil sample collection and preparation

A total of 71 top layer (0–15 cm) soil samples were collected from the Le'an River flood plain, China. The fieldwork lasted from 29 October 2009 to 1 November 2009. This data set includes 45 samples from agricultural land, 11 from forest land, 7 from pasture and 8 from river bench, ensuring a representative sample set of the Le'an River floodplain.

* Corresponding author. Yaolin Liu, Email: yaolin610@whu.edu.cn or yaolin610@163.com.

The soil samples were first air-dried in the laboratory at 20–25° for two days. The dried soil samples were then gently crushed in a porcelain mortar to break up large aggregates, and sieved using a 0.2 mm stainless steel sieve. It was assumed that the differences between samples, with and without pretreatment, only existed in soil particle size and water content.

2.2 VNIR spectra analyses

The reflectance of the soil samples was measured twice: once before sample preparation, generating a data set named Ori-01; and the other data set, Ori-02, was generated after the air-drying, grinding and 0.02 mm sieving process.

An ASD FieldSpec3 portable spectral radiometer with a wavelength of 350–2500 nm was used to measure the spectral reflectance of the soil samples. The sampling interval and spectral resolution were 1.4 nm and 3 nm for the 350–1000 nm range, and 2 nm and 10 nm for the 1000–2500 nm range (<http://www.asdi.com>). The spectra scanning procedure was carried out in a dark room at night, minimizing the influence of external light. A white light source matched with the spectroradiometer was used with a 45° incident angle. A soil sample of around 500 g, spanning a diameter of approximately 20 cm, was scanned by the spectroradiometer, with a distance of 12 cm from probe to sample surface and a zenith angle of 90°.

The spectral radiance over a standardized white Spectralon® panel was measured every ten samples. Then the spectral radiance over the soil was scanned (10 internal scans of 100 ms each). By dividing the radiance over the Spectralon® panel, the reflectance spectra of each soil sample was automatically derived and displayed on the screen of the laptop. Data were exported in ASCII format for further analyses.

2.3 Spectra pre-processing

This study applied seven spectra pre-processing/transformation techniques over both the Ori-01 and Ori-02 data sets. These were first and second derivative (Savitzky-Golay derivation with 11 points and a second-order polynomial), absorbance (log [1/reflectance]), first and second derivative of absorbance, standard normal variate (SNV), and continuum removal (CR). The CR technique has been widely used in VNIR spectroscopy for its ability to isolate particular absorption features in diffuse reflectance spectra (Gomez et al., 2008; Viscarra Rossel et al., 2009). Further details of these spectra transformations were reviewed by Rinnan (Reimann et al., 2001). The Unscrambler®X (<http://www.camo.com/>) was used to perform these spectra transformations, except for the CR process, which was calculated using ENVI Version 4.5 (www.itvvis.com). The algorithms of these pre-processing techniques were not listed for their wide applications and supports from a variety of software. Sixteen sets of spectra were finally available for distribution analysis, including the Ori-01 and Ori-02 data sets.

2.4 Distribution analyses

Methods for characterizing the distribution of a random variable can be divided into two categories: graphical methods, like histograms and the Q-Q plot; and numerical methods, such as skewness, kurtosis and the Lilliefors test of normality. Graphical methods usually compare the distribution of a random variable to a theoretical one by means of a bar plot (e.g. histogram) or a dot plot (e.g. Q-Q plot), both of which are

visually appealing. Numerical methods apply descriptive statistics (e.g. skewness) or statistical tests (e.g. the Lilliefors normality test) to the variables. These methods perform well when the number of variables for analyses is limited. However, they can either fail to present the results in a simple way or be inefficient in the case that variables numbering more than several thousand need to be presented. Therefore, one aim of this study is to find new solutions to overcome such difficulties. Two kinds of plots for the data distribution analysis of VNIR spectra are proposed. They are the kurtosis and skewness curve plot and the Lilliefors normality test plot of each wavelength.

2.4.1 Lilliefors normality test plot: The Lilliefors normality test is a kind of Kolmogorov-Smirnov test for normality when the mean and variance are unknown. There are two steps to draw a Lilliefors normality test plot. The first step is to test the normality of each wavelength variables. The second is to plot the results (y axis) over the wavelengths (x axis). This study utilizes the "lillietest" function in Matlab to test the normality of a wavelength variable, with 95% confidence interval. The results returned by the function can be either "zero" or "one". "Zero" means that the null hypothesis "the data are normally distributed", cannot be rejected at the 5% significance level, while "one" indicates that the null hypothesis can be rejected at the 5% level. A Matlab program was written to test the normality of each wavelength variable. The results (zero or one) were stored in a vector and afterwards plotted versus the wavelengths (350–2500 nm) using the "bar" function in Matlab® (R2008a).

2.4.2 Kurtosis and skewness curve plot: Kurtosis, based on the fourth central moment, is known for its ability to measure the thinness of tails or "peakedness" of a probability distribution. If kurtosis of a random variable is less than three, the distribution has thicker tails and a lower peak compared to a normal distribution. Kurtosis of larger than three indicates a higher peak and thinner tails.

$$\frac{E[(x-\mu)^4]}{\sigma^4} = \frac{\sum(x_i - \bar{x})^4}{s^4(n-1)} = \frac{(n-1)\sum(x_i - \bar{x})^4}{\left[\sum(x_i - \bar{x})^2\right]^2} \quad (1)$$

Skewness is a third standardized moment that measures the degree of symmetry of a probability distribution. If skewness is greater than zero, the distribution is skewed to the right, having more observations on the left. S is the third central moment of X, divided by the cube of its standard deviation.

$$\frac{E[(x-\mu)^3]}{\sigma^3} = \frac{\sum(x_i - \bar{x})^3}{s^3(n-1)} = \frac{\sqrt{n-1}\sum(x_i - \bar{x})^3}{\left[\sum(x_i - \bar{x})^2\right]^{\frac{3}{2}}} \quad (2)$$

A normally distributed random variable should have skewness near zero and kurtosis near three. Two steps are needed to draw a kurtosis and skewness curve plot. The first step is to calculate the kurtosis and skewness of each wavelength variable; this study used the kurtosis and skewness functions in Matlab® (R2008a). The second is to plot the calculated kurtosis and skewness (double y axes) over each wavelength (x axis). For this, the "line" function in Matlab® (R2008a) was used.

3. RESULTS AND DISCUSSIONS

3.1 Plots of the Lilliefors normality test

Figure 1 presents the Lilliefors normality test results of each wavelength variable from the 16 data sets mentioned in Section 2.3. Eight spectra transformation methods divided this figure into eight pairs. Each pair has two bar plots, the upper subplot represents the normality test result from the untreated soil sample, and the lower subplot is for the pretreated sample. In each bar plot, the value read from the y axis can be either zero or one, indicating either normality or non-normality of the responding wavelength variable from the x axis. The bar color is in blue. Thus, the areas in blue indicate the non-normality of corresponding wavelength variables, while the blank areas or the areas in white indicate that the corresponding wavelength variables take a normal distribution. By means of the Lilliefors normality test plot, one can simultaneously investigate the normality of thousands of wavelength variables from spectra with different transformations and from samples with different preparations.

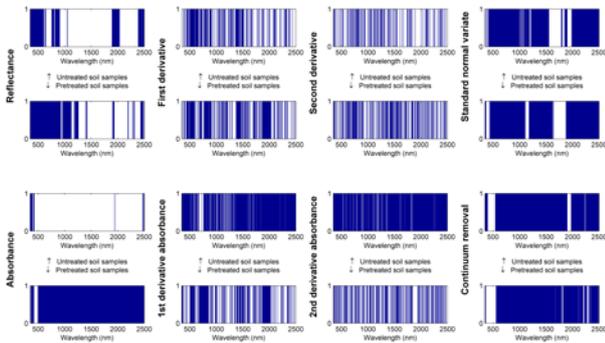


Figure 1. Lilliefors normality test plots (at a confidence interval of 0.05).

The normality test results of the case data show that wavelength variables from differently transformed spectra are not always normally distributed. The wavelength variables of SNV are normal in the region of 1700–1900 nm. They are non-normal in the regions 400–1000 nm, 1300–1600 nm and 2000–2500 nm. For the continuum removal spectra, normal wavelength variables exist mostly in the region of 400–550 nm. The normality test results of wavelength variables from the derivative-based transformations show a discrete pattern. It was noticed that for the pretreated sample, the wavelength variables in the region 1880–1950 nm are non-normal for the 1st derivative absorbance, whereas for the 2nd derivative absorbance, they are normal. After examining the bar plots one by one, we found that the region 1880–1950 nm is an important region where changes happen, either from non-normal to normal or normal to non-normal. As wavelengths around 1900 nm are known to be related to vibrations of hydroxyl (—OH) groups in water molecules, the normality test plot could help to select the spectra transformation methods when modelling the relationship between soil moisture and spectral wavelengths.

Soil sample preparation processes such as air-drying, grinding, and 0.2 mm sieving can drastically change the distribution of wavelength variables (Figure 1 and Figure 2). About 1/3 of the reflectance wavelength variables are non-normal whereas the number increases to 1/2 after the soil samples are pretreated, including the whole visible region. For the soil samples without preparation, the wavelength variables from the absorbance

spectra, in almost all the region of VNIR (350–2500 nm), are normal. For the soil samples with preparation, however, wavelength variables are normal only in a narrow region around 550 nm. The pattern for the wavelength variables from derivative-based transformations is like a barcode. The pattern changes, which result from soil sample preparations, are less clear. The signal noise magnified by derivative-based transformations might be a potential factor that adds to the pattern complexity.

The Lilliefors normality test plot is advantageous for its easy understanding and its ability to simultaneously present the normality test results of thousands of variables. This method, however, has some limitations, and the results depend on the confidence interval of the Lilliefors normality test. Different confidence intervals may result in different results. Thus, the bar plot could be different even for the same data set. Secondly, wavelength variables that have different distributions could have the same normality test result. In other words, the Lilliefors normality test plot does not provide any details about the distribution of the wavelength variables.

3.2 Plots of the kurtosis and skewness curves

For the wavelength variables of reflectance from the pretreated sample, the kurtosis curve reaches as high as 20 in the region of 350–500 nm, and decreases sharply in the region of 500–800 nm. In the region thereafter, the curve is even and approaches a value of 3. The skewness curve has a similar curve shape, but fluctuates a little around 1400 nm, 1900 nm and 2200 nm. In the region of 350–1400 nm, the wavelength variables are positively skewed; in the region of 1400–2500 nm, the wavelength variables skew to the left a little. For the wavelength variables of reflectance from the untreated sample, the kurtosis curve varies from around 12 at 350 nm to around 3 in the region of 600–2500 nm. The skewness curve decreases from 2.5 at 350 nm to approximately 0.8 in the region of 600–2500 nm. It then fluctuates in the opposite direction to that of the pretreated sample at around 1400 nm and 1900 nm. Although the orientation of the peaks is to the opposite, the absolute values of both skewness curves increase together at around 1400 nm and 1900 nm. Such phenomena might be related to the vibrations of hydroxyl (—OH) groups in water molecules. Valleys on the skewness curve for the pretreated sample are also found at 2200 nm and 2350 nm, whereas no peaks were found on the curve from the untreated sample. The differences of these two kurtosis curves in the region of 350–760 nm indicate that the soil sample preparation process increases the visual similarity of most of the 71 soil samples.

For the wavelength variables of absorbance from the pretreated sample, the kurtosis curve is higher in the visible region and has a peak at 500 nm. It decreases to 5 and becomes horizontal in the region of 750–2500 nm. The kurtosis curve from the untreated samples is quite even and parallel to the x axis at about 3 in almost the whole region of 350–2500 nm. The skewness curve for the pretreated sample is below 0 for most parts of the visible region and has two valleys at 400 nm and 500 nm. This curve rises to above 0 at 600 nm, approximately, and grows slowly and becomes even when it reaches 1 at around 1400–2500 nm. Two tiny peaks are visible at 1400 nm and 1900 nm. For the untreated sample, the skewness curve is about 0 at 350 nm, then increases slowly and intersects with that of the pretreated sample at 750 nm, where it reaches to about 0.5. The curve then vibrates around 0.5 in the region thereafter. Two tiny peaks at 1400 nm and 1900 nm are also visible.

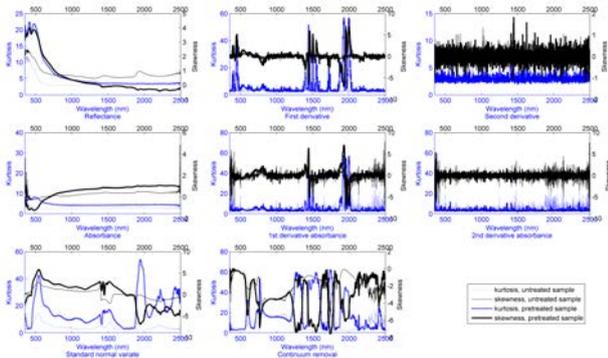


Figure 2. Kurtosis and skewness curve plots.

The patterns for both the kurtosis and skewness curves from the SNV-transformed spectra are not that flat in comparison with those of absorbance. For the pretreated sample, the kurtosis curve has the first valley at 400–450 nm, where the curve reaches as low as 3. It then upsurges to about 44 at 550 nm. The curve then decreases sharply all the way from 550 nm to 750 nm. In the region of 750–1200 nm, the curve fluctuates a little with a decreasing trend. In the region of 1200–1550 nm, a rising trend, accompanied with a spinous valley at 1400 nm, is visible; in the region of 1450–1550 nm, it fluctuates slightly and frequently. After the declivity in the region of 1550–1700 nm, the curve reaches to about 3. It then remains even in the region of 1700–1850 nm. The curve upsurges almost vertically from 3 at 1850 nm to about 55 at 1900 nm, and then drops sharply to about 13 at 2100 nm. It then increases to about 35 at 2250 nm, decreases to about 15 by 2350 nm, and then increases again to around 36 at 2450 nm. The kurtosis curve for the untreated sample generally has a similar curve shape as that of the pretreated sample, but fluctuates less severely with lower peaks, and has no peak at 1900 nm and 2450 nm. Large kurtosis usually indicates clustering of the data. From the two kurtosis curves in the "Standard normal variate" subplot, one can easily point out that at which wavelengths do these two curves differ from each other, and afterwards propose hypothesis or potential explanations.

The skewness curve in the "Standard normal variate" subplot for the pretreated sample is below 0 in the regions of 350–450 nm and 1850–2500 nm. It has a peak at 550 nm and three vibrations at 1400–1550 nm. In the region of 1850–2500 nm, it has three valleys and two peaks. The valley bottoms and peaks of the skewness curves correspond to the peaks and valley bottoms of the kurtosis curves, one by one. The skewness curve for the untreated sample is closer to 0, when compared with the curve for the pretreated sample. It also has a peak at 550 nm and vibrates at 1400–1550 nm, but has no evident peak at 1850 nm and is more even in the region of 1850–2500 nm.

The kurtosis and skewness curves for the derivative-based and continuum removal transformations fluctuate so severely that it is difficult to relate each peak and valley with certain spectral features of a soil parameter, or even to distinguish the curve between the pretreated and the untreated samples. For the first derivative based transformations, vibrations of the skewness curves are more evident at 1450 nm and 1850 nm. For the kurtosis curves, some values are extremely high at 1450 nm and 1850 nm. The skewness curves for the CR transformations are below zero. The complexity of the curve patterns might be due to the noise effects magnified by the spectra transformations, as well as useful spectral features.

A possible perspective for some of the peaks and valleys in all the eight subplots (Figure 2) could be that the non-normality of spectra-related soil parameters results in the non-normality of specific wavelength variables. Spectra transformations and soil sample preparation might amplify the differences of skewness and/or kurtosis between wavelength variables related and not related to soil parameters. Measurement of those spectra-related soil parameters might help to reveal the truth.

For the 71 soil samples considered, eight subplots in Figure 2 show how spectra transformations influence the kurtosis and skewness of each wavelength variable. Each subplot shows how the kurtosis and skewness curves fluctuate in the region of 350–2500 nm. The effects that sample preparations have on the wavelength variable can also be identified, and further represented with different line types. The left y axis in blue is the kurtosis axis; the y value on the blue curves should read from this y axis. A larger y value usually means the distribution of the corresponding wavelength variable has a higher peak. The right y axis in black is the skewness axis; the y value on the black curves should read from this y axis. A positive y value from the black curve indicates that the distribution of the corresponding wavelength variable is skewed to the right. The larger the y value, the more observations are there on the left while examining a histogram of specific wavelength. The advantage of plotting together the kurtosis and skewness of each wavelength variable is that we can easily see the trend of kurtosis and skewness curves at the same time over the wavelength axis. This can help to answer the questions of in which wavelength region do the kurtosis and skewness move at the same direction and in which at the opposite. Answering such questions could provide a statistical insight into the effects that soil sample preparation process has on the spectra.

3.3 Histograms for wavelength variables of specific interest

The Lilliefors normality test plot and the kurtosis and skewness curves plot can conveniently sketch the distributional characteristics of thousands of wavelength variables, simultaneously. The examples given above demonstrate the use of these plots for statistical understanding of the effects of the spectra transformations and sample preparations. Interesting patterns could also be found in Figure 1 and Figure 2. For further exploration of specific wavelength variables, a histogram is suggested. Figure 3 demonstrates the use of histograms in the distribution analyses of wavelength variables at 1450 nm and 1900 nm. In case one is not familiar with the meaning of kurtosis and skewness, or not used to such terms as "positively skewed", you can first use the Lilliefors normality test plot and/or the kurtosis and skewness curves plot to identify the wavelength variables that are of interest, or that you have problems with. After that, the histograms of specific wavelength variable(s) can be drawn, as shown in Figure 3.

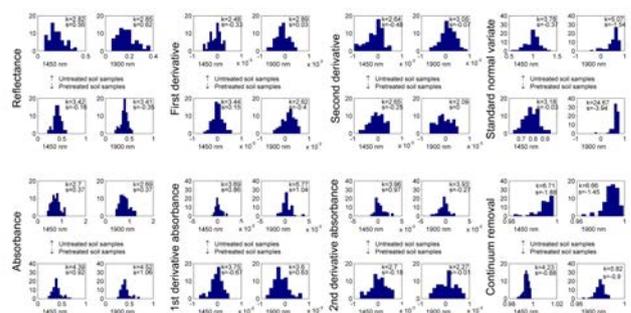


Figure 3. Histograms of wavelength variables at 1450 nm and 1900 nm from different spectra transformations and soil sample preparations (kurtosis denoted as “k”; skewness denoted as “s”)

Wavelength variables at 1450 nm and 1900 nm are selected as examples for their physical relations with the vibrations of hydroxyl (—OH) groups in water molecules (Paul J, 1997), and they are the wavelengths where peaks and valleys occurs in Figure 2. It can be seen from Figure 3 that the distribution of wavelength variables at both 1450 nm and 1900 nm are mostly different, and do not always take on a bell shape. The distribution can be dissymmetric, and sometimes has a steep peak. Figure 3 also shows that soil sample preparations could dramatically modify the histograms of wavelength variables at 1450 nm and 1900 nm. For examples, the wavelength variables at 1450 nm and 1900 nm in "Reflectance" and "Absorbance" subplots change from platykurtosis to leptokurtosis with soil sample preparations, showing that values are more clustered. Potential “outliers” could also be identified from the “Standard normal variate” and “Continuum removal” subplots.

It can be drawn from the above discussion that soil sample preparations can drastically change the distribution of wavelength variables at both 1450 nm and 1900 nm, both of which are absorption bands from hydrogen bonding. Therefore, when applying statistical methods with normality assumption in the wavelength variables, the distribution analyses should be a prior step.

4. CONCLUSIONS

Spectra transformations are necessary when estimating soil sample parameters from VNIR spectra. Our knowledge about the effects that spectra transformations have on the distribution of wavelength variables is limited. The recent trend of extending the scope of VNIR spectra to field use also requires a better understanding of how soil sample preparation can influence the distribution of each wavelength variable. Traditional distribution analyses are deficient for analyzing thousands of variables at a time. This study proposes two novel methods for simultaneously analyzing the normality and distributional characteristics of thousands of wavelength variables: (1) the Lilliefors normality test plot; (2) the kurtosis and skewness curve plot. The histograms of wavelength variables at 1450 nm and 1900 nm are also drawn for a comparative study between the classical distributional analysis and the proposed methods.

This study demonstrates the application of these two methods with 71 soil samples collected from the the Le'an River floodplain of China. Each bar plot in the Lilliefors normality test plot exhibits different patterns, indicating that the normality assumptions in the wavelength variables are selectively advocated. Interesting patterns could also be found in the kurtosis and skewness curves plot. A possible explanation for the vibrations on the skewness and kurtosis curves is proposed: the non-normality of spectra-related soil parameters results in the non-normality of specific wavelength variables. The differences of skewness and/or kurtosis between wavelength variables can be amplified by the spectra transformations and soil sample preparation. Whether the proposed explanation can be a statistical mechanism that makes the spectra transformations and soil sample preparation effective in the VNIR-based estimation of soil parameters deserves further study. The inspiration for the distributional analyses in the wavelengths variables could be a contribution of this paper.

Moreover, the kurtosis and skewness curve plots coupled with the histogram of specific wavelength variable could be a potential graphical tool for “outliers” searching and identification.

ACKNOWLEDGEMENTS

This research was supported by the National Key Technologies R & D Program of China during the 12th Five-Year Plan Period (2011BAB01B06). The authors would like to thank Dr. Zhuo Luo and Dr. Wu Jian for their participation in the field work, and Prof. Guofeng Wu for his valuable comments.

REFERENCES

- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry* 29, 1073-1081.
- Bilgili, A.V., Akbas, F., van Es, H.M., 2011. Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables spatially. *Precis Agric* 12, 395-420.
- Daszykowski, M., Walczak, B., Massart, D.L., 2003. Projection methods in chemistry. *Chemometrics and Intelligent Laboratory Systems* 65, 97-112.
- Ge, Y., Thomasson, J.A., Morgan, C.L., Searcy, S.W., 2007. VNIR diffuse reflectance spectroscopy for agricultural soil property determination based on regression-kriging. *T Asabe* 50, 1081-1092.
- Gomez, C., Lagacherie, P., Coulouma, G., 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma* 148, 141-148.
- Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* 158, 23-31.
- Paul J, G., 1997. Rugged spectroscopic calibration for process control. *Chemometrics and Intelligent Laboratory Systems* 39, 29-40.
- Reimann, C., Niskavaara, H., Kashulina, G., Filzmoser, P., Boyd, R., Volden, T., Tomilina, O., Bogatyrev, I., 2001. Critical remarks on the use of terrestrial moss (*Hylocomium splendens* and *Pleurozium schreberi*) for monitoring of airborne pollution. *Environmental Pollution* 113, 41-57.
- Rinnan, A., van den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trac-Trend Anal Chem* 28, 1201-1222.
- Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma* 158, 15-22.
- Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* In Press, Corrected Proof.
- Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma* 150, 253-266.
- Wold, S., Sjöström, M., Eriksson, L., 2001a. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58, 109-130.
- Wold, S., Trygg, J., Berglund, A., Antti, H., 2001b. Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems* 58, 131-150.

Wu, C.-Y., Jacobson, A.R., Laba, M., Baveye, P.C., 2009.
Accounting for surface roughness effects in the near-infrared
reflectance sensing of soils. *Geoderma* 152, 171-180.